Master's Thesis

Heavy-tailed Distributions and Financial Risk Measures

Julia Schauman

Fall 2022
University of Helsinki
Department of Mathematics and Statistics

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA
MATEMATISK-NATURVETENSKAPLIGA FAKULTETEN
FACULTY OF SCIENCE

Tiedekunta – Fakultet – Faculty Faculty of Science		Koulutusohjelma – Utbildningsprogram – Degree programme Master's Programme in Mathematics and Statistics
Opintosuunta – Studieinriktning – Study track Insurance and Financial Mathematics		
Tekijä – Författare – Author Julia Schauman		
Työn nimi – Arbetets titel – Title Heavy-tailed Distributions and Financial Risk Measures		
Työn laji – Arbetets art – Level Master's thesis	Aika – Datum – Month and year Fall 2022	Sivumäärä – Sidoantal – Number of pages 50 p.

Tiivistelmä – Referat – Abstract

In this thesis, we explore financial risk measures in the context of heavy-tailed distributions. Heavy-tailed distributions and the different classes of heavy-tailed distributions will be defined mathematically in this thesis but in more general terms, heavy-tailed distributions are distributions that have a tail or tails that are heavier than the exponential distribution. In other words, distributions which have tails that go to zero more slowly than the exponential distribution. Heavy-tailed distributions are much more common than we tend to think and can be observed in everyday situations. Most extreme events, such as large natural phenomena like large floods, are good examples of heavy-tailed phenomena.

Nevertheless, we often expect that most phenomena surrounding us are normally distributed. This probably arises from the beauty and effortlessness of the central limit theorem which explains why we can find the normal distribution all around us within natural phenomena. The normal distribution is a light-tailed distribution and essentially it assigns less probability to the extreme events than a heavy-tailed distribution. When we don't understand heavy tails, we underestimate the probability of extreme events such as large earthquakes, catastrophic financial losses or major insurance claims.

Understanding heavy-tailed distributions also plays a key role when measuring financial risks. In finance, risk measuring is important for all market participants and using correct assumptions on the distribution of the phenomena in question ensures good results and appropriate risk management. Value-at-Risk (VaR) and the expected shortfall (ES) are two of the best-known financial risk measures and the focus of this thesis. Both measures deal with the distribution and more specifically the tail of the loss distribution. Value-at-Risk aims at measuring the risk of a loss whereas ES describes the size of a loss exceeding the VaR. Since both risk measures are focused on the tail of the distribution, mistaking a heavy-tailed phenomena for a light-tailed one can lead to drastically wrong conclusions. The mean excess function is an important mathematical concept closely tied to VaR and ES as the expected shortfall is mathematically a mean excess function. When examining the mean excess function in the context of heavy-tails, it presents very interesting features and plays a key role in identifying heavy-tails. This thesis aims at answering the questions of what heavy-tailed distributions are and why are they are so important, especially in the context of risk management and financial risk measures.

Chapter 2 of this thesis provides some key definitions for the reader. In Chapter 3, the different classes of heavy-tailed distributions are defined and described. In Chapter 4, the mean excess function and the closely related hazard rate function are presented. In Chapter 5, risk measures are discussed on a general level and Value-at-Risk and expected shortfall are presented. Moreover, the presence of heavy tails in the context of risk measures is explored. Finally, in Chapter 6, simulations on the topics presented in previous chapters are shown to shed a more practical light on the presentation of the previous chapters.

Avainsanat – Nyckelord – Keywords
Heavy-tailed distributions, risk measures, mean excess function, hazard rate, Value-at-Risk, expected shortfall

Säilytyspaikka – Förvaringställe – Where deposited

Muita tietoja – Övriga uppgifter – Additional information

Contents

1 Introduction **3**

2 Preliminaries **5**

3 Heavy-tailed distributions and their properties **8**
3.1 The classes of heavy-tailed distributions 8
3.1.1 Heavy-tailed distributions 9
3.1.2 Long-tailed distributions 12
3.1.3 Regularly varying distributions 13
3.1.4 Dominatedly varying distributions 15
3.2 Subexponential distributions 16
3.3 Differences between heavy-tailed and light-tailed random variables 18

4 Mean excess function **20**
4.1 The mean excess function 20
4.2 The hazard rate function 25
4.3 The mean excess plot 33

5 Heavy tails and financial risk measures **35**
5.1 Risk measures in general 35
5.2 Value-at-Risk 36
5.3 Expected Shortfall 38

6 Simulation **43**
6.1 Summary of the data 43
6.2 Mean excess plots 44
6.3 Risk measures 46

7 Conclusions **49**

1 Introduction

We often expect and have been taught to think that most phenomena surrounding us are normally distributed. This probably arises from the beauty and effortlessness of the central limit theorem which explains why we can find the normal distribution all around us within natural phenomena. The normal distribution is very useful and often well applicable in many situations but not everything surrounding us follows a normal distribution or can be modelled by assuming a light-tailed normal distribution. Indeed, many phenomena, especially in the financial context, have a heavy-tailed distribution.

Heavy-tailed distributions will be defined mathematically in Chapter 3 of this thesis but in more general terms, heavy-tailed distributions are distributions that have a tail or tails that are heavier than the exponential distribution. In other words, distributions which have tails that go to zero more slowly than the exponential distribution. The tail functions for the exponential distribution with parameter 1 and the heavy-tailed Pareto (1,1) distribution are presented in Figure 1. The figure shows how the red curve corresponding to a heavy-tailed Pareto distribution remains above the black curve corresponding to the exponential distribution as x becomes larger.

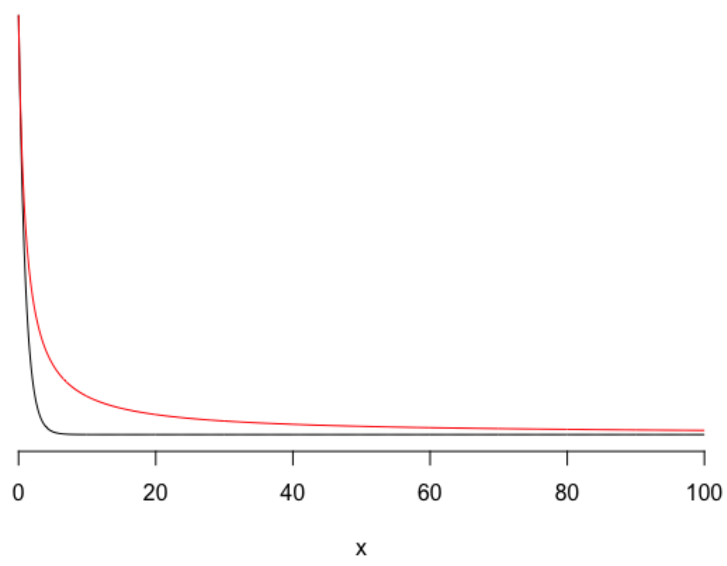

Figure 1: The tail functions for the exponential distribution with parameter 1 and the Pareto (1,1) distribution

The normal distribution is light-tailed and essentially it assigns less probability to the extreme events than a heavy-tailed distribution. When modelling, for example, a financial phenomenon it is in many cases considered more appropriate to use a heavy-tailed distribution as we know that many events, such as the returns of financial assets, have a heavy-tailed distribution. Heavy-tailed distributions are much more common than we tend to think and can be observed in every day situations. Most extreme events, such as large natural phenomena like large floods, are good examples of heavy-tailed phenomena. When we don't understand or acknowledge heavy tails, we underestimate the probability of these extreme events such as large earthquakes, catastrophic financial losses or major insurance claims.

Understanding heavy-tailed distributions also plays a key role when measuring financial risks. In finance, risk measuring is important for all market participants and using correct assumptions on the distribution of the phenomena in question ensures good results and appropriate risk management. Value-at-Risk (VaR) and the expected shortfall (ES) are two of the best known financial risk measures. Both measures deal with the distribution and more specifically the tail of the loss distribution. Value-at-Risk aims at measuring the risk of a loss whereas ES describes the size of a loss exceeding the VaR. Since both risk measures are focused on the tail of the distribution, mistaking a heavy-tailed phenomena for a light-tailed one can lead to drastically wrong conclusions. The mean excess function is an important mathematical concept closely tied to VaR and ES. When examining the mean excess function in the context of heavy tails, it presents very interesting features and plays a key role in identifying heavy tails.

This thesis aims at answering the questions of what are heavy-tailed distributions and why are they are so important, especially in the context of risk management and financial risk measures. Chapter 2 of this thesis provides some key definitions for the reader. In Chapter 3, the different classes of heavy-tailed distributions are defined and described. In Chapter 4, the mean excess function and the closely related hazard rate function are presented. In Chapter 5, risk measures are discussed on a general level and Value-at-Risk and expected shortfall are presented. Moreover, the presence of heavy tails in the context of risk measures is explored. Finally, in Chapter 6, simulations on the topics presented in previous chapters are shown to shed a more practical light on the presentation of the previous chapters.

2 Preliminaries

In this section some key definitions are presented. The definitions are based mainly on the definitions presented by Athreya et al. (2006), Foss et al. (2011) and Klugman et al. (2008).

Definition 2.1. (Probability space) Let Ω be a nonempty set and $\mathcal{F}$ a σ-algebra on Ω. Then, the pair $(\Omega, \mathcal{F})$ is called a measurable space. By adding a probability measure $\mathbb{P}$, we get $(\Omega, \mathcal{F}, \mathbb{P})$, which is called a probability space.

Definition 2.2. (Random variable) A function $X : \Omega \to \mathbb{R}$ is called a random variable if

$$\{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F}$$

for all $a \in \mathbb{R}$. Such a function is called $\mathcal{F}$-measurable.

Definition 2.3. For a random variable X with density function $f(x)$, the cumulative distribution function $F : [0, 1] \to \mathbb{R}$ is defined as follows:

$$F(x) = \mathbb{P}(X \leq x).$$

$F(x)$ is referred to as simply the distribution function in this thesis.

Definition 2.4. (Cumulative distribution function) F is the cumulative distribution function of a random variable X if and only if

(i) $0 \leq F(x) \leq 1$ for all x,

(ii) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$,

(iii) F is right-continuous on $\mathbb{R}$,

(iv) F is a non-decreasing function on $\mathbb{R}$.

The tails of a distribution function are the parts of the distribution function that correspond to the smallest and largest values of the random variable. The right tail, which we are focusing on in this thesis, corresponds to the largest values.

Definition 2.5. (Tail function) For the distribution function F, its tail function $\overline{F}$ is given by

$$\overline{F}(x) = \mathbb{P}(X > x) = 1 - F(x).$$

Definition 2.6. (The k^{th} moment) Provided that it exists, the k^{th} moment of a continuous random variable X with density function $f(x)$ is

$$\mathbb{E}(X^k) = \int_{-\infty}^{\infty} x^k f(x)\, dx.$$

Definition 2.7. (Moment generating function) The moment generating function $M_X : \mathbb{R} \to \mathbb{R}$ of a random variable X is defined by

$$M_X(t) = \mathbb{E}(e^{tX})$$

for all $t \in \mathbb{R}$. As e^{tX} is always non-negative, the moment generating function is always well defined although it can be infinite.

Definition 2.8. Given some distribution function F, the α-quantile of F for $\alpha \in (0, 1)$ is given by

$$q_\alpha(F) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}.$$

For a random variable X with distribution function F, the α-quantile can also be denoted by $q_\alpha(X)$.

Definition 2.9. (Asymptotic equivalence) Let f and g be positive functions on $[0, \infty)$. The functions are asymptotically equivalent if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

We denote asymptotic equivalence by $f \sim g$.

Definition 2.10. (Slowly varying function) The function $g(x)$ is slowly varying if

$$\lim_{x\to\infty} \frac{g(xy)}{g(x)} = 1 \quad \text{for all } y > 0.$$

Then, we denote $g \in \mathcal{R}_0$.

Definition 2.11. (The endpoint of a distribution) The endpoint T of a distribution F is given by

$$T = \inf\{x | F(x) = 1\} < \infty \quad \text{for some } x \in \mathbb{R}.$$

When $F(x) < 1$, then $T = \infty$.

Definition 2.12. (Convex cone) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{M} \subset L(\Omega, \mathcal{F}, \mathbb{P})$ which is the set of all random variables on $(\Omega, \mathcal{F})$ which are almost surely finite. Then, $\mathcal{M}$ is a convex cone if $X \in \mathcal{M}$ and $Y \in \mathcal{M}$ implies $X + Y \in \mathcal{M}$ and $\lambda X \in \mathcal{M}$ for every $\lambda > 0$.

Definition 2.13. (Laplace-Stieltjes transform) Let F be distribution function. Then, the Laplace-Stieltjes transform of F is given by

$$\hat{f}(t) = \int_{-\infty}^{\infty} e^{-tx} dF(x) = \mathbb{E}[e^{-tx}] \quad \text{for all } t > 0.$$

Lemma 2.1. *(Karamata's Theorem) Suppose $\mathcal{L} \in \mathcal{R}_0$ is locally bounded in $[x_0, \infty)$ for some $x_0 \geq 0$. Then,*

(i) for $\alpha > -1$

$$\int_{x_0}^{x} t^{\alpha} L(t)\, dt \sim (\alpha + 1)^{-1} x^{\alpha+1} L(x), \quad x \to \infty,$$

(ii) for $\alpha < -1$

$$\int_{x}^{\infty} t^{\alpha} L(t)\, dt \sim -(\alpha + 1)^{-1} x^{\alpha+1} L(x), \quad x \to \infty.$$

Proof. The proof can be found from Bingham et al. (1987). □

3 Heavy-tailed distributions and their properties

3.1 The classes of heavy-tailed distributions

Heavy-tailed distributions are an important group of classes of distributions with interesting and very useful properties. To avoid confusion we must first clarify what is meant by the tail of a distribution. In this thesis we focus on the right tail of distributions which can be defined as the portion of the distribution that holds the largest positive values of the random variable. In general terms, a heavy-tailed distribution is simply a distribution which assigns larger probabilities to the larger values. The definition of the tail function $\overline{F}$ is given in Definition 2.5.

