Methodologies in Factor Modeling

Zur Erlangung des akademischen Grades eines Doktors der
Wirtschaftswissenschaften

(Dr. rer. pol.)

von der Fakultät für Wirtschaftswissenschaften der Universität Fridericiana zu
Karlsruhe

genehmigte

DISSERTATION

von

Dipl.-Math., Dipl.-Inform. Sebastian Kring

Danksagung

Mein erster Dank gilt meiner Frau Susanne, die mich stets während der Dissertation unterstützt hat und mir in schwierigen Phasen meiner Arbeit Rückhalt gegeben hat. Ihr widme ich diese Arbeit.

Meinen Eltern danke ich für die Förderung, die sie mir auf meinem gesamten Lebensweg haben zukommen lassen.

Allen Professoren und Mitarbeitern des Lehrstuhls für Statistik, Ökonometrie und mathematische Finanzwirtschaft der Universität Karlsruhe danke ich für die gemeinsame Zeit während der vergangenen zweieinhalb Jahre. Insbesondere danke ich Prof. Kuno Egle und Prof. Wolfgang Stummer, die mich auf die Möglichkeit aufmerksam gemacht haben, bei Prof. Svetlozar T. Rachev zu promovieren. Meinem Freund und Kollegen Dr. Markus Höchstötter bin ich für die vielen konstruktiven Gespräche dankbar.

Prof. Frank J. Fabozzi von der Universität Yale danke ich für die Unterstützung beim Anfertigen meiner Dissertation.

Schließlich möchte ich meinem Doktorvater Prof. Svetlozar T. Rachev für die Auswahl des Themas und die mir gewährten Möglichkeiten danken.

Karlsruhe, Dezember 2007   Sebastian Kring

iv

Contents

Chapter 1

Overview about Factor Models

1.1 Introduction

Factor models are a fundamental approach in asset pricing and portfolio management. In asset pricing theory they can help to explain and determine the price and return of an asset. They offer parsimony, i.e. the ability to describe a large set of security returns in terms of few factors, and the capacity to identify common sources of correlation among securities. In portfolio management they are used to quantify a portfolio's return and risk characteristics. In particular, when dealing with large portfolios, only factor models can make a quantitative approach to portfolio analysis feasible, since they allow for financial modeling in much lower dimensions.

The literature about factor models starts with a one factor model introduced by Sharpe (1963). The purpose of this model was to simplify the estimation of the covariance matrix in the mean-variance approach developed by Markowitz (1952). One year later the famous Capital Asset Pricing Model (CAPM) was derived by Sharpe (1964) and Lintner (1965a), which is a one factor model. Ross (1976) developed the Arbitrage Pricing Theory (APT), implying a linear multifactor approach for asset pricing. In the last three decades various factor models have been developed. One of the first studies investigating and documentating the relationship between multiple factors and asset returns was the one by Chen et al. (1986) for the US financial market. Since then a score of studies, for example McElroy, and Burmeister (1988), Poon, and Taylor (1991), Clare, and Thomas (1994), Jagannathan, and Wang (1996), Reyfman (1997) and others, have identified the effects of such magnitudes as labor income, industrial production, inflation and other macroeconomic variables. Alternatively, Fama, and French (1993) and (1996) developed a three factor model (market, small market value minus big market portfolio, highbook/market minus low book/market portfolio). They found out that these factors successfully explain the average returns of size and book market sorted portfolios, and also of other strategies. Although there is no satisfactory

theory explaining this phenomenon, these findings may suggest that these fundamental factors are proxies for some macroeconomic 'distress' of 'recession' factor.

An economical justification for factor modeling can be achieved very elegantly by the stochastic discount factor approach (see Cochrane (2001) for a thorough discussion of this approach). This approach can be described by the following two equations

$$
\begin{aligned}
P_t &= E(m_{t+1}X_{t+1}|\mathcal{F}_t) \\
m_{t+1} &= f(data, parameters),
\end{aligned}
$$

where P_t is the asset's price at time t, X_{t+1} is the asset payoff at time $t+1$, $E(.|\mathcal{F}_t)$ denotes the conditional expectation operator at time t, f denotes some function, and m_{t+1} is the stochastic discount factor. The stochastic discount factor is a random variable that computes market prices today (time t) by discounting, state by state, the corresponding future payoff. Certainly, the advantages of the stochastic discount factor approach are its universality, unification of more specific theories, and simplicity. The fundamental obstacle in this approach is that we do not know m_{t+1}. In the consumption based theory one can show that the stochastic discount factor can be expressed in terms of

$$
m_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)},
$$

where c_t and c_{t+1} are consumption at time t and $t+1$, respectively, $u(.)$ is an increasing and concave utility function and the constant β captures the investor's impatience[1]. The term $\beta \frac{u'(c_{t+1})}{u'(c_t)}$ can be interpreted as marginal utility. It should be plausible that macroeconomic risk factors, which cannot be diversified and describe the state of the economy, determines marginal utility of an investor. This argument can be expressed mathematically by the following equation

$$
m_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)} \approx \sum_{i=1}^{n} b_i F_{t+1,i}. \tag{1.1}
$$

In the appendix we show that equation (1.1) leads to a factor model of the form

$$
E(R_{i,t+1}|\mathcal{F}_t) = R_f + \sum_{j=1}^{p} \beta_{ij} E(F_{t+1}|\mathcal{F}_t), \tag{1.2}
$$

where $\beta_{ij} \in \mathbf{R}$, $i = 1, ..., d$ and $j = 1, ..., p$, are called factor loadings and measure the sensitivity between the factor returns and the asset returns. Equation (1.2) says if one knows the exposures (factor loadings) and the conditional mean of the factor returns,

---

[1] A high β refers to an impatient investor.
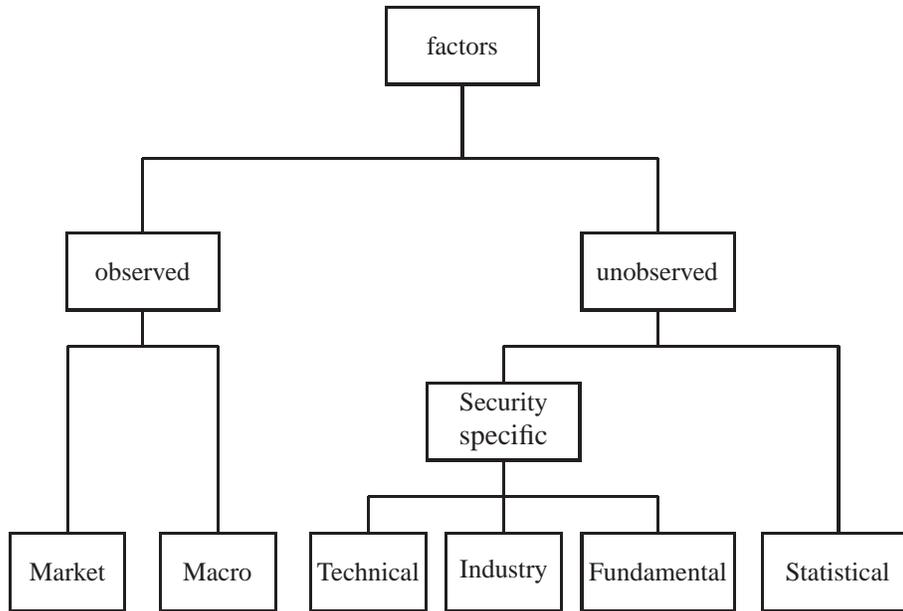
Figure 1.1: Hierachy of Factors

one can forecast the asset return and thus, the fair price of the asset.

In portfolio and risk management we are primarily focused on understanding the variances and covariances of the asset returns in the portfolio in order to understand the risk characteristics. For example, if one has a portfolio consisting of d assets one has to estimate $d(d+1)/2$ variances and covariances for the risk characteristics, i.e. in case of 200 and 2000 stocks in a portfolio, we have to estimate over 20 thousand and 20 million variances and covariances, respectively! In most cases available sample sizes are insufficient to estimate robustly a positive definite covariance matrix. Without a factor model describing the behavior of asset returns by much fewer factors, say up to 50, a quantitative approach to portfolio analysis is simply impossible. With a well working factor model we can estimate the covariance matrix of the factor returns and hence, the covariance matrix of the asset returns.

Equity risk factor models take a variety of forms. A basic categorization of factor models is whether the factor returns are observed or unobserved. Factor models that rely on observed factor returns include market and macroeconomic factor models. Alternatively, factor models exhibiting unobserved factors are statistical or fundamental factor models. Figure 1.1 depicts the hierarchy of the factors. In order to make the classification a bit less abstract, Table 1.1 depicts examples for each factor class. Figure 1.1 and Table 1.1 are adopted from Zangari (2003).

Factor Class	Examples
Market	S&P 500, Wishire 5000, MSCI World indexes
Macroeconomic	Industrial Production, unemployment rate, interest rate
Technical	Excess stock return on previous month, trading volumes
Sector	Energy, transportation, technology
Fundamental	Value, growth, return on equity
Statistical	Principal components

Table 1.1: Examples of Factors

The different classes of factor models require different kinds of estimation techniques.

This chapter is organized as follows. In Section 1.2 we deal with observed factor models. We present their basic properties, give examples and show how to calibrate them to data. In Section 1.3 we consider fundamental and statistical factor models. As mentioned before, both are unobserved factor models but they are based on different approaches. In the last section we give an outlook of the thesis and describe how our work contributes to the field of financial factor modeling.

1.2   Factor Models with Observed Factors

In this section we consider observed factor models, i.e. market and macroeconomic factor models. We assume that we have identified a set of observed factors explaining and determining asset returns. In practical work this identification process is driven by macroeconomic consideration and intensive data analysis. In the model construction process these factors are treated as exogenous variables.

After defining a factor model and presenting its mathematical basic properties we give two examples of observed factor models. The first one is the so-called market model that can be derived immediately from the CAPM. The second one is a macroeconomic factor model proposed by Chen et al. (1986), which can be considered to be an extension of the market model by adding additional macroeconomic factors. We conclude this section with statistical calibration of factor models.

1.2.1   Definition and Basic Properties

By using an observed factor model we try to explain and determine the randomness in the components of a d-dimensional vector X in terms of a smaller set of common factors $F = (F_1, ..., F_p)'$. If the components of $X = (X_1, ..., X_d)'$ represent asset returns, it is clear that a large part of their variation can be explained in terms of the variation of a smaller set of market or macroeconomic factors. Formally we define a factor model as follows according to McNeil, Frey, and Embrechts (2005).

Definition 1. *A d-dimensional stochastic variable X is said to follow a p-factor model if it can be decomposed as*

$$X = a + BF + \epsilon, \tag{1.3}$$

where

   (i) $F = (F_1, ..., F_p)'$ is a random vector of common factors with $p < d$ and a covariance matrix that is positive definite;

   (ii) $\epsilon = (\epsilon_1, ..., \epsilon_d)'$ is a random vector of idiosyncratic error terms*, which are uncorrelated and have mean zero;*

   (iii) $B \in \mathbf{R}^{d \times p}$ is a matrix of constant factor loadings and $a \in \mathbf{R}^d$ is a vector of constants; and

   (iv) $\mathrm{Cov}(F, \epsilon) = E((F - E(F))\epsilon') = 0.$

Definition 1 says if we know the factors F we can determine X up to an idiosyncratic error term ϵ. According to equation (1.3) the factor model induces a structure for the covariance matrix Σ of X. If we denote the covariance matrix of F by Ω and that of ϵ by the diagonal matrix Υ, it follows

$$\begin{aligned}
\Sigma &= \mathrm{Var}(X) = \mathrm{Var}(a + BF + \epsilon) \\
&= \mathrm{Var}(BF) + \mathrm{Var}(\epsilon) + B\,\mathrm{Cov}(F, \epsilon) + \mathrm{Cov}(\epsilon, F)B' \\
&= B\Omega B' + \Upsilon. \tag{1.4}
\end{aligned}$$

Hence, if we know the covariance matrix of F and the factor loading B (see appendix for a statistical calibration), we can approximate the covariance matrix of X.

If we transfer Definition 1 in a time series framework meaning we replace the random variable X by the process $(X_t)_{t \in \mathbf{Z}}$, F by $(F_t)_{t \in \mathbf{Z}}$ and ϵ by $(\epsilon_t)_{t \in \mathbf{Z}}$, we obtain

$$X_t = a + BF_t + \epsilon_t \text{ for all } t \in \mathbf{Z}, \tag{1.5}$$

where we assume that the process $(\epsilon_t)_{t \in \mathbf{Z}}$ is a martingale difference (see chapter 3 for a definition of this notion) and independent of $(F_t)_{t \in \mathbf{Z}}$. In particular, if we are capable to forecast the future values of factor returns – this can be done by economical expertise or an econometric model for the factor return – we can predict the future value of the assets and asset returns. Mathematically expressed, we have

$$E(X_{t+1}|\mathcal{F}_t) = a + BE(F_{t+1}|\mathcal{F}_t),$$

where $E(\epsilon_{t+1}|\mathcal{F}_t) = 0$, since e_{t+1} is independent from present and past state of the world. As stressed in the introduction of this chapter, forecasting expected values of asset returns using historical information is one of the main applications of factor models, especially, when we are dealing with long horizon return data. As shown by Kim, Malz, and Mina (1999) forecasts for a horizon of at least three months produce relatively accurate predictions of future returns.

On the contrary, when working with daily or weekly asset and factor return data we can apply factor models to predict the temporal dependence structure of asset returns and to explain systematic variations and comovements among stock returns. In mathematical terms we have

$$
\begin{aligned}
\Sigma_t &= \operatorname{Var}(X_{t+1}|\mathcal{F}_t) \\
&= B\operatorname{Var}(F_{t+1}|\mathcal{F}_t)B' + \operatorname{Var}(\epsilon_{t+1}|\mathcal{F}_t) \\
&= B\Omega_t B' + \Upsilon_t.
\end{aligned}
$$

Instead of modeling the process of the conditional covariance $(\Sigma_t)_{t\in\mathbf{Z}}$, we focus on modeling the process $(\Omega_t)_{t\in\mathbf{Z}}$. In case of $p << d$ this is a huge simplification of estimating and modeling the covariance process. These covariance processes can be modeled by multivariate GARCH processes and we will see in chapter 3 that the dimension of these processes is the bottle-neck in terms of fitting the process to data.

Summing up, for daily or weekly asset return data factor models are primarily applied to model the temporal dependence structure in terms of the conditional covariance matrix, while for three monthly return data and data based on even longer periods factor models are used to forecast expected return.

1.2.2   Two Examples

The first asset pricing model derived from economic theory is the Capital Asset Pricing Model[2]. The CAPM is given by the following formula

$$E(X_i) = X_f + \beta_i(E(X_M) - X_f), \tag{1.6}$$

where $E(X_i)$ = the expected return on asset i;

$E(X_M)$ = the expected return on the "market portfolio";

X_f = return on a risk-free security; and

β_i = measure of systematic risk of asset i relative to the "market portfolio".

One can further show that β_i in equation (1.6) is given by $\frac{\operatorname{Cov}(X_i,X_M)}{\operatorname{Var}(X_m)}$. The empir-

---

[2]For a derivation of this model see Sharpe (1964) and Lintner (1965a).

ical analogue for equation (1.6) is

$$X_{t,i} - X_{t,f} = \alpha_i + \beta_i(X_{t,M} - X_{t,f}) + \epsilon_{t,i}, \tag{1.7}$$

where $\epsilon_{t,i}$ is the idiosyncratic error term and $t = 1, .., n$. Equation (1.7) is called the characteristic line. Although simple, the CAPM may not offer the practitioner a useful way to measure and explain risk. A manager may mistakenly select the wrong or not sufficient proxy for the market portfolio in the analysis or may simply be interested in a richer model to help explain sources of risk and return. Fama, and French (1996) have shown that the market portfolio does a rather poor job at explaining movements in individual stock returns. Thus, the market return may not be the only factor explaining movements in excess stock returns, and therefore more factors are needed.

Chen et al. (1986) have investigated whether an extension of the CAPM can help to explain stock returns by adding a set of macroeconomic factors. Examples of the macroeconomic factors include: (1) the growth rate in monthly industrial production; (2) a measure of default premium, measured as the difference between the return on a high-yield bond index and the return on long-term government bonds.; (3) the real interest rate; (4) the maturity premium, measured as the difference between return on the long-term government bond and the one-month Treasury bill return; and (5) the change in monthly expected inflation.

We incorporate the macroeconomic factors into the market model as follows. We do not assume anymore that the expectation of the error term in equation 1.7 is zero, since it is driven by macroeconomic risk factors that cannot be diversified and thus, the investor should be rewarded for taking these risks.

$$E(X_i) = X_f + \beta_i(E(X_M) - X_f) + E(e_i) \tag{1.8}$$

$$E(e_i) = \sum_{j=1}^{K-1} \beta_{ij} E(F_j), \tag{1.9}$$

where β_{ij} measures the exposure of the jth factor on the ith asset and $E(F_j)$ is the expected return of the jth factor. The empirical analogue for equations (1.8) and (1.9) is

$$X_{t,i} - X_{t,f} = \alpha_i + \beta_i(X_{t,M} - X_{t,f}) + e_{t,i}$$

$$e_{t,i} = \sum_{j=1}^{K-1} \beta_{ij} F_{t,j} + \epsilon_{t,i}, \tag{1.10}$$

where $F_{t,j}$ =   return on the ith macroeconomic factor at time t;

β_{ij} =   exposure of the jth factor on the ith asset; and

$\epsilon_{t,i}$ = ith security's idiosyncratic return.

Note, that we have assumed the factor loadings β_{ij} and the asset return's α_i to be constant over time in both models. However, this assumption is questionable, since fundamental changes in a company, such as mergers and acquisitions or changes in the capital structure, should have considerable influence on the factor exposures.

1.3 Factor Models with Unobserved Factors

In this section we consider factor models whose factors are not observable. According to Table 1.1 we distinguish between statistical factor models and security specific factor models including technical, sector and fundamental factor models. The key issue in both classes of factor models is to estimate the unobserved values of the factors. We will see that we require fundamentally different estimation techniques for calibrating these two model classes.

1.3.1 Statistical Factor Models

In statistical factor modeling we extract a number of unobserved factors from a multivariate return series. The aim of statistical factor models is not to explain the returns of an asset but rather to reduce the dimensionality of the data by finding a small number of independent or at least uncorrelated factors that account for the systematic variations and comovements among asset returns. The identified factors are linear combinations of the asset returns and, certainly a drawback, they are difficult to interpret economically in most cases.

Principal Component Analysis

The most important statistical method to extract factors from data is the principal component analysis (PCA) and this for two reasons. First, many commercially available risk systems use this method as part of their risk models. Second, for many practitioners PCA is what often comes to mind when thinking about factors and factor models. In Chapter 2 we introduce PCA in a very general framework that has not been developed so far in this chapter. The concept of PCA is used throughout the thesis. After defining this method in Chapter 2, we apply it in the factor composed MGARCH model (see Chapter 3). In Chapter 4 we show that multi-tail elliptical distributions allow to model principal components with different tail indices. The PCA has been

extensively applied in the financial literature. For example, Alexander (2002) used the PCA in the so-called PC-GARCH model (see Chapter 3) and Borovkova (2006) used the PCA very successfully in the commodity future market.

Factor Analysis

This section follows Johnson, and Wichern (1982). The essential purpose of *factor analysis* is to describe, if possible, the covariance relationship among asset returns in terms of a few underlying, but unobservable, random quantities called factors. We assume the model described in Definition 1. Our objective is to find decomposition of Σ according to equation (1.4). In addition, we assume that the covariance matrix Ω of the factors equals the identity. This assumption leads to the decomposition

$$\Sigma = BB' + \Upsilon. \tag{1.11}$$

It is important to note, that equation (1.11) determines the unobserved factors F only up to an orthogonal transformation $\Gamma \in \mathbf{R}^{d \times d}$, since equation (1.3) in Definition 1 can be rewritten

$$\begin{aligned} X &= a + BF + \epsilon = aB\Gamma\Gamma'F + \epsilon \\ &= a + \tilde{B}\tilde{F} + \epsilon, \end{aligned}$$

where $\tilde{B} = B\Gamma$ and $\tilde{F} = \Gamma'F$. Thus, we have

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{Var}(a + \tilde{B}\tilde{F} + \epsilon) = \tilde{B}\,\mathrm{Var}(\tilde{F})\tilde{B}' + \Upsilon \\ &= B\Gamma\Gamma'\Gamma\Gamma'B' + \Upsilon = BB' + \Upsilon. \end{aligned}$$

In factor analysis there are two very popular methods to obtain such a decomposition. The first one is the principal component method and the second one the maximum likelihood method.

The principal component method

The principal component method is based on the spectral decomposition theorem (see appendix 1.5.2). Let Σ have eigenvalue-eigenvector pairs (λ_i, v_i) with $\lambda_1 \geq \lambda_2 \geq ...\lambda_d \geq 0$. Then we have due to the spectral decomposition theorem

$$\begin{aligned} \Sigma &= \sum_{i=1}^{d} \lambda_i v_i v_i' \\ &= (\sqrt{\lambda_1}v_1|...|\sqrt{\lambda_d}v_d)(\sqrt{\lambda_1}v_1|...|\sqrt{\lambda_d}v_d)'. \end{aligned} \tag{1.12}$$

Equation (1.12) fits in the covariance structure of equation (1.11). Since we are primarily interested in reducing the dimensionality of the factors we try to explain the covariance structure in terms of a few common factors. One approach, when the last $d-p$ eigenvalues are small, is to neglect the contribution of $\sum_{i=p+1}^{d} \lambda_i v_i v_i'$ to Σ. Thus, we obtain the approximation

$$\Sigma \approx \underbrace{(\sqrt{\lambda_1} v_1 |...| \sqrt{\lambda_p} v_p)}_{=B} \underbrace{(\sqrt{\lambda_1} v_1 |...| \sqrt{\lambda_p} v_p)'}_{=B'} \qquad (1.13)$$

This approximation assumes that the idiosyncratic error terms ϵ in Definition 1 are of minor importance and can be ignored.

If one is interested in including the error terms, one can take the diagonal elements of $\Sigma - BB'$, where BB' is defined in equation (1.13). Allowing for specific factors, the approximation becomes

$$\Sigma \approx BB' + \Upsilon$$

where $\Upsilon_{ii} = \sigma_{ii} - \sum_{j=1}^{p} B_{ij}$, $i = 1, 2, ..., p$ and $\Upsilon_{ij} = 0$, $i \neq j$.

The maximum likelihood method

If the common factors F and the idiosyncratic error terms are assumed to be normally distributed we can obtain the matrix B of factor loadings and the specific variances Υ through maximum likelihood estimates. When F_j and ϵ_j, $j = 1, ..., n$ are jointly normal, the observations $X_j - \mu = BF_j + \epsilon_j$ are then normal, and the likelihood is

$$L(\mu, \Sigma) = (2\pi)^{-np/2} |\Sigma|^{-n/2} e^{-\frac{1}{2} \sum_{j=1}^{n} (x_j - \bar{x}) \Sigma^{-1} (x_j - \bar{x})}$$

which depends on B and Υ through $\Sigma = BB' + \Upsilon$. This model is still not well defined because the solution is unique up to an orthogonal transformation as shown before.

We can achieve uniqueness of the solution by imposing the following computationally efficient uniqueness condition

$$B'\Upsilon B = \Delta \text{ is diagonal.} \qquad (1.14)$$

1.3.2   Fundamental Factor Models

This section follows Zangari (2003). The fundamental factor models relate to the situation where the factors are unobserved, but the loading or exposure matrix B is assumed to be known. More precisely, we group every asset in the considered universe by geographical or industrial sector, firm size or other important characteristics. For example, the return of a European technology company might have a high exposure to

an unobserved factor representing technology companies and to an unobserved factor representing European companies.

A very popular statistical approach for fundamental factor models is a linear cross sectional model that is defined by

$$X_t = B_{t-1}F_t + \epsilon_t, \tag{1.15}$$

where X_t $=$ d-dimensional vector of one-period asset (stock) returns;
        B_{t-1} $=$ $d \times p$ matrix of asset exposures to factors as of time $t-1$;
        F_t $=$ p-dimensional vector of one-period factor returns; and
        ϵ_t $=$ d-dimensional vector of one-period specific returns.

The model given in equation (1.15) can be interpreted as a regression model, where the asset returns are treated as dependent variables and the exposures are the independent variables. At any point in time we perform a cross-sectional regression leading to

$$F_t = (B'_{t-1}B_{t-1})^{-1}B_{t-1}X_t$$

where $t = 1, ..., n$ (see Murray (2006) for a thorough discussion of ordinary least square regression). In most applications of the cross sectional model it is too restrictive to assume that idiosyncratic error terms $\epsilon_{t,i}$, $i = 1, ..., d$, have the same variances. Instead, we assume that the covariance matrix of the specific returns, ϵ_t, $t = 1, ..., n$, is given by

$$\Upsilon_t = \begin{pmatrix} \sigma_{t,1}^2 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_{t,d}^2 \end{pmatrix},$$

where $\sigma_{t,i}^2 \neq \sigma_{t,j}^2$ for $i \neq j$, $i, j = 1, ..., d$. In this situation ordinary least square regression is not efficient and we have to apply generalized least square (GLS) regression which leads to

$$F_t = (B'_{t-1}\Upsilon_t B_{t-1})^{-1}B_{t-1}\Upsilon_t^{-1}X_t,$$

where $t = 1, ...n$ (see Murray (2006) for generalized least square regression).

Once, we have estimated the returns of the fundamental factors we can apply econometric models in order to predict factor returns or to identify the temporal dependence structure between them.

Asset Exposure

So far, we have not discussed the different types of factors in a fundamental factor model and how to determine the corresponding exposures. In particular, in a linear cross-sectional factor model determining exposures is one of the essential tasks to do, since they are treated as independent or exogenous variables in the model.

Typically, in a fundamental factor model an asset can have exposures to:

- itself

- a particular industry or sector

- a currency

- a country

- investment styles: volatility, market capitalization, value and others.

In the following we explain how to determine the raw exposure of an asset to a specific class of factors, e.g. industry or investment styles. Subsequently, we show how to standardize the raw exposures[3] in order to combine them consistently in a linear factor model.

Industry Exposures

Probably the easiest set of exposures to understand are industry exposures. There are two different classification schemes to allocate an asset's exposure to multiple industries or groups of industries (sectors).

(i) An asset's exposure to an industry is one if it is in that industry, otherwise zero.

This classification may be coarse for companies belonging to multiple industries or sectors. For example, a company can be assigned to two industries, where in one industry the company generates 90% of its earning. In this situation it is questionable to apply classification scheme (i).

(ii) If a company is exposed to more than one industry, we assign weights to these industries or sectors according to certain criterion, where the weights sum up to 1.

The question is of course what the right criteria to choose are. One criteria might be the percental profits or earnings gained in the industries.

---

[3]Note that raw exposures of different factor classes have different units and have to be normalized in some way.

These industry assignments are provided by various vendors. E.g., in the United States we have Barra and Russel among others and in Europe there are FTSE and Dow Jones STOXX and others. Note that each classification scheme groups industries in different sectors providing a coarser grouping of assets.

Country or Local Market Exposures

In the same way in the case of industry exposures we present two ways to define an asset's country exposure. The first approach takes the following form.

(i) An asset's country exposure takes a value of one if it belongs to a country, zero otherwise.

The crucial question in this approach is, what do we mean by the term *belongs*? Is it the country of domicile (the country where a company has been registered), the country of issuance (the country where the stock has been issued) or is it the country where the company sells its products? However, we see there are problems with using this approach. An alternative approach to define an asset's country exposure is to use local market betas, leading to the following four step algorithm.

(ii)   1) Assign each asset to a country or countries

       2) Identify the market portfolio corresponding to each country. This portfolio is referred to as the local market index.

       3) Regress the returns of the asset on the returns of the local market index to get the beta.

       4) The estimated value of beta is that asset's exposure to the country.

This four step process applies to estimating multiple country exposures (i.e., multiple betas) for a particular asset.

Investment Style/Risk Exposures

Investment style exposures capture an asset's sensitivity to a particular investment strategy. In the following we describe examples of very popular investment style asset exposures.

(i) The *volatility factor* captures the volatility of assets. Assets that have high (low) historical volatility have a high (low) exposure to the volatility factor. An asset's exposure to volatility may be computed as the standard deviation of its historical returns.

(ii) The *market capitalization* also known as the size factor explains the asset returns in terms of the company's market capitalization. Companies with large (small)

market capitalization have high (low) exposure to the size factor. An asset's raw exposure to this factor is defined as the square root of the observed market capitalization of the company.

(iii) The *value* factor captures a company's value orientation. An asset's exposure to value may be calculated as the ratio of its price to book value.

Note, that the investment style exposures are measured in different units. We call this exposure also raw exposure. In the next section we show how to standardize exposures.

1.3.3 Standardizing Exposures

In a fundamental factor model we have to standardize an asset's raw exposures to different investment style factors, since we have to make exposures across different investment styles comparable. For example, a company has a market capitalization of \$1 billion which gains a market size exposure of \$31,663 (see (ii) above). On the other hand the same asset has an exposure to volatility of 24% (historical volatility annualized). It is obvious that we have to convert them to standardized units, if we want to incorporate these raw exposures into a factor model using regression.

In order to incorporate these raw exposures into a factor models using regression, we have to convert them to standardized units.

We discuss two methodologies in order to standardize asset exposures. For the first approach we have to carry out the following steps.

1) Define the universe of assets over which a particular group of exposures will be standardized.

2) Compute the average raw exposure of this universe.

3) Compute the standard deviation of exposures for this universe.

4) An asset's standardized exposure is defined as

$$\beta_i = \frac{\beta r_j - \mu}{\sigma}, \tag{1.16}$$

where βr_j is the raw exposure of the jth asset, μ is the mean of the exposures in the considered universe, and σ is the standard deviation of the exposures.

The standardized exposures from this approach are measured in units of standard deviation. In practice, there are some variations to this methodology. For example, the universe used to standardize investment style exposures may be based on the company's industry classification. In this case we group the asset according to its industry

classification and we standardize the company's raw exposure, e.g. size factor, due to its industry group.

Chan, Karecedski, and Lakonishok (1998) suggest an alternative approach for standardization. They propose a three step algorithm.

1) Define the universe of assets over which a particular group of exposures is to be standardized.

2) Rank exposures.

3) Rescale the ranked exposures so that their values lie between 0 and 1 according to

$$\beta_i = \frac{\rho(\beta r_i) - 1}{\max_{j=1,\dots,n}(\rho(\beta r_j) - 1)},$$

where n is the number of assets in the universe, βr_j is the raw exposure of the jth asset, and the function $\rho(.)$ ranks the raw exposures.

1.4 Motivation and Outline of the Thesis

As we have seen in the preceding sections one of the key applications of factor models is to understand and to identify the risk and return characteristics of large asset portfolios that demands a thorough modeling of the factor returns. The contribution of this thesis to the field of factor models is to develop methods and models allowing to capture the important style facts about factor returns. In our work, we focus on statistical factor models which are based on principal component analysis.

In empirical work we observe that principal components exhibit following statistical properties.

(i) Heavy tailed return distributions.

(ii) Tail dependency.

(iii) Multiple tail indices.

(iv) Clustering of the volatility.

An important class of distributions dedicated to capture properties (i) to (iii) are α-stable distributions as shown by Rachev, and Mittnik (2000). Property (iv) is a time series property. This property has been modeled with overwhelming success by ARCH and GARCH processes.

One of the major objectives of the thesis is to combine these two concepts leading to an α-stable multivariate GARCH model (see Chapter 3).

In Chapter 2 we develop a moment type estimator for the dispersion matrix of an α-stable sub-Gaussian distribution. We generalize this estimator so that it is capable to estimate the dispersion matrix of any elliptical distribution. Such an estimator is important in order to conduct an α-stable Principal Component Analysis. The chapter corresponds to Kring et al. (2007).

In Chapter 3 we present a new class of multivariate GARCH models that we label composed and factor composed multivariate GARCH models. We show important statistical properties of these processes such as covariance stationarity and transformation invariance. Furthermore, we present estimation algorithms in order to calibrate these processes. We conclude the chapter by introducing an α-stable version of the composed and factor composed multivariate GARCH processes and applications. This chapter refers to Kring et al. (2007b).

In Chapter 4 we introduce a new class of distributions that we call multi-tail elliptical distributions. Multi-tail elliptical distributions are an extension of elliptical distributions (see Fang, Kotz and Ng (1987)) and a new subclass of generalized elliptical distributions (see Frahm (2004)). This chapter corresponds to Kring et al. (2007c).

Chapter 5 summarizes the results of the work.

1.5 Appendix

1.5.1 Economical Explanation for Factor Models

In the following theorem we show the relation between factor models in terms of beta pricing models and discount factor models. For a thorough treatment of this topic we refer to Cochrane (2001). With loss of generality we assume $E(F) = 0$, since we can always standardize the factor return through

$$F = \tilde{F} - E(\tilde{F}).$$

Theorem 1. *Let* $F = (F_1, ..., F_p)' \in \mathbf{R}^p$ *be common factors with zero mean and positive definite covariance matrix,* $a > 0$ *and* $b \in \mathbf{R}^p$. *Furthermore, let*

$$m = a + b'F, 1 = E(mX_i) \tag{1.17}$$

be a discount factor model, where X_i *is the* ith *asset return,* $i = 1, ..., d$. *Then there exists a beta pricing model of the form*

$$E(X_i) = \gamma + \sum_{j=1}^{p} \beta_{ij}\lambda_j, i = 1, ..., n, \tag{1.18}$$

where

   (i) $\beta_i = (\beta_{i1}, ..., \beta_{ip})' = E(FF')^{-1}E(X_iF)$ *are the multiple regression coeffi- cients of* X_i *on* F;

   (ii) $\gamma = \frac{1}{E(m)} = \frac{1}{a}$;

   (iii) $\lambda = -\frac{1}{a}E(mF)$.

Conversly, given $\gamma > 0$ *and* $\lambda \in \mathbf{R}^p$ *in a beta pricing model of the form (1.18) and the common factors* $F \in \mathbf{R}^p$ *as above, then there exists* $a > 0$ *and* $b \in \mathbf{R}^p$ *such that (1.17) holds.*

Proof. The idea is to construct a relation between (a, b) and (γ, λ). Due to the model (1.17) we have

$$1 = E(mX_i) = \text{Cov}(m, X_i) - E(m)E(X_i).$$

This is equivalent to

$$
\begin{aligned}
E(X_i) &= \frac{1}{E(m)} - \frac{\text{Cov}(m, X_i)}{E(m)} \\
&= \frac{1}{a} - \frac{\text{Cov}(a + b'F, X_i)}{E(m)} \quad \text{(since } a > 0) \\
&= \frac{1}{a} - \frac{1}{a}\text{Cov}(F', X_i)b \\
&= \frac{1}{a} - \frac{1}{a}E(X_i F')b \quad \text{(since } E(F) = 0).
\end{aligned}
$$

In a beta pricing model the coefficients β_i, $i = 1, ..., n$ are the multiple regression coefficients of X_i on F that are defined by

$$
\beta_i = E(FF')^{-1}E(FX_i). \tag{1.19}
$$

Since the last expression is equivalent to

$$
E(X_i F') = \beta_i' E(FF'),
$$

we obtain

$$
E(X_i) = \frac{1}{a} - \beta_i'\frac{E(FF')b}{a}.
$$

Hence, we define

$$
\begin{aligned}
\gamma &:= \frac{1}{E(m)} = \frac{1}{a} \\
\lambda &:= -\frac{E(FF')b}{a} = -\gamma E(mF). \tag{1.20}
\end{aligned}
$$

Using equations (1.20), we can easily go backwards from the expected return-beta representation to $m = a + b'F$. $\qquad\square$

1.5.2   Spectral Decomposition Theorem

Theorem 2 (Spectral Decomposition Theorem). *Let $A \in \mathbf{R}^{d \times d}$ be a symmetric matrix. Then there is an orthonormal basis of $\mathbf{R}^d$ consisting of eigenvectors of A. In particular, A can be written*

$$
A = \sum_{i=1}^{d} \lambda_i v_i v_i',
$$

where (λ_i, v_i), $i = 1, ..., d$, are eigenvalue-eigenvector pairs of A.

Chapter 2

Estimation of α-Stable Sub-Gaussian Distributions

2.1 Introduction

Classical models in financial risk management and portfolio optimization such as the Markowitz portfolio optimization approach are based on the assumption that risk factor returns and stock returns are normally distributed. Since the seminal work of Mandelbrot (1963) and further investigations by Fama (1965), Chen, and Rachev (1995), McCulloch (1996), and Rachev, and Mittnik (2000) there has been overwhelming empirical evidences that the normal distribution must be rejected. These investigations led to the conclusion that marginal distributions of risk factors and stock returns exhibit skewness and leptokurtosis, i.e., a phenomena that cannot be explained by the normal distribution.

Stable or α-stable distributions have been suggested by the authors above for modeling these peculiarities of financial time series. Beside the fact that α-stable distributions capture these phenomena very well, they have further attractive features which allow them to generalize Gaussian-based financial theory. First, they have the property of stability meaning, that a finite sum of independent and identically distributed (i.i.d.) α-stable distributions is a stable distribution. Second, this class of distribution allows for the generalized Central Limit Theorem: A normalized sum of i.i.d. random variables converges in distribution to an α-stable random vector.

A drawback of stable distributions is that, with a few exceptions, they do not know any analytic expressions for their densities. In the univariate case, this obstacle could be negotiated by numerical approximation based on new computational possibilities. These new possibilities make the α-stable distribution also accessible for practitioners in the financial sector, at least, in the univariate case. The multivariate α-stable case

is even much more complex, allowing for a very rich dependence structure, which is represented by the so-called spectral measure. In general, the spectral measure is very difficult to estimate even in low dimensions. This is certainly one of the main reasons why multivariate α-stable distributions have not been used in many financial applications.

In financial risk management as well as in portfolio optimization, all the models are inherently multivariate as stressed by McNeil, Frey, and Embrechts (2005). The multivariate normal distribution is not appropriate to capture the complex dependence structure between assets, since it does not allow for modeling tail dependencies between the assets and leptokurtosis as well as heavy tails of the marginal return distributions. In many models for market risk management multivariate elliptical distributions, e.g. t-distribution or symmetric generalized hyperbolic distributions, are applied. They model better than the multivariate normal distributions (MNDs) the dependence structure of assets and offer an efficient estimation procedure. In general, elliptical distributions (EDs) are an extension of MNDs since they are also elliptically contoured and characterized by the so-called dispersion matrix. The dispersion matrix equals the variance covariance matrix up to a scaling constants if second moments of the distributions exist, and has a similar interpretation as the variance-covariance matrix for MNDs. In empirical studies [1] it is shown that especially data of multivariate asset returns are roughly elliptically contoured.

In this chapter, we focus on multivariate α-stable sub-Gaussian distributions (MSSDs). In two aspects they are a very natural extension of the MNDs. First, they have the stability property and allow for the generalized Central Limit Theorem, important features making them attractive for financial theory. Second, they belong to the class of EDs implying that any linear combination of an α-stable sub-Gaussian random vector remains α-stable sub-Gaussian and therefore the Markowitz portfolio optimization approach is applicable to them.

We derive two methods to estimate the dispersion matrix of an α-stable sub-Gaussian random vector and analyze them empirically. The first method is based on the covariation and the second one is a moment-type estimator. We will see that the second one outperforms the first one. We conclude the chapter with an empirical analysis of the DAX30 using α-stable sub-Gaussian random vectors.

In Section 2.2 we introduce α-stable distributions and MSSDs, respectively. In Section 2.3 we provide background information about EDs and normal variance mixture distributions, as well as outline their role in modern quantitative market risk management and modeling. In Section 2.4 we present our main theoretical results: we derive two new moments estimators for the dispersion matrix of an MSSD and show

---

[1] For further information, see McNeil, Frey, and Embrechts (2005)

the consistency of the estimators. In Section 2.5 we analyze the estimators empirically using boxplots. In Section 2.6 we fit, as far as we know, for the first time an α-stable sub-Gaussian distribution to the DAX30 and conduct a principal component analysis of the stable dispersion matrix. We compare our results with the normal distribution case. In Section 2.7 we summarize our findings.

2.2 α-stable Distribution: Definitions and Properties

2.2.1 Univariate α-stable Distributions

The applications of α-stable distributions to financial data come from the fact that they generalize the normal (Gaussian) distribution and allow for the heavy tails and skewness, frequently observed in financial data.

There are several ways to define stable distribution.

Definition 2. *Let* $X, X_1, X_2, ..., X_n$ *be i.i.d. random variables. If the equation*

$$X_1 + X_2 + ... + X_n \overset{d}{=} c_n X + d_n$$

holds for all $n \in \mathbf{N}$ *with* $c_n > 0$ *and* $d_n \in \mathbf{R}$, *then we call* X *stable or* α-stable distributed.

The definition justifies the term stable because the sum of i.i.d. random variables has the same distribution as X up to a scale and shift parameter. One can show that the constant c_n in Definition 2 equals $n^{1/\alpha}$.

The next definition represents univariate α-stable distributions in terms of their characteristic functions and determines the parametric family which describes univariate stable distributions.

Definition 3. *A random variable is* α-stable if the characteristic function of X is

$$E(\exp(itX)) = \begin{cases} \exp\left(-\sigma^\alpha |t|^\alpha \left[1 - i\beta \left(\tan \frac{\pi\alpha}{2}\right)(sign\ t)\right] + i\mu t\right) & , \quad \alpha \neq 1 \\ \exp\left(-\sigma |t| \left[1 + i\beta \frac{\pi}{2}(sign\ \ln |t|)\right] + i\mu t\right) & , \quad \alpha = 1. \end{cases}$$

where $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\sigma \in (0, \infty)$ *and* $\mu \in \mathbf{R}$.

The probability densities of α-stable random variables exist and are continuous but, with a few exceptions, they are not known in closed forms. These exceptions are the Gaussian distribution for $\alpha = 2$, the Cauchy distribution for $\alpha = 1$, and the Lévy distribution for $\alpha = 1/2$. (For further information, see Samorodnitsky and Taqqu (1994), where the equivalence of these definitions is shown). The parameter α is called the index of the law, the index of stability or the characteristic exponent. The parameter

β is called skewness of the law. If $\beta = 0$, then the law is symmetric, if $\beta > 0$, it is skewed to the right, if $\beta < 0$, it is skewed to the left. The parameter σ is the scale parameter. Finally, the parameter μ is the location parameter. The parameters α and β determine the shape of the distribution. Since the characteristic function of an α-stable random variable is determined by these four parameters, we denote stable distributions by $S_\alpha(\sigma, \beta, \mu)$. $X \sim S_\alpha(\sigma, \beta, \mu)$, indicating that the random variable X has the stable distribution $S_\alpha(\sigma, \beta, \mu)$. The next definition of an α-stable distribution which is equivalent to the previous definitions is the generalized Central Limit Theorem:

Definition 4. *A random variable X is said to have a stable distribution if it has a domain of attraction, i.e., if there is a sequence of i.i.d. random variables $Y_1, Y_2, ...$ and sequences of positive numbers $(d_n)_{n \in \mathbf{N}}$ and real numbers $(a_n)_{n \in \mathbf{N}}$, such that*

$$\frac{Y_1 + Y_2 + ... + Y_n}{d_n} + a_n \xrightarrow{d} X.$$

The notation $\xrightarrow{d}$ denotes convergence in distribution. If we assume that the sequence of random variables $(Y_i)_{i \in \mathbf{N}}$ has second moments, we obtain the ordinary Central Limit Theorem (CLT). In classical financial theory, the CLT is the theoretical justification for the Gaussian approach, i.e., it is assumed that the price process (S_t) follows a log-normal distribution. If we assume that the log-returns $\log(S_{t_i}/S_{t_{i-1}})$, $i = 1, ..., n$, are i.i.d. and have second moments, we conclude that $\log(S_t)$ is approximately normally distributed. This is a result of the ordinary CLT since the stock price can be written as the sum of independent innovations,i.e.,

$$
\begin{aligned}
\log(S_t) &= \sum_{i=1}^{n} \log\left(S_{t_i}\right) - \log(S_{t_{i-1}}) \\
&= \sum_{i=1}^{n} \log\left(\frac{S_{t_i}}{S_{t_{i-1}}}\right),
\end{aligned}
$$

where $t_n = t$, $t_0 = 0$, $S_0 = 1$ and $t_i - t_{i-1} = 1/n$. If we relax the assumption that stock returns have second moments, we derive from the generalized CLT, that $\log(S_t)$ is approximately α-stable distributed. With respect to the CLT, α-stable distributions are the natural extension of the normal approach. The tail parameter α has an important meaning for α-stable distributions. First, α determines the tail behavior of a stable distribution, i.e.,

$$
\begin{aligned}
\lim_{\lambda \to \infty} \lambda^\alpha P(X > \lambda) &\rightarrow C_+ \\
\lim_{\lambda \to -\infty} \lambda^\alpha P(X < \lambda) &\rightarrow C_-.
\end{aligned}
$$

Second, the parameter α characterizes the distributions in the domain of attraction of a stable law. If X is a random variable with $\lim_{\lambda \to \infty} \lambda^{\alpha} P(|X| > \lambda) = C > 0$ for some $0 < \alpha < 2$, then X is in the domain of attraction of a stable law. Many authors claim that the returns of assets should follow an infinitely divisible law, i.e., for all $n \in \mathbf{N}$ there exists a sequence of i.i.d. random variable $(X_{n,k})_{k=1,...,n}$ satisfying

$$X \overset{d}{=} \sum_{k=1}^{n} X_{n,k}.$$

The property is desirable for models of asset returns in efficient markets since the dynamics of stock prices are caused from continuously arising but independent information. From Definition 4, it is obvious that α-stable distribution are infinitely divisible.

The next lemma is useful for deriving an estimator for the scale parameter σ.

Lemma 1. *Let $X \sim S_{\alpha}(\sigma, \beta, \mu)$, $1 < \alpha < 2$ and $\beta = 0$. Then for any $0 < p < \alpha$ there exists a constant $c_{\alpha, \beta}(p)$ such that:*

$$E(|X - \mu|^p)^{1/p} = c_{\alpha, \beta}(p)\sigma$$

where $c_{\alpha, \beta}(p) = (E|X_0|^p)^{1/p}$, $X_0 \sim S_{\alpha}(1, \beta, 0)$.

