MODELLING, ESTIMATING AND VALIDATING MULTIDIMENSIONAL DISTRIBUTION FUNCTIONS -WITH APPLICATIONS TO RISK MANAGEMENT-

By

Markus Junker

D 368

Vom Fachbereich Mathematik der Technischen Universität Kaiserslautern zur Verleihung des akademischen Grades Doktor der Naturwissenschaften (Doctor rerum naturalium, Dr. rer. nat.) genehmigte Dissertation

1. Gutachter: Prof. Dr. Jürgen Franke
2. Gutachter: Prof. Dr. Dietmar Pfeifer

Tag der Disputation: 15.12.2003

© Copyright by Markus Junker, 2003

To the dependence that never will be modelled,

estimated or validated,

LOVE

Table of Contents

Table of Contents **iv**

Abstract **viii**

Acknowledgements **ix**

Introduction **1**

1 A Probability Density Applying to Asymmetric Pareto Type Tails **4**
1.1 Introduction 5
1.2 A smooth distribution with Pareto tails 6
1.3 Model Specification 14
1.4 Estimation 15
1.4.1 The Homoscedasticity Model 16
1.4.2 The Conditional Heteroscedasticity Model 19
1.5 VaR and backtesting 21

2 Copula Theory **25**
2.1 The Copula Concept 25
2.1.1 Rotational and Permutational Symmetry 26
2.2 Dependence Concepts 28
2.2.1 Measures of Dependence 30
2.3 Parametric Copula Classes 31
2.4 Pitfalls in Copula Theory 33

3 Elliptical Copulas: Applicability and Limitations **37**
3.1 Introduction 37
3.2 Copula and Marginal Parameters 38
3.3 Elliptical Distributions 39
3.3.1 Characterizations 39
3.3.2 Dependence Structure 41

4 Archimedean Copulas-
2-dimensional Flexibility, Multidimensional Complexity 46
4.1 Properties 46
4.2 A Transformation Rule for Generators 48
4.2.1 A family covering the Gumbel and Frank families 50
4.3 The multidimensional case 55
4.3.1 More multidimensional extension rules 58

5 An asymmetric Extreme Value Copula 63
5.1 An Asymmetric extreme value copula 63

6 Multidimensional Goodness-of-Fit Tests 67
6.1 Multidimensional χ^2 tests 67
6.1.1 A mapping rule leading to joint uniform independence 68
6.1.2 χ^2 test with class separation 69
6.1.3 Estimation and χ^2 testing using the distribution of differences . . 73
6.2 More goodness-of-fit tests 76

7 Estimating the Tail-Dependence Coefficient 80
7.1 Introduction 80
7.2 Basic definitions 81
7.3 Estimation methods for the tail-dependence coefficient 83
7.3.1 Estimation from the entire sample 83
7.3.2 Estimation from the distribution's tail or extreme sample 88
7.3.3 A final remark about pitfalls 95
7.4 Simulation Study 97
7.4.1 Pre-simulation 98
7.4.2 Simulation and estimation 102
7.4.3 Discussion of the results 106
7.5 Conclusion 108

8 Copula Estimations for Risk Factor Portfolios 111
8.1 Pre-considerations 111
8.1.1 Pre-selection of the models 112
8.1.2 Estimation method 116
8.1.3 Diagnostics 116
8.1.4 Data sets, data preparation and filtering 118
8.2 Estimation results and discussion 120
8.2.1 2-dimensional data 121
8.2.2 4-dimensional data 123
8.3 Conclusion 126

9 Measurement of Aggregate Risk with Copulas **127**
9.1 Copula specification . . . 127
9.2 Value at Risk calculation under different distribution assumptions . . . 129

10 European Real Estate Firms in Crash Situations **137**
10.1 Introduction . . . 137
10.2 Mathematical Framework . . . 139
10.3 Empirical studies . . . 140
10.3.1 Single asset studies . . . 141
10.3.2 Portfolio investigations . . . 143
10.4 Conclusion . . . 146

11 Nonlinear Term Structure Dependence **151**
11.1 Introduction . . . 151
11.2 The Term Structure Model . . . 154
11.2.1 Term Structure Dynamics . . . 154
11.2.2 Nonlinear Term Structure Dependence . . . 157
11.3 Empirical Analysis of Nonlinear Term Structure Dependence . . . 158
11.3.1 The Sample . . . 158
11.3.2 Term Structure Model Estimation . . . 160
11.3.3 Derivation of Term Structure Innovations . . . 160
11.3.4 Analysis of the Term Structure Innovations . . . 162
11.4 Application: Measuring Bond Portfolio Risk . . . 174
11.5 Conclusion . . . 179

12 Generalized Elliptical Distributions - Models and Estimation **180**
12.1 Motivation . . . 180
12.2 Generalized Elliptical Distributions . . . 181
12.2.1 Definition . . . 181
12.2.2 Spectral Density . . . 184
12.2.3 Estimation . . . 186
12.3 Models . . . 192

Conclusion **197**

Bibliography **198**

A A Probability Density Applying to Asymmetric Pareto Type Tails **209**
A.1 Proof of Proposition 1.2.4: . . . 209
A.2 Lemma A.2.1 . . . 210

B Elliptical Copulas: Applicability and Limitations **211**
B.1 Multidimensional Sub-Gaussian Standard Density 211
B.2 Equivalent Representation of Equation (3.3.5) 212

C Estimating the Tail-Dependence Coefficient **214**

D Copula estimations for risk factor portfolios **217**

E Nonlinear Term Structure Dependence **232**
E.1 Affine Models 232
E.2 Yield-Factor Representation 234

Abstract

The question of how to model dependence structures between financial assets was revolutionized since the last decade when the copula concept was introduced in financial research. Even though the concept of splitting marginal behavior and dependence structure (described by a copula) of multidimensional distributions already goes back to Sklar (1955) and Hoeffding (1940), there were very little empirical efforts done to check out the potentials of this approach.

The aim of this thesis is to figure out the possibilities of copulas for modelling, estimating and validating purposes. Therefore we extend the class of Archimedean Copulas via a transformation rule to new classes and come up with an explicit suggestion covering the Frank and Gumbel family. We introduce a copula based mapping rule leading to joint independence and as results of this mapping we present an easy method of multidimensional χ^2 testing and a new estimate for high dimensional parametric distributions functions. Different ways of estimating the tail dependence coefficient, describing the asymptotic probability of joint extremes, are compared and improved. The limitations of elliptical distributions are carried out and a generalized form of them, preserving their applicability, is developed. We state a method to split a (generalized) elliptical distribution into its radial and angular part. This leads to a positive definite robust estimate of the dispersion matrix (here only given as a theoretical outlook). The impact of our findings is stated by modelling and testing the return distributions of stock-, currency-, oil related commodities- and LME metal baskets; showing the crash stability of real estate based firms and the existence of nonlinear dependence in between the yield curve.

Acknowledgements

First thanks are acknowledged to my supervisor Professor Dr. Jürgen Franke. Without his feeling for praxis relevant topics and his comprehension for the demands of an external "long distance" doctorate, it wouldn't have been possible to realize this thesis.

Science needs the discussion and team work with other researchers to be able to grow. So here is a capital thank for my coworkers for the fruitful collaboration THANKS: G. Frahm, Dr. A. May, Dr. R. Schmidt, Dr. A. Szimayer and Dr. N. Wagner

Further I would like to thank the center of advanced european studies and research (caesar) for giving me the opportunity of writing this thesis in a practical oriented surrounding and for the flexible working times (my credo "carpe noctem, salute Eos"). Special thanks are also given to our project partners WestLB and IVG AG for their data support.

Working in the Financial Engineering Group at caesar was from the beginning a pleasure. A very open atmosphere extended work to be fun and colleges to be friends. Thanks Alex for the philosophic discussions at 2 am, Alexei for the Vodka, Angelika for her support, Annett for the "sächsisch" lessons and her laughing, Gabriel for the hottest (not only mathematical) disputes, Marcus for showing me easy living and Stefan for listening and his mathematical exactness.

Finally, I wish to thank all my friends that relationship to this thesis was to stop me thinking and working *all* the time for it.

Bonn, Markus Junker
August 5, 2003

Introduction

The copula concept goes back to Sklar (1955) and Hoeffding (1940). The idea is to separate the univariate marginal behavior in each dimension, given by the univariate distribution functions $F_1, \ldots, F_n$, and the dependence structure of an n-dimensional random vector X, given by a copula C_X. Sklar has shown that the relation $H(x) = C_X\left(F_1\left(x_1\right), \ldots, F_n\left(x_n\right)\right)$ is fulfilled for all $(x_1, \ldots, x_n)$ in the support of H, and for a (unique, if $F_1, \ldots, F_d$ are continuous) copula C_X. With other words C_X can be defined as the distribution function of the n-dimensional random variable $\left(F_1\left(X_1\right), \ldots, F_n\left(X_n\right)\right)$, that has uniformly distributed margins $F_i\left(X_i\right) \sim U(0,1)$.