The main classes of heavy-tailed distributions are the class of heavy-tailed distributions ($\mathcal{K}$) itself, long-tailed distributions ($\mathcal{L}$) and subexponential distributions ($\mathcal{S}$). These classes are closely linked to one another and all long-tailed distributions are also heavy-tailed and all subexponential distributions are actually both long-tailed and heavy-tailed. In addition, there are regularly varying distributions ($\mathcal{R}$) and dominatedly varying distributions ($\mathcal{D}$). In Figure 2, the different classes and their relationships are presented in a similar way as presented by Embrechts et al. (1997).

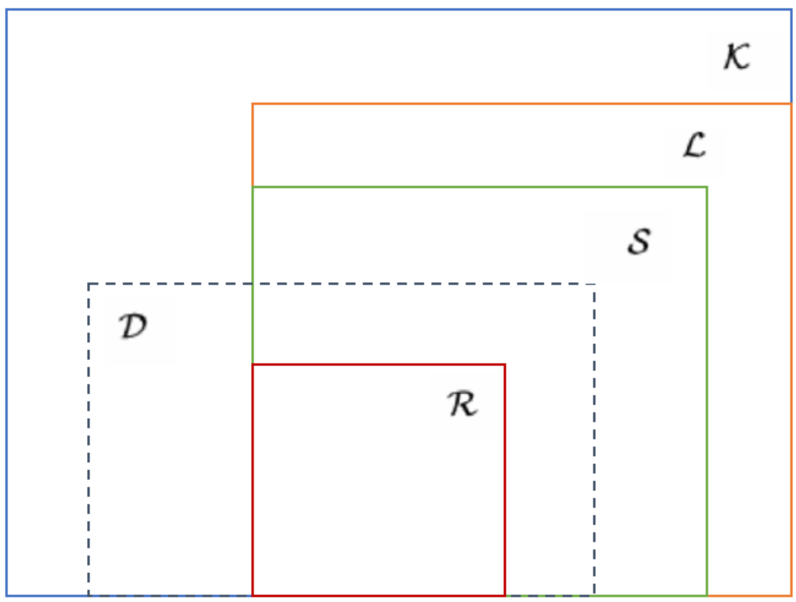

Figure 2: Different classes of heavy-tailed distributions

It can be argued that the concept of heavy-tailed distributions is relative, which means that a distribution can have a heavier tail than another distribution without really being a truly heavy-tailed distribution. To make a clear distinction, a heavy-tailed distribution is often defined with respect to the exponential distribution, which serves as a comparison point between the classes of heavy-tailed and light-tailed distributions. A heavy-tailed distribution is simply described as a distribution which has a tail that is heavier than the exponential distribution as presented in Figure 1. This means that there is more probability mass in the tail of a heavy-tailed distribution than there is in the tail of an exponential distribution after a certain point x. In this thesis we will focus on distributions which are right heavy-tailed.

In the following sections the definitions for the different classes of heavy-tailed distributions presented in Figure 2 are given. The class of heavy-tailed ($\mathcal{K}$) distributions is presented in Section 3.1.1, long-tailed ($\mathcal{L}$) distributions in Section 3.1.2, regularly varying distributions ($\mathcal{R}$) in Section 3.1.3, dominatedly-varying distributions ($\mathcal{D}$) in Section 3.1.4 and subexponential distributions ($\mathcal{S}$) are presented in Section 3.2.

3.1.1 Heavy-tailed distributions

The class of heavy-tailed distributions ($\mathcal{K}$) is the broadest of the classes presented in Figure 2 and consists of all different types of heavy-tailed distributions. Part of the reason why heavy-tailed distributions can seem unfamiliar to us is due to the fact that there are many ways to define a heavy-tailed distribution, which causes confusion. The definition presented here focuses in the right tail of the distribution and follows the presentation by Nair et al. (2022).

Definition 3.1. A distribution F is heavy-tailed if and only if

$$\limsup_{x \to \infty} \frac{\overline{F}(x)}{e^{-tx}} = \infty$$

for all $t > 0$.

A distribution is light-tailed if it isn't heavy-tailed. Moreover, a random variable X is heavy-tailed if its distribution function F is heavy-tailed. The definition also highlights that for a distribution to be heavy-tailed, one only needs to look at the tail of the distributions.

The following lemma shows that there are indeed many ways to define a heavy-tailed distribution and a typical definition is based on the moment generating function. Moreover, the lemma also sheds light on the differences between heavy-tailed and light-tailed distributions which are further explored in Subsection 3.3. Let's first define the corresponding risk function.

Definition 3.2. The risk function $R(x)$ for a distribution function F is given by

$$R(x) = -\ln \overline{F}(x).$$

In literature, the risk function is sometimes referred to as the (cumulative) hazard function. In Chapter 4, the risk function's first derivative is referred to as the hazard rate function.

Lemma 3.1. *The following statements are equivalent for any random variable X with distribution function F:*

(i) X *is heavy-tailed.*

(ii) $M_X(t) = \mathbb{E}[e^{tX}] = \infty$ *for all* $t > 0$.

(iii) It holds $\liminf_{x\to\infty} \frac{R(x)}{x} = 0$.

Proof. *(i) ⇒ (ii).* Suppose X is a heavy-tailed random variable with distribution function F. Based on Definition 3.1, this implies that there exists a strictly increasing sequence $(x_k)_{k\geq 1}$ such that $\lim_{x\to\infty} x_k = \infty$ and

$$\lim_{k\to\infty} e^{tx_k}\overline{F}(x_k) = \infty$$

when $t > 0$. Then, using the Laplace-Stieltjes transform we can show that for $M_X(t)$ it holds

$$\mathbb{E}[e^{tX}] = \int_0^\infty e^{tx} dF(x) \geq \int_{x_k}^\infty e^{tx} dF(x) \geq e^{tx_k}\overline{F}(x_k) \quad \text{for all } k.$$

Hence, $M_X(t) = \mathbb{E}[e^{tX}] = \infty$.

(ii) ⇒ (iii). Suppose Condition (ii) holds and Condition (iii) does not hold for X. This means that

$$\liminf_{x\to\infty} \frac{R(x)}{x} > 0$$

since $\frac{R(x)}{x} \geq 0$. This would further implies that there exists $\mu > 0$ and $x_0 > 0$ such that

$$\frac{R(x)}{x} = \frac{-\ln \overline{F}(x)}{x} \geq \mu \quad \Longleftrightarrow \quad \overline{F}(x) \leq e^{-\mu x}$$

for some $x \geq x_0$. If we now choose some t such that $0 < t < \mu$, we have

$$\mathbb{E}[e^{tX}] = \int_0^\infty \mathbb{P}(e^{tX} > x)\, dx = \int_0^{e^{tx_0}} \mathbb{P}(e^{tX} > x)\, dx + \int_{e^{tx_0}}^\infty \mathbb{P}\Big(X > \frac{-\ln x}{t}\Big)\, dx.$$

Furthermore, by noticing that $x \geq e^{tx_0} \iff \frac{\ln x}{t} \geq x_0$, we get

$$\begin{aligned}\mathbb{E}[e^{tX}] &\leq e^{tx_0} + \int_{e^{tx_0}}^{\infty} e^{-\mu \frac{\ln x}{t}}\, dx \\ &= e^{tx_0} + \int_{e^{tx_0}}^{\infty} x^{-\mu/t}\, dx.\end{aligned}$$

We chose t such that $0 < t < \mu$ meaning that $\mu/t > 1$. Then $\int_{e^{tx_0}}^{\infty} x^{-\mu/t} < \infty$. This further implies that $M_X(t) = \mathbb{E}[e^{tX}] < \infty$, which is a contradiction since we assumed (ii) holds for all $t > 0$. Hence, Condition (ii) implies Condition (iii).

(iii) ⇒ (i). To prove the last part, we assume that Condition (iii) holds for the random variable X. Then there exists a strictly increasing sequence $(x_k)_{k\geq 1}$ such that $\lim_{x\to\infty} x_k = \infty$ and

$$\liminf_{x\to\infty} \frac{R(x_k)}{x_k} = 0.$$

When $t > 0$, the previous implies that there exists t_0 such that

$$\begin{aligned} & \frac{R(x_k)}{x_k} < e^{-t/2} \\ \iff \quad & \overline{F}(x_k) > e^{-tx_k/2} \\ \iff \quad & \frac{\overline{F}(x_k)}{e^{tx_k}} > e^{tx_k/2} \end{aligned}$$

for all $t > t_0$, which implies $\lim_{t\to\infty} \frac{\overline{F}(x_k)}{e^{tx_k}} = \infty$ and further that $\limsup_{x\to\infty} \frac{\overline{F}(x)}{e^{tx}} = \infty$. Based on Definition 3.1 we conclude that Condition (iii) implies Condition (i). □

The following lemma shows an interesting property for continuous non-negative random variables with heavy-tailed distribution functions as a heavy-tailed distribution function implies a heavy-tailed density function. However, the inverse does not hold since a random variable with a heavy-tailed density function can still have a light-tailed distribution function (Foss et al. (2011)). A heavy-tailed function is a function which is not bounded by a decreasing exponential function and is defined similarly to a heavy-tailed distribution as presented in Definition 3.1. See Foss et al. (2011) for the exact definition of a heavy-tailed function.

Lemma 3.2. *Suppose a continuous distribution F has a density function f and F is heavy-tailed. Then the density function $f(x)$ is also heavy-tailed.*

Proof. The proof is presented by Foss et al. (2011). □

Example 3.1. The Pareto distribution has the distribution function

$$F(x) = 1 - \left(\frac{\theta}{x+\theta}\right)^{\alpha}$$

and, hence, the corresponding tail function for some $\theta > 0$ and $\alpha > 0$ is

$$\overline{F}(x) = \left(\frac{\theta}{x+\theta}\right)^{\alpha}.$$

Furthermore, we can see that

$$\limsup_{x\to\infty} \left(\frac{\theta}{x+\theta}\right)^{\alpha} e^{tx} = \infty \quad \text{for all } \lambda, t > 0$$

and we can conclude based on Definition 3.1 that the Pareto distribution is a heavy-tailed distribution.

Example 3.2. The Weibull distribution is an important distribution, which has a key role, among other things, when describing phenomena of nature or economics. The Weibull distribution has the following distribution function, when $x \geq 0$,

$$F(x) = 1 - e^{-\lambda x^{\alpha}}$$

and, hence, the corresponding tail function

$$\overline{F}(x) = e^{-\lambda x^{\alpha}}.$$

If $\alpha = 1$ the distribution is exponential. Suppose $0 < \alpha < 1$ and $\lambda > 0$. Then, we get

$$\liminf_{x\to\infty} \frac{-\ln e^{-\lambda x^{\alpha}}}{x} = \liminf_{x\to\infty} \frac{\lambda x^{\alpha}}{x} = 0$$

and, hence, we conclude based on Condition (iii) of Lemma 3.1 that the Weibull distribution with a shape parameter $0 < \alpha < 1$ is indeed heavy-tailed.

3.1.2 Long-tailed distributions

Long-tailed distributions ($\mathcal{L}$) are also heavy-tailed distributions but they are described as having a specific smoothness to their tail function distinguishing them from other heavy-tailed distributions. As Figure 2 suggests the class of long-tailed distributions is actually a broad class as it contains the classes of subexponential and regularly varying distributions, i.e. the relationships $\mathcal{S} \in \mathcal{L}$ and $\mathcal{R} \in \mathcal{L}$ hold. Hence, many of the commonly known heavy-tailed distributions are long-tailed. A long-tailed distribution is defined as follows:

Definition 3.3. Let F be the distribution function of a positive random variable X. Then F is long-tailed if $\overline{F} > 0$ and, for any fixed positive y, it holds that

$$\lim_{x\to\infty} \frac{\overline{F}(x+y)}{\overline{F}(x)} = 1$$

for all $x \in \mathbb{R}$.

The emergence of long-tailed distributions can be described through the concept of residual life. When thinking about waiting times, the long-tailed distributions fit a situation where one has already waited a very long time and the wait can be expected to last forever. This means that the residual life or expected waiting time is infinite. Intuitively this makes sense as presented in the examples by Nair et al. (2022). When waiting for a train, if the train hasn't arrived after a very long wait, it is likely there is a more severe problem, such as a major technical problem, and therefore the train cannot be expected to arrive soon. These type of situations can be modelled by long-tailed distributions.

Example 3.3. It was previously shown that the Pareto distribution is heavy-tailed. Let us now show that the Pareto distribution is also long-tailed. As presented previously, the Pareto distribution has the following tail function,

$$\overline{F}(x) = \left(\frac{\theta}{x+\theta}\right)^{\alpha}$$

where $\theta, \alpha > 0$ and F has support on $(0, \infty)$. Using Definition 3.3, we get for the Pareto distribution that

$$\lim_{x\to\infty} \frac{\overline{F}(x+y)}{\overline{F}(x)} = \lim_{x\to\infty} \frac{\left(\frac{\theta}{(x+y)+\theta}\right)^{\alpha}}{\left(\frac{\theta}{x+\theta}\right)^{\alpha}} = \lim_{x\to\infty} \left(\frac{x+\theta}{x+y+\theta}\right)^{\alpha} = 1.$$

We conclude that the Pareto distribution is long-tailed.

3.1.3 Regularly varying distributions

Another class of heavy-tailed distributions are the regularly varying distributions ($\mathcal{R}$). There are two important concepts closely related to regularly varying distributions, which are scale invariance and the power law property. Neither concept directly imply that a distribution is heavy-tailed but both are closely linked to regularly varying distributions which consequently are heavy-tailed.

