Quantitative Analysis of Extreme Risks and Extremal Dependence in Insurance and Finance

by

Hengxin Cui

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Liang Peng
Professor
Dept. of Risk Management and Insurance, Georgia State University

Supervisor(s): Ken Seng Tan
Professor

Fan Yang
Assistant Professor

Internal Member: Jun Cai
Professor

Alexander Schied
Professor

Internal-External Member: Xuyang Ma
Assistant Professor, School of Accounting and Finance

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In this thesis, we aim at a quantitative understanding of extreme risks and extremal dependence in insurance and finance. We use regularly varying distribution functions in extreme value theory (EVT) to model extreme risks, and apply various tools in multivariate extreme value theory (MEVT) to capture extremal dependence. We focus on developing asymptotics for certain risk measures.

We start with a portfolio diversification problem. In finance, investors usually construct a mixed portfolio in order to diversify away the individual risks. However, this is not always the case when heavy-tailedness and tail dependence of large losses are considered. Chapter 3 applies the multivariate regular variation (MRV) model to study this problem in an asymptotic sense and provides an applicable portfolio optimization strategy. A practical performance test for our strategy is also provided in this Chapter.

The mainstream of the literature on the limitation of portfolio diversification follows the assumption that risks have unbounded distribution support, i.e., no cap for potential loss. However, real-world firms usually have limited liability. Then a natural question arises whether the non-diversification effect strictly depends on the tail behaviour of the loss distribution. For risks with bounded support, will similar non-diversification results still exist? We answer this question in Chapter 4 and we argue that diversification is still possible to be inferior as long as the risks are truncated at sufficiently large threshold level.

In Chapter 5, we consider the risk of a large credit portfolio of multiple obligors subject to possible default. Contrary to the Gaussian and t copulas that are widely used in practice, we assume a portfolio dependence structure of Archimedean copula type. Under this setting, we derive sharp asymptotics for portfolio credit risk that highlight the impact of extremal dependence among obligors. By utilizing these asymptotic results, we propose two different algorithms that are shown to be asymptotically optimal and can be applied to efficiently estimate portfolio credit risk via Monte Carlo simulation. In order to capture hierarchical dependence structure among the obligors in a large credit portfolio, we also extend our asymptotic analysis to the structure of nested Gumbel copulas and an efficient algorithm of bounded relative error is also developed for this more complex structure. Numerical results are provided at the end of the chapter to illustrate the performance of our algorithms, as well as their respective merits.

Acknowledgements

First, I would like to express my sincere gratitude to my PhD supervisors Professor Ken Seng Tan and Professor Fan Yang. Considering that I am not hard working, in fact, they have spent more patience than regular during the last five years. Without their endless encouragement and inspiring guidance, I cannot imagine how long it will take to finish my PhD degree.

I would also like to thank my thesis committee members: Professor Liang Peng, Professor Jun Cai, Professor Alexander Schied and Professor Xuyang ma, for serving on my thesis committee and providing valuable comments on the thesis. In particular, I thank Professor Liang Peng for flying to Waterloo to attend my oral defence, although he has a very tight schedule. My sincere thanks also go to Professor David Landriault who is my master supervisor and Professor Changbao Wu who brings me to Waterloo.

University of Waterloo offers a great atmosphere for PhD student. I would thank to Ms Mary Lou Dufton and all administrative staff for their patience in helping me and my fellow colleagues in 4th floor for your kindness. I would also like to acknowledge the financial support from the Department of Statistics and Actuarial Science, and the Hickman Scholar program of the Society of Actuaries.

I am also indebted to my parents. Starting from junior high school, I have already been not on their side for almost 14 years. They never asked anything back from me, instead, always understand and support me. In my opinion, this is already more than any other help.

The last paragraph is devoted to my dearest Xinzhe. Thank you for your love and encouragement, and for all late nights. I am actually quiet and less disturbing when I was concentrating on my PhD work, especially when I worked on mathematical proofs or I was under crazy deadline pressure. So thank you for your extra patience on my silence and absences. Without you being aside, I believe I will definitely lose the entire colourful life. Your endless support is the fundamental reason that keeps me sane over the last few months.

Dedication

This is dedicated to my family.

Table of Contents

List of Tables

List of Figures

Abbreviations

2RV second-order regularly varying

DR diversification ratio

ERI extreme risk index

EVT Extreme Value Theory

FGM Farlie–Gumbel–Morgenstern

i.i.d. independent and identically distributed

L-S Laplace-Stieltjes

LGD loss given default

MDP most diversified portfolio

MRV multivariate regular variation

RV regular variation

VaR Value-at-Risk

Chapter 1

Introduction

1.1 Objectives of the Thesis

Loosely speaking, extreme risk is the risk of rare events that typically could lead to a disastrous financial and social consequence. From the 1987 stock market crash, to the 2008 financial crisis, to the 2011 Fukushima Daiichi nuclear disaster, it is evident that rare events substantially affect insurance and financial institutions. Responses to rare events are stressed in recent regulatory reform. For example, the minimal capital ratio of the Basel III has been doubled and banks are directed to hold excess capital as conservation, see BCBS (2011); a pillar in Solvency II requires insurers to hold risk capital to remain solvency at a confidence level of 99.5%.

Considering the severe consequences of rare events and motivated by the regulatory frameworks, we conduct asymptotic analysis of extreme risks in this thesis. The reason to perform asymptotic analysis is that the asymptotic expressions of risk measures usually are easier to compute compared with the expressions at finite levels. More importantly, the asymptotic expressions can provide us a better understanding of extreme risks, and of ways to managing and controlling them. In such analysis, it is important to realize that the aforementioned rare events often involve highly associated underlying risk across space and time, which leaves a strong possibility of large losses that could happen simultaneously. If one recall the 2008 financial crisis, company defaults occur in clusters; that is, if one company defaults, this may have a chain reaction of triggering other companies to a financial distress, or even to default. For this reason, the intricate dependence structure introduced by such events can be viewed as another source of risk, and it is of paramount importance to "correctly" model the extremal dependence among large individual losses.

In this thesis, we use heavy-tailed distribution functions to model potentially large losses; while to model dependence, we use tools such as copulas and multivariate regular variation (MRV) structures. An important difference between copulas and MRV structures in describing dependence is that copulas provide a complete description while MRV structures focus on the tail part only. Quantities of interest to us include:

- the amount of capital a financial institution or an insurance company has to hold as required by its regulator,
- characterization of diversification effects for extreme risks and the resulting portfolio optimization problem,
- the diversification effects for extreme risks with limited liability,
- the tail behaviour of the loss from defaults of a large portfolio.

1.2 Structure of the Thesis

In Chapter 2, we prepare some mathematical concepts and tools regarding heavy-tailed distributions and extremal dependence that will be widely used in the rest of the thesis.

In Chapter 3, we study a portfolio optimization problem. More specifically, we investigate the optimal portfolio construction aiming at extracting the most diversification benefit. We employ the diversification ratio based on the Value-at-Risk (VaR) as the measure of the diversification benefit. With modeling the dependence of risk factors by the multivariate regularly variation model, the most diversified portfolio is obtained by optimizing the asymptotic diversification ratio. Theoretically, we show that the asymptotic solution is a good approximation to the finite level solution. Our theoretical results are supported by extensive numerical examples. By applying our portfolio optimization strategy to real market data, we show that our strategy provides a fast algorithm for handling a large portfolio, while outperforming other peer strategies in out-of-sample risk analyses.

In Chapter 4, we revisit the limits of diversification for truncated risks. It is a known fact that diversification is not necessary a preferred risk mitigation strategy for extremely heavy-tailed (infinite first moment), independent and unbounded risks. This finding has important implications in the management of extreme risks, especially in catastrophe insurance market. However, in many real world applications, the extremely heavy-tailed risks are not just independent and unbounded; they can be dependent and often truncated. In this chapter, we provide a comprehensive study on how the truncation affects the diversification for extremely heavy-tailed risks with different dependence structures. For both real-valued and nonnegative risks, we derive the bounds of the truncation such that the diversification is suboptimal or optimal. We find that the diversification effect is much easier to become suboptimal for nonnegative risks than that for real-valued risks. For nonnegative risks, when the truncation level is sufficiently high, the diversification effect is not affected by the dependence structure or the heavy-tailedness of the marginals. Simulation studies are provided to highlight the key findings of our results.

In Chapter 5, we study the asymptotic behaviour of the loss from defaults of a large credit portfolio. Contrary to the widely used Gaussian copula, we assume the portfolio dependence structure of Archimedean copula family, in which latent variables governing individual defaults follow a mixture structure incorporating extremal dependence and asymmetry. Under the assumption that the mixing variable or the so-called systematic risk factor has a regularly varying tail, we derive sharp asymptotics for the tail probability of portfolio losses and the expected shortfall. The asymptotic results further help us design two different numerical algorithms that can efficiently estimate portfolio risk via simulation. In order to capture potential hierarchical structure in a large portfolio, we also extend our work to a special case of nested Archimedean copulas, namely partial nested Gumbel copulas. An efficient algorithm for this particular type is provided as well. At the end of this chapter, an extensive simulation study is conducted to justify the accuracy and variance reduction performance of our proposed algorithms.

Several potential topics for further research are presented in Chapter 6.

1.3 Notation and Conventions

In this section, we provide a list of notation and conventions to be used throughout the thesis. Some of the definitions may be repeatedly mentioned in the main text for reference.

A summary of notation used in this thesis is given in the following table

Table 1.1: Notations

$\mathbf{0}$	a vector with all components being 0
1_E	the indicator function of an event E
$\lvert I \rvert$	cardinality of set I
$\xrightarrow{\text{v}}$, $\xrightarrow{\text{w}}$	vague convergence, weak convergence
$\lfloor x \rfloor$	the greatest integer less than or equal to the real number x

Table 1.1: (continued): Notations

$\lceil x \rceil$	the smallest integer greater than or equal to the real number x
$\mathcal{B}$	Borel σ-field
$\mathbb{E}$	expectation
$f^{\leftarrow}(y)$	$\inf\{x \in (-\infty, \infty) : f(x) \geq y\}$ for a non-decreasing function f
$f(x) \sim g(x)$	$\lim_{x \to x_0} f(x)/g(x) = 1$, where x_0 is clear from the context
$f(x) = o(g(x))$	$\lim_{x \to x_0} f(x)/g(x) = 0$, where f and g are positive functions, and x_0 is clear from the context
$f(x) = O(g(x))$	$\limsup_{x \to x_0} f(x)/g(x) < \infty$, where f and g are positive functions, and x_0 is clear from the context
$f(x) \lesssim g(x)$	$\limsup_{x \to x_0} f(x)/g(x) \leq 1$, where f and g are positive functions, and x_0 is clear from the context
$f(x) \gtrsim g(x)$	$\liminf_{x \to x_0} f(x)/g(x) \geq 1$, where f and g are positive functions, and x_0 is clear from the context
$\bar{F}$	the tail function, $1 - F$, where F is a distribution function
γ	Euler's constant
$\mathbb{K}$	$[-\infty, \infty]^d \setminus \{\mathbf{0}\}$
λ_L, λ_U	lower tail dependence, upper tail dependence
$\mathcal{L}_V(s)$	Laplace-Stieltjes transform of random variable V
$\mathrm{MRV}_{-\alpha}(\Psi)$	the class of multivariate regularly varying tailed distribution functions with tail index α and spectral measure Ψ
$\mathbb{P}$	probability measure
ϕ	generator of an Archimedean copula
$\mathbb{R}$	$(-\infty, \infty)$
$\mathbb{R}_+$	$[0, \infty)$
$\mathbb{R}^d$	d-dimensional real vector space
$\mathbb{R}_+^d$	nonnegative d-dimensional real vector space
$\mathrm{RV}_{-\alpha}$	the class of regularly varying tailed distribution functions with tail index α
$\mathcal{S}^{d-1}$	$\left\{\mathbf{s} \in \mathbb{R}^d : \|\mathbf{s}\| = 1\right\}$
$\mathcal{S}_1^{d-1}$	$\left\{\mathbf{s} \in \mathbb{R}^d : \|\mathbf{s}\|_1 = 1\right\}$
$\mathcal{S}_+^{d-1}$	$\left\{\mathbf{s} \in \mathbb{R}_+^d : \|\mathbf{s}\| = 1\right\}$
tA	$\{tx : x \in A\}$ for a set A and a real number t
t_v	Student-t distribution with degree of freedom v
var	variance
$x \vee y$	$\max\{x, y\}$
$x \wedge y$	$\min\{x, y\}$
x_i	the i-th element of vector $\mathbf{x}$

Table 1.1: (continued): Notations

$X_{(k)}$	the k-th smallest order statistics

Some conventions that we shall follow are listed below:

- All limits are taken to ∞ unless otherwise stated;
- All regularly varying functions are studied at ∞ unless otherwise stated;
- All convergences are weak convergence unless otherwise stated;
- The right tail of a distribution function is of our interest unless otherwise stated;
- Let $\mathbf{x} = (x_1, \ldots, x_n)^\top$ and $\mathbf{y} = (y_1, \ldots, y_n)^\top$, then relation $\mathbf{x} \leq \mathbf{y}$ means that $x_i \leq y_i$ for all $i = 1, \ldots, n$.

Chapter 2

Preliminary

In this chapter, we prepare some mathematical concepts and tools necessary for the rest of the thesis.

2.1 Univariate Regular Variation

2.1.1 Definition

Note that there is still some discrepancy over the use of the term heavy-tailed in the literature. In this thesis, we follow Embrechts et al. (2013).

A distribution function F on $\mathbb{R}$ is said to be heavy-tailed, if it holds for every $\varepsilon > 0$ that

$$\int_0^\infty e^{\varepsilon x} \mathrm{d}F(x) = \infty.$$

Otherwise, the distribution F is referred to as light-tailed. Clearly, for any light-tailed distribution on $\mathbb{R}_+$, all moments are finite.

Table 2.1 below provides a list of common light-tailed and heavy-tailed distributions. In the given table, c is the scale parameter and α is the shape parameter for corresponding distributions. Apparently, the tail of a heavy-tailed distribution function F is not exponentially bounded, therefore it assigns a relatively large probability to the right tail.

According to the application area (finance or actuarial science), the loss variable X is of our interest, i.e., $X > 0$ quantifies losses and $X < 0$ quantifies gains. Thus, heavy-tailed

Light-tailed	Heavy-tailed
$\text{Exp}(\lambda)$	$\text{Pareto}(c, \alpha)$
$\text{N}(\mu, \sigma^2)$	t_v
$\text{Gamma}(\alpha, \beta)$	α-Stable ($\alpha < 2$)
$\text{Weibull}(c, \alpha)$, $\alpha \geq 1$	$\text{Weibull}(c, \alpha)$, $\alpha \in (0, 1)$

Table 2.1: Common light-tailed and heavy-tailed distributions

distribution functions can be considered as an effective tool to model loss variables that are likely to be extremely large.

The heavy-tailed property is strengthened by the assumption of (univariate) regular variation (RV). We first define regularly varying functions as follows. Let f denote a positive real-valued function on $\mathbb{R}_+$.

Definition 2.1.1 *The function f is said to be regularly varying at infinity with index $\alpha \in \mathbb{R}$, written as $f \in \text{RV}_\alpha(\infty)$, if $f(tx)/f(t) \to x^\alpha$ as $t \to \infty$, for all $x > 0$. If $f \in \text{RV}_0(\infty)$, then f is slowly varying at infinity.*

Intuitively, a function f is regularly varying at infinity if it behaves like a power law function near infinity. Interested readers are referred to Bingham et al. (1989) and Resnick (2013) for textbook treatments. The definition of regular variation at zero is a simple modification of the definition of regular variation at infinity.

Definition 2.1.2 *The function f is said to be regularly varying at zero with index $\alpha \in \mathbb{R}$, written as $f \in \text{RV}_\alpha(0)$, if $f(tx)/f(t) \to x^\alpha$ as $t \downarrow 0$, for all $x > 0$. If $f \in \text{RV}_0(0)$, then f is slowly varying at zero.*

One can easily check that the function $f \in \text{RV}_\alpha(0)$ if and only if the function $x \mapsto f(1/x)$ is in $\text{RV}_{-\alpha}(\infty)$. We also define the regular variation at one as follows.

Definition 2.1.3 *The function f is said to be regularly varying at one with index $\alpha \in \mathbb{R}$, written as $f \in \text{RV}_\alpha(1)$, if $f(1-tx)/f(1-t) \to x^\alpha$ as $t \downarrow 0$, for all $x > 0$. If $f \in \text{RV}_0(1)$, then f is slowly varying at one.*

Actually RV can be defined at any positive point x_0 other than 0, 1 or ∞. We often drop the argument x_0 as long as the meaning of regularity is clear in the context. Now, we are able to define regular variation of loss variable X with tail index $\alpha \in [0, \infty)$.

Definition 2.1.4 *A distribution function F_X corresponding to a loss variable X is said to be regularly varying with tail index α, if its tail is regularly varying at ∞ with index $-\alpha$, i.e., $\bar{F}_X \in \mathrm{RV}_{-\alpha}(\infty)$.*

The definition above immediately indicates that a smaller value of tail index α means a heavier tail of distribution function F. Consider two loss variables X and Y with unequal tail indices, it is well known that the contribution of lighter tail to the aggregated loss $X + Y$ is asymptotically negligible compared to that of heavier one. Consequently, the later study of portfolio diversification effect will be reduced to the non-trivial case by assuming identical tail index α for each component.

2.1.2 Properties of Regular Variation

In the following lemma, we list some important properties on regular variation.

Lemma 2.1.1 *Let $f \in \mathrm{RV}_\alpha(\infty)$ for some $0 < \alpha < \infty$. We have the following:*

1. *(Potter's bounds) For every $\varepsilon > 0$, there exists some $x_0 > 0$ such that for $x, y \geq x_0$,*

$$(1-\varepsilon)\left(\left(\frac{y}{x}\right)^{\alpha+\varepsilon} \wedge \left(\frac{y}{x}\right)^{\alpha-\varepsilon}\right) \leq \frac{f(y)}{f(x)} \leq (1+\varepsilon)\left(\left(\frac{y}{x}\right)^{\alpha+\varepsilon} \vee \left(\frac{y}{x}\right)^{\alpha-\varepsilon}\right).$$

2. *Let $f : \mathbb{R}_+ \to \mathbb{R}$ be a non-decreasing function with $f(\infty) = \infty$. Then $f \in \mathrm{RV}_\alpha(\infty)$ if and only if $f^{\leftarrow} \in \mathrm{RV}_{1/\alpha}(\infty)$.*

3. *(Karamata's theorem) Suppose there exists $x_0 > 0$ such that $f(x)$ is positive and locally bounded for $x \geq x_0$. If $\alpha \geq -1$ then*

$$\lim_{x\to\infty} \frac{xf(x)}{\int_{x_0}^{x} f(s)\mathrm{d}s} = \alpha + 1.$$

If $\alpha < -1$, or $\alpha = -1$ and $\int_0^\infty f(s)\mathrm{d}s < \infty$, then

$$\lim_{x\to\infty} \frac{xf(x)}{\int_{x}^{\infty} f(s)\mathrm{d}s} = -\alpha - 1.$$

Proof. 1. See Theorem 1.5.6 of Bingham et al. (1989).

2. See Proposition 0.8(V) of Resnick (2013).

3. See Theorem B.1.5 of de Haan and Ferreira (2006). ■

The applicability of regular variation can be further enhanced by Karamata's Tauberian theorem for Laplace-Stieltjes (L-S) transforms; see, e.g., Feller (1971), pp.442–446. Especially for a distribution function and its L-S transform, the relation between their asymptotic behaviours is given by Corollary 8.1.7 of Bingham et al. (1989).

Proposition 2.1.1 *Suppose F_X is a distribution function for a positive random variable X with Laplace-Stieltjes transform $\mathcal{L}_X$. For $0 \le \alpha < 1$ and $l \in \mathrm{RV}_0(\infty)$, the following are equivalent:*

(a) $1 - \mathcal{L}_X(s) \sim s^\alpha l(1/s), \qquad s \downarrow 0,$

(b) $\bar{F}_X(x) \sim \dfrac{x^{-\alpha} l(x)}{\Gamma(1-\alpha)}, \qquad x \to \infty.$

As we can see, not only the index α but even the slowly varying function l is preserved after taking L-S transform.

2.1.3 Second-order Condition

In this section, we introduce a refinement of the concept of regular variation, namely second-order regularly varying (2RV). This concept is especially useful for studying rates of convergence in extreme value theory. The definition below seems more complex with use of the second-order auxiliary function A, however, the meaning itself is still clear in that it measures the convergence rate of the first-order asymptotic.

Definition 2.1.5 *A positive measurable function f is said to be second-order regularly varying with first-order index $\alpha \in \mathbb{R}$ and second-order index $\rho \le 0$, denoted by $f \in 2\mathrm{RV}_{\alpha,\rho}$, if there exists an auxiliary function A, which does not change sign and $\lim_{x\to\infty} A(x) = 0$, such that, for all $y > 0$,*

$$\lim_{x\to\infty} \frac{\frac{f(xy)}{f(x)} - y^\alpha}{A(x)} = y^\alpha \frac{y^\rho - 1}{\rho}. \tag{2.1.1}$$

The function A, describing the rate of convergence in 2.1.1, is regularly varying with index ρ. Thus, ρ governs the speed of convergence, if $\rho < 0$, we have an algebraic speed of convergence; if $\rho = 0$, the convergence is logarithmic. More discussions about the second-order condition can be found in Resnick (2002), de Haan and Ferreira (2006) and Hua and Joe (2011).

2.2 Copulas

For modeling the dependence structures between heavy-tailed random variables, the concept of copula plays an important role and provides practitioners a promising tool to model the dependence structure independently of the marginal behaviors. See the monograph Nelsen (2007) for a complete reference on copula functions, and see Cherubini et al. (2004) and McNeil et al. (2015) for discussions on copula methods applied in finance and quantitative risk management.

A function $C : [0, 1]^n \to [0, 1]$ is called *copula* if C is a multivariate distribution function with uniform margins, i.e.,

$$C(u, 1, \dots, 1) = u, \qquad \forall u \in [0, 1].$$

It is useful because it can couple marginal distributions with the joint distribution. Sklar (1959) shows that for each joint distribution function with marginal distributions $F_1, \dots, F_n$, there exists an n-dimensional copula C such that

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)),$$

and the copula C is unique when the marginal distributions are continuous. By Sklar's theorem, it is also possible to derive the relationship between the survivor joint distribution and the survival copula distribution $\hat{C}(1 - u_1, \dots, 1 - u_n)$.

The survival joint function is given by

$$\bar{F}(x_1, \dots, x_n) = \mathbb{P}(X_1 > x_1, \dots, X_n > x_n),$$

and it follows that

$$\bar{F}(x_1, \dots, x_n) = \hat{C}(1 - F_1(x_1), \dots, 1 - F_n(x_n)).$$

Moreover, given $C(u_1, \dots, u_n)$ is a copula, the corresponding survival copula is defined as

$$\hat{C}(u_1, \dots, u_n) = 1 + \sum_{I \subset \{1,\dots,n\}} (-1)^{|I|} C_I(1 - u_i, i \in I),$$

where C_I is the I-margin of the copula C with $|I|$ the cardinality of the set I.

Commonly used copulas include elliptical copulas (Gaussian copulas, t copulas), Archimedean copulas, and Farlie–Gumbel–Morgenstern (FGM) copulas.

2.2.1 Archimedean Copulas

Exchangeable Case

Multivariate Archimedean copulas are widely used in insurance and financial risk analyses. Differently from the elliptical copulas, Archimedean copulas have a simple closed form and can be represented by a generator ϕ, rather than a multivariate distribution. We use the generator ϕ to define the class of Archimedean copulas as follows:

$$C(u_1, \ldots, u_n) = \phi^{-1}(\phi(u_1) + \ldots + \phi(u_n)). \tag{2.2.1}$$

The generator function $\phi : [0, 1] \to [0, \infty]$ is continuous, decreasing and convex such that $\phi(1) = 0$ and $\phi(0) = \infty$, and ϕ^{-1} is the inverse of ϕ. We further assume ϕ^{-1} is completely monotonic, i.e. $(-1)^i \left(\phi^{-1}\right)^{(i)} \geq 0$ for all $i \in \mathbb{N}$. These requirements ensure that C is a copula for all dimensions $n \geq 2$; see Kimberling (1974). By using Bernstein's theorem (see Feller (1971), pp. 439), ϕ^{-1} is completely monotonic if and only if ϕ^{-1} is a L-S transform of the distribution of some positive random variable V. We recall that the Laplace-Stieltjes transform of V is given by

$$\mathcal{L}_V(s) = \int_0^\infty e^{-sv} \mathrm{d}F_V(v) = \mathbb{E}\left[e^{-sV}\right].$$

Because of the importance of such copulas in our following analysis, we will call these copulas LT-Archimedean and make the following definition.

Definition 2.2.1 *An LT-Archimedean copula is a copula of the form (2.2.1), where ϕ^{-1} is the Laplace-Stieltjes transform of the distribution of some positive random variable V.*

For many popular Archimedean copulas, the random variable V has a known distribution. For example, V is Gamma distributed for Clayton copulas. While for Gumbel copulas, V is a one-sided Stable random variable. A detailed specification about the parameters can be found in Table 1 of Hofert (2008).

The following result providing a mixture representation for an LT-Archimedean copula is first proposed by Marshall and Olkin (1988) and later formally proved in McNeil et al. (2015).

Proposition 2.2.1 *Consider an LT-Archimedean copula C with generator ϕ. Let V be a positive random variable with Laplace-Sieltjes transform ϕ^{-1} and let $R_1, \ldots, R_n$ be a sequence of i.i.d. standard exponential random variables that are also independent of V. Then the random vector*

$$\mathbf{U} = \left(\phi^{-1}\left(\frac{R_1}{V}\right), \ldots, \phi^{-1}\left(\frac{R_n}{V}\right) \right) \tag{2.2.2}$$

is distributed according to copula C.

The above construct is especially useful in the field of credit risk. One can regard random variable V as a proxy for systematic risks. Conditioning on V, random variables $U_1, \ldots, U_n$ are independent with conditional distribution function $\mathbb{P}(U_i \leq u | V = v) = \exp(-v\phi(u))$ for $u \in [0, 1]$.

Non-exchangeable Case

In higher dimensions, the Archimedean copula functions we considered above suffer from the deficiency that they impose too much structure on the dependency. In particular, all uniform variables U_i are exchangeable. In applications such as portfolio credit risk discussed in Chapter 5, this means that we cannot have some groups of obligors with higher dependency, and others with less dependency. In order to capture the multi-level dependence structure of the underlying portfolio, many nested constructions are possible. One of them is to use a partially exchangeable construction with two levels of nesting. Let C_0 be an outer LT-Archimedean copula with generator ϕ_0 and C_j be inner LT-Archimedean copulas with generators ϕ_j. Then we can define a partially nested Archimedean copula as follows:

$$C(\mathbf{u}) = \phi_0^{-1}\left(\sum_{j=1}^{J} \phi_0 \circ \phi_j^{-1}\left(\sum_{l=1}^{n_j} \phi_j(u_{jl}) \right) \right), \tag{2.2.3}$$

where $\mathbf{u} = (u_{jl}), 1 \leq l \leq n_j$ and $1 \leq j \leq J$. One can think of the partially nested Archimedean copula of (2.2.3) as capturing the inner dependence structure by Archimedean copula $C_j, j = 1, \ldots, J$ and combining these inner copulas by an overall dependence structure, C_0. The corresponding tree structure is depicted in Figure 2.1.

In order for (2.2.3) to be a proper copula function, McNeil (2008) has shown that the composite function $\phi_0 \circ \phi_j^{-1}$ must have completely monotonic derivatives for any $1 \leq j \leq J$ (a condition that will later make sense in Chapter 5). Let

$$\phi_{0,j}^{-1}(\cdot; x) := \exp\left(-x\phi_0 \circ \phi_j^{-1}(\cdot)\right)$$

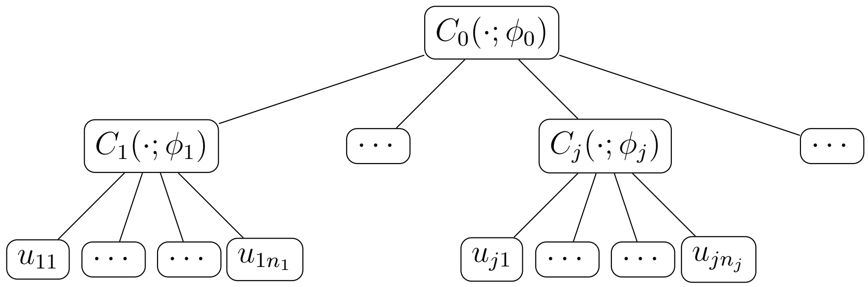


Figure 2.1: Tree structure of a partially nested Archimedean copula

for $1 \leq j \leq J$. The following proposition given by Hofert (2012) provides us a stochastic representation for the random vector $\mathbf{U}$ of a partially nested Archimedean copula defined by (2.2.3).

Proposition 2.2.2 *Let C be a nested Archimedean copula defined in (2.2.3). Further, let $\mathbf{U} \sim C$. Then, $\mathbf{U}$ admits the following stochastic representation,*

$$(U_{j1}, \dots U_{jn_j}) = \left(\phi_j^{-1}\left(\frac{R_{j1}}{V_j}\right), \dots \phi_j^{-1}\left(\frac{R_{jn_j}}{V_j}\right)\right), \qquad 1 \leq j \leq J, \tag{2.2.4}$$

where R_{jl}'s are i.i.d. standard exponential random variables for $1 \leq j \leq J$ and $1 \leq l \leq n_j$. The variables V_j are uniquely determined by its Laplace-Stieltjes transform $\psi_{0,j}^{-1}(\cdot; V_0)$ for $1 \leq j \leq J$ and V_0 follows the distribution function whose Laplace-Stieltjes transform is given by ϕ_0^{-1}.

Note that although nested Archimedean copulas involve multiple Archimedean generators and more complicated distributions for the variables V_j, there still exists some resemblance in the stochastic representation as compared with Archimedean copulas.

2.2.2 Tail Dependence

In this subsection, we focus on the dependence in the tail area, i.e., the tail dependence. The coefficients we describe as follows are defined in terms of limiting conditional probabilities of quantile exceedances.

Definition 2.2.2 *Let X_1 and X_2 be two random variables, distributed by F_1 and F_2, respectively. The coefficient of upper tail dependence of X_1 and X_2 is*

$$\lambda_U = \lim_{q \uparrow 1} \mathbb{P}\left(X_2 > F_2^{\leftarrow}(q) | X_1 > F_1^{\leftarrow}(q)\right),$$

provided a limit λ_U exists. Analogously, the coefficient of lower tail dependence is

$$\lambda_L = \lim_{q \downarrow 0} \mathbb{P}\left(X_2 \leq F_2^{\leftarrow}(q) | X_1 \leq F_1^{\leftarrow}(q)\right),$$

provided a limit λ_L exists.

The motivation for looking at these coefficients is that they provide measures of extremal dependence. Specifically, if $\lambda_U \in (0, 1]$, then X_1 and X_2 show upper tail dependence or extremal dependence in the upper tail.

When a multivariate distribution is characterized by an Archimedean copula, simple expressions for λ_U and λ_L are introduced in Charpentier and Segers (2009), where the coefficients of tail dependence are measures of pairwise dependence that depend only on the generator ϕ of the copula. Their result is summarized as follows.

Proposition 2.2.3 *Let C be a multivariate Archimedean copula with generator ϕ.*

1. *If $\phi \in \mathrm{RV}_\alpha(1)$ for some $\alpha \geq 1$, then the upper tail dependence is*

$$\lambda_U = \lim_{q \uparrow 1} \mathbb{P}(U_i > q | U_j > q) = 2 - 2^{1/\alpha}, \forall i \neq j,$$

2. *If $\phi \in \mathrm{RV}_{-\alpha}(0)$ for some $\alpha \geq 0$, then the lower tail dependence is*

$$\lambda_L = \lim_{q \downarrow 0} \mathbb{P}(U_i \leq q | U_j \leq q) = 2^{-1/\alpha}, \forall i \neq j.$$

It turns out that many Archimedean copulas used in practice have generators that are regularly varying at zero or one. Some well-known examples are provided below. A comprehensive list can be found in Table 4.1 of Nelsen (2007).

- Clayton copula: $\phi(t) = \frac{1}{\alpha}(t^{-\alpha} - 1), \alpha \in [-1, \infty] \backslash \{0\}$. Provided that $\alpha > 0$, ϕ^{-1} is completely monotonic and $\phi \in \mathrm{RV}_{-\alpha}(0)$.
- Gumbel copula: $\phi(t) = (-\ln(t))^\alpha, \alpha \in [1, \infty)$. ϕ^{-1} is completely monotonic and $\phi \in \mathrm{RV}_\alpha(1)$.

2.3 Multivariate Regular Variation

Besides copulas, the multivariate regular variation (MRV) structure introduced in this section, can also be used to model tail dependence. Compared with the copula approach, this approach provides a unifying framework for modeling both heavy-tailedness of marginal distributions and tail dependence among loss variables, and it actually strengthens the theory of (univariate) regular variation.

2.3.1 Vague Convergence of Radon Measures

Later we will see the concept of MRV is based on the interplay between regular variation and vague convergence of Radon measures. Hence, in this subsection, we first review the notions of vague convergence and radon measures. Consider a d-dimensional punctured space $\mathbb{K} = [-\infty, \infty]^d \setminus \{\mathbf{0}\}$ equipped with a Borel σ-field $\mathcal{B}$, a measure on such a space is called *Radon* if its value is finite for every compact subset of $\mathbb{K}$.

Remark 2.3.1 *Note in the punctured space $\mathbb{K}$, 0 is excluded and ∞ included. This is required since vague convergence is only able to define on relatively compact sets. By considering regular variation, sets of the form (x_i, ∞) are generally considered, yet are not bounded under the usual topology with ∞ excluded.*

Given a sequence of Radon measures $\{\mu_n, n = 1, 2, \ldots\}$ on $\mathbb{K}$, we say μ_n converges vaguely to μ, written as $\mu_n \xrightarrow{\text{v}} \mu$, if

$$\mu_n(f) := \int_{\mathbb{K}} f(x)\mu_n(\mathrm{d}x) \xrightarrow{\text{w}} \mu(f) := \int_{\mathbb{K}} f(x)\mu(\mathrm{d}x)$$

holds for every nonnegative continuous function f with compact support. By Portmanteau theorem, $\mu_n \xrightarrow{\text{v}} \mu$ on $\mathbb{K}$ if and only if the convergence

$$\mu_n(B) \xrightarrow{\text{w}} \mu(B)$$

holds for all relatively compact continuity set B.

From above equivalent definitions, the notion of vague convergence is closely related to that of weak convergence. For a full account of technical details related to the construction of punctured space $\mathbb{K}$ and the notion of vague convergence, the reader is referred to Resnick (2007).

2.3.2 Definition and Implication of MRV

For nonnegative random variables, multivariate regular variation can be defined based on MRV of the distribution function.

Definition 2.3.1 *A d-dimensional random vector $\mathbf{X}$ is* MRV *if there exists a normalizing function $b(t) \to \infty$ and a non-zero Radon measure ν on $\mathcal{B}(\mathbb{K})$, called the limit measure, such that $\nu\left([-\infty, \infty]^d \setminus \mathbb{R}^d\right) = 0$ and, as $t \to \infty$,*

$$t\mathbb{P}\left(\frac{\mathbf{X}}{b(t)} \in \cdot\right) \xrightarrow{\mathrm{v}} \nu(\cdot). \tag{2.3.1}$$

Additionally to (2.3.1), we assume that the limit measure ν is non-degenerate in the sense that

$$\nu\left(\left\{\mathbf{x} \in \mathbb{R}^d : x_i > \varepsilon\right\}\right) > 0, \qquad i = 1, \ldots, d,$$

for all $\varepsilon > 0$. This assumption ensures that all components of $\mathbf{X}$ are comparable in the upper tail. Therefore, by normalizing all components of $\mathbf{X}$ with the same function b, the phrasing of MRV in Definition 2.3.1 indicates the univariate regular variation of the marginals with the same tail index. One should also note that making different choices of function b does not change the limit measure ν except for a constant factor.

From Proposition 2.3 in Resnick (2007), relation (2.3.1) implies that the limit measure is homogeneous:

$$\nu(sA) = s^{-\alpha}\nu(A) \tag{2.3.2}$$

for some $\alpha \in (0, \infty)$ and all sets $A \in \mathcal{B}(\mathbb{K})$. Hence, we write $\mathbf{X} \in \mathrm{MRV}_{-\alpha}$.

The MRV structure can also be defined with the polar coordinate transformation. Given any arbitrary norm $\|\cdot\|$, in the restricted space $\mathbb{K}$, the polar coordinate transformation of a vector $\mathbf{x}$ is

$$T(\mathbf{x}) = \left(\|\mathbf{x}\|, \|\mathbf{x}\|^{-1}\mathbf{x}\right). \tag{2.3.3}$$

Note that

$$T : \mathbb{K} \mapsto (0, \infty) \times \mathcal{S}^{d-1},$$

where $\mathcal{S}^{d-1} = \left\{\mathbf{s} \in \mathbb{R}^d : \|\mathbf{s}\| = 1\right\}$. Thus, the scaling property in (2.3.2) leads to a decomposition of the induced measure $\nu^T := \nu \circ T^{-1}$, namely,

$$\nu^T = c \cdot \rho_\alpha \times \Psi \qquad \text{on} \quad (0, \infty) \times \mathcal{S}^{d-1} \tag{2.3.4}$$

with the constant factor

$$c = \nu\left(\left\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| > 1\right\}\right) > 0,$$

the measure ρ_α defined by

$$\rho_\alpha((x, \infty]) = x^{-\alpha}, \qquad x > 0,$$

and a probability measure Ψ on $\mathcal{S}^{d-1}$ with respect to $\|\cdot\|$; see Theorem 6.1 in Resnick (2007). The measure Ψ is often called the spectral or angular measure. Throughout the thesis, we denote that $\mathbf{X}$ is MRV with tail index α and spectral measure Ψ by $\mathbf{X} \in \mathrm{MRV}_{-\alpha}(\Psi)$, which implies the corresponding limit measure ν as in (2.3.4). The dimension d is suppressed in this notation as it is usually clear from the context.

Theoretically, it does not matter which norm is chosen in the polar representation (2.3.3). For simplicity, we consider the ℓ_1-norm $\|\cdot\|_1$ and let Ψ denote the spectral measure on $\mathcal{S}_1^{d-1}$ induced by $\|\cdot\|_1$. Further, by constraining the measure ν to the set $A_1 := \left\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_1 > 1\right\}$, the constant c is normalized to 1. With a proper choice of $b(t) = F_R^{\leftarrow}(1 - 1/t), R = ||\mathbf{X}||_1$, the vague convergence in (2.3.1) implies the weak convergence on $\mathcal{B}(A_1)$, as

$$\nu_t(\cdot)|_{A_1} = \frac{\mathbb{P}\left(t^{-1}\mathbf{X} \in \cdot\right)}{\mathbb{P}\left(\|\mathbf{X}\|_1 > t\right)} \xrightarrow{\mathrm{w}} \nu(\cdot)|_{A_1}, \qquad t \to \infty, \tag{2.3.5}$$

where $\nu|_{A_1}$ is the restriction of ν to the set A_1.

With the MRV structure, all the information of upper tail dependence is provided by the limit measure ν. For ease of explanation, suppose the random vector $\mathbf{X} = (X_1, \ldots, X_d)$ consists of identical nonnegative components with distribution function F. Let the normalizing function be $b(t) = \left(1/\bar{F}\right)^{\leftarrow}(t)$, then from (2.3.1) we have

$$\frac{1}{\bar{F}(t)}\mathbb{P}\left(\frac{\mathbf{X}}{t} \in \cdot\right) \xrightarrow{\mathrm{v}} \nu(\cdot) \qquad \text{on } \mathcal{B}([0,\infty]^d \backslash \{\mathbf{0}\}).$$

Hence, for each pair $(X_i, X_j), 1 \le i \ne j \le d$, the coefficient of upper tail dependence is given by

$$\lambda_U = \nu\left\{\mathbf{x} \in [0,\infty]^d \backslash \{\mathbf{0}\} : x_i \wedge x_j > 1\right\}.$$

It is easy to see that if the limit measure ν spreads mass onto axes, then $\lambda_U = 0$, i.e. X_i and X_j are pairwise tail independent for every $1 \le i \ne j \le d$. Otherwise, we say some components of $\mathbf{X}$ are pairwise tail dependent. For further details on the limit measure ν or the spectral measure Ψ, see Section 6.5 of Resnick (2007) for related discussions.

Chapter 3

Asymptotic Analysis of Portfolio Diversification

3.1 Introduction

In order to mitigating risks in portfolios of financial investment, a common tool used by risk managers is the diversification strategy. The benefit from a diversification strategy can be reflected in the reduction of tail risks in a diversified portfolio. Guided by regulation rules such as the Basel II and III Accords for banking regulation and the Solvency II Directive for insurance regulation, the Value-at-Risk (VaR) became the main concern of the regulators, and therefore is also adopted by risk managers as the main measure of risks. In this chapter, we investigate the optimal portfolio construction aiming at extracting the most diversification benefit based on the VaR measure.

A key difficulty in evaluating the diversification benefit based on the VaR measure is that there is often no explicit formula for calculating the portfolio VaR. Since a portfolio is a linear combination of the underlying risky assets, only if the asset returns follow sum-stable distributions such as the Gaussian distribution or the stable distributions, one can precisely calculate the distribution of the portfolio return, and derive the VaR therefore. As an alternative, Extreme Value Theory (EVT), in particular, the multivariate regular variation (MRV) model, may provide an explicit approximation to the tail of the distribution of the portfolio return; see e.g. Mainik and Rüschendorf (2010), Mainik and Embrechts (2013) and Zhou (2010). By inverting the approximation formula on the tail of the distribution, one may get an approximation for the VaR measure, when the probability level in VaR is considered to be close to 1. Therefore, the EVT approach opens a new door for investigating the diversification benefit based on the VaR measure.

Nevertheless, when applying the EVT approach, two difficulties remain to be handled. Both of them are due to the fact that the approximation holds only in the limit when the probability level in VaR tending to 1. Firstly, the EVT approach provides an approximation for "the VaR in the limit" when the probability level in VaR tends to 1. However, for heavy-tailed portfolio returns as assumed in the setup of the MRV, when the probability level in VaR tends to 1, the VaR converges to infinity. Consequently, the goal of portfolio optimization turns to be minimizing "the VaR in the limit", even if the limit is infinity. It is difficult to provide an economic interpretation for such a mathematical exercise. Secondly, the practical goal for risk managers is to minimize VaR at a given probability level, such as 99% (Basel II) or 99.5% (Solvency II), while "the VaR in the limit" is not of their concern. Further, it is not guaranteed that the optimal portfolio based on minimizing " the VaR in the limit" is also close to the practical goal.