Naturally the question *"What is the benefit of using copulas?"* arises. Well, roughly half a century later we are able to provide some answers. This thesis aims to add some new ones, as well as to show how to take advantage in the use of copulas for solving financial problems.
1. The copula theory provides a model kit for the construction of multivariate distribution functions. Suppose e.g. one wants to describe the 2-dimensional return distribution of a portfolio consisting of a stock and a large credit basket. The log-returns of the stock may assumed to be normal distributed, whereas the return of the credit basket follows e.g. a normal-inverse distribution. Without copulas it will be a hard task to couple these totally different univariate distributions to a joint 2-dimensional one. However, assuming the dependence structure between the stock and the credit basket to be normal, we can define as the joint return distribution $H(s, cb) := C_N\left(N(s), NID(cb)\right)$, where N, NID are the appropriate normal, and normal-inverse distributions respectively and C_N denotes the normal copula. So copulas help *modelling* financial problems.
2. Since we can split via copulas in marginal and dependence behavior, we can also split the estimation of the parameters regarding the marginals and the ones corresponding to the dependence structure. Suppose e.g. we have to estimate a 10-dimensional normal distribution. This leads to the estimation of $20 + 45$ parameters. 20 for the location and scale of the margins, and 45 for the correlation matrix. With the use of copulas we can first estimate 10 times 2 parameters of

every marginal and then use the pseudo observations $\left(\widehat{N}_1(x_1), \ldots, \widehat{N}_n(x_n)\right)$ to estimate C_N, what might have a stabilizing effect on the estimation. Furthermore, having the copula as a dependence function, we can measure and estimate special structures like e.g. the probability of joint extremes. So copulas can improve *estimating* procedures.
3. Copulas can also play an important rule in testing and *validating* assumed distribution models. With copulas it is possible to test every marginal and copula assumption independently. So we do not need to reject a complete model but can identify the "worst" components to be exchanged.

Mostly every chapter has its own more detailed introduction and short conclusion. So here we give a brief scheme of the thesis. In Chapter 1 we introduce an univariate smooth distribution function that allows for different tail indices at the lower and upper tail. Surprisingly it seem that such an asymmetry is not covered by the existing models like skewed t- or hyperbolic distribution.
Chapter 2 states the theoretical background about copulas as far as needed in this framework. For a broader introduction into the subject, we refer to monographs like Nelsen [92] and Joe [62].
The next two chapters outline the applicabilities and limitations of the most popular copula families. The elliptical one, i.e. the copulas of elliptical distributions, and the Archimedean one, i.e. copulas that can be described by a one dimensional function. Whereas Chapter 3 highlights the simplicity of multidimensional extension of elliptical copulas and show their nice analytical manner of presentation of the tail dependence coefficient. It also mentions the drawback of the class being restricted to rotationally symmetry $C(u,v) = C(1-u, 1-v)$ and hence to equal lower and upper tail dependence. The Archimedean copulas on the other side can provide very flexible models as described in Chapter 4, where we introduce a transformation rule extending this class. However, it is also remarked that there is no easy canonical extension to multidimensional models. In fact we conclude that this class is not appropriate for high dimensional problems.
Both, elliptical and Archimedean copulas, are symmetric in the sense $C(u,v) = C(v,u)$. In Chapter 5 we state for theoretical completion an asymmetric copula family. However, in our financial applications we have not found evidence for asymmetry in the former sense, but it may be of interest if coupling default credit data, e.g. for default probabilities.
The next Chapter 6 concentrates on multivariate testing procedures. We state some well known tests and give some numerical comments to them. The principal part of this chapter is the introduction of a copula based mapping to multivariate independence, that enables us to provide a simple method of high dimensional

χ^2 testing. Further we are able to present a new method to estimate multidimensional parametric models with an univariate maximum likelihood procedure.
Chapter 7 states the known estimation methods for the tail dependence coefficient (i.e. the probability of the joint occurrence of extremes) and provides some refinements to them. The methods are compared in an extensive simulation study.
The next four chapters build the empirical part of the thesis. First we fit different elliptical and Archimedean copulas to stock indices, foreign exchange, LME metal and oil related commodity baskets up to dimension four (Chapter 8). We are able to show evidence for the existence of different upper and lower tail dependence in financial data and therefore the need of non-elliptical models.
In Chapter 9 we apply the copula concept to measure the aggregate risk of portfolios in terms of Value at Risk and Expected Shortfall. To outline the procedure we use a two stock portfolio and exemplarily carry out all the steps from the rare data sets over the fitted joint (log)return distribution to the calculated risk measures.
With standard methods of univariate Extreme Value Theory and the use of the tail dependence concept we are in Chapter 10 able to certify real estate based firms (i.e. stock corporations that make their money by letting and reselling real estates) a higher crash resistance than common stocks. Especially possible diversification effects do not break down in crash situations as it can be observed for common stock portfolios.
In the last empirical survey (Chapter 11) we show the existence of nonlinear dependence structure in-between the USD yield curve. Therefore we utilize a two factor Vasicek model to achieve two abstract yield factors. These we couple with different copula models and check which of them reflect the pseudo observations best.
The last Chapter 12 introduces an extension of the class of elliptical copulas to a more generalized form. Precisely we allow the univariate random variable of the radial part to depend on the n-dimensional random variable including the directional information. Further a robust and always positive definite estimator for the VCV matrix (up to scale) is stated. This last chapter is an outlook to the forthcoming thesis of Gabriel Frahm, where further results will be given.

Chapter 1

A Probability Density Applying to Asymmetric Pareto Type Tails

Whenever we are talking about copulas of random vectors $(X_1, \ldots, X_n)$ we need the univariate marginal distribution functions $F_1, \ldots, F_n$ to achieve the random vector $U = (F_1(X_1), \ldots, F_n(X_n))$ with uniformly distributed margins. The copula C can be defined as the distribution function of U. Later on in Chapter 2 we see that C does not depend on the $F_1, \ldots, F_n$, but for empirics it is essential to guarantee for the estimated marginal models $\widehat{F}_i$ the convergence to the true marginals, i.e. $\widehat{F}_i \to F_i$ for all $i = 1, \ldots, n$. Otherwise we would not have $\widehat{F}_i(X_i) \sim U(0,1)$ what is the key assumption in copula theory. See Subsection 7.3.3 for pitfalls under misspecified margins. For model or estimation purposes a solution can be to use the empirical marginal distribution functions to achieve $\widehat{F}_i \to F_i$ but for simulation and stress test scenarios the use of parametric models has the favor of a (possibly) better description of extremal events. This was the motivation to start a thesis about multivariate distribution functions with a chapter on a univariate one. The content of this chapter is based on Junker and Szimayer [68].

Standard distributions—like normal or hyperbolic—are in some cases inadequate for describing financial data. Often heavy tailed distributions like the generalized Pareto distribution (GPD) turn out to be more appropriate. The GPD is designed for tail approximation and leads therefore to the problem of determining a threshold for both the left and the right tail. In this chapter we introduce an alternative distribution function that fits the entire data set and shows the behavior of a GPD in the tails. All parameters of the proposed distribution can be estimated via maximum likelihood and there are no thresholds to define. Therefore, the estimation algorithm yields a good computational tractability. To show the quality of the proposed distribution function we present a Value-at-Risk forecast and a backtesting utility.

1.1 Introduction

Many empirical studies like Danielsson and De Vries [26], Frey and McNeil [52] and Longin [79] have shown that empirical return distributions of market movements are heavy tailed. This is illustrated in Figure 1, where the QQ-plot of the empirical quantiles corresponding to 507 daily log-returns of IBM from Oct. 5, 1998 to Oct. 6, 2000 against standard normal quantiles is shown. Extreme losses

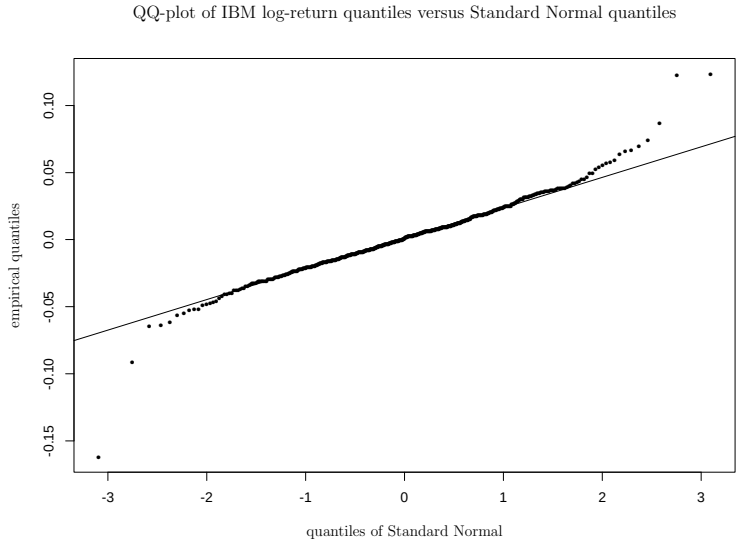

Figure 1.1: QQ-plot of IBM log-return quantiles versus Standard Normal quantiles.

and profits occur with higher probability than under the assumption of an underlying normal distribution. Common parametric distribution families like normal, hyperbolic or student's t have limited variability in describing the tail behavior of some random variable X. Complementary, extremal events are covered by Extreme Value Theory (EVT). As guidelines through EVT we refer to Resnick [100] and Embrechts et al. [38]. Nevertheless, EVT is primarily an asymptotic theory neglecting non-extremal events. Eberlein and Keller [36] propose the hyperbolic Lévy distribution. They exhibit a good statistical fit to log returns of DAX-stocks, but using the hyperbolic Lévy distribution they are limited to exponential tails. For calculating the Value-at-Risk (VaR) it is essential to model the tails. To accomplish this, the underlying stochastic model should exhibit a wide range of heavy tail behaviors. Frey and McNeil [52] suggest a method for estimating the VaR describing the tail of the conditional distribution of a heteroscedastic financial return series. Their approach combines pseudo-maximum-likelihood fitting of GARCH-models to estimate the current volatility, and EVT for estimating the

tail of the innovation distribution of the GARCH-model. For estimating the tails, they use the POT-method where they choose an arbitrary threshold.

We propose a parametric distribution function on $I\!R$ that incorporates EVT in the tails. Thus, we take a linear convex combination of suitable transformed generalized Pareto distributions. With such a construction, we get a flexible tail behavior—including light, exponential and heavy tailedness. We model the log-return series of financial data by applying the proposed distribution function where we additionally permit conditional heteroscedasticity. The estimates are used for Value-at-Risk forecasting. To examine the quality of the forecast, we apply some backtesting procedure.

