

Generalized Elliptical Distributions: Theory and Applications

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2004

vorgelegt
von

Dipl.-Kfm. Gabriel Frahm

aus

Naharia/Israel

Referent: Prof. Dr. F. Schmid
Korreferent: Prof. Dr. K. Mosler
Tag der Promotion: 2. Juli 2004

To my mother

*Mai am un singur dor :
În liniștea serii
Să mă lăsați să mor
La marginea mării;
Să-mi fie somnul lin
Și codrul aproape,
Pe-ntinsele ape
Să am un cer senin.
Nu-mi trebuie flamuri,
Nu voi sicriu bogat,
Ci-mi împlețiți un pat
Din tinere ramuri.*

*Și nime-n urma mea
Nu-mi plângă la creștet,
Doar toamna glas să dea
Frunzisului veșted.
Pe când cu zgomot cad
Izvoarele-ntr-una.
Alunece luna
Prin vârfuri lungi de brad.
Pătrunză talanga
Al serii rece vânt,
Deasupra-mi teiul sfânt
Să-și scuture creanga.*

*M-or troieni cu drag
Aduceri aminte.
Luceferi, ce răsar
Din umbră de cetini,
Fiindu-mi prieteni,
O să-mi zâmbescă iar.
Va geme de patemi
Al mării aspru cânt...
Ci eu voi fi pământ
In singurătate-mi.*

Mihai Eminescu (1850-1889)

Preface

In 1999 I was searching for an appropriate topic regarding my diploma thesis. My supervisor professor Friedrich Schmid made the proposal to focus on financial risk management applying univariate extreme value theory. That was really a pleasure such that I had an itch to work further on possible applications of extreme value theory in the multivariate context. More than 2 years after the diploma my wish came true and I have got an appointment at the Center of Advanced European Studies and Research in Bonn. I would like to thank Dr. Angelika May very much for the possibility to combine such an exciting work with a doctoral thesis.

Of course, the reason for my fascination was not only caused by the subject matter but also by the kind support of professor Friedrich Schmid. I am very grateful that he took on an ‘external’ Ph.D. student and I would like to thank him once again for supervising me. In the same manner I would like to thank professor Karl Mosler for his many constructive suggestions and nice talks.

I am indebted to the members of caesar’s financial engineering group. With the brilliant help of my colleague Dr. Markus Junker I had the chance to acquire important knowledge on complex dependence structures in the twinkling of an eye. But particularly, I learned that academic discussions indeed may be lively and funny. Also I would like to thank Stefan Hartmann for his painstaking reviews of my manuscripts and his endless patience when listening to open problems. I am also happy to collaborate with Annett Keller. In many useful discussions she showed me to see things from a different angle.

Without the suggestions of Dr. Christoph Memmel (Deutsche Bundesbank) and Dr. Uwe Jaekel (C&C Research Laboratories, NEC Europe Ltd.) the practical part of this thesis would have never been accomplished. Much of the material treated in the financial applications chapter is due to the delightful discussions with Christoph. I would like to thank also very much to Uwe who brought me into the world of mathematical physics. To carry out research with him is a pleasure and the chapter about random matrix theory is due to a joint work.

Many thanks belong to professor Robert Israel and professor Herman Rubin who kindly supported me with answers to important questions. I am also thankful to Marco Kriesche from Thomson Financial Datastream who breathed life into the practical part of this thesis by kindly providing sufficiently many stock market data.

During the seminar ‘Stochastic modelling and statistics in finance’ in Oberwolfach, 2003, I experienced how wonderful mathematics can be. Particularly, I refer to the nocturnal jam sessions in the piano room with Stefan Ankirchner and Hilmar Hauer who are exceptionally gifted jazz musicians. That was real fun.

Last but not least I thank my wonderful wife Franziska and my children Ilian and Jana. Franziska, you are the ‘driving factor’ in my life. Once again you successfully got over the time of my mental (and physical) absence.

Bonn, 5th November, 2004

Contents

Preface	i
Introduction	v
I Theory	1
1 Elliptically Symmetric Distributions	3
1.1 Definition and Characterization	3
1.2 Basic Properties	7
1.2.1 Density Functions	7
1.2.2 Symmetry	10
1.2.3 Moments	11
1.2.4 Affine Transformations and Marginal Distributions	13
1.2.5 Conditional Distributions	13
1.3 Additional Properties	15
1.3.1 Summation Stability	15
1.3.2 Infinite Divisibility	16
1.3.3 Self-decomposability	17
2 Extreme Values and Dependence Structures	19
2.1 Univariate Extremes	19
2.2 Multivariate Extremes and Copulas	22
2.3 Asymptotic Dependence of Meta-elliptical Distributions	27
2.3.1 Bivariate Asymptotic Dependence	27
2.3.2 Multivariate Asymptotic Dependence	31
2.4 Covariance Matrix Estimation in the Presence of Extreme Values	38
3 Generalized Elliptical Distributions	43
3.1 Motivation	43
3.2 Definition	46
3.3 Basic Properties	48
3.4 Models	51

4	Robust Estimation	57
4.1	Basics of M-estimation	57
4.2	Dispersion Matrix Estimation	58
4.2.1	Spectral Density Approach	58
4.2.2	Fixed-point Representation	63
4.2.3	Existence and Uniqueness	64
4.3	Location Vector Estimation	66
5	Statistical Properties of the Spectral Estimator	69
5.1	Information Matrix	69
5.2	Consistency and Asymptotic Efficiency	73
5.3	Asymptotic Covariance Matrix	75
II	Applications	79
6	Motivation	81
6.1	Empirical Evidence of Extremes	81
6.2	On the Curse of Dimensions	83
7	Applications in Finance	87
7.1	Modern Portfolio Theory	87
7.1.1	Portfolio Optimization	87
7.1.2	Portfolio Weights Estimation	89
7.2	Principal Component Analysis	93
8	Random Matrix Theory	99
8.1	Limiting Distributions of Eigenvalues	99
8.1.1	Wigner's Semi-circle Law	99
8.1.2	The Marčenko-Pastur Law	101
8.2	Separation of Signal and Noise	108
8.3	Application to Econophysics	110
	Summary	115
	List of Abbreviations	117
	List of Symbols	119
	Bibliography	125

Introduction

Motivation

A natural generalization of the multivariate normal (or ‘Gaussian’) distribution function is given by the broad class of elliptical distributions. These were introduced by Kelker (1970) and well investigated by Cambanis, Huang, and Simons (1981) and by Fang, Kotz, and Ng (1990). Every d -dimensional elliptical random vector X can be represented by $X =_{\text{d}} \mu + \mathcal{R}\Lambda U^{(k)}$, where $\mu \in \mathbb{R}^d$, $\Lambda \in \mathbb{R}^{d \times k}$, $U^{(k)}$ is a k -dimensional random vector uniformly distributed on the unit hypersphere, and \mathcal{R} is a nonnegative random variable independent of $U^{(k)}$. The distribution function of \mathcal{R} constitutes the particular elliptical distribution family of X and is called the ‘generating distribution function’. Suppose that the generating variate \mathcal{R} belongs to the maximum domain of attraction of the Fréchet distribution (Embrechts, Klüppelberg, and Mikosch, 2003, Section 3.3.1), i.e. $\overline{F}_{\mathcal{R}} = \lambda(x) \cdot x^{-\alpha}$ for all $x > 0$, where $\alpha > 0$ and λ is a slowly varying function (Resnick, 1987, p. 13). The parameter α is called the ‘tail index’ of the generating distribution function $F_{\mathcal{R}}$ which corresponds also to the tail index of the regularly varying random vector X (Hult and Lindskog, 2002). Hence the class of multivariate elliptical distributions allows for heavy tails though it remains the simple linear dependence structure known from the normal distribution family. In addition to the normal distribution function many other well-known and widely used multivariate distribution functions are elliptical too, e.g. the t -distribution (Fang, Kotz, and Ng, 1990, p. 32), the symmetric generalized hyperbolic distribution (Barndorff-Nielsen, Kent, and Sørensen, 1982), the sub-Gaussian α -stable distribution (Rachev and Mittnik, 2000, p. 437).

Elliptical distributions inherit a lot of nice Gaussian properties. This is because the characteristic function of the multivariate centered normal distribution, i.e. $t \mapsto \exp(-1/2 \cdot t' \Sigma t)$ is simply weakened to $t \mapsto \varphi(t' \Sigma t)$. Here $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$ (called the ‘characteristic generator’) is an arbitrary function only guaranteeing that $t \mapsto \varphi(t' \Sigma t)$ is a characteristic function. Any affinely transformed elliptical random vector is also elliptical. Furthermore, any marginal distribution function of an elliptical random vector is elliptically contoured, too. This holds even for the conditional distribution functions (Kelker, 1970). Moreover, the density function of an elliptical distribution can be simply derived from the density function of \mathcal{R} provided it is absolutely continuous.

From a practical point of view elliptical distributions are attractive in particular for the modeling of financial data. The theory of portfolio optimization developed by Markowitz (1952) and continued by Tobin (1958), Sharpe (1963, 1964) and Lintner (1965) is the basis of modern portfolio risk management. It relies on the Gaussian distribution hypothesis and its quintessence is that the portfolio diversification effect depends essentially on the covariance matrix, i.e. the linear dependence structure of the portfolio components. Generally, this information is not sufficient for elliptically contoured distributions (Embrechts, McNeil, and Straumann, 2002). The risk of extreme simultaneous losses, i.e. the ‘asymptotic dependence’ is not only determined by the correlation coefficient but also by the tail index of the multivariate elliptical distribution (Schmidt, 2002). Asymptotic dependence usually is quantified by the tail dependence coefficient (Joe, 1993). Loosely speaking, this is the

probability that the realization of a random variable is extremely negative (or positive) under the condition that the realization of another random variable is extremely negative (or positive), too. If an elliptical random vector is regularly varying, i.e. if the generating distribution function belongs to the maximum domain of attraction of the Fréchet distribution then the tail dependence coefficient of each bivariate marginal distribution is positive, provided that the linear dependence of the two random components is not perfectly negative. To seek a contrast, the generating distribution function of the multivariate normal distribution belongs to the maximum domain of attraction of the Gumbel distribution (Embrechts, Klüppelberg, and Mikosch, 2003, Section 3.3.3), i.e. the Gaussian distribution is not heavy tailed and the tail dependence coefficient of its bivariate marginal distributions corresponds to zero. Many authors show that the Gaussian distribution hypothesis cannot be justified for financial data, see Eberlein and Keller (1995), Fama (1965), and Mandelbrot (1963) concerning univariate financial time series, and Frahm, Junker, and Szimayer (2003) as well as Junker and May (2002) regarding the dependence structure of multivariate time series. Hence elliptical distributions are an acceptable alternative retaining the workability of the normal distribution, for the most part.

The covariance matrix of an elliptically distributed random vector X corresponds to the dispersion matrix $\Sigma := \Lambda\Lambda'$ up to a scaling constant, i.e. $\text{Var}(X) = E(\mathcal{R}^2)/k \cdot \Sigma$ provided the second moment of \mathcal{R} is finite (Cambanis, Huang, and Simons, 1981). But estimating the covariance matrix of elliptical random vectors via the method of moments, especially the correlation matrix by Pearson's correlation coefficient is dangerous when the underlying distribution is not normal (Lindskog, 2000). This is because Pearson's correlation coefficient is very sensitive to outliers and the smaller the distribution's tail index, i.e. the heavier the tails the larger the estimator's variance. Indeed, there are a lot of robust techniques to insulate from the 'bad influence' of outliers (see, e.g., Huber, 1981 and Visuri, 2001, pp. 31-51). But there may be 'bad' and 'good' outliers. Bad outliers are caused by sampling errors due to the measurement process whereas good outliers are data caused by true extremal events. The simplest approach is to eliminate every outlier and to apply the sample covariance matrix on the residual data. But from the viewpoint of extreme value theory this has the annoying effect of neglecting useful information contained in extremal realizations. In particular, estimating the tail index is impossible without outliers.

In this work the class of elliptical distributions is generalized to allow for asymmetry. All the ordinary components of elliptical distributions, i.e. the generating variate \mathcal{R} , the location vector μ and the dispersion matrix Σ remain for this new class of 'generalized elliptical distributions'. It is shown that the class of generalized elliptical distributions contains the class of skew-elliptical distributions (Branco and Dey, 2001). The basic properties of generalized elliptical distributions are derived and compared with those of elliptical distributions. The second aim of the thesis is to develop a robust estimator for the dispersion matrix Σ yet recognizing all the available data. This is called the 'spectral estimator'. It is shown that the spectral estimator is an ML-estimator. Nevertheless it is robust within the class of generalized elliptical distributions since it requires only the assumption that the generating variate has no atom at 0. Hence it is not disturbed neither by asymmetries nor by outliers and all the available data points can be used for estimation purposes. Given the estimates of location and dispersion the empirical generating distribution function can be extracted preserving the outliers. This can be used for tail index estimation regarding \mathcal{R} , for instance. Further, it is shown that the spectral estimator corresponds to the M-estimator for elliptical distributions developed by Tyler (1983, 1987a). In contrast to the more general M-approach used by Tyler (1987a) the spectral estimator can be derived on the basis of maximum-likelihood theory (Tyler, 1987b). Hence, desired properties like, e.g., asymptotic normality, consistency, and asymptotic efficiency follow in a straightforward manner.

A further goal of this thesis is to discuss the impact of high-dimensional (financial) data on statistical inference. Statistical theory usually presumes a constant number of dimensions or at least $n/d \rightarrow \infty$. The quantity $q := n/d$ can be interpreted as 'average sample size per

dimension' or as 'effective sample size'. Unfortunately, large sample properties of covariance matrix estimates which are based on the central limit theorem fail if q is small even if n is large. There is a branch of statistical physics called 'random matrix theory' dealing with this case of 'high-dimensional data'. Random matrix theory is mainly concerned with the distribution of eigenvalues of randomly generated matrices. An important result is that if one assumes independent and identically distributed matrix elements the distribution of the eigenvalues converges to a specified law which does not depend on the distribution of the matrix elements but primarily on q . Since the sample covariance matrix is a random matrix the results of random matrix theory can be applied in the case of normally distributed data. For data which is not normally but generalized elliptically distributed the results of random matrix theory are no longer applicable if one uses the sample covariance matrix. But it is shown that this vacancy can be filled easily by using the spectral estimator instead.

Possible applications are discussed in the context of modern portfolio theory and principal component analysis. More precisely, the spectral estimator can be used for portfolio optimization to obtain robust portfolio weights estimates. Further, it is shown how the 'driving' risk factors of stock prices can be identified, robustly. This depends essentially on the accuracy of the estimates of eigenvectors and eigenvalues of the dispersion matrix which belongs to the field of random matrix theory mentioned above. Therefore, some classical results of random matrix theory are given and it is shown how generalized elliptical distributions, random matrices, and the spectral estimator are related to each other.

Structure of the Thesis

The thesis is divided into two parts, a theoretical part ('Theory') and a practical part ('Applications'). The theoretical part begins with the traditional class of elliptically symmetric distributions. Apart from the definition and characterization of elliptical distributions their basic properties will be derived. The corresponding theorems (and their proofs) have a strong relationship to the theory of generalized elliptical distributions treated in Chapter 3.

The second chapter is about extreme value theory. Classical results from univariate extreme value theory as well as relatively new insights from multivariate extreme value theory are examined. This involves the theory of 'copulas'. Copulas are extremely useful for the analysis of complex dependence structures. They can be used also to describe the concept of asymptotic dependence. This will be done with a special emphasis on 'meta-elliptical' distributions which are discussed in Chapter 2. The chapter completes with some conclusions concerning covariance matrix estimation drawn from the consideration of extreme values.

In the third chapter the class of generalized elliptical distributions is introduced. This is motivated by empirical findings on financial markets. Particularly, we aim at robust covariance matrix estimation under the stylized facts of asymmetry and heavy tails. Further, the basic properties of generalized elliptical distributions are derived and compared with those of elliptically symmetric distributions. The chapter closes with the modeling of generalized elliptical distributions.

The fourth chapter focuses on the robust estimation of the dispersion matrix and the location vector of generalized elliptical distributions. The 'spectral density' of a multivariate normally distributed random vector projected to the unit hypersphere is derived and subsequently used for constructing a completely robust covariance matrix estimator for generalized elliptical distributions, namely the spectral estimator. Since the spectral estimator emerges as an M-estimator some basics of the M-estimation approach are presented and the corresponding fixed-point solution for the spectral estimator is derived. Also its positive definiteness, existency and uniqueness will be discussed. Furthermore, it is shown that the componentwise sample median is an appropriate estimator for the location vector in the context of angularly symmetric generalized elliptical distributions.

The last chapter of the first part concentrates on the statistical properties of the spectral estimator. Since the spectral estimator is not only an M-estimator but also an ML-estimator standard methods of maximum-likelihood theory are applied to derive its Fisher information. Furthermore, its consistency, asymptotic efficiency and normality are proved. At last the asymptotic covariance matrix of the spectral estimator in the case of $\Sigma = \sigma^2 I_d$ is derived in a closed form and compared with the asymptotic covariance matrix of the sample covariance matrix.

The second part of the thesis begins with some stylized facts of empirical finance. The results of the spectral estimator are demonstrated on an S&P 500 data set consisting of the current 500 stocks and ranging from 1980-01-02 to 2003-11-26. Since financial markets are characterized by a large number of risk factors the typical difficulties occurring with high-dimensional data sets are discussed. Some examples are constructed to show that the central limit theorem lose its effect if the effective sample size q is small even if n is very large.

Chapter 7 deals with applications in finance. The main results of modern portfolio theory are derived with an emphasis on portfolio optimization. It is shown how the key figures of portfolio risk management, namely the asset's 'Betas' can be estimated, robustly. This is explained in terms of principal component analysis.

The last chapter of the second part can be interpreted as a brief introduction to random matrix theory. Starting from Wigner's semi-circle law for symmetric random matrices we turn to a similar result for random projection matrices known as the Marčenko-Pastur law. The relationships between the Marčenko-Pastur law, the generating variate, and the spectral estimator are pointed out. It is shown how the Marčenko-Pastur law can be used for separating 'signal' from 'noise', i.e. detecting the main principal components or the 'driving risk factors' of financial markets. The spectral estimator emerges as a robust alternative to the sample covariance matrix not only in the case of $n/d \rightarrow \infty$ but also for $n/d \rightarrow q < \infty$, i.e. in the context of high-dimensional data.

Mathematical Notation and Abbreviations

Throughout the thesis I will deal only with real (random) scalars, vectors, and matrices unless otherwise noted. Vectors are supposed to be columns. Zero scalars, zero vectors as well as zero matrices are denoted by 0 whenever the dimension is clear. The d -dimensional identity matrix is always represented by I_d ($I_1 \equiv 1$). If x is a scalar then $|x|$ is its absolute value. If A is a set then $|A|$ denotes its cardinality. $\|\cdot\|$ is an arbitrary vector norm on \mathbb{R}^d whereas $\|\cdot\|_2$ denotes the Euclidean norm. If A is a matrix and $x \in \mathbb{R} \setminus \{0\}$ then A/x is defined as $x^{-1}A$. The transpose of a matrix A is denoted by A' . The inverse A^{-1} of a rectangular matrix A generally corresponds to the Moore-Penrose inverse (the 'pseudo-inverse') which is defined as (see, e.g., Schönfeld, 1971, p. 294)

$$A^{-1} := (A'A)^{-1} A',$$

where

$$(A'A)^{-1} := \mathcal{O}\mathcal{D}^{-1}\mathcal{O}'.$$

Here $\mathcal{O}\mathcal{D}\mathcal{O}'$ is the spectral decomposition of $A'A$, i.e. \mathcal{O} is an orthonormal square matrix and \mathcal{D} is a diagonal matrix containing the eigenvalues of $A'A$. Further, \mathcal{D}^{-1} is a diagonal matrix reciprocal to all positive main diagonal elements of \mathcal{D} whereas all zero elements of \mathcal{D} are retained unchanged.

Sometimes we will need to calculate the absolute value of the 'determinant' of a rectangular matrix $A \in \mathbb{R}^{d \times k}$ (e.g. the determinant of a rectangular Jacobian). For this case we define

$$|\det(A)| := \prod_{i=1}^k \sqrt{\mathcal{D}_{ii}},$$

where \mathcal{D}_{ii} is the i -th diagonal element of \mathcal{D} ($i = 1, \dots, k$). If $r(A) = k$ this quantity can be interpreted as the volume of the trapezoid generated by the column vectors of A . Note that both the pseudo-inverse and the absolute pseudo-determinant are generalizations of the corresponding non-pseudo functions.

In the following every positive (semi-)definite matrix is supposed to be symmetric. Let $A \in \mathbb{R}^{d \times d}$ be a positive semidefinite matrix with $r(A) = r$. The matrix A always has an LDL' -decomposition, i.e.

$$A = LDL',$$

where L is a lower triangular matrix and D is a diagonal matrix where the first r main diagonal entries are positive and the residual entries correspond to zero. Thus we can represent A as

$$A = \left(L\sqrt{D} \right) \left(L\sqrt{D} \right)',$$

where \sqrt{D} is diagonal, too, containing the roots of the main diagonal entries of D . Let $C \in \mathbb{R}^{d \times r}$ be the rectangular matrix of the first r columns of $L\sqrt{D}$. Thus $A = CC'$ and C is called the ‘generalized Cholesky root’ of A .

Further, a ‘measurable’ function is always ment to be Lebesgue measurable. An ‘increasing’ or ‘decreasing’ function is always supposed to be monotonic but not necessarily in the strict sense. The term ‘independence’ always means stochastic independence unless otherwise noted. The sample realizations of n independent copies of X are denoted by the matrix

$$S_n := \begin{bmatrix} x_{.1} & x_{.2} & \cdots & x_{.n} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{d1} & \cdots & \cdots & x_{dn} \end{bmatrix}.$$

Hence a ‘sample’ is always supposed to contain independent and identically distributed data.

A random vector which corresponds to a real number (almost surely) as well as its corresponding distribution function is called ‘degenerate’. The variance of a random vector X corresponds to its covariance matrix, i.e.

$$\text{Var}(X) := E \left((X - E(X)) (X - E(X))' \right).$$

The distribution function (‘cumulative density function’) of a random quantity is abbreviated by ‘c.d.f.’ (even if it is not absolutely continuous) whereas its (probability) density function is labeled by ‘p.d.f.’. The abbreviation ‘i.i.d.’ means ‘independent and identically distributed’ whereas ‘a.s.’ stands for ‘almost surely’. Lists of further notations and abbreviations can be found at the end of the thesis.

Part I

Theory

Chapter 1

Elliptically Symmetric Distributions

The class of elliptically symmetric distributions has been well investigated by Cambanis, Huang, and Simons (1981), Fang, Kotz, and Ng (1990), and Kelker (1970). In the following this class of distributions will be simply called ‘elliptical distributions’ without the additional attribute ‘symmetric’ whenever there is no much fear of confusion. The theory of elliptical distributions is the starting point for the definition and analysis of generalized elliptical distributions. This chapter examines the basic properties of elliptical distributions.

1.1 Definition and Characterization

Definition 1 (Spherical distribution) *Let X be a d -dimensional random vector. X is said to be ‘spherically distributed’ (or simply ‘spherical’) if and only if $X =_d \mathcal{O}X$ for every d -dimensional orthonormal matrix \mathcal{O} .*

Spherical distributions and the corresponding random vectors sometimes are also called ‘radial’ (Kelker, 1970) or ‘isotropic’ (Bingham and Kiesel, 2002). According to the definition above the class of spherical distributions corresponds to the class of rotationally symmetric distributions. Let $U^{(d)}$ be uniformly distributed on the unit hypersphere with $d - 1$ topological dimensions,

$$\mathcal{S}^{d-1} := \{x \in \mathbb{R}^d : \|x\|_2 = 1\},$$

where $\mathcal{S} := \mathcal{S}^1$. Then every d -dimensional random vector X which can be represented as $X =_d \mathcal{R}U^{(d)}$, where \mathcal{R} is a nonnegative random variable stochastically independent of $U^{(d)}$, is rotationally symmetric and thus spherical. The remaining question is if a spherical random vector X is necessarily representable by $\mathcal{R}U^{(d)}$.

Let $t \in \mathbb{R}^d$ and $\angle(t, X)$ be the angle between t and a d -dimensional spherical random vector X . Since $t'X = \|X\|_2 \cdot \|t\|_2 \cdot \cos(\angle(t, X))$ the characteristic function of X corresponds to

$$t \longmapsto \varphi_X(t) := E(\exp(it'X)) = E(\exp(i \cdot \|X\|_2 \cdot \|t\|_2 \cdot \cos(\angle(t, X)))).$$

Using the law of total expectations we find that

$$\begin{aligned} t \longmapsto \varphi_X(t) &= \int_0^\infty E(\exp(i \cdot r \|t\|_2 \cos(\angle(t, X)))) dF_{\|X\|_2}(r) \\ &= \int_0^\infty \varphi_{\cos(\angle(t, X))}(r \|t\|_2) dF_{\|X\|_2}(r), \end{aligned}$$

where $\varphi_{\cos(\angle(t, X))}$ is the characteristic function of $\cos(\angle(t, X))$ and $F_{\|X\|_2}$ is the c.d.f. of the Euclidean norm $\|X\|_2$.

Due to the rotational symmetry of X the stochastic equality

$$\cos(\angle(t, X)) \stackrel{d}{=} \cos\left(\angle\left(v, U^{(d)}\right)\right) \stackrel{d}{=} v'U^{(d)}$$

holds for every $v \in \mathcal{S}^{d-1}$ and $U^{(d)}$ being uniformly distributed on \mathcal{S}^{d-1} . Hence

$$\begin{aligned} s \longmapsto \varphi_{\cos(\angle(t, X))}(s) &= \varphi_{v'U^{(d)}}(s) = E\left(\exp\left(isv'U^{(d)}\right)\right) \\ &= E\left(\exp\left(i(sv)'U^{(d)}\right)\right) = \varphi_{U^{(d)}}(sv) \end{aligned}$$

for any arbitrary $v \in \mathcal{S}^{d-1}$ where $\varphi_{U^{(d)}}$ is the characteristic function of $U^{(d)}$. Thus

$$\varphi_{\cos(\angle(t, X))}(r\|t\|_2) = \varphi_{U^{(d)}}\left(r\|t\|_2 \cdot \frac{t}{\|t\|_2}\right) = \varphi_{U^{(d)}}(rt) = \varphi_{rU^{(d)}}(t),$$

for any $r \geq 0$ since $t/\|t\|_2 \in \mathcal{S}^{d-1}$. So we obtain

$$t \longmapsto \varphi_X(t) = \int_0^\infty \varphi_{rU^{(d)}}(t) dF_{\|X\|_2}(r), \quad t \in \mathbb{R}^d.$$

The right hand side of this equation corresponds to the characteristic function of a random vector $\mathcal{R}U^{(d)}$, where \mathcal{R} is a nonnegative random variable having the same distribution as $\|X\|_2$ and being independent of $U^{(d)}$. Thus every spherical random vector X is necessarily representable by $X =_d \mathcal{R}U^{(d)}$. We call \mathcal{R} the ‘generating random variable’ or ‘generating variate’ of X (Schmidt, 2002).

Example 1 (Generating variate of $X \sim N_d(0, I_d)$) Let $X \sim N_d(0, I_d)$ be represented by $X =_d \mathcal{R}U^{(d)}$. Since

$$\chi_d^2 \stackrel{d}{=} X'X \stackrel{d}{=} \mathcal{R}^2 U^{(d)'}U^{(d)} \stackrel{a.s.}{=} \mathcal{R}^2,$$

the generating variate of X corresponds to $\sqrt{\chi_d^2}$.

Now consider the characteristic function $\varphi_{U^{(d)}}$ of $U^{(d)}$. We know that $\varphi_{U^{(d)}}(sv)$ does not depend on the point v (provided $v \in \mathcal{S}^{d-1}$) but only on $s \in \mathbb{R}$. Moreover, since $\varphi_{U^{(d)}}((-s)v) = \varphi_{U^{(d)}}(s(-v))$ and $-v \in \mathcal{S}^{d-1}$ the considered quantity even does not depend on the sign of s but only on its absolute value $|s|$ or its square s^2 , alternatively. So we can find a function $\phi_{U^{(d)}}$ such that $\varphi_{U^{(d)}}(sv) = \phi_{U^{(d)}}(s^2)$ for every $s \in \mathbb{R}$. Since

$$\varphi_{U^{(d)}}(t) = \varphi_{U^{(d)}}\left(\|t\|_2 \cdot \frac{t}{\|t\|_2}\right) = \phi_{U^{(d)}}\left(\|t\|_2^2\right) = \phi_{U^{(d)}}(t't), \quad t \in \mathbb{R}^d,$$

and thus $\varphi_{rU^{(d)}}(t) = \phi_{U^{(d)}}(r^2 t't)$ we obtain

$$t \longmapsto \varphi_X(t) = \int_0^\infty \phi_{U^{(d)}}(r^2 t't) dF_{\mathcal{R}}(r), \quad t \in \mathbb{R}^d,$$

for the characteristic function of X .

The characteristic function $t \mapsto \phi_{U^{(d)}}(t't)$ depends only on d . To emphasize this we define $\Omega_d := \phi_{U^{(d)}}$ (Schoenberg, 1938). Hence, φ_X can be represented through

$$s \longmapsto \phi_X(s) = \int_0^\infty \Omega_d(r^2 s) dF_{\mathcal{R}}(r), \quad s \geq 0. \quad (1.1)$$

See Fang, Kotz, and Ng (1990, p. 70) for an analytic expression of Ω_d . Since $t \mapsto \varphi_X(t) = \phi_X(t't)$ the function ϕ_X is called the ‘characteristic generator’ of X . Note that ϕ_X is always real valued due to the rotational symmetry of $U^{(d)}$ (Schmidt, 2002).

Example 2 (Characteristic generator of $X \sim N_d(0, I_d)$) Since the characteristic function of an univariate standard normally distributed random variable corresponds to $t \mapsto \exp(-t^2/2)$ (see, e.g., Fisz, 1989, p. 136) and the components of $X \sim N_d(0, I_d)$ are mutually independent the characteristic function of X corresponds to

$$t = (t_1, \dots, t_d) \mapsto \prod_{i=1}^d \exp\left(-\frac{t_i^2}{2}\right) = \exp\left(-\frac{t't}{2}\right).$$

Thus the characteristic generator of X is $s \mapsto \phi_X(s) = \exp(-s/2)$.

Of course, every function ϕ of the form (1.1) is a characteristic generator. Conversely, every characteristic generator can be represented by Eq. 1.1. This theorem belongs to Schoenberg (1938). Note that the characteristic generator contains all information about the generating variate \mathcal{R} .

Proposition 1 Let X be a k -dimensional spherically distributed random vector with characteristic generator ϕ_X . Further, let $\Lambda \in \mathbb{R}^{d \times k}$ be an arbitrary matrix and $\mu \in \mathbb{R}^d$. Then the characteristic function φ_Y of $Y := \mu + \Lambda X$ corresponds to

$$t \mapsto \varphi_Y(t) = \exp(it'\mu) \cdot \phi_X(t'\Sigma t), \quad t \in \mathbb{R}^d,$$

where $\Sigma := \Lambda\Lambda'$.

Proof. The characteristic function of Y corresponds to

$$\begin{aligned} t \mapsto \varphi_Y(t) &= E(\exp(it'(\mu + \Lambda X))) = \exp(it'\mu) \cdot \varphi_X(\Lambda't) \\ &= \exp(it'\mu) \cdot \phi_X\left((\Lambda't)'(\Lambda't)\right) = \exp(it'\mu) \cdot \phi_X(t'\Sigma t). \end{aligned}$$

■

This is the basis for the classical definition of elliptical distributions (cf. Cambanis, Huang, and Simons, 1981) given below.

Definition 2 (Elliptical distribution) Let X be a d -dimensional random vector. X is said to be ‘elliptically distributed’ (or simply ‘elliptical’) if and only if there exist a vector $\mu \in \mathbb{R}^d$, a positive semidefinite matrix $\Sigma \in \mathbb{R}^{d \times d}$, and a function $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ such that the characteristic function $t \mapsto \varphi_{X-\mu}(t)$ of $X - \mu$ corresponds to $t \mapsto \phi(t'\Sigma t)$, $t \in \mathbb{R}^d$.

If a d -dimensional random vector X is elliptically distributed with the parameters specified in Definition 2 we write ‘ $X \sim \mathcal{E}_d(\mu, \Sigma, \phi)$ ’. Hence, a random vector $Y \sim \mathcal{E}_d(0, I_d, \phi)$ is spherically distributed. Due to Proposition 1 every affinely transformed spherical random vector is elliptically distributed. The following stochastic representation theorem shows that the converse is true if the transformation matrix has full rank.

Theorem 2 (Cambanis, Huang, and Simons, 1981) $X \sim \mathcal{E}_d(\mu, \Sigma, \phi)$ with $r(\Sigma) = k$ if and only if

$$X \stackrel{d}{=} \mu + \mathcal{R}\Lambda U^{(k)},$$

where $U^{(k)}$ is a k -dimensional random vector uniformly distributed on \mathcal{S}^{k-1} , \mathcal{R} is a non-negative random variable being stochastically independent of $U^{(k)}$, $\mu \in \mathbb{R}^d$, and $\Lambda \in \mathbb{R}^{d \times k}$ with $r(\Lambda) = k$.

Proof. The ‘if’ follows immediately from Proposition 1. For the ‘only if’ please recognize that every positive semidefinite matrix $\Sigma \in \mathbb{R}^{d \times d}$ with $r(\Sigma) = k$ has a root $\Lambda \in \mathbb{R}^{d \times k}$ such that $\Lambda\Lambda' = \Sigma$. Hence, we may define the random vector

$$Y := \Lambda^{-1}(X - \mu)$$

by using the pseudo-inverse $\Lambda^{-1} \in \mathbb{R}^{k \times d}$ of Λ . Note that $\Lambda^{-1}\Lambda = I_k$ as well as $\Lambda'\Lambda'^{-1} = I_k$. Thus the characteristic function of Y corresponds to

$$\begin{aligned} t \mapsto \varphi_Y(t) &= \varphi_{X-\mu}\left((\Lambda^{-1})'t\right) = \phi\left(t'\Lambda^{-1}\Sigma\Lambda'^{-1}t\right) \\ &= \phi\left(t'\Lambda^{-1}(\Lambda\Lambda')\Lambda'^{-1}t\right) = \phi(t't), \quad t \in \mathbb{R}^k, \end{aligned}$$

and so Y is spherically distributed with characteristic generator ϕ and can be represented by $\mathcal{R}U^{(k)}$, stochastically. Hence $\mu + \Lambda Y \stackrel{d}{=} \mu + \mathcal{R}\Lambda U^{(k)} \sim \mathcal{E}_d(\mu, \Sigma, \phi)$. ■

Due to the transformation matrix Λ the spherical random vector $U^{(k)}$ produces elliptically contoured density surfaces, whereas the generating random variable \mathcal{R} determines the distribution’s shape, in particular the tailedness of the distribution. Further, μ determines the location of the random vector X .

The stochastic representation of an elliptically distributed random vector is usually more convenient for practical purposes than its characteristic representation. Especially, due to the stochastic representation we see that elliptical random vectors can be easily simulated. Let $X \sim N_k(0, I_k)$, i.e. $X \stackrel{d}{=} \sqrt{\chi_k^2} U^{(k)}$. Then

$$\frac{X}{\|X\|_2} \stackrel{d}{=} \frac{\sqrt{\chi_k^2} U^{(k)}}{\|\sqrt{\chi_k^2} U^{(k)}\|_2} \stackrel{\text{a.s.}}{=} \frac{U^{(k)}}{\|U^{(k)}\|_2} \stackrel{\text{a.s.}}{=} U^{(k)}.$$

Hence the random vector $U^{(k)}$ can be simulated simply by dividing a standard normally distributed random vector by its length. Indeed, for simulating \mathcal{R} its c.d.f. must be known (at least approximately).

The matrix Σ is called the ‘dispersion matrix’ or ‘scatter matrix’ of X . So every elliptical distribution belongs to a location-scale-family (Kelker, 1970) defined by an underlying spherical ‘standard’ distribution. For $d = 1$ the class of elliptical distributions coincides with the class of univariate symmetric distributions (Cambanis, Huang, and Simons, 1981).

Example 3 (Multivariate normal distribution) Let $\mu \in \mathbb{R}^d$ and $\Lambda \in \mathbb{R}^{d \times k}$ such that $\Sigma := \Lambda\Lambda' \in \mathbb{R}^{d \times d}$ is positive definite. The random vector $X \sim N_d(\mu, \Sigma)$ is elliptically distributed since X is representable as

$$X \stackrel{d}{=} \mu + \sqrt{\chi_k^2} \Lambda U^{(k)}$$

(see, e.g., Hult and Lindskog, 2002). The underlying spherical standard distribution is the standard normal (see Example 1). Further, since $s \mapsto \exp(-s/2)$ is the characteristic generator for the class of normal distributions (see Example 2) the characteristic function of $X - \mu$ corresponds to $t \mapsto \varphi_{X-\mu}(t) = \exp(-t'\Sigma t/2)$, $t \in \mathbb{R}^d$.

Note that the generating variate of an elliptical location-scale family may vary under d . We will come back to this point in Section 1.2.3 and in Section 1.2.5. Nevertheless, the index ‘ d ’ on the generating variate is omitted for sake of simplicity as long as no confusion is in sight.

Example 4 (Multivariate t -distribution) Consider the random vector

$$Y \stackrel{d}{=} \frac{X}{\sqrt{\frac{\chi_\nu^2}{\nu}}}, \quad \nu \in \mathbb{N},$$

where $X \sim N_d(0, I_d)$ with χ_ν^2 and X being independent. Then Y is said to be ‘multivariate t -distributed with ν degrees of freedom’ (Fang, Kotz, and Ng, 1990, p. 32 and Peracchi, 2001, p. 87). X can be represented by $\sqrt{\chi_d^2} U^{(d)}$ (see Example 1), where U , χ_d^2 and χ_ν^2 are mutually independent. So Y can be represented by

$$Y \stackrel{d}{=} \frac{\sqrt{\chi_d^2}}{\sqrt{\frac{\chi_\nu^2}{\nu}}} \cdot U^{(d)} = \sqrt{d \cdot \frac{\chi_d^2/d}{\chi_\nu^2/\nu}} \cdot U^{(d)} \stackrel{d}{=} \sqrt{d \cdot F_{d,\nu}} \cdot U^{(d)},$$

where $F_{d,\nu}$ is an F -distributed random variable with d and ν degrees of freedom and independent of $U^{(d)}$. Further,

$$\frac{\sqrt{\chi_d^2}}{\sqrt{\frac{\chi_\nu^2}{\nu}}} \xrightarrow{d} \sqrt{\chi_d^2}, \quad \nu \longrightarrow \infty,$$

as a consequence of $\chi_\nu^2/\nu \xrightarrow{a.s.} 1$ due to the strong law of large numbers. Thus $Y \xrightarrow{d} N_d(0, I_d)$ for $\nu \rightarrow \infty$. Note that the random vector $\mu + \sqrt{d \cdot F_{d,\nu}} \Lambda U^{(d)}$ has a multivariate t -distribution with location vector μ and dispersion matrix $\Sigma = \Lambda \Lambda'$ provided Λ has full rank (see, e.g., Hult and Lindskog, 2002).

In the following we will generally allow for Σ being positive semidefinite also in the context of multivariate normal and t -distributions. Especially for the t -distribution the number ν of degrees of freedom is not required anymore to be an integer but a positive real number. The corresponding d -variate t -distribution will be denoted by $t_d(\mu, \Sigma, \nu)$.

It is somewhat surprising that the dispersion of an elliptically distributed random vector is uniquely determined by the matrix Σ , i.e. the particular matrix decomposition Λ is irrelevant even though Λ determines the support of $\Lambda U^{(d)}$. Consider the elliptical surface generated by a nonsingular matrix A , i.e.

$$\mathcal{E}_A = \{Au : u \in \mathcal{S}^{d-1}\},$$

and let $\Sigma := AA'$. Now focus an arbitrary point $x_0 = Au_0$ of the surface and let B be a nonsingular matrix satisfying $BB' = \Sigma$, too. Suppose there is a point v_0 such that $Bv_0 = Au_0 = x_0$. Then $v_0 = B^{-1}Au_0$ and

$$\begin{aligned} \|v_0\|_2 &= \|B^{-1}Au_0\|_2 = \sqrt{(B^{-1}Au_0)' B^{-1}Au_0} = \sqrt{u_0' A' (B^{-1})' B^{-1} Au_0} \\ &= \sqrt{u_0' A' \Sigma^{-1} Au_0} = \sqrt{u_0' A' A'^{-1} A^{-1} Au_0} = \sqrt{u_0' u_0} = \|u_0\|_2 = 1. \end{aligned}$$

Thus, any point $x = Au \in \mathcal{E}_A$ can be represented by a linear transformation B of a point v on the unit sphere surface \mathcal{S}^{d-1} (not necessarily $v = u$), i.e. $\mathcal{E}_A \subset \mathcal{E}_B$. Conversely, if $y_0 = Bv_0$ is an element of the elliptical surface \mathcal{E}_B generated by B then y_0 is also an element of \mathcal{E}_A because (by the same token) there is always a point $u_0 \in \mathcal{S}^{d-1}$ such that $Au_0 = y_0$. Hence \mathcal{E}_A corresponds to \mathcal{E}_B , that is the linear transformations A and B generate the same elliptical surfaces. Since $U^{(d)}$ is uniformly distributed on \mathcal{S}^{d-1} and the generating variate \mathcal{R} does not depend on $U^{(d)}$ the random vectors AU and BU have the same support.

1.2 Basic Properties

1.2.1 Density Functions

A nice property of an elliptical distribution function is the fact that its multivariate density function may be expressed via the density function of the generating variate, provided this is absolutely continuous. In the following and throughout the thesis density functions are allowed to be defined not only on \mathbb{R}^d but on certain lower dimensional linear subspaces and manifolds of \mathbb{R}^d .

Theorem 3 Let $X \sim \mathcal{E}_d(\mu, \Sigma, \phi)$ where $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$ is positive semidefinite with $\text{r}(\Sigma) = k$. Then X can be represented stochastically by $X =_d \mu + \mathcal{R}\Lambda U^{(k)}$ with $\Lambda\Lambda' = \Sigma$ according to Theorem 2. Further, let the c.d.f. of \mathcal{R} be absolutely continuous and \mathcal{S}_Λ be the linear subspace of \mathbb{R}^d spanned by Λ . Then the p.d.f. of X is given by

$$x \mapsto f_X(x) = |\det(\Lambda)|^{-1} \cdot g_{\mathcal{R}}((x - \mu)' \Sigma^{-1} (x - \mu)), \quad x \in \mathcal{S}_\Lambda \setminus \{\mu\},$$

where

$$t \mapsto g_{\mathcal{R}}(t) := \frac{\Gamma(\frac{k}{2})}{2\pi^{k/2}} \cdot \sqrt{t}^{-(k-1)} \cdot f_{\mathcal{R}}(\sqrt{t}), \quad t > 0,$$

and $f_{\mathcal{R}}$ is the p.d.f. of \mathcal{R} .

Proof. Since the c.d.f. of \mathcal{R} is absolutely continuous the joint p.d.f. of \mathcal{R} and $U^{(k)}$ exists and corresponds to

$$(r, u) \mapsto f_{(\mathcal{R}, U^{(k)})}(r, u) = \frac{\Gamma(\frac{k}{2})}{2\pi^{k/2}} \cdot f_{\mathcal{R}}(r), \quad r > 0, \quad u \in \mathcal{S}^{k-1},$$

where $f_{\mathcal{R}}$ is the density function of \mathcal{R} . Note that $\Gamma(\frac{k}{2}) / (2\pi^{k/2})$ corresponds to the uniform density on the unit hypersphere \mathcal{S}^{k-1} . To get the density of $\mathcal{R}U^{(k)} =: Y$ we define the transformation $h:]0, \infty[\times \mathcal{S}^{k-1} \rightarrow \mathbb{R}^k \setminus \{0\}$, $(r, u) \mapsto ru =: y$. Note that h is injective and the p.d.f. of Y is given by

$$y \mapsto f_Y(y) = f_{(\mathcal{R}, U^{(k)})}(h^{-1}(y)) \cdot |J_h|^{-1}, \quad y \neq 0,$$

where J_h is the Jacobian determinant of $\partial ru / \partial (r, u)'$. Let

$$\mathcal{S}_r^{k-1} := \{x \in \mathbb{R}^k : \|x\|_2 = r > 0\}$$

be the hypersphere with radius r . Since the partial derivative $\partial ru / \partial r$ has unit length and is orthogonal to each tangent plane $\partial ru / \partial u'$ on \mathcal{S}_r^{k-1} which has only $k - 1$ topological dimensions, the absolute value of the Jacobian determinant of $\partial ru / \partial (r, u)'$ is given by

$$|J_h| = \det \left(\begin{bmatrix} 1 & 0' \\ 0 & rI_{k-1} \end{bmatrix} \right) = r^{k-1} = \|y\|_2^{k-1}, \quad y \neq 0.$$

Further, $h^{-1}(y) = (\|y\|_2, y / \|y\|_2)$ and so the p.d.f. of Y corresponds to

$$\begin{aligned} y \mapsto f_Y(y) &= f_{(\mathcal{R}, U^{(k)})}(\|y\|_2, y / \|y\|_2) \cdot \|y\|_2^{-(k-1)} \\ &= \frac{\Gamma(\frac{k}{2})}{2\pi^{k/2}} \cdot \|y\|_2^{-(k-1)} \cdot f_{\mathcal{R}}(\|y\|_2), \quad y \neq 0, \end{aligned}$$

where $u = y / \|y\|_2$. Now we define the transformation $q: \mathbb{R}^k \setminus \{0\} \rightarrow \mathcal{S}_\Lambda \setminus \{\mu\}$, $y \mapsto \mu + \Lambda y =: x$. Note that since $\Lambda^{-1}\Lambda = I_k$ the transformation q is injective. The absolute value of the Jacobian determinant of $\partial(\mu + \Lambda y) / \partial y'$ corresponds to $|J_q| = |\det(\Lambda)|$, and thus the p.d.f. of $X =_d \mu + \Lambda Y = \mu + \mathcal{R}\Lambda U^{(k)}$ is given by

$$\begin{aligned} x \mapsto f_X(x) &= f_Y(q^{-1}(x)) \cdot |J_q|^{-1} \\ &= f_Y(\Lambda^{-1}(x - \mu)) \cdot |\det(\Lambda)|^{-1}, \quad x \in \mathcal{S}_\Lambda \setminus \{\mu\}. \end{aligned}$$

Hence the p.d.f. of X becomes

$$x \mapsto f_X(x) = |\det(\Lambda)|^{-1} \cdot \frac{\Gamma(\frac{k}{2})}{2\pi^{k/2}} \cdot \|\Lambda^{-1}(x - \mu)\|_2^{-(k-1)} \cdot f_{\mathcal{R}}(\|\Lambda^{-1}(x - \mu)\|_2),$$

with $x \in \mathcal{S}_\Lambda \setminus \{\mu\}$. Since

$$\|\Lambda^{-1}(x - \mu)\|_2 = \sqrt{(x - \mu)' \Lambda'^{-1} \Lambda^{-1} (x - \mu)},$$

and per definition

$$\Lambda'^{-1} \Lambda^{-1} = \left((\Lambda \Lambda')^{-1} \Lambda \right) \left((\Lambda \Lambda')^{-1} \Lambda \right)' = (\Sigma^{-1} \Lambda) (\Sigma^{-1} \Lambda)' = \Sigma^{-1} \Sigma \Sigma^{-1} = \Sigma^{-1},$$

we obtain the formula given in Theorem 3. ■

The function $g_{\mathcal{R}}$ is called ‘density generator’ or ‘p.d.f. generator’ (Fang, Kotz, and Ng, 1990, p. 35) of X (or of F_X , respectively). Note that the density contours produced by the density generator corresponds to elliptical surfaces. For this reason elliptical distributions are often referred to as ‘elliptically contoured’ distributions (Cambanis, Huang, and Simons, 1981).

The following corollary corresponds to the classical theorem for elliptically contoured density functions providing a nonsingular dispersion matrix (see, e.g., Fang, Kotz, and Ng, 1990, p. 46).

Corollary 4 *Let $X \sim \mathcal{E}_d(\mu, \Sigma, \phi)$ where $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$ is positive definite. Then X can be represented stochastically by $X =_d \mu + \mathcal{R} \Lambda U^{(d)}$ with $\Lambda \Lambda' = \Sigma$ according to Theorem 2. Further, let the c.d.f. of \mathcal{R} be absolutely continuous. Then the p.d.f. of X is given by*

$$x \mapsto f_X(x) = \sqrt{\det(\Sigma^{-1})} \cdot g_{\mathcal{R}}((x - \mu)' \Sigma^{-1} (x - \mu)), \quad x \neq \mu,$$

where

$$t \mapsto g_{\mathcal{R}}(t) := \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^{d/2}} \cdot \sqrt{t}^{-(d-1)} \cdot f_{\mathcal{R}}(\sqrt{t}), \quad t > 0,$$

and $f_{\mathcal{R}}$ is the p.d.f. of \mathcal{R} .

Proof. The corollary follows immediately from Theorem 3 after substituting k by d and considering that

$$|\det(\Lambda)|^{-1} = \sqrt{\det(\Lambda) \cdot \det(\Lambda')^{-1}} = \sqrt{\det(\Sigma)}^{-1} = \sqrt{\det(\Sigma^{-1})},$$

since Λ is nonsingular. ■

Given the p.d.f. $f_{\mathcal{R}}$ of the generating variate one can simply calculate the density generator of the corresponding elliptical distribution.

Example 5 (Density generator of $X \sim N_d(0, I_d)$) *The p.d.f. of χ_d^2 corresponds to*

$$x \mapsto f(x) = \frac{x^{\frac{d}{2}-1} \cdot e^{-\frac{x}{2}}}{2^{d/2} \cdot \Gamma\left(\frac{d}{2}\right)}, \quad x \geq 0,$$

(cf., e.g., Peracchi, 2001, p. 81). Thus the p.d.f. of $\mathcal{R} := \sqrt{\chi_d^2}$ is given by

$$r \mapsto f_{\mathcal{R}}(r) = 2r \cdot f(r^2),$$

and the density generator of $X =_d \sqrt{\chi_d^2} U^{(d)}$ equals to

$$t \mapsto g_{\sqrt{\chi_d^2}}(t) = \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^{d/2}} \cdot \sqrt{t}^{-(d-1)} \cdot 2\sqrt{t} \cdot f(t) = \frac{1}{(2\pi)^{d/2}} \cdot \exp\left(-\frac{t}{2}\right),$$

which corresponds to the generator of the multivariate normal distribution.

Conversely, given a density generator $g_{\mathcal{R}}$ one may derive the corresponding density function $f_{\mathcal{R}}$ by

$$r \mapsto f_{\mathcal{R}}(r) = \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})} \cdot r^{d-1} \cdot g_{\mathcal{R}}(r^2).$$

Example 6 ($f_{\mathcal{R}}$ of $X \sim t_d(\mu, \Sigma, \nu)$) *The density function of a multivariate t -distribution corresponds to*

$$x \mapsto f_X(x) = \frac{\Gamma(\frac{d+\nu}{2})}{\Gamma(\frac{\nu}{2})} \cdot \sqrt{\frac{\det(\Sigma^{-1})}{(\nu\pi)^d}} \cdot \left(1 + \frac{(x-\mu)' \Sigma^{-1} (x-\mu)}{\nu}\right)^{-\frac{d+\nu}{2}},$$

where $\nu > 0$ and Σ is assumed to be positive definite (see, e.g., Peracchi, 2001, p. 87). So the density generator of X is

$$t \mapsto g_{\mathcal{R}}(t) = \frac{\Gamma(\frac{d+\nu}{2})}{\Gamma(\frac{\nu}{2})} \cdot \frac{1}{(\nu\pi)^{d/2}} \cdot \left(1 + \frac{t}{\nu}\right)^{-\frac{d+\nu}{2}}.$$

After some algebra we find

$$\begin{aligned} r \mapsto f_{\mathcal{R}}(r) &= \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})} \cdot r^{d-1} \cdot g_{\mathcal{R}}(r^2) \\ &= \frac{2r}{d} \cdot \frac{\Gamma(\frac{d+\nu}{2})}{\Gamma(\frac{d}{2}) \cdot \Gamma(\frac{\nu}{2})} \cdot \left(\frac{d}{\nu}\right)^{\frac{d}{2}} \cdot \left(\frac{r^2}{d}\right)^{\frac{d}{2}-1} \\ &\quad \cdot \left(1 + \frac{d}{\nu} \cdot \frac{r^2}{d}\right)^{-\frac{d+\nu}{2}} = \frac{2r}{d} \cdot f_F\left(\frac{r^2}{d}\right), \end{aligned}$$

where f_F represents the p.d.f. of an $F_{d,\nu}$ -distributed random vector (see, e.g., Peracchi, 2001, p. 85). But $r \mapsto 2r/d \cdot f_F(r^2/d)$ is just the p.d.f. of the random variable $\sqrt{d} \cdot F_{d,\nu}$ (see Example 4).

1.2.2 Symmetry

There are several definitions of symmetry of multivariate distributions and random vectors. I am going to concentrate on four basic symmetries which are ‘rotational’, ‘permutational’, ‘radial’, and ‘angular’ symmetry. For a more advanced discussion of symmetry of distribution functions see, e.g., Fang, Kotz, and Ng (1990, pp. 1-10).