The first difficulty can be overcome by comparing the portfolio VaR to the VaRs of marginal risks. For that purpose, we employ the measure diversification ratio (DR), or sometimes with its alternative name: the risk concentration based on VaR; see, for example Degen et al. (2010) and Embrechts et al. (2009a). The diversification ratio is defined as follows. Let $\mathbf{X} := (X_1, \ldots X_d)^T$ be a random loss vector in $\mathbb{R}^d$, where $X_i > 0$ indicates losses and $X_i < 0$ represents gains. The portfolio loss is given by $\mathbf{w}^T\mathbf{X}$, where the weights satisfy $\mathbf{w} = (w_1, w_2, \ldots, w_d)^T \in \Sigma^d := \left\{\mathbf{x} \in [0,1]^d : x_1 + x_2 + \ldots + x_d = 1\right\}$. For this portfolio, the diversification ratio based on VaR at level $q \in (0,1)$ is defined as

$$\mathrm{DR}_{\mathbf{w},q} = \frac{\mathrm{VaR}_q(\mathbf{w}^T\mathbf{X})}{\sum_{i=1}^d w_i \mathrm{VaR}_q(X_i)}. \tag{3.1.1}$$

The DR is a measure of diversification benefit in the following sense. Consider the comonotonic case where all assets are completely dependent. Then DR is a constant one regardless how the portfolio is allocated. This is a special case in which any diversification strategy would not reduce the portfolio risk. Consequently, in a general case, $1 - \mathrm{DR}_{\mathbf{w},q}$ can be regarded as the diversification benefit.

The first result in this chapter is to show that the DR converges to a finite value for any portfolio as $q \to 1$ under the MRV model. More specifically, by modeling the joint distribution of the random vector $\mathbf{X}$ by MRV, we can derive an explicit formula for

$$\mathrm{DR}_{\mathbf{w},1} := \lim_{q\uparrow 1} \mathrm{DR}_{\mathbf{w},q}$$

with respect to the weight $\mathbf{w}$ and the two key elements characterizing the MRV model: the tail index of the marginals and the spectral measure for the tail dependence structure.[1]

This result overcomes the first difficulty regarding the interpretation: one may target minimizing the DR in the limit, which is at a finite level. We show that there exists a unique solution to the optimization problem

$$\mathbf{w}^* := \min_{\mathbf{w}\in\Sigma^d} \mathrm{DR}_{\mathbf{w},1}.$$

A portfolio that minimizes the DR is consequently extracting the most diversification benefit based on the VaR measure. It is also worth noticing that by taking the marginal VaRs in the denominator, the optimal portfolio based on the DR is mainly driven by the dependence structure across the risky assets, while is more robust to changes in marginal risks.

However, the second difficulty raised above remains valid after switching to minimizing the DR. Is the optimal solution based on minimizing the DR in the limit close to the practical goal of minimizing the DR at a given probability level? We formalize this question by the following notations.

Practically, with introducing the DR, risk managers aim at solving the following optimization problem:

$$\min_{\mathbf{w}\in\Sigma^d} \mathrm{DR}_{\mathbf{w},q}. \tag{3.1.2}$$

Denote the solution to (3.1.2) by $\mathbf{w}_q$.

We remark that solving (3.1.2) directly is computationally intensive. With observations on the joint distribution of the random vector $\mathbf{X}$, $\mathbf{w}_q$ can be estimated by conducting a numerical search. However, such a searching algorithm suffers from the dimensionality curse: the computational burden increases exponentially with respect to the dimension d.

The second main result of this chapter is to show how close the solution $\mathbf{w}^*$ is from the solution of the original optimization problem $\mathbf{w}_q$. First, we show theoretically that

$$\lim_{q\uparrow 1} \mathbf{w}_q = \mathbf{w}^*. \tag{3.1.3}$$

The convergence in (3.1.3) ensures that one may use the solution to the optimization problem in the limit as an approximation to the solution to the original problem with a

[1]As pointed out by Mainik and Embrechts (2013), under the MRV structure, when the tail index is great than 1, $\mathrm{DR}_{\mathbf{w},1} < 1$. In other words, the VaR measure possesses subadditivity as $q \to 1$. Hence, diversification is always optimal in this situation and the optimization problem (3.1.2) is well defined.

finite level q close to 1. Further, define the distance between $\mathbf{w}_q$ and $\mathbf{w}^*$, measured by $\|\mathbf{w}_q - \mathbf{w}^*\|$ with respect to an arbitrary norm as D_q. In other words, given a finite level of q close to 1, the solution $\mathbf{w}_q$ is within an area defined as a D_q radius circle around $\mathbf{w}^*$. For a special case of MRV, the Farlie–Gumbel–Morgenstern (FGM) copula, we explicitly determine D_q.

Empirically, with observations on the joint distribution of the random vector $\mathbf{X}$, one can estimate the two main components for the MRV: the marginal tail index and the spectral measure. By plugging in the estimates of these two elements, the solution $\mathbf{w}^*$ can be estimated using conventional convex optimization method. We show the consistency of the estimator. Notice that the computational burden is much lower than the aforementioned numerical approach for solving $\mathbf{w}_q$.

We use a few numerical examples to support our theoretical results and also apply our method to empirical data. We find that portfolio constructed using our approach possess the lowest DR and also suffers low losses in out–of–sample periods, compared to other portfolio optimization strategies,

One possible drawback of our portfolio optimization strategy (3.1.2) is that it only minimizes the risk without taking into account the upper side potential: portfolio returns. In fact, it is straightforward to consider the return components simultaneously. For example, consider the "safety–first" criterion proposed by Roy (1952), which aims at first constraining the downside risk to a given level and then maximizing the profit. This is equivalent to minimizing risk with a linear constraint on the returns. Comparing this optimization problem with the aforementioned unconstrained convex minimization problem, taking the return into consideration is just to impose an additional linear constraint. It is straightforward to verify that our current results remain valid for the constrained optimization problem. To avoid complicating the discussion, in this chapter we opt to focusing on the optimization of DR without considering the return side.

Our proposed portfolio optimization strategy is comparable to other strategies based on tail risk. Mainik and Rüschendorf (2010), proposed to minimize the so-called extreme risk index (ERI),

$$\mathrm{ERI} = \arg\min_{\mathbf{w}} \lim_{q \uparrow 1} \frac{\mathrm{VaR}_q(\mathbf{w}^T\mathbf{X})}{\mathrm{VaR}_q(||\mathbf{X}||_1)},$$

which essentially is minimizing the portfolio VaR. This strategy is more sensitive to marginal tail risks and consequently load high on marginals with a low VaR. On the contrary, minimizing DR in (3.1.1) scales off the effect of marginals and focuses more on the dependence structure. Another advantage of our DR strategy is that the DR measure is leverage invariant. For example, being 100% exposed to a risky portfolio is as diversified as being

50% exposed to such risky portfolio and leaving the rest in cash. However if we simply minimize portfolio VaR, the best strategy is to put more weight in cash, which does not increase diversification benefit.

Another closely related strategy is the so called most diversified portfolio (MDP)

$$\text{MDP} = \arg\min_{\mathbf{w}} \frac{\text{var}(\mathbf{w}^T\mathbf{X})}{\sum_{i=1}^{d} w_i \text{var}(X_i)},$$

proposed by Choueifaty and Coignard (2008). The MDP method shares the same structure with our approach: it considers the ratio between portfolio risk and the sum of individual risks measured by variances. Since variance is a measure of overall risk rather than focusing on the tail region, the MDP method may fail to capture the extreme risks.

This chapter is organized as follows. In Section 3.2, we provide our main results on the convergence of optimal portfolios. Section 3.3 discusses the convergence rate of the optimal portfolio. In Section 3.4, we demonstrate the empirical performance of our strategy based on two numerical examples. Section 3.5 and 3.6 provides the application of our strategy to real market data. Proofs are postponed to Section 3.7.

3.2 Convergence of Optimal Portfolios

3.2.1 Preliminaries

In this subsection, we give a general result on the convergence of minimizers. Throughout the chapter, for a function $g : Z \to \mathbb{R}$, we denote $M(g)$ the set of all the minimizers of g. That is,

$$M(g) = \left\{ x \in Z : g(x) = \inf_{y \in Z} g\left(y\right) \right\}.$$

A minimizer of g is denoted by $m_g \in M(g)$.

Lemma 3.2.1 *Suppose that $\{f_n\}$ is a sequence of lower semi-continuous functions from a compact metric space Z to $\overline{\mathbb{R}} = [-\infty, \infty]$, and f_n converges uniformly to a function f. If, in addition, assume that f has a unique minimum point in Z, then*

$$\lim_{n \to \infty} m_{f_n} = \arg\min f. \tag{3.2.1}$$

Proof. On the compact metric space Z, we have that the sequence $\{f_n\}$ is equi-coercive and gamma-converges to f under the conditions of Lemma 3.2.1. Then by Corollary 7.24 in Dal Maso (2012), the relation (3.2.1) holds. ■

3.2.2 Main Results

The first result regards the weak convergence of $\mathrm{DR}_{\mathbf{w},q}$ as $q \uparrow 1$, which is a direct consequence of known result in the literature.

Proposition 3.2.1 *Suppose the random vector* $\mathbf{X} \in \mathrm{MRV}_{-\alpha}(\Psi)$ *with* $\alpha > 0$. *Then for any* $\mathbf{w} \in \Sigma^d$, *we have*

$$\lim_{q\uparrow 1} \mathrm{DR}_{\mathbf{w},q} = \mathrm{DR}_{\mathbf{w},1},$$

where

$$\mathrm{DR}_{\mathbf{w},1} = \frac{\eta_{\mathbf{w}}^{1/\alpha}}{\sum_{i=1}^d w_i \eta_{\mathbf{e}_i}^{1/\alpha}} \tag{3.2.2}$$

with $\eta_{\mathbf{w}} = \int_{\mathcal{S}_1^{d-1}} (\mathbf{w}^T\mathbf{s})_+^\alpha \Psi(d\mathbf{s})$ *and* $\mathbf{e}_i = (0, ..., 1, ..., 0)^T$ *only the* i*th component being* 1 *for* $i = 1, .., d$.

Proof. Note that

$$\mathrm{DR}_{\mathbf{w},q} = \frac{\mathrm{VaR}_q(\mathbf{w}^T\mathbf{X})/\mathrm{VaR}_q(\|\mathbf{X}\|_1)}{\sum_{i=1}^d w_i \mathrm{VaR}_q(X_i)/\mathrm{VaR}_q(\|\mathbf{X}\|_1)}. \tag{3.2.3}$$

For $\mathbf{X} \in \mathrm{MRV}_{-\alpha}(\Psi)$ with $\alpha > 0$, it follows that

$$\lim_{q\uparrow 1} \frac{\mathrm{VaR}_q(\mathbf{u}^T\mathbf{X})}{\mathrm{VaR}_q(\|\mathbf{X}\|_1)} = \eta_{\mathbf{u}}^{1/\alpha}, \qquad \mathbf{u} \in \Sigma^d, \tag{3.2.4}$$

which can be found in e.g. Mainik and Rüschendorf (2010), Mainik and Embrechts (2013) and Zhou (2010). The proposition can be proved by letting $\mathbf{u} = \mathbf{w}$ and $\mathbf{u} = \mathbf{e}_i$ in (3.2.4). ■

In the following theorem, we develop the uniform convergence of $\mathrm{DR}_{\mathbf{w},q}$, which is essential for proving the convergence of minimizers. It is also an interesting result on its own. The proof is postponed to Section 3.7.

Theorem 3.2.1 *Suppose the random vector* $\mathbf{X}$ *is continuously distributed with a positive joint density function. Further assume that* $\mathbf{X} \in \mathrm{MRV}_{-\alpha}(\Psi)$ *with* $\alpha > 0$*. Then*

$$\lim_{q \uparrow 1} \sup_{\mathbf{w} \in \Sigma^d} |\mathrm{DR}_{\mathbf{w},q} - \mathrm{DR}_{\mathbf{w},1}| = 0. \tag{3.2.5}$$

The main result of this section, in the following theorem, shows that the convergence of a sequence of optimal solutions of $\mathrm{DR}_{\mathbf{w},q}$ to the unique minimizer of $\mathrm{DR}_{\mathbf{w},1}$.

Theorem 3.2.2 *Suppose the random vector* $\mathbf{X}$ *is continuously distributed with a positive joint density function. Further assume that*

- $\mathbf{X} \in \mathrm{MRV}_{-\alpha}(\Psi)$ *with* $\alpha > 1$,
- $\Psi\left(\left\{\mathbf{x} \in \mathcal{S}^{d-1} : \mathbf{a}^T\mathbf{x} = 0\right\}\right) = 0$ *for any* $\mathbf{a} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$,
- $\Psi\left(\left\{\mathbf{x} \in \mathcal{S}^{d-1} : \mathbf{w}^T\mathbf{x} \leq 0\right\}\right) = 1$ *for at most one vector* $\mathbf{w} \in \Sigma^d$.

Then $\mathbf{w}^* = \arg\min \mathrm{DR}_{\mathbf{w},1}$ *exists and is unique. Moreover,*

$$\lim_{q \uparrow 1} \mathbf{w}_q = \mathbf{w}^*, \tag{3.2.6}$$

where $\mathbf{w}_q$ *is a solution of* $\min_{\mathbf{w} \in \Sigma^d} \mathrm{DR}_{\mathbf{w},q}$.

Proof. The existence $\mathbf{w}^*$ is due to the continuity of $\mathrm{DR}_{\mathbf{w},1}$ and the compactness of Σ^d. To show the uniqueness, first note that the minimization problem $\min_{\mathbf{w} \in \Sigma^d} \mathrm{DR}_{\mathbf{w},1}$ is equivalent to

$$\begin{aligned} &\min_{\mathbf{w}} \eta_{\mathbf{w}}^{1/\alpha} \\ &\textbf{s.t.} \quad \textstyle\sum_{i=1}^d w_i \eta_{\mathbf{e}_i}^{1/\alpha} = 1 \text{ with } w_i \geq 0 \text{ for } i = 1, 2, \ldots, d. \end{aligned} \tag{3.2.7}$$

Since the set of constraints in (3.2.7) is nonempty, closed and bounded, it is compact. By Theorem 2.6 of Mainik and Embrechts (2013), $\eta_{\mathbf{w}}^{1/\alpha}$ is strictly convex when $\alpha > 1$. Suppose $\mathbf{w}_1$ and $\mathbf{w}_2$ are two different minimal points of the optimization problem. Let $\mathbf{w} = (\mathbf{w}_1 + \mathbf{w}_2)/2$. From the strictly convexity of the object function and compactness of the set of constraints, it follows that $\eta_{\mathbf{w}}^{1/\alpha} < \eta_{\mathbf{w}_1}^{1/\alpha} = \eta_{\mathbf{w}_2}^{1/\alpha}$, which yields a contradiction. Thus, $\mathbf{w}^*$ is unique.

Now we prove (3.2.6). In the proof of Theorem 3.7.2, we showed that $\mathrm{VaR}_q(\mathbf{w}^T\mathbf{X})$ is continuous with respect to $\mathbf{w} \in \Sigma^d$ for q large. Then there exists $q^* > 0$ such that $\mathrm{DR}_{\mathbf{w},q}$ is continuous with respect to $\mathbf{w} \in \Sigma^d$ for every $q^* < q < 1$. The desired result follows from Theorem 3.2.1, the uniqueness of $\mathbf{w}^*$ and Lemma 3.2.1. ■

Remark 3.2.1 *Endowment of the DR approach with a target return is straightforward. Analogously to the classical Markowitz problem, it suffices to add the linear constraint*

$$\mathbf{w}^T \boldsymbol{\mu} = -\bar{\mu}, \qquad \bar{\mu} > 0, \tag{3.2.8}$$

to the optimization problem

$$\min_{\mathbf{w} \in \Sigma^d} \mathrm{DR}_{\mathbf{w},1}.$$

The condition $\alpha > 1$ guarantees the existence of first moment for each X_i and the negative sign of the target return is due to the fact that $\mathbf{X}$ is a loss vector. Note that the subspace $\Sigma^d \cap \{\mathbf{w} : \mathbf{w}^T \boldsymbol{\mu} = -\bar{\mu}\}$ is again compact. Then by the continuity of $\mathrm{DR}_{\mathbf{w},1}$, at least one optimal solution exists. Since the objective function $\mathrm{DR}_{\mathbf{w},1}$ in (3.2.2) is the ratio of a convex function and an affine function, our problem is in the field of fractional programming as described in Schaible and Ibaraki (1983) and Avriel et al. (2010). A well-known result in the theory of fractional programming is that a local minimum is global and unique if the numerator is strictly convex, since in this case the objective function is strictly quasiconvex and can be related to a convex optimization problem through transformations. Therefore, we show the existence and uniqueness of $\mathbf{w}^*$ for the optimization problem with constraint in (3.2.8).

Remark 3.2.2 *In Theorem 3.2.2, $\mathbf{w}^*$ is the optimal portfolio obtained by minimizing the diversification ratio with respect to VaR at limit 1. Due to the Karamata's theorem (see Theorem B.1.5 in de Haan and Ferreira (2006)), for $\alpha > 1$, the following asymptotic relation always holds,*

$$\lim_{q \uparrow 1} \frac{\mathrm{ES}_q(\mathbf{w}^T \mathbf{X})}{\mathrm{VaR}_q(\mathbf{w}^T \mathbf{X})} = \frac{\alpha}{\alpha - 1}.$$

Consequently, $\mathbf{w}^*$ also minimizes the diversification ratio with respect to ES for $q \uparrow 1$. The asymptotic result can be also generalized to the class of spectral risk measures by imposing certain constraint on the admissible risk spectrum ϕ. Interested readers may refer to Section 5 in Mainik and Rüschendorf (2010).

3.2.3 Beyond the Main Theorem

In our main result, Theorem 3.2.2, some restrictions are imposed on the index α and spectral measure Ψ to make sure that the optimization problem is well defined. In fact, they are not necessary conditions. In the following several special cases, we show that the conditions can be relaxed.

The condition $\Psi\left(\left\{\mathbf{x} : \mathbf{a}^T\mathbf{x} = 0\right\}\right) = 0$ for any $\mathbf{a} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ means that the spectral measure Ψ does not concentrate on any linear subspace. It ensures the uniqueness of the solution $\mathbf{w}^*$ of the limiting problem $\mathrm{DR}_{\mathbf{w},1}$. But it excludes the special cases such as independent or comonotonic structure of $\mathbf{X}$. If $\mathbf{X}$ has an independent structure with regularly varying marginals, then it is not hard to show that

$$\mathrm{DR}_{\mathbf{w},1} = \sum_{k=1}^{d} w_k^\alpha.$$

By Jensen's inequality, $\mathrm{DR}_{\mathbf{w},1}$ is minimized when $w_k = 1/d$ for $k = 1, 2, ..., d$, which is unique. Therefore, Theorem 3.2.2 holds for the independent case. If $\mathbf{X}$ is comonotonic, then $\mathrm{DR}_{\mathbf{w},q} = 1$ for any $\mathbf{w}$ or q. There is no optimization problem to consider.

If we restrict ourselves to elliptical distributions, then Theorem 3.2.2 holds for any random vector $\mathbf{X} \in \mathbb{R}^d$ and any $\alpha > 0$, without any restriction on Ψ, or even without the MRV assumption. In the rest of the section, we focus on this special case.

A random vector $\mathbf{X}$ in $\mathbb{R}^d$ is elliptically distributed if it satisfies

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + YB\mathbf{U}, \tag{3.2.9}$$

where $\boldsymbol{\mu} \in \mathbb{R}^d, B \in \mathbb{R}^{d\times d}, \mathbf{U} = (U_1, ..., U_d)^T$ is uniformly distributed on the Euclidean sphere $\mathcal{S}^{d-1}$, and Y is a nonnegative random variable that is independent of $\mathbf{U}$. The matrix $C := BB^T$ is called ellipticity matrix of $\mathbf{X}$. To avoid degenerate cases, we assume throughout the following that C is positive definite.

It is well known that if $\mathbf{X}$ is elliptically distributed, then $\mathbf{X} \in \mathrm{MRV}_{-\alpha}(\Psi)$ if and only if $Y \in \mathrm{RV}_{-\alpha}$; for example, see Hult and Lindskog (2002). By Theorem 6.8 of McNeil et al. (2015), the subadditivity property of VaR always holds for $0.5 \le q < 1$. It then follows that $\mathrm{DR}_{\mathbf{w},q} \le 1$, which means that diversification is always optimal for $0.5 \le q < 1$ no matter what distribution Y follows and thus the optimization problem is well defined. In the general MRV case, to have $\mathrm{DR}_{\mathbf{w},q} \le 1$ is ensured by restricting $\alpha > 1$. In another word, if $\mathbf{X}$ is elliptically distributed and $Y \in \mathrm{RV}_{-\alpha}$, then Theorem 3.2.2 holds without any restriction on α.

Actually, elliptical distributions leads to the explicit expressions of $\mathrm{DR}_{\mathbf{w},q}$ and $\mathrm{DR}_{\mathbf{w},1}$. This enables us to further relax the assumption of MRV. As long as Y is unbounded, we are able to directly show the convergence of (3.2.6) without the assumption that Y is regularly varying. A direct calculation yields that

$$\mathrm{VaR}_q\left(\mathbf{w}^T\mathbf{X}\right) = \mathbf{w}^T\boldsymbol{\mu} + \left\|B^T\mathbf{w}\right\|_2 F_Z^{\leftarrow}(q), \tag{3.2.10}$$

where $Z \stackrel{d}{=} YU_1$. The diversification ratio for elliptical distributions can then be obtained as

$$\mathrm{DR}_{\mathbf{w},q} = \frac{\mathbf{w}^T \boldsymbol{\mu} + \left\| B^T \mathbf{w} \right\|_2 F_Z^{\leftarrow}(q)}{\mathbf{w}^T \boldsymbol{\mu} + \sum_{i=1}^d w_i \left\| B^T \mathbf{e}_i \right\|_2 F_Z^{\leftarrow}(q)}. \tag{3.2.11}$$

If the random variable Y is unbounded, then by $F_Z^{\leftarrow}(q) \to \infty$ as $q \uparrow 1$, we obtain

$$\lim_{q \uparrow 1} \mathrm{DR}_{\mathbf{w},q} = \frac{\left\| B^T \mathbf{w} \right\|_2}{\sum_{i=1}^d w_i \left\| B^T \mathbf{e}_i \right\|_2} := \mathrm{DR}_{\mathbf{w},1}. \tag{3.2.12}$$

In the following lemma, we first show that the convergence in (3.2.12) is indeed uniform, whose proof is postponed to the last section.

Lemma 3.2.2 *For elliptically distributed* $\mathbf{X}$*, if* $\|\boldsymbol{\mu}\|_1 < \infty$ *and random variable* Y *is unbounded, then the convergence in (3.2.12) is uniform for* $\mathbf{w} \in \Sigma^d$*. Moreover, the mapping* $\mathbf{w} \to \mathrm{DR}_{\mathbf{w},1}$ *is continuous.*

Now we are ready to show that Theorem 3.2.2 holds in the most general setting of elliptical distributions by dropping the MRV assumption.

Theorem 3.2.3 *Under the conditions of Lemma 3.2.2, we have*

$$\lim_{q \uparrow 1} \arg\min_{\mathbf{w} \in \Sigma^d} \frac{\mathrm{VaR}_q(\mathbf{w}^T \mathbf{X})}{\sum_{i=1}^d w_i \mathrm{VaR}_q(X_i)} = \arg\min_{\mathbf{w} \in \Sigma^d} \frac{\left\| B^T \mathbf{w} \right\|_2}{\sum_{i=1}^d w_i \left\| B^T \mathbf{e}_i \right\|_2}. \tag{3.2.13}$$

Proof. By Lemmas 3.2.1 and 3.2.2, we only need to show that the solutions of the minimization problems on both sides of (3.2.13) exist and are unique. To achieve it, first note that the minimization problem

$$\min_{\mathbf{w} \in \Sigma^d} \frac{\mathrm{VaR}_q(\mathbf{w}^T \mathbf{X})}{\sum_{i=1}^d w_i \mathrm{VaR}_q(X_i)}$$

is equivalent to a convex optimization problem

$$\begin{aligned} &\min_{\mathbf{w}} \mathbf{w}^T \boldsymbol{\mu} + \left\| B^T \mathbf{w} \right\|_2 F_Z^{\leftarrow}(q) \\ &\textbf{s.t.} \quad \mathbf{w}^T \boldsymbol{\mu} + \textstyle\sum_{i=1}^d w_i \left\| B^T \mathbf{e}_i \right\|_2 F_Z^{\leftarrow}(q) = 1 \text{ with } w_i \geq 0 \text{ for } i = 1, 2, \ldots, d. \end{aligned} \tag{3.2.14}$$

Similarly, the minimization problem

$$\min_{\mathbf{w}\in\Sigma^d} \frac{\left\|B^T\mathbf{w}\right\|_2}{\sum_{i=1}^d w_i \left\|B^T\mathbf{e}_i\right\|_2}$$

is equivalent to

$$\begin{aligned}&\min_{\mathbf{w}} \left\|B^T\mathbf{w}\right\|_2 \\ &\mathbf{s.t.} \quad \textstyle\sum_{i=1}^d w_i \left\|B^T\mathbf{e}_i\right\|_2 = 1 \text{ with } w_i \geq 0 \text{ for } i = 1, 2, \ldots, d.\end{aligned} \tag{3.2.15}$$

Denote the constraint sets in (3.2.14) and (3.2.15) by C_1 and C_2. It is obvious that C_1 and C_2 are nonempty, closed, convex and bounded. Hence, they are compact by the Heine–Borel theorem. By the triangle inequality and positive homogeneity of $\|\cdot\|_2$, the objective functions in (3.2.14) and (3.2.15) are convex over $\mathbb{R}^d$, and they are continuous over the constraint sets C_1 and C_2; see Rockafellar (2015). By the compactness of the constraint set and continuity of the objective functions, the solutions to (3.2.14) and (3.2.15) exist due to the Weierstrass extreme value theorem.

Next, we show the uniqueness of the solution to (3.2.15). Due to the convexity, we have for any $\lambda \in (0, 1)$,

$$\left\|B^T\left(\lambda\mathbf{w}_1 + (1-\lambda)\,\mathbf{w}_2\right)\right\|_2 \leq \lambda\left\|B^T\mathbf{w}_1\right\|_2 + (1-\lambda)\left\|B^T\mathbf{w}_2\right\|_2. \tag{3.2.16}$$

The equality in (3.2.16) holds only when $\mathbf{w}_1 = k\mathbf{w}_2$ for $k \in \mathbb{R}^+$ and $\mathbf{w}_1, \mathbf{w}_2$ nonzero. If both $\mathbf{w}_1$ and $\mathbf{w}_2$ belong to the constraint set C_1 or C_2, then k can only be 1. This means for any $\mathbf{w}_1 \neq \mathbf{w}_2$, the strictly inequality in (3.2.16) holds. Therefore, the objective function in (3.2.15) is strictly convex. The uniqueness of the solution then follows from the similarly arguments in the proof of Theorem 3.2.2. ■

3.2.4 Estimation of the Diversification Ratio

When the DR optimization strategy with MRV structure is applied in practice, the estimations of MRV structure and $\mathrm{DR}_{\mathbf{w},1}$ are required. In this section, we propose an estimation procedure and show the consistency of the estimators.

Assume $\mathbf{X} \in \mathrm{MRV}_{-\alpha}(\Psi)$ with $\alpha > 1$. Let $\mathbf{X}_1, \ldots \mathbf{X}_n$ be an i.i.d. sample of $\mathbf{X}$. By Theorem 3.2.1, we propose the following estimation procedure.

1. Estimate the tail index α by an estimator $\widehat{\alpha}$.

2. Estimate the spectral measure Ψ by an estimator $\widehat{\Psi}$.

3. Estimate $\eta_{\mathbf{w}}$ by
$$\widehat{\eta}_{\mathbf{w}} = \int_{\mathcal{S}^{d-1}} (\mathbf{w}^T\mathbf{s})_+^{\widehat{\alpha}} \widehat{\Psi}(d\mathbf{s}).$$

4. Estimate $\mathrm{DR}_{\mathbf{w},1}$ by
$$\widehat{\mathrm{DR}}_{\mathbf{w},1} = \frac{\widehat{\eta}_{\mathbf{w}}^{1/\alpha}}{\sum_{i=1}^d w_i \widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}}.$$

With the estimated diversification ratio, we can obtain an optimal portfolio by minimizing $\widehat{\mathrm{DR}}_{\mathbf{w},1}$. Denote the optimal portfolio weights following this procedure as $\hat{\mathbf{w}}^*$.

More specifically, in the first two steps, we use standard estimators for α and Ψ as follows. Let (R, S) and (R_i, S_i) denote the polar coordinates of $\mathbf{X}$ and $\mathbf{X}_i$ with respect to $||\cdot||_1$. That is,
$$(R, S) = \left(||\mathbf{X}||_1, \frac{\mathbf{X}}{||\mathbf{X}||_1} \right). \tag{3.2.17}$$
Assume in this section that the distribution function of R is continuous. Choose an intermediate sequence k such that
$$k\left(n\right) \to \infty, \qquad \frac{k\left(n\right)}{n} \to 0.$$
We use the observations corresponding to the top k order statistics of $R_1, \ldots, R_n$ for estimating α and Ψ. Denote the k upper order statistics of $R_1, \ldots, R_n$ by $R_{(n)} \geq \ldots \geq R_{(n-k+1)}$. The tail index α is estimated by some usual estimator as a function of these order statistics:
$$\widehat{\alpha} = \widehat{\alpha}\left(R_{(n)}, \ldots, R_{(n-k+1)}\right).$$
Many existing estimators can be applied here, see for example, Hill (1975), Pickands III (1975), Smith (1987), Dekkers et al. (1989), among others. They all possess consistency and asymptotic normality.

Next, let $\pi\left(1\right), \ldots, \pi\left(k\right)$ denote the indices corresponding to $R_{(n)}, \ldots, R_{(n-k+1)}$ in the original sequence $R_1, \ldots, R_n$. These indices are used to identify each "angle" $S_{\pi(j)}$ corresponding to $R_{(j)}$. The spectral measure Ψ is estimated by the empirical measure of the angular parts $S_{\pi(1)}, \ldots, S_{\pi(k)}$,
$$\widehat{\Psi} = \frac{1}{k}\sum_{j=1}^{k} \delta_{S_{\pi(j)}}, \tag{3.2.18}$$
where $\delta_{_{\pi(j)}}(\cdot)$ is the Dirac measure.

Lemma 3.2.3 *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *be an i.i.d. sample of* $\mathbf{X} \in \mathrm{MRV}_{-\alpha}(\Psi)$ *with* $\alpha > 1$. *Assume that the distribution function* F_R *of* R *in (3.2.17) is continuous. If the estimator* $\widehat{\alpha}$ *is consistent almost surely, and then the estimator* $\widehat{\mathrm{DR}}_{\mathbf{w},1}$ *is consistent uniformly in* $\mathbf{w} \in \Sigma^d$, *i.e.,*

$$\sup_{\mathbf{w}\in\Sigma^d} \left| \widehat{\mathrm{DR}}_{\mathbf{w},1} - \mathrm{DR}_{\mathbf{w},1} \right| \to 0, \qquad a.s. \tag{3.2.19}$$

Combining Theorem 3.2.1 and Lemma 3.2.3, we obtain the consistency in the optimal portfolio weights in the following theorem.

Theorem 3.2.4 *Under the conditions of Lemma 3.2.3 and* $\Psi\left(\left\{\mathbf{x} : \mathbf{a}^T\mathbf{x} = 0\right\}\right) = 0$ *for any* $\mathbf{a} \in \mathbb{R}^d \backslash \{\mathbf{0}\}$, *the estimator* $\widehat{\mathbf{w}}^*$ *and the estimated value* $\widehat{\mathrm{DR}}_{\mathbf{w}^*,1}$ *are consistent almost surely, i.e.,*

$$\widehat{\mathbf{w}}^* \to \mathbf{w}^*, \qquad a.s.; \qquad \widehat{\mathrm{DR}}_{\mathbf{w}^*,1} \to \mathrm{DR}_{\mathbf{w}^*,1}, \qquad a.s.$$

Here we only established consistency. Under some additional conditions, further asymptotic properties for the estimator of $\mathrm{DR}_{\mathbf{w},1}$ can be established in a straightforward way. For example, Theorem 4.5 of Mainik and Rüschendorf (2010) shows that, under some additional conditions, for any $\mathbf{w} \in \Sigma^d$, $\sqrt{k}\,(\widehat{\eta}_{\mathbf{w}} - \eta_{\mathbf{w}})$ converges to a multivariate Gaussian distribution $G_{\mathbf{w}}$. Then by the functional delta method (e.g. Theorem 20.8 in Van der Vaart (2000)), it is easy to show that $\sqrt{k}\left(\widehat{\mathrm{DR}}_{\mathbf{w},1} - \mathrm{DR}_{\mathbf{w},1}\right)$ converges to a Gaussian distribution as well. However, to establish the convergence in an uniform way is difficult and may be left for future research. Without a uniform asymptotic property on $\widehat{\mathrm{DR}}_{\mathbf{w},1}$ we cannot further investigate the asymptotic property of the optimal portfolio weights.

3.3 The Rate of Convergence to the Optimal Portfolio: An Example

In this section, we discuss how $\mathbf{w}^*$ approximates $\mathbf{w}_q$ by determining the convergence rate of (3.2.6) under some special dependence structure, such as the FGM copula.

The FGM copula was originally introduced by Morgenstern (1956) and investigated by Gumbel (1960) and Farlie (1960). The FGM copula is defined as

$$C(u, v) = uv(1 + \theta(1 - u)(1 - v)), \quad (u, v) \in [0, 1]^2, \tag{3.3.1}$$

where $\theta \in [-1, 1]$ is a dependence parameter. This model has been generalized in various ways, for example, from two dimensions to higher dimensions or with more general form of $(1-u)(1-v)$ in (3.3.1); see Cambanis (1977), Fischer and Klein (2007), among others. Here we focus on a high dimensional generalized FGM copula proposed by Cambanis (1977), which is defined as

$$C(u_1, \ldots, u_n) = \prod_{k=1}^{n} u_k \left(1 + \sum_{1 \le i < j \le n} a_{ij}(1 - u_i)(1 - u_j) \right), \quad (u_1, \ldots, u_n) \in [0, 1]^n. \quad (3.3.2)$$

The constants $a_{i,j}$, $1 \le i < j \le n$, are so chosen that $C(u_1, \ldots, u_n)$ is a proper copula. A necessary and sufficient condition on $a_{i,j}$'s is that they satisfy a set of 2^n inequalities

$$1 + \sum_{1 \le i < j \le n} \epsilon_i \epsilon_j a_{ij} \ge 0 \quad \text{for all } (\epsilon_1, \ldots, \epsilon_n) \in \{-1, 1\}^n.$$

An FGM copula defined as in (3.3.2) is asymptotically independent.

We intend to consider the random vector $\mathbf{X}$ following an FGM copula with identical regularly varying marginals. For that purpose we need a second-order convergence in Proposition 3.2.1. This further requires the second-order expansion of tail probabilities of the weighted sum

$$\overline{F}_{\mathbf{w}^T\mathbf{X}}(t) = \mathbb{P}\left(\mathbf{w}^T \mathbf{X} > t \right),$$

where $F_{\mathbf{w}^T\mathbf{X}} = 1 - \overline{F}_{\mathbf{w}^T\mathbf{X}}$ is the distribution function of $\mathbf{w}^T\mathbf{X}$. In the next subsection, we present this result.

3.3.1 Tail Expansion for the Weighted Sum

Assume that the random vector $\mathbf{X}$ has a common marginal distribution function $G = 1 - \overline{G}$. Further, assume $\overline{G}$ to be second-order regularly varying (2RV), denoted by $\overline{G} \in 2\mathrm{RV}_{-\alpha,\rho}$. That is, there exist some $\rho \le 0$ and a measurable function $A(\cdot)$, which does not change sign eventually and converges to 0, such that, for all $x > 0$,

$$\lim_{t \to \infty} \frac{\overline{G}(tx)/\overline{G}(t) - x^{-\alpha}}{A(t)} = x^{-\alpha} \frac{x^{\rho} - 1}{\rho} =: H_{-\alpha,\rho}(x). \quad (3.3.3)$$

When $\rho = 0$, $H_{-\alpha,\rho}(x)$ is understood as $x^{-\alpha} \log x$.

For simplicity, here we only consider the case $\alpha > 1$ which implies that $\mathbf{X}$ has a finite mean. The results for $0 < \alpha \le 1$ can be obtained in a similar way. The proof of the next lemma is postponed.

Lemma 3.3.1 *Let* $\mathbf{X}$ *be a nonnegative random vector with identically distributed marginal with common distribution function* G *satisfying that* $\overline{G} \in 2\mathrm{RV}_{-\alpha,\rho}$ *with* $\alpha > 1$, $\rho \leq 0$ *and auxiliary function* $A(\cdot)$. *Assume that* $\mathbf{X}$ *follows an* n*-dimensional generalized FGM copula given by (3.3.2). Then as* $t \to \infty$, *we have that*

$$\begin{aligned}
&\frac{\overline{F}_{\mathbf{w}^T\mathbf{X}}(t)}{\overline{G}(t)} - \sum_{k=1}^{d} w_k^{\alpha} \\
&= \begin{cases} \alpha t^{-1}\mu_G^*(1+o(1)), & \rho < -1, \\ (1+Q_{\mathbf{a}})\sum_{k=1}^{d} H_{-\alpha,\rho}\left(w_k^{-1}\right)A(t)(1+o(1)), & \rho \geq -1, \end{cases}
\end{aligned} \tag{3.3.4}$$

where $H_{-\alpha,\rho}(\cdot)$ *is given in (3.3.3),* $Q_{\mathbf{a}} = \sum_{1\leq i<j\leq n} a_{ij}$, $\mu_G = \int_0^\infty x dF(x)$, $\mu_{G^2} = \int_0^\infty x dF^2(x)$, *and*

$$\begin{aligned}
\mu_G^* = &(1+Q_{\mathbf{a}})\,\mu_G \sum_{k\neq l} w_k^{\alpha} w_l \\
&+ \sum_{i<j} a_{i,j}\left(\sum_{k,l=i,j}\left(\sum_{l\neq k}\mu_{G^2} w_k^{\alpha} w_l - \mu_G w_k \sum_{m\neq i,j} w_m^{\alpha} - 2\mu_G w_k^{\alpha} w_l - \mu_G w_k^{\alpha} w_l\right)\right) \\
&- \sum_{i<j} a_{i,j}\sum_{k\neq i,j}\sum_{l\neq k,i,j}\mu_G w_k^{\alpha} w_l.
\end{aligned}$$

Further, the convergence in (3.3.4) is uniform for all $\mathbf{w} \in \Sigma^d$.

3.3.2 Convergence Rate

We first show a general lemma regarding the convergence rate of minimizers under the setup of Lemma 3.2.1. Define the distance between f_n and f as $D_n = ||f_n - f||_\infty$, where $||\cdot||_\infty$ is the supremum norm. The distance between m_{f_n} and $\arg\min f$ is defined as $||m_{f_n} - \arg\min f||_\Box$ for a norm $||\cdot||_\Box$ on the space Z. Since Z is a metric space, all the norms on Z are equivalent in the sense that there exist constants c_1 and c_2 such that

$$c_1||x||_\Box \leq ||x||_\Diamond \leq c_2||x||_\Box, \qquad x \in Z,$$

for any two norms $||\cdot||_\Box$ and $||\cdot||_\Diamond$ on Z. In case no confusion arises, the norm index ∞ or $\Box$ is dropped in the rest of the chapter.

Lemma 3.3.2 *Under the assumptions of Lemma 3.2.1, we have for n large*

$$||m_{f_n} - \arg\min f|| < C\sqrt{D_n},$$

where $D_n = ||f_n - f||_\infty$ and C is a constant.

Lemma 3.2.1 shows that m_{f_n}, the minimizer of function f_n, can be approximated by the minimizer of the limiting function m_f, which is usually much easier to calculate. The result in Lemma 3.3.2 further explores how good the approximation is. In practice, if we can determine D_n, which is related to the second-order expansion of f_n, then the error of the approximation can be determined.

Now we are ready to determine the convergence rate of the optimal portfolio under the FGM copula.

Theorem 3.3.1 *Under the conditions of Lemma 3.3.1, we have that*

$$(1-q)^{(-1\vee\rho)/\alpha}\left\|\mathbf{w}_q - \mathbf{d}^{-1}\right\| = O(1),$$

where $\mathbf{w}_q$ is a solution of $\min_{\mathbf{w}\in\Sigma^d}\mathrm{DR}_{\mathbf{w},q}$, and $\mathbf{d}^{-1} = (1/d, ..., 1/d)^T$.

Proof. In this proof, all the limits are taken as $q \uparrow 1$. We first derive the second-order expansion of $\mathrm{DR}_{\mathbf{w},q}$. Similar to the proof of Theorem 4.6 in Mao and Yang (2015), we have that

$$U\left(\frac{1}{\overline{F}_{\mathbf{w}^T\mathbf{X}}(F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q))}\right) = G^{\leftarrow}(q) + o(A(G^{\leftarrow}(q))),$$

where $U(\cdot)$ is the tail quantile function of G defined as $U(\cdot) = (1/\overline{G})^{\leftarrow}(\cdot) = G^{\leftarrow}(1-1/\cdot)$. For simplicity, denote $t = F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q)$. It is easy to see that $t \to \infty$ as $q \uparrow 1$. Then noting that $U(1/\overline{G}(t)) = t + o(A(t))$ and by the uniform convergence of (3.3.3), it follows that

$$\begin{aligned}
\mathrm{DR}_{\mathbf{w},q} = \frac{F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q)}{G^{\leftarrow}(q)} &= \frac{U(1/\overline{G}(t))}{U(1/\overline{F}_{\mathbf{w}^T\mathbf{X}}(t))} + o(A(t)) \\
&= \left(\frac{\overline{F}_{\mathbf{w}^T\mathbf{X}}(t)}{\overline{G}(t)}\right)^{1/\alpha} + H_{1/\alpha,\rho/\alpha}\left(\frac{\overline{F}_{\mathbf{w}^T\mathbf{X}}(t)}{\overline{G}(t)}\right)\alpha^{-2}A(U(1/\overline{F}_{\mathbf{w}^T\mathbf{X}}(t)))(1+o(1)) \\
&= \begin{cases} \left(\sum_{k=1}^d w_k^\alpha\right)^{1/\alpha}\left(1 + \mu_G^*\left(\sum_{k=1}^d w_k^\alpha\right)^{-1/\alpha-1}\left(G^{\leftarrow}(q)\right)^{-1}(1+o(1))\right), & \rho < -1, \\ \left(\sum_{k=1}^d w_k^\alpha\right)^{1/\alpha}\left(1 + \tau_\alpha A(G^{\leftarrow}(q))(1+o(1))\right), & \rho > -1. \end{cases}
\end{aligned} \tag{3.3.5}$$

where

$$\tau_\alpha = \frac{(1+Q_a)\sum_{k=1}^d H_{-\alpha,\rho}\left(w_k^{-1}\right)}{\alpha\sum_{k=1}^d w_k^\alpha} + \frac{\left(\sum_{k=1}^d w_k^\alpha\right)^{\rho/\alpha}}{\rho\alpha}.$$

This gives the second-order expansion of $\mathrm{DR}_{\mathbf{w},q}$.