Scale invariance is an interesting property which can be observed with some heavy-tailed distributions although not all heavy-tailed distributions are scale invariant. In practice, a scale invariant distribution is a distribution for which the shape of the tail function $\overline{F}$ does not change when the scale at which it is examined changes. More precisely, a scale invariant distribution is defined as follows:

Definition 3.4. A distribution function F is scale invariant if there exists $a > 0$ and a positive continuous function $g(x)$ such that

$$\overline{F}(\lambda x) = g(\lambda)\overline{F}(x) \qquad \text{for all } \lambda x > a.$$

Scale invariant distributions are closely linked to power law distributions. To clarify, power law distributions are defined as distributions which have tails that match the Pareto distribution by some multiplicative constant. In fact, as explained by Nair et al. (2022), all scale invariant distributions are power law distributions. In the context of heavy-tailed distributions the tail of the distribution is of interest and, moreover, the further part of the tail. Hence, when scale invariance is examined as $x \to \infty$, we are introduced to asymptotic scale invariance. Let's formally define asymptotically scale invariant distributions.

Definition 3.5. A distribution function F is asymptotically scale invariant if there exists a positive continuous function $g(x)$ such that

$$\lim_{x\to\infty} \frac{\overline{F}(\lambda x)}{\overline{F}(x)} = g(x) \qquad \text{for all } \lambda > 0.$$

Scale invariance is quite a special property and not many distributions are scale invariant. However, as asymptotic scale invariance focuses only on the tail of the distribution, it is a much more common property since the body of the distribution can be ignored. In fact, asymptotically scale invariant distributions are regularly varying distributions. Regularly varying distributions can be described as asymptotically power law distributions.

Definition 3.6. A distribution function F is regularly varying with index $-\alpha$ if

$$\lim_{x\to\infty} \frac{\overline{F}(xy)}{\overline{F}(x)} = y^{-\alpha} \quad \text{for all } y > 0.$$

Due to the close link to scale invariance and the power law property, regularly varying distributions can be extremely useful as they behave in a beneficial way, similarly to the Pareto distribution. In fact, regularly varying distributions can in many cases be analysed as if they were Pareto distributions which gives beneficial computational advantages. Some of the best known regularly varying distributions are the Burr distribution and the Cauchy distribution.

Example 3.4. As presented previously, the Pareto distribution has the following tail function

$$\overline{F}(x) = \left(\frac{\theta}{x+\theta}\right)^{\alpha}$$

where $\theta, \alpha > 0$. By following Definition 3.6, we get

$$\lim_{x\to\infty} \frac{\left(\frac{\theta}{xy+\theta}\right)^\alpha}{\left(\frac{\theta}{x+\theta}\right)^\alpha} = \lim_{x\to\infty} \left(\frac{1+\frac{\theta}{x}}{y+\frac{\theta}{x}}\right)^\alpha = y^{-\alpha}$$

which shows that the Pareto distribution is a regularly varying distribution.

3.1.4 Dominatedly varying distributions

The next class of heavy-tailed distributions to be presented is the class of dominatedly varying distributions. The class of dominatedly varying distribution ($\mathcal{D}$) is defined as follows:

Definition 3.7. A distribution with the distribution function F is dominatedly varying if

$$\limsup_{x\to\infty} \frac{\overline{F}(x/2)}{\overline{F}(x)} < \infty.$$

As Figure 2 illustrates, regularly varying distributions $\mathcal{R}$ are also dominatedly varying, that is $\mathcal{R} \subset \mathcal{D}$. However, not all dominatedly varying distributions are subexponential or vice versa meaning that $\mathcal{D} \not\subset \mathcal{S}$ and $\mathcal{S} \not\subset \mathcal{D}$ hold. However, it holds that $\mathcal{L} \cap \mathcal{D} \subset \mathcal{S}$ as Figure 2 shows (see Embrechts et al. (1997)).

Example 3.5. In other examples, the focus has been on showing that a distribution belongs to the class of heavy-tailed distributions in question. As Figure 2 suggests the class of dominatedly varying distributions does not include all commonly known heavy-tailed distributions and, hence, let us show that the Weibull distribution is one of the distributions that does not belong to the class of dominatedly varying distributions.

The Weibull distribution has the tail function

$$\overline{F}(x) = e^{-cx^\beta}$$

where $c > 0$ and $0 < \beta < 1$ for the distribution to be heavy-tailed. Based on Definition 3.7 we have

$$\limsup_{x\to\infty} \frac{e^{-c(\frac{x}{2})^\beta}}{e^{-cx^\beta}} = \limsup_{x\to\infty} e^{cx^\beta - 2^{-\beta}cx^\beta} = \infty$$

as $c > 0$ and $\beta > 0$. We conclude that the heavy-tailed Weibull distribution is not dominatedly varying.

3.2 Subexponential distributions

Perhaps the most important class of heavy-tailed distributions is the class of subexponential distributions ($\mathcal{S}$). Subexponential distributions were first introduced in the 1960s and especially in the financial context, subexponential distributions have shown to be very useful. They are widely used for example when studying random walks and in the insurance context where it is important to use a distribution that allows for extremely large values to occur.

The class is named based on the property that the tail of a subexponential distribution decreases more slowly than the tail of an exponential distribution. In fact, most of the heavy-tailed distributions that are commonly known are subexponential distributions. Although many heavy-tailed distributions are subexponential distributions, not all heavy-tailed distributions are subexponential as presented in Figure 2. Any subexponential distribution is long-tailed and particularly any subexponential distribution is heavy-tailed. That is, $\mathcal{S} \subset \mathcal{L} \subset \mathcal{K}$ (see Embrechts et al. (1997)). The contrary, however, does not hold since there are long-tailed distributions which are not subexponential.

Let's look at the formal definition of a subexponential distribution. Consider the tail function $\overline{F}$ as given in Definition 2.5. Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables for all n. Denote the tail of the n-fold convolution of F by

$$\overline{F^{n*}}(x) = 1 - F^{n*}(x) = P(X_1 + \cdots + X_n > x).$$

Then, we get the following definition for a subexponential distribution.

Definition 3.8. A distribution F is subexponential if for all $n \geq 2$ independent random variables $X_1, \ldots, X_n$ with distribution F it holds

$$\mathbb{P}(X_1 + \cdots + X_n > x) \sim n\mathbb{P}(X_1 > x) \quad \text{as } x \to \infty,$$

which can also be presented as $\overline{F^{n*}}(x) \sim n\overline{F}(x)$.

The Definition 3.8 of subexponentiality states that it must hold for all $n \geq 2$ which in practice is not a very convenient definition as one would need to check all possible values of n to show subexponentiality. The definition can be presented in a simpler form as it has been proven that the limit presented in Definition 3.8 holds for all $n \geq 2$ if and only if it holds for $n = 2$. Then, the limit is equal to 2 as the following lemma presents.

Lemma 3.3. *A distribution F on $[0, \infty)$ is subexponential if*

$$\lim_{x \to \infty} \frac{\overline{F^{2*}}(x)}{\overline{F}(x)} = 2$$

for all $x \geq 0$.

Proof. The proof was originally presented by Chistyakov (1964). □

The following lemma formally presents the condition sufficient to show subexponentiality. The proof follows the version of the proof presented by Embrects et al. (1997) and Asmussen et al. (2010).

Lemma 3.4. *(Sufficient condition for subexponentiality). For any distribution F on $(0,\infty)$, F is subexponential if*

$$\limsup_{x\to\infty} \frac{\overline{F^{2*}}(x)}{\overline{F}(x)} \leq 2. \tag{3.1}$$

Proof. Let X_1 and X_2 be non-negative independent random variables with distribution function F. For the nominator it holds that

$$\overline{F^{2*}(x)} = \mathbb{P}(X_1 + X_2 > x) \geq \mathbb{P}(\max(X_1, X_2) > x)$$

for all $x \geq 0$. We can modify the previous further

$$\mathbb{P}(\max(X_1, X_2) > x) = \mathbb{P}(X_1 > x) + \mathbb{P}(X_2 > x) - \mathbb{P}(X_1 > x, X_2 > x)$$
$$= 2\overline{F}(x) - (\overline{F}(x))^2.$$

Then, it follows

$$\liminf_{x\to\infty} \frac{\overline{F^{2*}}(x)}{\overline{F}(x)} \geq \liminf_{x\to\infty} \frac{2\overline{F}(x) - (\overline{F}(x))^2}{\overline{F}(x)} \geq 2.$$

Given (3.1) we conclude that Lemma 3.3 holds. □

An important phenomenon related to subexponential distributions is called the *principle of a single big jump* which describes the probabilistic behaviour of sums of independent subexponential random variables. The principle characterizes the nature of heavy tails as it describes the typical behaviour of heavy-tailed random variables. Intuitively, the principle of a single big jump states that the sum of independent and identically distributed subexponential random variables is large if and only if the maximum of those random variables is large. In practice, a single large value can have a significant effect on the result of the sum. This could be for example the effect of a single large claim on the total claim amount in the context of insurance. This phenomenon explains the typical presence of really large values in observations of subexponentially distributed random variables.

Lemma 3.5. *(Principle of a single big jump). Let $X_1, \ldots, X_n$ be independent random variables which all have a common subexponential distribution F. Then*

$$\lim_{x\to\infty} \frac{\mathbb{P}(X_1 + \cdots + X_n > x)}{\mathbb{P}(\max(X_1, \ldots, X_n) > x)} = 1 \quad \textit{for all } n \geq 2.$$

Proof. The Lemma can be proved using Definition 3.8 of subexponentiality. Let $X_1, \dots X_n$ be non-negative, independent and subexponential random variables with distribution function F. Based on the definition we know that for all $n \geq 2$,

$$\mathbb{P}(X_1 + \cdots + X_n > x) \sim n\overline{F}(x) \quad \text{as } x \to \infty.$$

Now it suffices to show that the same holds for the denominator, i.e. that it holds

$$\mathbb{P}(\max(X_1, \dots, X_n) > x) \sim n\overline{F}(x)$$

for all $n \geq 2$ as $x \to \infty$. Let's examine the denominator in more detail. We observe that

$$\mathbb{P}(\max(X_1, \dots, X_n) > x) = 1 - F^n(x) = \overline{F}(x) \sum_{k=0}^{n-1} F^k(x).$$

Because $F \in [0, 1]$ for all $x \in \mathbb{R}$ and the sum is finite for a fixed n, it holds

$$\lim_{x\to\infty} \sum_{k=0}^{n-1} F^k(x) = \sum_{k=0}^{n-1} \lim_{x\to\infty} F^k(x) = \sum_{k=0}^{n-1} 1 = n.$$

Then as $x \to \infty$, it follows that

$$\lim_{x\to\infty} \overline{F}(x) \sum_{k=0}^{n-1} F^k(x) = \lim_{x\to\infty} n\overline{F}(x).$$

By combining both parts, we see that

$$\lim_{x\to\infty} \frac{\mathbb{P}(X_1 + \cdots + X_n > x)}{\mathbb{P}(\max(X_1, \dots, X_n) > x)} = \lim_{x\to\infty} \frac{n\overline{F}(x)}{n\overline{F}(x)} = 1 \quad \text{for all } n \geq 2.$$

□

3.3 Differences between heavy-tailed and light-tailed random variables

One might wonder why making the distinction between light-tailed and heavy-tailed distributions is so important. Essentially, the most important differences arises from the fact that heavy-tailed distributions describe a very different phenomena than light-tailed distributions and some of these differences are discussed here.

Firstly, in Section 3.1.3 the concept of scale invariance was presented. As stated it is a property related to the family of Pareto distributions and asymptotically to regularly varying distributions. It is in fact a property that is only present with heavy-tailed distributions. In every day life, we encounter phenomena, such as income, which can vary significantly within a population and do not have a typical scale, which means they are scale invariant. Income is a scale invariant variable and follows a heavy-tailed distribution. The height of the population, on the other hand, remains within certain lower and upper bounds and has a typical scale which we are all familiar with. Height is, hence, not scale invariant and is indeed a phenomena which has a light-tailed distribution.

The second interesting property is called the catastrophe principle by Nair et al. (2022). Essentially, the principle states that the explanation for a large sum is simply that one single large event occurred. The mathematical representation for this principle was already presented in Lemma 3.5 and it is commonly known as the principle of a single big jump. This property is strongly linked to heavy-tailed distributions, and more precisely to subexponential distributions. If the phenomena in question was light-tailed, the large sum would be explained, not by one single large event, but by many events slightly larger than average.

The last property, and perhaps the less intuitive, is related to the mean residual life of an event which was already briefly discussed in reference to long-tailed distributions. Intuitively one would think that after waiting for some time the expected remaining waiting time should be less than in the beginning since one has already waited for some time. For events that have a heavy-tailed distribution, this may not always be true. In fact, heavy-tailed distributions often have an increasing mean residual life as will be touched upon in the next chapter, which means that the waiting time increases although one has already waited for some time. This seemingly very counter-intuitive property only appears with heavy-tailed phenomena.

From a more mathematical perspective, there is also another significant difference between heavy-tailed and light-tailed distributions. We know based on Lemma 3.1 that heavy-tailed distributions have infinite moment generating functions. In practice this means that the moment generating function is not as a useful tool for heavy-tailed distributions as it is for light-tailed distributions and, hence, heavy-tailed distributions require a different set of mathematical tools.

4 Mean excess function

4.1 The mean excess function

The mean excess function, which is also known as the mean residual life function in the insurance context, is a useful tool when assessing the tail of a distribution. It was originally introduced in the 1960s and later studied by many others, such as Hall and Wellner (1981). The mean excess function for a random variable X is defined as follows.