Rotational symmetry was already mentioned by Definition 1. A weaker form of symmetry is called ‘permutational symmetry’ or ‘exchangeability’ (Fang, Kotz, and Ng, 1990, p. 5). That is a d -dimensional random vector X satisfying

$$X \stackrel{d}{=} \mathcal{P}X,$$

for every d -dimensional permutation matrix \mathcal{P} . Every rotationally symmetric random vector is also permutationally symmetric because every \mathcal{P} is orthonormal, but the converse is not true. Exchangeability is equivalent to $F_X(x) = F_X(\pi_x)$ for all permutations π_x of the vector (x_1, \dots, x_d) .

Example 7 (Exchangeability of independent random components) *Every random vector X with mutually independent and identically distributed components X_1, \dots, X_d is permutationally symmetric, since $F_X = \prod_{i=1}^d F_i$ and $F_1 = \dots = F_d$.*

Example 8 (Exchangeability of equicorrelated elliptical components) Consider a d -dimensional elliptical random vector X with zero location, i.e. $\mu = 0$ and equicovariance structure, i.e.

$$\Sigma = \begin{bmatrix} b & a & \cdots & a \\ a & b & & a \\ \vdots & & \ddots & \vdots \\ a & a & \cdots & b \end{bmatrix},$$

where $-b/(d-1) < a < b$. Now,

$$\mathcal{P}X \stackrel{d}{=} \mathcal{P}(\mathcal{R}\Lambda U^{(d)}) = \mathcal{R}\mathcal{P}\Lambda U^{(d)},$$

where $\Lambda\Lambda' = \Sigma$. Thus, the dispersion matrix of $\mathcal{P}X$ corresponds to $\mathcal{P}\Lambda\Lambda'\mathcal{P}' = \mathcal{P}\Sigma\mathcal{P}' = \Sigma$. So X and $\mathcal{P}X$ have the same distribution, i.e. X is permutationally symmetric.

A d -dimensional random vector X is called ‘radially symmetric’ or simply ‘symmetric (about $c \in \mathbb{R}^d$)’ (Fisz, 1989) if

$$X - c \stackrel{d}{=} -(X - c).$$

Of course, if X is rotationally symmetric then it is also symmetric about 0 since the matrix $-I_d$ is orthonormal and $X \stackrel{d}{=} -I_d X = -X$. From Theorem 3 we see that the density function of an elliptical distribution function F_X is symmetric with respect to its location, i.e.

$$f_X(\mu + x) = f_X(\mu - x), \quad \forall x \in \mathbb{R}^d,$$

provided F_X is absolutely continuous. That is X is radially symmetric about μ . But even if there is no density function an elliptical distribution is always radially symmetric about μ , since

$$-(X - \mu) \stackrel{d}{=} -\mathcal{R}\Lambda U^{(d)} = \mathcal{R}\Lambda(-U^{(d)}) \stackrel{d}{=} \mathcal{R}\Lambda U^{(d)} \stackrel{d}{=} X - \mu.$$

Another kind of symmetry is given by the property

$$\frac{X - c}{\|X - c\|_2} \stackrel{d}{=} -\frac{X - c}{\|X - c\|_2}.$$

Now, X is called ‘angularly symmetric (about $c \in \mathbb{R}^d$)’ (Liu, 1988). If X is radially symmetric it is also angularly symmetric provided X has no atom at its center c . The concept of angular symmetry will play a prominent role in the construction of a robust location vector estimator for generalized elliptical distributions (see Section 4.3).

Hence, spherical distributions are rotationally, permutationally, radially, and (provided $\mathcal{R} >_{\text{a.s.}} 0$) angularly symmetric. In contrast, elliptical distributions generally are only radially and if $\mathcal{R} >_{\text{a.s.}} 0$ also angularly symmetric. If the elliptical distribution has zero location and equicovariance structure then it is also permutationally symmetric.

1.2.3 Moments

The mean vector of a d -dimensional elliptical random vector X corresponds to

$$E(X) = E\left(\mu + \mathcal{R}\Lambda U^{(k)}\right) = \mu + \Lambda E(\mathcal{R}) \cdot E\left(U^{(k)}\right),$$

since \mathcal{R} and $U^{(k)}$ are supposed to be independent. Here we assume that $E(\mathcal{R})$ is finite. Since $E(U^{(k)}) = 0$ we obtain $E(X) = \mu$.

The covariance matrix of X is

$$\text{Var}(X) = E\left(\left(\mathcal{R}\Lambda U^{(k)}\right)\left(\mathcal{R}\Lambda U^{(k)}\right)'\right) = E(\mathcal{R}^2) \cdot \Lambda E\left(U^{(k)}U^{(k)'}\right)\Lambda',$$

provided $E(\mathcal{R}^2)$ is finite. Since $\sqrt{\chi_k^2} U^{(k)} \sim N_k(0, I_k)$ and therefore

$$I_k = E \left(\left(\sqrt{\chi_k^2} U^{(k)} \right) \left(\sqrt{\chi_k^2} U^{(k)} \right)' \right) = E(\chi_k^2) \cdot E(U^{(k)} U^{(k)'}) = k \cdot E(U^{(k)} U^{(k)'}),$$

we obtain $E(U^{(k)} U^{(k)'}) = I_k/k$ and thus

$$\text{Var}(X) = \frac{E(\mathcal{R}^2)}{k} \cdot \Sigma.$$

Note that k is not necessarily the rank of Σ or the dimension of X but the number of components of $U^{(k)}$. Further, the dispersion matrix generally does not coincide with the covariance matrix. The normal distribution is an exceptional case because $E(\mathcal{R}^2) = E(\chi_k^2) = k$ and thus $\text{Var}(X) = \Sigma$. Nevertheless, by multiplying \mathcal{R} with $\sqrt{k/E(\mathcal{R}^2)}$ we can always find a representation such that $\text{Var}(X) = \Sigma$ (cf. Bingham and Kiesel, 2002 and Hult and Lindskog, 2002).

It was mentioned in Section 1.1 that the generating distribution function of an elliptical location-scale family usually depends on its dimension d . Suppose the spherical random vector which is underlying to a location-scale family has the stochastic representation

$$X^{(d)} \stackrel{d}{=} \mathcal{R}^{(d)} U^{(d)}, \quad \forall d \in \mathbb{N},$$

where $U^{(d)}$ is uniformly distributed on \mathcal{S}^{d-1} and $\mathcal{R}^{(d)}$ is a generating variate such that $X^{(d)}$ has always the characteristic function $t \mapsto \phi(t't)$. That is to say the characteristic generator ϕ is supposed to be independent of d . Then the characteristic function of the marginal c.d.f. of an arbitrary component of $X^{(d)}$ is always (i.e. for $d = 1, \dots$) given by $s \mapsto \phi(s^2)$ where $s \in \mathbb{R}$. Hence, the marginal distribution functions and their existing moments do not depend on d . Consequently, the second moment of $\mathcal{R}^{(d)}$ must be proportional to d provided it is finite.

Example 9 (The 2nd moment of $\mathcal{R}^{(d)}$ for the normal distribution) Since the generating variate of $X^{(d)} \sim N_d(0, I_d)$ corresponds to $\sqrt{\chi_d^2}$ (see Example 1) we obtain

$$E((\mathcal{R}^{(d)})^2) = E(\chi_d^2) = d.$$

The following theorem emerges as very useful for calculating the asymptotic covariances of covariance matrix estimators of (generalized) elliptical distributions treated in the chapters below.

Theorem 5 (Dickey and Chen, 1985) Let $X = (X_1, \dots, X_d)$ be a spherically distributed random vector with stochastic representation $\mathcal{R}U^{(d)}$. Its mixed moment of order (m_1, \dots, m_d) corresponds to

$$E \left(\prod_{i=1}^d X_i^{m_i} \right) = \frac{E(\mathcal{R}^m)}{\left(\frac{d}{2}\right)^{(m/2)}} \cdot \prod_{i=1}^d \frac{m_i!}{2^{m_i} \left(\frac{m_i}{2}\right)!},$$

where $m := \sum_{i=1}^d m_i$ and every m_1, \dots, m_d is supposed to be an even nonnegative integer. Here $(\cdot)^{(k)}$ is the ‘rising factorial’, i.e. $(x)^{(k)} := x \cdot (x+1) \cdots (x+k-1)$ for $k \in \mathbb{N}$ and $(x)^{(0)} := 1$. If at least one of the m_i ’s is odd then the mixed moment vanishes.

Proof. Fang, Kotz, and Ng (1990), p. 73. ■

1.2.4 Affine Transformations and Marginal Distributions

Let $a \in \mathbb{R}^k$ and $A \in \mathbb{R}^{k \times d}$. Consider the transformed random vector

$$Y = a + AX,$$

where $X \stackrel{d}{=} \mu + \mathcal{R}\Lambda U^{(m)}$ with $\Lambda \in \mathbb{R}^{d \times m}$. So we obtain

$$Y \stackrel{d}{=} a + A \left(\mu + \mathcal{R}\Lambda U^{(m)} \right) = (a + A\mu) + \mathcal{R}A\Lambda U^{(m)}.$$

Hence, every affinely transformed and particularly every linearly combined elliptical random vector is elliptical, too. An interesting fact is that the generating variates of affinely transformed random vectors always remain constant. Thus affinely transformed random vectors not only are elliptical but even closed under the corresponding location-scale family. We say that Y is of the ‘same type’. Note that the characteristic function of $Y - (a + A\mu)$ corresponds to $t \mapsto \phi_X(t' A \Sigma A' t)$ where $\Sigma := \Lambda \Lambda'$ (Hult and Lindskog, 2002).

Let $\mathcal{P}_k \in \{0, 1\}^{k \times d}$ ($k \leq d$) be a ‘permutation and deletion’ matrix, i.e. \mathcal{P}_k has only binary entries of 0’s and 1’s and $\mathcal{P}_k \mathcal{P}_k' = I_k$. So the transformation $\mathcal{P}_k X =: Y$ permutes and deletes certain components of X such that Y is a k -dimensional random vector containing the remaining components of X and having a (multivariate) marginal distribution with respect to the joint distribution of X . According to the assertions above

$$Y \stackrel{d}{=} \mathcal{P}_k (\mu + \mathcal{R}\Lambda U) = \mathcal{P}_k \mu + \mathcal{R}\mathcal{P}_k \Lambda U,$$

i.e. Y is of the same type as X . Moreover, the characteristic function of $Y - \mathcal{P}_k \mu$ corresponds to $t \mapsto \phi_X(t' \mathcal{P}_k \Sigma \mathcal{P}_k' t)$. So both the location vector $\mathcal{P}_k \mu$ and the dispersion matrix $\mathcal{P}_k \Sigma \mathcal{P}_k'$ of Y exactly consist of the remaining entries of μ and Σ (Hult and Lindskog, 2002).

1.2.5 Conditional Distributions

The following theorems on the conditional distributions of spherical and elliptical random vectors belong to Kelker (1970) and Cambanis, Huang, and Simons (1981). The corresponding theorems for generalized elliptical distributions in Chapter 3 will heavily rely on the following derivations.

From now on the notation of a ‘conditional random vector’ $Y \mid X = x$ is frequently used. This is a standard notation in multivariate statistics (see, e.g., Bilodeau and Brenner, 1999, Section 5.5 and Fang, Kotz, and Ng, 1990, Section 2.4). The quantity $Y \mid X = x$ is simply a random vector possessing the c.d.f. of Y under the condition $X = x$.

Theorem 6 *Let $X \stackrel{d}{=} \mathcal{R}U^{(d)} \sim \mathcal{E}_d(0, I_d, \phi)$ and $X = (X_1, X_2)$ where X_1 is a k -dimensional sub-vector of X . Provided the conditional random vector $X_2 \mid X_1 = x_1$ exists it is also spherically distributed and can be represented stochastically by*

$$X_2 \mid (X_1 = x_1) \stackrel{d}{=} \mathcal{R}^* U^{(d-k)},$$

where $U^{(d-k)}$ is uniformly distributed on \mathcal{S}^{d-k-1} and the generating variate is given by

$$\mathcal{R}^* = \mathcal{R} \sqrt{1 - \beta} \mid \left(\mathcal{R} \sqrt{\beta} U^{(k)} = x_1 \right). \quad (1.2)$$

Here $U^{(k)}$ is uniformly distributed on \mathcal{S}^{k-1} and $\beta \sim \text{Beta}\left(\frac{k}{2}, \frac{d-k}{2}\right)$ where \mathcal{R} , β , $U^{(k)}$, and $U^{(d-k)}$ are supposed to be mutually independent.

Proof. Let

$$U^{(d)} = \begin{bmatrix} U_1^{(d)} \\ U_2^{(d)} \end{bmatrix} := \begin{bmatrix} \frac{\|Z_1\|_2}{\|Z\|_2} \cdot \frac{Z_1}{\|Z_1\|_2} \\ \frac{\|Z_2\|_2}{\|Z\|_2} \cdot \frac{Z_2}{\|Z_2\|_2} \end{bmatrix} = \begin{bmatrix} \sqrt{\beta} \cdot U^{(k)} \\ \sqrt{1-\beta} \cdot U^{(d-k)} \end{bmatrix}, \quad (1.3)$$

where $Z = (Z_1, Z_2) \sim N_d(0, I_d)$, and

$$U^{(k)} := \frac{Z_1}{\|Z_1\|_2}, \quad U^{(d-k)} := \frac{Z_2}{\|Z_2\|_2}, \quad \sqrt{\beta} := \frac{\|Z_1\|_2}{\|Z\|_2}.$$

Consider the random vector

$$X = (X_1, X_2) \stackrel{d}{=} \begin{bmatrix} \mathcal{R}\sqrt{\beta}U^{(k)} \\ \mathcal{R}\sqrt{1-\beta}U^{(d-k)} \end{bmatrix},$$

where the random quantities \mathcal{R} , β , $U^{(k)}$, and $U^{(d-k)}$ are mutually independent and $\beta \sim \text{Beta}(\frac{k}{2}, \frac{d-k}{2})$ (Cambanis, Huang, and Simons, 1981). ■

Theorem 7 Let $X \sim \mathcal{E}_d(\mu, \Sigma, \phi)$ where $\mu = (\mu_1, \mu_2) \in \mathbb{R}^d$ and the matrix $\Sigma \in \mathbb{R}^{d \times d}$ is positive semidefinite with $r(\Sigma) = r$. Then X can be represented stochastically by $X =_d \mu + \mathcal{R}CU^{(r)}$ according to Theorem 2, where

$$C = \begin{bmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{bmatrix} \in \mathbb{R}^{d \times r}$$

is the generalized Cholesky root of Σ with sub-matrices $C_{11} \in \mathbb{R}^{k \times k}$, $C_{21} \in \mathbb{R}^{(d-k) \times k}$, and $C_{22} \in \mathbb{R}^{(d-k) \times (r-k)}$, respectively. Further, let $X = (X_1, X_2)$ where X_1 is a k -dimensional ($k < r$) sub-vector of X . Provided the conditional random vector $X_2 \mid X_1 = x_1$ exists it is also elliptically distributed and can be represented stochastically by

$$X_2 \mid (X_1 = x_1) \stackrel{d}{=} \mu^* + \mathcal{R}^*C_{22}U^{(r-k)},$$

where $U^{(r-k)}$ is uniformly distributed on \mathcal{S}^{r-k-1} and the generating variate is given by

$$\mathcal{R}^* = \mathcal{R}\sqrt{1-\beta} \mid \left(\mathcal{R}\sqrt{\beta}U^{(k)} = C_{11}^{-1}(x_1 - \mu_1) \right),$$

whereas the location vector corresponds to

$$\mu^* = \mu_2 + C_{21}C_{11}^{-1}(x_1 - \mu_1).$$

Here $U^{(k)}$ is uniformly distributed on \mathcal{S}^{k-1} and $\beta \sim \text{Beta}(\frac{k}{2}, \frac{r-k}{2})$ where \mathcal{R} , β , $U^{(k)}$, and $U^{(r-k)}$ are supposed to be mutually independent.

Proof. Let $U^{(r)}$ be defined as in the proof of Theorem 6 where d is substituted by r . Further, consider

$$\begin{aligned} X = (X_1, X_2) &\stackrel{d}{=} \begin{bmatrix} \mu_1 + C_{11}\mathcal{R}U_1^{(r)} \\ \mu_2 + C_{21}\mathcal{R}U_1^{(r)} + C_{22}\mathcal{R}U_2^{(r)} \end{bmatrix} \\ &= \begin{bmatrix} \mu_1 + C_{11}\mathcal{R}\sqrt{\beta}U^{(k)} \\ \mu_2 + C_{21}\mathcal{R}\sqrt{\beta}U^{(k)} + C_{22}\mathcal{R}\sqrt{1-\beta}U^{(r-k)} \end{bmatrix}. \end{aligned}$$

Under the condition $X_1 = x_1$ the random vector $\mathcal{R}\sqrt{\beta}U^{(k)}$ degenerates to $\mathcal{R}\sqrt{\beta}U^{(k)} = C_{11}^{-1}(x_1 - \mu_1)$. That is

$$\mu^* := \mu_2 + C_{21}\mathcal{R}\sqrt{\beta}U^{(k)} = C_{21}C_{11}^{-1}(x_1 - \mu_1)$$

and the generating variate of $X_2 \mid (X_1 = x_1)$ is given by $\mathcal{R}\sqrt{1-\beta}$ under the specified condition. ■

The following corollary shows that the conditional distribution can be expressed in terms of the components of Σ without the need of its generalized Cholesky root (see Cambanis, Huang, and Simons, 1981 as well as Fang, Kotz, and Ng, 1990, p. 45).

Corollary 8 *Let $X \sim \mathcal{E}_d(\mu, \Sigma, \phi)$ where $\mu = (\mu_1, \mu_2) \in \mathbb{R}^d$ and the matrix $\Sigma \in \mathbb{R}^{d \times d}$ is positive semidefinite with $r(\Sigma) = r$. Then X can be represented stochastically by $X =_d \mu + \mathcal{R}\Lambda U^{(r)}$ with $\Lambda \in \mathbb{R}^{d \times r}$ according to Theorem 2. Let*

$$\Sigma = \Lambda\Lambda' = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \in \mathbb{R}^{d \times d}$$

with sub-matrices $\Sigma_{11} \in \mathbb{R}^{k \times k}$, $\Sigma_{21} \in \mathbb{R}^{(d-k) \times k}$, $\Sigma_{12} \in \mathbb{R}^{k \times (d-k)}$, and $\Sigma_{22} \in \mathbb{R}^{(d-k) \times (d-k)}$, respectively. Further, let $X = (X_1, X_2)$ where X_1 is a k -dimensional ($k < r$) sub-vector of X . Suppose that the conditional random vector $X_2 \mid X_1 = x_1$ exists. Then $X_2 \mid (X_1 = x_1) \sim \mathcal{E}_{d-k}(\mu^*, \Sigma^*, \phi^*)$ where

$$\begin{aligned} \mu^* &= \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1), \\ \Sigma^* &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, \end{aligned}$$

and ϕ^* corresponds to the characteristic generator of $\mathcal{R}^*U^{(r-k)}$ with

$$\mathcal{R}^* = \mathcal{R}\sqrt{1-\beta} \mid \mathcal{R}\sqrt{\beta}U^{(k)} = C_{11}^{-1}(x_1 - \mu_1).$$

Here C_{11} is the Cholesky root of Σ_{11} , $U^{(k)}$ is uniformly distributed on \mathcal{S}^{k-1} and $\beta \sim \text{Beta}(\frac{k}{2}, \frac{r-k}{2})$ where \mathcal{R} , β , $U^{(k)}$, and $U^{(r-k)}$ are supposed to be mutually independent.

Proof. Consider Theorem 7 and note that

$$C_{21}C_{11}^{-1} = (C_{21}C'_{11})(C'^{-1}_{11}C_{11}^{-1}) = \Sigma_{21}\Sigma_{11}^{-1},$$

and thus

$$\mu^* = \mu_2 + C_{21}C_{11}^{-1}(x_1 - \mu_1) = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1).$$

Further, $C_{11}C'_{11} = \Sigma_{11}$ and

$$\begin{aligned} \Sigma^* &= C_{22}C'_{22} = C_{21}C'_{21} + C_{22}C'_{22} - C_{21}C'_{21} \\ &= (C_{21}C'_{21} + C_{22}C'_{22}) - (C_{21}C'_{11})(C'^{-1}_{11}C_{11}^{-1})(C_{11}C'_{21}) \\ &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}. \end{aligned}$$

■

1.3 Additional Properties

1.3.1 Summation Stability

The sum of independent elliptically distributed random vectors X_1, \dots, X_n with the same dispersion matrix Σ is elliptical, too (Hult and Lindskog, 2002). This is because the characteristic function of $\sum_{i=1}^n X_i - \mu$ where $\mu := \sum_{i=1}^n \mu_i$ corresponds to

$$t \longmapsto E \left(\exp \left(it' \sum_{i=1}^n (X_i - \mu_i) \right) \right) = \prod_{i=1}^n E \left(\exp (it' (X_i - \mu_i)) \right) = \prod_{i=1}^n \phi_{X_i - \mu_i} (t' \Sigma t).$$

Especially the sum of i.i.d. elliptical random vectors is closed in the sense that it does not leave the class of elliptical distributions. But that does not mean that the sum is of the same type, i.e. it usually does not belong to the location-scale family of its addends. This is only given for the class of multivariate ‘(summation-)stable’ distributions (Embrechts, Klüppelberg, and Mikosch, 2003, pp. 522-526, Rachev and Mittnik, 2000, Section 7.1). Here the prefix ‘summation’ is usually ignored.

Every elliptical random vector X with characteristic function

$$t \longmapsto \varphi_{\text{sub}}(t; \alpha) = \exp(it' \mu) \cdot \exp\left(-\left(\frac{1}{2} \cdot t' \Sigma t\right)^{\alpha/2}\right), \quad 0 < \alpha \leq 2,$$

is stable. If $\alpha \in]0, 2[$ then X is called ‘sub-Gaussian α -stable’. The parameter α is called ‘characteristic exponent’ or ‘index of stability’ of X (Mikosch, 2003, p. 45). For $\alpha = 2$ we obtain the multivariate Gaussian distribution whereas for $\alpha = 1$ the random vector X is multivariate (non-isotropic) Cauchy-distributed (Embrechts, Klüppelberg, and Mikosch, 2003, p. 72).

The following theorem states that under some weak conditions even the sum of a series of *dependent* elliptical random vectors is elliptical, too.

Theorem 9 (Hult and Lindskog, 2001) *Let X_1 and X_2 be two d -dimensional elliptically distributed random vectors with stochastic representation*

$$X_1 \stackrel{d}{=} \mu_1 + \mathcal{R}_1 \Lambda U_1^{(k)}$$

and

$$X_2 \stackrel{d}{=} \mu_2 + \mathcal{R}_2 \Lambda U_2^{(k)},$$

respectively. Here $(\mathcal{R}_1, \mathcal{R}_2)$, $U_1^{(k)}$ and $U_2^{(k)}$ are mutually independent whereas \mathcal{R}_1 and \mathcal{R}_2 may depend on each other. Then $X_1 + X_2$ is also elliptically distributed with location vector $\mu_1 + \mu_2$ and dispersion matrix $\Sigma = \Lambda \Lambda'$.

Proof. Hult and Lindskog, 2002. ■

This closure property is very useful for time series analysis when assuming a sequence $\mathcal{R}_1, \mathcal{R}_2, \dots$ of dependent (e.g. heteroscedastic) generating variates. This point will be addressed in Section 7.1.2.

1.3.2 Infinite Divisibility

In the preceding section it was shown that the sum of independent elliptical random vectors is elliptical, too, provided every component has the same dispersion matrix Σ . For the modeling of financial time series some authors (see, e.g., Bingham and Kiesel, 2002 and Eberlein and Keller, 1995) demand also the property of ‘infinite divisibility’.

In empirical finance usually one investigates the ‘log-price process’ of several assets, i.e. $Y_t := (\log P_t)_{t \in \mathbb{S}}$ where \mathbb{S} is an arbitrary index set and P_t represents the price vector of the considered assets at time t . Let $\mathbb{S} = \mathbb{R}_+$ with $Y_0 =_{\text{a.s.}} 1$ and consider the log-price vector $Y_1 =: Y$ at time $t = 1$ (w.l.o.g.). Now, one may assume that Y can always be ‘decomposed stochastically’ by an arbitrary number of i.i.d. increments (namely the asset’s ‘log-returns’), i.e.

$$Y \stackrel{d}{=} \sum_{t=1}^n X_{t/n}^{(n)}, \quad \forall n \in \mathbb{N}.$$

Note that the c.d.f. of each addend depends essentially on n . This property is known as ‘infinite divisibility’ (Bingham and Kiesel, 2002, Eberlein and Hammerstein, 2003, and Embrechts, Klüppelberg, and Mikosch, 2003, p. 81). It can be interpreted as the assumption

that the dynamics of stock prices results from continuously evolving but independent informations over time. This is of particular interest for the modeling of financial time series by Lévy processes (Barndorff-Nielsen and Shephard, 2001).

An elliptical random vector Y (and its corresponding distribution function) is infinitely divisible if for every $n \in \mathbb{N}$ there exists a random vector $X^{(n)}$ such that $\varphi_Y = \varphi_{X^{(n)}}^n$. Indeed, there are some infinitely divisible elliptical distributions. For instance, both the Gaussian and the sub-Gaussian α -stable distributions belong to the class of infinitely divisible distributions (Embrechts, Klüppelberg, and Mikosch, 2003, p. 81). This is because for $0 < \alpha \leq 2$ the sub-Gaussian α -stable characteristic generator satisfies

$$\begin{aligned} s \longmapsto \phi_{\text{sub}}(s; \alpha) &= \exp \left(- \left(\frac{1}{2} \cdot s \right)^{\alpha/2} \right) \\ &= \left(\exp \left(- \left(\frac{1}{2} \cdot \frac{s}{n^{2/\alpha}} \right)^{\alpha/2} \right) \right)^n = \phi_{\text{sub}}^n \left(\frac{s}{n^{2/\alpha}}; \alpha \right), \end{aligned}$$

i.e. each sub-Gaussian α -stable random vector with location vector μ and dispersion matrix Σ can be divided into an arbitrary number of ‘smaller’ sub-Gaussian α -stable random vectors with location vector μ/n and dispersion matrix $\Sigma/n^{2/\alpha}$. For $t := 1/n$ one obtains $\mu(t) = \mu t$ and $\sqrt{\Sigma}(t) = \sqrt{\Sigma} t^{1/\alpha}$, i.e. location is proportional to time but $\sqrt{\Sigma}(t) \propto t^{1/\alpha}$. Indeed, short-term financial data often seem to have a ‘scaling exponent’ of $1/\alpha > 0.5$, i.e. the normal distribution hypothesis becomes less appropriate the higher the frequency of the data (Bouchaud, Cont, and Potters, 1998, Breymann, Dias, and Embrechts, 2003).

1.3.3 Self-decomposability

Suppose that (X_t) is a simple autoregressive process, i.e.

$$X_t = \rho X_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots,$$

where each ε_t is white noise independent of X_{t-1} . If (X_t) is stationary (i.e. $|\rho| < 1$) then

$$X_t \stackrel{d}{=} X_{t+1} \stackrel{d}{=} \rho X_t + \varepsilon_{t+1}, \quad t = 1, 2, \dots$$

If a random vector X can be represented stochastically by

$$X \stackrel{d}{=} \rho X + \varepsilon^{(\rho)}, \quad \forall \rho \in]0, 1[,$$

where $\varepsilon^{(\rho)}$ is independent of X then it is called ‘self-decomposable’ (Barndorff-Nielsen and Shephard, 2003, Section 1.2.2). Hence, a random vector X is self-decomposable if its characteristic function satisfies the property $t \mapsto \varphi_X(t) = \varphi_X(\rho t) \varphi^{(\rho)}(t)$, $\forall \rho \in]0, 1[$, where $\varphi^{(\rho)}$ denotes the characteristic function of $\varepsilon^{(\rho)}$ which is considered as white noise. Note that $\varphi^{(\rho)}$ depends essentially on the parameter ρ . The larger ρ the smaller the contribution of the white noise and vice versa. Any self-decomposable law is infinitely divisible (Barndorff-Nielsen and Shephard, 2003, p. 13).

Now, consider again the characteristic generator of the sub-Gaussian α -stable distribution. Since for every $s \geq 0$,

$$\begin{aligned} \phi_{\text{sub}}(s; \alpha) &= \exp \left(- \left(\frac{\rho^2}{2} \cdot s \right)^{\alpha/2} \right) \cdot \exp \left(\left(\frac{\rho^2}{2} \cdot s \right)^{\alpha/2} - \left(\frac{1}{2} \cdot s \right)^{\alpha/2} \right) \\ &= \phi_{\text{sub}}(\rho^2 s; \alpha) \cdot \exp \left((\rho^\alpha - 1) \cdot \left(\frac{1}{2} \cdot s \right)^{\alpha/2} \right) \end{aligned}$$

$$\begin{aligned}
&= \phi_{\text{sub}}(\rho^2 s; \alpha) \cdot \exp \left(- \left(\frac{(1 - \rho^\alpha)^{2/\alpha}}{2} \cdot s \right)^{\alpha/2} \right) \\
&= \phi_{\text{sub}}(\rho^2 s; \alpha) \cdot \phi_{\text{sub}}((1 - \rho^\alpha)^{2/\alpha} s; \alpha),
\end{aligned}$$

any sub-Gaussian α -stable random vector is self-decomposable. More precisely, if X is sub-Gaussian α -stable with dispersion matrix Σ then

$$X \stackrel{d}{=} \rho X + (1 - \rho^\alpha)^{1/\alpha} \varepsilon, \quad \forall \rho \in]0, 1[,$$

where the white noise $\varepsilon^{(\rho)} = (1 - \rho^\alpha)^{1/\alpha} \varepsilon$ is also sub-Gaussian α -stable possessing the same dispersion matrix as X .

Not only the Gaussian and sub-Gaussian α -stable distributions are self-decomposable (and thus infinitely divisible) but also the family of symmetric generalized hyperbolic distributions (Barndorff-Nielsen, Kent, and Sørensen, 1982, Bingham and Kiesel, 2002). This particular elliptical distribution family is extremely useful for the modeling of financial data and will be discussed in more detail in Section 3.1.

Chapter 2

Extreme Values and Dependence Structures

This chapter concentrates on univariate and multivariate extremes with special focus on elliptical distributions. For a more comprehensive treatment of extreme value theory see, e.g., Bingham, Goldie, and Teugels, 1987, Coles, 2001, Embrechts, Klüppelberg, and Mikosch, 2003, Mikosch, 1999, and Resnick, 1987.

2.1 Univariate Extremes

The probability distribution of extremal events is a priori of main interest for insurance and finance. Extreme value theory (EVT) is a special topic of probability theory and has become standard in risk theory and management (see, e.g., Embrechts, Klüppelberg, and Mikosch, 2003). In insurance it has been used e.g. to calculate the potential severity of losses caused by natural disasters like earthquakes, hurricanes, floods, etc. (Haan, 1990, McNeil, 1997, McNeil and Saladin, 2000, Resnick, 1997, as well as Rootzén and Tajvidi, 1997). Calculating the value-at-risk of asset portfolios on high confidence-levels has retained as a typical finance application of the theory of extreme values (Danielsson and Vries, 2000, Frahm, 1999, and Këllezi and Gilli, 2003).

The fundamental theorem of Fisher-Tippett (Embrechts, Klüppelberg, and Mikosch, 1997, p. 121 in connection with p. 152) preludes the transition from classical statistics to EVT.

Theorem 10 (Fisher and Tippett, 1928) *Let X_1, \dots, X_n ($n = 1, 2, \dots$) be sequences of i.i.d. random variables and $M_n := \max\{X_1, \dots, X_n\}$ the corresponding sample maximum. If there exist norming constants $a_n > 0$, $b_n \in \mathbb{R}$ and a non-degenerate c.d.f. H such that*

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} H, \quad n \longrightarrow \infty,$$

then there exist parameters $\sigma > 0$ and $\mu, \xi \in \mathbb{R}$ such that

$$H(x) = H_\xi\left(\frac{x - \mu}{\sigma}\right) = \begin{cases} \exp\left(-\left(1 + \xi \cdot \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}}\right), & \xi \neq 0, \\ \exp\left(-\exp\left(-\frac{x - \mu}{\sigma}\right)\right), & \xi = 0, \end{cases}$$

with support

$$\mathcal{D}_{\xi, \sigma, \mu}^{GEV} \equiv \begin{cases} x > \mu - \frac{\sigma}{\xi}, & \xi > 0, \\ x \in \mathbb{R}, & \xi = 0, \\ x < \mu - \frac{\sigma}{\xi}, & \xi < 0. \end{cases}$$

The limit law H is referred to as ‘generalized extreme value distribution’ (GEV).

Proof. Resnick, 1997, pp. 9-10, using the Jenkinson-von-Mises representation of the extreme value distributions (Embrechts, Klüppelberg, and Mikosch, 1997, p. 152). ■

The limit theorem of Fisher-Tippett can be interpreted as a ‘sample-maxima analogue’ to the classical central limit theorem which is based on sample sums (Embrechts, Klüppelberg, and Mikosch, 1997, p. 120). But rather than the probabilistic property of a sum of i.i.d. variables the stochastic behavior of the maximum plays the key role when investigating the tail of a distribution.

Definition 3 (MDA) *The c.d.f. F of X (or roughly speaking the random variable X) belongs to the maximum domain of attraction (MDA) of the GEV H_ξ if the conditions of the Fisher-Tippett theorem hold for F (for X). We write $F \in \text{MDA}(H_\xi)$ or $X \in \text{MDA}(H_\xi)$, respectively.*

The parameter ξ is constant under affine transformations so that both the scale parameter σ and the location parameter μ can be neglected.

Theorem 11 (Tail behavior of $\text{MDA}(H_0)$) *Let F be a c.d.f. with right endpoint $x_F \leq \infty$ and $\bar{F} := 1 - F$ its survival function. $F \in \text{MDA}(H_0)$ if and only if a constant $v < x_F$ exists, such that*

$$\bar{F}(x) = \gamma(x) \cdot \exp\left(-\int_v^x \frac{1}{f(t)} dt\right), \quad v < x < x_F,$$

where γ is a measurable function with $\gamma(x) \rightarrow \gamma > 0$, $x \nearrow x_F$, and f is a positive, absolutely continuous function with $f'(x) \rightarrow 0$, $x \nearrow x_F$.

Proof. Resnick (1987), Proposition 1.4 and Corollary 1.7. ■

Theorem 11 implies that every tail of a distribution $F \in \text{MDA}(H_0)$ with $x_F = \infty$ may be approximated by the exponential law

$$\bar{F}(x) \approx \gamma(v) \cdot \exp\left(-\frac{x-v}{f(v)}\right), \quad \forall x > v,$$

provided v is a sufficiently high threshold. We say that the distribution F is ‘light-’, ‘thin-’, or ‘exponential-tailed’. This is given e.g. for the normal-, lognormal-, exponential-, and the gamma-distribution. The following theorem can be found in Embrechts, Klüppelberg, and Mikosch (2003, p. 131).

Theorem 12 (Tail behavior of $\text{MDA}(H_{\xi>0})$) *$F \in \text{MDA}(H_{\xi>0})$ if and only if*

$$\bar{F}(x) = \lambda(x) \cdot x^{-\frac{1}{\xi}}, \quad x > 0,$$

where λ is a slowly varying function, i.e. λ is a measurable function $\mathbb{R}^+ \rightarrow \mathbb{R}^+$ satisfying

$$\frac{\lambda(tx)}{\lambda(x)} \rightarrow 1, \quad x \rightarrow \infty, \quad \forall t > 0.$$

Proof. Embrechts, Klüppelberg, and Mikosch, 2003, pp. 131-132. ■

In the following $\alpha := 1/\xi$ is called the ‘tail index’ of F . For $\xi = 0$ we define $\alpha = \infty$. A measurable function $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ with

$$\frac{f(tx)}{f(x)} \rightarrow t^{-\alpha}, \quad x \rightarrow \infty, \quad \forall t > 0,$$

is called ‘regularly varying (at ∞) with tail index $\alpha \geq 0$ ’ (Mikosch, 1999, p. 7). Note that a slowly varying function is regularly varying with tail index $\alpha = 0$.

A survival function \bar{F} is regularly varying if and only if $F \in \text{MDA}(H_{\xi>0})$ (Resnick, 1987, p. 13). Thus a regularly varying survival function exhibits a power law, that is to say

$$\bar{F}(x) \approx \lambda(v) \cdot x^{-\frac{1}{\xi}} = \lambda(v) \cdot x^{-\alpha}, \quad \xi, \alpha > 0, \quad \forall x > v > 0,$$

for a sufficiently high threshold v . Now the c.d.f. F is said to be ‘heavy-’, ‘fat-’, or ‘power-tailed’. This is the case e.g. for the Pareto-, Burr-, loggamma-, Cauchy-, and Student’s t -distribution. Clearly a power tail converges slower to zero than an exponential tail and therefore this class is of special interest to risk theory.

A random variable X is said to be regularly varying with tail index $\alpha > 0$ if both $X \in \text{MDA}(H_{1/\alpha})$ and $-X \in \text{MDA}(H_{1/\alpha})$ (Mikosch, 2003, p. 23). Then the tail index α has a nice property: X has no finite moment of orders larger than α . Conversely, for $\alpha = \infty$ every moment exists and is finite. Hence the smaller the tail index the bigger the weight of the tail. Therewith both ξ and α are well suited to characterize the tailedness of the underlying distribution (Embrechts, Klüppelberg, and Mikosch, 1997, p. 152):

$$\begin{aligned} \xi > 0 & \Leftrightarrow 0 < \alpha < \infty : & \text{Fréchet class,} \\ \xi = 0 & \Leftrightarrow \alpha = \infty : & \text{Gumbel class,} \\ \xi < 0 & \Leftrightarrow -\infty < \alpha < 0 : & \text{Weibull class.} \end{aligned}$$

Definition 4 (GPD) *The distribution function*

$$G_{\xi} \left(\frac{x - \mu}{\sigma} \right) = \begin{cases} 1 - (1 + \xi \cdot \frac{x - \mu}{\sigma})^{-\frac{1}{\xi}}, & \xi \neq 0, \\ 1 - \exp(-\frac{x - \mu}{\sigma}), & \xi = 0, \end{cases} \quad \sigma > 0,$$

with support

$$\mathcal{D}_{\xi, \mu, \sigma}^{GPD} \equiv \begin{cases} x \geq \mu, & \xi \geq 0, \\ \mu \leq x < \mu - \frac{\sigma}{\xi}, & \xi < 0, \end{cases}$$

is referred to as ‘generalized Pareto distribution’ (GPD).

The second crucial theorem of EVT is the following one.

Theorem 13 (Pickands, 1975) $F \in \text{MDA}(H_{\xi \in \mathbb{R}})$ if and only if a positive function σ exists with

$$\lim_{v \nearrow x_F} \sup_{0 \leq y < x_F - v} \left| F^{v \rightarrow}(y) - G_{\xi} \left(\frac{y}{\sigma(v)} \right) \right| = 0,$$

where

$$F^{v \rightarrow}(y) := P(X - v \leq y \mid X > v).$$

Proof. Embrechts, Klüppelberg, and Mikosch, 2003, pp. 165-166. ■

The c.d.f. $F^{v \rightarrow}$ is referred to as ‘excess distribution (over the threshold v)’. Thus the GPD is the limit distribution of an arbitrary excess distribution $F^{v \rightarrow}$ provided its threshold v is sufficiently high. Theorem 13 implies that the excess distribution depends asymptotically only on two parameters, a scale parameter σ and a shape parameter ξ , respectively. The scale parameter is determined by the selected threshold whereas the shape parameter is constant through all thresholds because it only depends on the MDA of F . So the probability of extremal events may be quantified without knowing the underlying distribution family and

the corresponding parameters may be estimated in a semi-parametric fashion allowing for heavy tails.

The advantage of non-parametric methods is that no information about the distribution law is required and unbiased estimates can be attained. That is to say the resulting estimates are robust regarding an error in the interpretation of the true law. The trade off is large variance of the estimates. If the true law were known clearly the corresponding parametric approach would be more efficient. But biased estimates appear if the predicted distribution law is wrong, i.e. parametric methods bear an awkward model risk. With semi-parametric methods one can take advantage of the particular attributes of both, the parametric and the non-parametric approaches. So EVT is a proper compromise between bias (parametric methods) and variance (non-parametric methods) provided the intention to quantify the probability of extremal events.

Several empirical studies (see, e.g., Haan, 1990, McNeil, 1997, and Rootzén and Tajvidi, 1997) in the context of insurance show that loss distributions belong significantly to the Fréchet class. So classical loss distributions like gamma and lognormal are not suitable for modeling extreme losses. There is also a huge number of empirical studies showing that log-returns of financial assets usually are heavy tailed, too (see, e.g., Eberlein and Keller, 1995, Embrechts, Klüppelberg, and Mikosch, 2003, Section 7.6, Fama, 1965, Mandelbrot, 1963, and Mikosch, 2003, Chapter 1). From our point of view we are interested in the class of distributions which can be properly applied for modeling the generating variate of elliptically distributed log-returns. Especially it is worth to know the consequences regarding the dependence structure of elliptical distributions.

2.2 Multivariate Extremes and Copulas

In the preceding section regularly varying random variables were discussed. The concept of regular variation can be extended to the multivariate case. The following definitions can be found in Hult and Lindskog (2002), Mikosch (1999), Resnick (1987), and Schmidt (2003a).

Definition 5 (Multivariate regular variation) *A random vector X is said to be ‘regularly varying’ with tail index $\alpha \geq 0$ if there exists a random vector S distributed on the unit hypersphere $\mathcal{S}_{\|\cdot\|}^{d-1}$ such that for any $t > 0$ and any Borel-set $B \subset \mathcal{S}_{\|\cdot\|}^{d-1}$,*

$$\frac{P\left(\|X\| > tx, \frac{X}{\|X\|} \in B\right)}{P(\|X\| > x)} \xrightarrow{v} P_S(B) \cdot t^{-\alpha}, \quad x \longrightarrow \infty. \quad (2.1)$$

Here $\|\cdot\|$ denotes an arbitrary norm on \mathbb{R}^d and

$$\mathcal{S}_{\|\cdot\|}^{d-1} := \{x \in \mathbb{R}^d : \|x\| = 1\}.$$

Further, $P_S(B) := P(S \in B)$ and v symbolizes vague convergence (Resnick, 1987, p. 140). The probability measure P_S is called the ‘spectral measure’ of X .

Vague convergence is equivalent to the usual convergence given some additional (but relatively weak) topological conditions for the considered Borel-sets (cf. Mikosch, 1999, p. 31 and 2003, p. 25, Resnick, 1987, p. 140, Hult and Lindskog, 2002, and Schmidt, 2003a, p. 28). The definition of multivariate regular variation indeed covers also the univariate case. Apart from (2.1) there are other equivalent definitions of multivariate regular variation (cf. Mikosch, 1999, p. 32). Note that the choice of the norm $\|\cdot\|$ does not matter because if a random vector is regularly varying with respect to a specific norm then it is also regularly varying with respect to any other norm (Hult and Lindskog, 2002). But the spectral measure

P_S indeed depends on the choice of the norm. Also, even if the norm is fixed the spectral measure may depend on the tail index α .

If $B = \mathcal{S}_{\|\cdot\|}^{d-1}$ then

$$\frac{P(\|X\| > tx)}{P(\|X\| > x)} \longrightarrow t^{-\alpha}, \quad x \longrightarrow \infty, \quad \forall t > 0.$$

That is to say a regularly varying random vector X exhibits a power tail in the sense of the corresponding vector norm. If $t = 1$ then

$$P\left(\frac{X}{\|X\|} \in B \mid \|X\| > x\right) \xrightarrow{v} P_S(B), \quad x \longrightarrow \infty.$$

Hence the ‘direction’ of X under its excess distribution, i.e. $X/\|X\|$ under the condition $\|X\| > x$ (x large), is distributed according to its spectral measure, approximately. Note that both the spectral measure and the power law are linked multiplicatively in Eq. 2.1. Thus if X is regularly varying the two events ‘direction of X ’ and ‘length of X ’ are asymptotically independent.

The following theorem is concerned with the regular variation of elliptically distributed random vectors.

Theorem 14 (Hult and Lindskog, 2001) *Let $X \stackrel{d}{=} \mu + \mathcal{R}\Lambda U^{(d)} \sim \mathcal{E}_d(\mu, \Sigma, \phi)$ where $\Sigma = \Lambda\Lambda'$ is positive definite. Further, let $F_{\mathcal{R}}$ be the generating distribution function of X . Then $F_{\mathcal{R}} \in \text{MDA}(H_{\xi>0})$ if and only if X is regularly varying with tail index $\alpha = 1/\xi > 0$.*

Proof. Consider $\Sigma = [\sigma_{ij}]$ and define $\sigma_i := \sqrt{\sigma_{ii}}$ for all $i = 1, \dots, d$. Due to the positive definiteness of Σ it is easy to show that $\sigma_i^2 > 0$, $i = 1, \dots, d$ and also $|\sigma_{ij}| \neq \sigma_i\sigma_j$, i.e. $|\rho_{ij}| < 1$ for any i and j with $i \neq j$. As stated by Theorem 4.3 of Hult and Lindskog (2002) this is sufficient for the assertion above. ■

Hence, the choice of the generating variate \mathcal{R} determines essentially the extremal behavior of the corresponding elliptical random vector. Due to Definition 5 the radial part ($\|X\|$) of a regularly varying elliptical random vector X is asymptotically independent of its angular part ($X/\|X\|$) under the condition that the radial part has an extreme outcome. Hence there is a sort of dependence between the components of X which cannot be explained only by linear dependence. This sort of dependence is referred to as ‘extremal’ or ‘asymptotic’ dependence.

Note that the components of a random vector $X \sim \mathcal{E}_d(\mu, \Sigma, \phi)$ are stochastically independent if and only if X has a multivariate normal distribution and its correlation matrix corresponds to the identity matrix (Fang, Kotz, and Ng, 1990, p. 106). Hence, even if a random vector is spherically distributed and not regularly varying another sort of nonlinear dependence is given, anyway.

Example 10 (Nonlinear dependence of $U \sim \mathcal{S}$) *Suppose $U = (U_1, U_2)$ is uniformly distributed on the unit circle. Then*

$$U_2 \stackrel{a.s.}{=} \pm \sqrt{1 - U_1^2},$$

i.e. U_2 depends strongly on U_1 .

Now, the question is if there exists a multivariate analogue to the Fisher-Tippett theorem in the univariate case. More precisely, is it possible to extend the concept of maximum domains of attraction to the multi-dimensional setting?

Definition 6 (Multivariate domain of attraction) Let X_1, \dots, X_n ($n = 1, 2, \dots$) be sequences of i.i.d. random vectors with common c.d.f. F and $M_{in} := \max \{X_{i1}, \dots, X_{in}\}$ the corresponding sample maximum of the i -th component. The c.d.f. F (or roughly speaking the corresponding random vector X) is said to be ‘in the domain of attraction of a multivariate extreme value distribution H ’ if there exist norming constants $a_{in} > 0$, $b_{in} \in \mathbb{R}$ ($i = 1, \dots, d$) and H has non-degenerate margins such that

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} H, \quad n \longrightarrow \infty,$$

where $(M_n - b_n)/a_n$ corresponds to the random vector $[(M_{in} - b_{in})/a_{in}]$.

For the definition see, e.g., Resnick (1987, p. 263). Since the componentwise maxima are invariant under strictly increasing transformations it is allowed to choose alternative representations of the marginal distribution functions. Coles (2001), for instance, refers to the bivariate case and assumes that each component of X has a standard Fréchet distribution, i.e. $F_i(x_i) = \exp(-1/x_i)$, $x_i > 0$, $i = 1, \dots, d$ (Coles, 2001, p. 144). This is convenient for his consecutive analysis of extremes.

Theorem 15 (Schmidt, 2003) Let $X \stackrel{d}{=} \mu + \mathcal{R}\Lambda U^{(d)} \sim \mathcal{E}_d(\mu, \Sigma, \phi)$ where $\Sigma = \Lambda\Lambda'$ is positive definite. Further, let $F_{\mathcal{R}}$ be the generating distribution function of X . If $F_{\mathcal{R}} \in \text{MDA}(H_{\xi>0})$ then X is in the domain of attraction of a multivariate extreme value distribution.

Proof. Schmidt (2003a). ■

Resnick (1987, p. 281) states that every regularly varying random vector is in the domain of attraction of an extreme value distribution H with

$$x \longmapsto H(x) = \exp\left(-v\left(\overline{[-\infty, x]}\right)\right),$$

where $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ and $\overline{[-\infty, x]}$ is the complement set of $[-\infty, x_1] \times \dots \times [-\infty, x_d]$, $x_1, \dots, x_d > 0$. Hence Theorem 15 follows also by Theorem 14. Here, v is a measure (called ‘exponent measure’) with the following property:

$$v\left(\left\{x \in \mathbb{R}^d \setminus \{0\} : \|x\| > r, \frac{x}{\|x\|} \in B\right\}\right) = P_S(B) \cdot r^{-\alpha}, \quad r > 0,$$

for any Borel-set $B \subset \mathcal{S}_{\|\cdot\|}^{d-1}$ (cf. Resnick, 1987, p. 281 and Schmidt, 2003a). For elliptical distributions the spectral measure P_S depends on the dispersion matrix. Hult and Lindskog (2002) give an analytic expression for bivariate elliptical distributions. Unfortunately, the exponent measure cannot be more specified for arbitrary regularly varying random vectors. Thus a closed form representation of the extreme value distribution as in the univariate case does not exist in the multivariate context.

For a better understanding of elliptical random vectors we have to take a closer look on their dependence structures. This can be done by the theory of copulas (Joe, 1997, Drouot Mari and Kotz, 2001, and Nelsen, 1999). An axiomatic definition of copulas can be found in Nelsen (1999, Section 2.2 and 2.10), for instance. According to this definition a copula is a d -variate distribution function

$$C : [0, 1]^d \rightarrow [0, 1].$$

Owing to our interest in copula families we have to study copulas generated by specific classes of distributions as follows:

Definition 7 (Copula of a random vector X) Let $X = (X_1, \dots, X_d)$ be a random vector with multivariate c.d.f. F and continuous margins F_1, \dots, F_d . The copula of X (or of the c.d.f. F , respectively) is the multivariate c.d.f. C of the random vector

$$U := (F_1(X_1), \dots, F_d(X_d)).$$

Due to the continuity of the margins F_1, \dots, F_d every random variable $F_i(X_i) = U_i$ is standard uniformly distributed, i.e. $U_i \sim U(0, 1)$. Thus the copula of a continuous random vector X has uniform margins and

$$C(u_1, \dots, u_d) = F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)), \quad \forall u = (u_1, \dots, u_d) \in]0, 1[^d \quad (2.2)$$

holds, where

$$F_i^{\leftarrow}(u_i) := \inf \{x : F_i(x) \geq u_i\}, \quad u_i \in]0, 1[, \quad i = 1, \dots, d$$

are the marginal quantile functions.

Theorem 16 (Sklar, 1959) Let F be a d -variate c.d.f. with margins F_1, \dots, F_d . Then there exists a copula C such that

$$x \mapsto F(x) = C(F_1(x_1), \dots, F_d(x_d)), \quad \forall x = (x_1, \dots, x_d) \in \overline{\mathbb{R}}^d, \quad (2.3)$$

where $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$. If all margins F_1, \dots, F_d are continuous then C is unique. Otherwise C is uniquely determined on the Cartesian product of the ranges of F_1, \dots, F_d . Conversely, if C is a copula and F_1, \dots, F_d are some univariate distribution functions then F given by Eq. 2.3 is a d -variate c.d.f. with marginal distribution functions F_1, \dots, F_d .

Proof. Nelsen, 1999, p. 41 and the corresponding references. ■

That is to say by the ‘marginal mapping’ only the dependence structure, i.e. the copula is extracted from F . This can be used for ‘coupling’ with another set of arbitrary marginal distribution functions in order to obtain a new multivariate c.d.f. but with the same dependence structure. This is the matter why we speak about a ‘copula’ (Nelsen, 1999, p. 2 and p. 15).

In general the multivariate c.d.f. F contains parameters that do not affect the copula of F , and other parameters affects the copula and possibly the margins. The latter type of parameters are called ‘copula parameters’. Let θ be a parameter vector $(\theta_1, \dots, \theta_m) \in \mathbb{R}^m$ and $F(\cdot; \theta)$ a continuous multivariate c.d.f. with copula $C(\cdot; \theta)$. Let $I_C \subset I = \{1, \dots, m\}$ be an index-set that contains all i for which at least one $u \in]0, 1[^d$ exists, such that

$$\frac{\partial C(u; \theta)}{\partial \theta_i} \neq 0.$$

So I_C contains all copula parameter indices.

Lemma 17 Suppose a distribution family is generated by a d -variate c.d.f. $F^*(\cdot; \theta_0)$ with continuous margins $F_1^*(\cdot; \theta_0), \dots, F_d^*(\cdot; \theta_0)$ and d continuous and strictly increasing marginal transformations $h_1(\cdot; \theta_1), \dots, h_d(\cdot; \theta_d)$, where the parameters $\theta_1, \dots, \theta_d$ may be some real-valued vectors:

$$F(x_1, \dots, x_d; \theta) = F^*(h_1(x_1; \theta_1), \dots, h_d(x_d; \theta_d); \theta_0), \quad (2.4)$$

with

$$\theta = (\theta_0, \theta_1, \dots, \theta_d).$$

Then only the vector θ_0 contains copula parameters.

