Multi-Tail Elliptical Distributions

Sebastian Kring, **Svetlozar T. Rachev**,
Markus Höchstötter, **Frank J. Fabozzi**

Sebastian Kring
Department of Econometrics, Statistics and Mathematical Finance
School of Economics and Business Engineering
University of Karlsruhe
Postfach 6980, 76128, Karlsruhe, Germany
E-mail: sebastian.kring@statistik.uni-karlsruhe.de

Svetlozar T. Rachev
Chair-Professor, Chair of Econometrics, Statistics and Mathematical Finance
School of Economics and Business Engineering
University of Karlsruhe
Postfach 6980, 76128 Karlsruhe, Germany
and
Department of Statistics and Applied Probability
University of California, Santa Barbara
CA 93106-3110, USA
E-mail: rachev@statistik.uni-karlsruhe.de

Markus Höchstötter
Department of Econometrics, Statistics and Mathematical Finance
School of Economics and Business Engineering
University of Karlsruhe
Postfach 6980, 76128, Karlsruhe, Germany

Frank J. Fabozzi
Professor in the Practice of Finance
School of Management
Yale University
New Haven, CT USA

Abstract

In this paper we present a new type of multivariate distributions for asset returns which we call the multi-tail elliptical distributions. Multi-tail elliptical distribution can be thought to be an extension of the elliptical distributions that allow for varying tail parameters. We present a two-step random mechanism leading to this new type of distributions. In particular, this mechanism is derived from typical behavior of financial markets. We determine the densities of multi-tail elliptical distribution and introduce a function which we label the tail function to describe the tail behavior of these distributions. We apply multi-tail elliptical distributions to logarithmic returns of German stocks included in the DAX index. Our empirical results indicate that the multi-tail model significantly outperforms the classical elliptical model and the null hypothesis of homogeneous tail behavior can be rejected.

Keywords and Phrases: Elliptical distributions, Operator stable distributions, α-stable distributions, t distributions, Varying tail parameter, Risk Management, DAX index, Likelihood ratio test.

JEL Classification: C12, C16

Acknowledgments: The authors would like to thank Stoyan Stoyanov and Borjana Racheva-Iotova from FinAnalytica Inc for providing ML-estimators encoded in MATLAB. For further information, see Stoyanov and Racheva-Iotova (2004).

1 Introduction

Since the seminal work of Mandelbrot (1963) there has been a great deal of empirical evidence supporting the existence of heavy-tailed models in finance (see Fama (1965), Jansen and de Vries (1991), Loretan and Phillips (1994), McCulloch (1996), and Rachev and Mittnik (2000)). In the finance literature different models have been proposed to model multivariate heavy-tailed return data. Rachev and Mittnik (2000) have suggested multivariate α-stable distributions to model multivariate asset returns since the α-stable distributions are the natural extension of the normal distribution in terms of the generalized Central Limit Theorem (see Samorodnitsky and Taqqu (1994) for further information) and they allow the modeling of the rich dependence structure of asset returns. Eberlein and Keller (1995) and Eberlein, Keller and Prause (1998) have popularized the generalized hyperbolic distribution as a model for financial returns. Kotz and Nadarajah (2004) have suggested the multivariate t distributions for modeling asset returns. In their papers they provide an extensive overview of the applications of the multivariate t distribution in finance.

Another important phenomenon or stylized fact that has been observed in multivariate asset returns is that extreme returns in one component, i.e. asset, often coincide with extreme returns in several other components (see McNeil, Frey and Embrechts (2005)). This phenomenon can be captured by choosing distributions in multivariate models that allow so-called tail dependence. Heavy-tailed elliptical distributions exhibit tail dependence and in the case of elliptical distributions this property has been extensively studied by Schmidt (2002). Elliptical distributions (e.g., t distributions, generalized hyperbolic distributions, and α-stable sub-Gaussian distributions) are radial symmetric (see Fang, Kotz and Ng (1987)). In empirical work we observe that the lower tail dependence is often much stronger than upper tail dependence (see McNeil, Frey and Embrechts (2005)). Thus, this property cannot be captured by elliptical distributions because of their radial symmetry.

Furthermore and not surprisingly, the research of Jansen and de Vries (1991), Mittnik and Rachev (1993), Loretan and Phillips (1994), and Rachev and Mittnik (2000) has found that the tail index α varies significantly between assets. Despite this well known fact, most existing research on heavy-tailed portfolios and factor models has assumed that the probability tails are the same in every direction. Rachev and Mittnik (2000) consider such a model based on the multivariate stable distribution for portfolio analysis so that α is the same for every direction and asset. The same approach has been applied to portfolio analysis by Fama (1965), Belkacem, Vehel and Walter (2000), and Rachev and Han (2000). In particular, the multivariate α-stable distributions as well as the heavy-tailed elliptical distributions cannot capture different tail thickness.

Mittnik and Rachev (1993), Rachev and Mittnik (2000), and Meerschaert and Scheffler (2003) have suggested operator stable random vectors in order to overcome the limitations of the established heavy-tailed distributions and to capture different tail thickness in assets. This approach leads to a more realistic and flexible representation of financial portfolios. The operator stable distributions can be considered to be a generalization of the multivariate α-stable distributions. Furthermore, they are capable of modeling complex dependence structures but they are incredibly difficult to estimate even in dimension two since with the exception of a few cases their densities are not known in closed form.

We propose a new class of distributions that we label multi-tail elliptical distributions that allow for modeling different tail thickness. These distributions can be considered to be an extension of the elliptical distributions since they are constructed by assuming a particular dependence between the radial random variable R and the random vector U being uniformly distributed on the unit hypersphere. In the classical elliptical case, R and U are independent (see Fang, Kotz and Ng (1987)). The idea of assuming dependence between R and U is also considered in Frahm (2004), leading to the class of generalized elliptical distributions. Furthermore, we give an interpretation of the random mechanism underlying the multi-tail elliptical distributions. We introduce a three-step estimation procedure to estimate the multi-tail elliptical distribution.

This paper is organized as follows. In Section 2 we provide an overview of the operator stable distributions. In Section 3 we introduce the multi-tail elliptical distributions, derive their basic properties, and give examples. In Section 4 we show how to estimate the multi-tail elliptical distributions. In Section 5 we apply the multi-tail elliptical distributions to return data and Section 6 summarizes our conclusions.

2 Operator Stable Distributions

2.1 Definitions and Basic Properties

As already noted, empirical studies suggest that tail behavior varies significantly between different assets or, more generally, different risk factors. If we assume the multivariate α-stable distributions or elliptical distributions as a model for asset returns, then all marginals (i.e., all assets) have the same tail parameter α. This is certainly a limitation in modeling asset returns. In order to overcome this limitation, Mittnik and Rachev (1993) suggested and applied the operator stable distributions in finance for the first time. The operator stable distributions generalize the multivariate α-stable distributions and follow from the generalized Central Limit Theorem by matrix scaling.

In this section we define the operator stable distributions and derive some of their basic properties. The presentation of this section follows Meerschaert and Scheffler (2003) and Stoyanov (2005). For a detailed discussion of the operator stable distributions, see Meerschaert and Scheffler (2001).

Definition 2.1. A random vector X is said to be an operator stable random vector in $\mathbb{R}^d$ if there exists a matrix $E \in \mathbb{R}^{d\times d}$ and a vector a_n such that

$$n^{-E}(X_1 + X_2 + ... + X_n - a_n) \stackrel{d}{=} X, \tag{1}$$

where $X_1, X_2, ..., X_n \in \mathbb{R}^d$ are independent copies of X.

The notion n^{-E} is defined as

$$n^E = I + \sum_{m=1}^{\infty} \frac{(n(-E))^m}{m!}.$$

For further information, see the appendix. The matrix E is called the exponent of the operator stable random vector X.

Rewriting equation (1), we obtain

$$X_1 + X_2 + ... + X_n \stackrel{d}{=} n^E X + a_n. \tag{2}$$

The characteristic function of the right hand-side equals

$$\begin{aligned}
\Phi_{n^E X + a_n}(t) &= E(e^{it' n^E X + a_n}) \\
&= e^{it' a_n} E(e^{it' n^E x}) \\
&= e^{it' a_n} E(e^{i(tn^E)' X}) \\
&= e^{it' a_n} \Phi_X((n^E)' t) \\
&= e^{it' a_n} \Phi_X((n^{E'}) t).
\end{aligned}$$

The characteristic function of the left hand-side of equation (2) is the n-th power of $\Phi_X(t)$ because the sum consists of independent and identically-distributed random vectors in this equation. Thus, equation (2) implies

$$(\Phi_X(t))^n = e^{it' a_n} \Phi_X(n^{E'} t).$$

The following definition is equivalent to Definition 2 and characterizes the operator stable random vectors in terms of their domain of attraction. In order to generalize Gaussian-based financial theories, it is important that the distribution of a random vector X has a domain of attraction. This property implies that the random vector X can be interpreted as an aggregation of many shocks caused by new arriving financial information.

Definition 2.2. A random variable X is said to have an operator stable distribution if it has a domain of attraction, i.e., if there is a sequence of independent and identically-distributed random vectors $Y_1, Y_2, ...$ and a sequence of $d \times d$ matrices $(A_n)_{n \in \mathbb{N}}$ and vectors $(b_n)_{n \in \mathbb{N}}$, such that

$$A_n(Y_1 + Y_2 + ... + Y_n - b_n) \stackrel{d}{\Rightarrow} X. \tag{3}$$

The limits in equation (3) are called operator stable (see Sharpe (1969) and Jurek and Mason (1993)). If $E(||Y_1||^2)$ exists in Definition 2.2 then the classical Central Limit Theorem shows that X is multivariate normal, a special case of operator stable. In this case, we can take $A_n = n^{-1/2}$ and $b_n = nE(X)$. But if $0 < \alpha < 2$ then $E(||X||^2) = \infty$ and the classical Central Limit Theorem does not apply. If, in this case, the tails of Y_1 fall off at the same rate α in every direction then equation (3) holds with $A_n = n^{-1/\alpha} I$ and the limit X is multivariate stable distributed due to the generalized Central Limit Theorem (see Rachev and Mittnik (2000) for a detailed discussion).

In general, the tail behavior of the operator stable random vector is determined by the eigenvalues of its exponent E. The Spectral Decomposition Theorem allows us to write the exponent of an operator stable distribution in the form $E = PBP^{-1}$, where P is a change of coordinate matrix and B is a block-diagonal matrix satisfying

$$B = \begin{pmatrix} B_1 & 0 & \dots & 0 \\ 0 & B_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & B_p \end{pmatrix},$$

where $B \in \mathbb{R}^{d_i \times d_i}, i = 1, ..., p$ and $\sum_{i=1}^{p} d_i = d$. One can show (see Meerschaert and Scheffler (2003)) that every eigenvalue of B_i has a real part equal to a_i with $a_1 > a_2 > ... > a_p \geq 1/2$[1]. Furthermore, the Spectral Decomposition Theorem says that we can decompose $\mathbb{R}^d$ in p E-invariant subspaces, i.e.,

$$\mathbb{R}^d = V_1 \oplus V_2 \oplus ... \oplus V_p \text{ and } EV_i \subset V_i, \tag{4}$$

where $V_i = \text{span}\{Pe_j : d_1 + ... + d_{i-1} < j \leq d_1 + d_2 + ... + d_i\}, i = 1, ..., p$.

Given a non-zero vector $v \in \mathbb{R}^d$, we can write its unique spectral decomposition $v = v_1 + ... + v_p$ according to equation (4). Let's define

$$\alpha(v) = \min_{i=1,...,p} \{\frac{1}{a_i} : v_i \in V_i \setminus \{0\}, i = 1, ..., p\}$$

that is, the number $\alpha(v)$ is the smallest $\frac{1}{a_i}$ of those eigenvalues that correspond to non-zero components in the spectral decomposition of the vector v. In Meerschaert and Scheffler (2001) it is shown that the following property holds

Proposition 2.1. Let X be an operator stable vector in $\mathbb{R}^d$ with exponent E and let $v \in \mathbb{R}^d$. Then for any small $\delta > 0$ we have

$$\lambda^{-\alpha(v)-\delta} < P(|v'X| > \lambda) < \lambda^{-\alpha(b)+\delta}$$

for all $\lambda > 0$ sufficiently large.

Proposition 2.1 says that the tail behavior of the linear combination $v'X$ is dominated by the component with the heaviest tail. In particular, this means that $E(|v'X|^s)$ exists for all $0 < s < \alpha(v)$ and diverges for $s > \alpha(v)$.

If we decompose an operator stable random vector X with exponent E according to equation (4) we obtain

$$X = \sum_{i=1}^{p} p_{V_i}(X)$$

where $p_{V_i}(X)$ is the projection of X into the subspace V_i , $i = 1, ..., p$. One can show (see Meerschaert and Scheffler (2001)) that $p_{V_i}(X)$ is operator stable with some exponent E_i whose eigenvalues have the same real part a_i. One says that $p_{V_i}(X)$ is spectrally simple with tail index $\alpha_i = 1/a_i$ (see Meerschaert and Scheffler (2001)).

In the next example we present an interesting subclass of operator stable random vectors that is important for financial applications.

Example 2.1. *Take $B = \text{diag}(a_1, ..., a_d)$ where $a_1 > a_2 > ... > a_d$ and P orthogonal. If $X = (X_1, ..., X_d)'$ is operator stable with exponent $E = PBP'$ then $B_i = a_i$ and $V_1, ..., V_d$ are the coordinate axes in the new coordinate system defined by the vectors $p_i = Pe_i$, $i = 1, ..., d$. One can show that the ith component $p_i'X$ in the new coordinate system is stable with index $\alpha_i = \frac{1}{a_i}$ (see Meerschaert and Scheffler (2003)). Whereas the original ith component X_i is not stable since it is a linear combination of stable laws of different indices. Hence, the change of coordinate matrix P*

[1] This condition is necessary for the existence of a generalized domain of attraction (see Definition 2.2).

rotates the coordinate axes to make the marginals stable. More mathematically, since $n^{PBP^{-1}} = Pn^BP^{-1}$ *it follows from Definition 2.1 that*

$$\begin{aligned} Pn^{-B}P^{-1}(X_1 + ... + X_n - b_n) &\stackrel{d}{=} X \\ n^{-B}(P^{-1}X_1 + ... + P^{-1}X_n - P^{-1}b_n) &\stackrel{d}{=} P^{-1}X, \end{aligned}$$

so that $Y = P^{-1}X$ *is operator stable with exponent* $B = \mathrm{diag}(a_1, ..., a_n)$.

2.2 Applications

Mittnik and Rachev (1993) seem to have been the first to apply operator stable models to problems in finance. Since empirical work suggests that the stable indices α_i, $i = 1, ..., d$ vary depending on the assets, they assume the univariate version of equation (3) holds

$$\frac{Y_{1i} + Y_{2i} + ... + Y_{ni} - b_{ni}}{n^{1/\alpha_i}} \Rightarrow X_i,$$

for each $i = 1, ..., d$, so that Y_i is stable with index α_i. Assuming the joint convergence one obtains

$$A_n(Y_1 + Y_2 + ... + Y_n - b_n) \Rightarrow X$$

with a diagonal norming matrix

$$A_n = \begin{pmatrix} n^{-1/\alpha_1} & 0 & \dots & 0 \\ 0 & n^{-1/\alpha_1} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & n^{-1/\alpha_d} \end{pmatrix}.$$

The matrix scaling is natural since it allows a more realistic portfolio model. In this model one can show that the ith marginal X_i of the operator stable limit vector X is α_i-stable distributed. This is consistent with the findings of empirical studies.

Meerschaert and Scheffler (2003) apply the model presented in example 2.1 to exchange rate log-returns for the German Deutsche mark $X_1(t)$ and $X_2(t)$, $t = 1, ..., n$, both taken against the U.S.-dollar. They argue that since the tail parameter α usually depends on the coordinate, the wrong coordinate system can mask variations in α because of the domination of the heaviest tail (see Proposition 2.1). They suggest that the coordinate system given by the principal components of the sample covariance matrix is a judicious choice, since the principal components can be interpreted as the directions of the largest dispersion. They assume that the random vector $Y(t) = P'X(t)$ is operator stable with exponent $\mathrm{diag}(a_1, a_2)$, implying that $Y(t)$ has stable marginals. Hence, $X(t)$ is operator stable with exponent $P\,\mathrm{diag}(a_1, a_2)P'$. In particular, they estimate $\alpha_1 = 1/a_1 \approx 1.656$ and $\alpha_2 = 1/a_2 \approx 1.996$. They conclude that $Z_1(t)$ is α_1-stable distributed and $Z_2(t)$ is normally distributed since a_2 is very close to $1/2$, indicating normality.

3 Multi-Tail Elliptical Distributions

We have seen that distributions allowing modeling of different tail thickness of risk factors or asset returns are desirable since they lead to more realistic portfolio models. Operator stable random vectors possess the property to capture variations in the tail parameter α, but they have the drawback that they are very difficult to estimate.

In this section we combine the elliptical distributions with the concept of varying tail thickness. This leads to a new class of distributions that we call the multi-tail elliptical distributions. But before offering a motivation and the definition of this new class of distributions we have to define additional notions.