Immediately from (3.3.5), the limiting function is

$$\lim_{q\uparrow 1}\mathrm{DR}_{\mathbf{w},q} = \left(\sum_{k=1}^d w_k^\alpha\right)^{1/\alpha} = \mathrm{DR}_{\mathbf{w},1}.$$

By Jensen's inequality, $\mathrm{DR}_{\mathbf{w},1}$ is uniquely minimized at $\mathbf{d}^{-1} = (1/d, ..., 1/d)^T$. If $\rho < -1$, then

$$\mathrm{DR}_{\mathbf{w},q} - \left(\sum_{k=1}^d w_k^\alpha\right)^{1/\alpha} = \mu_G^*\left(\sum_{k=1}^d w_k^\alpha\right)^{-1}\left(G^{\leftarrow}(q)\right)^{-1}(1+o(1)).$$

By Lemma 3.3.1, the above convergence is uniform. Hence, we have that for some constant $C > 0$

$$\left|\mathrm{DR}_{\mathbf{w},q} - \left(\sum_{k=1}^d w_k^\alpha\right)^{1/\alpha}\right| < C\left(G^{\leftarrow}(q)\right)^{-1}.$$

By Lemma 3.3.2, we get that

$$(1-q)^{-1/\alpha}\left\|\mathbf{w}_q - \mathbf{d}^{-1}\right\| = O(1).$$

Similarly, if $\rho > -1$, then

$$\mathrm{DR}_{\mathbf{w},q} - \left(\sum_{k=1}^d w_k^\alpha\right)^{1/\alpha} = \left(\sum_{k=1}^d w_k^\alpha\right)^{1/\alpha}\tau_\alpha A(G^{\leftarrow}(q))(1+o(1)).$$

Since for any $\mathbf{w} \in \Sigma^d$

$$\tau_\alpha \leq \frac{(1+Q_a)\,\rho d^{(\alpha-1)^2/\alpha} + d^{\rho(1-\alpha)/\alpha}}{\rho\alpha},$$

we obtain that for some constant $C > 0$

$$\left|\mathrm{DR}_{\mathbf{w},q} - \left(\sum_{k=1}^d w_k^\alpha\right)^{1/\alpha}\right| < CA(G^{\leftarrow}(q)).$$

By Lemma 3.3.2 we get that

$$(1-q)^{\rho/\alpha} \left\| \mathbf{w}_q - \mathbf{d}^{-1} \right\| = O(1).$$

This completes the proof. ∎

3.4 Numerical Examples

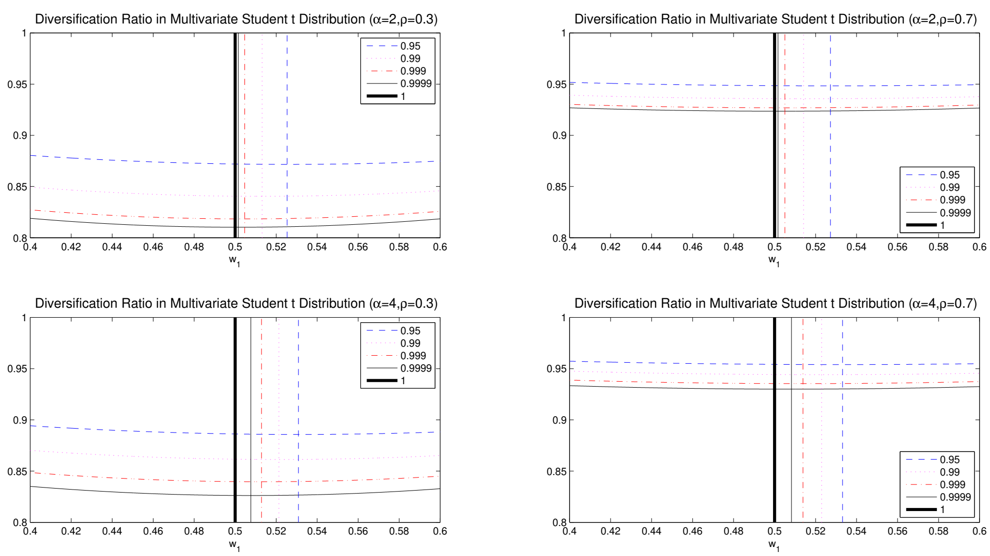
In this section, we conduct two numerical examples to examine our theoretical results. The first example is an elliptical distribution–the bivariate Student-t distribution, while the second one is a non-elliptical distribution.

Consider $\mathbf{X}$ follows a bivariate Student-t distribution $t_\alpha(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = (1,2)^T$ and the covariance matrix Σ is $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Then the marginals both follow Student-t distribution with the degree of freedom α but different shifts 1 and 2.

We construct portfolios as a linear combination of the two risk factors from $\mathbf{X}$ defined above. As discussed in Section 3.2.3, both $\mathrm{DR}_{\mathbf{w},q}$ and $\mathrm{DR}_{\mathbf{w},1}$ can be explicitly expressed for elliptical distributions as in (3.2.11) and (3.2.12), which are used in this example. In Figure 3.1, we plot the diversification ratio of such portfolios for various values of q against the weight w_1. For the parameters, we choose α and ρ at $\alpha = 2, 4$ and $\rho = 0.3, 0.7$, and plot the results for different pairs of (α, ρ) in the four subfigures in Figure 3.1. The level of q is set to 0.95, 0.99, 0.999 and 0.9999. For each q level, we indicates the optimal portfolio weight on w_1 by a vertical line, which is given at the lowest point of the convex diversification ratio curve. Notice that due to the different shifts, the optimal portfolio at a finite q level tends to load higher on the first dimension with a lower mean. However, as $q \to 1$, the difference in the mean plays no role in the limit of the diversification ratio. Therefore, due to symmetry, the optimal portfolio for $q = 1$ load equal weights on the two dimensions. We indicate this optimal solution for the limit diversification ratio by a thick vertical line located at 0.5.

First, we observe that $\mathbf{w}_q$ is converging to $\mathbf{w}_1$ as $q \uparrow 1$. This verifies our theoretical result in Theorem 3.2.2. Second, the absolute difference between $\mathbf{w}_q$ and $\mathbf{w}_1$ remains at a low level across all subfigures. For example, when focusing on approximating the optimal portfolio based on diversification ratio at $q = 0.99$ level, if one takes the optimal weight for the limit diversification ratio 0.5 as an approximation, then she makes an error for loading 2% less on the first dimension. Third, given the level of dependence (ρ), the heavier the

Figure 3.1: Optimal portfolio from elliptical distribution risk factors



Note: The portfolios are constructed as a linear combination of two risk factors from a bivariate Student-t distribution $t_\alpha(\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{\mu} = (1, 2)^T$ and Σ is $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The $\mathrm{DR}_{\mathbf{w},q}$ of such portfolios for various values of q against the weight w_1 are plotted for different pairs of (α, ρ) with $\alpha = 2, 4$ and $\rho = 0.3, 0.7$ in the four subfigures. The level of q is set to 0.95, 0.99, 0.999 and 0.9999. For each q level, the optimal portfolio weight on w_1 is indicated by a vertical line of different style. The optimal solution for $\mathrm{DR}_{\mathbf{w},1}$ is indicated by a thick vertical line.

marginal tails reflected in a lower α, the faster the convergence rate. This is in line with our finding in Theorem 3.3.1 α plays a role in the speed of convergence, the higher the α, the slower the speed of convergence. Lastly, when fixing the level of heavy–tailedness (α), the more dependence reflected in a higher ρ, the slower the convergence rate in the limit relation $\mathbf{w}_q \to \mathbf{w}_1$. Nevertheless, the slow convergence is not of a concern in practice. With a strong dependence at the first place, the room for diversification benefit is limited. As a result, the diversification ratio is in general at a high level and is less sensitive to the variation of the weights. Therefore, with a strong dependence, although the solution in the limit $(0.5, 0.5)^T$ might not be close to the optimal solution at a finite q, investing in the portfolio $(0.5, 0.5)^T$ would not result in a large increase in diversification ratio at a finite q level, compared to the actual optimal portfolio.

Next, we study a different numerical example based on a non-elliptical distribution. We construct the example using linear combinations of heavy-tailed random variables. Let Y_1 and Y_2 be two i.i.d. random variables with regularly varying tails. A random vector $\mathbf{X} = (X_1, X_2)^T$ is then defined as

$$\mathbf{X} = A\mathbf{Y}, \quad A := \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}, \tag{3.4.1}$$

where $\rho \in (-1, 1)$. Such random vector follows a non-elliptical distribution. In the case that the variance of Y_1 and Y_2 exists, ρ is the correlation coefficient between X_1 and X_2 Under this structure, the diversification ratio $\mathrm{DR}_{\mathbf{w},1}$ can be explicitly calculated. Following Mainik and Embrechts (2013), we have that

$$\frac{\eta_{\mathbf{w}}}{\eta_{\mathbf{e}_1}} = (w_1 + w_2\rho)^\alpha + \left(w_2\sqrt{1-\rho^2}\right)^\alpha,$$

and

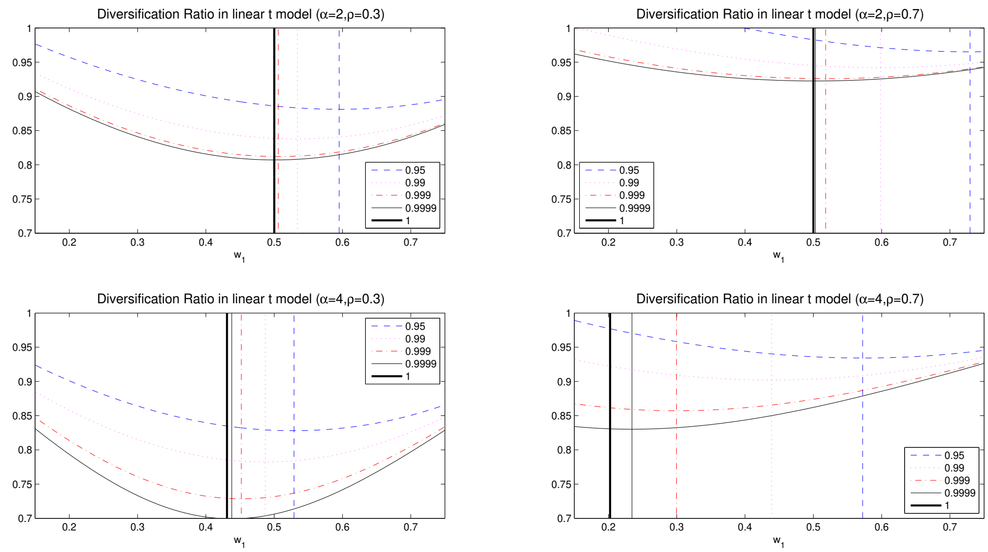
$$\frac{\eta_{\mathbf{w}}}{\eta_{\mathbf{e}_2}} = \frac{(w_1 + w_2\rho)^\alpha + \left(w_2\sqrt{1-\rho^2}\right)^\alpha}{\rho^\alpha + \sqrt{1-\rho^2}^\alpha}.$$

Hence,

$$\mathrm{DR}_{\mathbf{w},1} = \left(w_1 \left((w_1 + w_2\rho)^\alpha + \left(w_2\sqrt{1-\rho^2}\right)^\alpha \right)^{-\frac{1}{\alpha}} + w_2 \left(\frac{(w_1 + w_2\rho)^\alpha + \left(w_2\sqrt{1-\rho^2}\right)^\alpha}{\rho^\alpha + \sqrt{1-\rho^2}^\alpha} \right)^{-\frac{1}{\alpha}} \right)^{-1}.$$

We use this formula to determine $\mathrm{DR}_{\mathbf{w},1}$. Since the expression for $\mathrm{DR}_{\mathbf{w},q}$ is less explicit, its calculation is based on simulations.

Figure 3.2: Optimal portfolio from non–elliptical distribution risk factors



Note: The portfolios are constructed as a linear combination of two risk factors from a vector $\mathbf{X}$ defined in (3.4.1) with Y_1 and Y_2 following a standard Student-t distribution with degree of freedom $\alpha > 1$. The $\mathrm{DR}_{\mathbf{w},q}$ of such portfolios for various values of q against the weight w_1 are plotted for different pairs of (α, ρ) with $\alpha = 2, 4$ and $\rho = 0.3, 0.7$ in the four subfigures. The level of q is set to 0.95, 0.99, 0.999 and 0.9999. For each q level, the optimal portfolio weight on w_1 is indicated by a vertical line of different style. The optimal solution for $\mathrm{DR}_{\mathbf{w},1}$ is indicated by a thick vertical line.

Consider a special case where Y_1 and Y_2 follow a standard Student-t distribution with degree of freedom $\alpha > 1$. By choosing $\alpha = 2, 4$ and $\rho = 0.3, 0.7$, in Figure 3.2 we plot the calculated diversification ratios $\mathrm{DR}_{\mathbf{w},q}$ against the loading on X_1, w_1 for various values of q: 0.95, 0.99, 0.999 and 0.9999. The optimal weight for each q level is again marked by a corresponding vertical line, with thick vertical line indicating the optimal weight for the limit case $q = 1$.

All four observations in the elliptical case remain qualitatively valid for the non–elliptical case. Quantitatively, the distance between the optimal solutions for finite q and the limit case can be far apart. For example, in the worst case scenario when the lower tail index meets the stronger dependence (right bottom subfigure), the distance between the optimal weight for $q = 0.99$ and that for $q = 1$ is around 0.25. In this case, the optimal portfolio in the limit is not a good approximation for that based on a finite q. To summarize, we recommend using the optimal portfolio based on the limit diversification ratio particularly for the case with low cross–sectional dependence and heavy marginal tails.

3.5 Empirical Study

In the numerical examples, the limit diversification ratio $\mathrm{DR}_{\mathbf{w},1}$ can be calculated explicitly. With real data application, we need to estimate this function using historical data, and then consider the optimal portfolio based on the estimated diversification ratio. In Section 3.2.4, we discuss the estimation methodology for $\mathrm{DR}_{\mathbf{w},1}$. In this section, we apply our estimation method and the optimal portfolio construction procedure to real market data.

The dataset consists of underlying stocks in the S&P 500 index that have a full trading history throughout the period from January 2, 2002 to December 31, 2015. This results in 425 stocks. We construct the continuously compounded loss returns of these stocks. That is, if the price of asset i at time t is denoted by $P_i(t)$, then the log loss at time t for asset i, denoted by $X_i(t)$ is given by

$$X_i(t) = -\log\left(\frac{P_i(t)}{P_i(t-1)}\right).$$

We conduct three empirical studies. Firstly, we demonstrate the difference between the optimal portfolio constructed based on minimizing a diversification ratio at a finite q level and that based on minimizing the limit diversification ratio. Secondly, we show that our proposed methodology has the advantage of bearing less computational burden. Lastly, we evaluate the out–of–sample performance between our portfolio optimization procedure and those existing in the literature.

The first empirical study is set up as follows. To avoid dimensional curse in the numerical search strategy (see below), we select 10 stocks from the dataset that share a similar level of tail index. Notice that having the same marginal tail index is a necessary condition for MRV. We estimate the tail indices of the 425 stocks using the Hill estimator (Hill (1975)) as

$$\widehat{\alpha} = \frac{k}{\sum_{j=1}^{k} \log\left(R_{(n-j+1)}/R_{(n-k)}\right)}.$$

We select 10 stocks with the lowest estimates that are not significantly different from each other. Here, to test whether the 10 stocks have significantly different tail indices, we employ the test constructed in Moore et al. (2013) for testing tail index equivalence. In other words, we select 10 stocks with the lowest estimates while not being rejected by this test. The reason for selecting stocks with lower α follows from the numerical example: the approximation works better when α is lower. The selected stocks are given in Table 3.1, where the estimate of α and its standard deviation (std) for each stock are provided. From Table 3.1, we observe that the point estimates of the tail index range from 1.989 to 2.040.

Table 3.1: Tail index estimates for the 10 selected stocks

Stock	C	FRT	HST	LM	L	RF	TMK	VTR	VNO	XEL
$\hat{\alpha}$	1.989	2.000	2.002	2.007	2.012	2.014	2.019	2.036	2.036	2.040
std	0.168	0.169	0.169	0.170	0.170	0.170	0.171	0.172	0.172	0.172

Note: The table shows the tail index estimates for 10 selected stocks within the S&P 500 index based on their daily returns in the period from January 2, 2002 to December 31, 2015. The tail indices are estimated using the Hill estimator (Hill (1975)). The second row reports the standard deviations of the estimates.

Our empirical analysis is based on daily data in each five-year window, namely, 2002–2006, 2003–2007, etc. Within each window, for a given q level, we first construct the optimal portfolio that minimizes $\mathrm{DR}_{\mathbf{w},q}$ by a numerical search. This is achieved by assigning weights to the 10 stocks on a grid spanning the set Σ^{10}, evaluating $\mathrm{DR}_{\mathbf{w},q}$ at each grid point and taking the weights that corresponds to the minimum diversification ratio. Then we construct the optimal portfolio based on minimizing the estimated $\mathrm{DR}_{\mathbf{w},1}$ using the procedure laid out in Section 3.2.4.

The numerical search strategy gives a numerical optimal while our portfolio optimization strategy gives an approximation to that. To evaluate the difference between the two optimal portfolios, we use $||\mathbf{w}_q - \mathbf{w}^*||_1/10$. This distance indicates the average error made on the weight for one stock. We conduct this analysis for nine different windows and four different levels of q: 0.95, 0.975, 0.99 and 0.999.

In the estimation procedure, we need to select the intermediate sequence k. It should be chosen by balancing the bias and variance of the estimation. Here, we choose k to be 4% for estimating α and 10% for estimating the spectral measure $\widehat{\Psi}$. Moreover, since we only consider the loss, the estimator for $\eta_{\mathbf{w}}$ is slightly modified to

$$\widehat{\eta}_{\mathbf{w}} = \frac{1}{k} \sum \left(\mathbf{w}^T S_{\pi(j)}\right)_+^{\widehat{\alpha}} . \tag{3.5.1}$$

Table 3.2 shows the results on the error made using our optimization procedure. We observe that the distance is decreasing as q increases. This is in line with our theoretical result.

Table 3.2: Average error made on the weight for each stock

q	02-06	03-07	04-08	05-09	06-10	07-11	08-12	09-13	10-14
95%	0.1348	0.1091	0.125	0.0673	0.0868	0.0967	0.1447	0.1426	0.0941
97.50%	0.0838	0.0978	0.0967	0.0638	0.0795	0.0663	0.0985	0.0668	0.0802
99%	0.0837	0.0861	0.0858	0.0573	0.0636	0.0476	0.0834	0.0642	0.0731
99.9%	0.0442	0.0582	0.0688	0.0444	0.0397	0.0435	0.0435	0.0538	0.044

Note: Within in each five-year window, for a given q level, two portfolios are constructed. The numerical search strategy provides the first optimal portfolio that minimizes $\mathrm{DR}_{\mathbf{w},q}$. This is achieved by assigning weights to the 10 stocks on a grid spanning the set Σ^{10}, evaluating $\mathrm{DR}_{\mathbf{w},q}$ at each grid point and taking the weights that corresponds to the minimum diversification ratio. The second optimal portfolio minimizes the estimated $\mathrm{DR}_{\mathbf{w},1}$ using the procedure laid out in Section 3.2.4. The numbers reported are the distance calculated by $||\mathbf{w}_q - \mathbf{w}^*||_1/10$ between the two portfolios.

Next, we turn to analyzing the computation time for obtaining the optimal portfolio. For this analysis, we use only data in the most recent six windows and only consider $q = 0.95$. To show that the computational burden for the numerical search strategy largely depends on the number of stocks, we also perform the numerical search when using less stocks, namely the first 3, 5, and 8 stocks in Table 3.1. In contrast, we perform our portfolio optimization strategy always based on 10 stocks. The computation time of all the experiments run in Matlab 2013a on a Thinkpad T430 (dual core, 2.6GHz CPU, 4GB of memory) computer is reported in Table 3.3. We observe that as the number of stocks increasing, the computation time for $\mathbf{w}_{95\%}$ increases significantly. On the contrary, our portfolio optimization strategy for 10 stocks takes even less time than that using the numerical search for 3 stocks.

Finally, we perform an out–of–sample analysis comparing our portfolio optimization strategy with those in the literature. Within each five-year window, we perform our strategy to construct the optimal portfolio based on the 10 selected stocks in Table 3.1. Then we hold this portfolio for one year, and calculate the diversification ratio at 95% and the 95% VaR using the one-year out–of–sample data. We focus on $q = 95\%$ here because one-year loss data (roughly 250 daily observations) do not permit an accurate estimation of tail risk measures with a higher probability level. With a similar setup, we also apply the numerical search strategy laid out in the first empirical study which minimizes the $\mathrm{DR}_{\mathbf{w},95\%}$ within each five-year window, and evaluates the out–of–sample performance of this strategy. In addition, we apply four other strategies as competitors for out–of–sample performance,

Table 3.3: Computation time

Strategy		05-09	06-10	07-11	08-12	09-13	10-14
Numerical search	3 Stocks	0.350s	0.310s	0.261s	0.249s	0.231s	0.235s
Numerical search	5 Stocks	0.483s	0.402s	0.417s	0.391s	0.570s	0.612s
Numerical search	8 Stocks	1.226s	1.265s	1.594s	0.861s	1.463s	1.397s
Numerical search	10 Stocks	2.418s	2.799s	3.673s	2.022s	2.016s	2.383s
Minimizing $\mathrm{DR}_{\mathbf{w},1}$	10 Stocks	0.218s	0.189s	0.164s	0.175s	0.304s	0.166s

Note: Within each five-year window, the numerical search strategy is performed for minimizing the DR with $q = 0.95$ based on 3, 5, 8 and 10 stocks. The computation time are reported in the first four rows. The last row reports the computation time when performing the portfolio optimization strategy minimizing $\mathrm{DR}_{\mathbf{w},1}$ based on 10 stocks.

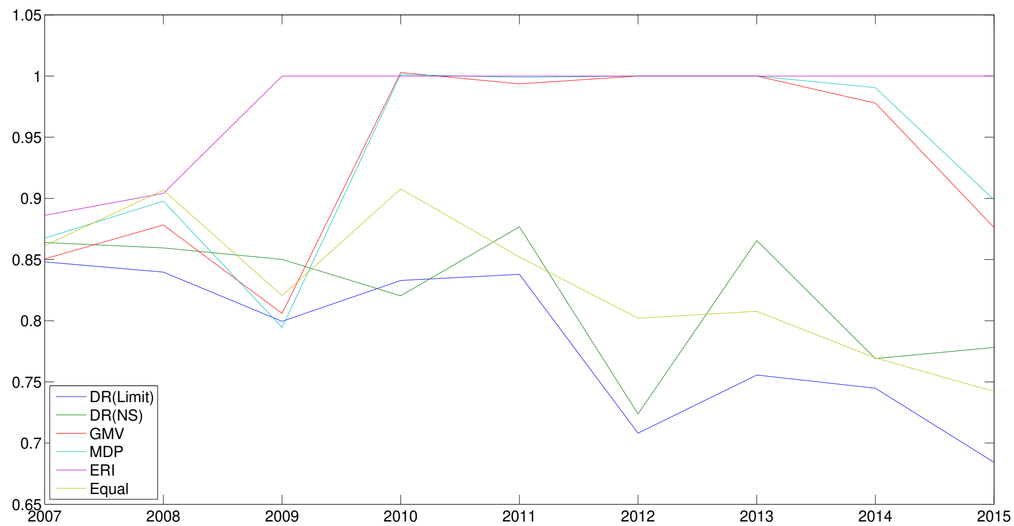
namely, the ERI, the MDP, global minimum variance (see, e.g. Merton (1972)), and lastly a simple equal weight strategy.

Figure 3.3 shows the results on the out–of–sample diversification ratios. Our strategy produces consistently the lowest diversification ratio only except in 2009, where our strategy yields a diversification ratio slightly above that derived from the MDP, and in 2010 slightly higher than that derived from the numerical research strategy. To achieve the tail diversification benefit measured by the diversification ratio, our portfolio optimization strategy gives the best out–of–sample performance.

Figure 3.4 shows the results on the out–of–sample VaR. Our portfolio optimization strategy produces the lowest VaR in 2007 and 2008, but not in the other years. Nevertheless, the VaR of the optimal portfolio from our strategy is never largely above ERI, which minimizes VaR among the six strategies. Furthermore, it matters the most to get an optimal portfolio with the lowest risk in the period ahead of the crisis. Therefore, we conclude that our strategy also gives the best out–of–sample performance in terms of risk management.

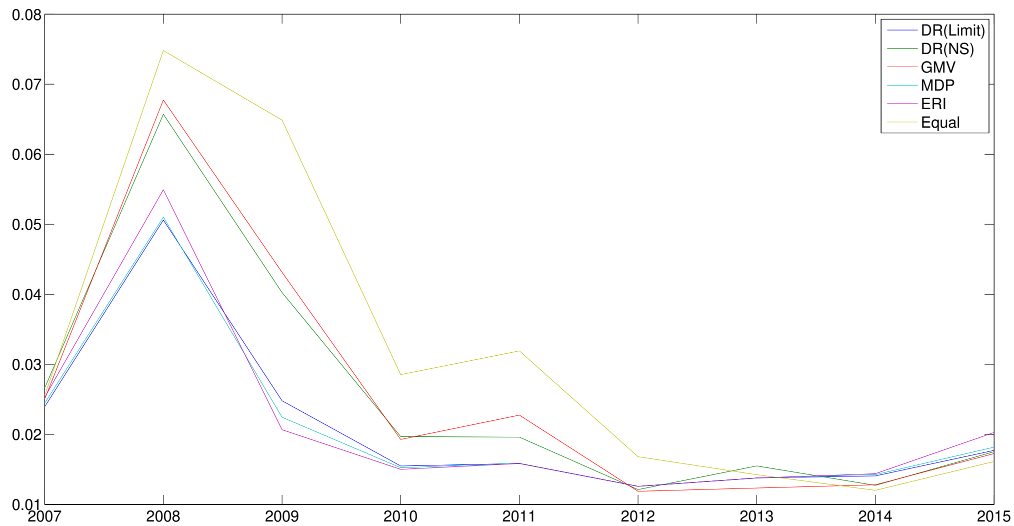
From all three empirical studies, we conclude that the computation burden of our portfolio optimization strategy is much lower than the numerical search. Although there is a moderate distance between the optimal portfolios obtained from our limit DR optimization strategy and the numerical search strategy, it turns out in the out–of–sample analysis that our strategy outperforms. It is therefore worth bearing the errors on the weights while using the fast and better performed algorithm derived from our limit DR optimization strategy.

Figure 3.3: Out–of–sample diversification ratio



Note: Within each five-year window, the optimal portfolio based on the 10 selected stocks in Table 3.1 is constructed by minimizing $\mathrm{DR}_{\mathbf{w},1}$. These weights are held for one year. The diversification ratio at 95% is reported using the one-year out–of–sample data and named as DR(Limit) in the figure. The same steps are repeated for five other strategies, the numerical search strategy for minimizing $\mathrm{DR}_{\mathbf{w},95\%}$ (DR(NS)), global minimum variance (GMV; see, e.g. Merton (1972)), the MDP, the ERI, and equal weight strategy (Equal).

Figure 3.4: Comparison of portfolio risks



Note: Within each five-year window, the optimal portfolio based on the 10 selected stocks in Table 3.1 is constructed by minimizing $\mathrm{DR}_{\mathbf{w},1}$. These weights are held for one year. The 95% VaR is reported using the one-year out–of–sample data and named as DR(Limit) in the figure. The same steps are repeated for five other strategies, the numerical search strategy for minimizing $\mathrm{DR}_{\mathbf{w},95\%}$ (DR(NS)), global minimum variance (GMV; see, e.g. Merton (1972)), the MDP, the ERI, and equal weight strategy (Equal).

3.6 Backtesting Study

Our dataset covers S&P 500 market index and its corresponding components, ranging from January 2, 2002 to July 20, 2015, which has a history of 3410 days. Here, we exclude certain stocks with missing data and we split 3410 days into two part. The first part serves as history, which has 1500 days. The second part is regarded as backtest period, containing 1910 days. Following the similar idea in Mainik et al. (2015), the estimated optimal portfolio is based on moving window of 1500 historical data.

Specifically, for every trading day $t > 1500$, we use 1500 data points in the historical observation window $t - 1500, \ldots, t - 1$ to estimate tail index α and spectral measure Ψ. From these estimates, we then calculate the estimator for DR based on steps provided in Section 3.2.4 and search for the optimal portfolio by minimizing the DR estimator. Finally, the estimated optimal portfolio is used to compose the portfolio for the trading day t. Such a procedure is repeated for all trading days $t > 1500$, so that the portfolio is rebalanced daily.

Following the procedure outlined above, the numerical result is shown in Figure 3.5. For benchmarking, the graph provides two additional curve. One corresponds to the S&P 500 raw index price and the other is the portfolio price obtained based on the global minimum variance (GMV) strategy (strategy that selects the portfolio with minimum risk, i.e., minimal variance). The GMV strategy is implemented similarly as the DR portfolio strategy involving rebalancing and over moving window of 1500 historical data. Over the backtest period, it can be seen that the DR strategy tends to stabilize the price, especially when price jumps with a big magnitude. Note that DR method takes into consideration tail risk and hence is a strategy that is particularly suited when the loss distribution is heavy-tailed and big jump of price exists over a certain period, DR strategy will smooth the price and make it more stable. Additionally, it should be noted that when price tends to increase, at a certain point, the GMV strategy will outperform the DR strategy due to the fact that only downside risk is considered under the DR strategy. However, in terms of stabilizing prices, GMV strategy does not perform better than DR strategy as confirmed in Table 3.4 which reports the cumulative returns and standard deviations for S&P 500 and portfolio strategies based on DR and GMV.

Firstly, it can be seen that the standard deviation of the DR is the smallest for the whole backtest period, which agrees with our anticipation and the model. Secondly, during 09/30/2011-07/20/2015 when the overall price tends to go upward, the cumulative return under the GMV strategy is better than that under the DR method, whereas during 12/17/2007-09/29/2011 it is the opposite. It is known that financial crisis does incur severe loss to the market. However, since tail risk is the main focus of the DR method, the

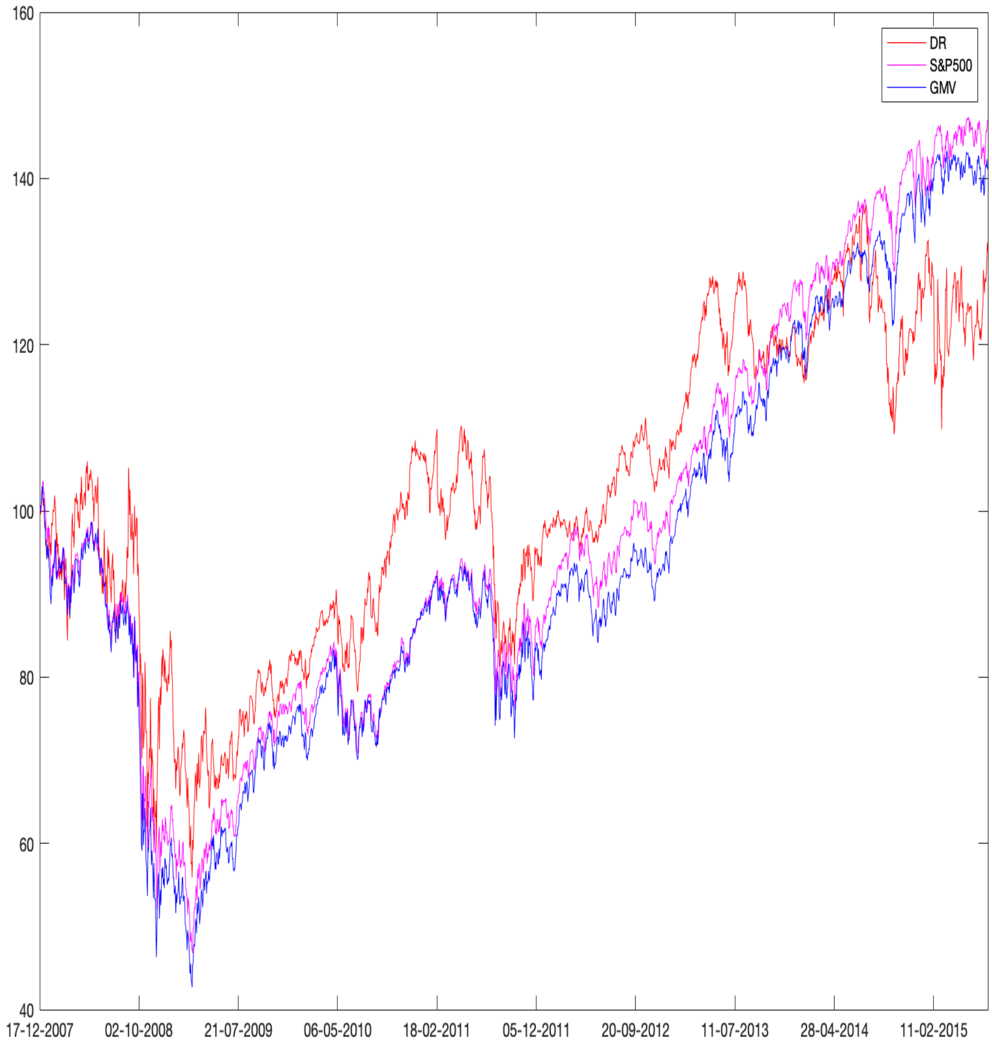


Figure 3.5: Portfolio optimization backtest for the DR minimization strategy and global minimal variance strategy. The resulting portfolio value is scaled to 100 for the first date of the backtest period.

Table 3.4: Cumulative Return and Standard Deviation

Cumulative Return	DR	GMV	S&P 500
12/17/2007-09/29/2011	-0.2132	-0.2262	-0.2453
09/30/2011-07/20/2015	0.4535	0.6018	0.6311
Standard Deviation	**DR**	**GMV**	**S&P 500**
12/17/2007-07/20/2015	18.3764	24.4042	25.0204

potential loss by DR approach should be slightly lower during this period, which accords with our previous finding.

3.7 Proofs

In this section, we first prove Theorem 3.2.1, which is the key and the most difficult part in the proof of Theorem 3.2.2, in two steps as Sections 3.7.1 and 3.7.2. Then the very last section contains all the proofs of lemmas from previous sections.

3.7.1 Uniform Convergence in Radon Measures

Define a family of mappings from $A_1 = \left\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_1 > 1\right\}$ to $\mathbb{R}_+$ as

$$M = \left\{ f_{\mathbf{w}}(\mathbf{x}) = \frac{1}{1 + (\mathbf{w}^T\mathbf{x})_+} : \mathbf{w} \in \Sigma^d,\, \mathbf{x} \in A_1 \right\}. \tag{3.7.1}$$

Note that the construction of the mappings in M is not unique. Let $A_{\mathbf{w},1}$ denotes the events where the portfolio loss $\mathbf{w}^T\mathbf{X}$ exceeds 1, namely for $\mathbf{w} \in \Sigma^d$,

$$A_{\mathbf{w},1} = \left\{\mathbf{x} \in \mathbb{R}^d : \mathbf{w}^T\mathbf{x} > 1\right\}.$$

Theorem 3.7.1 *If* $\mathbf{X} \in \mathrm{MRV}_{-\alpha}(\Psi)$ *with* $\alpha > 0$, *then*

$$\lim_{t\to\infty} \sup_{\mathbf{w}\in\Sigma^d} \left|\nu_t\left(A_{\mathbf{w},1}\right) - \nu\left(A_{\mathbf{w},1}\right)\right| = 0, \tag{3.7.2}$$

where ν_t *and* ν *are defined in (2.3.5).*

Proof. Since $A_{\mathbf{w},1} \in \mathcal{B}(A_1)$, by (2.3.5) we have that $\nu_t(A_{\mathbf{w},1})$ converges weakly to $\nu(A_{\mathbf{w},1})$. To further show the uniform convergence, we apply Theorem 3.4 of Rao (1962). That is we need to verify the following three conditions. (1) The mappings in M defined in (3.7.1) are continuous mappings from a separable metric space to $\mathbb{R}_+$. (2) The family M is relative compact; that is every sequence in M on a compact subset of A_1 has a subsequence that converges uniformly. (3) For each $f_{\mathbf{w}} \in M$, $vf_{\mathbf{w}}^{-1}$ has a continuous distribution. Next, we prove them separately.

(1) By Theorem 1.5 of Lindskog (2004), there exists a metric $\overline{\rho}$ such that $(A_1, \overline{\rho})$ is a locally compact, complete and separable metric space. It is easy to see that each $f_{\mathbf{w}} \in M$ is continuous.

(2) Note that for $\mathbf{x}, \mathbf{y} \in A_1$, we have $\left(\mathbf{w}^T\mathbf{x}\right)_+, \left(\mathbf{w}^T\mathbf{y}\right)_+ > 0$. Then, by Cauchy–Schwarz inequality,

$$|f_{\mathbf{w}}(\mathbf{x}) - f_{\mathbf{w}}(\mathbf{y})| \leq \left|\mathbf{w}^T(\mathbf{x} - \mathbf{y})\right| \leq \sqrt{d}\, \|\mathbf{x} - \mathbf{y}\|_2 .$$

For arbitrary $\varepsilon > 0$, we can choose $\delta < \varepsilon/\sqrt{d}$, which is independent of f, $\mathbf{x}$ and $\mathbf{y}$, such that when $\|\mathbf{x} - \mathbf{y}\|_2 < \delta$, we have $|f_{\mathbf{w}}(\mathbf{x}) - f_{\mathbf{w}}(\mathbf{y})| < \varepsilon$. This shows that M is equicontinuous at each $\mathbf{x} \in A_1$. Moreover, M is uniformly bounded as for each $\mathbf{x} \in A_1$,

$$\sup_{f_{\mathbf{w}} \in M} \{f_{\mathbf{w}}(\mathbf{x})\} = \sup_{\mathbf{w} \in \Sigma^d} \left\{\frac{1}{1 + \left(\mathbf{w}^T\mathbf{x}\right)_+}\right\} \leq 1.$$

Therefore, from the Arzelà-Ascoli theorem, we know M is relatively compact.

(3) For $x \in \mathbb{R}_+$, we have

$$\begin{aligned} vf^{-1}((0,x)) &= \int_{\mathcal{S}^{d-1}} \int_{\mathbb{R}_+} 1_{\{\mathbf{w}^T\mathbf{s}>0\}} 1_{\left\{r > \left(\frac{1}{x} - 1\right)/\mathbf{w}^T\mathbf{s}\right\}} \rho_\alpha(dr)\Psi(d\mathbf{s}) \\ &= \left(\frac{1}{x} - 1\right)^{-\alpha} \int_{\mathcal{S}^{d-1}} \left(\mathbf{w}^T\mathbf{s}\right)_+^\alpha \Psi(d\mathbf{s}), \end{aligned}$$

which is obviously continuous for any $0 < x < 1$. Furthermore, we have $\nu(A_1) = 1$.

By far, we have verified the three conditions. By the weak convergence in (2.3.5) and Theorem 3.4 of Rao (1962), we obtain

$$\lim_{t \to \infty} \sup_{\mathbf{w} \in \Sigma^d} |\nu_t(A_{\mathbf{w},1}) - \nu(A_{\mathbf{w},1})| = 0,$$

where the supremum is taken over all sets $A_{\mathbf{w},1}$ of the form $A_{\mathbf{w},1} = \left\{\mathbf{x} \in \mathbb{R}^d : f_{\mathbf{w}}(\mathbf{x}) < \frac{1}{2}\right\} = \left\{\mathbf{x} \in \mathbb{R}^d : \mathbf{w}^T\mathbf{x} > 1\right\}$ with $\mathbf{w} \in \Sigma^d$. ∎

Next corollary is a natural rewriting of relation (3.7.2). It yields a uniform convergence of the ratio $\mathbb{P}\left(\mathbf{w}^T\mathbf{X} > t\right)/\mathbb{P}\left(\|\mathbf{X}\|_1 > t\right)$ to $\eta_{\mathbf{w}}$. However, only the weak convergence of it is known in the literature.

Corollary 3.7.1

$$\lim_{t\to\infty} \sup_{\mathbf{w}\in\Sigma^d} \left| \frac{\mathbb{P}\left(\mathbf{w}^T\mathbf{X} > t\right)}{\mathbb{P}\left(\|\mathbf{X}\|_1 > t\right)} - \eta_{\mathbf{w}} \right| = 0, \tag{3.7.3}$$

where

$$\eta_{\mathbf{w}} = \int_{\mathcal{S}^{d-1}} (\mathbf{w}^T\mathbf{s})_+^{\alpha}\,\Psi(d\mathbf{s}).$$

Further, the mapping $\mathbf{w} \mapsto \eta_{\mathbf{w}}$ *is uniform continuous.*

Proof. First note that $A_{\mathbf{w},t} = tA_{\mathbf{w},1}$. Since $A_{\mathbf{w},1} \subset \mathcal{B}\left(A_1\right)$ for $\mathbf{w} \in \Sigma^d$, we have that

$$\nu_t\left(A_{\mathbf{w},1}\right) = \frac{\mathbb{P}\left(\frac{\mathbf{X}}{t} \in A_{\mathbf{w},1}\right)}{\mathbb{P}\left(\|\mathbf{X}\|_1 > t\right)} = \frac{\mathbb{P}\left(\mathbf{X} \in A_{\mathbf{w},t}\right)}{\mathbb{P}\left(\|\mathbf{X}\|_1 > t\right)}.$$

Moreover $\nu\left(A_{\mathbf{w},1}\right)$ is actually

$$\nu\left(A_{\mathbf{w},1}\right) = \int_{\mathcal{S}^{d-1}} (\mathbf{w}^T\mathbf{s})_+^{\alpha}\,\Psi(d\mathbf{s}) = \eta_{\mathbf{w}}.$$

The desired result (3.7.3) then follows. Lastly, since $\eta_{\mathbf{w}}$ is continuous on the compact set Σ^d, it implies the uniform continuity of $\eta_{\mathbf{w}}$. ■

3.7.2 Uniform Convergence in Quantiles

In order to show that the convergence in (3.2.4) is indeed uniform, we first prepare a key lemma. For notational simplicity, we denote

$$l(\mathbf{w}, q) := \frac{\mathrm{VaR}_q(\mathbf{w}^T\mathbf{X})}{\mathrm{VaR}_q(\|\mathbf{X}\|_1)} = \frac{F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q)}{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q)}, \tag{3.7.4}$$

where $F_{\mathbf{w}^T\mathbf{X}}$ is the distribution function of $\mathbf{w}^T\mathbf{X}$ and $F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q) = \mathrm{VaR}_q(\mathbf{w}^T\mathbf{X})$.