Proof. The lemma follows from the fact that any copula is invariant under continuous and strictly increasing transformations $h_1(\cdot; \theta_1), \dots, h_d(\cdot; \theta_d)$. Thus also the parameters $\theta_1, \dots, \theta_d$ cannot affect the copula. ■

So the parameters $\theta_1, \dots, \theta_d$ are canceled down through copula separation and only θ_0 remains. We call the c.d.f. $F^*(\cdot; \theta_0)$ the ‘underlying distribution’ of the corresponding copula $C(\cdot; \theta_0)$. In particular, the copula of an elliptical distribution function will be called ‘elliptical copula’.

Let h_1, \dots, h_d be some continuous and strictly increasing functions. If $Y \sim F$, where F is a multivariate elliptical c.d.f., then $x \mapsto G(x) := F(h_1(x_1), \dots, h_d(x_d))$ is the multivariate c.d.f. of the random vector $X := (h_1^{-1}(Y_1), \dots, h_d^{-1}(Y_d))$. This can be used for modeling new distribution functions based on the class of elliptical distributions. Conversely, if X is arbitrarily distributed, nevertheless one possibly finds some strictly increasing functions h_1, \dots, h_d such that the random vector $h(X) := (h_1(X_1), \dots, h_d(X_d))$ is elliptically distributed with multivariate c.d.f. F . Then the multivariate c.d.f. of X is given by $x \mapsto G(x) = F(h(x))$. This can be used for pre-processing, i.e. estimation techniques for elliptical distributions can be applied on the transformed data $h(x_1), h(x_2), \dots, h(x_n)$ of a sample x_1, x_2, \dots, x_n so as to obtain \hat{F} and thus $\hat{G} = \hat{F} \circ h$.

Suppose that both the marginal distribution functions G_1, \dots, G_d of X and the marginal distribution functions F_1, \dots, F_d of Y are continuous and strictly increasing. Further suppose that the copula C of X is generated by Y , i.e.

$$P(G_1(X_1) \leq u_1, \dots, G_d(X_d) \leq u_d) = P(F_1(Y_1) \leq u_1, \dots, F_d(Y_d) \leq u_d)$$

for all $u = (u_1, \dots, u_d) \in]0, 1]^d$. According to Eq. 2.2 we obtain

$$C(u) = G(G_1^{\leftarrow}(u_1), \dots, G_d^{\leftarrow}(u_d)),$$

i.e.

$$C(G_1(x_1), \dots, G_d(x_d)) = G(G_1^{\leftarrow}(G_1(x_1)), \dots, G_d^{\leftarrow}(G_d(x_d))) = G(x).$$

Since

$$\begin{aligned} x \mapsto G(x) &= C(G_1(x_1), \dots, G_d(x_d)) \\ &= P(G_1(X_1) \leq G_1(x_1), \dots, G_d(X_d) \leq G_d(x_d)) \\ &= P(F_1(Y_1) \leq G_1(x_1), \dots, F_d(Y_d) \leq G_d(x_d)) \\ &= P(Y_1 \leq F_1^{\leftarrow}(G_1(x_1)), \dots, Y_d \leq F_d^{\leftarrow}(G_d(x_d))) \\ &= F(F_1^{\leftarrow}(G_1(x_1)), \dots, F_d^{\leftarrow}(G_d(x_d))), \end{aligned}$$

the corresponding transformations h_1, \dots, h_d are given by $h_i = F_i^{\leftarrow} \circ G_i$, $i = 1, \dots, d$.

Definition 8 (Meta-elliptical distribution) A random vector $X = (X_1, \dots, X_d)$ (or its corresponding multivariate c.d.f. G) is said to be ‘meta-elliptically distributed’ if the copula of X is elliptical, i.e. if there exists a random vector $Y \sim \mathcal{E}_d(\mu, \Sigma, \phi)$ with c.d.f. F such that

$$P(G_1(X_1) \leq u_1, \dots, G_d(X_d) \leq u_d) = P(F_1(Y_1) \leq u_1, \dots, F_d(Y_d) \leq u_d),$$

for all $u \in [0, 1]^d$. This is denoted by $X \sim \mathcal{ME}_d(\mu, \Sigma, \phi)$.

A treatment on meta-elliptical distributions can be found in Abdous, Genest, and Rémillard (2004) as well as in Fang, Fang, and Kotz (2002). See also Embrechts, Frey, and McNeil (2004, pp. 89-90) for a discussion of meta-Gaussian and meta- t -distributions. Hence, estimation procedures for elliptical distributions can be applied even if the observed data is not elliptically but meta-elliptically distributed provided the transformations h_1, \dots, h_d are known.

2.3 Asymptotic Dependence of Meta-elliptical Distributions

In the following section some common properties of the dependence structures of meta-elliptically distributed random vectors are examined with special focus on asymptotic dependence. Even though the following statements concerning measures for asymptotic dependence refer to elliptically distributed random vectors the results can be easily extended to the class of meta-elliptical distributions. This is because the considered measures act only on the copula and thus it is sufficient to provide elliptical copulas, i.e. meta-elliptical distributions.

2.3.1 Bivariate Asymptotic Dependence

Affine marginal transformations are often applied for constructing distribution families, more precisely location-scale-families. The location-scale-family generated by the multivariate distribution F^* contains all distributions

$$(x_1, \dots, x_d) \longmapsto F(x_1, \dots, x_d; \theta) = F^*\left(\frac{x_1 - \mu_1}{\sigma_1}, \dots, \frac{x_d - \mu_d}{\sigma_d}; \theta_0\right),$$

with given parameter vector θ_0 , variable location parameters μ_1, \dots, μ_d and scale parameters $\sigma_1, \dots, \sigma_d$. So this distribution family is generated by affine marginal transformations and the location and scale parameters are not copula parameters according to Lemma 17.

Let us turn towards the dependence structure in $F(\cdot; \theta)$. Kendall's τ is an appropriate dependence measure for bivariate monotonic dependence.

Definition 9 (Kendall's τ) *Let the bivariate random vector (\tilde{X}, \tilde{Y}) be an independent copy of (X, Y) . Kendall's τ of X and Y is defined as*

$$\tau(X, Y) := P\left(\left(\tilde{X} - X\right)\left(\tilde{Y} - Y\right) > 0\right) - P\left(\left(\tilde{X} - X\right)\left(\tilde{Y} - Y\right) < 0\right).$$

Kendall's τ is a rank correlation, so

$$\tau(X, Y) = \tau(F_X(X), F_Y(Y))$$

holds, i.e. it is completely determined by the copula of (X, Y) and thus it depends only on the copula parameters of the c.d.f. of (X, Y) .

Now let

$$\Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1d} \\ \vdots & \ddots & \vdots \\ \sigma_{d1} & \cdots & \sigma_{dd} \end{bmatrix}, \quad \sigma := \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_d \end{bmatrix}, \quad \rho := \begin{bmatrix} \rho_{11} & \cdots & \rho_{1d} \\ \vdots & \ddots & \vdots \\ \rho_{d1} & \cdots & \rho_{dd} \end{bmatrix},$$

with

$$\sigma_i := \sqrt{\sigma_{ii}}, \quad i = 1, \dots, d,$$

and

$$\rho_{ij} := \frac{\sigma_{ij}}{\sigma_i \sigma_j}, \quad i, j = 1, \dots, d,$$

so that $\Sigma = \sigma \rho \sigma$ and $\varphi(\cdot; \mu, \Sigma, \vartheta) \equiv \varphi(\cdot; \mu, \sigma, \rho, \vartheta)$. A d -dimensional elliptical random vector with characteristic function $\varphi(\cdot; \mu, \sigma, \rho, \vartheta)$ is denoted by $X \sim \mathcal{E}_d(\mu, \sigma, \rho, \phi(\cdot; \vartheta))$. Especially, the d -variate t -distribution (cf. Example 4) with ν degrees of freedom will be

denoted by $t_d(\mu, \sigma, \rho, \nu)$ and the d -variate sub-Gaussian α -stable distribution (cf. Section 1.3.1) is symbolized by $N_d^{\text{sub}}(\mu, \sigma, \rho, \alpha)$. Note that ρ is only the correlation matrix of X if $E(\mathcal{R}^2) < \infty$ (cf. Section 1.2.3). Therefore ρ will be called ‘pseudo-correlation matrix’ (cf. Lindskog, 2000).

With the reparametrization above we obtain the equation $\Lambda = \sigma\sqrt{\rho}$, where $\sqrt{\rho} \in \mathbb{R}^{d \times d}$ with $\sqrt{\rho}\sqrt{\rho}' = \rho$, and thus

$$X \stackrel{d}{=} \mu + \sigma \mathcal{R}\sqrt{\rho}U^{(d)}.$$

Hence for studying the copulas of elliptical random vectors it is sufficient to analyze

$$X^* := \mathcal{R}\sqrt{\rho}U^{(d)},$$

or the corresponding characteristic generator $\phi(\cdot; 0, I_d, \rho, \vartheta)$.

Example 11 (Sub-Gaussian α -stable copula) *The density function of the random vector $X \sim N_d^{\text{sub}}(0, I_d, \rho, \alpha)$, i.e. the ‘standard density’ of a sub-Gaussian α -stable random vector, can be obtained through multivariate Fourier-transformation (Frahm, Junker, and Szimayer, 2003) and corresponds to*

$$\begin{aligned} f_{\alpha, \rho}^*(x) &= \frac{1}{(2\pi)^d} \cdot \int_{\mathbb{R}^d} \varphi_{\text{sub}}(t; 0, I_d, \rho, \alpha) \cdot \exp(-it'x) \, dt \\ &= \frac{1}{(2\pi)^d} \cdot \int_{\mathbb{R}^d} \exp\left(-\left(\frac{1}{2} \cdot t' \rho t\right)^{\alpha/2}\right) \cdot \cos(t'x) \, dt, \quad 0 < \alpha \leq 2. \end{aligned}$$

The copula generated by a d -variate sub-Gaussian α -stable distribution is

$$C_\alpha(u_1, \dots, u_d) = F_{\alpha, \rho}^*(F_\alpha^{*\leftarrow}(u_1), \dots, F_\alpha^{*\leftarrow}(u_d)),$$

where $F_{\alpha, \rho}^*$ is the multivariate standard distribution function

$$F_{\alpha, \rho}^*(x) := \int_{]-\infty, x]} f_{\alpha, \rho}^*(s) \, ds, \quad s \in \mathbb{R}^d,$$

with $]-\infty, x] :=]-\infty, x_1] \times \dots \times]-\infty, x_d]$, and $F_\alpha^{*\leftarrow}$ is the inverse of the univariate standard distribution function

$$F_\alpha^*(x) := \int_{-\infty}^x f_{\alpha, 1}^*(s) \, ds, \quad s \in \mathbb{R}.$$

For continuous elliptical distributions there is a straight link between Kendall’s τ and the matrix ρ (Lindskog, McNeil, and Schmock, 2003):

Theorem 18 (Lindskog, McNeil and Schmock, 2001) *Let $X \sim \mathcal{E}_d(\mu, \sigma, \rho, \phi)$, having continuous and non-degenerate components. For any two components of X , X_i and X_j , Kendall’s τ corresponds to*

$$\tau(X_i, X_j) = \frac{2}{\pi} \cdot \arcsin(\rho_{ij}). \quad (2.5)$$

Proof. Lindskog, McNeil, and Schmock, 2003. ■

That is to say Kendall’s τ depends only on ρ and neither the characteristic generator nor location and scale affect the rank correlation. This is due to the linear dependence structure of elliptical distributions. Note also that Kendall’s τ remains the same if X is not elliptically distributed but meta-elliptically distributed with the same copula parameter ρ .

In addition to bivariate monotonic dependence, which is measured by rank correlation, financial data usually is likely to exhibit bivariate lower asymptotic dependence (Frahm, Junker, and Schmidt, 2003 and Junker and May, 2002), that is to say a relatively large probability of extreme simultaneous losses.

Definition 10 (Tail dependence coefficient) *Let C be the copula of (X, Y) , whereas F_X is the marginal c.d.f. of X and F_Y is the marginal c.d.f. of Y , respectively. The lower tail dependence coefficient of X and Y is defined as*

$$\lambda_L(X, Y) := \lim_{t \searrow 0} P(F_Y(Y) \leq t \mid F_X(X) \leq t) = \lim_{t \searrow 0} \frac{C(t, t)}{t}, \quad (2.6)$$

whereas the upper tail dependence coefficient of X and Y is defined as

$$\lambda_U(X, Y) := \lim_{t \nearrow 1} P(F_Y(Y) > t \mid F_X(X) > t) = \lim_{t \nearrow 1} \frac{1 - 2t + C(t, t)}{1 - t},$$

provided the corresponding limits exist. If $\lambda_L(X, Y) > 0$ or $\lambda_U(X, Y) > 0$ the random vector (X, Y) (or the corresponding random components X and Y) is said to be ‘lower tail dependent’ or ‘upper tail dependent’, respectively.

Loosely speaking, this is the probability that the realization of a random variable is extremely negative (or positive) under the condition that the realization of another random variable is extremely negative (or positive), too. Note that in the elliptical framework the lower tail dependence coefficient equals to the upper tail dependence coefficient due to the radial symmetry. Since asymptotic dependence is defined by means of the copula, beside Kendall’s τ also λ_L and λ_U depend only on the copula parameters. Coherently, a dependence measure which is frequently used for *any* kind of distributions (like, e.g., Pearson’s correlation coefficient) should be invariant under marginal transformations. Unfortunately, the correlation coefficient does not have this property. An interesting investigation of possible mistakes due to ignoring this fact can be found in Embrechts, McNeil, and Straumann (2002).

Note that a sub-Gaussian α -stable random vector with $\alpha < 2$ is regularly varying with tail index α (Mikosch 2003, p. 45). Further, a multivariate t -distributed random vector with ν degrees of freedom ($\nu > 0$) is regularly varying with tail index ν (Mikosch 2003, p. 26). The following theorem connects the tail index with the tail dependence coefficient of elliptical distributions:

Theorem 19 (Schmidt, 2002) *Let $X \sim \mathcal{E}_d(\mu, \Sigma, \phi)$ be regularly varying with tail index $\alpha > 0$ and $\Sigma = \sigma\sqrt{\rho}$ a positive definite dispersion matrix where σ and ρ are defined as described in Section 2.3.1. Then every pair of components of X , say X_i and X_j , is tail dependent and the tail dependence coefficient corresponds to*

$$\lambda(X_i, X_j; \alpha, \rho_{ij}) = \frac{\int_0^{f(\rho_{ij})} \frac{s^\alpha}{\sqrt{s^2-1}} ds}{\int_0^1 \frac{s^\alpha}{\sqrt{s^2-1}} ds}, \quad f(\rho_{ij}) = \sqrt{\frac{1+\rho_{ij}}{2}}. \quad (2.7)$$

Proof. Schmidt, 2002. ■

So the tail dependence coefficient is only a function $\rho \mapsto \lambda$ whereas the tail index α of the elliptical random vector results from its specific characteristic generator. Given the matrix ρ the tail dependence is a function $\alpha \mapsto \lambda$, and due to Theorem 18 also the relation $\alpha \mapsto \tau$ holds for a given matrix of Kendall’s τ .

Note that the regular variation and thus the tail index come from the joint distribution function, whereas the tail dependence concerns particularly the copula. By Sklar’s theorem

(see Theorem 16) it is possible to construct new multivariate distributions with arbitrary margins, providing a specific copula. Especially this is done by constructing meta-elliptical distributions. In this case α is generally no longer the tail index of the new distributions but still a copula parameter.

Substituting the integration variable s in Eq. 2.7 by $\cos(v)$ leads to the following equivalent representation of the tail dependence coefficient of two elliptically distributed random variables X_i and X_j (this is observed by Hult and Lindskog, 2002, see Frahm, Junker, and Szimayer, 2003 for the details):

$$\lambda(X_i, X_j; \alpha, \rho_{ij}) = \frac{\int_{g(\rho_{ij})}^{\pi/2} \cos^\alpha(v) dv}{\int_0^{\pi/2} \cos^\alpha(v) dv}, \quad g(\rho_{ij}) = \arccos\left(\sqrt{\frac{1+\rho_{ij}}{2}}\right).$$

Due to relation (2.5) ρ_{ij} can be substituted by $\sin(\tau_{ij} \cdot \frac{\pi}{2})$ so that

$$\lambda(X_i, X_j; \alpha, \tau_{ij}) = \frac{\int_{h(\tau_{ij})}^{\pi/2} \cos^\alpha(v) dv}{\int_0^{\pi/2} \cos^\alpha(v) dv}, \quad h(\tau_{ij}) = \frac{\pi}{2} \left(\frac{1-\tau_{ij}}{2} \right). \quad (2.8)$$

Thus for the limiting case $\alpha = 0$ the tail dependence coefficient is an affine function of Kendall's τ :

$$\lim_{\alpha \searrow 0} \lambda(X_i, X_j; \alpha, \tau_{ij}) = \frac{1 + \tau_{ij}}{2}. \quad (2.9)$$

Since the tail index α of an elliptical random vector is given by the generating random variable \mathcal{R} , the tail dependence coefficient λ_{ij} of each bivariate combination (X_i, X_j) is uniquely determined by τ_{ij} . Thus modeling the tail dependence structure of elliptical copulas especially for higher dimensions is strongly restricted by the set

$$\{(\lambda, \tau) \in [0, 1] \times [-1, 1] : \lambda = \lambda(\alpha, \tau)\}$$

given the tail index parameter α .

The tail dependence coefficient of a bivariate t -distributed random vector (X, Y) with ν degrees of freedom corresponds to

$$\begin{aligned} \lambda &= 2 \cdot \bar{t}_{\nu+1} \left(\sqrt{\nu+1} \cdot \sqrt{\frac{1-\rho}{1+\rho}} \right) \\ &= 2 \cdot \bar{t}_{\nu+1} \left(\sqrt{\nu+1} \cdot \sqrt{\frac{1 - \sin(\tau \cdot \frac{\pi}{2})}{1 + \sin(\tau \cdot \frac{\pi}{2})}} \right), \quad \nu > 0, \end{aligned} \quad (2.10)$$

where $\bar{t}_{\nu+1}$ is the survival function of Student's univariate t -distribution with $\nu + 1$ degrees of freedom (cf. Embrechts, McNeil, and Straumann, 2002).

Since Eq. 2.10 holds for all $\nu > 0$, where ν corresponds to the tail index α of X and Y , and Theorem 19 states that the tail dependence coefficient of two elliptically distributed random variables depends only on ρ_{ij} and α , Eq. 2.7 can be replaced by

$$\begin{aligned} \lambda_{ij} &= 2 \cdot \bar{t}_{\alpha+1} \left(\sqrt{\alpha+1} \cdot \sqrt{\frac{1-\rho_{ij}}{1+\rho_{ij}}} \right) \\ &= 2 \cdot \bar{t}_{\alpha+1} \left(\sqrt{\alpha+1} \cdot \sqrt{\frac{1 - \sin(\tau_{ij} \cdot \frac{\pi}{2})}{1 + \sin(\tau_{ij} \cdot \frac{\pi}{2})}} \right), \quad \alpha > 0. \end{aligned} \quad (2.11)$$

Student's t -distribution is a default routine in statistics software and is tabulated in many textbooks (see, e.g., Johnson, Kotz, and Balakrishnan, 1995). So it is more convenient to use Eq. 2.11 rather than Eq. 2.7 for practical purposes.

In the following figure the upper-bound of the tail dependence coefficient as a function of ρ for any elliptical copula allowing for $\alpha > 0$ is plotted. The range of possible tail dependence in the special case $\alpha < 2$, which holds for the sub-Gaussian α -stable copula, is marked explicitly.

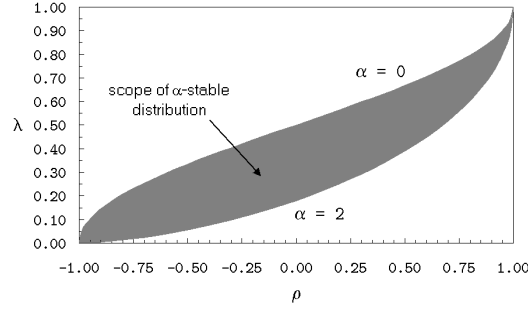


Figure 2.1 Tail dependence barriers for elliptical copulas as a function of ρ . The range of possible tail dependence for $\alpha < 2$ is marked dark-grey.

An empirical investigation (Junker, 2002) of several stocks from the German and the US market shows that the lower tail dependence ranges from 0 to 0.35, whereas Kendall's τ takes values in between 0 to 0.4, approximately. With Formula 2.8 we can plot the tail dependence barriers as a function of Kendall's τ .

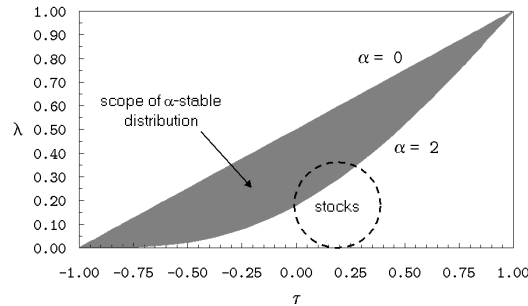


Figure 2.2 Tail dependence barriers for elliptical copulas as a function of τ . The range of possible tail dependence for $\alpha < 2$ is marked dark-grey.

Note that for $\alpha = \infty$ (i.e. if the corresponding random vector is not regularly varying) the tail dependence coefficient equals to 0 (except the comonotone case $\rho_{ij} = 1$) whereas for the limit case $\alpha = 0$ the tail dependence coefficient is an affine function of τ , as stated by Eq. 2.9. Hence the sub-Gaussian α -stable copula restricts the scope of possible tail dependence too much. The dependence structure generated by the sub-Gaussian α -stable distribution is not suitable for modeling financial risk because the provided range of λ has only a small intersection with the empirical results. Arguments against the α -stable hypothesis for financial data can also be found in the univariate case (Mikosch, 2003, p. 5).

2.3.2 Multivariate Asymptotic Dependence

The previous section dealt with the concept of bivariate asymptotic dependence. A natural generalization of the tail dependence coefficient to the multivariate case is given by

$$\lambda_L(I) := \lim_{t \searrow 0} P \left(\bigwedge_{i \in I} (F_i(X_i) \leq t) \mid \bigwedge_{j \in J} (F_j(X_j) \leq t) \right),$$

and

$$\lambda_U(I) := \lim_{t \nearrow 1} P \left(\bigwedge_{i \in I} (F_i(X_i) > t) \mid \bigwedge_{j \in J} (F_j(X_j) > t) \right),$$

where $I \cup J = \{1, \dots, d\}$, $I \cap J = \emptyset$, and \bigwedge denotes the logical conjunction. But the remaining question is how to partition the index-set $\{1, \dots, d\}$. Since the tail dependence coefficient always depends on a certain partition the generalization of bivariate asymptotic dependence to the multivariate case is not obvious. Hence, an alternative definition of multivariate asymptotic dependence is attempted.

Definition 11 (Extremal dependence coefficient) *Let X be a d -dimensional random vector with c.d.f. F and marginal distribution functions F_1, \dots, F_d . Furthermore, let $F_{\min} := \min \{F_1(X_1), \dots, F_d(X_d)\}$ and $F_{\max} := \max \{F_1(X_1), \dots, F_d(X_d)\}$. The lower extremal dependence coefficient of X is defined as*

$$\varepsilon_L := \lim_{t \searrow 0} P(F_{\max} \leq t \mid F_{\min} \leq t),$$

whereas the upper extremal dependence coefficient of X is defined as

$$\varepsilon_U := \lim_{t \nearrow 1} P(F_{\min} > t \mid F_{\max} > t),$$

provided the corresponding limits exist.

Thus the lower extremal dependence coefficient can be interpreted as the probability that the best performer of X is ‘attracted’ by the worst one provided this one has an extremely bad performance. This interpretation holds vice versa regarding the upper extremal dependence coefficient.

Note that this aspect of multivariate extremes does not correspond to the classical one by taking the componentwise maxima into consideration (cf. Definition 6). Usually, classical methods of extreme value theory can be applied even if the margins of a multivariate time series stem from completely different periods (Coles, 2001, p. 143). So the classical approach does not necessarily account for the probability of *simultaneous* extremes but only for the dependence structure of marginal extremes. That is to say there is no information about the concomitance of extremal events. But from our perception it seems to be worth to study the probability distribution of extremes which occur simultaneously.

Equation

$$P(F_{\max} \leq t \mid F_{\min} \leq t) = \frac{P(F_{\min} \leq t, F_{\max} \leq t)}{P(F_{\min} \leq t)} = \frac{P(F_1(X_1) \leq t, \dots, F_d(X_d) \leq t)}{1 - P(F_1(X_1) > t, \dots, F_d(X_d) > t)}$$

holds for the lower case and

$$P(F_{\min} > t \mid F_{\max} > t) = \frac{P(F_{\min} > t, F_{\max} > t)}{P(F_{\max} > t)} = \frac{P(F_1(X_1) > t, \dots, F_d(X_d) > t)}{1 - P(F_1(X_1) \leq t, \dots, F_d(X_d) \leq t)}$$

holds for the upper case, respectively. Thus

$$\varepsilon_L = \lim_{t \searrow 0} \frac{P(F_1(X_1) \leq t, \dots, F_d(X_d) \leq t)}{1 - P(F_1(X_1) > t, \dots, F_d(X_d) > t)} = \lim_{t \searrow 0} \frac{C(t, \dots, t)}{1 - \tilde{C}(1 - t, \dots, 1 - t)},$$

and

$$\varepsilon_U = \lim_{t \nearrow 1} \frac{P(F_1(X_1) > t, \dots, F_d(X_d) > t)}{1 - P(F_1(X_1) \leq t, \dots, F_d(X_d) \leq t)} = \lim_{t \nearrow 1} \frac{\tilde{C}(1 - t, \dots, 1 - t)}{1 - C(t, \dots, t)},$$

where C is the copula of X and \tilde{C} is the survival copula corresponding to C (cf., e.g., Junker, 2003, p. 27), i.e.

$$u \mapsto \tilde{C}(u) := \sum_{I \subset M} (-1)^{|I|} \cdot C\left((1 - u_1)^{\mathbb{1}_{1 \in I}}, \dots, (1 - u_d)^{\mathbb{1}_{d \in I}}\right), \quad (2.12)$$

where $u = (u_1, \dots, u_d) \in [0, 1]^d$ and $M := \{1, \dots, d\}$. Note that the (multivariate) survival function of C is defined as $u \mapsto \bar{C}(u) := \tilde{C}(\underline{1} - u)$ and is not a copula. Also $\bar{C} \neq 1 - C$.

Let C be a symmetric copula in the sense that $C(u) = \tilde{C}(u)$ for all $u \in [0, 1]^d$. Then $\varepsilon_L = \varepsilon_U$, since

$$\varepsilon_L = \lim_{t \searrow 0} \frac{C(t \cdot \underline{1})}{1 - \tilde{C}((1 - t) \cdot \underline{1})} = \lim_{t \nearrow 1} \frac{C((1 - t) \cdot \underline{1})}{1 - \tilde{C}(t \cdot \underline{1})} = \lim_{t \nearrow 1} \frac{\tilde{C}((1 - t) \cdot \underline{1})}{1 - C(t \cdot \underline{1})} = \varepsilon_U.$$

Thus, for elliptical distributions the lower extremal dependence coefficient equals to the upper extremal dependence coefficient.

If the dependence between the components of a random vector X is perfectly positive (not necessarily in a linear manner) X is said to be ‘comonotonic’.

Definition 12 (Comonotonicity) *Two random variables X and Y are said to be ‘comonotonic’ if there exist a random variable Z and two strictly increasing functions $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ such that $X \stackrel{\text{a.s.}}{=} f(Z)$ and $Y \stackrel{\text{a.s.}}{=} g(Z)$. Further, a d -dimensional random vector $X = (X_1, \dots, X_d)$ is said to be comonotonic if there exist a random variable Z and d strictly increasing functions $f_i : \mathbb{R} \rightarrow \mathbb{R}$, $i = 1, \dots, d$, such that $X_i \stackrel{\text{a.s.}}{=} f_i(Z)$ for $i = 1, \dots, d$.*

If X and Y are comonotonic and f and g are continuous then $X \stackrel{\text{a.s.}}{=} f(g^{-1}(Y))$, i.e. X is a strictly increasing function of Y (a.s.) and vice versa.

Proposition 20 *If a random vector is comonotonic then both the lower extremal dependence coefficient and the upper extremal dependence coefficient correspond to 1.*

Proof. If a random vector X is comonotonic then obviously its copula corresponds to the ‘minimum copula’ $\wedge_d : u \mapsto \min\{u_1, \dots, u_d\}$. \wedge_d is called the ‘Fréchet-Hoeffding upper bound’ (Nelsen, 1999, p. 9). Note that $\wedge_d = \tilde{\wedge}_d$ and thus the lower extremal dependence coefficient of X corresponds to

$$\varepsilon_L = \lim_{t \searrow 0} \frac{\wedge_d(t \cdot \underline{1})}{1 - \tilde{\wedge}_d((1 - t) \cdot \underline{1})} = \lim_{t \searrow 0} \frac{t}{1 - (1 - t)} = \lim_{t \searrow 0} \frac{t}{t} = 1,$$

Analogously, for the upper extremal dependence we obtain

$$\varepsilon_U = \lim_{t \nearrow 1} \frac{\tilde{\wedge}_d((1 - t) \cdot \underline{1})}{1 - \wedge_d(t \cdot \underline{1})} = \lim_{t \nearrow 1} \frac{1 - t}{1 - t} = 1. \quad \blacksquare$$

Proposition 21 *If the components of a random vector are mutually independent then both the lower extremal dependence coefficient and the upper extremal dependence coefficient correspond to 0.*

Proof. It is obvious that the copula of a random vector X with independent components X_1, \dots, X_d corresponds to the ‘product copula’ $\Pi_d : u \mapsto u_1 \cdot \dots \cdot u_d$ and also $\Pi_d = \tilde{\Pi}_d$. Applying l’Hôpital’s rule we obtain for the lower extremal dependence coefficient

$$\varepsilon_L = \lim_{t \searrow 0} \frac{\Pi_d(t \cdot \underline{1})}{1 - \tilde{\Pi}_d((1 - t) \cdot \underline{1})} = \lim_{t \searrow 0} \frac{t^d}{1 - (1 - t)^d} = \lim_{t \searrow 0} \frac{t^{d-1}}{(1 - t)^{d-1}} = 0.$$

The upper extremal dependence coefficient becomes also

$$\varepsilon_U = \lim_{t \nearrow 1} \frac{\tilde{\Pi}_d((1-t) \cdot \underline{1})}{1 - \Pi_d(t \cdot \underline{1})} = \lim_{t \nearrow 1} \frac{(1-t)^d}{1-t^d} = \lim_{t \nearrow 1} \frac{(1-t)^{d-1}}{t^{d-1}} = 0.$$

■

Note that within the class of elliptical distributions this holds only for normally distributed random vectors whose correlation matrix correspond to the identity matrix.

If two random variables depends on each other in a perfectly negative manner then they are said to be ‘countermonotonic’.

Definition 13 (Countermonotonicity) *Two random variables X and Y are said to be ‘countermonotonic’ if there exist a random variable Z , a strictly increasing function $f : \mathbb{R} \rightarrow \mathbb{R}$, and a strictly decreasing function $g : \mathbb{R} \rightarrow \mathbb{R}$ such that $X \stackrel{a.s.}{=} f(Z)$ and $Y \stackrel{a.s.}{=} g(Z)$.*

The copula of two countermonotonic random variables X and Y corresponds to $W : u \mapsto \max\{u_1 + u_2 - 1, 0\}$. This is called the ‘Fréchet-Hoeffding lower bound’ (Nelsen, 1999, p. 9).

Proposition 22 *If two random variables are countermonotonic then both the lower extremal dependence coefficient and the upper extremal dependence coefficient correspond to 0.*

Proof. Note that $W = \widetilde{W}$ and

$$W(1-t, 1-t) = \max\{2(1-t) - 1, 0\} = \max\{1 - 2t, 0\}.$$

Once again applying l’Hospital’s rule the lower extremal dependence of X and Y becomes

$$\varepsilon_L = \lim_{t \searrow 0} \frac{W(t, t)}{1 - \widetilde{W}(1-t, 1-t)} = \lim_{t \searrow 0} \frac{\max\{2t - 1, 0\}}{1 - \max\{1 - 2t, 0\}} = \frac{0}{2} = 0,$$

whereas the upper extremal dependence corresponds to

$$\varepsilon_U = \lim_{t \nearrow 1} \frac{\widetilde{W}(1-t, 1-t)}{1 - W(t, t)} = \lim_{t \nearrow 1} \frac{\max\{1 - 2t, 0\}}{1 - \max\{2t - 1, 0\}} = \frac{0}{-2} = 0.$$

■

Proposition 23 *Let λ_L and λ_U be the tail dependence coefficients of a pair of random variables. Further, let ε_L and ε_U be the corresponding extremal dependence coefficients. Then*

$$\varepsilon_L = \frac{\lambda_L}{2 - \lambda_L}$$

and

$$\varepsilon_U = \frac{\lambda_U}{2 - \lambda_U}.$$

Proof. Consider

$$\begin{aligned} \varepsilon_L &= \lim_{t \searrow 0} \frac{C(t, t)}{1 - \widetilde{C}(1-t, 1-t)} = \lim_{t \searrow 0} \frac{C(t, t)}{2t - C(t, t)} \\ &= \lim_{t \searrow 0} \frac{C(t, t)/t}{(2t - C(t, t))/t} = \lim_{t \searrow 0} \frac{C(t, t)/t}{2 - C(t, t)/t}, \end{aligned}$$

and note that

$$\lambda_L = \lim_{t \searrow 0} \frac{C(t, t)}{t}.$$

Similarly

$$\begin{aligned} \varepsilon_U &= \lim_{t \nearrow 1} \frac{\tilde{C}(1-t, 1-t)}{1-C(t, t)} = \lim_{t \nearrow 1} \frac{1-2t+C(t, t)}{1-C(t, t)} = \lim_{t \nearrow 1} \frac{1-2t+C(t, t)}{2(1-t)-(1-2t+C(t, t))} \\ &= \lim_{t \nearrow 1} \frac{(1-2t+C(t, t))/(1-t)}{(2(1-t)-(1-2t+C(t, t)))/(1-t)} = \lim_{t \nearrow 1} \frac{(1-2t+C(t, t))/(1-t)}{2-(1-2t+C(t, t))/(1-t)}, \end{aligned}$$

and note that

$$\lambda_U = \lim_{t \nearrow 1} \frac{1-2t+C(t, t)}{1-t}.$$

■

Hence the extremal dependence coefficient is a convex function of the tail dependence coefficient. Given a small (upper/lower) tail dependence coefficient λ the (upper/lower) extremal dependence coefficient ε is approximatively $\lambda/2$.

Proposition 24 *Let $\varepsilon_L(X)$ be the lower extremal dependence coefficient of a d -dimensional random vector X and \overline{X} be an arbitrary $(d-1)$ -dimensional sub-vector of X . Then*

$$\varepsilon_L(\overline{X}) \geq \varepsilon_L(X).$$

The same holds concerning the upper extremal dependence coefficient, i.e.

$$\varepsilon_U(\overline{X}) \geq \varepsilon_U(X).$$

Proof. Let $F_{\min}^{(d)}$ be the minimum of the mapped components of X , i.e. the minimum of $F_1(X_1), \dots, F_d(X_d)$ and $F_{\min}^{(d-1)}$ be the minimum of the mapped components of \overline{X} , respectively. Analogously, define $F_{\max}^{(d)}$ and $F_{\max}^{(d-1)}$. Since

$$P\left(F_{\max}^{(d)} \leq t \mid F_{\min}^{(d)} \leq t\right) = \frac{P\left(F_{\min}^{(d)} \leq t, F_{\max}^{(d)} \leq t\right)}{P\left(F_{\min}^{(d)} \leq t\right)} = \frac{P\left(F_{\max}^{(d)} \leq t\right)}{P\left(F_{\min}^{(d)} \leq t\right)},$$

but

$$P\left(F_{\max}^{(d-1)} \leq t\right) \geq P\left(F_{\max}^{(d)} \leq t\right),$$

and

$$P\left(F_{\min}^{(d-1)} \leq t\right) \leq P\left(F_{\min}^{(d)} \leq t\right),$$

inevitably

$$\begin{aligned} P\left(F_{\max}^{(d-1)} \leq t \mid F_{\min}^{(d-1)} \leq t\right) &= \frac{P\left(F_{\max}^{(d-1)} \leq t\right)}{P\left(F_{\min}^{(d-1)} \leq t\right)} \\ &\geq \frac{P\left(F_{\max}^{(d)} \leq t\right)}{P\left(F_{\min}^{(d)} \leq t\right)} = P\left(F_{\max}^{(d)} \leq t \mid F_{\min}^{(d)} \leq t\right). \end{aligned}$$

Since $P(F_{\max}^{(d)} \leq t \mid F_{\min}^{(d)} \leq t)$ is a lower bound of $P(F_{\max}^{(d-1)} \leq t \mid F_{\min}^{(d-1)} \leq t)$, the lower extremal dependence coefficient of \overline{X} is also bounded by the lower extremal dependence coefficient of X . The same argument holds for the upper extremal dependence coefficients.

■

So if one removes a random component of X then the remaining random vector generally exhibits higher risk of extremes. Conversely, if one adds a random component to a given random vector then the new random vector has lower risk of extremes which can be interpreted as diversification effect.

Corollary 25 *Let $X = (X_1, \dots, X_d)$ be a random vector with lower extremal dependence coefficient $\varepsilon_L > 0$. Then each lower tail dependence coefficient $\lambda_L(X_i, X_j)$ of two arbitrary components X_i and X_j of X is positive. Similarly, if $\varepsilon_U > 0$ then $\lambda_U(X_i, X_j) > 0$ for arbitrary components X_i and X_j .*

Proof. Since $\varepsilon_L(X)$ is a lower bound for $\varepsilon_L(X_i, X_j)$ also $\varepsilon_L(X_i, X_j)$ must be positive and due to Proposition 23 this holds also for the lower tail dependence coefficient $\lambda_L(X_i, X_j)$. The same argument holds for the upper tail dependence coefficients. ■

But what is the ‘driving factor’ of the extremal dependence of elliptical distributions? For the sake of simplicity we are going to focus on the multivariate t -distribution.

Lemma 26 *Let $X = (X_1, \dots, X_d) \sim t_d(0, I_d, \rho, \nu)$ with $\nu > 0$ degrees of freedom and ρ be positive definite. Let \overline{X}_i be the $(d-1)$ -dimensional sub-vector of X without the i -th component. Further, let*

$$\rho = \begin{bmatrix} 1 & \cdots & \rho_{1i} & \cdots & \rho_{1d} \\ \vdots & \ddots & \vdots & & \vdots \\ \rho_{i1} & \cdots & 1 & \cdots & \rho_{id} \\ \vdots & & \vdots & \ddots & \vdots \\ \rho_{d1} & \cdots & \rho_{di} & \cdots & 1 \end{bmatrix},$$

and let $\overline{\rho}_i$ be the sub-matrix of ρ without the i -th row and the i -th column, whereas γ_i corresponds to the i -th row of ρ without the i -th element $\rho_{ii} = 1$. Then

$$\overline{X}_i \mid (X_i = x) \sim t_{d-1} \left(\gamma_i x, \sqrt{\frac{\nu + x^2}{\nu + 1}} \cdot I_{d-1}, \rho_i, \nu + 1 \right),$$

where $\rho_i := \overline{\rho}_i - \gamma_i \gamma_i'$.

Proof. It is known (Bilodeau and Brenner, 1999, p. 239 in connection with p. 63) that if $X = (X_1, X_2) \sim t_d(\mu, \Sigma, \nu)$ where X_1 is a k -dimensional sub-vector of X and

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

then

$$X_2 \mid (X_1 = x_1) \sim t_{d-k}(\mu^*, h(x_1) \cdot \Sigma^*, \nu + k),$$

where

$$\begin{aligned} \mu^* &= \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (x_1 - \mu_1), \\ \Sigma^* &= \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}, \end{aligned}$$

and

$$h(x_1) = \frac{\nu + (x_1 - \mu_1)' \Sigma_{11}^{-1} (x_1 - \mu_1)}{\nu + k}.$$

Regarding \overline{X}_i we may assume w.l.o.g. that $i = 1$. Then the lemma follows immediately after setting $k = 1$, $\mu = 0$, $\Sigma_{11} = 1$, $\Sigma_{12} = \Sigma'_{21} = \gamma_1$, and $\Sigma_{22} = \overline{\rho}_1$. ■

Theorem 27 Let $X \sim t_d(\mu, \sigma, \rho, \nu)$ with $\nu > 0$ degrees of freedom and positive definite dispersion matrix $\Sigma = \sigma\sqrt{\rho}$ where σ and ρ are defined as described in Section 2.3.1. Then both the lower and the upper extremal dependence coefficients of X correspond to

$$\varepsilon = \frac{\sum_{i=1}^d t_{d-1, \nu+1}(-\sqrt{\nu+1} \cdot \sqrt{\rho_i}^{-1}(\mathbf{1} - \gamma_i))}{\sum_{i=1}^d t_{d-1, \nu+1}(\sqrt{\nu+1} \cdot \sqrt{\rho_i}^{-1}(\mathbf{1} - \gamma_i))},$$

where $t_{d-1, \nu+1}$ denotes the c.d.f. of the $(d-1)$ -variate t -distribution with $\nu+1$ degrees of freedom, γ_i and ρ_i are defined as in Lemma 26, and $\sqrt{\rho_i}$ is such that $\sqrt{\rho_i}\sqrt{\rho_i}' = \rho_i$.

Proof. The lower and the upper extremal dependence coefficients coincide due to the radial symmetry of X and also $C = \tilde{C}$. So taking the lower extremal dependence coefficient, for instance, leads to

$$\varepsilon = \lim_{t \searrow 0} \frac{C(t, \dots, t)}{1 - \tilde{C}(1-t, \dots, 1-t)} = \lim_{t \searrow 0} \frac{C(t, \dots, t)}{1 - C(1-t, \dots, 1-t)}.$$

Since ε depends only on the copula of X we may consider a standardized version of X , say $X^* = (X_1^*, \dots, X_d^*)$ (cf. Section 2.3.1). Then we obtain

$$\begin{aligned} \varepsilon &= \lim_{t \searrow 0} \frac{P(F^*(X_1^*) \leq t, \dots, F^*(X_d^*) \leq t)}{1 - P(F^*(X_1^*) \leq 1-t, \dots, F^*(X_d^*) \leq 1-t)} \\ &= \lim_{x \searrow -\infty} \frac{P(X_1^* \leq x, \dots, X_d^* \leq x)}{1 - P(X_1^* \leq -x, \dots, X_d^* \leq -x)}, \end{aligned}$$

where F^* is a standardized marginal c.d.f. of X . Applying l'Hospital's rule we find

$$\begin{aligned} \varepsilon &= \lim_{x \searrow -\infty} \frac{dP(X_1^* \leq x, \dots, X_d^* \leq x)/dx}{dP(X_1^* \leq -x, \dots, X_d^* \leq -x)/dx} \\ &= \lim_{x \searrow -\infty} \frac{\sum_{i=1}^d \partial P(X_1^* \leq x, \dots, X_i^* \leq x, \dots, X_d^* \leq x)/\partial x_i}{\sum_{i=1}^d \partial P(X_1^* \leq -x, \dots, X_i^* \leq -x, \dots, X_d^* \leq -x)/\partial x_i}. \end{aligned}$$

Note that $\partial P(X_1^* \leq x, \dots, X_i^* \leq x, \dots, X_d^* \leq x)/\partial x_i$ corresponds to

$$f_{X_i^*}(x) \cdot P(\overline{X}_i^* \leq x \cdot \mathbf{1} \mid X_i^* = x),$$

where \overline{X}_i^* is the $(d-1)$ -dimensional sub-vector of X^* without the i -th component and $f_{X_i^*}$ is the (standard) density function of X_i^* . From Lemma 26 we know that

$$\overline{X}_i^* \mid (X_i^* = x) \sim t_{d-1} \left(\gamma_i x, \sqrt{\frac{\nu+x^2}{\nu+1}} \cdot I_{d-1}, \rho_i, \nu+1 \right).$$

Thus

$$\varepsilon = \lim_{x \searrow -\infty} \frac{f_{X_i^*}(x) \cdot \sum_{i=1}^d t_{d-1, \nu+1} \left(x \cdot \sqrt{\frac{\nu+1}{\nu+x^2}} \cdot \sqrt{\rho_i}^{-1}(\mathbf{1} - \gamma_i) \right)}{f_{X_i^*}(-x) \cdot \sum_{i=1}^d t_{d-1, \nu+1} \left(-x \cdot \sqrt{\frac{\nu+1}{\nu+x^2}} \cdot \sqrt{\rho_i}^{-1}(\mathbf{1} - \gamma_i) \right)}.$$

Note that $f_{X_i^*}$ is symmetric, so $f_{X_i^*}(x)$ and $f_{X_i^*}(-x)$ are canceled down. Hence,

$$\varepsilon = \frac{\sum_{i=1}^d t_{d-1, \nu+1}(-\sqrt{\nu+1} \cdot \sqrt{\rho_i}^{-1}(\mathbf{1} - \gamma_i))}{\sum_{i=1}^d t_{d-1, \nu+1}(\sqrt{\nu+1} \cdot \sqrt{\rho_i}^{-1}(\mathbf{1} - \gamma_i))}.$$

■

In the following figure the extremal dependence coefficient of the multivariate t -distribution is plotted for different dimensions and degrees of freedom by assuming an equicorrelation structure.

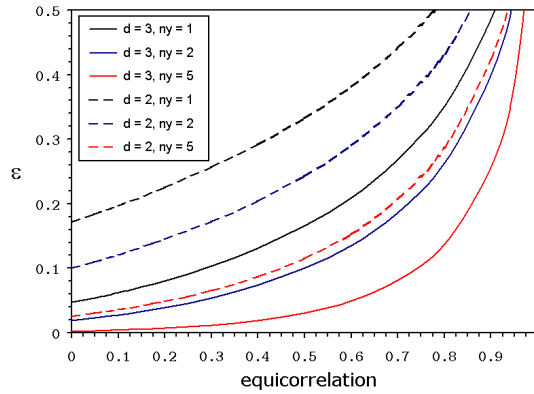


Figure 2.3 Extremal dependence coefficient of the multivariate t -distribution for $d = 2$ (dotted lines) and $d = 3$ (solid lines) where $\nu = 1$ (black lines), $\nu = 2$ (blue lines), and $\nu = 5$ (red lines).

Hence also the extremal dependence of a multivariate t -distributed random vector X is determined essentially by its number ν of degrees of freedom. Note that the multivariate normal distribution ($\nu = \infty$) has no extremal dependence. The smaller ν the larger ε (given a certain dispersion of X), i.e. the probability that each component of X is attracted by the outperformer. Since ν corresponds to the tail index of X or equivalently of its generating variate it may be expected that the extremal dependence coefficient of any elliptical random vector is mainly determined by its tail index. Moreover, following the arguments given in Section 2.3.1 concerning the tail dependence coefficient it is obvious that the extremal dependence coefficient of any other elliptical distribution can be computed also by the formula given in Theorem 27. This is part of a forthcoming work.

2.4 Covariance Matrix Estimation in the Presence of Extreme Values

In the previous section it was pointed out that the tail index of the generating variate of an elliptical random vector bears the essential information about the probability of extreme outcomes. If the realizations of \mathcal{R} are known then standard methods of EVT (cf. Coles, 2001 and Embrechts, Klüppelberg, and Mikosch, 2003) can be used for estimating the tail index. For instance, this can be done simply by fitting a GPD to the empirical excess distribution over a sufficiently high threshold (cf. Theorem 13).

Suppose $X \stackrel{d}{=} \mu + \mathcal{R}\Lambda U^{(k)} \sim \mathcal{E}_d(\mu, \Sigma, \phi)$ where μ and the positive definite matrix Σ are known. Then \mathcal{R} is given by

$$\mathcal{R} \stackrel{\text{a.s.}}{=} \|\mathcal{R}U^{(d)}\|_2 \stackrel{d}{=} \|\Lambda^{-1}(X - \mu)\|_2.$$

This is equivalent to the Mahalanobis distance of X from its center μ because

$$\|\Lambda^{-1}(X - \mu)\|_2 = \sqrt{(\Lambda^{-1}(X - \mu))' (\Lambda^{-1}(X - \mu))} = \sqrt{(X - \mu)' \Sigma^{-1} (X - \mu)}.$$

But if μ and Σ are unknown then the corresponding parameters must be replaced by some estimates $\hat{\mu}$ and $\hat{\Sigma}$, respectively. The resulting random variable

$$\hat{\mathcal{R}} = \sqrt{(X - \hat{\mu})' \hat{\Sigma}^{-1} (X - \hat{\mu})}$$

is only an estimate of \mathcal{R} . Thus even the realizations of \mathcal{R} are unknown and must be estimated before applying extreme value statistics. Other nonparametric methods like, e.g., kernel density estimation can be used (Schmidt, 2003b, pp. 159-160) if not (only) the tail behavior of \mathcal{R} is of interest but (also) its entire distribution.

It is well-known that the sample covariance matrix corresponds both to the moment estimator and to the ML-estimator for the dispersion matrix of normally distributed data. But note that for any other elliptical distribution family the dispersion matrix usually does not correspond to the covariance matrix (cf. Section 1.2.3). So the dispersion matrix is also referred to as ‘pseudo-covariance matrix’ or ‘scatter matrix’ (Visuri, 2001, p. 39).

There exist many applications like, e.g., principal component analysis, canonical correlation analysis, linear discriminant analysis, and multivariate regression for which the dispersion matrix is demanded only up to a scaling constant (cf. Oja, 2003). Further, by Tobin’s Two-fund Separation Theorem (Tobin, 1958) the optimal portfolio of risky assets does not depend on the scale of the covariance matrix and this holds also in the context of random matrix theory (cf. Part II of this thesis). If $\sigma := \sqrt{\text{tr}(\Sigma)/d}$ is defined as the ‘scale’ of Σ then the ‘shape matrix’ is obtained by Σ/σ^2 (Oja, 2003). The trace of the shape matrix (and the sum of its eigenvalues) corresponds to the dimension of Σ . Alternatively, the normalization could be done also by the determinant of Σ or simply by its upper left element. We will come back to the latter point in Section 4.2.1.

If \mathcal{R} is regularly varying with tail index $\alpha > 0$ then the survival function of $\sigma\mathcal{R}$ is regularly varying with the same tail index. Hence, also for tail index estimation it is sufficient to observe \mathcal{R} merely up to scale. In the following it is presumed that the statistician’s goal is to estimate the shape matrix of an elliptical random vector or only the corresponding (pseudo-)correlation matrix. In the case of shape matrix estimation we will loosely speak about ‘covariance matrix estimation’, anyway. Note that the shape matrix generally has more structural information than the corresponding pseudo-correlation matrix because the shape matrix preserves the variances of each component (up to scale).

Estimating the shape matrix via the sample covariance matrix, especially the correlation matrix by Pearson’s correlation coefficient is dangerous when the underlying distribution is not normal (Lindskog, 2000 and Oja, 2003). This is because Pearson’s correlation coefficient is very sensitive to outliers. Especially, if the data stem from a regularly varying random vector the smaller the tail index, i.e. the heavier the tails the larger the estimator’s variance.

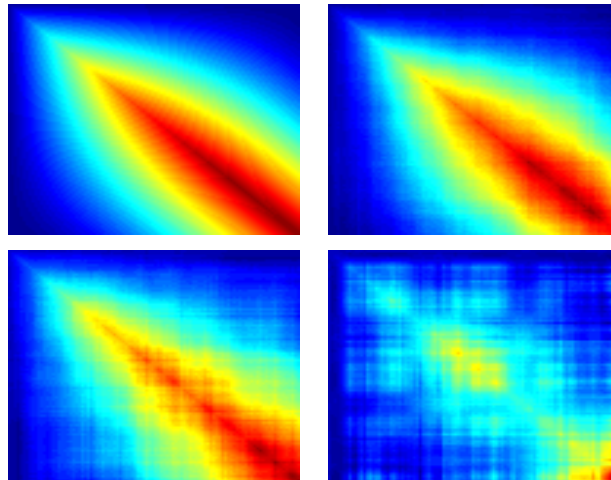


Figure 2.4 True dispersion matrix (upper left) and sample covariance matrices of samples drawn from a multivariate t -distribution with $\nu = \infty$ (i.e. the normal distribution, upper right), $\nu = 5$ (lower left), and $\nu = 2$ (lower right) degrees of freedom.

In Figure 2.4 we see sample covariance matrices with sample size 500 drawn from a centered multivariate t -distribution with 100 dimensions where the true dispersion matrix is given by the upper left image. Every cell corresponds to a matrix element. The blue colored cells represent small numbers whereas the red colored cells stand for large numbers. The true dispersion matrix as well as every covariance matrix estimate is normalized by $\Sigma_{11} = 1$. For correlation matrices a similar result is obtained which can be seen in Figure 2.5.