Proof. See Samorodnitsky and Taqqu (1994). □

To get a first feeling for the sort of data we are dealing with, we display in Figure 2.1 the kernel density plots of the empirical returns, the Gaussian fit and the α-stable fit of some representative stocks. We can clearly discern the individual areas in the plot where the normal fit causes problems. It is around the mode where the empirical peak is too high to be captured by the Gaussian parameters. Moreover, in the mediocre parts of the tails, the empirical distribution attributes less weight than the Gaussian distribution. And finally, the tails are underestimated, again. In contrast to the Gaussian, the stable distribution appears to account for all these features of the empirical distribution quite well.

Another means of presenting the aptitude of the stable class to represent stock returns is the quantile plot. In Figure 2.2, we match the empirical stock return percentiles of Adidas AG with simulated percentiles for the normal and stable distributions, for the respective estimated parameter tuples. The stable distribution is liable to produce almost absurd extreme values compared to the empirical data. Hence, we need to discard the most extreme quantile pairs. However, the overall position of the line of the joint empirical-stable percentiles with respect to the interquartile line appears quite
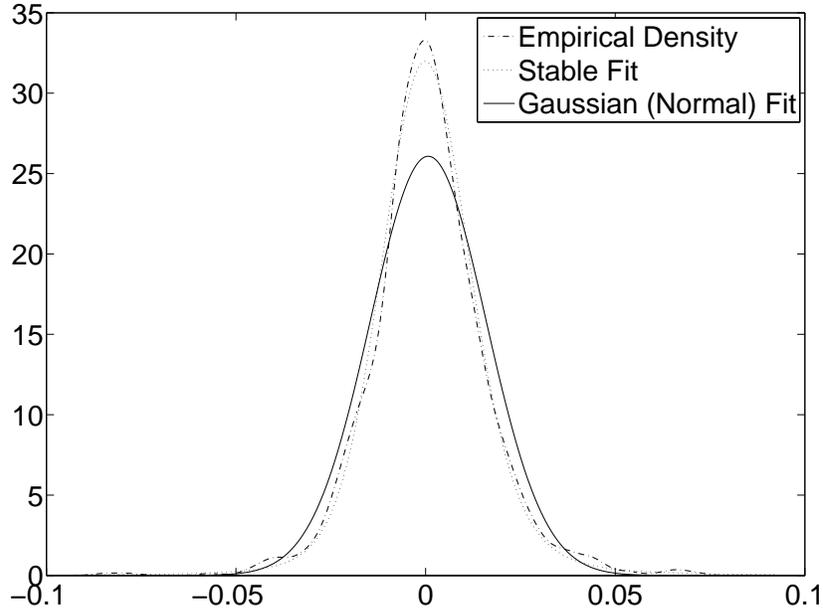
Figure 2.1: Kernel density plots of Adidas AG: empirical, normal, and stable fits.

convincingly in favor of the stable distribution. [2]

2.2.2 Multivariate α-Stable Distributions

Multivariate stable distributions are the distributions of stable random vectors. They are defined by simply extending the definition of stable random variables to $\mathbf{R}^d$. As in the univariate case, multivariate Gaussian distribution is a particular case of multivariate stable distributions. Any linear combination of stable random vectors is a stable random variate. This is an important property in terms of portfolio modeling. Multivariate stable cumulative distribution functions or density functions are usually not known in closed form and therefore, one works with their characteristic functions. The representation of these characteristic functions include a finite measure on the unit sphere, the so-called *spectral measure*. This measure describes the dependence structure of the stable random vector. In general, stable random vectors are difficult to use for financial modeling, because the spectral measure is difficult to estimate even in low dimensions. For stable financial model building, one has to focus on certain subclasses of stable random vectors where the spectral measure has an easier representation. Such a subclass is the multivariate α-stable *sub-Gaussian* law. They are obtained by multiplying a *Gaussian* vector by $W^{1/2}$ where W is a stable random variable totally skewed

---

[2]In Figure 2.2 we remove the two most extreme points in the upper and lower tails, respectively.
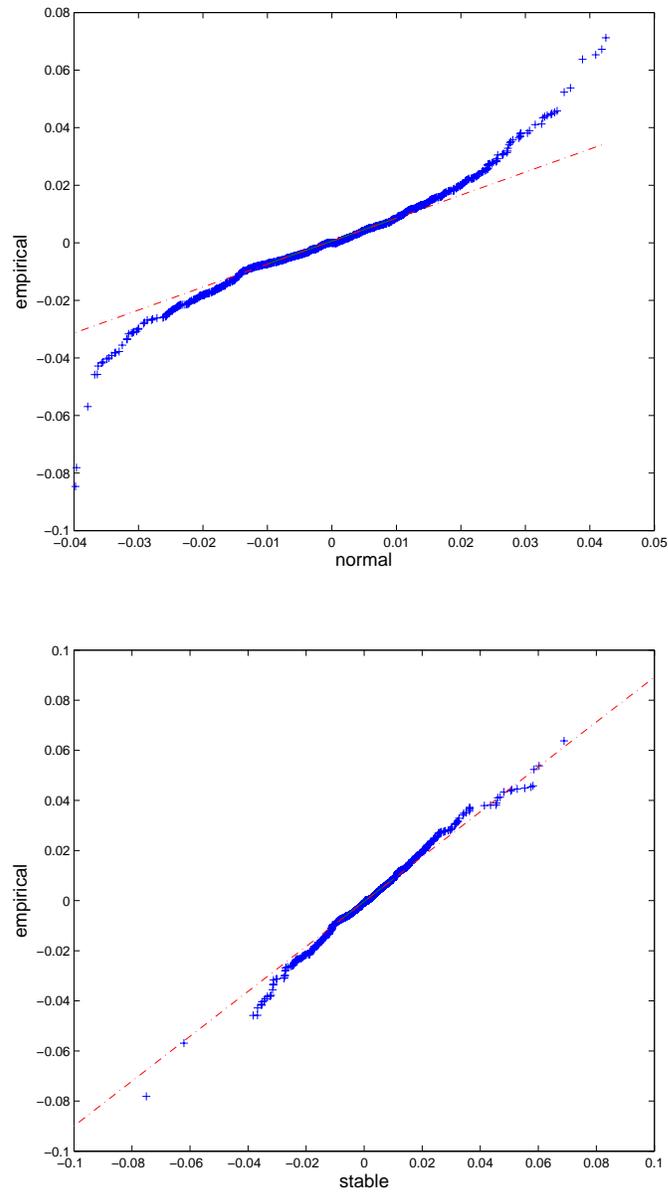
Figure 2.2: Adidas AG quantile plots of empirical return percentiles vs normal (top) and stable (bottom) fits.

to the right. Stable sub-Gaussian distributions inherit their dependence structure from the underlying Gaussian vector. In the next section we will see that the distribution of multivariate stable sub-Gaussian random vectors belongs to the class of elliptical distributions. The definition of stability in $\mathbf{R}^d$ is analogous to that in $\mathbf{R}$.

Definition 5. *A random vector $X = (X_1, ..., X_d)$ is said to be a* stable random vector *in $\mathbf{R}^d$ if for any positive numbers A and B there is a positive number C and a vector $D \in \mathbf{R}^d$ such that*

$$AX^{(1)} + BX^{(2)} \stackrel{d}{=} CX + D$$

where $X^{(1)}$ and $X^{(2)}$ are independent copies of X.

Note, that an α-stable random vector X is called symmetric stable if X satisfies

$$P(X \in A) = P(-X \in A)$$

for all Borel-sets A in $\mathbf{R}^d$.

Theorem 3. *Let X be a stable (respectively symmetric stable) vector in $\mathbf{R}^d$. Then there is a constant $\alpha \in (0, 2]$ such that in Definition 5, $C = (A^\alpha + B^\alpha)^{1/\alpha}$. Moreover, any linear combination of the components of X of the type $Y = \sum_{i=1}^{d} b_k X_k = b'X$ is an α-stable (respectively symmetric stable) random variable.*

Proof. A proof is given in Samorodnitsky and Taqqu (1994). $\square$

The parameter α in Theorem 3 is called the index of stability. It determines the tail behavior of a stable random vector, i.e., the α-stable random vector is regularly varying with tail index α[3]. For portfolio analysis and risk management, it is very important that stable random vectors are closed under linear combinations of the components due to Theorem 3. In the next section we will see that elliptically distributed random vectors have this desirable feature as well.

The next theorem determines α-stable random vectors in terms of the characteristic function. Since there is a lack of formulas for stable densities and distribution functions, the characteristic function is the main device to fit stable random vectors to data.

Theorem 4. *The random vector $X = (X_1, ..., X_d)$ is an α-stable random vector in $\mathbf{R}^d$ iff there exists an unique finite measure Γ on the unit sphere $\mathcal{S}^{d-1}$, the so-called* spectral measure, *and an unique vector $\mu \in \mathbf{R}^d$ such that:*

---

[3]For further information about regularly varying random vectors, see Resnick (1987).

(i) If $\alpha \neq 1$,

$$E(e^{it'X}) \ = \ \exp\{-\int_{\mathcal{S}^{d-1}} |(t,s)|^{\alpha}(1 - i\operatorname{sign}((t,s))\tan\frac{\pi\alpha}{2})\Gamma(ds) + i(t,\mu)\}$$

(ii) If $\alpha = 1$,

$$E(e^{it'X}) \ = \ \exp\{-\int_{\mathcal{S}^{d-1}} |(t,s)|(1 + i\frac{2}{\pi}\operatorname{sign}((t,s))\ln|(t,s)|)\Gamma(ds) + i(t,\mu)\}$$

In contrast to the univariate case, stable random vectors have not been applied frequently in financial modeling. The reason is that the spectral measure, as a measure on the unit sphere $\mathcal{S}^{d-1}$, is extremely difficult to estimate even in low dimensions. (For further information see Rachev, and Mittnik (2000) and Nolan, Panorska, and McCulloch (2001).)

Another way to describe stable random vectors is in terms of linear projections. We know from Theorem 3 that any linear combination

$$(b, X) = \sum_{i=1}^{d} b_i X_i$$

has an α-stable distribution $S_{\alpha}(\sigma(b), \beta(b), \mu(b))$. By using Theorem 4 we obtain for the parameters $\sigma(b), \beta(b)$ and $\mu(b)$

$$\sigma(b) \ = \ \left(\int_{\mathcal{S}^{d-1}} |(b,s)|^{\alpha}\Gamma(ds)\right)^{1/\alpha},$$

$$\beta(b) \ = \ \frac{\int_{\mathcal{S}^{d-1}} |(b,s)|^{\alpha}\operatorname{sign}(b,s)\Gamma(ds)}{\int_{\mathcal{S}^{d-1}} |(b,s)|^{\alpha}\Gamma(ds)}$$

and

$$\mu(b) = \begin{cases} (b,\mu) & \text{if} \quad \alpha \neq 1 \\ (b,\mu) - \frac{2}{\pi}\int_{\mathcal{S}^{d-1}}(b,s)\ln|(b,s)|\Gamma(ds) & \text{if} \quad \alpha = 1. \end{cases}$$

The parameters $\sigma(b), \beta(b)$, and $\mu(b)$ are also called the projection parameters and $\sigma(.)$, $\beta(.)$ and $\mu(.)$ are called the projection parameter functions. If one knows the values of the projection functions for several directions, one can reconstruct approximatively the dependence structure of an α-stable random vector by estimating the spectral measure. Because of the complexity of this measure, the method is still not very efficient. But for specific subclasses of stable random vectors where the spectral measure has a much simpler form, we can use this technique to fit stable random vectors to data.

Another quantity for characterizing the dependence structure between two stable random vectors is the *covariation*.

Definition 6. *Let X_1 and X_2 be jointly symmetric stable random variables with $\alpha > 1$ and let Γ be the spectral measure of the random vector $(X_1, X_2)'$. The* covariation *of X_1 on X_2 is the real number*

$$[X_1, X_2]_\alpha = \int_{\mathcal{S}_1} s_1 s_2^{<\alpha-1>} \Gamma(ds), \qquad (2.1)$$

where the signed power $a^{<p>}$ equals

$$a^{<p>} = |a|^p \operatorname{sign} a.$$

The covariance between two normal random variables X and Y can be interpreted as the inner product of the space $L_2(\Omega, \mathcal{A}, \mathbf{P})$. The covariation is the analogue of two α-stable random variables X and Y in the space $L_\alpha(\Omega, \mathcal{A}, \mathbf{P})$. Unfortunately, $L_\alpha(\Omega, \mathcal{A}, \mathbf{P})$ is not a Hilbert space and this is why it lacks some of the desirable and strong properties of the covariance. It follows immediately from the definition that the covariation is linear in the first argument. Unfortunately, this statement is not true for the second argument. In the case of $\alpha = 2$, the covariation equals the covariance.

Proposition 1. Let (X, Y) be jointly symmetric stable random vectors with $\alpha > 1$. Then for all $1 < p < \alpha$,

$$\frac{EXY^{<p-1>}}{E|Y|^p} = \frac{[X, Y]_\alpha}{||Y||_\alpha^\alpha},$$

where $||Y||_\alpha$ denotes the scale parameter of Y.

Proof. For the proof, see Samorodnitsky and Taqqu (1994). $\qquad\qquad\square$

In particular, we apply Proposition 1 in Section 2.4.1 in order to derive an estimator for the dispersion matrix of an α-stable sub-Gaussian distribution.

2.2.3 α-Stable Sub-Gaussian Random Vectors

In general, as pointed out in the last section, α-stable random vectors have a complex dependence structure defined by the spectral measure. Since this measure is very difficult to estimate even in low dimensions, we have to retract to certain subclasses, where the spectral measure becomes simpler. One of these special classes is the multivariate α-stable sub-Gaussian distribution.

Definition 7. *Let Z be a zero mean Gaussian random vector with variance covariance matrix Σ and $W \sim S_{\alpha/2}((\cos\frac{\pi\alpha}{4})^{2/\alpha}, 1, 0)$ a totally skewed stable random variable*

independent of Z. The random vector

$$X = \mu + \sqrt{W} Z$$

is said to be a sub-Gaussian α-stable random vector. The distribution of X is called multivariate α-stable sub-Gaussian distribution.

An α-stable sub-Gaussian random vector inherits its dependence structure from the underlying Gaussian random vector. The matrix Σ is also called the dispersion matrix. The following theorem and proposition show properties of α-stable sub-Gaussian random vectors. We need these properties to derive estimators for the dispersion matrix.

Theorem 5. *The sub-Gaussian α-stable random vector X with location parameter $\mu \in \mathbf{R}^d$ has the characteristic function*

$$E(e^{it'X}) \quad = \quad e^{it'\mu} e^{-(\frac{1}{2}t'\Sigma t)^{\alpha/2}},$$

where $\Sigma_{ij} = EZ_iZ_j$, $i,j = 1,...,d$ are the covariances of the underlying Gaussian random vector $(Z_1,...,Z_d)'$.

For α-stable sub-Gaussian random vectors, we do not need the spectral measure in the characteristic functions. This fact simplifies the calculation of the projection functions.

Proposition 2. Let $X \in \mathbf{R}^d$ be an α-stable sub-Gaussian random vector with location parameter $\mu \in \mathbf{R}^d$ and dispersion matrix Σ. Then, for all $a \in \mathbf{R}^d$, we have $a'X \sim S_\alpha(\sigma(a), \beta(a), \mu(a))$, where

(i) $\sigma(a) = (\frac{1}{2}a'\Sigma a)^{1/2}$

(ii) $\beta(a) = 0$

(iii) $\mu(a) = a'\mu.$

Proof. It is well known that the distribution of dX is determined by its characteristic function.

$$
\begin{aligned}
E(\exp(it(a'X))) \quad &= \quad E(\exp(i(ta')X))) \\
&= \quad \exp(ita'\mu)\exp(-|\frac{1}{2}(ta)'\Sigma(ta)|^{\alpha/2}|) \\
&= \quad \exp(ita'\mu)\exp(-|\frac{1}{2}t^2 a'\Sigma a|^{\alpha/2}|) \\
&= \quad \exp(-|t|^\alpha|(\frac{1}{2}a'\Sigma a)^{\frac{1}{2}}|^\alpha + ita'\mu)
\end{aligned}
$$

If we choose $\sigma(a) = (\frac{1}{2}a'\Sigma a)^{1/2}$, $\beta(a) = 0$ and $\mu(a) = a'\mu$, then for all $t \in \mathbf{R}$ we have

$$E(\exp(it(a'X))) = \exp\left(-\sigma(a)^\alpha |t|^\alpha \left[1 - i\beta(a)\left(\tan\frac{\pi\alpha}{2}\right)(\text{sign } t)\right] + i\mu(a)t\right).$$

$\square$

In particular, we can calculate the entries of the dispersion matrix directly.

Corollary 1. *Let $X = (X_1, ..., X_n)'$ be an α-stable sub-Gaussian random vector with dispersion matrix Σ. Then we obtain*

(i) $\sigma_{ii} = 2\sigma(e_i)^2$

(ii) $\sigma_{ij} = \frac{\sigma^2(e_i+e_j) - \sigma^2(e_i-e_j)}{2}$.

Since α-stable sub-Gaussian random vectors inherit their dependence structure of the underlying Gaussian vector, we can interpret σ_{ii} as the quasi-variance of the component X_i and σ_{ij} as the quasi-covariance between X_i and X_j.

Proof. It follows from Proposition 2 that $\sigma(e_i) = \frac{1}{2}\sigma_{ii}^2$. Furthermore, if we set $a = e_i + e_j$ with $i \neq j$, we yield $\sigma(e_i + e_j) = (\frac{1}{2}(\sigma_{ii} + 2\sigma_{ij} + \sigma_{jj}))^{1/2}$ and for $b = e_i - e_j$, we obtain $\sigma(e_i - e_j) = (\frac{1}{2}(\sigma_{ii} - 2\sigma_{ij} + \sigma_{jj}))^{1/2}$. Hence, we have

$$\sigma_{ij} = \frac{\sigma^2(e_i + e_j) - \sigma^2(e_i - e_j)}{2}.$$

$\square$

Proposition 3. *Let $X = (X_1, ..., X_n)'$ be a zero mean α-stable sub-Gaussian random vector with dispersion matrix Σ. Then it follows*

$$[X_i, X_j]_\alpha = 2^{-\alpha/2}\sigma_{ij}\sigma_{jj}^{(\alpha-2)/2}.$$

Proof. For a proof see Samorodnitsky and Taqqu (1994). $\square$

2.3   Elliptical Distributions

Many important properties of α-stable sub-Gaussian distributions with respect to risk management, portfolio optimization, and principal component analysis can be understood very well, if we regard them as elliptical or normal variance mixture distributions. Elliptical distributions are a natural extension of the normal distribution which is a special case of this class. They obtain their name because of the fact that, their

densities are constant on ellipsoids. Furthermore, they constitute a kind of ideal environment for standard risk management, see Embrechts, McNeil, and Strautmann (1999). First, correlation and covariance have a very similar interpretation as in the Gaussian world and describe the dependence structure of risk factors. Second, the Markowitz optimization approach is applicable. Third, value-at-risk is a coherent risk measure. Fourth, they are closed under linear combinations, an important property in terms for portfolio optimization. And finally, in the elliptical world minimizing risk of a portfolio with respect to any coherent risk measures leads to the same optimal portfolio.

Empirical investigations have shown that multivariate return data for groups of similar assets often look roughly elliptical and in market risk management the elliptical hypothesis can be justified. Elliptical distributions cannot be applied in credit risk or operational risk, since hypothesis of elliptical risk factors are found to be rejected.

2.3.1   Elliptical Distributions and their Basic Properties

Definition 8. *A random vector $X = (X_1, ..., X_d)'$ has*

(i) a spherical distribution iff, for every orthogonal matrix $U \in \mathbf{R}^{d \times d}$,

$$UX \overset{d}{=} X.$$

(ii) an elliptical distribution if

$$X \overset{d}{=} \mu + AY,$$

where Y is a spherical random variable and $A \in \mathbf{R}^{d \times K}$ and $\mu \in \mathbf{R}^d$ are a matrix and a vector of constants, respectively.

Elliptical distributions are obtained by multivariate affine transformations of spherical distributions. Figure 2.3.1 (a) and (b) depict a bivariate scatterplot of BMW versus Daimler Chrysler and Commerzbank versus Deutsche Bank log-returns. Both scatterplots are roughly elliptical contoured.

Theorem 6. *The following statements are equivalent*

(i) X is spherical.

(ii) There exists a function ψ of a scalar variable such that, for all $t \in \mathbf{R}^d$,

$$\phi_X(t) = E(e^{it'X}) = \psi(t't) = \psi(t_1^2 + ... + t_d^2).$$

(a)   (b)
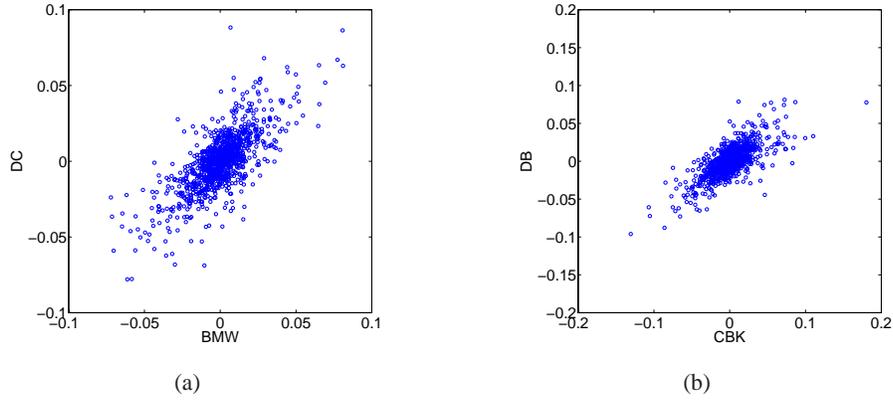
Figure 2.3: Bivariate scatterplot of BMW versus DaimlerChrysler and Commerzbank versus Deutsche Bank. Depicted are daily log-returns from May 6, 2002 through March 31, 2006.

(iii) For all $a \in \mathbf{R}^d$, we have

$$a'X \stackrel{d}{=} ||a||X_1$$

(iv) X can be represented as

$$X \stackrel{d}{=} RS$$

where S is uniformly distributed on $\mathcal{S}^{d-1} = \{x \in \mathbf{R}^d : x'x = 1\}$ and $R \geq 0$ is a radial random variable independent of S.

Proof. See McNeil, Frey, and Embrechts (2005) $\square$

ψ is called the characteristic generator of the spherical distribution and we use the notation $X \in S_d(\psi)$.

Corollary 2. *Let X be a d-dimensional elliptical distribution with $X \stackrel{d}{=} \mu + AY$, where Y is spherical and has the characteristic generator ψ. Then, the characteristic function of X is given by*

$$\phi_X(t) := E(e^{it'X}) = e^{it'\mu}\psi(t'\Sigma t),$$

where $\Sigma = AA'$.

Furthermore, X can be represented by

$$X = \mu + RAS,$$

where S is the uniform distribution on $\mathcal{S}^{d-1}$ and $R \geq 0$ is a radial random variable.

Proof. We notice that

$$
\begin{aligned}
\phi_X(t) &= E(e^{it'X}) = E(e^{it'(\mu+AY)}) = e^{it'\mu}E(e^{i(A't)'Y}) = e^{it'\mu}\psi((A't)'(A't)) \\
&= e^{it'\mu}\psi(t'AA't)
\end{aligned}
$$

$\square$

Since the characteristic function of a random variate determines the distribution, we denote an elliptical distribution by

$$
X \sim E_d(\boldsymbol{\mu}, \Sigma, \psi).
$$

Because of

$$
\mu + RAS = \mu + cR\frac{A}{c}S,
$$

the representation of the elliptical distribution in equation (2.2) is not unique. We call the vector μ the location parameter and Σ the dispersion matrix of an elliptical distribution, since first and second moments of elliptical distributions do not necessarily exist. But if they exist, the location parameter equals the mean and the dispersion matrix equals the covariance matrix up to a scale parameter. In order to have uniqueness for the dispersion matrix, we demand $\det(\Sigma) = 1$.

If we take any affine linear combination of an elliptical random vector, then, this combination remains elliptical with the same characteristic generator ψ. Let $X \sim E_d(\mu, \Sigma, \psi)$, then it can be shown with similar arguments as in Corollary 2 that

$$
BX + b \;\; \sim \;\; E_k(B\mu + b, B\Sigma B', \psi)
$$

where $B \in \mathbf{R}^{k\times d}$ and $b \in \mathbf{R}^d$.

Let X be an elliptical distribution. Then the density $f(x)$, $x \in \mathbf{R}^d$, exists and is a function of the quadratic form

$$
f(x) = \det(\Sigma)^{-1/2}g(Q) \text{ with } Q := (x - \mu)'\Sigma^{-1}(x - \mu).
$$

g is the density of the spherical distribution Y in Definition 8. We call g the density generator of X. As a consequence, since Y has an unimodal density, so is the density of X and clearly, the joint density f is constant on hyperspheres $H_c = \{x \in \mathbf{R}^d : Q(x) = c\}$, $c > 0$. These hyperspheres H_c are elliptically contoured.

Example 1. *An α-stable sub-Gaussian random vector is an elliptical random vector. The random vector $\sqrt{W}Z$ is spherical, where $W \sim S_\alpha((\cos\frac{\pi\alpha}{4})^{2/\alpha}, 1, 0)$ and $Z \sim$*

$N(0, 1)$ *because of*

$$\sqrt{W}Z \stackrel{d}{=} U\sqrt{W}Z$$

for any orthogonal matrix. The equation is true, since Z is rotationally symmetric. Hence any linear combination of $\sqrt{W}Z$ is an elliptical random vector. The characteristic function of an α-stable sub-Gaussian random vector is given by

$$E(e^{it'X}) \quad = \quad e^{it'\mu}e^{-(\frac{1}{2}t'\Sigma t)^{\alpha/2}}$$

due to Theorem 5. Thus, the characteristic generator of an α-stable sub-Gaussian random vector equals

$$\psi_{sub}(s, \alpha) = e^{-(\frac{1}{2}s)^{2/\alpha}}.$$

Using the characteristic generator, we can derive directly that an α-stable sub-Gaussian random vector is infinitely divisible, since we have

$$\psi_{sub}(s, \alpha) \quad = \quad e^{-(\frac{1}{2}s)^{\alpha/2}} = \left(e^{-\left(\frac{1}{2}\frac{s}{n^{2/\alpha}}\right)^{\alpha/2}} \right)^n$$

$$= \quad \left(\psi_{sub}\left(\frac{s}{n^{2/\alpha}}, \alpha\right) \right)^n.$$

2.3.2 Normal Variance Mixture Distributions

Normal variance mixture distributions are a subclass of elliptical distributions. We will see that they inherit their dependence structure from the underlying Gaussian random vector. Important distributions in risk management such as the multivariate t-, generalized hyperbolic, or α-stable sub-Gaussian distribution belong to this class of distributions.

Definition 9. *The random vector X is said to have a (multivariate) normal variance mixture distribution (NVMD) if*

$$X = \mu + W^{1/2}AZ$$

where

(i) $Z \sim N_d(0, I_d)$;

(ii) $W \geq 0$ *is a non-negative, scalar-valued random variable which is independent of G, and*

(iii) $A \in \mathbf{R}^{d \times d}$ *and $\mu \in \mathbf{R}^d$ are a matrix of constants, respectively.*

We call a random variable X with NVMD a normal variance mixture (NVM). We observe that $X_w = (X|W = w) \sim N_d(\mu, w\Sigma)$, where $\Sigma = AA'$. We can interpret the distribution of X as a composite distribution. According to the law of W, we take normal random vectors X_w with mean zero and covariance matrix $w\Sigma$ randomly. In the context of modeling asset returns or risk factor returns with normal variance mixtures, the mixing variable W can be thought of as a shock that arises from new information and influences the volatility of all stocks.

Since $U\sqrt{W}Z \stackrel{d}{=} \sqrt{W}Z$ for all $U \in O(d)$ every normal variance mixture distribution is an elliptical distribution. The distribution F of X is called the mixing law. Normal variance mixture are closed under affine linear combinations, since they are elliptical. This can also be seen directly by

$$
\begin{aligned}
BX + \mu_1 & \stackrel{d}{=} B(\sqrt{W}AZ + \mu_0) + \mu = \sqrt{W}BAZ + (B\mu_0 + \mu_1) \\
& = \sqrt{W}\tilde{A}Z + \tilde{\mu}.
\end{aligned}
$$

This property makes NVMDs and, in particular, MSSDs applicable to portfolio theory. The class of NVMD has the advantage that structural information about the mixing law W can be transferred to the mixture law. This is true, for example, for the property of infinite divisibility. If the mixing law is infinitely divisible, then so is the mixture law. (For further information see Bingham, Kiesel and Schmidt (2003).) It is obvious from the definition that an α-stable sub-Gaussian random vector is also a normal variance mixture with mixing law $W \sim S_\alpha((\cos \frac{\pi\alpha}{4})^{2/\alpha}, 1, 0)$.

2.3.3   Market Risk Management with Elliptical Distributions

In this section, we discuss the properties of elliptical distributions in terms of market risk management and portfolio optimization. In risk management, one is mainly interested in modeling the extreme losses which can occur. From empirical investigations, we know that an extreme loss in one asset very often occurs with high losses in many other assets. We show that this market behavior cannot be modeled by the normal distribution but, with certain elliptical distributions, e.g. α-stable sub-Gaussian distribution, we can capture this behavior.

The Markowitz's portfolio optimization approach which is originally based on the normal assumption can be extended to the class of elliptical distributions. Also, statistical dimensionality reduction methods such as the principal component analysis are applicable to them. But one must be careful, in contrast to the normal distribution, these principal components are not independent.

Let F be the distribution function of the random variable X, then we call

$$F^{\leftarrow}(\alpha) = \inf\{x \in \mathbf{R} : F(x) \geq \alpha\}$$

the quantile function. $F^{\leftarrow}$ is also the called generalized inverse, since we have

$$F(F^{\leftarrow}(\alpha)) = \alpha,$$

for any df F.

Definition 10. *Let X_1 and X_2 be random variables with dfs F_1 and F_2. The coefficient of the upper tail dependence of X_1 and X_2 is*

$$\lambda_u := \lambda_u(X_1, X_2) := \lim_{q \to 1^-} P(X_2 > F_2^{\leftarrow}(q)|X_1 > F_1^{\leftarrow}(q)), \qquad (2.2)$$

provided a limit $\lambda_u \in [0,1]$ exists. If $\lambda_u \in (0,1]$, then X_1 and X_2 are said to show upper tail dependence; if $\lambda_u = 0$, they are asymptotically independent in the upper tail. Analogously, the coefficient of the lower tail dependence is

$$\lambda_l = \lambda_l(X_1, X_2) = \lim_{q \to 0^+} P(X_2 \leq F^{\leftarrow}(q)|X_1 \leq F_1^{\leftarrow}(q)), \qquad (2.3)$$

provided a limit $\lambda_l \in [0,1]$ exists.

For a better understanding of tail dependence we introduce the concept of copulas.

Definition 11. *A d-dimensional copula is a distribution function on $[0,1]^d$.*

It is easy to show that for $U \sim U(0,1)$, we have $P(F^{\leftarrow}(U) \leq x) = F(x)$ and if the random variable Y has a continuous df G, then $G(Y) \sim U(0,1)$. The concept of copulas gained its importance because of Sklar's Theorem.

Theorem 7. *Let F be a joint distribution function with margins $F_1, ..., F_d$. Then, there exists a copula $C : [0,1]^d \to [0,1]$ such that for all $x_1, ..., x_d$ in $\overline{\mathbf{R}} = [\infty, \infty]$,*

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)). \qquad (2.4)$$

If the margins are continuous, then C is unique; otherwise C is uniquely determined on $F_1(\overline{R}) \times F_2(\overline{R}) \times .... \times F_d(\overline{R})$. Conversely, if C is a copula and $F_1, ..., F_d$ are univariate distribution functions, the function F defined in (2.4) is a joint distribution function with margins $F_1, ..., F_d$.

This fundamental theorem in the field of copulas, shows that any multivariate distribution F can be decomposed in a copula C and the marginal distributions of F. Vice

versa, we can use a copula C and univariate dfs to construct a multivariate distribution function.

With this short excursion in the theory of copulas we obtain a simpler expression for the upper and the lower tail dependencies, i.e.,

$$
\begin{aligned}
\lambda_l &= \lim_{q \to 0^+} \frac{P(X_2 \leq F^{\leftarrow}(q), X_1 \leq F_1^{\leftarrow}(q))}{P(X_1 \leq F_1^{\leftarrow}(q))} \\
&= \lim_{q \to 0^+} \frac{C(q, q)}{q}.
\end{aligned}
$$

Elliptical distributions are radially symmetric, i.e., $\mu - X \overset{d}{=} \mu + X$, hence the coefficient of lower tail dependence λ_l equals the coefficient of upper tail dependence λ_u. We denote with λ the coefficient of tail dependence.

We call a measurable function $f : \mathbf{R}_+ \to \mathbf{R}_+$ *regularly varying* (at ∞) with index $\alpha \in \mathbf{R}$ if, for any $t > 0$, $\lim_{x \to \infty} f(tx)/f(x) = t^{\alpha}$. It is now important to notice that regularly varying functions with index $\alpha \in \mathbf{R}$ behave asymptotically like a power function. An elliptically distributed random vector $X = RAU$ is said to be regularly varying with tail index α, if the function $f(x) = P(R \geq x)$ is regularly varying with tail index α. (see Resnick (1987).) The following theorem shows the relation between the tail dependence coefficient and the tail index of elliptical distributions.

Theorem 8. *Let $X \sim E_d(\mu, \Sigma, \psi)$ be regularly varying with tail index $\alpha > 0$ and Σ a positive definite dispersion matrix. Then, every pair of components of X, say X_i and X_j, is tail dependent and the coefficient of tail dependence corresponds to*

$$
\lambda(X_i, X_j; \alpha, \rho_{ij}) = \frac{\int_0^{f(\rho_{ij})} \frac{s^{\alpha}}{\sqrt{1-s^2}} ds}{\int_0^1 \frac{s^{\alpha}}{\sqrt{1-s^2}} ds} \tag{2.5}
$$

where $f(\rho_{ij}) = \sqrt{\frac{1+\rho_{ij}}{2}}$ and $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$.

Proof. See Schmidt (2002). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

It is not difficult to show that an α-stable sub-Gaussian distribution is regularly varying with tail index α. The coefficient of tail dependence between two components, say X_i and X_j, is determined by equation (2.5) in Theorem 8. In the next example, we demonstrate that the coefficient of tail dependence of a normal distribution is zero.

Example 2. *Let (X_1, X_2) be a bivariate normal random vector with correlation $\rho \in (-1, 1)$ and standard normal marginals. Let C_ρ be the corresponding Gaussian copula*

due to Sklar's theorem, then, by the L'Hôpital rule,

$$
\begin{aligned}
\lambda &= \lim_{q\to 0^+} \frac{C_\rho(q,q)}{q} \overset{l'H}{=} \lim_{q\to 0^+} \frac{dC_\rho(q,q)}{dq} = \lim_{q\to 0^+}\lim_{h\to 0^+} \frac{C_\rho(q+h,q+h) - C_\rho(q,q)}{h} \\
&= \lim_{q\to 0^+}\lim_{h\to 0} \frac{C_\rho(q+h,q+h) - C_\rho(q+h,q) + C_\rho(q+h,q) - C_\rho(q,q)}{h} \\
&= \lim_{q\to 0^+}\lim_{h\to 0} \frac{P(U_1 \le q+h, q \le U_2 \le q+h))}{P(q \le U_2 \le q+h)} \\
&+ \lim_{q\to 0^+}\lim_{h\to 0} \frac{P(q \le U_1 \le q+h, U_2 \le q)}{P(q \le U_1 \le q+h)} \\
&= \lim_{q\to 0^+} P(U_2 \le q | U_1 = q) + \lim_{q\to 0^+} P(U_1 \le q | U_2 = q) \\
&= 2 \lim_{q\to 0^+} P(U_2 \le q | U_1 = q) \\
&= 2 \lim_{q\to 0^+} P(\Phi^{-1}(U_2) \le \Phi^{-1}(q) | \Phi^{-1}(U_1) = \Phi^{-1}(q)) \\
&= 2 \lim_{x\to -\infty} P(X_2 \le x | X_1 = x)
\end{aligned}
$$

Since we have $X_2 | X_1 = x \sim N(\rho x, 1 - \rho^2)$, we obtain

$$
\lambda = 2 \lim_{x\to -\infty} \Phi(x\sqrt{1-\rho}/\sqrt{1+\rho}) = 0. \tag{2.6}
$$

Equation (2.6) shows that beside the fact that a normal distribution is not heavy tailed the components are asymptotically independent. This, again, is a contradiction to empirical investigations of market behavior. Especially, in extreme market situations, when a financial market declines in value, market participants tend to behave homogeneously, i.e. they leave the market and sell their assets. This behavior causes losses in many assets simultaneously. This phenomenon can only be captured by distributions which are asymptotically dependent.

Markowitz (1952) optimizes the risk and return behavior of a portfolio based on the expected returns and the covariances of the returns in the considered asset universe. The risk of a portfolio consisting of these assets is measured by the variance of the portfolio return. In addition, he assumes that the asset returns follow a multivariate normal distribution with mean μ and covariance Σ. This approach leads to the following optimization problem

$$
\min_{w\in \mathbf{R}^d} w'\Sigma w,
$$

subject to

$$
\begin{aligned}
w'\mu &= \mu_p \\
w'\underline{1} &= 1.
\end{aligned}
$$

This approach can be extended in two ways. First, we can replace the assumption of normally distributed asset returns by elliptically distributed asset returns and second, instead of using the variance as the risk measure, we can apply any positive-homogeneous, translation-invariant measure of risk to rank risk or to determine the optimal risk-minimizing portfolio. In general, due to the work of Artzner et al. (1999), a risk measure is a real-valued function $\varrho : \mathcal{M} \to \mathbf{R}$, where $\mathcal{M} \subset L^0(\Omega, \mathcal{F}, P)$ is a convex cone. $L^0(\Omega, \mathcal{F}, P)$ is the set of all almost surely finite random variables. The risk measure ϱ is *translation invariant* if for all $L \in \mathcal{M}$ and every $l \in \mathbf{R}$, we have $\varrho(L + l) = \varrho(L) + l$. It is *positive-homogeneous* if for all $\lambda > 0$, we have $\varrho(\lambda L) = \lambda \varrho(L)$. Note, that value-at-risk (VaR) as well as conditional value-at-risk (CVaR) fulfill these two properties.

Theorem 9. *Let the random vector of asset returns X be $E_d(\mu, \Sigma, \psi)$. We denote by $\mathcal{W} = \{w \in \mathbf{R}^d : \sum_{i=1}^d w_i = 1\}$ the set of portfolio weights. Assume that the current value of the portfolio is V and let $L(w) = V \sum_{i=1}^d w_i X_i$ be the (linearized) portfolio loss. Let ϱ be a real-valued risk measure depending only on the distribution of a risk. Suppose ϱ is positive homogeneous and translation invariant and let $\mathcal{Y} = \{w \in \mathcal{W} : -w'\mu = m\}$ be the subset of portfolios giving expected return m. Then, $\mathrm{argmin}_{w \in \mathcal{Y}} \varrho(L(w)) = \mathrm{argmin}_{w \in \mathcal{Y}} w'\Sigma w$.*

Proof. See McNeil, Frey, and Embrechts (2005). $\square$

The last theorem stresses that the dispersion matrix contains all the information for the management of risk. In particular, the tail index of an elliptical random vector has no influence on optimizing risk. Of course, the index has an impact on the value of the particular risk measure like VaR or CVaR, but not on the weights of the optimal portfolio, due to the Markowitz approach.

In risk management, we have very often to deal with portfolios consisting of many different assets. In many of these cases it is important to reduce the dimensionality of the problem in order to not only understand the portfolio's risk but also to forecast the risk. A classical method to reduce the dimensionality of a portfolio whose assets are highly correlated is principal component analysis (PCA). PCA is based on the spectral decomposition theorem. Any symmetric or positive definite matrix Σ can be decomposed in

$$\Sigma = PDP',$$

where P is an orthogonal matrix consisting of the eigenvectors of Σ in its columns and D is a diagonal matrix of the eigenvalues of Σ. In addition, we demand $\lambda_i \geq \lambda_{i-1}$, $i = 1, ..., d$ for the eigenvalues of Σ in D. If we apply the spectral decomposition

theorem to the dispersion matrix of an elliptical random vector X with distribution $E_d(\mu, \Sigma, \psi)$, we can interpret the principal components which are defined by

$$Y_i = P_i'(X - \mu), i = 1, ..., d, \tag{2.7}$$

as the main statistical risk factors of the distribution of X in the following sense

$$P_1'\Sigma P_1 = \max\{w'\Sigma w : w'w = 1\}. \tag{2.8}$$

More generally,

$$P_i'\Sigma P_i = \max\{w'\Sigma w : w \in \{P_1, ..., P_{i-1}\}^{\perp}, w'w = 1\}.$$

From equation (2.8), we can derive that the linear combination $Y_1 = P_1'(X - \mu)$ has the highest dispersion of all linear combinations and $P_i'(X - \mu)$ has the highest dispersion in the linear subspace $\{P_1, ..., P_{i-1}\}^{\perp}$. If we interpret $\operatorname{trace} \Sigma = \sum_{j=1}^{d} \sigma_{ii}$ as a measure of total variability in X and since we have

$$\sum_{i=1}^{d} P_i'\Sigma P_i = \sum_{i=1}^{d} \lambda_i = \operatorname{trace} \Sigma = \sum_{i=1}^{d} \sigma_{ii},$$

we can measure the ability of the first principal component to explain the variability of X by the ratio $\sum_{j=1}^{k} \lambda_j / \sum_{j=1}^{d} \lambda_j$.

Furthermore, we can use the principal components to construct a statistical factor model. Due to equation (2.7), we have

$$Y = P'(X - \mu),$$

which can be inverted to

$$X = \mu + PY.$$

If we partition Y due to $(Y_1, Y_2)'$, where $Y_1 \in \mathbf{R}^k$ and $Y_2 \in \mathbf{R}^{d-k}$ and also P leading to (P_1, P_2), where $P_1 \in \mathbf{R}^{d \times k}$ and $P_2 \in \mathbf{R}^{d \times (d-k)}$, we obtain the representation

$$X = \mu + P_1 Y_1 + P_2 Y_2 = \mu + P_1 Y_1 + \epsilon.$$

But one has to be careful. In contrast to the normal distribution case, the principal components are only quasi-uncorrelated but not independent. Furthermore, we obtain for the coefficient of tail dependence between two principal components, say Y_i and

Y_j,

$$\lambda(Y_i, Y_j, 0, \alpha) = \frac{\int_0^{\sqrt{1/2}} \frac{s^\alpha}{\sqrt{1-s^2}} ds}{\int_0^1 \frac{s^\alpha}{\sqrt{1-s^2}} ds}.$$

2.4 Estimation of α-Stable Sub-Gaussian Distributions

In contrast to the general case of multivariate α-stable distributions, we show that the estimation of the parameters of an α-stable sub-Gaussian distribution is feasible. As shown in the last section, α-stable sub-Gaussian distributions belong to the class of elliptical distributions. In general, one can apply a two-step estimation procedure for the elliptical class. In the first step, we estimate independently the location parameter $\mu \in \mathbf{R}^d$ and the positive definite dispersion matrix Σ up to a scale parameter. In the second step, we estimate the parameter of the radial random variable W.

We apply this idea to α-stable sub-Gaussian distributions. In Sections 2.4.1 and 2.4.2 we present our main theoretical results, deriving estimators for the dispersion matrix and proving their consistency. In Section 2.4.3 we present a new procedure to estimate the parameter α of an α-stable sub-Gaussian distribution.

2.4.1 Estimation of the Dispersion Matrix with Covariation

In Section 2.2.1, we introduced the covariation of a multivariate α-stable random vector. This quantity allows us to derive a consistent estimator for an α-stable dispersion matrix. In order to shorten the notation we denote with $\sigma_j = \sigma(e_j)$ the scale parameter of the jth component of an α-stable random vector $X = (X_1, ..., X_d)' \in \mathbf{R}^d$.

Proposition 4.   (a) Let $X = (X_1, ..., X_d)' \in \mathbf{R}^d$ be a zero mean α-stable sub-Gaussian random vector with positive definite dispersion matrix $\Sigma \in \mathbf{R}^{d \times d}$. Then, we have

$$\sigma_{ij} = \frac{2}{c_{\alpha,0}(p)^p} \sigma(e_j)^{2-p} E(X_i X_j^{<p-1>}), \tag{2.9}$$

where $p \in (1, \alpha)$, $c_{\alpha,0}(p) = E(|Y|^p)^{1/p} > 0$ and $Y \sim S_\alpha(1, 0, 0)$.

(b) Let $X_1, X_2, ..., X_n$ be independent and identically distributed samples with the same distribution as the random vector X. Let $\hat{\sigma}_j$ be a consistent estimator for σ_j, the scale parameter of the jth component of X, then, the estimator $\hat{\sigma}_{ij}^{(2)}(n, p)$, defined as

$$\hat{\sigma}_{ij}^{(2)}(n, p) = \frac{2}{c_{\alpha,0}(p)^p} \hat{\sigma}_j^{2-p} \frac{1}{n} \sum_{t=1}^n X_{ti} X_{tj}^{<p-1>}, \tag{2.10}$$

is a consistent estimator for σ_{ij}, where X_{ti} refers to the ith entries of the observation X_t, $t = 1, ..., n$, $c_{\alpha,0}(p) = E(|Y|^p)^{1/p}$ and $Y \sim S_\alpha(1, 0, 0)$.