In Section 1.2, we introduce a smooth distribution function with Pareto like tails and show some properties of this distribution function. In Section 1.3, we specify time series models allowing for conditional heteroscedasticity, and in Section 1.4 we apply the proposed distribution to financial datasets. The results of the estimation are used in Section 1.5 for Value-at-Risk forecast and backtesting.

1.2 A smooth distribution with Pareto tails

In this section we construct a parametric distribution function with a high degree of tail variability. To capture different types of tail behaviors, we use an appropriately transformed generalized Pareto distribution that is well known from EVT for describing extremal events. By this, we circumvent the drawback of most other parametric distribution functions that are a priori limited to a certain class of tail behavior. Tail variability plays a key role when measuring the risk of investments, since the risk is situated in the loss tail of the investment's distribution function.

For an overview about Extreme Value Theory we refer to Embrechts et al. [38]. Here we only summarize the definitions needed. Some of the following results are characterized by regularly varying functions with (tail)-index ζ. A measurable function $f : I\!R_+ \to I\!R_+$ is called *regularly varying* (at ∞) with (tail)-index $\zeta \in I\!R$ if for any $t > 0$

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^{-\zeta}, \tag{1.2.1}$$

and it is called *rapidly varying* with index $-\infty$ if

$$\lim_{x\to-\infty} \frac{f(tx)}{f(x)} = \begin{cases} 0 & if \quad t > 1 \\ \infty & if \quad 0 < t < 1 \end{cases} \tag{1.2.2}$$

Now we recall the definition of the generalized Pareto distribution. The generalized Pareto distribution with parameters $\xi \in I\!R$, $\beta > 0$ is defined by

$$G_{\xi,\beta}(x) \;=\; \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right), & \xi = 0, \end{cases} \tag{1.2.3}$$

where $x \geq 0$ for $\xi \geq 0$ and $0 \leq x \leq -\frac{\beta}{\xi}$ for $\xi < 0$.

The shape parameter ξ models the tail behavior and is therefore called the tail index. If $\xi > 0$ we say the df $G_{\xi,\beta}$ is heavy tailed, for $\xi = 0$ exponential tailed, and $\xi < 0$ light tailed. If the scaling parameter $\beta = 1$, we write G_ξ as an abbreviation for $G_{\xi,1}$.

The mean excess function is defined by

$$e(u) = E[X - u \mid X > u], \quad u \in I\!R,$$

where X is a random variable (rv). If the law of X is GPD, the mean excess function is linear.

The GPD is defined on the positive half axis. Often we need to shift the distribution to some new starting point u that is called threshold. In general, the GPD might only reflect the tail behavior of a given rv. In this case we can determine the threshold u by graphical data analysis. We choose u such that the empirical mean excess function $\widehat{e}(x)$ is approximately linear for $x \geq u$, see Embrechts et al. [39], pp. 352.

There are two main disadvantages following this approach.

- There is no estimator known for the threshold and therefore no sufficiently manageable algorithm for determining the threshold, even if we are only interested in the tails.

- Fitting a distribution on $I\!R$ with GPD in the tails, we face the problem of non differentiability of the density at the thresholds.

The GPD is designed for approximating a given distribution function in the tail, say in the right tail. In contrast to tail approximation where a threshold is chosen a priori or estimated, we extend the GPD support to $I\!R$ by an appropriate transformation h. Here, an appropriate transformation preserves the tail behavior on the right tail and creates a light or exponential left tail of $G_{\xi,\beta} \circ h$. The same argument can be applied for modelling the left tail. In the end, we build a convex linear combination of two $G_{\xi,\beta} \circ h$ terms. This results into a distribution function on $I\!R$, where each tail is determined by a function $G_{\xi,\beta} \circ h$.

A natural requirement for the transformation h is that $G_{\xi,\beta} \circ h$ is a distribution function on $I\!R$, i.e. h is an increasing function on $I\!R$ and maps on $I\!R^+$. For the

maximum likelihood estimation we want $G_{\xi,\beta} \circ h \in \mathrm{C}^3(I\!R, [0,1])$, so $G_{\xi,\beta} \circ h$ has a two times continuously differentiable density function.

To preserve the limit behavior of the generalized Pareto distribution for $G_{\xi,\beta} \circ h$ on the right tail we require

$$h \in \mathcal{R}_1^+ \,, \tag{1.2.4}$$

where $\mathcal{R}_1^+$ is the class of regularly varying functions of index 1 at $+\infty$.

To ensure that the right tail describing function $G_{\xi,\beta} \circ h$ exhibits a light (or exponential) left tail we demand the following.

$$h \in \mathcal{R}_{-\infty}^- \,, \tag{1.2.5}$$

where $\mathcal{R}_{-\infty}^-$ is the class of rapidly varying functions at $-\infty$.

Proposition 1.2.1. *Let $h \in \mathrm{C}(I\!R, I\!R^+)$ be an increasing function, $h \in \mathcal{R}_1^+$ and $h \in \mathcal{R}_{-\infty}^-$. Let X be a rv with $X \sim G_{\xi,\beta} \circ h$ for $\xi > 0$ and $\beta > 0$. Then X is heavy tailed on the right with tail index $\frac{1}{\xi}$, and light (or exponential) tailed on the left, i.e.*

$$\overline{G_{\xi,\beta} \circ h} \in \mathcal{R}_{-\frac{1}{\xi}}^+ \quad \textit{and} \quad G_{\xi,\beta} \circ h \in \mathcal{R}_{-\infty}^- \,. \tag{1.2.6}$$

Proof. With $h \in \mathcal{R}_1^+$ and $G_{\xi,\beta} \in \mathcal{R}_{-\frac{1}{\xi}}^+$, we observe for $t > 0$

$$\begin{aligned}
\lim_{x\to+\infty} \frac{\overline{G_{\xi,\beta} \circ h}(tx)}{\overline{G_{\xi,\beta} \circ h}(x)} &= \lim_{x\to+\infty} \frac{\left(1 + \frac{\xi\, h(tx)}{\beta}\right)^{-\frac{1}{\xi}}}{\left(1 + \frac{\xi\, h(x)}{\beta}\right)^{-\frac{1}{\xi}}} = \left(\lim_{x\to+\infty} \frac{1 + \frac{\xi\, h(tx)}{\beta}}{1 + \frac{\xi\, h(x)}{\beta}}\right)^{-\frac{1}{\xi}} \\
&= \left(\lim_{x\to+\infty} \frac{\frac{\xi\, h(tx)}{\beta}}{\frac{\xi\, h(x)}{\beta}}\right)^{-\frac{1}{\xi}} = \left(\lim_{x\to+\infty} \frac{h(tx)}{h(x)}\right)^{-\frac{1}{\xi}} \\
&= \left(t^1\right)^{-\frac{1}{\xi}} = t^{-\frac{1}{\xi}} \,,
\end{aligned}$$

since $h(x) \to \infty$ for $x \to \infty$. The latter follows from $h \in \mathcal{R}_1^+$ and Corollary A3.4 in Embrechts et al. [39], p. 566. With this we have shown $\overline{G_{\xi,\beta} \circ h} \in \mathcal{R}_{-\frac{1}{\xi}}^+$.

We proceed with the second statement $G_{\xi,\beta} \circ h \in \mathcal{R}_{-\infty}^-$. By elementary calculation we obtain $\lim_{y\searrow 0} \frac{G_{\xi,\beta}(y)}{y} = \frac{1}{\beta}$. Furthermore, $\lim_{x\to-\infty} h(x) = 0$ since $h \in \mathcal{R}_{-\infty}^-$ with Theorem A3.12 (cf. Embrechts et al. [39], p. 570).

$$\begin{aligned}
\lim_{x\to-\infty}\frac{G_{\xi,\beta}\circ h(tx)}{G_{\xi,\beta}\circ h(x)} &= \lim_{x\to-\infty}\left(\frac{G_{\xi,\beta}(h(tx))}{h(tx)}\,\frac{h(x)}{G_{\xi,\beta}(h(x))}\,\frac{h(tx)}{h(x)}\right)\\
&= \lim_{x\to-\infty}\frac{G_{\xi,\beta}(h(tx))}{h(tx)}\lim_{x\to-\infty}\frac{h(x)}{G_{\xi,\beta}(h(x))}\lim_{x\to-\infty}\frac{h(tx)}{h(x)}\\
&= \lim_{y\searrow 0}\frac{G_{\xi,\beta}(y)}{y}\lim_{y\searrow 0}\frac{y}{G_{\xi,\beta}(y)}\lim_{x\to-\infty}\frac{h(tx)}{h(x)}\\
&= \frac{1}{\beta}\,\beta\lim_{x\to-\infty}\frac{h(tx)}{h(x)}\\
&= \lim_{x\to-\infty}\frac{h(tx)}{h(x)}\,,
\end{aligned}$$

and consequently $G_{\xi,\beta}\circ h\in\mathcal{R}^-_{-\infty}$, because $h\in\mathcal{R}^-_{-\infty}$. □

Remark 1. *(i) We can deduce the same result in the case $\xi=0$. Here, the right tail behavior will also be preserved, since the transformation h is in $\mathcal{R}^+_1$. The df $G_{\xi,\beta}\circ h$ is then exponentially tailed with $\overline{G_{\xi,\beta}\circ h}\in\mathcal{R}^+_{-\infty}$.*

(ii) For a light tailed df with $\xi<0$, we face some technical problems because of the compact support of a GPD with negative tail index $[0,-\frac{\beta}{\xi}]$. The transformation h must not exceed the value $-\frac{\beta}{\xi}$, otherwise $G_{\xi,\beta}\circ h$ is no longer a df. Later on, we will show how to deal with that kind of problems.