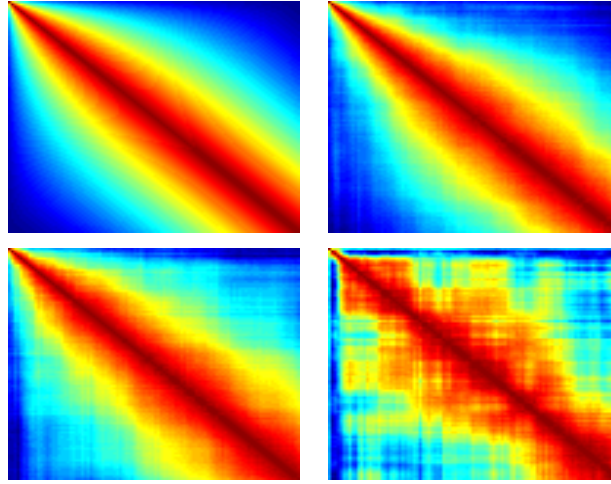


Figure 2.5 True pseudo-correlation matrix (upper left) and sample correlation matrices of samples drawn from a multivariate t -distribution with $\nu = \infty$ (i.e. the normal distribution, upper right), $\nu = 5$ (lower left), and $\nu = 2$ (lower right) degrees of freedom.

Hence the tail index by itself determines the quality of the data which is used for its estimation. Consequently, one has to rely on robust covariance estimators. Indeed, there are a lot of robust techniques to insulate from the ‘bad influence’ of outliers (cf. Visuri, 2001, pp. 31-51 and the subsequent references). But there may be ‘bad’ and ‘good’ outliers. Bad outliers are caused by sampling errors due to the measurement process whereas good outliers are data caused by true extremal events. The aim is to preserve good outliers particularly from the perspective of EVT.

For a nice overview of robust covariance matrix estimation see, e.g., Visuri (2001, Chapter 3). The simplest approach is to eliminate outliers (which is called ‘trimming’) and to apply the sample covariance matrix on the residual data (Gnanadesikan and Kettenring, 1972 and Lindskog, 2000). From the viewpoint of extreme value theory this has the annoying effect of neglecting useful information contained in extremes. In particular, estimating the tail index is impossible without outliers.

Instead of detecting outliers to eliminate them one may specify another more subtle ‘penalty’ or ‘weight’ function applying to extreme realizations. This is done by the M-estimation approach (Maronna, 1976). M-estimation can be interpreted as a generalization of the ML-estimation approach (Oja, 2003). Indeed, the ‘weight’ used implicitly by ML-estimation results from the density function of the generating variate. If one knows the true model the weights are clear otherwise they must be chosen in a more or less arbitrary manner. Nevertheless, Maronna (1976) and Huber (1981) considered criteria for existency, uniqueness, consistency, and asymptotic normality of M-estimators. But it has to be pointed out that the theoretical conditions particularly for asymptotic normality and consistency are not trivial (Visuri, 2001, p. 40). Further, the robustness of an M-estimator depends on how far the chosen weight function deviates from the optimal weight function which is given by the corresponding ML-estimator (Oja, 2003). The more nonparametric the weight function, i.e. the more compatible with alternative laws the more robust the resulting M-estimator.

Another kind of robust estimators are given by some geometrical methods invented by Rousseeuw (1985) called the ‘minimum volume ellipsoid’ (MVE-)estimator and the ‘minimum covariance determinant’ (MCD-)estimator. The MVE-estimator minimizes the volume of an ellipsoid encompassing a certain number of data points (usually more than half of the sample). Similarly, the MCD-estimator minimizes the covariance determinant (which is the squared volume of the trapezoid generated by the columns of the transformation matrix Λ). These estimators are popular and has been investigated by a number of authors (cf. Peña and Prieto, 2001). MVE- and MCD-estimators can attain very high contamination breakdown points dependent of the number of considered data (Lopuhaä and Rousseeuw, 1991). But there is a trade-off between variance and breakdown point. If the number of the considered data is small the estimator indeed has a high breakdown point but also large variance. Moreover, these kind of estimators become computationally expensive in higher dimensions because of the fact that the minimization algorithm acts on a nonconvex and nondifferentiable function created by the empirical data points (Peña and Prieto, 2001). For this case numerical approximations have to be used to obtain reasonable computational times (Rousseeuw and Driessen, 1999).

An extension of Rousseeuw’s MVE-estimator is given by the class of S-estimators (Lopuhaä, 1989). Similarly to the MVE-estimator one tries to minimize the volume of an ellipsoid but under the constraint that a number of *weighted* data points are considered. If the weight function reduces to an indicator function then the MVE-estimator occurs as a special case.

For determining the ‘outlyingness’ of a data point without the need of multivariate methods one may consider the orthogonal projections of the data onto each direction $s \in \mathcal{S}^{d-1}$. Then the outlyingness or say alternatively the ‘depth’ (Mosler, 2003) of the data point is determined by the direction which maximizes the distance of this data point relative to the others. For the purpose of comparison the data points must be standardized on each direction. Since the projected data are univariate this can be simply done by robust standard estimators for univariate location and scale (Visuri, 2001, p. 44). After knowing the depth of each data point one may define a robust covariance matrix estimator as an M-estimator where the weight of each data point is a function of its depth. This approach was invented by Stahel (1981) and Donoho (1982). Unfortunately, this method is not appropriate for high-dimensional problems, too (Peña and Prieto, 2001 and Visuri, 2001, p. 44).

Some estimators try to solve the curse of dimensions by estimating each element of the shape matrix, separately. This is nothing else but considering each projection of the data onto their bivariate subspaces. As a drawback positive definiteness cannot be guaranteed. So one has to transform the original estimate to the ‘next possible’ positive definite alternative, i.e. a matrix which is close to the original one (Lindskog, 2000). This is done, for instance, by a spectral decomposition of the original matrix and replacing its (hopefully few) negative eigenvalues by small positive ones.

Of course, every covariance matrix estimator can be used for estimating the pseudo-correlation matrix, too. But it was mentioned before that the covariance matrix has more structural information than the corresponding correlation matrix. So if one is interested only in the correlation structure why burden the estimator with needless tasks? A more efficient way of robust correlation matrix estimation in the context of elliptical distributions is described by Lindskog, McNeil, and Schmock (2003). This is simply done by inverting Eq. 2.5 in order to obtain $\rho_{ij} = \sin(\tau_{ij} \cdot \pi/2)$ for each pair of random components. Then a robust estimator of ρ_{ij} is given by $\hat{\rho}_{ij} = \sin(\hat{\tau}_{ij} \cdot \pi/2)$ where $\hat{\tau}_{ij}$ is the sample analogue of Kendall’s τ . This is given by $\hat{\tau}_{ij} = (c - d) / (c + d)$ where c is the number of concordant pairs of the realizations of X_i and X_j and d is the complementary number of discordant pairs (Lindskog, 2000). Note that this estimator depends only on the rank correlation of the data. Hence it is invariant under strictly increasing transformations and thus more robust than Pearson’s correlation coefficient. But it is not positive definite for the reasons mentioned above.

Chapter 3

Generalized Elliptical Distributions

In the following chapter the class of generalized elliptical distributions will be introduced. First, some motivation is given. Then, corresponding to the first chapter of this thesis the basic properties of the class of generalized elliptical distributions are derived. The chapter will close by examining some techniques for the construction of generalized elliptical distributions.

3.1 Motivation

Financial data usually neither are light tailed nor symmetrically distributed in the sense of radial symmetry (cf. Section 1.2.2). This holds both for the univariate case (Eberlein and Keller, 1995, Fama, 1965, Mandelbrot, 1963, Mikosch, 2003, Chapter 1) as well as for the multivariate case (Breyman, Dias, and Embrechts, 2003, Costinot, Roncalli, and Teïletche, 2000, Junker, 2002, Junker and May, 2002).

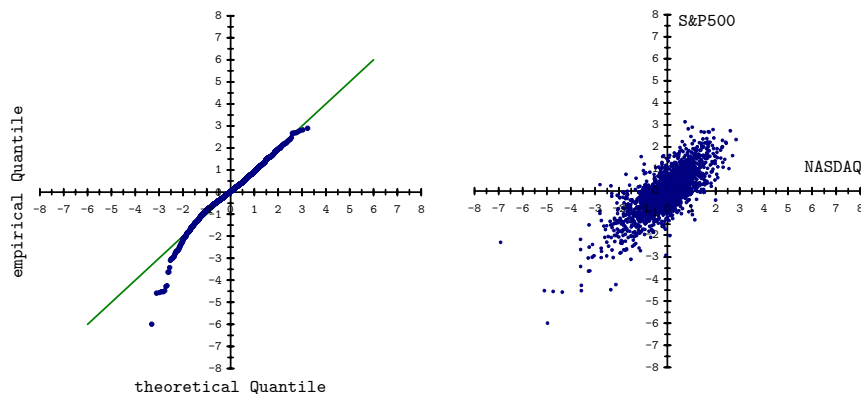


Figure 3.1 GARCH(1,1)-residuals of daily log-returns of NASDAQ and S&P 500 from 1993-01-01 to 2000-06-30 (right hand). QQ-plot of the S&P 500 residuals only (left hand).

But elliptical distributions are radially symmetric. So the question is how to model radial asymmetry without losing too much of the basic properties of elliptical distributions (cf. Section 1.2). On the one hand one should aim for parsimony regarding the parametrization

of a model for multivariate asymmetry especially in the high-dimensional case. On the other hand all the ordinary components of elliptical distributions, i.e. the generating variate \mathcal{R} , the location vector μ and the dispersion matrix Σ (which contains the linear dependence of each pair of components) should remain for the new class of asymmetric distributions.

Definition 14 (Elliptical variance-mean mixture) *A d -dimensional random vector X is called ‘elliptical variance-mean mixture’ if it can be represented by*

$$X \stackrel{d}{=} \mu + \mathcal{R}\beta + \sqrt{\mathcal{R}}Y,$$

where $\mu \in \mathbb{R}^d$, $\beta \in \mathbb{R}^d$, $Y \sim \mathcal{E}_d(0, \Sigma, \phi)$, $\Sigma \in \mathbb{R}^{d \times d}$ is positive definite, and \mathcal{R} is a nonnegative random variable being independent of Y . If $\beta = 0$ then X is an ‘elliptical variance mixture’.

Since

$$X \mid (\mathcal{R} = r) \stackrel{d}{=} \mu + r\beta + \sqrt{r}Y,$$

the c.d.f. of X is given as a mixture of $X \mid \mathcal{R} = r$ with mixing distribution $r \mapsto F_{\mathcal{R}}(r)$. This is artificially denoted by

$$x \mapsto F_X(x) = \int_0^\infty \mathcal{E}_d(\mu + r\beta, r\Sigma, \phi) dF_{\mathcal{R}}(r).$$

The vector β is not a location vector but determines the skewness of the elliptical variance-mean mixture. Elliptical variance mixtures of course are elliptically distributed.

Example 12 (Generalized hyperbolic distribution) *If $Y \sim N_d(0, \Sigma)$ then X belongs to the class of ‘normal variance-mean mixtures’ (Barndorff-Nielsen, Kent, and Sørensen, 1982). Additionally, suppose \mathcal{R} has a generalized inverse Gaussian distribution, i.e. its density function corresponds to*

$$r \mapsto f_{\mathcal{R}}(r) = \frac{(\sqrt{\kappa/\delta})^\lambda}{2K_\lambda(\sqrt{\kappa\delta})} \cdot r^{\lambda-1} \cdot \exp\left(-\frac{1}{2} \cdot (\kappa r + \delta r^{-1})\right), \quad r > 0, \quad (3.1)$$

where K_λ is the modified Bessel function of the third kind with index λ (Prause, 1999, p. 3 and Appendix B) and the parameter space corresponds to

$$\begin{cases} \kappa > 0, \delta \geq 0, & \lambda > 0, \\ \kappa > 0, \delta > 0, & \lambda = 0, \\ \kappa \geq 0, \delta > 0, & \lambda < 0. \end{cases}$$

Then X is said to be ‘generalized hyperbolic distributed’ (Barndorff-Nielsen, Kent, and Sørensen, 1982). The cases $\kappa = 0$ and $\delta = 0$ are to be interpreted as $\kappa \searrow 0$ and $\delta \searrow 0$, respectively. Note that the density of a generalized inverse Gaussian distribution can be interpreted as a mixture of power and exponential laws. This is often referred to as ‘semi-heavy’ tails (Barndorff-Nielsen and Shephard, 2003, p. 164). For $\lambda < 0$, $\kappa = 0$ and by defining $\nu := -2\lambda$ we obtain

$$\frac{(x/\delta)^\lambda}{2K_\lambda(x)} \longrightarrow \frac{(2/\delta)^\lambda}{\Gamma(-\lambda)} = \frac{\delta^{\nu/2}}{2^{\nu/2}\Gamma(\frac{\nu}{2})}, \quad x \searrow 0.$$

Then (3.1) becomes

$$\begin{aligned} r \mapsto f_{\mathcal{R}}(r) &= \frac{1}{2^{\nu/2}\Gamma(\frac{\nu}{2})} \cdot \delta^{\frac{\nu}{2}} \cdot \left(\frac{1}{r}\right)^{\frac{\nu}{2}+1} \cdot \exp\left(-\frac{1}{2} \cdot \delta \cdot \left(\frac{1}{r}\right)\right) \\ &= \frac{1}{2^{\nu/2}\Gamma(\frac{\nu}{2})} \cdot \frac{1}{\delta^{-1}} \cdot \left(\frac{1}{\delta^{-1}} \cdot \frac{1}{r}\right)^{\frac{\nu}{2}-1} \cdot \exp\left(-\frac{1}{2} \cdot \left(\frac{1}{\delta^{-1}} \cdot \frac{1}{r}\right)\right) \cdot \frac{1}{r^2}. \end{aligned}$$

This is the density function of the reciprocal of χ_{ν}^2/δ . Hence, by setting $\delta = \nu$ and the skewness parameter $\beta = 0$ we obtain the multivariate t -distribution with ν degrees of freedom (cf. Example 4) as a special case of a generalized hyperbolic distribution. Similarly, many other distributions are representable as generalized hyperbolic distributions. A nice overview is given in Prause (1999, Section 1.1).

Hence, the generalized inverse Gaussian distribution is complex and because of the possibility of combining power and exponential tails an attractive candidate for modeling the generating variate. Additionally, in Section 1.3 it was mentioned that the class of symmetric generalized hyperbolic distributions is infinitely divisible and self-decomposable.

Definition 15 (Elliptical location-scale mixture) A d -dimensional random vector X is called ‘elliptical location-scale mixture’ if it can be represented by

$$X \stackrel{d}{=} \mu + \mathcal{R}Y,$$

where $\mu \in \mathbb{R}^d$, $Y \sim \mathcal{E}_d(\beta, \Sigma, \phi)$, $\beta \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d \times d}$ is positive definite, and \mathcal{R} is a nonnegative random variable being independent of Y . If $\beta = 0$ then X is an ‘elliptical scale mixture’.

Now, the c.d.f. of X can be represented by

$$x \mapsto F_X(x) = \int_0^{\infty} \mathcal{E}_d(\mu + r\beta, r^2\Sigma, \phi) dF_{\mathcal{R}}(r).$$

If $Y \sim N_d(\beta, \Sigma)$ we may call X a ‘normal location-scale mixture’. Neither normal variance-mean mixtures nor normal location-scale mixtures are elliptically distributed if $\beta \neq 0$. Nevertheless, both classes are characterized by the ordinary components of elliptical random vectors. Only the additional parameter vector β determines the skewness, i.e. the radial asymmetry.

Another way for incorporating skewness into the elliptical framework is given by the technique of ‘hidden truncation’ (Arnold and Beaver, 2002).

Definition 16 (Skew-elliptical distribution) Let $(Y_0, Y) \sim \mathcal{E}_{d+1}(\mu^*, \Sigma^*, \phi)$ where $\mu \in \mathbb{R}^d$, $\mu^* := (0, \mu)$, $\beta \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d \times d}$, and

$$\Sigma^* := \begin{bmatrix} 1 & \beta' \\ \beta & \Sigma \end{bmatrix}.$$

Then the d -dimensional random vector $X := Y \mid Y_0 > 0$ is said to be ‘skew-elliptically distributed’ (Branco and Dey, 2001) which is denoted by $X \sim \mathcal{SE}_d(\mu, \beta, \Sigma, \phi)$.

Again β serves as a skewness parameter. If ϕ corresponds to the characteristic generator of the normal distribution then X is called ‘multivariate skew-normally distributed’ (Azzalini and Dalla Valle, 1996). A nice overview on the literature concerning skew-elliptical distributions can be found in Azzalini (2003).

For the modeling of multivariate asymmetric distributions one should guarantee the existence of a robust covariance matrix estimator. This is the quintessence of the previous chapter. More precisely, a robust covariance matrix estimator should not depend on \mathcal{R} . The main idea of this thesis is as follows: Let X be a d -dimensional elliptical location-scale mixture $X =_d \mu + \mathcal{R}Y$ with generating variate $\mathcal{R} >_{\text{a.s.}} 0$ and $Y =_d \beta + \mathcal{Q}\Lambda U^{(d)}$ where $\mathcal{Q} >_{\text{a.s.}} 0$, too. Further, let the location vector μ be known and the dispersion matrix $\Sigma = \Lambda\Lambda'$ be positive definite. The random vector

$$V := \frac{X - \mu}{\|X - \mu\|_2} \stackrel{d}{=} \frac{\mathcal{R}Y}{\|\mathcal{R}Y\|_2} \stackrel{\text{a.s.}}{=} \frac{Y}{\|Y\|_2}, \quad Y \sim \mathcal{E}_d(\beta, \Sigma, \phi),$$

does not depend on \mathcal{R} but only on β , Σ , and ϕ . Moreover,

$$V \stackrel{d}{=} \frac{Y}{\|Y\|_2} \stackrel{d}{=} \frac{\beta + \mathcal{Q}\Lambda U^{(d)}}{\|\beta + \mathcal{Q}\Lambda U^{(d)}\|_2} \stackrel{\text{a.s.}}{=} \frac{\beta/\mathcal{Q} + \Lambda U^{(d)}}{\|\beta/\mathcal{Q} + \Lambda U^{(d)}\|_2}.$$

Note that V is supported by \mathcal{S}^{d-1} and that the density function $\psi_d(\cdot; \gamma, \Lambda)$ of the random vector

$$\frac{\gamma + \Lambda U^{(d)}}{\|\gamma + \Lambda U^{(d)}\|_2}, \quad (3.2)$$

exists for all $\gamma \in \mathbb{R}^d$. Similarly to the spectral measure (cf. Section 2.2) ψ_d is a ‘spectral density function’ acting on the unit hypersphere. Now, also the density function of V exists and corresponds to

$$v \mapsto \tilde{\psi}(v) = \int_0^\infty \psi_d\left(v; \frac{\beta}{q}, \Lambda\right) dF_{\mathcal{Q}}(q), \quad v \in \mathcal{S}^{d-1}.$$

This can be used for a maximum-likelihood estimation of β and Σ . It is to be pointed out that this estimation procedure is robust against the generating distribution function $F_{\mathcal{R}}$ (provided it has no atom at zero) and it works even if \mathcal{R} would depend on Y because \mathcal{R} is canceled down anyway. The remaining problem is that it is necessary to specify not the ‘mixing distribution’ $F_{\mathcal{R}}$ but the ‘mixed distribution’ \mathcal{E}_d , i.e. the corresponding elliptical distribution family of the location-scale mixture. Indeed for the most interesting case $Y \sim N_d(\beta, \Sigma)$ an analytic expression of the density function of V is derived in Section 4.2.1.

So this approach is not completely robust. But note that the underlying elliptical distribution family must be specified only if $\beta \neq 0$ since otherwise

$$V \stackrel{d}{=} \frac{Y}{\|Y\|_2} \stackrel{d}{=} \frac{\mathcal{Q}\Lambda U^{(d)}}{\|\mathcal{Q}\Lambda U^{(d)}\|_2} \stackrel{\text{a.s.}}{=} \frac{\Lambda U^{(d)}}{\|\Lambda U^{(d)}\|_2}.$$

Now the random vector V even does not depend on \mathcal{Q} . So it is plausible to define the class of multivariate asymmetric distributions according to the stochastic representation of elliptical random vectors but allowing the generating variate \mathcal{R} to depend on the unit random vector $U^{(d)}$. This extended class of elliptical distributions allows both for asymmetry and for robust covariance matrix estimation.

3.2 Definition

Definition 17 (Generalized elliptical distribution) *The d -dimensional random vector X is said to be ‘generalized elliptically distributed’ if and only if*

$$X \stackrel{d}{=} \mu + \mathcal{R}\Lambda U^{(k)},$$

where $U^{(k)}$ is a k -dimensional random vector uniformly distributed on \mathcal{S}^{k-1} , \mathcal{R} is a random variable, $\mu \in \mathbb{R}^d$, and $\Lambda \in \mathbb{R}^{d \times k}$.

In contrast to elliptical distributions the generating variate \mathcal{R} may become negative and even it may depend stochastically on the direction determined by $U^{(k)}$. Hence the dependence structure of \mathcal{R} and $U^{(k)}$ constitutes the multivariate c.d.f. of X , essentially. In particular, X has not to be radially symmetric anymore, and its covariance matrix is not necessarily equal to $E(\mathcal{R}^2)/k \cdot \Sigma$. Moreover, μ does not correspond to the vector of expected values, generally. Unfortunately, the assertions made in Section 2.3 concerning the asymptotic dependence of meta-elliptical distributions are no longer valid for the class of generalized elliptical distributions because the copula of a generalized elliptical random vector needs not to be elliptical, anymore.

In Section 1.1 it was mentioned that the dispersion of an elliptically distributed random vector is uniquely determined via the matrix $\Sigma = \Lambda\Lambda'$, i.e. the particular matrix decomposition is irrelevant. Due to the possible dependence between \mathcal{R} and $U^{(k)}$ this is not true for generalized elliptical distributions and the transformation matrix Λ must be specified explicitly. Note that in the definition above it is not presumed that Λ has full rank. Nevertheless, if $\mathcal{R} \geq_{\text{a.s.}} 0$ and \mathcal{R} and $U^{(k)}$ are stochastically independent then (due to Proposition 1) X is elliptically symmetric distributed. Conversely, if a random vector X is elliptically symmetric distributed then (due to Theorem 2) X is always representable as in Definition 17, with \mathcal{R} and $U^{(k)}$ being independent. Hence the class of generalized elliptical distributions includes the class of elliptical distributions.

Of course, the class of elliptically symmetric distributions forms an intersection of both the class of meta-elliptical and generalized elliptical distributions. But how far meta-elliptical distributions can be represented by generalized elliptical (and vice versa) is not obvious. Fortunately, it can be shown that the class of generalized elliptical distributions contains the class of skew-elliptical distributions.

Theorem 28 *If $X \sim \mathcal{SE}_d(\mu, \beta, \Sigma, \phi)$ then X is generalized elliptically distributed with location vector μ and dispersion matrix Σ .*

Proof. Per definition X may be represented by $Y \mid Y_0 > 0$, where

$$\begin{bmatrix} Y_0 \\ Y \end{bmatrix} \stackrel{\text{d}}{=} \begin{bmatrix} 0 \\ \mu \end{bmatrix} + \mathcal{R} \cdot \sqrt{\begin{bmatrix} 1 & \beta' \\ \beta & \Sigma \end{bmatrix}} U^{(d+1)}$$

Let ϕ be the characteristic generator of $\mathcal{R}U^{(d+1)}$ where $U^{(d+1)}$ is uniformly distributed on \mathcal{S}^d and \mathcal{R} is a nonnegative random variable being stochastically independent of $U^{(d+1)}$. Consider the root

$$\sqrt{\begin{bmatrix} 1 & \beta' \\ \beta & \Sigma \end{bmatrix}} = \begin{bmatrix} 1 & 0 \\ \beta & \sqrt{\Sigma - \beta\beta'} \end{bmatrix}.$$

Further, let the generating variate \mathcal{R}^* be defined as

$$\mathcal{R}^* = \begin{cases} \mathcal{R}, & U_1^{(d+1)} > 0, \\ -\mathcal{R}, & U_1^{(d+1)} \leq 0. \end{cases}$$

Now, X can be represented by

$$X \stackrel{\text{d}}{=} \mu + \mathcal{R}^* \cdot \begin{bmatrix} \beta & \sqrt{\Sigma - \beta\beta'} \end{bmatrix} \cdot U^{(d+1)}.$$

Hence the dispersion matrix of X corresponds to Σ . ■

A d -dimensional generalized elliptically distributed random vector X can be simulated after specifying a location vector $\mu \in \mathbb{R}^d$, a transformation matrix $\Lambda \in \mathbb{R}^{d \times k}$, and the conditional

distribution functions $r \mapsto F_{\mathcal{R}|U^{(k)}=u}(r) := P(\mathcal{R} \leq r | U^{(k)} = u)$ for every $u \in \mathcal{S}^{k-1}$. Using the conditional quantile function

$$p \mapsto F_{\mathcal{R}|U^{(k)}=u}^{\leftarrow}(p) := \inf \{r : F_{\mathcal{R}|U^{(k)}=u}(r) \geq p\}, \quad 0 < p < 1,$$

the random vector X results from

$$X := \mu + F_{\mathcal{R}|U^{(k)}=\tilde{U}^{(k)}}^{\leftarrow}(Z) \Lambda \tilde{U}^{(k)},$$

where $\tilde{U}^{(k)}$ is uniformly distributed on \mathcal{S}^{k-1} and can be simulated as described in Section 1.1. Further, $Z \sim U(0, 1)$ and stochastically independent of $\tilde{U}^{(k)}$.

3.3 Basic Properties

In Section 1.2.4 it was shown that affinely transformed elliptical random vectors are also elliptical and even that the generating variate of the transformed random vector remains constant. This is because the generating variate is not bothered by the transformation and the same argument holds even if \mathcal{R} probably takes values in \mathbb{R} or if it depends on the unit random vector. So generalized elliptical distributions are also closed under affine transformations and marginalizations.

Since generalized elliptical distributions are made to allow for asymmetries consequently they do not satisfy any of the symmetry properties described in Section 1.2.2, generally. But for the quite general case $\mathcal{R} >_{\text{a.s.}} 0$ indeed they are angularly symmetric. This is because

$$\frac{X - \mu}{\|X - \mu\|_2} \stackrel{d}{=} \frac{\mathcal{R} \Lambda U^{(k)}}{\|\mathcal{R} \Lambda U^{(k)}\|_2} \stackrel{\text{a.s.}}{=} \frac{\Lambda U^{(k)}}{\|\Lambda U^{(k)}\|_2}$$

neither depends on the particular c.d.f. of \mathcal{R} nor on the dependence structure of \mathcal{R} and $U^{(k)}$. Since the random vector $\Lambda U^{(k)} / \|\Lambda U^{(k)}\|_2$ is radially symmetric the same holds for $(X - \mu) / \|X - \mu\|_2$, i.e. X is angularly symmetric about μ .

In the following it will be shown that generalized elliptical distributions fortunately are similar to elliptically symmetric distributions also concerning their density functions and conditional distributions.

Theorem 29 *Let $X \stackrel{d}{=} \mu + \mathcal{R} \Lambda U^{(k)}$ with $\Sigma := \Lambda \Lambda'$ be a d -dimensional generalized elliptically distributed random vector where $\mu \in \mathbb{R}^d$ and $\Lambda \in \mathbb{R}^{d \times k}$ with $\text{r}(\Lambda) = k$. Further, let the joint c.d.f. of \mathcal{R} and $U^{(k)}$ be absolutely continuous and \mathcal{S}_Λ be the linear subspace of \mathbb{R}^d spanned by Λ . Then the p.d.f. of X is given by*

$$x \mapsto f_X(x) = |\det(\Lambda)|^{-1} \cdot g_{\mathcal{R}}((x - \mu)' \Sigma^{-1} (x - \mu); u), \quad x \in \mathcal{S}_\Lambda \setminus \{\mu\},$$

where

$$u := \frac{\Lambda^{-1}(x - \mu)}{\sqrt{(x - \mu)' \Sigma^{-1} (x - \mu)}},$$

$$t \mapsto g_{\mathcal{R}}(t; u) := \frac{\Gamma(\frac{k}{2})}{2\pi^{k/2}} \cdot \sqrt{t}^{-(k-1)} \cdot \left(f_{\mathcal{R}|U^{(k)}=-u}(-\sqrt{t}) + f_{\mathcal{R}|U^{(k)}=u}(\sqrt{t}) \right), \quad t > 0,$$

and $f_{\mathcal{R}|U^{(k)}=u}$ is the conditional p.d.f. of \mathcal{R} under $U^{(k)} = u \in \mathcal{S}^{k-1}$.

Proof. Since the joint c.d.f. of \mathcal{R} and $U^{(k)}$ is absolutely continuous the joint p.d.f. $(r, u) \mapsto f_{(\mathcal{R}, U^{(k)})}(r, u)$ exists. Consider the conditional density function of \mathcal{R} , i.e.

$$r \mapsto f_{\mathcal{R}|U^{(k)}=u}(r) := \frac{f_{(\mathcal{R}, U^{(k)})}(r, u)}{f_{U^{(k)}}(u)},$$

where $f_{U^{(k)}}(u) = \Gamma\left(\frac{k}{2}\right) / (2\pi^{k/2})$ is the uniform density on the unit hypersphere \mathcal{S}^{k-1} . Thus the joint p.d.f. of \mathcal{R} and $U^{(k)}$ corresponds to

$$(r, u) \mapsto f_{(\mathcal{R}, U^{(k)})}(r, u) = \frac{\Gamma\left(\frac{k}{2}\right)}{2\pi^{k/2}} \cdot f_{\mathcal{R}|U^{(k)}=u}(r).$$

We define a similar transformation as in the proof of Theorem 3, i.e. $h: \mathbb{R} \setminus \{0\} \times \mathcal{S}^{k-1} \rightarrow \mathbb{R}^k \setminus \{0\}$, $(r, u) \mapsto ru =: y$. But now h is no longer injective since $r = \|y\|_2$ and $u = y/\|y\|_2$ lead to the same result as $r = -\|y\|_2$ and $u = -y/\|y\|_2$. So let

$$h^\leftarrow(y) := (\|y\|_2, y/\|y\|_2).$$

Hence the p.d.f. of Y is given by

$$\begin{aligned} y \mapsto f_Y(y) &= \left(f_{(\mathcal{R}, U^{(k)})}(-h^\leftarrow(y)) + f_{(\mathcal{R}, U^{(k)})}(h^\leftarrow(y)) \right) \cdot |J_h|^{-1}, \\ &= \frac{\Gamma\left(\frac{k}{2}\right)}{2\pi^{k/2}} \cdot \|y\|_2^{-(k-1)} \cdot \left(f_{\mathcal{R}|U^{(k)}=-u}(-\|y\|_2) + f_{\mathcal{R}|U^{(k)}=u}(\|y\|_2) \right), \quad y \neq 0, \end{aligned}$$

where $u = y/\|y\|_2$. Analogously to the proof of Theorem 3 we obtain the formula given in Theorem 29. \blacksquare

Corollary 30 *Let $X \stackrel{d}{=} \mu + \mathcal{R}\Lambda U^{(d)}$ with $\Sigma := \Lambda\Lambda'$ be a d -dimensional generalized elliptically distributed random vector where $\mu \in \mathbb{R}^d$ and $\Lambda \in \mathbb{R}^{d \times d}$ has full rank. Further, let the joint c.d.f. of \mathcal{R} and $U^{(d)}$ be absolutely continuous. Then the p.d.f. of X is given by*

$$x \mapsto f_X(x) = \sqrt{\det(\Sigma^{-1})} \cdot g_{\mathcal{R}}((x - \mu)' \Sigma^{-1}(x - \mu); u), \quad x \neq \mu,$$

where

$$u := \frac{\Lambda^{-1}(x - \mu)}{\sqrt{(x - \mu)' \Sigma^{-1}(x - \mu)}},$$

$$t \mapsto g_{\mathcal{R}}(t; u) := \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^{d/2}} \cdot \sqrt{t}^{-(d-1)} \cdot \left(f_{\mathcal{R}|U^{(d)}=-u}(-\sqrt{t}) + f_{\mathcal{R}|U^{(d)}=u}(\sqrt{t}) \right), \quad t > 0,$$

and $f_{\mathcal{R}|U^{(d)}=u}$ is the conditional p.d.f. of \mathcal{R} under $U^{(d)} = u \in \mathcal{S}^{d-1}$.

Proof. See the proof of Corollary 4. \blacksquare

Theorem 31 *Let $X \stackrel{d}{=} \mathcal{R}U^{(d)}$ be a d -dimensional generalized elliptically distributed random vector and $X = (X_1, X_2)$ where X_1 is a k -dimensional sub-vector of X . Provided the conditional random vector $X_2 | X_1 = x_1$ exists it is also generalized elliptically distributed and can be represented stochastically by*

$$X_2 | (X_1 = x_1) \stackrel{d}{=} \mathcal{R}^* U^{(d-k)},$$

where $U^{(d-k)}$ is uniformly distributed on \mathcal{S}^{d-k-1} and the generating variate is given by

$$\mathcal{R}^* = \mathcal{R} \sqrt{1 - \beta} \mid \left(\mathcal{R} \sqrt{\beta} U^{(k)} = x_1 \right).$$

Here $U^{(k)}$ is uniformly distributed on \mathcal{S}^{k-1} and $\beta \sim \text{Beta}\left(\frac{k}{2}, \frac{d-k}{2}\right)$ where β , $U^{(k)}$, and $U^{(d-k)}$ are supposed to be mutually independent. Further, \mathcal{R} may depend on $U^{(d)}$ which is given by

$$U^{(d)} = \left(\sqrt{\beta} \cdot U^{(k)}, \sqrt{1 - \beta} \cdot U^{(d-k)} \right).$$

Proof. Consider the proof of Theorem 6 but note that \mathcal{R} is no longer independent of β , $U^{(k)}$, and $U^{(d-k)}$, generally. \blacksquare

Theorem 32 Let $X \stackrel{d}{=} \mu + \mathcal{R}\Lambda U^{(r)}$ be a d -dimensional generalized elliptically distributed random vector where $\mu = (\mu_1, \mu_2) \in \mathbb{R}^d$, $\Lambda \in \mathbb{R}^{d \times r}$ with $r(\Lambda) = r$ and $\Sigma := \Lambda\Lambda'$. Let

$$C = \begin{bmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{bmatrix} \in \mathbb{R}^{d \times r}$$

be the generalized Cholesky root of Σ with sub-matrices $C_{11} \in \mathbb{R}^{k \times k}$, $C_{21} \in \mathbb{R}^{(d-k) \times k}$, and $C_{22} \in \mathbb{R}^{(d-k) \times (r-k)}$, respectively. Further, let $X = (X_1, X_2)$ where X_1 is a k -dimensional ($k < r$) sub-vector of X and let

$$\mathcal{R}_\Lambda \mid (U^{(r)} = u) := \mathcal{R} \mid (U^{(r)} = \Lambda^{-1}Cu),$$

for all $u \in \mathcal{S}^{r-1}$. Provided the conditional random vector $X_2 \mid X_1 = x_1$ exists it is also generalized elliptically distributed and can be represented stochastically by

$$X_2 \mid (X_1 = x) \stackrel{d}{=} \mu^* + \mathcal{R}_\Lambda^* C_{22} U^{(r-k)},$$

where $U^{(r-k)}$ is uniformly distributed on \mathcal{S}^{r-k-1} and the generating variate is given by

$$\mathcal{R}_\Lambda^* = \mathcal{R}_\Lambda \sqrt{1-\beta} \mid \left(\mathcal{R}_\Lambda \sqrt{\beta} U^{(k)} = C_{11}^{-1} (x_1 - \mu_1) \right),$$

whereas the location vector corresponds to

$$\mu^* = \mu_2 + C_{21} C_{11}^{-1} (x_1 - \mu_1).$$

Here $U^{(k)}$ is uniformly distributed on \mathcal{S}^{k-1} and $\beta \sim \text{Beta}\left(\frac{k}{2}, \frac{r-k}{2}\right)$ where β , $U^{(k)}$, and $U^{(r-k)}$ are supposed to be mutually independent. Further, \mathcal{R}_Λ may depend on $U^{(d)}$ which is given by

$$U^{(d)} = \left(\sqrt{\beta} \cdot U^{(k)}, \sqrt{1-\beta} \cdot U^{(d-k)} \right).$$

Proof. Note that $\mu + \mathcal{R}\Lambda U^{(r)} \neq_d \mu + \mathcal{R}CU^{(r)}$ if \mathcal{R} depends on $U^{(r)}$. But consider

$$X \stackrel{d}{=} \mu + \mathcal{R}\Lambda U^{(r)} = \mu + \mathcal{R}\Lambda\Lambda^{-1}\Lambda U^{(r)}.$$

Since

$$\Lambda\Lambda^{-1} = (\Lambda\Lambda') (\Lambda'^{-1}\Lambda^{-1}) = \Sigma\Sigma^{-1} = (CC') (C'^{-1}C^{-1}) = CC^{-1},$$

(cf. the proof of Theorem 3) we obtain the stochastic representation

$$X \stackrel{d}{=} \mu + \mathcal{R}CC^{-1}\Lambda U^{(r)} = \mu + \mathcal{R}CU_\Lambda^{(r)}$$

where $U_\Lambda^{(r)} := C^{-1}\Lambda U^{(r)}$. Note that $U_\Lambda^{(r)}$ is also uniformly distributed on \mathcal{S}^{r-1} since

$$(C^{-1}\Lambda) (C^{-1}\Lambda)' = C^{-1}\Sigma C'^{-1} = C^{-1}CC'C'^{-1} = I_r,$$

i.e. $(C^{-1}\Lambda)^{-1} = (C^{-1}\Lambda)'$ and thus $(C^{-1}\Lambda)'(C^{-1}\Lambda) = I_r$, too. So the transformation $C^{-1}\Lambda$ only rotates the random vector $U^{(r)}$. Thus we may replace $U_\Lambda^{(r)}$ by $U^{(r)}$ if we define

$$\mathcal{R}_\Lambda \mid (U^{(r)} = u) := \mathcal{R} \mid (U^{(r)} = \Lambda^{-1}Cu)$$

such that

$$X \stackrel{d}{=} \mu + \mathcal{R}_\Lambda CU^{(r)}.$$

With this representation we are able to follow the arguments in the proofs of Theorem 7 and Theorem 31 in order to obtain the formulas given in Theorem 32. ■

3.4 Models

In the following a feasible method for the modeling of asymmetric generalized elliptical distributions will be developed. Let $v_1, \dots, v_m \in \mathcal{S}^{d-1}$ be some fixed ‘reference vectors’ on the unit hypersphere. Assume that the conditional c.d.f. of \mathcal{R} is a function of some ‘distances’ $\delta(u, v_1), \dots, \delta(u, v_m)$ between u and the reference vectors v_1, \dots, v_m , i.e.

$$r \longmapsto F_{\mathcal{R}|U=u}(r) = H(r, \delta(u, v_1), \dots, \delta(u, v_m)),$$

where $H(\cdot, \delta_1, \dots, \delta_m)$ is a c.d.f. for all $(\delta_1, \dots, \delta_m) \in [0, 1]^m$.

For an adequate definition of the reference vectors v_1, \dots, v_m we may diagonalize the dispersion matrix $\Sigma \in \mathbb{R}^{d \times d}$, i.e. $\Sigma = \mathcal{O}\mathcal{D}\mathcal{O}'$, where \mathcal{O} is the orthonormal basis of the eigenvectors and \mathcal{D} is the diagonal matrix of the eigenvalues of Σ . Hence we obtain the diagonal root $\Lambda = \mathcal{O}\sqrt{\mathcal{D}}$ of Σ . Here $\sqrt{\mathcal{D}}$ is a diagonal matrix containing the square roots of the main diagonal entries of \mathcal{D} . Define $Y := \sqrt{\mathcal{D}}\mathcal{R}U^{(d)}$ such that $X =_d \mu + \mathcal{O}Y$. We can interpret the components of $Y = (Y_1, \dots, Y_d)$ as uncorrelated ‘risk factors’ of X . We will come back to this idea in Section 7.2. The variance of each factor is determined by the corresponding eigenvalue whereas its direction is determined by the associated eigenvector. Note that if w is an eigenvector of Σ then w can be substituted by its negative conjugate $-w$. Now we define both the eigenvectors v_1^+, \dots, v_d^+ and their negative conjugates v_1^-, \dots, v_d^- as reference vectors.

The next goal is to attain an adequate definition of the distance between two vectors on the unit hypersphere.

Theorem 33 *Let the d -dimensional random vector $U^{(d)}$ be uniformly distributed on the unit hypersphere. The c.d.f. of the angle $\angle(U^{(d)}, v)$ between $U^{(d)}$ and a given reference vector $v \in \mathcal{S}^{d-1}$ corresponds to*

$$a \longmapsto P\left(\angle(U^{(d)}, v) \leq a\right) = \frac{1}{2} + \frac{1}{2} \cdot \operatorname{sgn}\left(a - \frac{\pi}{2}\right) \cdot F_{\text{Beta}}\left(\cos^2(a); \frac{1}{2}, \frac{d-1}{2}\right),$$

where $a \in [0, \pi]$, $d > 1$, and $\angle(\cdot, v) := \arccos(\langle \cdot, v \rangle)$.

Proof. Since $U^{(d)} = (U_1^{(d)}, \dots, U_d^{(d)})$ is uniformly distributed on \mathcal{S}^{d-1} it can be assumed w.l.o.g. that $v = (-1, 0, \dots, 0)$. Thus $\angle(U^{(d)}, v) = \arccos(\langle U^{(d)}, v \rangle) = \arccos(-U_1^{(d)}) = \pi - \arccos(U_1^{(d)})$ and

$$P\left(\angle(U^{(d)}, v) \leq a\right) = P\left(U_1^{(d)} \leq \cos(\pi - a)\right) = P\left(U_1^{(d)} \leq -\cos(a)\right).$$

The p.d.f. of $U_1^{(d)}$ corresponds to (Fang, Kotz, and Ng, 1990, p. 73)

$$u \longmapsto f(u) = \frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{d-1}{2}\right)} \cdot (1-u^2)^{\frac{d-1}{2}-1}, \quad -1 < u < 1,$$

If $0 \leq a < \pi/2$, after substituting u by $-\sqrt{t}$ we get

$$\begin{aligned} P\left(U_1^{(d)} \leq -\cos(a)\right) &= \int_{-1}^{-\cos(a)} f(u) du = \int_1^{\cos^2(a)} f(-\sqrt{t}) \cdot \left(-\frac{1}{2}\right) \cdot t^{\frac{1}{2}-1} dt \\ &= \frac{1}{2} \cdot \int_{\cos^2(a)}^1 f(-\sqrt{t}) \cdot t^{\frac{1}{2}-1} dt \\ &= \frac{1}{2} \cdot \left(1 - F_{\text{Beta}}\left(\cos^2(a); \frac{1}{2}, \frac{d-1}{2}\right)\right). \end{aligned}$$

Similarly, if $\pi/2 \leq a \leq \pi$ we set $u = \sqrt{t}$ so that

$$\begin{aligned} P\left(U_1^{(d)} \leq -\cos(a)\right) &= \frac{1}{2} + \int_0^{-\cos(a)} f(u) du = \frac{1}{2} + \int_0^{\cos^2(a)} f(\sqrt{t}) \cdot \frac{1}{2} \cdot t^{\frac{1}{2}-1} dt \\ &= \frac{1}{2} + \frac{1}{2} \cdot F_{\text{Beta}}\left(\cos^2(a); \frac{1}{2}, \frac{d-1}{2}\right). \end{aligned}$$

■

Now we define

$$\begin{aligned} \delta(u, v) &:= P\left(\angle(U^{(d)}, v) \leq \angle(u, v)\right) \\ &= \frac{1}{2} - \frac{1}{2} \cdot \text{sgn}(\langle u, v \rangle) \cdot F_{\text{Beta}}\left(\langle u, v \rangle^2; \frac{1}{2}, \frac{d-1}{2}\right), \quad u, v \in \mathcal{S}^{d-1}, \end{aligned}$$

and propose it as a distance measure taking the number of dimensions adequately into consideration. Note that $\delta(u, v)$ is the area of the spherical cap on \mathcal{S}^{d-1} spanned by u and v divided by the surface area of \mathcal{S}^{d-1} . For $d = 2$ the distance $\delta(u, v)$ becomes simply $\arccos \langle u, v \rangle / \pi$. So one can interpret δ as a probabilistic generalization of the radian measure for d dimensions. But consider that if $d > 2$, δ is not a metric since there always exist some $u, v, w \in \mathcal{S}^{d-1}$ such that $\delta(u, v) + \delta(v, w) \not\geq \delta(u, w)$. This is because δ is a convex function of the angle between u and v provided $\angle(u, v) < \pi/2$.

A surprising relationship between the tail dependence coefficient λ of regularly varying elliptical random pairs and their tail index α is stated as follows.

Corollary 34 *Let $X \sim \mathcal{E}_d(\mu, \Sigma, \phi)$ be regularly varying with tail index $\alpha \in \mathbb{N}$ and $\Sigma = \sigma\sqrt{\rho}$ be a positive definite dispersion matrix where σ and ρ are defined as described in Section 2.3.1. Then the tail dependence coefficient of two arbitrary components of X corresponds to*

$$\lambda_{ij} = 2 \cdot P\left(\angle(U^{(\alpha+2)}, v) \leq \arccos\left(\sqrt{\frac{1-\rho_{ij}}{2}}\right)\right), \quad \rho_{ij} \in [-1, 1],$$

where $U^{(\alpha+2)}$ is uniformly distributed on the unit hypersphere $\mathcal{S}^{\alpha+1}$.

Proof. Consider Student's univariate t -distribution with ν degrees of freedom, i.e.

$$x \mapsto t_\nu(x) = \int_{-\infty}^x \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} \cdot \frac{1}{\sqrt{\nu}} \cdot \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt, \quad \nu \in \mathbb{N}.$$

Substituting t by $-\sqrt{\nu((1-s)^{-1}-1)}$ ($0 \leq s < 1$) leads to

$$\begin{aligned} t_\nu(x) &= \int_1^{l_\nu(x)} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} \cdot \frac{1}{\sqrt{\nu}} \cdot (1-s)^{\frac{\nu+1}{2}} \cdot \left(-\frac{1}{2}\right) \cdot \sqrt{\frac{\nu}{(1-s)^{-1}-1}} \cdot (1-s)^{-2} ds \\ &= \frac{1}{2} \cdot \int_{l_\nu(x)}^1 \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} \cdot s^{\frac{1}{2}-1} \cdot (1-s)^{\frac{\nu}{2}-1} ds \\ &= \frac{1}{2} \cdot \bar{F}_{\text{Beta}}\left(l_\nu(x); \frac{1}{2}, \frac{\nu}{2}\right) = P\left(\angle(U^{(\nu+1)}, v) \leq \arccos\left(\sqrt{l_\nu(x)}\right)\right), \end{aligned}$$

where

$$l_\nu(x) := 1 - \frac{1}{1 + \frac{x^2}{\nu}} = \frac{x^2}{x^2 + \nu}, \quad x \leq 0.$$

Further, the tail dependence coefficient (cf. Eq. 2.11) corresponds to

$$\lambda_{ij} = 2 \cdot \bar{t}_{\alpha+1}(z) = 2 \cdot t_{\alpha+1}(-z) = 2 \cdot P\left(\angle\left(U^{(\alpha+2)}, v\right) \leq \arccos\left(\sqrt{l_{\alpha+1}(-z)}\right)\right),$$

where

$$z := \sqrt{\alpha+1} \cdot \sqrt{\frac{1-\rho_{ij}}{1+\rho_{ij}}}, \quad \rho_{ij} \in]-1, 1],$$

such that

$$l_{\alpha+1}(-z) = \frac{(\alpha+1) \cdot \frac{1-\rho_{ij}}{1+\rho_{ij}}}{(\alpha+1) \cdot \frac{1-\rho_{ij}}{1+\rho_{ij}} + \alpha+1} = \frac{1-\rho_{ij}}{2}.$$

Note that in the limiting case $\rho_{ij} = -1$ the tail dependence coefficient always corresponds to 0 due to the countermonotonicity of X_i and X_j . ■

A nice geometrical interpretation of the previous corollary is as follows. Consider the limiting case $\alpha = 0$ so that $U^{(\alpha+2)}$ is distributed on the unit circle \mathcal{S} . Then the tail dependence coefficient corresponds to the probability that the angle between $U^{(2)}$ and an arbitrary point $v \in \mathcal{S}$ lies either within the cap

$$C_\rho(v) := \left\{ u \in \mathcal{S} : \angle(u, v) \leq \arccos\left(\sqrt{\frac{1-\rho}{2}}\right) \right\}$$

or within the ‘opposite cap’ $C_\rho(-v)$. Note that for all $\alpha \in \mathbb{N} \cup \{0\}$ this is given with probability 1 if $\rho = 1$ (i.e. the two caps merges to the unit sphere and $\lambda = 1$) whereas with probability 0 if $\rho = -1$ (i.e. the two caps degenerate to poles and $\lambda = 0$). But for all $\rho \in]-1, 1[$ the tail dependence coefficient depends essentially on the number of topological dimensions $\alpha + 1$ (cf. Figure 2.1).

Now, with the definitions above we can give some examples of generalized elliptical distributions.

Example 13 (Conditional scale distribution) Let $\Sigma = \Lambda\Lambda'$ be positive definite and the conditional c.d.f. of \mathcal{R} be $r \mapsto P(\mathcal{R} \leq r | U^{(d)} = u) = P(\gamma(u) \cdot \mathcal{R}^* \leq r)$, where the scale function γ is given by

$$u \mapsto \gamma(u) = \gamma_0 + \sum_{i=1}^d \alpha_i \left(\delta\left(\frac{\Lambda u}{\|\Lambda u\|_2}, v_i^+\right) \right)^{\vartheta_i} + \sum_{i=1}^d \beta_i \left(\delta\left(\frac{\Lambda u}{\|\Lambda u\|_2}, v_i^-\right) \right)^{\theta_i},$$

with $\gamma_0 > 0$, $\alpha_1, \dots, \alpha_d, \beta_1, \dots, \beta_d \geq 0$, $\vartheta_1, \dots, \vartheta_d, \theta_1, \dots, \theta_d > 0$. Further, \mathcal{R}^* is a positive random variable possessing a p.d.f. and being stochastically independent of $U^{(d)}$. Hence $r \mapsto f_{\mathcal{R}|U^{(d)}=u}(r) = f_{\mathcal{R}^*}(r/\gamma(u))/\gamma(u)$ and due to Corollary 30 the multivariate p.d.f. of the random vector X is given by

$$x \mapsto f_X(x) = \sqrt{\det(\Sigma^{-1})} \cdot \sigma^{-d}(x) \cdot g_{\mathcal{R}^*}\left(\frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{\sigma^2(x)}\right), \quad x \neq \mu,$$

where $g_{\mathcal{R}^*}$ is the density generator corresponding to \mathcal{R}^* , and $\sigma(x)$ is the conditional scaling factor, i.e.

$$\sigma(x) := \gamma(\Lambda^{-1}(x-\mu)).$$

Note that for the degenerate case $\alpha_1, \dots, \alpha_d, \beta_1, \dots, \beta_d = 0$ the resulting distribution becomes elliptical.

Example 14 (Generalized t -distribution) Consider Example 6 and let the conditional c.d.f. of \mathcal{R} be $r \mapsto P(\mathcal{R} \leq r \mid U^{(d)} = u) = P(\mathcal{R}^2/d \leq r^2/d \mid U^{(d)} = u) = F_{d, \gamma(u)}(r^2/d)$, where γ is the scaling function defined in Example 13. Similarly to the p.d.f. of the multivariate t -distribution the p.d.f. of X is given by

$$x \mapsto f_X(x) = \frac{\Gamma\left(\frac{d+\nu(x)}{2}\right)}{\Gamma\left(\frac{\nu(x)}{2}\right)} \cdot \sqrt{\frac{\det(\Sigma^{-1})}{(\nu(x) \cdot \pi)^d}} \cdot \left(1 + \frac{(x - \mu)' \Sigma^{-1} (x - \mu)}{\nu(x)}\right)^{-\frac{d+\nu(x)}{2}},$$

where $x \neq \mu$ and $\nu(x) \equiv \sigma(x)$. For the degenerate case $\alpha_1, \dots, \alpha_d, \beta_1, \dots, \beta_d = 0$ we obtain the d -variate t -distribution with location μ , dispersion matrix Σ and γ_0 degrees of freedom. Moreover, for $\gamma_0 \rightarrow \infty$ the d -variate normal distribution $N_d(\mu, \Sigma)$ appears.

In Figure 3.2 we see some density contour lines of Example 13 (left hand) and of Example 14 (right hand) where $d = 2$, $\mu = 0$, $\Sigma_{11} = \Sigma_{22} = 1$, and $\Sigma_{12} = \Sigma_{21} = 0.5$. The density generator of Example 13 corresponds to the density generator of the bivariate t -distribution with 100 degrees of freedom. For each example there is only one reference vector, more precisely $v_1^+ = (\cos(\pi/4), \sin(\pi/4))$ for Example 13 and $v_1^- = (-\cos(\pi/4), -\sin(\pi/4))$ for Example 14. The parameters are $\gamma_0 = 1$, $\alpha_1 = 0.25$, and $\vartheta_1 = 1$, as well as $\gamma_0 = 2$, $\beta_1 = 98$, and $\theta_1 = 2$, respectively. The residual parameters are set to zero, i.e. $\beta_1 = 0$ in Example 13 and $\alpha_1 = 0$ in Example 14, and especially $\alpha_2 = \beta_2 = 0$ in both examples. The dashed contour lines symbolize the density of the bivariate t -distribution with 100 degrees of freedom with the location and dispersion given above. This corresponds to the degenerate cases $\alpha_1 = 0$ (Example 13) and $\phi_0 = 100$, $\beta_1 = 0$ (Example 14).