Definition 4.1. Suppose X is a strictly positive random variable with a right-continuous distribution function F and $\mathbb{P}(X > t) > 0$ for all $t > 0$. We assume $F(0) = 0$, then the mean excess function for X is given by

$$e_X(t) = \mathbb{E}[X - t | X > t] \tag{4.1}$$

for $t \geq 0$ and where $e_X(t) = 0$ when $\overline{F}(t) = 0$.

The mean excess function has many applications and can be interpreted as a function which describes the mean or expected remaining lifetime of an item which has survived to time x. The mean excess variable is a left truncated and shifted variable since it disregards the observations below x truncating the observations and since x is subtracted from the values of X it becomes a shifted variable.

Lemma 4.1. *Suppose X is a strictly positive random variable with the density function $f(x)$ and $\mathbb{E}(X) < \infty$. Then, the mean excess function can be presented as*

$$e_X(t) = \frac{1}{\overline{F}(t)} \int_t^\infty \overline{F}(x)\, dx. \tag{4.2}$$

Proof. The Definition 4.1 can be expressed as

$$e_X(t) = \frac{\mathbb{E}[(X - t)\mathbb{1}_{X>t}]}{\mathbb{P}(X > t)} = \frac{\int_t^\infty (x - t) f(x)\, dx}{1 - F(t)}.$$

Now when taking $f(x) = -\frac{d}{dx}\overline{F}(x)$ and integrating by parts, we get

$$\begin{aligned} e_X(t) &= \frac{\int_t^\infty (x - t) f(x)\, dx}{\overline{F}(t)} \\ &= \frac{-\big|_t^\infty (x - t)\overline{F}(x) + \int_t^\infty \overline{F}(x)\, dx}{\overline{F}(t)} \\ &= \frac{1}{\overline{F}(t)} \int_t^\infty \overline{F}(x)\, dx. \end{aligned}$$

as $-\Big|_t^\infty (x - t)\overline{F}(x)$ is equal to zero. □

Example 4.1. By using the tail function of the Pareto distribution presented in previous examples we can derive the mean excess function using the presentation given in Lemma 4.1 as follows:

$$\begin{aligned} e(x) &= \int_0^\infty \frac{\overline{F}(x+y)}{\overline{F}(x)}\, dy \\ &= \int_0^\infty \frac{\left(\frac{\theta}{x+y+\theta}\right)^\alpha}{\left(\frac{\theta}{x+\theta}\right)^\alpha}\, dy \\ &= (x+\theta)^\alpha \int_0^\infty (x+y+\theta)^{-\alpha}\, dy \\ (x+\theta)^\alpha &\Big|_0^\infty \frac{(x+y+\theta)^{-\alpha+1}}{-\alpha+1} \\ (x+\theta)^\alpha &\frac{(x+\theta)^{-\alpha+1}}{\alpha-1} = \frac{x+\theta}{\alpha-1}. \end{aligned}$$

The mean excess function has many interesting properties. Hall and Wellner (1981) presented the many properties in their article from 1981. Some of these properties and their proofs are presented in Proposition 4.1 following the presentation by Hall and Wellner.

Proposition 4.1. *(Properties of the mean excess function) Suppose X is a strictly positive random variable with density function f and right-continuous distribution function F. Suppose $\mathbb{E}(X) = \mu < \infty$. T is the endpoint of the distribution as given in Definition 2.11.*

(i) *$e(x)$ is non-negative and right-continuous and $e(0) = \mu > 0$.*

(ii) *$v(x) = e(x) + x$ is non-decreasing.*

(iii) *$e(x)$ has left limits everywhere in $(0, \infty)$ and jumps upwards at discontinuities.*

(iv) *If $T < \infty$, then $e(x-) > 0$ for $x \in (0, T)$, $e(T-) = 0$ and $e(x)$ is continuous at T.*

(v) $\overline{F}(x) = \frac{e(0)}{e(x)} \exp\big\{ -\int_0^x \frac{1}{e(u)}\, du\big\}$.

(vi) $\int_0^x \frac{1}{e(u)}\, du \to \infty$ *as* $x \to T$.

Proof. (i) Part (i) of Proposition 4.1 follows from Equation (4.2). In the equation both the numerator and the denominator are positive or non-negative which leads to $e(x)$ being non-negative. Moreover, the numerator in Equation (4.2) is continuous as an indefinite integral and the denominator is right-continuous as the inverse of the tail function. Hence, $e(x)$ is right-continuous. The last part can be proved by using the presentation given in Lemma 4.1. Knowing that X is a strictly positive random variable, we get

$$e(0) = \frac{1}{\overline{F}(0)} \int_0^\infty \overline{F}(x)\, dx = \int_0^\infty \overline{F}(x)\, dx = \mathbb{E}(X) = \mu.$$

(ii) To prove part (ii) we can look at the differential $v(x+t) - v(x)$ for $t > 0$. By rewriting the differential using the presentation in Equation (4.2), we get

$$\begin{aligned} v(x+t) - v(x) &= e(x+t) + x + t - e(x) - x \\ &= \frac{1}{\overline{F}(x+t)} \int_{x+t}^\infty \overline{F}(u)\, du + x + t - \left(\frac{1}{\overline{F}(x)} \int_x^\infty \overline{F}(u)\, du + x \right) \\ &= \frac{1}{\overline{F}(x+t)} \int_{x+t}^\infty \overline{F}(u)\, du + t - \frac{1}{\overline{F}(x)} \int_x^\infty \overline{F}(u)\, du \\ &\geq \frac{1}{\overline{F}(x)} \left(\int_{x+t}^\infty \overline{F}(u)\, du - \int_x^\infty \overline{F}(u)\, du \right) + t \\ &= -\frac{1}{\overline{F}(x)} \int_x^{x+t} \overline{F}(u)\, du + t \end{aligned}$$

which is always ≥ 0. Hence, $v(x)$ is non-decreasing.

(iii) To prove part (iii) we need to look at the right side of Equation (4.2). We know that the tail function $\overline{F}(x)$ has left limits in $(0, \infty)$ and similarly has the inverse $\frac{1}{\overline{F}(x)}$. Moreover, since $\overline{F}(x)$ has left limits in $(0, \infty)$, then the integral $\int_x^\infty \overline{F}(t)\, dt$ also has left limits. By combining these, we know that $e(x)$ also has left limits in $(0, \infty)$.

When a function jumps upwards in the discontinuity points, it means that the function has positive increments at discontinuities. This means that if one approaches a discontinuity point u from the left and from the right, the difference between these values is positive. In a more mathematical way, this can be proved using the proof for part (ii). Let $u > 0$ be the discontinuity point.

$$\lim_{x \to u+} e(x) - \lim_{x \to u-} e(x) = \lim_{x \to u+} e(x) + u - \lim_{x \to u-} e(x) - u = \lim_{x \to u+} v(x) - \lim_{x \to u-} v(x) \geq 0.$$

(iv) When $x \in (0,T)$ and $T < \infty$, then the limit from the left is equal to

$$e(x-) = \frac{1}{\overline{F}(x-)} \int_{x-}^{\infty} \overline{F}(u)\,du = \frac{1}{\overline{F}(x-)} \int_{x-}^{T} \overline{F}(u)\,du > 0.$$

Moreover, if $x \leq T$, then $v(x) \leq v(T)$ since $v(x)$ is a non-decreasing function as proved in part (ii). Then it follows that $e(x) \leq T - x$. When $x \to T$, then $e(x) = 0$. We know that $e(T) = 0$ and since $e(x)$ is a non-negative and right-continuous function as proved in part (i), it must hold that $e(T-) = 0$.

The function $e(x)$ is continuous at T since the one-sided limits are the same, that is

$$e(T+) = 0 = e(T-).$$

as $e(T+) = e(T)$.

(v) To prove part (v), we write $k(x) = \int_x^{\infty} \overline{F}(t)\,dt = \overline{F}(x)e(x)$ and $\overline{F}(x) = -\frac{d}{dx}\int_x^{\infty} \overline{F}(t)dt$. Then, it follows

$$\begin{aligned}
\int_0^u \frac{1}{e(x)}\,dx &= \int_0^u \frac{\overline{F}(x)}{\int_x^{\infty} \overline{F}(t)\,dt}\,dx \\
&= -\int_0^u \frac{\frac{d}{dx}\int_x^{\infty} \overline{F}(t)dt}{\int_x^{\infty} \overline{F}(t)\,dt}\,dx \\
&= -\int_0^u \mathrm{d}\log k(t)\,dt \\
&= -(\log k(u) - \log k(0)) \\
&= -\log\left(\frac{k(u)}{k(0)}\right) \\
&= -\log\left(\frac{\overline{F}(u)e(u)}{e(0)}\right)
\end{aligned}$$

when $\overline{F}(0) = 1$.

If we take the exponent on both sides of the equation

$$\int_0^u \frac{1}{e(x)}\,dx = -\log\left(\frac{\overline{F}(u)e(u)}{e(0)}\right)$$

and rearrange the equation, we get

$$\overline{F}(u) = \frac{e(0)}{e(u)} \exp\left\{ -\int_0^u \frac{1}{e(x)}\,dx \right\}.$$

(vi) In the proof for part (v), we used the equation

$$\int_0^u \frac{1}{e(x)}\,dx = -\log\left(\frac{k(u)}{k(0)}\right).$$

When $u \to T$, then $k(u) \to 0$ since $k(u) = \int_u^\infty \overline{F}(x)\,dx$. Consequently,

$$\lim_{u\to T} \int_0^u \frac{1}{e(x)}\,dx = -\log(0) = \infty.$$

□

The mean excess function for a regularly varying distribution is especially interesting and explains why the mean excess function is a popular tool in identifying heavy-tailed distributions. The following proposition sheds light into the useful asymptotic property of the mean excess function for regularly varying random variables.

Proposition 4.2. *Suppose a positive random variable X has a regularly varying distribution F at index $\alpha > 1$. Then,*

$$e_X(x) \sim \frac{x}{\alpha - 1} \qquad \text{as } x \to \infty.$$

Proof. Since F is regularly varying, we can use the fact that the tail $\overline{F}$ is also regularly varying with the same index α (Cooke et al. (2011)). This means that there exists a slowly varying function $l(x)$, as defined in Chapter 2, such that

$$\overline{F}(x) = x^{-\alpha} l(x).$$

By using the presentation given in Definition 4.1, we can write

$$\begin{aligned} e_X(x) &= \frac{1}{\overline{F}(x)} \int_x^\infty \overline{F}(t)\,dt \\ &= \frac{1}{x^{-\alpha} l(x)} \int_x^\infty t^{-\alpha} l(t)\,dt. \end{aligned}$$

By applying Karamata's theorem, we get

$$\int_x^\infty t^{-\alpha} l(t)\,dt \sim -(-\alpha+1)^{-1} t^{-\alpha+1} l(t) \quad \text{as} \quad x \to \infty$$

and, consequently, as $x \to \infty$ we can write

$$\begin{aligned} e_X(x) &= \frac{x^{\alpha}}{l(x)} \times (\alpha - 1)^{-1} x^{-\alpha+1} l(x) \\ &= \frac{x^{\alpha-\alpha+1}}{(\alpha - 1)} \\ &= \frac{x}{\alpha - 1}. \end{aligned}$$

Hence, we conclude

$$e_X(x) \sim \frac{x}{\alpha - 1} \qquad \text{as } x \to \infty.$$

□

Proposition 4.2 implies that regularly varying distributions can be used to model empirical distributions with linearly increasing or approximately linearly increasing mean excess functions which can be very useful information when choosing a model. The Pareto distribution is regularly varying as shown in Example 3.4 and has the mean excess function $\frac{\theta+x}{\alpha-1}$ as shown in Example 4.1. This is in line with what Proposition 4.2 suggests and it means that the Pareto distribution has a linear mean excess function with slope $\frac{1}{\alpha-1}$.

4.2 The hazard rate function

Another useful tool to identify a heavy-tailed distribution or examine the thickness of the tail of a distribution is the hazard rate function. The hazard rate function, or failure rate as it is sometimes referred to, is defined as the ratio between the density function and the tail function. In the insurance context, the hazard rate can be interpreted as the rate at which the risk of large claims decreases when x becomes larger.

Definition 4.2. Suppose X is a non-negative random variable with density function $f(x)$ and tail function $\overline{F}(x) > 0$. The hazard rate function $h(x)$ for X is given by

$$h(x) = \frac{f(x)}{\overline{F}(x)} \qquad \text{for } x \geq 0.$$

Example 4.2. The Pareto distribution will be used again in this example. From previous examples, we remember that the density function and tail function of the Pareto distribution are defined as

$$f(x) = \frac{\alpha\theta^{\alpha}}{(x+\theta)^{\alpha+1}} \quad \text{and} \quad \overline{F}(x) = \left(\frac{\theta}{x+\theta}\right)^{\alpha}$$

respectively, where $\alpha > 0$ is the shape parameter and $\theta > 0$ is the scale parameter. By applying Definition 4.2, the hazard rate function of the Pareto distribution is given by

$$h(x) = \frac{f(x)}{\overline{F}(x)} = \frac{\alpha\theta^{\alpha}(x+\theta)^{-\alpha-1}}{\theta^{\alpha}(x+\theta)^{-\alpha}} = \frac{\alpha}{x+\theta}.$$

The hazard rate function is closely related to the risk function $R(x)$ presented in Definition 3.2. The hazard rate function can be obtained as the derivative of the risk function as follows

$$R'(x) = -\frac{d}{dx}\log\overline{F}(x) = -\frac{1}{\overline{F}(x)}\overline{F}'(x) = \frac{f(x)}{\overline{F}(x)} = h(x).$$

The hazard rate function is also closely related to the mean excess function. In fact, the hazard rate function can be presented using the mean excess function. By rewriting the mean excess function of Lemma 4.1, we get

$$\overline{F}(x)e(x) = \int_x^{\infty} \overline{F}(t)\,dt.$$

Then by differentiating both sides and rearranging the the equation, we can write

$$\begin{aligned}
\overline{F}'(x)e(x) + \overline{F}(x)e'(x) &= -\overline{F}(x) \\
\frac{-\overline{F}'(x)}{\overline{F}(x)} &= \frac{1+e'(x)}{e(x)} \\
h(x) &= \frac{1+e'(x)}{e(x)}.
\end{aligned}$$

In many cases a distribution can be shown to be heavy-tailed with the help of the mean excess function and the hazard rate function. In fact, the mean excess function and the hazard rate function are closely linked in two ways. The first relationship is presented in the following proposition and the second will be presented in Proposition 4.4.