Lemma 3.7.1 *Suppose the random vector* $\mathbf{X}$ *is continuously distributed with a positive joint density function. Further assume that* $\mathbf{X} \in \mathrm{MRV}_{-\alpha}(\Psi)$ *with* $\alpha > 0$*. Given* $\mathbf{w} \in \Sigma^d$*, for any* $\varepsilon > 0$ *there exist* $0 < \tilde{q} < 1$ *and* δ *such that for all* $\tilde{q} < q < 1$ *and* $\mathbf{z} \in \Sigma^d$ *satisfying* $\|\mathbf{w} - \mathbf{z}\| < \delta$*, we have*

$$|l(\mathbf{w}, q) - l(\mathbf{z}, q)| < \varepsilon. \tag{3.7.5}$$

Proof. We start by showing that for any $\varepsilon_1 > 0$, there exist $t_0(\varepsilon_1)$ and $\delta(\varepsilon_1)$ such that for all $t > t_0$ and all $\mathbf{w}, \mathbf{z} \in \Sigma^d$ with $\|\mathbf{w} - \mathbf{z}\| < \delta$, we have

$$\left|\overline{F}_{\mathbf{w}^T\mathbf{X}}(t) - \overline{F}_{\mathbf{z}^T\mathbf{X}}(t)\right| < \varepsilon_1 \overline{F}_{\mathbf{w}^T\mathbf{X}}(t). \tag{3.7.6}$$

Note that $\eta_{\mathbf{w}} > 0$ for every $\mathbf{w} \in \Sigma^d$. Since Σ^d is compact, there exists $\underline{\eta} > 0$ such that $\eta_{\mathbf{w}} > \underline{\eta} > 0$. Further, $\eta_{\mathbf{w}}$ is uniform continuous by Corollary 3.7.1. That is, for any $\varepsilon_1 > 0$, there exists $\delta(\varepsilon_1)$ such that for all $\mathbf{w}, \mathbf{z} \in \Sigma^d$ with $\|\mathbf{w} - \mathbf{z}\| < \delta$, we have

$$|\eta_{\mathbf{w}} - \eta_{\mathbf{z}}| < \frac{\underline{\eta}}{6}\varepsilon_1. \tag{3.7.7}$$

Again, by Corollary 3.7.1, there exists $t_0(\varepsilon_1)$ such that for all $t > t_0$ and all $\mathbf{w} \in \Sigma^d$

$$\left|\frac{\overline{F}_{\mathbf{w}^T\mathbf{X}}(t)}{\overline{F}_{\|\mathbf{X}\|_1}(t)} - \eta_{\mathbf{w}}\right| < \frac{\underline{\eta}}{6}\varepsilon_1 \wedge \frac{\underline{\eta}}{2}, \tag{3.7.8}$$

which implies that

$$\frac{\overline{F}_{\mathbf{w}^T\mathbf{X}}(t)}{\overline{F}_{\|\mathbf{X}\|_1}(t)} > \eta_{\mathbf{w}} - \frac{\underline{\eta}}{2} > \frac{\underline{\eta}}{2}. \tag{3.7.9}$$

Then, combining (3.7.7), (3.7.8) and (3.7.9), for all $t > t_0$ and all $\mathbf{w}, \mathbf{z} \in \Sigma^d$ with $\|\mathbf{w} - \mathbf{z}\| < \delta$,

$$\begin{aligned}
\left|\frac{\overline{F}_{\mathbf{w}^T\mathbf{X}}(t) - \overline{F}_{\mathbf{z}^T\mathbf{X}}(t)}{\overline{F}_{\mathbf{w}^T\mathbf{X}}(t)}\right| &= \left|\frac{\overline{F}_{\mathbf{w}^T\mathbf{X}}(t) - \overline{F}_{\mathbf{z}^T\mathbf{X}}(t)}{\overline{F}_{\|\mathbf{X}\|_1}(t)}\right| \cdot \frac{\overline{F}_{\|\mathbf{X}\|_1}(t)}{\overline{F}_{\mathbf{w}^T\mathbf{X}}(t)} \\
&\le \left(\left|\frac{\overline{F}_{\mathbf{w}^T\mathbf{X}}(t)}{\overline{F}_{\|\mathbf{X}\|_1}(t)} - \eta_{\mathbf{w}}\right| + |\eta_{\mathbf{w}} - \eta_{\mathbf{z}}| + \left|\eta_{\mathbf{z}} - \frac{\overline{F}_{\mathbf{z}^T\mathbf{X}}(t)}{\overline{F}_{\|\mathbf{X}\|_1}(t)}\right|\right) \cdot \frac{\overline{F}_{\|\mathbf{X}\|_1}(t)}{\overline{F}_{\mathbf{w}^T\mathbf{X}}(t)} \\
&< \left(\frac{\underline{\eta}}{6}\varepsilon_1 + \frac{\underline{\eta}}{6}\varepsilon_1 + \frac{\underline{\eta}}{6}\varepsilon_1\right)\frac{2}{\underline{\eta}} = \varepsilon_1,
\end{aligned}$$

which yields (3.7.6).

Next, for the chosen $t_0(\varepsilon_1)$, let us denote $q_0 = \sup_{\mathbf{z}\in\Sigma^d} F_{\mathbf{z}^T\mathbf{X}}(t_0(\varepsilon_1))$. Then for any $q_0 < q < 1$ and all $\mathbf{z} \in\Sigma^d$, we have

$$F^{\leftarrow}_{\mathbf{z}^T\mathbf{X}}(q) \geq F^{\leftarrow}_{\mathbf{z}^T\mathbf{X}}(q_0) \geq t_0. \tag{3.7.10}$$

By (3.7.6) and (3.7.10), it leads to that for all $q > q_0$ and $\|\mathbf{w} - \mathbf{z}\| < \delta$,

$$\left|\overline{F}_{\mathbf{w}^T\mathbf{X}}(F^{\leftarrow}_{\mathbf{z}^T\mathbf{X}}(q)) - (1-q)\right| < \varepsilon_1(1-q).$$

By the monotonicity of $F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q)$, we obtain

$$F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}\left(q(1+\varepsilon_1) - \varepsilon_1\right) < F^{\leftarrow}_{\mathbf{z}^T\mathbf{X}}(q) < F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}\left(q(1-\varepsilon_1) + \varepsilon_1\right). \tag{3.7.11}$$

Finally we handle $|l(\mathbf{w}, q) - l(\mathbf{z}, q)|$ in (3.7.5). We only discuss the upper bound of $l(\mathbf{w}, q) - l(\mathbf{z}, q)$ in this step as the lower bound can be derived in a similar way. By (3.7.11),

$$\begin{aligned}
&l(\mathbf{w}, q) - l(\mathbf{z}, q)\\
\leq&\frac{F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q)}{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q)} - \frac{F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}\left(q(1+\varepsilon_1) - \varepsilon_1\right)}{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q)}\\
=&\left(\frac{F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q)}{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q)} - \frac{F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}\left(q(1+\varepsilon_1) - \varepsilon_1\right)}{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q(1+\varepsilon_1) - \varepsilon_1)}\right) + \frac{F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}\left(q(1+\varepsilon_1) - \varepsilon_1\right)}{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q(1+\varepsilon_1) - \varepsilon_1)}\left(1 - \frac{F^{\leftarrow}_{\|\mathbf{X}\|_1}\left(q(1+\varepsilon_1) - \varepsilon_1\right)}{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q)}\right)\\
:=&I_1 + I_2,
\end{aligned}$$

where

$$I_1 = l(\mathbf{w}, q) - l(\mathbf{w}, q(1+\varepsilon_1) - \varepsilon_1),$$

and

$$I_2 = l(\mathbf{w}, q(1+\varepsilon_1) - \varepsilon_1)\left(1 - \frac{F^{\leftarrow}_{\|\mathbf{X}\|_1}\left(q(1+\varepsilon_1) - \varepsilon_1\right)}{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q)}\right).$$

We show that $I_1 < \varepsilon/2$ and $I_2 < \varepsilon/2$.

For I_1, note the random vector $\mathbf{X}$ is continuously distributed with a positive joint density function. By using the change of variables, the density functions for random variables $\|\mathbf{X}\|_1$ and $\mathbf{w}^T\mathbf{X}$ can be shown to be positive as well, which implies that $F_{\|\mathbf{X}\|_1}(t)$ and $F_{\mathbf{w}^T\mathbf{X}}(t)$ are strictly increasing in t. By Proposition 1 (7) in Embrechts and Hofert (2013), we have that $F^{\leftarrow}_{\|\mathbf{X}\|_1}(q)$ and $F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q)$ are both continuous in q for any fixed $\mathbf{w}$. Moreover, from (3.2.4), $l(\mathbf{w}, 1)$ can be continuously defined as $\eta_{\mathbf{w}}^{1/\alpha}$. Thus, given $\mathbf{w}$, $l(\mathbf{w}, q)$

is uniformly continuous in q when $q \in [1/2, 1]$. That is, there exists $\lambda_1(\mathbf{w}, \varepsilon)$ such that when $1/2 \leq p, q \leq 1$ and $|p - q| < \lambda_1$, we have

$$|l(\mathbf{w}, q) - l(\mathbf{w}, p)| < \frac{\varepsilon}{2}.$$

Then, for $|q - (q(1 + \varepsilon_1) - \varepsilon_1)| < \lambda_1$ or $q > 1 - \lambda_1/\varepsilon_1$, we obtain that $I_1 < \varepsilon/2$.

For I_2, we first show that $l(\mathbf{w}, q(1+\varepsilon_1)-\varepsilon_1)$ is bounded. Since $\lim_{q\to 1} l(\mathbf{w}, q) = l(\mathbf{w}, 1)$, there exists $\lambda_2(\mathbf{w})$ such that when $1-(q(1+\varepsilon_1)-\varepsilon_1) < \lambda_2$ or $1 > q > 1-\lambda_2/(1+\varepsilon_1)$, we have

$$|l(\mathbf{w}, q(1 + \varepsilon_1) - \varepsilon_1) - l(\mathbf{w}, 1)| < 1.$$

Denote $M_0 = \sup_{\mathbf{w}\in\Sigma^d} l(\mathbf{w}, 1)$. We obtain

$$l(\mathbf{w}, q(1 + \varepsilon_1) - \varepsilon_1) < M_0 + 1, \qquad \text{for } q > 1 - \lambda_2/(1 + \varepsilon_1). \tag{3.7.12}$$

Finally, we consider $1 - \frac{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q(1+\varepsilon_1)-\varepsilon_1)}{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q)}$ in term I_2. It is known that if $\mathbf{X} \in \mathrm{MRV}_{-\alpha}(\Psi)$ then $\|\mathbf{X}\|_1 \in \mathrm{RV}_{-\alpha}$; e.g. see Basrak et al. (2002). Thus,

$$\lim_{q\to 1} \frac{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q(1 + \varepsilon_1) - \varepsilon_1)}{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q)} = (1 + \varepsilon_1)^{1/\alpha}.$$

By Proposition B.1.10 of de Haan and Ferreira (2006), there exists $q_3(\varepsilon) < 1$ such that for all $q > q_3(\varepsilon)$ we have

$$\left| \frac{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q(1 + \varepsilon_1) - \varepsilon_1)}{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q)} - (1 + \varepsilon_1)^{1/\alpha} \right| < \frac{1}{M_0 + 1}\frac{\varepsilon}{4}.$$

Moreover, when ε_1 is so chosen that

$$\left|1 - (1 + \varepsilon_1)^{1/\alpha}\right| < \frac{1}{M_0 + 1}\frac{\varepsilon}{4}, \tag{3.7.13}$$

it leads to that

$$\left| \frac{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q(1 + \varepsilon_1) - \varepsilon_1)}{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q)} - 1 \right| < \frac{\varepsilon}{2\,(M_0 + 1)}, \qquad \text{for } q > q_3(\varepsilon). \tag{3.7.14}$$

Combining (3.7.12) and (3.7.14), $I_2 < \varepsilon/2$ for $q > 1 - \lambda_2/(1 + \varepsilon_1) \vee q_3(\varepsilon)$.

To sum up, given $\mathbf{w}$, for arbitrary $\varepsilon > 0$, and for any ε_1 so chosen that (3.7.13) holds, there exist δ, $q_0, \lambda_1, \lambda_2$, and q_3 such that for all $\mathbf{z} \in \Sigma^d$ with $\|\mathbf{w} - \mathbf{z}\| < \delta$ and for all q satisfying that

$$1 > q > q_0 \vee \left(1 - \frac{\lambda_1}{\varepsilon_1}\right) \vee \left(1 - \frac{\lambda_2}{1 + \varepsilon_1}\right) \vee q_3,$$

we have $l(\mathbf{w}, q) - l(\mathbf{z}, q) < \varepsilon$. The other side of the inequality can be derived similarly. ■

Now we are ready to show that the convergence in (3.2.4) is uniform.

Theorem 3.7.2 *Suppose the random vector* $\mathbf{X}$ *is continuously distributed with a positive joint density function. Further assume that* $\mathbf{X} \in \mathrm{MRV}_{-\alpha}(\Psi)$ *with* $\alpha > 0$. *Then*

$$\lim_{q \uparrow 1} \sup_{\mathbf{w} \in \Sigma^d} \left| \frac{\mathrm{VaR}_q(\mathbf{w}^T \mathbf{X})}{\mathrm{VaR}_q(\|\mathbf{X}\|_1)} - \eta_{\mathbf{w}}^{1/\alpha} \right| = 0. \tag{3.7.15}$$

Proof. Consider the decomposition for some $\mathbf{v} \in \Sigma^d$

$$\left| l(\mathbf{w}, q) - \eta_{\mathbf{w}}^{1/\alpha} \right| \le \left| l(\mathbf{w}, q) - l\left(\mathbf{v}, q\right) \right| + \left| l(\mathbf{v}, q) - \eta_{\mathbf{v}}^{1/\alpha} \right| + \left| \eta_{\mathbf{v}}^{1/\alpha} - \eta_{\mathbf{w}}^{1/\alpha} \right|, \tag{3.7.16}$$

where $l(\mathbf{w}, q)$ is defined as in (3.7.4). By properly choosing $\mathbf{v}$, if the three terms can be shown to be arbitrarily small for any $\mathbf{w} \in \Sigma^d$ as q close to 1, then (3.7.15) is proved. In the following we show how $\mathbf{v}$ can be determined.

By Lemma 3.7.1 and the uniform continuity of $\eta_{\mathbf{w}}$, for any $\varepsilon > 0$, there exist $\delta(\mathbf{w}) > 0$ and $0 < \tilde{q}(\mathbf{w}) < 1$ such that for any $\mathbf{w}, \mathbf{z} \in \Sigma^d$ satisfying $\|\mathbf{w} - \mathbf{z}\| < \delta(\mathbf{w})$ and all $q \ge \tilde{q}(\mathbf{w})$, we have

$$\left| l(\mathbf{w}, q) - l(\mathbf{z}, q) \right| < \varepsilon. \tag{3.7.17}$$

and

$$\left| \eta_{\mathbf{w}}^{1/\alpha} - \eta_{\mathbf{z}}^{1/\alpha} \right| < \varepsilon. \tag{3.7.18}$$

That is, $\delta(\mathbf{w})$ is so chosen that both (3.7.17) and (3.7.18) hold. Now we are ready to determine $\mathbf{v}$ in (3.7.16) by constructing open coverings. Let $B_{\mathbf{w},\delta(\mathbf{w})}$ denote the open ball of $\mathbf{w}$; that is $B_{\mathbf{w},\delta(\mathbf{w})} = \{\mathbf{z} \in \Sigma^d : \|\mathbf{w} - \mathbf{z}\| < \delta(\mathbf{w})\}$. Then the collection of all the sets $B_{\mathbf{w},\delta(\mathbf{w})}$ for each $\mathbf{w}$ is an open cover of Σ^d. By the compactness, there exists a finite subcover denoted by $B_{\mathbf{w}_1,\delta(\mathbf{w}_1)}, \ldots, B_{\mathbf{w}_m,\delta(\mathbf{w}_m)}$. For each selected $\mathbf{w}_i$, by the limit relation in (3.2.4), there exists $0 < q_i < 1$ such that

$$\left| l(\mathbf{w}_i, q) - \eta_{\mathbf{w}_i}^{1/\alpha} \right| < \varepsilon,$$

for all $q_i \leq q < 1$. Let $q^* = \max\{\tilde{q}(\mathbf{w}_1), \ldots, \tilde{q}(\mathbf{w}_m), q_1, \ldots, q_m\}$. For any $\mathbf{w} \in \Sigma^d$, one can find i such that $\mathbf{w} \in B_{\mathbf{w}_i,\delta(\mathbf{w}_i)}$, which means $\|\mathbf{w} - \mathbf{w}_i\| < \delta(\mathbf{w}_i)$. This $\mathbf{w}_i$ is the proper choice of $\mathbf{v}$ in (3.7.16) since each term on the right-hand side of (3.7.16) is smaller than ε for all $q^* \leq q < 1$. This completes the proof. ■

Now we are ready to prove Theorem 3.2.1.

Proof of Theorem 3.2.1. Since the convergence $\lim_{q\uparrow 1} \mathrm{VaR}_q(X_i)/\mathrm{VaR}_q(\|\mathbf{X}\|_1) = \eta_{e_i}^{1/\alpha}$ is independent of $\mathbf{w}$, applying Theorem 3.7.2 to the rewriting in (3.2.3) we obtain the desired result. ■

3.7.3 Proofs of Lemmas

Lastly, we present the proofs of lemmas from previous sections.

Proof of Lemma 3.2.2. To prove $\mathrm{DR}_{\mathbf{w},q} \xrightarrow{\text{unif}} \mathrm{DR}_{\mathbf{w},1}$, we need to show for any given $\varepsilon > 0$, there exists a number q_0 such that $|\mathrm{DR}_{\mathbf{w},q} - \mathrm{DR}_{\mathbf{w},1}| < \varepsilon$ for every $q > q_0$ and for every $\mathbf{w}$ in Σ^d. Note the rewriting

$$|\mathrm{DR}_{\mathbf{w},q} - \mathrm{DR}_{\mathbf{w},1}| = \left| \frac{\mathbf{w}^T\boldsymbol{\mu}\left(\sum_{i=1}^d w_i \left\|B^T\mathbf{e}_i\right\|_2 - \left\|B^T\mathbf{w}\right\|_2\right)}{\left(\mathbf{w}^T\boldsymbol{\mu} + \sum_{i=1}^d w_i \left\|B^T\mathbf{e}_i\right\|_2 F_Z^{\leftarrow}(q)\right)\sum_{i=1}^d w_i \left\|B^T\mathbf{e}_i\right\|_2} \right|.$$

For every $\mathbf{w} \in \Sigma^d$, since $\|\boldsymbol{\mu}\|_1 < \infty$, there exists $N_1 > 0$ such that $\mathbf{w}^T\boldsymbol{\mu} < \|\boldsymbol{\mu}\|_1 < N_1$. Let $\lambda_{max}, \lambda_{min}$ denote the largest and smallest eigenvalues of $C = BB^T$, then there exists $N_2, N_3 > 0$ such that

$$0 < \sum_{i=1}^d w_i \left\|e_i^T B\right\|_2 < \sum_{i=1}^d \left\|B^T\mathbf{e}_i\right\|_2 \leq d\sqrt{\lambda_{max}} < N_2,$$

and

$$\left\|B^T\mathbf{w}\right\|_2 \leq \sqrt{\lambda_{max}} < N_3.$$

Since Y is unbounded, there exists $0 < q_0 < 1$ such that

$$F_Z^{\leftarrow}(q) > \frac{N_1(N_2 + N_3)}{N_2^2\varepsilon} - \frac{N_1}{N_2},$$

for every $q > q_0$. Combining the above analysis, the desired result $|\mathrm{DR}_{\mathbf{w},q} - \mathrm{DR}_{\mathbf{w},1}| < \varepsilon$ for every $q > q_0$ and for every $\mathbf{w}$ in Σ^d follows.

Next, we show that $\mathrm{DR}_{\mathbf{w},1}$ is continuous. For $\mathbf{w}, \mathbf{v} \in \Sigma^d$, we have that

$$
\begin{aligned}
&\left|\mathrm{DR}_{\mathbf{w},1} - \mathrm{DR}_{\mathbf{v},1}\right| \\
\leq &\left| \frac{\left\|B^T(\mathbf{w}-\mathbf{v})\right\|_2 \sum_{i=1}^d v_i \left\|B^T\mathbf{e}_i\right\|_2 + \left\|B^T\mathbf{v}\right\|_2 \left(\sum_{i=1}^d v_i \left\|B^T\mathbf{e}_i\right\|_2 - \sum_{i=1}^d w_i \left\|B^T\mathbf{e}_i\right\|_2\right)}{\left(\sum_{i=1}^d w_i \left\|B^T\mathbf{e}_i\right\|_2\right)\left(\sum_{i=1}^d v_i \left\|B^T\mathbf{e}_i\right\|_2\right)} \right| \\
\leq &\frac{\sqrt{\lambda_{max}} \left\|\mathbf{w}-\mathbf{v}\right\|_1 \sum_{i=1}^d v_i \left\|B^T\mathbf{e}_i\right\|_2}{\left(\sum_{i=1}^d w_i \left\|B^T\mathbf{e}_i\right\|_2\right)\left(\sum_{i=1}^d v_i \left\|B^T\mathbf{e}_i\right\|_2\right)} + \frac{\left\|B^T\mathbf{v}\right\|_2 \left\|\mathbf{w}-\mathbf{v}\right\|_1 \max\limits_{1\leq i\leq d} \left\|B^T\mathbf{e}_i\right\|_2}{\left(\sum_{i=1}^d w_i \left\|B^T\mathbf{e}_i\right\|_2\right)\left(\sum_{i=1}^d v_i \left\|B^T\mathbf{e}_i\right\|_2\right)} \\
= &\left\|\mathbf{w}-\mathbf{v}\right\|_1 \frac{\sqrt{\lambda_{max}} \sum_{i=1}^d v_i \left\|B^T\mathbf{e}_i\right\|_2 + \left\|B^T\mathbf{v}\right\|_2 \max\limits_{1\leq i\leq d} \left\|B^T\mathbf{e}_i\right\|_2}{\left(\sum_{i=1}^d w_i \left\|B^T\mathbf{e}_i\right\|_2\right)\left(\sum_{i=1}^d v_i \left\|B^T\mathbf{e}_i\right\|_2\right)} \\
\leq &\left\|\mathbf{w}-\mathbf{v}\right\|_1 \frac{\sqrt{\lambda_{max}}\sqrt{\lambda_{max}} + \sqrt{\lambda_{max}}\sqrt{\lambda_{max}}}{\sqrt{\lambda_{min}}\sqrt{\lambda_{min}}}.
\end{aligned}
$$

Therefore for fixed $\mathbf{w}$, when $\left\|\mathbf{w}-\mathbf{v}\right\|_1$ is small enough, we have $\left|\mathrm{DR}_{\mathbf{w},1} - \mathrm{DR}_{\mathbf{v},1}\right| < \varepsilon$. This proves the mapping $\mathbf{w} \to \mathrm{DR}_{\mathbf{w},1}$ is continuous. ■

Proof of Lemma 3.2.3. First note that

$$
\begin{aligned}
&\sup_{\mathbf{w}\in\Sigma^d} \left|\widehat{\mathrm{DR}}_{\mathbf{w},1} - \mathrm{DR}_{\mathbf{w},1}\right| \\
= &\sup_{\mathbf{w}\in\Sigma^d} \left| \frac{\widehat{\eta}_{\mathbf{w}}^{1/\alpha} \sum_{i=1}^d w_i \eta_{\mathbf{e}_i}^{1/\alpha} - \eta_{\mathbf{w}}^{1/\alpha} \sum_{i=1}^d w_i \widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}}{\left(\sum_{i=1}^d w_i \widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}\right)\left(\sum_{i=1}^d w_i \eta_{\mathbf{e}_i}^{1/\alpha}\right)} \right| \\
\leq &\sup_{\mathbf{w}\in\Sigma^d} \left| \frac{\left(\widehat{\eta}_{\mathbf{w}}^{1/\alpha} - \eta_{\mathbf{w}}^{1/\alpha}\right) \sum_{i=1}^d w_i \eta_{\mathbf{e}_i}^{1/\alpha}}{\left(\sum_{i=1}^d w_i \widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}\right)\left(\sum_{i=1}^d w_i \eta_{\mathbf{e}_i}^{1/\alpha}\right)} \right| + \sup_{\mathbf{w}\in\Sigma^d} \left| \frac{\eta_{\mathbf{w}}^{1/\alpha} \left(\sum_{i=1}^d w_i \eta_{\mathbf{e}_i}^{1/\alpha} - \sum_{i=1}^d w_i \widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}\right)}{\left(\sum_{i=1}^d w_i \widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}\right)\left(\sum_{i=1}^d w_i \eta_{\mathbf{e}_i}^{1/\alpha}\right)} \right|.
\end{aligned}
\tag{3.7.19}
$$

Thus, to show that (3.7.19) converges to 0 almost surely, the key is the strong consistency of $\widehat{\eta}_{\mathbf{w}}$ uniformly in $\mathbf{w}$. This is ensured by Theorem 4.4 of Mainik (2010). Further, by the continuity of the mapping $\widehat{\eta}_{\mathbf{w}} \longmapsto \widehat{\eta}_{\mathbf{w}}^{1/\alpha}$, we have

$$
\sup_{\mathbf{w}\in\Sigma^d} \left|\widehat{\eta}_{\mathbf{w}}^{1/\alpha} - \eta_{\mathbf{w}}^{1/\alpha}\right| \to 0, \qquad a.s.,
$$

and

$$\sup_{\mathbf{w}\in\Sigma^d}\left|\sum_{i=1}^d w_i\eta_{\mathbf{e}_i}^{1/\alpha}-\sum_{i=1}^d w_i\widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}\right|=\sup_{\mathbf{w}\in\Sigma^d}\left|\sum_{i=1}^d w_i\left(\eta_{\mathbf{e}_i}^{1/\alpha}-\widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}\right)\right|\to 0,\qquad a.s.$$

Further notice that $\sum_{i=1}^d w_i\eta_{\mathbf{e}_i}^{1/\alpha}$ and $\sum_{i=1}^d w_i\widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}$ are uniformly bounded away from 0 because both the empirical measure $\widehat{\Psi}$ and the limit measure Ψ are non–degenerated. Combining all these, we obtain that (3.7.19) converges to 0 almost surely, which yields the desired result. ■

Proof of Lemma 3.3.1. In this proof the limit is taken as $t\to\infty$. For $t>0$, denote the region $S_t=\{(x_1,\dots,x_d)\in\mathbb{R}_+^d:\sum_{i=1}^d w_ix_i\geq t\}$. We can split $\overline{F}_{\mathbf{w}^T\mathbf{X}}(t)$ as

$$\begin{aligned}\overline{F}_{\mathbf{w}^T\mathbf{X}}(t)&=\int_{S_t}\mathrm{d}\left(\prod_{k=1}^d G_k(x_k)\right)+\sum_{i<j}a_{i,j}\int_{S_t}\mathrm{d}\left(\left(1-G_i(x_i)\right)\left(1-G_j(x_j)\right)\prod_{k=1}^d G_k(x_k)\right)\\&=I(t)+\sum_{i<j}a_{i,j}J_{i,j}(t),\end{aligned}$$

where $G_k(x)=G(x/w_k)$ for $k=1,...,d$. The term $I(t)$ can be understood as the survival distribution function of $w_1X_1^*+\cdots+w_dX_d^*$, where $X_1^*,\dots,X_d^*$ are i.i.d. with common distribution function G. For $I(t)$, it follows from Theorems 4.7 of Mao and Ng (2015) that,

$$\frac{I(t)}{\overline{G}(t)}=\sum_{k=1}^d w_k^\alpha+\sum_{k=1}^d H_{-\alpha,\rho}\left(w_k^{-1}\right)A(t)(1+o(1))+\alpha t^{-1}\mu_G\sum_{k\neq l}w_k^\alpha w_l(1+o(1)).$$

For $J_{i,j}(t)$'s, note that it suffices to study $J_{1,2}(t)$ by symmetry. Then we have

$$\begin{aligned}J_{1,2}(t)=&I(t)-\int_{S_t}\mathrm{d}\left(G_1^2(x_1)\prod_{k=2}^d G_k(x_k)\right)-\int_{S_t}\mathrm{d}\left(G_2^2(x_2)\prod_{k\neq 2}G_k(x_k)\right)\\&+\int_{S_t}\mathrm{d}\left(G_1^2(x_1)G_2^2(x_2)\prod_{k=3}^n G_k(x_k)\right)\\=&I(t)-J_{1,2}^{(1)}(t)-J_{1,2}^{(2)}(t)+J_{1,2}^{(3)}(t).\end{aligned}$$

Note that $\overline{G}_k(x)=\overline{G}(x/w_k)\sim w_k^\alpha\overline{G}(t)$ and $\overline{G_1^2}(t)/\overline{G}_1(t)\to 2$. Since $\alpha\geq 1$, by regarding

$G_1^2(\cdot)$ as a distribution function, Proposition 3.7 of Mao and Ng (2015) leads to

$$
\begin{aligned}
&J_{1,2}^{(1)}(t)\\
&= (2w_1^\alpha + w_2^\alpha + \cdots + w_d^\alpha)\,\overline{G}(t) + o\left(\overline{G}(t)A(t)\right)\\
&+ \alpha t^{-1}\left(2w_1^\alpha \sum_{k=2}^{d} w_k\mu_G + w_1\mu_{G^2}\sum_{k=2}^{d} w_k^\alpha + \sum_{k,l\geq 2,k\neq l} w_k^\alpha w_l \mu_G\right)\overline{G}(t)(1+o(1)).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&J_{1,2}^{(2)}(t)\\
&= (w_1^\alpha + 2w_2^\alpha + \cdots + w_d^\alpha)\,\overline{G}(t) + o\left(\overline{G}(t)A(t)\right).\\
&+ \alpha t^{-1}\left(2w_2^\alpha \sum_{k\neq 2} w_k\mu_G + w_2\mu_{G^2}\sum_{k\neq 2} w_k^\alpha + \sum_{k,l\neq 2,k\neq l} w_k^\alpha w_l \mu_G\right)\overline{G}(t)(1+o(1)).
\end{aligned}
$$

and

$$
\begin{aligned}
J_{1,2}^{(3)}(t) &= (2w_1^\alpha + 2w_2^\alpha + \cdots + w_d^\alpha)\,\overline{G}(t) + o\left(\overline{G}(t)A(t)\right)\\
&+ \alpha t^{-1}\left(2\sum_{l=1}^{2}\sum_{k\neq l} w_l^\alpha w_k \mu_{G^2} + 2\sum_{l=1}^{2}\sum_{k=3}^{d} w_l^\alpha w_k\mu_G\right)\overline{G}(t)(1+o(1))\\
&+ \alpha t^{-1}\left(\sum_{l=1}^{2}\sum_{k=3}^{d} w_k^\alpha w_l \mu_{G^2} + \sum_{k,l\geq 3,k\neq l} w_k^\alpha w_l\mu_G\right)\overline{G}(t)(1+o(1)).
\end{aligned}
$$

Combining all the asymptotics for $I(t)$, $J_1(t)$, $J_2(t)$ and $J_3(t)$ yields that

$$
\begin{aligned}
\frac{\overline{F}_{\mathbf{w}^T\mathbf{X}}(t)}{\overline{G}(t)} - \sum_{k=1}^{d} w_k^\alpha &= (1+Q_{\mathbf{a}})\frac{I(t)}{\overline{G}(t)} + \frac{\sum_{i<j} a_{i,j}\left(-J_{ij}^{(1)}(t) - J_{ij}^{(2)}(t) + J_{ij}^{(3)}(t)\right)}{\overline{G}(t)} - \sum_{k=1}^{d} w_k^\alpha\\
&= \begin{cases} \alpha t^{-1}\mu_G^*(1+o(1)), & \rho < -1,\\ (1+Q_{\mathbf{a}})\sum_{k=1}^{d} H_{-\alpha,\rho}\left(w_k^{-1}\right)A(t)(1+o(1)), & \rho \geq -1. \end{cases}
\end{aligned}
$$

This completes the proof of (3.3.4).

The uniform convergence of (3.3.4) follows immediately from checking that for the limit relations in Proposition 3.7 and Theorems 4.7 of Mao and Ng (2015). The details are omitted here but are available upon request. ■

Proof of Lemma 3.3.2. In this proof we denote $\arg\min f$ by m_f for notational simplicity. By the definition of D_n, for any n, $|f_n(m_f) - f(m_f)| < D_n$. It follows that

$$f_n(m_{f_n}) \le f_n(m_f) < f(m_f) + D_n.$$

Again, by $|f_n(m_{f_n}) - f(m_{f_n})| < D_n$ we have

$$f(m_{f_n}) < f_n(m_{f_n}) + D_n < f(m_f) + 2D_n.$$

Deriving the similar inequalities for the other side yields that

$$|f(m_{f_n}) - f(m_f)| < 2D_n. \tag{3.7.20}$$

By the Taylor's theorem, for any $\mathbf{x}$ in a small neighborhood of m_f we obtain that

$$f(\mathbf{x}) = f(m_f) + \frac{1}{2}(\mathbf{x} - m_f)^T \nabla^2 f(m_f)(\mathbf{x} - m_f) + o\left(||\mathbf{x} - m_f||_2^2\right), \tag{3.7.21}$$

where we used the multi-index notation and $\nabla^2 f(m_f)$ is the Hessian matrix of f at m_f. Since $D_n \to 0$ as $n \to \infty$ by Lemma 3.2.1, m_{f_n} is in a small neighborhood of m_f for large n. It then follows from the expansion in (3.7.21) that

$$|f(m_{f_n}) - f(m_f)| > \frac{c}{2}||m_{f_n} - m_f||_2^2, \tag{3.7.22}$$

where c is the smallest eigenvalue of $\nabla^2 f(m_f)$. Combining (3.7.20) and (3.7.22) leads to the desired result. ■

Chapter 4

Diversification of Truncated Risks

4.1 Introduction

One of the fundamental techniques in mitigating portfolio risks is by diversification. However, diversification is not always preferred. For a wide class of extremely heavy-tailed risks with infinite first moment, the diversification is suboptimal even when they are independent; see e.g., Embrechts et al. (2002), Ibragimov and Walden (2007) and Ibragimov (2009). This phenomenon is related to the risk measure Value-at-Risk (VaR) being not subadditive for extremely heavy-tailed risks. For a risk X with distribution function F, the VaR at level $q \in (0, 1)$ of X is defined as

$$\mathrm{VaR}_q(X) = \inf\{x : F(x) \geq q\}.$$

With no surprise, the same non-diversification effect or the non-subadditivity of VaR is observed for extremely heavy-tailed dependent risks. For example, Wüthrich (2003), Alink et al. (2004) and Embrechts et al. (2009b) investigated the effect under the dependence structure of Archimedean copulas; Ibragimov and Prokhorov (2016) studied the power-type copulas, which includes Farlie–Gumbel–Morgenstern (FGM) copulas; Ibragimov and Walden (2011) also considered dependence arising from common multiplicative and additive shocks. These studies indicate that investing in a single extremely heavy-tailed asset has lower risk than investing in a portfolio of them.

The above results are obtained for unbounded risks. Nevertheless, in many real world applications, risks may be truncated. For example, the reinsurance, especially in the catastrophe insurance market, covers the last layer of loss. Borch (1960) and Arrow (1978) showed that if the reinsurance premium is calculated by the expected value principle, the stop-loss reinsurance treaty is the optimal strategy. Thus, risks considered here are large but truncated at a very high level. Intuitively, if the extremely heavy-tailed risks are truncated, then they should behave like light-tailed risks so that the diversification may become optimal. This is verified by Ibragimov and Walden (2007): for independent and identically distributed (i.i.d.) risks, only when the truncation level is sufficiently large, the diversification becomes suboptimal. Based on this result, the diversification is expected to be suboptimal for dependent and truncated risks, just like the i.i.d. case, but rigorous investigation has yet to be conducted. Besides the real-valued risks, nonnegative risks are also practically needed in the modeling, for example the catastrophic risks, where the "profit" does not have a meaningful explanation. Thus, the study of the diversification effect for the truncated nonnegative risks is particularly crucial for catastrophe insurance. In this chapter, we aim to provide a comprehensive study of the diversification effect for the extremely heavy-tailed and truncated risks with different dependence structures for risks that are either real-valued or nonnegative.

Next, we formalize our proposed line of inquiry. For a risk X, the truncated risk $X^{(k)}$ is defined as

$$X^{(k)} = \begin{cases} X, & X \leq k, \\ k, & X > k. \end{cases}$$

Let $S_n^{(k)} = \frac{1}{n}\left(X_1^{(k)} + \cdots + X_n^{(k)}\right)$ denote the aggregated risk. To simplify the analysis, we assume the risks are identically distributed. In order to study the diversification effect of a random vector $\mathbf{X}^{(k)} = \left(X_1^{(k)}, \ldots, X_n^{(k)}\right)^T$, we define the diversification ratio (DR) at level q as follows:

$$\mathrm{DR}_q(\mathbf{X}^{(k)}) = \frac{\mathrm{VaR}_q\left(S_n^{(k)}\right)}{\mathrm{VaR}_q\left(X_1^{(k)}\right)}. \tag{4.1.1}$$

Since we are interested in extreme risks, we focus on the behavior of DR when q is close to 1. We say that the diversification is asymptotically optimal if $\lim_{q\uparrow 1} \mathrm{DR}_q(\mathbf{X}^{(k)}) \leq 1$, and the diversification is asymptotically suboptimal if $\lim_{q\uparrow 1} \mathrm{DR}_q(\mathbf{X}^{(k)}) \geq 1$. The study of the diversification at a high level of q (close to 1) is critically important. The regulation rules such as Basel II and III Accords and the Solvency II Directive have required to use a higher confidence level of VaR, which is close to 1. Moreover, the diversification may be optimal at a lower level of q, but this can be violated for extreme risks evaluated at a high level of q. Thus, by studying the asymptotic behaviour of DR, we investigate the diversification effect for all high levels of q close to 1. This provides crucial insights for risk management, especially for catastrophe insurance market.

The heavy-tailed risks are defined through regular variation. A risk X with distribution function $F = 1 - \bar{F}$ is regularly varying with tail index α if for $x > 0$,

$$\lim_{t\to\infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha},$$

denoted by $\bar{F} \in \mathrm{RV}_{-\alpha}$. The tail index α represents the heavy-tailedness of X; the smaller the α, the heavier the X. The extremely heavy-tailedness corresponds to the case $0 < \alpha < 1$. The dependence structure in this chapter is modeled through copulas. More specifically we use Archimedean copulas, which covers a wide range of dependence from independence to comonotonicity. Such modeling allows us to flexibly study how heavy-tailedness and the dependence structure affect the diversification.

For real-valued risks, under both independent and dependent structures, we obtain a lower bound of truncation when the diversification is suboptimal. For nonnegative risks, we obtain a finer result that shows how the diversification is transitioned from optimal to suboptimal by changing the truncation level. An interesting observation is that when the truncation level is sufficiently high, the diversification stays suboptimal for any degree of extremely heavy-tailedness of the marginal distribution or the dependence structure. Vice versa, when the truncation is in a relatively low level, the diversification stays optimal, regardless of the heavy-tailedness or the dependence structure.

This chapter is organized as follows. In Section 4.2 and Section 4.3, we discuss our main results for truncated risks under independence and dependence structures. Simulation studies are carried out in Section 4.4 to further illustrate our results. Section 4.5 makes some concluding remarks. Proofs are postponed to Section 4.6.

4.2 Diversification Effect Under Independence

For i.i.d. regularly varying risks with tail index α, it is well known that VaR is non-subadditive if and only if $0 < \alpha < 1$. In the following theorem, we show that similar results hold for i.i.d. truncated risks as long as the truncation level k is large enough.

Theorem 4.2.1 *Let $X_1, \ldots, X_n$ be n i.i.d. risks with continuous distribution function F such that $\bar{F} \in \mathrm{RV}_{-\alpha}$ with $0 < \alpha < 1$. If $\lim_{q\uparrow 1} \frac{k(q)}{\mathrm{VaR}_q(X_1)} = c > \left(\frac{n^{1-\alpha}-1}{n}\right)^{-1/\alpha}$, then diversification is asymptotically suboptimal for truncated risks $X_1^{(k)}, \ldots, X_n^{(k)}$.*

The above theorem offers the same insight as Theorem 1 in Ibragimov and Walden (2007) that VaR may also be non-subadditive for truncated risks. Our result is obtained in an asymptotic way in a sense that for all q close enough to 1 and the truncation k growing with $\mathrm{VaR}_q(X_1)$, the diversification is suboptimal; while their result is obtained for all $q \in (1/2, 1)$ and the truncation level k also depends on q. It is noteworthy in Ibragimov and Walden (2007) that they assume distribution F falls into the class of symmetric stable distributions. In establishing Theorem 4.2.1, we relax this technical assumption to regularly varying distributions.

The above result holds for real-valued risks. However, when dealing with catastrophic events, the "profit" does not have a meaningful interpretation. In the following theorem, we focus on nonnegative risks and examine the asymptotic diversification behaviour for different truncation levels.

Theorem 4.2.2 *Let $X_1, \ldots, X_n$ be i.i.d. nonnegative risks with continuous distribution function F such that $\bar{F} \in \mathrm{RV}_{-\alpha}$ with $0 < \alpha < 1$. Assume $\lim_{q \uparrow 1} \frac{k(q)}{\mathrm{VaR}_q(X_1)} = c > 0$.*

- *If $c > n$, the diversification is asymptotically suboptimal for $X_1^{(k)}, \ldots, X_n^{(k)}$.*
- *If $c < n$, the diversification is asymptotically optimal for $X_1^{(k)}, \ldots, X_n^{(k)}$.*

For nonnegative risks, Theorem 4.2.2 gives the basis of a qualitative picture for the asymptotic diversification effect. It clearly highlights the switch from optimal to suboptimal when moving from a small truncation level ($c < n$) to a large truncation level ($c > n$). Compared with the result in Theorem 4.2.1, we notice that diversification becomes suboptimal at a much lower level of k. This is intuitive since the influence of the negative tails is removed, positive extreme values can no longer be compensated by negative ones, which results in less diversification benefit. Another interesting observation is that when the truncation level is high enough, the diversification stays suboptimal for any degree of the extremely heavy-tailedness of the marginal distribution.

A by-product of Theorem 4.2.2 is the following result on the asymptotic behaviour of aggregated risks.