Let $\Sigma \in \mathbb{R}^{d\times d}$ be a positive definite matrix. Then we denote the *Cholesky factor* of Σ with $\Sigma^{1/2}$ and the inverse by $\Sigma^{-1/2}$ (see Hamilton (1994) for the Cholesky factorization). With the random vector $S \in \mathbb{R}^d$ we denote the uniform distribution on the unit hypersphere $\mathcal{S}^{d-1} = \{x \in \mathbb{R}^d : ||x|| = 1\}$. Let $x \in \mathbb{R}^d$, then we call $s(x) = x/||x|| \in \mathcal{S}^{d-1}$ the spectral projection of x. Every elliptical random vector X can be written in the form

$$X = \mu + RAS,$$

where R is a non-negative random variable independent of S, $A \in \mathbb{R}^{d\times d}$ a matrix, and μ a location parameter. We denote a random variable R with tail parameter $\alpha > 0$ by R_α and the density of a random vector X by f_X.

Let the elliptical random vector $X = (X_1, X_2, ..., X_d)'$ describe the log-returns of a portfolio with d assets. AS determines the direction of the portfolio development, while R independent of direction AS impacts the volatilities of all stocks. It is important to note that AS and $s(AS)$ show in the same direction and if AS independent of R so is $s(AS)$. However, it seems questionable to model R independent of $s(AS)$. Rather R should depend on $s(AS)$ in the sense that it determines the tail behavior of R.

Let us assume that the daily log-returns of a portfolio modeled by the random variable X follow an elliptical model $X = RAS$.[2] Given the the partial information that we observe *negative* signs in *all* components of X (hence, all components of AS and $s(AS)$ are negative since R is positive) we may conclude that such an observation is caused by an important macroeconomic event or financial crisis (i.e., where markets are in a stress situation). The markets tend to be extreme and it is very likely to observe large losses in many components of X. In such a market stress scenario one says that the correlations approach unity (see McNeil, Frey and Embrechts (2005)). However, instead of assuming a higher correlation between assets, we suppose lower tail parameters of R in these directions (e.g, $AS = (-1, ..., -1)$). These lower tail parameters incorporate a higher tail dependence in these directions and cause simultaneous high losses.

In this section we define multi-tail elliptical distributions being predestinated to capture the market behavior described above. We derive their basic properties, give examples of multi-tail elliptical distribution families, and discuss how the tail parameter α varies subject to the direction AS.

[2] For daily returns one can assume $\mu = 0$. See McNeil, Frey and Embrechts (2005) for further information.

3.1 Definition and Basic Properties

Definition 3.1 (**Multi-tail Elliptical Distributions**). Let $S \in \mathbb{R}^d$ be the uniform distribution on the unit hypersphere $\mathcal{S}^{d-1}$, I an interval of tail parameters and $(R_\alpha)_{\alpha \in I}$ a family of positive random variables with tail parameter α. The random vector $X = (X_1, X_2, ..., X_d)' \in \mathbb{R}^d$ has a multi-tail elliptical distribution, if X satisfies

$$X \stackrel{d}{=} \mu + R_{\alpha(s(AS))} AS, \tag{5}$$

where $A \in \mathbb{R}^{d \times d}$ is a regular matrix, μ a location parameter and $\alpha : \mathcal{S}^{d-1} \to I$ a function.

In particular, we call the function

$$\alpha : \mathcal{S}^{d-1} \to I$$

the tail function of a multi-tail elliptical distribution. In Section 3.2 we discuss several examples of the tail function and its impact on the distribution.

From Definition 3.1 it is obvious that the random vector AS determines the scaling properties of X (i.e., the proportion of the components of X among each other) while $R_{\alpha(AS)}$ characterizes the tail behavior. More mathematically expressed, we have for the scaling property

$$s(X - \mu) = \frac{R_{\alpha(s(AS))} AS}{||R_{\alpha(s(AS))} AS||} = \frac{R_{\alpha(s(AS))} AS}{R_{\alpha(s(AS))} ||AS||} = \frac{AS}{||AS||} = s(AS)$$

and for the tail behavior

$$||X - \mu|| = ||R_{\alpha(s(AS))} AS|| = R_{\alpha(s(AS))} ||AS|| \leq c R_{\alpha(s(AS))},$$

where $c > 0$ is an appropriate constant.

The trick in Definition 3.1 is to make the tail parameter α depend on AS and not directly on S. In empirical work the random vector S is not observable and A is not unique because of the rotational symmetry of S, implying

$$S \stackrel{d}{=} PS \text{ and } AS \stackrel{d}{=} A(PS) \tag{6}$$

where $P \in \mathbb{R}^{d \times d}$ is orthogonal.

Lemma 3.1. *Let $S \in \mathbb{R}^d$ be a random vector with uniform distribution on the hypersphere $\mathcal{S}^{d-1}$. Then its density satisfies*

$$f_S(s) = \frac{\Gamma(d/2)}{2\pi^{d/2}} 1_{\mathcal{S}^{d-1}}(s),$$

where $\Gamma(.)$ is the gamma function[3].

Proof. Note, that the surface of $\mathcal{S}^{d-1}$ equals $2\pi^{d/2}/\Gamma(d/2)$. Since S is uniformly distributed we obtain

$$f_S(s) = \frac{1}{2\pi^{d/2}/\Gamma(d/2)} 1_{\mathcal{S}^{d-1}}(s) = \frac{\Gamma(d/2)}{2\pi^{d/2}} 1_{\mathcal{S}^{d-1}}(s).$$

□

[3] See Definition 7.1 in the appendix for further information about Gamma function.

An advantage of multi-tail elliptical distributions is that we have an analytic expression for their densities.

Theorem 3.1. *Let $X = R_{\alpha(s(AS))}AS$ be a multi-tail elliptical distribution, where $A \in \mathbb{R}^{d\times d}$ is a regular matrix, I an interval of tail parameters, $(R_\alpha)_{\alpha\in I}$ a family of positive random variables, $(f_{R_\alpha})_{\alpha\in I}$ its family of densities, S the uniform distribution on the hypersphere $\mathcal{S}^{d-1}$ and $\alpha : \mathcal{S}^{d-1} \to I$ a tail function. Then the density of X is given by*

$$f_X(x) = |det(\Sigma)|^{-1/2} g_{\alpha(s(x-\mu))}((x-\mu)'\Sigma^{-1}(x-\mu)) \tag{7}$$

and

$$g_{\alpha(u)}(r^2) = \frac{\Gamma(d/2)}{2\pi^{d/2}} r^{-d+1} f_{R_{\alpha(u)}}(r),$$

where $\Sigma = AA'$ is the dispersion matrix and $g_{\alpha(.)}$ is called the density generator of the multi-tail elliptical distribution.

We note that the matrix A is not part of the density term in equation (7). This is important for empirical work, since the matrix A is difficult to observe and not unique due to equation (6).

Proof. Due to Definition 3.1 we have

$$R_{\alpha(s(AS))}|S = s \quad = \quad R_{\alpha(s(As))}$$

and thus, the density of $R_{\alpha(s(AS))}$ given $S = s$ satisfies

$$f_{R_{\alpha(s(AS))}|S}(r|s) = f_{R_{\alpha(s(As))}}(r).$$

Hence the joint density exists and corresponds to

$$\begin{aligned} f_{(R,S)}(r,s) &= f_S(s) f_{R|S}(r|s) \\ &= f_S(s) f_{R_{\alpha(s(As))}}(r) \\ &= \frac{\Gamma(d/2)}{2\pi^{d/2}} f_{R_{\alpha(s(As))}}(r) \end{aligned} \tag{8}$$

We define the transformation

$$g : \begin{cases} \mathbb{R} \times \mathcal{S}^{d-1} & \to \quad \mathbb{R}^d \\ (r,s) & \mapsto \quad \mu + rAs. \end{cases}$$

One can show that the inverse function g^{-1} is given by

$$g^{-1} : \begin{cases} \mathbb{R}^d & \to \quad \mathbb{R}_+ \times \mathcal{S}^{d-1} \\ x & \mapsto \quad (g_1^{-1}(x), g_2^{-1}(x)) = (\sqrt{(x-\mu)'\Sigma^{-1}(x-\mu)}, \frac{A^{-1}(x-\mu)}{||A^{-1}(x-\mu)||}) \end{cases},$$

where $\Sigma = AA'$.
Since we have for all $A \in \mathcal{B}^d$

$$\begin{aligned}
\int_A f_X(x)dx &= P[X \in A] \\
&= P[g(R,S) \in A] \\
&= P[(R,S) \in g^{-1}(A)] \\
&= \int_{g^{-1}(A)} f_{(R,S)}(r,s)drds \\
&\overset{(1)}{=} \int_A f_{(R,S)}(g^{-1}(x))|\det((Dg^{-1})(x))|dx, \qquad (9)
\end{aligned}$$

it follows

$$f_X(x) = f_{(R,S)}(g^{-1}(x))|\det(Dg^{-1})(x)|.$$

Note that (1) holds in equation (9) because of Theorem 7.1 in the appendix. Furthermore, Corollary 7.1 in the appendix says that we have

$$\det(D(g^{-1})(x)) = \frac{((x-\mu)'\Sigma^{-1}(x-\mu))^{(-d+1)/2}}{\det(\Sigma)^{1/2}} \qquad (10)$$

and furthermore, we obtain by using equation (8)

$$f_{(R,S)}(g^{-1}(x)) = \frac{\Gamma(d/2)}{2\pi^{d/2}} f_{R_{\alpha(s(Ag_2^{-1}(x)))}}(g_1^{-1}(x)). \qquad (11)$$

It is important to note that we have

$$\begin{aligned}
s(Ag_2^{-1}(x)) &= s\left(\frac{x-\mu}{||A^{-1}(x-\mu)||}\right) = \frac{x-\mu}{||A^{-1}(x-\mu)||}\frac{||A^{-1}(x-\mu)||}{||x-\mu||} \\
&= \frac{x-\mu}{||x-\mu||} = s(x-\mu). \qquad (12)
\end{aligned}$$

Hence, we derive with equations (11) and (12)

$$f_{(R,S)}(g^{-1}(x)) = \frac{\Gamma(d/2)}{2\pi^{d/2}} f_{R_{\alpha(s(x-\mu))}}(\sqrt{(x-\mu)'\Sigma^{-1}(x-\mu)}). \qquad (13)$$

Finally, adding equations (10) and (13) leads to

$$\begin{aligned}
f_X(x) &= \frac{\Gamma(d/2)}{2\pi^{d/2}} \det(\Sigma)^{-1/2} \left(\sqrt{(x-\mu)'\Sigma^{-1}(x-\mu)}\right)^{-d+1} \\
&\quad f_{R_{\alpha(s(x-\mu))}}(\sqrt{(x-\mu)'\Sigma^{-1}(x-\mu)})
\end{aligned}$$

□

Corollary 3.1. *Let $X = \mu + R_{\alpha(s(AS))}AS \in \mathbb{R}^d$ be a multi-tail elliptical random vector. Then we have*

$$\sqrt{(X-\mu)'\Sigma^{-1}(X-\mu)}|(s(X-\mu) = s) \overset{d}{=} R_{\alpha(s)}$$

or equivalently,

$$(X-\mu)'\Sigma^{-1}(X-\mu)|(s(X-\mu) = s) \overset{d}{=} R^2_{\alpha(s)},$$

where $\Sigma = AA'$.

Proof. We have

$$\begin{aligned}\sqrt{(X-\mu)'\Sigma^{-1}(X-\mu)} &= ||\Sigma^{-1/2}(X-\mu)|| \\ &= ||\Sigma^{-1/2}(\mu+R_{\alpha(s(AS)}AS-\mu)|| \\ &= ||\Sigma^{-1/2}AS||R_{\alpha(s(AS))} \\ &\stackrel{d}{=} ||S||R_{\alpha(s(AS))} \\ &= R_{\alpha(s(AS))} \qquad (14)\end{aligned}$$

Note that in the third line of the last equation we have equality in distribution because $(\Sigma^{-1/2})'A$ is orthogonal[4]. $\Sigma^{-1/2}A$ is orthogonal since we have

$$\begin{aligned}(\Sigma^{-1/2}A)(\Sigma^{-1/2}A)' &= \Sigma^{-1/2}AA'(\Sigma^{-1/2})' \\ &= \Sigma^{-1/2}\Sigma(\Sigma^{-1/2})' \\ &= \Sigma^{-1/2}\Sigma^{1/2}(\Sigma^{1/2})'(\Sigma^{-1/2})'. \\ &= \mathrm{Id}\,.\end{aligned}$$

From equation (14) we conclude

$$\sqrt{(X-\mu)'\Sigma^{-1}(X-\mu)}|(s(X-\mu)=s) = R_{\alpha(s)}.$$

Analogously, we can derive

$$(X-\mu)'\Sigma^{-1}(X-\mu)|(s(X-\mu)=s) \stackrel{d}{=} R^2_{\alpha(s)}.$$

□

In the following we discuss the connection between multi-tail elliptical random vectors and elliptical random vectors.

Definition 3.2. Let $(R_\alpha)_{\alpha\in I}$ be a family of positive random variables with tail parameter α and $Y_\alpha = R_\alpha AS$, $\alpha \in I$, a family of elliptical random vectors. We call the random vector X a *corresponding* multi-tail elliptical random vector if it is given by $X = R_{\alpha(s(AS))}$AS, where $\alpha : \mathcal{S}^{d-1} \to I$ is a tail function.

Theorem 3.2. *Let $X = R_\alpha AS$ be an elliptical random vector, where $\mu \in \mathbb{R}^d$ is a location vector, $A \in \mathbb{R}^{d\times d}$ a regular matrix, R_α a positive random variable with tail parameter α and the random vector $S \in \mathbb{R}^d$ uniformly distributed on $\mathcal{S}^{d-1}$. Then X possesses a density f_X if and only if R_α has a density f_{R_α}. The relationship between f_X and f_{R_α} is as follows*

$$f_X(x) = |det(\Sigma)|^{-1/2}g((x-\mu)'\Sigma^{-1}(x-\mu))$$

and

$$g(r^2) \;=\; \frac{\Gamma(d/2)}{2\pi^{d/2}}r^{d-1}f_{R_\alpha}(r),$$

where $\Sigma = AA'$ is the dispersion matrix.

[4]See also equation (6).

Proof. See Fang, Kotz and Ng (1987). □

If follows from Theorems 3.1 and 3.2 that the density of a multi-tail elliptical random vector can be obtained by substituting the constant tail parameter α through the tail function $\alpha(.)$ in the density of the corresponding family of elliptical random vectors. It is obvious that multi-tail elliptical distributions are a generalization of elliptical distributions. Vice versa, if we choose the tail function to be constant, R and S are independent and we obtain the classical case. Summing up, we obtain

Corollary 3.2. *Densities of a family of elliptical random vectors differ from a density of a corresponding multi-tail elliptical random vector by substituting the constant* α *through a tail function* $\alpha(.)$.

Normal variance mixtures are a particular subclass of elliptical distributions (see Fang, Kotz and Ng (1987)). They are given by

$$Y = \mu + W_\alpha^{1/2} AZ, \tag{15}$$

where $\mu \in \mathbb{R}^d$ is a location parameter, W_α a positive random variable with tail parameter $\alpha/2$,[5] $A \in \mathbb{R}^{d\times d}$ a regular matrix, and $Z \sim N(0, Id)$. We can rewrite equation (15) in the form

$$Y = \sqrt{W_\alpha ||Z||^2} A \left(\frac{Z}{||Z||} \right). \tag{16}$$

It is important to note that we have[6]

(i) $Z/||Z||$ is uniformly distributed on the d-dimensional hypersphere $\mathcal{S}^{d-1}$.

(ii) $||Z||^2$ is chi-squared distributed with d degrees of freedom.

(iii) $||Z||^2$ and $Z/||Z||$ are independent.

Thus, identifying $R_\alpha = \sqrt{W_\alpha ||Z||^2}$ and $Z/||Z|| = S$, we see that normal variance mixtures are elliptical random vectors according to equation (16).

Furthermore, multi-tail normal variance mixtures are well defined and we obtain

$$\begin{aligned} X &= \mu + R_{\alpha(s(AS))} AS = \mu + W^{1/2}_{\alpha(s(AZ/||Z||))} ||Z|| A \frac{Z}{||Z||} \\ &= \mu + W^{1/2}_{\alpha(s(AZ))} AZ. \end{aligned}$$

We introduce this ”mixture” notation because most well-known elliptical distribution families such as α-stable sub-Gaussian distributions, t distributions, and generalized hyperbolic distributions are defined in the mixture notation. In the next section we introduce their corresponding multi-tail versions.

[5] One can show that if W_α has tail parameter $\alpha/2$ then $W_\alpha^{1/2}$ has tail parameter α (see Samorodnitsky and Taqqu (1994)).

[6] A thorough discussion of items (i), (ii) and (iii) can be found in Fang, Kotz and Ng (1987).

3.2 Principal Component Tail Functions

In this section we take a closer look at the tail functions $\alpha : \mathcal{S}^{d-1} \to I$ of multi-tail elliptical distributions. Just as the tail parameter α varies between assets[7] it varies also between linear combinations of asset returns (see Meerschaert and Scheffler (2003) and Rachev and Mittnik (2000)). It is reasonable to explore the principal components of asset returns[8] since they determine the directions of the largest dispersions.[9] In particular, Meerschaert and Scheffler (2003) argue that the coordinate system given by the principal components can reveal the tail behavior of random vectors with varying tail thickness and an inappropriate coordinate system even masks the right tail behavior.

In the following we introduce tail functions based on the principal components of the dispersion matrix $\Sigma = AA'$ of a multi-tail elliptical random vector. We denote eigenvector eigenvalue pairs of the dispersion matrix $\Sigma = AA'$ of a multi-tail elliptical random vector [10] with (v_1, λ_1), (v_2, λ_2),...,(v_d, λ_d), where $\lambda_1 \geq \lambda_2 ... > \lambda_d > 0$ and $||v_1|| = ... = ||v_d|| = 1$.