Proposition 4.3. *If the hazard rate function $h(x)$ is decreasing for all $x \geq 0$, then the mean excess function $e(x)$ is increasing for all $x \geq 0$.*

Proof. Firstly, we note that the mean excess function can be written as

$$e(x) = \int_0^{\infty} \exp\left[-\int_x^{x+t} h(y)\,dy\right] dt$$

when assuming both $e(x)$ and $h(x)$ exist (see Nair et al. (2022)). Then, we note that

$$\frac{\overline{F}(y+u)}{\overline{F}(u)} = \frac{\exp\left[-\int_0^{y+u} h(x)\,dx\right]}{\exp\left[-\int_0^{u} h(x)\,dx\right]} = \exp\left[-\int_u^{y+u} h(x)\,dx\right] = \exp\left[-\int_0^{y} h(u+t)\,dt\right].$$

Hence, if $h(x)$ is a decreasing function for a fixed y, then $\int_0^y h(u+t)\,dt$ is a decreasing function and, consequently, $\overline{F}(y+u)/\overline{F}(u)$ is an increasing function of d. According to Lemma 4.1, the mean excess function can be written as

$$e(u) = \frac{\int_u^\infty \overline{F}(x)\,dx}{\overline{F}(u)} = \int_0^\infty \frac{\overline{F}(y+u)}{\overline{F}(u)}\,dy.$$

Therefore, if the hazard rate function $h(u)$ is a decreasing function, then the mean excess function $e(u)$ is an increasing function of u since we showed that it holds with $\overline{F}(y+u)/\overline{F}(u)$ for a fixed y. Similarly, it holds that if the hazard rate function is an increasing function for a fixed y, then the mean excess function is a decreasing function. □

However, the inverse does not necessarily hold as there are distributions that have decreasing mean excess functions but do not have increasing hazard rate functions for all values of the variable as presented in the following example.

Example 4.3. Let the random variable X have a density function of

$$f(x) = (1+2x^2)e^{-2x}$$

for all $x \geq 0$. Consequently, the distribution function for X is equal to

$$F(x) = \int_0^x (1+2t^2)e^{-2t}\,dt = 1 - e^{-2x}(x^2+x+1)$$

and the corresponding tail function is

$$\overline{F}(x) = 1 - F(x) = e^{-2x}(x^2+x+1).$$

Now using the Lemma 4.1 and Definition 4.2, we can solve the mean excess function and hazard rate function for X as follows:

$$e(x) = \frac{1}{e^{-2x}(x^2+x+1)} \int_x^\infty e^{-2t}(t^2+t+1)\,dt = \frac{x^2+2x+2}{2x^2+2x+2}$$

and

$$h(x) = \frac{f(x)}{\overline{F}(x)} = \frac{e^{-2x}(1+2x^2)}{e^{-2x}(x^2+x+1)} = \frac{1+2x^2}{x^2+x+1}.$$

Moreover, we see that the first derivative of the mean excess function is

$$e'(x) = -\frac{x^2 + 2x}{2(x^2 + x + 1)^2}$$

which gets negative values for all $x \geq 0$. This implies that the mean excess function is a strictly decreasing function when $x \geq 0$. However, the first derivative for the hazard rate function is equal to

$$h'(x) = \frac{2x^2 + 2x - 1}{x^2 + x + 1}.$$

We notice that $h'(x)$ does not get positive values for all $x \geq 0$ and, hence, $h(x)$ is not strictly increasing. The two functions are presented in Figure 3 and it is clearly visible from the graph that $h(x)$ is not increasing for all values of x although $e(x)$ is strictly decreasing.

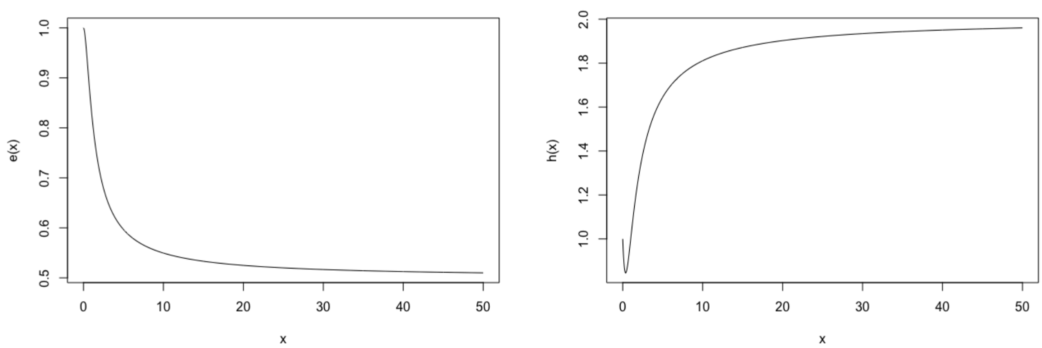

Figure 3: The mean excess function and hazard rate function presented in Example 4.3

In the context of heavy-tailed distributions, we are interested in understanding how the mean excess function and hazard rate function behave for heavy-tailed distributions. We would be tempted to think a decreasing hazard rate function and an increasing mean excess function proves a distribution to be heavy-tailed but this statement is not true for all distributions and all values of the random variable (Nair et al. (2022)). Nevertheless, it is true or asymptotically true that most distributions with decreasing hazard rate functions and increasing mean excess functions are heavy-tailed. In literature, this relationship is widely applied and even used to define heavy-tailed distributions. The following example shows that indeed all heavy-tailed distributions do not have a decreasing hazard rate function and increasing mean excess function.

Example 4.4. The Burr distribution is a commonly known heavy-tailed distribution and more specifically a regularly varying distribution with index $-ck$ (see Nair et al. (2022)). The Burr distribution's density function and distribution function are given by

$$f(x) = \frac{ck(x/\lambda)^c}{x[1 + (x/\lambda)^c]^{k+1}} \quad \text{and} \quad \overline{F}(x) = (1 + (x/\lambda)^c)^{-k}$$

where $x > 0$ and λ, c and k are positive parameters. Suppose that $c = 2, k = 2$ and $\lambda = 1$. Following Definition 4.2, the corresponding hazard rate function is given by

$$h(x) = \frac{4x^2}{x(1 + x^2)^3}.$$

The distribution is heavy-tailed but as $c > 1$, the hazard rate function is not decreasing for all $x > 0$ as Figure 4 presents.

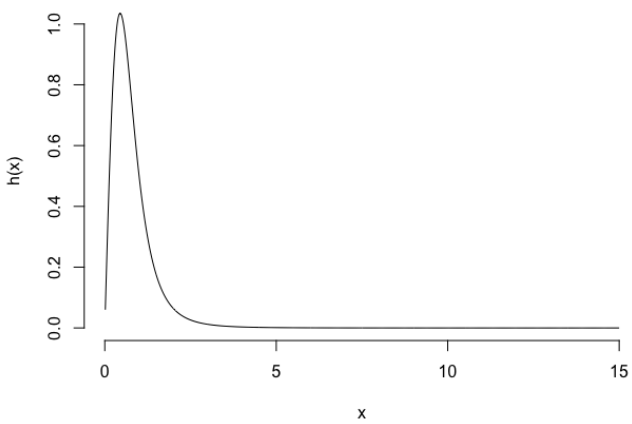

Figure 4: The hazard rate function for the Burr distribution with parameters $c = k = 2$ and $\lambda = 1$.

The mean excess function is more complex and is therefore not presented here. Nonetheless, as Nair et al. (2022) state, when $c > 1$ the mean excess function is not increasing for all $x > 0$. Hence, a distribution can be heavy-tailed although it does not have a decreasing hazard rate function and an increasing mean excess function.

The hazard rate function's property can be visually confirmed with the help of the risk function. As stated previously, the hazard rate function is the derivative of the risk function. If the hazard rate function is decreasing, it has a negative derivative. This consequently means that the risk function has a negative second derivative and is a concave function. Figure 5 presents the risk functions for heavy-tailed, light-tailed and exponential distributions. The heavy-tailed distributions chosen for Figure 5 are the Weibull distribution with shape parameter equal to 0.5 and scale parameter equal to 1 and the Pareto distribution with shape and scale parameters equal to 1. The light-tailed distribution is the Weibull distribution with shape parameter equal to 1.5 and scale parameter equal to 1.

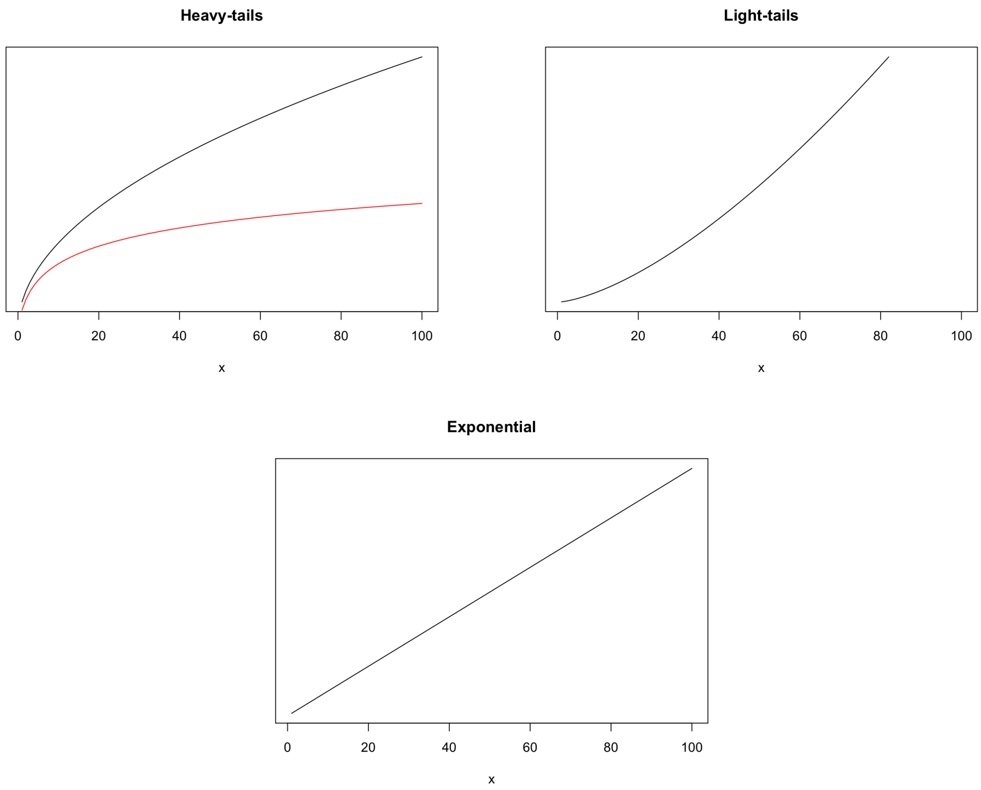

Figure 5: Risk functions for the heavy-tailed, light-tailed and exponential distributions.

The graph clearly shows that the heavy-tailed distributions presented in the graph on the left side, have concave risk functions and in the graph on the right the light-tailed distribution has a convex risk function. Furthermore, the exponential distribution, which is neither heavy-tailed nor light-tailed has a risk function which is neither convex nor concave. The statement can also be justified by examining the asymptotic behaviour of the mean excess function and the hazard rate for different distributions. When looking at the asymptotic behaviour of random variables, there is a clear link between the mean excess function and the hazard rate function as presented by Calabria and Pulcini (1987).

Proposition 4.4. *Suppose X is a non-negative random variable with the mean excess function $e(x)$ and the hazard rate function $h(x)$. When $x \to \infty$, it holds*

$$\lim_{x\to\infty} e(x) = \lim_{x\to\infty} \frac{1}{h(x)}. \tag{4.3}$$

Proof. Suppose $f(x)$ is the density function and $\overline{F}(x)$ the tail function of X. The proposition can be proved by applying l'Hôpital's rule to the mean excess function as it is presented in Equation (4.2). Then,

$$\lim_{x\to\infty} e(x) = \lim_{x\to\infty} \frac{\int_x^\infty \overline{F}(t)\,dt}{\overline{F}(x)} = \lim_{x\to\infty} \frac{\frac{d}{dx}\int_x^\infty \overline{F}(t)\,dt}{\frac{d}{dx}\overline{F}(x)} = \lim_{x\to\infty} \frac{-\overline{F}(x)}{-f(x)} = \lim_{x\to\infty} \frac{1}{h(x)}.$$

□

In their article Calabria and Pulcini (1987) presented the following table which clearly shows the asymptotic values of the mean excess function $e(x)$ and the hazard rate function $h(x)$ being inverse for all continuous distributions presented in Table 1.

Distribution	$\lim_{x\to\infty} h(x)$	$\lim_{x\to\infty} e(x)$
Exponential	α	$1/\alpha$
Weibull ($\beta > 1$)	∞	0
Weibull ($\beta < 1$)	0	∞
Gamma	α	$1/\alpha$
Normal	∞	0
Log-normal	0	∞
Inverse Gaussian	$\beta/2\alpha^2$	$2\alpha^2/\beta$

Table 1: Asymptotic values of the hazard rate function and the mean excess function

It can be seen from the table that $e(x) \to \infty$ and $h(x) \to 0$ as $x \to \infty$ for all heavy-tailed distributions, i.e. the log-normal distribution and the Weibull distribution with

$\beta < 1$ as Proposition 4.4 implies. And conversely, $e(x) \to 0$ and $h(x) \to \infty$ as $x \to \infty$ for all light-tailed distributions such as the normal distribution and the Weibull distribution with $\beta > 1$.