Proof.   a) Due to the Proposition 3 we have

$$
\begin{aligned}
\sigma_{ij} &\overset{\text{Prop.3}}{=} 2^{\alpha/2}\sigma_{jj}^{(2-p)/2}[X_i, X_j]_\alpha \\
&\overset{\text{Prop.1}}{=} 2^{\alpha/2}\sigma_{jj}^{(2-\alpha)/2}E(X_iX_j^{<p-1>})\sigma_j^\alpha/E(|X_j|^p) \\
&\overset{\text{Lemma 1}}{=} 2^{\alpha/2}\sigma_{jj}^{(2-\alpha)/2}E(X_iX_j^{<p-1>})\sigma_j^\alpha/(c_{\alpha,0}(p)^p\sigma_j^p) \\
&\overset{\text{Cor.1 (i)}}{=} 2^{p/2}\sigma_{jj}^{(2-p)/2}E(X_iX_j^{<p-1>})/(c_{\alpha,0}(p)^p)
\end{aligned}
$$

b) The estimator $\hat{\sigma}_j$ is consistent and $f(x) = x^{2-p}$ is continuous. Then, the estimator $\hat{\sigma}_j^{2-p}$ is consistent for σ_j^{2-p}. $\frac{1}{n}\sum_{k=1}^n X_{ki}X_{kj}^{<p-1>}$ is consistent for $E(X_iX_j^{<p-1>})$ due to the law of large numbers. Since the product of two consistent estimators is consistent, the estimator

$$
\hat{\sigma}_{ij}^{(2)} = = \frac{2}{c_{\alpha,0}(p)^p}\hat{\sigma}_j^{2-p}\frac{1}{n}\sum_{t=1}^n X_{ti}X_{tj}^{<p-1>}
$$

is consistent.

$\square$

2.4.2   Estimation of the Dispersion Matrix with Moment-Type Estimators

In this section, we present an approach of estimating the dispersion matrix up to a scale parameter which is applicable to the class of normal variance mixtures. In particular, we will see that if we know the tail parameter of an α-stable sub-Gaussian random vector $X \in \mathbf{R}^d$, this approach allows us to estimate the dispersion matrix of X.

We denote with $(W_\theta)_{\theta \in \Theta}$ a parametric family of positive random variables.

Lemma 2. *Let $Z \in \mathbf{R}^d$ a normally distributed random vector with mean zero and positive definite dispersion matrix $\Sigma \in \mathbf{R}^{d\times d}$ and let $X_\theta = \mu + \sqrt{W_\theta}Z$, $\theta \in \Theta$, be a d-dimensional normal variance mixture with location parameter $\mu \in \mathbf{R}^d$. Furthermore, we assume that $\sqrt{W_\theta}$ has tail parameter $\alpha(\theta)$, $\theta \in \Theta$.[4] Then, there exists a function $c : \{(\theta, p) \in \Theta \times \mathbf{R} : p \in (0, \alpha(\theta))\} \to (0, \infty)$ such that, for all $a \in \mathbf{R}^d \setminus \{0\}$, we have*

$$
E(|a'(X_\theta - \mu)|^p) = c(\theta, p)^p(a'\Sigma a)^{p/2}. \tag{2.11}
$$

---

[4]Note, if the ranodom variable X has tail parameter α then $E(|X|^p) < \infty$ for all $p < \alpha$ and $E(|X|^p) = \infty$ for all $p \geq \alpha$ (see Samorodnitsky and Taqqu (1994)).

The function c is defined by

$$c(\theta, p) = E(W_\theta^{p/2}) E(|\tilde{Z}|^p),$$

where the random vector $\tilde{Z} \in \mathbf{R}^d$ is standard normally distributed. Furthermore, c satisfies

$$\lim_{p \to 0} c(\theta, p) = 1, \tag{2.12}$$

for all $\theta \in \Theta$.

We see from equation (2.11) that the covariance matrix of Z determines the dispersion matrix of X_θ up to a scaling constant.

Proof. Let $\theta \in \Theta$, $p \in (0, \alpha(\theta))$ and $a \in \mathbf{R}^d \setminus \{0\}$, then we have

$$
\begin{aligned}
E(|a'(X_\theta - \mu)|^p) &= E(|a'W_\theta^{1/2} Z|^p) \\
&= \underbrace{E(W_\theta^{p/2}) E(|a'Z/(a'\Sigma a)^{1/2}|^p)}_{=:c(\theta,p)} (a'\Sigma a)^{p/2}.
\end{aligned}
$$

Note that $\tilde{Z} = a'Z/(a'\Sigma a)^{1/2}$ is standard normally distributed, hence $c(\theta, p)$ is independent of a. Since $E(W_\theta^{p/2}) > 0$ and $E(|a'Z/(a'\Sigma a)^{1/2}|^p) > 0$, so $c(\theta, p) > 0$.

Since we have $x^p \leq \max\{1, x^{\alpha(\theta)}\}$ for $p \in (0, \alpha(\theta))$ and $x > 0$, it follows from Lebesque's Theorem

$$
\begin{aligned}
\lim_{p \to 0} c(\theta, p) &= \lim_{p \to 0} E(\sqrt{W_\theta}^p) \lim_{p \to 0} E(|\tilde{Z}|^p) \\
&= E(\lim_{p \to 0} \sqrt{W_\theta}^p) E(\lim_{p \to 0} |\tilde{Z}|^p) \\
&= E(1)E(1) \\
&= 1.
\end{aligned}
$$

$\square$

Theorem 10. *Let Z, X_θ, $\theta \in \Theta$, and $c : \{(\theta, p) \in \Theta \times \mathbf{R} : p \in (0, \alpha(\theta))\} \to (0, \infty)$ be as in Lemma 2. Let $X_1, ..., X_n \in \mathbf{R}^d$ be i.i.d. samples with the same distribution as X_θ. The estimator*

$$\hat{\sigma}_n(p, a) = \frac{1}{n} \sum_{i=1}^n \frac{|a'(X_i - \mu)|^p}{c(\theta, p)^p} \tag{2.13}$$

(i) is unbiased, i.e.,

$$E(\hat{\sigma}_n(p, a)) = (a'\Sigma a)^{p/2} \text{ for all } a \in \mathbf{R}^d$$

(ii) is consistent, i.e.,

$$P(|\hat{\sigma}_n(p, a) - (a'\Sigma a)^{p/2}| > \epsilon) \to 0 \ (n \to \infty),$$

if $p < \alpha(\theta)/2$.

Proof. (i) follows directly from Lemma 2. For statement (ii), we have show that

$$P(n) := P(|\hat{\sigma}_n(p, a) - (a'\Sigma a)^{p/2}| > \epsilon) \to 0 \ (n \to \infty).$$

But this holds because of

$$
\begin{aligned}
P(n) &\overset{(*)}{\leq} \frac{1}{\epsilon^2} \operatorname{Var}(\hat{\sigma}_n(p, a)) \\
&= \frac{1}{\epsilon^2 n^2 c(\theta, p)^{2p}} \operatorname{Var}\left(\sum_{i=1}^{n} |a(X_i - \mu)|^p\right) \\
&= \frac{1}{\epsilon^2 n c(\theta, p)^{2p}} \operatorname{Var}(|a'(X - \mu)|^p) \\
&= \frac{1}{\epsilon^2 n c(\theta, p)^{2p}} (E(|a'(X - \mu)|^{2p}) - E(|a'(X - \mu)|^p)^2) \\
&= \frac{1}{\epsilon^2 n c(\theta, p)^{2p}} \left(c(\theta, 2p)^{2p}(a'\Sigma a)^{2p} - c(\theta, p)^{2p}(a'\Sigma a)^{2p}\right) \\
&= \frac{1}{\epsilon^2 n} \left(\left(\frac{c(\theta, 2p)}{c(\theta, p)}\right)^{2p} - 1\right) (a'\Sigma a)^{2p} \to 0 \ (n \to \infty).
\end{aligned}
$$

The inequation $(*)$ holds because of the Chebyshev's inequality and we have $E(|\acute{d}(X - \mu)|^{2p}) < \infty$ because of the assumption $p < \alpha(\theta)/2$. $\square$

Note, that $\hat{\sigma}_n(p, a)^{2/p}$, $a \in \mathbf{R}^d$, is a biased, but consistent estimator for $(a\Sigma d)$. However, since we cannot determine $c(\theta, p) > 0$ we have to use

$$\hat{\sigma}_n(p, a)c(\theta, p)^p = \frac{1}{n}\sum_{i=1}^{n} |a'(X_i - \mu)|^p \tag{2.14}$$

as the estimator. But then, Theorem 10 allows us the estimate the dispersion matrix only up to a scaling constant by using linear combinations $dX_1, ..., a'X_n$, $a \in \mathbf{R}^d$ of the observations $X_1, ..., X_n$. We can apply two different approaches to do this.

The first approach is based on the fact that the following equation holds

$$\sigma_{ij} = \frac{(e_i + e_j)'\Sigma(e_i + e_j) - (e_i - e_j)'\Sigma(e_i - e_j)}{4}$$

for all $1 \leq i < j \leq d$. Then, we can conclude that the estimator

$$\hat{\sigma}_{ij}(n,p) := \frac{c(\theta,p)\hat{\sigma}_n(p, e_i + e_j)^{2/p} - c(\theta,p)\hat{\sigma}_n(p, e_i - e_j)^{2/p}}{4} \quad (2.15)$$

is a consistent estimator for σ_{ij} up to the scaling constant $c(\theta, p)$, that is the same for all $1 \leq i < j \leq d$.

For the second approach we use different linear projections $a_i'X_1, ..., a_i'X_n$, $a_i \in \mathbf{R}^d$, $i = 1, ..., m$, of the observations in order to reconstruct Σ through the following optimization problem

$$\hat{\Sigma}(n,p) = \text{argmin}_{\Sigma \in \mathbf{R}^{d \times d}:\text{sym.}} \sum_{i=1}^{m}(c(\theta,p)\hat{\sigma}_n(p, a_i)^{2/p} - a_i\Sigma a_i)^2. \quad (2.16)$$

It is important to note that the optimization problem (2.16) can be solved by ordinary least squares regression.

In the next theorem, we present an estimator that is based on the following observation. Letting $X_\theta, X_1, X_2, X_3, ...$ be a sequence of i.i.d. normal variance mixtures, then we have

$$\lim_{n\to\infty}\lim_{p\to 0}\left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{a'X_i - \mu(a)}{c(\theta,p)}\right|^p\right)^{1/p} \overset{(*)}{=} \lim_{n\to\infty}\prod_{i=1}^{n}|a'X_i - \mu(a)|^{1/n}$$
$$= (a'\Sigma a)^{1/2}.$$

The last equation is true because of (ii) of the following theorem. The proof of the equality (*) can be found in Stoyanov (2005).

Theorem 11. *Let Z, X_θ, $\theta \in \Theta$, and $c : \{(\theta,p) \in \Theta \times \mathbf{R} : p \in (0, \alpha(\theta))\} \to (0,\infty)$ be as in Lemma 2 and let $X_1, ..., X_n \in \mathbf{R}^d$ be i.i.d. samples with the same distribution as X_θ. The estimator*

$$\hat{\sigma}_n(a) = \frac{1}{c(\theta, 1/n)}\prod_{i=1}^{n}|a'(X_i - \mu)|^{1/n}$$

(i) is unbiased, i.e.,

$$E(\hat{\sigma}_n(a)) = (a'\Sigma a)^{1/2} \text{ for all } a \in \mathbf{R}$$

(ii) is consistent, i.e.,

$$P(|\hat{\sigma}_n(a) - (a'\Sigma a)^{1/2}| > \epsilon) \to 0 \ (n \to \infty).$$

Proof. (i) follows directly from Lemma 2. For statement (ii), we have show that

$$P(n) := P(|\hat{\sigma}_n(a) - (a'\Sigma a)^{p/2}| > \epsilon) \to 0 \ (n \to \infty).$$

But this holds because of

$$
\begin{aligned}
P(n) & \overset{(*)}{\leq} \frac{1}{\epsilon^2} \operatorname{Var}(\hat{\sigma}_n(a)) \\
& = \frac{1}{\epsilon^2 c(\theta, 1/n)^2} \operatorname{Var}\left(\prod_{i=1}^n |a'(X_i - \mu)|^{1/n}\right) \\
& = \frac{1}{\epsilon^2 c(\theta, 1/n)^2} \left(\prod_{i=1}^n E(|a'(X_i - \mu)|^{2/n}) - \prod_{i=1}^n E(|a'(X_i - \mu)|^{1/n})^2\right) \\
& = \frac{1}{\epsilon^2 c(\theta, 1/n)^2} (E(|a'(X - \mu)|^{2/n})^n - E(|a'(X - \mu)|^{1/n}))^{2n}) \\
& = \frac{1}{\epsilon^2 c(\theta, 1/n)^2} (c(\theta, 2/n)^2 (a'\Sigma a)^2 - (c(\theta, 1/n)^2 (a'\Sigma a)^2)) \\
& = \frac{1}{\epsilon^2} \left(\left(\frac{c(\theta, 2/n)}{c(\theta, 1/n)}\right)^2 - 1\right) (a'\Sigma a)^2 \to 0 \ (n \to \infty).
\end{aligned}
$$

The inequation $(*)$ holds because of the Chebyshev's inequality. Then (ii) follows from equation (2.12) in Lemma 2. $\qquad\square$

Note, that $\hat{\sigma}_n^2(a)$, $a \in \mathbf{R}^d$, is a biased but consistent estimator for $(d'\Sigma a)$.

For the rest of this section we concentrate on α-stable sub-Gaussian random vectors. In this case, the family of positive random variables $(W_\theta)_{\theta \in \Theta}$ is given by

$$(W_\alpha)_{\alpha \in (0,2)} \text{ and } W_\alpha \sim S_{\alpha/2}(\cos(\frac{\pi\alpha}{4}), 1, 0).$$

Furthermore, the scaling function $c(.,.)$ defined in Lemma 2 satisfies

$$
\begin{aligned}
c(\alpha, p)^p & = 2^p \frac{\Gamma(\frac{p+1}{2})\Gamma(1 - p/\alpha)}{\Gamma(1 - p/2)\sqrt{\pi}} \\
& = \frac{2}{\pi} \sin\left(\frac{\pi p}{2}\right) \Gamma(p) \Gamma\left(1 - \frac{p}{\alpha}\right), \quad (2.17)
\end{aligned}
$$

where $\Gamma(.)$ is the Gamma-function. For the proof of equation (2.17), see Hardin (1984) and Stoyanov (2005). With Theorems 10 and 11, we derive two estimators for the scale parameter $\sigma(a)$ of the linear projection $d'X$ fo an α-stable sub-Gaussian random

vector X. The first one is

$$\hat{\sigma}_n(p,a) \quad = \quad \frac{1}{n}\left(\frac{2}{\pi}\sin\left(\frac{\pi p}{2}\right)\Gamma(p)\Gamma\left(1-\frac{p}{\alpha}\right)\right)^{-1}\sum_{i=1}^{n}|a'X_i-\mu(a)|^p$$

based on Theorem 10. The second one is

$$\hat{\sigma}_n(a) \quad = \quad \frac{1}{c(\alpha,1/n)}\prod_{i=1}^{n}(|a'X_i-\mu(a)|)^{1/n}$$

$$= \quad \left(\frac{2}{\pi}\sin\left(\frac{\pi}{2n}\right)\Gamma\left(\frac{1}{n}\right)\Gamma\left(1-\frac{1}{n\alpha}\right)\right)^{-n}.$$

$$\prod_{i=1}^{n}(|a'X_i-\mu(a)|)^{1/n}.$$

based on Theorem 11. We can reconstruct the stable dispersion matrix from the linear projections as shown in the equations (2.15) and (2.16).

2.4.3   Estimation of the Parameter α

We assume that the data $X_1,...,X_n \in \mathbf{R}^d$ follow a sub-Gaussian α-stable distribution. We propose the following algorithm to obtain the underlying parameter α of the distribution.

(i) Generate i.i.d. samples $u_1,u_2,...,u_n$ according to the uniform distribution on the unit hypersphere $\mathcal{S}^{d-1}$.

(ii) For all i from 1 to n estimate the index of stability α_i with respect to the data $u_i'X_1, u_i'X_2, ..., u_i'X_n$, using an unbiased and fast estimator $\hat{\alpha}$ for the index.

(iii) Calculate the index of stability of the distribution by

$$\hat{\alpha} = \frac{1}{n}\sum_{k=1}^{n}\hat{\alpha}_k.$$

The algorithm converges to the index of stability α of the distribution. (For further information we refer to Rachev, and Mittnik (2000).)

2.4.4   Simulation of α-Stable Sub-Gaussian Distributions

Efficient and fast multivariate random number generators are indispensable for modern portfolio investigations. They are important for Monte-Carlo simulations for VaR, which have to be sampled in a reasonable time frame. For the class of elliptical distributions we present a fast and efficient algorithm which will be used for the simulation

of α-stable sub-Gaussian distributions in the next section. We assume the dispersion matrix Σ to be positive definite. Hence we obtain for the Cholesky decomposition $\Sigma = AA'$ a unique full-rank lower-triangular matrix $A \in \mathbf{R}^{d \times d}$. We present a generic algorithm for generating multivariate elliptically-distributed random vectors. The algorithm is based on the stochastic representation of Corollary 2. For the generation of our samples, we use the following algorithm:

Algorithm for $EC_r(\mu, R; \psi_{sub})$ simulation

  (i) Set $\Sigma = AA'$, via Cholesky decomposition.

  (ii) Sample a random number from W.

  (iii) Sample d independent random numbers $Z_1, ..., Z_d$ from a $N_1(0,1)$ law.

  (iv) Set $U = Z/||Z||$ with $Z = (Z_1, ..., Z_d)$.

  (v) Return $X = \mu + \sqrt{W}AU$

If we want to generate random number with a $E_d(\mu, \Sigma, \psi_{sub})$ law with the algorithm, we choose $W \overset{d}{=} S_{\alpha/2}(\cos(\frac{\pi\alpha}{4})^{2/\alpha}, 1, 0)||Z||^2$, where Z is $N_d(0, Id)$ distributed. It can be shown that $||Z||^2$ is independent of both W as well as $Z/||Z||$.

2.5 Empirical Analysis of the Estimators

In this section, we evaluate two different estimators for the dispersion matrix of an α-stable sub-Gaussian distribution using boxplots. We are primarily interested in estimating the off-diagonal entries, since the diagonal entries σ_{ii} are essentially only the square of the scale parameter σ. Estimators for the scale parameter σ have been analyzed in numerous studies. Due to Corollary 1 and Theorem 11, the estimator

$$\hat{\sigma}_{ij}^{(1)}(n) = \frac{(\hat{\sigma}_n(e_i + e_j))^2 - (\hat{\sigma}_n(e_i - e_j))^2}{2} \tag{2.18}$$

is a consistent estimator for σ_{ij} and the second estimator

$$\hat{\sigma}_{ij}^{(2)}(n,p) = \frac{2}{c_{\alpha,0}(p)^p} \hat{\sigma}_n(e_j)^{2-p} \frac{1}{n} \sum_{k=1}^{n} X_{ki} X_{kj}^{<p-1>}. \tag{2.19}$$

is consistent because of Proposition 4 for $i \neq j$. We analyze the estimators empirically.

For an empirical evaluation of the estimators described above, it is sufficient to exploit the two-dimensional sub-Gaussian law since for estimating σ_{ij} we only need
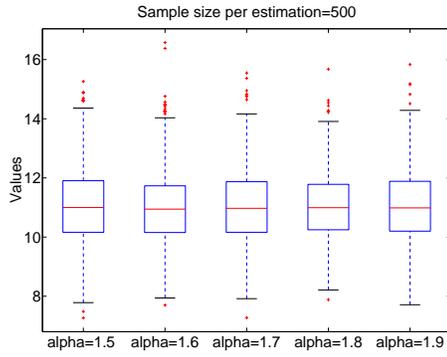
Figure 2.4: Sample size 500   Figure 2.5: Sample size 1000
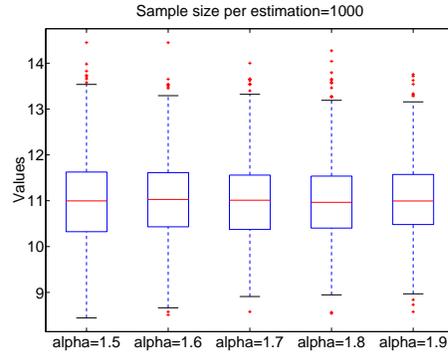
the ith and jth component of the data $X_1, X_2, ..., X_n \in \mathbf{R}^d$. For a better understanding of the speed of convergence of the estimators, we choose different sample sizes ($n = 100, 300, 500, 1000$). Due to the fact that asset returns exhibit an index of stability in the range between 1.5 and 2, we only consider the values $\alpha = 1.5, 1.6, ..., 1.9$. For the empirical analysis of the estimators, we choose the matrix

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

The corresponding dispersion matrix is

$$\Sigma = AA' = \begin{pmatrix} 5 & 11 \\ 11 & 25 \end{pmatrix}.$$

2.5.1 Empirical Analysis of $\hat{\sigma}_{ij}^{(1)}(n)$

For the empirical analysis of $\hat{\sigma}_{ij}^{(1)}(n)$, we generate samples as described in the previous paragraph and use the algorithm described in Section 2.4.4. The generated samples follow an α-stable sub-Gaussian distribution, i.e., $X_i \sim E_2(0, \Sigma, \psi_{sub}(., \alpha))$, $i = 1, ..., n$, where A is defined above. Hence, the value of the off-diagonal entry of the dispersion matrix σ_{12} is 11.

In Figures 2.6 through 2.5, we illustrate the behavior of the estimator $\hat{\sigma}_{ij}^{(1)}(n)$ for several sample sizes and various values for the tail index, i.e., $\alpha = 1.5, 1.6, ..., 1.9$. We demonstrate the behavior of the estimator using boxplots based on 1,000 sample runs for each setting of sample length and parameter value.

In general, one can see that for all values of α the estimators are median-unbiased. By analyzing the figures, we can additionally conclude that all estimators are slightly skewed to the right. Turning our attention to the rate of convergence of the estimates towards the median value of 11, we examine the boxplots. Figure 2.6 reveals that for a

sample size of $n = 100$ the interquartile range is roughly equal to four for all values of α. The range diminishes gradually for increasing sample sizes until which can be seen in Figures 2.6 to 2.5. Finally in Figure 2.5, the interquartile range is equal to about 1.45 for all values of α. The rate of decay is roughly $n^{-1/2}$. Extreme outliers can be observed for small sample sizes larger than twice the median, regardless of the value of α. For $n = 1,000$, we have a maximal error around about 1.5 times the median. Due to right-skewness, extreme values are observed mostly to the right of the median.

2.5.2   Empirical Analysis of $\hat{\sigma}_{ij}^{(2)}(n,p)$

We examine the consistency behavior of the second estimator as defined in (2.19) again using boxplots. In Figures 2.7 through 2.12 we depict the statistical behavior of the estimator. For generating independent samples of various lengths for $\alpha = 1.5, 1.6, 1.7, 1.8,$ and 1.9, and two different values of p we use the algorithm described in Section 2.4.4.[5] For the values of p, we select 1.0001 and 1.3, respectively. A value for p closer to one leads to improved properties of the estimator as will be seen.

In general, we can observe that the estimates are strongly skewed. This is more pronounced for lower values of α while skewness vanishes slightly for increasing α. All figures display a noticeable bias in the median towards low values. Finally, as will be seen, $\hat{\sigma}_{ij}^{(1)}(n)$ seems more appealing than $\hat{\sigma}_{ij}^{(2)}(n,p)$.

For a sample length of $n = 100$, Figures 2.8 and 2.9 show that the bodies of the boxplots which are represented by the innerquartile ranges are as high as 4.5 for a lower value of p and α. As α increases, this effect vanishes slightly. However, results are worse for $p = 1.3$ as already indicated. For sample lengths of $n = 300$, Figures 2.10 and 2.11 show interquartile ranges between 1.9 and 2.4 for lower values of p. Again, results are worse for $p = 1.3$. For $n = 500$, Figures 2.12 and 2.13 reveal ranges between 1.3 and 2.3 as α increases. Again, this worsens when p increases. And finally for samples of length $n = 1,000$, Figures 2.14 and 2.15 indicate that for $p = 1.00001$ the interquartile ranges extend between 1 for $\alpha = 1.9$ and 1.5 for $\alpha = 1.5$. Depending on α, the same pattern but on a worse level is displayed for $p = 1.3$.

It is clear from the statistical analysis that concerning skewness and median bias, the estimator $\hat{\sigma}_{ij}^{(1)}(n)$ has properties superior to estimator $\hat{\sigma}_{ij}^{(2)}(n,p)$ for both values of p. Hence, we use estimator $\hat{\sigma}_{ij}^{(1)}(n)$.

---

[5]In most of these plots, extreme estimates had to be removed to provide for a clear display of the boxplots.
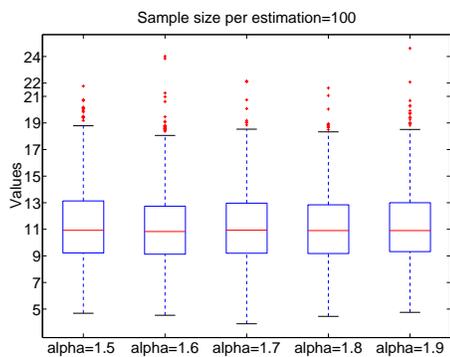
Figure 2.6: Sample size 100
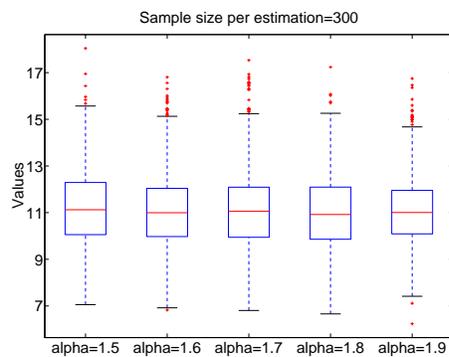
Figure 2.7: Sample size 300
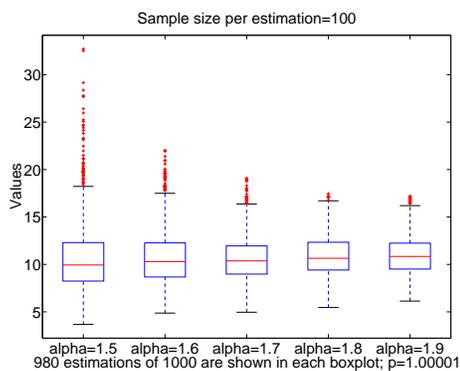
Figure 2.8: Sample size 100, p=1.00001
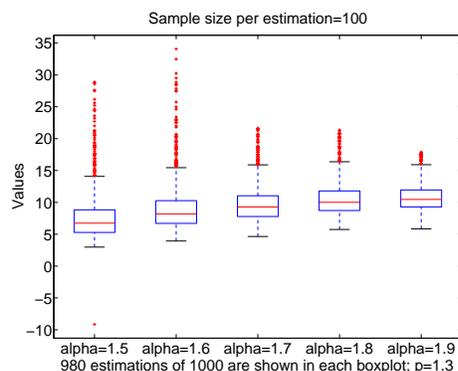
Figure 2.9: Sample size 100, p=1.3
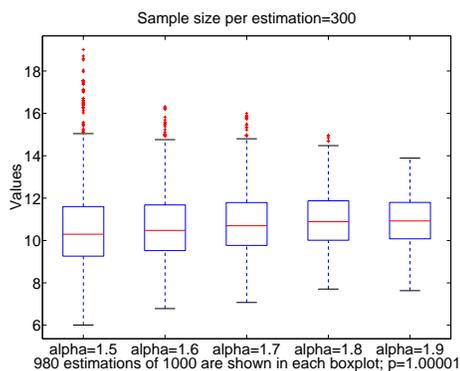
Figure 2.10: Sample size 300, p=1.00001
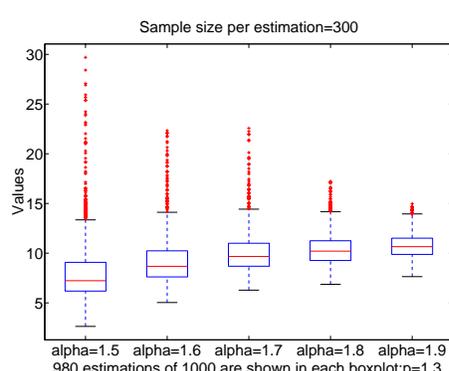
Figure 2.11: Sample size 300,p=1.3
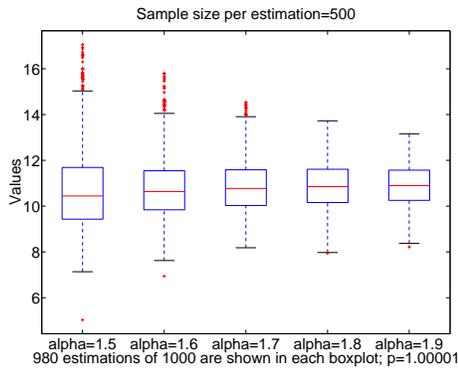
Figure 2.12: Sample size 500
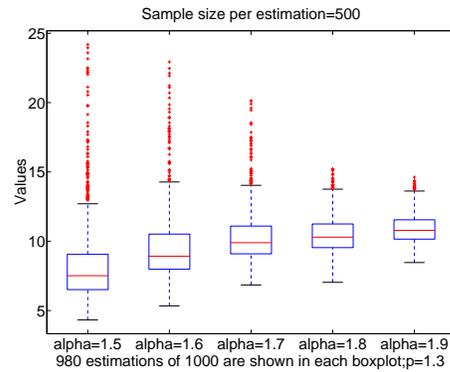

Figure 2.13: Sample size 500
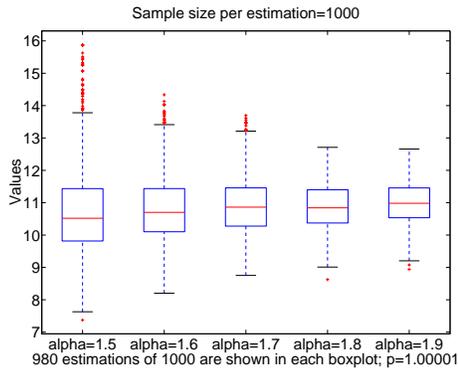
Figure 2.14: Sample size 1000
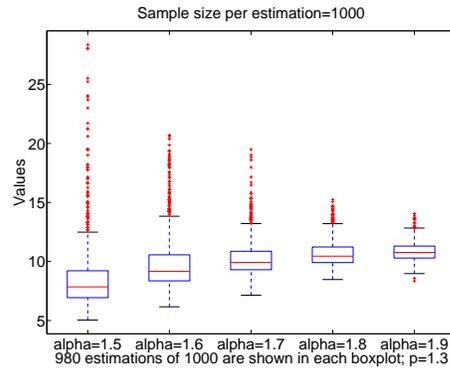
Figure 2.15: Sample size 1000

2.6   Application to the DAX 30

For the empirical analysis of the DAX30 index, we use the data from the Karlsruher Kapitaldatenbank. We analyze data from May 6, 2002 to March 31, 2006. For each company listed in the DAX30, we consider $1,000$ daily log-returns in the study period.[6]

2.6.1   Model Check and Estimation of the Parameter α

Before fitting an α-stable sub-Gaussian distribution, we assessed if the data are appropriate for a sub-Gaussian model. This can be done with at least two different methods. In the first method, we analyze the data by pursuing the following steps (also Nolan (2005)):

(i) For every stock X_i, we estimate $\hat{\theta} = (\hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i, \hat{\mu}_i)$, $i = 1, ..., d$.

---

[6]During our period of analysis Hypo Real Estate Holding AG was in the DAX for only 630 days. Therefore we exclude this company from further treatment leaving us with 29 stocks.
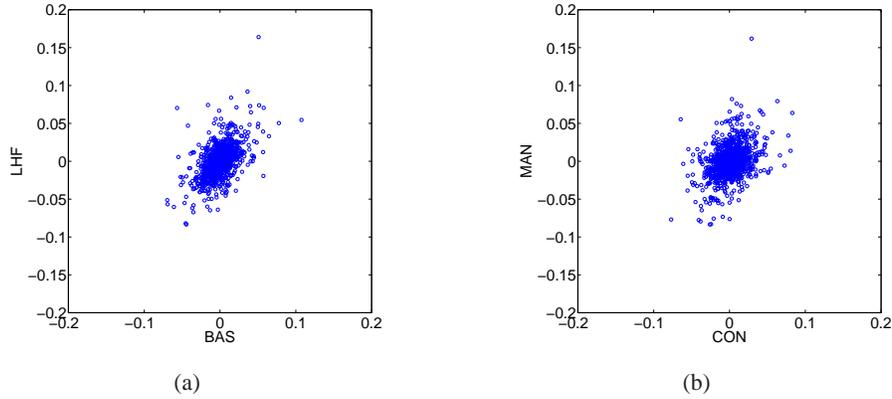
Figure 2.16: Bivariate Scatterplots of BASF and Lufthansa in (a); and of Continental and MAN in (b).

(ii) The estimated $\hat{\alpha}_i$'s should not differ much from each other.

(iii) The estimated $\hat{\beta}_i$'s should be close to zero.

(iv) Bivariate scatterplots of the components should be elliptically contoured.

(v) If the data fulfill criteria (ii)-(iv), a sub-Gaussian model can be justified. If there is a strong discrepancy to one of these criteria we have to reject a sub-Gaussian model.

In Table 2.1, we depict the maximum likelihood estimates for the DAX30 components. The estimated $\hat{\alpha}_i$, $i = 1, ..., 29$, are significantly below 2, indicating leptokurtosis. We calculate the average to be $\bar{\alpha} = 1.6$. These estimates agree with earlier results from Höchstötter, Rachev, and Fabozzi (2005). In that work, stocks of the DAX30 are analyzed during the period 1988 through 2002. Although using different estimation procedures, the results coincide in most cases. The estimated $\hat{\beta}_i$, $i = 1, ..., 29$, are between -0.1756 and 0.1963 and the average, $\bar{\beta}$, equals -0.0129. Observe the substantial variability in the α's and that not all β's are close to zero. These results agree with Nolan (2005) who analyzed the Dow Jones Industrial Average. Concerning item (iv), it is certainly not feasible to look at each bivariate scatterplot of the data. Figure 2.16 depicts randomly chosen bivariate plots. Both scatterplots are roughly elliptical contoured.

The second method to analyze if a dataset allows for a sub-Gaussian model is quite similar to the first one. Instead of considering the components of the DAX30 directly, we examine randomly chosen linear combinations of the components. We only demand that the Euclidean norm of the weights of the linear combination is 1. Due to the theory of α-stable sub-Gaussian distributions, the index of stability is invariant under linear combinations. Furthermore, the estimated $\hat{\beta}$ of linear combination should

Name	Ticker Symbol	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\mu}$
Addidas	ADS	1.716	0.196	0.009	0.001
Allianz	ALV	1.515	-0.176	0.013	-0.001
Atlanta	ALT	1.419	0.012	0.009	0.000
BASF	BAS	1.674	-0.070	0.009	0.000
BMW	BMW	1.595	-0.108	0.010	0.000
Bayer	BAY	1.576	-0.077	0.011	0.000
Commerzbank	CBK	1.534	0.054	0.012	0.001
Continental	CON	1.766	0.012	0.011	0.002
Daimler-Chryser	DCX	1.675	-0.013	0.011	0.000
Deutsch Bank	DBK	1.634	-0.084	0.011	0.000
Deutsche Brse	DB1	1.741	0.049	0.010	0.001
Deutsche Post	DPW	1.778	-0.071	0.011	0.000
Telekom	DTE	1.350	0.030	0.009	0.000
Eon	EOA	1.594	-0.069	0.009	0.000
FresenMed	FME	1.487	0.029	0.010	0.001
Henkel	HEN3	1.634	0.103	0.008	0.000
Infineon	IFX	1.618	0.019	0.017	-0.001
Linde	LIN	1.534	0.063	0.009	0.000
Lufthansa	LHA	1.670	0.030	0.012	-0.001
Man	MAN	1.684	-0.074	0.013	0.001
Metro	MEO	1.526	0.125	0.011	0.001
MncherRck	MUV2	1.376	-0.070	0.011	-0.001
RWE	RWE	1.744	-0.004	0.010	0.000
SAP	SAP	1.415	-0.093	0.011	-0.001
Schering	SCH	1.494	-0.045	0.009	0.000
Siemens	SIE	1.574	-0.125	0.011	0.000
Thyssen	TKA	1.650	-0.027	0.011	0.000
Tui	TUI	1.538	0.035	0.012	-0.001
Volkswagen	VOW	1.690	-0.024	0.012	0.000
Average values		$\bar{\alpha} = 1,6$	$\bar{\beta} = -0,0129$		

Table 2.1: Stable parameter estimates using the maximum likelihood estimator

be close to zero under the sub-Gaussian assumption. These considerations lead us to the following model check procedure:

(i) Generate i.i.d. samples $u_1, ..., u_n \in \mathbf{R}^d$ according to the uniform distribution on the hypersphere $\mathcal{S}_{d-1}$.

(ii) For each linear combination $u_i'X$, $i = 1, ..., n$, estimate $\theta_i = (\hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i, \hat{\mu}_i)$.

(iii) The estimated $\hat{\alpha}_i$'s should not differ much from each other.

(iv) The estimated $\hat{\beta}_i$'s should be close to zero.

(v) Bivariate scatterplots of the components should be elliptically contoured.

(vi) If the data fulfill criteria (ii)-(v) a sub-Gaussian model can be justified.

If we conclude after the model check that our data are sub-Gaussian distributed, we estimate the α of the distribution by taking the mean $\bar{\alpha} = \frac{1}{n} \sum_{i=1}^{n} \hat{\alpha}_i$. This approach has the advantage compared to the former one that we incorporate more information from the dataset and we can generate more sample estimates $\hat{\alpha}_i$ and $\hat{\beta}_i$. In the former approach, we analyze only the marginal distributions.

Figure 2.17 depicts the maximum likelihood estimates for 100 linear combinations due to (ii). We observe that the estimated $\hat{\alpha}_i$, $i = 1, ..., n$, range from 1.5 to 1.84. The average, $\bar{\alpha}$, equals 1.69. Compared to the first approach, the tail indices increase, meaning less leptokurtosis, but the range of the estimates decreases. The estimated $\hat{\beta}_i$'s, $i = 1, ..., n$, lie in a range of -0.4 and 0.4 and the average, $\bar{\beta}$, is -0.0129. In contrast to the first approach, the variability in the β's increases. It is certainly not to be expected that the DAX30 log-returns follow a pure i.i.d. α stable sub-Gaussian model, since we do not account for time dependencies of the returns. The variability of the estimated $\hat{\alpha}$'s might be explained with GARCH-effects such as clustering of volatility. The observed skewness in the data[7] cannot be captured by a sub-Gaussian or any elliptical model. Nevertheless, we observe that the mean of the β's is close to zero.

2.6.2 Estimation of the Stable DAX30 Dispersion Matrix

In this section, we focus on estimating the sample dispersion matrix of an α-stable sub-Gaussian distribution based on the DAX30 data. For the estimation procedure, we use the estimator $\hat{\sigma}_{ij}^{(1)}(n)$, $i \neq j$ presented in Section 2.5. Before applying this estimator, we center each time series by subtracting its sample mean. Estimator $\hat{\sigma}_{ij}^{(1)}(n)$ has the disadvantage that it cannot handle zeros. But after centering the data, there are no zero

---

[7]The estimated $\hat{\beta}$'s differ sometimes significantly from zero.
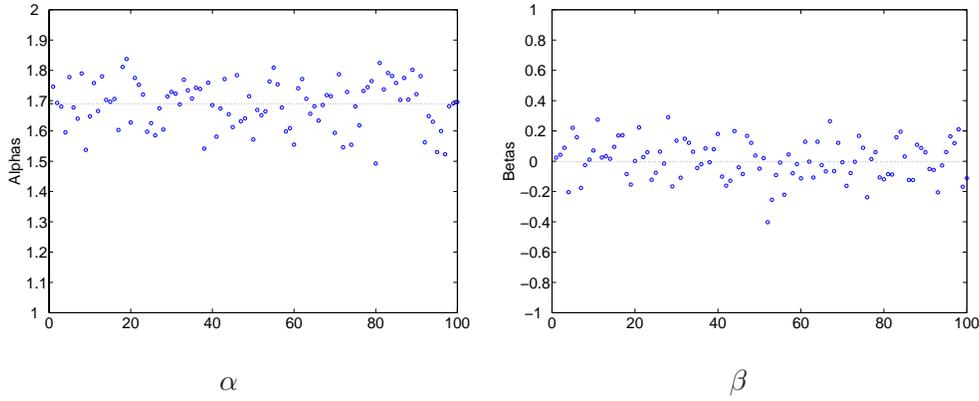
Figure 2.17: Scatterplot of the estimated α's and β's for 100 linear combinations.

log-returns in the time series. In general, this is a point which has to be considered carefully.

For the sake of clarity, we display the sample dispersion matrix and covariance matrix as heat maps, respectively. Figure 2.18 is a heat map of the sample dispersion matrix of the α-stable sub-Gaussian distribution. The sample dispersion matrix is positive definite and has a very similar shape and structure as the sample covariance matrix which is depicted in Figure 2.19. Dark blue colors correspond to low values, whereas dark red colors depict high values.

Figure 2.20 (a) and (b) illustrate the eigenvalues λ_i, $i = 1, ...29$, of the sample dispersion matrix and covariance matrix, respectively. In both Figures, the first eigenvalue is significantly larger than the others. The amounts of the eigenvectors decline in similar fashion.

Figures 2.21 (a) and (b) depict the cumulative proportion of the total variability explained by the first k principal components corresponding to the k largest eigenvalues. In both figures, more than 50% is explained by the first principal component. We observe that the first principal component in the stable case explains slightly more variability than in the ordinary case, e.g., 70% of the total amount of dispersion is captured by the first six stable components whereas in the normal case, only 65% is explained. In contrast to the normal PCA the stable components are not independent but quasi-uncorrelated. Furthermore, in the case of $\alpha = 1.69$, the coefficient of tail dependence for two principal components, say Y_i and Y_j, is

$$\lambda(Y_i, Y_j, 0, 1.69) = \frac{\int_0^{\sqrt{1/2}} \frac{s^{1.69}}{\sqrt{1-s^2}} ds}{\int_0^1 \frac{s^{1.69}}{\sqrt{1-s^2}} ds} \approx 0.21$$

due to Theorem 8 for all $i \neq j$, $i, j = 1, ..., 29$.
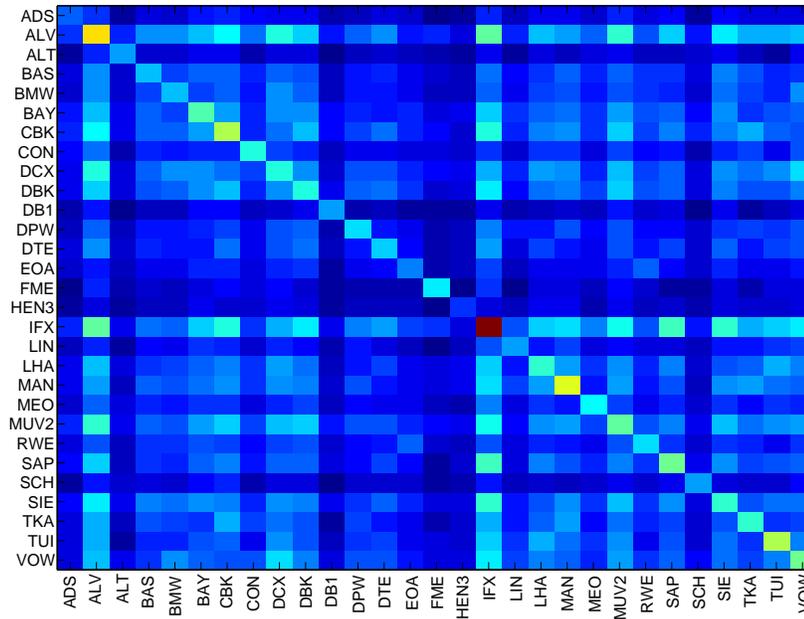
Figure 2.18: Heat map of the sample dispersion matrix. Dark blue colors corresponds to low values (min=0.0000278), to blue, to green, to yellow, to red for high values (max=0,00051)

In Figure 2.22 (a), (b),(c), and (d) we show the first four eigenvectors of the sample dispersion matrix, the so-called vectors of loadings. The first vector is positively weighted for all stocks and can be thought of as describing a kind of index portfolio. The weights of this vector do not sum to one but they can be scaled to be so. The second vector has positive weights for technology titles such as Deutsche Telekom, Infineon, SAP, Siemens and also to the non-technology companies Allianz, Commerzbank, and Tui. The second principal component can be regarded as a trading strategy of buying technology titles and selling the other DAX30 stocks except for Allianz, Commerzbank, and Tui. The first two principal components explain around 56% of the total variability. The vectors of loadings in (c) and (d) correspond to the third and fourth principal component, respectively. It is slightly difficult to interpret this with respect to any economic meaning, hence, we consider them as pure statistical quantities. In conclusion, the estimator $\hat{\sigma}_{ij}(n)$, $i \neq j$, offers a simple way to estimate the dispersion matrix in an i.i.d. α-stable sub-Gaussian model. The results delivered by the estimator are reasonable and consistent with economic theory. Finally, we stress that a stable PCA is feasible.
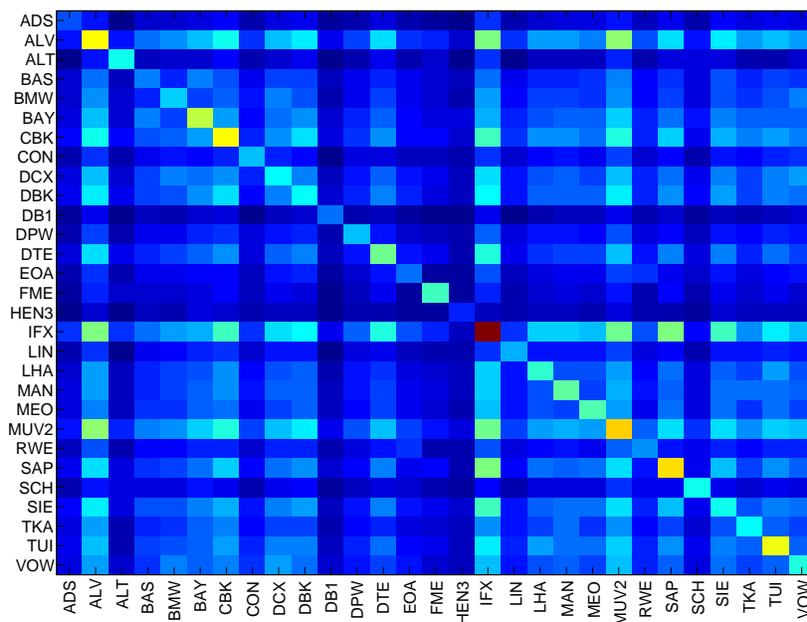
Figure 2.19: Heat map of the sample covariance matrix. Dark blue colors corresponds to low values (min=0.000053), to blue, to green, to yellow, to red for high values (max=0,00097)
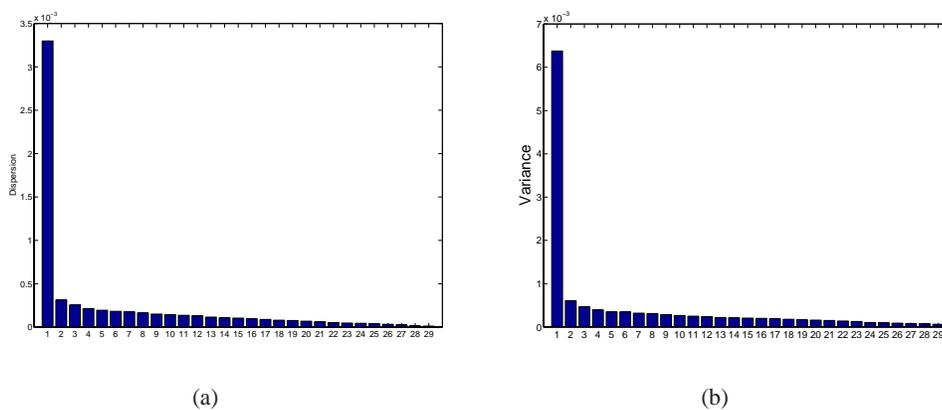
Figure 2.20: Barplots (a) and (b) depict the eigenvalues of the sample dispersion matrix and the sample covariance matrix.