Definition 3.3. Let X be a multi-tail elliptical random vector with dispersion matrix Σ. We call the tail function $\alpha : \mathcal{S}^{d-1} \to I$ of X a principal component tail function (pc-tail function) if it satisfies

$$\alpha(s) = \sum_{i=1}^{d} w_i^+(< s, v_i >)\alpha_i^+ + w_i^-(< s, v_i >)\alpha_i^-,$$

where $s \in \mathcal{S}^{d-1}$, $w_i^+, w_i^- : [-1, 1] \to [0, 1]$, $i = 1, ..., d$, are weighting functions with

$$\sum_{i=1}^{d} w_i^+(< s, v_i >) + w_i^-(< s, v_i >) = 1$$

and

$$w_i^+(0) = w_i^-(0) = 0.$$

Note that for all pc-tail functions $\alpha(.)$ we have $\alpha(s) \in I$, $s \in \mathcal{S}^{d-1}$ since I is an interval and $\alpha(s)$ can be interpreted as a convex combination of $\alpha_i^-, \alpha_i^+ \in I$, $i = 1, ..., d$.

Because we assume that we cannot explain the different dispersion intensities only by different scaling factors, we assign to every principal component a tail index $\alpha_i^{(+)}$ or $\alpha_i^{(-)}$, depending on the sign of $< s, v_i >$, $i = 1, ..., d$. The principal components $(v_i)_{i=1,...,d}$ are an orthogonal decomposition of the main directions of dispersion, hence we obtain

$$\alpha(v_i) = w_i^+(1)\alpha_i^+ + w_i^-(1)\alpha_i^-$$

[7] For references see Introduction.

[8] Note that principal components are just linear combinations of asset returns.

[9] See Kring et al. (2007), Johnson and Wichern (1982) or McNeil, Frey and Embrechts (2005)

[10] We will see in Section 4 that we can estimate Σ without knowing the tail function. This is of course important for practical work.

and

$$\alpha(-v_i) = w_i^+(-1)\alpha_i^+ + w_i^-(-1)\alpha_i^-$$

for all $i = 1, ..., d$.

In the context of financial risk factor modeling, this means that if our portfolio moves in the direction v_i or $-v_i$ the size of this movement is determined by a radial random variable $R_{\alpha(\pm v_i)}$ with tail parameters $\alpha(\pm v_i)$. For an arbitrary direction $s, -s \in \mathcal{S}^{d-1}$ we weight every tail parameter α_i^+ and α_i^- according to $w_i^+(< s, v_i >)$ and $w_i^-(< s, v_i >)$, $i = 1, ..., d$.

Note that if the pc-tail function of a multi-tail elliptical random vector X satisfies $\alpha(s) = \alpha(-s)$, then X is radial symmetric with respect to the location parameter μ. That is

$$X - \mu \stackrel{d}{=} \mu - X.$$

In the following we give four examples of the pc-tail functions. The first pc-tail function $\alpha_1(.)$ is given by

$$\alpha_1 : \begin{cases} \mathcal{S}^{d-1} & \rightarrow \quad I \\ s & \mapsto \quad \sum_{i=1}^{d} < u, v_i >^2 \alpha_i \end{cases}$$

with weighting functions $w_i^+(< s, v_i >) = w_i^-(< s, v_i >) = \frac{1}{2} < s, v_i >^2$ for all $i = 1, ..., d$. This tail function assigns to every principal component a tail parameter α_i, $i = 1, ..., d$. In particular, we have $\alpha(v_i) = \alpha(-v_i) = \alpha_i$ for $i = 1, ..., d$. In any other direction $s \in \mathcal{S}^{d-1}$ $\alpha(s)$ is a convex combination of the tail parameters α_i. In fact, $\alpha_1(.)$ is a pc-tail function since we have for all $s \in \mathcal{S}^{d-1}$

$$\begin{aligned} \sum_{i=1}^{d} \frac{1}{2} < s, v_i >^2 + \frac{1}{2} < s, v_i >^2 &= \sum_{i=1}^{d} < s, v_i >^2 \\ &= < s, \sum_{i=1}^{d} < s, v_i > v_i > \\ &= < s, s > \\ &= 1. \end{aligned}$$

A refinement of the pc-tail function $\alpha_1(.)$ is the pc-tail function

$$\alpha_2 : \begin{cases} \mathcal{S}^{d-1} & \rightarrow \quad I \\ s & \mapsto \quad \sum_{i=1}^{d} < u, v_i >^2 I_{(0,\infty)}(< u, v_i >)\alpha_i^+ \\ & \qquad + \sum_{i=1}^{d} < u, v_i >^2 I_{(-\infty,0)}(< u, v_i >)\alpha_i^- . \end{cases}$$

$\alpha_2(.)$ allows for different tail parameters for each direction. This is also plausible since movements of the portfolio in the direction s might have a tail parameter different to that in the opposite direction, $-s$. The weighting functions of $\alpha_2(.)$ are defined by

(i) $w_i^+(< s, v_i >) = < s, v_i >^2 I_{(0,\infty)}(< s, v_i >)$

(ii) $w_i^-(< s, v_i >) = < s, v_i >^2 I_{(-\infty,0)}(< s, v_i >)$.

It is obvious that

$$\sum_{i=1}^{d} < s, v_i >^2 I_{(0,\infty)}(< s, v_i >) + < s, v_i >^2 I_{(-\infty,0)}(< s, v_i >) = 1. \qquad (17)$$

Thus, $w_i^+(< s, v_i >)$ and $w_i^-(< u, v_i >)$ are weighting functions.

In high dimensional financial risk factor modeling it is certainly too complicated to assign to each principal component v_i tail parameters α_i^+ and α_i^-. It is sufficient to match different tail indices α_i^+ and α_i^- to the first principal components $(v_i)_{i=1,...,d_1<<d}$ and an overall tail index α_0 to the additional principal components $(v_i)_{i=d_1+1,...,d}$. This simplification leads to functions $\alpha_3(.)$ and $\alpha_4(.)$ presented in the following.

Definition 3.4. Let $A \subset \mathbb{R}^d$ be a linear subspace. A ϵ-cone $C_\epsilon(A)$ is the set defined by

$$C_\epsilon(A) \quad = \quad \{x \in \mathbb{R}^d | \angle(x, A) < \epsilon\}.$$

where $\angle(x, A)$ is given by

$$\angle(x, A) := \angle(x, p_A(x)) := \frac{x}{||x||} \cdot \frac{p_A(x)}{||p_A(x)||}.$$

$P_A(x)$ is the orthogonal projection of x in the linear subspace A.

The pc-tail function $\alpha_3(.)$ is given by

$$\alpha_3 : \begin{cases} \mathcal{S}^{d-1} & \rightarrow \quad I \\ s & \mapsto \quad \sum_{i=1}^{d} < s, v_i >^2 \alpha_i + \sum_{i=d_1+1}^{d} < s, v_i >^2 \alpha_0. \end{cases}$$

Only for the principal components with the larges eigenvalues do we use different tail parameters α_i, since in the ϵ-cone $C_\epsilon(\text{span}\{F_1, ..., F_{d_1}\})$ we observe most of the volatility.

A natural refinement of the pc-tail function $\alpha_3(.)$ is defined by

$$\alpha_4 : \begin{cases} \mathcal{S}^{d-1} & \rightarrow \quad I \\ s & \mapsto \quad \sum_{i=1}^{d} < s, v_i >^2 I_{(0,\infty)}(< s, v_i >)\alpha_i^+ \\ & \qquad + \sum_{i=1}^{d} < s, v_i >^2 I_{(-\infty,0)}(< s, v_i >)\alpha_i^- \\ & \qquad + \sum_{i=d_1+1}^{d} < s, v_i >^2 \alpha_0. \end{cases}$$

It is easy to see that $\alpha_3(.)$ and $\alpha_4(.)$ are pc-tail functions.

3.3 Two Examples of Multi-Tail Elliptical Distributions

We introduce two examples of multi-tail elliptical distributions. The first one is derived from α-stable sub-Gaussian distributions and the second one from multivariate t distributions.

Definition 3.5 (Multi-tail α-stable sub-Gaussian random vectors). Let $Z \in \mathbb{R}^d$ be a standard normal random vector and $(W_\alpha)_{\alpha\in(0,2)}$ a family of positive α-stable random variables with tail parameter $\alpha/2$ satisfying

$$W_\alpha \sim S_{\alpha/2}(\cos\left(\frac{\pi\alpha}{4}\right)^{2/\alpha}, 1, 0).$$

Then the random vector

$$X = \mu + W^{1/2}_{\alpha(s(AZ))} AZ$$

is a multi-tail α-stable sub-Gaussian random vector where $\alpha : \mathcal{S}^{d-1} \to (0,2) = I$ is a tail function.

In particular, multi-tail α-stable sub-Gaussian random vectors allow us to combine the concepts of α-stable distributions and varying tail thickness such as operator stable random vectors. Figure 1 depicts scatterplots of 1000 samples of a multi-tail α-stable sub-Gaussian distribution with dispersion matrix $10^{-4}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. $F_1 = s((1,1)')$ is the first principal component of Σ and $F_2 = s((1,-1)')$ the second one. In scatterplot (a) in Figure 1 we have the pc-tail function

$$\alpha(s) =< s, F_1 >^2 \cdot 1.5 + < s, F_2 > \cdot 1.9$$

and in scatterplot (b) the pc-tail function equals

$$\alpha(s) =< s, F_1 >^2 \cdot 1.6 + < s, F_2 > \cdot 1.9.$$

We observe in Figure 1 that we have much more outliers in the cones around F_1 and $-F_1$ than in the ones around F_2 and $-F_2$ because of the pc-tail function.

Figure 2 shows scatterplot of 1000 samples of a multi-tail sub-Gaussian distribution with the same dispersion matrix as in Figure 1 but with different tail function. In scatterplot (a) in Figure 2 we have the pc-tail function

$$\begin{aligned}\alpha(s) &= < s, F_1 >^2 I_{(0,\infty)}(< s, F_1 >) \cdot 1.9 + < s, F_1 >^2 I_{(-\infty,0)}(< s, F_1 >) \cdot 1.5 \\ &\quad + < s, F_2 >^2 \cdot 1.9\end{aligned}$$

and in scatterplot (b)

$$\begin{aligned}\alpha(s) &= < s, F_1 >^2 I_{(0,\infty)}(< s, F_1 >) \cdot 1.9 + < s, F_1 >^2 I_{(-\infty,0)}(< s, F_1 >) \cdot 1.6 \\ &\quad + < s, F_2 >^2 \cdot 1.9\end{aligned}$$

In Figure 2 there are many more outliers in the cone around $-F_1$ than in the one around F_1. This is because of the smaller tail parameters in the directions in the cone around $-F_1$.

A d-dimensional multivariate t distributed random vector Y with ν degrees of freedom is given by

$$Y \stackrel{d}{=} \mu + W_\nu^{1/2} AZ,$$

where $W_\nu \sim \text{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$[11] and $\nu \in (0, \infty)$. Furthermore, the density of a t distributed random vector Y is given by

$$f_Y(x) \;=\; \frac{\Gamma(\frac{1}{2}(\nu+d))}{\Gamma(\frac{1}{2}\nu)(\pi\nu)^{d/2}\det(c\Sigma_0)^{1/2}}\left(1+\frac{(x-\mu)'(c\Sigma_0)^{-1}(x-\mu)}{\nu}\right)^{-(\nu+d)/2},$$

where $\nu > 0$ is the tail parameter, $c > 0$ a scaling parameter, and Σ_0 the normalized dispersion matrix.[12] The corresponding multi-tail t distribution is defined as follows.

Definition 3.6 (Multi-tail t-distributed random vectors). Let $Z \in \mathbb{R}^d$ be a standard normal random vector and $(W_\nu)_{\nu\in(0,\infty)}$ a family of inverse gamma distributed random vectors with tail parameter $\nu/2$, i.e. $W_\nu \sim \text{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$. The random vector

$$X = W_{\nu(s(AZ))}^{1/2} AZ$$

is a multi-tail t distributed random vector, where $\nu : \mathcal{S}^{d-1} \to (0, \infty) = I$ is a tail function.

According to Corollary 3.2 and the knowledge about the density f_Y of a t distributed random vector we can derive

Corollary 3.3. *Let $X \in \mathbb{R}^d$ be a multi-tail t distributed random vector with tail parameter function $\nu : \mathcal{S}^{d-1} \to (0, \infty)$ and a normalized dispersion matrix Σ_0, a scaling parameter $c > 0$, and a location parameter $\mu \in \mathbb{R}^d$. The density f_X satisfies*

$$\begin{aligned} f_X(x) \;=\; & \frac{\Gamma(\frac{1}{2}(\nu(s(x-\mu))+d))}{\Gamma(\frac{1}{2}\nu(s(x-\mu)))(\pi\nu(s(x-\mu)))^{d/2}\det(c\Sigma_0)^{1/2}} \\ & \cdot\left(1+\frac{(x-\mu)'(c\Sigma_0)^{-1}(x-\mu)}{\nu(s(x-\mu))}\right)^{-(\nu(s(x-\mu))+d)/2}. \end{aligned}$$

Figure 3 depicts the density contour lines of two multi-tail t distributions with dispersion matrix $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. In Figure 3 (a) we have the pc-tail function

$$\begin{aligned} \nu(s) \;=\; & < s, F_1 >^2 I_{(0,\infty)}(< s, F_1 >) \cdot 5 + < s, F_1 >^2 I_{(-\infty,0)}(< s, F_1 >) \cdot 2.5 \\ & + < s, F_2 >^2 \cdot 5 \end{aligned}$$

and in Figure 3 (b)

$$\begin{aligned} \nu(s) \;=\; & < s, F_1 >^2 I_{(0,\infty)}(< s, F_1 >) 5 + < s, F_1 >^2 I_{(-\infty,0)}(< s, F_1 >) \cdot 3 \\ & + < s, F_2 >^2 \cdot 5. \end{aligned}$$

We observe in Figure 3 that the shape of the contour lines around the mean $(0,0)'$ is determined by the dispersion matrix Σ. But in the tails (i.e., far away of $(0,0)'$) the influence of the tail function starts to dominate. In particular, outliers in a cone

[11] Ig(α,β) is the inverse gamma distribution for further information (see Rachev and Mittnik (2000) or McNeil, Frey and Embrechts (2005)).

[12] We introduce normalization criteria for a dispersion matrix Σ in Section 4.

around $-F_1$ are much more likely than in any other direction. Figure 4 shows again the densities contour lines of two multi-tail t distributions with the same dispersion matrices as in Figure 3. In Figure 3 (a) the tail function is

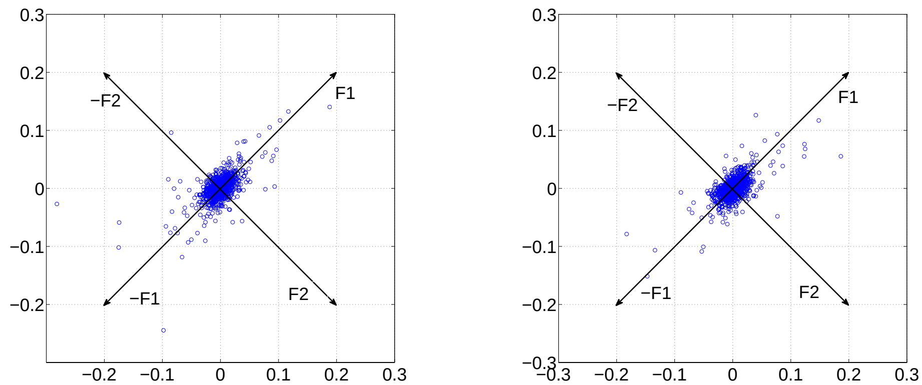
$$\nu(s) = < s, F_1 >^2 \cdot 3 + < s, F_2 >^2 \cdot 6$$

and in Figure 3 (b) we have

$$\begin{aligned} \nu(s) &= < s, F_1 >^2 \cdot 6 + < s, F_2 >^2 I_{(0,\infty)}(< s, F_2 >) \cdot 6 \\ &\quad + < s, F_2 >^2 I_{(-\infty,0)} \cdot 6. \end{aligned}$$

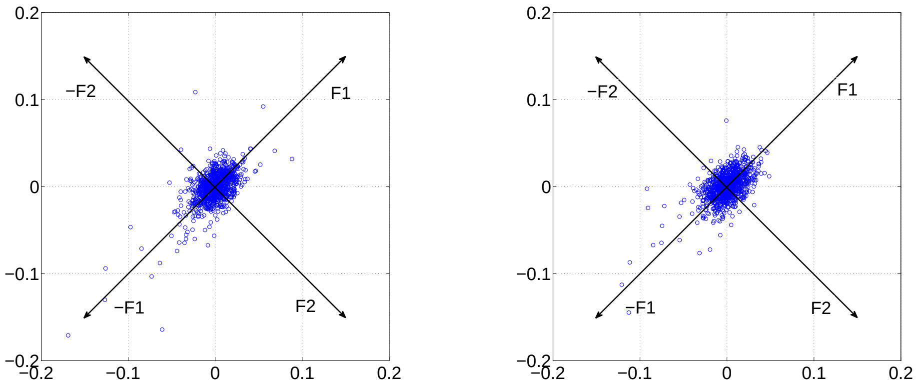
In Figure 3 (a) the tail function supports the scaling properties of the dispersion matrix induced by Σ. In Figure 3 (b) we have the opposite effect in the cone around $-F_2$.