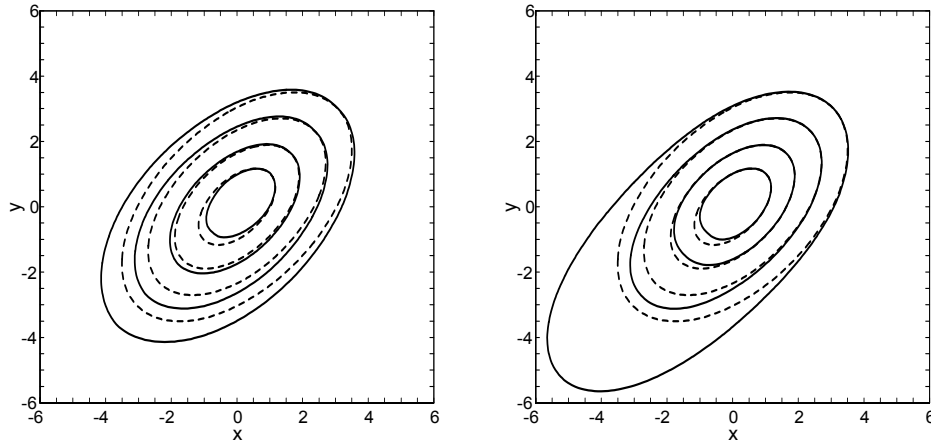


Figure 3.2 Density contours of Example 13 (left hand) and of Example 14 (right hand). The degenerate cases are represented by the dashed contour lines.

The next figure shows once again the joint distribution of the random noise of the NASDAQ and S&P 500 log-returns from 1993-01-01 to 2000-06-30 (see the right hand of Figure 3.1). On the right hand of Figure 3.3 we see a simulated distribution of GARCH(1,1)-residuals on the basis of the generalized elliptical distribution function defined in Example 14 where the pseudo-correlation coefficient corresponds to 0.78 and the location vector equals to 0. The reference vector is $v_1^- = (-\cos(\pi/4), -\sin(\pi/4))$ and the parameters are given by $\gamma_0 = 4$, $\beta_1 = 1000$, and $\theta_1 = 3$. Further, the residual parameters are set to zero.

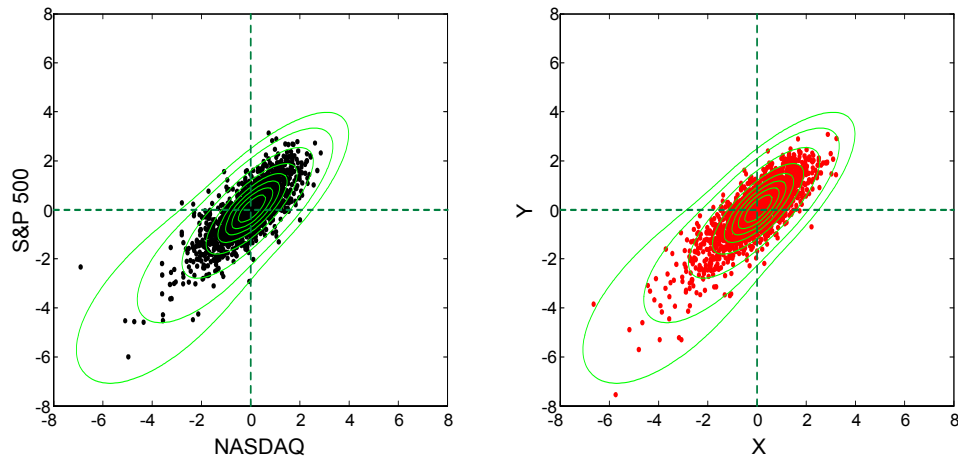


Figure 3.3 Joint distribution of NASDAQ and S&P 500 GARCH (1, 1)-residuals (left hand) and simulated generalized t -distributed data (right hand). The density contours of the corresponding generalized t -distribution are marked green.

Obviously, both asymmetry and heavy tails of financial data can be reproduced satisfactorily by an appropriate generalized elliptical distribution function.

Chapter 4

Robust Estimation

Now we come to the estimation procedures for generalized elliptically distributed random vectors motivated in Section 3.1. A robust estimator for the dispersion matrix Σ is derived presuming that the latter is positive definite. If the location vector μ is known the dispersion matrix can be estimated only provided $P(\mathcal{R} = 0) = 0$. If μ is unknown it is shown that the parameters μ and Σ may be estimated properly provided the data is angularly symmetric.

4.1 Basics of M-estimation

Let the random vector $X \stackrel{\text{d}}{=} \mu + \mathcal{R}\Lambda U^{(d)}$ be elliptically symmetric distributed with positive definite dispersion matrix Σ and absolutely continuous generating variate \mathcal{R} . Then the density function of X (cf. Corollary 4) corresponds to

$$\begin{aligned} x \mapsto f_X(x) &= \sqrt{\det(\Sigma^{-1})} \cdot g_{\mathcal{R}}((x - \mu)' \Sigma^{-1} (x - \mu)), & x \neq \mu, \\ &= \sqrt{\det(\Sigma^{-1})} \cdot g_{\mathcal{R}}(z), & z > 0, \end{aligned}$$

where $z := (x - \mu)' \Sigma^{-1} (x - \mu)$. Now, let $z \mapsto h_{\mathcal{R}}(z) := -\log(g_{\mathcal{R}}(z))$, $z > 0$. Then the log-density function of X corresponds to

$$x \mapsto \log(f_X(x)) = \frac{1}{2} \cdot \log \det(\Sigma^{-1}) - h_{\mathcal{R}}(z), \quad z > 0.$$

By applying matrix derivatives we obtain

$$\frac{\partial \log(f_X(x))}{\partial \Sigma^{-1}} = \frac{1}{2} \cdot (2\Sigma - \text{diag}(\Sigma)) - \frac{dh_{\mathcal{R}}(z)}{dz} \cdot (2z' - \text{diag}(z')), \quad (4.1)$$

where

$$z' := (x - \mu)(x - \mu)',$$

and

$$\frac{\partial \log(f_X(x))}{\partial \mu} = 2 \cdot \frac{dh_{\mathcal{R}}(z)}{dz} \cdot \Sigma^{-1} (x - \mu),$$

respectively. Note that not only Σ but Σ^{-1} is symmetric, too. This is the reason for the diag's within Eq. 4.1.

The ML-estimator for Σ is given by the root of

$$\sum_{j=1}^n \frac{\partial \log(f_X(x_{.j}))}{\partial \Sigma^{-1}} = 0,$$

i.e. by the fixed-point equation

$$\widehat{\Sigma} = \frac{1}{n} \cdot \sum_{j=1}^n 2h'_{\mathcal{R}}(z_j) \cdot z'_j = \frac{1}{n} \cdot \sum_{j=1}^n 2h'_{\mathcal{R}} \left((x_{\cdot j} - \hat{\mu})' \widehat{\Sigma}^{-1} (x_{\cdot j} - \hat{\mu}) \right) \cdot (x_{\cdot j} - \hat{\mu}) (x_{\cdot j} - \hat{\mu})',$$

where $g_{\mathcal{R}}$ is supposed to be differentiable and $h'_{\mathcal{R}}(z_j) := dh_{\mathcal{R}}(z_j)/dz_j$, $j = 1, \dots, n$. Further, the ML-estimator for μ is given by the root of

$$\sum_{j=1}^n \frac{\partial \log(f_X(x_{\cdot j}))}{\partial \mu} = 0,$$

i.e. by

$$\hat{\mu} = \frac{\sum_{j=1}^n 2h'_{\mathcal{R}}(z_j) \cdot x_{\cdot j}}{\sum_{j=1}^n 2h'_{\mathcal{R}}(z_j)} = \frac{\sum_{j=1}^n 2h'_{\mathcal{R}} \left((x_{\cdot j} - \hat{\mu})' \widehat{\Sigma}^{-1} (x_{\cdot j} - \hat{\mu}) \right) \cdot x_{\cdot j}}{\sum_{j=1}^n 2h'_{\mathcal{R}} \left((x_{\cdot j} - \hat{\mu})' \widehat{\Sigma}^{-1} (x_{\cdot j} - \hat{\mu}) \right)}.$$

Example 15 (ML-estimators if $X \sim N_d(\mu, \Sigma)$) Suppose that $X \sim N_d(\mu, \Sigma)$. Since the density generator of the class of normal distributions is given by $z \mapsto (2\pi)^{-d/2} \cdot \exp(-z/2)$ we obtain $z \mapsto h_{\mathcal{R}}(z) \propto z/2$ and thus $z \mapsto h'_{\mathcal{R}}(z) = 1/2$ for every $z > 0$. Hence the ML-estimator for Σ simply corresponds to the sample covariance matrix

$$\widehat{\Sigma} = \frac{1}{n} \cdot \sum_{j=1}^n (x_{\cdot j} - \hat{\mu}) (x_{\cdot j} - \hat{\mu})',$$

whereas the ML-estimator for μ is given by the sample mean vector $\hat{\mu} = \frac{1}{n} \cdot \sum_{j=1}^n x_{\cdot j}$.

Let $u_{\cdot j} := \sqrt{\widehat{\Sigma}}^{-1} (x_{\cdot j} - \hat{\mu}) / \sqrt{z_j}$, $j = 1, \dots, n$. Now, the ML-estimation approach described above can be represented compactly by

$$\sum_{j=1}^n 2\sqrt{z_j} h'_{\mathcal{R}}(z_j) \cdot u_{\cdot j} = 0, \quad \frac{1}{n} \cdot \sum_{j=1}^n 2z_j h'_{\mathcal{R}}(z_j) \cdot u_{\cdot j} u'_{\cdot j} = I_d.$$

Here, the terms $2\sqrt{z_j} h'_{\mathcal{R}}(z_j)$ and $2z_j h'_{\mathcal{R}}(z_j)$ can be interpreted as some weights applying to the squared Mahalanobis distances z_j . By taking other suitable weight functions (cf. Maronna, 1976 and Huber, 1981), say w_1 for estimating the location vector and w_2 for estimating the dispersion matrix, and solving the system of equations

$$\sum_{j=1}^n w_1(z_j) \cdot u_{\cdot j} = 0, \quad \frac{1}{n} \cdot \sum_{j=1}^n w_2(z_j) \cdot u_{\cdot j} u'_{\cdot j} = I_d,$$

one leaves the framework of maximum-likelihood estimation and gets to the domain of ‘maximum-likelihood-type’ (M-)estimation (Oja, 2003).

4.2 Dispersion Matrix Estimation

4.2.1 Spectral Density Approach

Definition 18 (Unit random vector) Let $\Lambda \in \mathbb{R}^{d \times k}$ with $\det(\Lambda \Lambda') \neq 0$ and $U^{(k)}$ be uniformly distributed on the unit hypersphere \mathcal{S}^{k-1} . The random vector

$$S := \frac{\Lambda U^{(k)}}{\|\Lambda U^{(k)}\|_2} \tag{4.2}$$

is called the ‘unit random vector generated by Λ ’.

Let X be a d -dimensional generalized elliptically distributed random vector where the location vector μ is assumed to be known. Further, let the transformation matrix Λ be defined as in Definition 18 and suppose that the generating variate \mathcal{R} is positive (a.s.). In Section 3.3 it was already mentioned that

$$\frac{X - \mu}{\|X - \mu\|_2} \stackrel{d}{=} \frac{\mathcal{R}\Lambda U^{(k)}}{\|\mathcal{R}\Lambda U^{(k)}\|_2} \stackrel{\text{a.s.}}{=} \frac{\Lambda U^{(k)}}{\|\Lambda U^{(k)}\|_2} = S, \quad (4.3)$$

neither depends on the particular c.d.f. of \mathcal{R} nor on the dependence structure of \mathcal{R} and $U^{(k)}$. Thus S is invariant under the choice of \mathcal{R} .

Theorem 35 *The spectral density function of the unit random vector generated by $\Lambda \in \mathbb{R}^{d \times k}$ corresponds to*

$$s \mapsto \psi(s) = \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}} \cdot \sqrt{\det(\Sigma^{-1})} \cdot \sqrt{s' \Sigma^{-1} s}^{-d}, \quad \forall s \in \mathcal{S}^{d-1}, \quad (4.4)$$

where $\Sigma := \Lambda \Lambda'$.

Proof. Due to the invariance property described above it can be assumed w.l.o.g. that

$$X \stackrel{d}{=} \sqrt{\chi_k^2} \Lambda U^{(k)} \sim N_d(0, \Sigma)$$

(cf. Example 3). The p.d.f. of X under the condition $\|X\|_2 = r > 0$ is

$$x \mapsto f_r(x) := c_r^{-1} \cdot f_X(x), \quad x \in \mathcal{S}_r^{d-1},$$

where f_X is the Gaussian p.d.f. of X and $c_r := \int_{\mathcal{S}_r^{d-1}} f_X(x) dx$. To obtain the spectral density of

$$S \stackrel{d}{=} \frac{X}{\|X\|_2}, \quad X \sim N_d(0, \Sigma),$$

we define the transformation $h: \mathbb{R}^d \setminus \{0\} \rightarrow \mathcal{S}^{d-1}$, $x \mapsto x/\|x\|_2 =: s$. Further, let ψ_r be defined as the p.d.f. of the random vector $h(X) = X/\|X\|_2$ under the condition $\|X\|_2 = r > 0$, i.e.

$$s \mapsto \psi_r(s) = f_r(h^{-1}(s)) \cdot |J_{h^{-1}}| = c_r^{-1} \cdot f_X(rs) \cdot |J_{h^{-1}}|, \quad s \in \mathcal{S}^{d-1}.$$

Here $J_{h^{-1}}$ is the Jacobian determinant of $\partial h^{-1}/\partial s'$ for a given radius r , i.e.

$$|J_{h^{-1}}| = \left| \det \left(\frac{\partial rs}{\partial s'} \right) \right|.$$

Since the tangent plane on the hypersphere \mathcal{S}_r^{d-1} has only $d-1$ dimensions (cf. the Proof of Theorem 3) we obtain

$$|J_{h^{-1}}| = \det(rI_{d-1}) = r^{d-1}.$$

Now the p.d.f. of S is given by

$$\begin{aligned} s \mapsto \psi(s) &= \int_0^\infty \psi_r(s) \cdot c_r dr = \int_0^\infty f_X(rs) \cdot r^{d-1} dr \\ &= \int_0^\infty \sqrt{\frac{\det(\Sigma^{-1})}{(2\pi)^d}} \cdot \exp\left(-\frac{1}{2} \cdot (rs)' \Sigma^{-1} (rs)\right) \cdot r^{d-1} dr. \end{aligned}$$

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

$$\begin{aligned} s \mapsto \psi(s) &= \int_0^\infty \sqrt{\frac{\det(\Sigma^{-1})}{(2\pi)^d}} \cdot \exp(-t) \sqrt{2t}^{d-2} \sqrt{s'\Sigma^{-1}s}^{-d} dt \\ &= \frac{\sqrt{\det(\Sigma^{-1})}}{\sqrt{2}^d \sqrt{\pi}^d} \cdot \sqrt{2}^{d-2} \cdot \sqrt{s'\Sigma^{-1}s}^{-d} \cdot \int_0^\infty \exp(-t) \cdot t^{\frac{d}{2}-1} dt \\ &= \frac{\sqrt{\det(\Sigma^{-1})}}{2\pi^{d/2}} \cdot \sqrt{s'\Sigma^{-1}s}^{-d} \cdot \Gamma\left(\frac{d}{2}\right), \quad s \in \mathcal{S}^{d-1}. \end{aligned}$$

■

Note that if $\Sigma = I_d$ then $\psi(s) = \Gamma(d/2) / (2\pi^{d/2})$ for every $s \in \mathcal{S}^{d-1}$. Thus the unit random vector generated by I_d is uniformly distributed on \mathcal{S}^{d-1} and $2\pi^{d/2}/\Gamma(d/2)$ is the surface area of \mathcal{S}^{d-1} . Further, the spectral density function ψ is invariant under the scale transformation $\Sigma \mapsto \sigma\Sigma$, $\sigma > 0$, since

$$\sqrt{\det((\sigma\Sigma)^{-1})} = \sigma^{-\frac{d}{2}} \sqrt{\det(\Sigma^{-1})}$$

and

$$\sqrt{s'(\sigma\Sigma)^{-1}s}^{-d} = \sigma^{\frac{d}{2}} \sqrt{s'\Sigma^{-1}s}^{-d}.$$

The distribution represented by ψ is called ‘angular central Gaussian distribution’ (Kent and Tyler, 1988, Tyler 1987b). If S belongs to the unit circle the same distribution albeit given by polar coordinates arises as a special case of the so called ‘offset normal distribution’ (Mardia, 1972, Section 3.4.7).

Proposition 36 *The unit random vector generated by Λ is generalized elliptically distributed.*

Proof. By defining $\mu := 0$ and $\mathcal{R} := \|\Lambda U^{(k)}\|_2^{-1}$ the unit random vector can be represented by $S = \mu + \mathcal{R}\Lambda U^{(k)}$. ■

Note that even though the distribution of S does not depend on the particular matrix decomposition $\Sigma = \Lambda\Lambda'$ this is not satisfied for generalized elliptical distributions (cf. Section 3.2), generally.

The spectral density of the projection of a multivariate normally distributed random vector can be derived even if $\mu \neq 0$ (cf. Section 3.1). But we need the following lemma before.

Lemma 37 *Let $a \in \mathbb{R}$, $b > 0$, and $x > 0$. Then*

$$\begin{aligned} \int_0^\infty \exp(-(at + bt^2)) t^{x-1} dt &= \frac{b^{-\frac{x}{2}}}{2} \cdot \Gamma\left(\frac{x}{2}\right) \cdot {}_1F_1\left(\frac{a^2}{4b}; \frac{x}{2}, \frac{1}{2}\right) \\ &\quad - \frac{ab^{-\frac{x+1}{2}}}{2} \cdot \Gamma\left(\frac{x+1}{2}\right) \cdot {}_1F_1\left(\frac{a^2}{4b}; \frac{x+1}{2}, \frac{3}{2}\right), \end{aligned}$$

where $z \mapsto {}_1F_1(z; \alpha, \beta)$ is the confluent hypergeometric function (Hassani, 1999, p. 420)

$$z \mapsto {}_1F_1(z; \alpha, \beta) := \frac{\Gamma(\beta)}{\Gamma(\alpha)} \cdot \sum_{k=0}^\infty \frac{\Gamma(\alpha + k)}{\Gamma(\beta + k) \cdot \Gamma(k+1)} \cdot z^k.$$

Proof. Let

$$J(a, b) := \int_0^\infty \exp(-(at + bt^2)) t^{x-1} dt, \quad a \in \mathbb{R}, b > 0.$$

After substituting t by s/\sqrt{b} we obtain $J(a, b) = b^{-x/2} \cdot J(a/\sqrt{b}, 1)$. Hence it suffices to consider the integral

$$J(c, 1) = \int_0^\infty \exp(-ct - t^2) t^{x-1} dt, \quad c \in \mathbb{R},$$

where $c = a/\sqrt{b}$. Note that by Taylor expansion $e^{-ct} = \sum_{k=0}^\infty (-ct)^k / k!$ and, after substituting $t = \sqrt{s}$,

$$\int_0^\infty e^{-t^2} t^y dt = \frac{1}{2} \cdot \int_0^\infty e^{-s} s^{\frac{y+1}{2}-1} ds = \frac{\Gamma(\frac{y+1}{2})}{2}.$$

Thus

$$J(c, 1) = \int_0^\infty e^{-ct} e^{-t^2} t^{x-1} dt = \int_0^\infty \sum_{k=0}^\infty \frac{(-ct)^k}{k!} \cdot e^{-t^2} t^{x-1} dt.$$

Using Lebesgue's Dominated Convergence Theorem we get

$$\begin{aligned} J(c, 1) &= \sum_{k=0}^\infty \frac{(-c)^k}{k!} \int_0^\infty e^{-t^2} t^{x+k-1} dt = \sum_{k=0}^\infty \frac{(-c)^k}{k!} \frac{\Gamma(\frac{x+k}{2})}{2} \\ &= \frac{1}{2} \cdot \left(\sum_{k=0,2,\dots}^\infty \frac{c^k}{k!} \cdot \Gamma\left(\frac{x+k}{2}\right) - \sum_{k=1,3,\dots}^\infty \frac{c^k}{k!} \cdot \Gamma\left(\frac{x+k}{2}\right) \right). \end{aligned}$$

Note that

$$\begin{aligned} \sum_{k=0,2,\dots}^\infty \frac{c^k}{k!} \cdot \Gamma\left(\frac{x+k}{2}\right) &= \Gamma\left(\frac{x}{2}\right) + \frac{c^2}{2!} \cdot \Gamma\left(\frac{x}{2}\right) \cdot \frac{x}{2} + \frac{c^4}{4!} \cdot \Gamma\left(\frac{x}{2}\right) \cdot \frac{x}{2} \cdot \left(\frac{x}{2} + 1\right) + \dots \\ &= \Gamma\left(\frac{x}{2}\right) \left(1 + \frac{\frac{x}{2}}{\frac{2!}{4}} \cdot \frac{c^2}{4} + \frac{\frac{x}{2} \cdot (\frac{x}{2} + 1)}{\frac{4!}{4^2}} \cdot \left(\frac{c^2}{4}\right)^2 + \dots \right) \\ &= \Gamma\left(\frac{x}{2}\right) \left(1 + \frac{\frac{x}{2}}{\frac{1}{2} \cdot 1!} \cdot \frac{c^2}{4} + \frac{\frac{x}{2} \cdot (\frac{x}{2} + 1)}{\frac{1}{2} \cdot \frac{3}{2} \cdot 2!} \cdot \left(\frac{c^2}{4}\right)^2 + \dots \right) \\ &= \Gamma\left(\frac{x}{2}\right) \cdot \sum_{k=0}^\infty \frac{(\frac{x}{2})^{(k)}}{(\frac{1}{2})^{(k)} k!} \cdot \left(\frac{c^2}{4}\right)^k = \Gamma\left(\frac{x}{2}\right) \cdot {}_1F_1\left(\frac{c^2}{4}; \frac{x}{2}, \frac{1}{2}\right). \end{aligned}$$

A similar argument holds for the odd part of $J(c, 1)$. Hence we obtain

$$J(c, 1) = \frac{1}{2} \cdot \left(\Gamma\left(\frac{x}{2}\right) \cdot {}_1F_1\left(\frac{c^2}{4}; \frac{x}{2}, \frac{1}{2}\right) - \frac{c}{2} \cdot \Gamma\left(\frac{x+1}{2}\right) \cdot {}_1F_1\left(\frac{c^2}{4}; \frac{x+1}{2}, \frac{3}{2}\right) \right),$$

and after inserting $c = a/\sqrt{b}$ and multiplying by $b^{-x/2}$ the formula given in the Lemma. ■

Theorem 38 Consider the random vector $X \sim N_d(\mu, \Sigma)$ where $\mu \in \mathbb{R}^d$ and $\Sigma = \Lambda \Lambda'$ is positive definite. The spectral density function of

$$S = \frac{X}{\|X\|_2},$$

corresponds to

$$s \mapsto \tilde{\psi}(s) = \exp\left(-\frac{1}{2} \cdot \mu' \Sigma^{-1} \mu\right) \cdot \omega(s) \cdot \psi(s), \quad s \in \mathcal{S}^{d-1},$$

where ψ is the spectral density function of the unit random vector generated by Λ (cf. Theorem 35) and

$$\omega(s) := \omega_1(s) + 2 \cdot \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \cdot \omega_2(s),$$

with

$$\begin{aligned} \omega_1(s) &:= {}_1F_1\left(z; \frac{d}{2}, \frac{1}{2}\right), \\ \omega_2(s) &:= \sqrt{z} \cdot {}_1F_1\left(z; \frac{d+1}{2}, \frac{3}{2}\right), \end{aligned}$$

and

$$z := \frac{1}{2} \cdot \frac{(s' \Sigma^{-1} \mu)^2}{s' \Sigma^{-1} s}.$$

Proof. Consider the proof of Theorem 35. By applying the change of variable formula once again we obtain

$$\begin{aligned} s \mapsto \tilde{\psi}(s) &= \int_0^\infty \psi_r(s) \cdot c_r \, dr = \int_0^\infty f_X(rs) \cdot r^{d-1} \, dr \\ &= \int_0^\infty \sqrt{\frac{\det(\Sigma^{-1})}{(2\pi)^d}} \cdot \exp\left(-\frac{1}{2} \cdot (rs - \mu)' \Sigma^{-1} (rs - \mu)\right) \cdot r^{d-1} \, dr. \\ &= \sqrt{\frac{\det(\Sigma^{-1})}{(2\pi)^d}} \cdot \exp\left(-\frac{1}{2} \cdot \mu' \Sigma^{-1} \mu\right) \\ &\quad \cdot \int_0^\infty \exp\left(s' \Sigma^{-1} \mu \cdot r - \frac{s' \Sigma^{-1} s}{2} \cdot r^2\right) \cdot r^{d-1} \, dr. \end{aligned}$$

Setting $a = -s' \Sigma^{-1} \mu$, $b = s' \Sigma^{-1} s/2$, $x = d$, and applying Lemma 37 for the integral leads to the formula given in Theorem 38. \blacksquare

The function ω in Theorem 38 determines the skewness of $\tilde{\psi}$. Of course, if $\mu = 0$ then $\omega(s) = 1$ and $\tilde{\psi}(s) = \psi(s)$, $\forall s \in \mathcal{S}^{d-1}$, i.e. we obtain the radially symmetric density function from Theorem 35. Even though ω looks a little obscure it can be shown that both functions $z \mapsto \omega_1(z)$ and $z \mapsto \omega_2(z)$ solve the so called ‘confluent hypergeometric differential equation’ (Hassani, 1999, p. 420)

$$z \cdot \frac{\partial^2 \omega_i}{\partial z^2} + \left(\frac{1}{2} - z\right) \cdot \frac{\partial \omega_i}{\partial z} - \frac{d}{2} \cdot \omega_i = 0, \quad i = 1, 2,$$

and thus also the linear combination $z \mapsto \omega(z)$ is a solution of this differential equation. Indeed, $\tilde{\psi}$ may be interesting for some theoretical reasons but for the subsequent derivation of a robust covariance matrix estimator in the context of generalized elliptical distributions only ψ will be used.

At first, it is assumed that the location vector μ is known. Motivated by the discussion in Section 3.1 and by Theorem 35 we may estimate the dispersion matrix $\Sigma \in \mathbb{R}^{d \times d}$ of X up to a scaling constant by maximizing the log-likelihood function $\sum_{j=1}^n \log \psi(s_{\cdot j})$, i.e.

$$\hat{\Sigma} := \arg \max_{\tilde{\Sigma} \in M} n \cdot \log \det(\tilde{\Sigma}^{-1}) - d \cdot \sum_{j=1}^n \log \left(s'_{\cdot j} \tilde{\Sigma}^{-1} s_{\cdot j} \right), \quad (4.5)$$

where

$$s_{\cdot j} := \frac{x_{\cdot j} - \mu}{\|x_{\cdot j} - \mu\|_2}, \quad j = 1, \dots, n, \quad (4.6)$$

and M represents the set of all positive definite matrices with dimension d . Since the log-likelihood function given by (4.5) is invariant under scale transformations of $\tilde{\Sigma}$ an additional constraint must be embedded to get an unique solution of the maximum-likelihood problem. A simple running constraint is given by $\tilde{\Sigma}_{11} = 1$. Alternative constraints are given by fixing the trace or the determinant of $\tilde{\Sigma}$ (cf. Section 2.4).

Note that if \mathcal{R} is not restricted to be positive but only $\mathcal{R} \neq_{\text{a.s.}} 0$, then

$$\frac{X - \mu}{\|X - \mu\|_2} \stackrel{\text{d}}{=} \frac{\mathcal{R} \Lambda U^{(k)}}{\|\mathcal{R} \Lambda U^{(k)}\|_2} \stackrel{\text{a.s.}}{=} \pm \frac{\Lambda U^{(k)}}{\|\Lambda U^{(k)}\|_2} = \pm S, \quad (4.7)$$

where $\pm := \text{sgn}(\mathcal{R})$. The random vector $\pm S$ does not depend on the absolute value of \mathcal{R} . But the sign of \mathcal{R} still remains and this may depend on $U^{(k)}$, anymore. So \mathcal{R} cannot be cancelled down ‘without a trace’ and thus $\pm S$ is not angularly symmetric, generally. Particularly, the density function of $\pm S$ usually does not correspond to ψ . Nevertheless, since ψ is a symmetric density function the sign of $\pm S$ does not matter at all, i.e. the ML-estimation approach considered above works even if the data is not angularly symmetric. This is given by skew-elliptically distributed data, for instance.

Even though this is a true ML-procedure there is no need for information about the generating distribution function. In particular, the estimator does not depend on the finiteness and even not on the existence of the moments of X . This is due to the separation of the radial and the angular part of X caused by relation (4.7). Note that the dispersion matrix Σ is estimated only up to an unknown scaling constant. Nevertheless, the pseudo-correlation matrix can be estimated robustly by

$$\hat{\rho} := \left[\frac{\hat{\sigma}_{ij}}{\hat{\sigma}_i \hat{\sigma}_j} \right], \quad (4.8)$$

where $\hat{\sigma}_i := \sqrt{\hat{\sigma}_{ii}}$ ($i = 1, \dots, d$). We call both $\hat{\Sigma}$ and $\hat{\rho}$ ‘spectral estimators’, respectively.

4.2.2 Fixed-point Representation

Even though a unit random vector generated by Λ is not elliptical its spectral density function ψ can be represented by

$$s \mapsto \psi(s) = \sqrt{\det(\Sigma^{-1})} \cdot g(s' \Sigma^{-1} s), \quad \forall s \in \mathcal{S}^{d-1},$$

where

$$z \mapsto g(z) := \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^{d/2}} \cdot \sqrt{z}^{-d}.$$

The same argument as described above leads further to

$$z \mapsto h(z) := -\log(g(z)) \propto \frac{d}{2} \cdot \log(z), \quad z > 0,$$

and $h'(z) = d/2 \cdot z^{-1}$. Hence the weight function for estimating the dispersion matrix becomes $z \mapsto w_2(z) = 2zh'(z) = d$. Now, the spectral estimator corresponds to the root of

$$\widehat{\Sigma} = \frac{d}{n} \cdot \sum_{j=1}^n \frac{s_{\cdot j} s_{\cdot j}'}{s_{\cdot j}' \widehat{\Sigma}^{-1} s_{\cdot j}}. \quad (4.9)$$

Note that due to (4.6) we simply obtain

$$\widehat{\Sigma} = \frac{d}{n} \cdot \sum_{j=1}^n \frac{(x_{\cdot j} - \mu)(x_{\cdot j} - \mu)'}{(x_{\cdot j} - \mu)' \widehat{\Sigma}^{-1} (x_{\cdot j} - \mu)} \quad (4.10)$$

for the spectral estimator. This is a fixed-point solution of the maximum-likelihood problem given by (4.5). Note that we fix $\widehat{\Sigma}$ by the additional constraint $\widehat{\Sigma}_{11} = 1$ for the purpose of uniqueness.

It is somewhat surprising that even though the spectral estimator can be represented by means of the original data x_1, \dots, x_n instead of the projected data s_1, \dots, s_n it is completely independent of the generating distribution function. The ‘trick’ is to find a proper weight function (see Section 4.1) such that the information about the generating variate is completely eliminated. This is exactly given by the weight function $z_j \mapsto w_2(z_j) = d$.

The estimator given by Eq. 4.10 was already proposed by Tyler (1983, 1987a). Tyler (1987a) derives the corresponding estimator as an M-estimator. More precisely, he considers the so called ‘Huber-type’ weight function

$$z \mapsto w_{\text{Huber}}(z) := \begin{cases} az, & z \leq r^2, \\ ar^2, & z > r^2, \end{cases}$$

for a fixed number $r > 0$. The number a is determined such that (cf. Tyler, 1987a)

$$E(w_{\text{Huber}}(\chi_d^2)) = d.$$

Tyler (1987a) notes that

$$w_{\text{Huber}}(z) \longrightarrow d, \quad r \searrow 0.$$

for every $z > 0$. Hence the weight function of the spectral estimator is a limiting form of the Huber-type weight function. Actually, Tyler’s estimator is not only an M-estimator on the original sample but even more an ML-estimator on the sample of elliptically distributed data which are projected to the unit hypersphere. This is also observed by Tyler (1987b). But it must be pointed out that the statement holds not only for the traditional class of elliptically symmetric distributions but even more for the extended class of generalized elliptical distributions.

4.2.3 Existence and Uniqueness

Regarding Eq. 4.10 we see that the spectral estimator $\widehat{\Sigma}$ is a fixed point or, provided $\widehat{\Sigma}$ is not fixed (for instance by $\widehat{\Sigma}_{11} = 1$), rather a ‘fixed line’. Therefore, a very simple and effective iterative algorithm for finding $\widehat{\Sigma}$ is given by $\widehat{\Sigma}^{(i+1)} = f(\widehat{\Sigma}^{(i)})$, $i = 0, 1, 2, \dots, N$, where

$$f(\widehat{\Sigma}^{(i)}) := \frac{d}{n} \cdot \sum_{j=1}^n \frac{(x_{\cdot j} - \mu)(x_{\cdot j} - \mu)'}{(x_{\cdot j} - \mu)' \widehat{\Sigma}^{(i)-1} (x_{\cdot j} - \mu)},$$

and N is a large number. For the initial estimate one may choose $\widehat{\Sigma}^{(0)} = I_d$.

During the iteration any additional requirement such as $\widehat{\Sigma}_{11}^{(i)} = 1$ has not to be considered. For applying the results of maximum-likelihood theory it is sufficient to do the normalization $\widehat{\Sigma} = \widehat{\Sigma} / \widehat{\Sigma}_{11}$ merely at the end of N iterations.

Proposition 39 *Let the d -dimensional random vector $X \stackrel{d}{=} \mu + \mathcal{R}\Lambda U^{(d)}$ be generalized elliptically distributed with $P(\mathcal{R} = 0) = 0$ and positive definite dispersion matrix $\Sigma = \Lambda\Lambda'$. Further, suppose that the location vector μ is known. If $n > d$ the spectral estimator which is obtained numerically after a finite number of iterations is positive definite (a.s.) provided its initial estimate is positive definite, too.*

Proof. Let X_1, \dots, X_n be n independent copies of X where $n > d$. Then

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

is positive definite (a.s.). Thus $\sum_{j=1}^n (x_{\cdot j} - \mu)(x_{\cdot j} - \mu)'$ is positive definite, too. Further, the quantity $w_j := d / (x_{\cdot j} - \mu)' \tilde{\Sigma}^{-1} (x_{\cdot j} - \mu)$ must be positive if $\tilde{\Sigma}$ is positive definite. Thus

$$f(\tilde{\Sigma}) = \frac{1}{n} \cdot \sum_{j=1}^n (\sqrt{w_j} (x_{\cdot j} - \mu)) (\sqrt{w_j} (x_{\cdot j} - \mu))'$$

is positive definite, too. By complete induction we conclude that the spectral estimator obtained numerically after a finite number of iterations is always positive definite provided the initial estimate is positive definite, too. ■

The positive definiteness of the initial estimate can be ensured simply by $\tilde{\Sigma}^{(0)} = I_d$ or by taking the sample covariance matrix as an initial estimate. Of course, the fact that the spectral estimator obtained after a finite number of numerical iterations is always positive definite does not guarantee that the theoretical solution to the fixed-point equation is positive definite, too. But Tyler (1987a) states that a positive definite fixed point almost surely exists and is unique (up to a scaling constant) if $n > d(d-1)$ and, additionally, the sample is drawn from a continuous distribution which is not necessarily generalized elliptical. Clearly the continuity condition is superfluous if one presumes that the data is generalized elliptically distributed with positive definite dispersion matrix and without an atom at μ . This is because the spectral estimator can be written in terms of the data projected to the unit hypersphere (cf. Eq. 4.9) and this is always continuously distributed if the dispersion matrix is positive definite. It should be pointed out that $n > d(d-1)$ is a sufficient condition for the existency of the spectral estimator. Fortunately, in practice the spectral estimator exists in most cases when n is already slightly larger than d .

Since the standard conditions for the existence of M-estimators (Maronna, 1976 and Huber, 1981, Chapter 8) do not apply on Tyler's estimator he rather gives a constructive proof. That is to say both existency and uniqueness are established via an iterative algorithm similar to the method discussed above. Nevertheless, if one needs the existence and uniqueness of the spectral estimator only for proving its asymptotic properties (given a constant number of dimensions) like, e.g., consistency then obviously the critical sample size does not matter.

Tyler (1987a) also derives several properties of the spectral estimator like, e.g., the asymptotic normality and strong consistency. But this is not straightforward since due to the limiting behavior of Huber's weight function some standard results of M-estimation theory (Huber, 1981) cannot be applied. In Chapter 5 it is shown that the desired statistical properties, i.e. consistency, asymptotic efficiency and normality can be derived on the basis of standard maximum-likelihood theory, instead.

Now we may compare the spectral estimators on the simulated data described in Section 2.4 with the sample covariance matrix (cf. Figure 2.4) and the sample correlation matrix (cf. Figure 2.5).

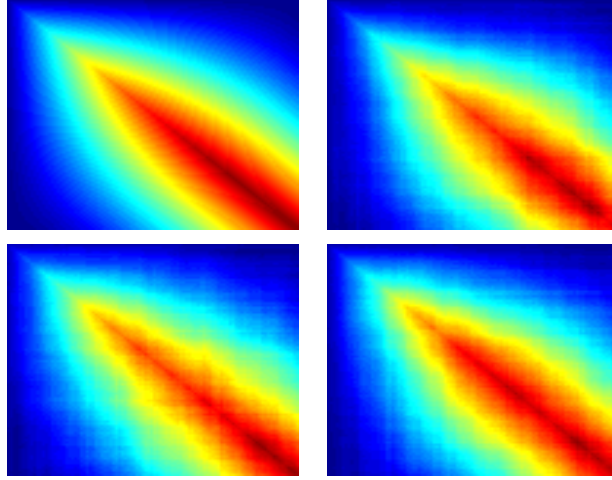


Figure 4.1 True dispersion matrix (upper left) and spectral estimator of samples drawn from a multivariate t -distribution with $\nu = \infty$ (i.e. the normal distribution, upper right), $\nu = 5$ (lower left), and $\nu = 2$ (lower right) degrees of freedom.

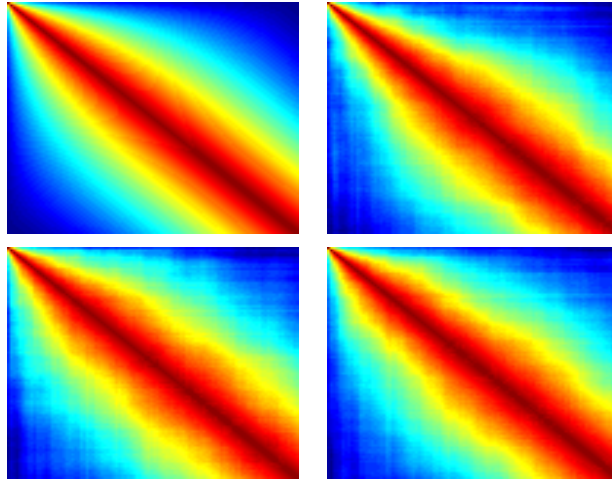


Figure 4.2 True pseudo-correlation matrix (upper left) and spectral estimator of samples drawn from a multivariate t -distribution with $\nu = \infty$ (i.e. the normal distribution, upper right), $\nu = 5$ (lower left), and $\nu = 2$ (lower right) degrees of freedom.

4.3 Location Vector Estimation

Now, let the location vector μ be unknown. Hence, it must be substituted in Eq. 4.10 by an adequate estimate $\hat{\mu}$, i.e.

$$\hat{\Sigma} = \frac{d}{n} \cdot \sum_{j=1}^n \frac{(x_{\cdot j} - \hat{\mu})(x_{\cdot j} - \hat{\mu})'}{(x_{\cdot j} - \hat{\mu})' \hat{\Sigma}^{-1} (x_{\cdot j} - \hat{\mu})}. \quad (4.11)$$

Unfortunately, the location vector μ cannot be estimated robustly by the spectral density approach presented in Section 4.2.1. This is because if X is generalized elliptically distributed then the distribution of the random vector

$$\frac{X - \theta}{\|X - \theta\|_2} \stackrel{d}{=} \frac{(\mu - \theta) + \mathcal{R}\Lambda U^{(k)}}{\|(\mu - \theta) + \mathcal{R}\Lambda U^{(k)}\|_2}$$

is no longer unaffected by \mathcal{R} . Even if \mathcal{R} is independent of $U^{(k)}$ one has to specify the distribution of \mathcal{R} for calculating the p.d.f. of $(X - \theta) / \|X - \theta\|_2$ (see Eq. 3.2 and Theorem 38).

Let $x_{i1}^*, \dots, x_{in}^*$ be a permutation of the observations x_{i1}, \dots, x_{in} such that $x_{i1}^* \leq \dots \leq x_{in}^*$, $i = 1, \dots, d$. Further, let $x_{\cdot j}^*$ be the j -th column vector of the matrix $S_n^* := [x_{ij}^*]$. The componentwise sample median is defined as

$$\hat{x}_{0.5, n} := \begin{cases} \frac{1}{2} \cdot (x_{\cdot, 0.5 \cdot n}^* + x_{\cdot, 0.5 \cdot n + 1}^*), & n \text{ even,} \\ x_{\cdot, 0.5 \cdot (n+1)}^*, & n \text{ odd.} \end{cases}$$

In the following the affix ‘componentwise’ will be dropped for the sake of simplicity.

Let $\hat{\mu}$ be an arbitrary estimator for μ like, e.g., the sample mean or the sample median. Then $\hat{\mu}$ may be adopted in order to estimate $\hat{\Sigma}$ by using Eq. 4.11. If $n > d(2d - 1)$ then the spectral estimate $\hat{\Sigma}$ both exists and is unique provided those observations where $x_{\cdot j} = \hat{\mu}$ are skipped in Eq. 4.11 (Tyler, 1987a). If $\hat{\mu} \rightarrow_{\text{a.s.}} \mu$ ($n \rightarrow \infty$) then $\hat{\Sigma} \rightarrow_p \Sigma$ as $n \rightarrow \infty$. But note that the sample mean is not consistent since $E(X) \neq \mu$, generally (see Section 3.2). In the following it is shown that the sample median is an appropriate robust alternative to the sample mean provided the data is angularly symmetric.

Definition 19 (Strict median) Let F_i^{\leftarrow} be the i -th marginal quantile function (cf. Definition 7) of an arbitrary d -dimensional random vector X ($i = 1, \dots, d$). The median of X is defined as the vector

$$x_{0.5} := \left[\frac{1}{2} \cdot (F_i^{\leftarrow}(0.5^-) + F_i^{\leftarrow}(0.5^+)) \right].$$

If $F_i^{\leftarrow}(0.5^-) = F_i^{\leftarrow}(0.5^+)$ for all $i = 1, \dots, d$, then we say that X has a ‘strict median’.

Proposition 40 Let the d -dimensional random vector X be generalized elliptically distributed with location vector μ . If X is angularly symmetric possessing a strict median then $x_{0.5} = \mu$.

Proof. Since the median of X is supposed to be strict it is sufficient to show that

$$P(X_i - \mu_i \leq 0) \geq 0.5 \geq P(X_i - \mu_i < 0), \quad i = 1, \dots, d.$$

Due to the angular symmetry of X ,

$$\begin{aligned} P(X_i - \mu_i \leq 0) &= P\left(\frac{X_i - \mu_i}{\|X - \mu\|_2} \leq 0\right) = P\left(-\frac{X_i - \mu_i}{\|X - \mu\|_2} \leq 0\right) \\ &= 1 - P\left(\frac{X_i - \mu_i}{\|X - \mu\|_2} < 0\right) = 1 - P(X_i - \mu_i < 0), \quad i = 1, \dots, d, \end{aligned}$$

i.e. the assertion holds. ■

Proposition 40 implies that if X is angularly symmetric then its location vector may be properly estimated by the sample median. Since the median of X is supposed to be strict the sample median converges strongly to the theoretical one (Pestman, 1998, p. 320), i.e. $\hat{x}_{0.5, n} \rightarrow_{\text{a.s.}} x_{0.5} = \mu$ ($n \rightarrow \infty$).

Alternatively, the location vector may be estimated robustly by

$$\sum_{j=1}^n \frac{x_{\cdot j} - \hat{\mu}}{\sqrt{(x_{\cdot j} - \hat{\mu})' \hat{\Sigma}^{-1} (x_{\cdot j} - \hat{\mu})}} = 0,$$

i.e. by the fixed-point equation

$$\hat{\mu} = \frac{\sum_{j=1}^n x_{\cdot j} / \sqrt{(x_{\cdot j} - \hat{\mu})' \hat{\Sigma}^{-1} (x_{\cdot j} - \hat{\mu})}}{\sum_{j=1}^n 1 / \sqrt{(x_{\cdot j} - \hat{\mu})' \hat{\Sigma}^{-1} (x_{\cdot j} - \hat{\mu})}}. \quad (4.12)$$

This is suggested by Tyler (1987a) and indeed $\hat{\mu}$ corresponds to the M-estimator obtained by taking the constant weight function already discussed in Section 4.2.2. Obviously, this estimation procedure is reasonable if the data is angularly symmetric. If the estimates $\hat{\mu}$ and $\hat{\Sigma}$ are calculated simultaneously by the fixed-point equations 4.11 and 4.12 their existency is not easy to show (Tyler, 1987a). Nevertheless, the latter approach for estimating the location vector has been found to be useful and reliable in practice.

Chapter 5

Statistical Properties of the Spectral Estimator

In the following the statistical properties of the spectral estimator are examined by applying classical maximum-likelihood theory. We consider a sample of i.i.d. realizations. For the sake of simplicity it is assumed that the location vector μ is known.

5.1 Information Matrix

Note that every d -dimensional generalized elliptical random vector can be represented by

$$X \stackrel{d}{=} \mu + \mathcal{R}\Lambda U^{(k)} = \mu + c\mathcal{R} \left(\frac{\Lambda}{c} \right) U^{(k)} = \mu + \tilde{\mathcal{R}}\tilde{\Lambda}U^{(k)},$$

where $\tilde{\mathcal{R}} := c\mathcal{R}$, $\tilde{\Lambda} := \Lambda/c$, and $c := \|\Lambda_1\|_2$ is the Euclidean norm of the first row of Λ . Clearly, if $\Lambda_1 \neq 0$ then $\|\Lambda_1\|_2 > 0$. Hence, the normalized dispersion matrix $\tilde{\Sigma} := \tilde{\Lambda}\tilde{\Lambda}'$ has the property $\tilde{\Sigma}_{11} = 1$ which is used for fixing the spectral estimator (cf. Section 4.2). But since the spectral estimator is invariant under scale transformations the latter property is without restriction of any kind.

The following derivation focuses on the log-likelihood of the unit random vector S . But actually only $\pm S$ is observable. Nevertheless, due to the radial symmetry of the spectral density function ψ we may proceed on the assumption that each realization of S is known (cf. Section 4.2.1). To obtain the Fisher information matrix we have to calculate $\partial \log(\psi(S)) / \partial \Sigma$ rather than $\partial \log(\psi(S)) / \partial \Sigma^{-1}$ (cf. Section 4.1). Unfortunately, notation becomes cumbersome once matrix derivatives and especially higher moments of them (i.e. expected tensors) are involved. For the purpose of keeping the transparency as high as possible let

$$\text{vec}(\lfloor A), \quad A \in \mathbb{R}^{d \times d},$$

be the vector of the lower triangular part of A without its upper left element, i.e.

$$\text{vec}(\lfloor A) := (A_{22}, A_{33}, \dots, A_{dd}, A_{21}, A_{32}, \dots, A_{d,(d-1)}, A_{31}, \dots, A_{d1}).$$

Proposition 41 *Let the d -dimensional random vector $X \stackrel{d}{=} \mu + \mathcal{R}\Lambda U^{(d)}$ be generalized elliptically distributed with $P(\mathcal{R} = 0) = 0$ and positive definite dispersion matrix $\Sigma = \Lambda\Lambda'$. Further, let*

$$S := \frac{X - \mu}{\|X - \mu\|_2}$$

be the unit random vector generated by Λ and let

$$\xi := \text{vec} \left(\left[\frac{\partial \log \psi(S; \Sigma)}{\partial \Sigma} \right] \right)$$

denote the score of the sample element S with respect to Σ . It has the following distribution

$$\xi \stackrel{d}{=} \text{vec} \left(\left[\left(d \cdot ZZ' - \Sigma^{-1} \right) - \frac{1}{2} \cdot \text{diag} \left(d \cdot ZZ' - \Sigma^{-1} \right) \right] \right),$$

where $Z := \tilde{\Lambda}'^{-1} U^{(d)}$ with $\tilde{\Lambda} := \Lambda / \|\Lambda_1\|_2$. Further, $E(\xi) = 0$ and the elementary information matrix of the spectral estimator

$$\mathcal{J} := E(\xi \xi')$$

is finite and nonsingular.

Proof. We are searching for

$$\frac{\partial \log \psi(S; \Sigma)}{\partial \Sigma} = \left(\frac{\partial \text{vec}(\Sigma^{-1})'}{\partial \Sigma} \right) \left(I_d \otimes \frac{\partial \left(\frac{1}{2} \cdot \log \det(\Sigma^{-1}) - \frac{d}{2} \cdot \log(S' \Sigma^{-1} S) \right)}{\partial \text{vec}(\Sigma^{-1})} \right).$$

Since the spectral density is invariant under scale transformations of $\Sigma = \Lambda \Lambda'$ we may assume $\Sigma_{11} = 1$ w.l.o.g. such that $\|\Lambda_1\|_2 = 1$, too. Note that

$$\frac{\partial \Sigma^{-1}}{\partial \Sigma} = - (I_d \otimes \Sigma^{-1}) \frac{\partial \Sigma}{\partial \Sigma} (I_d \otimes \Sigma^{-1}),$$

which has the effect that every $\partial \Sigma / \partial \Sigma_{ij}$ ($i, j = 1, \dots, d$) is multiplied by Σ^{-1} from the left and from the right, i.e.

$$\frac{\partial \Sigma^{-1}}{\partial \Sigma} = \left[\frac{\partial \Sigma^{-1}}{\partial \Sigma_{ij}} \right] = \left[-\Sigma^{-1} \cdot \frac{\partial \Sigma}{\partial \Sigma_{ij}} \cdot \Sigma^{-1} \right].$$

Let A, B and Σ be symmetric elements of $\mathbb{R}^{d \times d}$. After a little thought we obtain the relation

$$\left[\text{vec} \left(-A \cdot \frac{\partial \Sigma}{\partial \Sigma_{ij}} \cdot A \right)' \right] \cdot \left(I_d \otimes \text{vec} \left(\frac{1}{2} \cdot B \right) \right) = -ABA - \frac{1}{2} \cdot \text{diag}(-ABA).$$

Therefore

$$\begin{aligned} \frac{\partial \log \psi(S; \Sigma)}{\partial \Sigma} &= -\Sigma^{-1} \left(\Sigma - d \cdot \frac{SS'}{S' \Sigma^{-1} S} \right) \Sigma^{-1} \\ &\quad - \frac{1}{2} \cdot \text{diag} \left(-\Sigma^{-1} \left(\Sigma - d \cdot \frac{SS'}{S' \Sigma^{-1} S} \right) \Sigma^{-1} \right) \\ &= \left(d \cdot \Sigma^{-1} \cdot \frac{SS'}{S' \Sigma^{-1} S} \cdot \Sigma^{-1} - \Sigma^{-1} \right) \\ &\quad - \frac{1}{2} \cdot \text{diag} \left(d \cdot \Sigma^{-1} \cdot \frac{SS'}{S' \Sigma^{-1} S} \cdot \Sigma^{-1} - \Sigma^{-1} \right). \end{aligned}$$

Hence

$$S' \Sigma^{-1} S \stackrel{d}{=} \left(\frac{\Lambda U^{(d)}}{\|\Lambda U^{(d)}\|_2} \right)' (\Lambda')^{-1} \Lambda^{-1} \left(\frac{\Lambda U^{(d)}}{\|\Lambda U^{(d)}\|_2} \right) = \frac{U^{(d)'} U^{(d)}}{(\Lambda U^{(d)})' (\Lambda U^{(d)})} = \frac{1}{(\Lambda U^{(d)})' (\Lambda U^{(d)})},$$

and also

$$SS' \stackrel{d}{=} \left(\frac{\Lambda U^{(d)}}{\|\Lambda U^{(d)}\|_2} \right) \left(\frac{\Lambda U^{(d)}}{\|\Lambda U^{(d)}\|_2} \right)' = \frac{\Lambda U^{(d)} U^{(d)'} \Lambda'}{(\Lambda U^{(d)})' (\Lambda U^{(d)})},$$

such that

$$\frac{SS'}{S'\Sigma^{-1}S} \stackrel{d}{=} \Lambda U^{(d)} U^{(d)'} \Lambda'.$$

Note that due to the positive definiteness of Σ the quantity $S'\Sigma^{-1}S$ never becomes 0. Furthermore,

$$d \cdot \Sigma^{-1} \cdot \frac{SS'}{S'\Sigma^{-1}S} \cdot \Sigma^{-1} \stackrel{d}{=} d \cdot \Lambda'^{-1} U^{(d)} U^{(d)'} \Lambda^{-1} \stackrel{d}{=} d \cdot ZZ',$$

where $Z := \Lambda'^{-1} U^{(d)}$. Thus the score of each sample element with respect to Σ equals to

$$\left(d \cdot ZZ' - \Sigma^{-1} \right) - \frac{1}{2} \cdot \text{diag} \left(d \cdot ZZ' - \Sigma^{-1} \right),$$

in distribution. Note that only the upper left element exceptionally equals to 0. This is suppressed for notional convenience. Since

$$E(Z) = 0,$$

and

$$\text{Var}(Z) = E(ZZ') = \Lambda'^{-1} \text{Var}(U) \Lambda^{-1} = \frac{\Sigma^{-1}}{d},$$

we conclude $E(\partial \log \psi(S; \Sigma) / \partial \Sigma) = 0$. Further, the elementary information is given by the covariance matrix of the lower triangular elements of

$$d \cdot \left(ZZ' - \frac{1}{2} \cdot \text{diag}(ZZ') \right),$$

but without the upper left element. Obviously, the elementary information is finite because Z is bounded. Note that the number of parameters of the spectral estimator corresponds to $m := \binom{d+1}{2} - 1$. Because Λ is supposed to have full rank the support of the random vector $\text{vec}(\lfloor ZZ' \rfloor)$ has also m dimensions and so the elementary information is nonsingular. ■

Lemma 42 *Let the d -dimensional random vector $X \stackrel{d}{=} \mu + \mathcal{R} \Lambda U^{(d)}$ be generalized elliptically distributed with $P(\mathcal{R} = 0) = 0$ and positive definite dispersion matrix $\Sigma = \Lambda \Lambda'$ with $\Sigma_{11} = 1$ (w.l.o.g.). Further, let X_1, \dots, X_n ($n = 1, 2, \dots$) be sequences of independent copies of X and let ξ_n be the sample score. Then*

$$\xi_n / \sqrt{n} \longrightarrow N_m(0, \mathcal{J}), \quad n \longrightarrow \infty,$$

where $m := \binom{d+1}{2} - 1$ and \mathcal{J} is the elementary information matrix of the spectral estimator given by Proposition 41.