Corollary 4.2.1 *Let $X_1, \ldots, X_n$ be i.i.d. nonnegative risks with continuous distribution function F such that $\bar{F} \in \mathrm{RV}_{-\alpha}$ with $0 < \alpha < 1$. Assume $\lim_{t \to \infty} \frac{k(t)}{t} = c > 0$.*

- *If $c > n$, we have*

$$\lim_{t\to\infty} \frac{\mathbb{P}\left(S_n^{(k)} > t\right)}{\mathbb{P}\left(X_1^{(k)} > t\right)} = n^{1-\alpha}.$$

- *If $1 < c < n$, we have*

$$\mathbb{P}\left(S_n^{(k)} > t\right) = o\left(\mathbb{P}\left(X_1^{(k)} > t\right)\right).$$

4.3 Diversification Effect Under Dependence

In this section, we investigate the diversification effect for dependent risks. The dependence structure is modeled by Archimedean copulas, which cover from independence to strong dependence. Typical examples of Archimedean copulas are Clayton, Gumbel and Frank copulas. An Archimedean copula is defined as

$$C(u_1, \ldots, u_n) = \phi^{-1}(\phi(u_1) + \ldots + \phi(u_n)),$$

where the generator function $\phi : [0, 1] \to [0, \infty]$ is continuous, decreasing and convex such that $\phi(1) = 0$ and $\phi(0) = \infty$, and ϕ^{-1} is the inverse of ϕ. We further assume ϕ^{-1} is completely monotonic, i.e. $(-1)^i \left(\phi^{-1}\right)^{(i)} \geq 0$ for all $i \in \mathbb{N}$. These requirements ensure that C is a copula for all dimensions $n \geq 2$.

Since we focus on the tail risks, an Archimedean survival copula $\widehat{C}$ is assumed, that is,

$$\mathbb{P}(X_1 > x_1, \ldots, X_n > x_n) = \widehat{C}\left(\bar{F}_1(x_1), \ldots, \bar{F}_n(x_n)\right),$$

where $\widehat{C}$ is an Archimedean copula.

The diversification effect of truncated dependent risks is studied in the following theorem.

Theorem 4.3.1 *Let $X_1, \ldots, X_n$ be identical risks with continuous distribution function F such that $\bar{F} \in \mathrm{RV}_{-\alpha}$ with $0 < \alpha < 1$ and follow an Archimedean survival copula with $\phi \in \mathrm{RV}_{-\beta}$ at 0^+ with $\beta > 0$. If $\lim_{q\uparrow 1} \frac{k(q)}{\mathrm{VaR}_q(X_1)} = c > \left(\frac{q_n(\alpha,\beta)n^{-\alpha}-1}{c_n(\beta)}\right)^{-1/\alpha}$, with $c_n(\beta) = \sum_{i=1}^n \binom{n}{n-i}(-1)^{i-1} i^{-1/\beta}$ and $q_n(\alpha, \beta) = \int_{\mathbb{R}_+^n} 1_{\left\{\sum_{i=1}^n \frac{1}{x_i} > 1\right\}} \frac{d^n}{dx_1 \cdots dx_n} \left(\sum_{i=1}^n x_i^{-\alpha\beta}\right)^{-1/\beta} \mathrm{d}x_1 \cdots \mathrm{d}x_n$, then the diversification is asymptotically suboptimal for $X_1^{(k)}, \ldots, X_n^{(k)}$.*

The tail index β of the generator function ϕ measures the upper tail dependence, i.e. extreme losses tend to occur together. It is known from Embrechts et al. (2009b) that $\lim_{\beta\to 0} q_n(\alpha,\beta) = \lim_{\beta\to 0} c_n(0) = n$ and $\lim_{\beta\to\infty} q_n(\alpha,\beta) = n^\alpha$, $\lim_{\beta\to\infty} c_n(\beta) = 1$. Then one can easily check when $\beta \to 0$, the result in Theorem 4.3.1 is consistent with that in Theorem 4.2.1. On the other hand, the above result also implies the diversification is always suboptimal for truncated risks when $\beta \to \infty$. Actually this reduces to a comonotonic case, where the DR is 1 for any q.

If we restrict ourselves to nonnegative risks, more informative results can obtained as follows.

Theorem 4.3.2 *Let $X_1, \ldots, X_n$ be identical nonnegative risks with continuous distribution function F such that $\bar{F} \in \mathrm{RV}_{-\alpha}$ with $0 < \alpha < 1$ and follow an Archimedean survival copula with $\phi \in \mathrm{RV}_{-\beta}$ at 0^+ with $\beta > 0$. Assume $\lim_{q\uparrow 1} \frac{k(q)}{\mathrm{VaR}_q(X_1)} = c > 0$.*

- *If $c > n$, the diversification is asymptotically suboptimal for $X_1^{(k)}, \ldots, X_n^{(k)}$.*
- *If $c < n$, the diversification is asymptotically optimal for $X_1^{(k)}, \ldots, X_n^{(k)}$.*

For nonnegative risks, Theorem 4.3.2 shows how the diversification is transitioned from optimal to suboptimal by changing the truncation level. From the above result, the transition of truncation level does not depend on α or β. That is to say, having less heavy marginals or less dependent structure on $\mathbf{X}$ will not make the diversification optimal, when the truncation level is sufficiently high.

Similarly, a by-product of Theorem 4.3.2 is given as follows.

Corollary 4.3.1 *Let $X_1, \ldots, X_n$ be identical nonnegative risks with continuous distribution function F such that $\bar{F} \in \mathrm{RV}_{-\alpha}$ with $0 < \alpha < 1$ and follow an Archimedean survival copula with $\phi \in \mathrm{RV}_{-\beta}$ at 0^+ with $\beta > 0$. Assume $\lim_{t\to\infty} \frac{k(t)}{t} = c > 0$.*

- *If $c > n$, we have*
$$\lim_{t\to\infty} \frac{\mathbb{P}\left(S_n^{(k)} > t\right)}{\mathbb{P}\left(X_1^{(k)} > t\right)} = q_n(\alpha,\beta) n^{-\alpha}.$$

- *If* $1 < c < n$, *we have*

$$\lim_{t\to\infty} \frac{\mathbb{P}\left(S_n^{(k)} > t\right)}{\mathbb{P}\left(X_1^{(k)} > t\right)} = q_n(\alpha,\beta)n^{-\alpha} - \sum_{i=1}^{\lfloor \frac{n}{c} \rfloor} (-1)^{i-1} \binom{n}{i} (\Delta_{1,i}(\alpha,\beta,c) - \Delta_{2,i}(\alpha,\beta,c)), \quad \textit{with}$$

$$\Delta_{1,i}(\alpha,\beta,c) = \int_{\mathbb{R}_+^n} 1_{\left\{\sum_{j=1}^n \frac{1}{x_j} > n, x_1 < 1/c, \ldots, x_i < 1/c\right\}} \frac{d^n}{dx_1 \cdots dx_n} \left(\sum_{j=1}^n x_j^{-\alpha\beta}\right)^{-1/\beta} \mathrm{d}x_1 \cdots \mathrm{d}x_n, \quad \textit{and}$$

$$\Delta_{2,i}(\alpha,\beta,c) = \int_{\mathbb{R}_+^n} 1_{\left\{\sum_{j=i+1}^n \frac{1}{x_j} > n - ic, x_1 < 1/c, \ldots, x_i < 1/c\right\}} \frac{d^n}{dx_1 \cdots dx_n} \left(\sum_{j=1}^n x_j^{-\alpha\beta}\right)^{-1/\beta} \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

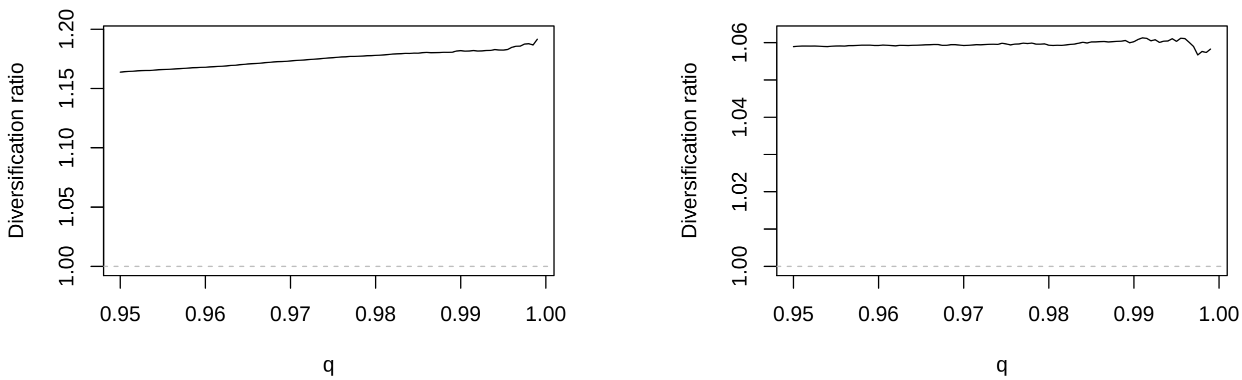
4.4 Simulation Study

In this section, we perform simulations to illustrate the theoretical results obtained in Sections 4.2 and 4.3. For dependence model, we consider the Clayton copula, which is a special case of Archimedean copula. The generator function of the Clayton copula is $\phi(t) = \frac{1}{\beta}(t^{-\beta} - 1)$ with $\beta > 0$. One can check that this generator function is regularly varying at 0. By varying β, the Clayton copula covers from independence ($\beta \to 0$) to comonotonicity ($\beta \to \infty$). This enables us to study how dependence plays a role in the diversification effect. In all our reported experiments, we consider a portfolio consisting of two risks. We simulate 1,000,000 times of such portfolio. After truncation at level k, we compute the DR in (4.1.1) at level q by using the following approximation of VaR_q of a risk X

$$\mathrm{VaR}_q(X) \approx X_{[1000000q]},$$

where $X_{[i]}$ is the ith largest observation. We iterate above approach 100 times and obtain 100 DR estimates. The average of these estimates is then used to characterize the diversification behaviour.

In the first example, we consider real-valued risks. Assume each risk follows a Student-t distribution, where the degree of freedom is exactly the same as the tail index α. This example is designed to illustrate the results in Theorems 4.2.1 and 4.3.1. We keep $\alpha = 0.8$. By Theorem 4.2.1, the sufficient condition for diversification being suboptimal is approximately equivalent to $c > 25.76$. Thus, we pick $k = 26\mathrm{VaR}_q(X_1)$ in our simulation (Figure 4.1(a)). When dependence is introduced (Figure 4.1(b)), choosing $\beta = 1$, by numerical integration and Theorem 4.3.1, the sufficient condition truncation for diversification being not preferred is approximately equivalent to $c > 79.67$. Thus, we pick $k = 80\mathrm{VaR}_q(X_1)$. As seen in Figure 4.1, the DR always exceeds 1 under different levels of q starting from 0.95, which suggests that the diversification is indeed suboptimal under a high enough truncation level. A point worthy of mention is that Theorems 4.2.1 and 4.3.1 only give a sufficient condition on the truncation level such that the diversification is suboptimal, which means the bound for c is not sharp.

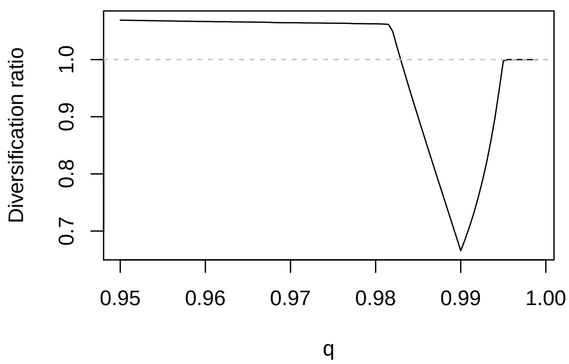


(a) $\alpha = 0.8, \beta = 0$

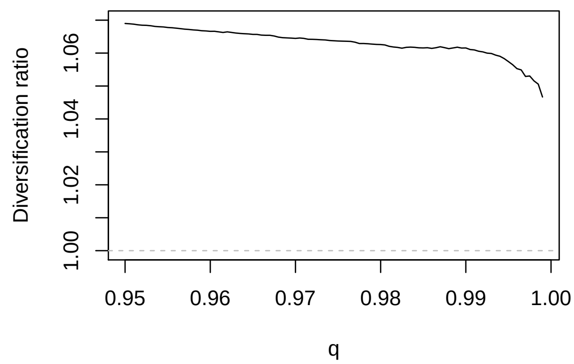
(b) $\alpha = 0.8, \beta = 1$

Figure 4.1: DR for the portfolio with two Student-t risks is plotted against q. The left graph is for independent risks with tail index $\alpha = 0.8$. The truncation level k is chosen as $26\mathrm{VaR}_q(X_1)$. The right graph is for dependent risks with tail index $\alpha = 0.8$ and $\beta = 1$. The truncation level k is chosen as $80\mathrm{VaR}_q(X_1)$. The grey dotted line corresponds to $\mathrm{DR} = 1$.

In the second example, we consider nonnegative risks. Assume the two risks follow a Clayton copula and each risk follows a Pareto distribution with tail index α as $F(x) = 1 - (1+x)^{-\alpha}$. The model parameters are chosen to be $\alpha = 0.8, \beta = 1$. In Figure 4.2(a), we keep $k = \mathrm{VaR}_{0.99}(X_1)$ fixed. By varying q from 0.95 to 0.999, the ratio $k/\mathrm{VaR}_q(X_1)$ experiences a transition from greater than 2 to less than 2, and eventually close to 0. We see a clear pattern in Figure 4.2(a) that, the DR first drops below 1 at lower levels of q, then increases after $q > 0.99$ and finally approaches to 1. This finding supports our result in Theorem 4.3.2 that there is a switch from suboptimal to optimal when moving from a high enough truncation level to a relatively small truncation level. On the other hand, if we let k grow with q, by selecting proper ratio, a large enough truncation level can always be achieved for diversification to be suboptimal. This is confirmed in Figure 4.2(b) that the DR is above 1 for all q by taking $k = 3\mathrm{VaR}_q(X_1)$.



(a) $\alpha = 0.8, \beta = 1$



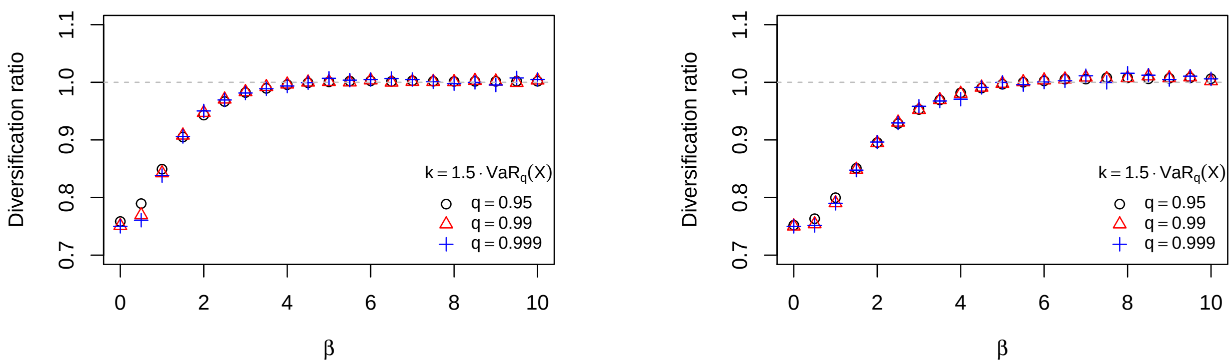
(b) $\alpha = 0.8, \beta = 1$

Figure 4.2: DR for the portfolio with two Pareto risks following a Clayton copula is plotted against q. The model parameters are chosen to be $\alpha = 0.8, \beta = 1$. The truncation level k in the left graph is fixed at $\mathrm{VaR}_{0.99}(X_1)$; while in the right graph, k is set to be $3\mathrm{VaR}_q(X_1)$. The grey dotted line corresponds to $\mathrm{DR} = 1$.

We use the last example to highlight the implication of our results that the transition of truncation level does not depend on α and β. We first fix $k = 1.5\mathrm{VaR}_q(X_1)$. By varying the tail index β of the generator function ϕ and the level of q for VaR, we plot the value of DR in Figure 4.3. The left graph of Figure 4.3 is plotted for risks with tail index $\alpha = 0.8$, and the right graph of Figure 4.3 is plotted for risks with tail index $\alpha = 0.5$. The level of q is set to 0.95, 0.99 and 0.999. Then we proceed in a similar way by letting k equal $3\mathrm{VaR}_q(X_1)$. Our results shown in Figure 4.3 and Figure 4.4 closely follow Theorem 4.2.2 and Theorem 4.3.2 that the diversification is asymptotically suboptimal (resp. optimal) when $c > 2$ (resp. $c < 2$), regardless of the dependence structure and the heavy-tailedness of the marginals. Besides, one can notice that in Figure 4.3, the DR is an increasing function of β; while in Figure 4.4, the DR decreases as the dependence becomes stronger. This finding is actually verified in our proof for Theorem 4.3.2.

4.5 Conclusion

In this chapter, we study the effects of diversification for truncated extremely heavy-tailed risks with different dependence structures through the diversification ratio defined

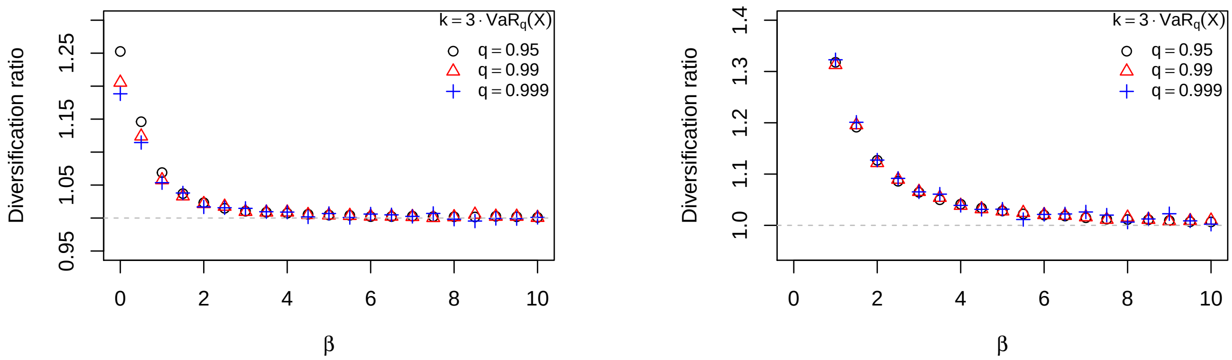


(a) $\alpha = 0.8$

(b) $\alpha = 0.5$

Figure 4.3: DR for the portfolio with two Pareto risks following a Clayton copula is plotted by varying the tail index β of the generator function. The left graph is for the risks with tail index $\alpha = 0.8$. The right graph is for the risks with tail index $\alpha = 0.5$. The level of q is set to 0.95, 0.99 and 0.999. The truncation level k is chosen as $1.5\mathrm{VaR}_q(X_1)$. The grey dotted line corresponds to DR $= 1$.

in (4.1.1). Intuitively, truncated extremely heavy-tailed risks behave like light-tailed risks, which results in that diversification is always optimal. However, for the catastrophic risks, the truncation of the risks are usually far in the tail. To this end, we investigate the diversification effect when the truncation level increases asymptoticly with the VaR of the risk. Under both structures, our finding suggests that the diversification effect is much easier to become suboptimal for nonnegative risks (catastrophic risks) than that for real-valued risks. For nonnegative risks with a sufficiently large truncation level, a finer result is obtained on how the diversification effect transits from suboptimal to optimal by varying the truncation level. It also implies that the transition of diversification effect does not depend on the heavy-tailedness of the marginal distribution or the dependence structure, but the number of risks matters. The simulation studies further illustrate our main results. From the experiments on the real-valued risks, the bound on the truncation level is not sharp and future improvement is needed.



(a) $\alpha = 0.8$

(b) $\alpha = 0.5$

Figure 4.4: DR for the portfolio with two Pareto risks following a Clayton copula is plotted by varying the tail index β of the generator function. The left graph is for the risks with tail index $\alpha = 0.8$. The right graph is for the risks with tail index $\alpha = 0.5$. The level of q is set to 0.95, 0.99 and 0.999. The truncation level k is chosen as $3\mathrm{VaR}_q(X_1)$. The grey dotted line corresponds to $\mathrm{DR} = 1$.

4.6 Proofs

To make the notation simplified, we denote $S_n = \frac{1}{n}(X_1 + \cdots + X_n)$ throughout this section.

First note that for $k < \mathrm{VaR}_q(X_1)$, $\mathbb{P}\left(X_1^{(k)} = k\right) = \mathbb{P}(X_1 \geq k) > 1 - q$. Then $\mathrm{VaR}_q\left(X_1^{(k)}\right) = k$. Since $\mathrm{VaR}_q\left(S_n^{(k)}\right) \leq k$, it immediately implies that $\mathrm{DR}_q(\mathbf{X}^{(k)}) \leq 1$ for all q. Thus, the diversification is asymptotically optimal for all $c \leq 1$.

Our remaining proof focuses on the case where $c > 1$. Note that, for any two random variables X and Y, if $\mathbb{P}(X > t) > \mathbb{P}(Y > t)$ for all $t \geq t_0$, then $\mathrm{VaR}_q(X) > \mathrm{VaR}_q(Y)$ for $q_0 \leq q < 1$. Thus, if the truncation level $k > t$, to prove diversification is asymptotically sub-optimal or optimal, it suffices to study

$$\lim_{t\to\infty} \frac{\mathbb{P}\left(S_n^{(k)} > t\right)}{\mathbb{P}\left(X_1^{(k)} > t\right)} \geq (\text{or }) \leq 1. \tag{4.6.1}$$

Moreover, by letting $t = \mathrm{VaR}_q(X_1)$, the condition $\lim_{q \uparrow 1} k(q)/\mathrm{VaR}_q(X_1) = c$ is equivalent to that

$$\lim_{t\to\infty} \frac{k(t)}{t} = c.$$

These are how we start the proofs in the following.

Proof of Theorem 4.2.1. Since $c > \left(\frac{n^{1-\alpha}-1}{n}\right)^{-1/\alpha} > 1$, there exists t_0 such that $k(t) > t$ for all $t \geq t_0$. First we consider the numerator of (4.6.1). Depending on on the number of risks that exceed the truncation level k, we have

$$\begin{aligned}
\mathbb{P}\left(S_n^{(k)} > t\right) &= \mathbb{P}\left(S_n > t,\ \text{none of } X_i > k\right) + \mathbb{P}\left(S_n^{(k)} > t,\ \text{at least one of } X_i > k\right) \\
&= \mathbb{P}\left(S_n > t\right) - I,
\end{aligned}$$

where

$$I = \mathbb{P}\left(S_n > t, \text{at least one of } X_i > k\right) - \mathbb{P}\left(S_n^{(k)} > t, \text{at least one of } X_i > k\right). \tag{4.6.2}$$

By noting that $I \leq \mathbb{P}\left(\text{at least one of } X_i > k\right)$, it leads to

$$\mathbb{P}\left(S_n^{(k)} > t\right) \geq \mathbb{P}\left(S_n > t\right) - \mathbb{P}\left(\text{at least one of } X_i > k\right). \tag{4.6.3}$$

Next, since $k(t) > t$ when $t > t_0$, the denominator of (4.6.1) is indeed $\mathbb{P}\left(X_1^{(k)} > t\right) = \mathbb{P}(X_1 > t)$. Now combining the above analysis yields that

$$\begin{aligned}
\lim_{t\to\infty} \frac{\mathbb{P}\left(S_n^{(k)} > t\right)}{\mathbb{P}\left(X_1^{(k)} > t\right)} &\geq \lim_{t\to\infty} \frac{\mathbb{P}\left(S_n > t\right) - \mathbb{P}\left(\text{at least one of } X_i > k\right)}{\mathbb{P}(X_1 > t)} \\
&= \lim_{t\to\infty} \frac{\mathbb{P}\left(S_n > t\right)}{\mathbb{P}(X_1 > t)} - \lim_{t\to\infty} \frac{1 - (1 - \bar{F}(k))^n}{\bar{F}(t)} \\
&= n^{1-\alpha} - nc^{-\alpha} > 1,
\end{aligned}$$

where the third step is due to Corollary 1.3.2 in Embrechts et al. (2013). ■

Proof of Theorem 4.2.2. First we consider the case $c > n$. In this case, there exists t_0 such that $k > nt$ for all $t \geq t_0$. Since $X_1, \ldots, X_n$ are nonnegative, it follows that $\mathbb{P}\left(S_n > t, X_1 > k\right) = \mathbb{P}(X_1 > k)$. Then, $I = 0$ for $t \geq t_0$, where I is defined in (4.6.2). Under the same decomposition in the proof of Theorem 4.2.1, we have

$$\lim_{t\to\infty} \frac{\mathbb{P}\left(S_n^{(k)} > t\right)}{\mathbb{P}\left(X_1^{(k)} > t\right)} = \lim_{t\to\infty} \frac{\mathbb{P}\left(S_n > t\right)}{\mathbb{P}\left(X_1 > t\right)} = n^{1-\alpha} > 1.$$

Now we turn to the case $1 < c < n$. In this case, there exists t_0 such that $t < k < nt$ for all $t \geq t_0$. Depending on the number of risks that exceed k, we further split I as

$$
\begin{aligned}
I &= \mathbb{P}\left(S_n > t, \text{only one of } X_i > k\right) + \mathbb{P}\left(S_n > t, \text{more than one of } X_i > k\right) \\
&\quad - \mathbb{P}\left(S_n^{(k)} > t, \text{only one of } X_i > k\right) - \mathbb{P}\left(S_n^{(k)} > t, \text{more than one of } X_i > k\right) \\
&:= J_1 + J_2 - J_3 - J_4.
\end{aligned}
$$

For $\varepsilon > 0,$ by symmetry we can rewrite J_1 as

$$
\begin{aligned}
J_1 &= n\mathbb{P}\left(S_n > t, X_1 > k, X_2 \leq k, \ldots, X_n \leq k\right) \\
&= n\mathbb{P}\left(S_n > t, X_1 > nt, X_2 \leq k, \ldots, X_n \leq k\right) + n\mathbb{P}\left(S_n > t, (n-\varepsilon)t < X_1 \leq nt, X_2 \leq k, \ldots, X_n \leq k\right) \\
&\quad + n\mathbb{P}\left(S_n > t, k < X_1 \leq (n-\varepsilon)t, X_2 \leq k, \ldots, X_n \leq k\right) \\
&:= n(J_{11} + J_{12} + J_{13}).
\end{aligned}
$$

By the independence of X_i's and $k \to \infty$ as $t \to \infty$, we have

$$
\lim_{t\to\infty} \frac{J_{11}}{\bar{F}(t)} = \lim_{t\to\infty} \frac{\mathbb{P}\left(X_1 > nt, X_2 \leq k, \ldots, X_n \leq k\right)}{\bar{F}(t)} = n^{-\alpha},
$$

and

$$
0 \leq \lim_{t\to\infty} \frac{J_{12}}{\bar{F}(t)} \leq \lim_{t\to\infty} \frac{\bar{F}((n-\varepsilon)t) - \bar{F}(nt)}{\bar{F}(t)} = (n-\varepsilon)^{-\alpha} - n^{-\alpha}.
$$

For J_{13}, applying Corollary 1.3.2 in Embrechts et al. (2013), we have

$$
\begin{aligned}
0 \leq \lim_{t\to\infty} \frac{J_{13}}{\bar{F}(t)} &\leq \lim_{t\to\infty} \frac{\mathbb{P}(X_2 + \ldots + X_n > nt - X_1, k < X_1 \leq (n-\varepsilon)t)}{\bar{F}(t)} \\
&\leq \lim_{t\to\infty} \frac{\mathbb{P}\left(X_2 + \ldots + X_n > \varepsilon t, X_1 > k\right)}{\bar{F}(t)} \\
&\leq (n-1)\varepsilon^{-\alpha} \lim_{t\to\infty} \bar{F}(k) = 0.
\end{aligned}
$$

By the arbitrariness of ε, we obtain $\lim_{t\to\infty} J_1/\bar{F}(t) = n^{1-\alpha}$.

For J_2, if there are exactly two of X_i that are greater than k, without loss of generality we assume they are X_1 and X_2. Then,

$$
\mathbb{P}\left(S_n > t, X_1 > k, X_2 > k\right) \leq \left(\bar{F}(k)\right)^2.
$$

which implies that $J_2 = o(\bar{F}(t))$ as $t \to \infty$. Similarly, $J_4 = o(\bar{F}(t))$.

For J_3, by the symmetry, the independence and non-negativity of X_i's, and $nt-k \to \infty$ as $t \to \infty$, we have

$$\begin{aligned}
0 \le \lim_{t\to\infty} \frac{J_3}{\bar{F}(t)} &= \lim_{t\to\infty} \frac{n\bar{F}(k)\mathbb{P}\left(X_2 + \ldots + X_n > nt - k, X_2 \le k, \ldots, X_n \le k\right)}{\bar{F}(t)} \\
&\le \lim_{t\to\infty} \frac{n\bar{F}(k)\mathbb{P}\left(X_2 + \ldots + X_n > nt - k\right)}{\bar{F}(t)} \\
&\le n(n-1)(n-c)^{-\alpha} \lim_{t\to\infty} \bar{F}(k) = 0.
\end{aligned}$$

Combining I, J_1, J_2, J_3, and J_4 yields the desired result. ■

Proof of Theorem 4.3.1. Recall the proof in Theorem 4.2.1,

$$\begin{aligned}
\lim_{t\to\infty} \frac{I_1}{\bar{F}(t)} &\le \lim_{t\to\infty} \frac{\mathbb{P}(\text{at least one of } X_i > k)}{\bar{F}(t)} \\
&= \lim_{t\to\infty} \frac{1 - \sum_{i=0}^{n} \left\{ \binom{n}{n-i} (-1)^i \left[\phi^{-1}\left(i\Phi(\bar{F}(k)) \right) \right] \right\}}{\bar{F}(t)} \\
&= \lim_{t\to\infty} \frac{\sum_{i=1}^{n} \left\{ \binom{n}{n-i} (-1)^{i-1} \left[\phi^{-1}\left(i\Phi(\bar{F}(k)) \right) \right] \right\}}{\bar{F}(t)} \\
&= \lim_{t\to\infty} \frac{\sum_{i=1}^{n} \left\{ \binom{n}{n-i} (-1)^{i-1} \left[i^{-1/\beta} \bar{F}(k) \right] \right\}}{\bar{F}(t)} \\
&= \sum_{i=1}^{n} \left\{ \binom{n}{n-i} (-1)^{i-1} i^{-1/\beta} \right\} c^{-\alpha} \\
&= c_n(\beta) c^{-\alpha},
\end{aligned}$$

where $c_n(\beta) = \sum_{i=1}^{n} \binom{n}{n-i} (-1)^{i-1} i^{-1/\beta}$. The fourth step and fifth step is because $\phi^{-1} \in \mathrm{RV}_{-1/\beta}$ and $\bar{F}(\cdot) \in \mathrm{RV}_{-\alpha}$, respectively.

Now consider the following limiting ratio, since $c > \left(\frac{q_n(\alpha,\beta) n^{-\alpha} - 1}{c_n(\beta)} \right)^{-1/\alpha}$, we prove that

$$\begin{aligned}
\lim_{t\to\infty} \frac{\mathbb{P}\left(S_n^{(k)} > t \right)}{\mathbb{P}\left(X_1^{(k)} > t \right)} &= \lim_{t\to\infty} \frac{\mathbb{P}\left(S_n > t\right) - I_1}{\mathbb{P}(X_1 > t)} \\
&\ge q_n(\alpha, \beta) n^{-\alpha} - c_n(\beta) c^{-\alpha} \\
&> 1.
\end{aligned}$$

This completes the proof for the asymptotic sub-optimality of diversification. ■

Proof of Theorem 4.3.2. First we consider the case $c > n$. The proof is similar to that of Theorem 4.2.2 by noting that $\mathbb{P}\left(S_n > t, X_1 > k\right) = \mathbb{P}(X_1 > k)$ if $X_1, \ldots, X_n$ are nonnegative. Then, we have

$$\lim_{t\to\infty} \frac{\mathbb{P}\left(S_n^{(k)} > t\right)}{\mathbb{P}\left(X_1^{(k)} > t\right)} = \lim_{t\to\infty} \frac{\mathbb{P}\left(S_n > t\right)}{\mathbb{P}\left(X_1 > t\right)} = q_n(\alpha, \beta) n^{-\alpha} \geq 1.$$

The last inequality holds for all $\alpha \in (0,1)$ and $\beta > 0$, see Lemma 3.1 (d) in Embrechts et al. (2009b).

Now we turn to the case $1 < c < n$. For arbitrary small $\varepsilon < c - \frac{n}{\lfloor \frac{n}{c} \rfloor + 1}$, there exists t_0 such that $(c - \varepsilon)t < k < nt$ for all $t \geq t_0$. Define $A_i = \{X_i > k\}$. Applying the inclusion-exclusion principle, the first part of I can be further split as

$$\begin{aligned}
\mathbb{P}\left(S_n > t, \text{at least one of } X_i > k\right) &= \mathbb{P}\left(X_1 + \ldots + X_n > nt, \bigcup_{i=1}^{n} A_i\right) \\
&= \sum_{i=1}^{n} \left((-1)^{i-1} \sum_{\substack{\mathcal{Z} \subset \{1,\ldots,n\} \\ |\mathcal{Z}| = i}} \mathbb{P}\left(X_1 + \ldots + X_n > nt, \bigcap_{j\in\mathcal{Z}} A_j\right)\right) \\
&= \sum_{i=1}^{\lfloor \frac{n}{c} \rfloor} \left((-1)^{i-1} \binom{n}{i} \mathbb{P}\left(X_1 + \ldots + X_n > nt, \bigcap_{j=1}^{i} A_j\right)\right) \\
&+ \sum_{i=\lfloor \frac{n}{c} \rfloor + 1}^{n} \left((-1)^{i-1} \binom{n}{i} \mathbb{P}\left(\bigcap_{j=1}^{i} A_j\right)\right).
\end{aligned} \tag{4.6.4}$$

The first part of the third equality is by symmetry and the second part is due to the fact that

$$\begin{aligned}
X_1 + \ldots + X_n &> \left(\left\lfloor \frac{n}{c} \right\rfloor + 1\right) k \\
&> \left(\left\lfloor \frac{n}{c} \right\rfloor + 1\right) (c - \varepsilon) t \\
&> \left(\left\lfloor \frac{n}{c} \right\rfloor + 1\right) \frac{n}{\lfloor \frac{n}{c} \rfloor + 1} t \\
&> nt.
\end{aligned}$$

Similarly, the second part of I can be written as

$$\mathbb{P}\left(S_n^{(k)} > t, \text{at least one of } X_i > k\right) = \sum_{i=1}^{\lfloor \frac{n}{c} \rfloor} \left((-1)^{i-1} \binom{n}{i} \mathbb{P}\left(X_{i+1} + \ldots + X_n > nt - ik, \bigcap_{j=1}^{i} A_j \right) \right)$$
$$+ \sum_{i=\lfloor \frac{n}{c} \rfloor + 1}^{n} \left((-1)^{i-1} \binom{n}{i} \mathbb{P}\left(\bigcap_{j=1}^{i} A_j \right) \right). \tag{4.6.5}$$

Then it follows from (4.6.4) and (4.6.5), I in (4.6.2) can be reformulated as

$$\begin{aligned} I = &\sum_{i=1}^{\lfloor \frac{n}{c} \rfloor} \left((-1)^{i-1} \binom{n}{i} \mathbb{P}\left(X_1 + \ldots + X_n > nt, \bigcap_{j=1}^{i} A_j \right) \right) \\ &- \sum_{i=1}^{\lfloor \frac{n}{c} \rfloor} \left((-1)^{i-1} \binom{n}{i} \mathbb{P}\left(X_{i+1} + \ldots + X_n > nt - ik, \bigcap_{j=1}^{i} A_j \right) \right) \\ &:= \sum_{i=1}^{\lfloor \frac{n}{c} \rfloor} (-1)^{i-1} \binom{n}{i} (J_{1,i} - J_{2,i}). \end{aligned}$$

For $J_{1,i}$, the key idea is to connect $\mathbb{P}\left(X_1 + \ldots + X_n > nt | X_1 > k, \ldots, X_i > k\right)$ together with $\mathbb{P}\left(X_j > \frac{t}{x_j}, j = 1, \ldots, n | X_1 > k, \ldots, X_i > k\right)$. By similar methods as in the proof of Theorem 2.2 in Alink et al. (2004), we can take random variables $\left(Y_1^{(t)}, \ldots, Y_n^{(t)}\right)$ with the following distribution function:

$$\mathbb{P}(Y_1^{(t)} \le x_1, \ldots, Y_n^{(t)} \le x_n) = \mathbb{P}\left(X_j > \frac{t}{x_j}, j = 1, \ldots, n | X_1 > k, \ldots, X_i > k \right).$$

Then $\left(Y_1^{(t)}, \ldots, Y_n^{(t)}\right)$ converges weakly to $Y_1, \ldots, Y_n$, defined on $(0, 1/c)^i \times (0, \infty)^{n-i}$ with distribution function $G_{i,c}^{\beta,\alpha}$, where

$$\begin{aligned}
G_{i,c}^{\beta,\alpha}(x_1,\ldots,x_n) &= \lim_{t\to\infty} \frac{\mathbb{P}\left(X_j > \frac{t}{x_j}, j=1,\ldots,n\right)}{\mathbb{P}(X_1 > k,\ldots,X_i > k)} \\
&= \lim_{t\to\infty} \frac{\phi^{-1}(\phi(\bar{F}(t/x_1) + \ldots + \bar{F}(t/x_n)))}{\phi^{-1}(i\phi(\bar{F}(k)))} \\
&= \lim_{t\to\infty} \frac{\phi^{-1}\left(\left(\sum_{j=1}^n x_j^{-\alpha\beta}\right)\phi(\bar{F}(t)\right)}{\phi^{-1}(ic^{\alpha\beta}\phi(\bar{F}(t)))} \\
&= \left(\sum_{j=1}^n x_j^{-\alpha\beta}\right)^{-1/\beta} i^{1/\beta} c^{\alpha}. \qquad (4.6.6)
\end{aligned}$$

Therefore, we can find that

$$\mathbb{P}\left(\sum_{j=1}^n \frac{1}{Y_j^{(t)}} > n\right) = \mathbb{P}\left(\sum_{j=1}^n X_j > nt | X_1 > k,\ldots,X_i > k\right)$$

converges (again as $t\to\infty$) to

$$\begin{aligned}
\mathbb{P}\left(\sum_{j=1}^n \frac{1}{Y_j} > n\right) &= i^{1/\beta}c^{\alpha}\int_{\mathbb{R}_+^n} 1_{\left\{\sum_{j=1}^n \frac{1}{x_j} > n, x_1<1/c,\ldots,x_i<1/c\right\}} \frac{d^n}{dx_1\cdots dx_n}\left(\sum_{j=1}^n x_j^{-\alpha\beta}\right)^{-1/\beta} \mathrm{d}x_1\cdots\mathrm{d}x_n \\
&= i^{1/\beta}c^{\alpha}\Delta_{1,i}(\alpha,\beta,c). \qquad (4.6.7)
\end{aligned}$$

(4.6.7) tells us that

$$\lim_{t\to\infty}\frac{J_{1,i}}{\bar{F}(t)} = \lim_{t\to\infty}\mathbb{P}\left(\sum_{j=1}^n X_j > nt | X_1 > k,\ldots,X_i > k\right)\frac{\mathbb{P}(X_1 > k,\ldots,X_i > k)}{\bar{F}(t)} = \Delta_{1,i}(\alpha,\beta,c).$$

The same argument applied to $J_{2,i}$ yields that

$$\begin{aligned}
\lim_{t\to\infty}\frac{J_{2,i}}{\bar{F}(t)} &= \int_{\mathbb{R}_+^n} 1_{\left\{\sum_{j=i+1}^n \frac{1}{x_j} > n-ic, x_1<1/c,\ldots,x_i<1/c\right\}} \frac{d^n}{dx_1\cdots dx_n}\left(\sum_{j=1}^n x_j^{-\alpha\beta}\right)^{-1/\beta} \mathrm{d}x_1\cdots\mathrm{d}x_n \\
&= \Delta_{2,i}(\alpha,\beta,c).
\end{aligned}$$

Combining $I, J_{1,i}, J_{2,i}$, we have

$$\begin{aligned}\lim_{t\to\infty}\frac{\mathbb{P}\left(S_n^{(k)}>t\right)}{\mathbb{P}\left(X_1^{(k)}>t\right)}&=\lim_{t\to\infty}\frac{\mathbb{P}\left(S_n>t\right)}{\mathbb{P}\left(X_1>t\right)}-\lim_{t\to\infty}\frac{I}{\bar{F}(t)}\\&=q_n(\alpha,\beta)n^{-\alpha}-\sum_{i=1}^{\left\lfloor\frac{n}{c}\right\rfloor}(-1)^{i-1}\binom{n}{i}\left(\Delta_{1,i}(\alpha,\beta,c)-\Delta_{2,i}(\alpha,\beta,c)\right)\\&=\Delta_n(\alpha,\beta,c).\end{aligned}\tag{4.6.8}$$

By arguing as above, the limit (4.6.1) exists. In the following, we derive the monotonicity of $\Delta_n(\alpha,\beta,c)$ as a function of β by using the concept of supermodular ordering. Such an ordering is used to compare the dependence of multivariate distributions with equal marginals; see Müller (1997) for details, definitions and further properties.

We choose two random vectors $\mathbf{X}$ and $\mathbf{Y}$ as follows: they have Archimedean survival copula with parameters $0<\beta_X<\beta_Y$ and identical marginal functions. With Theorem 3.1 in Wei and Hu (2002) and equation (4.3) in Embrechts et al. (2009b), we know that $\mathbf{X}$ is smaller than $\mathbf{Y}$ in supermoduar ordering. That is to say, for all supermodular functions f, we have $\mathbb{E}f(\mathbf{X})\le\mathbb{E}f(\mathbf{Y})$. Note that the function $\mathbf{x}\in\mathbb{R}^n\mapsto S_n(\mathbf{x})$ is convex and supermodular, so is the function $\mathbf{x}\mapsto(S_n(\mathbf{x})-t)_+$ by preservation of convexity. It also follows from Theorem 2.5(b) in Müller (1997) that the supermodular property is preserved under coordinatewise increasing transformations. This immediately implies that the function $\mathbf{x}\mapsto\left(S_n^{(k)}(\mathbf{x})-t\right)_+$ is supermodular. Hence for two real numbers $t<k$

$$\begin{aligned}&\mathbb{E}\left(S_n^{(k)}(\mathbf{X})-t\right)_+\le\mathbb{E}\left(S_n^{(k)}(\mathbf{Y})-t\right)_+\\\Longrightarrow&\int_0^{k-t}\mathbb{P}\left(S_n^{(k)}(\mathbf{Y})>t+y\right)\mathrm{d}y-\int_0^{k-t}\mathbb{P}\left(S_n^{(k)}(\mathbf{X})>t+y\right)\mathrm{d}y\ge0.\end{aligned}$$

Choose $\varepsilon>0$. For sufficiently large t, one can let $k\in((c-\varepsilon)t,(c+\varepsilon)t)$. By (4.6.8) and the monotonicity of $\Delta_n(\alpha,\beta,c)$ in c, we have that

$$(1+\varepsilon)\Delta_n(\alpha,\beta_Y,c+\varepsilon)\int_0^{k-t}\bar{F}(t+y)\mathrm{d}y-(1-\varepsilon)\Delta_n(\alpha,\beta_X,c-\varepsilon)\int_0^{k-t}\bar{F}(t+y)\mathrm{d}y\ge0.\tag{4.6.9}$$

As $t\to\infty$, the integral $\int_0^{k-t}\bar{F}(t+y)\mathrm{d}y$ tends to infinity, and therefore (4.6.9) forces $(1+\varepsilon)\Delta_n(\alpha,\beta_Y,c+\varepsilon)$ is at least equal to $(1-\varepsilon)\Delta_n(\alpha,\beta_X,c-\varepsilon)$. It then follows that the limiting ratio (4.6.1) is an increasing function of β because ε was arbitrary.