Summing up: The matrix Σ determines the shape of the distribution around the mean while the influence of the tail functions increases in the tails of a multi-tail elliptical distribution.



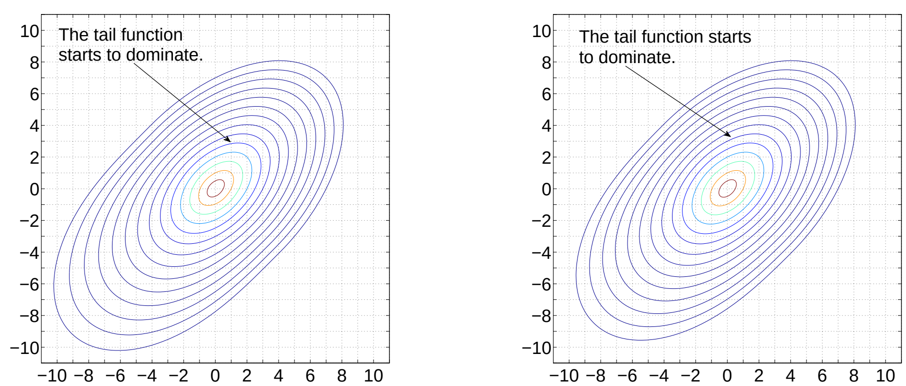
(a) (b)

Figure 1: Scatterplots of a multi-tail α-stable sub-Gaussian distributions with symmetric pc-tail functions



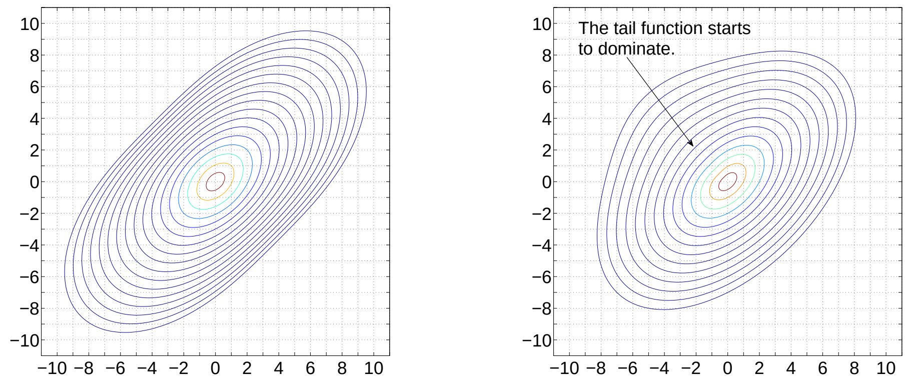
(a) (b)

Figure 2: Scatterplots of a multi-tail α-stable sub-Gaussian distributions with asymmetric pc-tail functions



(a) (b)

Figure 3: Contour lines of the densities of two asymmetric multi-tail t distributions



(a) (b)

Figure 4: (a) depicts the density contour lines of a symmetric multi-tail t distribution; (b) the density contour lines of a asymmetric multi-tail t distribution

4 Estimation of Multi-Tail Elliptical Distributions

In this section we explain how to estimate the parameters of a multi-tail elliptical random vector X. In Section 3 we explained how a multi-tail elliptical random vector $X = \mu + R_{\alpha(s(AS))}AS$ depends on the location parameter μ, the dispersion matrix $\Sigma = AA'$, the tail function $\alpha(.)$, and the family of random variables $(R_\alpha)_{\alpha \in I}$ that is a one parametric family in most applications. We present a three-step estimation procedure to fit the multi-tail elliptical random vectors to data.

In the first step, we estimate the location vector $\mu \in \mathbb{R}^d$ with some robust method. In the second step, we estimate the dispersion matrix $\Sigma \in \mathbb{R}^{d\times d}$ up to a scaling constant $c > 0$ using the spectral estimator that was developed by Tyler (1987a,1987b) and also investigated by Frahm (2004); in the third step, we estimate the scaling constant c and the tail function $\alpha(.)$ applying again the maximum likelihood method. Since we have an analytic expression for the density of a multi-tail elliptical distribution, we could in principle estimate all parameters in a single optimization step. But this approach is not recommended, at least in higher dimensions, because it leads to an extremely complex optimization problem.

As in the classical elliptical case (see McNeil, Frey and Embrechts (2005)), a dispersion matrix of a multi-tail elliptical random vector is only determined up to a scaling constant because of

$$X = \mu + cR_{\alpha(s(AS))}\frac{A}{c}S,$$

for $c > 0$. Hence we have to normalize it. If second moments exist, one can normalize the dispersion matrix by the the covariance matrix (see McNeil, Frey and Embrechts (2005)). In general, the following normalization schemes are always applicable

$$\text{(i) } \Sigma_{11} = 1 \quad \text{(ii) } \det(\Sigma) = 1 \quad \text{(iii) } \operatorname{tr}(\Sigma) = 1, \tag{18}$$

even though second moments do not exist. For the remainder of this section we denote a normalized dispersion matrix by Σ_0.

In the third step we have to estimate the scale parameter c and the tail function $\alpha(.)$. Since we assume a pc-tail function, we have to evaluate the tail parameters $(\alpha_1, ..., \alpha_k) \in I^k$, $k \in \mathbb{N}$, of the pc-tail function. In the last step we determine the parameters from the set $\Theta = \mathbb{R}_+ \times I^k$, where I is the interval of tail parameters.

4.1 Estimation of the Dispersion Matrix

We present a robust estimator of the dispersion matrix of multi-tail elliptical distribution based on the work of Tyler (1987a,1987b) and Frahm (2004). The so-called spectral estimator estimates the dispersion matrix up to a scaling constant. Furthermore, we assume the location parameter μ to be known.

Definition 4.1 (Unit random vector). Let $A \in \mathbb{R}^{d\times d}$ be a regular matrix and S be uniformly distributed on the hypersphere $\mathcal{S}^{d-1}$. We call the random vector

$$S_A := \frac{AS}{||AS||}$$

the "unit random vector generated by A".

Let $X = \mu + R_{\alpha(s(AS))}AS$ be a multi-tail elliptical random vector where the location vector is assumed to be known. Then we obtain

$$\begin{aligned} \frac{X-\mu}{||X-\mu||} &= \frac{R_{\alpha(s(AS))}AS}{||R_{\alpha(s(AS))}AS||} \\ &= \frac{AS}{||AS||} \\ &= S_A \end{aligned} \tag{19}$$

The family of random variables $(R_\alpha)_{\alpha\in I}$ and the tail function $\alpha(.)$ have no influence $s(X-\mu) = S_A$. This is important since it allows for a robust estimation of the dispersion matrix Σ up to a scaling constant. In particular, it follows from equation (19) immediately that we have $S_{cA} = S_A$ for all $c > 0$.

Theorem 4.1. *The spectral density function of the unit random vector generated by $A \in \mathbb{R}^{d\times d}$ satisfies*

$$f_{S_A}(s) = \frac{\Gamma(d/2)}{2\pi^{d/2}} \frac{(s'\Sigma^{-1}s)^{-d/2}}{\det(\Sigma)^{1/2}}$$

for all $s \in \mathcal{S}^{d-1}$, where $\Sigma = AA'$ is the dispersion matrix. The distribution f_{S_A} is called angular central Gaussian distribution (see Kent and Tyler (1988)).

Proof. Because of the invariance property described in equation (19) we can assume without loss of generality that X is normally distributed with location parameter 0 and covariance matrix Σ, i.e. $X \sim N(0, \Sigma)$, and we have

$$X = RAS \sim N_d(0, \Sigma)$$

where $R^2 \sim \chi^2_d$ and S is uniformly distributed on the hypersphere $\mathcal{S}^{d-1}$.[13] The density of X under the condition $||X|| = r > 0$ is

$$f_{X|(||X||)}(x|r) = \frac{f_X(x)}{f_{||X||}(r)} 1_{\mathcal{S}_r^{d-1}}(x), \tag{20}$$

where f_X is the Gaussian density with parameters $(0, \Sigma)$, $\mathcal{S}_r^{d-1} = \{x \in \mathbb{R}^d : ||x|| = r\}$ and $f_{||X||}(r) = \int_{\mathcal{S}_r^d} f_X(x)dx$. In order to obtain the spectral density of

$$S_{\Sigma^{1/2}} = \frac{X}{||X||},\ X \sim N_d(0, \Sigma),$$

we define the transformation $h_r : \mathcal{S}_r^d \to \mathcal{S}^{d-1}$, $x \mapsto x/||x|| = s$. Further, we have that $h_r^{-1} : \mathcal{S}^{d-1} \to \mathcal{S}_r^d$, $s \mapsto rs$ and $Dh_r^{-1} = rId_{d-1}$. Hence we obtain $|\det(Dh_r^{-1})| = r^{d-1}$. Let f_r be defined by

$$f_r(s) = f_{h_r^{-1}(X)|(||X||)}(s|r).$$

Due to Theorem 7.1 given in the appendix we have

$$\begin{aligned} f_r(s) &= f_{X|(||X||}(h_r^{-1}(s))|\det(Dh_r^{-1})| \\ &= \frac{f_X(rs)r^{d-1}}{f_{||X||}(r)}. \end{aligned}$$

Thus, the density of $S_{\Sigma^{1/2}}$ is given by

$$\begin{aligned} f_{S_{\Sigma^{1/2}}}(s) &= \int_0^\infty f_r(s) f_{||X||}(r)dr = \int_0^\infty f_X(rs)r^{d-1}dr \\ &= \int_0^\infty \frac{\det(\Sigma)^{-1/2}}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}(rs)'\Sigma(rs)\right) r^{d-1}dr. \end{aligned}$$

Substituting r by $\sqrt{2t/s'\Sigma^{-1}s}$ leads to

$$\begin{aligned} f_{S_{\Sigma^{1/2}}}(s) &= \int_0^\infty \frac{\det(\Sigma)^{-1/2}}{(2\pi)^{d/2}} \exp(-t)(s'\Sigma^{-1}s)^{-d/2}dt \\ &= \frac{\det(\Sigma)^{-1/2}}{(2\pi)^{d/2}}(s'\Sigma^{-1}s)^{-d/2} \int_0^\infty \exp(-t)t^{d/2-1}dt \\ &= \frac{\det(\Sigma)^{-1/2}}{2\pi^{d/2}}(s'\Sigma^{-1}s)^{-d/2}\Gamma(d/2), \end{aligned}$$

where $s \in \mathcal{S}^{d-1}$ and $\Gamma(.)$ is the gamma function (see Definition 7.1). □

Corollary 4.1. *Let $X = \mu + R_{\alpha(s(AS))}AS \in \mathbb{R}^d$ be a multi-tail elliptical random vector, where $\Sigma = AA'$. Then the joint distribution of $R = \sqrt{(X-\mu)\Sigma^{-1}(X-\mu)}$ and S_A is given by*

$$f_{(R,S_A)}(r, s) = \frac{\Gamma(d/2)}{(2\pi)^{d/2}} \frac{(s'\Sigma^{-1}s)^{-d/2}}{(\det \Sigma)^{1/2}} f_{R_{\alpha(s)}}(r).$$

[13] χ^2_d is the d-dimensional chi-square distribution (see Feller (1971)); see Fang, Kotz and Ng (1987) for the representation of the normal distribution in terms of elliptical distributions.

Proof. We observe

$$\begin{aligned} f_{(R,S_A)}(r,s) &= f_{S_A}(s) f_{R|S_A}(r|s) \\ &= \frac{\Gamma(d/2)}{2\pi^{d/2}} \frac{(s'\Sigma^{-1}s)^{-d/2}}{(\det \Sigma)^{-1/2}} f_{R_{\alpha(s)}}(r). \end{aligned}$$

□

Since we know the density of the random vector S_A we can apply the maximum likelihood method to estimate the normalized dispersion matrix Σ_0 of a multi-tail elliptical random vector.

Let $X_1, ..., X_n \in \mathbb{R}^d$ be a sample of identically distributed multi-tail elliptical data vectors. Thus, the data vectors $S_i = (X_i - \mu)$, $i = 1, ..., n$, are angular central Gaussian distributed. The log-likelihood function of the sample $(S_1, S_2, ..., S_n)$ equals

$$\begin{aligned} \log\left(\prod_{i=1}^n f_{S_A}(s_i)\right) &= \sum_{i=1}^n \log(f_{S_A}(S_i)) \\ &= \sum_{i=1}^n \Big(\log(\Gamma(d/2)) - \log(2\pi^{d/2}) \\ &\quad + \log((S_i'\Sigma^{-1}S_i)^{-d/2}) - \log(\det(\Sigma)^{1/2})\Big) \\ &= n\log(\Gamma(d/2)) - \frac{nd}{2}\log(2\pi) - \frac{n}{2}\log(\det(\Sigma)) \\ &\quad - \frac{d}{2}\sum_{i=1}^n \log(S_i'\Sigma^{-1}S_i). \end{aligned} \tag{21}$$

Since in equation (21) $n\log(\Gamma(d/2))$ and $nd\log(2\pi)/2$ are constants subject to Σ we can neglect them in the following maximum likelihood optimization problem

$$\hat{\Sigma} = \operatorname{argmax}_{\Sigma \in D_d^2} -n \log\det(\Sigma) - d\sum_{i=1}^n \log(S_i'\Sigma^{-1}S_i), \tag{22}$$

where $D_d^2 \subset \mathbb{R}^{d\times d}$ denotes the set of positive definite $d \times d$ matrices. Note that equation (21) determines $\hat{\Sigma}$ up to a scale parameter $c > 0$. If $\hat{\Sigma}$ satisfies this equation, it can be seen immediately that it holds for $c\hat{\Sigma}$ as well.

Definition 4.2. The maximum likelihood estimator $\hat{\Sigma}$ defined by equation (21) is also called the spectral estimator.

In Theorem 4.2 we interpret D_d^2 as an open subset of $\mathbb{R}^{d(d+1)/2}$ vector space. Thus, we interpret the matrix Σ as a vector in this space. Note that if $\Sigma \in D_d^2$ then also $\Sigma^{-1} \in D_d^2$. Let $f : \mathbb{R}^{d(d+1)/2} \to \mathbb{R}$ be differentiable in the point Σ. Then, we denote the Jacobian with $Df(\Sigma)$.

Theorem 4.2. *Let $X \in \mathbb{R}^d$ be a random vector with density*

$$f_X : \begin{cases} \mathbb{R}^d & \to \quad \mathbb{R} \\ x & \mapsto \quad \sqrt{\det(\Sigma^{-1})}\, g((x-\mu)'\Sigma^{-1}(x-\mu)) \end{cases} \tag{23}$$

Then the function

$$h_x : \left\{ \begin{array}{lcl} D_d^2 \subset \mathbb{R}^{d(d+1)/2} & \to & \mathbb{R} \\ \Sigma & \mapsto & \log\left(\sqrt{\det(\Sigma)}g((x-\mu)'\Sigma(x-\mu))\right) \end{array} \right. \qquad (24)$$

has the Jacobian

$$\begin{aligned} Dh_x(\Sigma) &= \frac{1}{2}(2\Sigma^{-1} - \mathrm{diag}(\Sigma^{-1})) \\ &\quad + \frac{d(\log(g(z)))}{dz}(2(x-\mu)'(x-\mu) - \mathrm{diag}((x-\mu)'(x-\mu))) \end{aligned} \qquad (25)$$

where $z = (x-\mu)'\Sigma(x-\mu)$.

Note $Dh_x(.) \in \mathbb{R}^{1\times d(d+1)/2}$.

Proof. The canonical basis of the vector space of all symmetric matrices in $\mathbb{R}^{d\times d}$, denoted $S^{d\times d}$, is

$$f_{ij} = \left\{ \begin{array}{ccc} e_{ii} & , & i=j \\ e_{ij}+e_{ji} & , & i<j \end{array} \right.$$

where $1 \le i \le j \le d$.