Example 4.5. In Example 4.2, the hazard rate function for the Pareto distribution was equal to

$$h(x) = \frac{\alpha}{x + \theta}.$$

It is a decreasing function when $\alpha > 1$ since

$$h'(x) = -\frac{\alpha}{(x + \theta)^2} < 0$$

for all x. If we look at the limit when $x \to \infty$, we get

$$\lim_{x\to\infty} h(x) = \lim_{x\to\infty} \frac{\alpha}{x + \theta} = 0.$$

The mean excess function for the Pareto distribution is

$$e(x) = \frac{\theta + x}{\alpha - 1}$$

which is an increasing function when $\alpha > 1$ since

$$e'(x) = \frac{1}{\alpha - 1} > 0.$$

If we look at the limit when $x \to \infty$, we notice that

$$\lim_{x\to\infty} e(x) = \lim_{x\to\infty} \frac{\theta + x}{\alpha - 1} = \frac{\infty}{\alpha - 1} = \infty$$

when $\alpha > 1$. Hence, the Pareto distribution behaves just as expected from a heavy-tailed distribution. The hazard rate function is strictly decreasing and the mean excess function is strictly increasing. As stated previously this is the typical behaviour of a heavy-tailed distribution and as we know the Pareto distribution is heavy-tailed based previous examples.

Example 4.6. The exponential distribution is considered as being neither heavy-tailed nor light-tailed and both the mean excess function and the hazard rate function support this notion. Following Definition 4.2, the hazard rate function for the exponential distribution with parameter $\lambda > 0$ is given by

$$h(x) = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda.$$

The mean excess function for the exponential distribution is

$$e(x) = \frac{1}{\lambda}.$$

We notice that both the hazard rate function and the mean excess function for the exponential distribution are constants which do not depend on x, which supports the notion that the exponential distribution is neither heavy-tailed nor light-tailed. Moreover, it is commonly known that the exponential distribution is a memoryless distribution which implies that the mean excess function and hazard rate function do not depend on x. We also notice that

$$\lim_{x\to\infty} e(x) = \frac{1}{\lambda} = \lim_{x\to\infty} \frac{1}{h(x)}.$$

as stated in Proposition 4.4.

4.3 The mean excess plot

The mean excess plot is a graphical tool which can be used to portray and understand the tail of the distribution of the data. It also serves as a tool to visually distinguish heavy-tailed distributions from light-tailed distributions. The mean excess plot is essentially a graphical, empirical presentation of the mean excess function of a random variable. The mean excess plot is typically used for data of independent and identically distributed random variables to visually confirm that the data follows a generalized Pareto distribution or any other heavy-tailed distribution.

Definition 4.3. A random variable X with generalized Pareto distribution has the following distribution function

$$F(x) = \begin{cases} 1 - (1 + \frac{\xi x}{\beta})^{\frac{-1}{\xi}}, & \xi \neq 0 \\ 1 - e^{\frac{1}{\beta}}, & \xi = 0 \end{cases}$$

where ξ and β are the shape and scale parameters respectively.

Now for a random variable X with a generalized Pareto distribution, when $\xi < 1$ and consequently $\mathbb{E}(X) < \infty$, we have the following mean excess function,

$$M(u) = \frac{\beta}{1-\xi} + \frac{\xi}{1-\xi} u$$

where $u \in [0, \infty)$ if $\xi \in [0, 1)$ and $u \in [0, -\beta/\xi)$ if $\xi < 0$. The mean excess plot is formed by plotting the values of the i.i.d. random variable X from the sample of size n with the values of the empirical mean excess function $\hat{M}(u)$, which is defined as

$$\hat{M}(u) = \frac{\sum_{i=1}^{n} (X_i - u) I_{[X_i > u]}}{\sum_{i=1}^{n} I_{[X_i > u]}}, \quad u \geq 0.$$

If the data follows a generalized Pareto distribution, the mean excess plot looks approximately linear (see Carmona (2014), Ghosh et al. (2011)). Linearity appears especially with high values of the threshold. This linearity is a property that only the family of generalized Pareto distributions have.

Although the mean excess plot is strongly linked to the generalized Pareto distribution, it can generally be used to distinguish heavy-tailed distributions from light-tailed distributions. It was stated previously that many heavy-tailed random variables have increasing mean excess functions and this can be seen from mean excess plots as the plots for heavy-tailed random variables tend to have a positive slope.

5 Heavy tails and financial risk measures

5.1 Risk measures in general

In the financial sector, managing risks plays a key role for many reasons. Firstly, sound risk management provides security, stability and smoothness to the financial sector which is highly desired to ensure that the banking and insurance sectors work as expected. Secondly, proper risk management, when adequately organized, can increase the value of a company which naturally is in the shareholders' interest. Finally, risk management plays a key role in companies maintaining an adequate amount of economic capital. Moreover, modern financial theory focuses strongly on the trade-off between risk and return and understanding risks is key for achieving the desired returns.

Risk management can be both qualitative and quantitative. In quantitative risk management, risk measures are highly important. Institutions such as banks, insurance companies or some times asset managers are especially interested in the downside risk of their asset portfolios. This means that the companies, as well as other market participants or regulators, aim to model and measure the magnitude and probability of potential loss related to their portfolios as risk is defined as the chance of a loss. For this purpose, quantitative risk management plays a key role and quantitative risk measures are needed. Let's first present the definition of a risk measure as given by McNeil et al. (2005).

Definition 5.1. A risk measure $\rho(X)$ is a function $\rho : \mathcal{M} \to \mathbb{R}$ where $X \in \mathcal{M}$ is a random variable representing the risks or possible losses of a portfolio and $\mathcal{M}$ is a convex cone.

Risks are measured for many reasons. On one hand, companies or financial institutions can use risk measures as tools for the management to understand the risk exposure of the company or financial institution and, on the other hand, regulators want risks to be measured in order to determine how much capital financial institutions need to hold to attain a certain level of capital adequacy to be able to bear the risks they are exposed to. No matter what the reason for measuring risk is, the risk measure used should have certain desirable properties for it to be fit for the purpose. Such a risk measure is called a coherent risk measure.

Definition 5.2. A risk measure ρ is coherent if it satisfies the following axioms of coherence:

(i) *Monotonicity*: $\rho(X) \leq \rho(Y)$ when $X \leq Y$.

(ii) *Positive homogeneity*: $\rho(aX) = a\rho(X)$ for all $a > 0$.

(iii) *Translation invariance*: $\rho(X + a) = \rho(X) + a$ for all $a \in \mathbb{R}$.

(iv) *Subadditivity*: $\rho(X+Y) \leq \rho(X) + \rho(Y)$.

The axioms of a coherent risk measure are quite intuitive as McNeil et al. (2005) explain. Monotonicity is certainly the one many feel comfortable with. If a position X contains more risk than Y, then the risk measure $\rho(X)$ must give a greater value than $\rho(Y)$ and consequently the company should hold more capital against the greater risk exposure. In the same way, positive homogeneity seems logical since when the risks of a portfolio are multiplied, the same is expected to happen to the risk measure. In other words, the risk must scale with the size of the risk position. Subadditivity, on the other hand, supports the notion that risk can be reduced by diversifying one's position which is a fundamental principal in the financial context. In practice, if subadditivity would not be required from a risk measure, financial institutions could be split into smaller companies to optimize the regulatory capital requirements through their group structure as it would hold that $\rho(X+Y) > \rho(X) + \rho(Y)$. In practice, this makes no sense and does not explain in any way a decreased risk position. Translation invariance also presents an important notion of financial risk. Translation invariance states that by adding the amount a to the portfolio, the capital requirement should change by the exact same amount.

There are many different types of risk measures, and in this thesis we focus on risk measures which are based on loss distributions. Many problems can emerge if a company relies solely on a single risk measure which aims at summarizing the risk contained in a portfolio. Nevertheless, risk measures that are based on loss distributions can be extremely useful when used in the right way. Losses are of great interest in risk management and using the distribution of losses directly as the basis of the risk measure is very natural. Moreover, the use of loss distributions is a good choice for analysing aggregated risk, taking into account diversification effects and when comparing different portfolios. In the next sections two popular risk measures, Value-at-Risk (VaR) and expected shortfall (ES), are presented. Both are measures that are based on the loss distribution.

5.2 Value-at-Risk

Especially in the banking sector, VaR has become a standard risk measure and it has been included in the Basel II regulatory framework. VaR has also gained vast popularity in the entire financial sector. The popularity of VaR is mainly linked to its simplicity and its ease of calculation. VaR is a risk measure that provides a single figure which is easy to communicate, for example, to management. One cannot really talk about VaR without its important companion, the expected shortfall. VaR has its shortcomings and therefore ES has gained an important role in the financial sector. Essentially, VaR presents the maximum loss of a portfolio for a given time period assuming a certain fixed confidence level. Suppose X represents the portfolio losses and $\alpha \in (0,1)$ the confidence level. To clarify, we consider the distribution of losses or returns to contain the losses in the right tail of the distribution and portfolio losses to be positive values. We define VaR as follows.

Definition 5.3. When $x \in \mathbb{R}$ and $\alpha \in (0, 1)$, then

$$VaR_\alpha(x) = \inf\{x|\mathbb{P}(X > x) \leq 1 - \alpha\} = \inf\{x|\mathbb{P}(X \leq x) \geq \alpha\}.$$

We can see from the definition that VaR is simply a quantile of the loss distribution. VaR is most commonly calculated at a 95%, 99% or even higher confidence level and the time horizon can vary. The interpretation of VaR should be that one can be α percent certain that the portfolio will not lose more than the amount of money given by VaR in time T. Figure 6 shows the density function of a standard normal distribution and illustrates VaR and ES calculated for the distribution. Although VaR is typically calculated for higher confidence levels, in Figure 6 a 75% confidence level for both VaR and ES is chosen for illustrative purposes.

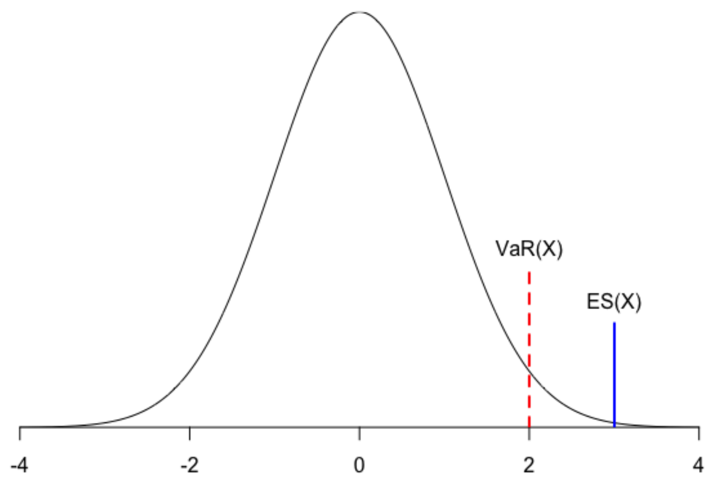

Figure 6: VaR and ES for the standard normal distribution with $\alpha = 0.75$

Example 5.1. Let F be the distribution function for the exponential distribution such that

$$F(x) = 1 - e^{-\lambda x}.$$

Mathematically in this case the quantile function is equal to the inverse of the distribution function, which can be expressed as $Q_x(p) = F^{-1}(x)$. By solving x from the previous equation with respect to p we get

$$\begin{aligned} p &= 1 - e^{-\lambda x} \\ e^{-\lambda x} &= 1 - p \\ -\lambda x &= \ln(1-p) \\ x &= \frac{\ln(1-p)}{-\lambda} \end{aligned}$$

which corresponds to the VaR of the exponential distribution at a confidence level of p.

Value-at-Risk can be calculated using three different methods. The historical simulation method is perhaps the most simple method as it is based purely on using an empirical data sample. The VaR is simply read as the order statistic of the distribution of the empirical sample. The second method for calculating VaR is the parametric method. The method assumes a certain distribution and the parameters required to calculate VaR as the α quantile of the assumed distribution are either estimated based on historical data or a more forward looking method. The final method for calculating VaR is the Monte Carlo simulation method. The method relies on simulating portfolio returns based on suitable risk factors and using the distribution of simulated returns to get the VaR value.

Despite being a commonly used risk measure, VaR has many shortcomings. Firstly, VaR is not subadditive and therefore is not a coherent risk measure as Definition 5.2 states. This can cause significant problems in practice since the VaR of a portfolio is not smaller or equal to the sum of the VaR values of its subportfolios. In other words, VaR does not aggregate risk in a logical way. Secondly, VaR undermines the losses beyond the VaR value as it represents the lowest bound of the losses. In other words, a VaR figure does not tell anything about the magnitude of losses exceeding the VaR level. This could potentially lead to catastrophic outcomes if a financial institution is not aware of this shortcoming. In practice, two portfolios can have the same VaR although the tails of the loss distributions of the portfolios are drastically different.

5.3 Expected Shortfall

Due to the shortcomings of VaR, there was a great need for a risk measure that is not only coherent but also easy to understand and calculate. This need was fulfilled by the expected shortfall. The expected shortfall is a risk measure that addresses the losses beyond the VaR level by definition and is therefore strongly related to VaR. The expected shortfall gives the conditional expectation of losses beyond the VaR level and, hence is more suitable than VaR for examining tail risk. The expected shortfall became popular due to its better properties as a risk measure but also due to the fact that ES looks further into the tail of the loss distribution which is especially important in relation to heavy-tailed loss distributions. The expected shortfall has many variations and can be found from literature under different names such as the Tail Condition Expectation or Conditional VaR (cVaR). The expected shortfall at the confidence level α is defined as follows.