Due to the monotonicity of $\Delta_n(\alpha,\beta,c)$ in β, to establish the desired result, it suffices to show that $\lim_{\beta\to\infty}\Delta_n(\alpha,\beta,c)=1$. Using the transform $x_j\mapsto x_j^{-1/\alpha}$, we obtain

$$\Delta_{1,i}(\alpha,\beta,c)=\int_{\mathbb{R}_+^n}1_{\left\{\sum_{j=1}^n x_j^{1/\alpha}>n, x_1^{-1/\alpha}<1/c,\ldots,x_i^{-1/\alpha}<1/c\right\}}g_\beta(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{4.6.10}$$

where $\mathbf{x}=(x_1,\ldots,x_n)$ and $g_\beta(\mathbf{x})=\frac{d^n}{dx_1\cdots dx_n}\left(\sum_{j=1}^n x_j^\beta\right)^{-1/\beta}$. On $\mathbb{R}_+^n$, we reformulate the problem in polar coordinates. Let $\mathbf{x}\mapsto(\|\mathbf{x}\|_1,\mathbf{x}/\|\mathbf{x}\|_1)$. We rewrite (4.6.10) as follows.

$$\begin{aligned}\Delta_{1,i}(\alpha,\beta,c)&=\int_{[0,\infty)\times\mathcal{S}_+^{n-1}}1_{\left\{r>\max\left\{n^\alpha\left(\sum_{j=1}^n w_j^{1/\alpha}\right)^{-\alpha},\frac{c^\alpha}{w_1},\ldots,\frac{c^\alpha}{w_i}\right\}\right\}}r^{-2}g_\beta(\mathbf{w})\mathrm{d}r\mathrm{d}\mathbf{w}\\&=\int_{\mathcal{S}_+^{n-1}}\min\left\{n^{-\alpha}\left(\sum_{j=1}^n w_j^{1/\alpha}\right)^\alpha, c^{-\alpha}w_1,\ldots,c^{-\alpha}w_i\right\}g_\beta(\mathbf{w})\mathrm{d}\mathbf{w},\end{aligned}$$

where $\mathcal{S}_+^{n-1}=\left\{\mathbf{w}\in\mathbb{R}_+^n:\|\mathbf{w}\|_1=1\right\}$ is the unit simplex and $g_\beta(\cdot)/n$ is a probability density on $\mathcal{S}_+^{n-1}$; see equation (3.7) of Kotz and Nadarajah (2000). By similar approach, we obtain

$$\Delta_{2,i}(\alpha,\beta,c)=\int_{\mathcal{S}_+^{n-1}}\min\left\{(n-ic)^{-\alpha}\left(\sum_{j=i+1}^n w_j^{1/\alpha}\right)^\alpha, c^{-\alpha}w_1,\ldots,c^{-\alpha}w_i\right\}g_\beta(\mathbf{w})\mathrm{d}\mathbf{w}.$$

Note that when $\beta\to\infty$, $\left(\sum_{j=1}^n x_j^\beta\right)^{-1/\beta}$ converges to $\min\limits_{j=1,\ldots,n}\frac{1}{x_j}$. Then by Proposition 5.26 on p.294 of Resnick (2013), the probability density function $g_\beta(\mathbf{w})/n$ can only place all its mass on the point $(1/n,\ldots,1/n)$, which means

$$\begin{aligned}&\lim_{\beta\to\infty}\Delta_{1,i}(\alpha,\beta,c)=n\cdot\min\left\{1/n,c^{-\alpha}/n,\ldots,c^{-\alpha}/n\right\}=c^{-\alpha},\\&\lim_{\beta\to\infty}\Delta_{2,i}(\alpha,\beta,c)=n\cdot\min\left\{((n-ic)/(n-i))^{-\alpha}/n,c^{-\alpha}/n,\ldots,c^{-\alpha}/n\right\}=c^{-\alpha}.\end{aligned}\tag{4.6.11}$$

Additionally, Lemma 3.3 of Embrechts et al. (2009b) shows that $\lim_{\beta\to\infty}q_n(\alpha,\beta)=n^\alpha$. It then follows by (4.6.8) and (4.6.11)

$$\lim_{\beta\to\infty}\Delta_n(\alpha,\beta,c)=1-\sum_{i=1}^{\left\lfloor\frac{n}{c}\right\rfloor}(-1)^{i-1}\binom{n}{i}\left(c^{-\alpha}-c^{-\alpha}\right)=1. \tag{4.6.12}$$

The desired result holds by the monotonicity of $\Delta_n(\alpha,\beta,c)$ in β and (4.6.12). ■

Chapter 5

Portfolio Credit Risk with Archimedean Copulas: Asymptotic Analysis and Efficient Simulation

5.1 Introduction

Credit risk in banking and trading book is by far the largest financial risk exposure for many financial institutions. One of the most prominent examples of the importance of credit risk management and supervision is of course the 2007-2009 financial crisis. After experiencing the financial crisis, regulators have been very active in the development of new, stricter solvency guidelines. In particular, they recommend the banking and insurance industry carefully choose the model for the dependence structure of the default events. Default dependence has a direct impact on the upper tail of the credit loss distribution for a large portfolio. Therefore, analyzing and modelling the dependence among a large number of defaultable obligors is especially helpful for the description of credit portfolios and of large insurance portfolios in the actuarial sciences.

The event of default for an individual obligor within the portfolio is often captured using the so-called threshold models. These models can be viewed as multivariate extensions of the Merton's seminal firm value model, see Merton (1974). The idea is that default occurs for an obligor i if some critical random variable X_i, usually called a latent variable, exceeds (or falls below) a pre-specfied threshold. The dependence among defaults then stems from the dependence among those latent variables. It has been found that the copula representation is a useful concept for studying the dependence structures. Specifically, the copula of the latent variables determines the link between marginal default probabilities for individual obligors and the joint default probabilities for groups of obligors. Although there are numerous tractable copula functions, most threshold models used in industry are based explicitly or implicitly on the Gaussian copula, for example, CreditMetrics (Gupton et al. 1997) and Moody's KMV system (Kealhofer and Bohn 2001). Under the Gaussian copula framework, default dependence is usually induced from a set of common factors affecting multiple obligors. These factors are typically interpreted as systematic risks and with these factors, many threshold models have convenient representations as mixture models, see the monograph by McNeil et al. (2015).

In terms of the performance measures of portfolio credit risk, a main focus is on the probability of large portfolio loss over a fixed time horizon, e.g., a one-year horizon. Credit portfolios are often large; however, the default probabilities of high-quality obligors are extremely small. These features then spurred a vast literature on the asymptotic study of large portfolio loss. Vasicek (1987, 1991) show how to derive a simple closed-form solution for the loss distribution of an asymptotically large, homogeneous portfolio. By assuming a Gaussian copula between different borrowers, the losses from default are conditionally independent and identically distributed (i.i.d.), the limiting loss distribution is therefore immediate from the law of large numbers. By similar approach, Lucas et al. (2001) and Gordy (2003) study the loss distribution for a large heterogeneous portfolio. The tail behaviour of the loss for a large heterogeneous portfolio can also be analyzed using large deviation arguments, see Dembo et al. (2004) and Glasserman et al. (2007). Despite its popularity, the Gaussian copula-based models are criticized for lacking flexibility of modelling the tail dependence. Regarding the problem of large homogeneous portfolio approximation, Vasicek's result is respectively extended to the case of the Archimedean copula in Schönbucher (2002) and the t copula in Schloegl and OKane (2005). Bush et al. (2011) further provide a dynamic extension of Vasicek's model under the setting that the common risk factor follows a Brownian motion and study the loss distribution through a stochastic partial differential equation.

This chapter is concerned about portfolio credit risk with extremal dependence. The model we develop builds on the threshold approach and assumes an Archimedean copula to model the dependence structure among latent variables. Our main objective is to derive sharp asymptotics for the tail of the portfolio loss distribution, in contrast to the existing logarithmic asymptotics provided in Maier and Wüthrich (2009). It is also worth noting that unlike the work of Schönbucher (2002), our model relies on a more general semiparametric assumption (generally speaking, a single parameter is used to capture the extent of extremal dependence present in the portfolio). In addition, we develop sharp asymptotics for expected shortfall, a risk measure that is widely used in both risk management and credit derivatives pricing. Another contribution of this work is that we construct two quite different algorithms to efficiently estimate the portfolio risk via simulation. The first is an algorithm based on hazard rate twisting (see Juneja and Shahabuddin (2002)), which is shown to be asymptotically optimal. The second algorithm uses the idea of conditional Monte Carlo (see, e.g., Asmussen and Kroese (2006)) and has bounded relative error. This suggests that the second algorithm should outperform the first algorithm, and we indeed verify this to be the case in our simulation study. However, the key contribution of the first algorithm lies in estimating more general risk measures, such as expected shortfall, whereas the second algorithm is specifically designed to estimate loss probabilities.

In order to capture a more realistic dependence structure, we also study portfolio credit risk under nested Archimedean copulas. Nested Archimedean copulas are constructed by recursive application of Archimedean geneartors; see McNeil (2008), Hofert (2008, 2011) for technical details. Such a generalization provides a promising tool for multi-level dependence modelling. Especially for a large credit portfolio, it allows us to classify obligors based on a certain attribute such as industry sector, geographic location, or quality of risk. In this chapter, we focus on a nested Archimedean copula with a Gumbel generator. Empirically, it has been shown by many authors that Gumbel family is preferable for the modelling of portfolio credit risk and the pricing of credit derivatives; see, e.g., Hofert and Scherer (2011), Choroś-Tomczyk et al. (2013) and Jakob and Fischer (2014). With respect to the probability of large portfolio losses, we provide an asymptotic lower bound on its decay rate. An algorithm based on conditional Monte Carlo is given for the nested Gumbel copula and it also exhibits bounded relative error.

The rest of the chapter is organized as follows. In Section 5.2, we formulate our problem and link the portfolio structure with the LT-Archimedean copula and nested Archimedean copula. Section 5.3, 5.4 and 5.5 exhibit the main results: the former section derives the sharp asymptotics and the latter two sections provide efficient algorithms and investigates their performance. The performance of the proposed algorithms is further demonstrated via an extensive simulation study in Section 5.6. Proofs are relegated to Section 5.7.

5.2 Problem Formulation

5.2.1 General Portfolio Structure and Relation to Copulas

Consider a large credit portfolio of n obligors. Similarly to Bassamboo et al. (2008), we employ a static structural model for portfolio loss by introducing latent variables $\{X_1, \ldots, X_n\}$

so that each obligor defaults if its latent variable exceeds some pre-specified threshold x_i. In the structural modelling framework, X_i can be interpreted as the loss on the assets of obligor i. The threshold x_i is implied from the marginal obligor default probability p_i. The associated risk exposure at default is denoted as $c_i > 0$. The loss incurred from defaults is then given by

$$L_n = \sum_{i=1}^{n} c_i 1_{\{X_i > x_i\}},$$

where 1_A is the indicator function of an event A. Such a threshold model can first go back to Merton (1974). We denote by $F_i(x) = \mathbb{P}(X_i \leq x)$ the marginal distribution function of X_i. It is clear that the marginal default probability of obligor i is given by $p_i = \bar{F}_i(x_i)$.

When dealing with the above credit portfolios, dependence structure of the latent variables serves a critical role. Most threshold models popular in the financial industry are based explicitly or implicitly on the Gaussian copula, where the vector of latent variables follows a multivariate normal distribution. The underlying dependence structure is often specified through a linear factor model

$$X_i = \beta_i \mathbf{a}_i' \mathbf{F} + \sqrt{1 - \beta_i^2} \varepsilon_i, i = 1, \ldots, n, \tag{5.2.1}$$

where $\mathbf{F}$ is a vector of common factors satisfying $\mathbf{F} \sim N_p(\mathbf{0}, \Omega)$ with $p < n$. These factors typically measure global, country and industry effects impacting all obligors and ε_i are idiosyncratic variables. It follows that β_i as a constant, can be viewed as a measure of systematic risk of X_i. Usually $\mathbf{a}_i$ is chosen by imposing the constraint $\mathbf{a}_i' \Omega \mathbf{a}_i = 1$ for all i. The model therefore assumes that $\mathbf{a}_i' \mathbf{F}$ and $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. standard normal random variables.

However, in the credit risk context, it has been argued that the main source of risk in large credit portfolios is the occurrence of many near simultaneous defaults. The Gaussian copula models, although can accommodate a wide range of different correlation structures, become inadequate to model extremal dependence between the latent variables. Alternatives have been proposed, most prominently the t copula and the large family of Archimedean copula. In this chapter, we restrict ourselves to Archimedean copulas that are able to capture upper tail dependence of latent variables. Formally, the upper tail depedence is defined as

$$\lambda_U = \lim_{q \uparrow 1} \mathbb{P}(X_2 > F_2^{\leftarrow}(q) | X_1 > F_1^{\leftarrow}(q)),$$

where $F^{\leftarrow}(q) = \inf\{x \in \mathbb{R} : F(x) \geq q\}$. This means that latent variables may simultaneously take on very large values (we focus on loss distribution) with non-negligible probability. The importance of copulas in a threshold model is highlighted by Lemma 11.2 of McNeil et al. (2015). Their result shows that in a threshold model, the copula of the latent variables determines the link between marginal default probabilities and portfolio default probabilities, which is of our interest. Let $U_i = F_i(X_i)$ for $i = 1, \ldots, n$. Lemma 11.2 immediately implies that $(U_i, p_i)_{1\le i\le n}$ and $(X_i, x_i)_{1\le i\le n}$ are two equivalent threshold models. Therefore, in our subsequent analysis, the credit portfolio loss is modelled as

$$L_n = \sum_{i=1}^{n} c_i 1_{\{U_i > 1-p_i\}}. \tag{5.2.2}$$

Note that one can also use survival copulas to describe the dependence structure among latent variables. The only difference is that for those survival copulas, the property of lower tail dependence is required.

5.2.2 Large Portfolio Loss and Low Default Probability

Following the idea of Bassamboo et al. (2008), the main focus of this chapter is on an asymptotic regime where the credit portfolio is consisted of a large number of obligors and each obligor has low default probability. Those rare but significant large loss events are of our interest. As we target a large credit portfolio with low-default probability, the probability of large portfolio loss should diminish as n increases. To set the stage, we assume that the individual default probability equals $l_i f_n$ for $i = 1, \ldots, n$, where f_n is a decreasing function converging to 0 as $n \to \infty$ and $\{l_1, \ldots, l_n\}$ are strictly positive constants accounting for variations effect on different obligors. In this way, we rewrite the overall portfolio loss as

$$L_n = \sum_{i=1}^{n} c_i 1_{\{U_i > 1-l_i f_n\}}. \tag{5.2.3}$$

To characterize the potential heterogeneity among obligors, we adopt the same assumption on the sequence $\{(c_i, l_i) : i \ge 1\}$ as in Bassamboo et al. (2008).

Assumption 5.2.1 *Let the positive sequence $((c_i, l_i) : i \ge 1)$ take values in a finite set $\mathcal{W}$. Denote n_j by the number of each element $(c_j, l_j) \in \mathcal{W}$ in the portfolio. Further assume that n_j/n converges to $w_j > 0$, for each $j \le |\mathcal{W}|$ as $n \to \infty$.*

In practice, such an assumption can be interpreted as a heterogeneous credit portfolio is comprised of a finite number of homogeneous sub-portfolios based on risk type and exposure sizes. We note that it is easy to relax this assumption to the case where c_i and l_i are random variables; see Tong et al. (2016) and Tang et al. (2019) for recent discussions.

Later in this chapter, we attempt to develop sharp asymptotics and efficient simulation techniques for the tail probability of large portfolio losses, $\mathbb{P}(L_n > nb)$ and the expected shortfall, as $n \to \infty$, where b is an arbitrarily fixed number smaller than $\bar{c} := \sum_{j \leq |\mathcal{W}|} c_j w_j$, i.e. the limiting average loss when all obligors default.

5.2.3 Archimedean Copulas

LT-Archimedean Copulas

Now consider the threshold model (5.2.2), and assume $\mathbf{U} = (U_1, \ldots, U_n)$ has an LT-Archimedean copula with generator ϕ. By Proposition 2.2.1, there exists a mixture representation of U_i, i.e., $U_i = \phi^{-1}\left(\frac{R_i}{V}\right)$, where $R_1, \ldots, R_n$ is a sequence of i.i.d. standard exponential random variables that are also independent of V. By such a construction, threshold models that we consider can be represented as one-factor Bernoulli mixture models with mixing variable V.

As notably mentioned in McNeil et al. (2015), Bernoulli mixture models lend themselves to practical advantages in Monte Carlo simulations. In order to simulate from a Bernoulli mixture model, one can first simulate a realization v of V and then conduct independent Bernoulli experiments with conditional default probabilities $p_i(v)$. Moreover, Bernoulli mixture models offer more convenience for asymptotic analysis of large portfolio loss. In particular, we will later see that in one-factor models the tail of the loss distribution is essentially determined by the mixing distribution of V or its L-S transform ϕ^{-1}.

Nested Archimedean Copulas

Following the structure of a partially nested Archimedean copula introduced in (2.2.3), in this chapter, we have

$$C(\mathbf{u}) = \phi_0^{-1}\left(\sum_{j=1}^{|\mathcal{W}|} \phi_0 \circ \phi_j^{-1}\left(\sum_{l=1}^{n_j} \phi_j(u_{jl})\right)\right),$$

where $\mathbf{u} = (u_{jl}), 1 \leq l \leq n_j$ and $1 \leq j \leq |\mathcal{W}|$. In conjunction with Assumption 5.2.1, C_j actually describes the dependence structure within sub-portfolio j whose portfolio size is n_j. The number of inner copulas is equivalent to the number of sub-portfolios, which is $|\mathcal{W}|$.

5.3 Asymptotic Analysis

In this section, an asymptotic analysis is performed on a regime where the number of obligors is large, each individual obligor has an excellent credit rating (with small default probability), and the focus is on large portfolio losses.

5.3.1 Discussion of the Assumption on f_n

To get a rough idea on the tail behaviour of L_n, we take the form (5.2.3) as an example and consider a simplified case with $f_n \equiv f$ being a constant. Furthermore, we assume the latent variables follow an LT-Archimedean copula, see Definition 2.2.1.

Let

$$\begin{aligned} p_0(v,i) &:= \mathbb{P}\left(U_i > 1 - l_i f | V = v\right) \\ &= \mathbb{P}\left(\frac{R_i}{V} < \phi(1 - l_i f)\middle| V = v\right) \\ &= \mathbb{P}\left(R_i < v\phi(1 - l_i f)\middle| V = v\right) \\ &= 1 - \exp(-v\phi(1 - l_i f)). \end{aligned} \tag{5.3.1}$$

Note that $p_0(v,i)$ is strictly increasing in v. Under the condition that $V = v$, due to Kolmogorov's strong law of large numbers, almost surely

$$\frac{L_n}{n} \to r_0(v) := \sum_{j \leq |\mathcal{W}|} c_j w_j p_0(v,j), \qquad \text{as } n \to \infty,$$

where the limit follows from Assumption 5.2.1 and $r_0(v)$ denotes the limiting average portfolio loss when $V = v$. Clearly, $r_0(v)$ is also strictly increasing in v. Let v_0^* be defined as the unique solution to

$$r_0(v) = b.$$

It then follows that for $v > v_0^*$, the limiting average portfolio loss is greater than b, and hence the event of large loss $\{L_n > nb\}$ happens with probability one. For $v \leq v_0^*$, the limiting average portfolio loss is less than or equal to b, and hence the probability of large losses vanishes as $n \to \infty$. Thus, for any $b \in (0, \bar{c})$,

$$\lim_{n\to\infty} \mathbb{P}(L_n > nb) = \bar{F}_V(v_0^*). \tag{5.3.2}$$

As we can see, the probability of large losses is no longer small. Above example therefore explains the need for the assumption that f_n diminishes to 0 as $n \to \infty$, to account for the rarity of large loss. For the case that $f_n \to 0$, $p_0(v,i)$ in (5.3.1) converges to 0, so is L_n/n, and then, the set $\{r_0(v) > b\}$ is empty. In order to develop an asymptotic approximation for $\mathbb{P}(L_n > nb)$, we have to rely on the tail behaviour of V.

5.3.2 Sharp Asymptotics under LT-Archimedean Copulas

Consider the portfolio loss model given in (5.2.3). We restrict ourselves to LT-Archiemdean copulas to fully take advantage of the Bernoulli mixture structure explained in Section 5.2.3. Inspired by Proposition 2.2.3, we impose an additional assumption on ϕ such that $\phi \in \mathrm{RV}_\alpha(1)$ to account for the upper tail dependence. Moreover, by convexity of the generator ϕ, the condition $\alpha > 1$ necessarily holds.

Motivated by the heuristic analysis in Section 5.3.1, by conditioning on $V = \dfrac{v}{\phi(1-f_n)}$, we have

$$\begin{aligned} p(v,i) &:= \mathbb{P}\left(U_i > 1 - l_i f_n \middle| V = \frac{v}{\phi(1-f_n)}\right) \\ &= \mathbb{P}\left(R_i < v\frac{\phi(1-l_i f_n)}{\phi(1-f_n)}\right) \\ &= 1 - \exp\left(-v\frac{\phi(1-l_i f_n)}{\phi(1-f_n)}\right). \end{aligned} \tag{5.3.3}$$

Since $\phi \in \mathrm{RV}_\alpha(1)$, we immediately obtain that

$$\lim_{n\to\infty} p(v,i) = 1 - \exp\left(-v l_i^\alpha\right) := \tilde{p}(v,i).$$

Similarly to the derivation in Section (5.3.1), under the condition $V = \dfrac{v}{\phi(1-f_n)}$, by Kolmogorov's strong law of large numbers it follows that, almost surely

$$\frac{L_n}{n} \to r(v) := \sum_{j \le |\mathcal{W}|} c_j w_j \tilde{p}(v,j), \qquad \text{as } n \to \infty. \tag{5.3.4}$$

Note that $r(v)$ is strictly increasing in v and attains its upper bound $\bar{c} = \sum_{j\le|\mathcal{W}|} c_j w_j$ at infinity. Thus, for each $b \in (0,\bar{c})$, we let v^* denote the unique solution to

$$r(v) = b.$$

Essentially, v^* represents the threshold value so that for $V \in (0, v^*/\phi(1-f_n))$, the limiting average portfolio loss is less than b; for $V \in (v^*/\phi(1-f_n), \infty)$, the limiting average portfolio loss is greater than b. The following theorem derives a sharp asymptotic for the probability of large portfolio losses.

Theorem 5.3.1 *Consider the portfolio loss (5.2.3) and assume the following:*

- *Assumption 5.2.1 holds true,*
- $\mathbf{U} = (U_1, \ldots, U_n)$ *follows an LT-Archimedean copula with generator* ϕ,
- $\phi \in \mathrm{RV}_\alpha(1)$ *for some* $\alpha > 1$,
- ϕ^{-1} *is the Laplace-Stieltjes transform for some continuous random variable,*
- $\exp(-n\beta) = o(f_n)$ *for any* $\beta > 0$.

Then the relation

$$\mathbb{P}(L_n > nb) \sim f_n \frac{(v^*)^{-1/\alpha}}{\Gamma(1-1/\alpha)} \tag{5.3.5}$$

holds true for any fixed $b \in (0, \bar{c})$.

Our Theorem 5.3.1 shows that the probability of large portfolio loss diminishes to zero at the same rate as f_n. From expression (5.3.5), the asymptotic behaviour of the portfolio loss is mostly governed by f_n and α. The assumption that f_n decays at a subexponential rate ensures that a large portfolio loss occurs primarily when V takes large values, whereas $R_i, i = 1, \ldots, n$ generally does not play any role in its occurrence. For α, it controls the likelihood that obligors tend to default simultaneously.

Next we use an example to further illustrate the implications of our results.

Example 5.3.1 Assume a fully homogeneous portfolio, that is $l_i \equiv l, c_i \equiv c$. Under this assumption, (5.3.4) can be simplified to

$$r(v) = c\left(1 - \exp\left(-vl^\alpha\right)\right).$$

Thus, $v^* = l^{-\alpha} \ln \dfrac{c}{c-b}$ is the unique solution to $r(v) = b$. It immediately follows from relation (5.3.5), that

$$\mathbb{P}(L_n > nb) \sim l f_n \frac{\left(\ln \frac{c}{c-b}\right)^{-1/\alpha}}{\Gamma(1-1/\alpha)} \tag{5.3.6}$$

holds for all $b \in (0, c)$. Straight calculation further shows that the right-hand side of (5.3.6) is an increasing function of α if $\ln \frac{c}{c-b} \geq \exp(-\gamma)$, i.e., $b/c \geq 1 - e^{-e^{-\gamma}}$, where γ denotes the Euler's constant. This monotonic result can be interpreted in an intuitive way. A larger α corresponds to a stronger upper tail dependence, therefore a joint default of obligors is more likely to occur. However, the monotonicity fails if b is not large, in which a large portfolio loss may occur due to a single default, therefore the result will not entirely based on the upper tail dependence.

Note that the tail probability $\mathbb{P}(L_n > nb)$ appears as the denominator in the calculation of expected shortfall. Therefore, Theorem 5.3.1 becomes the key to establishing an asymptotic for the expected shortfall in Theorem 5.3.2.

Theorem 5.3.2 *Under the same assumption as in Theorem 5.3.1, the following relation*

$$\mathbb{E}\left[L_n | L_n > nb\right] \sim n\psi(\alpha, b) \tag{5.3.7}$$

holds true for any fixed $b \in (0, \bar{c})$, *where*

$$\psi(\alpha, b) := b + \frac{\int_{v^*}^{\infty} r'(v) v^{-1/\alpha} \mathrm{d}v}{(v^*)^{-1/\alpha}}.$$

The theorem above states that the expected shortfall grows almost linearly with the size of the portfolio n.

5.3.3 Sharp Asymptotics under Nested Gumbel Copulas

In this section, we choose the partially nested structure in (2.2.3) with Gumbel families for its ability and convenience to capture the hierarchical structure in the random vector $\mathbf{U} = (U_{jl})_{1 \leq l \leq n_j, 1 \leq j \leq |\mathcal{W}|}$. Recall that the sufficient condition for a proper nested Archimedean copula is $\phi_0 \circ \phi_j^{-1}$ has completely monotonic derivatives for any $1 \leq j \leq |\mathcal{W}|$. If we restrict ourselves to generators from Gumbel families, i.e., $\phi_0(\cdot) = \phi(\cdot; \alpha_0)$ and $\phi_j(\cdot) = \phi(\cdot; \alpha_j)$, it has been verified that the condition is automatically fulfilled if $\alpha_0 \leq \alpha_j, 1 \leq j \leq |\mathcal{W}|$; see Joe (1997). This result implies that the upper tail dependence is increasing in the depth of nesting, which means obligors in a sub-portfolio are more likely to default than obligors belonging to different sub-portfolios.

We have a mixture representation for random vector $\mathbf{U}$ in (2.2.4), where $\{U_{jl}\}_{1 \leq l \leq n_j, 1 \leq j \leq |\mathcal{W}|}$ are conditionally independent given mixing variables $V_j, 1 \leq j \leq |\mathcal{W}|$. Under Gumbel copula families, the mixing variables V_0 and V_j of the outer copula C_0 and the inner copulas

	Outer	Inner
Parameter	$\alpha_0 \geq 1$	$\alpha_j/\alpha_0 \geq 1$
Latent variable	$V_0 \sim S\left(\frac{1}{\alpha_0}, 1, \left(cos\frac{\pi}{2\alpha_0}\right)^{\alpha_0}, 0; 1\right)$	$V_j(V_0) \sim S\left(\frac{\alpha_0}{\alpha_j}, 1, \left(V_0 cos\frac{\pi\alpha_0}{2\alpha_j}\right)^{\alpha_j/\alpha_0}, 0; 1\right)$
L-S transform	$\phi_0^{-1}(s) = \exp\left(-s^{1/\alpha_0}\right)$	$\psi_{0,j}^{-1}(\cdot; V_0) = \exp(-V_0 s^{\alpha_0/\alpha_j})$

Table 5.1: Partially nested Archimedean copulas defined by (2.2.3) for Gumbel families

C_j have same stable distributions but different parameters. Details are listed in Table 5.1; see Nolan (2003) for the stable parametrization $S(\alpha, \beta, \gamma, \delta; 1)$.

Note that γ is the scale parameter in above parametrization, then $V_j(V_0)$ can be further represented as $V_0^{\frac{\alpha_j}{\alpha_0}} \tilde{V}_j$, where $\tilde{V}_j \sim S\left(\frac{\alpha_0}{\alpha_j}, 1, \left(cos\frac{\pi\alpha_0}{2\alpha_j}\right)^{\alpha_j/\alpha_0}, 0; 1\right)$ and is independent of V_0 for every $1 \leq j \leq |\mathcal{W}|$. Besides, for Gumbel generators, the following relation always holds, $\phi_j(x) = \phi_0(x)^{\alpha_j/\alpha_0}$ for all j.

Similarly to the derivation in Section 5.3.2, by conditioning on $\tilde{V}_j, 1 \leq j \leq |\mathcal{W}|$ and V_0, it holds that,

$$
\begin{aligned}
p(v_0, v_j, j) &:= \mathbb{P}\left(U_{jl} > 1 - l_j f_n \middle| \tilde{V}_j = v_j, V_0 = \frac{v_0}{\phi_0(1-f_n)}\right) \\
&= \mathbb{P}\left(R_{jl} < V_j \phi_j(1 - l_j f_n) \middle| \tilde{V}_j = v_j, V_0 = \frac{v_0}{\phi_0(1-f_n)}\right) \\
&= \mathbb{P}\left(R_{jl} < v_j v_0^{\alpha_j/\alpha_0} \frac{\phi_j(1-l_j f_n)}{\phi_j(1-f_n)}\right) \\
&= 1 - \exp\left(-v_j v_0^{\alpha_j/\alpha_0} \frac{\phi_j(1-l_j f_n)}{\phi_j(1-f_n)}\right), \qquad 1 \leq l \leq n_j.
\end{aligned}
$$

Since Gumbel generator $\phi_j \in \mathrm{RV}_{\alpha_j}(1)$, we immediately obtain that

$$
\lim_{n\to\infty} p(v_0, v_j, j) = 1 - \exp\left(-v_j v_0^{\alpha_j/\alpha_0} l_j^{\alpha_j}\right) := \tilde{p}(v_0, v_j, j).
$$

Therefore, by Kolmogorov's strong law of large numbers it follows that, almost surely

$$
\frac{L_n}{n} \to r(v_0, v_1, \ldots, v_{|\mathcal{W}|}) := \sum_{j \leq |\mathcal{W}|} c_j w_j \tilde{p}(v_0, v_j, j), \qquad \text{as } n \to \infty, \tag{5.3.8}
$$

given the conditions $V_0 = \dfrac{v_0}{\phi_0(1-f_n)}$ and $\tilde{V}_j = v_j$ for all $1 \le j \le |\mathcal{W}|$. Note that for the LT-Archimedean copulas discussed in Section 5.3.2, there is only one single direction in which shifting the underlying factor could increase the limiting average portfolio loss, i.e., increase the value for V. However, for nested Archimedean copulas, there might be many different directions depending on the values of b and $c_j w_j, 1 \le j \le |\mathcal{W}|$, which adds substantial complexity to the asymptotic analysis of large portfolio loss.

In the following theorem, we provide an asymptotic lower bound for the probability of large portfolio losses, which gives a general idea on the decay rate.

Theorem 5.3.3 *Consider the portfolio loss* (5.2.3) *and assume the following:*

- *Assumption 5.2.1 holds true,*
- $\mathbf{U} = (U_1, \ldots, U_n)$ *follows a partially nested Gumbel copulas defined by* (2.2.3)*, where* $\phi_j = (-\ln(t))^{\alpha_j}$*,*
- $\alpha_j > \alpha_0 > 1$*, for each* $1 \le j \le |\mathcal{W}|$*,*
- $\exp(-n\beta) = o(f_n)$ *for any* $\beta > 0$*.*

Then the relation

$$\liminf_{n\to\infty} \mathbb{P}(L_n > nb)/f_n \ge K \tag{5.3.9}$$

holds true for any fixed $b \in (0, \bar{c})$*, where* K *is a positive constant.*

5.4 Importance Sampling Simulations for Large Portfolio Loss under LT-Archimedean Copulas

In order to conduct an asymptotic analysis in Section 5.3, we assume the number of obligors in a credit portfolio tends to infinity. However, as we later show in the numerical examples section, the asymptotics given in Theorem 5.3.1 can produce inaccurate estimates unless the portfolio size is very large. Hence, Monte Carlo methods become a practical alternative to handle the estimation problem, although the problem of rare event simulation arises. Since the estimated default probability for large portfolio losses is usually small, naive Monte Carlo estimator is unstable and subject to high variability, unless the sample size is large enough. To generate more scenarios with large losses in simulation, we provide two different algorithms. In this section, we focus on the first method, which is an importance sampling (IS) algorithm based on a hazard rate twisting; see Juneja and Shahabuddin (2002) for an introduction on hazard rate twisting. In the next section, we discuss the second algorithm, which uses the idea on conditional Monte Carlo; see, e.g., Asmussen and Kroese (2006) and Asmussen (2018).

5.4.1 Preliminary of Importance Sampling

The probability of our interest is $\mathbb{P}(L_n > nb)$, where L_n can be considered as a linear combination of conditionally independent Bernoulli random variables $\{1_{\{U_i > 1 - l_i f_n\}}, i = 1, \ldots, n\}$. For each Bernoulli variable, the associated probability is denoted as $p_i, i = 1, \ldots, n$, which is a function of the generated variable V (We suppress this dependence for ease of notation, the explicit form for p_i is displayed in (5.3.3)). The simulation of above probability is usually conducted in two steps. In step 1, we simulate the common factor V using the density function $f_V(\cdot)$ and in step 2, we generate the corresponding Bernoulli random variables. Since the default probability p_i for each Bernoulli variable is small, estimation by naive Monte Carlo simulation becomes impractical due to the large number of samples needed, and therefore, one has to resort to variance reduction techniques.

As in other rare event simulation problems, importance sampling is often used as a technique that gets around this problem by placing further probability mass on the rare event of interest and then suitably unbiasing the resulting output. In our context, as discussed in the previous section, the tail behaviour of large portfolio loss highly depends on the tail distribution of V, i.e., the key to the occurrence of the large loss event corresponds to V taking large value. Then a good biasing distribution for random variable V should be more heavy-tailed than its original distribution, so that a larger probability could be assigned to the event that the average portfolio loss conditioned on V exceeds the level b. Let $\tilde{f}_V(\cdot)$ denote the new density function for V after IS. Besides applying IS to the distribution of V, one can also improve the efficiency of the numerical method by applying IS on the conditional probabilities. For this particular type of problem, there is a fairly well-established approach, see, e.g., Glasserman and Li (2005), which eventually increases the default probabilities by exponentially twisting the corresponding Bernoulli random variables.

Now suppose samples of V are drawn according to density function $\tilde{f}_V(\cdot)$, and each Bernoulli success probability p_i is replaced by some other probability $\tilde{p}_i$, for $i = 1, \ldots, n$. Let $\tilde{\mathbb{P}}$ denote the corresponding probability measure and $\tilde{\mathbb{E}}$ be the expectation under the measure $\tilde{\mathbb{P}}$. Then the following identity holds,

$$\mathbb{P}(L_n > nb) = \mathbb{E}\left[1_{\{L_n>nb\}}\right] = \tilde{\mathbb{E}}\left[1_{\{L_n>nb\}}\tilde{L}\right],$$

where $\tilde{L} = \frac{d\mathbb{P}}{d\tilde{\mathbb{P}}}$ is the Radon-Nikodym derivative of $\mathbb{P}$ and equals

$$\frac{f_V(V)}{\tilde{f}_V(V)} \prod_{j \le |\mathcal{W}|} \left(\frac{p_j}{\tilde{p}_j}\right)^{n_j Y_j} \left(\frac{1-p_j}{1-\tilde{p}_j}\right)^{n_j(1-Y_j)},$$

where $Y_j = 1_{\{U_j > 1 - l_j f_n\}}$, $n_j Y_j$ denotes the number of defaults in sub-portfolio j. We refer to $\tilde{\mathbb{P}}$ as the IS measure and $\tilde{L}$ as the unbiasing likelihood ratio.

The further questions arise are how to characterize the performance of the produced IS estimators and how to choose the IS measure appropriately. For rare event simulation, estimators are not only expected to provide a small variance but also a small relative error. Asymptotically, the best performance commonly observed in realistic situations is a bounded relative error; see, e.g., Asmussen and Kroese (2006) and McLeish (2010). We say a sequence of estimators $(1_{\{L_n>nb\}}\tilde{L} : n \ge 1)$ under probability measure $\tilde{\mathbb{P}}$ has bounded relative error if

$$\limsup_{n\to\infty} \frac{\sqrt{\tilde{\mathbb{E}}\left[1_{\{L_n>nb\}}\tilde{L}^2\right]}}{\mathbb{P}(L_n > nb)} < \infty.$$

A slightly weaker form criterion called asymptotically optimal is also widely used (see, e.g., Glasserman and Li (2005), Glasserman et al. (2007) and Glasserman et al. (2008)), if the following condition holds,

$$\lim_{n\to\infty} \frac{\log \tilde{\mathbb{E}}\left[1_{\{L_n>nb\}}\tilde{L}^2\right]}{\log \mathbb{P}(L_n > nb)} = 2.$$

This condition is equivalent to saying that $\lim_{n\to\infty} \tilde{\mathbb{E}}\left[1_{\{L_n>nb\}}\tilde{L}^2\right] / \mathbb{P}(L_n > nb)^{2-\varepsilon} = 0$, for every $\varepsilon > 0$. It is readily to check that bounded relative error implies asymptotically optimality.

5.4.2 Two-Step Importance Sampling

The First Step

As a first step in providing our IS algorithm for LT-Archimedean copulas, we apply IS to the distribution of random variable V. In Theorem 5.3.1, we assume the generator

$\phi \in \mathrm{RV}_\alpha(1)$, where ϕ^{-1} is the Laplace-Stieltjes transform of random variable V. Then by Karamata's Tauberian Theorem, see, e.g., Feller (1971), pp. 442–446, V is actually heavy-tailed with tail index $1/\alpha$. As noted in Asmussen et al. (2000), traditional exponential twisting approach cannot work directly for distributions with heavy tails, since a finite cumulant generating function in (5.4.6) does not exist when a positive twisting parameter is required. So an alternative method must be used. In this section we describe an IS algorithm to assign a larger probability to the event $\left\{V > \frac{v^*}{\phi(1-f_n)}\right\}$ by hazard rate twisting the original distribution of V. We prove this leads to an estimator that is asymptotically optimal.

Let us define the hazard function associated to the random variable V as

$$\mathcal{H}(x) = -\log(\bar{F}_V(x)).$$

By changing $\mathcal{H}(x)$ to $(1-\theta)\mathcal{H}(x)$ for some $0 < \theta < 1$, the tail distribution is changed to

$$\bar{F}_{V,\theta}(x) = (\bar{F}_V(x))^{1-\theta} = \exp((\theta - 1)\mathcal{H}(x)), \tag{5.4.1}$$

and the p.d.f. turns out to be

$$f_{V,\theta}(x) = (1-\theta)(\bar{F}_V(x))^{-\theta} f_V(x) = (1-\theta)\exp(\theta\mathcal{H}(x)) f_V(x). \tag{5.4.2}$$

This is similar to exponential twisting, except that the twisting rate is $\theta\mathcal{H}(x)$ rather than θx. By (5.4.1), one can also note that the tail of random variable V becomes heavier after the twisting.

The key, then, is finding the best parameter θ. By (5.4.2), the associated likelihood ratio for $f_V(x)/f_{V,\theta}(x)$ is $\frac{1}{1-\theta}\exp(-\theta\mathcal{H}(x))$, and this is upper bounded by

$$\frac{1}{1-\theta}\exp\left(-\theta\mathcal{H}\left(\frac{v^*}{\phi(1-f_n)}\right)\right) \tag{5.4.3}$$

on the set $\left\{V > \frac{v^*}{\phi(1-f_n)}\right\}$. We search for $\tilde{\theta}$ by minimizing the upper bound on the likelihood ratio, since this also minimizes the upper bound of the second moment of the estimator $1_{\{L_n > nb\}}\frac{f_V(V)}{f^*_{V,\theta}(V)}$. By taking the derivative on the upper bound (5.4.3) w.r.t. θ, we obtain

$$\tilde{\theta} = 1 - \frac{1}{\mathcal{H}\left(\frac{v^*}{\phi(1-f_n)}\right)}.$$

Then, the tail distribution in (5.4.1) corresponding to hazard rate twisting by $\tilde{\theta}$ equals

$$\bar{F}_{V,\tilde{\theta}}(x) = \exp\left(-\frac{\mathcal{H}(x)}{\mathcal{H}\left(\frac{v^*}{\phi(1-f_n)}\right)}\right). \tag{5.4.4}$$

Explicit form of (5.4.4) is usually difficult to derive, because the tail distribution for random variable V is only specified in a semiparametric way. Alternatively, we can replace the hazard function $\mathcal{H}(x)$ by $\tilde{\mathcal{H}}(x)$ where $\mathcal{H}(x) \sim \tilde{\mathcal{H}}(x)$ and $\tilde{\mathcal{H}}(x)$ is available in a closed form, Juneja et al. (2007) prove that estimators derived by such "asymptotic" hazard rate twisting method can achieve asymptotic optimality.