Our goal is to prove equation (25). If Σ is positive definite, then Σ^{-1} is positive definite, too. Due to Theorem 7.2 in the appendix, the equation

$$D\det(A) = \det(A)A^{-1} \qquad (26)$$

holds of all symmetric and regular matrices. Note that we have according to equation (24)

$$h_x(\Sigma) = \frac{1}{2}\log\det(\Sigma) - \log(g((x-\mu)'\Sigma(x-\mu))).$$

We obtain the following partial derivatives with respect to $(f_{ij})_{1\le i\le j\le d}$.

$$\begin{aligned} \frac{\partial h_x(\Sigma)}{\partial f_{ij}} &= \frac{1}{2}\frac{1}{\det(\Sigma)}\frac{\partial\det(\Sigma)}{\partial f_{ij}} + \frac{\partial\log(g(z))}{\partial z}\frac{\partial(x-\mu)'\Sigma(x-\mu)}{\partial f_{ij}} \\ &= \frac{1}{2}\frac{1}{\det(\Sigma)}D\det(\Sigma)(e_{ij}+e_{ji}) + 2\frac{\partial\log(g(z))}{\partial z}(x_i-\mu_i)(x_j-\mu_j) \\ &\overset{(26)}{=} \frac{1}{2}\frac{1}{\det(\Sigma)}\det(\Sigma)\Sigma^{-1}(e_{ij}+e_{ji}) + 2\frac{\partial\log(g(z))}{\partial z}(x_i-\mu_i)(x_j-\mu_j) \\ &= \Sigma_{ij}^{-1} + 2\frac{\partial\log(g(z))}{\partial z}(x_i-\mu_i)(x_j-\mu_j), \end{aligned}$$

where $1 \le i < j \le d$.

$$\begin{aligned} \frac{\partial h_x(\Sigma)}{\partial f_{ii}} &= \frac{1}{2}\frac{1}{\det(\Sigma)}\frac{\partial\det(\Sigma)}{\partial f_{ii}} + \frac{\partial\log(g(z))}{\partial z}\frac{\partial(x-\mu)'\Sigma(x-\mu)}{\partial f_{ii}} \\ &\overset{(26)}{=} \frac{1}{2}\frac{1}{\det(\Sigma)}\det(\Sigma)\Sigma^{-1}e_{ii} + \frac{\partial\log(g(z))}{\partial z}(x_i-\mu_i)^2 \\ &= \frac{1}{2}\Sigma_{ii}^{-1} + \frac{\partial\log(g(z))}{\partial z}(x_i-\mu_i)^2, \end{aligned}$$

where $i = 1, ..., d$. Hence, equation (25) follows immediately. □

Let $X, X_1, ..., X_n \in \mathbb{R}^d$ be a sample of identically distributed data vectors with a density of the form

$$f_X(x) \quad = \quad \sqrt{\det(\Sigma^{-1})} g((x-\mu)'\Sigma^{-1}(x-\mu))$$

then the maximum likelihood estimator $\hat{\Sigma}$ of the dispersion matrix Σ satisfies the following equation

$$\sum_{i=1}^{n} Dh_{X_i}(\Sigma^{-1}) \stackrel{!}{=} 0, \tag{27}$$

since it is a necessary condition for the maximum.

Theorem 4.3. *Let $X, X_1, ..., X_n \in \mathbb{R}^d$ be a sample of identically distributed data vectors with a density of the form*

$$f_X(x) = \sqrt{\det(\Sigma^{-1})} g((x-\mu)'\Sigma^{-1}(x-\mu)).$$

Then the solution to the optimization problem in equation (22) can be characterized by the following fix point equation

$$\hat{\Sigma} = -\frac{2}{n} \sum_{k=1}^{n} \frac{\partial \log(g(z_k))}{\partial z_k} (X_k - \mu)(X_k - \mu)', \tag{28}$$

where $z_k = (X_k - \mu)'\hat{\Sigma}^{-1}(X_k - \mu)$.

Equation (28) is only a necessary condition for the solution of the optimization problem stated in equation (22).

Proof. We show that equation (28) is true for every entry $\hat{\Sigma}_{ij}$ with $1 \le i \le j \le d$. We denote the ith component of the random vector X_k with X_{ki}.

(i) In the case $i = j$ we observe

$$\begin{aligned}
0 \quad &\stackrel{!}{=} \quad \sum_{k=1}^{n} Dh_{X_k}(\hat{\Sigma}^{-1})_{ii} \\
&= \quad \sum_{k=1}^{n} \frac{1}{2}(2\hat{\Sigma}_{ii} - \hat{\Sigma}_{ii}) - \frac{\partial \log(g(z_k)))}{\partial z_k}(2(X_{ki} - \mu_i)^2 - (X_{ki} - \mu_i)^2) \\
&= \quad \frac{1}{2} \sum_{k=1}^{n} \hat{\Sigma}_{ii} - \sum_{k=1}^{n} \frac{\partial \log(g(z_k))}{\partial z_k}(X_{ki} - \mu_i)^2 \\
&\Leftrightarrow \quad \hat{\Sigma}_{ii} = \frac{2}{n} \sum_{k=1}^{n} \frac{\partial \log(g(z_k))}{\partial z_k}(X_{ki} - \mu_i)^2.
\end{aligned}$$

where $z_k = (X_k - \mu)'\hat{\Sigma}^{-1}(X_k - \mu)$.

(ii) In the case $i < j$ we obtain

$$
\begin{aligned}
0 &\overset{!}{=} \sum_{k=1}^{n} D_{X_k}(\hat{\Sigma}^{-1})_{ij} \\
&= \sum_{k=1}^{n} \frac{1}{2}(2\hat{\Sigma}_{ij} - 0) - \frac{\partial \log(g(z_k))}{\partial z_k}(2(X_{ki} - \mu_i)(X_{kj} - \mu_j) - 0) \\
&= \sum_{k=1}^{n} \hat{\Sigma}_{ij} - 2\sum_{k=1}^{n} \frac{\partial \log(g(z_k))}{\partial z_k}(X_{ki} - \mu_i)(X_{kj} - \mu_j) \\
\Leftrightarrow &\quad \hat{\Sigma}_{ij} = \frac{2}{n}\sum_{k=1}^{n} \frac{\partial \log(g(z_k))}{\partial z_k}(X_{ki} - \mu_i)(X_{kj} - \mu_j),
\end{aligned}
$$

where $z_k = (X_k - \mu)'\hat{\Sigma}^{-1}(X_k - \mu)$ and $1 \leq i < j \leq d$. Hence, we obtain

$$
\hat{\Sigma} = -\frac{2}{n}\sum_{k=1}^{n} \frac{\partial \log(g(z_k))}{\partial z_k}(X_k - \mu)(X_k - \mu).
$$

□

Let S_A be a unit random vector generated by A with density f_{S_A}. The density f_{S_A} is of the form given in equation (23) and satisfies

$$
g(z) = \frac{\Gamma(d/2)}{2\pi^{d/2}} z^{-d/2}. \tag{29}
$$

The derivative of $\log(g(.))$ satisfies

$$
\begin{aligned}
\log(g(z))' &= \left(\log\left(\frac{\Gamma(d/2)}{2\pi^{d/2}}\right) - \frac{d}{2}\log(z)\right)' \\
&= -\frac{d}{2}z^{-1},
\end{aligned} \tag{30}
$$

where $z > 0$.

Let $S_1, S_2, ..., S_n$ be a sample of identically distributed unit data vectors generated by A. According to Theorem 4.3 and equation (30), the maximum likelihood estimator $\hat{\Sigma}$ satisfies the following fix point equation

$$
\begin{aligned}
\hat{\Sigma} &= -\frac{2}{n}\sum_{i=1}^{n}(-1)\frac{d}{2}(S_i'\hat{\Sigma}^{-1}S_i)^{-1}S_iS_i'. \\
&= \frac{d}{n}\sum_{i=1}^{n}\frac{S_iS_i'}{S_i'\hat{\Sigma}^{-1}S_i}.
\end{aligned}
$$

Furthermore, let $X_1, X_2, ..., X_n$ be a sample of identically distributed multi-tail elliptical data vectors. Due to equation (19) the sample $s(X_1 - \mu), ..., s(X_n - \mu)$

consists of unit data vectors generated by A. Hence, we obtain

$$\begin{aligned}
\hat{\Sigma} &= \frac{d}{n}\sum_{i=1}^{n}\frac{s(X_i-\mu)s(X_i-\mu)'}{s(X_i-\mu)'\Sigma^{-1}s(X_i-\mu)} \\
&= \frac{d}{n}\sum_{i=1}^{n}\frac{((X_i-\mu)/||X_i-\mu||)((X_i-\mu)/||X_i-\mu||)'}{((X_i-\mu)/||X_i-\mu||)'\hat{\Sigma}^{-1}((X_i-\mu)/||X_i-\mu||)} \\
&= \frac{d}{n}\sum_{i=1}^{n}\frac{(X_i-\mu)(X_i-\mu)'}{(X_i-\mu)'\hat{\Sigma}^{-1}(X_i-\mu)}. \qquad (31)
\end{aligned}$$

Equation (31) is the fixed point representation of the maximum likelihood problem given in equation (22) in terms of the original data $X_1,...,X_n$. It is important to note that the family $(R_\alpha)_{\alpha\in I}$ of random variables has no influence on the fix point representation.

It can be seen immediately that the fix point equation determines $\hat{\Sigma}$ only up to a scale parameter since we have for all $c>0$

$$\begin{aligned}
c\hat{\Sigma} &= c\frac{d}{n}\sum_{i=1}^{n}\frac{(X_i-\mu)(X_i-\mu)'}{(X_i-\mu)'\hat{\Sigma}^{-1}(X_i-\mu)} \\
&= \frac{d}{n}\sum_{i=1}^{n}\frac{(X_i-\mu)(X_i-\mu)'}{(X_i-\mu)'\frac{1}{c}\hat{\Sigma}^{-1}(X_i-\mu)} \\
&= \frac{d}{n}\sum_{i=1}^{n}\frac{(X_i-\mu)(X_i-\mu)'}{(X_i-\mu)'(c\hat{\Sigma})^{-1}(X_i-\mu)}.
\end{aligned}$$

In order to have uniqueness of the optimization problem, we apply one of the optimization criteria given in equation (18).

The next theorem is very important since it says when a solution to the optimization problem stated in equation (22) exists and when it is unique. Furthermore, it shows an iterative algorithm to approximate the solution.

Theorem 4.4. *Let $X_1,...,X_n$ be a sample of identically distributed multi-tail elliptical data vectors with $n>d(d-1)$. Then a fix point $\hat{\Sigma}$ of the equation*

$$\hat{\Sigma} = \frac{d}{n}\sum_{i=1}^{n}\frac{(X_i-\mu)(X_i-\mu)'}{(X_i-\mu)'\hat{\Sigma}^{-1}(X_i-\mu)}$$

exists and it is unique up to a scale parameter. In particular, the normalized sequence $(\tilde{\Sigma}_0^{(i)})_{i\in\mathbf{N}}$ defined by

$$\begin{aligned}
\tilde{\Sigma}_0 &= \mathrm{Id} \\
\dot{\Sigma}^{(i+1)} &= \frac{d}{n}\sum_{i=1}^{n}\frac{(X_i-\mu)'(X_i-\mu)}{(X_i-\mu)'(\tilde{\Sigma}_0^{(i)})^{-1}(X_i-\mu)} \\
\tilde{\Sigma}_0^{(i+1)} &= \dot{\Sigma}_0^{(i+1)} \qquad (32)
\end{aligned}$$

converges a.s. to $\hat{\Sigma}_0$, that is a normalized version of $\hat{\Sigma}$.

Proof. Since a multi-tail elliptical distribution is absolute continuous to the Lebesque measure λ^d, the probability that two vectors in the sample $\{X_i : i = 1, .., n\}$ are linear dependent is 0. So the Condition 2.1 in Tyler (1987a, p.236) holds and we can apply corollary 2.2 in the same publication. □

Theorem 4.4 says that the maximum likelihood optimization problem stated in equation (22) has a solution that is unique up to a scale parameter. We can use an iterative algorithm to find the solution $\hat{\Sigma}$ if we have more than $d(d-1)$ observations. Fortunately, the algorithm works also quite well if we have only a sample size n which is slightly larger than d.

Finally, we remark that

Corollary 4.2. *Let $X_1, ..., X_n$ be a d-dimensional multi-tail elliptical distributed sample with known location parameter μ and $n \geq d$. Then the sequence $(\tilde{\Sigma}^{(i)})_{i\in\mathbb{N}}$ is positive definite for all $i \in \mathbb{N}$.*

Proof. Since the sample $X_1, ..., X_n$ is multi-tail elliptically distributed with $n \geq d$ the matrix

$$\sum_{j=1}^{n}(X_j - \mu)(X_j - \mu)'$$

is positive definite almost sure. Since $\tilde{\Sigma}^{(i)}$ is positive definite, the quantity $w_j = \frac{d}{(X_j-\mu)\hat{\Sigma}^{(i)}(X_j-\mu)'}$ is positive. We obtain

$$\dot{\Sigma}^{(i+1)} = \frac{1}{n}\sum_{j=1}^{n}(\sqrt{w_j}(X_j - \mu))(\sqrt{w_j}(X_j - \mu))'.$$

From the last equation it is obvious that $\dot{\Sigma}^{(i+1)}$ is positive definite. Hence,

$$\tilde{\Sigma}^{(i+1)} = \dot{\Sigma}_0^{(i+1)}$$

is positive definite. Finally, Corollary 4.2 follows with induction. □

4.1.1 Empricial Analysis of the Spectral Estimator

For the empirical analysis of the spectral estimator we generate samples from a multi-tail α-stable sub-Gaussian distribution. In particular, we choose the location parameter μ to be $(0,0)'$ and the dispersion matrix Σ to be $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$. The tail function $\alpha(.)$ satisfies

$$\alpha(s) =< s, F_1 >^2 \cdot 1.7 + < s, F_2 >^2 \cdot 1.9$$

where $F_1 = s((1,1)')$ and $F_2 = s((1,-1)')$ are the first two principal components of Σ. We denote with $\hat{\Sigma}_0(n)$ the estimate of the spectral estimator with n iterations. Since the spectral estimator is unique up to a scaling constant, we normalize it by the condition $\sigma_{11} = 1$.

Figures 5 to 8 depict boxplots of the spectral estimator $\hat{\Sigma}_0(n_i)$, $n_i = 20, 50, 100$ and $i = 1, 2, 3$. Each boxplot illustrated in these figures consists of 1000 estimates. All figures have in common that increasing the amount of iterations has only a minor influence on the accuracy of the estimator, while increasing the sample size improves its behavior. Furthermore, we observe that the median $q_{0.5}$ is very close to the true values for all sample sizes, see Tables 1 and 2 and Figures 5 to 8, respectively. Tables 1 and 2 depicts several quantiles of the empirical distribution of the spectral estimator for different samples sizes. The relative error is defined as

$$e_{rel}^{1-2\alpha} = \max\left\{ \left| \frac{q_\alpha - c}{c} \right|, \left| \frac{q_{1-\alpha} - c}{c} \right| \right\} \tag{33}$$

while c denotes the true value and $q_\alpha \in \mathbb{R}$ the α-quantile. In particular, $e_{rel}^{1-2\alpha}$ tells us that the relative error of an estimate is smaller than $e_{rel}^{1-2\alpha}$ with probability $1 - 2\alpha$.[14] We observe that the relative errors in both tables are approximately equal. Hence, estimators $\hat{\sigma}_{12}$ and $\hat{\sigma}_{22}$ have practically the same accuracy. Furthermore, the quantiles are nearly symmetric around the true value for larger samples sizes, i.e. 1000 and 2000.

size	$q_{0.05}$	$q_{0.1}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.9}$	$q_{0.95}$	$e_{rel}^{0.5}$	$e_{rel}^{0.9}$
100	0.285	0.335	0.406	0.498	0.584	0.656	0.713	18%	43%
500	0.406	0.431	0.463	0.498	0.533	0.567	0.584	7%	18%
1000	0.436	0.450	0.473	0.498	0.526	0.551	0.562	5.4%	13%
2000	0.453	0.466	0.481	0.499	0.517	0.536	0.546	3.8%	10%

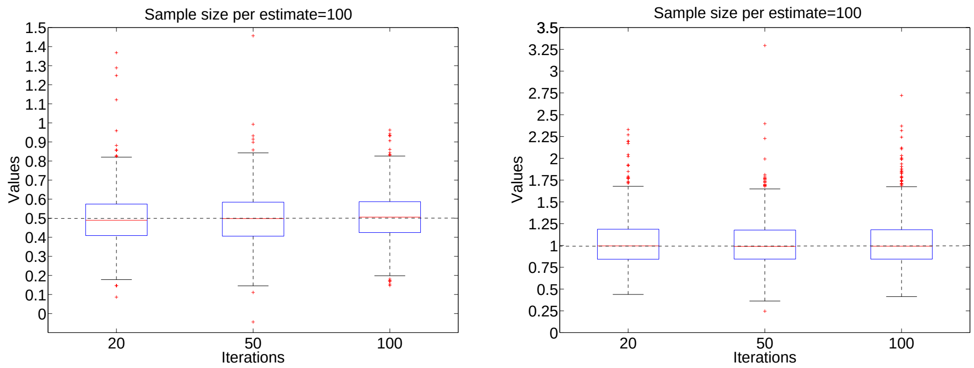
Table 1: The table depicts different quantiles of the spectral estimator $\hat{\sigma}_{12}$ and its relative errors.

size	$q_{0.05}$	$q_{0.1}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.9}$	$q_{0.95}$	$e_{rel}^{0.5}$	$e_{rel}^{0.9}$
100	0.664	0.719	0.844	0.989	1.177	1.367	1.498	17%	50%
500	0.826	0.866	0.924	1	1.074	1.151	1.198	7%	20%
1000	0.877	0.902	0.949	1	1.054	1.102	1.13	5.4%	13%
2000	0.904	0.926	0.960	0.996	1.034	1.068	1.09	4%	10%

Table 2: The table depicts different quantiles of the spectral estimator $\hat{\sigma}_{22}$ and its relative errors.