Definition 5.4. For portfolio losses X with the distribution function F and $\mathbb{E}(X) < \infty$, we define the expected shortfall at a certain $\alpha \in (0,1)$ as

$$ES_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 q_u(F)\,du,$$

where $q_u(F)$ is the quantile function of the loss distribution F.

With reference to Definition 5.3 in which VaR was defined as a quantile of the loss distribution, we can present the expected shortfall in the following form:

$$ES_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(X)\,du.$$

For a continuous loss distribution, the expected shortfall is the conditional expectation of losses given that the portfolio losses exceed the VaR level. For a continuous loss distributions with X representing the portfolio losses and $VaR_\alpha(X) \in \mathbb{R}$ being the corresponding VaR, ES can be presented in an even more intuitive manner which shows that ES is indeed the expected loss over the VaR. The proof for Lemma 5.1 will be presented later in this chapter.

Lemma 5.1. *For the losses X with a continuous distributions function F and any $\alpha \in (0,1)$, the expected shortfall corresponds to*

$$ES_\alpha(X) = \mathbb{E}[X|X \geq VaR_\alpha(X)].$$

Remark 5.1. It follows directly from Lemma 5.1 that for positive losses X at any confidence level $\alpha \in (0,1)$, it holds that

$$ES_\alpha(X) \geq VaR_\alpha(X).$$

when α is the same for both VaR_α and $ES_\alpha(X)$. The expected shortfall is defined as a conditional expectation given that the losses X exceed $VaR_\alpha(X)$. Since the losses are positive, the expected value of Lemma 5.1 is always positive, and the inequality holds.

Example 5.2. This simple empirical example aims at showing how VaR and ES work in practice to better understand the two risk measures. This example follows closely the example presented by Koller (2011). Let's assume we have a data sample of 1000 observations. These observations could be for example losses of a portfolio of assets in thousands or millions of euros. In this case the losses are presented as positive values and the worst 10 observations are presented in the following table.

Number	**Observation**
1000	210
999	205
998	200
997	192
996	190
995	188
994	180
993	175
992	170
991	155

Now, let's say we want to determine the 99.6% VaR. Following Definition 5.3 we get that the VaR is 192, which is simply the 4^{th} observation out of 1000 starting from the worst observation. To get the 99.6% expected shortfall we need to look at the values within the worst 0.4% of the observations. The expected shortfall corresponds to the average of these values. Based on Lemma 5.1, we can calculate the 99.6% ES as follows

$$ES_{99.6\%} = \frac{210 + 205 + 200 + 192}{4} = 201.8.$$

The expected shortfall has become a popular risk measure partly due to the fact that ES is a superior risk measure to VaR with respect to tail properties since ES is more sensitive to the tail of the loss distribution than VaR. Moreover, ES is also a subadditive risk measure, and in fact a coherent risk measure unlike VaR.

Proposition 5.1. *The expected shortfall is a coherent risk measure.*

Proof. The proof can be found from Artzner et al. (1997). □

Despite ES being a coherent risk measure, it also has its shortcomings. Yamai at al. (2005) suggest that the accuracy of the results yielded by ES is highly dependent on the estimation method. They suggest that ES generates worse results than VaR if a simulation method is used for the estimation. For heavy-tailed distributions the estimation error is even larger. This is due to a practical matter. When simulating data from a heavy-tailed distribution, the large and unlikely events are more rare and, hence, the ES results vary more. VaR, on the other hand, is not as strongly affected by this phenomenon, as it completely disregards the tail of the distribution. In general, the larger the sample, the smaller the estimation error for ES is.

As has already been shown through the definitions presented in this chapter, VaR and ES are closely related to one another. A third interesting and closely related risk measure is the excess loss over the VaR. It can be defined as follows.

Definition 5.5. The excess loss $e_\alpha(X)$ over VaR is given by

$$e_\alpha(X) = \mathbb{E}[X - VaR_\alpha(X)|X \geq VaR_\alpha(X)].$$

The definition of excess loss is very similar to the definition of the expected shortfall. Indeed, the excess loss can be written using VaR and ES.

Proposition 5.2. *Let $ES_\alpha(X)$ and $VaR_\alpha(X)$ be the expected shortfall and Value-at-Risk for the random variable X at the confidence level α. Then, the excess loss is equal to*

$$e_\alpha(X) = ES_\alpha(X) - VaR_\alpha(X).$$

Proof. The proof follows directly from the Definitions 5.5 and 5.1 and the rules of conditional expectation.

$$\begin{aligned} e_\alpha(X) &= \mathbb{E}[X - VaR_\alpha(X)|X \geq VaR_\alpha(X)] \\ &= \mathbb{E}[X|X \geq VaR_\alpha(X)] - \mathbb{E}[VaR_\alpha(X)|X \geq VaR_\alpha(X)] \\ &= \mathbb{E}[X|X \geq VaR_\alpha(X)] - VaR_\alpha(X) \\ &= ES_\alpha(X) - VaR_\alpha(X). \end{aligned}$$

□

Moreover, we observe that the expected shortfall can be obtained by adding the mean of the excess losses to VaR meaning that

$$ES_\alpha(X) = VaR_\alpha(X) + e_\alpha(X).$$

Hence, the expected shortfall is larger than VaR by the average of all losses exceeding VaR. The excess loss is also linked to the mean excess function presented in Chapter 4. Indeed, mathematically the excess loss is a mean excess function as can be seen from Definition 5.5.

Proposition 5.3. *Let e_X be the mean excess function for a random variable X, then the excess loss can be written as follows*

$$e_\alpha(X) = e_X(VaR_\alpha(X))$$

Proof. The proof follows directly from the definitions given previously. The Definition 5.5 gives the first equality and the second equality follows directly from the Definition 4.1 of the mean excess function in the following manner

$$e_\alpha(X) = \mathbb{E}[X - VaR_\alpha(X)|X \geq VaR_\alpha(X)] = e_X(VaR_\alpha(X)).$$

□

Now that the excess loss has been formally defined, we can finally present the proof for Lemma 5.1.

Proof. (Proof for Lemma 5.1).

By definition, we know that ES corresponds to the average of the losses exceeding VaR. With the help of the mean excess function, we can write

$$ES_\alpha(X) = VaR_\alpha(X) + e_X(VaR_\alpha(X)).$$

Furthermore, by following Lemma 4.1, we can further write

$$\begin{aligned} &VaR_\alpha(X) + e_X(VaR_\alpha(X)) \\ &= VaR_\alpha(X) + \frac{\int_{x_\alpha}^{\infty}(x - VaR_\alpha(X))f(x)\,dx}{1-\alpha} \\ &= \mathbb{E}[X|X \geq VaR_\alpha(X)]. \end{aligned}$$

□

Both VaR and ES are risk measures based on the loss distribution. According to McNeil et al. (2005) these types of risk measures have two major problems. Firstly, the measures rely on past data and, hence, work under the assumption that the past is predictive of the future. Perhaps on average it is a satisfactory assumption, but when focusing on extreme events and the tail of the loss distribution, as both VaR and ES do, this can be a very deceptive assumption. Secondly, it is in general difficult to accurately estimate the loss distribution. This usually leads to the practical decision to assume some distribution, such as the normal distribution. This in turn often leads to model risk, which means that, in practice, there is a risk or chance of losses because the financial institution or company uses a risk measure which is based on a poorly specified model. Since VaR and ES are directly based on the loss distribution, assumptions related to the loss distribution are typically a source of model risk.

6 Simulation

In this section we simulate data from several different distributions, both light-tailed and heavy-tailed, to explore the topics of this thesis in a more practical light. Moreover, we aim to give some additional insights into the risk measures presented in Chapter 5 by calculating the measures based on the simulated data.

6.1 Summary of the data

Data samples of the size of 100 000 observations were simulated from the following distributions:

(i) Normal distribution with mean 1 and standard deviation 1;

(ii) Exponential distribution with parameter 1;

(iii) Log-normal distribution with parameters 0 and 1; and

(iv) Weibull distribution with shape parameter 0.5 and scale parameter equal to 1.

The normal distribution is a light-tailed distribution whereas the exponential distribution is typically categorized as neither light-tailed nor heavy-tailed as previously stated. The log-normal distribution and the Weibull distribution with a shape parameter of 0.5 are both commonly known heavy-tailed distributions. The random samples were generated using the R software's functions for generating random samples. Histograms of the data samples are presented in Figure 7. In order to have sound and sensible results, the data samples were normalized by dividing each data point by the mean of the random sample in order to have samples which are independent of their units of measurement. The same method was used for all samples and each sample has a mean of 1.

The histograms presented in Figure 7 give an adequate representation of each distribution as a whole and gives some insight on the shapes of light-tailed and heavy-tailed distributions. However, the larger values in the right tails of the distributions do not appear because there are so few large values compared to other values and, hence, based on histograms it is difficult to infer anything regarding the tails of the distributions. To better understand the tails of the sample data, we will next look at the mean excess plots generated based on the sample data.

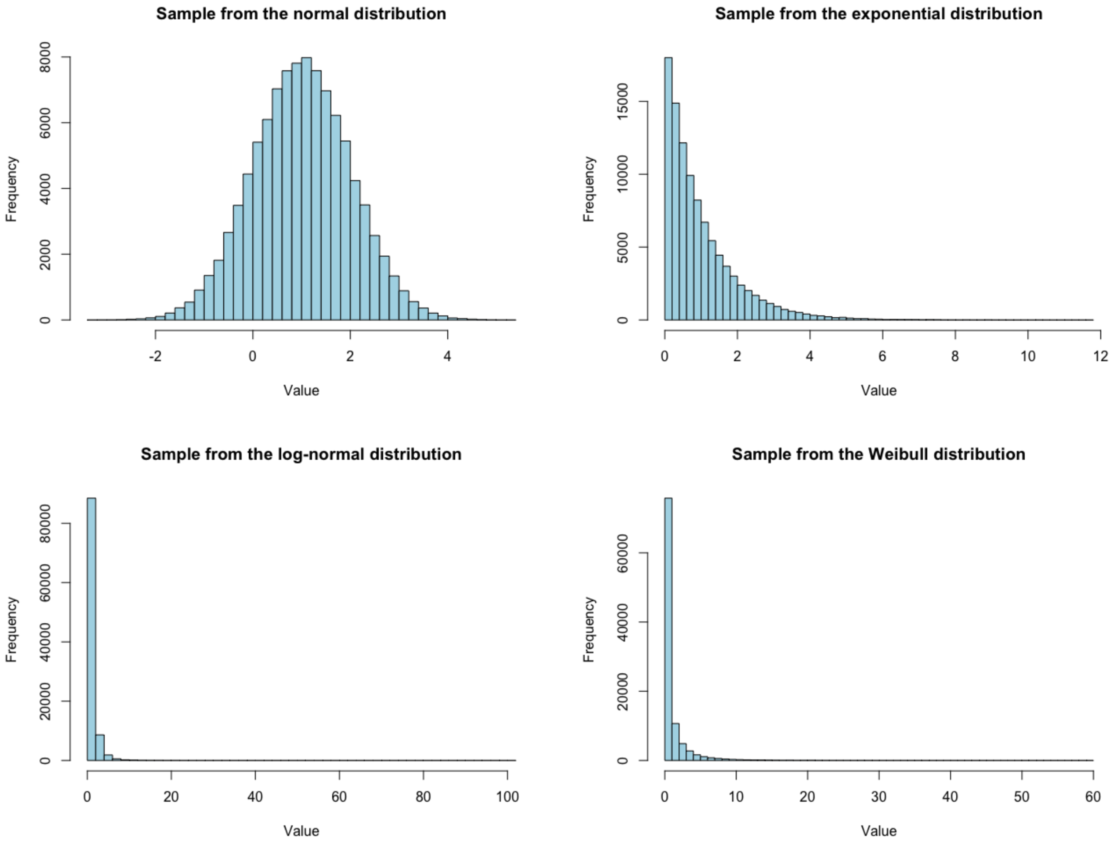

Figure 7: Histograms of data samples

6.2 Mean excess plots

In Chapter 4, it was stated that the mean excess function can be used to distinguish heavy-tailed distributions from light-tailed distributions. In this subsection, the mean excess plots for the different samples presented in the previous subsection are shown. The plots were calculated using R software's *qrmtools* package (Hofert et al.(2022)). Typically as the threshold value increases the mean excess plot scatters due to the small number of observations in the tails of the distributions. It is common practice to omit a few of the largest observations in order to obtain a graph which is easier to interpret. In this case, 10 of the largest values were omitted for each sample data. Let's first explore the mean excess plots of the data samples generated from the normal distribution and the exponential distribution. The plots are presented in Figure 8.

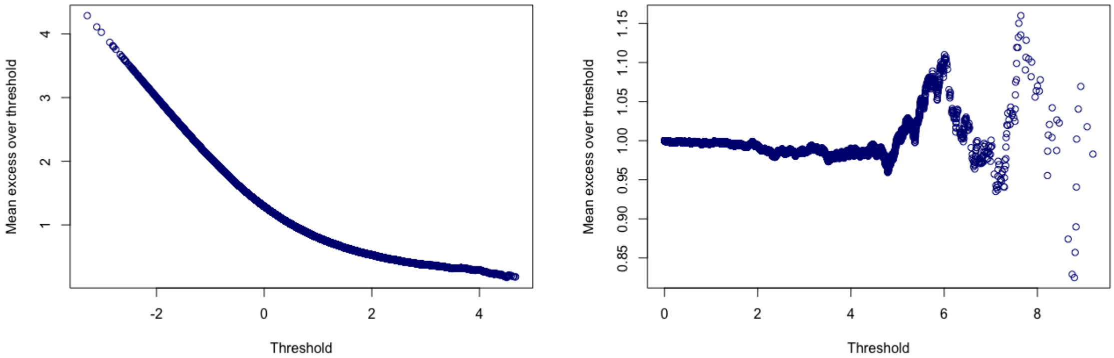

Figure 8: The mean excess plots of the normal (left) and exponential sample data (right).