(a)   (b)
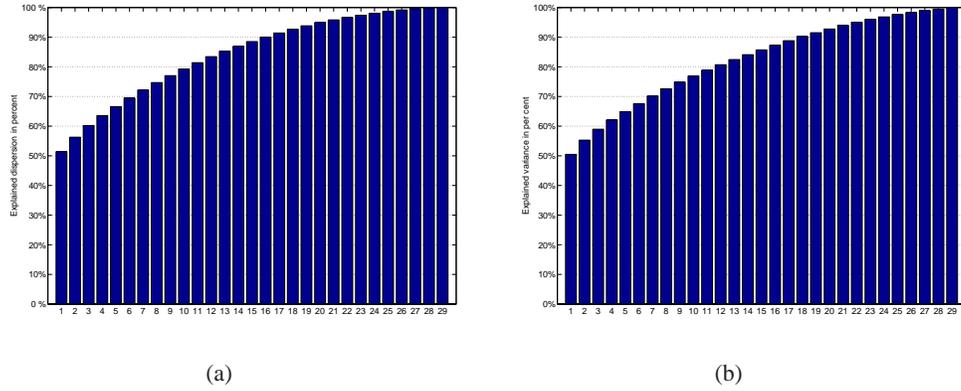
Figure 2.21: Barplot (a) and (b) show the cumulative proportion of the total dispersion and variance explained by the components, i.e., $\sum_{i=1}^{k} \lambda_i / \sum_{i=1}^{29} \lambda_i$.

(a)   (b)

(c)   (d)
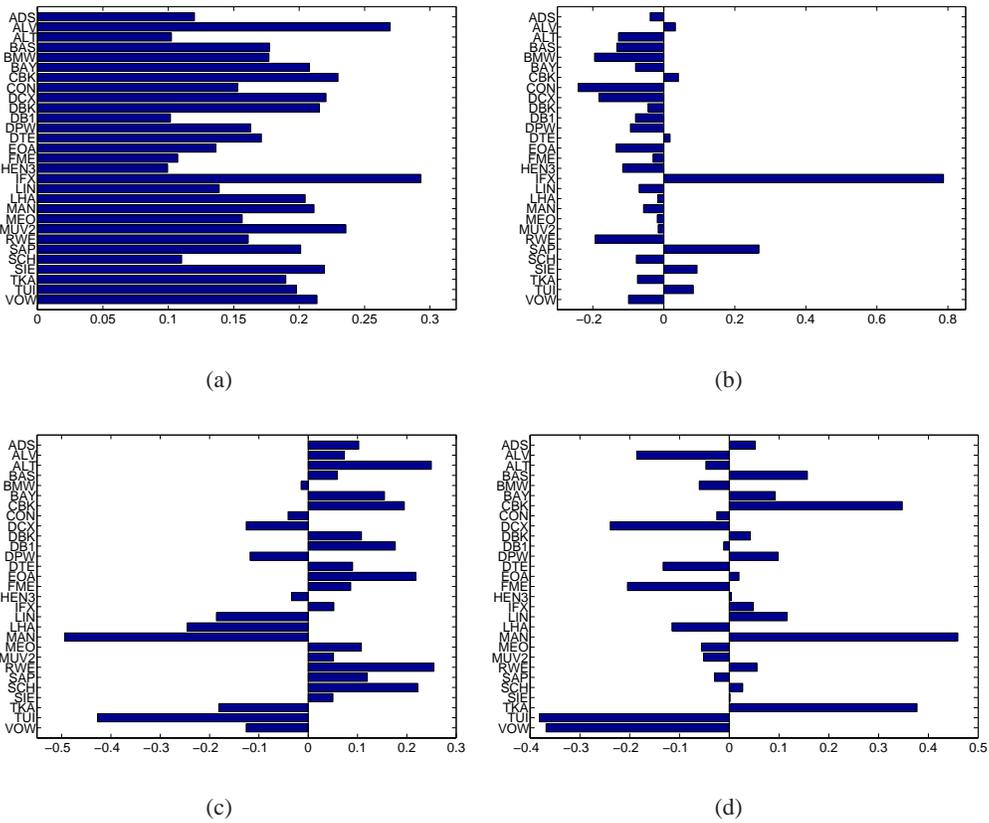
Figure 2.22: Barplot summarizing the loadings vectors g_1, g_2, g_3 and g_4 defining the first four principal components:(a) factor 1 loadings; (b) factor 2 loadings; (c) factor 3 loadings; and (d) factor 4 loadings

2.7   Conclusion

In this chapter we present different estimators which allow one to estimate the dispersion matrix of any normal variance mixture distribution. We analyze the estimators theoretically and show their consistency. We find empirically that the estimator $\hat{\sigma}_{ij}^{(1)}(n)$ has better statistical properties than the estimator $\hat{\sigma}_{ij}^{(2)}(n,p)$ for $i \neq j$. We fit an α-stable sub-Gaussian distribution to the DAX30 components for the first time. The sub-Gaussian model is certainly more realistic than a normal model, since it captures tail dependencies. But it has still the drawback that it cannot incorporate time dependencies.

Chapter 3

Composed and Factor Composed Multivariate GARCH Models

3.1 Introduction

In modern risk management and factor modeling it is important to understand and to predict the temporal dependence structure of assets and risk factor returns in a multivariate time series framework. It is now widely accepted, and some researchers call it even a stylized fact (see McNeil, Frey, and Embrechts (2005)), that the conditional volatilities and the conditional correlation of multivariate financial time series vary over time and occur in clusters.

These style facts are very well understood and modeled by univariate GARCH models for one-dimensional time series. It is straightforward to generalize univariate GARCH models to multivariate GARCH models. Although they are the natural candidate to capture these stylized facts about multivariate financial time series, multivariate GARCH modeling has not been applied very often in the financial industry. The reason is that the implementation of these models is extremely difficult even in low dimensions, their major problem being that the number of parameters tends to explode with the dimension of the model. Because of this, the maximum likelihood function becomes very flat and optimization of the likelihood is practically impossible in higher dimensions, as stressed by Alexander (2002). But from the an asset manager's perspective, a multivariate modeling framework is desirable since it opens the door to better decision tools in various areas, such as asset pricing, portfolio selection, factor modeling, and risk management.

Multivariate GARCH models were introduced by Bollerslev, Engle, and Wooldridge (1988). At the beginning of the 1990s new models were developed such as the constant conditional correlation (CCC) GARCH model by Bollerslev (1990), the principal com-

ponent GARCH model by Ding, and Engle (1994), the BEKK model of Baba, Engle, Kroner and Kraft (1995), and others. At the beginning of the 2000s Christodoulakis, and Satchell (2002), Engle (2002), and Tse, and Tsui (2002) developed the dynamic conditional correlation (DCC)-GARCH model that can be considered to be an extension of the CCC-GARCH model. Furthermore, Patton (2000) and Jondeau, and Rockinger (2006) introduced copula-GARCH models.

The most common applications of multivariate GARCH models are for the study of the conditional covariance and correlation between several markets. Multivariate GARCH models can help asset managers understand if the volatility of one market (e.g., the Dow Jones 30), leads the volatilities of several other markets (such as Euro Stoxx 50, DAX 30 or Nikkei).

In asset pricing theory, the asset excess returns are modeled as linear combinations of factors (e.g., market return). In arbitrary approaches, the coefficients of the factors are assumed to be constant and estimated by an ordinary least squares (OLS-regression. Since these coefficients are the covariance between the asset excess return and the factor returns divided by the variance of the factor returns, these coefficients can be modeled as time varying by a multivariate GARCH model.

In asset management it is not recommended modeling directly all assets in a large portfolio by a multivariate GARCH model since the parameters of the model explode as noted above. Instead, an asset manager should use factor-model strategies in order to reduce the overall dimension of the time series modeling problem. After that the factors obtained can be modeled thoroughly by a multivariate GARCH or, even better, VARMA-MGARCH model.

In this chapter we introduce two new multivariate GARCH models, which we refer to as the composed and factor composed MGARCH models. The idea behind these models comes from a common technique in portfolio risk management: Risk managers of large portfolios have to forecast risk functionals such as value-at-risk (VaR) or expected shortfall of the underlying portfolio. A common approach is to generate a univariate return series from the current asset shares and the multivariate return series of the assets in the portfolio. A univariate model such as a GARCH or ARMA-GARCH model is fitted to this time series allowing the calculation of these risk functionals. However, the univariate model is only valid for the current weights. Since the weights change daily, we have to repeat this procedure every day. Furthermore, the univariate model does not provide any information about the dependence structure of the assets, which is important for the portfolio risk manager. The basic idea behind the composed and factor composed MGARCH models is to use many linear combinations of the multivariate asset return series in the portfolio in order to reconstruct the conditional covariance matrix Σ_t. The matrix Σ_t can be reconstructed by solving an optimization

problem ensuring the positivity of Σ_t.

We extend the composed and factor composed MGARCH model to an α-stable version with multivariate α-stable sub-Gaussian innovations. According to Rachev, and Mittnik (2000), there are empirical as well as theoretical evidences that α-stable laws are the fundamental "building blocks" (i.e., innovations) that drive asset return series in many sectors of the financial market.

This chapter is organized as follows. In Section 3.2 we provide a short review of ARMA-GARCH models since they are the key device for the composed MGARCH models. In Section 3.3 we present and discuss the advantages and disadvantages of the most common multivariate GARCH models. In Section 3.4 we introduce the composed and factor composed GARCH models, put them in the context of the former models, and propose methods to fit composed and factor composed MGARCH models to data. We introduce an α-stable version of the composed and factor composed MGARCH model, that is based on the α-stable power GARCH processes introduced by Mittnik, Paolella, and Rachev (2002) in Section 3.5. In Section 3.6, the α-stable composed MGARCH model is applied to the returns on four German stocks included in the DAX index. We compare the performance of the proposed model with the exponentially weighted moving average (EWMA) model of RiskMetrics. Section 3.7 concludes the chapter.

3.2   Univariate GARCH Models

The univariate generalized autoregressive conditional heteroscedasticity (GARCH) models have been successfully applied in financial econometrics since their introduction by Engle (1982) and Bollerslev (1986). They have been used with great success in volatility forecasting in several financial markets.

The voluminous literature related to GARCH models spans modeling exchange rates, equity returns, convergent term structure volatility forecast, and stochastic volatility models for option pricing and hedging. For a survey of ARCH-type models, see Bollerslev, Chou, and Kroner (1992), Bera, and Higgins (1993), Shephard (1996), Alexander (2001), among others.

In this section we review the basic definitions and properties in the field of univariate GARCH models. We do so because they are the fundamental device for multivariate GARCH modeling. We denote with $\mu_t = E(X_t|\mathcal{F}_{t-1})$, $t \in \mathbf{Z}$ the conditional mean of the time series $(X_t)_{t\in\mathbf{Z}}$, where $\mathcal{F}_t = \sigma(\{X_s : s \leq t\})$, $t \in \mathbf{Z}$, is the sigma field generated by the past and present values of $(X_t)_{t\in\mathbf{Z}}$.

Definition 12. $(Z_t)_{t\in\mathbf{Z}}$ *is a strict white noise (SWN) process if it is a series of identically distributed, finite-variance random variables.*

An important property of financial return series is whether they are *strictly* stationary or *covariance* stationary (see, e.g., McNeil, Frey, and Embrechts (2005) for a definition of this properties). Both of these definitions attempt to formalize the notion that the behavior of a time series is similar in any epoch in which we might observe it. Systematic changes in mean, variance, or the covariances between equally spaced observations are inconsistent with stationarity. We require these notions for the next definition.

Definition 13. *Let $(Z_t)_{t\in\mathbf{Z}}$ be $SWN(0,1)$. The process $(X_t)_{t\in\mathbf{Z}}$ is a GARCH(p,q) process if it is strictly stationary and if it satisfies, for all $t \in \mathbf{Z}$ and some strictly positive-valued process $(\sigma_t)_{t\in\mathbf{Z}}$, the equations*

$$X_t = \sigma_t Z_t, \ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, ..., p$, and $\beta_j \geq 0$, $j = 1, ..., q$.

We demand strictly stationarity of univariate GARCH processes, since most financial return series seem to have this property.

It is straightforward to generalize GARCH processes to so-called ARMA-GARCH processes $(X_t)_{t\in\mathbf{Z}}$ satisfying the equation

$$X_t = \mu_t + \sigma_t Z_t,$$

where $(X_t - \mu_t)_{t\in\mathbf{Z}}$ follows a GARCH(p, q) process and $(\mu_t)_{t\in\mathbf{Z}}$ an ARMA process. (For an introduction to ARMA processes, see Hamilton (1994).) In daily return series volatility effects captured by the GARCH part are much more important than the mean effects modeled by the ARMA part of the model.[1] Because of this fact and for notational ease we do not consider ARMA processes in this chapter.

In the next theorem we give sufficient and necessary conditions for covariance stationarity of GARCH processes.

Theorem 12. *A GARCH(p,q) process is a covariance-stationary white noise process if and only if $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. The variance of the covariance-stationary process is given by $\alpha_0/(1 - \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j)$.*

Proof. See McNeil, Frey, and Embrechts (2005). □

We will see in Section 3.4 that Theorem 12 is very useful to ensure covariance stationarity of composed and factor composed MGARCH processes.

---

[1]For a more detailed discussion of this issue, see RiskMetrics (1996) and McNeil, Frey, and Embrechts (2005).

3.3 Multivariate GARCH Models

In this section we give an historical overview[2] of the more important multivariate GARCH models. As stressed in Section 3.1 we always have to consider a trade-off between the complexity of the model (i.e., amount of parameters) and its applicability for financial modeling. A sophisticated multivariate GARCH specification might have the capability to capture all the phenomenons in the underlying multivariate time series, but if there may not exist an estimation procedure to fit the model to data, the model is not applicable. On the other hand, if the model has a too parsimonious parametrization, we can fit it easily to data but it might be worthless since it does not model the data appropriately.

Additional important properties of multivariate GARCH models are if the definition of the model ensures the positive definiteness of the conditional covariance matrix, the covariance stationarity of the process and the invariance of the model under linear transformation. A positive definite conditional covariance matrix can be achieved in most models, whereas the covariance stationarity is difficult to derive. For practical purposes, the former property is more important because we require a positive definite covariance matrix for the *Cholesky decomposition* in the definition of MGARCH processes.[3]

At the end of this section we describe how to integrate MGARCH models into factor models. This is an important issue in risk management since it is still not possible to model all risk factors of a large portfolio in one MGARCH model. Instead, we have to identify the common underlying risk factors of the portfolio and thoroughly model them by an MGARCH process.

But before beginning with an historical overview of the most common MGARCH models, we present the basic definitions and properties of this model class.

3.3.1 Basic Definitions and Properties

Consider a d-dimensional multivariate time series $(X_t)_{t \in \mathbf{Z}}$ defined on some probability space $(\Omega, \mathcal{F}, P)$. We assume for the rest that $(X_t)_{t \in \mathbf{Z}}$ is always a d-dimensional time series. We denote with $\mathcal{F}_t = \sigma_t(\{X_s : s \leq t\})$ the sigma field generated by the past and present values of the time series $(X_t)_{t \in \mathbf{Z}}$. Based on the efficient market hypothesis (see Fama (1991)) the sigma field $\mathcal{F}_t$ can be interpreted as representing the publicly available information at time t. Furthermore, we refer to

$$\mu_t = E_{t-1}(X_t) = E(X_t | \mathcal{F}_{t-1})$$

---

[2]Detailed surveys about multivariate GARCH models can be found in Bauwens, Laurent, and Rombouts (2006) and McNeil, Frey, and Embrechts (2005).

[3]See Hamilton (1994) for a discussion of the Cholesky decomposition.

as the conditional mean and to

$$\text{Var}_{t-1}(X_t) = E((X_t - \mu_t)(X_t - \mu_t)'|\mathcal{F}_{t-1})$$

as the conditional covariance matrix. The conditional covariance matrix P_t is defined by

$$P_t = \mathcal{P}(\Sigma_t) = \Delta(\Sigma_t)^{-1}\Sigma_t\Delta(\Sigma_t)^{-1},$$

where the operator $\mathcal{P}(.)$ extracts the correlation matrix from the covariance matrix and the operator Δ satisfies

$$\Delta(\Sigma) = \text{diag}(\sqrt{\sigma_{11}}, ..., \sqrt{\sigma_{dd}}).$$

A d-dimensional time series $(Z_t)_{t\in\mathbf{Z}}$ is called *multivariate strict white noise*, denoted by SNW(μ,Σ), if it is a series of independent elliptically distributed random vectors with mean μ and covariance matrix Σ.

Definition 14. *A process $(X_t)_{t\in\mathbf{Z}}$ has a multivariate martingale difference property with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbf{Z}}$ if it satisfies*

$$E|X_t| < \infty \text{ and } E(X_t|\mathcal{F}_{t-1}) = 0,$$

for all $t \in \mathbf{Z}$.

The martingale difference property corresponds to the stylized fact about daily financial return series that conditional expected returns are close to zero.[4] We will see below that a multivariate GARCH process fulfills this property.

Further important properties of multivariate time series can be captured by the following definitions.

Definition 15. *The multivariate time series $(X_t)_{t\in\mathbf{Z}}$ is strictly stationary if*

$$(X'_{t_1}, ..., X'_{t_n}) \overset{d}{=} (X'_{t_1+k}, ..., X'_{t_n+k}),$$

for all $t_1, ..., t_n, k \in \mathbf{Z}$ and for all $n \in \mathbf{N}$.

Definition 16. *The multivariate time series $(X_t)_{t\in\mathbf{Z}}$ is covariance stationary if the first two moments exist and satisfy*

$$\begin{aligned} E(X_t) &= \mu(t) = \mu, \ t \in \mathbf{Z} \\ \text{Cov}(X_t, X_s) &= \text{Cov}(X_{t+k}, X_{s+k}) \ t, s, k \in \mathbf{Z}. \end{aligned}$$

---

[4]For a more detail treatment of this topic see McNeil, Frey, and Embrechts (2005).

Definitions 15 and 16 formalize the notion that the behavior of a time series is similar in any epoch in which we might observe it. Systematic changes in mean, variance or covariances between equally spaces observation are inconsistent with stationarity.

We turn now our attention to the definition of a multivariate GARCH process.

Definition 17. *Let $(Z_t)_{t \in \mathbf{Z}}$ be SWN$(0, Id)$. The process $(X_t)_{t \in \mathbf{Z}}$ is said to be a multivariate GARCH process if it is strictly stationary and satisfies equations of the form*

$$X_t = \Sigma_t^{1/2} Z_t, \ t \in \mathbf{Z},$$

where $\Sigma^{1/2}$ is the Cholesky factor of a positive-definite matrix Σ_t which isH measurable with respect to $\mathcal{F}_{t-1}$.

The most important property about multivariate GARCH models is that the conditional covariance Σ_t is measurable with respect to $\mathcal{F}_{t-1}$. This means that the covariance matrix of tomorrow's asset returns is known today. Note that in contrast to the definition of univariate GARCH processes, there is no functional specification of Σ_t in Definition 17. The functional form will depend on the specific model we define. Because of missing specification we cannot derive the general conditions that are necessary or sufficient for covariance stationarity of a multivariate GARCH process.

It is an immediate conclusion that any pure MGARCH process $(X_t)_{t \in \mathbf{Z}}$ has the martingale difference property, since we have

$$E(X_t | F_t) = E(\Sigma_t^{1/2} Z_t | \mathcal{F}_{t-1}) = \Sigma_t^{1/2} E(Z_t) = 0.$$

Furthermore, Σ_t is the conditional covariance matrix of any MGARCH process, since we have

$$\text{Var}_{t-1}(X_t) = E(X_t X_t' | \mathcal{F}_{t-1}) = \Sigma_t^{1/2} E(Z_t Z_t')(\Sigma_t^{1/2})' = \Sigma_t^{1/2}(\Sigma_t^{1/2})' = \Sigma_t.$$

In particular, in the context of MGARCH models we use Σ_t interchangeably with conditional covariance $\text{Var}(X_t | \mathcal{F}_{t-1})$.

Proposition 5. Let $(X_t)_{t \in \mathbf{Z}}$ be a multivariate GARCH process with conditional covariance matrix process $(\Sigma_t)_{t \in \mathbf{Z}}$. Then the univariate process $(a' X_t)_{t \in \mathbf{Z}}$ has a conditional variance process $(a' \Sigma_t a)_{t \in \mathbf{Z}}$ that is conditioned on the filtration $(\mathcal{F}_t)_{t \in \mathbf{Z}}$ for all $a \in \mathbf{R}^d$.

Proof.

$$\text{Var}_{t-1}(a' X_t) = \text{Var}(a' X_t | \mathcal{F}_{t-1}) = a' \text{Var}(X_t | \mathcal{F}_{t-1}) a = a' \Sigma_t a,$$

for all $a \in \mathbf{R}^d$. $\qquad \square$

As in the univariate case, one can extend the definition of an multivariate GARCH process to a VARMA-GARCH process satisfying equations of the form

$$X_t = \mu_t + \Sigma_t^{1/2} Z_t,$$

where $(X_t - \mu_t)_{t \in \mathbf{Z}}$ follows a MGARCH process and $(\mu_t)_{t \in \mathbf{Z}}$ a VARMA process (see Hamilton (1994)). But such as in the univariate case volatility effects are much more important than mean effects with respect to daily return series. Because of this and the notational ease, we do not model the conditional mean process $(\mu_t)_{t \in \mathbf{Z}}$ in this chapter.

3.3.2   MGARCH Models - an Historical Overview

Multivariate GARCH models were introduced by Bollerslev, Engle, and Wooldridge (1988) in the familiar half-vec *(vech)* form, providing a general framework for multivariate volatility models. In their paper they suggest the vector GARCH or VEC model.

Definition 18. *The process $(X_t)_{t \in \mathbf{Z}}$ is a VEC process if it has the general structure given in Definition 17, and the dynamics of the conditional covariance matrix Σ_t are given by the equations*

$$\operatorname{vech} \Sigma_t = a_0 + \sum_{i=1}^{p} \tilde{A}_i \operatorname{vech}(X_{t-i} X'_{t-i}) + \sum_{j=1}^{q} \tilde{B}_j \operatorname{vech}(\Sigma_{t-j}),$$

for $a_0 \in \mathbf{R}^{d(d+1)/2}$ and matrices $\tilde{A}_i$ and $\tilde{B}_j$ in $\mathbf{R}^{(d(d+1)/2) \times (d(d+1)/2)}$.

The operator vech in Definition 18 stacks the columns of the lower triangle of a symmetric matrix in a single column vector of the length $d(d+1)/2$. In this general definition each element of Σ_t is a linear function of the lagged squared errors and cross-products of errors and the values of the lagged conditional covariance matrices. The fully unrestricted VEC model requires $O(d^4)$ parameters to be estimated by maximum likelihood, where d denotes the dimension of the underlying multivariate time series. The VEC model is certainly the most general MGARCH model, but it has too many parameters for practical purposes and is only of theoretical interest. It is also difficult to ensure the positive definiteness of the conditional covariance matrix. In order to overcome the drawbacks of the VEC model, Bollerslev, Engle, and Wooldbridge proposed the diagonal VEC or DVEC model in the same paper. The DVEC model is essential in the VEC model, but with the additional restriction that the matrices $\tilde{A}_i$ and $\tilde{B}_j$ in Definition 18 have to be diagonal. The DVEC model can be formulated elegantly in terms of the *Hadamard product*, denoted $\circ$, which signifies element-by-element multiplication of two matrices of the same size.

Definition 19. *The process $(X_t)_{t \in \mathbf{Z}}$ is called DVEC process if it has the general struc-*
ture given in Definition 17 and satisfies equations of the form

$$\Sigma_t = A_0 + \sum_{i=1}^{p} A_i \circ (X_{t-i} X'_{t-i}) + \sum_{j=1}^{q} B_j \circ \Sigma_{t-j}$$

where A_0, A_i and B_j are symmetric matrices in $\mathbf{R}^{d \times d}$ such that A_0 has positive diag-
onal elements and all others matrices have non-negative diagonal elements.

The conditional covariance matrix Σ_t is a linear combination of own lagged squared
errors and cross-products of errors. The advantage of the model compared to former
ones is that only $O(d^2)$ parameters needed to be estimated by maximum likelihood.
Furthermore, because of the Hadamard representation of the model it is easy to guar-
antee that Σ_t is positive definite for all t: Provided that A_0, A_i, B_j and the initial
covariance matrix Σ_0 are positive definite for all t Attansio (1991) showed that Σ_t is
positive definite for all t. Certainly, a disadvantage of the DVEC specification is that,
in contrast to the VEC model, the volatility of a single component series cannot be
affected directly by large lagged values in other time series. It should be mentioned
that the DVEC model is still highly parameterized and large-scale systems are difficult
to estimate in practice.

Bollerslev (1990) proposed the constant conditional correlation (CCC) multivari-
ate GARCH specification. The CCC-GARCH model is the simplest representative of
the class of MGARCH processes, where the marginals and the dependence structure
of the multivariate time series are modeled separately. In this class, the marginals are
modeled by univariate GARCH processes, whereas the dependence structure is defined
model specific. In the case of the CCC-GARCH model, the dependence structure is
captured by a constant correlation matrix leading to the following definition.

Definition 20. *The process $(X_t)_{t \in \mathbf{Z}}$ is called a CCC-GARCH process if it is a process*
with the general structure given in Definition 17. The conditional covariance matrix
is of the form $\Delta_t P_c \Delta_t$, where

  (i) P_c is a constant, positive definite correlation matrix; and

  (ii) Δ_t is a diagonal volatility matrix with elements $\sigma_{t,k}$ satisfying

$$\sigma_{t,k}^2 = \alpha_{k0} + \sum_{i=1}^{p_k} \alpha_{ki} X_{t-i,k}^2 + \sum_{j=1}^{q_k} \beta_{kj} \sigma_{t-j,k}^2, \ \ k = 1, ..., d, \qquad (3.1)$$

  where $\alpha_{k0} > 0$, $\alpha_{ki} \geq 0$, $i = 1, ..., p_k$, $\beta_{kj} \geq 0$, $j = 1, ..., q_k$.

It is easy to show that the design of the models guarantees a positive definite con-
ditional covariance matrix. Because of the separation of the marginals and the depen-

dence structure, an efficient two-step estimation procedure is available. In the first step we fit univariate GARCH models to the marginals and in the second step we use the devolatized residuals $Y_t = \Delta^{-1} X_t$ to estimate the constant correlation matrix P_c. This approach has the advantage that it opens the door to modeling large-scale systems. On the other hand, the model has a very parsimonious specification and the assumption of constant correlation may seem to be questionable in empirical work. In particular, Tsui, and Yu (1999) have found that constant correlation can be rejected for certain multivariate time series. However, the CCC-GARCH model is more popular in the financial industry than the models described before and because of its simplicity it is a good starting point for MGARCH modeling.

The BEKK model of Baba, Engle, Kroner and Kraft was published in Engle, and Kroner (1995). The model was also named after the two other authors who co-authored an earlier unpublished manuscript.

Definition 21. *The process $(X_t)_{t \in \mathbf{Z}}$ is a BEKK process if it has the general structure given in Definition 17 and if the conditional covariance matrix Σ_t follows the specification*

$$\Sigma_t = A_0 + \sum_{k=1}^{K} \sum_{i=1}^{p} A'_{k,i} X_{t-i} X'_{t-i} A_{k,i} + \sum_{k=1}^{K} \sum_{i=1}^{q} B_{k,j} \Sigma_{t-j} B_{k,j},$$

where $t \in \mathbf{Z}$, all matrices $A_{k,i}$ and $B_{k,j}$ are in $\mathbf{R}^{d \times d}$ and A_0 is symmetric and positive definite.

The advantage of the model is that it guarantees the positivity of the conditional covariance matrix Σ_t without imposing further restrictions. This is because of the general quadratic structure of the model. One can show that the BEKK model is a special case of the VEC model. The parameter K determines the generality of the process and one can show that the BEKK model covers all DVEC models. In practical applications the parameter K equals 1; even in this case the model is difficult to fit to data and it is rarely used in dimension larger than 3 or 4. In the most common version of the BEKK model $O(d^2)$ parameters have to be estimated. Certainly, a further disadvantage of the model is that the exact interpretation of the individual parameters is not obvious.

Ding, and Engle (1994) described the principal component GARCH (PC-GARCH) model for the first time. This model was extensively investigated by Alexander (2002) under the name orthogonal GARCH.

Definition 22. *The process $(X_t)_{t \in \mathbf{Z}}$ follows a PC-GARCH model if it has the general structure of the process described in Definition 17 and if there exists some orthogonal matrix $\Gamma \in \mathbf{R}^{d \times d}$ with $\Gamma\Gamma' = \Gamma'\Gamma = Id$ such that $(\Gamma' X_t)_{t \in \mathbf{Z}}$ follows a pure diagonal*

GARCH model. The conditional covariance matrix Σ_t satisfies for all t

$$\Sigma_t = \Gamma \Delta_t \Gamma',$$

where Δ_t is defined as in Definition 20.

It can be seen that the model ensures a positive definite covariance matrix Σ_t for all t without imposing further constrains. The strength of this approach is its simplicity and the possibility for dimensionality reduction. The model allows the estimation of large conditional covariance matrices since we have a straightforward estimation technique: In the first step we estimate the sample covariance matrix and by using the Spectral Decomposition Theorem, we calculate the sample principal components. In a second step we fit univariate GARCH models to the principal components. Furthermore, if certain components do not contribute much to the variability of the whole system, they can be neglected, leading to a dimensionality reduction. As Alexander (2002) stresses, the strength of the approach relies on modeling highly correlated systems such as the term structure of commodities futures or interest rates, where only a few principal components capture the behavior of the underlying multivariate time series. On the other hand, the simplicity of the model permits only a very limited evolution of the time series $(\Sigma_t)_{t \in \mathbf{Z}}$. If we have in mind that there is a one-to-one correspondence between a covariance matrix and an ellipsoid, we can visualize the evolution of $(\Sigma_t)_{t \in \mathbf{Z}}$: The corresponding ellipsoid can only be diluted and edged along its principal components, a rotation of the ellipsoid is not possible. As a result, the model only works well in those time series where the directions of the components do not vary over time since the principal components vary their directions over time. This is why the model reveals its weakness in modeling conditional correlation of asset returns.

In 1996 RiskMetrics suggested the exponentially weighted moving average scheme for modeling the conditional covariance matrix in RiskMetrics (1996).

Definition 23. *The process $(X_t)_{t \in \mathbf{Z}}$ is an exponentially weighted moving average (EWMA) process if it is a VEC process satisfying the updating scheme*

$$\Sigma_t = (1 - \lambda) X_{t-1} X'_{t-1} + \lambda \Sigma_{t-1},$$

or equivalent,

$$\Sigma_t = (1 - \lambda) \sum_{i=-\infty}^{t-1} X_i X'_i,$$

for all t.

The RiskMetrics model is widely used in industry, especially for portfolio VaR and is now considered to be an industry standard for market risk. The primary advantage of the RiskMetrics model is that it is extremely easy to estimate, since it has no parameters to be estimated. RiskMetrics suggested the smoothing factor λ to be 0.94 for daily log-returns and $\lambda = 0.97$ for monthly log-returns based on extensive data analysis in various markets and countries. Since in practice we use only the last M observations, we have to rescale the updating scheme in Definition 23, leading to

$$\Sigma_t = \frac{1 - \lambda}{1 - \lambda^{M+1}} \sum_{i=1}^{M} \lambda^i X_{t-i} X'_{t-i}.$$

The obvious drawback of the model is that it has no estimated parameters, and that it forces all assets to have the same decay coefficient irrespective of the asset type. It is necessary to assume the same decay coefficient for all assets to guarantee a positive definite conditional covariance matrix. The EWMA model of RiskMetrics can be regarded as the benchmark model that all other MGARCH models have to outperform.

As mentioned earlier, the assumption of constant correlation in the CCC-GARCH model seems unrealistic in empirical application. Christodoulakis, and Satchell (2002), Engle, and Sheppard (2001), and Tse, and Tsui (2002) suggest a generalization of the CCC-GARCH model the so-called dynamic conditional correlation (DCC) model. There are different versions of the DCC model, the two most common being those of Tse, and Tsui (2002), denoted DCC$_T$, and the one of Engle (2002), denoted DCC$_E$.

Definition 24. *The process $(X_t)_{t \in \mathbf{Z}}$ is a DCC$_T$-GARCH process if it is a process with the general structure given in Definition 17. The conditional covariance matrix is of the form $\Sigma_t = \Delta_t P_t \Delta_t$. The volatility matrix Δ_t is defined as in Definition 20 and P_t satisfies*

$$P_t = (1 - \alpha - \beta) P_c + \alpha \Psi_{t-1} + \beta P_{t-1} \tag{3.2}$$

where $\alpha \geq 0$, $\beta \geq 0$ and $\alpha + \beta < 1$, $P_c \in \mathbf{R}^{d \times d}$ is a positive definite matrix and $\Psi_{t-1} \in \mathbf{R}^{d \times d}$ is the correlation matrix of $(Y_{t-1}, ..., Y_{t-M})$, where $(Y_t)_{t \in \mathbf{Z}} = (\Delta_t^{-1} X_t)_{t \in \mathbf{Z}}$ is the devolatized process.

Definition 25. *The process $(X_t)_{t \in \mathbf{Z}}$ is a DCC$_E$-GARCH process if it has the structure of the process given in Definition 24, but P_t satisfies*

$$P_t = \mathcal{P}(Q_t) \text{ and } Q_t = (1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j) Q_c + \sum_{i=1}^{p} \alpha_i Y_{t-i} Y'_{t-i} + \sum_{j=1}^{q} \beta_j Q_{t-j} \tag{3.3}$$

for all t, where Q_c is the unconditional covariance matrix of the time series $(Y_t)_{t \in \mathbf{Z}} =$

$(\Delta_t^{-1} X_t)_{t \in \mathbf{Z}}$ and the coefficients satisfy $\alpha_i \geq 0$, $\beta_j \geq 0$ and $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$.

The two versions of the DCC-GARCH model permit estimating large conditional covariance matrices since we have the same two-step estimation procedure as in the CCC-GARCH model. The chief difference is that the dependence structure is modeled by a time dependent correlation matrix which is defined by equation (3.2) and equation (3.3), respectively. In particular, we can divide the second step into sub-steps. In the first sub-step, we estimate the matrices P_c and Q_c and in the second sub-step we estimate the scalars α and β and α_i and β_j, respectively. The DCC-GARCH model guarantees the positive definiteness of the sample covariance matrix without imposing further constraints. Since in the DCC$_{\mathrm{T}}$ model P_c, Ψ_{t-i}, and P_{t-1} are positive definite, so is P_t and since in the DCC$_{\mathrm{E}}$ model Q_c, $Y_{t-i}Y'_{t-i}$ and Q_{t-i}, so is Q_t and P_t. If we set $\alpha = \beta = \alpha_i = \beta_i = 0$ we observe that the DCC$_{\mathrm{T}}$ and the DCC$_{\mathrm{E}}$ reduce to the CCC-GARCH model. It can be tested if $\alpha = \beta = 0$ and $\alpha_i = \beta_j = 0$ in order to check whether imposing constant correlation is empirically relevant. Certainly, a drawback of the DCC model is that α and β in the DCC$_{\mathrm{T}}$ model and α_i and β_j in the DCC$_{\mathrm{E}}$ model are scalars instead of matrices. Hence, all entries of the conditional correlation matrix are influenced by the same coefficients which might not be realistic in empirical work. However, these conditions are necessary in order to maintain the positivity of the conditional correlation matrix. In the literature there are extensions of the DCC-GARCH model to overcome the scalar problem. For a further discussion see Billio, Caporin, and Gobbo (2003), Engle (2002) and Pelletier (2003).

Patton (2000) and Jondeau, and Rockinger (2006) were the first to propose a copula-factor model. These models are specified in such a way that the marginals follow GARCH processes and their time varying dependence structure is modeled by a copula. The following definition formalizes this class of processes.

Definition 26. *The process $(X_t)_{t \in \mathbf{Z}}$ is a copula-GARCH model iff it is a process satisfying*

   (i) *the marginals $(X_{t,k})_{t \in \mathbf{Z}}$, $k = 1, ..., d$, follow a GARCH(p_k, q_k) process;*

  (ii) *the dependence structure of the marginals is modeled by a copula*

$$C(u_1, ..., u_d | R_t),$$

   where R is the parameter set defining the copula C and R_t follows an updating scheme $R_t = f(X_{t-1}, X_{t-2}, ...)$;

 (iii) *the conditional distribution is given by*

$$X_t | \mathcal{F}_{t-1} = C(F^{-1}_{X_{t,1} | \mathcal{F}_{t-1}}(X_{t,k}), ..., F^{-1}_{X_{t,d} | \mathcal{F}_{t-1}}(X_{t,d}) | R_t).$$

Model	Parameter count	≥ 10	2	5	10
VEC	$d(d+1)(1+(p+q)d(d+1)/2)/2$	No	21	465	6105
BEKK	$d(d+1)/2 + Kd^2(p+q)$	No	11	65	255
DVEC	$d(d+1)/2(1+p+q)$	No	9	45	165
CCC	$d(d+1)/2 + d(p+q)$	Yes	7	25	77
DCC_T	$d(d+1)/2 + 2 + d(p+q)$	Yes	9	27	77
DCC_E	$d(d+1)/2 + (p+q)(d+1)$	Yes	9	27	77
PC	$d(d+1)/2 + (p+q)d$	Yes	7	25	75
EWMA	$d(d+1)/2$	Yes	3	15	55

Table 3.1: Summary of numbers of parameters in various multivariate GARCH models: in the CCC, DCC_T and DCC_E it is assumed that the numbers of GARCH terms are p and q; in the DCC_T we assume that the conditional correlation matrix has 2 parameters and in the DCC_E we suppose that the conditional correlation matrix has $p + q$ parameters. The second column gives the general formula. The final columns give the numbers for models of dimensions $2, 5$, and 10 when $p = q = 1$.

Note that R_t is measurable with respect to $\mathcal{F}_{t-1}$ and time varying. Patton and Jondeau and Rockinger highlighted in both papers the need to allow for a time-variation in the conditional copula function. The copula function is rendered time varying through its parameters, which can be functions of past data. The copula-MGARCH model can be viewed as an extension of the CCC and DCC-GARCH model.

In Table 3.1[5] we show an overview of the number of parameters used in the models presented in this section. The VEC, BEKK and DVEC models are only applied in low dimensions ($d \leq 10$) and the VEC is purely of theoretical interest. The CCC-, DCC- and PC-GARCH models are implemented in dimensions larger than 10 in the financial industry.

3.3.3   Factor Modeling with MGARCH Models

The material presented in this section follows McNeil, Frey, and Embrechts (2005). It is still not recommended to model all financial risk factors with general multivariate GARCH models. Rather, these models have to be combined with factor-model strategies to reduce the overall dimension of the time series modeling problem.

A fundamental consideration is whether factors are identified *a priori* and treated as exogenous variables, or whether they are treated as endogenous variables and statistical factors manufactured from the observed data.

Suppose we adopt the former approach and identify a small number of common factors F_t to explain the variation in many equity returns X_t. These common factors can be modeled by multivariate GARCH models. The dependence of the individual returns on the factor returns can then be modeled by calibrating a factor model of the

---

[5]Table 3.1 is adopted from McNeil, Frey, and Embrechts (2005).

type

$$X_t = a + BF_t + \epsilon, \ t = 1, ..., n.$$

We assume then that, conditional on the factors F_t, the errors form a multivariate white noise process with GARCH volatility structure.

The latter approach is based on a linear transformation of the equity returns X_t to define factors

$$F_t \ = \ (F_{t,1}, ..., F_{t,k})' = \Gamma_1' X_t,$$

where $\Gamma_1 \in \mathbf{R}^{d \times k}$ and $k << d$. The factors F_t can be modeled by a transformation invariant multivariate GARCH model and should explain most of the variability of the equity returns X_t. This approach leads to a factor model of the form

$$X_t = \Gamma_1 F_t + \epsilon_t, \ t = 1, ..., n,$$

where the error term is usually ignored in practice.

3.4   Composed and Factor Composed MGARCH Models

For high dimensional multivariate GARCH modeling it is indispensable that the model definition permits an efficient estimation procedure. In the previous section we have seen that only those models that allow for a multi-step estimation procedure can be applied in higher dimensions ($d \geq 10$). For example, the specification of the CCC-, DCC-, and copula-GARCH model admits a two-step estimation procedure to estimate the dynamics of the marginals and the temporal dependence structure separately. Similar, in the PC-GARCH model the temporal dependence structure is captured by modeling the principal components of the unconditional covariance matrix through univariate GARCH processes. In addition, since we are interested in statistical factor modeling it is essential that the presented models are invariant under linear transformation. In this section we show that composed and factor composed MGARCH models exhibit the invariance property and allow for two-step estimation procedures.

3.4.1   Definitions and Properties

The key idea behind the specification of the composed and factor composed MGARCH model introduced in this section is to identify the temporal dependence structure of the multivariate time series $(X_t)_{t \in \mathbf{Z}}$ through linear combinations of this time series. These linear combinations are modeled by univariate GARCH processes. In a second step the

dependence structure is reconstructed by solving an optimization problem. But before defining these two models, we have to introduce additional notions.

We assume in the following that all processes $(X_t)_{t\in\mathbf{Z}}$ exhibit unconditional and conditional second moments. Let $(X_t)_{t\in\mathbf{Z}}$ be a d- dimensional process and we denote by $\mathcal{F}_t(a) = \sigma(\{a'X_s : s \leq t\})$ the sigma field generated by the past and present values of the univariate time series $(d\,X_t)_{t\in\mathbf{Z}}$. If $(a'X_t)_{t\in\mathbf{Z}}$ follows a GARCH(p,q) process, we write $\sigma_t^2(a)$ for the conditional variance $\mathrm{Var}(d\,X_t|\mathcal{F}_{t-1}(a))$.

The sigma field $\mathcal{F}_t$ (defined in Section 3.3) includes more information than $\mathcal{F}_t(a)$. It is important to note that mathematically

$$\mathrm{Var}(a'X_t|\mathcal{F}_{t-1}) = \mathrm{Var}(a'X_t|\mathcal{F}_{t-1}(a)) \tag{3.4}$$

is *not* true since we are dealing with different sigma fields.

But we reasonable assume that equation (3.4) holds for many multivariate financial return series at least approximately. In the univariate case, we know that GARCH models based on the filtration $\mathcal{F}(e_i)$ have been successfully applied in volatility forecasting, implying immediately

$$\begin{aligned}
\mathrm{Var}(e_i'X_t|\mathcal{F}_{t-1}) &= e_i'\Sigma_t e_i = \sigma_{t,ii} \approx \mathrm{Var}(e_i'X_t|\mathcal{F}_{t-1}(e_i)) \\
&= \alpha_0 + \sum_{i=1}^{p} X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j,i}^2.
\end{aligned} \tag{3.5}$$

Hence, equation (3.5) justifies equation (3.4) for the marginals, i.e. $a = e_i, i = 1,...,d$, from an empirical point of view. Another argument as to why equation (3.4) holds is the efficient market hypothesis (see Fama (1991)) which asserts that all relevant information of an asset is represented by the past and present values of the time series $(X_{t,i})_{t\in\mathbf{Z}}$. Hence, we obtain

$$\mathrm{Var}(e_i'X_t|\mathcal{F}_{t-1}) = \mathrm{Var}(e_i'X_t|\mathcal{F}_{t-1}(e_i)).$$

Concerning non-trivial linear combinations of multivariate financial time series, we observe that a very widespread and successfully applied technique in the risk management of large portfolios is to take the current weights $w_t \in \mathbf{R}^d$, where d denotes the number of assets in the portfolio, and to generate a univariate time series $(d\,X_t)_{t\in\mathbf{Z}}$ from the portfolio's log-returns $(X_t)_{t\in\mathbf{Z}}$, where $w_t = a \in \mathbf{R}^d$. Then a univariate GARCH process is fitted to the time series $(a'X_t)_{t\in\mathbf{Z}}$. The success of this technique

is based on the empirical fact that we have at least approximately

$$
\begin{aligned}
\mathrm{Var}(a'X_t|\mathcal{F}_{t-1}) &= a'\Sigma_t a \approx \sigma_t^2(a) = \mathrm{Var}(a'X_t|\mathcal{F}_{t-1}(a)) \\
&= \alpha_0(a) + \sum_{i=1}^{p}\alpha_i(a'X_{t-i})^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2(a),
\end{aligned}
$$

where $a \in \mathbf{R}^d$. This consideration is evidence for equation (3.4).