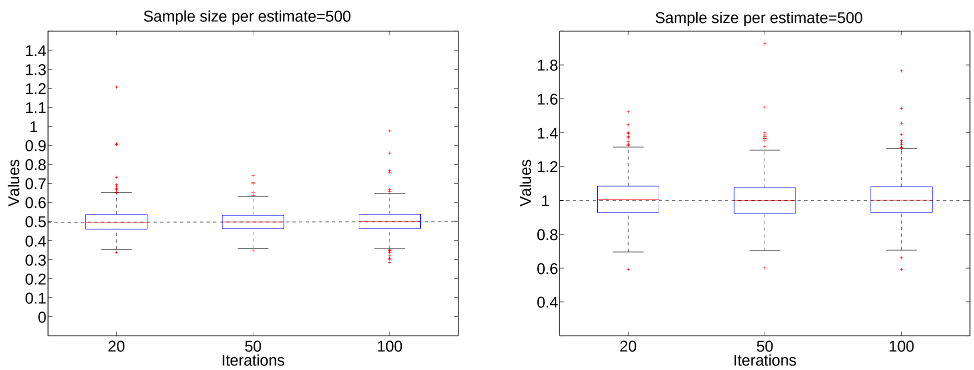
In Figure 5, we depict all estimate of spectral estimator lying in range between -0.0439 and 1.45 for and 0.246 and 3.2942. In Figure 6 we do not depict one upper outlier (value=2.645) in (a) and in (b) three upper outliers (values=$7.1, 7.0, 3.8$). In Figure 7 (b) we do not show three upper outliers (values=1.8,2.69,2.3) and in Figure 8 we do not illustrate one upper outlier (value=1.14) and one lower (value=0.24) in (a) and in (b) one upper outlier (value=2.76). We are a bit surprised to observe such large outliers when the sample size is relatively large, i.e. 1000 and 2000. The reason might be numerical instabilities of the spectral estimator.

[14] In order to be consistent, we require α to be in $(0, 0.5)$.



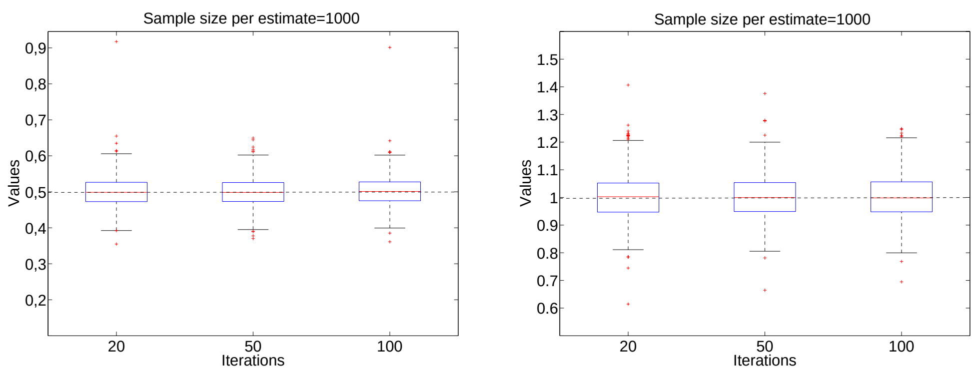
(a) (b)

Figure 5: Boxplots of the spectral estimator with sample size 100 per estimate. (a) depicts the estimates of σ_{12} and (b) of σ_{22}.



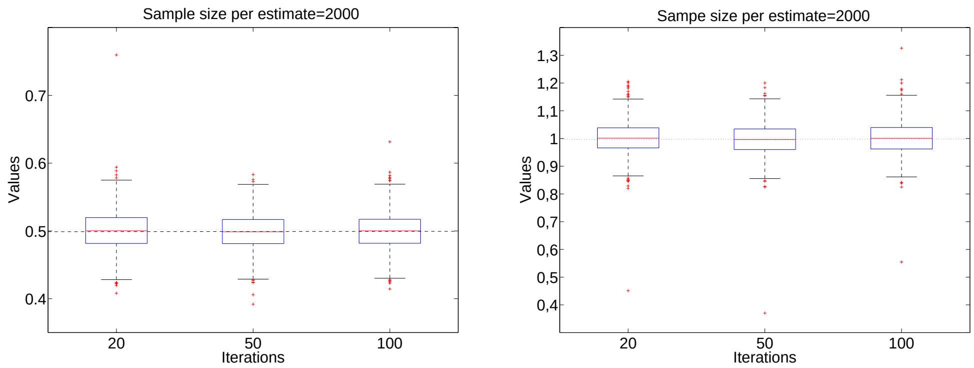
(a) (b)

Figure 6: Boxplots of the spectral estimator with sample size 500 per estimate. (a) depicts the estimates of σ_{12} and (b) of σ_{22}.



(a) (b)

Figure 7: Boxplots of the spectral estimator with sample size 1000 per estimate. (a) depicts the estimates of σ_{12} and (b) of σ_{22}.



(a) (b)

Figure 8: Boxplots of the spectral estimator with sample size 2000 per estimate. (a) depicts the estimates of σ_{12} and (b) of σ_{22}.

4.2 Estimation of the Parameter Set Θ

At the beginning of this section we explained that a multi-tail elliptical random vector can be estimated by a three-step procedure. In the following we deal with the third step. We assume that we have already estimated the location parameter μ and the normalized dispersion matrix Σ_0. We can write the multi-tail elliptical random vector X in the form

$$X = \mu + cR_{\alpha(s(\Sigma_0^{1/2}S))}\Sigma_0^{1/2}S, \tag{34}$$

where $c > 0$ is a thus far an unknown scale parameter[15] and $\alpha(.)$ is the tail function. Since we assume $\alpha(.)$ to be a pc-tail function (see Section 3.2) it can be determined by the tail parameters $(\alpha_1, ..., \alpha_k) \in I^k$, $k \in \mathbb{N}$. Hence, we have to estimate

$$(c, \alpha_1, \alpha_2, ..., \alpha_k) \in \mathbb{R}_+ \times I^k = \Theta.$$

In the following we present two equivalent methods to estimate the parameters $(c, \alpha_1, ..., \alpha_k) \in \Theta$

4.2.1 Radial Variate ML-Approach

In this approach it follows from Corollary 3.1 and equation (34) that we have

$$\sqrt{(X-\mu)'\Sigma_0^{-1}(X-\mu)}|(s(X-\mu)=s) \overset{d}{=} cR_{\alpha(s)}.$$

Hence for the density of $\sqrt{(X-\mu)'\Sigma_0^{-1}(X-\mu)}$ given $s(X-\mu)=s$ we obtain

$$f_{\sqrt{(X-\mu)'\Sigma_0^{-1}(X-\mu)}|s(X-\mu)}(r|s) = f_{cR_{\alpha(s)}}(r). \tag{35}$$

[15]Note that we cannot estimate c in the second step because in that step the dispersion matrix Σ can only be determined up to this scale parameter.

Proposition 4.1. Let $X \in \mathbb{R}$ be a random variable with density f_X and $Y = cX$, $c > 0$. Then we have

$$f_Y(y) = \frac{1}{c} f_X(y/c),$$

where f_Y denotes the density of Y.

According to Proposition 4.1 we obtain

$$f_{cR_{\alpha(s)}}(r) = \frac{1}{c} f_{R_{\alpha(s)}}(r/c).$$

Let $X_1, ..., X_n \in \mathbb{R}^d$ be a sample of identically distributed multi-tail elliptical data vectors. We assume the location parameter μ and the normalized dispersion matrix Σ_0 to be known. We define the samples

$$R_i = \sqrt{(X_i - \mu)' \Sigma_0^{-1} (X_i - \mu)}, i = 1, ..., n$$

and

$$S_i = s(X_i - \mu), i = 1, ..., n.$$

Then the data vectors $R_1|S_1, R_2|S_2..., R_n|S_n$ are independent since we assume μ and Σ_0 to be known. According to equation (35), the log-likelihood function of this sample satisfies

$$\begin{aligned} \log\left(\prod_{i=1}^{n} f_{cR_{(S_i)}}(R_i)\right) &= \sum_{i=1}^{n} \log\left(\frac{1}{c} f_{R_{\alpha(S_i)}}(R_i/c)\right) \\ &= -n\log(c) + \sum_{i=1}^{n} \log(f_{R_{\alpha(S_i)}}(R_i/c)). \end{aligned}$$

Finally, this leads to the optimization problem

$$\hat{\theta} = \mathrm{argmax}_{\theta\in\Theta} -n\log(c) + \sum_{i=1}^{n} \log\left(f_{R_{\alpha(S_i)}}(R_i/c)\right).$$

Summing up, we obtain

$$\begin{aligned} \hat{\theta}(X_1, ..., X_n) &= \mathrm{argmax}_{\theta\in\Theta} -n\log(c) \\ &\quad + \sum_{i=1}^{n} \log\left(f_{R_{\alpha(s(X_i-\mu))}}\left(\sqrt{(X_i-\mu)'\Sigma_0^{-1}(X_i-\mu)}/c\right)\right). \end{aligned}$$

4.2.2 The Density Generator ML-Approach

The second approach uses directly the density f_X of a multi-tail elliptical random vector. Due to Theorem 3.1, we know that the density satisfies

$$f_X(x) = |\det(c\Sigma_0)|^{-1/2} g_{\alpha(s(x-\mu))}((x-\mu)'(c\Sigma_0)^{-1}(x-\mu)).$$

Since we assume Σ_0 and μ to be known, we obtain the following log-likelihood function

$$
\begin{aligned}
\log\left(\prod_{i=1}^{n} f_X(X_i)\right) &= \sum_{i=1}^{n} \log f_X(X_i) \\
&= \sum_{i=1}^{n} \log\Big(|c^d \det(\Sigma_0)|^{-1/2} \\
&\quad g_{\alpha(s(X_i-\mu))}\left(\frac{||\Sigma^{-1/2}(X_i-\mu)||}{c}\right)\Big) \\
&= \sum_{i=1}^{n} -\frac{d}{2}\log(c) - \frac{1}{2}\log(\det(\Sigma_0)) \\
&\quad + \sum_{i=1}^{n} \log\left(g_{\alpha(s(X_i-\mu))}\left(\frac{(X_i-\mu)\Sigma_0^{-1}(X_i-\mu)}{c}\right)\right) \\
&= -\frac{dn}{2}\log(c) - \frac{n}{2}\log(\det(\Sigma_0)) \\
&\quad + \sum_{i=1}^{n} \log\left(g_{\alpha(s(X_i-\mu))}\left(\frac{(X_i-\mu)'\Sigma_0^{-1}(X_i-\mu)}{c}\right)\right)
\end{aligned}
$$

Since the term $-\frac{n}{2}\log(\det(\Sigma_0))$ is a constant subject to Θ, we can neglect this term in the optimization problem. Thus, we obtain the following log-likelihood optimization problem

$$
\hat{\theta} = \operatorname{argmax}_{\theta\in\Theta} -\frac{dn}{2}\log(c) + \sum_{i=1}^{n} \log\left(g_{\alpha(s(X_i-\mu))}\left(\frac{(X_i-\mu)\Sigma_0^{-1}(X_i-\mu)}{c}\right)\right).
$$

4.2.3 Equivalence of the Approaches

The Radial Variate ML-Approach and the Density Generator ML-Approach are equivalent. This is the case because we have

$$
\begin{aligned}
\log(f_X(x)) &= -\frac{1}{2}\log(\det(c\Sigma_0)) + \log\left(g_{\alpha(s(x-\mu))}\left(\sqrt{(x-\mu)'c\Sigma_0(x-\mu)}\right)\right) \\
&= -\frac{1}{2}\log(\det(\Sigma_0)) \\
&\quad -\frac{d}{2}\log(c) + \log\left(g_{\alpha(s(x-\mu))}\left(\sqrt{(x-\mu)'c\Sigma_0(x-\mu)}\right)\right) \qquad (36) \\
&= -\frac{1}{2}\log(\det(\Sigma_0)) - \log\left(\frac{\Gamma(d/2)}{2\pi^{d/2}}\right) + \log\left(\sqrt{(x-\mu)'c\Sigma_0(x-\mu)}\right) \\
&\quad -\frac{d}{2}\log(c) + \log\left(f_{R_{\alpha(s(x-\mu)}}\left(\sqrt{(x-\mu)'c\Sigma_0(x-\mu)}\right)\right). \qquad (37)
\end{aligned}
$$

Since the terms $-\frac{1}{2}\log(\det(\Sigma_0))$,$-\log\left(\frac{\Gamma(d/2)}{2\pi^{d/2}}\right)$ and $\log\left(\sqrt{(x-\mu)'c\Sigma_0(x-\mu)}\right)$ are constants with subject to parameter set Θ, it is equivalent to optimize equation (36) or (37) with respect to Θ. Of course, one should use the approach which leads to a simpler term in the optimization process.

4.2.4 Statistical Analysis

For the empirical analysis of the spectral estimator we generate samples from a multi-tail t distribution. In particular, we choose the location parameter μ to be $(0,0)'$ and the dispersion matrix Σ to be $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. The tail function $\alpha(.)$ satisfies

$$\alpha(s) = < s, F_1 >^2 \cdot 3 + < s, F_2 >^2 \cdot 6$$

where $F_1 = s((1,1)')$ and $F_2 = s((1,-1)')$ are the first two principal components of Σ. We assume knowledge of the location parameter and the dispersion matrix up to a scaling constant, i.e.,

$$\Sigma_0 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

Hence, we have to estimate the scaling parameter $c = 2$, the first tail parameter $\nu_1 = 3$, and the second tail parameter $\nu_2 = 6$. Since we know the density of muti-tail t distribution, we apply the Density Generator ML-Approach to estimate these parameters.

In Tables 3 to 5 and Figures 9 to 11 we observe that medians are close to the corresponding true values. The empirical distributions of $\hat{c}$, $\hat{\nu}_1$ and $\hat{\nu}_2$ are skewed to the left for small samples sizes and become more symmetric for large sample size per estimate. The accuracy of the estimator $\hat{c}$ is higher than the one of $\hat{\nu}_1$ and the one of $\hat{\nu}_1$ is higher than the one of $\hat{\nu}_2$ (compare the relative errors of Tables 9 to 11). It is not surprising that $\hat{c}$ performs better than $\hat{\nu}_1$ and $\hat{\nu}_2$ since the parameters ν_1 and ν_2 determine the tail behavior of the distribution and naturally we do not have many observations in the tails. In particular, the accuracy of $\hat{\nu}_1$ and especially $\hat{\nu}_2$ is very poor for small sample sizes per estimate, i.e. sample size < 1000. For reliable estimates we need at least a sample size of 3000 because of the relative error $e^{0.9}_{rel}$ of Tables 10 and 11.[16]

sample size	$q_{0.05}$	$q_{0.1}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.9}$	$q_{0.95}$	$e^{0.5}_{rel}$	$e^{0.9}_{rel}$
100	1.514	1.618	1.804	2.04	2.321	2.578	2.719	16%	36%
250	1.680	1.746	1.869	2.008	2.171	2.301	2.422	8.5%	21%
500	1.763	1.812	1.902	2.011	2.111	2.224	2.300	5.5%	15%
1000	1.838	1.871	1.933	2.004	2.076	2.146	2.194	3.8%	10%
3000	1.897	1.92	1.96	2.004	2.045	2.086	2.11	2.3%	5%
6000	1.927	1.942	1.970	2.002	2.032	2.057	2.073	1.6%	3.6%

Table 3: Quantiles of the estimator $\hat{c}$ for different sample sizes per estimate and the corresponding relative errors.

[16] $e^{1-2\alpha}_{rel}$ means the the relative error of the estimator is smaller than $e^{1-2\alpha}_{rel}$ with probability $1-2\alpha$ measured by the empirical distribution of the estimator. For a definition, see equation (33).

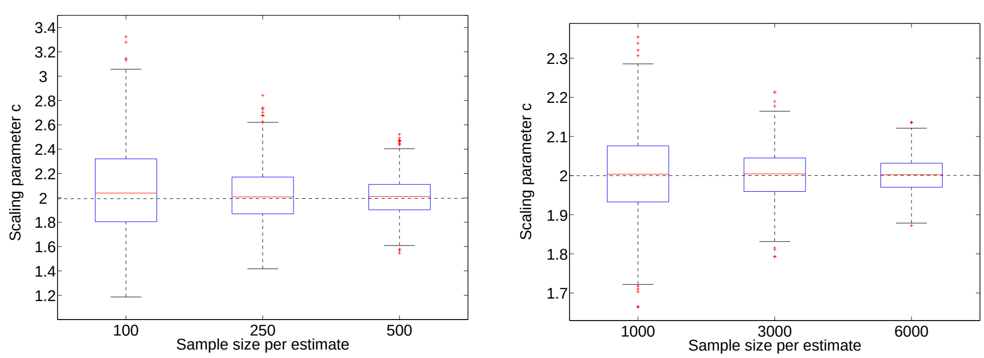
sample size	$q_{0.05}$	$q_{0.1}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.9}$	$q_{0.95}$	$e_{rel}^{0.5}$	$e_{rel}^{0.9}$
100	1.679	1.871	2.326	2.978	4.046	5.733	7.263	34%	142%
250	2.112	2.267	2.576	3.034	3.638	4.406	5.002	21%	67%
500	2.321	2.48	2.714	3.004	3.378	3.827	4.165	13%	39%
1000	2.482	2.588	2.789	3.01	3.279	3.56	3.725	9%	24%
3000	2.681	2.761	2.878	3.015	3.169	3.299	3.394	6%	13%
6000	2.772	2.822	2.896	2.999	3.093	3.195	3.267	3%	9%

Table 4: Quantiles of the estimator $\hat{\nu}_1$ for different sample sizes per estimate and the corresponding relative errors.

size	$q_{0.05}$	$q_{0.1}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.9}$	$q_{0.95}$	$e_{rel}^{0.5}$	$e_{rel}^{0.9}$
100	2.955	3.423	4.571	6.634	11.365	19.314	29.931	89%	399%
250	3.669	4.06	4.935	6.148	7.954	11.04	13.463	36%	124%
500	4.216	4.585	5.21	6.183	7.361	8.878	9.897	23%	64%
1000	4.479	4.7617	5.376	6.049	6.867	7.802	8.492	14%	41%
3000	5.086	5.279	5.61	6.069	6.514	6.971	7.222	9%	20%
6000	5.356	5.482	5.735	6.062	6.363	6.672	6.868	6%	14%

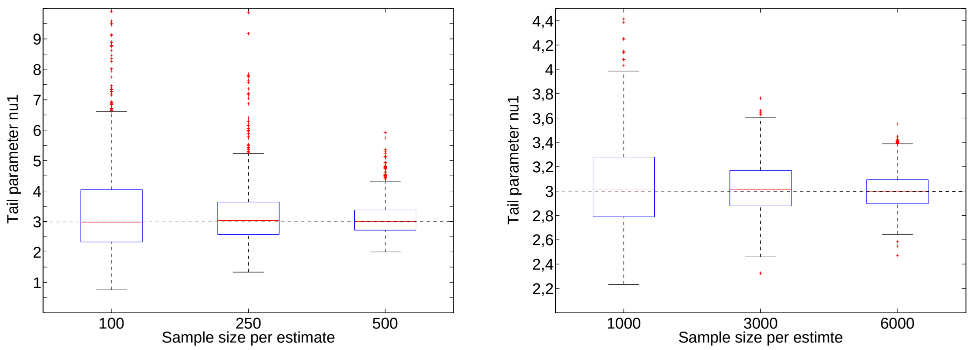
Table 5: Quantiles of the estimator $\hat{\nu}_2$ for different sample sizes per estimate and the corresponding relative errors.