Proof. Schönfeld, 1971, p. 316. ■

Corollary 43 *Let the d -dimensional random vector $X \stackrel{d}{=} \mu + \mathcal{R}(\sigma I_d) U^{(d)}$ be generalized elliptically distributed with $P(\mathcal{R} = 0) = 0$ and $\sigma > 0$. Then the elementary information matrix of the spectral estimator corresponds to*

$$\mathcal{J}_0 \equiv [\mathcal{J}_{0,ij}] := \begin{cases} \frac{1}{2} \cdot \frac{d-1}{d+2}, & i = j = 1, \dots, d-1, \\ -\frac{1}{2} \cdot \frac{1}{d+2}, & i, j = 1, \dots, d-1, i \neq j, \\ \frac{d}{d+2}, & i = j = d, \dots, \binom{d+1}{2} - 1, \\ 0, & \text{else.} \end{cases}$$

Proof. We have to calculate the covariances between the elements of

$$\text{vec} \left(\left[d \cdot \left(U^{(d)} U^{(d)'} - \frac{1}{2} \cdot \text{diag} \left(U^{(d)} U^{(d)'} \right) \right) \right] \right),$$

which can be done by using

$$\text{Cov} \left(U_i^{(d)} U_j^{(d)}, U_k^{(d)} U_l^{(d)} \right) = E \left(U_i^{(d)} U_j^{(d)} U_k^{(d)} U_l^{(d)} \right) - E \left(U_i^{(d)} U_j^{(d)} \right) \cdot E \left(U_k^{(d)} U_l^{(d)} \right),$$

for $i, j, k, l = 1, \dots, d$ and applying Theorem 5, extensively. ■

Note that the elementary information of the spectral estimator depends essentially on the number of dimensions. Surprisingly, the information increases with the number of dimensions. This fact leads to remarkable properties of the spectral estimator in higher dimensions which were partially investigated by Dümbgen (1998). Some new results in the context of random matrix theory are presented in Chapter 8.

By using the same argument as in the proof of Proposition 41 we obtain the elementary information of the sample covariance matrix if $X \sim N_d(0, I_d)$ simply by the covariance matrix of the elements of

$$\frac{\partial \log f(X; \Sigma)}{\partial \Sigma} \stackrel{d}{=} X X' - \frac{1}{2} \cdot \text{diag}(X X').$$

Here f denotes the Gaussian density function. The elementary information matrix is given by

$$\mathcal{I}'_0 \equiv [\mathcal{I}'_{0,ij}] := \begin{cases} \frac{1}{2}, & i = j = 1, \dots, d, \\ 1, & i = j = d+1, \dots, \binom{d+1}{2}, \\ 0, & \text{else,} \end{cases}$$

which can be easily verified. Let \mathcal{I}_0 be the elementary information matrix of the sample covariance matrix after deleting the first column and the first row of \mathcal{I}'_0 for the purpose of comparison. Then obviously

$$\mathcal{J}_0 \longrightarrow \mathcal{I}_0, \quad d \longrightarrow \infty.$$

Hence the elementary information of the spectral estimator providing only a generalized elliptical random vector with $\Sigma = \sigma^2 I_d$ converges to the elementary information of the sample covariance matrix providing $X \sim N_d(0, I_d)$.

In the convergence above the number of dimensions and not the sample size grows to infinity. But here one has to be very careful. For applying classical maximum-likelihood theory we must at least guarantee that $n/d \rightarrow \infty$ as $n \rightarrow \infty$ and $d \rightarrow \infty$, i.e. $d = o(n)$. The quantity $q := n/d$ can be interpreted as ‘average sample size per dimension’ or as ‘effective sample size’.

Dümbgen (1998) shows that under the conditions of Corollary 43 the condition number γ , i.e. the ratio between the largest and the smallest eigenvalue of Tyler’s M-estimator (i.e. the spectral estimator) has the property

$$\gamma = 1 + \frac{4}{\sqrt{q}} + o_P \left(\frac{1}{\sqrt{q}} \right) = 1 + O_P \left(\frac{1}{\sqrt{q}} \right).$$

Note that γ is a random variable and $q \rightarrow \infty$ implies $n \rightarrow \infty$ but $d = o(n)$. Now, the same convergence holds also for the sample covariance matrix providing a standard normally distributed random vector with uncorrelated (i.e. independent) components (Dümbgen, 1998). Because the results of maximum-likelihood theory are particularly based on the central limit theorem many large sample properties of covariance matrix estimates fail if the effective sample size q is small even if n is large. A more detailed discussion of this sort of ‘high-dimensional problems’ follows in Part II of this thesis.

5.2 Consistency and Asymptotic Efficiency

The subsequent proofs rely on standard maximum-likelihood theory as given, e.g., by Schönfeld (1971, Appendix D). For the following it is convenient to decompose the parameter $\tilde{\Sigma}$ (see Eq. 4.5) by its Cholesky root, i.e. $\tilde{\Sigma} = CC'$, where C is a lower triangular matrix. Note that $\tilde{\Sigma}$ is required to be positive definite. Thus C is nonsingular, i.e. there are no zero elements on the main diagonal of C . Changing the sign of an arbitrary column vector of C does not make any difference regarding $CC' = \tilde{\Sigma}$. Thus, by convention, C is required to have only positive elements on its main diagonal. It is well-known that any positive definite matrix $\tilde{\Sigma}$ has a full rank Cholesky decomposition. Conversely, any lower triangular matrix C with positive main diagonal elements leads to a positive definite matrix $CC' = \tilde{\Sigma}$. Thus $\tilde{\Sigma}$ may be represented properly by its Cholesky root.

Theorem 44 (Consistency) *Let the d -dimensional random vector $X \stackrel{d}{=} \mu + \mathcal{R}\Lambda U^{(d)}$ be generalized elliptically distributed with $P(\mathcal{R} = 0) = 0$ and positive definite dispersion matrix $\Sigma = \Lambda\Lambda'$ with $\Sigma_{11} = 1$ (w.l.o.g.). Further, let X_1, \dots, X_n ($n = 1, 2, \dots$) be sequences of independent copies of X and let $\hat{\Sigma}_n$ be the corresponding spectral estimator. Then the spectral estimator is weakly consistent, i.e.*

$$\hat{\Sigma}_n \xrightarrow{p} \Sigma, \quad n \longrightarrow \infty.$$

Proof. For any positive definite dispersion matrix Σ it can be shown that

1. the parameter space is an open interval of the Euclidean space;
 2. the log-likelihood function is continuously differentiable (with respect to each parameter of the dispersion matrix) up to the third derivative;
 3. the first, second, and the third partial derivatives of the log-likelihood function can be obtained under the integral sign;
 4. further, the third partial derivatives have finite expectations;
 5. the elementary information matrix is finite and nonsingular;
 6. the root of the log-likelihood equation always exists and is unique.
- ad 1. Consider the Cholesky decomposition $\tilde{\Sigma} = CC'$. The lower triangular part of $C \in \mathbb{R}^{d \times d}$ stems from an open interval of the $\binom{d+1}{2}$ -dimensional Euclidean space since the main diagonal entries of C are required to be positive. Moreover, since we fix $\tilde{\Sigma}_{11} = 1$, i.e. $C_{11} = 1$ the parameter space is an open interval of the Euclidean space with $\binom{d+1}{2} - 1$ dimensions.
- ad 2. By virtue of Proposition 41 we see that not only the first but also the second and the third derivatives of the log-likelihood function exist and are continuous for any given observation.
- ad 3. The first derivative of the log-likelihood function is given by the elementary score derived in the proof of Proposition 41. We see that the score depends essentially on the product of each two components of an observation. Thus it is continuous with respect to each observation and one easily may find continuous (and thus integrable) upper bounds for the absolute values of the partial derivatives so that the Dominated Convergence Theorem holds. Moreover, the second and third derivatives do not depend on the observations at all and the Dominated Convergence Theorem holds, too.
- ad 4. Because the third derivatives are not random their expectations are trivially finite.

- ad 5. The finiteness and regularity of the elementary information matrix was shown already in Proposition 41.
- ad 6. The spectral estimator, i.e. the root of the log-likelihood equation exists and is unique if $n > d(d-1)$ (Tyler, 1987a). Hence existence and uniqueness is guaranteed, asymptotically. ■

Theorem 45 (Asymptotic efficiency) *Let the conditions of Theorem 44 be fulfilled. Then the spectral estimator is asymptotically efficient in the sense of Rao (1962) and*

$$\sqrt{n} \cdot \left(\text{vec} \left(\left[\hat{\Sigma}_n \right] \right) - \text{vec} \left(\left[\Sigma \right] \right) \right) \xrightarrow{d} N_m \left(0, \mathcal{J}^{-1} \right), \quad n \longrightarrow \infty,$$

where $m := \binom{d+1}{2} - 1$ and \mathcal{J} is the elementary information matrix given in Proposition 41.

Proof. Due to the conditions stated in the proof of Theorem 44 the spectral estimator is asymptotically efficient (Rao, 1962), i.e.

$$\text{plim}_{n \rightarrow \infty} \sqrt{n} \cdot \left(\left(\text{vec} \left(\left[\hat{\Sigma}_n \right] \right) - \text{vec} \left(\left[\Sigma \right] \right) \right) - \mathcal{J}^{-1} \xi_n / n \right) = 0,$$

where ξ_n is the sample score. Further, due to Lemma 42, $\xi_n / \sqrt{n} \rightarrow N_m(0, \mathcal{J})$ and thus $\sqrt{n} \cdot \mathcal{J}^{-1} \xi_n / n \rightarrow N_m(0, \mathcal{J}^{-1})$. ■

It must be pointed out that the asymptotic efficiency of the spectral estimator does only hold for generalized elliptically distributed data which is projected to the unit hypersphere. Once the original data is used for estimation of course a parametric maximum-likelihood approach is ‘more’ efficient provided the true model is known. For instance, if one knows that the data are multivariate normally distributed the asymptotic (co-)variance can be reduced by using the sample covariance matrix. This is due to the fact that the original data contains not only angular but also radial information which can be utilized if the true model is known. For a nice discussion of the interplay between robustness and efficiency of covariance matrix M-estimators see Oja (2003).

Hence the spectral estimator is a completely robust alternative if nothing is known except that the data is generalized elliptically distributed. Under this assumption fortunately not only weak consistency but also strong consistency can be established.

Theorem 46 (Strong consistency) *Let the conditions of Theorem 44 be fulfilled. Then the spectral estimator is strongly consistent, i.e.*

$$\hat{\Sigma}_n \xrightarrow{a.s.} \Sigma, \quad n \longrightarrow \infty.$$

Proof. Tyler (1987a) proves the strong consistency under the assumption that the sample stems from an arbitrary continuous multivariate distribution. Then Σ is to be interpreted as the solution of the fixed-point equation

$$\Sigma = d \cdot E \left(\frac{XX'}{X'\Sigma^{-1}X} \right),$$

rather than as a dispersion matrix. But in the case of generalized elliptical distributions Σ corresponds to the dispersion matrix. Recall that the spectral estimator can be represented by the projected data (cf. Eq. 4.9), i.e.

$$\hat{\Sigma} = \frac{d}{n} \cdot \sum_{j=1}^n \frac{SS'}{S'\hat{\Sigma}^{-1}S},$$

where

$$S = \frac{X - \mu}{\|X - \mu\|_2},$$

which is also generalized elliptically distributed (cf. Proposition 36). Due to the positive definiteness of Σ the random vector S has a continuous distribution given by the spectral density function ψ . ■

Of course, the weak consistency follows immediately by the strong consistency. But the former one can be proved straightforward using classical maximum-likelihood theory whereas the latter one is non-trivial.

5.3 Asymptotic Covariance Matrix

In Section 5.1 it was shown that the elementary information matrix of the spectral estimator providing only a generalized elliptical random vector with $\Sigma = \sigma^2 I_d$ converges to the elementary information matrix of the sample covariance matrix given $X \sim N_d(0, I_d)$. Since the convergence refers to $d \rightarrow \infty$ it is not clear whether $\mathcal{J}_0 \rightarrow \mathcal{I}_0$ implies $\mathcal{J}_0^{-1} \rightarrow \mathcal{I}_0^{-1}$, i.e. if not only the information matrices but also the asymptotic covariance matrices converge. This is an inverse problem and thus it is appropriate to calculate \mathcal{J}_0^{-1} , explicitly.

Lemma 47 *Let $M \in \mathbb{R}^{n \times n}$ be of the form*

$$M = \begin{bmatrix} 1 & a & \cdots & a \\ a & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ a & \cdots & \cdots & 1 \end{bmatrix},$$

where $a \neq -1/(n-1)$ and $a \neq 1$. Then the inverse of M corresponds to

$$M^{-1} = \begin{bmatrix} x & y & \cdots & y \\ y & x & & \vdots \\ \vdots & & \ddots & \vdots \\ y & \cdots & \cdots & x \end{bmatrix}, \quad (5.1)$$

where

$$x = \frac{1 + (n-2) \cdot a}{1 + (n-2) \cdot a - (n-1) \cdot a^2},$$

and

$$y = -\frac{a}{1 + (n-2) \cdot a - (n-1) \cdot a^2}.$$

Proof. Assume that M^{-1} has the form (5.1) with $x, y \in \mathbb{R}$. Necessarily $MM^{-1} = I_n$, i.e.

$$\begin{aligned} x + (n-1) \cdot ay &= 1, \\ ax + (1 + (n-2) \cdot a) \cdot y &= 0. \end{aligned}$$

This is a system of linear equations with the solutions x and y given in the lemma. ■

Proposition 48 *Let the d -dimensional random vector $X \stackrel{d}{=} \mu + \mathcal{R}(\sigma I_d) U^{(d)}$ be generalized elliptically distributed with $P(\mathcal{R} = 0) = 0$ and $\sigma > 0$. Then the asymptotic covariance*

matrix of the spectral estimator corresponds to

$$\mathcal{J}_0^{-1} \equiv [\mathcal{J}_{0,ij}^{-1}] := \begin{cases} 4 \cdot \frac{d+2}{d}, & i = j = 1, \dots, d-1, \\ 2 \cdot \frac{d+2}{d}, & i, j = 1, \dots, d-1, i \neq j, \\ \frac{d+2}{d}, & i = j = d, \dots, \binom{d+1}{2} - 1, \\ 0, & \text{else.} \end{cases}$$

Proof. Due to Corollary 43 we know that the information matrix corresponds to

$$\mathcal{J}_0 = \begin{bmatrix} \begin{bmatrix} b & a & \cdots & a \\ a & b & & \vdots \\ \vdots & & \ddots & \vdots \\ a & \cdots & \cdots & b \end{bmatrix} & 0 \\ 0 & cI_{\binom{d}{2}} \end{bmatrix} =: \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix},$$

where

$$a = -\frac{1}{2} \cdot \frac{1}{d+2}, \quad b = \frac{1}{2} \cdot \frac{d-1}{d+2}, \quad c = \frac{d}{d+2},$$

with $A \in \mathbb{R}^{(d-1) \times (d-1)}$ and $B \in \mathbb{R}^{\binom{d}{2} \times \binom{d}{2}}$. The inverse of B is simply

$$B^{-1} = \frac{1}{c} \cdot I_{\binom{d}{2}} = \frac{d+2}{d} \cdot I_{\binom{d}{2}}.$$

Consider the matrix

$$M^* := 2 \cdot \frac{d+2}{d-1} \cdot A = \begin{bmatrix} 1 & -\frac{1}{d-1} & \cdots & -\frac{1}{d-1} \\ -\frac{1}{d-1} & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ -\frac{1}{d-1} & \cdots & \cdots & 1 \end{bmatrix}.$$

For the inverse of M^* apply Lemma 47 so as to obtain

$$M^{*-1} = \frac{d-1}{d} \cdot \begin{bmatrix} 2 & 1 & \cdots & 1 \\ 1 & 2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 1 & \cdots & \cdots & 2 \end{bmatrix},$$

and thus

$$A^{-1} = \frac{d+2}{d} \cdot \begin{bmatrix} 4 & 2 & \cdots & 2 \\ 2 & 4 & & \vdots \\ \vdots & & \ddots & \vdots \\ 2 & \cdots & \cdots & 4 \end{bmatrix}.$$

Now bringing A^{-1} and B^{-1} together leads to the asymptotic covariance matrix given in the proposition. \blacksquare

In contrast, it is easy to verify that the asymptotic covariance matrix of the sample covariance matrix corresponds to

$$\mathcal{I}_0^{-1} \equiv [\mathcal{I}_{0,ij}^{-1}] = \begin{cases} 2, & i = j = 1, \dots, d-1, \\ 1, & i = j = d, \dots, \binom{d+1}{2} - 1, \\ 0, & \text{else,} \end{cases} \quad (5.2)$$

provided $X \sim N_d(\mu, \sigma^2 I_d)$. Hence,

$$\mathcal{J}_0^{-1} \not\rightarrow \mathcal{I}_0^{-1}, \quad d \longrightarrow \infty,$$

since the asymptotic variances of the main diagonal entries of the spectral estimator equal to 4 (and not to 2). Further, an interesting fact is that the asymptotic covariances of the main diagonal entries of the spectral estimator equal to 2 and do not vanish as $d \rightarrow \infty$. Only concerning the asymptotic covariances of the off diagonal elements the spectral estimator behaves like the sample covariance matrix.

Tyler (1987a) gives a representation of the asymptotic covariance matrix of the spectral estimator with arbitrary (but positive definite) dispersion matrix Σ . But it should be mentioned that Tyler derives the asymptotic covariance matrix of

$$\text{vec} \left(d \cdot \frac{\widehat{\Sigma}}{\text{tr}(\Sigma^{-1} \widehat{\Sigma})} \right)$$

and not of $\text{vec}(\widehat{\Sigma})$ or its relevant elements, i.e. only the one contained in a triangular part of $\widehat{\Sigma}$. Hence the considered asymptotic covariance matrix is not positive definite and cannot be compared directly with the results given in Proposition 48.

Further, Tyler (1983) compares the asymptotic variances of several M-estimators for different elliptical populations and dimensions. He also refers to a Monte Carlo simulation study of robust covariance matrix estimators done by Devlin, Gnanadesikan, and Kettenring (1981) which can be used for comparing the finite sample properties rather than the large sample behavior. Kent and Tyler (1988) adopt the spectral estimator for the parameters of a wrapped Cauchy distribution. Further, in Tyler (1987b) the spectral estimator is demonstrated on testing for uniformity and circularity. Dümbgen and Tyler (2004) show that under very general distributional assumptions the contamination breakdown point of the spectral estimator corresponds to $1/d$. Further properties of the spectral estimator are examined by Maronna and Yohai (1990) and by Adrover (1998).

Part II

Applications

Chapter 6

Motivation

‘Certain people find it intellectually stimulating that, eventually, there is a formula which explains a difficult procedure such as the pricing of an option very much in the same way as Newton’s law gives a quantitative description of gravitation.’

(T. Mikosch, 2003)

6.1 Empirical Evidence of Extremes

Financial data usually exhibit similar properties which are called ‘stylized facts’, i.e. heavy tails, extremal dependence, distributional asymmetry, volatility clustering, etc.; especially if the log-price changes (called the ‘log-returns’) of stocks, stock indices, and foreign exchange rates are considered (see, e.g., Eberlein and Keller, 1995, Embrechts, Klüppelberg, and Mikosch, 2003, Section 7.6, Fama, 1965, Mandelbrot, 1963, Mikosch, 2003, Chapter 1). Enumerating all the empirical studies on this topic would go beyond the scope of this preamble.

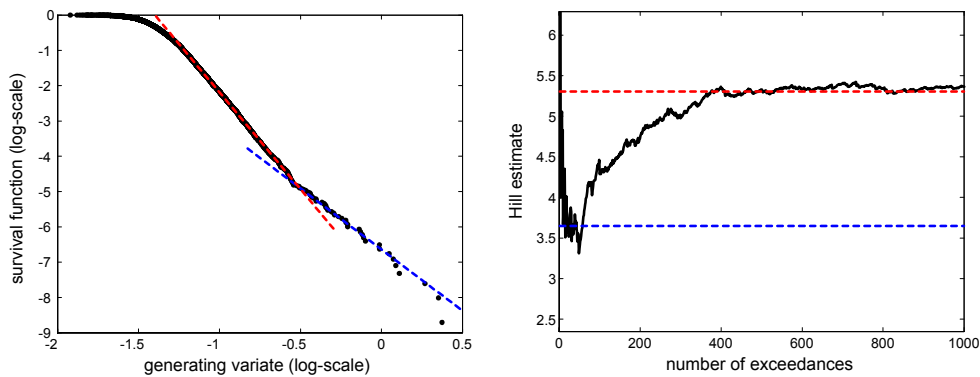


Figure 6.1 Empirical survival function of the generating variate of S&P 500 daily log-returns on a log-log-scale for the sample period 1980-01-02 to 2003-11-26 (left hand) and corresponding Hill-plot for the largest 1000 data points (right hand).

On the left hand of Figure 6.1 we see the empirical survival function (on a log-log-scale) of the generating variate of the daily log-returns of the current 285 S&P 500 stocks which had an IPO date before 1980-01-02. The sample period ranges from 1980-01-02 to 2003-11-26.

The generating variate is estimated according to the method described in Section 2.4 where the spectral estimator is used as covariance matrix estimator. We see that the decay is linear on the log-log-scale. This corresponds to the regular variation property of financial data described in Section 2.1, i.e.

$$\overline{F}(x) \approx \lambda \cdot x^{-\alpha} \quad \Rightarrow \quad \log \overline{F}(x) \approx \log \lambda - \alpha \cdot \log x, \quad \forall x > v > 0,$$

for a sufficiently high threshold v . Remarkably, we find two different slopes given by the dashed red line and the dashed blue line, respectively. The right hand side of Figure 6.1 shows the corresponding two different Hill estimates (Embrechts, Klüppelberg, and Mikosch, 2003, p. 330) of the tail index α . The dashed blue line corresponds to the tail index estimated by the 68 largest extremes whereas the dashed red line seems to indicate a ‘long-run tail index’ and corresponds to the usual size observed for daily log-returns. This indicates that there may be a sort of ‘mixture of tail indices’ in the S&P 500. An investigation of time-varying tail indices of the German DAX can be found in Wagner (2003). Wagner (2004) proposes a model of time-varying tail indices not only for stock returns but also for daily changes of government bond yield spreads.

This thesis concentrates on the extremal behavior of log-returns and their asymmetry. This was shown also in Section 3.1 for the NASDAQ and S&P 500. Further evidence against the Gaussian distribution hypothesis concerning the dependence structure of single stocks can be found in Junker (2002) and Junker and May (2002) whereas the dependence structures between foreign stock markets are investigated in Costinot, Roncalli, and Teïletche, 2000. In Section 2.4 the negative impact of extremal events on covariance matrix estimation was discussed. Moreover, Section 2.3.2 dealt with *simultaneous* extremes which are characteristic for financial markets. In the following figure we see the total numbers of S&P 500 stocks whose absolute values of daily log-returns exceeded 10% for each trading day during 1980-01-02 to 2003-11-26. Actually, on the 19th October 1987 (i.e. the ‘Black Monday’) there occurred 239 extremes. This is suppressed for the sake of transparency.

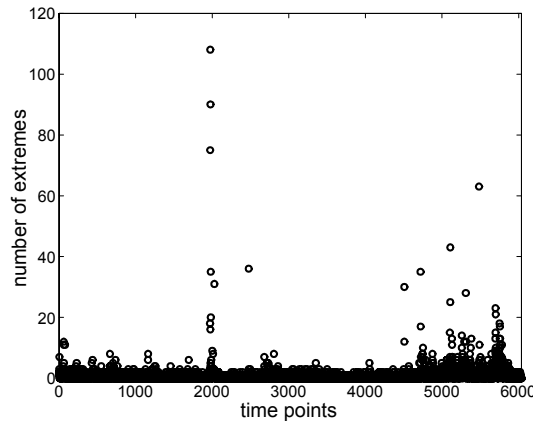


Figure 6.2 Number of extremes in the S&P 500 index during 1980-01-02 to 2003-11-26.

The latter figure shows the concomitance of extremes. If extremes would occur independently then the number of extremal events (no matter if losses or profits) should be small and all but constant over time. Obviously, this is not the case. In contrast we see the October crash of 1987 and several extremes which occur permanently since the beginning of the bear market in the middle of 2000.

The next figure serves to exemplify the effect of simultaneous extremes in more detail. We compare General Motors and Hilton to point out that simultaneous extremes even may occur if the linear dependence between two assets is rather small because of completely different lines of business.

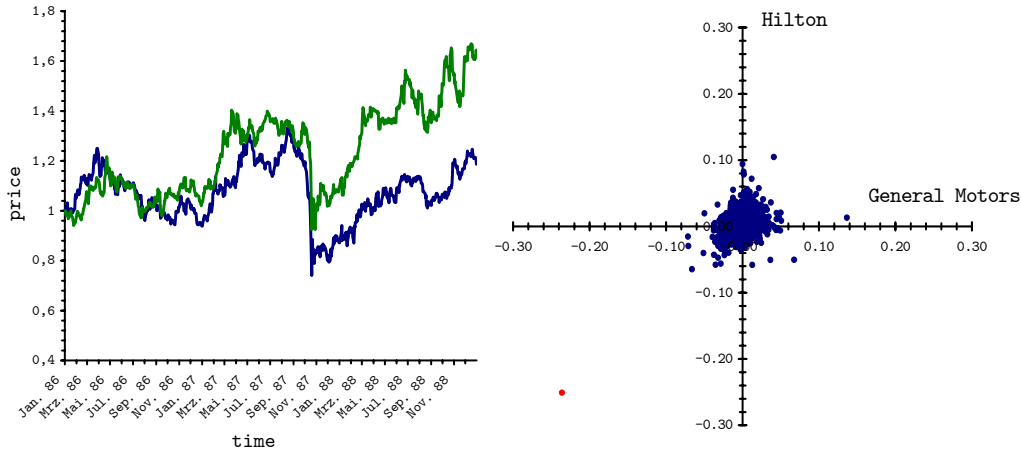


Figure 6.3 Daily prices of General Motors (blue line) and Hilton (green line) from 1986-01-01 to 1988-12-30 (left hand) and the corresponding log-returns (right hand). The red point corresponds to the October crash, 1987.

For a profound discussion of dependence structures on financial markets see Breymann, Dias, and Embrechts (2003), Junker (2003), and Schmidt (2003b).

Financial markets are characterized not only by the presence of extreme losses but also by a tremendous amount of financial instruments and risk factors like, e.g., shares, options, futures, foreign exchanges rates, interest rates, etc. Hence one of the main goals of this work is to develop an estimation procedure both suitable for the analysis of high-dimensional data and robust against extreme price fluctuations.

6.2 On the Curse of Dimensions

Consider a d -dimensional random vector U which is uniformly distributed on the unit hypercube $[0, 1]^d$ and let U_1, \dots, U_n be a sequence of independent copies of U . Hence, the expected number of data points of an arbitrary component U_i of U lying in an interval of length π , i.e.

$$0 \leq a < U_i \leq a + \pi \leq 1$$

equals to $n\pi$ and is independent of d . But the expected number of data points lying within a sub-cube of $[0, 1]^d$ with length $\pi \in]0, 1[$ corresponds to $n\pi^d$.

Let the length of a sub-cube for which the expected number of realizations corresponds to 1 be denoted by $\tilde{\pi} := n^{-1/d}$. The smaller $\tilde{\pi}$ the more dense the data. If one does not know the true law of U and tries to estimate its joint distribution function with a given accuracy $\tilde{\pi}$ then the minimum sample size corresponds to $n = \tilde{\pi}^{-d}$. Hence the sample size has to increase exponentially with the number of dimensions to prevent a loss of accuracy. This is usually called the ‘curse of dimensions’.

In Section 5.1 the special role of the effective sample size $q = n/d$ was prementioned. Consider the following problem. Let X be a d -dimensional t -distributed random vector with 3 degrees of freedom, i.e. $X \sim t_d(0, I_d, 3)$ (see Example 4). Since its generating variate corresponds to $\sqrt{d \cdot F_{d,3}}$ the variance of each component of X equals to $\text{Var}(X_i) = E(d \cdot F_{d,3})/d = 3/(3-2) = 3$. Note that the components of X are uncorrelated. Now, let $Y := X/\sqrt{3}$ and Y_1, \dots, Y_n be a sequence of independent copies of Y . Due to the multivariate central limit theorem (see, e.g., Hayashi, 2000, p. 96) we expect

$$Z_n := \frac{1}{\sqrt{n}} \cdot \sum_{j=1}^n Y_{\cdot j} \xrightarrow{d} N(0, I_d), \quad n \longrightarrow \infty,$$

and consequently the sum of the squared components of Z_n should be χ_d^2 -distributed, asymptotically.

The subsequent figures show histograms of Z_n for $n = 100$ and different numbers of dimension. The density function of the corresponding χ_d^2 -law is represented by the green line. Obviously, the finite sample property of Z_{100} depends essentially on d , i.e. the smaller the number of dimensions the better the central limit theorem works.

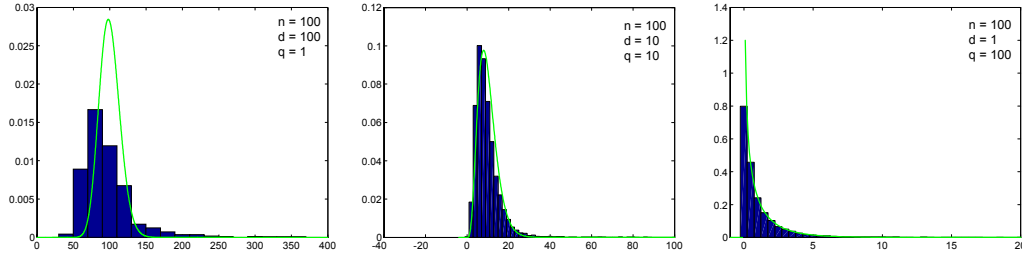


Figure 6.4 Histograms of Z_{100} for different numbers of dimension. The corresponding χ_d^2 -law is represented by the green line.

In the next figure consider Z_n for a sample size of $n = 1000$. Remarkably, even for this large sample size the central limit theorem does not apply in the case of $d = 1000$ (upper left). This is the same as for the left picture of Figure 6.4. The smaller d , i.e. the larger $q = n/d$ the better the fit of the corresponding χ_d^2 -law. For $q \geq 100$ we see that the central limit theorem makes an impact.

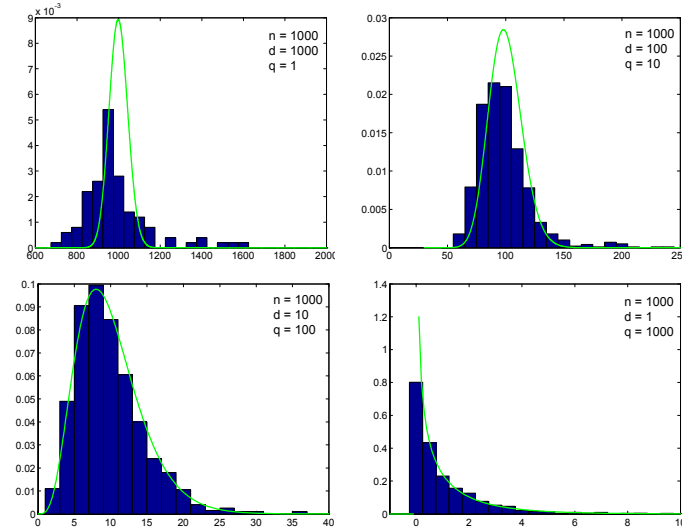


Figure 6.5 Histograms of Z_{1000} for different numbers of dimension. The corresponding χ_d^2 -law is represented by the green line.

The next figure shows the same effect even for the sample size of $n = 10000$. For $d = 1000$ the central limit theorem does not apply again due to the small effective sample size $q = 10$. But if $q \rightarrow \infty$ the central limit theorem holds as before.

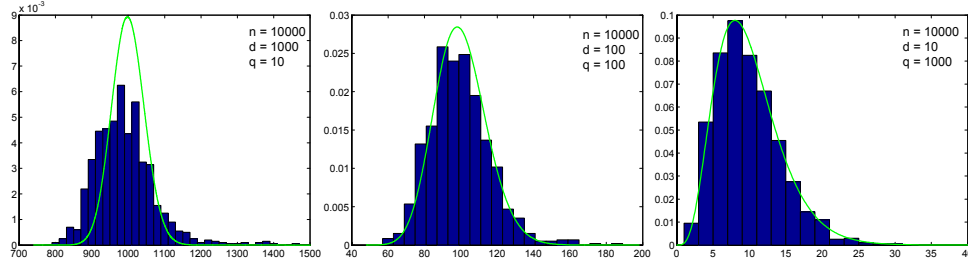


Figure 6.6 Histograms of Z_{10000} for different numbers of dimension. The corresponding χ_d^2 -law is represented by the green line.

It is somewhat surprising that the finite sample property of Z_n depends on the number of dimensions even though every component of Z_n , i.e. Z_{in} , $i = 1, \dots, d$, is asymptotically normal distributed where its finite sample property a priori does not depend on d . Further on, the components of Z_n are uncorrelated, not only asymptotically but even in the finite samples. But for all that the random vector Z_n is not normally distributed, approximately, for small q . This is because normality of the margins do not imply joint normality. Moreover, uncorrelated normally distributed random components are not necessarily independent. In practical situations this may be a typical source of misunderstandings (Embrechts, McNeil, and Straumann, 2002).

With the help of copulas one may easily construct distribution functions with normal margins and uncorrelated components (cf. Section 2.2). For instance, let the multivariate c.d.f. be defined by

$$x \mapsto F(x) = C(\Phi(x_1), \dots, \Phi(x_d)),$$

where Φ is the univariate standard normal c.d.f. and C is the copula of a spherical distribution. Note that due to the copula also the components of $X \sim F$ are uncorrelated.

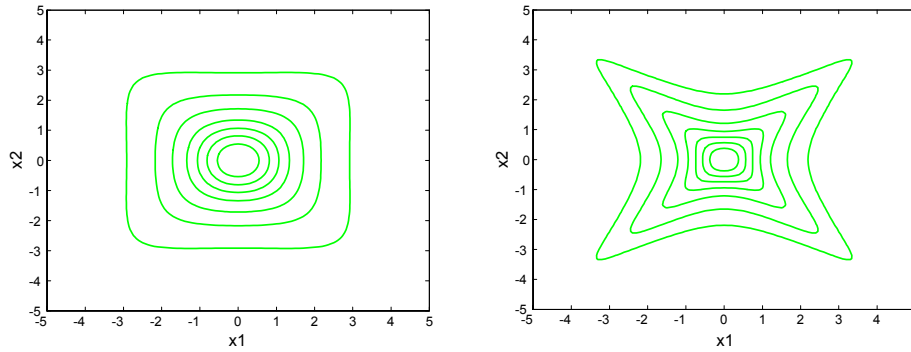


Figure 6.7 Density contours of bivariate distribution functions with standard normal margins but $t_2(0, I_2, \nu)$ -copulas. The degrees of freedom are $\nu = 5$ (left hand) and $\nu = 1$ (right hand).

Thus for a proper statistical analysis of multidimensional problems one should distinguish the following asymptotics:

- $n \rightarrow \infty, d \text{ const.}:$ classical asymptotics,
- $n \rightarrow \infty, d \rightarrow \infty, \text{ and } n/d \rightarrow \infty:$ classical asymptotical results may hold anymore,
- $n \rightarrow \infty, d \rightarrow \infty, \text{ but } n/d \rightarrow q < \infty:$ non-classical asymptotics (random matrix theory).

The latter case belongs to the domain of random matrix theory which will be discussed in Chapter 8.

Chapter 7

Applications in Finance

Now the methodology developed so far will be related to financial applications like, e.g., portfolio optimization and Beta estimation. The particular properties of the spectral density approach are compared with the conventional approach which means using the sample covariance matrix.

7.1 Modern Portfolio Theory

We start with modern portfolio theory (MPT) developed by Markowitz (1952) and continued by Tobin (1958), Sharpe (1964) and Lintner (1965).

7.1.1 Portfolio Optimization

Consider a frictionless market with a constant number of d risky financial assets having elliptically distributed daily log-returns $(X_{it})_{t \in \mathbb{Z}}$, i.e.

$$X_{it} := \log P_{it} - \log P_{i,t-1}, \quad i = 1, \dots, d, \quad \forall t \in \mathbb{Z},$$

where P_{it} symbolizes the price of asset i at time t . Further, it is assumed that there exists a riskless bond with a constant log-return (the ‘risk free interest rate’). The price of an asset only vanishes when the corresponding company becomes bankrupt. This event is assumed to be impossible, particularly because it would be an absorbing state. Thus the log-returns are well-defined (a.s.).

Further, it is assumed that the vectors of log-returns $(X_t, \forall t \in \mathbb{Z})$ have finite cross moments of second order and that the centered log-returns $X_t - \mu$ are ergodic stationary martingale difference sequences (Hayashi, 2000, p. 104), i.e. $E(X_t) = \mu$ and particularly

$$E(X_t | X_{t-1}, X_{t-2}, \dots) \stackrel{\text{a.s.}}{=} \mu,$$

for each $t \in \mathbb{Z}$. Ergodic stationarity (Hayashi, 2000, p. 101) refers to the property of (X_t) to be stationary and, additionally, for any two bounded functions $f : \mathbb{R}^k \rightarrow \mathbb{R}$ and $g : \mathbb{R}^l \rightarrow \mathbb{R}$,

$$\begin{aligned} |E(f(X_t, \dots, X_{t+k-1})g(X_{t+n}, \dots, X_{t+n+l-1}))| &\longrightarrow \\ |E(f(X_t, \dots, X_{t+k-1}))| \cdot |E(g(X_t, \dots, X_{t+l-1}))|, &\quad n \longrightarrow \infty. \end{aligned}$$

Thus, two cut-outs of a (multivariate) ergodic time series become more uncorrelated the larger the lag between each other. This is given for a stationary ARCH(1)-process (cf. Hayashi, 2000, p. 106), for instance.

Suppose that Σ is the positive definite *covariance* matrix of X_t . Then

$$\sqrt{T} \cdot \left(\frac{1}{T} \sum_{t=1}^T X_{it} - \mu \right) \xrightarrow{d} N_d(0, \Sigma), \quad T \longrightarrow \infty.$$

This is the central limit theorem for ergodic stationary martingale differences (Hayashi, 2000, p. 106). Hence, given a sufficiently long target horizon, say at least one year ($T \approx 252$), the sum of log-returns is approximately normally distributed. Note that the sum of log-returns $\sum_{t=1}^T X_{it}$ ($i = 1, \dots, d$) coincides with the log-return of investment i over the target period. Thus one may justify the Gaussian distribution assumption regarding *long-term* log-returns even under the relatively weak condition of ergodicity.

Unfortunately, for MPT we need to consider discretely compounded returns

$$R_i := \frac{P_{iT} - P_{i0}}{P_{i0}}, \quad i = 1, \dots, d,$$

rather than log-returns. Commonly, also discretely compounded returns are assumed to be multivariate normally distributed. But since asset prices cannot become negative there is a ‘natural’ infimum of -1 for discrete returns. Hence the Gaussian distribution hypothesis can only serve as a kluge. But we will follow the classical argument of MPT for the sake of simplicity.

The return of a portfolio $w = (w_0, w_1, \dots, w_d)$ is given by

$$R_P := \sum_{i=0}^d w_i R_i,$$

where $\sum_{i=0}^d w_i = 1$ but the weights may become negative. Here $R_0 \equiv r \geq 0$ symbolizes the risk free interest rate. Consequently, the expected portfolio return is given by $\mu_P := \sum_{i=0}^d w_i \mu_i$, where $\mu_i := E(R_i)$, $i = 0, 1, \dots, d$.

Suppose that each investor has an exponential utility function

$$x \longmapsto u(x) = -\exp(-\gamma x), \quad \gamma > 0, \quad (7.1)$$

where γ is an individual risk aversion parameter. Hence the expected utility is given by

$$E(u(R_P)) = -E(\exp(-\gamma R_P)).$$

Consider that $E(\exp(-\gamma R_P))$ is the moment generating function of R_P at $-\gamma$, i.e.

$$E(u(R_P)) = -\exp\left(-\gamma \cdot \left(\mu_P + \frac{1}{2} \cdot (-\gamma) \cdot \sigma_P^2\right)\right),$$

where σ_P^2 represents the portfolio variance. Define the ‘certainty equivalent’ $\zeta(R_P)$ by the solution of the equation

$$u(\zeta(R_P)) = E(u(R_P)).$$

Hence the certainty equivalent corresponds to a riskless portfolio return which gives the same (expected) utility as the risky return R_P . We see that

$$\zeta(R_P) = \mu_P - \frac{1}{2} \cdot \gamma \sigma_P^2.$$

This is the well-known objective function of portfolio optimization (the ‘mean-variance utility function’).

Due to the budget constraint $\sum_{i=0}^d w_i = 1$ we can substitute w_0 by $w_0 = 1 - \sum_{i=1}^d w_i$. Thus

$$\begin{aligned}\zeta(R_P) &= \left(1 - \sum_{i=1}^d w_i\right) \cdot r + \sum_{i=1}^d w_i \mu_i - \frac{1}{2} \cdot \gamma \cdot \sum_{i,j=1}^d w_i w_j \text{Cov}(R_i, R_j) \\ &= r + \sum_{i=1}^d w_i (\mu_i - r) - \frac{1}{2} \cdot \gamma \cdot \sum_{i,j=1}^d w_i w_j \text{Cov}(R_i, R_j).\end{aligned}$$

If we define the ‘excess return’ $\Delta\mu_i := \mu_i - r$, $i = 1, \dots, d$, and the vector of stock weights $\tilde{w} := (w_1, \dots, w_d)$ we obtain

$$\zeta(R_P) = r + \tilde{w}' \Delta\mu - \frac{\gamma}{2} \cdot \tilde{w}' \Sigma \tilde{w}, \quad (7.2)$$

where Σ symbolizes the covariance matrix of $R = (R_1, \dots, R_d)$. Thus, maximizing the mean variance utility function is a simple quadratic optimization problem which has the solution $\Delta\mu - \gamma \Sigma \tilde{w} = 0$, i.e. $\tilde{w} = \Sigma^{-1} \Delta\mu / \gamma$. Note that the sum of the components of \tilde{w} generally does not coincide with 1. Indeed, the sum of the stock weights depends essentially on the investor’s individual risk aversion. But the optimal stock portfolio is always given by

$$\omega := \frac{\tilde{w}}{\mathbf{1}' \tilde{w}} = \frac{\Sigma^{-1} \Delta\mu}{\mathbf{1}' \Sigma^{-1} \Delta\mu}, \quad (7.3)$$

i.e. the optimal portfolio of risky assets does not depend on the particular risk aversion of the investor, provided there is a money market (Tobin, 1958). This is known as ‘Tobin’s (Two-fund) Separation Theorem’. Regard also that ω does not depend on the scale of Σ , too. Hence, the optimal capital allocation can be estimated by

$$\hat{\omega} = \frac{\hat{\Sigma}^{-1} (\hat{\mu} - r \mathbf{1})}{\mathbf{1}' \hat{\Sigma}^{-1} (\hat{\mu} - r \mathbf{1})}, \quad (7.4)$$

where $\hat{\mu}$ and $\hat{\Sigma}$ may be some robust estimates of μ and Σ .

If $\mu_1 = \dots = \mu_d > r$ then the optimal solution is simply given by

$$\omega_0 := \frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}' \Sigma^{-1} \mathbf{1}}. \quad (7.5)$$

In that case maximizing the quadratic function (7.2) is equivalent to minimizing the portfolio risk (which is given by $\tilde{w}' \Sigma \tilde{w}$) since the expected return of the stock portfolio is not affected by changing the portfolio weights. Indeed, ω_0 is the optimal solution if the investor per se is not interested in portfolio optimization but risk minimization no matter if the expected returns are equal or not. Thus ω_0 is called ‘global minimum variance portfolio’ (Kempf and Memmel, 2002).

7.1.2 Portfolio Weights Estimation

Now it is assumed that the time series of log-returns (X_t) is not only ergodic stationary but i.i.d. for the sake of simplicity. From now on we will reconsider continuously compounded returns, i.e.

$$R_i := \sum_{t=1}^T X_{it}, \quad i = 1, \dots, d,$$

where the target horizon T is assumed to be large. Suppose that the mean vector μ is known and that the positive definite matrix Σ is estimated by the sample covariance matrix. Thus we can assume $\mu = 0$ (w.l.o.g.).

Since (X_t) is supposed to be i.i.d. the covariance matrix of $R = (R_1, \dots, R_d)$ corresponds to

$$\Sigma = n \cdot \text{Var}(X_t).$$

Thus Σ can be estimated directly by using the daily observations, i.e.

$$\widehat{\Sigma} = n \cdot \widehat{\text{Var}}(X_t).$$

Suppose that $\widehat{\Sigma}$ is estimated via the sample covariance matrix. Note that it does not matter if we insert $\widehat{\Sigma}$ or $\widehat{\text{Var}}(X_t)$ in Eq. 7.4. So let $\widehat{\Sigma} = [\hat{\sigma}_{ij}]$ denote the estimated covariance matrix on the daily basis and $\Sigma = [\sigma_{ij}]$ be the corresponding true covariance matrix.

If the daily log-returns have finite fourth order cross moments then the sample covariance matrix is asymptotically (matrix-valued) normally distributed. This is a direct consequence of the central limit theorem. More precisely, each component of $\widehat{\Sigma}$ exhibits

$$\sqrt{n} \cdot (\hat{\sigma}_{ij} - \sigma_{ij}) \xrightarrow{d} N(0, \text{Var}(X_{it}X_{jt})), \quad n \longrightarrow \infty.$$

Here n is the sample size, i.e. the length of the observed time series (X_t) and not the target horizon T . Further,

$$\text{Var}(X_{it}X_{jt}) = E(X_{it}^2X_{jt}^2) - E^2(X_{it}X_{jt}) = E(X_{it}^2X_{jt}^2) - \sigma_{ij}^2,$$

where $\sigma_{ij} := \text{Cov}(X_{it}, X_{jt})$. We see that the asymptotic variance of each component of $\widehat{\Sigma}$ depends essentially on the fourth order cross moments of the components of X_t . One can interpret the term $E(X_{it}^2X_{jt}^2)$ as ‘cross kurtosis’.

Not only the asymptotic variances but also the asymptotic covariances depend particularly on the kurtosis of the components of X_t since (Praag and Wesselman, 1989 and Tyler, 1983)

$$\text{Cov}(X_{it}X_{jt}, X_{kt}X_{lt}) = (1 + \kappa) \cdot (\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) + \kappa \cdot \sigma_{ij}\sigma_{kl},$$

where

$$\kappa := \frac{1}{3} \cdot \frac{E(X_{it}^4)}{E^2(X_{it}^2)} - 1$$

is called ‘kurtosis parameter’. Note that the kurtosis parameter is the same for every i because it does not depend on the scale of X_{it} . It is well-known that in the case of normality $\kappa = 0$. A distribution with positive (or even infinite) κ is called ‘leptokurtic’. Particularly, regularly varying distributions are leptokurtic.

Suppose for the sake of simplicity that X_t is spherically distributed, i.e. $\Sigma = \sigma^2 I_d$. Since the vector of optimal portfolio weights is invariant under scale transformations we may assume w.l.o.g. that $\Sigma = I_d$. From Theorem 5 we know that

$$E(X_{it}^4) = E\left(\left(\mathcal{R}_t U_{it}^{(d)}\right)^4\right) = E(\mathcal{R}_t^4) \cdot E\left(\left(U_{it}^{(d)}\right)^4\right) = \frac{3 \cdot E(\mathcal{R}_t^4)}{d(d+2)},$$

and

$$\begin{aligned} E(X_{it}^2 X_{jt}^2) &= E\left(\left(\mathcal{R}_t U_{it}^{(d)}\right)^2 \left(\mathcal{R}_t U_{jt}^{(d)}\right)^2\right) \\ &= E(\mathcal{R}_t^4) \cdot E\left(\left(U_{it}^{(d)}\right)^2 \left(U_{jt}^{(d)}\right)^2\right) = \frac{E(\mathcal{R}_t^4)}{d(d+2)}, \end{aligned}$$

for $i \neq j$. Note that $E(\mathcal{R}_t^2) = d$ since we assume that Σ represents the covariance matrix (cf. Section 1.2.3).

Example 16 (Asymptotic variances if $X_{\cdot t} \sim N_d(0, I_d)$) Let $\mathcal{R}_t = \sqrt{\chi_d^2}$, that is to say $X_{\cdot t} \sim N_d(0, I_d)$. Then

$$E(\mathcal{R}_t^4) = E((\chi_d^2)^2) = \text{Var}(\chi_d^2) + E^2(\chi_d^2) = 2d + d^2 = d(d+2).$$

Hence $E(X_{it}^4) = 3$, i.e. $\text{Var}(X_{it}^2) = 3 - 1 = 2$, and $\text{Var}(X_{it}X_{jt}) = 1$, $i \neq j$ (see also Eq. 5.2).

Now, let $X_{\cdot t}$ be multivariate t -distributed with covariance matrix I_d and $\nu > 4$ degrees of freedom. Since the generating variate must satisfy $E(\mathcal{R}_t^2) = d$ for all ν we obtain

$$\mathcal{R}_t = \sqrt{\frac{\nu-2}{\nu}} \cdot \sqrt{d \cdot F_{d,\nu}}, \quad \forall \nu > 4,$$

rather than $\sqrt{d \cdot F_{d,\nu}}$ (cf. Example 4). Then

$$E(\mathcal{R}_t^4) = \left(\frac{\nu-2}{\nu}\right)^2 \cdot d^2 \cdot E(F_{d,\nu}^2).$$

The second moment of $F_{d,\nu}$ corresponds to (cf. Johnson, Kotz, and Balakrishnan, 1995, p. 325)

$$E(F_{d,\nu}^2) = \left(\frac{\nu}{d}\right)^2 \cdot \frac{d(d+2)}{(\nu-2)(\nu-4)}.$$

Hence

$$E(\mathcal{R}_t^4) = d(d+2) \cdot \frac{\nu-2}{\nu-4},$$

and $E(X_{it}^4) = 3 \cdot (\nu-2) / (\nu-4)$, i.e.

$$\text{Var}(X_{it}^2) = 3 \cdot \frac{\nu-2}{\nu-4} - 1 = 2 \cdot \frac{\nu-1}{\nu-4},$$

as well as

$$\text{Var}(X_{it}X_{jt}) = \frac{\nu-2}{\nu-4}, \quad i \neq j.$$

Since it is assumed that the covariance matrix of $X_{\cdot t}$ corresponds to I_d the kurtosis parameter is simply given by

$$\kappa = \frac{1}{3} \cdot E(X_{it}^4) - 1 = \frac{\nu-2}{\nu-4} - 1 = \frac{2}{\nu-4}, \quad \nu > 4.$$

Even though the true covariance matrix remains the same under the variation of ν both the asymptotic variances of the main diagonal entries and the asymptotic variances of the off diagonal entries of the sample covariance matrix depend essentially on ν . For $\nu \rightarrow \infty$ we see that the asymptotic variances tend to the values given for the normal distribution hypothesis. But for $\nu \searrow 4$ the asymptotic variances tend to infinity and if $0 < \nu \leq 4$ the sample covariance matrix is no longer normally distributed, asymptotically.