In contrast to the marginals, it is at least not immediately clear how to justify

$$
\mathrm{Var}(a'X_t|\mathcal{F}_{t-1}) = \mathrm{Var}(a'X_t|\mathcal{F}_{t-1}(a))
$$

with the efficient market hypothesis, since $(d\,X_t)_{t\in\mathbf{Z}}$ is an artificial time series that is not observable and traded in financial markets.

These semi-theoretical considerations are summed up in the following definitions.

Definition 27. *A multivariate time series* $(X_t)_{t\in\mathbf{Z}}$ *is projection-efficient if it satisfies*

$$
a'X_t|\mathcal{F}_{t-1} \overset{d}{=} a'X_t|\mathcal{F}_{t-1}(a), \tag{3.6}
$$

for all $t \in \mathbf{Z}$ *and* $a \in \mathbf{R}^d$.

Note that the last definition implies

$$
\mathrm{Var}(a'X_t|\mathcal{F}_{t-1}(a)) = \mathrm{Var}(a'X_t|\mathcal{F}_{t-1}) = a'\Sigma_t a.
$$

The next definition tells us how to model $\mathrm{Var}(d\,X_t|\mathcal{F}_{t-1})$.

Definition 28. *A multivariate time series* $(X_t)_{t\in\mathbf{Z}}$ *is called GARCH-projection-efficient if it is projection-efficient and satisfies*

$$
\mathrm{Var}(a'X_t|\mathcal{F}_{t-1}) = \alpha_0 + \sum_{i=1}^{p}\alpha_i(a'X_{t-i})^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2(a),
$$

for all $t \in \mathbf{Z}$ *and* $a \in \mathbf{R}^d$.

The notion of projection-efficient is derived from the efficient market hypothesis in the sense that all information about the projective time series $(d\,X_t)_{t\in\mathbf{Z}}$ included in the sigma algebra $\mathcal{F}_{t-1}$ equals the information in the sigma algebra $\mathcal{F}_{t-1}(a)$. Furthermore, the term GARCH-projection-efficient stresses that all information about the volatility $a'X_t$ is captured by $\mathcal{F}_{t-1}(a)$ and can be modeled by a univariate GARCH(p,q) process. In particular, a GARCH-projection-efficient time series $(X_t)_{t\in\mathbf{Z}}$ has the property that the variance of $a'X_t|\mathcal{F}_{t-1}(a)$ can be modeled by a GARCH process. Due to the consideration above there should be many multivariate financial time series which are at

least approximately projection-efficient and GARCH-projection efficient. It is now straightforward and consistent to define the following multivariate GARCH process.

Definition 29. *The process* $(X_t)_{t \in Z}$ *follows a composed MGARCH (CMGARCH) process if it is a process with the general structure given in Definition 17 and the conditional matrix* $\Sigma_t = (\sigma_{t,ij})$*satisfies*

(i) $(\sigma_{t,k}^2)_{t \in \mathbf{Z}}$ *follows a univariate GARCH(p_k,q_k) process for* $k = 1, ..., d$.

(ii) *For all* $i, j = 1, ..., d$, $i \neq j$ *we have*

$$\sigma_{t,ij} = \frac{1}{4}(\sigma_t^2(e_i + e_j) - \sigma_t^2(e_i - e_j)),$$

where $(\sigma_t^2(e_i + e_j))_{t \in \mathbf{Z}}$ *and* $\sigma_t^2(e_i - e_j))_{t \in \mathbf{Z}}$ *follow univariate GARCH(p_{ij}^+, q_{ij}^+) and GARCH(p_{ij}^-, q_{ij}^-) processes, respectively.*

The composed MGARCH model does not impose any explicit functional form of the conditional covariance matrix Σ_t such as the other models in Section 3.3. We only have to assume that $(X_t)_{t \in \mathbf{Z}}$ follows a multivariate GARCH process which is GARCH-projection-efficient in order to be consistent. This idea is formalized in the next theorem.

Theorem 13. *Let* $(X_t)_{t \in \mathbf{Z}}$ *be a GARCH-projection-efficient MGARCH process with conditional covariance time series* $(\Sigma_t)_{t \in \mathbf{Z}}$*, then* $(\Sigma_t)_{t \in \mathbf{Z}}$ *can be modeled by a composed MGARCH process.*

Proof. Let $i, j = 1, ..., d$, then we have

$$
\begin{aligned}
\sigma_{t,ij} &= \mathrm{Cov}(X_{t,i}, X_{t,j}|\mathcal{F}_{t-1}) \\
&= E((X_{t,i} - \mu_{t,i})(X_{t,j} - \mu_{t,j})|\mathcal{F}_{t-1}) \\
&= E\left(\frac{1}{4}(X_{t,i} + X_{t,j} - (\mu_{t,i} + \mu_{t,j})^2)) \right. \\
&\quad \left. -\frac{1}{4}(X_{t,i} - X_{t,j} - (\mu_{t,i} - \mu_{t,j}))^2)|\mathcal{F}_{t-1}\right) \\
&= \frac{1}{4}E((X_{t,i} + X_{t,j} - (\mu_{t,i} + \mu_{t,j}))^2|\mathcal{F}_{t-1}) \\
&\quad -\frac{1}{4}E((X_{t,i} - X_{t,j} - (\mu_{t,i} - \mu_{t,j}))^2)|\mathcal{F}_{t-1}) \\
&= \frac{1}{4}(\mathrm{Var}((e_i + e_j)'X_t|\mathcal{F}_{t-1}) - \mathrm{Var}((e_i - e_j)'X_t|\mathcal{F}_{t-1})) \\
&\overset{(*)}{=} \frac{1}{4}(\mathrm{Var}((e_i + e_j)'X_t|\mathcal{F}_{t-1}(e_i + e_j)) - \mathrm{Var}((e_i - e_j)'X_t|\mathcal{F}_{t-1}(e_i - e_j))) \\
&= \frac{1}{4}(\sigma_t^2(e_i + e_j) - \sigma_t^2(e_i - e_j))
\end{aligned}
$$

(*) holds in the last equation, since the process is projection-efficient. In particular, $(\sigma_t^2(e_i + e_j))_{t\in\mathbf{Z}}$ and $(\sigma_t^2(e_i - e_j))_{t\in\mathbf{Z}}$ follow GARCH processes, since $(X_t)_{t\in\mathbf{Z}}$ is GARCH-projection-efficient. The same arguments hold also for diagonal entries $\sigma_{t,ii}$ of Σ_t. $\square$

The property GARCH-projection-efficient is essential for CMGARCH processes since it determines the class of processes that can be modeled by them. In contrast to many applied MGARCH models, we have a motivation for the CMGARCH models.[6] The CMGARCH model resembles the PC-GARCH (see Definition 22) in the sense that we use linear combinations of the multivariate time series $(X_t)_{t\in\mathbf{Z}}$ to model the conditional covariance matrix Σ_t and hence, the temporal dependence structure. But in the CMGARCH approach we extend this idea to a new level, since the conditional covariance matrix is determined solely by univariate processes, which is not the case in the PC-GARCH model. In contrast to most of the other multivariate GARCH models we reviewed in Section 3.3 we can easily derive sufficient conditions for the covariance-stationarity of a CMGARCH model.

Theorem 14. *Let the time series $(X_t)_{t\in\mathbf{Z}}$ follow a composed MGARCH process. The process is covariance stationary if all GARCH processes $(\sigma_{t,i}^2)_{t\in\mathbf{Z}}$, $(\sigma_t^2(e_i + e_j))_{t\in\mathbf{Z}}$ and $(\sigma_t^2(e_i - e_j))_{t\in\mathbf{Z}}$ are covariance stationary, or equivalently, if the coefficients of the GARCH processes $(X_{t,i})_{t\in\mathbf{Z}}$, $((e_i + e_j)'X_t))_{t\in\mathbf{Z}}$ and $((e_i - e_j)'X_t)_{t\in\mathbf{Z}}$ satisfy*

$$\sum_{k=1}^{p_i} \alpha_k^{(i)} + \sum_{k=1}^{q_i} \beta_k^{(i)} < 1,$$

$$\sum_{k=1}^{p_{ij}^+} \alpha_k^{(ij+)} + \sum_{k=1}^{q_{ij}^+} \beta_k^{(ij+)} < 1, \text{ and}$$

$$\sum_{k=1}^{p_{ij}^+} \alpha_k^{(ij-)} + \sum_{k=1}^{q_{ij}^-} \beta_k^{(ij-)} < 1,$$

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

Proof. Since $(\sigma_{t,i})_{t\in\mathbf{Z}}$ is covariance stationary, we obtain

$$E(X_{t,i}^2) = E(\sigma_{t,i}^2 Z_t^2) = E(\sigma_{t,i}^2) = \sigma_i^2,$$

---

[6]See the arguments leading to Definitions 27 and 28.

for $1 = 1, ..., d$. Furthermore, for the unconditional covariances we have

$$
\begin{aligned}
E(X_{t,i}X_{t,j}) &= E(\mathrm{Cov}(X_{t,i}, X_{t,j}|\mathcal{F}_{t-1})) \\
&= E(\sigma_{t,ij}) \\
&= E(\frac{1}{4}(\sigma_t^2(e_i + e_j)) - \sigma_t^2(e_i - e_j)) \\
&= \frac{1}{4}(E(\sigma_t^2(e_i + e_j)) - E(\sigma_t^2(e_i - e_j))) \\
&= \frac{1}{4}(\sigma^2(e_i + e_j) - \sigma^2(e_i - e_j))
\end{aligned}
$$

for all $i, j = 1, ..., d$. In particular, we have $\sigma_{ij} = E(X_{t,i}X_{t,j})$ for all $t \in \mathbf{Z}$. Since $(X_t)_{t\in\mathbf{Z}}$ is a multivariate martingale difference with finite, non-time-dependent second moments $\sigma_{ij}, i, j = 1, ..., d$, it is covariance-stationary white noise. $\square$

One further important property of a multivariate GARCH model is the invariance of the model with respect to linear combinations, that is, the times series $(Y_t)_{t\in\mathbf{Z}} = (FX_t)_{t\in\mathbf{Z}}$ belongs to the same model class, where $F \in \mathbf{R}^{k\times d}$. If $(X_t)_{t\in\mathbf{Z}}$ is a time series of asset returns, a linear transformation $(FX_t)_{t\in\mathbf{Z}}$ corresponds to new assets (portfolios combining the original assets). It seems sensible that a model should be invariant, otherwise the question arises as to which basic assets should be modeled. This aspect becomes very important when we are interested in statistical factor modeling in order to reduce the dimensionality of the portfolio. Statistical risk factors are linear combinations of the underlying assets. Modeling the factors and the assets, respectively, should lead to the same results.

Theorem 15. *Let $(X_t)_{t\in\mathbf{Z}}$ follow a GARCH-projection-efficient CMGARCH process. Then the CMGARCH process is invariant under linear transformation, i.e., the process $(Y_t)_{t\in\mathbf{Z}} = (FX_t)_{t\in\mathbf{Z}}$ follows a CMGARCH process in terms of $(\mathcal{G}_t)_{t\in\mathbf{Z}}$ and*

$$
\mathrm{Var}(Y_t|\mathcal{F}_{t-1}) = F\Sigma_t F' = \mathrm{Var}(Y_t|\mathcal{G}_{t-1}), \tag{3.7}
$$

where $F \in \mathbf{R}^{k\times d}$, $k \in \mathbf{N}$, and $\mathcal{G}_t$ is the sigma field generated by $\sigma(\{Y_s : s \leq t\})$.

Proof. Note, that we have $\mathcal{G}_t \subset \mathcal{F}_t$ for all $t \in \mathbf{Z}$. First we show equation (3.7).

$$
\begin{aligned}
\mathrm{Cov}(Y_{t,i}, Y_{t,j}|\mathcal{G}_{t-1}) &= \frac{1}{4}(\mathrm{Var}((e_i + e_j)'Y_t|\mathcal{G}_{t-1}) - \mathrm{Var}((e_i - e_j')Y_t|\mathcal{G}_{t-1})) \\
&= \frac{1}{4}(\underbrace{E(((e_i + e_j)'Y_t)^2|\mathcal{G}_{t-1})}_{(1)} - \underbrace{(E((e_i + e_j)'Y_t|\mathcal{G}_{t-1}))^2}_{(2)} \\
&\quad - \underbrace{E(((e_i - e_j)'Y_t)^2|\mathcal{G}_{t-1})}_{(3)} - \underbrace{(E((e_i - e_j)'Y_t|\mathcal{G}_{t-1}))^2}_{(4)})
\end{aligned}
$$
$$\tag{3.8}$$

Since we have $\mathcal{F}_t((e_i + e_j)'F) \subset \mathcal{G}_t \subset F_t$ we can derive from term (1) in equation (3.8) that we have

$$
\begin{aligned}
E(((e_i + e_j)'Y_t)^2|\mathcal{G}_{t-1}) &= E(E(((e_i + e_j)'Y_t)^2|\mathcal{F}_{t-1})|\mathcal{G}_{t-1})) \\
&= E(E(((e_i + e_j)'Y_t)^2|\mathcal{F}_{t-1}((e_i + e_j)'F))|\mathcal{G}_{t-1})) \\
&= E(((e_i + e_j)'Y_t)^2|\mathcal{F}_{t-1}((e_i + e_j)'F)) \\
&= E(((e_i + e_j)'Y_t)^2|\mathcal{F}_{t-1}). \tag{3.9}
\end{aligned}
$$

Analogously, equation (3.9) holds also for terms (2), (3), and (4) in equation (3.8). Hence, we obtain

$$
\begin{aligned}
\text{Cov}(Y_{t,i}, Y_{t,j}|\mathcal{G}_{t-1}) &= \frac{1}{4}(E(((e_i + e_j)'Y_t)^2|\mathcal{F}_{t-1}) - E((e_i + e_j)'Y_t|\mathcal{F}_{t-1})^2 \\
&\quad - E(((e_i - e_j)'Y_t)^2|\mathcal{F}_{t-1}) - E((e_i - e_j)'Y_t|\mathcal{F}_{t-1})^2) \\
&= \frac{1}{4}(\text{Var}((e_i + e_j)'Y_t|\mathcal{F}_{t-1}) - \text{Var}((e_i - e_j')Y_t|\mathcal{F}_{t-1})) \\
&= \text{Cov}(Y_{t,i}, Y_{t,j}|\mathcal{F}_{t-1}).
\end{aligned}
$$

Furthermore, we observe that we have

$$
\text{Cov}(Y_{t,i}, Y_{t,j}|\mathcal{F}_{t-1}) = e_i'F\Sigma_t F'e_j
$$

and we have proved equation (3.7). We can derive from equations (3.8) and (3.9) that we have

$$
\begin{aligned}
\text{Cov}(Y_{t,i}, Y_{t,j}|\mathcal{G}_{t-1}) &= \frac{1}{4}(\text{Var}((e_i + e_j)'Y_t|\mathcal{G}_{t-1}(e_i + e_j)) \\
&\quad - \text{Var}((e_i + e_j)'Y_t|\mathcal{G}_{t-1}(e_i - e_j))),
\end{aligned}
$$

where $\mathcal{G}_t(e_i + e_j) = \mathcal{F}_t((e_i + e_j)'F)$ and $\mathcal{G}_t(e_i - e_j) = \mathcal{F}_t((e_i - e_j)'F)$. Since the process $(X_t)_{t\in\mathbf{Z}}$ is GARCH-projection-efficient, $\text{Var}((e_i + e_j)'Y_t|\mathcal{G}_{t-1}(e_i + e_j))$ and $\text{Var}((e_i - e_j)'Y_t|\mathcal{G}_{t-1}(e_i - e_j))$ follow a GARCH process. Hence, the time series $(Y_t)_{t\in\mathbf{Z}}$ is a CMGARCH process. $\square$

Because of Theorem 17 we can consistently define an extension of the composed multivariate GARCH model what we call the factor composed multivariate GARCH (FCMGARCH) model.

Definition 30. *The process $(X_t)_{t\in\mathbf{Z}}$ follows a factor composed MGARCH (FCM-GARCH) process, if there exists some orthogonal matrix $\Gamma \in \mathbf{R}^{d\times d}$ satisfying $\Gamma\Gamma' = $ Id such that $(\Gamma'X_t)_{t\in\mathbf{Z}}$ follows a composed MGARCH process.*

The definition of the factor composed MGARCH model resembles the definition

of the PC-GARCH model. As with the PC-GARCH model we are interested in modeling the principal components $(Y_t) = (\Gamma' X_t)_{t \in \mathbf{Z}}$ of the unconditional covariance matrix $\mathrm{Cov}(X_t)$. If the multivariate time series $(X_t)_{t \in \mathbf{Z}}$ is highly correlated, this approach has the advantage that we can model the system through $d_1 << d$ principal components appropriately. But in contrast to the PC-GARCH model, the factor composed GARCH model is more flexible because not only are the principal components modeled by univariate GARCH processes but also the conditional covariance between these factors. Furthermore, the FCMGARCH model offers the opportunity of reducing the dimensionality of the estimation problem since the number of parameters needed to be estimated is proportional to d^2. This follows from the next proposition.

Proposition 6. Let $(X_t)_{t \in \mathbf{Z}}$ be a d-dimensional CMGARCH process and the time series $(X_{t,i})_{t \in \mathbf{Z}}$, $((e_i + e_j)' X_t)_{t \in \mathbf{Z}}$ and $((e_i - e_j)' X_t)_{t \in \mathbf{Z}}$ follow GARCH(p, q) processes, then we have to estimate $(p + q + 1)d^2$ parameters.

Proof. Since the conditional volatility process $(\sigma_{t,ii})_{t \in \mathbf{Z}}$ of $(X_{t,i})_{t \in \mathbf{Z}}$ follows a GARCH(p, q) process, we have to estimate $1 + p + q$ coefficients for one GARCH process. Since there are d different marginal processes we have to estimate $d(1+p+q)$ parameters. For each conditional covariance process $(\sigma_{t,ij})_{t \in \mathbf{Z}}$, $i \neq j$, we estimate $2(1 + p + q)$ parameters due to the formula

$$\sigma_{t,ij} = \frac{1}{4}(\sigma_t^2(e_i + e_j) - \sigma_t^2(e_i - e_j)),$$

for $i, j = 1, ..., d$. Since we have

$$\sigma_{t,ij} = \sigma_{t,ji}$$

we have $d(d - 1)/2$ different conditional covariance processes. Hence we have to estimate $d(d - 1)(1 + p + q)$ parameters for the conditional covariance processes $(\sigma_{t,ij})_{t \in \mathbf{Z}}$. In conclusion, we estimate $(1 + p + q)d^2$ parameters.   □

In many financial applications, it is sufficient to use a GARCH$(1, 1)$ to model the linear combinations. In the case of a d-dimensional CMGARCH process $(X_t)_{t \in \mathbf{Z}}$, we have to estimate $3d^2$ coefficients. In the CMGARCH model, we have to estimate more parameters than in the CCC-GARCH, DCC-GARCH or the PC-GARCH model as can be seen from Table 3.1. But we stress, that the CMGARCH model is more flexible than these models since we do not have a restrictive functional form which allows only a very constrained evolution of the conditional covariance process $(\Sigma_t)_{t \in \mathbf{Z}}$. A further advantage of this model is that we estimate Σ_t only through univariate GARCH processes. This approach allows us to circumvent the problem of applying maximum

likelihood estimation in high dimension. As stressed in Alexander (2001), this is the fundamental problem of multivariate GARCH modeling.

Nevertheless, in multivariate GARCH modeling for large portfolios several researchers such as Alexander (2001) and McNeil, Frey, and Embrechts (2005) recommend a factor model approach in order to reduce the dimensionality of the portfolio. According to Theorem 17, the FCMGARCH model is consistent with the CMGARCH and, as mentioned before, is predestinated to model the principal components of the unconditional covariance matrix $\mathrm{Cov}(X_t)$.[7] In addition, in many financial time series we observe the so-called "80/20 rule" or "Pareto principle" which says that 20% of the largest eigenvalues account for 80% of the overall variability. Hence, if we model 20% of the "largest" principal components we can decrease the parameters needed to be estimated by 96%.

Certainly, a drawback of the CMGARCH model is that its definition does not ensure a positive definite conditional covariance matrix Σ_t, meaning that if the estimation error of $\hat{\Sigma}_t$ becomes too large the matrix is not necessarily positive definite. In the next section we present a method to deal with this problem.

3.4.2 Estimation of the Models

In this section, we introduce two approaches to estimate the CMGARCH and FCMGARCH model. The two approaches have in common that in the first step the problem of estimating the conditional covariance matrix Σ_t is decomposed into $n \in \mathbf{N}$ simpler estimation problems. For these estimation problems, efficient solving algorithms are available. In the following steps we apply these solutions of the n estimation problems to reconstruct the conditional covariance matrix Σ_t. A similar approach has been successfully applied to estimate multivariate α-stable sub-Gaussian distributions (see Nolan (2005) and Kring et al. (2007)).

Since the CMGARCH and FCMGARCH model specification does not guarantee the positivity of the conditional covariance matrix Σ_t, fortunately, the second presented estimation procedure ensures a positive definite estimate $\hat{\Sigma}_t$ of the conditional covariance matrix by applying the Cholesky Decomposition Theorem.

The first estimation approach is immediately derived from the definition of the CMGARCH process.

(1) Let $X_1, ..., X_t$ be a sequence of return data. Fit univariate GARCH processes to the projective data sets $X_{1,i}, ..., X_{t,i}$, where $(e_i + e_j)'X_1, ..., (e_i + e_j)'X_t$ and $(e_i - e_j)'X_1, ..., (e_i - e_j)'X_t$, $1 \leq i < j \leq d$. Denote the corresponding volatility estimates with $\hat{\sigma}_t^2(e_i)$, $\hat{\sigma}_t^2(e_i + e_j)$ and $\hat{\sigma}_t^2(e_i - e_j)$.

---

[7]Due to Theorem 14, we can assume $(X_t)_{t \in \mathbf{Z}}$ to be covariance stationary.

(2) Reconstruct the conditional covariance matrix $\hat{\Sigma}_t$ by

$$
\begin{aligned}
\hat{\Sigma}_{t,ii} &= \hat{\sigma}_t^2(e_i), \\
\hat{\Sigma}_{t,ij} &= \frac{1}{4}(\hat{\sigma}_t^2(e_i + e_j) + \hat{\sigma}_t^2(e_i - e_j)),
\end{aligned}
$$

or alternatively,

(2') Reconstruct the conditional covariance matrix $\hat{\Sigma}_t$ by

$$
\begin{aligned}
\hat{\Sigma}_{t,ii} &= \hat{\sigma}_t^2(e_i), \\
\hat{\Sigma}_{t,ij} &= \frac{1}{2}(\hat{\sigma}_t^2(e_i + e_j) - \hat{\sigma}_t^2(e_i) - \hat{\sigma}_t^2(e_j)),
\end{aligned}
$$

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

This approach has the advantage that it is computationally straightforward. In step (1) we have to fit d^2 GARCH processes in order to calculate $\hat{\Sigma}$ in step (2) and for the alternative approach we have to fit only $d(d+1)/2$ GARCH processes in step (1) since we do not have to estimate $\hat{\sigma}_t^2(e_i - e_j)$ in step (2'). This method has the drawback that we cannot ensure the positivity of the conditional covariance matrix $\hat{\Sigma}$. Hence, we always have to check whether $\hat{\Sigma}_t$ is positive definite. One way of doing this is to apply the Spectral Decomposition Theorem.

If all eigenvalues are positive, then $\hat{\Sigma}_t$ is positive definite.

The second estimation approach is called the *regression approach* since we reconstruct the conditional covariance matrix Σ_t using a regression.

(1) Let $X_1, ..., X_t$ be a sequence of return data. Fit univariate GARCH processes to the projective time series $u_i'X_1, ..., u_i'X_t$, where $u_i \in \mathbf{R}^d$ and $i = 1, ..., n$. Denote the volatility estimates with $\hat{\sigma}_t^2(u_i)$, $i = 1, ..., n$.

(2) Reconstruct the conditional covariance matrix by

$$
\hat{\Sigma}_t^{(2)} = \operatorname{argmin}_{\Sigma \in S^{d \times d}} \sum_{i=1}^{n} (u_i'\Sigma u_i - \sigma_t^2(u_i))^2,
$$

where $S^{d \times d} = \{\Sigma | \Sigma \in \mathbf{R}^{d \times d}, \Sigma' = \Sigma\}$.

The regression approach may be more accurate than the former approach because it uses multiple directions, whereas the first method only uses the directions e_i, e_j, $e_i + e_j$ and $e_i - e_j$. In addition, this approach allows for more flexibility since the directions u_i are not predefined and their number n is also variable. For example, it might be better to model the principal components Y_i, $i = 1, ..., d$ of the sample covariance matrix Σ and their linear combinations than the marginals in order to estimate $\hat{\Sigma}_t$. Or,

one might increase the accuracy of the estimate for Σ_t by increasing the number of directions. But still, so far we cannot ensure $\hat{\Sigma}_t^{(2)}$ to be positive definite. We overcome this drawback through an additional step (3).

(3) In case $\det(\hat{\Sigma}_t^{(2)}) > 0$, set

$$\hat{\Sigma}_t = \hat{\Sigma}_t^{(2)}.$$

Otherwise reconstruct the conditional covariance matrix $\hat{\Sigma}_t^{(3)}$ by

$$\begin{aligned}
\hat{\Delta}_t &= \operatorname{argmin}_{\Delta \in D_d} \sum_{i=1}^{n} (u_i' \Delta\Delta' u_i - \hat{\sigma}_t^2(u_i))^2, \text{ and} \\
\hat{\Sigma}_t^{(3)} &= \hat{\Delta}_t \hat{\Delta}_t',
\end{aligned}$$

where $D_d = \{\Delta | \Delta \in \mathbf{R}^{d \times d}, \Delta \text{ regular upper triangular matrix}\}$. Finally set

$$\hat{\Sigma}_t = \hat{\Sigma}_t^{(3)}.$$

Due to the Cholesky Decomposition Theorem, the optimization problem in step (3) is equivalent to

$$\hat{\Sigma}_t = \operatorname{argmin}_{\Sigma \in D_d^2} \sum_{i=1}^{n} (u_i' \Sigma u_i - \hat{\sigma}_t^2(u_i))^2$$

where D_d^2 is the set of all positive definite $d \times d$ matrices. Hence, the conditional step (3) guarantees $\hat{\Sigma}_t$ to be positive definite. While the optimization problem in step (3) is computationally much more involved than in the one in step (2)[8], it is important to note that if $\hat{\Sigma}_t^{(2)}$ is positive definite then $\hat{\Sigma}_t^{(2)}$ equals $\hat{\Sigma}_t^{(3)}$.

It is straightforward to show that $\hat{\Sigma}_t^{(3)} = \hat{\Delta}_t \hat{\Delta}_t'$ is positive definite, since we have

$$\begin{aligned}
u' \hat{\Sigma}_t u &= u' \hat{\Delta}_t \hat{\Delta}_t' u \\
&= ||\hat{\Delta}_t' u||^2 \\
&> 0,
\end{aligned}$$

for all $u \in \mathbf{R}^d \setminus \{0\}$ and $\hat{\Delta}_t \in D_d$.

---

[8]The optimization problem in step (2) can be solved by OLS regression.

3.5 α-Stable Composed and Factor Composed MGARCH Models

3.5.1 α-Stable Power-GARCH Processes

It is often observed when fitting GARCH models to financial time series that univariate GARCH residuals tend to be heavy tailed. To accommodate this, GARCH models with heavier conditional innovation distributions than those of the normal have been proposed, among them the Student's t and the Generalized Hyperbolic Distribution. To allow for heavy-tailed, conditional distributions, GARCH processes with α-stable error distributions have been considered by McCulloch (1985), Panorska, Mittnik, and Rachev (1995), Mittnik, Paolella, and Rachev (1998), Rachev, and Mittnik (2000), among others.

An objection against the use of the α-stable distribution is that it has no second moments. This seems to contradict empirical studies suggesting the existence of third or fourth moments for various financial return data. But as Mittnik, Paolella, and Rachev (2002) stressed, these findings had been almost exclusively obtained by the use of the Hill (1975) or related tail estimators, which are known to be highly unreliable.

In this section, we present α-stable power-GARCH processes which were originally introduced by Rachev, and Mittnik (2000) and Mittnik, Paolella, and Rachev (2002).

Definition 31. *An univariate process $(X_t)_{t \in \mathbf{Z}}$ is called an α-stable Paretian power-GARCH process, in short, an $S_{\alpha,\beta,\delta}GARCH(r,s)$ process, if it is described by*

$$X_t = \mu_t + \sigma_t Z_t, \text{ where } Z_t \sim S_\alpha(1,\beta,0), \tag{3.10}$$

and

$$\sigma_t^\delta = \alpha_0 + \sum_{i=1}^p \alpha_i |X_{t-i} - \mu_{t-i}|^\delta + \sum_{j=1}^q \beta_j \sigma_{t-j}^\delta,$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, ..., r$, $\beta_j \geq 0$, $j = 1, ..., q$, $0 < \delta < \alpha$ and $S_\alpha(1,\beta,0)$ denotes the α-stable distribution with tail index $\alpha \in (1,2]$, skewness parameter $\beta \in [-1,1]$, zero location parameter and unit scale parameter. The location parameter process $(\mu_t)_{t \in \mathbf{Z}}$ in (3.10) follows an ARMA process.

Since for $\alpha < 2$ Z_t in Definition 31 does not possess moments of order α or higher, we restrict α to be in the set $(1,2]$ in order to possess first moments.[9] This restriction is consistent with financial return data (see among others Höchstötter, Rachev, and

---

[9]See Samorodnitsky and Taqqu (1994) for the existence of moments of an α-stable random variable.

Fabozzi (2005) and Rachev, and Mittnik (2000)), where we observe α to be in the same range. Furthermore, it is important to note, that σ_t is just a time-varying scaling parameter, implying $\sigma_t Z_t | \mathcal{F}_{t-1} \sim S_\alpha(\sigma_t, \sigma_t \beta, 0)$. Hence, in an α-stable power-GARCH process we forecast the scale parameter of the α-stable innovation distribution.

Mittnik, Paolella, and Rachev (2002) show the following proposition.

Proposition 7. The $S_{\alpha,\beta,\delta}$GARCH process has unique, strictly stationary solution if

$$\lambda_{\alpha,\beta,\delta} \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j \leq 1, \tag{3.11}$$

where $\lambda_{\alpha,\beta,\delta} = E(|\mathbf{Z}_t|^\delta)$ and $Z_t \sim S_\alpha(1, \beta, 0)$.

Proposition 7 allows us to guarantee a unique, strictly stationary solution of $S_{\alpha,\beta,\delta}$GARCH process by imposing equation (3.11) during estimation.

3.5.2 A Multivariate α-stable GARCH Model

In this section, we propose an α-stable version of the composed and factor composed MGARCH models, allowing similar estimation procedures as in the ordinary versions of these processes. In particular, we are dealing with processes with α-stable innovation we believe, according to Rachev, and Mittnik (2000), that these are the fundamental "building blocks" that drive asset return processes. But the main problem we face in defining an α-stable MGARCH model with multivariate α-stable innovations is that we do not possess second moments and the conditional covariance matrix or any covariance matrix are not defined. [10] We overcome this problem by choosing the α-stable sub-Gaussian distribution for the innovations. In this particular case, we obtain a substitute for the covariance matrix, the so-called dispersion matrix. The dispersion matrix has the same interpretation in terms of the scaling properties of the distribution (see Samorodnitsky and Taqqu (1994) and Kring et al. (2007) for a discussion of this issue). But before defining these processes, we have to introduce additional notions.

Definition 32. $(Z_t)_{t \in \mathbf{Z}}$ *is multivariate α-stable strict white noise if it is a series of independent and identically distributed α-stable sub-Gaussian random vectors with dispersion matrix Σ.*

An α-stable strict white noise process with mean μ and covariance matrix Σ will be denoted by α-SWN(μ,Σ). It can be shown easily that a dispersion matrix of an α-stable sub-Gaussian random vector has to be non-negative definite.

---

[10]See also Doganoglu, Hartz, and Mittnik (2006) for a multivariate model with conditionally varying and heavy-tailed risk factors.

Definition 33. *Let $(Z_t)_{t \in \mathbf{Z}}$ be α-stable strict white noise α-SWN$(0, \mathrm{Id})$. The process $(X_t)_{t \in \mathbf{Z}}$ is said to be an α-stable multivariate GARCH process if it is strictly stationary and satisfies equations of the form*

$$X_t = \Sigma_t^{1/2} Z_t, t \in \mathbf{Z},$$

where $\Sigma^{1/2}$ is the Cholesky factor of a positive-definite matrix Σ_t which is measurable with respect to $\mathcal{F}_{t-1}$.

As in Section 3.4, we take no account of the conditional mean vector for notational ease. It is usually specified as function of the past, through a vectorial autoregressive moving average (VARMA) representation.

Due to Kring et al. (2007, pp. 12-13), it follows immediately that we have

$$\Sigma_t^{1/2} Z_t | \mathcal{F}_{t-1} \sim E_d(0, \Sigma_t, \psi_{sub}(., \alpha)) \tag{3.12}$$

In order to shorten the notation we introduce the dispersion operator

$$\mathrm{Disp}(X) = \Sigma,$$

where X is an α-stable sub-Gaussian random vector with dispersion matrix Σ. In particular, we define by

$$\mathrm{Disp}(X_t | \mathcal{F}_{t-1}) = \Sigma_t$$

the conditional dispersion matrix of X_t given $\mathcal{F}_{t-1}$. This notion is well defined because of equation (3.12). Furthermore, we have

$$\mathrm{Disp}(a' X_t | \mathcal{F}_{t-1}) = a' \Sigma_t a,$$

since we have $a' X_t | \mathcal{F}_{t-1} \sim E_1(0, a' \Sigma_t a, \psi_{sub}(., \alpha))$ (see Kring et al. (2007, p. 13) or Samorodnitsky and Taqqu (1994, p. 77 et seq.)).

In Section 3.4 we argued that equation (3.4) holds at least approximately for many financial return time series possessing second moments. We can now repeat those arguments for processes with α-stable sub-Gaussian innovations. Hence, we obtain

$$\mathrm{Disp}(a' X_t | \mathcal{F}_{t-1}) \quad = \quad \mathrm{Disp}(a' X_t | \mathcal{F}_{t-1}(a))$$

holds at least approximately. We can now rephrase Definition 27.

Definition 34. *An α-stable multivariate GARCH process $(X_t)_{t \in \mathbf{Z}}$ is α-stable projection-*

efficient if it satisfies for all $t \in \mathbf{Z}$ and $a \in \mathbf{R}^d$

$$a'X_t|\mathcal{F}_{t-1} \stackrel{d}{=} a'X_t|\mathcal{F}_{t-1}(a) \sim E_1(0, a'\Sigma_t a, \psi_{sub}(., \alpha)),$$

where Σ_t is the conditional dispersion matrix.

According to Kring et al. (2007) Proposition 2 and Samorodnitsky and Taqqu (1994, p. 77 et seq.) the following equation holds for the scaling and dispersion parameter of $a'X$

$$\sigma(a) = \left(\frac{1}{2}a'\Sigma a\right)^{1/2} = (\text{Disp}(a'X))^{1/2}, \tag{3.13}$$

where $X \sim E_d(0, \Sigma, \psi_{sub}(., \alpha))$. Note that in the classical case where second moments exist, we have

$$\sigma(a) = (a\Sigma a)^{1/2}, \tag{3.14}$$

where $\sigma(a)$ and $(a\Sigma a)^{1/2}$ can be considered as the standard deviation and variance of $a'X$, respectively. In the α-stable case, we have to take the factor $1/2$ in the relation between scaling parameter and dispersion parameter of dX due to equation (3.13). In particular, we can write

$$E_1(0, a'\Sigma a, \psi_{sub}(., \alpha)) \text{ or } S_\alpha\left(\left(\frac{1}{2}a'\Sigma a\right)^{1/2}, 0, 0\right)$$

for the distribution of $d'X$. In addition, if $a'X_t|\mathcal{F}_{t-1}(a)$ is α-stable distributed, we denote the scaling parameter with $\sigma_t(a)$ and we can write

$$\sigma_t(a) = \left(\frac{1}{2}\text{Disp}(a'X_t|\mathcal{F}_{t-1}(a))\right)^{1/2}. \tag{3.15}$$

We restate Definition 28 in terms of α-stable power GARCH processes.

Definition 35. *An α-stable multivariate GARCH process $(X_t)_{t \in \mathbf{Z}}$ is called power-GARCH-projection-efficient if it is α-stable projection efficient and satisfies*

$$(\sigma_t(a))^\delta = \alpha_0 + \sum_{i=1}^p \alpha_i |a'X_{t-i}|^\delta + \sum_{j=1}^q \beta_j \sigma_{t-j}^\delta(a), \tag{3.16}$$

for all $t \in \mathbf{Z}$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, ..., r$, $\beta_j \geq 0$, $j = 1, ..., q$, $0 < \delta < \alpha$.

We have now all notions to define an α-stable version of the CMGARCH model.

Definition 36. *The process $(X_t)_{t \in \mathbf{Z}}$ follows an α-stable composed MGARCH process*

if it is a process with the general structure given in Definition 33 and the conditional dispersion matrix $\Sigma_t = (\sigma_{t,ij})$ satisfies

(i) *$(\sigma_t(e_i))_{t \in \mathbf{Z}} = ((\frac{1}{2}\sigma_{t,ii})^{1/2})_{t \in \mathbf{Z}}$ follows a $S_{\alpha,0,\delta}GARCH(p_i, q_i)$ process for $i = 1, ..., d$.*

(ii) *For all $i, j = 1, ..., d$ $i \neq j$ we have*

$$\sigma_{t,ij} = \frac{1}{2}(\sigma_t^2(e_i + e_j) - \sigma_t^2(e_i - e_j)),$$

where $(\sigma_t(e_i + e_j))_{t \in \mathbf{Z}}$ and $(\sigma_t(e_i - e_j))_{t \in \mathbf{Z}}$ follow a $S_{\alpha,0,\delta}GARCH(p_{ij}^+, q_{ij}^+)$ and $S_{\alpha,0,\delta}GARCH(p_{ij}^-, q_{ij}^-)$ processes, respectively.

An α-stable CMGARCH model does not impose any explicit functional form of the conditional dispersion matrix Σ_t. The next theorem shows which α-stable MGARCH processes can be modeled by α-stable CMGARCH processes.

Theorem 16. *Let $(X_t)_{t \in \mathbf{Z}}$ be an α-stable multivariate GARCH process which is power-GARCH-projection-efficient with conditional dispersion time series $(\Sigma_t)_{t \in \mathbf{Z}}$, then this time series can be modeled by an α-stable composed MGARCH process.*

Proof.

$$
\begin{aligned}
\sigma_{t,ij} \quad &= \quad \frac{1}{4}((e_i + e_j)\Sigma_t(e_i - e_j) - (e_i - e_j)\Sigma_t(e_i - e_j)) \\
&= \quad \frac{1}{4}(\text{Disp}((e_i + e_j)'X_t|\mathcal{F}_{t-1}) - \text{Disp}((e_i - e_j)'X_t|\mathcal{F}_{t-1})) \\
&\overset{(*)}{=} \quad \frac{1}{4}(\text{Disp}((e_i + e_j)'X_t|\mathcal{F}_{t-1}(e_i + e_j)) \\
&\qquad - \text{Disp}((e_i - e_j)'X_t|\mathcal{F}_{t-1}(e_i - e_j))) \\
&\overset{\text{equation (3.15)}}{=} \quad \frac{1}{4}(2\sigma_t^2(e_i + e_j) - 2\sigma_t^2(e_i - e_j)) \\
&= \quad \frac{1}{2}(\sigma_t^2(e_i + e_j) - \sigma_t^2(e_i - e_j))
\end{aligned}
$$

equation $(*)$ holds because the process is α-stable-projection-efficient. $\sigma_t^2(e_i + e_j)$ and $\sigma_t^2(e_i - e_j)$ are modeled by power-GARCH-processes. $\quad\square$

Unfortunately, we cannot rephrase Theorem 14 for an α-stable CMGARCH process. This is because of the fact that we do not know the unconditional distribution of X_t, $t \in \mathbf{Z}$, so we cannot ensure if the dispersion operator is well defined. But it seems reasonable to impose that the univariate $S_{\alpha,\beta,\delta}GARCH$ processes of an α-stable CMGARCH model should be strictly stationary, i.e., $\lambda_{\alpha,\beta,\delta} \sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j \leq 1$.

We show that an α-stable CMGARCH model is invariant under linear transformation. This is essential, since this result enables us to define consistently an α-stable factor composed MGARCH model.

Theorem 17. *Let $(X_t)_{t \in \mathbf{Z}}$ follow an α-stable GARCH-projection-efficient CMGARCH process. Then the CMGARCH process is invariant under linear transformation, i.e. the process $(Y_t)_{t \in \mathbf{Z}} = (FX_t)_{t \in \mathbf{Z}}$ follows an α-stable CMGARCH process in terms of the filtration $(\mathcal{G})_{t \in \mathbf{Z}}$, and we have*

$$\text{Disp}(Y_t|\mathcal{F}_{t-1}) = F\Sigma_t F' = \text{Disp}(Y_t|\mathcal{G}_{t-1}) \tag{3.17}$$

where $F \in \mathbf{R}^{k \times d}$, $k \in \mathbf{N}$, and $\mathcal{G}_t$ is the sigma field generated by $\sigma(\{Y_s : s \leq t\})$.

Proof. We show $\text{Disp}(FX_t|\mathcal{F}_{t-1}) = F\Sigma_t F' = \text{Disp}(FX_t|\mathcal{G}_{t-1})$. We know

$$\text{Disp}(a'X_t|\mathcal{F}_{t-1}) \quad = \quad \text{Disp}(a'FX_t|\mathcal{F}_{t-1}(a'F)) = a'F\Sigma_t F'a. \tag{3.18}$$

for all $a \in R^d$. Hence, by using the characteristic function ψ_{sub} of an α-stable sub-Gaussian distribution, we conclude

$$
\begin{aligned}
\psi_{sub}(x^2(a'F\Sigma_t F'a), \alpha) &= E(e^{ix(a'FX_t)}|\mathcal{F}_{t-1}) \\
&= E(e^{ix(a'FX_t)}|\mathcal{F}_{t-1}(a'F))
\end{aligned}
$$

for all $a \in \mathbf{R}^d$ and $x \in \mathbf{R}$. Since we have $\mathcal{F}_t(a'F) \subset \mathcal{G}_t \subset \mathcal{F}_t$ for all $a \in \mathbf{R}^d$ and $t \in \mathbf{Z}$, we obtain

$$\psi_{sub}(x^2(a'F\Sigma_t F'a), \alpha) \quad = \quad E(e^{ix(a'FX_t)}|\mathcal{G}_{t-1}).$$

Since $a \in \mathbf{R}^d$ and $x \in \mathbf{R}^d$ are arbitrary, we follow

$$\psi_{sub}(s'F\Sigma_t F's, \alpha) \quad = \quad E(e^{is'(FX_t)}|\mathcal{G}_{t-1})$$

for all $s \in \mathbf{R}^d$. Hence, we can conclude

$$FX_t|\mathcal{G}_{t-1} \sim E_k(0, F\Sigma_t F', \psi_{sub}(., \alpha))$$

and we have $\text{Disp}(FX_t|\mathcal{F}_{t-1}) = F\Sigma_t F' = \text{Disp}(FX_t|\mathcal{G}_{t-1})$. We write for $F\Sigma_t F'$ shortly Σ_t^Y. Since $FX_t|\mathcal{G}_{t-1}$ is sub-Gaussian, we can write

$$
\begin{aligned}
\sigma_{t,ij}^Y &= \frac{1}{4}(\text{Disp}((e_i + e_j)'FX_t|\mathcal{G}_{t-1}) - \text{Disp}((e_i - e_j)'FX_t|\mathcal{G}_{t-1}) \\
&= \frac{1}{2}(\sigma_t^2((e_i + e_j)'F) - \sigma_t^2((e_i - e_j)'F),
\end{aligned}
$$

where $(\sigma_t((e_i + e_j)'F))_{t \in \mathbf{Z}}$ and $(\sigma_t((e_i - e_j)'F))_{t \in \mathbf{Z}}$ follow $S_{\alpha,0,\delta}GARCH(p_{ij}^+, q_{ij}^+)$ and $S_{\alpha,0,\delta}GARCH(p_{ij}^-, q_{ij}^-)$. Hence, we have demonstrated that $(Y_t)_{t \in \mathbf{Z}}$ follows an α-stable CMGARCH process. $\qquad\square$

We can consistently define an extension of the α-stable composed multivariate GARCH model, what we label the α-stable factor composed multivariate GARCH model.

Definition 37. *The process $(X_t)_{t \in \mathbf{Z}}$ follows an α-stable factor composed MGARCH (α-FCMGARCH) process, if there exists some orthogonal matrix $\Gamma \in \mathbf{R}^{d \times d}$ satisfying $\Gamma\Gamma' = \mathrm{Id}$ such that $(\Gamma'X_t)_{t \in \mathbf{Z}}$ follows an α-stable composed MGARCH process.*

As with the FCMGARCH model, the α-stable version allows for statistical factor modeling and dimensionality reduction. Again, we estimate the sample dispersion matrix of the process $(X_t)_{t \in \mathbf{Z}}$ and model its principal components $(Y_t)_{t \in \mathbf{Z}}$ by an α-stable FCMGARCH model. In contrast to a α-stable version of the PC-GARCH model, we have the advantage that we can capture the conditional dependence of the components.

As with Proposition 6 in Section 3.3, we have to estimate $(p+q+1)d^2$ parameters for a d-dimensional α-stable CMGARCH process whose projective time series follow α-stable power-GARCH(p, q) processes.

3.5.3 Estimation of the Models

In principle, we can employ the same two estimation procedures as presented in Section 3.4.2. In the first step, we have to use algorithms fitting α-stable power-GARCH processes to data in both algorithms.