In Figure 9 all estimates of c are depicted and this highlights again that the scale parameter can be estimated with high accuracy. In the case of ν_1 and sample size 100, 23 estimates ranging from 10 to 127 are not shown in Figure 10 and for sample size 250 only one estimate (value=16.09) is not depicted. Finally, in Figure 11 49 estimate are not illustrated ranging from 30.2 to 2794 for sample size 100 and for sample size 250 three estimates (values=$36.6, 33.7, 38.47$) are not depicted.



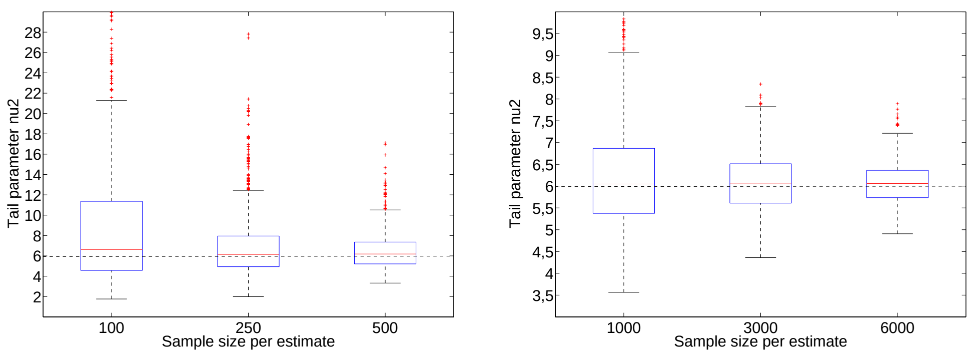
(a) (b)

Figure 9: Boxplots of the estimate of $\hat{c}$. Each boxplot consists of 1000 estimates.



(a) (b)

Figure 10: Boxplots of estimates of $\hat{\nu}_1$. Each boxplot consists of 1000 estimates.



(a) (b)

Figure 11: Boxplots of estimates of $\hat{\nu}_2$. Each boxplot consists of 1000 estimates.

5 Applications

For an empirical analysis of multi-tail elliptical distributions we investigated the daily logarithmic return series for the 29 German stocks included in the DAX index on March 31, 2006. The period covered is May 6, 2002 through March 31, 2006 (1,000 daily observations for each stock). In particular, the main focus of our analysis is to empirically assess whether a multi-tail model is superior to a classical elliptical one in the analysis of asset-return behavior.

5.1 Two-Dimensional Analysis

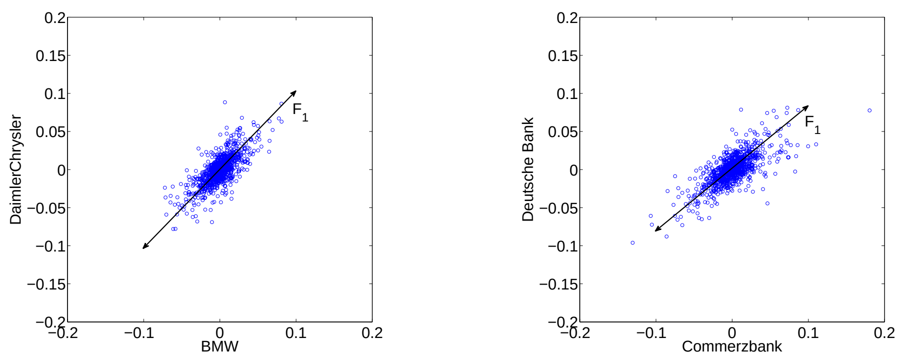
Figure 12 (a) depicts the two-dimensional scatterplots of BMW versus DaimlerChrysler and (b) the scatterplots of Commerzbank versus Deutsche Bank. In both figures we can see that there are more outliers in the directions around the first principal component F_1 motivating a multi-tail model. Applying the spectral estimator for both samples we obtain the normalized dispersion matrix ($\hat{\sigma}_{11} = 1$)

$$\hat{\Sigma}_0(X_1, ..., X_{1000}) = \begin{pmatrix} 1.000 & 0.762 \\ 0.762 & 1.204 \end{pmatrix}$$

for BMW versus DaimlerChrysler, and

$$\hat{\Sigma}_0(Y_1, ..., Y_{1000}) = \begin{pmatrix} 1.000 & 0.568 \\ 0.568 & 0.745 \end{pmatrix}$$

for Commerzbank versus Deutsche Bank representing the first step of the estimation procedure described in the previous section. Note that due to the properties of the spectral estimator, these normalized dispersion matrices are valid for the elliptical as well as for the multi-elliptical model.



(a) (b)

Figure 12: Bivariate scatterplot of (a) BMW versus DaimlerChrysler and (b) Commerzbank versus Deutsche Bank. Depicted are daily log-returns from May 6, 2002 through March 31, 2006.

For the second step we have to make a concrete distributional assumption for estimating the scale parameters and the tail parameters. In our analysis we choose the t- and multi-tail t-distribution. Note that the t-distribution has a constant tail function $\nu : \mathcal{S}^1 \to I = R^+$, $\nu(s) = \nu_0$, whereas for the multi-tail model we specify the tail function satisfying

$$\nu : \mathcal{S}^1 \to R^+, \nu(s) =< s, F_1 >^2 \nu_1 + < s, F_2 >^2 \nu_2,$$

where F_1 and F_2 are the first and second principal components of $\hat{\Sigma}_0$. Besides estimating the scale parameter c and the tail parameters ν_0, ν_1, and ν_2 in both models, we apply the Akaike information criterion and likelihood ratio test to identify the superior model.

Table 6 shows the estimates for the scale parameter and tail parameters in both models. In both models the scale parameters are the same while the tail parameters differ. This result is to be expected from our discussion in Section 4 because the scaling properties expressed by Σ_0 and c and the tail behavior captured by the tail parameters and the specified tail function are fairly independent for larger sample sizes.

	#par	$\hat{c}$	$\hat{\nu}_1$	$\hat{\nu}_2$	$\ln L$	AIC	p-value
t	6	$1.6 \cdot 10^{-4}$	3.7	3.7	5595.6	$-11,176.2$	-
multi-t	7	$1.6 \cdot 10^{-4}$	3.1	5.7	5598.3	-11,182.6	$< 2.5\%$

Table 6: Depicted are the likelihood estimates of the scale parameter and tail parameters for the BMW-DC returns in both models. The table shows the number of parameters (#par), the value of the log-likelihood at the maximum ($\ln L$), the value of the Akaike information criterion (AIC), and the p-value for a likelihood ratio test against the elliptical model. The period investigated is May 6, 2002 through March 31, 2006.

The Akaike information criterion[17] prefers the multi-tail model since we observe the smaller value in the multi-tail model. For the maximum likelihood ratio test,[18] we have $\Theta = \{(\nu_1, \nu_2) \in \mathbb{R}^+ : 0 < \nu_1 \leq \nu_2\}$ and the null hypothesis $H_0 : \theta \in \Theta_0 = \{(\nu_1, \nu_2) \in \mathbb{R}^+ : 0 < \nu_1 = \nu_2\}$ against the alternative $H_0 : \theta \in \Theta_0 = \{(\nu_1, \nu_2) \in \mathbb{R}^+ : 0 < \nu_1 < \nu_2\}$. According to Table 6, the p-value in this test is less than 2.5%, so it reasonable to reject the elliptical model.

Table 7 shows that we obtain basically the same results as in the previous case. The returns for Commerzbank and Deutsche Bank demand even more of a multi-tail model. The spread in the first and second tail parameters is larger than before. The difference between the log-likelihood values and Akaike information criterion is also greater and finally, the p-value of the maximum likelihood ratio test is practically equal to zero. Again, the scaling parameters are close and the tail parameters differ, indicating that the maximum likelihood estimator $\hat{c}$ for the scale parameter is fairly independent of the maximum likelihood estimates for ν_1 and ν_2.

	#par	$\hat{c}$	$\hat{\nu}_1$	$\hat{\nu}_2$	$\ln L$	AIC	p-value
t	6	$2.41 \cdot 10^{-4}$	3.4	3.4	5331.5	$-10,651$	-
multi-t	7	$2.39 \cdot 10^{-4}$	2.6	6.1	5337.8	-10,662	0

Table 7: Depicted are the maximum likelihood estimates of the scale parameter and tail parameters for the BMW-DC returns in both models. The table shows the number of parameters (#par), the value of the log-likelihood at the maximum ($\ln L$), the value of the Akaike information criterion (AIC), and the p-value for a likelihood ratio test against the elliptical model.

In particular, the results depicted in Tables 6 and 7 coincide with satterplots in Figure 12 (a) and (b) since in (b) we observe more pronounced outliers along the first principal component than in (a).

[17] In Akaike's approach we choose the model $M_1, ..., M_m$ minimizing

$$AIC(M_j) = -2\ln(L_j(\hat{\theta}_j; X) + 2k_j,$$

where $\hat{\theta}_j$ denotes the MLE of θ_j and k_j the number of parameters in model M_j.

[18] The likelihood ration test statistic satisfies

$$\lambda(X) = \frac{\sup_{\theta \in \Theta_0} L(\theta, X)}{\sup_{\theta \in \Theta} L(\theta, X)}.$$

Under the null hypothesis it can be shown that $-2\ln\lambda(X) \sim \chi^2_q$, where q is the difference between the free parameters in Θ and Θ_0.

5.2 Multi-Tail Elliptical Model Check for the DAX index

The investigated return data $X_1, X_2, ..., X_{1000} \in \mathbb{R}^{29}$ are the 29 German stocks[19] included in the DAX index. The period covered is May 6, 2002 to March 31, 2006. We start our analysis by estimating the normalized dispersion matrix $\hat{\Sigma}_0(X_1, ..., X_{1000})$ using the spectral estimator is depicted in Figure 13.

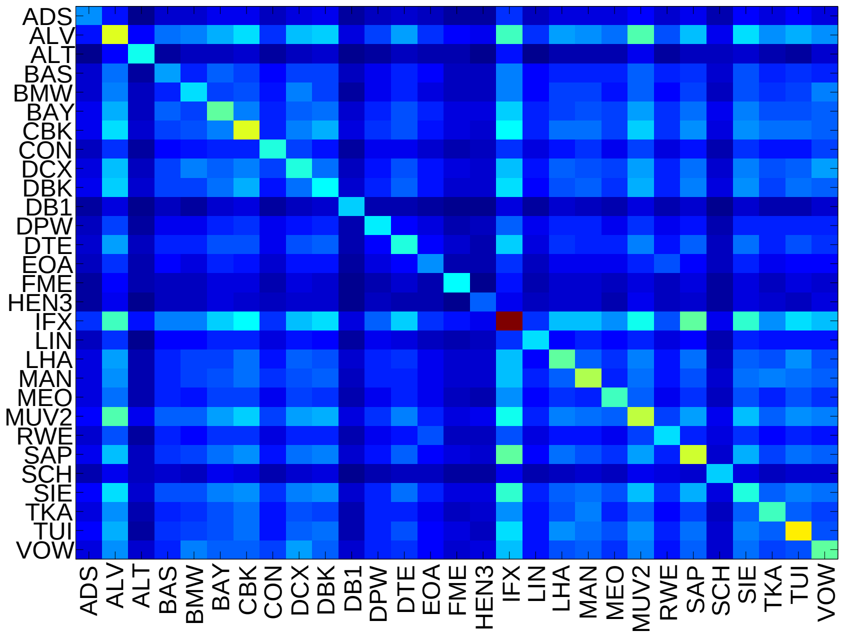


Figure 13: Heat map of the sample dispersion matrix estimated by the spectral estimator.

Figure 15 (a), (b), and (c) show the factor loadings (eigenvectors) g_1, g_2 and g_3 of the first three principal components. Figure 15 (d) depicts the eigenvalues of the normalized dispersion matrix obtained by the spectral estimator. We see that the eigenvalue of the first principal component is significantly larger than the others. The first vector of loadings is positively weighted for all stocks and can be thought of as describing a kind of index portfolio.

[19]We excluded HypoRealEstate Bank because we did not have sufficient return data for this stock for the period covered.

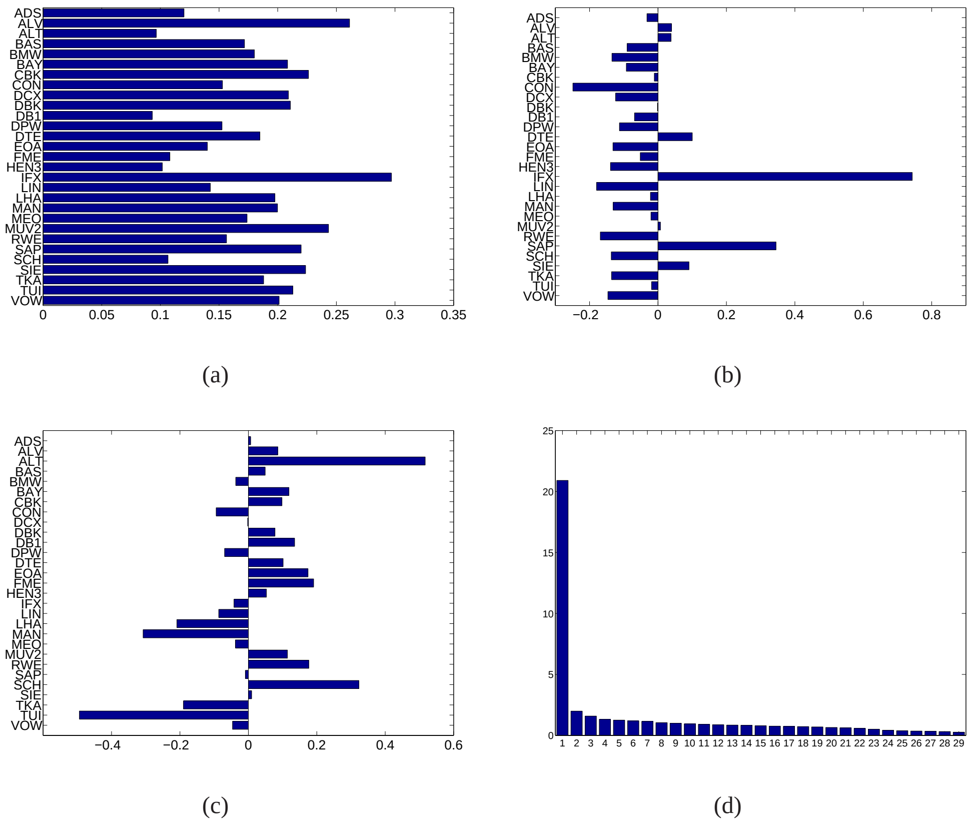
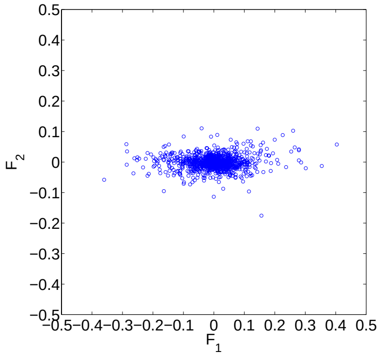


Figure 14: Barplot summarizing the loadings vectors g_1, g_2, g_3 and g_4 defining the first four principal components: (a) factor 1 loadings; (b) factor 2 loadings; (c) factor 3 loadings. (d) depicts the eigenvalues of the normalized dispersion matrix.

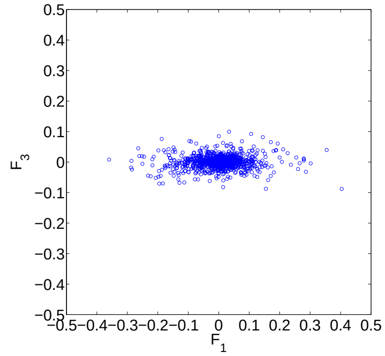
We obtain the corresponding time series of principal components through

$$F_{i,t} = g_i' X_t, t = 1, 2, ..., 1000.$$

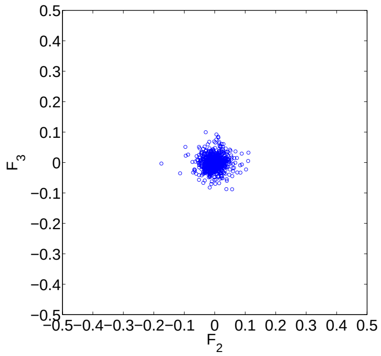
Figure 15 (a), (b), and (c) illustrate the pairwise scatterplots of these components. We can see in Figure 15 (a) and (b) that the scatterplots are stretched along the first principal component. This scaling behavior is caused by the large first eigenvalue of F_1. Moreover, it is important to note that we observe many outliers along F_1. This phenomenon may be attributed to smaller tail parameters in the directions around F_1 and $-F_1$. In Figure 15 (c) both principal components have fundamentally the same scale and the outliers are not so pronounced as in the former ones, suggesting a similar tail behavior.