In Section 5.3 (Proposition 48) it was shown that the asymptotic variance of the main diagonal entries of the spectral estimator in the case of isotropy corresponds to $4 \cdot (d+2)/d$, whereas the asymptotic variance of its off diagonal elements equals to $(d+2)/d$. Now, one may ask when the sample covariance matrix is dominated (in a componentwise manner) by the spectral estimator provided the data is multivariate t -distributed. Regarding the main diagonal entries this is given by

$$4 \cdot \frac{d+2}{d} < 2 \cdot \frac{\nu-1}{\nu-4},$$

i.e. if

$$\nu < 4 + 3 \cdot \frac{d}{d+4}.$$

Concerning the off diagonal entries we obtain

$$\frac{d+2}{d} < \frac{\nu-2}{\nu-4},$$

i.e. whenever if $\nu < 4 + d$ the variance of the off diagonal elements of the spectral estimator is smaller, asymptotically.

Until now we have investigated the asymptotic variances of the covariance matrix estimators using daily log-returns. Now we turn to the finite sample variances. Theoretically one could draw on intraday data or even high-frequency data to estimate the covariance matrix. Suppose that for each day there are m intraday observations available and denote the corresponding time series by $(Y_{\tau})_{\tau=1, \dots, nm}$. Further on we assume that (Y_{τ}) has i.i.d. innovations. Consequently,

$$Y_{\tau} \stackrel{d}{=} \frac{\mathcal{R}_t}{\sqrt{m}} \cdot U_t^{(d)}, \quad (7.6)$$

since on the basis of the i.i.d. assumption the covariance matrix of Y_{τ} must correspond to Σ/m . Let $\hat{\Sigma}_Y$ be the covariance matrix estimator applied on the high-frequency data (Y_{τ}) . Then the covariance matrix of the daily log-returns can be estimated by $m\hat{\Sigma}_Y$. The asymptotic variance of $m\hat{\Sigma}_Y$ corresponds to m^2 times the asymptotic variance of $\hat{\Sigma}_Y$. But due to (7.6) the asymptotic variance of $\hat{\Sigma}_Y$ corresponds to the asymptotic variance of $\hat{\Sigma}$ divided by m^2 , where $\hat{\Sigma}$ is the covariance matrix estimator applied on the daily data. Thus the asymptotic variance of $\hat{\Sigma}$ equals to the asymptotic variance of $m\hat{\Sigma}_Y$. But when using intraday data we have not only n but nm data points. So actually the finite sample variance of $m\hat{\Sigma}_Y$ is m times smaller than that of $\hat{\Sigma}$.

That is to say the higher the frequency of the available financial data the better the results of covariance matrix estimation. But that is only half the truth. Unfortunately, the components of high-frequency financial time series are usually not independent and even not stationary (Breyman, Dias, and Embrechts, 2003). Furthermore several empirical studies show that the tail index of financial data depends strongly on the time horizon (see, e.g., Breyman, Dias, and Embrechts, 2003, and Bouchaud, Cont, and Potters, 1998).

Theorem 9 states that the sum of elliptically distributed random vectors are elliptical, too, if the unit random vectors of each addend are independent and the dispersion matrix is constant. This holds even if the generating variates of the addends depend on each other. Hence the stylized facts of high-frequency data could be covered by a time series of dependent generating variates (\mathcal{R}_{τ}) where the tail index of each variate depends on its time horizon. Now, in contrast to the sample covariance matrix the spectral estimator leads to robust estimates because it does not depend on the generating variates but solely on the unit random vectors. But these are assumed to be i.i.d., anymore. Especially, the spectral estimator keeps its properties obtained by maximum-likelihood theory. Hence the spectral estimator is proposed for high-frequency data in order to minimize the finite sample variance.

Now we come to the estimation of μ . Suppose for the sake of simplicity that μ is estimated by the sample mean

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n X_t.$$

Following the same argument as for covariance matrix estimation we may use high-frequency data to estimate the expected daily log-return. This is done by

$$m\hat{\mu}_Y = m \cdot \left(\frac{1}{nm} \sum_{\tau=1}^{nm} Y_{\tau} \right),$$

which has an asymptotic variance of $m^2 \cdot \text{Var}(Y_\tau)$. The variance of Y_τ corresponds to the variance of X_t divided by m . So the asymptotic variance of the sample mean of high-frequency data is m times the asymptotic variance of $\hat{\mu}$. But since there are m times more high-frequency data points m is canceled down regarding the finite sample variance. We conclude that in contrast to covariance matrix estimation it does not matter if we use high-frequency data or not. In the same manner one may also use log-returns corresponding to the target horizon of portfolio optimization, say one year.

On a target horizon of one year the average mean log-return of the S&P 500 stocks considered so far correspond to 10%, approximately, whereas the average variance equals to 12%. Thus the average volatility (i.e. the average standard deviation) of the S&P 500 log-returns equals to 35%, approximately. Suppose for the sake of simplicity that the number N of observed years is an integer. Obviously, the finite sample variance of $\hat{\mu}$ corresponds to σ^2/N . Approximately 95% of the mean estimates lie within the interval

$$\left[\mu - 2 \cdot \frac{\sigma}{\sqrt{N}}, \mu + 2 \cdot \frac{\sigma}{\sqrt{N}} \right] = \left[0.1 - \frac{0.7}{\sqrt{N}}, 0.1 + \frac{0.7}{\sqrt{N}} \right].$$

So estimating μ accurately is practically impossible. Even if we have 50 years of data approximately 5% of the estimated expected returns still lie outside the interval 0.1 ± 0.1 .

We see that the main problem of portfolio optimization is given by expected return estimation rather than by covariance matrix estimation. This is pointed out also by Chopra and Ziemba (1993) and Kempf, Kreuzberg, and Memmel (2002). Using the sample median instead of the sample mean or, alternatively, the estimator given by the fixed-point solution of Eq. 4.12 seems not to give essentially better results.

One way to overcome this problem is to make certain restrictions about the expected returns like, e.g., allowing only for positive excess returns (Merton, 1980). Another way is to take additional data into consideration like, e.g., fundamental data, expertise, etc. (Mommel, 2004, Section 4.2). A somewhat tricky alternative is to assume that the expected returns of the risky assets are equal or that the investor is only interested in risk minimization. Then μ has not to be estimated at all and the optimal portfolio weights are given by

$$\hat{\omega}_0 := \frac{\hat{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}' \hat{\Sigma}^{-1} \mathbf{1}},$$

according to Eq. 7.5. In that case robust covariance matrix estimators like the spectral estimator lead to substantially better results than the sample covariance matrix estimator.

7.2 Principal Component Analysis

Now we presume that every investor on the market is risk averse according to the exponential utility function given by (7.1). Thus the optimal capital allocation of each investor is given by the corresponding formula (7.3). Within a market equilibrium per definition demand and supply must coincide. Hence the market portfolio ω_M , i.e. the vector of relative market capitalizations corresponds to the optimal portfolio of risky assets (Sharpe, 1964)

$$\omega_M = \omega = \frac{\Sigma^{-1} \Delta \mu}{\mathbf{1}' \Sigma^{-1} \Delta \mu}.$$

Consider that

$$\Delta \mu_M := \omega'_M \Delta \mu = \frac{\Delta \mu' \Sigma^{-1} \Delta \mu}{\mathbf{1}' \Sigma^{-1} \Delta \mu}$$

is the excess return of the market portfolio whereas

$$\sigma_M^2 := \omega'_M \Sigma \omega_M = \frac{\Delta \mu' \Sigma^{-1} \Delta \mu}{(\mathbf{1}' \Sigma^{-1} \Delta \mu)^2}$$

corresponds to the variance of the market portfolio return.

Define

$$\beta := \frac{\Sigma \omega_M}{\sigma_M^2} = \frac{\Delta \mu}{\underline{1}' \Sigma^{-1} \Delta \mu} \cdot \frac{(\underline{1}' \Sigma^{-1} \Delta \mu)^2}{\Delta \mu' \Sigma^{-1} \Delta \mu} = \Delta \mu \cdot \frac{\underline{1}' \Sigma^{-1} \Delta \mu}{\Delta \mu' \Sigma^{-1} \Delta \mu} = \frac{\Delta \mu}{\Delta \mu_M}.$$

Then

$$\Delta \mu = \Delta \mu_M \cdot \beta. \quad (7.7)$$

This is the fundamental equation of Sharpe's capital asset pricing model called the 'securities market line'. Let $R = (R_1, \dots, R_d)$ be the vector of asset returns. Note that β is nothing else but the vector of covariances between each asset return R_1, \dots, R_d and the market return $R_M := \omega'_M R$, divided by σ_M^2 . Thus Eq. 7.7 is more transparent in the form

$$\mu_i = r + (\mu_M - r) \cdot \beta_i, \quad (7.8)$$

where

$$\beta_i = \frac{Cov(R_i, R_M)}{Var(R_M)},$$

for $i = 0, 1, \dots, d$. If we define $\alpha_i := (1 - \beta_i)r = \mu_i - \beta_i \mu_M$ and $\varepsilon_i := R_i - (\alpha_i + \beta_i R_M)$, then

$$R_i = \alpha_i + \beta_i R_M + \varepsilon_i, \quad (7.9)$$

with $E(\varepsilon_i) = 0$ and $Cov(\varepsilon_i, R_M) = 0$. This is a regression function which describes the relationship between the asset returns and the market return where β_i is the regression coefficient of the i -th asset.

The regression function implies that every asset return consists mainly of two components, a market risk component and a firm specific, i.e. 'idiosyncratic' component. The firm specific risk is neither correlated with the market return nor it is systematic in the sense that it occurs with positive or negative expectation. One could interpret the market return as a common macro-economic risk factor whereas the idiosyncratic component represents the possibility that only a firm's captive property goes bust. Note that the idiosyncratic components $\varepsilon_1, \dots, \varepsilon_d$ may depend on each other.

So the return of an asset is particularly driven by its market sensitivity, i.e. its 'Beta'. The smaller the market sensitivity the smaller the expected return and vice versa. This is because the Beta specifies how far the corresponding asset can be used for the purpose of portfolio diversification. The larger the Beta, i.e. the bigger the firm's market exposure, the worse its diversification effect. This is 'compensated' by a larger expected return. Assets with vanishing or even negative Beta must be rare because otherwise they would constitute the market, i.e. their Beta would be positive which is a contradiction.

The Beta of an asset is of crucial importance for the pricing of corporates in practice. Since financial analysts usually try to forecast the future earnings of a company there must be a reasonable basis on which the future pay-offs shall be discounted. According to Eq. 7.8 this is given by the Beta. Indeed, Eq. 7.9 can be used for estimating the Beta by standard methods of multivariate analysis. This could be done for each asset separately. But since the idiosyncratic risks may depend on each other this is not an efficient way. It is more reasonable to estimate the Betas by linear regression with multiple equations (Hayashi, 2000, Chapter 4). Thus even though Eq. 7.9 is a short and sweet description of the relationship between each asset and the market portfolio the interdependences between the assets are tacitly shipped to the ε 's. But in fact there is no complexity or dimension reduction.

Instead of the market return R_M one could assume that there exists another random variable Y (and other noise variables $\varepsilon_1, \dots, \varepsilon_d$) such that

$$R_i = \alpha_i + \beta_i Y + \varepsilon_i,$$

holds for each asset where the noise variables $\varepsilon_1, \dots, \varepsilon_d$ are uncorrelated. This is known as Sharpe's single-index model (Sharpe, 1963). But for a better understanding of the interactions between each asset a multi-index model

$$R_i = \alpha_i + \beta_{i1}Y_1 + \dots + \beta_{ik}Y_k + \varepsilon_i, \quad i = 1, \dots, d \quad (7.10)$$

is more appropriate. Here we have k indices or 'risk factors' which are assumed to be uncorrelated, too. The first index can be interpreted as a market index whereas the other indices may represent segmental, sectoral, and regional risk factors. For the purpose of dimension reduction k shall not be too large. Note that the components of $Y = (Y_1, \dots, Y_k)$ can always be standardized because of the affine structure of Eq. 7.10. Additionally, if we replace α_i by r then (7.10) is known from Ross' (1976) arbitrage pricing theory.

Analogously to the simple linear regression model (7.9) we may find a reasonable model for R similar to (7.10) but even allowing for correlated idiosyncratic risks. Since $R \sim N_d(\mu, \Sigma)$ (approximately),

$$R \stackrel{d}{=} \mu + \mathcal{O}\sqrt{\mathcal{D}}Y, \quad (7.11)$$

where $Y \sim N_d(0, I_d)$ and $\Sigma = \mathcal{O}\mathcal{D}\mathcal{O}'$ is a spectral decomposition or 'diagonalization' of the covariance matrix Σ (cf. Section 3.4). Thus the main diagonal elements of \mathcal{D} correspond to the eigenvalues $\lambda_1, \dots, \lambda_d$ of Σ . We assume that the eigenvalues are given in a descending order and that the first k eigenvalues are large whereas the residual eigenvalues are small. Hence the first k elements of \mathcal{D} represent the variances of the driving risk factors contained in the 'first part' of Y , i.e. (Y_1, \dots, Y_k) . Because the $d - k$ residual risk factors contained in (Y_{k+1}, \dots, Y_d) are supposed to have small variances they can be interpreted as the components of the idiosyncratic risks of each asset, i.e.

$$\varepsilon_i := \sum_{j=k+1}^d \sqrt{\lambda_j} \mathcal{O}_{ij} Y_j, \quad i = 1, \dots, d.$$

The elements of Y are called 'principal components' of R (Chatfield and Collins, 2000, Chapter 4). The direction of each principal component is given by the corresponding column of \mathcal{O} , i.e. by the associated eigenvector. Since \mathcal{O} is orthonormal the distribution of the principal components remains up to a rotation in \mathbb{R}^d .

Now, regarding Eq. 7.10 the Betas are given by

$$\beta_{ij} := \sqrt{\lambda_j} \mathcal{O}_{ij},$$

for $i = 1, \dots, d$ and $j = 1, \dots, k$. Hence, we obtain the Beta estimates after the diagonalization

$$\hat{\Sigma} = \hat{\mathcal{O}}\hat{\mathcal{D}}\hat{\mathcal{O}}',$$

where $\hat{\Sigma}$ is an estimate of the covariance matrix of R . Now the Betas can be estimated by

$$\hat{\beta}_{\cdot j} = \sqrt{\hat{\lambda}_j} \hat{\mathcal{O}}_{\cdot j}, \quad j = 1, \dots, k.$$

Because $\text{tr}(\hat{\mathcal{D}}) = \text{tr}(\hat{\Sigma})$, the sum of the variances of Y_1, \dots, Y_d coincides with the sum of the variances of R_1, \dots, R_d . By dividing the sum of the first k eigenvalues of $\hat{\Sigma}$, i.e. the first k elements of $\hat{\mathcal{D}}$ by the sum of all eigenvalues of $\hat{\Sigma}$ or, equivalently, by its trace one may quantify the relative contribution of the main principal components to the overall dispersion. The considered quotient can be interpreted as the main principal components' 'proportion of the total variation' (Chatfield and Collins, 2000, p. 61). Further, each $\hat{\beta}_{\cdot j}$ can be interpreted as the direction of the j -th principal component. Note that the estimated Beta's are given only up to a scaling constant if Σ is estimated robustly. But that does not matter if one is only interested in analyzing both the relative contribution and the direction of each principal component.

A nice feature of principal component analysis (PCA) is that after estimating \mathcal{O} and \mathcal{D} one may extract the principal components easily by inverting Eq. 7.11 and substituting the parameters by their corresponding estimates. Hence the latent variables (Y_1, \dots, Y_k) can be represented by a weighted average of the asset returns, i.e. as stock indices. Indeed, the larger the weight of a particular asset the more the considered risk factor is characterized by the asset's line of business. For instance, if the risk factor Y_j is characterized mainly by computer firms then it can be interpreted as the relative change of semiconductor prices.

The purpose of PCA is to reduce the complexity caused by the number of dimensions. This can be done successfully only if there is indeed a number of principal components accountable for the most part of the distribution. Additionally, the covariance matrix estimator which is used for extracting the principal components should be robust against outliers. For example, let the daily data be multivariate t -distributed with ν degrees of freedom and suppose that $d = 500$ and $n = 1000$. The black lines in the following figure show the true proportion of the total variation for a set of 500 eigenvalues. We see that the largest 20% of the eigenvalues accounts for 80% of the overall variance. This constellation is known as '80/20 rule' or 'Pareto's principle'. The estimated eigenvalue proportions obtained by the sample covariance matrix are represented by the red lines whereas the corresponding estimates based on the spectral estimator are given by the green lines. Each line is an average over 100 concentration curves drawn from samples of the corresponding multivariate t -distribution.

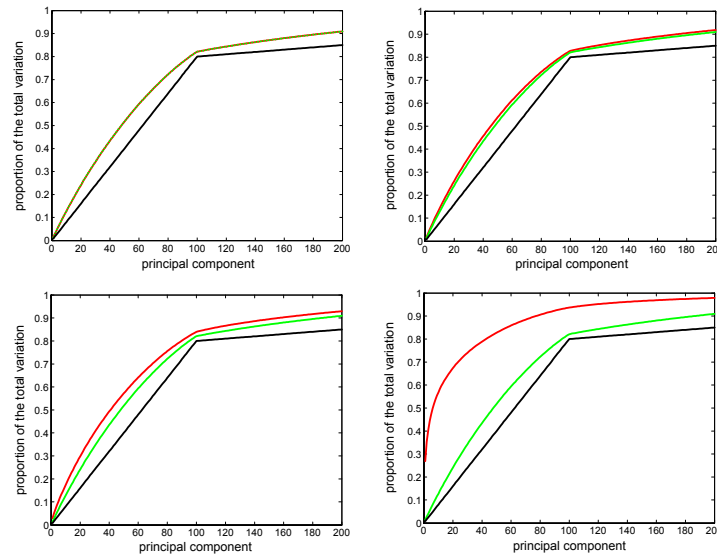


Figure 7.1 True proportion of the total variation (black line) and proportions obtained by the sample covariance matrix (red lines) and by the spectral estimator (green lines). The samples are drawn from a multivariate t -distribution with $\nu = \infty$ (upper left), $\nu = 10$ (upper right), $\nu = 5$ (lower left), and $\nu = 2$ (lower right) degrees of freedom.

The goal of PCA is to find the true number of driving factors. If the data is normally distributed (i.e. $\nu = \infty$) then both the spectral estimator and the sample covariance matrix come to similar results concerning the number of main principal components. But the smaller the tail index the more differences between the covariance matrix estimators occur. An interesting fact is that the sample covariance matrix tends to underestimate the number of driving factors if the data is regularly varying (i.e. $\nu < \infty$). This is due to the fact that an extreme value is confused with a strong realization of a driving factor.

In the following figure we see the concentration curves obtained by the sample covariance matrix (red line) and by the spectral estimator (green line) of the current 285 stocks of the S&P 500 whose IPO date is before 1980-01-02.

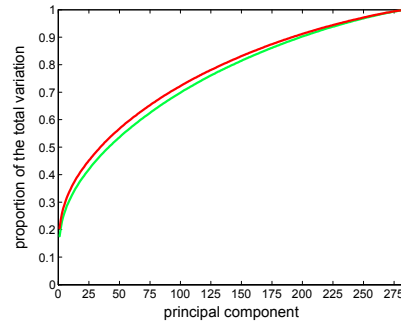


Figure 7.2 Eigenvalue proportions for 285 stocks of the S&P 500 for the sample period 1980-01-02 to 2003-11-26. Concentration curves obtained by the sample covariance matrix (red line) and by the spectral estimator (green line).

In the simulated example of Figure 7.1 it is assumed that the small eigenvalues are equal. This is equivalent to the assumption that the ‘small’ principal components are spherically distributed, i.e. that they contain no more information about the linear dependence structure of R . This is similar (but not equivalent) to the assumption that the noise variables $\varepsilon_1, \dots, \varepsilon_d$ are uncorrelated. Of course, even if the true eigenvalues are equal the corresponding estimates will not share this property. But it is important to know whether the residual principal components have structural information or the differences between the estimates are only caused by random noise. This is not an easy task (Chatfield and Collins, 2000, p. 65), especially if the data is not normally distributed and the number of dimensions is large which is the issue of the next chapter.

Chapter 8

Random Matrix Theory

Not only in finance but also in physics one is sometimes concerned with data which is not Gaussian. But typically the Gaussian distribution hypothesis is favored not only for the sake of simplicity but particularly because this is often a reasonable assumption for natural phenomena. But here one should act with caution. Especially when working with high-dimensional data there may be some areas of application like, e.g., noise reduction, filtering and image restoration in image processing, image object and pattern recognition, analysis of turbulent flows in (magneto-)hydrodynamics, where the Gaussian distribution hypothesis may lead to wrong conclusions. More than ever if methods of physics are applied to economics then the ‘normality assumption’ generally cannot be justified if one is interested in short-term phenomena. This is due to the fact that socio-economic systems are highly characterized by interacting individuals, information asymmetry, and overshooting.

More than 50 years ago the analysis of energy levels of complex nuclei led to a branch of statistical physics called ‘random matrix theory’ (RMT). In the meantime RMT emerged rapidly and is used in different areas like, e.g., chaotic systems, elastodynamics of structural materials, conductivity in disordered metals, and even for analyzing the distribution of the zeros of Riemann’s zeta function (Mehta, 1991). For a historical review of RMT and its applications see Forrester, Snaith, and Verbaarschot (2003). The standard reference to RMT is Mehta (1991). A relatively new treatment of this topic is given by Hiai and Petz (2000).

8.1 Limiting Distributions of Eigenvalues

8.1.1 Wigner’s Semi-circle Law

RMT is concerned with the distribution of eigenvalues and eigenvectors of randomly generated matrices. A random matrix is simply a matrix of random variables. We will consider only symmetric random matrices. Thus the corresponding eigenvalues are always real. The empirical distribution function of eigenvalues is defined as follows.

Definition 20 (Empirical distribution function of eigenvalues) *Let $S_d \in \mathbb{R}^{d \times d}$ be a symmetric random matrix with realized eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_d \in \mathbb{R}$. Then the function*

$$\lambda \mapsto \widehat{W}_d(\lambda) := \frac{1}{d} \cdot \sum_{i=1}^d \mathbb{1}_{\lambda_i \leq \lambda}$$

is called the ‘empirical distribution function of the eigenvalues’ of S_d .

Note that an eigenvalue of a random matrix in fact is random but per se not a random variable since there is no single-valued mapping $S_d \mapsto \lambda_i$ ($i \in \{1, \dots, d\}$) but rather $S_d \mapsto \lambda(S_d)$ where $\lambda(S_d)$ denotes the set of all eigenvalues of S_d . This can be simply fixed by assuming that the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_d$ are sorted either in an increasing or decreasing order.

Proposition 49 (Semi-circle law) *Let $S_d \equiv [S_{ij,d}] \in \mathbb{R}^{d \times d}$ ($d = 1, 2, \dots$) be sequences of symmetric random matrices with i.i.d. elements $S_{ij,d}$. Further, let $S_{ii,d} \sim G$ ($i = 1, \dots, d$) and $S_{ij,d} \sim F$ ($i, j = 1, \dots, d, i \neq j$) with $E(S_{ij,d}) = 0$, $i, j = 1, \dots, d$. It is assumed that*

$$\begin{aligned} E(S_{ii,d}^2) &< \infty, & i = 1, \dots, d, \\ E(S_{ij,d}^4) &< \infty, & i, j = 1, \dots, d, i \neq j, \end{aligned}$$

and $\sigma^2 := E(S_{ij,d}^2)$ ($i \neq j$). Further, let \widehat{W}_d be the empirical distribution function of the eigenvalues of $S_d/(2\sqrt{\sigma^2 d})$. Then

$$\widehat{W}_d(\lambda) \xrightarrow{p} W(\lambda), \quad d \longrightarrow \infty,$$

for every $\lambda \in \mathbb{R}$ where W is an absolutely continuous c.d.f. with density function

$$\lambda \mapsto w(\lambda) = \begin{cases} \frac{2}{\pi} \cdot \sqrt{1 - \lambda^2}, & |\lambda| \leq 1, \\ 0, & |\lambda| > 1. \end{cases}$$

Proof. Arnold (1967). ■

This theorem goes back to Wigner (1958) who proved $E(\widehat{W}_d) \rightarrow W$ ($d \rightarrow \infty$) rather than the convergence in probability. Wigner required the existence and finiteness of all moments and even the symmetry of the random components. But Arnold (1967) showed that under the weaker conditions above even the convergence in probability holds. Moreover, Arnold (1967) states that actually $\widehat{W}_d \xrightarrow{\text{a.s.}} W$ as $d \rightarrow \infty$ if the fourth moment of G and the sixth moment of F are finite.

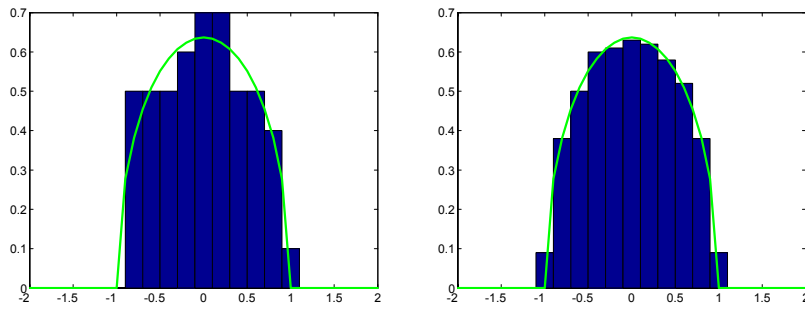


Figure 8.1 Histogram of the eigenvalues of a symmetric matrix containing i.i.d. variables which are uniformly distributed on $[-0.5, 0.5]$ for $d = 50$ (left hand) and $d = 500$ (right hand). Wigner's semi-circle law is represented by the green line.

The semi-circle law is often represented (see, e.g., Bouchaud and Potters, 2000, p. 41) by

$$\lambda \mapsto w(\lambda) = \frac{1}{2\pi\sigma^2} \cdot \sqrt{4\sigma^2 - \lambda^2}, \quad |\lambda| \leq 2\sigma.$$

This is the limit law for the empirical distribution function of the eigenvalues of S_d/\sqrt{d} , i.e. if the random matrix S_d is not normalized before.

8.1.2 The Marčenko-Pastur Law

Now, consider the random matrix $S_d S_d' / d$. The eigenvalues of $S_d S_d'$ are given by the squared eigenvalues of S_d . Consequently, the limit law for the eigenvalues of $S_d S_d' / d$ corresponds to (cf. Bouchaud and Potters, 2000, p. 41)

$$\lambda \mapsto \frac{w(\sqrt{\lambda})}{\sqrt{\lambda}} = \frac{1}{2\pi\sigma^2} \cdot \sqrt{\frac{4\sigma^2 - \lambda}{\lambda}}, \quad 0 < \lambda \leq 4\sigma^2.$$

Note that $S_d S_d' / d$ can be interpreted as a sample covariance matrix where S_d is a sample of d centered random vectors. Of course, the observations contained in S_d have a strange dependence structure because S_d is supposed to be symmetric. But actually, one may show that the same formula holds for the eigenvalues of sample covariance matrices of non-symmetric but only square sample matrices S_d . More generally, if the sample size n exceeds the number of dimensions d then

$$\lambda \mapsto \frac{q}{2\pi\sigma^2} \cdot \frac{\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})}}{\lambda}, \quad \lambda_{\max, \min} = \sigma^2 \left(1 \pm \frac{1}{\sqrt{q}}\right)^2 \quad (8.1)$$

is the limit law for the empirical distribution function of eigenvalues of the sample covariance matrix $\hat{\Sigma} = S_n S_n' / n$ as $n \rightarrow \infty$, $d \rightarrow \infty$ and $n/d \rightarrow q \geq 1$ where $S_n \in \mathbb{R}^{d \times n}$ represents a matrix of i.i.d. centered elements with finite variance σ^2 (Yin, 1986). Now, the effective sample size q comes into play. A nice feature is that the limit law holds also for $0 < q < 1$ but with an additional atom at 0 with probability $1 - q$ (Bouchaud and Potters, 2000, p. 42, Marčenko and Pastur, 1967). Formula 8.1 is called the ‘Marčenko-Pastur law’.

The Marčenko-Pastur law can be used exploratively to analyze the eigenspectrum of high-dimensional distributions. The key idea is that this can be done even if the sample size is small compared to the number of dimensions whereas classical statistical analysis fails in that case. Moreover, the eigenspectrum can be analyzed even if the number of data points falls below the number of dimensions.

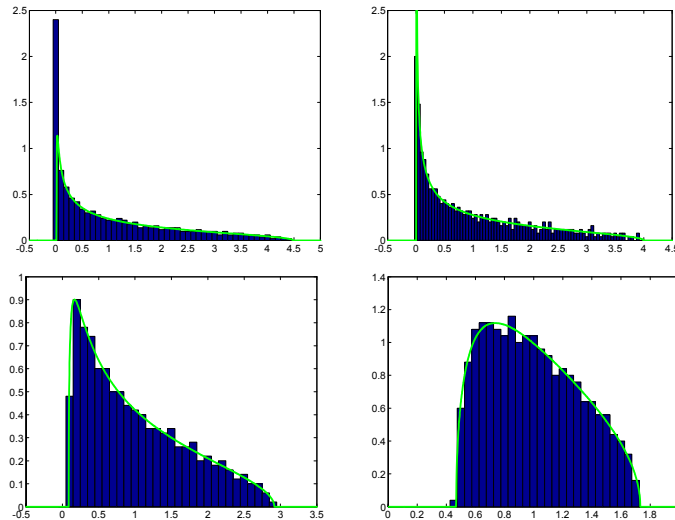


Figure 8.2 Histograms of the eigenvalues of sample covariance matrices from samples with independent and standardized uniformly distributed random elements where $d = 500$ and $n = 400$ (i.e. $q = 0.8$, upper left), $n = 500$ (i.e. $q = 1$, upper right), $n = 1000$ (i.e. $q = 2$, lower left), and $n = 5000$ (i.e. $q = 10$, lower right). The corresponding Marčenko-Pastur laws are represented by the green lines.

Up to now we have presumed that the random entries in S_n are centered and i.i.d. (see Figure 8.2). But in the context of elliptically symmetric distributions this is given only for the normal distribution. More precisely, if the random entries in S_n are i.i.d. standardized Gaussian random variables then the random matrix $S_n S_n' \in \mathbb{R}^{d \times d}$ corresponds to a d -dimensional standard ‘white’ Wishart matrix (Johnstone, 2001) with n degrees of freedom. The term ‘white’ refers to the fact that the components of each column of S_n are supposed to be uncorrelated. If the random vectors contained in S_n are not Gaussian but only spherical then the Marčenko-Pastur law is no longer valid for the eigenvalues of the sample covariance matrix. This is illustrated by the figure below.

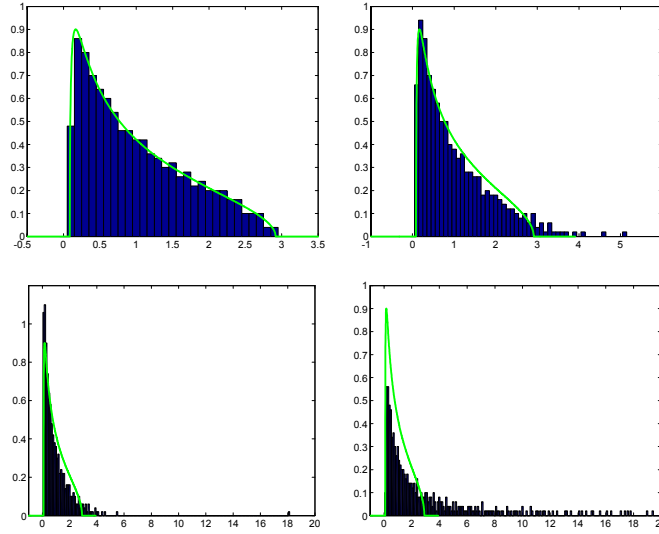


Figure 8.3 Histograms of the eigenvalues of sample covariance matrices from samples of standardized t_{500} -distributed random vectors ($n = 1000$, i.e. $q = 2$) with uncorrelated components and $\nu = \infty$ (i.e. Gaussian random vectors, upper left), $\nu = 10$ (upper right), $\nu = 5$ (lower left), and $\nu = 2$ (lower right). In the latter case the largest eigenvalues even go far beyond 20. This is suppressed for the sake of transparency. The Marčenko-Pastur law is represented by the green lines.

Hence it is obvious that the eigenspectrum generated by the sample covariance matrix can lead to misinterpretations concerning the true dispersion matrix if the generating variate is regularly varying and the tail index is small. This is due to the fact that the components of spherical random vectors are uncorrelated but generally not independent (cf. Section 2.2). It must be pointed out that the sample covariance matrix is not an appropriate tool not only if the data is regularly varying but also if it is strongly asymmetric (cf. Section 3.2).

Instead of the sample covariance matrix alternatively the sample correlation matrix may be considered for extracting the eigenspectrum. This has mainly two effects:

1. Under the elliptical distribution hypothesis every d -dimensional random vector can be represented by (cf. Theorem 2)

$$X \stackrel{\text{d}}{=} \mu + \sigma \sqrt{\rho} \mathcal{R} U^{(k)},$$

where σ and $\sqrt{\rho}$ are defined as described in Section 2.3.1. Suppose w.l.o.g. (cf. Section 1.2.3) that $E(\mathcal{R}^2) = k$ such that the main diagonal elements of σ represent the standard deviations of X . Now consider the standardized random vector

$$Y := \sigma^{-1} (X - \mu) \stackrel{\text{d}}{=} \sqrt{\rho} \mathcal{R} U^{(k)}.$$

Pearson's correlation coefficient is an estimator for ρ , i.e. the dispersion matrix of Y . By studying its eigenvalues one concentrates on the linear dependence structure of X without considering location and scales. This is similar to the trade off between joint distribution functions and copulas discussed in Section 2.2.

2. Provided μ and σ are known the components of Y have unit variance. Due to the law of large numbers the main diagonal elements of the sample covariance matrix (of a sample drawn from Y) converge strongly to 1. That is to say the sample covariance matrix behaves like the corresponding sample correlation matrix, asymptotically. Thus we can substitute the sample covariance matrix by the sample correlation matrix. Since the main diagonal elements of the sample correlation matrix always correspond to 1 there is no need to standardize the data and σ^2 can immediately be set to 1 in the Marčenko-Pastur law.

Let the dispersion matrix Σ be defined as

$$\Sigma := \mathcal{O}\mathcal{D}\mathcal{O}', \quad (8.2)$$

where \mathcal{O} is an orthonormal random matrix and \mathcal{D} is a diagonal matrix with entries $1, 2, \dots, d$. Obviously, Σ has a linear eigenspectrum given by the main diagonal elements of \mathcal{D} but randomly generated eigenvectors. The expected eigenspectrum of the corresponding pseudo-correlation matrix is also linear. In the following figure we see the histogram of the eigenvalues of the sample correlation matrix from a sample ($n = 1000$) of standardized t_{500} -distributed random vectors with uncorrelated components and $\nu = 5$ degrees of freedom. For comparison we see also a histogram of the eigenvalues of the sample correlation matrix from a sample of random vectors $X \sim N_d(0, \Sigma)$ where Σ is obtained by the diagonalization considered above. Note that the true eigenspectrum of the former sample is constant whereas the true eigenspectrum of the latter sample is linear.

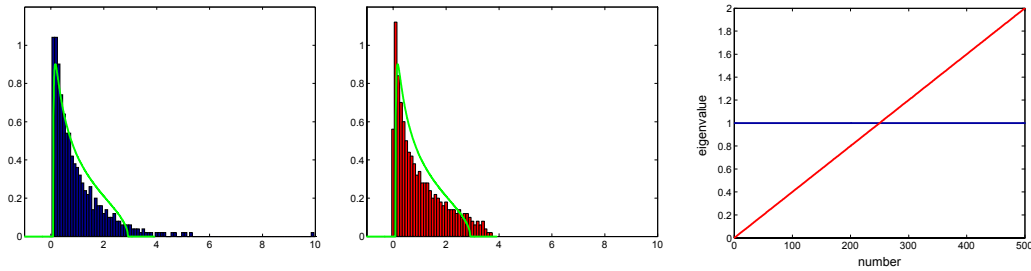


Figure 8.4 Histogram of the eigenvalues of the sample correlation matrix of standardized t_{500} -distributed random vectors with $\nu = 5$ degrees of freedom (left hand). Histogram of the eigenvalues of the sample correlation matrix of normally distributed random vectors (middle). The true eigenspectrum of the former sample (right hand, blue line) and the true eigenspectrum of the latter sample (right hand, red line).

It is difficult to distinguish between the eigenspectra of both samples. Moreover, in practice one would actually suggest that the true eigenspectrum of the first sample is wider than the true eigenspectrum of the second sample. But in fact the latter eigenspectrum is uniformly distributed on $[0, 2]$ whereas the former one has Dirac mass at 1. We conclude that the sample covariance matrix or, alternatively, the sample correlation matrix is not an appropriate tool for applying the Marčenko-Pastur law provided the data are elliptically distributed or even generalized elliptically distributed.

It is worth to stress that if the considered data is assumed to be power tailed in the context of elliptical distributions, then the data necessarily depend on each other. Conversely, assuming the data to be power tailed but independent contradicts the elliptical distribution hypothesis.

The Marčenko-Pastur law as given originally in the paper of 1967 is much weaker than usually stated. The relationship between the Marčenko-Pastur law and spherical distributions should be clarified in the following.

Theorem 50 (Marčenko and Pastur, 1967) *Let $U^{(d)}, U_1^{(d)}, U_2^{(d)}, \dots, U_n^{(d)}$ ($n = 1, 2, \dots$) be sequences of independent random vectors uniformly distributed on the unit hypersphere \mathcal{S}^{d-1} . Further, let $\mathcal{T}, \mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_n$ ($n = 1, 2, \dots$) be sequences of i.i.d. random variables where each variable is independent of the corresponding unit random vector. Consider the random matrix*

$$\sum_{i=1}^n \mathcal{T}_i U_i^{(d)} U_i^{(d)'},$$

where its empirical distribution function of the eigenvalues is denoted by \widehat{W}_d . Then, as $n \rightarrow \infty$, $d \rightarrow \infty$, $n/d \rightarrow q < \infty$,

$$\widehat{W}_d(\lambda) \xrightarrow{p} F_{MP}(\lambda; q),$$

at all points where F_{MP} is continuous. F_{MP} is a c.d.f. given by

$$\lambda \mapsto F_{MP}(\lambda; q) = \lim_{x \rightarrow -\infty} \lim_{y \searrow 0} \frac{1}{\pi} \cdot \int_x^\lambda \operatorname{Im}(T(t + iy; q)) dt. \quad (8.3)$$

Here $T(\cdot; q)$ is the Stieltjes transform of $F_{MP}(\cdot; q)$ which is given by the solution of the equation

$$T(x; q) = - \left(x - q \cdot E \left(\frac{\mathcal{T}}{1 + T(x; q) \cdot \mathcal{T}} \right) \right)^{-1}$$

on the region where $\operatorname{Im}(x) > 0$. Further, the Stieltjes transform exists and is unique.

Proof. Marčenko and Pastur, 1967. ■

Hence the Marčenko-Pastur law allows for negative \mathcal{T} , i.e. complex valued generating variates which is not covered by the traditional theory of elliptical distributions. If $\mathcal{T} \geq_{\text{a.s.}} 0$ then $\sqrt{\mathcal{T}} U^{(d)}$ corresponds to a spherically distributed random vector. Of course, if the generating variate $\sqrt{\mathcal{T}}$ is regularly varying its tail index makes an impact on the Stieltjes transform. But Marčenko and Pastur (1967) state that the Stieltjes transform generally cannot be given in a closed form. A loophole which turns out to be very useful for analyzing the spectral estimator is given by the following corollary.

Corollary 51 *Let the conditions of Theorem 50 be fulfilled. Additionally, let \mathcal{T} be a degenerated random variable corresponding to $\sigma^2 > 0$. Then $\lambda \mapsto F_{MP}(\lambda; q) = F_{MP}^{Dir}(\lambda; q) + F_{MP}^{Leb}(\lambda; q)$ where the Dirac part is given by*

$$\lambda \mapsto F_{MP}^{Dir}(\lambda; q) = \begin{cases} 1 - q, & \lambda \geq 0, 0 \leq q < 1, \\ 0, & \text{else,} \end{cases}$$

and the Lebesgue part $\lambda \mapsto F_{MP}^{Leb}(\lambda; q) = \int_{-\infty}^\lambda f_{MP}^{Leb}(x; q) dx$ is determined by the density function

$$\lambda \mapsto f_{MP}^{Leb}(\lambda; q) = \begin{cases} \frac{1}{2\pi\sigma^2} \cdot \frac{\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})}}{\lambda}, & \lambda_{\min} \leq \lambda \leq \lambda_{\max}, \\ 0, & \text{else,} \end{cases}$$

where

$$\lambda_{\min, \max} := \sigma^2 (1 \pm \sqrt{q})^2.$$

Proof. The Stieltjes transform is given by the solution of

$$T(x; q) = - \left(x - q \cdot \frac{\sigma^2}{1 + T(x; q) \cdot \sigma^2} \right)^{-1},$$

i.e. (cf. Marčenko and Pastur, 1967)

$$T(x; q) = - \frac{(1-q) + |1-q|}{2x} + \frac{-x + |1-q| \cdot \sigma^2 + \sqrt{(x - q\sigma^2 + \sigma^2)^2 - 4x\sigma^2}}{2x\sigma^2}. \quad (8.4)$$

Now, the limit law F_{MP} can be obtained by taking the derivative of (8.3) with respect to λ , i.e.

$$\lim_{y \searrow 0} \frac{1}{\pi} \cdot \text{Im} (T(\lambda + iy; q)).$$

Note that the first term of (8.4) vanishes for $q \geq 1$. But if $0 \leq q < 1$ it becomes

$$-\frac{1-q}{x} = -\frac{1-q}{\lambda + iy} = -(1-q) \cdot \frac{\lambda - iy}{\lambda^2 + y^2},$$

and its imaginary part corresponds to $(1-q) \cdot y / (\lambda^2 + y^2)$. Further,

$$\lim_{y \searrow 0} \frac{1-q}{\pi} \cdot \frac{y}{\lambda^2 + y^2} = (1-q) \cdot \delta(\lambda),$$

where δ denotes the delta function. The second term of (8.4) leads to the Lebesgue density function

$$\lambda \mapsto f_{\text{MP}}^{\text{Leb}}(\lambda; q) = \frac{1}{2\pi\sigma^2} \cdot \frac{\sqrt{4q\sigma^4 - (\lambda - q\sigma^2 - \sigma^2)^2}}{\lambda}, \quad (\lambda - q\sigma^2 - \sigma^2)^2 \leq 4q\sigma^4.$$

Note that

$$\begin{aligned} 4q\sigma^4 - (\lambda - q\sigma^2 - \sigma^2)^2 &= (2\sqrt{q}\sigma^2 + \lambda - q\sigma^2 - \sigma^2)(2\sqrt{q}\sigma^2 - \lambda + q\sigma^2 + \sigma^2) \\ &= (\lambda - \sigma^2(1 - \sqrt{q})^2)(\sigma^2(1 + \sqrt{q})^2 - \lambda). \end{aligned}$$

■

Now, suppose that $\mathcal{T} \stackrel{\text{a.s.}}{=} 1$ and consider the random matrix

$$\widehat{\Sigma}_{\text{MP}} := \frac{d}{n} \cdot \sum_{i=1}^n U_i^{(d)} U_i^{(d)'} ,$$

which will be called in the following ‘Marčenko-Pastur operator’. It is clear that $\widehat{\Sigma}_{\text{MP}}$ and

$$\frac{1}{q} \cdot \sum_{i=1}^n U_i^{(d)} U_i^{(d)'}$$

are asymptotically equivalent for P -almost all realizations since $n/d \rightarrow q$.

The Marčenko-Pastur law already given by (8.1) is simply obtained by setting $\mathcal{T} \stackrel{\text{a.s.}}{=} \sigma^2/q$ in Corollary 51. Surprisingly, formula (8.1) was given for the eigenvalues of sample covariance matrices of i.i.d. centered elements like, e.g., for normally distributed random variables rather than for the Marčenko-Pastur operator. But due to the strong law of large numbers $\chi_d^2/d \stackrel{\text{a.s.}}{\rightarrow} 1$ holds and we obtain the asymptotic equivalence of the Marčenko-Pastur operator and the random matrix

$$\frac{d}{n} \cdot \sum_{i=1}^n \frac{\chi_{d,i}^2}{d} \cdot U_i^{(d)} U_i^{(d)'} = \frac{1}{n} \cdot \sum_{i=1}^n \left(\sqrt{\chi_{d,i}^2} U_i^{(d)} \right) \left(\sqrt{\chi_{d,i}^2} U_i^{(d)} \right)',$$

which is just the sample covariance matrix of multivariate standard normally distributed random vectors with uncorrelated components.

Note that the Marčenko-Pastur law can always be used in the simple form with $\sigma^2 = 1$ if the trace of the covariance matrix estimator which is used for extracting the eigenvalues corresponds (asymptotically) to the dimension of the data. This is a priori given for the sample correlation matrix. Further, every other covariance matrix estimator can be simply normalized such that the trace corresponds to its dimension (cf. Section 2.4). Another way to obtain the same result is given by normalizing the estimated eigenvalues such that their sum corresponds to their quantity.

Now, by virtue of the argument given in the first part of this thesis one may expect that the Marčenko-Pastur operator is a better choice than the sample covariance matrix for analyzing the eigenspectra of generalized elliptically distributed data where $U^{(d)}$ is simply taken from the projections to the unit hypersphere. Indeed, if the data are spherically distributed or only generalized elliptically distributed with dispersion matrix $\sigma^2 I_d$ and positive generating variate then applying the Marčenko-Pastur operator leads to the desired result. This holds independent of the generating variate. But if the data has a linear dependence structure then the Marčenko-Pastur operator is a biased estimator. This is illustrated in the following figure.

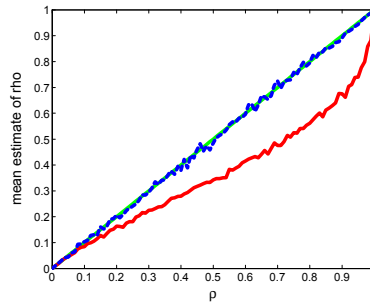


Figure 8.5 Mean off diagonal elements of the Marčenko-Pastur operator (red line) of 1000 independent, standardized, and equicorrelated t_{500} -distributed random vectors with $\nu = 5$. The true correlation is given by the green line. For comparison the mean off diagonal elements of Pearson's correlation coefficient is represented by the blue dotted line.

This can be explained as follows. Consider that the sample covariance matrix corresponds to the ML-estimator for normally distributed data. Applying the sample covariance matrix or the sample correlation matrix, alternatively, means trying to fit the Gaussian density to the realized data. But if the data is not normally distributed this approach may lead to wrong conclusions. Now, the Marčenko-Pastur operator is nothing else but the sample covariance matrix (up to an additional multiplication with d) applied to the data projected to the unit hypersphere. But this data a priori suggest that there are spherical rather than elliptical density contours and thus a bias towards the identity matrix occurs. Thus we conclude that the Marčenko-Pastur operator is not appropriate for estimating the linear dependence structure of generalized elliptically distributed data.

In Part I we investigated the statistical properties of the spectral estimator. It was shown that the spectral estimator is a robust alternative to the sample covariance matrix. Recall (cf. Section 4.2.2) that the spectral estimator corresponds to the solution of the fixed-point equation

$$\widehat{\Sigma} = \frac{d}{n} \cdot \sum_{j=1}^n \frac{s_{\cdot j} s_{\cdot j}'}{s_{\cdot j}' \widehat{\Sigma}^{-1} s_{\cdot j}},$$

where $s_{\cdot j}$ ($j = 1, \dots, n$) are the data points projected to the unit hypersphere. Due to Theorem 46 the spectral estimator converges strongly to the true dispersion matrix Σ . That

means (for P -almost all realizations)

$$\sum_{j=1}^n \frac{s_{\cdot j} s'_{\cdot j}}{s'_{\cdot j} \widehat{\Sigma}^{-1} s_{\cdot j}} \longrightarrow \sum_{j=1}^n \frac{s_{\cdot j} s'_{\cdot j}}{s'_{\cdot j} \Sigma^{-1} s_{\cdot j}}, \quad n \longrightarrow \infty, \quad d \text{ const.}$$

Consequently, if $\Sigma = I_d$ (up to a scaling constant) then

$$\sum_{j=1}^n \frac{s_{\cdot j} s'_{\cdot j}}{s'_{\cdot j} \widehat{\Sigma}^{-1} s_{\cdot j}} \longrightarrow \sum_{j=1}^n s_{\cdot j} s'_{\cdot j}, \quad n \longrightarrow \infty, \quad d \text{ const.}$$

Hence the spectral estimator and the Marčenko-Pastur operator are asymptotically equivalent. If the strong convergence holds anymore for $n \rightarrow \infty$, $d \rightarrow \infty$, $n/d \rightarrow q > 1$ then

$$\widehat{\Sigma} \longrightarrow \frac{1}{q} \cdot \sum_{j=1}^n s_{\cdot j} s'_{\cdot j}, \quad n \rightarrow \infty, \quad d \rightarrow \infty, \quad n/d \rightarrow q > 1,$$

for P -almost all realizations where $\widehat{\Sigma}$ exists. Recall that $n > d$ is a necessary condition whereas $n > d(d-1)$ is a sufficient condition for the existence of the spectral estimator (cf. Section 4.2.3). Several numerical experiments indicate that indeed the latter convergence holds. Hence the spectral estimator seems to be a robust alternative to the sample covariance matrix not only in the case of classical asymptotics but also in the context of random matrix theory.

Figure 8.6 can be compared directly with Figure 8.4. We see the histogram of the eigenvalues of the spectral estimator for the pseudo-correlation matrix of standardized t_{500} -distributed random vectors with $\nu = 5$ degrees of freedom on the upper left. Analogously, the histogram of its eigenvalues for the normally distributed random vectors used in Figure 8.4 is plotted on the upper right. Note that the true eigenspectrum of the former sample is given by the blue line of the right hand side of Figure 8.4 whereas the true eigenspectrum of the latter sample is given by the red line, respectively. On the lower left and the lower right of Figure 8.6 are the corresponding empirical eigenvalue distributions obtained by the Marčenko-Pastur operator.

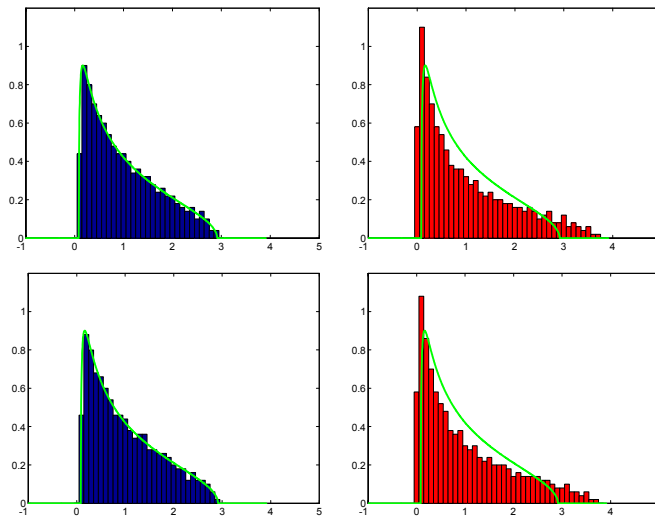


Figure 8.6 Histogram of the eigenvalues obtained by the spectral estimator (upper part) and by the Marčenko-Pastur operator (lower part). The data is the same as in Figure 8.4.

Hence the spectral estimator is proposed for a robust analysis of the eigenspectrum of high-dimensional generalized elliptically distributed data.

8.2 Separation of Signal and Noise

Let $\Sigma = \mathcal{O}\mathcal{D}\mathcal{O}' \in \mathbb{R}^{d \times d}$ be the dispersion matrix of an elliptical random vector. Here \mathcal{O} and \mathcal{D} are defined as in the previous section (Eq. 8.2). But now \mathcal{D} shall be a diagonal matrix containing a ‘bulk’ of equally small eigenvalues and a few large (but not necessarily equal) eigenvalues. For the sake of simplicity suppose

$$\mathcal{D} = \begin{bmatrix} cI_k & 0 \\ 0 & bI_{d-k} \end{bmatrix},$$

where $b, c > 0$, $c > b$, and k/d small. The k large eigenvalues can be interpreted as variances of the driving risk factors (cf. Section 7.2) of

$$X \stackrel{d}{=} \mu + \mathcal{O}\sqrt{\mathcal{D}}\mathcal{R}U^{(d)},$$

whereas the $d - k$ small eigenvalues are the variances of the residual risk factors. Suppose that one is interested in estimating the ‘signal’, i.e. the number and the amount of the large eigenvalues. This is demonstrated in the following.