In the first algorithm we reconstruct the conditional dispersion matrix by

$$\begin{aligned}
\hat{\Sigma}_{t,ii} &= 2\hat{\sigma}_t^2(e_i), \\
\hat{\Sigma}_{t,ij} &= \frac{1}{2}(\hat{\sigma}_t^2(e_i + e_j) + \hat{\sigma}_t^2(e_i - e_j)).
\end{aligned}$$

In the second approach we apply the univariate α-stable process $(\sigma_t(u_i))_{t \in \mathbf{Z}}$, $u_i \in \mathcal{S}^{d-1}$, $i = 1, ..., n$ to reconstruct the conditional dispersion. This leads to the following optimization problem.

$$\hat{\Sigma}_t = \mathrm{argmin}_{\Sigma \in S^{d \times d}} \sum_{i=1}^{d} (u_i'\Sigma u_i - 2\hat{\sigma}_t^2(u_i))^2$$

in step (2) and

$$\hat{\Sigma}_t = \mathrm{argmin}_{\Delta \in D_d} \sum_{i=1}^{n} (u_i'\Delta\Delta' u_i - 2\hat{\sigma}_t^2(u_i))^2$$

in step (3).

When fitting α-stable power-GARCH processes to the projective time series $(u_i'X_t)_{t\in\mathbf{Z}}$, $i=1,...,n$ via the standard maximum likelihood method, we need a global tail parameter α in order to be consistent with the model specification. The simplest way to obtain such an α is to estimate the unconditional tail parameter $\alpha(u_i)$ of the projective time series $(u_i'X_t)_{t\in\mathbf{Z}}$, $i=1,...,n$. Then, the global tail parameter α is defined by

$$\alpha = \frac{1}{n}\sum_{i=1}^{n}\alpha(u_i).$$

This is certainly a very heuristic method, since we do not estimate the global tail parameter of the innovations Z_t but of the returns X_t. But since we estimate the parameter $\alpha_0(u_i),...,\alpha_r(u_i)$ and $\beta_1(u_i),...,\beta_s(u_i)$ of the power-GARCH(r,s)-process $(u_i'X_t)_{t\in\mathbf{Z}}$ via the classical ML-method, these estimates are robust under misspecification of the tail parameter α. For larger sample sizes, $u_i'X_1,...,u_i'X_{t_0}$ (t_0 large) the estimates for the scale parameters and the tail parameter are nearly independent (see DuMouchel (1973) for further information). Thus the estimates of the power-GARCH parameters are nearly independent of the tail parameter α and hence the time series of scale parameters $(\sigma_t(u_i))_{t\in\mathbf{Z}}$.

In addition, after having estimated the time series of conditional dispersion matrices $(\hat{\Sigma}_t)_{t\in\mathbf{Z}}$ we can use the residuals

$$\hat{Z}_t = \hat{\Sigma}_t^{-1/2}X_t$$

to estimate the tail parameter of the innovations where $\hat{\Sigma}_t^{1/2}$ is the inverse of the Cholesky factor of $\hat{\Sigma}_t$.

3.6   Applications

For the empirical analysis of the α-stable CMGARCH model, we investigated the daily logarithmic return series for four German stocks included in the DAX index: Adidas, Allianz, Altana, and BASF. The period analyzed is January 2, 2001 through March 31, 2006 (1,338 daily observations for each stock). For the estimation of our model, we selected the first $1,000$ returns i.e. the period form January 2, 2001 until December 7, 2004. The balance of the observed returns are held out for an out-of-sample analysis of the model.

The plots of the individual returns series in the estimation period for the four stocks are shown in Figure 3.1. One can easily detect times of intense and less pronounced volatility which is to be attributed to the well-known effect of volatility clustering. In Table 3.2 the maximum likelihood estimates of the four return time series are listed.
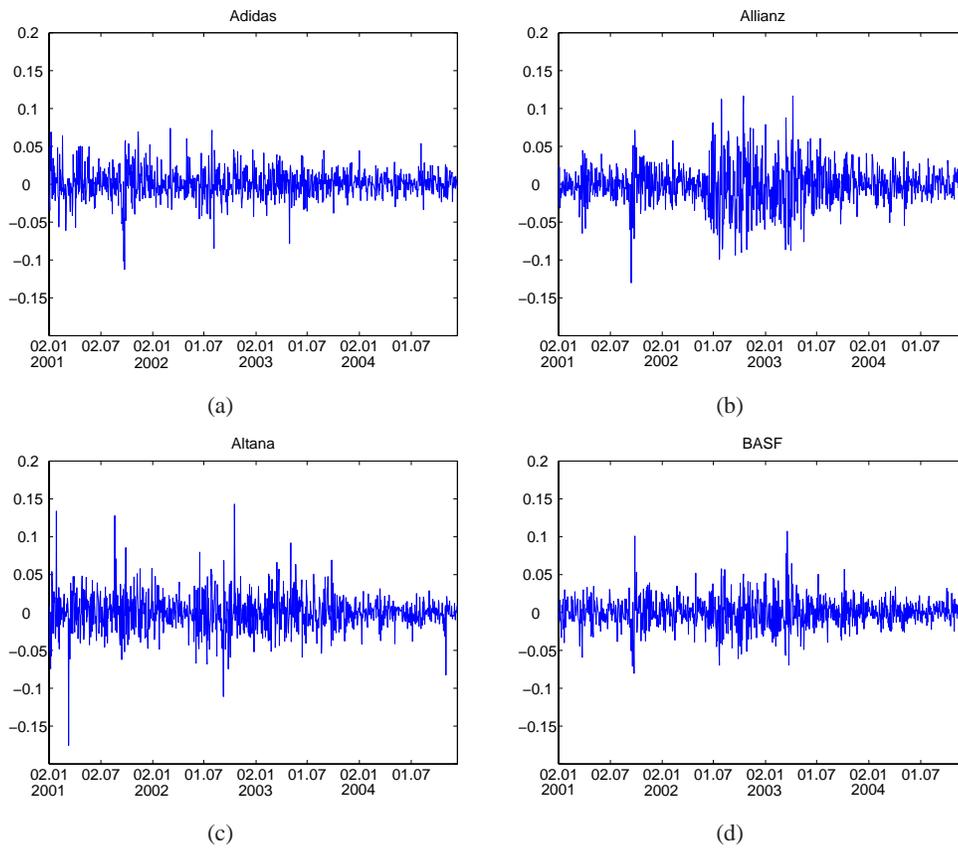
Figure 3.1: 1,000 daily returns during the period from January 2, 2001 until December 7, 2004.

The tail parameters α appear to be in a tight range around 1.69. Hence, assuming the same parameter α is justifiable. The scale parameter is within in a range of roughly 0.1 and 0.15. The location parameters μ are close to zero.

We assume that the four univariate return series follow $S_{\alpha,0,1}GARCH(1,1)$ processes. More precisely,

$$\sigma_t(e_i) = \alpha_0(e_i) + \alpha_1(e_i)|X_{t-1} - \mu(e_i)| + \beta_1(e_i)\sigma_{t-1}(e_i)$$

and

$$X_{t,i} = \mu(e_i) + \sigma_t(e_i)Z_t,$$

where $Z_t \sim S_\alpha(1,0,0)$ and $\mu(e_i)$ is the unconditional mean of the ith time series, $i = 1,...,4$ and $\alpha = 1.69$.[11] The estimated parameters based on the period from January 2, 2007 until December 7, 2007 are reported in the right half of Table 3.2. As

---

[11] For daily log-returns it is not necessary to model the daily mean process $(\mu_t)_{t \in \mathbf{z}}$ (see RiskMetrics (1996)).

Stock	α	σ	μ	α_0	α_1	β_1	LL
Adidas	1.667	0.0109	0.0001	0.0002	0.0413	0.9316	2617.30
Allianz	1.6672	0.0155	-0.0011	0.0002	0.0841	0.8854	2356.2
Altana	1.7337	0.0138	-0.0003	0.0004	0.0706	0.8837	2423.4
BASF	1.6992	0.011	0.0001	0.0003	0.0661	0.8966	2650.6

Table 3.2: The left half of the table depicts the unconditional stable estimates of the returns time series. The right half shows the estimated parameters of the univariate stable GARCH(1,1) processes for the returns. The time period is January 2, 2001 to December 7, 2004.

discussed in Section 3.5.3, a misspecification of the tail parameter α has only a minor influence on the estimated power-GARCH parameters. The absolute mean $\lambda_{\alpha,0}$ of a centered unit scale variable $Z_t \sim S_\alpha(1, 0, 0)$ is given by

$$\lambda_{\alpha,0} = \frac{2}{\pi}\Gamma\left(1 - \frac{1}{\alpha}\right) = 1.38.$$

Due to Proposition 7, all processes are strict stationary since we have $\alpha_0(e_i)$, $\alpha_1(e_i)$, $\beta_1(e_i) > 0$, and $\lambda_{\alpha,0}\alpha_1(e_i) + \beta_1(e_i) < 1$, $i = 1, ..., 4$. This purely univariate analysis of the return data does not reveal any contradiction to an α-stable CMGARCH modeling.

3.6.1 In-Sample Analysis of the α-stable CMGARCH Model

In the following we assume that the return data $X_1, ..., X_{1000}$ of the four stocks in our study are power-GARCH-projection-efficient with $\delta = 1$ and follow an α-stable CM-GARCH model. In order to estimate the α-stable CMGARCH process we generate random vectors $u_i \in \mathcal{S}^3$, $i = 1, ..., 100$, that are uniformly distributed on $\mathcal{S}^3$. The projective time series $(u_i'X_t)_{t\in\mathbf{Z}}$ follow again an α-stable power-GARCH process. According to the estimation procedure described in Sections 3.4.2 and 3.5.3, we have to estimate the parameters $\alpha_0(u_i)$, $\alpha_1(u_i)$, and $\beta_1(u_i)$ of the corresponding power-GARCH$(1, 1)$-processes. Figure 3.2 (a), (b), and (c) illustrate the estimates. The parameter estimates $\hat{\beta}_1(u_i)$ are tightly scattered around 0.91; $\hat{\alpha}_1(u_i)$ ranges from 0.038 to 0.082 and for $\hat{\alpha}_0(u_i)$ from $0.7 \cdot 10^{-4}$ to $3 \cdot 10^{-4}$. All projective time series $(u_i'X_t)_{t\in\mathbf{Z}}$ have a low market reaction ($\hat{\alpha}_1(u_i)$ small) but a high persistence ($\hat{\beta}_1(u_i)$ high).

In particular, we see that each projective time series is strict stationary. Figure 3.4 depicts the 100 corresponding time series $(\sigma_t(u_i))$ of the conditional, time-varying scale parameters. The effects of volatility clustering can be seen.

In order to obtain the time series of the conditional dispersion matrices $\Sigma_t \in \mathbf{R}^{4\times 4}$, we have to apply steps (2) and (3) of the estimation algorithm given in Sections 3.4.2 and 3.5.3, respectively. Figure 3.2 (d) shows the number of not positive definite ma-

(a)

(b)

(c)
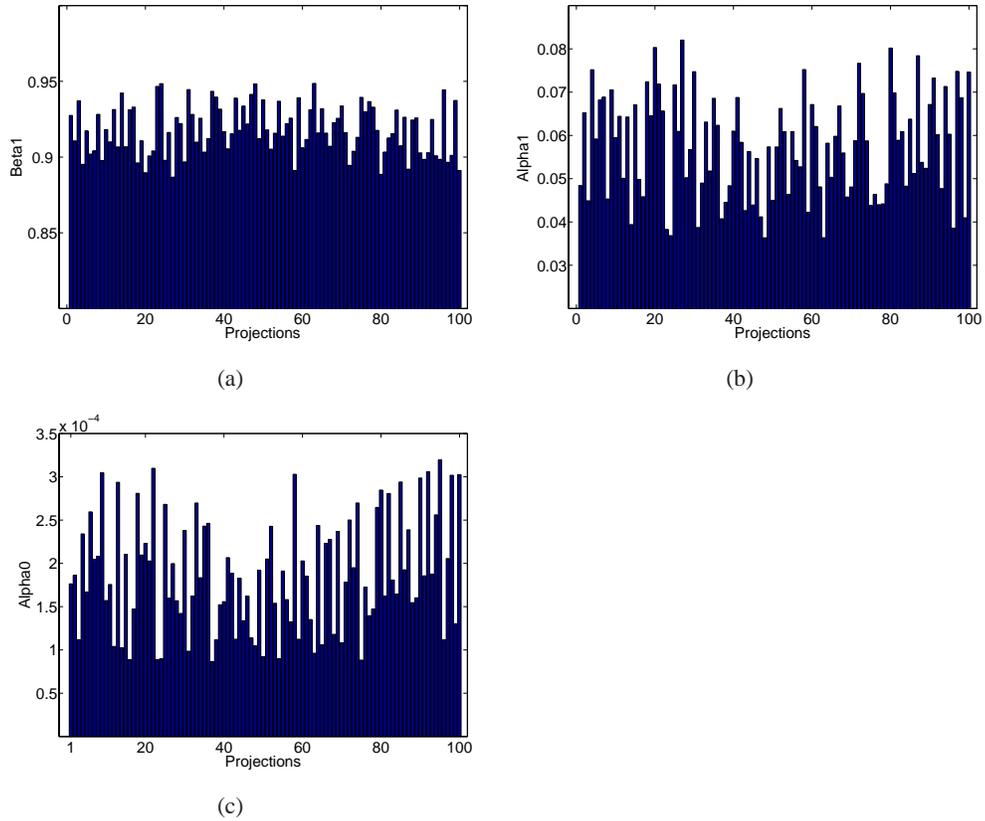
Figure 3.2: (a),(b),(c) show the stable GARCH(1,1) estimates for the 100 projective time series $(u_i' X_t)_{t \in I}$.
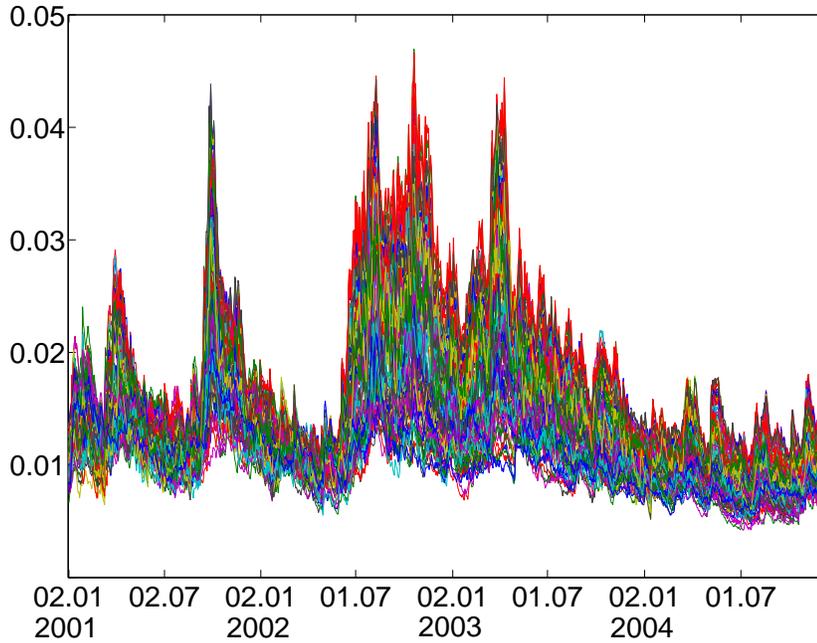
Figure 3.3: The figure shows the 100 different time series $(\sigma_t(u_i))_{t \in I}$, $i = 1, ..., 100$, of the stable GARCH$(1, 1)$-Processes.

trices Σ_t, $t = 1, ..., 1000$, obtained after applying step (2) subject to the number of projective time series $(\sigma_t(u_i))_{t \in \mathbf{Z}}$ used in the regression. By increasing the number of projections in step (2), the number of these matrices that exhibit this characteristic decreases fast. We notice the last not positive definite matrix when applying 35 projections. In the range from 36 until 100 projections, all matrices are positive definite. Furthermore, we observed a fast stabilization of the entries of the time series (Σ_t) subject to the number of projections used in step (2). This observation supports the assumption that the considered multivariate time series is GARCH-projection efficient.

In order to obtain a very high accuracy of our estimates, we use 100 projections for the reconstruction of the time series (Σ_t). In particular, we do not need to apply the optional step (3) since all conditional dispersion matrices are positive definite.

Figure 3.5 shows the 2-dimensional scatterplots between the different returns pairs. We find that Adidas and Altana as well as Altana and BASF exhibit a very low cross dispersion due to the scatterplots, while the ones of BASF-Allianz and Adidas-BASF illustrate stronger cross dispersion. Figure 3.6 depicts the time series of the conditional dispersion matrices $(\Sigma_t)_{t=1,...,1000}$ for the period January 2, 2001 to December 7, 2004.

In particular, the conditional dispersions $(\sigma_{t,13})$ and $(\sigma_{t,34})$ corresponding to Adidas-Altana and Altana-BASF are low, consistent with the observation made about Figure 3.5. Definitely the highest conditional dispersion can be observed between Allianz
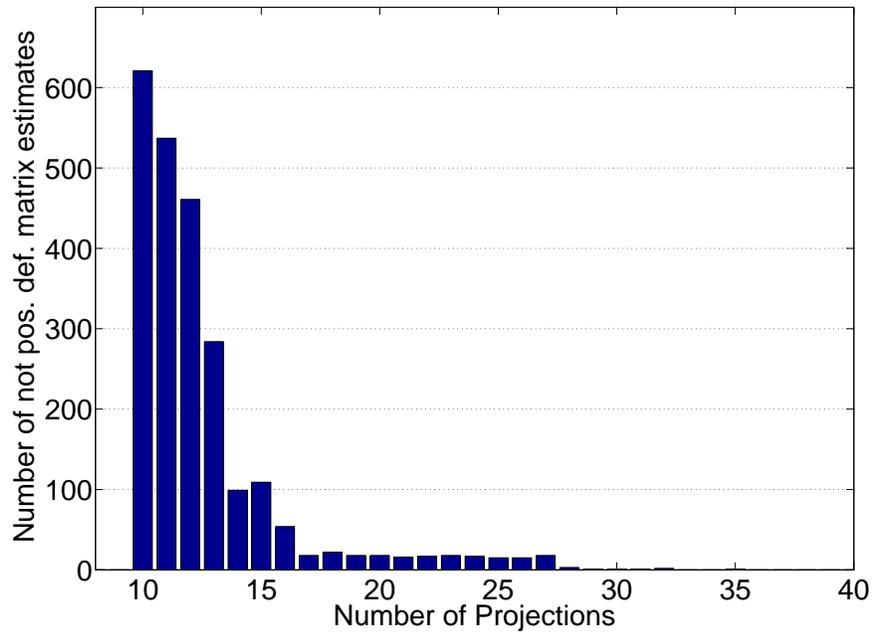
Figure 3.4: The figure illustrates the number of not positive definite matrices subject to the number of projection used in the regression.
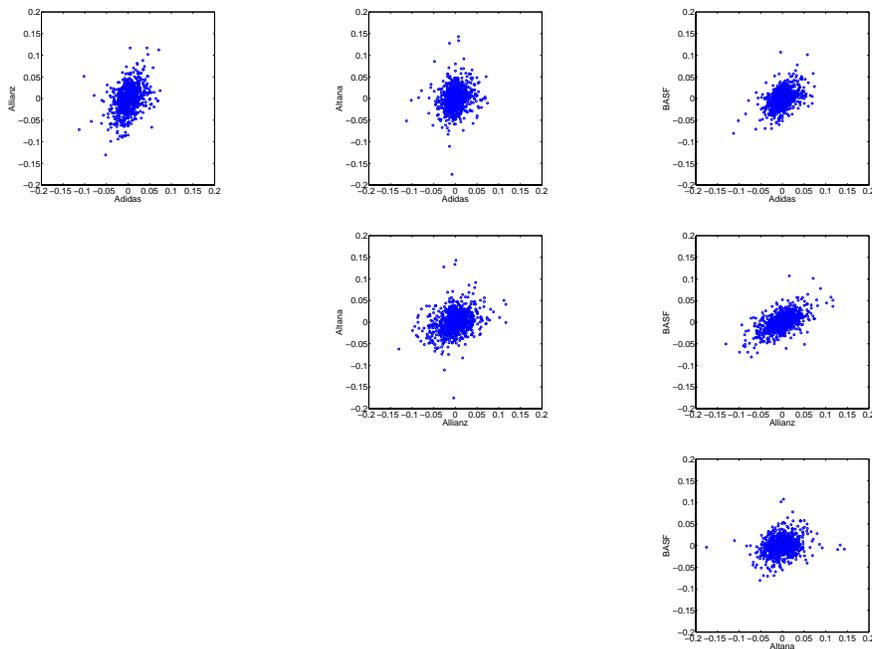

Figure 3.5: Two dimensional scatterplots of the returns in the period January 2, 2001 until December 7, 2004.
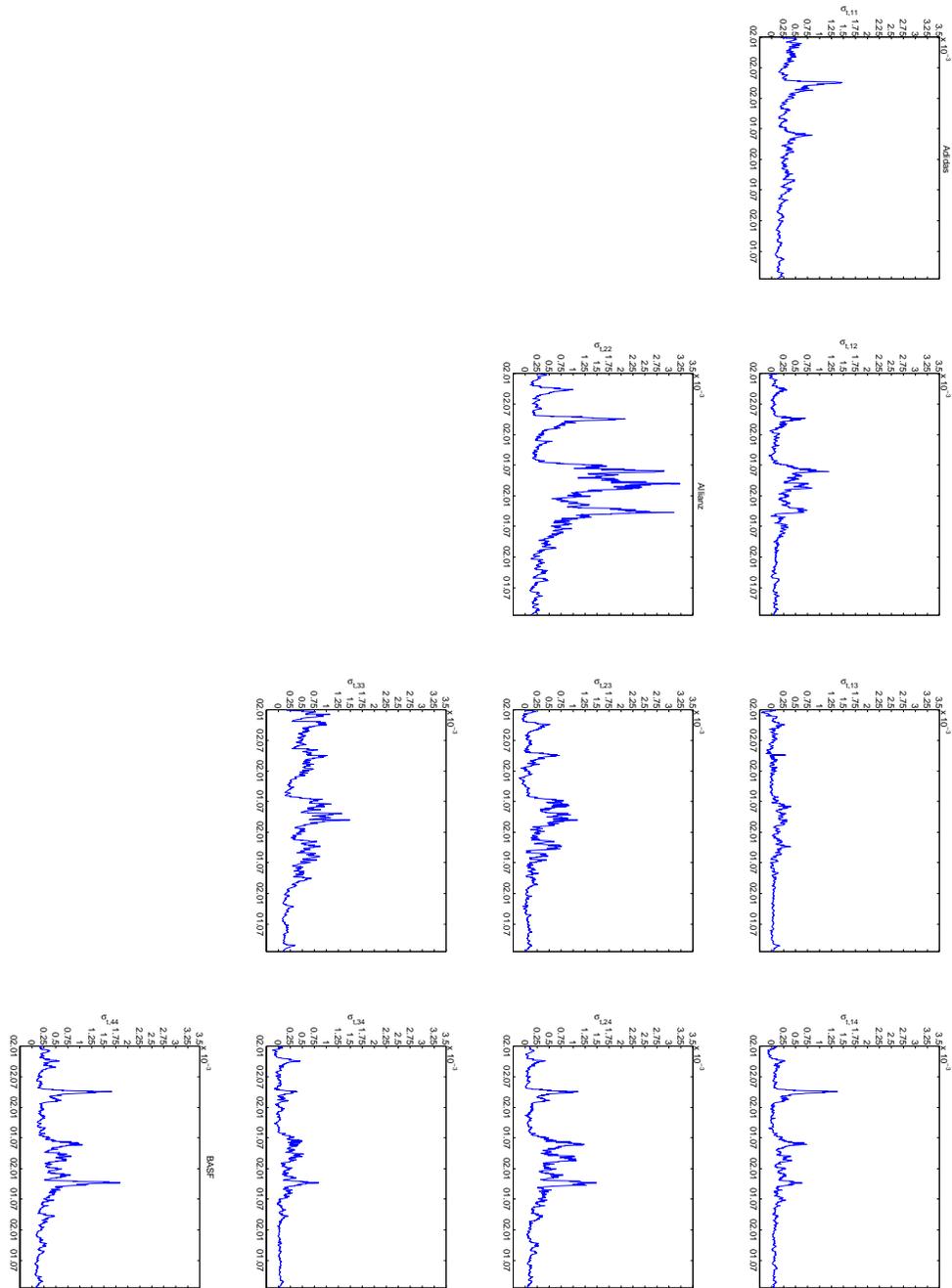
Figure 3.6: The figure depicts the estimated conditional dispersion matrices Σ_t in the period January 2, 2001 until December 7, 2004.

and BASF, $(\sigma_{t,34})$, which is also the case in the unconditional graphical consideration of Figure 3.5. For the time period investigated, the returns of Allianz are the most volatile especially in the period July 2002 until July 2003. Besides Altana, we observe an increase in the conditional dispersion and cross dispersion in all stocks because of September 11, 2001.

For a quantitative analysis of the α-stable CMGARCH model we examine its residuals given by

$$\hat{Z}_t = \hat{\Sigma}_t^{-1/2}(X_t - \hat{\mu}),$$

where $t = 1, ..., 1000$, $\hat{\mu} \in \mathbf{R}^4$ the unconditional mean of the four individual return time series, and $\hat{\Sigma}_t^{-1/2}$ the inverse of the Cholesky factor of $\hat{\Sigma}_t$. To test whether the generated innovations are strict white noise (SNW(0,Id)) (see Section 3.3 and Definition 32), we estimate their unconditional dispersion matrix $\hat{\Sigma}_Z$ by using the spectral estimator.[12] The spectral estimator is a robust estimator of the dispersion and covariance matrix up to a scaling constant. To have a unique dispersion matrix we demand $\hat{\sigma}_{11}$ to be 1.

For comparison, we first list the normalized dispersion matrix of the original returns given by

$$\hat{\Sigma}_0(X_1, ..., X_{1000}) = \begin{pmatrix} 1.0000 & 0.5294 & 0.1929 & 0.3724 \\ 0.5294 & 1.8865 & 0.5637 & 0.8195 \\ 0.1929 & 0.5637 & 1.4942 & 0.3121 \\ 0.3724 & 0.8195 & 0.3121 & 0.9499 \end{pmatrix}.$$

One can clearly detect the non-zero cross dispersion between the stock returns, because $\hat{\sigma}_{ij}(X_1, ..., X_{1000})$, $i, j = 1, ..., 4, i \neq j$, deviate significantly from zero. Moreover, the diagonal entries reveal non-standardized quantities, since $\hat{\sigma}_{ii}(X_1, ..., X_{1000})$, $i = 1, ..., 4$, differ significantly from 1. However, the residuals $\hat{Z}_1, ..., \hat{Z}_{1000}$ of the α-stable CMGARCH model yield the unconditional normalized dispersion matrix

$$\hat{\Sigma}_0(\hat{Z}_1, ..., \hat{Z}_{1000}) = \begin{pmatrix} 1.0000 & 0.0381 & -0.0223 & 0.0046 \\ 0.0381 & 1.0392 & 0.0305 & 0.0440 \\ -0.0223 & 0.0305 & 1.1323 & 0.0146 \\ 0.0046 & 0.0440 & 0.0146 & 1.0991 \end{pmatrix},$$

which indicates that the scales of the individual stocks are close to one and zero, respectively. In addition, the cross dispersion between the returns are definitely minimized if not removed.

---

[12]See Tyler (1987a) for further information about the spectral estimator.

We compare these residuals with those of the EWMA model introduced by Risk-Metrics.[13] The EWMA model is still the industry standard for multivariate conditional risk modeling (see the discussion in Section 3.3 and RiskMetrics (1996)). In the EWMA updating scheme

$$\Sigma_t = \frac{1-\lambda}{1-\lambda^{M+1}} \sum_{i=1}^{M} \lambda^i X_{t-i} X'_{t-i}$$

we choose M to be $112, 50$, and 20. For daily returns, RiskMetrics (1996) recommends an optimal decay factor $\lambda = 0.94$. By using $\lambda = 0.94$, 99.9% of the information is contained in the last 112 days and the classical RiskMetrics updating scheme

$$\Sigma_t = (1-\lambda)X_{t-1}X'_{t-1} + \lambda\Sigma_{t-1},$$

is captured very well. In this case, the residuals are denoted by $\hat{Z}_1^{E(112)}, ..., \hat{Z}_{1000}^{E(112)}$ and their unconditional normalized dispersion matrix satisfies

$$\hat{\Sigma}_0(\hat{Z}_1^{E(112)}, ..., \hat{Z}_{1000}^{E(112)}) = \begin{pmatrix} 1.0000 & 0.1562 & -0.2340 & -0.2519 \\ 0.1562 & 1.1071 & -0.1848 & -0.2198 \\ -0.2340 & -0.1848 & 1.5471 & 0.6278 \\ -0.2519 & -0.2198 & 0.6278 & 1.8321 \end{pmatrix}.$$

We see that the diagonal elements of this matrix deviate significantly from one and cross-dispersion is definitely not zero. It is obvious that the results of the α-stable CMGARCH are superior. Incorporating the last 112 might be too much.

In our statistical analysis we test different values of M (i.e., $M = 10, 20, ..., 100$). We obtain the best results for $M = 20$. In this case, the dispersion matrix of the residuals satisfies

$$\hat{\Sigma}_0(\hat{Z}_1^{E(20)}, ..., \hat{Z}_{1000}^{E(20)}) = \begin{pmatrix} 1.0000 & 0.0408 & -0.0162 & 0.0335 \\ 0.0408 & 1.2124 & -0.0036 & -0.0287 \\ -0.0162 & -0.0036 & 1.2769 & 0.0647 \\ 0.0335 & -0.0287 & 0.0647 & 1.5789 \end{pmatrix}.$$

The cross dispersions are similar to those in the α-stable CMGARCH model but the diagonal elements of the normalized dispersion matrix differ significantly from one. This behavior is expected since we know from univariate GARCH- and EWMA-modeling that volatility processes of univariate return series are captured much better by a GARCH process. Moreover, we have one decay factor $\lambda = 0.94$ that should be valid for all stocks simultaneously, which is not realistic. In the case of $M = 50$, we

---

[13]See Definition 23.

Stock	α	Disp	σ
Adidas	1.76	0.94	0.68
Allianz	1.92	1.12	0.75
Altana	1.8	1.11	0.74
BASF	1.9	1.18	0.77

Table 3.3: The table depicts the stable estimates of the α-stable CMGARCH residuals.

obtain

$$\hat{\Sigma}_0(\hat{Z}_1^{E(50)}, ..., \hat{Z}_{1000}^{E(50)}) = \begin{pmatrix} 1.0000 & 0.0680 & -0.0406 & -0.0480 \\ 0.0680 & 1.1271 & 0.0404 & -0.0197 \\ -0.0406 & 0.0404 & 1.1219 & 0.0771 \\ -0.0480 & -0.0197 & 0.0771 & 1.7043 \end{pmatrix},$$

which does not significantly differ from the case $M = 20$.

The maximum likelihood estimates of the residuals $\hat{Z}_1, ..., \hat{Z}_{1000}$ are depicted in Table 3.3. We see that the innovations have a larger tail parameter than the unconditional returns depicted in Table 3.2. This phenomenon is to be expected because the leptokurtosis in the unconditional distribution of the process $(X_t)_{t \in \mathbf{Z}}$ is attributed to the GARCH structure of the process (see RiskMetrics (1996) and McNeil, Frey, and Embrechts (2005) for further information). By removing the MGARCH effects from the data, we decrease the leptokurtosis and thereby increase the tail parameter.

In particular, $\hat{Z}_{t,1}, ..., \hat{Z}_{t,1000}$ has a scale parameter of 0.9440, thus the residuals have an estimated dispersion matrix satisfying

$$\hat{\Sigma}(\hat{Z}_1, ..., \hat{Z}_{1000}) = \begin{pmatrix} 0.9440 & 0.0359 & -0.0211 & 0.0044 \\ 0.0359 & 0.9810 & 0.0288 & 0.0416 \\ -0.0211 & 0.0288 & 1.0689 & 0.0138 \\ 0.0044 & 0.0416 & 0.0138 & 1.0375 \end{pmatrix},$$

that is close the identity. However, there is a difference between the estimates $\hat{\sigma}_{ii}(Z)$ of the spectral estimator and those of the univariate estimates depicted in Table 3.3. We have greater confidence in the spectral estimates for two reasons. First, the spectral estimator uses a four dimensional sample for its estimates whereas in the other case we use only the univariate time series. Second, the spectral estimator estimates $\hat{\sigma}_{ii}$, $i = 1, ..., 4$, immediately, whereas in the other method we estimate the scale parameter $\hat{\sigma}(e_i)$ and then we calculate the dispersion by the formula $\hat{\sigma}_{ii} = 2\hat{\sigma}(e_i)^2$ increasing the estimation error.

In order to complete our sample analysis, we analyze the autocorrelation of the squared returns and squared residuals in the different models. While one might cor-

rectly object that due to the model specification second moments of the residuals and returns do not exist, nevertheless the estimators of the autocorrelation function (ACF) have distributions that have lower and upper confidence bounds under the independent and identically distributed (i.i.d.) assumption. The range of these bounds is larger than the ones in the case where second moments exist. Hence, if the residuals are within the normal bounds, they are also in model specific confidence bounds. The results are reported in Table 3.4. The lower and upper 95% confidence bounds are -0.0632 and 0.0632, respectively. The squared returns significantely violate these bounds (29 violations) and the hypothesis of independence can be rejected. In the case of the α-stable CMGARCH residuals we find three violations of these bounds. In the classical EWMA model ($M = 112$) there are eight violations. The EWMA model using the 20 last observations only violates these bounds twice and for $M = 50$ we observe five values out of these bounds. All the models work well and remove a lot of autocorrelation in the squared return data. The α-stable CMGARCH model seems to be superior to the classical EWMA model since the estimates are closer to zero in most cases and there are less violations of the confidence bounds.

3.6.2   Out-Of-Sample Analysis of the α-stable CMGARCH Model

For the out-of-sample analysis we use the period from December 8, 2004 to March 31, 2006 (observations $1,001$ to $1,338$). The normalized dispersion matrix of the observed returns satisfies

$$\hat{\Sigma}_0(X_{1001},...,X_{1338}) = \begin{pmatrix} 1.0000 & 0.3756 & 0.0896 & 0.3487 \\ 0.3756 & 1.1663 & 0.2344 & 0.5109 \\ 0.0896 & 0.2344 & 0.7438 & 0.1429 \\ 0.3487 & 0.5109 & 0.1429 & 0.9180 \end{pmatrix}$$

and the α-stable CMGARCH residuals is given by

$$\hat{\Sigma}_0(\hat{Z}_{1001},...,\hat{Z}_{1338}) = \begin{pmatrix} 1.0000 & -0.0419 & 0.0954 & -0.0118 \\ -0.0419 & 1.0558 & -0.0354 & 0.0240 \\ 0.0954 & -0.0354 & 0.9994 & 0.0025 \\ -0.0118 & 0.0240 & 0.0025 & 1.1954 \end{pmatrix}.$$

The normalized dispersion matrix of the returns once again exhibit a significant cross dispersion. Furthermore, the diagonal entries of the normalized dispersion matrix are not close to one, suggesting that the univariate return series exhibit different scale properties. In contrast, the normalized dispersion matrix of the CMGARCH residuals are much closer to the identity matrix. This means that the forecasted conditional

Stock	1	2	3	4	5	6	7
$ACF((X_{t,1}^2))$	0.2	0.305	0.119	0.188	0.159	0.116	0.083
$ACF((X_{t,2}^2))$	0.236	0.232	0.23	0.23	0.353	0.19	0.233
$ACF((X_{t,3}^2))$	0.046	0.016	0.082	0.012	0.028	0.004	0.021
$ACF((X_{t,4}^2))$	0.176	0.246	0.263	0.355	0.133	0.2	0.13
$ACF((X_{t,1}^2))$	0.2	0.305	0.119	0.188	0.159	0.116	0.083
$ACF((\hat{Z}_{t,1}^2))$	0.028	0.054	0.004	0.046	0.041	0.04	-0.03
$ACF((\hat{Z}_{t,2}^2))$	0.101	0.045	0.073	0.059	0.081	-0.006	-0.015
$ACF((\hat{Z}_{t,3}^2))$	0.002	-0.009	0.033	-0.015	-0.009	-0.018	-0.021
$ACF((\hat{Z}_{t,4}^2))$	-0.009	0.01	0.042	-0.003	0.083	0.001	-0.018
$ACF((\hat{Z}_{t,1}^{2,E(112)}))$	0.043	0.058	0.012	0.066	0.059	0.098	-0.004
$ACF((\hat{Z}_{t,2}^{2,E(112)}))$	0.155	0.067	0.137	0.12	0.14	0.03	0.023
$ACF((\hat{Z}_{t,3}^{2,E(112)}))$	0.009	-0.016	0.062	0.01	0	0.004	-0.007
$ACF((\hat{Z}_{t,4}^{2,E(112)}))$	0.049	0.024	0.095	0.032	0.072	0.037	-0.008
$ACF((\hat{Z}_{t,1}^{2,E(20)}))$	0.011	0.003	-0.018	0.013	0.018	0.044	-0.037
$ACF((\hat{Z}_{t,2}^{2,E(20)}))$	0.097	0.007	0.026	0.028	0.051	-0.014	0.01
$ACF((\hat{Z}_{t,3}^{2,E(20)}))$	-0.01	-0.012	0.005	-0.008	-0.01	-0.014	-0.015
$ACF((\hat{Z}_{t,4}^{2,E(20)}))$	0.038	-0.018	0.041	0.021	0.033	-0.016	-0.006
$ACF((\hat{Z}_{t,1}^{2,E(50)}))$	0.023	0.034	-0.001	0.05	0.038	0.082	-0.022
$ACF((\hat{Z}_{t,2}^{2,E(50)}))$	0.128	0.048	0.086	0.079	0.1	0.018	0.045
$ACF((\hat{Z}_{t,3}^{2,E(50)}))$	0.003	-0.024	0.054	0.006	-0.008	-0.003	-0.014
$ACF((\hat{Z}_{t,4}^{2,E(50)}))$	0.036	0.01	0.077	0.018	0.059	0.027	-0.014

Table 3.4: The table depicts the autocorrelation of the squared returns and the squared residuals in the different models. The lower and upper 95% confidence bounds are -0.0632 and 0.0632, respectively. The period covered is January 2, 2001 to December 7, 2004.

dispersion matrix explains fairly well the common scaling properties of the returns.

Again, comparing the model with the classical EWMA model ($M = 112$) we see that the cross dispersion is explained well by the model but exhibits weakness on the diagonal entries

$$
\hat{\Sigma}_0(\hat{Z}_{1001}^{E(112)}, ..., \hat{Z}_{1338}^{E(112)}) = \begin{pmatrix} 1.0000 & 0.0335 & 0.0576 & 0.0555 \\ 0.0335 & 1.3439 & -0.0199 & -0.0647 \\ 0.0576 & -0.0199 & 0.8381 & 0.0503 \\ 0.0555 & -0.0647 & 0.0503 & 1.4686 \end{pmatrix}.
$$

The two alternative EWMA-models ($M = 20, 50$) show similar behavior: The off-diagonal entries are close to zero while the entries on the diagonal exhibit poor behavior. In particular, the standardized dispersion matrices of these residuals satisfy

$$
\hat{\Sigma}_0(\hat{Z}_{1001}^{E(20)}, ..., \hat{Z}_{1338}^{E(20)}) = \begin{pmatrix} 1.0000 & -0.0106 & 0.0980 & 0.0122 \\ -0.0106 & 1.3794 & -0.0501 & -0.0509 \\ 0.0980 & -0.0501 & 1.2486 & 0.0194 \\ 0.0122 & -0.0509 & 0.0194 & 1.8280 \end{pmatrix}
$$

and

$$
\hat{\Sigma}_0(\hat{Z}_{1001}^{E(50)}, ..., \hat{Z}_{1338}^{E(50)}) = \begin{pmatrix} 1.0000 & 0.0260 & 0.0671 & 0.0495 \\ 0.0260 & 1.3679 & -0.0200 & -0.0717 \\ 0.0671 & -0.0200 & 0.9335 & 0.0309 \\ 0.0495 & -0.0717 & 0.0309 & 1.5179 \end{pmatrix}.
$$

As already mentioned, this is due to the well-known univariate phenomenon that the volatility structure of univariate return series is captured better by GARCH processes than EWMA processes.

Again, we consider the autocorrelation of the squared returns and squared residuals in the period of December 8, 2004 until March 31, 2006. The results are depicted in Table 3.5. The lower and upper 95% confidence bounds are -0.109 and 0.109. The observed squared returns do not exhibit significant autocorrelation. This might be explained by the low volatility in this period. We observe only three violations of the 95% confidence bounds. The out-of-sample residuals of the α-stable CMGARCH model violate the these bounds only once. In the classical EWMA model ($M = 112$), we have three autcorrelation estimates out of this range. In the EWMA model using only the last 20 observations, there are no the violation of these bounds and in the last model we observe one violation.

Stock	1	2	3	4	5	6	7
$ACF((X_{t,1}^2))$	-0.006	-0.018	-0.012	-0.025	-0.03	-0.048	-0.043
$ACF((X_{t,2}^2))$	0.112	0.023	0.123	0.106	0.121	0.062	0.041
$ACF((X_{t,3}^2))$	-0.007	0.02	-0.003	-0.015	-0.012	-0.01	-0.01
$ACF((X_{t,4}^2))$	0.028	0.043	0.098	-0.083	0.015	-0.092	0.011
$ACF((\hat{Z}_{t,1}^2))$	-0.025	-0.009	-0.025	-0.026	-0.021	-0.04	-0.04
$ACF((\hat{Z}_{t,2}^2))$	-0.013	-0.041	0.007	-0.015	0.012	0.055	-0.015
$ACF((\hat{Z}_{t,3}^2))$	-0.009	-0.005	-0.006	-0.011	-0.007	-0.01	-0.011
$ACF((\hat{Z}_{t,4}^2))$	0.132	-0.077	-0.095	-0.061	-0.015	-0.043	-0.029
$ACF((\hat{Z}_{t,1}^{2,E(112)}))$	0.043	0.058	0.012	0.066	0.059	0.098	-0.004
$ACF((\hat{Z}_{t,2}^{2,E(112)}))$	0.155	0.067	0.137	0.12	0.14	0.03	0.023
$ACF((\hat{Z}_{t,3}^{2,E(112)}))$	0.009	-0.016	0.062	0.01	0	0.004	-0.007
$ACF((\hat{Z}_{t,4}^{2,E(112)}))$	0.049	0.024	0.095	0.032	0.072	0.037	-0.008
$ACF((\hat{Z}_{t,1}^{2,E(20)}))$	0.011	0.003	-0.018	0.013	0.018	0.044	-0.037
$ACF((\hat{Z}_{t,2}^{2,E(20)}))$	0.097	0.007	0.026	0.028	0.051	-0.014	0.01
$ACF((\hat{Z}_{t,3}^{2,E(20)}))$	-0.01	-0.012	0.005	-0.008	-0.01	-0.014	-0.015
$ACF((\hat{Z}_{t,4}^{2,E(20)}))$	0.038	-0.018	0.041	0.021	0.033	-0.016	-0.006
$ACF((\hat{Z}_{t,1}^{2,E(50)}))$	0.023	0.034	-0.001	0.05	0.038	0.082	-0.022
$ACF((\hat{Z}_{t,2}^{2,E(50)}))$	0.128	0.048	0.086	0.079	0.1	0.018	0.045
$ACF((\hat{Z}_{t,3}^{2,E(50)}))$	0.003	-0.024	0.054	0.006	-0.008	-0.003	-0.014
$ACF((\hat{Z}_{t,4}^{2,E(50)}))$	0.036	0.01	0.077	0.018	0.059	0.027	-0.014

Table 3.5: The table depicts the autocorrelation of the squared returns and the squared residuals for the different models. The lower and upper 95% confidence bounds are -0.109 and 0.109, respectively. The period covered is December 8, 2004 to March 31, 2006.

3.6.3 Summary of the Results

Summing up, the α-stable CMGARCH model outperforms the EWMA models because the normalized dispersion matrix of its residuals is closer to strict white noise than the ones in the other EWMA models. The EWMA models, in particular, reveal its weakness in estimating the diagonal entries. Furthermore, the autocorrelation in the squared return data is captured better by the α-stable CMGARCH model than by the classical EWMA model ($M = 112$) because we observe in the former one less violation of the confidence bounds and, in general, the estimates are closer to zero.

These observations hold for the in-sample as well as the out-of-sample analysis. The good empirical performance of the α-stable CMGARCH model is clear evidence for the GARCH-projection-efficiency of the return series $(X_t)_{t\in\mathbf{Z}}$ investigated.

3.7 Conclusion

In this chapter we introduce a new class of multivariate GARCH models that is flexible enough to model multivariate time series appropriately and allow for estimation procedures that are applicable even in higher dimensions. We motivate these models by introducing the notions of projection-efficient and GARCH-projection-efficient that are fundamental for the working of these models.

Moreover, in this chapter we demonstrate that α-stable multivariate GARCH modeling is feasible. To do so, we develop α-stable versions of the CMGARCH and FCM-GARCH model. We demonstrate the applicability of the model and report empirical evidence that indicates that it outperforms the classical EWMA model introduced by RiskMetrics.

Chapter 4

Multi-Tail Elliptical Distributions

4.1 Introduction

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

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

distributions because of their radial symmetry.

Furthermore and not surprisingly, the research of Jansen, and de Vries (1991), Mittnik, and Rachev (1993), Loretan, and Phillips (1994), and Rachev, and Mittnik (2000) has found that the tail index α varies significantly between assets. Despite this well known fact, most existing research on heavy-tailed portfolios and factor models has assumed that the probability tails are the same in every direction.

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

We propose a new class of distributions that we label multi-tail elliptical distributions. Multi-tail elliptical distributions allow for modeling varying tail thickness leading to more realistic models and have the further advantage that a three-step estimation procedure is available; the estimation procedure is applicable even in higher dimensions. The random mechanism underlying the multi-tail elliptical distributions can be motivated by typical financial market behavior.

This chapter is organized as follows. In Section 4.3 we introduce the multi-tail elliptical distributions, derive their basic properties, and give examples. In Section 4.4 we develop a three-step estimation procedure to estimate the multi-tail elliptical distribution. In Section 4.5 we apply the multi-tail elliptical distributions to return data and Section 4.6 summarizes our conclusions.