Corollary 1.2.2. *In the situation of Proposition 1.2.1, we find for $X\sim G_{\xi,\beta}\circ h$*

$$\mathbf{E}\left[\left|X^+\right|^\kappa\right] < \infty, \quad \text{if } \kappa<\frac{1}{\xi} \tag{1.2.7}$$

$$\mathbf{E}\left[\left|X^+\right|^\kappa\right] = \infty, \quad \text{if } \kappa>\frac{1}{\xi}\,, \tag{1.2.8}$$

where $X^+=X\circ\mathbb{1}_{\{X\geq 0\}}$. For $X^-=-X\circ\mathbb{1}_{\{X\leq 0\}}$ all power moments are finite.

Proof. The results follow directly from the previous Proposition 1.2.1, applying Proposition A3.8 (d) on p. 568 for X^+ and Remark 7) on p. 570 for X^-, Embrechts et al. [39]. □

In the following, we propose a transformation h motivated by the considerations above, i.e. $h\in\mathcal{R}^+_1$ and $h\in\mathcal{R}^-_{-\infty}$. Let $h_{\delta,s,\beta}:\mathbb{R}\to\mathbb{R}^+$ be defined by

$$h_{\delta,s,\beta}(x)=\frac{\frac{x}{\beta}+\sqrt{\left(\frac{x}{\beta}\right)^2+\left(\frac{s}{\beta}\right)^2}}{2\left(1+\exp\left(-\delta\,\frac{x}{\beta}\right)\right)}, \tag{1.2.9}$$

where δ, s and β are positive real constants.

To fulfill the condition $h_{\delta,s,\beta} \in \mathcal{R}_1^+$, we make use of the fact that the following function is approximately linear for large x

$$x \mapsto \frac{\frac{x}{\beta} + \sqrt{\left(\frac{x}{\beta}\right)^2 + \left(\frac{s}{\beta}\right)^2}}{2} \approx \frac{x}{\beta}\,. \tag{1.2.10}$$

On the left tail, $h_{\delta,s,\beta} \in \mathcal{R}_{-\infty}^-$ is ensured by the function

$$x \mapsto \frac{1}{\left(1 + \exp\left(-\delta\,\frac{x}{\beta}\right)\right)} \approx \exp\left(\delta\,\frac{x}{\beta}\right)\,, \tag{1.2.11}$$

that exhibits approximately an exponential decay for large negative x.

So $h_{\delta,s,\beta}$ is a "minimal" transformation—minimal with respect to the number of parameters—that fulfills the requirements given in equations (1.2.4) and (1.2.5).

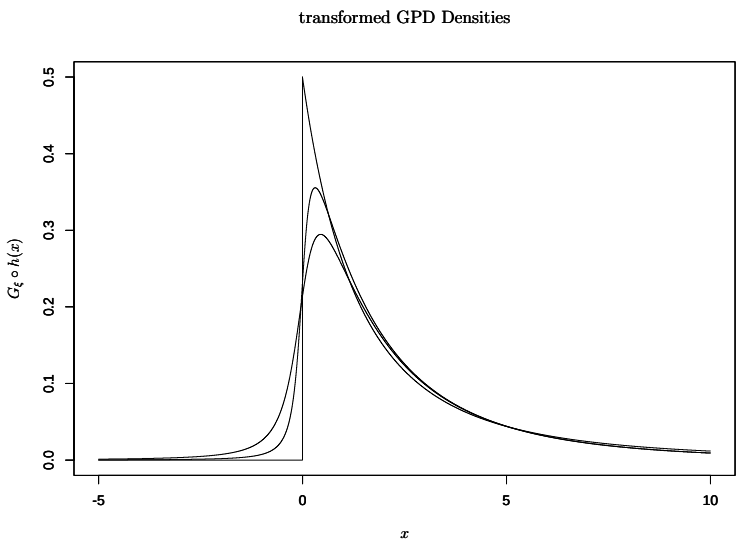

Figure 1.2: The transformation of a GPD density with $\xi = 0.5$ and $\beta = 1$, where the transformation $h_{\delta,s,\beta}$ is used with $\delta = 0.1$, $s = 0.25$ and $\delta = 0.1$, $s = 0.5$.

Figure 1.2 illustrates the effect of the transformation $h_{\delta,s,\beta}$ is illustrated. The density of $G_{\xi,\beta}$ is not differentiable in the threshold $t = 0$. Even for small s, e.g. $s = 0.25$, we have a C^∞-density of $G_\xi \circ h_{\delta,s,\beta}$, that is very close to the original density of $G_{\xi,\beta}$. With increasing s, e.g. $s = 0.5$, the density function $G_\xi \circ h_{\delta,s,\beta}$ becomes smoother.

Proposition 1.2.1 states that $G_{\xi,\beta}$ determines the right tail of $G_{\xi,\beta} \circ h$, where the tail index $\frac{1}{\xi}$ is preserved by the transformation h. By symmetry we can use

a second transformed GPD to describe the left tail. Gluing them together by a convex linear combination, we end up with a distribution function with tail variability in both tails.

Definition 1.2.3. *Let $F_\theta : I\!R \to I\!R$ be a function with $\theta = (\xi_l, \delta_l, s_l, \beta, \rho, \xi_r, \delta_r, s_r, \alpha)$ where $\xi_l, \xi_r, \rho \in I\!R$, $\beta, \delta_l, \delta_r, s_l, s_r > 0$, $0 \le \alpha \le 1$.*

For $\xi_l, \xi_r \ge 0$ and $x \in I\!R$ we define

$$F_\theta(x) := \alpha \cdot \left(1 - G_{\xi_l}\left(h_{\delta_l,s_l,\beta}\left(\rho - x\right)\right)\right) + (1-\alpha) \cdot G_{\xi_r}\left(h_{\delta_r,s_r,\beta}(x-\rho)\right), \quad (1.2.12)$$

where $h_{\delta,s,\beta}$ is given by Equation (1.2.9).

Remark 2. *(i) In the case $\xi < 0$ the domain of the GPD is bounded above by $-\frac{\beta}{\xi}$. For technical reasons, we have to modify the transformation h to $h^\star_{\delta,s,\beta,\xi} : I\!R \to I\!R^+$ given by*

$$h^\star_{\delta,s,\beta,\xi}(x) = \begin{cases} h_{\delta,s,\beta}(x), & x \le h^{-1}_{\delta,s,\beta}\left(-\frac{\beta}{\xi}\right), \\ -\frac{\beta}{\xi}, & x > h^{-1}_{\delta,s,\beta}\left(-\frac{\beta}{\xi}\right), \end{cases} \quad (1.2.13)$$

where $s \in \left(0, \left(\frac{4\beta}{\xi}\right)\right)$. The corresponding modification of F_θ is defined for $x \in I\!R$ by

$$F^\star_\theta(x) := \alpha \cdot \left(1 - G_{\xi_l}(h^\star_{\delta_l,s_l,\beta,\xi_l}(\rho - x))\right) + (1-\alpha) \cdot G_{\xi_r}(h^\star_{\delta_r,s_r,\beta,\xi_r}(x-\rho)). \quad (1.2.14)$$

(ii) Suppose, we a priori have only one *heavy tail, say the right tail. Then we are free to choose another mixture of distributions, e.g.*

$$F_\theta(x) := \alpha \cdot H(x) + (1-\alpha) \cdot G_\xi(h_{\delta,s,\beta}(x)),$$

where H is some arbitrary continuous distribution.

The following proposition summarizes some properties of F_θ. The proof is given in Appendix A.

Proposition 1.2.4. *Suppose F_θ is given as in Definition 1.2.3. Then the following properties hold.*

(a) The function F_θ is a continuous distribution function.

(b) For $\xi_l, \xi_r \ge 0$, we have $F_\theta \in \mathrm{C}^\infty(I\!R, [0, 1])$.
If we at least have one negative ξ, i.e. ξ_l or $\xi_r < 0$, then the preceding differentiability statement holds for $F^\star_\theta$ as well, except for $x_l = \rho - h^{-1}_{\delta_l,s_l,\beta}(-\frac{\beta}{\xi_l})$ and $x_r = \rho + h^{-1}_{\delta_r,s_r,\beta}(-\frac{\beta}{\xi_r})$, respectively.

Remark 3. *(i) For a more transparent interpretation of the smoothing parameter s, we introduce the following reparametrization. We define the parameters $c_l \in (0,\, \alpha)$ and $c_r \in (0,\, 1-\alpha)$ by*

$$c_l = \alpha G_{\xi_l}(h_{\delta_l,s_l,\beta}(0)) \tag{1.2.15}$$

$$c_r = (1-\alpha)\cdot G_{\xi_r}(h_{\delta_r,s_r,\beta}(0)). \tag{1.2.16}$$

Hence, c_r is the probability mass that is extracted by the transformation h from the original domain $[\rho,\, \infty)$ of $G_{\xi,\beta}$ to $(-\infty,\, \rho]$, where we assume without loss of generality $\xi \geq 0$. The same argumentation holds for c_l.

Solving equations (1.2.15) and (1.2.16) with respect to the smoothing parameters s_l, s_r, they can be replaced by

$$s_l = \frac{4\beta}{\xi_l}\left(\left(1-\frac{c_l}{\alpha}\right)^{-\xi_l} - 1\right) \tag{1.2.17}$$

$$s_r = \frac{4\beta}{\xi_r}\left(\left(1-\frac{c_r}{1-\alpha}\right)^{-\xi_r} - 1\right). \tag{1.2.18}$$

(ii) The parameter ρ describes the location of F_θ and does not influence the shape. Applying the reparametrization we get

$$F_\theta(\rho) = \begin{cases} \alpha - c_l + c_r\,, & \alpha \in (0,\, 1) \\ 1 - c_l\,, & \alpha = 1 \\ c_r\,, & \alpha = 0. \end{cases}$$

The location parameter ρ is the median iff $\alpha = \frac{1}{2}$ and $c_l = c_r$ or $\alpha \in \{0,1\}$ and $c_l = \frac{1}{2}$ and $c_r = \frac{1}{2}$, respectively.