It was stated in Chapter 4 that a heavy-tailed distribution typically has an increasing mean excess function and, hence, a mean excess plot with positive slope. It can be seen from Figure 8 that the mean excess plot for the sample data generated from a normal distribution behaves as we expect based on what was presented in Chapter 4. The plot has a distinctively negative slope which is typical for a light-tailed distribution. It was also previously stated that the exponential distribution is neither heavy-tailed nor light-tailed which is also clearly visible in the mean excess plot. Despite the values scattering at higher threshold values, most of the plot resembles a flat line with neither positive nor negative slope. In the following Figure 9, the mean excess plots calculated from the samples generated from the log-normal and Weibull distributions are presented.

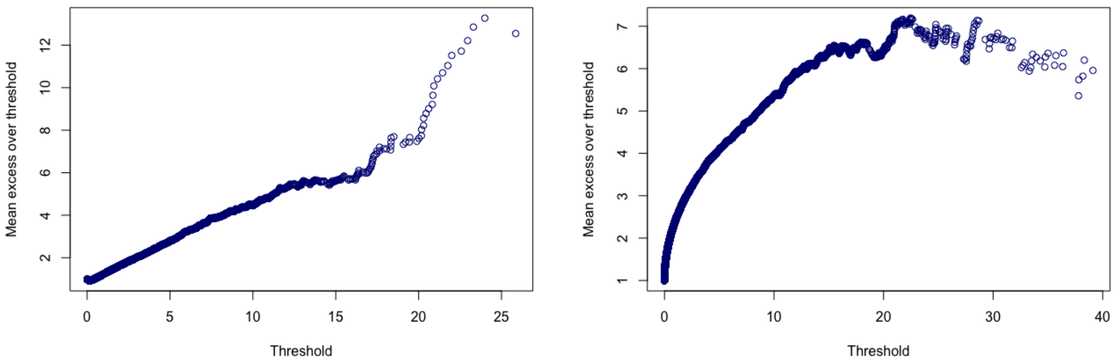

Figure 9: The mean excess plots of the log-normal (left) and Weibull sample data (right).

Again, the plots are in line with what was presented in Chapter 4. We know that both distributions are heavy-tailed and both plots have a clearly positive slope as expected. In all four plots, the lack of observations with higher threshold values creates a dispersion to the plot. Interestingly, this is barely visible in the light-tailed normal distribution presented in Figure 8 but a very present feature in the plots of the heavy-tailed distributions presented in Figure 9.

In this particular case, we know which distributions the random samples have been generated from and we can use the plots to confirm empirically what was stated in Chapter 4. In practice, one often doesn't know what distribution some empirical data follows and in those situations mean excess plots serve as a good tool to deduce whether the data follows a heavy-tailed, a light-tailed or an exponential distribution.

6.3 Risk measures

In this subsection we analyze the differences in the risk measures presented in Section 5 of this thesis. The values are calculated from the data samples described in Subsection 6.1. The aim is to highlight how large an effect the heavy-tail can have on VaR values and, hence, how significant a problem it is if one assumes the wrong distribution in practice.

When looking at VaR and ES in the context of heavy-tailed distributions, one interesting ratio to focus on is the ratio between the two risk measures, often referred to as the shortfall-to-quantile ratio.

Definition 6.1. The shortfall-to-quantile ratio is defined as

$$\lim_{\alpha \to 1} \frac{ES_\alpha}{VaR_\alpha}$$

where $\alpha \in (0, 1)$ represents the confidence level.

For a light-tailed distribution such as the normal distribution, the ratio is close to one meaning that the difference between the risk measures is small. For a heavy-tailed distribution the difference between VaR and ES is larger and the shortfall-to-quantile ratio is also larger. This result is explained by the fact that for heavy-tailed distributions the mean of the values exceeding VaR will be larger due to the extreme values in the tail. For a light-tailed distribution, the ratio will be close to one since there are no values far in the tail of the distribution by definition. Since ES_α is always larger than VaR_α when $\alpha \in (0, 1)$ is the same for both measures, the ratio will always be greater than one.

In the following tables, VaR and ES values are presented for all four distributions. Both risk measures are calculated with 95% and 99% confidence levels from the sample data. Moreover, the shortfall-to-quantile ratio is shown in both tables and is denoted by "SQ ratio". The VaR values were calculated according to Definition 5.3 by simply taking the quantile for each α of each data sample's distribution. The ES values correspond to the mean of the values exceeding VaR at each confidence level α as presented in Lemma 5.1.

Distribution	**VaR 95%**	**ES 95%**	**SQ ratio 95 %**
Normal	2.65	3.07	1.16
Exponential	2.99	3.98	1.33
Log-normal	3.12	5.18	1.66
Weibull	4.51	8.47	1.88

Table 2: Risk measures from sample data at a 95% confidence level.

Table 2 presents the risk measures at a 95% confidence level. One can see significant differences between the values calculated from data generated from different distributions and the differences between the light-tailed and heavy-tailed distributions are clear. For heavy-tailed distributions, both VaR and ES values are significantly larger than for light-tailed distributions. For the light-tailed normal distribution, the difference between the VaR and ES figures is not large and consequently the shortfall-to-quantile ratio is close to 1. For the heavy-tailed distributions, we notice the opposite and shortfall-to-quantile ratios approach 2. Moreover, it can be seen that for all values $ES_\alpha \geq VaR_\alpha$ as stated in Remark 5.1. In table 3, the risk measures at a 99% confidence level are presented.

Distribution	**VaR 99%**	**ES 99%**	**SQ ratio 99%**
Normal	3.32	3.67	1.10
Exponential	4.57	5.56	1.22
Log-normal	6.09	9.36	1.54
Weibull	10.59	16.00	1.51

Table 3: Risk measures from sample data at a 99% confidence level.

As the confidence level α increases, the results become even clearer. The shortfall-to-quantile ratios show the distinct difference between the heavy-tailed and light-tailed distributions. For the normal distribution, the 95 % shortfall-to-quantile ratio is close to 1 meaning that the 95% ES value is not very large compared to the 95% VaR. This means that the mean of all the values in the sample data in the last 5% are not very dispersed. This effect is even clearer in the 99% value presented in Table 3 than in the 95% value. For the Weibull distribution, on the other hand, the shortfall-to-quantile ratios are clearly larger which is explained by the heavier tail of the Weibull distribution compared to the normal distribution. The large differences in the shortfall-to-quantile ratios also highlights the practical problem of using only VaR as the sole risk measure to support decision-making. The heavier the tail the more important information about the tail risk is lost if one does not use, for example, both VaR and ES.

In the previous chapter it was mentioned that ES is a risk measure which is highly influenced by the size of the sample it is estimated from, especially for heavy-tailed random variables. This is due to the fact that a small sample size might not capture the extreme values from the tail of the distributions as these values are typically rare although large. The larger the sample, the better these values are included in the sample and, hence, the more accurate the ES estimate is. In the following Table 4, the 99% ES values for all four distributions are calculated with samples of 10 000, 50 000 and 100 000 observations simulated from the distributions in a similar way as described in the Subsection 6.1.

Distribution	ES_{10}	ES_{50}	ES_{100}
Normal	3.62	3.62	3.67
Exponential	5.62	5.65	5.56
Log-normal	9.07	9.00	9.36
Weibull	17.76	15.90	16.00

Table 4: The 99% expected shortfall with different sample sizes.

It can be seen from Table 4 that the sample size seems to affect especially the ES values calculated from the Weibull distribution and the log-normal distribution whereas the values based on the normal distribution and exponential distribution are quite close to one another. In this empirical analysis, it seems that the sample size has a stronger effect on the accuracy of ES figures derived from heavy-tailed distributions as stated by Yamai et al. (2005). When calculating ES figures from data simulated from the heavy-tailed distributions for the smaller data samples of 10 000 and 50 000 observations, the values had significant variation from one simulation to another. As mentioned previously, simulated data might not capture the tail of the distribution accurately and, hence, there can be significant differences in different data samples. With the data sample of 100 000 observations, the ES figures appeared more stable.

7 Conclusions

The aim of this thesis was to explain heavy-tailed distributions and examine financial risk measures, focusing on the two most popular ones VaR and ES, in the context of heavy-tailed distributions. The aim was also to shed light on why it is of fundamental importance to understand the differences between light-tailed and heavy-tailed distributions when carrying out any type of analysis, especially in the context of risk management.

Once the broad class of heavy-tailed distributions was thoroughly introduced and definitions and important properties for the subclasses were reviewed, it was highlighted how essentially different types of situations and phenomena light-tailed and heavy-tailed distributions describe. Chapter 4 presented the mean excess function and the hazard rate function which are both very interesting mathematical concepts when looking at heavy-tailed distributions and the relationship between the two functions was presented. It became clear that both the mean excess function and the hazard rate function behave in opposite ways for heavy-tailed and light-tailed distributions. In the following chapter, the two risk measures were thoroughly presented and it was shown how closely the mean excess function is tied to the two risk measures. Finally the measures were examined in the presence of heavy tails. In the last chapter the topics presented in previous chapters were examined in a more practical light through mean excess plots and risk measures were calculated based on data samples simulated from both light-tailed and heavy-tailed distributions.

In the context of financial markets and financial risk management, heavy-tailed phenomena are very common. It goes without saying that measuring and understanding risk in a correct manner is at the center of the trade-off between risk and return which plays a key role in modern financial theory. Therefore understanding the distribution of the returns, or losses, of financial assets is crucial. For both Value-at-Risk and the expected shortfall, disregarding the heavy-tailness of the distribution in question can cause devastatingly faulty conclusions as both risk measures are based on the loss distribution. Moreover, VaR does not take into account the tail of the distribution and, hence, with heavy-tailed phenomena, VaR can truly undermine risk. Expected shortfall, on the other hand, focuses strongly on the tail of the distribution and, hence, when heavy tails are not adequately accounted for, can ES values be very wrong. In conclusion, using both risk measures or a broader mix of methods and not assuming a distribution can help avoiding the largest pitfalls when dealing with heavy-tailed data.

Finally it should be noted that although in this thesis the focus was on financial losses and risk measures, the potential pitfalls of not understanding heavy-tailed distributions are also present in other applications in different subjects such as in the context of natural disasters.

References

[1] Artzner, P., Delbaen, F., Eber, J. and Heath, D. (1997). Thinking coherently. Risk, 10(11), pp. 68-71.

[2] Asmussen, S. and Albrecher, H. (2010). *Ruin Probabilities*. World Scientific, Singapore. 2^{nd} edition.

[3] Athreya, K. and Lahiri, S. (2006). *Measure Theory and Probability Theory*. Springer, New York.

[4] Bingham, N., Goldie, C. and Teugels, J. (1987) *Regular Variation*. Cambridge University Press, Cambridge.

[5] Bradley, B. and Taqqu, M. (2003). "Financial risk and heavy tails" in *Handbook of heavy tailed distributions in finance*, pp. 35-103. North-Holland.

[6] Calabria, R. and Pulcini, G. (1987). On the asymptotic behaviour of mean residual life function. Reliability engineering, 19(3), pp.165-170.

[7] Carmona, R. (2014). *Statistical analysis of financial data in R*. Springer, New York.

[8] Chistyakov, V. (1964). A theorem on sums of independent positive random variables and its applications to branching random processes. Theory of Probability & Its Applications, 9(4), pp. 640-648.

[9] Cooke, R. and Nieboer, D. (2011). Heavy-tailed distributions: Data, diagnostics, and new developments. Resources for the Future Discussion Paper, (11-19).

[10] Foss, S., Korshunov, D., and Zachary, S. (2011). *An introduction to heavy-tailed and subexponential distributions*. Springer, New York.

[11] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer, Heidelberg.

[12] Ghosh, S. and Resnick, S. (2011). When does the mean excess plot look linear? Stochastic models, 27(4), pp. 705-722.

[13] Goldie, C. (1978) Subexponential distributions and dominated-variation tails. Journal of Applied Probability, 15(2), pp. 440-442.

[14] Goldie, C. and Klüppelberg, C. (1998). Subexponential distributions. A practical guide to heavy tails: statistical techniques and applications, pp. 435-459.

[15] Gupta, R. and Bradley, D. (2003). Representing the mean residual life in terms of the failure rate. Mathematical and Computer Modelling, 37(12-13), pp. 1271-1280.

[16] Hall, W. and Wellner, J. (1981). Mean residual life. Statistics and related topics, pp. 169-184.

[17] Hofert, M., Hornik, K. and McNeil, A. (2022). *qrmtools: Tools for Quantitative Risk Management*. R package version 0.0-16.

[18] Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley & Sons, New Jersey.

[19] Klugman, S., Panjer, H. and Willmot G. (2008) *Loss Models: from Data to Decisions*. John Wiley & Sons, New Jersey. 3^{rd} edition.

[20] Koller, M. (2011). *Life Insurance Risk Management Essentials*. Springer, Heidelberg.

[21] McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative risk management: concepts, techniques and tools*. Princeton University Press, Princeton.

[22] Nair, J., Wierman, A., and Zwart, B. (2022). *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation*. Cambridge University Press, Cambridge.

[23] Panjer, H. (2006) *Operational Risk: Modeling Analytics*. John Wiley & Sons.

[24] R Core Team (2022). *The R Stats Package*. R Foundation for Statistical Computing. R package version 4.1.3.

[25] Yamai, Y. and Yoshiba, T. (2005). Value-at-risk versus expected shortfall: A practical perspective. Journal of Banking & Finance, 29(4), pp. 997-1015.