4.2   Operator Stable Distributions

4.2.1   Definitions and Basic Properties

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

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

Definition 38. *A random vector X is said to be an operator stable random vector in $\mathbf{R}^d$ if there exists a matrix $E \in \mathbf{R}^{d \times d}$ and a vector a_n such that*

$$n^{-E}(X_1 + X_2 + ... + X_n - a_n) \stackrel{d}{=} X, \qquad (4.1)$$

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

The notion n^{-E} is defined as

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

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

Rewriting equation (4.1), we obtain

$$X_1 + X_2 + ... + X_n \stackrel{d}{=} n^E X + a_n. \qquad (4.2)$$

The characteristic function of the right hand-side equals

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

The characteristic function of the left hand-side of equation (4.2) is the n-th power of $\Phi_X(t)$ because the sum consists of independent and identically-distributed random vectors in this equation. Thus, equation (4.2) implies

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

The following definition is equivalent to Definition 4.2 and characterizes the operator stable random vectors in terms of their domain of attraction. In order to generalize Gaussian-based financial theories, it is important that the distribution of a random vector X has a domain of attraction. This property implies that the random vector X

can be interpreted as an aggregation of many shocks caused by new arriving financial information.

Definition 39. *A random variable X is said to have an operator stable distribution if it has a domain of attraction, i.e., if there is a sequence of independent and identically-distributed random vectors $Y_1, Y_2, ...$ and a sequence of $d \times d$ matrices $(A_n)_{n \in \mathbf{N}}$ and vectors $(b_n)_{n \in \mathbf{N}}$, such that*

$$A_n(Y_1 + Y_2 + ... + Y_n - b_n) \overset{d}{\Rightarrow} X. \tag{4.3}$$

The limits in equation (4.3) are called operator stable (see Sharpe (1969) and Jurek, and Mason (1993)). If $E(||Y_1||^2)$ exists in Definition 39 then the classical Central Limit Theorem shows that X is multivariate normal, a special case of operator stable. In this case, we can take $A_n = n^{-1/2}$ and $b_n = nE(X)$. But if $0 < \alpha < 2$ then $E(||X||^2) = \infty$ and the classical Central Limit Theorem does not apply. If, in this case, the tails of Y_1 fall off at the same rate α in every direction then equation (4.3) holds with $A_n = n^{-1/\alpha}I$ and the limit X is multivariate stable distributed due to the generalized Central Limit Theorem (see Rachev, and Mittnik (2000) for a detailed discussion).

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

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

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

$$\mathbf{R}^d = V_1 \oplus V_2 \oplus ... \oplus V_p \text{ and } EV_i \subset V_i, \tag{4.4}$$

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

Given a non-zero vector $v \in \mathbf{R}^d$, we can write its unique spectral decomposition

---

[1]This condition is necessary for the existence of a generalized domain of attraction (see Definition 39).

$v = v_1 + ... + v_p$ according to equation (4.4). Let's define

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

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

Proposition 8. Let X be an operator stable vector in $\mathbf{R}^d$ with exponent E and let $v \in \mathbf{R}^d$. Then for any small $\delta > 0$ we have

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

for all $\lambda > 0$ sufficiently large.

Proposition 8 says that the tail behavior of the linear combination $v'X$ is dominated by the component with the heaviest tail. In particular, this means that $E(|v'X|^s)$ exists for all $0 < s < \alpha(v)$ and diverges for $s > \alpha(v)$.

If we decompose an operator stable random vector X with exponent E according to equation (4.4) we obtain

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

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

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

Example 3. *Take $B = \text{diag}(a_1, ..., a_d)$ where $a_1 > a_2 > ... > a_d$ and P orthogonal. If $X = (X_1, ..., X_d)'$ is operator stable with exponent $E = PBP'$ then $B_i = a_i$ and $V_1, ..., V_d$ are the coordinate axes in the new coordinate system defined by the vectors $p_i = Pe_i$, $i = 1, ..., d$. One can show that the ith component $p_i'X$ in the new coordinate system is stable with index $\alpha_i = \frac{1}{a_i}$ (see Meerschaert, and Scheffler (2003)). Whereas the original ith component X_i is not stable since it is a linear combination of stable laws of different indices. Hence, the change of coordinate matrix P rotates the coordinate axes to make the marginals stable. More mathematically, since $n^{PBP^{-1}} =$*

$Pn^B P^{-1}$ it follows from Definition 38 that

$$Pn^{-B}P^{-1}(X_1 + ... + X_n - b_n) \overset{d}{=} X$$
$$n^{-B}(P^{-1}X_1 + ... + P^{-1}X_n - P^{-1}b_n) \overset{d}{=} P^{-1}X,$$

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

4.2.2 Applications

Mittnik, and Rachev (1993) seem to have been the first to apply operator stable models to problems in finance. Since empirical work suggests that the stable indices α_i, $i = 1, ..., d$ vary depending on the assets, they assume the univariate version of equation (4.3) holds

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

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

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

with a diagonal norming matrix

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

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

Meerschaert, and Scheffler (2003) apply the model presented in example 3 to exchange rate log-returns for the German Deutsche mark $X_1(t)$ and $X_2(t)$, $t = 1, ..., n$, both taken against the U.S.-dollar. They argue that since the tail parameter α usually depends on the coordinate, the wrong coordinate system can mask variations in α because of the domination of the heaviest tail (see Proposition 8). They suggest that the coordinate system given by the principal components of the sample covariance matrix is a judicious choice, since the principal components can be interpreted as the directions of the largest dispersion. They assume that the random vector $Y(t) = PX(t)$ is operator stable with exponent $\text{diag}(a_1, a_2)$, implying that $Y(t)$ has stable marginals.

Hence, $X(t)$ is operator stable with exponent $P \operatorname{diag}(a_1, a_2) P'$. In particular, they estimate $\alpha_1 = 1/a_1 \approx 1.656$ and $\alpha_2 = 1/a_2 \approx 1.996$. They conclude that $Z_1(t)$ is α_1-stable distributed and $Z_2(t)$ is normally distributed since a_2 is very close to $1/2$, indicating normality.

4.3 Multi-Tail Elliptical Distributions

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

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

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

$$X = \mu + RAS,$$

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

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

Let us assume that the daily log-returns of a portfolio modeled by the random variable X follow an elliptical model $X = RAS$.[2] Given the the partial information that we observe *negative* signs in *all* components of X (hence, all components of AS and

---

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

$s(AS)$ are negative since R is positive) we may conclude that such an observation is caused by an important macroeconomic event or financial crisis (i.e., where markets are in a stress situation). The markets tend to be extreme and it is very likely to observe large losses in many components of X. In such a market stress scenario one says that the correlations approach unity (see McNeil, Frey, and Embrechts (2005)). However, instead of assuming a higher correlation between assets, we suppose lower tail parameters of R in these directions (e.g, $AS = (-1, ..., -1)$). These lower tail parameters incorporate a higher tail dependence in these directions and cause simultaneous high losses.

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

4.3.1   Definition and Basic Properties

Definition 40 (**Multi-tail Elliptical Distributions**). *Let $S \in \mathbf{R}^d$ be the uniform distribution on the unit hypersphere $\mathcal{S}^{d-1}$, I an interval of tail parameters and $(R_\alpha)_{\alpha \in I}$ a family of positive random variables with tail parameter α. The random vector $X = (X_1, X_2, ..., X_d)' \in \mathbf{R}^d$ has a multi-tail elliptical distribution, if X satisfies*

$$X \stackrel{d}{=} \mu + R_{\alpha(s(AS))} AS, \tag{4.5}$$

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

In particular, we call the function

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

the tail function of a multi-tail elliptical distribution. In Section 4.3.2 we discuss several examples of the tail function and its impact on the distribution.

The multi-tail elliptical distributions are a generalization of the elliptical distributions (see Fang, Kotz and Ng (1987)) and a new subclass of the generalized elliptical distributions (Frahm (2004)).

The trick in Definition 40 is to make the tail parameter α depend on AS and not directly on S. In empirical work the random vector S is not observable and A is not unique because of the rotational symmetry of S, implying

$$S \stackrel{d}{=} PS \text{ and } AS \stackrel{d}{=} A(PS) \tag{4.6}$$

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

Lemma 3. *Let $S \in \mathbf{R}^d$ be a random vector with uniform distribution on the hypersphere $\mathcal{S}^{d-1}$. Then its density satisfies*

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

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

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

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

$\square$

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

Theorem 18. *Let $X = R_{\alpha(s(AS))}AS$ be a multi-tail elliptical distribution, where $A \in \mathbf{R}^{d \times d}$ is a regular matrix, I an interval of tail parameters, $(R_\alpha)_{\alpha \in I}$ a family of positive random variables, $(f_{R_\alpha})_{\alpha \in I}$ its family of densities, S the uniform distribution on the hypersphere $\mathcal{S}^{d-1}$ and $\alpha : \mathcal{S}^{d-1} \to I$ a tail function. Then the density of X is given by*

$$f_X(x) = |det(\Sigma)|^{-1/2} g_{\alpha(s(x-\mu))}((x - \mu)'\Sigma^{-1}(x - \mu)) \tag{4.7}$$

and

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

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

We note that the matrix A is not part of the density term in equation (4.7). This is important for empirical work, since the matrix A is difficult to observe and not unique due to equation (4.6).

Proof. Due to Definition 40 we have

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

---

[3]See Definition 48 in the appendix for further information about Gamma function.

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

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

Hence the joint density exists and corresponds to

$$
\begin{aligned}
f_{(R,S)}(r,s) &= f_S(s)f_{R|S}(r|s) \\
&= f_S(s)f_{R_{\alpha(s(As))}}(r) \\
&= \frac{\Gamma(d/2)}{2\pi^{d/2}}f_{R_{\alpha(s(As))}}(r) \quad\quad\quad (4.8)
\end{aligned}
$$

We define the transformation

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

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

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

where $\Sigma = AA'$.

Since we have for all $A \in \mathcal{B}^d$

$$
\begin{aligned}
\int_A f_X(x)dx &= P[X \in A] \\
&= P[g(R,S) \in A] \\
&= P[(R,S) \in g^{-1}(A)] \\
&= \int_{g^{-1}(A)} f_{(R,S)}(r,s)drds \\
&\overset{(1)}{=} \int_A f_{(R,S)}(g^{-1}(x))|\det((Dg^{-1})(x))|dx, \quad\quad (4.9)
\end{aligned}
$$

it follows

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

Note that (1) holds in equation (4.9) because of Theorem 24 in the appendix. Furthermore, Corollary 8 in the appendix says that we have

$$
\det(D(g^{-1})(x)) = \frac{((x-\mu)'\Sigma^{-1}(x-\mu))^{(-d+1)/2}}{\det(\Sigma)^{1/2}} \quad\quad (4.10)
$$

and furthermore, we obtain by using equation (4.8)

$$f_{(R,S)}(g^{-1}(x)) = \frac{\Gamma(d/2)}{2\pi^{d/2}} f_{R_{\alpha(s(Ag_2^{-1}(x)))}}(g_1^{-1}(x)). \tag{4.11}$$

It is important to note that we have

$$
\begin{aligned}
s(Ag_2^{-1}(x)) &= s\left(\frac{x-\mu}{||A^{-1}(x-\mu)||}\right) = \frac{x-\mu}{||A^{-1}(x-\mu)||}\frac{||A^{-1}(x-\mu)||}{||x-\mu||} \\
&= \frac{x-\mu}{||x-\mu||} = s(x-\mu). \tag{4.12}
\end{aligned}
$$

Hence, we derive with equations (4.11) and (4.12)

$$f_{(R,S)}(g^{-1}(x)) = \frac{\Gamma(d/2)}{2\pi^{d/2}} f_{R_{\alpha(s(x-\mu))}}(\sqrt{(x-\mu)'\Sigma^{-1}(x-\mu)}). \tag{4.13}$$

Finally, adding equations (4.10) and (4.13) leads to

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

$$\square$$

Corollary 3. *Let* $X = \mu + R_{\alpha(s(AS))}AS \in \mathbf{R}^d$ *be a multi-tail elliptical random vector. Then we have*

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

or equivalently,

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

where $\Sigma = AA'$.

Proof. We have

$$
\begin{aligned}
\sqrt{(X-\mu)'\Sigma^{-1}(X-\mu)} &= ||\Sigma^{-1/2}(X-\mu)|| \\
&= ||\Sigma^{-1/2}(\mu + R_{\alpha(s(AS))}AS - \mu)|| \\
&= ||\Sigma^{-1/2}AS||R_{\alpha(s(AS))} \\
&\overset{d}{=} ||S||R_{\alpha(s(AS))} \\
&= R_{\alpha(s(AS))} \tag{4.14}
\end{aligned}
$$

Note that in the third line of the last equation we have equality in distribution because

$(\Sigma^{-1/2})'A$ is orthogonal[4]. $\Sigma^{-1/2}A$ is orthogonal since we have

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

From equation (4.14) we conclude

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

Analogously, we can derive

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

$\square$

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

Definition 41. *Let $(R_\alpha)_{\alpha \in I}$ be a family of positive random variables with tail parameter α and $Y_\alpha = R_\alpha AS$, $\alpha \in I$, a family of elliptical random vectors. We call the random vector X a corresponding multi-tail elliptical random vector if it is given by $X = R_{\alpha(s(AS))}AS$, where $\alpha : \mathcal{S}^{d-1} \to I$ is a tail function.*

Theorem 19. *Let $X = R_\alpha AS$ be an elliptical random vector, where $\mu \in \mathbf{R}^d$ is a location vector, $A \in \mathbf{R}^{d \times d}$ a regular matrix, R_α a positive random variable with tail parameter α and the random vector $S \in \mathbf{R}^d$ uniformly distributed on $\mathcal{S}^{d-1}$. Then X possesses a density f_X if and only if R_α has a density f_{R_α}. The relationship between f_X and f_{R_α} is as follows*

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

and

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

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

Proof. See Fang, Kotz and Ng (1987). $\square$

---

[4]See also equation (4.6).

If follows from Theorems 18 and 19 that the density of a multi-tail random vector can be obtained by substituting the constant tail parameter α through the tail function $\alpha(.)$ in the density of the corresponding family of elliptical random vectors. It is obvious that multi-tail elliptical distributions are a generalization of elliptical distributions. Vice versa, if we choose the tail function to be constant, R and S are independent and we obtain the classical case. Summing up, we obtain

Corollary 4. *Densities of a family of elliptical random vectors differ from a density of a corresponding multi-tail elliptical random vector by substituting the constant α through a tail function $\alpha(.)$.*

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

$$Y = \mu + W_\alpha^{1/2} A Z, \qquad (4.15)$$

where $\mu \in \mathbf{R}^d$ is a location parameter, W_α a positive random variable with tail parameter $\alpha/2$,[5] $A \in \mathbf{R}^{d \times d}$ a regular matrix, and $Z \sim N(0, Id)$. We can rewrite equation (4.15) in the form

$$Y = \sqrt{W_\alpha ||Z||^2} A \left(\frac{Z}{||Z||} \right). \qquad (4.16)$$

It is important to note that we have[6]

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

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

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

Thus, identifying $R_\alpha = \sqrt{W_\alpha ||Z||^2}$ and $Z/||Z|| = S$, we see that normal variance mixtures are elliptical random vectors according to equation (4.16).

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

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

We introduce this "mixture" notation because most well-known elliptical distribution families such as α-stable sub-Gaussian distributions, t distributions, and generalized

---

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

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

hyperbolic distributions are defined in the mixture notation. In the next section we introduce their corresponding multi-tail versions.

4.3.2 Principal Component Tail Functions

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

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

Definition 42. *Let X be a multi-tail elliptical random vector with dispersion matrix Σ. We call the tail function $\alpha : \mathcal{S}^{d-1} \to I$ of X a principal component tail function (pc-tail function) if it satisfies*

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

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

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

and

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

Note that for all pc-tail functions $\alpha(.)$ we have $\alpha(s) \in I$, $s \in \mathcal{S}^{d-1}$ since I

---

[7]For references see Introduction.

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

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

[10]We will see in Section 4.4 that we can estimate Σ without knowing the tail function. This is of course important for practical work.

is an interval and $\alpha(s)$ can be interpreted as a convex combination of $\alpha_i^-, \alpha_i^+ \in I$, $i = 1, ..., d$.

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

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

and

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

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

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

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

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

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

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

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

fact, $\alpha_1(.)$ is a pc-tail function since we have for all $s \in \mathcal{S}^{d-1}$

$$
\sum_{i=1}^{d} \frac{1}{2} < s, v_i >^2 + \frac{1}{2} < s, v_i >^2 \quad = \quad \sum_{i=1}^{d} < s, v_i >^2
$$

$$
= \quad < s, \sum_{i=1}^{d} < s, v_i > v_i >
$$

$$
= \quad < s, s >
$$

$$
= \quad 1.
$$

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

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

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

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

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

It is obvious that

$$
\sum_{i=1}^{d} < s, v_i >^2 I_{(0,\infty)}(< s, v_i >) + < s, v_i >^2 I_{(-\infty,0)}(< s, v_i >) = 1. \quad (4.17)
$$

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

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

Definition 43. *Let $A \subset \mathbf{R}^d$ be a linear subspace. A ϵ-cone $C_\epsilon(A)$ is the set defined by*

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

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

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

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

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

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

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

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

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

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

4.3.3   Two Examples of Multi-Tail Elliptical Distributions

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

Definition 44 (Multi-tail α-stable sub-Gaussian random vectors). *Let $Z \in \mathbf{R}^d$ be a standard normal random vector and $(W_\alpha)_{\alpha \in (0,2)}$ a family of positive α-stable random variables with tail parameter $\alpha/2$ satisfying*

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

Then the random vector

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

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

In particular, multi-tail α-stable sub-Gaussian random vectors allow us to combine the concepts of α-stable distributions and varying tail thickness such as operator stable random vectors. Figure 4.1 depicts scatterplots of 1000 samples of a multi-tail α-stable sub-Gaussian distribution with dispersion matrix $10^{-4} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. $F_1 = s((1,1)')$ is

the first principal component of Σ and $F_2 = s((1, -1)')$ the second one. In scatterplot (a) in Figure 4.1 we have the pc-tail function

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

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

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

We observe in Figure 4.1 that we have much more outliers in the cones around F_1 and $-F_1$ than in the ones around F_2 and $-F_2$ because of the pc-tail function.

Figure 4.2 shows scatterplot of 1000 samples of a multi-tail sub-Gaussian distribution with the same dispersion matrix as in Figure 4.1 but with different tail function. In scatterplot (a) in Figure 4.2 we have the pc-tail function

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

and in scatterplot (b)

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

In Figure 4.2 there are many more outliers in the cone around $-F_1$ than in the one around F_1. This is because of the smaller tail parameters in the directions in the cone around $-F_1$.

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

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

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

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

where $\nu > 0$ is the tail parameter, $c > 0$ a scaling parameter, and Σ_0 the normalized

---

[11]$\mathrm{Ig}(\alpha,\beta)$ is the inverse gamma distribution for further information (see Rachev, and Mittnik (2000) or McNeil, Frey, and Embrechts (2005)).

dispersion matrix.[12] The corresponding multi-tail t distribution is defined as follows.

Definition 45 (Multi-tail t-distributed random vectors). *Let $Z \in \mathbf{R}^d$ be a standard normal random vector and $(W_\nu)_{\nu \in (0,\infty)}$ a family of inverse gamma distributed random vectors with tail parameter $\nu/2$, i.e. $W_\nu \sim Ig(\frac{1}{2}\nu, \frac{1}{2}\nu)$. The random vector*

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

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

According to Corollary 4 and the knowledge about the density f_Y of a t distributed random vector we can derive

Corollary 5. *Let $X \in \mathbf{R}^d$ be a multi-tail t distributed random vector with tail parameter function $\nu : \mathcal{S}^{d-1} \to (0,\infty)$ and a normalized dispersion matrix Σ_0, a scaling parameter $c > 0$, and a location parameter $\mu \in \mathbf{R}^d$. The density f_X satisfies*

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

Figure 4.3 depicts the density contour lines of two multi-tail t distributions with dispersion matrix $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. In Figure 4.3 (a) we have the pc-tail function

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

and in Figure 4.3 (b)

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

We observe in Figure 4.3 that the shape of the contour lines around the mean $(0,0)'$ is determined by the dispersion matrix Σ. But in the tails (i.e., far away of $(0,0)'$) the influence of the tail function starts to dominate. In particular, outliers in a cone around $-F_1$ are much more likely than in any other direction. Figure 4.4 shows again the densities contour lines of two multi-tail t distributions with the same dispersion

---

[12]We introduce normalization criteria for a dispersion matrix Σ in Section 4.4.
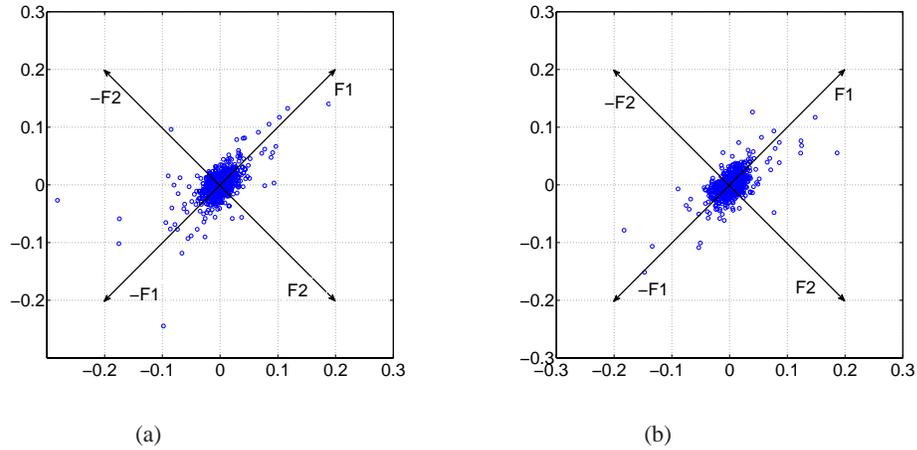
Figure 4.1: Scatterplots of a multi-tail α-stable sub-Gaussian distributions with symmetric pc-tail functions

matrices as in Figure 4.3. In Figure 4.3 (a) the tail function is

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

and in Figure 4.3 (b) we have

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

In Figure 4.3 (a) the tail function supports the scaling properties of the dispersion matrix induced by Σ. In Figure 4.3 (b) we have the opposite effect in the cone around $-F_2$.

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

4.4   Estimation of Multi-Tail Elliptical Distributions

In this section we explain how to estimate the parameters of a multi-tail elliptical random vector X. In Section 4.3 we explained how a multi-tail elliptical random vector $X = \mu + R_{\alpha(s(AS))}AS$ depends on the location parameter μ, the dispersion matrix $\Sigma = AA'$, the tail function $\alpha(.)$, and the family of random variables $(R_\alpha)_{\alpha \in I}$ that is a one parametric family in most applications. We present a three-step estimation procedure to fit the multi-tail elliptical random vectors to data.

In the first step, we estimate the location vector $\mu \in \mathbf{R}^d$ with some robust method.
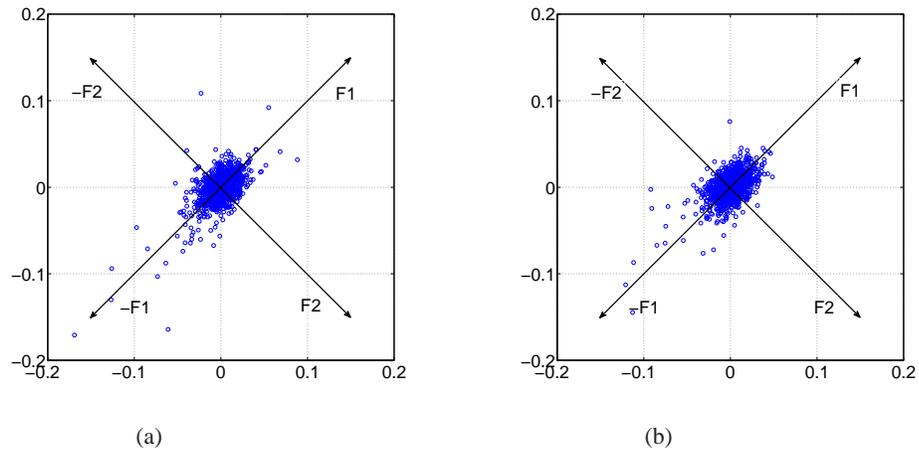
Figure 4.2: Scatterplots of a multi-tail α-stable sub-Gaussian distributions with asymmetric pc-tail functions
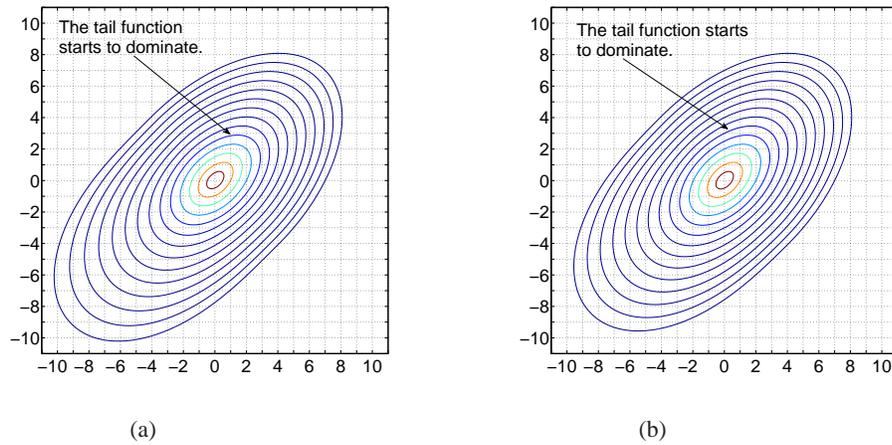
Figure 4.3: Contour lines of the densities of two asymmetric multi-tail t distributions
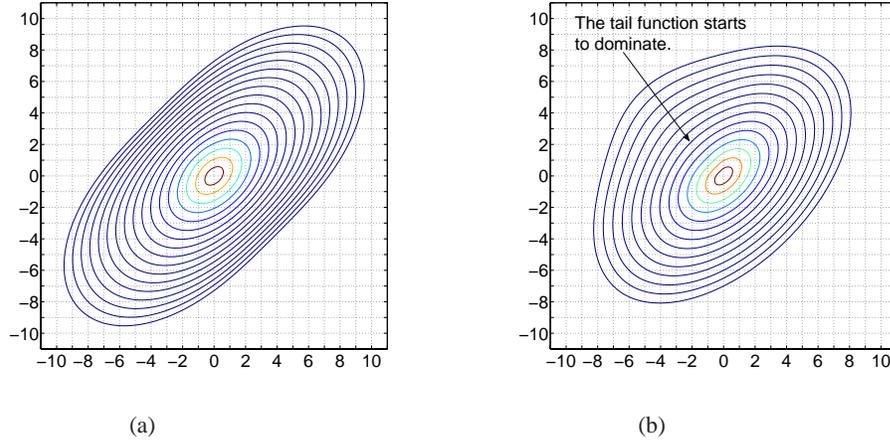
Figure 4.4: (a) depicts the density contour lines of a symmetric multi-tail t distribution; (b) the density contour lines of a asymmetric multi-tail t distribution

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

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

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

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

$$\text{(i) } \Sigma_{11} = 1 \quad \text{(ii) } \det(\Sigma) = 1 \quad \text{(iii) } \text{tr}(\Sigma) = 1, \tag{4.18}$$

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

In the third step we have to estimate the scale parameter c and the tail function $\alpha(.)$. Since we assume a pc-tail function, we have to evaluate the tail parameters $(\alpha_1, ..., \alpha_k) \in I^k$, $k \in \mathbf{N}$, of the pc-tail function. In the last step we determine the

parameters from the set $\Theta = \mathbf{R}_+ \times I^k$, where I is the interval of tail parameters.

4.4.1 Estimation of the Dispersion Matrix

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

Definition 46 (**Unit random vector**). *Let $A \in \mathbf{R}^{d \times d}$ be a regular matrix and S be uniformly distributed on the hypersphere $\mathcal{S}^{d-1}$. We call the random vector*

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

the "unit random vector generated by A".

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

$$
\begin{aligned}
\frac{X - \mu}{||X - \mu||} &= \frac{R_{\alpha(s(AS))} AS}{||R_{\alpha(s(AS))} AS||} \\
&= \frac{AS}{||AS||} \\
&= S_A
\end{aligned}
\tag{4.19}
$$

The family of random variables $(R_\alpha)_{\alpha \in I}$ and the tail function $\alpha(.)$ have no influence $s(X - \mu) = S_A$. This is important since it allows for a robust estimation of the dispersion matrix Σ up to a scaling constant. In particular, it follows from equation (4.19) immediately that we have $S_{cA} = S_A$ for all $c > 0$.

Theorem 20. *The spectral density function of the unit random vector generated by $A \in \mathbf{R}^{d \times d}$ satisfies*

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

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

Proof. Because of the invariance property described in equation (4.19) we can assume without loss of generality that X is normally distributed with location parameter 0 and covariance matrix Σ, i.e. $X \sim N(0, \Sigma)$, and we have

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

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

$$f_{X|(||X||)}(x|r) = \frac{f_X(x)}{f_{||X||}(r)} 1_{\mathcal{S}_r^{d-1}}(x), \qquad (4.20)$$

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

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

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

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

Due to Theorem 24 given in the appendix we have

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

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

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

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

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

where $s \in \mathcal{S}^{d-1}$ and $\Gamma(.)$ is the gamma function (see Definition 48). $\qquad \square$

---

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

Corollary 6. *Let $X = \mu + R_{\alpha(s(AS))}AS \in \mathbf{R}^d$ be a multi-tail elliptical random vector, where $\Sigma = AA'$. Then the joint distribution of $R = \sqrt{(X - \mu)\Sigma^{-1}(X - \mu)}$ and S_A is given by*

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

Proof. We observe

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

$\square$

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

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

$$
\begin{aligned}
\log\left(\prod_{i=1}^n f_{S_A}(s_i)\right) &= \sum_{i=1}^n \log(f_{S_A}(S_i)) \\
&= \sum_{i=1}^n \Big(\log(\Gamma(d/2)) - \log(2\pi^{d/2}) \\
&\quad + \log((S_i'\Sigma^{-1}S_i)^{-d/2}) - \log(\det(\Sigma)^{1/2})\Big) \\
&= n\log(\Gamma(d/2)) - \frac{nd}{2}\log(2\pi) - \frac{n}{2}\log(\det(\Sigma)) \\
&\quad - \frac{d}{2}\sum_{i=1}^n \log(S_i'\Sigma^{-1}S_i). \qquad\qquad (4.21)
\end{aligned}
$$

Since in equation (4.21) $n\log(\Gamma(d/2))$ and $nd\log(2\pi)/2$ are constants subject to Σ we can neglect them in the following maximum likelihood optimization problem

$$\hat{\Sigma} = \mathrm{argmax}_{\Sigma \in D_d^2} -n\log\det(\Sigma) - d\sum_{i=1}^n \log(S_i'\Sigma^{-1}S_i), \qquad (4.22)$$

where $D_d^2 \subset \mathbf{R}^{d\times d}$ denotes the set of positive definite $d \times d$ matrices. Note that equation (4.21) determines $\hat{\Sigma}$ up to a scale parameter $c > 0$. If $\hat{\Sigma}$ satisfies this equation, it can be seen immediately that it holds for $c\hat{\Sigma}$ as well.

Definition 47. *The maximum likelihood estimator $\hat{\Sigma}$ defined by equation (4.21) is also called the spectral estimator.*

In Theorem 21 we interpret D_d^2 as an open subset of $\mathbf{R}^{d(d+1)/2}$ vector space. Thus, we interpret the matrix Σ as a vector in this space. Note that if $\Sigma \in D_d^2$ then also $\Sigma^{-1} \in D_d^2$. Let $f : \mathbf{R}^{d(d+1)/2} \to \mathbf{R}$ be differentiable in the point Σ. Then, we denote the Jacobian with $Df(\Sigma)$.

Theorem 21. *Let $X \in \mathbf{R}^d$ be a random vector with density*

$$f_X : \begin{cases} \mathbf{R}^d & \to & \mathbf{R} \\ x & \mapsto & \sqrt{\det(\Sigma^{-1})}g((x-\mu)'\Sigma^{-1}(x-\mu)) \end{cases} \tag{4.23}$$

Then the function

$$h_x : \begin{cases} D_d^2 \subset \mathbf{R}^{d(d+1)/2} & \to & \mathbf{R} \\ \Sigma & \mapsto & \log\left(\sqrt{\det(\Sigma)}g((x-\mu)'\Sigma(x-\mu))\right) \end{cases} \tag{4.24}$$

has the Jacobian

$$\begin{aligned} Dh_x(\Sigma) & = & \frac{1}{2}(2\Sigma^{-1} - \operatorname{diag}(\Sigma^{-1})) \\ & & + \frac{d(\log(g(z)))}{dz}(2(x-\mu)'(x-\mu) - \operatorname{diag}((x-\mu)'(x-\mu))) \end{aligned} \tag{4.25}$$

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

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

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

$$f_{ij} = \begin{cases} e_{ii} & , & i = j \\ e_{ij} + e_{ji} & , & i < j \end{cases}$$

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

Our goal is to prove equation (4.25). If Σ is positive definite, then Σ^{-1} is positive definite, too. Due to Theorem 25 in the appendix, the equation

$$D \det(A) = \det(A)A^{-1} \tag{4.26}$$

holds of all symmetric and regular matrices. Note that we have according to equation (4.24)

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

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

$$
\begin{aligned}
\frac{\partial h_x(\Sigma)}{\partial f_{ij}} &= \frac{1}{2}\frac{1}{\det(\Sigma)}\frac{\partial \det(\Sigma)}{\partial f_{ij}} + \frac{\partial \log(g(z))}{\partial z}\frac{\partial (x-\mu)'\Sigma(x-\mu)}{\partial f_{ij}} \\
&= \frac{1}{2}\frac{1}{\det(\Sigma)}D\det(\Sigma)(e_{ij}+e_{ji}) + 2\frac{\partial \log(g(z))}{\partial z}(x_i-\mu_i)(x_j-\mu_j) \\
&\stackrel{(4.26)}{=} \frac{1}{2}\frac{1}{\det(\Sigma)}\det(\Sigma)\Sigma^{-1}(e_{ij}+e_{ji}) + 2\frac{\partial \log(g(z))}{\partial z}(x_i-\mu_i)(x_j-\mu_j) \\
&= \Sigma_{ij}^{-1} + 2\frac{\partial \log(g(z))}{\partial z}(x_i-\mu_i)(x_j-\mu_j),
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{\partial h_x(\Sigma)}{\partial f_{ii}} &= \frac{1}{2}\frac{1}{\det(\Sigma)}\frac{\partial \det(\Sigma)}{\partial f_{ii}} + \frac{\partial \log(g(z))}{\partial z}\frac{\partial (x-\mu)'\Sigma(x-\mu)}{\partial f_{ii}} \\
&\stackrel{(4.26)}{=} \frac{1}{2}\frac{1}{\det(\Sigma)}\det(\Sigma)\Sigma^{-1}e_{ii} + \frac{\partial \log(g(z))}{\partial z}(x_i-\mu_i)^2 \\
&= \frac{1}{2}\Sigma_{ii}^{-1} + \frac{\partial \log(g(z))}{\partial z}(x_i-\mu_i)^2,
\end{aligned}
$$

where $i = 1, ..., d$. Hence, equation (4.25) follows immediately. $\qquad\square$

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

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

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

$$
\sum_{i=1}^{n} Dh_{X_i}(\Sigma^{-1}) \stackrel{!}{=} 0, \tag{4.27}
$$

since it is a necessary condition for the maximum.

Theorem 22. *Let $X, X_1, ..., X_n \in \mathbf{R}^d$ be a sample of identically distributed data vectors with a density of the form*

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

Then the solution to the optimization problem in equation (4.22) can be characterized by the following fix point equation

$$
\hat{\Sigma} = -\frac{2}{n}\sum_{k=1}^{n}\frac{\partial \log(g(z_k))}{\partial z_k}(X_k-\mu)(X_k-\mu)', \tag{4.28}
$$

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

Equation (4.28) is only a necessary condition for the solution of the optimization problem stated in equation (4.22).

Proof. We show that equation (4.28) is true for every entry $\hat{\Sigma}_{ij}$ with $1 \leq i \leq j \leq d$. We denote the ith component of the random vector X_k with X_{ki}.

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

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

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

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

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

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

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

$\square$

Let S_A be a unit random vector generated by A with density f_{S_A}. The density f_{S_A}

is of the form given in equation (4.23) and satisfies

$$g(z) = \frac{\Gamma(d/2)}{2\pi^{d/2}} z^{-d/2}. \tag{4.29}$$

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

$$\begin{aligned}
\log(g(z))' &= \left(\log\left(\frac{\Gamma(d/2)}{2\pi^{d/2}} \right) - \frac{d}{2}\log(z) \right)' \\
&= -\frac{d}{2} z^{-1},
\end{aligned} \tag{4.30}$$

where $z > 0$.

Let $S_1, S_2, ..., S_n$ be a sample of identically distributed unit data vectors generated by A. According to Theorem 22 and equation (4.30), the maximum likelihood estimator $\hat{\Sigma}$ satisfies the following fix point equation

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

Furthermore, let $X_1, X_2, ..., X_n$ be a sample of identically distributed multi-tail elliptical data vectors. Due to equation (4.19) the sample $s(X_1 - \mu), ..., s(X_n - \mu)$ consists of unit data vectors generated by A. Hence, we obtain

$$\begin{aligned}
\hat{\Sigma} &= \frac{d}{n} \sum_{i=1}^{n} \frac{s(X_i - \mu)s(X_i - \mu)'}{s(X_i - \mu)'\Sigma^{-1}s(X_i - \mu)} \\
&= \frac{d}{n} \sum_{i=1}^{n} \frac{((X_i - \mu)/\|X_i - \mu\|)((X_i - \mu)/\|X_i - \mu\|)'}{((X_i - \mu)/\|X_i - \mu\|)'\hat{\Sigma}^{-1}((X_i - \mu)/\|X_i - \mu\|)} \\
&= \frac{d}{n} \sum_{i=1}^{n} \frac{(X_i - \mu)(X_i - \mu)'}{(X_i - \mu)'\hat{\Sigma}^{-1}(X_i - \mu)}.
\end{aligned} \tag{4.31}$$

Equation (4.31) is the fixed point representation of the maximum likelihood problem given in equation (4.22) in terms of the original data $X_1, ..., X_n$. It is important to note that the family $(R_\alpha)_{\alpha \in I}$ of random variables has no influence on the fix point representation.

It can be seen immediately that the fix point equation determines $\hat{\Sigma}$ only up to a

scale parameter since we have for all $c > 0$

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

In order to have uniqueness of the optimization problem, we apply one of the optimization criteria given in equation (4.18).

The next theorem is very important since it says when a solution to the optimization problem stated in equation (4.22) exists and when it is unique. Furthermore, it shows an iterative algorithm to approximate the solution.

Theorem 23. *Let $X_1, ..., X_n$ be a sample of identically distributed multi-tail elliptical data vectors with $n > d(d - 1)$. Then a fix point $\hat{\Sigma}$ of the equation*

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

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

$$
\begin{aligned}
\tilde{\Sigma}_0 &= \mathrm{Id} \\
\dot{\Sigma}^{(i+1)} &= \frac{d}{n}\sum_{i=1}^{n}\frac{(X_i - \mu)'(X_i - \mu)}{(X_i - \mu)'(\tilde{\Sigma}_0^{(i)})^{-1}(X_i - \mu)} \\
\tilde{\Sigma}_0^{(i+1)} &= \dot{\Sigma}_0^{(i+1)}
\end{aligned}
\tag{4.32}
$$

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

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

Theorem 23 says that the maximum likelihood optimization problem stated in equation (4.22) has a solution that is unique up to a scale parameter. We can use an iterative algorithm to find the solution $\hat{\Sigma}$ if we have more than $d(d - 1)$ observations. Fortunately, the algorithm works also quite well if we have only a sample size n which is slightly larger than d.

Finally, we remark that

Corollary 7. *Let $X_1, ..., X_n$ be a d-dimensional multi-tail elliptical distributed sample with known location parameter μ and $n \geq d$. Then the sequence $(\tilde{\Sigma}^{(i)})_{i \in \mathbf{N}}$ is positive definite for all $i \in \mathbf{N}$.*

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

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

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

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

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

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

is positive definite. Finally, Corollary 7 follows with induction. □

Empricial Analysis of the Spectral Estimator

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

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

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

Figures 4.5 to 4.8 depict boxplots of the spectral estimator $\hat{\Sigma}_0(n_i)$, $n_i = 20, 50, 100$ and $i = 1, 2, 3$. Each boxplot illustrated in these figures consists of 1000 estimates. All figures have in common that increasing the amount of iterations has only a minor influence on the accuracy of the estimator, while increasing the sample size improves

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

Table 4.1: The table depicts different quantiles of the spectral estimator $\hat{\sigma}_{12}$ and its relative errors.

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

Table 4.2: The table depicts different quantiles of the spectral estimator $\hat{\sigma}_{22}$ and its relative errors.

its behavior. Furthermore, we observe that the median $q_{0.5}$ is very close to the true values for all sample sizes, see Tables 4.1 and 4.2 and Figures 4.5 to 4.8, respectively. Tables 4.1 and 4.2 depicts several quantiles of the empirical distribution of the spectral estimator for different samples sizes. The relative error is defined as

$$e_{rel}^{1-2\alpha} = \max\left\{ \left| \frac{q_\alpha - c}{c} \right|, \left| \frac{q_{1-\alpha} - c}{c} \right| \right\} \qquad (4.33)$$

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

In Figure 4.5, we depict all estimate of spectral estimator lying in range between -0.0439 and 1.45 for and 0.246 and 3.2942. In Figure 4.6 we do not depict one upper outlier (value=2.645) in (a) and in (b) three upper outliers (values=7.1, 7.0, 3.8). In Figure 4.7 (b) we do not show three upper outliers (values=1.8,2.69,2.3) and in Figure 4.8 we do not illustrate one upper outlier (value=1.14) and one lower (value=0.24) in (a) and in (b) one upper outlier (value=2.76). We are a bit surprised to observe such large outliers when the sample size is relatively large, i.e. 1000 and 2000. The reason might be numerical instabilities of the spectral estimator.

---

[14]In order to be consistent, we require α to be in $(0, 0.5)$.
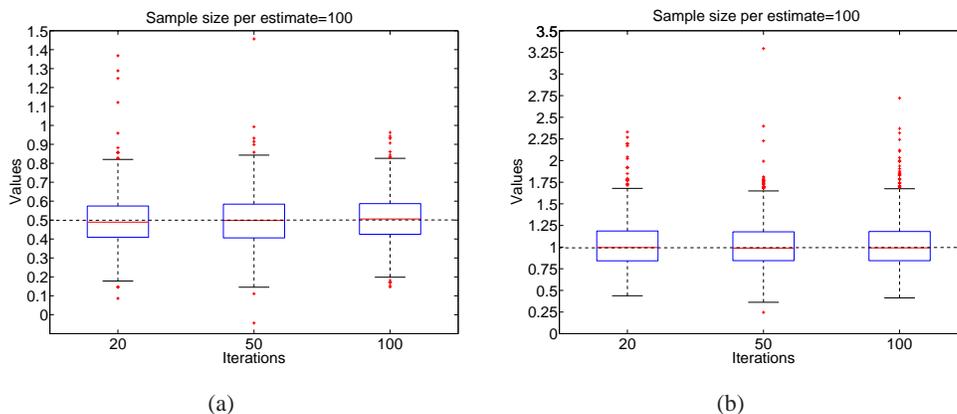
Figure 4.5: Boxplots of the spectral estimator with sample size 100 per estimate. (a) depicts the estimates of σ_{12} and (b) of σ_{22}.
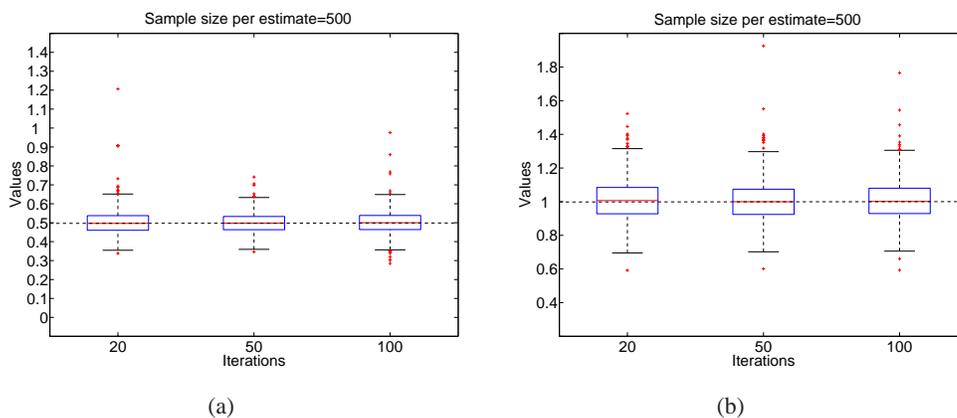
Figure 4.6: Boxplots of the spectral estimator with sample size 500 per estimate. (a) depicts the estimates of σ_{12} and (b) of σ_{22}.
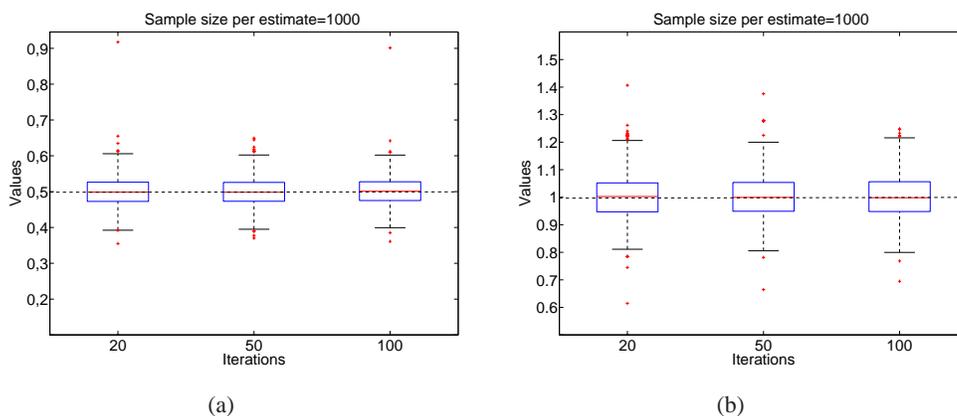
Figure 4.7: Boxplots of the spectral estimator with sample size 1000 per estimate. (a) depicts the estimates of σ_{12} and (b) of σ_{22}.
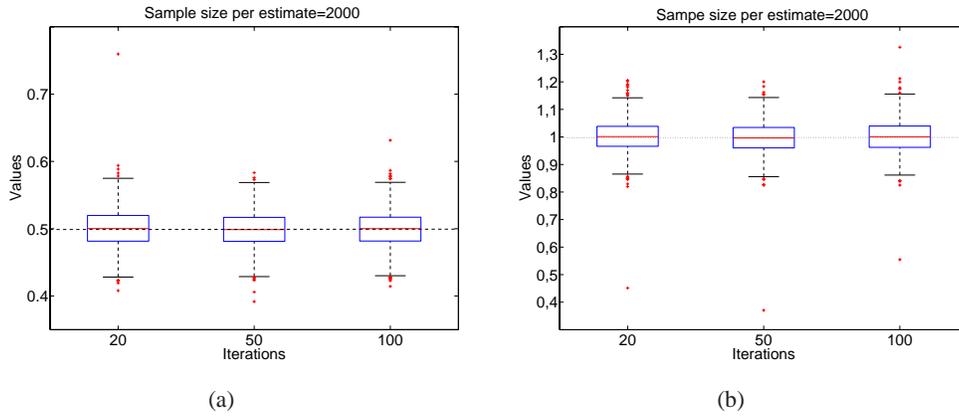
Figure 4.8: Boxplots of the spectral estimator with sample size 2000 per estimate. (a) depicts the estimates of σ_{12} and (b) of σ_{22}.