The distribution function F_θ is designed for capturing the tail behavior of both tails. The tail indices ξ_l and ξ_r describe the left and the right tail separately. With this construction, we are no longer limited to some a priori given tail behavior. F_θ exhibits tail variability and also good numerical tractability, where the latter is meant in the sense of smoothness. For estimation purposes, it is often useful to have a density function which is at least two times differentiable. This allows for a maximum likelihood estimation procedure.

Theorem 1.2.5. *Let $h_l, h_r \in \mathrm{C}(I\!R, I\!R^+)$ be increasing functions, $h_l, h_r \in \mathcal{R}_1^+$, $h_l, h_r \in \mathcal{R}_{-\infty}^-$ and $G_{\xi_l,\beta_l}, G_{\xi_r,\beta_r}$ GPDs with $\xi_l, \beta_l, \xi_r, \beta_r > 0$. Suppose, we are given a rv X with $X \sim F$, where the df F is defined by*

$$F = \alpha\, \overline{G_{\xi_l,\beta_l} \circ h_l} + (1-\alpha) G_{\xi_r,\beta_r} \circ h_r\,,$$

for some $\alpha \in (0,1)$.

Then the tails are determined by the GPD's tail indices ξ_l, ξ_r, i.e.

$$F \in \mathcal{R}_{-\frac{1}{\xi_l}}^- \quad \text{and} \quad \overline{F} \in \mathcal{R}_{-\frac{1}{\xi_r}}^+\,. \tag{1.2.19}$$

Proof. For proving the theorem, we define $F_l = \overline{G_{\xi_l,\beta_l} \circ h_l}$ and $F_r = G_{\xi_r,\beta_r} \circ h_r$, hence F is the convex linear combination of F_l and F_r given by

$$F = \alpha\, F_l + (1-\alpha) F_r\,.$$

With Proposition 1.2.1 we conclude $\overline{F_l} \in \mathcal{R}^+_{-\infty}$, $F_l \in \mathcal{R}^-_{-\frac{1}{\xi_l}}$, $\overline{F_r} \in \mathcal{R}^+_{-\frac{1}{\xi_r}}$ and $F_r \in \mathcal{R}^-_{-\infty}$. First, we take a look at the right tail of F. The result for the left tail then follows by symmetry. For $t > 0$ we have

$$\begin{aligned}
\lim_{x\to+\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} &= \lim_{x\to+\infty} \frac{\alpha\,\overline{F_l}(tx) + (1-\alpha)\overline{F_r}(tx)}{\alpha\,\overline{F_l}(x) + (1-\alpha)\overline{F_r}(x)} \\
&= \lim_{x\to+\infty} \frac{\overline{F_l}(tx)}{\overline{F_l}(x) + \frac{1-\alpha}{\alpha}\overline{F_r}(x)} + \lim_{x\to+\infty} \frac{\overline{F_r}(tx)}{\overline{F_r}(x) + \frac{\alpha}{1-\alpha}\overline{F_l}(x)}\,.
\end{aligned}$$

The first expression is positive and bounded above by $\frac{\alpha}{1-\alpha}\frac{\overline{F_l}(tx)}{\overline{F_r}(x)}$. $\overline{F_l} \in \mathcal{R}^+_{-\infty}$ and $\overline{F_r} \in \mathcal{R}^+_{-\frac{1}{\xi_r}}$, hence $\frac{\overline{F_l(t\cdot)}}{\overline{F_r}} \in \mathcal{R}^+_{-\infty}$ and

$$0 \le \lim_{x\to+\infty} \frac{\overline{F_l}(tx)}{\overline{F_l}(x) + \frac{1-\alpha}{\alpha}\overline{F_r}(x)} \le \lim_{x\to+\infty} \frac{\alpha}{1-\alpha}\frac{\overline{F_l}(tx)}{\overline{F_r}(x)} = 0\,,$$

by Lemma A.2.1 from the Appendix. Applying the same arguments to the second expression, we find

$$\begin{aligned}
\lim_{x\to+\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} &= \lim_{x\to+\infty} \frac{\overline{F_r}(tx)}{\overline{F_r}(x) + \frac{\alpha}{1-\alpha}\overline{F_l}(x)} \\
&= \lim_{x\to+\infty} \left\{ \frac{\overline{F_r}(tx)}{\overline{F_r}(x)} \left(1 + \frac{\alpha}{1-\alpha}\frac{\overline{F_l}(x)}{\overline{F_r}(x)}\right)^{-1} \right\} \\
&= \lim_{x\to+\infty} \frac{\overline{F_r}(tx)}{\overline{F_r}(x)} \left(1 + \frac{\alpha}{1-\alpha} \lim_{x\to+\infty} \frac{\overline{F_l}(x)}{\overline{F_r}(x)}\right)^{-1} \\
&= t^{-\frac{1}{\xi_r}}\,,
\end{aligned}$$

thus $\overline{F} \in \mathcal{R}^+_{-\frac{1}{\xi_r}}$ and the proof is finished. □

The previous theorem provides a very general statement. For the proposed transformation $h_{\delta,s,\beta}$ and the df F_θ, we deduce the following corollary.

Corollary 1.2.6. *Let X be a rv with df F_θ. If $\xi_l > 0$ then $F_\theta \in \mathcal{R}^-_{-\frac{1}{\xi_l}}$ and*

$$E\left[\left|X^-\right|^\kappa\right] < \infty \quad \textit{for } \kappa < \frac{1}{\xi_l} \quad \textit{and} \quad E\left[\left|X^-\right|^\kappa\right] = \infty \quad \textit{for } \kappa > \frac{1}{\xi_l},$$

and for $\xi_l \leq 0$ all power moments of X^- exist and $F_\theta \in \mathcal{R}^-_{-\infty}$.

If $\xi_r > 0$ then $\overline{F_\theta} \in \mathcal{R}^+_{-\frac{1}{\xi_r}}$ and

$$\mathbf{E}\left[\left|X^+\right|^\kappa\right] < \infty \quad \textit{for } \kappa < \frac{1}{\xi_r} \quad \textit{and} \quad \mathbf{E}\left[\left|X^+\right|^\kappa\right] = \infty \quad \textit{for } \kappa > \frac{1}{\xi_r},$$

and for $\xi_r \leq 0$ again all power moments of X^+ exist and $\overline{F_\theta} \in \mathcal{R}^+_{-\infty}$.

Proof. The corollary is a direct consequence of Theorem 1.2.5 and Corollary 1.2.2. □

Theorem 1.2.5 and Corollary 1.2.6 state that F_θ is well defined, i.e. each transformed GPD $G_{\xi,\beta} \circ h$ describes its own tail and does not interfere with its counterpart. With this, the desired tail variability of the parametric df F_θ is established. In the following, we apply F_θ to financial data and measure the risk of an investment. Hence the shape of the loss tail plays a crucial role.

1.3 Model Specification

In Section 1.2, we have constructed the parametric distribution function F_θ. This distribution function exhibits tail variability and, moreover, allows for different tail behavior in the left and the right tail. It is well known that empirical distributions of log-returns of financial time series are asymmetric, especially in the tails. Thus, F_θ is apparently an appropriate choice for fitting financial data. We model the log-return series of financial data by applying the proposed df F_θ, where we additionally permit conditional heteroscedasticity.

The time-series $(X_t)_{t\geq 1}$ that is modelling the log-returns, is an autoregressive process of order one with independent innovations that are described by a GARCH(1,1) variance structure, see Bollerslev [13].

$$X_t = \mu + \alpha\, X_{t-1} + \sigma_t\, \varepsilon_t \quad \text{for } t \geq 2\,, \tag{1.3.1}$$

where $\mu \in I\!R$ and $|\alpha| < 1$, and the innovations $(\varepsilon_t)_{t\geq 2}$ are iid with expectation zero and unit variance. The conditional heteroscedasticity is specified by

$$\sigma_t^2 = \omega_0 + \omega_1\,(\sigma_{t-1}\varepsilon_{t-1})^2 + \omega_2\,\sigma_{t-1}^2 \quad \text{for } t \geq 2\,, \tag{1.3.2}$$

where $\omega_0 > 0$, $\omega_1, \omega_2 \geq 0$.

The model described in Equation (1.3.1) and (1.3.2) is a common set-up for financial time-series and especially for Value-at-Risk models. For example, Frey and McNeil [52] utilize this framework where they allow for heavy tailed innovations (ε_t). They estimate the tails of the innovations by fitting a generalized Pareto distribution to tails of the empirical innovations, i.e. the AR(1)-GARCH(1,1) residuals. On this basis they perform VaR calculations.

Subsequently, we use a similar approach as Frey and McNeil [52]. However, we do not reduce the examination to the lower and upper tail behavior of the innovations, but estimate the entire distribution of the innovations by fitting the proposed df F_θ to the empirical innovations. Additionally, we investigate the problem that GARCH-type models are overspecified when the data exhibits no conditional homoscedasticity. For such data, exponential Lévy models are applicable, see Eberlein and Keller [36].

May and Szimayer [87] furthermore show that GARCH models break down when they are applied to conditional homoscedastic data. They suggest to test the data for reduction to the white noise model, that is given by the equation

$$X_t = \mu + \sigma\,\varepsilon_t \quad \text{for } t \geq 1\,, \tag{1.3.3}$$

where $\mu \in I\!R$, $\sigma > 0$, and the innovations $(\varepsilon_t)_{t\geq 1}$ are iid with expectation zero and unit variance. They consider pseudo-likelihood ratio test and provide the critical values for this specific test, i.e. 3.80, 5.13, and 8.28 for significance levels of 90%, 95%, and 99%.