By Proposition B.1.9(1) of de Haan and Ferreira (2006), $\bar{F}_V \in \mathrm{RV}_{-1/\alpha}(\infty)$ implies $\mathcal{H}(x) \sim \frac{1}{\alpha}\log(x)$ as $x \to \infty$. This, along with (5.4.4), suggests that the tail distribution $\bar{F}_{V,\tilde{\theta}}$ should be close to

$$\bar{F}_{V,\tilde{\theta}}(x) \approx x^{-1/(\log v^* - \log\phi(1-f_n))}.$$

For considerably large n, we can even ignore the term $\log(v^*)$ to achieve further simplification. Hence, the corresponding p.d.f. can be taken as

$$\frac{1}{-\log\phi(1-f_n)} x^{\frac{1}{\log\phi(1-f_n)}-1},$$

which is indeed a Pareto distribution with shape parameter $-1/\log\phi(1-f_n)$. Now we define

$$f_V^*(x) = \begin{cases} f_V(x), & x < c_1, \\ \bar{F}_V(c_1) c_1^{-1/\log\phi(1-f_n)} \frac{1}{-\log\phi(1-f_n)} x^{\frac{1}{\log\phi(1-f_n)}-1} & x \geq c_1, \end{cases} \tag{5.4.5}$$

where c_1 is chosen to remain the ratio $f_V(x)/f_V^*(x)$ upper bounded by a constant for all x. Thus, the tail part of random variable V becomes heavier from the twisting, but the probability for small values remains unchanged.

Remark 5.4.1 *The role of c_1 is crucial for showing the asymptotic optimality of the algorithm, which is later seen in the proof of Lemma 5.4.1. Theoretically, its value relies on the explicit expression of the p.d.f. $f_V(x)$. For ease of implementation, one may fix c_1 to an arbitrary constant in realizing the algorithm, numeric results are not sensitive to its choice.*

The Second Step

We now proceed to apply exponential twisting to Bernoulli random variables $\{1_{\{U_i>1-l_if_n\}}, i = 1, \ldots, n\}$ conditional on the common factor V. A measure $\tilde{\mathbb{P}}$ is said to be an exponentially twisted measure of $\mathbb{P}$ by parameter θ, for some random variable X, if

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = \exp(\theta X - \Lambda_X(\theta)), \tag{5.4.6}$$

where $\Lambda_X(\theta) = \log \mathbb{E}[\exp(\theta X)]$ represents the cumulant generating function. Suppose random variable X has p.d.f. $f_X(x)$, then the exponential twisted density has the form $\exp(\theta x - \Lambda_X(\theta))f_X(x)$.

Now we deal with $p(v,i)$ as defined in (5.3.3) by conditioning on $V = \dfrac{v}{\phi(1-f_n)}$. In order to increase the conditional default probabilities, followed by the idea in Glasserman and Li (2005), we apply an exponential twist by choosing a parameter θ and taking

$$p_\theta(v,i) = \frac{p(v,i)e^{\theta c_i}}{1 + p(v,i)\left(e^{\theta c_i} - 1\right)},$$

where $p_\theta(v,i)$ denotes the θ-twisted probability conditional on $V = \dfrac{v}{\phi(1-f_n)}$. Note that $p_\theta(v,i)$ is a strictly increasing function if $\theta > 0$. With this new choice of conditional default probabilities $\{p_\theta(v,j) : j \le |\mathcal{W}|\}$, straight calculation shows that the likelihood ratio conditioning on V simplifies to

$$\prod_{j \le |\mathcal{W}|} \left(\frac{p(v,j)}{p_\theta(v,j)}\right)^{n_j Y_j} \left(\frac{1-p(v,j)}{1-p_\theta(v,j)}\right)^{n_j(1-Y_j)} = \exp\left(-\theta L_n|V + \Lambda_{L_n|V}(\theta)\right), \tag{5.4.7}$$

where

$$\Lambda_{L_n|V}(\theta) = \log \mathbb{E}\left[e^{\theta L_n}\middle| V = \frac{v}{\phi(1-f_n)}\right] = \sum_{j \le |\mathcal{W}|} n_j \log\left(1 + p(v,j)\left(e^{\theta c_j} - 1\right)\right) \tag{5.4.8}$$

is the cumulant generating function of L_n conditional on V. For any θ, the estimator

$$1_{\{L_n > nb\}} e^{-\theta L_n|V + \Lambda_{L_n|V}(\theta)}$$

is unbiased for $\mathbb{P}\left(L_n > nb \middle| V = \frac{v}{\phi(1-f_n)}\right)$ if probabilities $\{p_\theta(v,j) : j \le |\mathcal{W}|\}$ are used to generate L_n. Equation (5.4.7) actually shows that applying an exponential twist on the probabilities is equivalent to applying an exponential twist to L_n itself.

It remains to choose the parameter θ. A standard practice in important sampling is to select a parameter θ that minimizes the upper bound of the second moment of the estimator to reduce the variance. As we can see,

$$\mathbb{E}_\theta \left[1_{\{L_n > nb\}} e^{-2\theta L_n|V + 2\Lambda_{L_n|V}(\theta)} \middle| V = \frac{v}{\phi(1-f_n)} \right] \le e^{-2nb\theta + 2\Lambda_{L_n|V}(\theta)},$$

where $\mathbb{E}_\theta$ denotes expectation using the θ-twisted probabilities. The problem is then identical to find a parameter θ that maximizes $nb\theta - \Lambda_{L_n|V}(\theta)$. Straightforward calculation shows that

$$\Lambda'_{L_n|V}(\theta) = \sum_{j \le |\mathcal{W}|} n_j c_j p_\theta(v, j) = \mathbb{E}_\theta \left[L_n \middle| V = \frac{v}{\phi(1-f_n)} \right]. \tag{5.4.9}$$

By the strictly increasing property of $\Lambda'_{L_n|V}(\theta)$, the maximum is attained at

$$\theta^* = \begin{cases} \text{unique solution to } \Lambda'_{L_n|V}(\theta) = nb, & nb > \Lambda'_{L_n|V}(0), \\ 0, & nb \le \Lambda'_{L_n|V}(0). \end{cases} \tag{5.4.10}$$

By (5.4.9), the two cases in (5.4.10) are distinguished by the value of $\mathbb{E}\left[L_n \middle| V = \frac{v}{\phi(1-f_n)} \right]$. For the former case, our choice of twisting parameter θ^* shift the distribution of L_n so that the average portfolio loss is b; while for the latter case, the event $\{L_n > nb\}$ is not rare, so we use the original probabilities.

Algorithm

Now we are ready to present the algorithm. It consists of three stages. First, a sample of V is generated using hazard rate twisting. Depending on the value of V, samples of the Bernoulli variables $1_{\{U_i > 1 - l_i f_n\}}$ are generated in the second step, using either naive simulation (original probabilities) or importance sampling. The details on how to adjust conditional default probabilities have already been discussed in previous part. Finally we compute the portfolio loss L_n and return the estimator by incorporating the likelihood ratio.

Importance Sampling (IS) Algorithm

Step 1. Generate a sample of V using the density f_V^*.

Step 2. If the average portfolio loss under V, is greater than b, for each $i \leq n$, generate samples of $1_{\{U_i > 1 - l_i f_n\}}$ independent of each other using the unchanged probability $p_i^* = p_i = p\left(V\phi(1-f_n), i\right)$. Otherwise, we use $p_i^* = p_{\theta^*}(V\phi(1-f_n), i)$.

Step 3. Calculate the portfolio loss L_n and return the estimator

$$1_{\{L_n > nb\}} \frac{f_V(V)}{f_V^*(V)} \prod_{j \leq |\mathcal{W}|} \left(\frac{p_j}{p_j^*}\right)^{n_j Y_j} \left(\frac{1-p_j}{1-p_j^*}\right)^{n_j(1-Y_j)}, \tag{5.4.11}$$

where $n_j Y_j$ denotes the number of defaults in sub-portfolio j within in a single simulation run.

Let $\mathbb{P}^*$ and $\mathbb{E}^*$ denote the probability measure and expectation corresponding to this algorithm. Besides, the likelihood ratio is given by

$$L^* = \frac{f_V(V)}{f_V^*(V)} \prod_{j \leq |\mathcal{W}|} \left(\frac{p_j}{p_j^*}\right)^{n_j Y_j} \left(\frac{1-p_j}{1-p_j^*}\right)^{n_j(1-Y_j)}.$$

The following lemma is important in showing the efficiency of our IS Algorithm.

Lemma 5.4.1 *Under the same assumptions as in Theorem 5.3.1, we have*

$$\frac{\log \mathbb{E}^*\left[1_{\{L_n > nb\}} {L^*}^2\right]}{\log f_n} \to 2, \quad \text{as } n \to \infty.$$

In view of Theorem 5.3.1, which provides the asymptotic estimate of the tail probability $\mathbb{P}\left(L_n > nb\right)$, we conclude in the following theorem that our proposed algorithm is asymptotically optimal.

Theorem 5.4.1 *Under the same assumptions as in Theorem 5.3.1, we have*

$$\lim_{n \to \infty} \frac{\log \mathbb{E}^*\left[1_{\{L_n > nb\}} {L^*}^2\right]}{\log \mathbb{P}\left(L_n > nb\right)} = 2.$$

In other words, the importance sampling estimator in (5.4.11) *achieves zero variance on the logarithmic scale.*

Importance Sampling for Expected Shortfall

In risk management, one is usually interested in estimating the expected shortfall at a confidence level close to 1, which is again a rare event simulation. In this subsection, we discuss how to apply our proposed IS algorithm to estimate the expected shortfall.

First, note that the expected shortfall can be understood as follows,

$$\mathbb{E}\left[L_n | L_n > nb\right] = nb + \frac{\mathbb{E}\left[(L_n - nb)_+\right]}{\mathbb{P}\left(L_n > nb\right)}. \tag{5.4.12}$$

By involving the unbiasing likelihood ratio L^*, (5.4.12) now is equivalent to

$$nb + \frac{\mathbb{E}^*\left[(L_n - nb)_+ L^*\right]}{\mathbb{E}^*\left[1_{\{L_n > nb\}} L^*\right]},$$

where $\mathbb{E}^*$ is the expectation corresponding to the IS algorithm in Section 5.4.2. Suppose m i.i.d. samples $(L_n^1, \ldots, L_n^m)$ are generated under measure $\mathbb{P}^*$. Let L_i^* denote the corresponding likelihood ratio for each sample i. Then the IS estimator of the expected shortfall is given as

$$nb + \frac{\sum_{i=1}^m (L_n^i - nb)_+ L_i^*}{\sum_{i=1}^m 1_{\{L_n^i > nb\}} L_i^*}. \tag{5.4.13}$$

Note that the samples generated to estimate the numerator in (5.4.13) take positive value only when large losses occur. Therefore, one can expect the IS algorithm that works for estimating the probability of the event $\{L_n > nb\}$ should also work well in estimating $\mathbb{E}[L_n - nb]_+$. This is later confirmed by our numerical results.

5.5 Conditional Monte Carlo Simulations

In this section, we introduce the conditional Monte Carlo approach, which is another variance reduction technique. The broad motivation for the following efficient conditional Monte Carlo algorithm is given in Chan and Kroese (2010), where the authors derive a simple simulation algorithm to estimate the probability of large portfolio losses under the t copula.

5.5.1 Conditional Monte Carlo for Large Portfolio Loss under LT-Archimedean Copulas

By utilizing the stochastic representation (2.2.2) for LT-Archimedean and the asymptotic description in Theorem 5.3.1, the rare event $\{L_n > nb\}$ happens primarily when the random variable V takes large value, while $\mathbf{R} = (R_1, \ldots, R_n)$ generally has little influence on the occurrence of the rare event. This simply suggests that an approach by integrating out V analytically could lead to a substantial variance reduction.

Consider the following algorithm. Define

$$O_i = \frac{R_i}{\phi(1 - l_i f_n)}, i = 1, \ldots, n. \tag{5.5.1}$$

Recall that the individual obligor defaults if $U_i > 1 - l_i f_n$, i.e., $V > O_i$. Then, the portfolio loss in (5.2.3) can be rewritten as,

$$L_n = \sum_{i=1}^{n} c_i 1_{\{V > O_i\}}.$$

Let $O_{(1)}, \ldots, O_{(n)}$ be the order statistics of $O_1, \ldots, O_n$, and let $c_{(i)}$ denote the associated exposure at default with $O_{(i)}$. Then, one can check that the event $\{L_n > nb\}$ happens if and only if $V > O_{(k)}$, where $k = \min\{l : \sum_{i=1}^{l} c_{(i)} > nb\}$. Particularly, if $c_i \equiv c$ for all $i = 1, \ldots, n$, then $k = \lceil nb/c \rceil$. Now conditional on $\mathbf{R}$, we have

$$\mathbb{P}\left(L_n > nb|\mathbf{R}\right) = \mathbb{P}\left(V > O_{(k)}|\mathbf{R}\right) := S(\mathbf{R}). \tag{5.5.2}$$

We summarize above procedure in the following algorithm.

Conditional Monte Carlo (CondMC) Algorithm 1

Step 1. Generate independent standard exponential random variables $R_1, \ldots, R_n$.

Step 2. For $i = 1, \ldots, n$, transform R_i to O_i according to (5.5.1).

Step 3. Find $O_{(k)}$ and return the conditional Monte Carlo estimator $S(\mathbf{R})$ in (5.5.2).

We now show that the conditional Monte Carlo estimator has bounded relative error, a stronger notion of asymptotic optimality than that established in Theorem 5.4.1.

Lemma 5.5.1 *Under the same assumptions as in Theorem 5.3.1 except that $\frac{1}{n} = O(f_n)$, we have*

$$\limsup_{n\to\infty} \frac{\mathbb{E}\left[S^2(\mathbf{R})\right]}{f_n^2} < \infty.$$

In view of Theorem 5.3.1, we immediately obtain the following theorem concerning algorithm efficiency.

Theorem 5.5.1 *Under the same assumptions as in Lemma 5.5.1, we have*

$$\limsup_{n\to\infty} \frac{\sqrt{\mathbb{E}\left[S^2(\mathbf{R})\right]}}{\mathbb{P}\left(L_n > nb\right)} < \infty.$$

In other words, the conditional Monte Carlo estimator in (5.5.2) has bounded relative error.

5.5.2 Conditional Monte Carlo for Large Portfolio Loss under Nested Gumbel Copulas

For nested Gumbel copulas, the stochastic representation in (2.2.4) can be further simplified as follows,

$$U_{jl} = \phi_j^{-1}\left(\frac{R_{jl}}{\tilde{V}_j V_0^{\alpha_j/\alpha_0}}\right), \quad l = 1, \ldots, n_j \text{ and } j = 1, \ldots, |\mathcal{W}|.$$

Similarly to the idea for LT-Archimedean copulas, we define

$$O_{jl} = \frac{\left(R_{jl}/\tilde{V}_j\right)^{\alpha_0/\alpha_j}}{\phi_0(1 - l_j f_n)}, \quad l = 1, \ldots, n_j \text{ and } j = 1, \ldots, |\mathcal{W}|. \tag{5.5.3}$$

Note that the individual obligor defaults if $U_{jl} > 1 - l_j f_n$, i.e., $V_0 > O_{jl}$. Let $Q_{(1)} \le \cdots \le Q_{(n)}$ be the order statistics of O_{jl} for $l = 1, \ldots, n_j$ and $j = 1, \ldots, |\mathcal{W}|$. Then, the event $\{L_n > nb\}$ happens if and only if $V > Q_{(k)}$, where $k = \min\{l : \sum_{i=1}^{l} c_{(i)} > nb\}$. For notational convenience, let $\mathbf{R} = \left(R_{jl}\right)_{1\le l\le n_j, 1\le j\le|\mathcal{W}|}$ and $\mathbf{Y} = \left(\mathbf{R}, \tilde{V}_1, \ldots, \tilde{V}_{|\mathcal{W}|}\right)$. Conditional on $\mathbf{Y}$, we have

$$\mathbb{P}\left(L_n > nb|\mathbf{Y}\right) = \mathbb{P}\left(V_0 > Q_{(k)}|\mathbf{Y}\right) := S(\mathbf{Y}). \tag{5.5.4}$$

Above procedure is summarized in the following algorithm.

Conditional Monte Carlo (CondMC) Algorithm 2

Step 1. Generate independent standard exponential random variables $R_{jl}, l = 1, \ldots, n_j$ and $j = 1, \ldots, |\mathcal{W}|$.

Step 2. For each j and l, transform R_{jl} to O_{jl} according to (5.5.3).

Step 3. Sort O_{jl}, find $Q_{(k)}$ and return the conditional Monte Carlo estimator $S(\mathbf{Y})$ in (5.5.4).

Again we show the conditional Monte Carlo estimator for nested Gumbel copulas has bounded relative error.

Lemma 5.5.2 *Under the same assumptions as in Theorem 5.3.3 except that* $\frac{1}{n} = O(f_n)$, *we have*

$$\limsup_{n\to\infty} \frac{\mathbb{E}\left[S^2(\mathbf{Y})\right]}{f_n^2} < \infty.$$

In view of Theorem 5.3.3, we immediately obtain the following theorem concerning algorithm efficiency.

Theorem 5.5.2 *Under the same assumptions as in Lemma 5.5.2, we have*

$$\limsup_{n\to\infty} \frac{\sqrt{\mathbb{E}\left[S^2(\mathbf{Y})\right]}}{\mathbb{P}\left(L_n > nb\right)} < \infty.$$

In other words, the conditional Monte Carlo estimator in (5.5.4) *has bounded relative error.*

5.6 Numerical Results

In this section, we demonstrate the performance of our proposed estimators via simulations, and investigate its sensitivity to α, n and b. The broad conclusions drawn from following experiments are that our algorithms provide considerable variance reduction when compared to naive simulation, especially the CondMC algorithms, which supports our theoretical result that all proposed algorithms are all asymptotically optimal.

5.6.1 Gumbel Copulas

Motivated by the assumption that $\phi \in \mathrm{RV}_\alpha(1)$, we consider the Gumbel copula in our numerical experiment. The generator function of Gumbel copula is $\phi(t) = (-\ln(t))^\alpha$ with $\alpha > 1$. By varying α, the Gumbel copula covers from independence ($\alpha \to 1$) to comonotonicity ($\alpha \to \infty$).

In all the experiments below, only homogeneous portfolio are considered. However, it is worth to emphasize that the performance of our algorithms is not affected by assuming an inhomogeneous portfolio. Actually, Theorem 5.4.1 and Theorem 5.5.1 are proved under the setting of an inhomogeneous portfolio. To access the accuracy of the estimators, for each set of specified parameters, we generate 50,000 samples for our proposed algorithms, estimate the probability of large portfolio loss, and provide the relative error (in %), which is defined as the ratio of the estimator's standard deviation to the estimator. To be precise, if $\hat{p}$ is an unbiased estimator of $\mathbb{P}(L_n > nb)$, its relative error is defied as $\sqrt{\mathrm{Var}(\hat{p})}/\hat{p}$. We also report the variance reduction achieved by our proposed algorithms compared with naive simulation. For naive simulation, it is highly possible that the rare event would not be observed in any sample path with only 50,000 samples. Therefore, variance under naive simulation is estimated indirectly by realizing the fact that variance for Bernoulli(p) equals $p(1-p)$.

Table 5.2 shows the comparison of our IS algorithm and CondMC algorithm with naive simulation as α changes. The model parameters are chosen to be $n = 500, f_n = 1/n, b = 0.8, l_i = 0.5$ and $c_i = 1$ for each i. As can be seen in Table 5.2, the CondMC algorithm performs significantly better than the IS algorithm, and both algorithms outperform the naive simulation, especially when α is small and the probability of large portfolio losses becomes smaller.

	Prob. estimate		Relative error (%)		Variance reduction	
α	IS	CondMC	IS	CondMC	IS	CondMC
1.1	6.112×10^{-5}	6.208×10^{-5}	1.468	0.023	1,519	6,248,304
1.5	2.652×10^{-4}	2.726×10^{-4}	1.554	0.017	312	2,658,936
2	4.436×10^{-4}	4.457×10^{-4}	1.542	0.012	189	2,910,515
5	7.706×10^{-4}	7.815×10^{-4}	1.575	0.005	105	10,338,790

Table 5.2: Performance of the proposed algorithms for Gumbel copula under different values of α.

In Table 5.3, we perform the same comparison but now we vary b while keeping α

fixed at 1.5. Under the setting that $c = 1$, the parameter b actually controls the level of the proportion of obligors who default. As is clear from the table, when b increases, the estimated probability decreases and the variance reduction becomes larger.

	Prob. estimate		Relative error (%)		Variance reduction	
b	IS	CondMC	IS	CondMC	IS	CondMC
0.3	7.415×10^{-4}	7.437×10^{-4}	1.414	0.024	135	447,754
0.5	4.714×10^{-4}	4.776×10^{-4}	1.462	0.019	198	1,130,242
0.7	3.293×10^{-4}	3.306×10^{-4}	1.506	0.017	268	2,129,103
0.9	2.101×10^{-4}	2.151×10^{-4}	1.569	0.017	386	3,090,169

Table 5.3: Performance of the proposed algorithms for Gumbel copula under different values of b.

Table 5.4 provides the relative error and variance reduction of our algorithms compared with naive simulation as the number of obligors changes. All other parameters are identical to previous experiments by fixing $\alpha = 1.5, b = 0.8$. In the last column, we also derive the sharp asymptotic for the desired probability of large portfolio loss based on the expression in (5.3.5). Note that as n increases, both the accuracy of the sharp asymptotic and the variance reduction improve.

	Prob. estimate		Relative error (%)		Variance reduction		
n	IS	CondMC	IS	CondMC	IS	CondMC	Asymptotic
100	1.373×10^{-3}	1.381×10^{-3}	1.398	0.037	74	105,710	1.359×10^{-3}
250	5.372×10^{-4}	5.470×10^{-4}	1.487	0.023	168	670,052	5.436×10^{-4}
500	2.723×10^{-4}	2.727×10^{-4}	1.529	0.017	314	2,671,423	2.718×10^{-4}
1,000	1.356×10^{-4}	1.361×10^{-4}	1.640	0.012	582	10,608,750	1.359×10^{-4}

Table 5.4: Performance of the proposed algorithms for Gumbel copula together with the sharp asymptotic derived in Theorem 5.3.1 under different values of n.

5.6.2 Nested Gumbel Copulas

In this section, we assume the dependence structure follows a nested Gumbel copula. For simplicity, we restrict ourselves to a large portfolio with exactly two different types and each type has same number of obligors. In the following example, $\alpha_0 = 1.5$ is the parameter for the outer copula, and α_1, α_2 are the parameters for the inner copulas. For each set of specified parameters, again we generate 50,000 samples for our proposed algorithm.

(α_1, α_2)	Prob. estimate	Relative error (%)	Variance reduction
(3,3)	3.391×10^{-4}	0.289	7,081
(3,5)	3.554×10^{-4}	0.329	5,182
(3,7)	3.611×10^{-4}	0.335	4,943

Table 5.5: Performance of the proposed CondMC algorithm for nested Gumbel copula under different values of α_1 and α_2.

Table 5.5 provides the default probability under the nested Gumbel copula for varying values of α_1 and α_2. Other model parameters are identical to the case in Table 5.2 for direct comparison, i.e., $n = 500, b = 0.8, l_i = 0.5$ and $c_i = 1$. If one looks at the second row corresponding to $\alpha = 1.5$ in Table 5.2, the estimated probability of large portfolio losses is around 2.652×10^{-4}. Once stronger dependence is allowed within each sub-portfolio, the default probability is expected to increase, which can be seen from Table 5.5. From simulation point of view, our conditional Monte Carlo algorithm still significantly outperforms the naive simulation.

5.6.3 Constant Function f_n

By fixing the function $f_n \equiv 0.1$, we will redo the numerical studies in Table 5.2, Table 5.3 and Table 5.4 within this part. Other model parameters remain unchanged, i.e., $l_i = 0.5, c_i = 1$. Therefore, each obligor has identical default probability, i.e., $l_i f_n = 0.05$.

	Prob. estimate		Relative error (%)		Variance reduction	
α	IS	CondMC	IS	CondMC	IS	CondMC
1.1	3.271×10^{-3}	3.272×10^{-3}	1.192	0.023	43	111,634
1.5	1.423×10^{-2}	1.415×10^{-2}	1.389	0.017	8	49,453
2	2.308×10^{-2}	2.285×10^{-2}	1.373	0.012	5	55,376
5	3.881×10^{-2}	3.936×10^{-2}	1.302	0.005	3	204,492

Table 5.6: Performance of the proposed algorithms for Gumbel copula under different values of α.

	Prob. estimate		Relative error (%)		Variance reduction	
b	IS	CondMC	IS	CondMC	IS	CondMC
0.3	3.915×10^{-2}	3.938×10^{-2}	1.091	0.025	4	7,562
0.5	2.491×10^{-2}	2.501×10^{-2}	1.215	0.020	5	20,384
0.7	1.725×10^{-2}	1.721×10^{-2}	1.326	0.017	7	39,003
0.9	1.081×10^{-2}	1.113×10^{-2}	1.498	0.018	8	57,484

Table 5.7: Performance of the proposed algorithms for Gumbel copula under different values of b.

As we have mentioned in Section 5.3.1, with a constant function f_n, the probability of large portfolio losses is no longer rare. However, according to Table 5.6 and Table 5.7, our CondMC algorithm can still achieve significant variance reduction compared to naive simulation. This is mainly because the CondMC algorithm itself does not rely on our asymptotic result. In contrast, our IS algorithm seems not to have a remarkable computational advantage.

In Table 5.8, by changing the number of obligors in the credit portfolio, we provide the relative error and variance reduction of our algorithms compared with naive simulation. With the increase of n, a further improvement of the performance for our CondMC algorithm is noticed. While for our IS algorithm, similar observation does not exist. At the end of Table 5.8, based on (5.3.2), the limiting default probability when $n \to \infty$ is also provided. One may note that the convergence speed for our condMC algorithm is very fast.

	Prob. estimate		Relative error (%)		Variance reduction	
n	IS	CondMC	IS	CondMC	IS	CondMC
100	1.448×10^{-2}	1.430×10^{-2}	1.281	0.037	8	9,935
250	1.414×10^{-2}	1.419×10^{-2}	1.304	0.024	8	24,618
500	1.376×10^{-2}	1.415×10^{-2}	1.332	0.017	8	49,352
1,000	1.418×10^{-2}	1.413×10^{-2}	1.315	0.012	8	99,089
∞	1.411×10^{-2}					

Table 5.8: Performance of the proposed algorithms for Gumbel copula together with the limiting result derived in (5.3.2) under different values of n.

5.7 Proofs

5.7.1 Proofs for LT-Archimedean Copulas

We now prepare a series of lemmas for proving Theorem 5.3.1 and Theorem 5.3.2. The following is a restatement of Theorem 2 of Hoeffding (1963).

Lemma 5.7.1 *If $X_1, X_2, \ldots, X_n$ are independent and $a_i \le X_i \le b_i$ for $i = 1, \ldots, n$, then for $\varepsilon > 0$*

$$\mathbb{P}\left(\left|\bar{X}_n - \mathbb{E}\left[\bar{X}_n\right]\right| \ge \varepsilon\right) \le 2\exp\left(-\frac{2n^2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right),$$

with $\bar{X}_n = (X_1 + X_2 + \ldots + X_n)/n$.

Applying Lemma 5.7.1, we obtain the following inequality:

Lemma 5.7.2 *For any $\varepsilon > 0$ and any large M, there exists a constant $\beta > 0$ such that*

$$\mathbb{P}_v\left(\left|\frac{1}{n}\sum_{i=1}^n c_i 1_{\{U_i > 1 - l_i f_n\}} - r(v)\right| \ge \varepsilon\right) \le \exp(-n\beta),$$

uniformly for all $0 < v \le M$ and for all sufficiently large n, where $\mathbb{P}_v$ denotes the original probability measure conditioned on $V = \frac{v}{\phi(1-f_n)}$.

Proof. Note that U_i are conditionally independent on V. Then by Lemma 5.7.1, for every n,

$$\mathbb{P}_v\left(\left|\frac{1}{n}\sum_{i=1}^n c_i 1_{\{U_i>1-l_if_n\}} - \frac{1}{n}\sum_{i=1}^n c_i p(v,i)\right| \geq 2\varepsilon\right) \leq 2\exp\left(-\frac{8n^2\varepsilon^2}{\sum_{i=1}^n c_i^2}\right) \leq \exp(-n\beta), \tag{5.7.1}$$

where β is some unimportant constant not depending on n and v.

Using (5.7.1), to obtain the desired result, it suffices to show the existence of N, such for all $n \geq N$,

$$\left|\frac{1}{n}\sum_{i=1}^n c_i p(v,i) - r(v)\right| \leq \varepsilon \tag{5.7.2}$$

holds uniformly for all $v \leq M$. Recall that $r(v) = \sum_{j\leq|\mathcal{W}|} c_j w_j \tilde{p}(v,j)$.

Recall that n_j denotes the number of obligors in sub-portfolio j. Thus,

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^n c_i p(v,i) - r(v)\right| &= \left|\sum_{j\leq|\mathcal{W}|} c_j\left(p(v,j)\frac{n_j}{n} - \tilde{p}(v,j)w_j\right)\right| \\ &\leq \sum_{j\leq|\mathcal{W}|} c_j p(v,j)\left|\frac{n_j}{n} - w_j\right| \\ &+ \sum_{j\leq|\mathcal{W}|} c_j w_j\left|p(v,j) - \tilde{p}(v,j)\right| \\ &\leq \sum_{j\leq|\mathcal{W}|} c_j\left|\frac{n_j}{n} - w_j\right| + \bar{c}\max_{j\leq|\mathcal{W}|}\left|p(v,j) - \tilde{p}(v,j)\right| \end{aligned}\tag{5.7.3}$$

By Assumption 5.2.1, there exists N_1 satisfying $\sum_{j\leq|\mathcal{W}|} c_j\left|\frac{n_j}{n} - w_j\right| \leq \frac{\varepsilon}{2}$ for all $n \geq N_1$. For the second part of (5.7.3), by noting that $e^x \geq 1 + x$ for all $x \in \mathbb{R}$, we have

$$\begin{aligned}\left|p(v,j) - \tilde{p}(v,j)\right| &= \exp\left(-v\left(\frac{\phi(1-l_jf_n)}{\phi(1-f_n)} \wedge l_j^\alpha\right)\right)\left(1 - \exp\left(-v\left|\frac{\phi(1-l_jf_n)}{\phi(1-f_n)} - l_j^\alpha\right|\right)\right) \\ &\leq v\left|\frac{\phi(1-l_jf_n)}{\phi(1-f_n)} - l_j^\alpha\right| \\ &\leq M\left|\frac{\phi(1-l_jf_n)}{\phi(1-f_n)} - l_j^\alpha\right|. \end{aligned}\tag{5.7.4}$$

Since $\phi \in \mathrm{RV}_\alpha(1)$, there exists N_2 such that for all $n \geq N_2$, $\bar{c}\max_{j\leq|\mathcal{W}|, v\in A}\left|p(v,j) - \tilde{p}(v,j)\right| \leq \frac{\varepsilon}{2}$.

Combining the upper bound for both parts in (5.7.3) and letting $N = \max\{N_1, N_2\}$, (5.7.2) holds uniformly for all $v \le M$. The proof is then completed. ■

Proof of Theorem 5.3.1. To perform asymptotic analysis, let v_δ^* denote the unique solution to the equation $r(v) = b - \delta$. By using continuity and monotonicity of $r(v)$ in v, we have

$$v_\delta^* \to v^*$$

as $\delta \to 0$.

Fix $\delta > 0$. We decompose the probability of the event $\{L_n > nb\}$ into two terms as

$$\begin{aligned}\mathbb{P}\left(L_n > nb\right) &= \mathbb{P}\left(L_n > nb, V \le \frac{v_\delta^*}{\phi(1-f_n)}\right) + \mathbb{P}\left(L_n > nb, V > \frac{v_\delta^*}{\phi(1-f_n)}\right) \\ &= I_1 + I_2.\end{aligned}$$

The remaining part of proof will be divided into three steps. We first show that I_1 is asymptomatically negligible. Then we develop upper and lower bounds for I_2 with the second and third step.

Step 1. We show $I_1 = o(f_n)$. Note that for any $v \le v_\delta^*$, $r(v) \le b - \delta$. Thus, by Lemma 5.7.2, for all sufficiently large n, there exists a constant $\beta > 0$ such that

$$\mathbb{P}_v\left(L_n > nb\right) \le \mathbb{P}_v\left(\frac{1}{n}\sum_{i=1}^{n} c_i 1_{\{U_i > 1 - l_i f_n\}} - r(v) > \delta\right) \le \exp(-n\beta)$$

uniformly for all $v \le v_\delta^*$. So the same upper bound holds for I_1. Due to the condition on f_n, $I_1 = o(f_n)$.

Step 2. We now develop an asymptotic upper bound for I_2. Note that

$$I_2 \le \mathbb{P}\left(V > \frac{v_\delta^*}{\phi(1-f_n)}\right) = \bar{F}_V\left(\frac{v_\delta^*}{\phi(1-f_n)}\right).$$

Recall that ϕ^{-1} is the L-S transform for random variable V. Then from the condition $\phi \in \mathrm{RV}_\alpha(1)$ and Karamatas Tauberian theorem, we obtain

$$\begin{aligned}I_2 &\lesssim \bar{F}_V\left(\frac{v_\delta^*}{\phi(1-f_n)}\right), && \bar{F}_V \in \mathrm{RV}_{-1/\alpha}(\infty) \\ &\sim \frac{1-\phi^{-1}\left(\frac{\phi(1-f_n)}{v_\delta^*}\right)}{\Gamma(1-1/\alpha)}, && 1-\phi^{-1} \in \mathrm{RV}_{1/\alpha}(0) \\ &\sim f_n \frac{(v_\delta^*)^{-1/\alpha}}{\Gamma(1-1/\alpha)}.\end{aligned}$$

Letting $\delta \downarrow 0$, we obtain

$$I_2 \lesssim f_n \frac{(v^*)^{-1/\alpha}}{\Gamma(1-1/\alpha)}. \tag{5.7.5}$$

Step 3. We now develop an asymptotic lower bound for I_2. Denote $v^*_{\hat{\delta}}$ as the unique solution to the equation $r(v) = b + \delta$. Similarly, we have $v^*_{\hat{\delta}} \to v^*$ as $\delta \to 0$. It also follows from the monotonicity of $r(v)$ that $v^*_{\hat{\delta}} \geq v^*_{\delta}$. Thus,

$$I_2 \geq \mathbb{P}\left(L_n > nb, V > \frac{v^*_{\hat{\delta}}}{\phi(1-f_n)}\right).$$

Note that for any large $M > 0$, applying Lemma 5.7.2, it holds uniformly for $v \in \left[v^*_{\hat{\delta}}, M\right]$ that

$$\mathbb{P}_v\left(L_n > nb\right) \geq \mathbb{P}_v\left(\frac{1}{n}\sum_{i=1}^n c_i 1_{\{U_i > 1 - l_i f_n\}} - r(v) > -\delta\right) \to 1.$$

Hence,

$$\begin{aligned}
I_2 &\gtrsim \bar{F}_V\left(\frac{v^*_{\hat{\delta}}}{\phi(1-f_n)}\right) - \bar{F}_V\left(\frac{M}{\phi(1-f_n)}\right) \\
&\sim f_n \frac{(v^*_{\hat{\delta}})^{-1/\alpha}}{\Gamma(1-1/\alpha)} - f_n \frac{M^{-1/\alpha}}{\Gamma(1-1/\alpha)}.
\end{aligned}$$

Taking $M \to \infty$ followed by $\delta \to 0$, we get

$$I_2 \gtrsim f_n \frac{(v^*)^{-1/\alpha}}{\Gamma(1-1/\alpha)}. \tag{5.7.6}$$

Combining (5.7.5), (5.7.6) with Step 1 completes the proof of the theorem. ■

Proof of Theorem 5.3.2. We first note that the expected shortfall can be rewritten as follows:

$$\mathbb{E}\left[L_n | L_n > nb\right] = nb + n \frac{\int_b^\infty \mathbb{P}\left(L_n > nx\right) \mathrm{d}x}{\mathbb{P}\left(L_n > nb\right)}. \tag{5.7.7}$$

Using Theorem 5.3.1, in order to get the desired result, it suffices to show that

$$\int_b^\infty \mathbb{P}\left(L_n > nx\right) \mathrm{d}x \sim f_n \frac{\int_{v^*}^\infty r'(v) v^{-1/\alpha} \mathrm{d}v}{\Gamma(1-1/\alpha)}. \tag{5.7.8}$$

We decompose the left hand side of (5.7.8) into the following two terms

$$\begin{aligned}\int_b^\infty \mathbb{P}\left(L_n > nx\right)\mathrm{d}x &= \int_b^{\bar{c}} \mathbb{P}\left(L_n > nx\right)\mathrm{d}x + \int_{\bar{c}}^\infty \mathbb{P}\left(L_n > nx\right)\mathrm{d}x \\ &= J_1 + J_2,\end{aligned}$$

where $\bar{c} = \sum_{j\le|\mathcal{W}|} c_j w_j$. The remaining part of proof will be divided into three steps. We first show $\mathbb{P}\left(L_n > n\bar{c}\right)$ and J_2 are asymptomatically negligible with the first and second step. Then we develop the asymptotic for J_1.

Step 1. In this step, we show

$$\mathbb{P}\left(L_n > n\bar{c}\right) = o(f_n). \tag{5.7.9}$$

Fix an arbitrarily small $\delta > 0$. Proceeding in the same way as in step 1 in the proof of Theorem 5.3.1, for all sufficiently large n, there exists a constant $\beta > 0$ such that

$$\mathbb{P}\left(L_n > n\bar{c}, V \le \frac{r^{\leftarrow}(\bar{c}-\delta)}{\phi(1-f_n)}\right) \le \exp(-n\beta).$$

Due to the condition on f_n and letting $\delta \downarrow 0$, we have the desired result in (5.7.9).

Step 2. In this step, we show $J_2 = o(f_n)$. Note that J_2 can be rewritten as follows,

$$\begin{aligned}J_2 &= \mathbb{E}\left[\left(\frac{L_n}{n} - \bar{c}\right)_+\right] \\ &= \mathbb{E}\left[\left(\frac{L_n}{n} - \bar{c}\right) 1_{\{L_n > n\bar{c}\}}\right].\end{aligned}$$

Since $\frac{L_n}{n} < \max\limits_{j\le|\mathcal{W}|} c_j$, we have

$$J_2 \le \left(\max_{j\le|\mathcal{W}|} c_j - \bar{c}\right)\mathbb{P}\left(L_n > n\bar{c}\right).$$

It follows from (5.7.9) that $J_2 = o(f_n)$.

Step 3. To this end, we show

$$\lim_{n\to\infty} \int_b^{\bar{c}} \frac{\Gamma(1-1/\alpha)}{f_n}\mathbb{P}\left(L_n > nx\right)\mathrm{d}x = \int_{v^*}^\infty r'(v) v^{-1/\alpha}\mathrm{d}v.$$

First note that, for any $x \in [b, \bar{c}]$, by Theorem 5.3.1 and Step 1, there exists $\varepsilon > 0$ such that the following inequality

$$\frac{\Gamma(1-1/\alpha)}{f_n}\mathbb{P}(L_n > nx) \leq \frac{\Gamma(1-1/\alpha)}{f_n}\mathbb{P}(L_n > nb) \leq (r^{\leftarrow}(b))^{-1/\alpha} + \varepsilon$$

holds for all sufficiently large n. Applying the dominated convergence theorem, which is justified by the inequality above and the compactness of the interval $[b, \bar{c}]$, we obtain

$$\begin{aligned}
\lim_{n\to\infty}\int_b^{\bar{c}} \frac{\Gamma(1-1/\alpha)}{f_n}\mathbb{P}(L_n > nx)\,\mathrm{d}x &= \int_b^{\bar{c}} \left(\lim_{n\to\infty}\frac{\Gamma(1-1/\alpha)}{f_n}\mathbb{P}(L_n > nx)\right)\mathrm{d}x \\
&= \int_b^{\bar{c}} (r^{\leftarrow}(x))^{-1/\alpha}\mathrm{d}x \\
&= \int_{v^*}^{\infty} r'(v)v^{-1/\alpha}\mathrm{d}v.
\end{aligned}$$

The last equality is by changing the variable and let $v = r^{\leftarrow}(x)$.

Combing Step 2 and Step 3 completes the proof of the theorem. ■

5.7.2 Proofs for Nested Gumbel Copulas

We now prepare a key lemma for proving Theorem 5.3.3.

Lemma 5.7.3 *For any $\varepsilon > 0$ and any large M, there exists a constant $\beta > 0$ such that*

$$\mathbb{P}_{v_0,v_1,\ldots,v_{|\mathcal{W}|}}\left(\left|\frac{1}{n}\sum_{j\leq|\mathcal{W}|} c_j \sum_{l=1}^{n_j} 1_{\{U_{jl}>1-l_j f_n\}} - r(v_0,v_1,\ldots,v_{|\mathcal{W}|})\right| \geq \varepsilon\right) \leq \exp(-n\beta),$$

uniformly for all $v_j \leq M, j = 0,1,\ldots,|\mathcal{W}|$ and for all sufficiently large n, where $\mathbb{P}_{v_0,v_1,\ldots,v_{|\mathcal{W}|}}$ denotes the original probability measure conditioned on $V_0 = \frac{v_0}{\phi_0(1-f_n)}$ and $\tilde{V}_j = v_j$ for every $j = 1,\ldots,|\mathcal{W}|$.

Proof. Note that U_{jl} are conditionally independent given V_0 and $\tilde{V}_j, j = 1,\ldots,|\mathcal{W}|$. Then by Lemma 5.7.1, for every n,

$$\begin{aligned}
&\mathbb{P}_{v_0,v_1,\ldots,v_{|\mathcal{W}|}}\left(\left|\frac{1}{n}\sum_{j\leq|\mathcal{W}|} c_j \sum_{l=1}^{n_j} 1_{\{U_{jl}>1-l_j f_n\}} - \frac{1}{n}\sum_{j\leq|\mathcal{W}|} c_j n_j p(v_0,v_j,j)\right| \geq 2\varepsilon\right) \\
\leq& 2\exp\left(-\frac{8n^2\varepsilon^2}{\sum_{i=1}^n c_i^2}\right) \leq \exp(-n\beta), \qquad (5.7.10)
\end{aligned}$$

where β is some constant not depending on n, v_0 and v_j for $j = 1, \dots, |\mathcal{W}|$.

Recall that $r(v_0, v_1, \dots, v_{|\mathcal{W}|}) = \sum_{j \le |\mathcal{W}|} c_j w_j \tilde{p}(v_0, v_j, j)$. Consequently, to obtain the desired result, it suffices to show the existence of N, such for all $n \ge N$,

$$\left| \frac{1}{n} \sum_{j \le |\mathcal{W}|} c_j n_j p(v_0, v_j, j) - r(v_0, v_1, \dots, v_{|\mathcal{W}|}) \right| \le \varepsilon \tag{5.7.11}$$

holds uniformly for all $v_j \le M, j = 0, \dots, |\mathcal{W}|$.