4.4.2 Estimation of the Parameter Set Θ

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

$$X = \mu + cR_{\alpha(s(\Sigma_0^{1/2}S))}\Sigma_0^{1/2}S, \tag{4.34}$$

where $c > 0$ is a thus far an unknown scale parameter[15] and $\alpha(.)$ is the tail function. Since we assume $\alpha(.)$ to be a pc-tail function (see Section 4.3.2) it can be determined by the tail parameters $(\alpha_1, ..., \alpha_k) \in I^k$, $k \in \mathbf{N}$. Hence, we have to estimate

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

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

Radial Variate ML-Approach

In this approach it follows from Corollary 3 and equation (4.34) that we have

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

---

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

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

$$f_{\sqrt{(X-\mu)'\Sigma_0^{-1}(X-\mu)}|s(X-\mu)}(r|s) = f_{cR_\alpha(s)}(r). \tag{4.35}$$

Proposition 9. Let $X \in \mathbf{R}$ be a random variable with density f_X and $Y = cX, c > 0$. Then we have

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

where f_Y denotes the density of Y.

According to Proposition 9 we obtain

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

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

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

and

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

Then the data vectors $R_1|S_1, R_2|S_2..., R_n|S_n$ are independent since we assume μ and Σ_0 to be known. According to equation (4.35), the log-likelihood function of this sample satisfies

$$\log\left(\prod_{i=1}^n f_{cR_{(S_i)}}(R_i)\right) = \sum_{i=1}^n \log\left(\frac{1}{c}f_{R_{\alpha(S_i)}}(R_i/c)\right)$$

$$= -n\log(c) + \sum_{i=1}^n \log(f_{R_{\alpha(S_i)}}(R_i/c)).$$

Finally, this leads to the optimization problem

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

Summing up, we obtain

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

The Density Generator ML-Approach

The second approach uses directly the density f_X of a multi-tail elliptical random vector. Due to Theorem 18, we know that the density satisfies

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

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

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

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

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

Equivalence of the Approaches

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

$$
\begin{aligned}
\log(f_X(x)) &= -\frac{1}{2}\log(\det(c\Sigma_0)) + \log\left(g_{\alpha(s(x-\mu))}\left(\sqrt{(x-\mu)'c\Sigma_0(x-\mu)}\right)\right) \\
&= -\frac{1}{2}\log(\det(\Sigma_0)) \\
&\quad -\frac{d}{2}\log(c) + \log\left(g_{\alpha(s(x-\mu))}\left(\sqrt{(x-\mu)'c\Sigma_0(x-\mu)}\right)\right) \qquad (4.36) \\
&= -\frac{1}{2}\log(\det(\Sigma_0)) - \log\left(\frac{\Gamma(d/2)}{2\pi^{d/2}}\right) + \log\left(\sqrt{(x-\mu)'c\Sigma_0(x-\mu)}\right) \\
&\quad -\frac{d}{2}\log(c) + \log\left(f_{R_{\alpha(s(x-\mu))}}\left(\sqrt{(x-\mu)'c\Sigma_0(x-\mu)}\right)\right). \qquad (4.37)
\end{aligned}
$$

Since the terms $-\frac{1}{2}\log(\det(\Sigma_0)), -\log\left(\frac{\Gamma(d/2)}{2\pi^{d/2}}\right)$ and $\log\left(\sqrt{(x-\mu)'c\Sigma_0(x-\mu)}\right)$ are constants with subject to parameter set Θ, it is equivalent to optimize equation (4.36) or (4.37) with respect to Θ. Of course, one should use the approach which leads to a simpler term in the optimization process.

Statistical Analysis

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

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

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

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

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

In Tables 4.3 to 4.5 and Figures 4.9 to 4.11 we observe that medians are close to the corresponding true values. The empirical distributions of $\hat{c}$, $\hat{\nu}_1$ and $\hat{\nu}_2$ are skewed to the left for small samples sizes and become more symmetric for large sample size per estimate. The accuracy of the estimator $\hat{c}$ is higher than the one of $\hat{\nu}_1$ and the one

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

Table 4.3: Quantiles of the estimator $\hat{c}$ for different sample sizes per estimate and the corresponding relative errors.

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

Table 4.4: Quantiles of the estimator $\hat{\nu}_1$ for different sample sizes per estimate and the corresponding relative errors.

of $\hat{\nu}_1$ is higher than the one of $\hat{\nu}_2$ (compare the relative errors of Tables 4.9 to 4.11). It is not surprising that $\hat{c}$ performs better than $\hat{\nu}_1$ and $\hat{\nu}_2$ since the parameters ν_1 and ν_2 determine the tail behavior of the distribution and naturally we do not have many observations in the tails. In particular, the accuracy of $\hat{\nu}_1$ and especially $\hat{\nu}_2$ is very poor for small sample sizes per estimate, i.e. sample size < 1000. For reliable estimates we need at least a sample size of 3000 because of the relative error $e_{rel}^{0.9}$ of Tables 4.10 and 4.11.[16]

In Figure 4.9 all estimates of c are depicted and this highlights again that the scale parameter can be estimated with high accuracy. In the case of ν_1 and sample size

---

[16]$e_{rel}^{1-2\alpha}$ means the the relative error of the estimator is smaller than $e_{rel}^{1-2\alpha}$ with probability $1 - 2\alpha$ measured by the empirical distribution of the estimator. For a definition, see equation (4.33).

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

Table 4.5: Quantiles of the estimator $\hat{\nu}_2$ for different sample sizes per estimate and the corresponding relative errors.
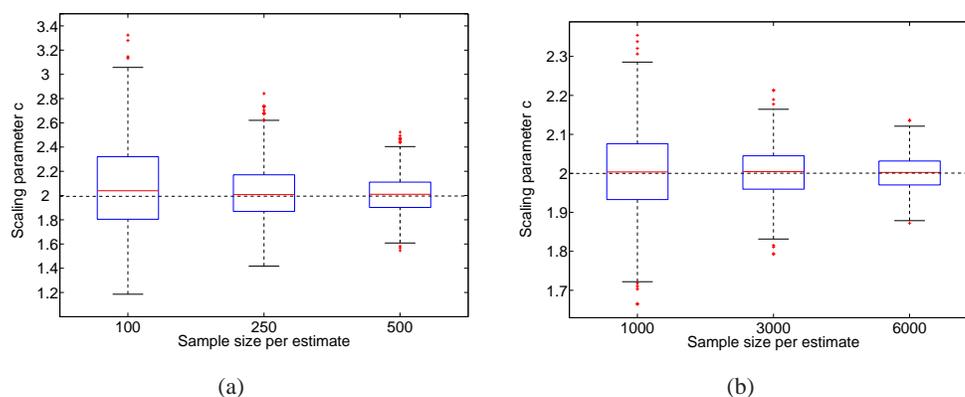
Figure 4.9: Boxplots of the estimate of $\hat{c}$. Each boxplot consists of 1000 estimates.
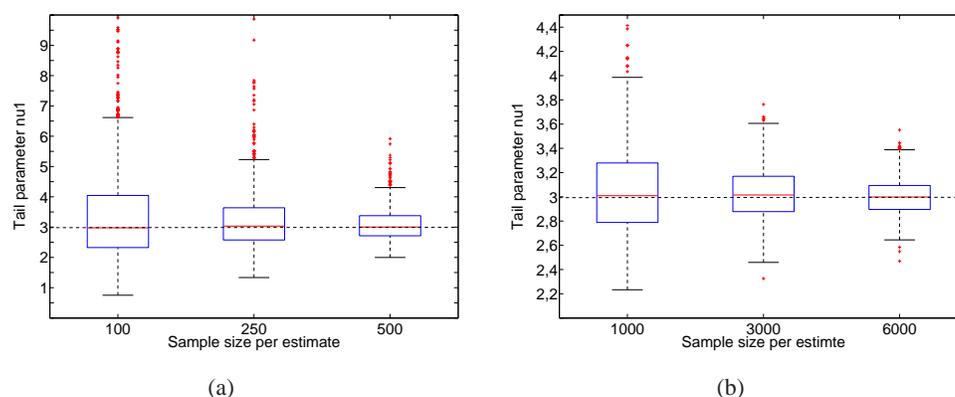
Figure 4.10: Boxplots of estimates of $\hat{\nu}_1$. Each boxplot consists of 1000 estimates.

100, 23 estimates ranging from 10 to 127 are not shown in Figure 4.10 and for sample size 250 only one estimate (value=16.09) is not depicted. Finally, in Figure 4.11 49 estimate are not illustrated ranging from 30.2 to 2794 for sample size 100 and for sample size 250 three estimates (values=$36.6, 33.7, 38.47$) are not depicted.

4.5 Applications

For an empirical analysis of multi-tail elliptical distributions we investigated the daily logarithmic return series for the 29 German stocks included in the DAX index on March 31, 2006. The period covered is May 6, 2002 through March 31, 2006 ($1,000$ daily observations for each stock). In particular, the main focus of our analysis is to empirically assess whether a multi-tail model is superior to a classical elliptical one in the analysis of asset-return behavior.
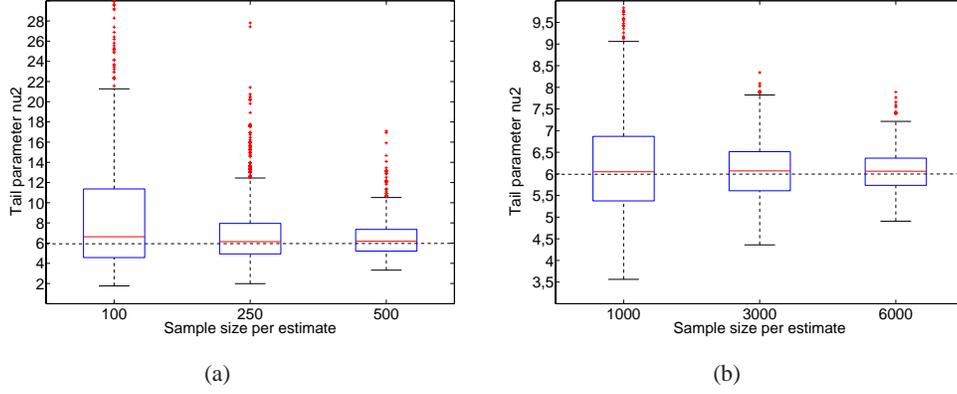
Figure 4.11: Boxplots of estimates of $\hat{\nu}_2$. Each boxplot consists of 1000 estimates.

4.5.1 Two-Dimensional Analysis

Figure 4.12 (a) depicts the two-dimensional scatterplots of BMW versus Daimler-Chrysler and (b) the scatterplots of Commerzbank versus Deutsche Bank. In both figures we can see that there are more outliers in the directions around the first principal component F_1 motivating a multi-tail model. Applying the spectral estimator for both samples we obtain the normalized dispersion matrix ($\hat{\sigma}_{11} = 1$)

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

for BMW versus DaimlerChrysler, and

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

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

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

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

(a)   (b)
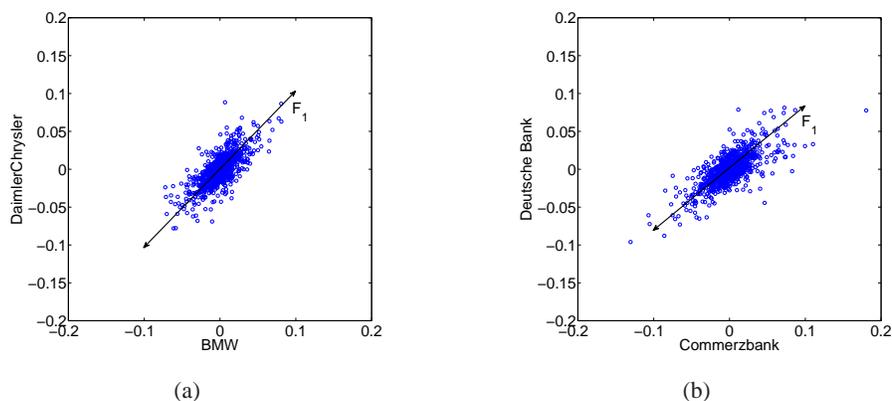
Figure 4.12: Bivariate scatterplot of (a) BMW versus DaimlerChrysler and (b) Commerzbank versus Deutsche Bank. Depicted are daily log-returns from May 6, 2002 through March 31, 2006.

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

Table 4.6: Depicted are the likelihood estimates of the scale parameter and tail parameters for the BMW-DC returns in both models. The table shows the number of parameters (#par), the value of the log-likelihood at the maximum ($\ln L$), the value of the Akaike information criterion (AIC), and the p-value for a likelihood ratio test against the elliptical model. The period investigated is May 6, 2002 through March 31, 2006.

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

Table 4.6 shows the estimates for the scale parameter and tail parameters in both models. In both models the scale parameters are the same while the tail parameters differ. This result is to be expected from our discussion in Section 4.4 because the scaling properties expressed by Σ_0 and c and the tail behavior captured by the tail parameters and the specified tail function are fairly independent for larger sample sizes. The Akaike information criterion[17] prefers the multi-tail model since we observe the

---

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

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

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

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

Table 4.7: Depicted are the maximum likelihood estimates of the scale parameter and tail parameters for the CB-DB returns in both models. The table shows the number of parameters (#par), the value of the log-likelihood at the maximum ($\ln L$), the value of the Akaike information criterion (AIC), and the p-value for a likelihood ratio test against the elliptical model.

smaller value in the multi-tail model. For the maximum likelihood ratio test,[18] we have $\Theta = \{(\nu_1, \nu_2) \in \mathbf{R}^+ : 0 < \nu_1 \leq \nu_2\}$ and the null hypothesis $H_0 : \theta \in \Theta_0 = \{(\nu_1, \nu_2) \in \mathbf{R}^+ : 0 < \nu_1 = \nu_2\}$ against the alternative $H_0 : \theta \in \Theta_0 = \{(\nu_1, \nu_2) \in \mathbf{R}^+ : 0 < \nu_1 < \nu_2\}$. According to Table 4.6, the p-value in this test is less than 2.5%, so it reasonable to reject the elliptical model.

Table 4.7 shows that we obtain basically the same results as in the previous case. The returns for Commerzbank and Deutsche Bank demand even more of a multi-tail model. The spread in the first and second tail parameters is larger than before. The difference between the log-likelihood values and Akaike information criterion is also greater and finally, the p-value of the maximum likelihood ratio test is practically equal to zero. Again, the scaling parameters are close and the tail parameters differ, indicating that the maximum likelihood estimator $\hat{c}$ for the scale parameter is fairly independent of the maximum likelihood estimates for ν_1 and ν_2. In particular, the results depicted in Tables 4.6 and 4.7 coincide with satterplots in Figure 4.12 (a) and (b) since in (b) we observe more pronounced outliers along the first principal component than in (a).

4.5.2 Multi-Tail Elliptical Model Check for the DAX index

The investigated return data $X_1, X_2, ..., X_{1000} \in \mathbf{R}^{29}$ are the 29 German stocks[19] included in the DAX index. The period covered is May 6, 2002 to March 31, 2006. We start our analysis by estimating the normalized dispersion matrix $\hat{\Sigma}_0(X_1, ..., X_{1000})$ using the spectral estimator is depicted in Figure 4.13.

Figure 4.15 (a), (b), and (c) show the factor loadings (eigenvectors) g_1, g_2 and g_3 of the first three principal components. Figure 4.15 (d) depicts the eigenvalues

---

[18]The likelihood ratio test statistic satisfies

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

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

[19]We excluded HypoRealEstate Bank because we did not have sufficient return data for this stock for the period covered.
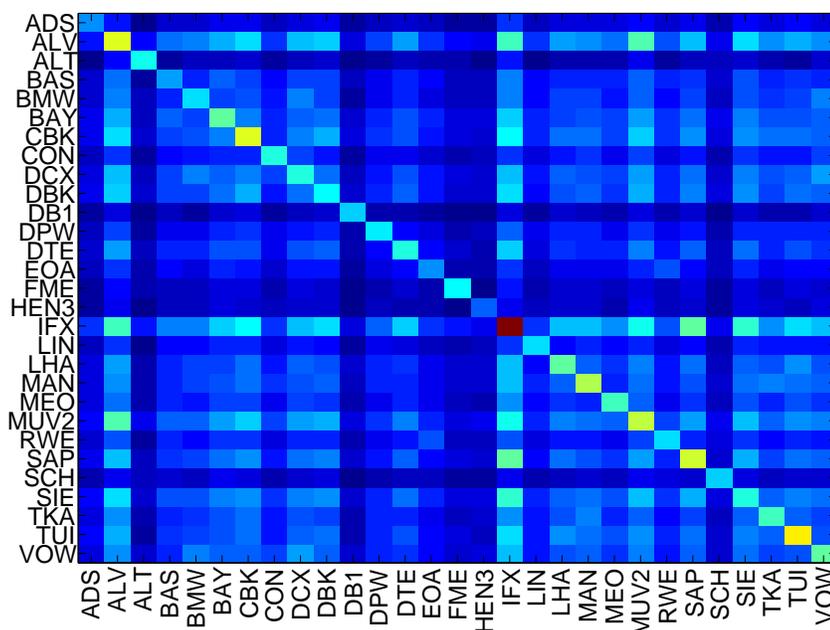
Figure 4.13: Heat map of the sample dispersion matrix estimated by the spectral estimator.

of the normalized dispersion matrix obtained by the spectral estimator. We see that the eigenvalue of the first principal component is significantly larger than the others. The first vector of loadings is positively weighted for all stocks and can be thought of as describing a kind of index portfolio. We obtain the corresponding time series of principal components through

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

Figure 4.15 (a), (b), and (c) illustrate the pairwise scatterplots of these components. We can see in Figure 4.15 (a) and (b) that the scatterplots are stretched along the first principal component. This scaling behavior is caused by the large first eigenvalue of F_1. Moreover, it is important to note that we observe many outliers along F_1. This phenomenon may be attributed to smaller tail parameters in the directions around F_1 and $-F_1$. In Figure 4.15 (c) both principal components have fundamentally the same scale and the outliers are not so pronounced as in the former ones, suggesting a similar tail behavior.

Figure 4.16 depicts a three-dimensional scatterplot of the first three principal components. This figure confirms the scaling properties and tail behavior observed in Figure 4.15.

The visual analysis of Figures 4.15 and 4.16 motivates a multi-tail model for the
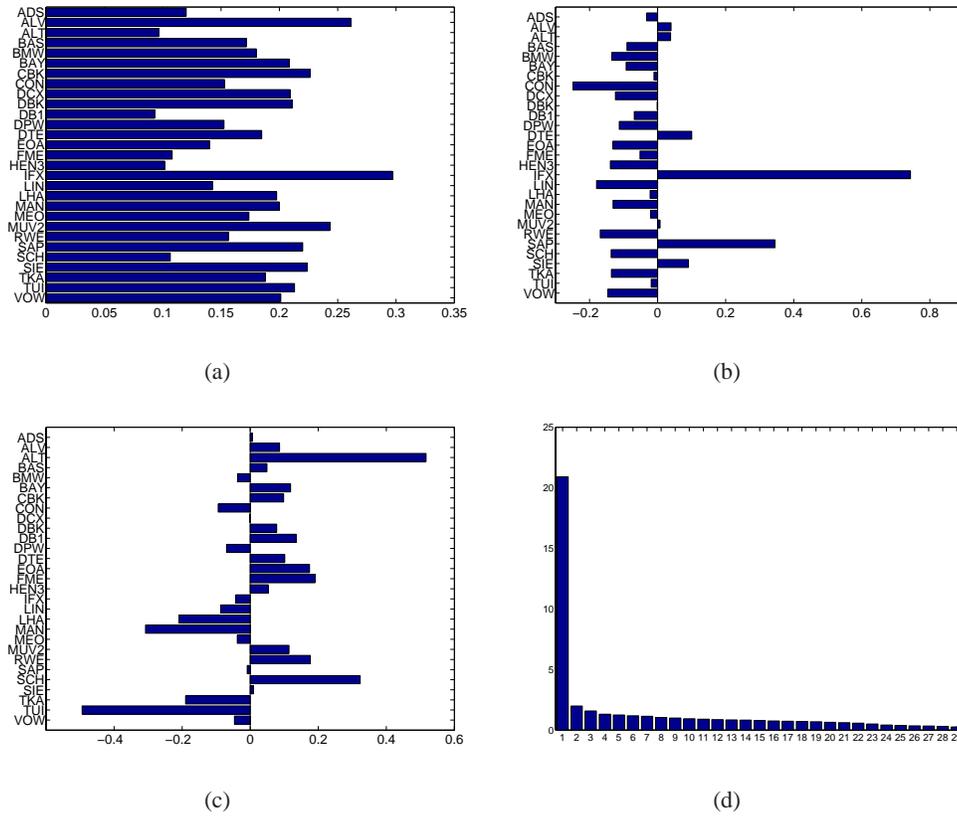
Figure 4.14: Barplot summarizing the loadings vectors g_1, g_2, g_3 and g_4 defining the first four principal components: (a) factor 1 loadings; (b) factor 2 loadings; (c) factor 3 loadings. (d) depicts the eigenvalues of the normalized dispersion matrix.

(a)

(b)

(c)
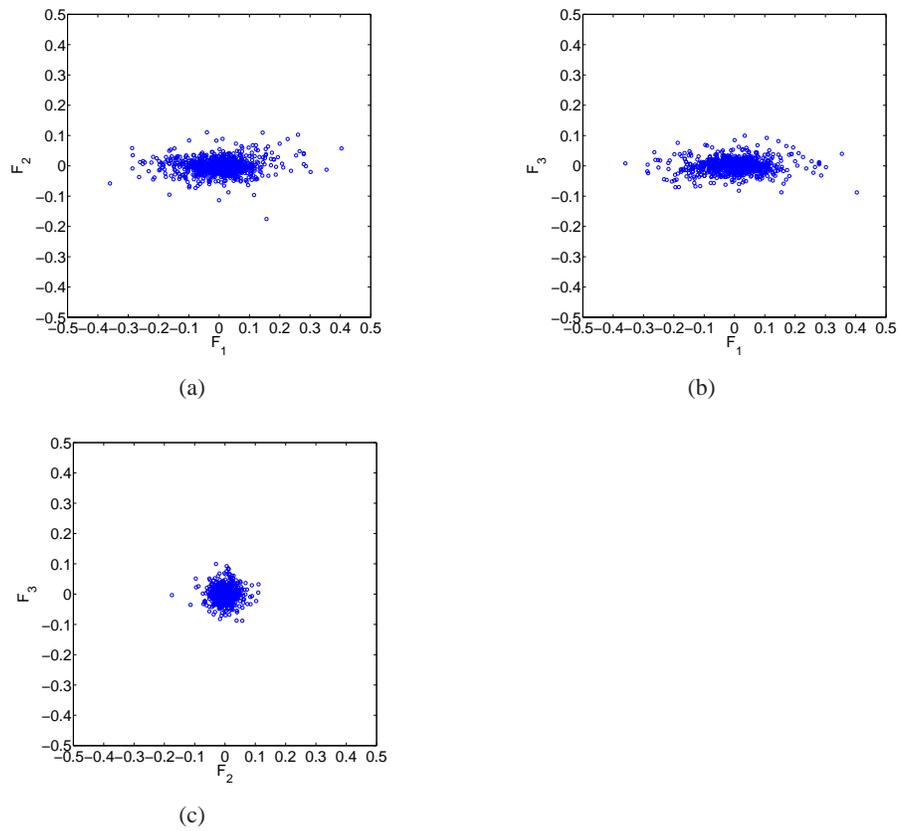
Figure 4.15: Figure (a),(b), and (c) show the pairwise, two-dimensional scatterplots of the first three principal components. The period covered is May 6, 2002 through March 31, 2006.
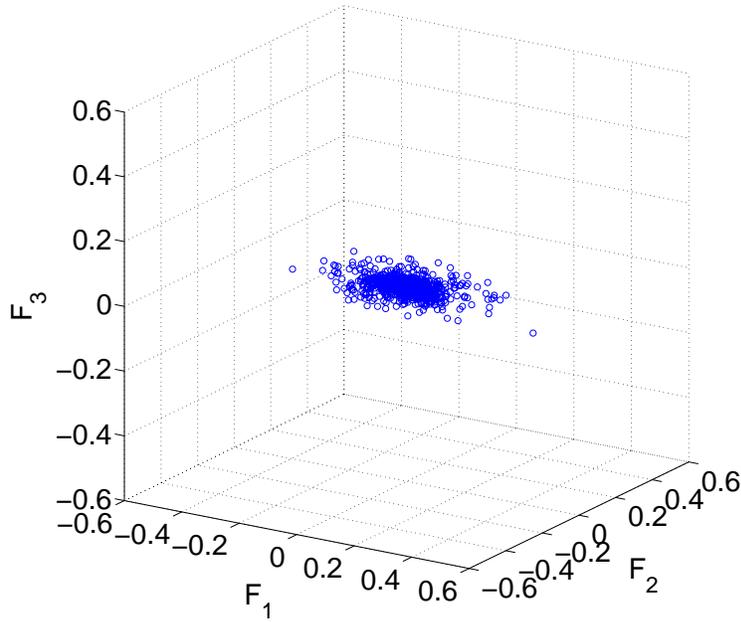
Figure 4.16: Figure depicts a three-dimensional scatterplot of the first three principal components.

logarithmic returns investigated. Since the first principal component differ from the others, we propose the tail function

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

As we did in Section 4.5.1, we compare this multi-tail model with an elliptical one which, per definition, has a constant tail function ($\nu(s) = \nu_0$, $s \in \mathcal{S}^{28}$). We conduct the same statistical analysis as in the previous section. In particular, we fit a t and multi-tail t distribution to the data. The results are reported in Table 4.8. The elliptical model has 29 location parameters, $29 \cdot 15 = 435$ dispersion parameters, and one tail parameter, while the multi-tail model has two tail parameters. Thus, we have 465 parameters in the elliptical and 466 in the multi-tail model. In the first step we estimate the dispersion parameter with the spectral estimator up to a scaling constant. The scale parameter $\hat{c}$ and tail parameters $\hat{\nu}_1$ and $\hat{\nu}_2$ are estimated in a second step. Table 4.8 shows that in both models the scale parameter $\hat{c}$ are almost the same, whereas the tail parameters differ significantly. The Akaike criterion as well as the likelihood ratio test do favor the multi-tail model. A p-value of less than 0.05% for this test indicates that we can reject the null hypothesis of an elliptical model at a very high confidence level.

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

Table 4.8: Depicted are the maximum likelihood estimates of the scale parameter and tail parameters for the DAX returns in both models. The table shows the number of parameters (#par), the value of the log-likelihood at the maximum ($\ln L$), the value of the Akaike information criterion (AIC), and the p-value for a likelihood ratio test against the elliptical model.

4.5.3   Summary of the Results

In the statistical analysis reported in Sections 4.5.1 and 4.5.2 we see that the simplest genuine multi-tail elliptical model significantly outperforms the classical elliptical model. The hypothesis of homogeneous tail behavior can be rejected in any of the data sets investigated. In particular, we see in the analysis of the stocks included in the DAX index that the first principal component is heavier tailed than the others.

4.6   Conclusion

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

4.7   Appendix

Definition 48. *The* gamma function Γ *is defined by*

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

where $x > 0$.

Definition 49. *Let $A \in \mathbf{R}^{d\times d}$ be a matrix and $t \in \mathbf{R}$. Then t^A is defined by*

$$t^A = I + \sum_{m=1}^\infty \frac{(\log(t)A)^m}{m!}. \tag{4.38}$$

It can be shown that the limit in equation (4.38) exists for all matrices $A \in \mathbf{R}^{d\times d}$. The matrix A in t^A is called the exponent of t. Furthermore, we have the following properties.

Proposition 10. Let $t \in \mathbf{R}$ and $A, B \in \mathbf{R}^{d\times d}$. Then we have

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

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

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

The proof can be found in Rubin (1986).

Theorem 24 (**Change-of-variables theorem**). *Suppose that*

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

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

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

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

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

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

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

Lemma 4. *Let the functions g, g_1 and g_2 be defined by*

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

where $A \in \mathbf{R}^{d \times d}$ is regular and $\mu \in \mathbf{R}^d$. Then we have

(i) $g = g_2 \circ g_1$

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

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

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

Proof.   (i) This is obvious.

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

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

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

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

(iv) We have

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

$\square$

Lemma 5. *Let $V, W \subset \mathbf{R}^d$ be open and $f : V \to W$ differentiable with inverse f^{-1}. Then we have*

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

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

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

The lemma follows now immediately because of

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

$\square$

Corollary 8. *Let g, g_1 and g_2 as in Lemma 4. Then we have*

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

Proof. The inverse of g satisfies

$$g^{-1} : \mathbf{R}^d \to \mathbf{R}_{>0} \times \mathcal{S}^{d-1}, x \mapsto \left(\sqrt{(x-\mu)'\Sigma^{-1}(x-\mu)}, \frac{A^{-1}(x-\mu)}{||A^{-1}(x-\mu)||} \right) \quad (4.39)$$

We have

$$
\begin{aligned}
\det(D(g_1 \circ g_2)^{-1}(x)) &= \det(D(g_1^{-1} \circ g_2^{-1})(x)) \\
&= \det(Dg_1^{-1}(g_2^{-1}(x))Dg_2^{-1}(x)) \\
&= \det(Dg_1^{-1}(g_2^{-1}(x))) \det(Dg_2^{-1}(x)) \\
&\overset{\text{Lemma 5}}{=} \frac{1}{\det(Dg_1(g_1^{-1} \circ g_2^{-1}(x)))} \frac{1}{\det(Dg_2(g_2^{-1}(x)))} \\
&= \frac{1}{\det(Dg_1(g^{-1}(x)))} \frac{1}{\det(A)} \\
&\overset{(1)}{=} \frac{\left(\sqrt{(x-\mu)'\Sigma^{-1}(x-\mu)}^{d-1}\right)^{-1}}{\det(\Sigma)^{1/2}}. \quad (4.40)
\end{aligned}
$$

(1) holds in equation (4.40), since $Dg_1((r,s)) = r^{d-1}$ and equation (4.39). $\square$

Theorem 25. *Let $M \subset \mathbf{R}^{1 \times d^2}$ be the open subset of d-dimensional regular matrices. Then the function $h : M \to \mathbf{R}\ A \mapsto \det(A)$ has the Jacobian*

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

$$(4.41)$$

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

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

$$(4.42)$$

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

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

$\square$

Chapter 5

Conclusion

Chapter 1 gives an overview about different kinds of factor models.

Chapter 2 deals with the estimation of α-stable sub-Gaussian distributions for asset returns. Fitting multivariate α-stable distributions to data is still not feasible in higher dimensions since the (non-parametric) spectral measure of the characteristic function is extremely difficult to estimate in dimensions higher than 2. This was shown by Nolan, Panorska, and McCulloch (2001). α-stable sub-Gaussian distributions are a particular (parametric) subclass of the multivariate α-stable distributions. We present and extend a method based on Nolan (2005) to estimate the dispersion matrix of an α-stable sub-Gaussian distribution and estimate the tail index α of the distribution. In particular, we develop an estimator for the off-diagonal entries of the dispersion matrix that has statistical properties superior to the normal off-diagonal estimator based on the covariation. Furthermore, this approach allows estimation of the dispersion matrix of any normal variance mixture distribution up to a scale parameter. We demonstrate the behaviour of these estimators by fitting an α-stable sub-Gaussian distribution to the DAX30 components. Finally, we conduct a stable principal component analysis and calculate the coefficient of tail dependence of the prinipal components.

In chapter 3 we present a new type of multivariate GARCH model which we call the composed MGARCH and factor composed MGARCH models. We show sufficient conditions for the covariance stationarity of these processes and proof of the invariance of the models under linear combinations, an important property for factor modeling. Furthermore, we introduce an α-stable version of these models and fit a four dimensional α-stable composed MGARCH process to the returns on four German stocks included in the DAX index. We show in an in-sample analysis as well as in an out-of-sample analysis that the model outperforms the classical exponentially weighted moving average (EWMA) model introduced by RiskMetrics.

In chapter 4 we present a new type of multivariate distributions for asset returns which we call the multi-tail elliptical distributions. Multi-tail elliptical distribution

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

List of Figures

List of Tables

Bibliography

Artzner, P., F. Delbaen, J.M. Eber, and D. Heath. (1999). "Coherent measure of risk." *Mathematical Finance* 9, 203-228.

Alexander, C.O. (2001). *Market Models: A Practitioner's Guide to Financial Data Analysis.* Chichester, West Sussex: Wiley.

Alexander, C.O. (2002). "Principal Component Models for Generating Large GARCH Covariance Matrices." *Economic Notes* 31(2), 337-359.

Attansio, O. (1991). "Risk, time-varying second moments and market efficiency." *Review of Economic Studies* 58, 479-494.

Bauwens, L., S. Laurent, and J.V.K. Rombouts. (2006). "Multivariate GARCH models: A Survey." *Journal of Applied Econometrics* 21, 79-109.

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

Bera, A.K., and M.L. Higgins. (1993). "ARCH models:properties, estimation and testing." *Journal of Economic Surveys* 7, 305-362.

Beutelspacher, A. (2003). *Lineare Algebra.* Wiesbaden, Hessen: Vieweg.

Billio M., M. Caporin, and M. Gobbo. (2003). "Block dynamic conditional correlation multivariate GARCH models." Greta Working Paper n.03.03, University Ca' Foscari, Venice.

Bingham, N. H., R. Kiesel, and R. Schmidt. (2003). "A semi-parametric approach to risk management." *Quantitative Finance* 3(6), 426-441.

Bollerslev, T. (1986). "Generalized autoregressive conditional heteroscedasticity." *Journal of Econometrics* 31, 307-327.

Bollerslev, T., F. Engle, and J. M. Wooldrige. (1988). "A capital asset pricing model with time-varying covariances." *Journal of Political Economy* 96, 116-131.

Bollerslev, T. (1990). "Modelling the coherence in short run nominal exchange rates: A multivariate generalized arch model." *The Review of Economics and Statistics* 72, 498-505.

Bollerslev, T., R.Y. Chou, and K.F. Kroner. (1992). "ARCH modeling in finance: a review of the theory and empirical evidence." *Journal of Econometrics* 52,5-59.

Borovkova, S. (2006). "Detecting Market Transitions and Energy Futures Risk Management using Principal Components." *The European Journal of Finance* 12(6-7), 495-512.

Chan, L.K.C., J. Kareceski, and J. Lakonishok. (1998). "Risk and Return from Factors." *Journal of Financial and Quantitative Analysis* 33, 159-188.

Chen, B.N., and S.T. Rachev (1995). "Multivariate stable securities in financial markets." *Mathematical Finance*.

Chen, N.F., R. Roll, and S.A. Ross. (1986). "Economic forces and the stock market." *Jorunal of Business* 59(3),383-403.

Christodoulakis G.A., and S.E. Satchell. (2002). "Correlated ARCH: modeling the time-varying correlation between financial asset returns." *European Journal of Operations* 139, 351-370.

Clare, A.D., and S.H. Thomas. (1994). "Macroeconomic factors, the APT and the UK stock market." *Journal of Business Finance and Accounting* 21(3), 309-330.

Cochrane, J.H. (2001). *Asset Pricing*. Princeton, New Jersey: Princeton University Press.

DuMouchel, W.H. (1973). "On the asymptotic normality of the Maximum-Likelihood Estimate when Sampling from a Stable Distribution." *Annals of Statistics* 1, 948-57.

Doganoglu, T., C. Hartz, and S. Mittnik. (2006). "Portfolio Optimization when Risk Factors are Conditionally Varying and Heavy Tailed." CFS working paper 2004-26, University of Frankfurt.

Ding, Z. (1994). Ding, Z., and R. Engle. (1994). "Large Scale Conditional Covariance Matrix Modelling, Estimation and Testing." UCSD Working Paper, San Diego, CA.

Eberlein, E., and U. Keller. (1995). "Hyperbolic distribution in finance." *Bernoulli* 1(3), 281-299.

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

Embrechts, P., A. McNeil, and D. Straumann. (1999). "Correlation: Pitfalls and Alternatives." Preprint, ETH Zrich.

Engle, R. F. (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation." *Econometrica* 50, 987-1007.

Engle, R. F., and K. F. Kroner. (1995). "Multivariate simultaneous generalized arch." *Econometric Theory* 11, 122-150.

Engle, R. F., and K. Sheppard. (2001). "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH." Mimeo, UCSD.

Engle, R.F. (2002). "Dynamic conditional correlation-a simple class of multivariate GARCH models." *Journal of Business and Economic Statistics* 20, 339-350.

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

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

Fama, E. (1991). "Efficient Capital Markets:II." *Journal of Finance* 46(5), 1575-1617.

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

Fama, E., and K. French. (1993). "Common ris factors in the returns on stocks and bonds." *Journal of Financial Economics* 33, 3-56.

Fama, E., and K. French. (1996). "Multifactor explanations of asset-pricing anomalies." *Journal of Finance* 47, 426-465.

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

Frahm, G. (2004). *Generalized elliptical distributions: Theory and Applications.* PhD-thesis, University of Cologne, Germany.

Hamilton, J.D. (1994). *Time Series Analysis.* Princeton, New Jersey: Princeton University Press.

Hardin Jr., C. D. (1984). "Skewed stable variables and processes." Technical report 79, Center for Stochastics Processes at the University of North Carolina, Chapel Hill.

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

Hill, B.M. (1975). "A simple general approach to inference about the tail of a distribution." *Annals of Statistics* 3, 1163-1174.

Höchstötter, M., S.T. Rachev, and F. J. Fabozzi. (2005). "Distributional Analysis of the Stocks Comprising the DAX 30." *Probability and Mathematical Statistics* 25, 363-383.

Jagannaghan, R., and Z. Wang. (1996). "The conditional CAPM and the cross-section of expected returns." *Journal of Finance* 51, 3-53.

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

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

Jondeau, E., and M. Rockinger. (2006). "The copula-GARCH model of conditional dependencies: an international stock-market application." *Journal of International Money and Finance* 25(5), 827-853.

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

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

Kim, J., A. M. Malz, and J. Mina. (1999). *Long Run Technical Document*. RiskMetrics Group.

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

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

Kring, S., S.T. Rachev, M. Höchstötter, and F.J. Fabozzi. (2007). "Composed and factor composed multivariate GARCH models." submitted to *Journal of Financial Econometrics*.

Kring, S., S.T. Rachev, M. Höchstötter, F.J. Fabozzi. (2007). "Multi-tail elliptical distributions." submitted to *The Econometrics Journal*.

Lintner, J. (1965a). "Security proces, risk and maximal gains form diversification." *Journal of Finance* 20, 587-615.

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

Mandelbrot, B. B. (1963a). "New Methods in Statistical Economics." *Journal of Political Economy* 71, 421-440.

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

Mandelbrot, B. B. (1963c). "New methods in statistical economics." *Journal of Political Economy* 71,421-440.

Markowitz, H. M. (1952). "Portfolio Selection." *The Journal of Finance* 7(1), 77-91.

McCulloch, J.H. (1985). "Interest-risk sensitive deposit insurance premia: stable ACH estimates." *Journal of Banking and Finance* 9(1), 137-156.

McCulloch, J. H. (1996). "Financial Applications of Stable Distributions." In G.S. Maddala, C.R. Rao (eds.), *Hanbbook of Statistics*, vol. 14. Amsterdam, Netherland: Elsevier.

McElroy, M., and E. Burmeister. (1988). "Arbitrage Pricing Theory as a restricted non-linear multivariate regression model." *Journal of Business and Economic Statistics* 6(1), 29-42.

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

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

Meerschaert, M.M., and H.-P. Scheffler. (2003). "Portfolio modeling with heavy tailed random vectors." In S.T. Rachev (eds.), *Handbook of Heavy Tailed Distributions in Finance*, vol. 1. North-Holland: Elsevier.

Mittnik, S., M.S. Paolella, and S.T. Rachev. (1998). "A tail estimator for the index of the stable Paretian distribution." *Communications in Statistics-Theory and Methods* 27, 1239-1262.

Mittnik, S., M.S. Paolella, and S.T. Rachev. (2002). "Stationarity of stable power-GARCH processes." *Journal of Econometrics* 106, 97-107.

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

Murray, M.P. (2006). *Econometrics A Modern Approach.* Boston, MA: Pearson International Edition.

Nolan, J. P., A.K. Panorska, and J.H. McCulloch. (2001). "Estimation of stable spectral measure." *Mathematical and Computer Modelling* 34, 1113-1122.

Nolan, J. P. (2005). "Multivariate stable densities and distribution functions: general and elliptical case." Deutsche Bundesbank's 2005 Annual Fall Conference.

Patton, A. 2000. "Modelling time-varying exchange rate dependence using the conditional copula." UCSD Discussion Paper 01-09, San Diego, CA.

Panorska, A., S. Mittnik, and S.T. Rachev. (1995). "Stable GARCH Models for Financial Time Series." *Applied Mathematics Letters* 8, 33-37.

Pelletier, D. (2003). "Regime switching for dynamic correlations." *Journal of Econometrics* 131, 445-473.

Pitsillis, M. (2005). "Review of literature on multifactor asset pricing models." In J. Knight and S. Satchell (eds.), *Linear Factor Models in Finance* Burlingtion, MA: Elsevier Finance.

Poon, S., and J. Taylor. (1991). "Macroeconomic factors and the UK stock market." *Journal of Business Finance and Accounting* 18(5), 619-636.

Rachev, S. T., and S. Mittnik. (2000). *Stable Paretian Models in Finance.* Chichester, West Sussex: Wiley.

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

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

Reyfman, A. (1997). *Labour market risk and expected asset returns.* PhD thesis, University of Chicago. Cited in Cochrane.

RiskMetrics. (1996). *Riskmetrics Technical Document.* 4th edn. J.P. Morgan: New York.

Ross, S.A. (1976). "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13, 341-360.

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

Shephard, N. (1996). "Statistical aspects of ARCH and stochastic volatility." In O.E. Barndorff-Nielsen, D.R. Cox, and D.V. Hinkley (eds.), *Time Series Models in Econometrics, Finance and Other Fields.* London: Chapman & Hall.

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

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

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

Sharpe, W. (1963). "A Simplified Model for Portfolio Analysis." *Management Science* 9(3), 277-293.

Sharpe, W. (1964). "Capital asset prices: A theory of market equilibrium under conditions of risk." *Journal of Finance* 19, 425-442.

Stoyanov, S., V. (2005). *Optimal Portfolio Management in Highly Volatile Markets.* PhD thesis, University of Karlsruhe, Germany.

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

Stoyanov, S., V., and B. Racheva-Iotova. (2004). "Univariate stable laws in the fields of finance-parameter estimation." *Journal of Concrete and Applicable Mathematics.* 2(4), 24-49.

Tse, Y. K., and A.K.C. Tsui. (2002). "A multivariate GARCH model with time-varying correlations." *Journal of Business and Economic Statistics* 20, 351-362.

Tsui, A.K., and Q. Yu. (1999). "Constant conditional correlation in a bivariate garch model: Evidence from the stock market in china." *Mathematics and Computers in Simulation* 48. 503-509.

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

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

Zangari, P. (2003). "Equity Risk Factor Models." *Modern Investment Management-An Equilibrium Approach*. Hoboken, New Jersey: Wiley Finance.

Lebenslauf

Vorname		Kring
Name		Sebastian
Adress		Veilchenstrae 17
		76131 Karlsruhe
Geburtsdatum		25.05.1978
Geburtsort		Düsseldorf

Schule

Grundschule	$1985 - 89$	Grundschule Haiger-Dillbrecht
Gynmnasium	$1989 - 98$	Wilhelm von Oranien Schule Dillenburg
(Abschluss mit Abitur)		

Studium

Mathematik (Diplom)	$1999 - 2005$	Universität Karlsruhe (TH) Fakultät für Mathematik
Informatik (Diplom)	$1999 - 2004$	Universität Karlsruhe (TH) Fakultät für Informatik

Promotion

Wissenschaftlicher Mitarbeiter	Aug 2005–	Lehrstuhl für Statistik u. math. Wirtschaftstheorie

Nebentätigkeiten

Hilfswissenschaftlicher Mitarbeiter	Juni/Juli 2005	Lehrstuhl für Statistik u. math. Wirtschaftstheorie

Sebastian Kring
Veilchenstrae 17
76131 Karlsruhe

Erklärung

(gem 4, Abs. 4 der Promotionsordnung vom 21.4.1989)

Ich versichere wahrheitsgemäß, die Dissertation bis auf die in der Abhandlung angegebene Hilfe selbständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und genau kenntlich gemacht zu haben, was aus Arbeiten anderer und aus eigenen Veröffentlichungen unverändert oder mit Abänderungen entnommen wurde.

Karlsruhe, den 23.10.2007