(a)



(b)



(c)

Figure 15: Figure (a),(b), and (c) show the pairwise, two-dimensional scatterplots of the first three principal components. The period covered is May 6, 2002 through March 31, 2006.

Figure 16 depicts a three-dimensional scatterplot of the first three principal components. This figure confirms the scaling properties and tail behavior observed in Figure 15.

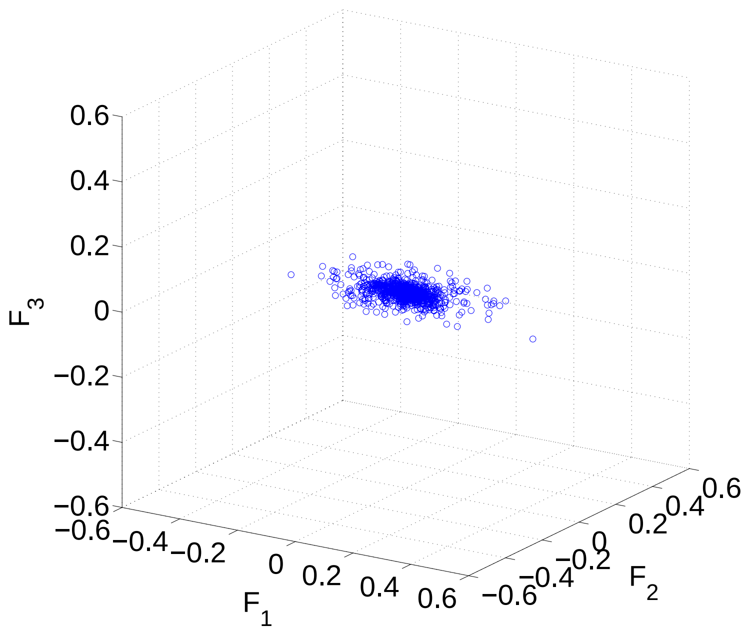


Figure 16: Figure depicts a three-dimensional scatterplot of the first three principal components.

The visual analysis of Figures 15 and 16 motivates a multi-tail model for the logarithmic returns investigated. Since the first principal component differ from the others, we propose the tail function

$$\nu : \mathcal{S}^{28} \to R^{+}, s \mapsto < s, F_1 >^2 \nu_1 + \sum_{i=2}^{29} < s, F_i >^2 \nu_2 .$$

As we did in Section 5.1, we compare this multi-tail model with an elliptical one which, per definition, has a constant tail function ($\nu(s) = \nu_0,\ s \in \mathcal{S}^{28}$). We conduct the same statistical analysis as in the previous section. In particular, we fit a t and multi-tail t distribution to the data. The results are reported in Table 8. The elliptical model has 29 location parameters, $29 \cdot 15 = 435$ dispersion parameters, and one tail parameter, while the multi-tail model has two tail parameters. Thus, we have 465 parameters in the elliptical and 466 in the multi-tail model. In the first step we estimate the dispersion parameter with the spectral estimator up to a scaling constant. The scale parameter $\hat{c}$ and tail parameters $\hat{\nu}_1$ and $\hat{\nu}_2$ are estimated in a second step.

	#par	c	ν_1	ν_2	$\ln L$	AIC	p-value
t	465	$1.45 \cdot 10^{-4}$	4	4	84,711.5	$-168,493$	-
multi-t	466	$1.44 \cdot 10^{-4}$	2.7	4.7	84,716.3	$-168,500.6$	$<< 0.05\%$

Table 8: Depicted are the maximum likelihood estimates of the scale parameter and tail parameters for the BMW-DC returns in both models. The table shows the number of parameters (#par), the value of the log-likelihood at the maximum ($\ln L$), the value of the Akaike information criterion (AIC), and the p-value for a likelihood ratio test against the elliptical model.

Table 8 shows that in both models the scale parameter $\hat{c}$ are almost the same, whereas the tail parameters differ significantly. The Akaike criterion as well as the likelihood ratio test do favor the multi-tail model. A p-value of less than 0.05% for this test indicates that we can reject the null hypothesis of an elliptical model at a very high confidence level.

5.3 Summary of the Results

In the statistical analysis reported in Sections 5.1 and 5.2 we see that the simplest genuine multi-tail elliptical model significantly outperforms the classical elliptical model. The hypothesis of homogeneous tail behavior can be rejected in any of the data sets investigated. In particular, we see in the analysis of the stocks included in the DAX index that the first principal component is heavier tailed than the others.

6 Conclusion

In this paper we introduce a new class of multivariate distributions that is flexible enough to capture a varying tail behavior of the underlying multivariate return data. We motivate this new type of distributions from typical behavior of financial markets. By introducing the notion of tail function we show how to capture varying tail behavior and present examples for tail functions. By applying the Akaike information criterion and likelihood ratio test, we find empirical evidence that a simple multi-tail elliptical model significantly outperforms common elliptical models. Moreover, the hypothesis of homogeneous tail behavior must be rejected.

7 Appendix

Definition 7.1. The *gamma function* Γ is defined by

$$\Gamma(x) = \int_0^\infty r^{x-1} e^{-r} dr,$$

where $x > 0$.

Definition 7.2. Let $A \in \mathbb{R}^{d\times d}$ be a matrix and $t \in \mathbb{R}$. Then t^A is defined by

$$t^A = I + \sum_{m=1}^{\infty} \frac{(\log(t)A)^m}{m!}. \tag{38}$$

It can be shown that the limit in equation (38) exists for all matrices $A \in \mathbb{R}^{d\times d}$. The matrix A in t^A is called the exponent of t. Furthermore, we have the following properties.

Proposition 7.1. Let $t \in \mathbb{R}$ and $A, B \in \mathbb{R}^{d\times d}$. Then we have

(i) $t^A t^B = t^{A+B}$;

(ii) $t^A t^{-A} = \mathrm{Id}$;

(iii) $(t^A)' = t^{A'}$.

The proof can be found in Rubin (1986).

Theorem 7.1 (Change-of-variables theorem). *Suppose that*

(i) $X \subset V \subset \mathbb{R}^d$, *$V$ is open,* $T : V \to \mathbb{R}^d$ *is continuous;*

(ii) X *is Lebesque measurable,* T *is one-to-one on* X*, and* T *is differentiable at every point of* X*;*

(iv) $\lambda(T(V \cap X^c)) = 0$.

Then, setting $Y = T(X)$,

$$\int_Y f(y)dy = \int_X f(T(x))|\det((DT)(x))|dx$$

for every measurable $f : \mathbb{R}^d \to [0, \infty]$.

For proof of this theorem we refer to Rubin (1986).

Lemma 7.1. *Let the functions* g, g_1 *and* g_2 *be defined by*

$$\begin{aligned} g &: \mathbb{R}_+ \times \mathcal{S}^{d-1} \to \mathbb{R}^d, (r, s) \mapsto \mu + rAs, \\ g_1 &: \mathbb{R}_+ \times \mathcal{S}^{d-1} \to \mathbb{R}^d, (r, s) \mapsto rs, \\ g_2 &: \mathbb{R}^d \to \mathbb{R}^d, x \mapsto \mu + Ax, \end{aligned}$$

where $A \in \mathbb{R}^{d\times d}$ *is regular and* $\mu \in \mathbb{R}^d$*. Then we have* *(i)* $g = g_2 \circ g_1$

(ii) $\det((Dg_1)(r,s)) = r^{d-1}$

(iii) $D(g_2)(y) = A$

(iv) $|\det(Dg_2)(y)| = |\det(A)| = |\det(\Sigma)^{1/2}|$, *where* $\Sigma = AA'$.

Proof. (i) This is obvious.

(ii) Sketch of the proof: We have to parameterize g_1 in terms of polar coordinates leading to a function

$$\tilde{g}_1 : \mathbb{R}_+ \times I \subset \mathbb{R}^{d-1} \to \mathbb{R}^d, (r, \phi_1, ..., \phi_{d-1}) \mapsto r\psi(\phi_1, ..., \phi_{d-1})$$

Calculating $\det(D\tilde{g}_1(r, \phi_1, ..., \phi_d))$ leads to r^{d-1}. For a detailed proof, see Fang, Kotz and Ng (1987).

(iii) The Jacobian of a linear function $g(x) = Ax + \mu$ is A. For example, see Rubin (1986).

(iv) We have

$$\begin{aligned} |\det(A)| &= (\det(A)\det(A'))^{1/2} \\ &= \det(AA')^{1/2} \\ &= \det(\Sigma)^{1/2}. \end{aligned}$$

□

Lemma 7.2. *Let* $V, W \subset \mathbb{R}^d$ *be open and* $f : V \to W$ *differentiable with inverse* f^{-1}. *Then we have*

$$\det((Df)(x)) = \frac{1}{\det((Df^{-1})(f(x)))}.$$

Proof. For all $x \in V$ we have

$$(D(f^{-1} \circ f)(x)) = Df^{-1}(f(x))D(f(x)).$$

The lemma follows now immediately because of

$$\begin{aligned} \det(Df^{-1}(f(x))D(f(x))) &= \det((D(f^{-1} \circ f)(x))) \\ &= \det(Id) \\ &= 1. \end{aligned}$$

□

Corollary 7.1. *Let* g, g_1 *and* g_2 *as in Lemma 7.1. Then we have*

$$\det(Dg^{-1}(x)) = \frac{((x-\mu)'\Sigma^{-1}(x-\mu))^{(-d+1)/2}}{\det(\Sigma)^{1/2}}.$$

Proof. The inverse of g satisfies

$$g^{-1}: \mathbb{R}^d \to \mathbb{R}_{>0} \times \mathcal{S}^{d-1}, x \mapsto \left(\sqrt{(x-\mu)'\Sigma^{-1}(x-\mu)}, \frac{A^{-1}(x-\mu)}{||A^{-1}(x-\mu)||} \right) \tag{39}$$

We have

$$\begin{aligned}
\det(D(g_1 \circ g_2)^{-1}(x)) &= \det(D(g_1^{-1} \circ g_2^{-1})(x)) \\
&= \det(Dg_1^{-1}(g_2^{-1}(x))Dg_2^{-1}(x)) \\
&= \det(Dg_1^{-1}(g_2^{-1}(x)))\det(Dg_2^{-1}(x)) \\
&\overset{\text{Lemma 7.2}}{=} \frac{1}{\det(Dg_1(g_1^{-1} \circ g_2^{-1}(x)))} \frac{1}{\det(Dg_2(g_2^{-1}(x)))} \\
&= \frac{1}{\det(Dg_1(g^{-1}(x)))} \frac{1}{\det(A)} \\
&\overset{(1)}{=} \frac{\left(\sqrt{(x-\mu)'\Sigma^{-1}(x-\mu)}^{d-1}\right)^{-1}}{\det(\Sigma)^{1/2}}.
\end{aligned} \tag{40}$$

(1) holds in equation (40), since $Dg_1((r,s)) = r^{d-1}$ and equation (39). ☐

Theorem 7.2. *Let $M \subset \mathbb{R}^{1\times d^2}$ be the open subset of d-dimensional regular matrices. Then the function $h: M \to \mathbb{R}$ $A \mapsto \det(A)$ has the Jacobian*

$$Dh(A) = \det(A)A^{-1} \tag{41}$$

Proof. It can be shown for all $i, j = 1, ..., n$ that

$$\frac{\partial \det(A)}{\partial e_{ij}} = (-1)^{i+j}\det(A_{ij}) \tag{42}$$

where A_{ij} is the matrix one obtains if the ith row and jth column are canceled from A. Hence we obtain by Cramer's rule

$$Dh(A) = \det(A)(A^{-1})'.$$

☐

References

Belkacem, L., J. Vehel and C. Walter 2000. *CAPM, risk and portfolio selection in α-stable markets.* Fractals 8 (1), 99-115.

Beutelspacher, A. 2003. *Lineare Algebra*. Vieweg

Eberlein, E. and U. Keller. 1995. *Hyperbolic distribution in finance*. Bernoulli, 281-299.

Eberlein, E., U. Keller and K. Prause. 1998. *New insights into smile, mispricing, and value at risk: the hyperbolic model.* Journal of Business 38:371-405.

Fama, E. 1965. *The Behavior of Stock Market Prices*. Journal of Business, 38, 34-105.

Fama, E. 1965. *Portfolio Analysis in a Stable Paretian market*. Management Science, 11, 404-419.

Fang, K.-T., S. Kotz and K.-W. Ng. 1987. *Symmetric Multivariate and Related Distributions.* London: Chapman & Hall.

Feller, W. 1971. *An Introduction to Probability Theory and Its Applications.* John Wiley & Sons, Inc.

Frahm, G. 2004. *Generalized elliptical distributions: Theory and Applications*. Ph.D. thesis, University of Cologne, Germany.

Hamilton, J.D. 1994 *Time Series Analysis.* Princetion University Press 1994.

Heuser, H. 1991. *Lehrbuch der Analysis, Teil 2.* B. G. Teubner

Jansen, D., C. de Vries 1991. *On the frquency of large stock market returns: Putting booms and busts into perspective.* Review of Economical Statistics 23, 18-24.

Johnson, R.A., D.W. Wichern 1983. *Applied Multivariate Statistical Analysis.* Prentice-Hall, Inc.

Jurek, Z. and J.D. Mason 1993. *Operator-Limit Distributions in Probability Theory.* Wiley, New York.

Kent, J.T. and D.E. Tyler 1988. Maximum likelihood estimation for the wrapped Cauchy distribution. *Journal of Applied Statistics* 15: pp. 247-254.

Kotz, S. and S. Nadarajah. 2004. *Multivariate t Distributions and Their Applications.* Cambridge University Press.

Kring, S., S.T. Rachev, M. Höchstötter, F.J. Fabozzi 2007. *Estimation of α-Stable Sub-Gaussian Distributions for Asset Returns.* Risk Assessment: Decisions in Banking and Finance (eds. G. Bol, S.T. Rachev, and R. Wrth), Physica Verlag Heidelberg 2007.

Loretan, M., P. Phillips 1994. *Testing the covariance stationarity of heavy tailed time series.* Journal of Empirical Finance 1, 211-248.

Mandelbrot, B. B. 1963b. *The Variation of Certain Speculative Prices*. Journal of Business, 36, 394-419.

McCulloch, J. H. 1952. *Financial Applications of Stable Distributions*. Hanbbook of Statistics-Statistical Methods in Finance, 14, 393-425. Elsevier Science B.V, Amsterdam.

McNeil, A. J., R. Frey and P. Embrechts. 2005. *Quantitative Risk Management,* Princeton University Press.

Meerschaert, M.M. and H.-P. Scheffler 2003. *Portfolio modeling with heavy tailed random vectors.* In: Handbook of Heavy Tailed Distributions in Finance, Editor: S.T. Rachev Elsevier/North-Holland, pp. 595-640.

Meerschaert, M.M. and H.-P. Scheffler 2001. *Limit distributions for sums of independent random vectors: Heavy tails in theory and practice.* Wiley, New York.

Mittnik, S. and S.T. Rachev 1993. *Reply to comments on Modeling asset returns with alternative stable distributions and some extensions.* Econometric Reviews 12, 347-389.

Nolan, J. P., A.K. Panorska and J.H. McCulloch 1996. *Estimation of stable spectral measure.* Mathematical and Computer Modelling 34, 1113-1122.

Rachev, S. T. and S. Mittnik 2000. *Stable Paretian Models in Finance*. Wiley.

Rachev, S. T. and S. Han 2000. *Portfolio management with stable distributions.* Mathematical Methods of Operations Research 51 (2), 341-352.

Resnick, S.I. 1987. *Extreme values, regular variation, and point processes*. Springer.

Rubin, W. 1986. *Real and Complex Analysis.*, McGraw-Hill international editions.

Samorodnitsky, G. and M. Taqqu. 1994. *Stable Non-Gaussian Random Processes,* Chapmann & Hall

Schmidt, R. 2002. *Tail dependence for elliptically contoured distributions*. Mathematical Methods of Operations Research 55: pp. 301-327.

Sharpe, M. 1969. *Operator-stable probability distributions on vector groups.* Translation of the American Mathematical Society 136,51-65.

Stoyanov, S., V. 2005. *Optimal Portfolio Management in Highly Volatile Markets.* Ph.D. thesis, University of Karlsruhe, Germany.

Stoyanov, S., V., B. Racheva-Iotova 2004. *Univariate stable laws in the fields of finance-approximations of density and distribution functions*. Journal of Concrete and Applicable Mathematics, 2/1, 38-57.

Tyler, D.E. 1987a. *A distribution-free M-estimator of multivariate scatter*. The Annals of Statistics 15: pp. 234-251.

Tyler, D.E. 1987b. *Statistical analysis for the angular central Gaussian distribution on the sphere*. Biometrika 74: pp. 579-589.