Similarly to the derivation of (5.7.3) in the proof of Lemma 5.7.2, we obtain that

$$\left| \frac{1}{n} \sum_{j \le |\mathcal{W}|} c_j n_j p(v_0, v_j, j) - r(v_0, v_1, \dots, v_{|\mathcal{W}|}) \right| \le \sum_{j \le |\mathcal{W}|} c_j \left| \frac{n_j}{n} - w_j \right| + \bar{c} \max_{j \le |\mathcal{W}|} |p(v_0, v_j, j) - \tilde{p}(v_0, v_j, j)| . \tag{5.7.12}$$

By Assumption 5.2.1, there exists N_1 satisfying $\sum_{j \le |\mathcal{W}|} c_j \left| \frac{n_j}{n} - w_j \right| \le \frac{\varepsilon}{2}$ for all $n \ge N_1$. For the second part of (5.7.12), the upper bound is easy to establish in a similar manner as (5.7.4), that is,

$$|p(v_0, v_j, j) - \tilde{p}(v_0, v_j, j)| \le M^{1+\alpha_j/\alpha_0} \left| \frac{\phi_j(1 - l_j f_n)}{\phi_j(1 - f_n)} - l_j^{\alpha_j} \right| .$$

Since $\phi_j \in \mathrm{RV}_{\alpha_j}(1)$, there exists N_2 such that for all $n \ge N_2$, $\bar{c} \max_{j \le |\mathcal{W}|} |p(v_0, v_j, j) - \tilde{p}(v_0, v_j, j)| \le \frac{\varepsilon}{2}$.

Combining the upper bound for both parts in (5.7.12) and letting $N = \max\{N_1, N_2\}$, (5.7.11) holds uniformly for all $v_j \le M, j = 0, \dots, |\mathcal{W}|$. The proof is then completed. ■

Proof of Theorem 5.3.3. Recall $v_{\hat{\delta}}^*$ is defined in Step 3 of the proof of Theorem 5.3.1, where

$$r(v_{\hat{\delta}}^*) = \sum_{j \le |\mathcal{W}|} c_j w_j \left(1 - \exp\left(-v_{\hat{\delta}}^* l_j^{\alpha_0}\right)\right) = b + \delta.$$

Compared with $r(v_0, v_1, \dots, v_{|\mathcal{W}|})$ defined in (5.3.8), one can easily check that if $v_j v_0^{\alpha_j/\alpha_0} l_j^{\alpha_j} > v_{\hat{\delta}}^* l_j^{\alpha_0}$, $r(v_0, v_1, \dots, v_{|\mathcal{W}|}) > b + \delta$ as well. A trivial case is $v_0 > v_{\hat{\delta}}^*$ and $v_j > \left(v_{\hat{\delta}}^*\right)^{1-\alpha_j/\alpha_0} l_j^{\alpha_0 - \alpha_j}$ for every $j = 1, \dots, |\mathcal{W}|$.

Then the default probability of the event $\{L_n > nb\}$ has the following lower bound,

$$\mathbb{P}\left(L_n > nb, V_0 \in \left(\frac{v_{\hat{\delta}}^*}{\phi_0(1 - f_n)}, \frac{M}{\phi_0(1 - f_n)} \right), \tilde{V}_j \in \left(\left(v_{\hat{\delta}}^*\right)^{1-\alpha_j/\alpha_0} l_j^{\alpha_0 - \alpha_j}, M \right), j = 1, \dots, |\mathcal{W}| \right).$$

Taking limit and by Lemma 5.7.3, we obtain

$$\mathbb{P}\left(L_n > nb\right) \gtrsim f_n \left(\frac{(v_{\hat{\delta}}^*)^{-1/\alpha_0}}{\Gamma(1-1/\alpha_0)} - \frac{M^{-1/\alpha_0}}{\Gamma(1-1/\alpha_0)} \right) \prod_{j \le |\mathcal{W}|} \left(\bar{F}_j \left(\left(v_{\hat{\delta}}^*\right)^{1-\alpha_j/\alpha_0} l_j^{\alpha_0 - \alpha_j} \right) - \bar{F}_j(M) \right),$$

where $\bar{F}_j$ is the tail probability for $\tilde{V}_j, j = 1, \ldots, |\mathcal{W}|$. Let $M \to \infty$ followed by $\delta \to 0$. The following inequality holds,

$$\mathbb{P}\left(L_n > nb\right) \gtrsim f_n \frac{(v^*)^{-1/\alpha_0}}{\Gamma(1-1/\alpha_0)} \prod_{j \le |\mathcal{W}|} \left(\bar{F}_j \left(\left(v^*\right)^{1-\alpha_j/\alpha_0} l_j^{\alpha_0 - \alpha_j} \right) \right) \gtrsim K f_n,$$

where $K > 0$ is some constant. ■

5.7.3 Proofs for Algorithm Efficiency

Lemma 5.7.4 and 5.7.5 will be used in proving Lemma 5.4.1.

Lemma 5.7.4 *For sufficiently large n, there exists a constant C such that*

$$\frac{f_V(x)}{f_V^*(x)} \le C\left(-\log \phi(1-f_n)\right) \tag{5.7.13}$$

for all x, where $f_V^*(x)$ is defined in (5.4.5).

Proof. By definition of $f_V^*(x)$, the ratio $\frac{f_V(x)}{f_V^*(x)}$ equals 1 for $x < c_1$. Hence, to show (5.7.13), it suffices to show the existence of a constant C for all $x \ge c_1$.

Note that when $x \ge c_1$,

$$\frac{f_V(x)}{f_V^*(x)} = \frac{f_V(x)}{\bar{F}_V(c_1)} c_1^{1/\log \phi(1-f_n)} \left(-\log \phi(1-f_n)\right) x^{1 - \frac{1}{\log \phi(1-f_n)}}. \tag{5.7.14}$$

For any small $\varepsilon < 1/\alpha$, by properly selecting c_1 based on ε and Potter's bounds (see Lemma

2.1.1(1)), it holds for all $x \geq c_1$ such that

$$
\begin{aligned}
\frac{\bar{F}_V(c_1)}{f_V(x)} &= \int_{c_1}^{\infty} \frac{f_V(t)}{f_V(x)} \mathrm{d}t \\
&> (1-\varepsilon) \left(\left(\int_{c_1}^{\infty} (t/x)^{-\frac{1}{\alpha}-1-\varepsilon} \mathrm{d}t \right) \wedge \left(\int_{c_1}^{\infty} (t/x)^{-\frac{1}{\alpha}-1+\varepsilon} \mathrm{d}t \right) \right) \\
&> (1-\varepsilon) x^{1/\alpha+1} c_1^{-1/\alpha} \left(\left(\frac{(x/c_1)^{\varepsilon}}{1/\alpha+\varepsilon} \right) \wedge \left(\frac{(x/c_1)^{-\varepsilon}}{1/\alpha-\varepsilon} \right) \right) \\
&> (1-\varepsilon) x^{1/\alpha+1} c_1^{-1/\alpha} \left(\frac{(x/c_1)^{-\varepsilon}}{1/\alpha+\varepsilon} \right). \qquad (5.7.15)
\end{aligned}
$$

From (5.7.15), an upper bound for (5.7.14) is derived as follows,

$$
\begin{aligned}
\frac{f_V(x)}{f_V^*(x)} &< \frac{1/\alpha+\varepsilon}{1-\varepsilon} \left(-\log \phi(1-f_n)\right) \left(\frac{x}{c_1} \right)^{-1/\alpha - \frac{1}{\log \phi(1-f_n)} + \varepsilon} \qquad (5.7.16) \\
&\leq \frac{1/\alpha+\delta}{1-\delta} \left(-\log \phi(1-f_n)\right),
\end{aligned}
$$

which yields our desired result by letting $C = \frac{1/\alpha+\varepsilon}{1-\varepsilon}$. The last inequality is due to the fact that $x/c_1 \geq 1$. ■

Lemma 5.7.5 *If $\phi \in \mathrm{RV}_\alpha(1)$ for some $\alpha > 1$ and f_n is a positive deterministic function converging to 0 as $n \to \infty$, then*

$$
\log \phi(1-f_n) \sim \alpha \log(f_n). \qquad (5.7.17)
$$

Proof. By Proposition B.1.9(1) of de Haan and Ferreira (2006), $\phi \in \mathrm{RV}_\alpha(1)$ implies that

$$
\log \phi(1-x) \sim \alpha \log(x)
$$

as $x \to 0$. This completes the proof. ■

Proof of Lemma 5.4.1. Let

$$
\hat{L} = \prod_{j \leq |\mathcal{W}|} \left(\frac{p_j}{p_j^*} \right)^{n_j Y_j} \left(\frac{1-p_j}{1-p_j^*} \right)^{n_j(1-Y_j)}.
$$

Note that if $\mathbb{E}\left[L_n \middle| V = \frac{v}{\phi(1-f_n)}\right] < nb$, $p_j^* = p_{\theta^*}(V\phi(1-f_n), j)$ where θ^* is chosen by solving $\Lambda'_{L_n|V}(\theta) = nb$; otherwise $p_j^* = p\,(V\phi(1-f_n), j)$ by setting $\theta^* = 0$. Besides, (5.4.7) shows $\hat{L}$ can be written as follows.

$$\hat{L} = \exp(-\theta^* L_n|V + \Lambda_{L_n|V}(\theta^*)). \tag{5.7.18}$$

Then it follows that, for any v,

$$1_{\left\{L_n>nb, V=\frac{v}{\phi(1-f_n)}\right\}}\hat{L} \leq 1_{\left\{L_n>nb, V=\frac{v}{\phi(1-f_n)}\right\}}\exp(-\theta^* nb + \Lambda_{L_n|V}(\theta^*)) \qquad \text{a.s.}$$

Since $\Lambda_{L_n|V}(\theta)$ is a strictly convex function, one can observe that $-\theta nb + \Lambda_{L_n|V}(\theta)$ is minimized at θ^* and equals 0 at $\theta = 0$. Hence, the following relation

$$1_{\left\{L_n>nb, V=\frac{v}{\phi(1-f_n)}\right\}}\hat{L} \leq 1_{\left\{L_n>nb, V=\frac{v}{\phi(1-f_n)}\right\}} \qquad \text{a.s.} \tag{5.7.19}$$

holds for any v.

To prove the theorem, now we re-express

$$\begin{aligned}\mathbb{E}^*\left[1_{\{L_n>nb\}}{L^*}^2\right] &= \mathbb{E}^*\left[1_{\left\{L_n>nb, V\leq\frac{v_\delta^*}{\phi(1-f_n)}\right\}}{L^*}^2\right] + \mathbb{E}^*\left[1_{\left\{L_n>nb, V>\frac{v_\delta^*}{\phi(1-f_n)}\right\}}{L^*}^2\right]\\ &= K_1 + K_2,\end{aligned}$$

where v_δ^* is the unique solution to the equation $r(v) = b - \delta$.

The remaining part of proof will be divided into three steps.

Step 1. In this step, we show

$$K_1 = o(f_n^2). \tag{5.7.20}$$

By Lemma 5.7.4, for sufficiently large n, there exists a finite positive constant C such that

$$\frac{f_V(v)}{f_V^*(v)} \leq C\left(-\log\phi(1-f_n)\right)$$

for all v. From (5.7.19), it then follows that

$$1_{\left\{L_n>nb, V\leq\frac{v_\delta^*}{\phi(1-f_n)}\right\}}{L^*}^2 \leq C\left(-\log\phi(1-f_n)\right)\left(1_{\left\{L_n>nb, V\leq\frac{v_\delta^*}{\phi(1-f_n)}\right\}}L^*\right) \qquad \text{a.s.}$$

Therefore, K_1 is upper bounded by

$$\begin{aligned}\mathbb{E}^*\left[1_{\left\{L_n>nb,V\leq\frac{v_\delta^*}{\phi(1-f_n)}\right\}}{L^*}^2\right] &\leq C\left(-\log\phi(1-f_n)\right)\left(\mathbb{E}^*\left[1_{\left\{L_n>nb,V\leq\frac{v_\delta^*}{\phi(1-f_n)}\right\}}L^*\right]\right)\\ &= C\left(-\log\phi(1-f_n)\right)\left(\mathbb{P}\left(L_n>nb, V\leq\frac{v_\delta^*}{\phi(1-f_n)}\right)\right)\\ &\leq C\left(-\log\phi(1-f_n)\right)\exp(-\beta n).\end{aligned}$$

The last step is due to step 1 in the proof of Theorem 5.3.1. And (5.7.20) is verified by Lemma 5.7.5 and noting that f_n has a sub-exponential decay rate.

Step 2. To this end, we show that

$$\limsup_{n\to\infty}\frac{\log K_2}{\log f_n}\leq 2. \tag{5.7.21}$$

By Jensen's inequality,

$$\begin{aligned}\mathbb{E}^*\left[1_{\left\{L_n>nb,V>\frac{v_\delta^*}{\phi(1-f_n)}\right\}}{L^*}^2\right] &\geq \left(\mathbb{E}^*\left[1_{\left\{L_n>nb,V>\frac{v_\delta^*}{\phi(1-f_n)}\right\}}L^*\right]\right)^2\\ &= \left(\mathbb{P}\left(L_n>nb, V>\frac{v_\delta^*}{\phi(1-f_n)}\right)\right)^2\\ &\sim f_n^2\left(\frac{(v^*)^{-1/\alpha}}{\Gamma(1-1/\alpha)}\right)^2.\end{aligned}$$

Then (5.7.21) follows by applying the logarithm function on both sides and using the fact that $\log(f_n)<0$ for all sufficiently large n.

Step 3. To this end, we show that

$$\liminf_{n\to\infty}\frac{\log K_2}{\log f_n}\geq 2. \tag{5.7.22}$$

First note that, on the set $\left\{L_n>nb, V>\frac{v_\delta^*}{\phi(1-f_n)}\right\}$, the likelihood ratio L^* is upper bounded by $\frac{f_V(v)}{f_V^*(v)}$ and hence by (5.7.16), with sufficiently large n, it holds for all $v>\frac{v_\delta^*}{\phi(1-f_n)}$ that

$$\begin{aligned}\frac{f_V(v)}{f_V^*(v)} &< C\left(-\log\phi(1-f_n)\right)\left(\frac{v_\delta^*/\phi(1-f_n)}{c_1}\right)^{-1/\alpha-\frac{1}{\log\phi(1-f_n)}+\varepsilon}\\ &< C\left(-\log\phi(1-f_n)\right)\left(\frac{v_\delta^*}{c_1}\right)^{-1/\alpha+\varepsilon}\phi(1-f_n)^{1/\alpha-\varepsilon}.\end{aligned}$$

Multiplying it with the indicator and taking expectation under $\mathbb{E}^*$, we obtain

$$\mathbb{E}^*\left[1_{\left\{L_n>nb,V>\frac{v_\delta^*}{\phi(1-f_n)}\right\}}L^{*^2}\right] \le C^2\left(-\log\phi(1-f_n)\right)^2\left(\frac{v_\delta^*}{c_1}\right)^{-2/\alpha+2\varepsilon}\phi(1-f_n)^{2/\alpha-2\varepsilon}. \tag{5.7.23}$$

Then, taking logarithms on both sides, dividing by $\log f_n$ and by Lemma 5.7.5, we obtain

$$\liminf_{n\to\infty}\frac{\log\mathbb{E}^*\left[1_{\left\{L_n>nb,V>\frac{v_\delta^*}{\phi(1-f_n)}\right\}}L^{*^2}\right]}{\log f_n} \ge 2-2\alpha\varepsilon.$$

Finally, (5.7.22) is yield by letting $\varepsilon\downarrow 0$.

Combining Step 1, Step 2 and Step 3, the desired result asserted in the theorem is obtained. ■

Lemma 5.7.6 below will be used in proving Lemma 5.5.1 and Lemma 5.5.2.

Lemma 5.7.6 *Let $R_1,\dots,R_n$ be an i.i.d. sequence of standard exponential random variables. Suppose $R_{(k)}$ is the kth order statistic and $\lim_{n\to\infty}\frac{k}{n}=a<1$. Then, for every $\varepsilon>0$, there exists a constant $\beta>0$ such that the following inequality*

$$\mathbb{P}\left(\left|R_{(k)}-\log\left(\frac{1}{1-a}\right)\right|\ge\varepsilon\right)\le\frac{\beta}{n}.$$

holds for all sufficiently large n.

Proof. For i.i.d. standard exponential random variables $R_i, i=1,\dots,n$, it follows from Rényi (1953) that

$$R_{(k)}\overset{d}{=}\sum_{j=1}^{k}\frac{R_j}{n-j+1}.$$

Then,

$$\mathbb{E}[R_{(k)}]=\sum_{j=1}^{k}\frac{1}{n-j+1}\to\log\left(\frac{1}{1-a}\right),\quad\text{as } n\to\infty, \tag{5.7.24}$$

and

$$\mathrm{Var}[R_{(k)}]=\sum_{j=1}^{k}\left(\frac{1}{n-j+1}\right)^2\sim\frac{a}{1-a}\frac{1}{n},\quad\text{as } n\to\infty. \tag{5.7.25}$$

By Chebyshev's inequality, it follows that, for every $n > 0$,

$$\mathbb{P}\left(|R_{(k)} - \mathbb{E}[R_{(k)}]| \geq \varepsilon\right) \leq \frac{\mathrm{Var}[R_{(k)}]}{\varepsilon^2}.$$

Due to (5.7.24) and (5.7.25), there exists N, such that for all $n \geq N$,

$$\mathbb{P}\left(\left|R_{(k)} - \log\left(\frac{1}{1-a}\right)\right| \geq \varepsilon\right) \leq \frac{\beta}{n},$$

where β only depends on ε and a. ∎

Proof of Lemma 5.5.1. Recall that $O_i = \frac{R_i}{\phi(1-l_i f_n)}$, for all $i = 1, \dots, n$. Then the order statistic $O_{(k)}$ is almost surely lower bounded by

$$\frac{R_{(k)}}{\phi\left(1 - \max\limits_{j \leq |\mathcal{W}|} l_j f_n\right)}.$$

Since $k = \min\{l : \sum_{i=1}^{l} c_{(i)} > nb\}$, we have

$$\liminf_{n\to\infty} \frac{k}{n} \geq \frac{b}{\max\limits_{j \leq |\mathcal{W}|} c_j} := b'. \tag{5.7.26}$$

Fix $\varepsilon > 0$. For all sufficiently large n, $\mathbb{E}\left[S^2(\mathbf{R})\right]$ can be bounded as follows,

$$\begin{aligned}
\mathbb{E}\left[S^2(\mathbf{R})\right] &\leq \mathbb{E}\left[\mathbb{P}\left(V > \frac{R_{(\lfloor nb' \rfloor)}}{\phi\left(1 - \max\limits_{j \leq |\mathcal{W}|} l_j f_n\right)}\right)^2\right] \\
&\leq \mathbb{E}\left[\left(\mathbb{P}\left(V > \frac{R_{(\lfloor nb' \rfloor)}}{\phi\left(1 - \max\limits_{j \leq |\mathcal{W}|} l_j f_n\right)}, R_{(\lfloor nb' \rfloor)} \geq \log\left(\frac{1}{1-b'}\right) - \varepsilon\right)\right.\right. \\
&\quad \left.\left. + \mathbb{P}\left(V > \frac{R_{(\lfloor nb' \rfloor)}}{\phi\left(1 - \max\limits_{j \leq |\mathcal{W}|} l_j f_n\right)}, R_{(\lfloor nb' \rfloor)} < \log\left(\frac{1}{1-b'}\right) - \varepsilon\right)\right)^2\right] \\
&\leq \left(\mathbb{P}\left(V > \frac{\log\left(\frac{1}{1-b'}\right) - \varepsilon}{\phi\left(1 - \max\limits_{j \leq |\mathcal{W}|} l_j f_n\right)}\right) + \mathbb{P}\left(R_{(\lfloor nb' \rfloor)} < \log\left(\frac{1}{1-b'}\right) - \varepsilon\right)\right)^2 .
\end{aligned}$$

Then,

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\mathbb{E}\left[S^2(\mathbf{R})\right]}{f_n^2} &\leq \left(\limsup_{n\to\infty} \frac{\mathbb{P}\left(V > \frac{\log\left(\frac{1}{1-b'}\right)-\varepsilon}{\phi\left(1 - \max\limits_{j \leq |\mathcal{W}|} l_j f_n\right)}\right)}{f_n} + \limsup_{n\to\infty} \frac{\mathbb{P}\left(R_{(\lfloor nb' \rfloor)} < \log\left(\frac{1}{1-b'}\right) - \varepsilon\right)}{f_n}\right)^2 \\
&\leq \left(\max_{j \leq |\mathcal{W}|} l_j \frac{\left(\log\left(\frac{1}{1-b'}\right) - \varepsilon\right)^{-1/\alpha}}{\Gamma(1 - 1/\alpha)} + M\right)^2 < \infty.
\end{aligned}$$

The last step is due to the regular variation of V, Lemma 5.7.6 and the condition that $\frac{1}{n} = O(f_n)$. ■

Proof of Lemma 5.5.2. Recall that

$$O_{jl} = \frac{\left(R_{jl}/\tilde{V}_j\right)^{\alpha_0/\alpha_j}}{\phi_0(1 - l_j f_n)},$$

for all $l = 1, \ldots, n_j$ and $j = 1, \ldots, |\mathcal{W}|$. Let $R_{(j,\lfloor n_j b' \rfloor)}$ be the $\lfloor n_j b' \rfloor$-th order statistic of $(R_{jl})_{1 \le l \le n_j}$ for any $j = 1, \ldots, |\mathcal{W}|$, where b' is defined in (5.7.26), and let $1/\tilde{V} = \min_{j \le |\mathcal{W}|} \left(1/\tilde{V}_j\right)^{\alpha_0/\alpha_j}$. The order statistic $Q_{(k)}$ is therefore almost surely lower bounded by

$$\frac{\min_{j \le |\mathcal{W}|} \left\{ R_{(j,\lfloor n_j b' \rfloor)}^{\alpha_0/\alpha_j} \right\} / \tilde{V}}{\phi_0 \left(1 - \max_{j \le |\mathcal{W}|} l_j f_n)\right)}.$$

Fix $\varepsilon > 0$. For all sufficiently large n, $\mathbb{E}\left[S^2(\mathbf{Y})\right]$ can be bounded as follows,

$$\begin{aligned}
\mathbb{E}\left[S^2(\mathbf{Y})\right] &\le \mathbb{E}\left[\mathbb{P}\left(V_0 > \frac{\min_{j \le |\mathcal{W}|} \left\{ R_{(j,\lfloor n_j b' \rfloor)}^{\alpha_0/\alpha_j} \right\} / \tilde{V}}{\phi_0 \left(1 - \max_{j \le |\mathcal{W}|} l_j f_n)\right)}\right)^2\right] \\
&= \mathbb{E}\left[\mathbb{P}\left(V_0 \tilde{V} > \frac{\min_{j \le |\mathcal{W}|} \left\{ R_{(j,\lfloor n_j b' \rfloor)}^{\alpha_0/\alpha_j} \right\}}{\phi_0 \left(1 - \max_{j \le |\mathcal{W}|} l_j f_n)\right)}\right)^2\right] \qquad (5.7.27) \\
&\le \left(\mathbb{P}\left(V_0 \tilde{V} > \frac{\log\left(\frac{1}{1-b'}\right)^{\alpha_0/\alpha_{min}} \wedge \log\left(\frac{1}{1-b'}\right)^{\alpha_0/\alpha_{max}} - \varepsilon}{\phi_0 \left(1 - \max_{j \le |\mathcal{W}|} l_j f_n)\right)}\right)\right. \\
&\quad \left. + \mathbb{P}\left(\min_{j \le |\mathcal{W}|} \left\{ R_{(j,\lfloor n_j b' \rfloor)}^{\alpha_0/\alpha_j} \right\} < \log\left(\frac{1}{1-b'}\right)^{\alpha_0/\alpha_{min}} \wedge \log\left(\frac{1}{1-b'}\right)^{\alpha_0/\alpha_{max}} - \varepsilon\right)\right)^2, \qquad (5.7.28)
\end{aligned}$$

where $\alpha_{min}, \alpha_{max}$ denotes the minimum and maximum of the set $\{\alpha_j, j = 1, \ldots, |\mathcal{W}|\}$.

Since $V_0 \in \mathrm{RV}_{-1/\alpha_0}(\infty), \tilde{V}_j \in \mathrm{RV}_{-\alpha_0/\alpha_j}(\infty)$, we observe that $\tilde{V} \in \mathrm{RV}_{-1}(\infty)$ and therefore the well-known Breiman's result implies

$$\mathbb{P}(V_0 \tilde{V} > x) \sim \mathbb{E}(\tilde{V}^{1/\alpha_0}) \mathbb{P}(V_0 > x). \qquad (5.7.29)$$

See Cline and Samorodnitsky (1994) for a complete proof and Jessen and Mikosch (2006) for details about the elementary functions of regularly varying random variables, such as products, minima, maxima. By (5.7.29), as $n \to \infty$, the first part of (5.7.28) is equivalent to

$$\mathbb{E}(\tilde{V}^{1/\alpha_0})\mathbb{P}\left(V_0 > \frac{\log\left(\frac{1}{1-b'}\right)^{\alpha_0/\alpha_{min}} \wedge \log\left(\frac{1}{1-b'}\right)^{\alpha_0/\alpha_{max}} - \varepsilon}{\phi_0\left(1 - \max\limits_{j \le |\mathcal{W}|} l_j f_n)\right)} \right) \sim M_1 f_n,$$

where M_1 is some constant. Also note that $n_j \to \infty$ as $n \to \infty$, then by Lemma 5.7.6 and the condition that $\frac{1}{n} = O(f_n)$, the second part of (5.7.28) is upper bounded by $M_2 f_n$.

Hence,

$$\limsup_{n\to\infty} \frac{\mathbb{E}\left[S^2(\mathbf{Y})\right]}{f_n^2} \le (M_1 + M_2)^2 < \infty.$$

The desired result is then obtained. ■

Chapter 6

Future Work

In this chapter, we conclude with some potential extensions to consider in the future.

- In Chapter 3, we only provide the second-order result regarding the convergence speed for the FGM copula, which may not be that popular in the literature. One possible direction of future research is to investigate similar results involving Gaussian copulas or mixture of Gaussian copulas. Besides, the main input for Chapter 3 lies in the theoretical treatments of multivariate regular variation (MRV), which, however, is criticized for its restrictive assumption of equality of tail indices. In our conducted empirical study, the validity of the MRV model is simply checked by comparing the tail indices of all marginal distributions. Since the tail index α_i for each component is estimated separately, one would hardly obtain identical values at the same time. Therefore, as long as the confidence interval for α_i overlap, we conjecture that the MRV model is close enough to reality. A formal goodness-of-fit test of the MRV model is recently proposed by Einmahl et al. (2018), where the authors compare the tail indices of the radial component conditional on the angular component falling in different, disjoint subsets. However, it remains an open question whether we can apply their formal test to datasets that exhibit obvious serial dependence.

- In Chapter 5, we consider an Archimedean copula-based model for measuring portfolio credit risk. We first derive sharp asymptotics for estimating the probability of large portfolio losses and the expected shortfall. Using this as a stepping stone, we develop an importance sampling algorithm based on hazard rate twisting and another algorithm based on conditional Monte Carlo. The assumption of an Archimedean copula may be more suitable for a large homogeneous portfolio. In order to capture hierarchical dependence structure among the obligors in a large credit portfolio, the nested Gumbel copula is taken into consideration due to its particular structure in the stochastic representation. For this special case, we provide an efficient algorithm for estimating the portfolio credit risk. A potential direction of future research is to consider other Archimedean families in a hierarchical structure.

Also note that in our portfolio loss model (5.2.3), we assume the loss given default (LGD) is 100%, which is impractical. To handle this issue, we can follow the literature to model the default probability and LGD jointly by linking them with some common risk factors; see, e.g., Rösch and Scheule (2014) and Betz et al. (2018). Alternatively, we can follow the idea in Shi et al. (2017) to link the LGD to the severity of default through a loss settlement function. Inspired by the work of Choroś-Tomczyk et al. (2013), we can also model the relation between the joint defaults and the LGD using nested Archimedean copulas by letting the LGD uniformly distributed on [0,1], see Figure 6.1. In this way, we introduce heterogeneous LGDs for each sub-portfolio and every LGD is further linked with the defaults in that sub-portfolio with a copula.

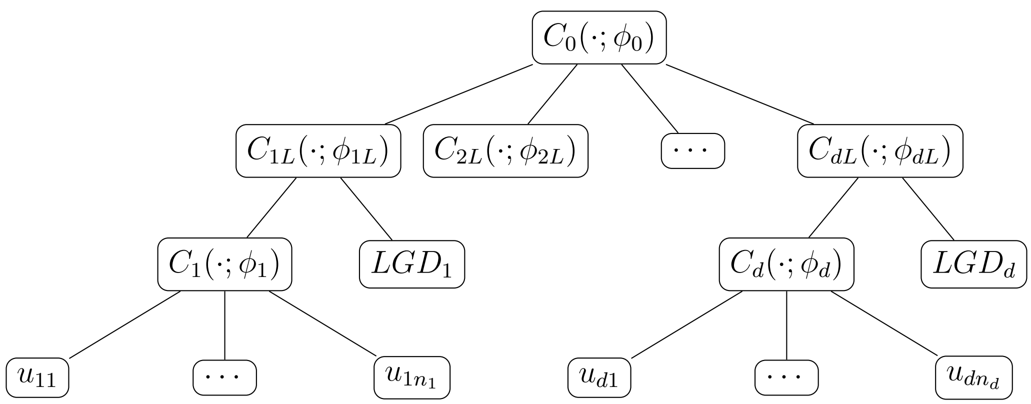


Figure 6.1: Tree structure of the partially nested Archimedean copula with random middle-level LGDs involved

References

Alink, S., Löwe, M. and Wüthrich, M. V. (2004). Diversification of aggregate dependent risks. *Insurance: Mathematics and Economics*, **35**(1), 77–95.

Arrow, K. J. (1978). Uncertainty and the welfare economics of medical care. In *Uncertainty in Economics*, pages 345–375. Elsevier.

Asmussen, S. (2018). Conditional monte carlo for sums, with applications to insurance and finance. *Annals of Actuarial Science*, **12**(2), 455–478.

Asmussen, S., Binswanger, K., Højgaard, B. et al. (2000). Rare events simulation for heavy-tailed distributions. *Bernoulli*, **6**(2), 303–322.

Asmussen, S. and Kroese, D. P. (2006). Improved algorithms for rare event simulation with heavy tails. *Advances in Applied Probability*, **38**(2), 545–558.

Avriel, M., Diewert, W. E., Schaible, S. and Zang, I. (2010). *Generalized concavity*, volume 63. SIAM.

Basrak, B., Davis, R. A. and Mikosch, T. (2002). A characterization of multivariate regular variation. *The Annals of Applied Probability*, **12**(3), 908–920.

Bassamboo, A., Juneja, S. and Zeevi, A. (2008). Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation. *Operations Research*, **56**(3), 593–606.

BCBS (2011). Basel III: A global regulatory framework for more resilient banks and banking systems. *Bank for International Settlements*.

Betz, J., Kellner, R. and Rösch, D. (2018). Systematic effects among loss given defaults and their implications on downturn estimation. *European Journal of Operational Research*, **271**(3), 1113–1144.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989). *Regular variation*, volume 27. Cambridge University Press.

Borch, K. (1960). *An attempt to determine the optimum amount of stop loss reinsurance.*

Bush, N., Hambly, B. M., Haworth, H., Jin, L. and Reisinger, C. (2011). Stochastic evolution equations in portfolio credit modelling. *SIAM Journal on Financial Mathematics*, **2**(1), 627–664.

Cambanis, S. (1977). Some properties and generalizations of multivariate eyraud-gumbel-morgenstern distributions. *Journal of Multivariate Analysis*, **7**(4), 551–559.

Chan, J. C. and Kroese, D. P. (2010). Efficient estimation of large portfolio loss probabilities in t-copula models. *European Journal of Operational Research*, **205**(2), 361–367.

Charpentier, A. and Segers, J. (2009). Tails of multivariate archimedean copulas. *Journal of Multivariate Analysis*, **100**(7), 1521–1537.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula methods in finance.* John Wiley & Sons.

Choroś-Tomczyk, B., Härdle, W. K. and Okhrin, O. (2013). Valuation of collateralized debt obligations with hierarchical archimedean copulae. *Journal of Empirical Finance*, **24**, 42–62.

Choueifaty, Y. and Coignard, Y. (2008). Toward maximum diversification. *Journal of Portfolio Management*, **35**(1), 40.

Cline, D. B. and Samorodnitsky, G. (1994). Subexponentiality of the product of independent random variables. *Stochastic Processes and their Applications*, **49**(1), 75–98.

Dal Maso, G. (2012). *An introduction to* Γ*-convergence*, volume 8. Springer Science & Business Media.

de Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction.* Springer Science & Business Media.

Degen, M., Lambrigger, D. D. and Segers, J. (2010). Risk concentration and diversification: second-order properties. *Insurance: Mathematics and Economics*, **46**(3), 541–546.

Dekkers, A. L., Einmahl, J. H. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, **17**(4), 1833–1855.

Dembo, A., Deuschel, J.-D. and Duffie, D. (2004). Large portfolio losses. *Finance and Stochastics*, **8**(1), 3–16.

Einmahl, J. H., Yang, F. and Zhou, C. (2018). Testing the multivariate regular variation model.

Embrechts, P. and Hofert, M. (2013). A note on generalized inverses. *Mathematical Methods of Operations Research*, **77**(3), 423–432.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media.

Embrechts, P., Lambrigger, D. D. and Wüthrich, M. V. (2009a). Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. *Extremes*, **12**(2), 107–127.

Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk Management: Value at Risk and Beyond*, **1**, 176–223.

Embrechts, P., Nešlehová, J. and Wüthrich, M. V. (2009b). Additivity properties for value-at-risk under archimedean dependence and heavy-tailedness. *Insurance: Mathematics and Economics*, **44**(2), 164–169.

Farlie, D. J. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, **47**(3/4), 307–323.

Feller, W. (1971). An introduction to probability theory and its applications john wiley. *New York*.

Fischer, M. and Klein, I. (2007). Constructing generalized fgm copulas by means of certain univariate distributions. *Metrika*, **65**(2), 243–260.

Glasserman, P., Kang, W. and Shahabuddin, P. (2007). Large deviations in multifactor portfolio credit risk. *Mathematical Finance*, **17**(3), 345–379.

Glasserman, P., Kang, W. and Shahabuddin, P. (2008). Fast simulation of multifactor portfolio credit risk. *Operations Research*, **56**(5), 1200–1217.

Glasserman, P. and Li, J. (2005). Importance sampling for portfolio credit risk. *Management Science*, **51**(11), 1643–1656.

Gordy, M. B. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation*, **12**(3), 199–232.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, **55**(292), 698–707.

Gupton, G. M., Finger, C. C. and Bhatia, M. (1997). *Creditmetrics: technical document*. JP Morgan & Co.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, pages 1163–1174.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, **58**(301), 13–30.

Hofert, M. (2008). Sampling archimedean copulas. *Computational Statistics & Data Analysis*, **52**(12), 5163–5174.

Hofert, M. (2011). Efficiently sampling nested archimedean copulas. *Computational Statistics & Data Analysis*, **55**(1), 57–70.

Hofert, M. (2012). A stochastic representation and sampling algorithm for nested archimedean copulas. *Journal of Statistical Computation and Simulation*, **82**(9), 1239–1255.

Hofert, M. and Scherer, M. (2011). Cdo pricing with nested archimedean copulas. *Quantitative Finance*, **11**(5), 775–787.

Hua, L. and Joe, H. (2011). Second order regular variation and conditional tail expectation of multiple risks. *Insurance: Mathematics and Economics*, **49**(3), 537–546.

Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied probability*, **34**(3), 587–608.

Ibragimov, R. (2009). Portfolio diversification and value at risk under thick-tailedness. *Quantitative Finance*, **9**(5), 565–580.

Ibragimov, R. and Prokhorov, A. (2016). Heavy tails and copulas: Limits of diversification revisited. *Economics Letters*, **149**, 102–107.

Ibragimov, R. and Walden, J. (2007). The limits of diversification when losses may be large. *Journal of Banking & Finance*, **31**(8), 2551–2569.

Ibragimov, R. and Walden, J. (2011). Value at risk and efficiency under dependence and heavy-tailedness: models with common shocks. *Annals of Finance*, **7**(3), 285–318.

Jakob, K. and Fischer, M. (2014). Quantifying the impact of different copulas in a generalized creditrisk+ framework an empirical study. *Dependence Modeling*, **2**(1).

Jessen, A. H. and Mikosch, T. (2006). Regularly varying functions. *Publications de L'institut Mathematique*, (94).

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Juneja, S., Karandikar, R. and Shahabuddin, P. (2007). Asymptotics and fast simulation for tail probabilities of maximum of sums of few random variables. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **17**(2), 7.

Juneja, S. and Shahabuddin, P. (2002). Simulating heavy tailed processes using delayed hazard rate twisting. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **12**(2), 94–118.

Kealhofer, S. and Bohn, J. (2001). Portfolio management of credit risk. *Technical Report*.

Kimberling, C. H. (1974). A probabilistic interpretation of complete monotonicity. *Aequationes Mathematicae*, **10**(2), 152–164.

Kotz, S. and Nadarajah, S. (2000). *Extreme value distributions: theory and applications*. World Scientific.

Li, D. X. (2000). On default correlation: A copula function approach. *The Journal of Fixed Income*, **9**(4), 43–54.

Lindskog, F. (2004). *Multivariate extremes and regular variation for stochastic processes*. Ph.D. thesis, ETH Zurich.

Lucas, A., Klaassen, P., Spreij, P. and Straetmans, S. (2001). An analytic approach to credit risk of large corporate bond and loan portfolios. *Journal of Banking & Finance*, **25**(9), 1635–1664.

Maier, R. and Wüthrich, M. V. (2009). Law of large numbers and large deviations for dependent risks. *Quantitative Finance*, **9**(2), 207–215.

Mainik, G. (2010). *On asymptotic diversification effects for heavy-tailed risks*. Ph.D. thesis, PhD thesis, University of Freiburg.

Mainik, G. and Embrechts, P. (2013). Diversification in heavy-tailed portfolios: properties and pitfalls. *Annals of Actuarial Science*, **7**(1), 26–45.

Mainik, G., Mitov, G. and Rüschendorf, L. (2015). Portfolio optimization for heavy-tailed assets: Extreme risk index vs. markowitz. *Journal of Empirical Finance*, **32**, 115–134.

Mainik, G. and Rüschendorf, L. (2010). On optimal portfolio diversification with respect to extreme risks. *Finance and Stochastics*, **14**(4), 593–623.

Mao, T. and Ng, K. W. (2015). Second-order properties of tail probabilities of sums and randomly weighted sums. *Extremes*, **18**(3), 403–435.

Mao, T. and Yang, F. (2015). Risk concentration based on expectiles for extreme risks under fgm copula. *Insurance: Mathematics and Economics*, **64**, 429–439.

Marshall, A. W. and Olkin, I. (1988). Families of multivariate distributions. *Journal of the American Statistical Association*, **83**(403), 834–841.

McLeish, D. L. (2010). Bounded relative error importance sampling and rare event simulation. *ASTIN Bulletin: The Journal of the IAA*, **40**(1), 377–398.

McNeil, A. J. (2008). Sampling nested archimedean copulas. *Journal of Statistical Computation and Simulation*, **78**(6), 567–581.

McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools-revised edition*. Princeton University Press.

Merton, R. C. (1972). An analytic derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis*, **7**(4), 1851–1872.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, **29**(2), 449–470.

Moore, K., Sun, P., de Vries, C. G. and Zhou, C. (2013). The cross-section of tail risks in stock returns. *Available at SSRN 2240131*.

Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitteilingsblatt fur Mathematische Statistik*, **8**, 234–235.

Müller, A. (1997). Stop-loss order for portfolios of dependent risks. *Insurance: Mathematics and Economics*, **21**(3), 219–223.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Nolan, J. (2003). *Stable distributions: models for heavy-tailed data*. Birkhauser Boston.

Owadally, I. (2012). An improved closed-form solution for the constrained minimization of the root of a quadratic functional. *Journal of Computational and Applied Mathematics*, **236**(17), 4428–4435.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, **3**(1), 119–131.

Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, pages 659–680.

Rényi, A. (1953). On the theory of order statistics. *Acta Mathematica Hungarica*, **4**(3), 191–231.

Resnick, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, **5**(4), 303–336.

Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.

Resnick, S. I. (2013). *Extreme values, regular variation and point processes*. Springer.

Rockafellar, R. T. (2015). *Convex analysis*. Princeton University Press.

Rogge, E. and Schönbucher, P. J. (2003). Modelling dynamic portfolio credit risk. *Department of Mathematics, Imperial College and ABN AMRO Bank, London and Department of Mathematics, ETH Zurich, Zurich*.

Rösch, D. and Scheule, H. (2014). Forecasting probabilities of default and loss rates given default in the presence of selection. *Journal of the Operational Research Society*, **65**(3), 393–407.

Roy, A. D. (1952). Safety first and the holding of assets. *Econometrica*, pages 431–449.

Schaible, S. and Ibaraki, T. (1983). Fractional programming. *European Journal of Operational Research*, **12**(4), 325–338.

Schloegl, L. and OKane, D. (2005). A note on the large homogeneous portfolio approximation with the student-t copula. *Finance and Stochastics*, **9**(4), 577–584.

Schönbucher, P. (2002). Taken to the limit: simple and not-so-simple loan loss distributions. *The Best of Wilmott*, **1**, 143–160.

Shi, X., Tang, Q. and Yuan, Z. (2017). A limit distribution of credit portfolio losses with low default probabilities. *Insurance: Mathematics and Economics*, **73**, 156–167.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, **8**, 229–231.

Smith, R. L. (1987). Estimating tails of probability distributions. *The Annals of Statistics*, **15**(3), 1174–1207.

Tang, Q., Tang, Z. and Yang, Y. (2019). Sharp asymptotics for large portfolio losses under extreme risks. *European Journal of Operational Research*.

Tong, E. N., Mues, C., Brown, I. and Thomas, L. C. (2016). Exposure at default models with and without the credit conversion factor. *European Journal of Operational Research*, **252**(3), 910–920.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge University Press.

Vasicek, O. (1991). Limiting loan loss probability distribution. *KMV corporation*.

Vasicek, O. A. (1987). *Probability of loss on loan portfolio*. KMV.

Vasicek, O. A. (2002). The distribution of loan portfolio value. *Risk*, **15**(12), 160–162.

Wei, G. and Hu, T. (2002). Supermodular dependence ordering on a class of multivariate copulas. *Statistics & Probability Letters*, **57**(4), 375–385.

Wüthrich, M. V. (2003). Asymptotic value-at-risk estimates for sums of dependent random variables. *ASTIN Bulletin: The Journal of the IAA*, **33**(1), 75–92.

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