Assume that $n = 1000$, $d = 500$ (i.e. $q = 2$) and a sample consists of normally distributed random vectors with dispersion matrix $\Sigma = \mathcal{O}\mathcal{D}\mathcal{O}'$, where $b = 1$, $c = 20$, and $k = 25$. By using the sample covariance matrix and normalizing the eigenvalues one obtains exemplarily the eigenspectrum and histogram of eigenvalues given in Figure 8.7. Clearly, one may separate the 25 largest eigenvalues from the bulk of small ones. The bulk is characterized by the property that it falls below the Marčenko-Pastur upper bound. But this is not sufficient for assuming that the small 475 eigenvalues are equal, i.e. that the bulk represents ‘noise’. In almost the same manner the residual eigenvalues could be linearly increasing, for instance. Testing the bulk for noise is simply done by fitting the Marčenko-Pastur law to the residual (but re-normalized) eigenvalues (Figure 8.8).

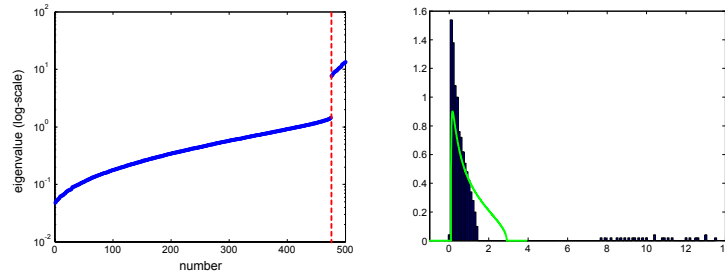


Figure 8.7 Estimated eigenvalues on a log-scale (left-hand) and histogram with corresponding Marčenko-Pastur law (right hand) for $c = 20$.

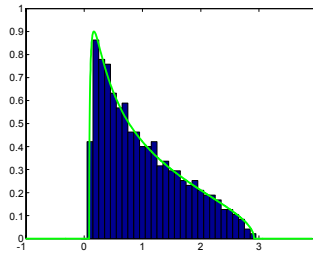


Figure 8.8 Histogram of the bulk eigenvalues with the corresponding Marčenko-Pastur law for $c = 20$.

We see that the bulk of eigenvalues indeed are due to random noise and conclude that the signal consists of 25 eigenvalues.

Now, consider $c = 2$, i.e. the signal is close to noise. In the next figure we see that now it is impossible to distinguish between small and large eigenvalues only by the eigenspectrum.

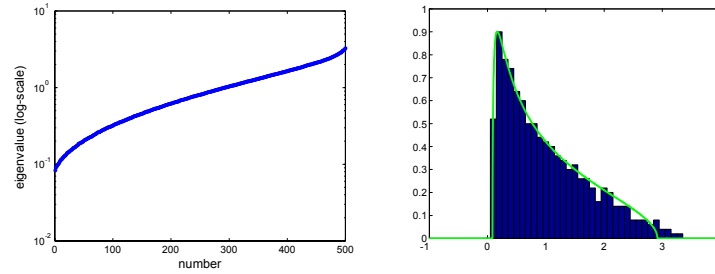


Figure 8.9 Estimated eigenvalues on a log-scale (left-hand) and histogram with corresponding Marčenko-Pastur law (right hand) for $c = 2$.

But one may separate the eigenvalues which exceed the Marčenko-Pastur upper bound. More precisely, one pitches on the largest eigenvalues, iteratively, until there are no more which exceed the upper bound. But note that the residual eigenvalues must be permanently re-normalized and that the upper bound depends in each iteration on the number of residual eigenvalues. It is not sufficient to take a look only on the original plot, especially if there are very large eigenvalues relative to the spectrum. At the end the residuum should be compared with the Marčenko-Pastur law. This can be done by the following figure.

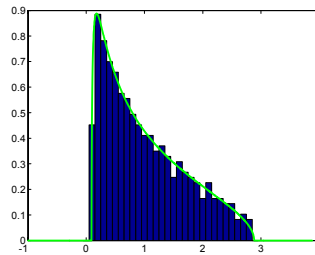


Figure 8.10 Histogram of the bulk eigenvalues with the corresponding Marčenko-Pastur law for $c = 2$.

The number of separated eigenvalues corresponds only to 14 (instead of 25). This is due to the fact that signal ($c = 2$) and noise ($b = 1$) are close to each other.

Now consider the same experiment but with multivariate t -distributed data with $\nu = 5$ degrees of freedom. Applying the iterative method again leads to 122 driving risk factors and 378 residuals. Thus the signal is overestimated, tremendously, what is also indicated by the relatively bad fit of the Marčenko-Pastur law (right hand of Figure 8.11). This is due to the effect of regular variation of the multivariate t -distribution. In contrast, applying the spectral estimator for the purpose of signal-noise separation leads only to 15 driving risk factors vs. 485 residuals (see Figure 8.12).

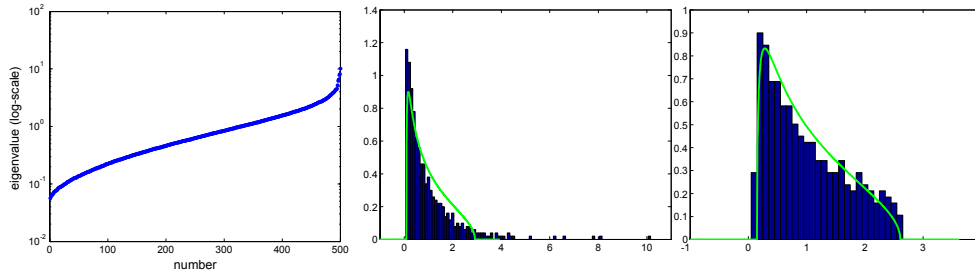


Figure 8.11 Estimated eigenvalues on a log-scale (left-hand) and histogram with corresponding Marčenko-Pastur law (middle) by using the sample covariance matrix for $c = 2$ and t -distributed data. Histogram of the bulk eigenvalues after separation of signal and noise (right hand).

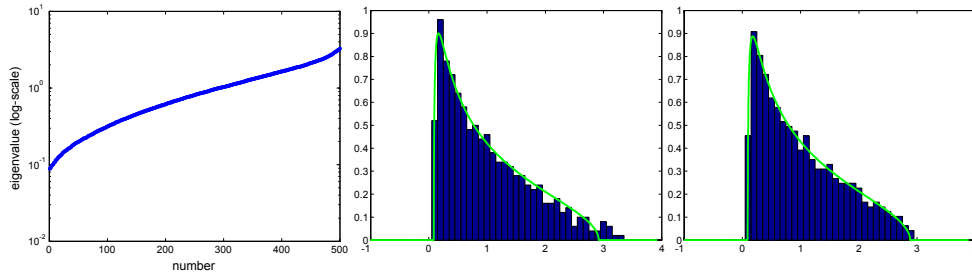


Figure 8.12 Estimated eigenvalues on a log-scale (left-hand) and histogram with corresponding Marčenko-Pastur law (middle) by using the spectral estimator for $c = 2$ and t -distributed data. Histogram of the bulk eigenvalues after separation of signal and noise (right hand).

We conclude that the spectral estimator is a robust alternative to the sample covariance matrix also for signal-noise separation leading to a better understanding of high-dimensional linear dependence structures if the data are elliptically distributed and regularly varying.

8.3 Application to Econophysics

For applying MPT (cf. Section 7.1) or PCA (cf. Section 7.2) on high-dimensional financial data it is suggested to consider the eigenspectrum of the corresponding covariance matrix estimate. This was done recently by many authors from physics (see, e.g., Amaral et al., 2002, Bouchaud et al., 2000, Bouchaud and Potters, 2000, Section 2.7, Gebbie and Wilcox, 2004, Kondor, Pafka, and Potters, 2004, Malevergne and Sornette, 2002, and Utsugi, Ino, and Oshikawa, 2003). In most cases the authors take the sample correlation matrix for extracting the eigenspectra. In the previous section it was shown that this may lead to misinterpretations provided the data is regularly varying and the tail index is small. This is usually the case for financial data as it was shown in Section 6.1.

In the following the S&P 500 data considered so far are used to compare the results of the sample covariance matrix with those of the spectral estimator. Note that only the current 285 stocks whose IPO date is before 1980-01-02 are taken into account. The corresponding portfolio ('S&P 285') is normalized to 1 on the 1st January, 1980.

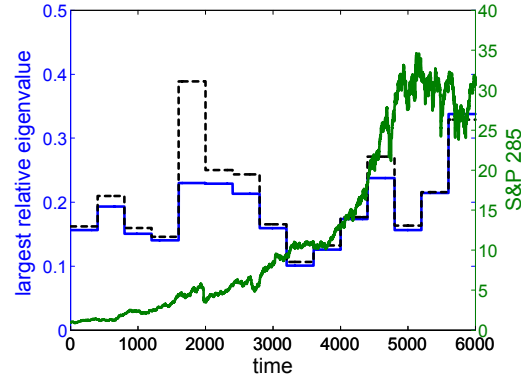


Figure 8.13 Index of the ‘S&P 285’ (green line) from 1980-01-02 to 2003-10-06. Largest relative eigenvalue estimated with the spectral estimator (blue line) and with the sample covariance matrix (dashed line) over 15 time intervals each containing 400 daily log-returns.

The largest relative eigenvalue can be interpreted as the part of the price movements which is due to the common market risk. We see that the sample covariance matrix overestimates the largest relative eigenvalue (i.e. the largest eigenvalue divided by the sum of all eigenvalues) during the 5th period which contains the Black Monday. Generally, the largest eigenvalue obtained by the sample covariance matrix lies above the corresponding result of the spectral estimator except for the last period. Nevertheless, also the spectral estimator indicates that the influence of the main principal component, i.e. the largest eigenvalue varies over time. Therefore it is reasonable to cut off signal from noise for each period separately by using the method discussed above. For the sake of convenience we concentrate on the 5th (which contains the October crash 1987) and on the 9th (the smallest eigenvalue period) time interval.

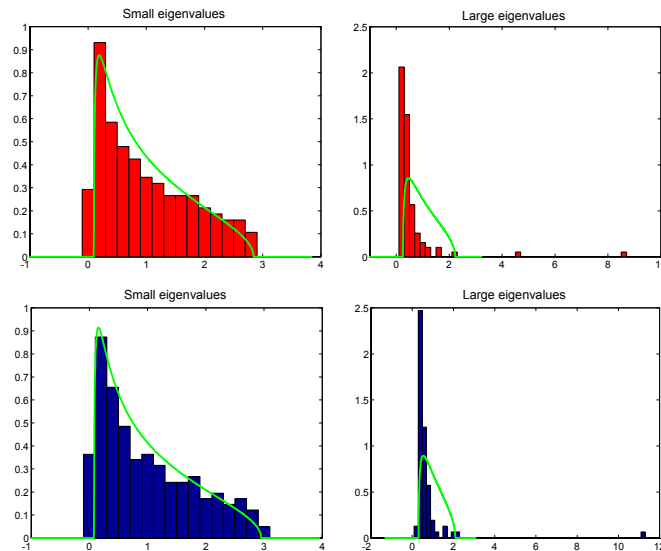


Figure 8.14 Histogram of the small (‘bulk’) eigenvalues (upper left) and of the large eigenvalues (upper right) within the 5th period obtained by the sample covariance matrix after signal-noise separation. The same obtained by the spectral estimator is represented in the lower part. The largest eigenvalues (upper right: 42.23, lower right: 22.85) are suppressed for the sake of transparency.

In the 5th period (see Figure 8.14) the spectral estimator detects 79 large eigenvalues and 206 small eigenvalues (the ‘bulk’). The sample covariance matrix leads to 97 vs. 188 eigenvalues. The left pictures of Figure 8.14 actually indicate that the random noise hypothesis is justifiable. This is not the case for the large eigenvalue part of the data, i.e. the signal seems to be heterogenous.

In the 9th period (see Figure 8.15) the spectral estimator detects 156 large eigenvalues and 129 residuals whereas the sample covariance matrix produces 272 ‘genuine eigenvalues’ and only 13 ‘noise driven eigenvalues’. Even though the spectral estimator has the ability to separate signal from noise more precisely there is no much evidence that the bulk of eigenvalues is purely noise driven. Note that the results of the sample covariance matrix are defective and allow no conclusion.

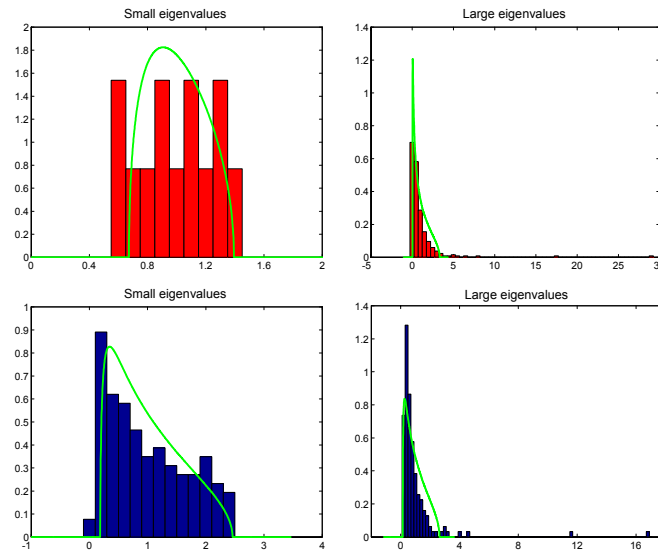


Figure 8.15 Histogram of the small (‘bulk’) eigenvalues (upper left) and of the large eigenvalues (upper right) within the 9th period obtained by the sample covariance matrix after signal-noise separation. The same obtained by the spectral estimator is represented in the lower part.

Hence the validity of the ‘signal/noise model’ discussed in Section 8.2 depends strongly on the considered period. It seems as if the signal/noise paradigm is only justifiable if the market participants agree about the state of the market which is obviously the case when a crash occurs. In a similar empirical study of the S&P 500 Amaral et al. (2002) find only a few driving risk factors (approximately 2% of the number of dimensions) using the sample correlation matrix. They argue that the bulk of eigenvalues can be simply determined by the set of eigenvalues lying within the Marčenko-Pastur bounds λ_{\min} and λ_{\max} already in the original histogram. Indeed, all the empirical studies mentioned above show that financial data exhibit a few yet very large eigenvalues. In this context more than ever it is important to re-normalize the bulk of eigenvalues before drawing any conclusion. This is a possible reason for the different findings presented here.

After the investigations above in the context of financial data it is summarized that

1. the spectral estimator generally leads to smaller estimates of the largest eigenvalue,
2. similarly, brings the driving risk factors into a sharper focus, but
3. even though the largest eigenvalue lies many times over the upper Marčenko-Pastur bound as a contradiction to other empirical findings it cannot be confirmed that there are only a few large but many small and equal eigenvalues.

Nevertheless, it should be pointed out that for dimension reduction via PCA or similar methods it is mainly important that the data can be described properly by a small number of variables. Fortunately, this seems to hold for financial data despite the question whether the bulk of information is pure noise or not.

Summary

The thesis recalls the traditional theory of elliptically symmetric distributions. Their basic properties are derived in detail and some important additional properties are mentioned. Further, the thesis concentrates on the dependence structures of elliptical or even meta-elliptical distributions using extreme value theory and copulas. Some recent results concerning regular variation and bivariate asymptotic dependence of elliptical distributions are presented.

For measuring multivariate asymptotic dependence a new measure called ‘extremal dependence coefficient’ is introduced and calculated explicitly for the multivariate t -distribution. It is pointed out that the probability of simultaneous extremes depends essentially on the heavyness of the tail of the generating distribution function. The tail index is an appropriate measure for the heavyness of the tail. It is shown that for a proper estimation of the tail index one should rely on robust covariance matrix estimation. Therefore, a compact overview of methods for robust covariance matrix estimation is given together with a discussion of their pros and cons.

The traditional class of elliptically symmetric distributions is extended to a new class of ‘generalized elliptical distributions’ to allow for asymmetry. This is motivated by observations of financial data. All the ordinary components of elliptical distributions, i.e. the generating variate \mathcal{R} , the location parameter μ and the dispersion matrix Σ remain. Particularly, it is proved that skew-elliptical distributions belong to the class of generalized elliptical distributions. The basic properties of generalized elliptical distributions are derived and compared with those of elliptically symmetric distributions. It is shown that the essential properties of elliptical distributions hold also within the broader class of generalized elliptical distributions and some models are presented.

Motivated by heavy tails and asymmetries observed in financial data the thesis aims at the construction of a robust covariance matrix estimator in the context of generalized elliptical distributions. A ‘spectral density approach’ is used for eliminating the generating variate. It is shown that the ‘spectral estimator’ is an ML-estimator provided the location vector is known. Nevertheless, it is robust within the class of generalized elliptical distributions since it requires only the assumption that the generating variate has no atom at 0.

The spectral estimator can be used for estimating the empirical generating distribution function, robustly, but preserving the outliers. Thus it is suitable for tail index estimation. By deriving a fixed-point representation of the spectral estimator it is concluded that it corresponds to an M-estimator developed 1983 by Tyler. But in contrast to the more general M-approach used by Tyler (1987a) the spectral estimator is derived on the basis of classical maximum-likelihood theory. Hence, desired properties like, e.g., consistency, asymptotic efficiency and normality follow in a straightforward manner. Both the Fisher information matrix and the asymptotic covariance matrix are derived under the null hypothesis $\Sigma = \sigma^2 I_d$ and compared with the statistical properties of the sample covariance matrix in the case of normally distributed data.

Not only caused by the empirical evidence of extremes but also due to the inferential problems occurring for high-dimensional data the performance of the spectral estimator is inves-

tigated in the context of modern portfolio theory and principal component analysis. The spectral estimator makes an impact especially for risk minimization and principal component analysis when the data is sufficiently heavy tailed.

Further, methods of random matrix theory are discussed. These are suitable for analyzing high-dimensional covariance matrix estimates, i.e. given a small sample size compared to the number of dimensions. It is shown that classical results of random matrix theory fail if the sample covariance matrix is used in the context of elliptically or even generalized elliptically distributed and heavy tailed data. Substituting the sample covariance matrix by the spectral estimator resolves the problem and the classical arguments of random matrix theory remain valid.

The thesis has mainly three contributions listed as follows.

1. The class of elliptically symmetric distributions is generalized to allow for asymmetry and its basic properties are derived,
2. a completely robust covariance matrix estimator is developed and its properties are obtained by maximum-likelihood theory and further,
3. it is shown that the corresponding estimator is a canonical random matrix for applying random matrix theory in the context of generalized elliptical distributions.

List of Abbreviations

a.s.	almost surely
c.d.f.	cumulative density function
e.g.	exempli gratia (for example)
EVT	extreme value theory
GED	generalized extreme value distribution
GPD	generalized Pareto distribution
i.e.	id est (that is)
i.i.d.	independent and identically distributed
MCD	minimum covariance determinant
MDA	maximum domain of attraction
MPT	modern portfolio theory
MVE	minimum volume ellipsoid
PCA	principal component analysis
p.d.f.	probability density function
RMT	random matrix theory
w.l.o.g.	without loss of generality

List of Symbols

0	zero scalar, zero vector, or zero matrix (depending on the context)
$\underline{1}$	vector of ones, i.e. $\underline{1} := (1, \dots, 1)$
$\mathbb{1}_{x \in M}$	indicator function, i.e. $\mathbb{1}_{x \in M} := \begin{cases} 1, & x \in M, \\ 0, & x \notin M. \end{cases}$
$\angle(u, v)$	angle between u and v
$\ \cdot\ $	arbitrary vector norm on \mathbb{R}^d
$\ \cdot\ _2$	Euclidean norm
$\wedge_d(\cdot)$	d -variate minimum copula (cf. Section 2.3.2)
$A \otimes B$	Kronecker product, i.e. if $A \in \mathbb{R}^{q \times r}$ and $B \in \mathbb{R}^{s \times t}$ then $A \otimes B \in \mathbb{R}^{qs \times rt}$ is the matrix obtained by multiplying each element of A with B
A'	transpose of the matrix A
A^{-1}	Moore-Penrose inverse of the rectangular matrix A (cf. Section ‘Mathematical Notation’)
A/x	if A is a matrix and $x \in \mathbb{R} \setminus \{0\}$ then $A/x := x^{-1}A$
$\text{Beta}(\alpha, \beta)$	Beta distribution with parameters α and β
$C(\cdot)$	copula, i.e. a d -variate distribution function $C : [0, 1]^d \rightarrow [0, 1]$ (cf. Section 2.2)
$\tilde{C}(\cdot)$	survival copula corresponding to $C(\cdot)$ (cf. Section 2.3.2)
$\overline{C}(\cdot)$	survival function of $C(\cdot)$, i.e. $u \mapsto \overline{C}(u) := \tilde{C}(\underline{1} - u)$ (cf. Section 2.3.2)
d	number of dimensions
$ \det(A) $	absolute pseudo-determinant of a rectangular matrix A (cf. Section ‘Mathematical Notation’)
$\text{diag}(A)$	diagonal part of a square matrix A , i.e. $\text{diag}(A)$ is a diagonal matrix containing the main diagonal elements of A
\mathcal{D}	diagonal matrix with nonnegative elements
$\mathcal{E}_d(\mu, \Sigma, \phi)$	d -variate elliptical distribution with location vector μ , dispersion matrix Σ , and characteristic generator ϕ

$x_n = o(y_n)$	$x_n/y_n \rightarrow 0, n \rightarrow \infty$
$x_n = O(y_n)$	$\limsup x_n/y_n < \infty, n \rightarrow \infty$
$X_n = o_P(y_n)$	$X_n/y_n \xrightarrow{p} 0, n \rightarrow \infty$
$X_n = O_P(y_n)$	$\text{plim sup } X_n/y_n < \infty, n \rightarrow \infty$
$f(a^-)$	limit ‘from the left’, i.e. $f(a^-) := \lim_{x \nearrow a} f(x)$
$f(a^+)$	limit ‘from the right’, i.e. $f(a^+) := \lim_{x \searrow a} f(x)$
$F_i(\cdot)$	marginal c.d.f. of the i -th random component of a random vector
$F_X(\cdot)$	c.d.f. of the random vector (or variable) X
$F^*(\cdot)$	standard c.d.f., only containing copula parameters (cf. Section 2.3.1)
$F_{\mathcal{R}}(\cdot)$	generating distribution function (cf. Section 1.1)
$F_{\text{Beta}}(\cdot; \alpha, \beta)$	c.d.f. of the Beta distribution with parameters α and β
$\overline{F}_X(\cdot)$	survival function of the random variable X , i.e. $\overline{F}_X := 1 - F_X$
$F^{\leftarrow}(\cdot)$	quantile function, i.e. $p \mapsto F^{\leftarrow}(p) := \inf \{x : F(x) \geq p\}$, $p \in]0, 1[$
$F_X \in \text{MDA}(H_{\xi})$	the c.d.f. F_X belongs to the maximum domain of attraction of the GEV H_{ξ} (cf. Section 2.1)
$F \in \text{MDA}(H_0)$	the c.d.f. F belongs to the Gumbel class (cf. Section 2.1)
$F \in \text{MDA}(H_{\xi>0})$	the c.d.f. F belongs to the Frechet class (cf. Section 2.1)
F_{\min}	minimum of mapped random components, i.e. $F_{\min} := \min \{F_1(X_1), \dots, F_d(X_d)\}$ (cf. Section 2.3.2)
F_{\max}	maximum of mapped random components, i.e. $F_{\max} := \max \{F_1(X_1), \dots, F_d(X_d)\}$ (cf. Section 2.3.2)
${}_1F_1(\cdot; \alpha, \beta)$	confluent hypergeometric function with parameters α and β (cf. Section 4.2.1)
\hat{F}	estimate of the c.d.f. F
$g_{\mathcal{R}}(\cdot)$	density generator (given by the generating variate \mathcal{R}) (cf. Section 1.2)
I_d	d -dimensional identity matrix
\mathcal{I}_0	elementary information matrix of $X \sim N_d(0, I_d)$ after deleting the first column and the first row (cf. Section 5.1)
\mathcal{J}	elementary information matrix of a unit random vector (cf. Section 5.1)
\mathcal{J}_0	elementary information matrix of a unit random vector in the case $\Sigma = \sigma^2 I_d$ (cf. Section 5.1)
$K_{\lambda}(\cdot)$	modified Bessel function of the third kind with index λ

M_n	sample maximum of a sequence of i.i.d. random variables (or vectors) X_1, \dots, X_n
$\mathcal{ME}_d(\mu, \Sigma, \phi)$	d -variate meta-elliptical distribution with underlying elliptical distribution $\mathcal{E}_d(\mu, \Sigma, \phi)$ (cf. Section 2.2)
$N_d(\mu, \Sigma)$	d -variate normal distribution with location vector μ and covariance matrix Σ
$N_d^{\text{sub}}(\mu, \Sigma, \alpha)$	d -variate sub-Gaussian α -stable distribution with location vector μ , covariance matrix Σ , and tail index α
\mathbb{N}	set of natural numbers
\mathcal{O}	orthonormal square matrix
$P_S(\cdot)$	spectral measure (cf. Section 2.2)
q	effective sample size of high-dimensional random matrices, i.e. $n \rightarrow \infty$, $d \rightarrow \infty$, $n/d \rightarrow q < \infty$
$r(A)$	rank of the matrix A
R	vector of returns (cf. Section 7.1)
R_P	portfolio return (cf. Section 7.1.1)
\mathbb{R}	set of real numbers
\mathbb{R}_+	$\mathbb{R}_+ := \{x \in \mathbb{R} : x \geq 0\}$
\mathbb{R}^+	$\mathbb{R}^+ := \{x \in \mathbb{R} : x > 0\}$
$\overline{\mathbb{R}}$	$\mathbb{R} \cup \{-\infty, \infty\}$
\mathcal{R}	generating variate (cf. Section 1.1)
$\text{sgn}(x)$	sign of x , i.e. $\text{sgn}(x) := \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases}$
$\mathcal{SE}_d(\mu, \beta, \Sigma, \phi)$	d -variate skew-elliptical distribution with skewness parameter β (cf. Section 3.1)
S	unit random vector (cf. Section 4.2.1)
S_n	sample of n realizations (cf. Section ‘Mathematical Notation’)
\mathcal{S}^{d-1}	unit hypersphere with $d - 1$ topological dimensions (cf. Section 1.1)
\mathcal{S}_r^{d-1}	Hypersphere with radius r and $d - 1$ topological dimensions (cf. Section 1.2.1)
\mathcal{S}	$\mathcal{S} := \mathcal{S}^1$, i.e. the unit circle
\mathcal{S}_Λ	linear subspace of \mathbb{R}^d spanned by a full rank matrix $\Lambda \in \mathbb{R}^{d \times k}$
$\text{tr}(A)$	trace of the matrix A
$t_d(\mu, \Sigma, \nu)$	d -variate t -distribution with location vector μ , dispersion matrix Σ , and $\nu > 0$ degrees of freedom

$t_{d,\nu}(\cdot)$	c.d.f. of the d -variate t -distribution with ν degrees of freedom
$t_\nu(\cdot)$	c.d.f. of Student's univariate t -distribution with ν degrees of freedom
$\bar{t}_\nu(\cdot)$	survival function of Student's univariate t -distribution with ν degrees of freedom
T	target horizon (cf. Section 7.1.1)
$U^{(d)}$	random vector uniformly distributed on the unit hypersphere \mathcal{S}^{d-1}
$U(0, 1)$	standard uniform distribution
$u(\cdot)$	utility function (cf. Section 7.1.1)
$\text{vec}(A)$	vector which is obtained by stacking the columns of the matrix A
$\text{vec}(\lfloor A)$	vector of the lower triangular part of A without its upper left element (cf. Section 5.1)
$\text{Var}(X)$	covariance matrix of the random vector X (cf. Section 'Mathematical Notation')
$w(\cdot)$	weight function (cf. Section 4.1)
w	portfolio (cf. Section 7.1.1)
$\widehat{W}_d(\cdot)$	empirical distribution function of eigenvalues (cf. Section 8.1.1)
$ x $	absolute value (if x is a scalar) or cardinality (if x is a set)
$x := y$	x is defined as y
$x \propto y$	x is proportional to y
$(x)^{(k)}$	rising factorial, i.e. $(x)^{(k)} := x \cdot (x+1) \cdot \dots \cdot (x+k-1)$ for $k \in \mathbb{N}$ and $(x)^{(0)} := 1$
$x_{0.5}$	median of a random vector X (cf. Section 4.3)
$\widehat{x}_{0.5}$	sample median (cf. Section 4.3)
x_F	right endpoint of the c.d.f. F
$X \in \text{MDA}(H_\xi)$	the same as $F_X \in \text{MDA}(H_\xi)$
\mathbb{Z}	set of integers
α	tail index
β	Beta distributed random variable (cf. Section 1.2.5 and Section 3.3) or vector of asset Betas (cf. Section 7.2)
$\Delta\mu$	vector of excess returns (cf. Section 7.1.1)
$\Delta\mu_M$	excess return of the market portfolio (cf. Section 7.2)
$\delta(u, v)$	radian measure for $u, v \in \mathcal{S}^{d-1}$ (cf. Section 3.4)

ε	vector of idiosyncratic risks (cf. Section 7.2)
$\varepsilon_L, \varepsilon_U$	lower/upper extremal dependence coefficient (cf. Section 2.3.2)
$\zeta(\cdot)$	certainty equivalent (cf. Section 7.1.1)
θ	parameter (scalar, vector, or matrix)
$\hat{\theta}$	estimate of the parameter θ
θ_0, ϑ	vector of copula parameters (cf. Section 2.2)
κ	kurtosis parameter (cf. Section 7.1.2)
λ_L, λ_U	lower/upper tail dependence coefficient (cf. Section 2.3.1)
$\lambda_{\min}, \lambda_{\max}$	Marčenko-Pastur bounds (cf. Section 8.1.2)
μ	location vector
$\hat{\mu}$	location vector estimator
μ_P	expected portfolio return (cf. Section 7.1.1)
ν	degrees of freedom of the t -distribution
ξ	score of a sample element (i.e. the elementary score) (cf. Section 5.1)
ξ_n	sample score (cf. Section 5.1)
$\Pi_d(\cdot)$	d -variate product copula (cf. Section 2.3.2)
ρ	pseudo-correlation coefficient or pseudo-correlation matrix (depending on the context)
σ_P^2	portfolio variance (cf. Section 7.1.1)
σ_M^2	variance of the market portfolio (cf. Section 7.2)
Σ	dispersion matrix, i.e. a positive (semi-)definite matrix
$\hat{\Sigma}$	dispersion matrix estimator
$\hat{\Sigma}_{\text{MP}}$	Marčenko-Pastur operator (cf. Section 8.1.2)
τ	Kendall's τ (cf. Section 2.3.1)
$\phi_X(\cdot)$	characteristic generator of X
$\varphi_X(\cdot)$	characteristic function of X
$\psi(\cdot)$	spectral density function (cf. Section 4.2.1)
$\tilde{\psi}(\cdot)$	skewed spectral density function (cf. Section 4.2.1)
ω	optimal portfolio (cf. Section 7.1.1)
ω_0	global minimum variance portfolio (cf. Section 7.1.1)
ω_M	market portfolio (cf. Section 7.2)
$\Omega_d(\cdot)$	characteristic generator of the uniform distribution on the unit hypersphere \mathcal{S}^{d-1}

Bibliography

- [1] Abdous, B., Genest, C., and Rémillard, B. (2004). ‘Dependence properties of meta-elliptical distributions.’ In: Duchesne, P. and Rémillard, B. (Eds.), *Statistical Modeling and Analysis for Complex Data Problems*, Kluwer.
- [2] Adrover, J.G. (1998). ‘Minimax bias-robust estimation of the dispersion matrix of a multivariate distribution.’ *The Annals of Statistics* **26**: pp. 2301-2320.
- [3] Amaral, L.A.N., Gopikrishnan, P., Guhr, T., Plerou, V., Rosenow, B., and Stanley, E. (2002). ‘Random matrix approach to cross correlations in financial data.’ *Physical Review E* **65**: no. 066126.
- [4] Arnold, B.C. and Beaver, R.J. (2002). ‘Skewed multivariate models related to hidden truncation and/or selective reporting.’ *Test* **11**: pp. 7-54.
- [5] Azzalini, A. (2003). ‘References on the skew-normal distribution and related ones.’ Retrieved 2004-10-08 from <http://azzalini.stat.unipd.it/SN/list-publ.pdf>.
- [6] Azzalini, A. and Dalla Valle, A. (1996). ‘The multivariate skew-normal distribution.’ *Biometrika* **83**: pp. 715-726.
- [7] Barndorff-Nielsen, O.E., Kent, J., and Sørensen, M. (1982). ‘Normal variance-mean mixtures and z distributions.’ *International Statistical Review* **50**: pp. 145-159.
- [8] Barndorff-Nielsen, O.E. and Shephard, N. (2001). ‘Modelling by Lévy processes for financial econometrics.’ In: Barndorff-Nielsen, O.E., Mikosch, T., and Resnick, S. (Eds.), *Lévy Processes - Theory and Applications*, Birkhäuser.
- [9] Barndorff-Nielsen, O.E. and Shephard, N. (2003). ‘Mathematical chapters.’ Lecture notes for the workshop ‘Stochastic Modelling and Statistics in Finance, with Applications’, Mathematisches Forschungsinstitut Oberwolfach, Germany.
- [10] Bilodeau, M. and Brenner, D. (1999). ‘Theory of multivariate statistics.’ Springer.
- [11] Bingham, N.H., Goldie, C.M., and Teugels, J.L. (1987). ‘Regular variation.’ Cambridge University Press.
- [12] Bingham, N.H. and Kiesel, R. (2002). ‘Semi-parametric modelling in finance: theoretical foundation.’ *Quantitative Finance* **2**, pp. 241-250.
- [13] Bouchaud, J.P., Cizeau, P., Laloux, L., and Potters, M. (2000). ‘Random matrix theory and financial correlations.’ *International Journal of Theoretical and Applied Finance* **3**: pp. 391-397.
- [14] Bouchaud, J.P., Cont, R., and Potters, M. (1998). ‘Scaling in stock market data: stable laws and beyond.’ In: Dubrulle, B., Graner, F., and Sornette, D. (Eds.), *Scale Invariance and Beyond*, Proceedings of the CNRS Workshop on Scale Invariance, Les Houches, March 1997, Springer.

- [15] Bouchaud, J.P. and Potters, M. (2000). 'Theory of financial risks - From statistical physics to risk management.' Cambridge University Press.
- [16] Branco, M.D. and Dey, D.K. (2001). 'A general class of multivariate skew-elliptical distributions.' *Journal of Multivariate Analysis* **79**: pp. 99-113.
- [17] Breymann, W., Dias, A., and Embrechts, P. (2003). 'Dependence structures for multivariate high-frequency data in finance.' *Quantitative Finance* **3**: pp. 1-14.
- [18] Cambanis, S., Huang, S., and Simons, G. (1981). 'On the theory of elliptically contoured distributions.' *Journal of Multivariate Analysis* **11**: pp. 368-385.
- [19] Chatfield, C. and Collins, A.J. (2000). 'Introduction to multivariate analysis.' Chapman & Hall / CRC.
- [20] Chopra, V.K. and Ziemba, W.T. (1993). 'The effect of errors in means, variances, and covariances on optimal portfolio choice.' *The Journal of Portfolio Management*, Winter 1993: pp. 6-11.
- [21] Coles, S. (2001). 'An introduction to statistical modeling of extreme values.' Springer.
- [22] Costinot, A., Roncalli, T., and Teïletche, J. (2000). 'Revisiting the dependence between financial markets with copulas.' Working paper, Credit Lyonnais, France. Retrieved 2004-10-14 from <http://gro.creditlyonnais.fr/content/wp/copula-contagion.pdf>.
- [23] Danielsson, J. and Vries, C.G. de (2000). 'Value-at-risk and extreme returns.' Working paper, London School of Economics, United Kingdom. Retrieved 2004-10-14 from <http://www.smartquant.com/references/VaR/var18.pdf>.
- [24] Devlin, S.J., Gnanadesikan, R. and Kettenring J.R. (1981). 'Robust estimation of dispersion matrices and principal components.' *Journal of the American Statistical Association* **76**: pp. 354-362.
- [25] Dickey, J.M. and Chen, C.H. (1985). 'Direct subjective-probability modelling using ellipsoidal distributions.' In: Bernardo, J.M. et al. (Eds.), *Bayesian Statistics 2*, Elsevier and Valencia University Press.
- [26] Donoho, D.L. (1982). 'Breakdown properties of multivariate location estimators.' Ph.D. thesis, Harvard University, United States of America.
- [27] Drouet Mari, D. and Kotz, S. (2001). 'Correlation and dependence.' Imperial College Press.
- [28] Dümbgen, L. (1998). 'On Tyler's M-functional of scatter in high dimension.' *Annals of the Institute of Statistical Mathematics* **50**: pp. 471-491.
- [29] Dümbgen, L. and Tyler, D.E. (2004). 'On the breakdown properties of some multivariate M-functionals.' Preprint, Institute of Mathematical Statistics and Actuarial Science, University of Bern, Switzerland. Retrieved 2004-10-15 from <http://www.stat.unibe.ch/~duembgen/abstracts/Breakdown.html>.
- [30] Eberlein, E. and Hammerstein, E.A. von (2003). 'Generalized hyperbolic and inverse Gaussian distributions: limiting cases and approximation of processes.' Proceedings of the 4th Ascona Conference, Birkhäuser.
- [31] Eberlein, E. and Keller, U. (1995). 'Hyperbolic distributions in finance.' *Bernoulli* **1**: pp. 281-299.

- [32] Embrechts, P., Frey, R., and McNeil, A.J. (2004). ‘Quantitative methods for financial risk management.’ In progress, but various chapters are retrievable from <http://www.math.ethz.ch/~mcneil/book.html>.
- [33] Embrechts, P., Klüppelberg, C., and Mikosch, T. (2003). ‘Modelling extremal events (for insurance and finance).’ Corrected 4th printing, Springer.
- [34] Embrechts, P., McNeil, A.J., and Straumann, D. (2002). ‘Correlation and dependency in risk management: properties and pitfalls.’ In: Dempster, M. (Ed.), *Risk management: value-at-risk and beyond*, Cambridge University Press.
- [35] Fama, E.F. (1965). ‘The behavior of stock market prices.’ *Journal of Business* **38**: pp. 34-105.
- [36] Fang, H.B., Fang, K.T., and Kotz, S. (2002). ‘The meta-elliptical distributions with given marginals.’ *Journal of Multivariate Analysis* **82**: pp. 1-16.
- [37] Fang, K.T., Kotz, S., and Ng, K.W. (1990). ‘Symmetric multivariate and related distributions.’ Chapman & Hall.
- [38] Fisher, R.A. and Tippett, L.H.C. (1928). ‘Limiting forms of the frequency distribution of the largest or smallest member of a sample.’ *Proceedings of the Cambridge Philosophical Society* **24**: pp. 180-190.
- [39] Fisz, M. (1989). ‘Wahrscheinlichkeitsrechnung und mathematische Statistik.’ VEB Deutscher Verlag der Wissenschaften, 11th edition.
- [40] Forrester, P.J., Snaith, N.C., and Verbaarschot, J.J.M. (2003). ‘Developments in random matrix theory.’ arXiv.org e-print archive, Cornell University. Retrieved 2004-10-14 from <http://www.arxiv.org/pdf/cond-mat/0303207>.
- [41] Frahm, G. (1999). ‘Ermittlung des Value-at-Risk von Finanzportefeuilles mit Methoden der Extremwerttheorie.’ Diploma thesis, University of Cologne. Retrievable from <http://home.arcor.de/gfrahm/>.
- [42] Frahm, G., Junker, M., and Schmidt, R. (2003). ‘Estimating the tail-dependence coefficient: properties and pitfalls.’ Working paper, London School of Economics, United Kingdom. Retrievable from <http://stats.lse.ac.uk/schmidt/FrahmJunkerSchmidt.pdf>.
- [43] Frahm, G., Junker, M., and Szimayer, A. (2003). ‘Elliptical copulas: applicability and limitations.’ *Statistics and Probability Letters* **63**: pp. 275-286.
- [44] Gebbie, T. and Wilcox, D. (2004). ‘An analysis of cross-correlations in South African market data.’ arXiv.org e-print archive, Cornell University. Retrieved 2004-10-14 from <http://www.arxiv.org/pdf/cond-mat/0402389>.
- [45] Gnanadesikan, R. and Kettenring J.R. (1972). ‘Robust estimates, residuals, and outlier detection with multiresponse data.’ *Biometrika* **28**: pp. 81-124.
- [46] Haan, L. de (1990). ‘Fighting the arch-enemy with mathematics.’ *Statistica Neerlandica* **44**: pp. 45-68.
- [47] Hassani, S. (1999). ‘Mathematical physics - A modern introduction to its foundations.’ Springer.
- [48] Hayashi, F. (2000). ‘Econometrics.’ Princeton University Press.
- [49] Hiai, F. and Petz, D. (2000). ‘The semicircle law, free random variables and entropy.’ American Mathematical Society.

- [50] Huber, P.J. (1981). ‘Robust statistics.’ Wiley.
- [51] Hult, H. and Lindskog, F. (2002). ‘Multivariate extremes, aggregation and dependence in elliptical distributions.’ *Advances in Applied Probability* **34**: pp. 587-608.
- [52] Joe, H. (1993). ‘Parametric families of multivariate distributions with given margins.’ *Journal of Multivariate Analysis* **46**: pp. 262-282.
- [53] Joe, H. (1997). ‘Multivariate models and dependence concepts.’ Chapman & Hall.
- [54] Johnson, N.L., Kotz, S., and Balakrishnan, N. (1995). ‘Continuous univariate distributions - Volume 2.’ Wiley, 2nd edition.
- [55] Johnstone, I.M. (2001). ‘On the distribution of the largest eigenvalue in principal components analysis.’ *The Annals of Statistics* **29**: pp. 295-327.
- [56] Junker, M. (2002). ‘European real estate firms in crash situations.’ Working paper, CAESAR, Bonn, Germany. Retrieved 2004-10-14 from http://www.caesar.de/uploads/media/cae_pp_0026_junker_2002-08-23_01.pdf.
- [57] Junker, M. (2003). ‘Modelling, estimating, and validating multidimensional distribution functions - with applications to risk management.’ Ph.D. thesis, University of Kaiserslautern, Germany.
- [58] Junker, M. and May, A. (2002). ‘Measurement of aggregate risk with copulas.’ Working paper, CAESAR, Bonn, Germany. Retrieved 2004-10-14 from http://www.caesar.de/uploads/media/cae_pp_0021_junker_2002-05-09.pdf.
- [59] Kelker, D. (1970). ‘Distribution theory of spherical distributions and a location-scale parameter generalization.’ *Sankhya A* **32**: pp. 419-430.
- [60] K llezi, E. and Gilli, M. (2003). ‘An application of extreme value theory for measuring risk.’ Preprint, University of Geneva, Switzerland. Retrieved 2004-10-14 from <http://www.unige.ch/ses/metri/gilli/evtrm/evtrm.pdf>.
- [61] Kempf, A., Kreuzberg, K., and Memmel, C. (2002). ‘How to incorporate estimation risk into Markowitz optimization.’ In: Chamoni, P. et al. (Eds.), *Operations Research Proceedings 2001*, Berlin et al. 2002.
- [62] Kempf, A. and Memmel, C. (2002). ‘On the estimation of the global minimum variance portfolio.’ Working paper, University of Cologne, Germany. Retrieved 2004-10-14 from <http://www.wiso.uni-koeln.de/finanzierung/forschung/veroeffentlichungen/gmvp.pdf>.
- [63] Kent, J.T. and Tyler, D.E. (1988). ‘Maximum likelihood estimation for the wrapped Cauchy distribution.’ *Journal of Applied Statistics* **15**: pp. 247-254.
- [64] Kondor, I., Pafka, S., and Potters, M. (2004). ‘Exponential weighting and random-matrix-theory-based filtering of financial covariance matrices for portfolio optimization.’ arXiv.org e-print archive, Cornell University. Retrieved 2004-10-19 from http://www.arxiv.org/PS_cache/cond-mat/pdf/0402/0402573.pdf.
- [65] Lindskog, F. (2000). ‘Linear correlation estimation.’ Working paper, Risklab, Switzerland. Retrieved 2004-10-14 from <http://www.risklab.ch/Papers.html#LCELindskog>.
- [66] Lindskog, F., McNeil, A.J., and Schmock, U. (2003). ‘Kendall’s tau for elliptical distributions.’ In: Bol, G. et al. (Eds.), *Credit Risk - Measurement, Evaluation and Management*, Physica-Verlag.

- [67] Lintner, J. (1965). 'The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets.' *Review of Economics and Statistics* **47**: pp. 13-37.
- [68] Liu, R.Y. (1988). 'On a notion of simplicial depth.' *Proceedings of the National Academy of Sciences of the United States of America* **85**: pp. 1732-1734.
- [69] Lopuhaä, H.P. (1989). 'On the relation between S-estimators and M-estimators of multivariate location and covariance.' *The Annals of Statistics* **17**: pp. 1662-1683.
- [70] Lopuhaä, H.P. and Rousseeuw, P.J. (1991). 'Breakdown points of affine equivariant estimators of multivariate location and covariance matrices.' *The Annals of Statistics* **19**: pp. 229-248.
- [71] Malevergne, Y. and Sornette, D. (2002). 'Collective origin of the coexistence of apparent RMT noise and factors in large sample correlation matrices.' arXiv.org e-print archive, Cornell University. Retrieved 2004-10-14 from <http://www.arxiv.org/pdf/cond-mat/0210115>.
- [72] Mandelbrot, B. (1963). 'The variation of certain speculative prices.' *Journal of Business* **36**: pp. 394-419.
- [73] Marčenko, V.A. and Pastur, L.A. (1967). 'Distribution of eigenvalues for some sets of random matrices.' *Mathematics of the USSR Sbornik* **72**: pp. 457-483.
- [74] Mardia, K.V. (1972). 'Statistics of directional data.' Academic Press.
- [75] Markowitz, H. (1952). "Portfolio selection." *Journal of Finance* **7**: pp. 77-91.
- [76] Maronna, R.A. (1976). 'Robust M-estimators of multivariate location and scatter.' *The Annals of Statistics* **4**: pp. 51-67.
- [77] Maronna, R.A. and Yohai, V.J. (1990). 'The maximum bias of robust covariances.' *Communications in Statistics: Theory and Methods* **19**: pp. 3925-3933.
- [78] McNeil, A.J. (1997). 'Estimating the tails of loss severity distributions using extreme value theory.' *ASTIN Bulletin* **27**: pp. 117-137.
- [79] McNeil, A.J. and Saladin, T. (2000). 'Developing scenarios for future extreme losses using the POT method.' In: Embrechts, P. (Ed.), *Extremes and Integrated Risk Management*, RISK books.
- [80] Mehta, M.L. (1991). 'Random matrices.' Academic Press, 2nd edition.
- [81] Memmel, C. (2004). 'Schätzrisiken in der Portfoliotheorie - Auswirkungen und Möglichkeiten der Reduktion.' Josef Eul Verlag.
- [82] Merton, R.C. (1980). 'On estimating the expected return on the market: An exploratory investigation.' *Journal of Financial Economics* **8**: pp. 323-361.
- [83] Mikosch, T. (1999). 'Regular variation, subexponentiality and their applications in probability theory.' Lecture notes for the workshop 'Heavy Tails and Queues', EURANDOM, Eindhoven, Netherlands. Retrieved 2004-10-14 from <http://www.math.ku.dk/~mikosch/Preprint/Eurandom/r.ps.gz>.
- [84] Mikosch, T. (2003). 'Modeling dependence and tails of financial time series.' In: Finkenstaedt, B. and Rootzén, H. (Eds.), *Extreme Values in Finance, Telecommunications, and the Environment*, Chapman & Hall.
- [85] Mosler, K. (2003). 'Central regions and dependency.' *Methodology and Computing in Applied Probability* **5**: pp. 5-21.

- [86] Nelsen, R.B. (1999). 'An introduction to copulas.' Springer.
- [87] Oja, H. (2003). 'Multivariate M-estimates of location and shape.' In: Höglund, R., Jäntti, M., and Rosenqvist, G. (Eds.), *Statistics, Econometrics and Society. Essays in Honor of Leif Nordberg*, Statistics Finland.
- [88] Peña, D. and Prieto, F.J. (2001). 'Multivariate outlier detection and robust covariance matrix estimation.' *Technometrics* **43**: pp. 286-301.
- [89] Peracchi, F. (2001). 'Introductory statistics.' Lecture notes, New York University, United States of America. Retrieved 2004-10-14 from <http://www.econ.nyu.edu/dept/courses/peracchi/statistics.html>.
- [90] Pestman, W.R. (1998). 'Mathematical statistics.' De Gruyter.
- [91] Pickands, L. III (1975). 'Statistical inference using extreme order statistics.' *The Annals of Statistics* **3**: pp. 119-131.
- [92] Praag, B.M.S. van and Wesselman, B.M. (1989). 'Elliptical multivariate analysis.' *Journal of Econometrics* **41**: pp. 189-203.
- [93] Prause, K. (1999). 'The generalized hyperbolic model: estimation, financial derivatives, and risk measures.' Ph.D. thesis, University of Freiburg, Germany.
- [94] Rachev, S. and Mitnik, S. (2000). 'Stable Paretian models in finance.' Wiley.
- [95] Rao, C.R. (1962). 'Efficient estimates and optimum inference procedures in large samples.' *Journal of the Royal Statistical Society, Series B* **24**: pp. 46-72.
- [96] Resnick, S.I. (1987). 'Extreme values, regular variation, and point processes.' Springer.
- [97] Resnick, S.I. (1997). 'Discussion of the danish data on large fire insurance losses.' *ASTIN Bulletin* **27**: pp. 139-151.
- [98] Rootzén, H. and Tajvidi, N. (1997). 'Extreme value statistics and wind storm losses: a case study.' *Scandinavian Actuarial Journal* **1**: pp. 70-94.
- [99] Ross, S.A. (1976). 'The arbitrage theory of capital asset pricing.' *Journal of Economic Theory* **13**: pp. 341-360.
- [100] Rousseeuw, P. (1985). 'Multivariate estimation with high breakdown point.' In: Grossmann, W. et al. (Eds.), *Mathematical Statistics and Applications*, Reidel.
- [101] Rousseeuw, P.J. and Driessen, K. van (1999). 'A fast algorithm for the minimum covariance determinant estimator.' *Technometrics* **41**: pp. 212-223.
- [102] Schmidt, R. (2002). 'Tail dependence for elliptically contoured distributions.' *Mathematical Methods of Operations Research* **55**: pp. 301-327.
- [103] Schmidt, R. (2003a). 'Credit risk modelling and estimation via elliptical copulae.' In: Bol, G. et al. (Eds.), *Credit Risk: Measurement, Evaluation and Management*, Physica.
- [104] Schmidt, R. (2003b). 'Dependencies of extreme events in finance - Modelling, statistics, and data analysis.' Ph.D. thesis, University of Ulm, Germany.
- [105] Schoenberg, I.J. (1938). 'Metric spaces and completely monotone functions.' *The Annals of Mathematics* **39**: pp. 811-841.
- [106] Schönfeld, P. (1971). 'Methoden der Ökonometrie - Band II.' Vahlen.
- [107] Sharpe, W.F. (1963). 'A simplified model for portfolio analysis.' *Management Science* **9**: pp. 277-293.

- [108] Sharpe, W.F. (1964). 'Capital asset prices: A theory of market equilibrium under conditions of risk.' *Journal of Finance* **19**: pp. 425-442.
- [109] Sklar, A. (1959). 'Fonctions de répartition à n dimensions et leurs marges.' *Publications de l'Institut de Statistique de l'Université de Paris* **8**: pp. 229-231.
- [110] Stahel, W.A. (1981). 'Breakdown of covariance estimators.' Research Report 31, Fachgruppe für Statistik, ETH Zurich, Switzerland.
- [111] Tobin, J. (1958). 'Liquidity preference as behavior towards risk.' *Review of Economic Studies* **25**: pp. 65-86.
- [112] Tyler, D.E. (1983). 'Robustness and efficiency properties of scatter matrices.' *Biometrika* **70**: pp. 411-420.
- [113] Tyler, D.E. (1987a). 'A distribution-free M -estimator of multivariate scatter.' *The Annals of Statistics* **15**: pp. 234-251.
- [114] Tyler, D.E. (1987b). 'Statistical analysis for the angular central Gaussian distribution on the sphere.' *Biometrika* **74**: pp. 579-589.
- [115] Utsugi, A., Ino, K., and Oshikawa, M. (2003). 'Random matrix theory analysis of cross correlations in financial markets.' arXiv.org e-print archive, Cornell University. Retrieved 2004-10-14 from http://arxiv.org/PS_cache/cond-mat/pdf/0312/0312643.pdf.
- [116] Visuri, S. (2001). 'Array and multichannel signal processing using nonparametric statistics.' Ph.D. thesis, Helsinki University of Technology, Signal Processing Laboratory, Finland.
- [117] Wagner, N. (2003). 'Estimating financial risk under time-varying extremal return behaviour.' *OR Spectrum* **25**: pp. 317-328.
- [118] Wagner, N. (2004). 'Auto-regressive conditional tail behavior and results on government bond yield spreads.' Working paper, Technische Universität München, Germany. Retrieved 2004-10-12 from <http://www.gloriamundi.org/picsresources/nwact.pdf>.
- [119] Wigner, E.P. (1958). 'On the distribution of the roots of certain symmetric matrices.' *The Annals of Mathematics* **67**: pp. 325-327.
- [120] Yin, Y.Q. (1986). 'Limiting spectral distribution for a class of random matrices.' *Journal of Multivariate Analysis* **20**: pp. 50-68.