1.4 Estimation

In the previous section, we have specified a time-series setup to model the log-returns of asset prices. We have included the effect of conditional heteroscedasticity and we also have allowed for heavy tailed and possibly asymmetric innovations by applying the df F_θ proposed in Section 1.2. In the following, we carry out a statistical data analysis for the daily log-returns of Allianz, BASF, VW, Apple, IBM and Microsoft. The observed time period ranges from Oct. 5, 1998 to Oct. 6, 2000, including 510 data points for the German stocks and 507 for the US stocks, respectively.

data	Allianz	BASF	VW	Apple	IBM	Microsoft
d_T	$37.70^{\star\star}$	2.73	$41.62^{\star\star}$	$148.32^{\star\star}$	3.48	$32.92^{\star\star}$

Table 1.1: The deviance for the analysed data. $^\star/^{\star\star}$ denotes the significance of the GARCH effects for the level of 90%/99%.

Along the lines of May and Szimayer [87] we test the considered data for conditional heteroscedasticity, see Section 1.3 for details. Table 1.1 displays the test results. The hypothesis of white noise is rejected for Allianz, VW, Apple and Microsoft at the 99% significance level. For these log-return series GARCH effects are of strong statistical evidence, and hence we have to apply the set-up given by Equation (1.3.1) and (1.3.2). The BASF and IBM sample show hardly any heteroscedasticity, and we can not reject the null hypothesis of white noise even at the 90% level. For BASF and IBM data we use the white noise model in Equation (1.3.3) and estimate F_θ directly from the log-return series.

For both model specifications, the parameters are estimated where we apply maximum likelihood. The white noise model/homoscedastic model is associated with the df F_θ, where we first simplify the parametrization. Via the approach presented in Section 1.2, we are enabled to model a different behavior of the left and the right tail. We accomplish this by tail indices ξ_l and ξ_r. The asymmetry in the tail behavior is not affected by the constraint of a symmetric transformation of the GPD's, i.e. $\delta = \delta_l = \delta_r$ and $s = s_l = s_r$. Furthermore, we assume $\delta = 1$ what implies a uniform scaling in the transformation, see Equations (1.2.10) and (1.2.11). With this specification for F_θ, we have to estimate the parameter vector $\theta_6 = (\xi_l, \beta, s, \rho, \xi_r, \alpha)$ of dimension 6. For the conditional heteroscedastic model/AR(1)-GARCH(1,1) model the reduced parametrization is applied to the empirical innovations, i.e. AR(1)-GARCH(1,1) residuals.

The quality of the estimation is investigated, especially the fit of the df F_θ to data. We perform a χ^2-goodness-of-fit test where we follow Moore [88]. In addition, we apply the integrated Anderson Darling test. The statistics is defined by

$$AD = N \int_{-\infty}^{\infty} \frac{(F_N(x) - F_\theta(x))^2}{F_\theta(x)\ (1 - F_\theta(x))}\, dF_\theta(x),$$

where N is the number of observations and F_N denotes the empirical distribution function (see also Chapter 6). The critical value for the significance level of 95% is given by 2.49, see Giles [54].

1.4.1 The Homoscedasticity Model

The data showing no significant conditional heteroscedasticity are the BASF and IBM log-return series. A maximum likelihood estimation is carried out. And beyond statistical significance, we illustrate the results by means of graphical tools. We provide a QQ-plot of the estimated distribution and compare the empirical and estimated density.

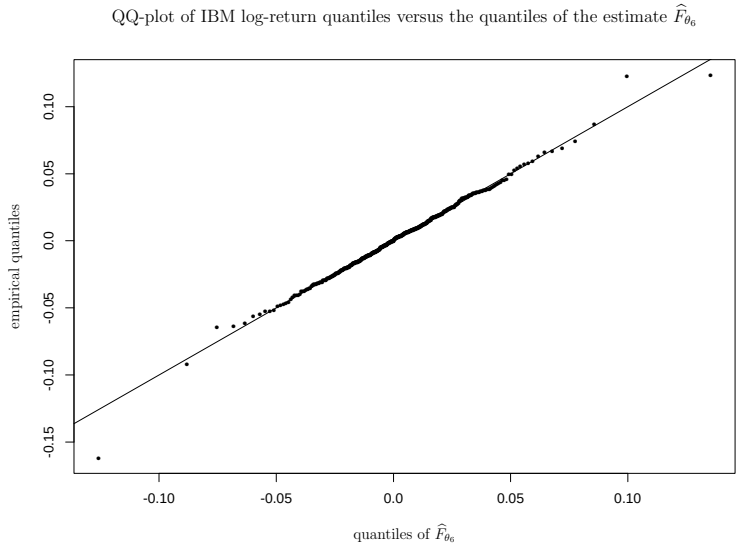

Figure 1.3: QQ-plot of IBM log-return quantiles versus the quantiles of $\widehat{F}_{\theta_6}$.

The QQ-plot of the IBM log-returns against an estimated F_{θ_6}-model is displayed in Figure 1.3, and demonstrates graphically the improvement of the model fit compared to the normal distributed setup, see Figure 1.1. The impression given by the QQ-plot is confirmed by the empirical density and the log-density of the IBM data that are compared with the density of the estimated distribution function $\widehat{F}_{\theta_6}$, see Figure 1.4. The estimated density clearly reflects the behavior of the data in the tails. This can be inferred by taking a look at the log-densities on the right in Figure 1.4. Most other parametric distribution functions are limited to an a priori given tail behavior. For a normal distribution, the log-density is a parabola. In the case of the hyperbolic Lévy distribution, the log density is approximately linear in the tails, since it is asymptotically exponential. Figure 1.4 clearly reflects that a parametrization with either normal or hyperbolic Lévy can not capture the characteristics of the given data.

The estimation results are listed in Table 1.2. The location and the scaling parameter of the distribution F_{θ_6} are given by ρ and β. Since we analyze daily log-returns of stock prices, a reasonable location is in the region around zero. In our sample, $\hat{\rho}$ takes values of 0.0073 and 0.0042 for BASF and IBM, respectively. The scaling parameter β corresponds to the standard deviation σ for normal distributions, and may be interpreted as "daily volatility". For BASF and IBM we find 0.0113 and 0.0126, both quite accurate estimated what is indicated by t-values of 12.71 and 10.80. The transformation $h_{\delta,s,\beta}$ is characterized by s, since we choose a natural scaling with $\delta = 1$. The parameter s is estimated with 0.0746 and 0.0717 for BASF and IBM. The estimates are both in the same range and

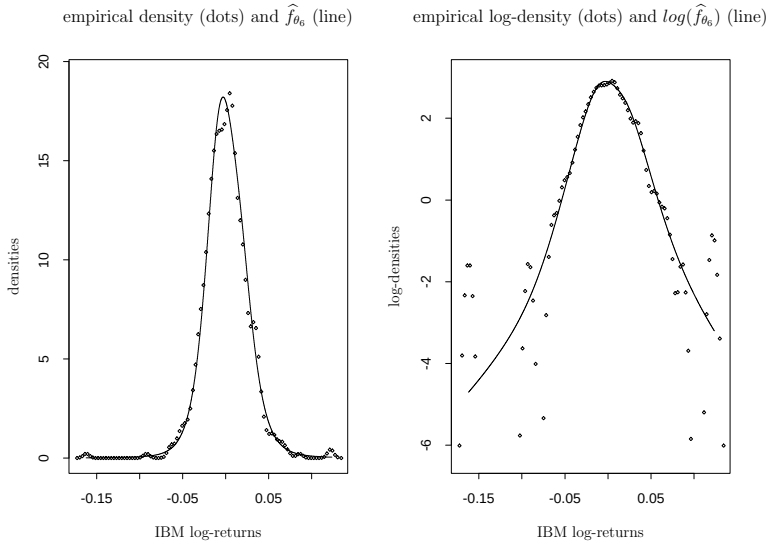

Figure 1.4: Empirical density and log-density of the IBM log-returns, compared with the estimated $\widehat{F}_{\theta_6}$ density and log-density, respectively.

reveal statistical significance with t-values of 5.63 and 7.10. The weights of the tails are described by α, where the parameter $\alpha \in [0, 1]$ is the weight of the left transformed GPD-tail and $1 - \alpha$ the weight of the right tail. For both stocks, α differs significantly from 0 and 1, i.e. we can not model the df with only one single transformed GPD.

data	$\widehat{\xi_l}$	$\widehat{\beta}$	$\widehat{s}$	$\widehat{\rho}$	$\widehat{\xi_r}$	$\widehat{\alpha}$	$\widehat{\xi_l} - \widehat{\xi_r}$	χ^2_{18}	AD
BASF	-0.0468	0.0113	0.0746	0.0073	-0.0483	0.2541	0.0015	19.22	0.17
s.e.	(0.6096)	(0.0009)	(0.0133)	(0.0038)	(0.2161)	(0.1686)	(0.8257)		
t-value	(-0.08)	(12.71)	(5.63)	(1.93)	(-0.22)	(1.51)	(0.00)		
IBM	0.1942	0.0126	0.0717	0.0042	0.1529	0.3539	0.0413	15.83	0.13
s.e.	(0.1011)	(0.0012)	(0.0101)	(0.0040)	(0.0935)	(0.1464)	(0.1946)		
t-value	(1.92)	(10.80)	(7.10)	(1.06)	(1.64)	(2.42)	(0.21)		

Table 1.2: Parameter estimates with standard errors (s.e.) and t-values in parentheses, and the estimated χ^2_{18}-statistics and Anderson Darling statistics for $\widehat{F}_{\theta_6}$ and the log-returns of iid samples.

The tails of the distribution function F_θ are determined by ξ_l and ξ_r. The estimates $\hat{\xi}_l$ and $\hat{\xi}_r$ reflect the tail behavior what can be inferred from the empirical log-density in Figure 1.4. For IBM we observe heavy tailed losses, whereas for BASF the hypothesis of exponential tailed losses can not be rejected, since we have a t-value of -0.08. On the right tail where the profits are situated, we find similar results. Signifanctly heavy tailed gains of IBM with a t-value of 1.64 can be observed, and exponential tailed gains for BASF. In terms of $\xi_l - \xi_r$ we can not reject the hypothesis of $\xi_l = \xi_r$, since the absolute t-values of $\xi_l - \xi_r$ are too

small for statistical significance.

To evaluate the quality of the estimation, we perform a χ^2-goodness-of-fit test on the estimated distribution functions and additionally we apply the Anderson Darling test. For the χ^2-goodness-of-fit test the number of cells is set to $k = 25$. The 95%-quantile of the χ^2-distribution with $k - 7$ degrees of freedom is $\chi^2_{18;\,0.95} = 28.87$. Table 2 represents the results for the analyzed time series. The test statistics ranging from 15.83 to 19.22 are around the expected value of 18. The χ^2-test indicates a good fit of the estimated distribution $\widehat{F}_\theta$ to the given empirical data. These findings are strengthened by the Anderson Darling test results. With values of 0.13 and 0.17 the statistics are clearly below 2.49 representing the critical value for a significance level of 95%. However, here we have to be careful, since we had run the tests in sample. Due to a possible change of the distribution of the statistics by in sample testing, test results should be taken with caution. Especially the Anderson Darling test does not penalize the number of used parameters.

1.4.2 The Conditional Heteroscedasticity Model

In this subsection we analyze the data showing conditional heteroscedasticity, i.e. Allianz, VW, Apple and Microsoft. We use an AR(1)-GARCH(1,1) setup to model the data, where the innovation process is described by the df F_θ. This is a similar approach as used by Frey and McNeil [52], but we model and estimate the entire distribution of the innovations, whereas they apply the so-called peaks-over-threshold (POT) method. The parameters of the AR(1)-GARCH model are estimated by pseudo-maximum likelihood. We omit these results, since they are quite standard and we want to emphasize on the innovations and their tail behavior.

The estimation results are summarized in Table 1.3. The particular estimates provide information about the characteristics of the innovations of the AR(1)-GARCH(1,1) model. Naturally, the innovations are expected to have zero mean and unit variance. Asymmetry in the center of the distribution is primarily influenced by the parameters ρ and α that are describing the location and the weights of the tails. For Allianz and Microsoft ρ is estimated significantly larger than zero indicating skewed innovations. The scaling parameter β takes values around 0.5, where the estimation in all cases is very accurate (all t-values larger than 10). We remind that β scales the distribution and usually does not equal the standard deviation. The parameter s is smoothing the generalized Pareto distribution. For the sample of this subsection we find estimates ranging from 2.48 to 3.21. These values are quite significant what can be inferred from the t-values, but are of different order compared to the results in the homoscedastic case. The reason is

the parametrization of the transformation h given in Equation (1.2.9). For related transformations/distributions we expect the ratio s/β to be roughly in the same range. All over the entire sample (including the homoscedastic model for BASF and IBM), the ratio s/β is relatively stable with values in-between 5 and 7.

data	$\widehat{\xi_l}$	$\widehat{\beta}$	$\widehat{s}$	$\widehat{\rho}$	$\widehat{\xi_r}$	$\widehat{\alpha}$	$\widehat{\xi_l}-\widehat{\xi_r}$	χ^2_{18}	AD
Allianz	0.1505	0.4737	2.4759	0.0266	0.2315	0.3998	-0.0810	17.35	0.18
s.e.	(0.0969)	(0.0346)	(0.3693)	(0.0125)	(0.0704)	(0.0440)	(0.1673)		
t-value	(1.55)	(13.71)	(6.70)	(2.12)	(3.29)	(9.09)	(-0.48)		
VW	0.0801	0.5751	3.1623	0.2721	0.0217	0.2389	-0.1588	26.27	0.33
s.e.	(0.1888)	(0.0550)	(1.0590)	(0.2518)	(0.0894)	(0.1033)	(0.2921)		
t-value	(0.42)	(10.45)	(2.99)	(1.08)	(0.24)	(2.31)	(-0.54)		
Apple	0.2249	0.4417	3.2135	-0.0343	0.1500	0.6041	0.0749	22.63	0.20
s.e.	(0.0752)	(0.0338)	(0.2994)	(0.0306)	(0.0982)	(0.0450)	(0.1734)		
t-value	(2.99)	(13.05)	(10.73)	(-1.12)	(1.53)	(13.41)	(0.43)		
Microsoft	0.2013	0.5058	3.1738	0.1604	0.0743	0.3744	0.1270	24.51	0.16
s.e.	(0.1034)	(0.0449)	(0.3692)	(0.1000)	(0.0945)	(0.0812)	(0.1979)		
t-value	(1.95)	(11.26)	(8.60)	(1.60)	(0.79)	(4.61)	(0.64)		

Table 1.3: Parameter estimates with standard errors (s.e.) and t-values in parentheses, and the estimated χ^2_{18}-statistics and Anderson Darling statistics for $\widehat{F}_{\theta_6}$ and the GARCH residuals for the heteroscedastic samples.

The tail indices ξ_l and ξ_r model the tail behavior and indicate possible asymmetry in the tails. For Allianz and Apple we observe significantly heavy tailed innovations, and there is also evidence for asymmetry in both cases. Nevertheless, for Allianz and Apple it is possible to suggest Student's t-distribution, here with 5 or 6 degrees of freedom, see Frey and McNeil [52] for discussion of this specification. In contrast to the latter findings, the time series of VW exhibits an exponential tail behavior. For the lower and upper tail the estimates of the tail indices are not bounded away significantly from zero, and hence the null hypothesis of exponential tails can not be rejected. The Microsoft data show an ambiguous behavior in the tails. The lower tail is significant heavy tailed with a t-value of 1.95, but the upper tail appears to be exponential (t-value of 0.79), and the null hypothesis $\xi_r = 0$ can not be rejected at any usual significance level. However, again we are not able to reject $\xi_l = \xi_r$, because of the low absolute t-values.

The statistical quality of the estimation is examined by performing a χ^2-goodness-of-fit test for the entire distribution, and the Anderson Darling test, see the homoscedastic model for details. The χ^2-test statistics are ranging from 17.35 to 26.27 and are below 28.87, representing the critical value for 95% significance level. With a maximal value of 0.33 the Anderson Darling test is below the critical value of 2.49, for 95% significance level (again this should be taken with caution, since we had tested in sample). Both tests demonstrate that the estimated distribution $\widehat{F}_\theta$ fits very well to the nature of the innovations, and this is once more strengthened by graphical evidence, see Figure 1.5.

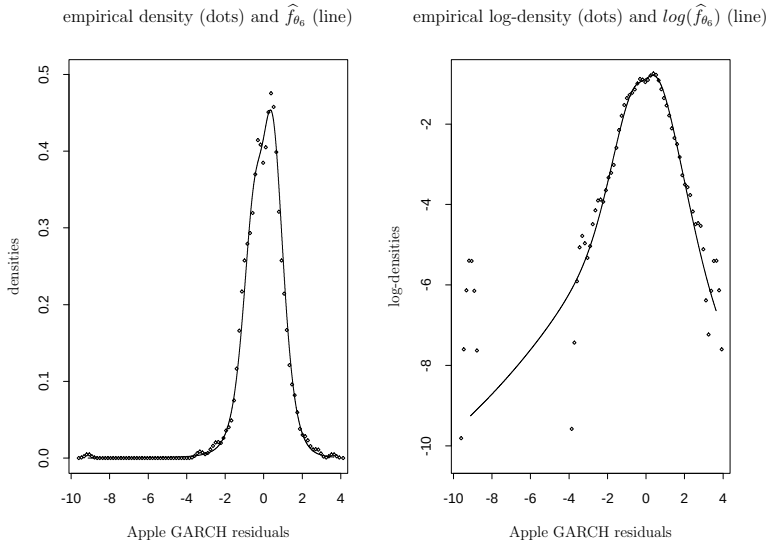

Figure 1.5: Empirical density and log-density of the Apple GARCH residuals, compared with the estimated $\widehat{F}_{\theta_6}$ density and log-density, respectively.

In this section, we showed that the proposed df F_θ fits very well to financial data. Especially, F_θ captured possible asymmetry in the tail behavior of the investigated time series, ranging from exponential to heavy tailedness for the loss tail and the profit tail. With a reduced set of 6 parameters, the estimation results revealed statistical significance. On the other hand it was not possible to reject $\xi_l = \xi_r$. Even we found evidence for $\xi_l \neq \xi_r$, we were not able to attain statistical significance in that point. The χ^2-goodness-of-fit test and the Anderson Darling test were clearly accepted for all financial time series. In the next section, the estimation results are used for Value-at-Risk forecasts and backtesting.

1.5 VaR and backtesting

For risk management purposes Value-at-Risk calculation and backtesting are essential components. From the perspective of statistics, financial risk is situated in the loss tail of a return distribution, i.e. in the left tail. We apply the df F_θ constructed in this chapter to the issues posed by risk management. Since we are concerned with losses, the tail variability plays an important role in quantifying and analyzing financial risk.

In the following, we compute VaR's (estimated negative quantiles of the log-returns time series) for the different stocks given in Section 1.4. For an overview of Value-at-Risk, we refer to Duffie and Pan [34] and Jorion [63]. In addition, we perform a standard backtest on the VaR calculated for IBM. The standard

backtesting procedure is a criterion for measuring the quality of VaR predictions imposed by the regulators. As guideline we refer to Cassidy and Gizycki [19].

The estimation procedure is based on the entire data of Allianz, BASF, VW, Apple, IBM and Microsoft from Oct. 5, 1998 to Oct. 6, 2000. For the homoscedastic time series of BASF and IBM, the calculation of the one-day γ-VaR requires the $(1-\gamma)$-quantile of F_θ. This is accomplished by numerical inversion of the estimated $\widehat{F_\theta}$. Here, we use the estimated parameters of F_θ given in Section 4 applied to the losses, i.e. the negative log-returns. The conditional heteroscedastic model is applied to Allianz, VW, Apple and Microsoft. For this data, we follow Jorion [63] and compute the so-called conditional Value-at-Risk. The conditional VaR is the specific quantile of the one-day prediction conditioned on the entire information. We use the variance forecast implied by the GARCH structure and as well the statistical analysis of the residuals for the estimation of the distribution of the innovations, see e.g. Frey and McNeil [52] for details. The results for various levels γ are listed in Table 1.4.

data	95% VaR	99% VaR	99.5% VaR	99.9% VaR
BASF	0.0306	0.0472	0.0542	0.0706
IBM	0.0370	0.0612	0.0750	0.1206
Allianz	0.0263	0.0420	0.0503	0.0772
VW	0.0369	0.0549	0.0630	0.0834
Apple	0.0478	0.0853	0.1114	0.2043
Microsoft	0.0357	0.0590	0.0724	0.1232

Table 1.4: The (conditional) VaR's for different levels computed by applying $\widehat{F}_{\theta_6}$ and the AR(1)-GARCH(1,1) parameters.

The calculated VaR's in Table 1.4 are closely connected to the estimation results of Section 1.4, see Table 1.2 and Table 1.3. For the homoscedastic time series the VaR predictions are primarily determined by β and ξ_l. The scaling parameter β has a linear influence on the VaR calculation. For IBM, β is estimated larger than for BASF. Accordingly, the VaR's of IBM are higher than the VaR's of BASF, but the difference increases dramatically, when we look at more extreme VaR's like the 99.9% VaR. This rise results from the tail parameter ξ_l that is estimated 0.1942 for IBM and -0.0468 for BASF. VaR calculation is more complex for data that incorporates conditional heteroscedasticity. The parameters of the AR(1)-GARCH(1,1) model have an effect on the VaR, the autoregressive part influences the mean and the stochastic variance caused by the GARCH equation determines the scale. The nature of the innovations becomes apparent when we study the 99.5% and the 99.9% VaR. Here, the two time series with most heavy tailed innovations (Apple and Microsoft) with $\widehat{\xi_l} = 0.2249$ and $\widehat{\xi_l} = 0.2013$ have

clearly the highest VaR's. Additionally, we observe that the 95% VaR of Microsoft is below the one of VW, but the relation is turned around for the 99% VaR, what can be explained by the heavy tailedness of the innovations of Microsoft compared to the exponential tailed innovations of VW.

We perform a standard backtest for the IBM stock, where we use a 250-day history for estimating the VaR. This is carried out for the last 250 days in our sample. In Figure 1.6, the 95% and the 99% VaR of IBM are plotted against time and furthermore, the negative log-returns are also added in this picture. An exceedance occurs whenever a loss is larger than the predicted VaR. For the 99% VaR, we expect 2.5 exceedances, since we have a sample size of 250. Here, we observe 1 exceedances, and in the case of the 95% VaR we have 16 exceedances, where the expected value is 12.5. To study the quality of the IBM backtest results, we conduct Kupiec's Proportion of Failures (POF) Test. The critical value of LR_{POF} for the 95% test level is given by 3.84, for further details see Kupiec (1995). We have observed 1 exceedance for the 99% VaR implying $LR_{POF} = 1.18$, and 16 exceedances for the 95% VaR, what results in $LR_{POF} = 0.95$. So for both VaR levels the statistics are well below the critical value and indicate an accurate VaR estimation.

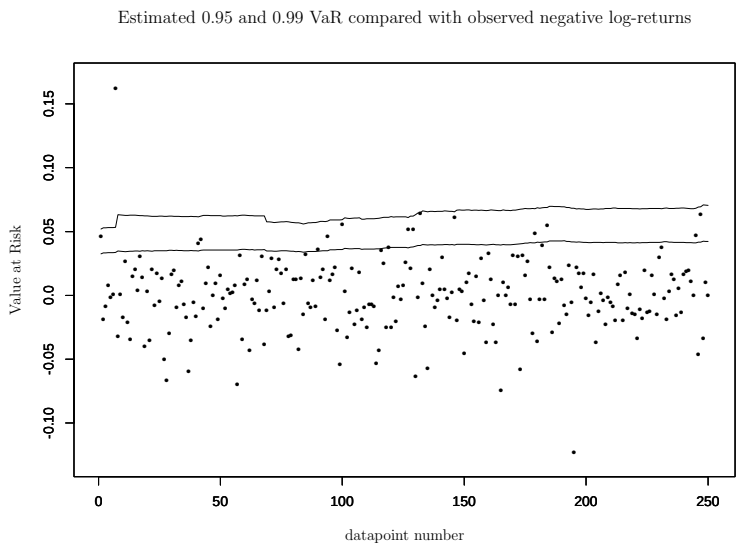

Figure 1.6: IBM backtesting result, where the upper line represents the estimated 99% VaR and the lower the 95% VaR respectively. The dots are the observed negative log-returns.

The quality of VaR estimation depends highly on good predictions of extremal events, since the VaR is calculated from the lowest returns. Naturally, the shape of the loss-tail plays an important role for VaR estimates, especially for "extreme"

VaR's like the 99.5% or the 99.9%-VaR. As discussed in Section 1.4, the log-density estimates given in Figure 1.4 and Figure 1.5 exhibit a good fit to the observed data, particularly in the tails. This indicates a good quality of the VaR prediction, what is validated by the backtesting procedure carried out for the IBM stock. Furthermore, with an estimated one-day log-return distribution we are enabled to calculate VaR's for different time horizons like the 10-day VaR's using a Monte-Carlo-Simulation. This is an advantage compared to VaR calculations with EVT (like the usual POT-method) where only the tail and scaling parameters over some threshold are estimated and not the entire distribution.

Chapter 2

Copula Theory

In this chapter, partially based on Junker and May [67], we give a very brief introduction in the copula theory. We state the basic definitions and well known results we will need in the further chapters. Additionally we derive some smaller calculation tools for survival copulas.
For a detailed course about copulas and multivariate distribution functions we highly recommend one of the recent monographs, e.g. Joe [62] and Nelsen [92]

2.1 The Copula Concept

The copula concept is based on a separate statistical treatment of dependence and marginal behavior. The mathematical idea goes back to Sklar (1955) and Hoeffding (1940).
Let $X = (X_1, \ldots, X_n)$ be an n-dimensional random vector with joint distribution function F and continuous marginal distribution functions $F_{X_1}, \ldots, F_{X_n}$. A copula is a multivariate distribution function defined on the unit cube $[0,1]^n$, with uniformly distributed marginals. For shorter writing we will sometimes use the notation $F_{X_j}(X_j) = U_j$ and $F_{X_j}(x_j) = u_j$, analogously.

Definition 2.1.1. *Let $X = (X_1, \ldots, X_n)$ be a random vector with multivariate distribution F and continuous marginals $F_1, ..., F_n$. The copula of X (of the distribution F, respectively) is the multivariate distribution C of the random vector*

$$U = (F_1(X_1), \ldots, F_n(X_n)).$$

Now the n-dimensional joint distribution function F for $(X_1, \ldots, X_n)$ can be written as follows

$$F(x_1, ..., x_n) = C(F_{X_1}(x_1), ..., F_{X_n}(x_n)), \tag{2.1.1}$$

and hence the copula C describes the dependence between the univariate random variables $X_1, ..., X_n$. Vice versa the copula of an n-dimensional distribution

function F with continuous marginals can be expressed in terms of

$$C\left(u_1,\ldots,u_n\right)=F\left(F_1^{\leftarrow}\left(u_1\right),\ldots,F_n^{\leftarrow}\left(u_n\right)\right), \tag{2.1.2}$$

where

$$F_i^{\leftarrow}\left(u_i\right):=\inf\left\{x:F_i\left(x\right)>u_i\right\}, \qquad i=1,...,n$$

are the marginal quantile functions. Equations (2.1.1), (2.1.2) and Definition 2.1.1, respectively, is known in literature as Sklar's Theorem and implies that for continuous multivariate distribution functions the univariate marginals and the dependence structure (encoded into the copula) can be separated in a unique way. The definition of a copula can be extended to the case of non-continuous marginals $F_1,...,F_n$. Equations (2.1.1) and (2.1.2) remain true but we loose the uniqueness of the copula and hence we are not able to define a copula like in Definition 2.1.1. However, the more general definition needs the introduction of further not needed concepts, and for most financial applications it is justified to assume continuous marginals. So for the sake of simplicity we skip this point here and refer the reader to the above stated monographs.
Note that for any n-dimensional copula each k-dimensional margin of the copula is a k-dimensional copula itself ($1\leq k\leq n$).

The set of copulas is convex in the sense that every convex linear combination of copulas is a copula itself (Nelsen [92], Ex. 2.3, p. 12]).

Set $\Pi(u)=u_1u_2\cdots u_n$, then Π is a copula for all $n\geq 2$, the so called independence copula. The function $M(u)=\min\{u_1,\ldots,u_n\}$ is a copula for all $n\geq 2$, whereas $W(u)=\max\{u_1+u_2+\ldots u_n-n+1,0\}$ is a copula only for $n=2$, but not for $n>2$. The functions W and M are known as Fréchet-Hoeffding bounds since

$$W(u)\leq C(u)\leq M(u)\,.$$

Note that both bounds are (pointwise) sharp.

2.1.1 Rotational and Permutational Symmetry

We start with proving some general results that will turn out to be useful in practical work later on. By $\overline{F}\colon x\mapsto P[X>x], x\in I\!R^n$ we denote the survival function of F. Since C is a distribution function itself, $\overline{C}$ is defined analogously. Furthermore, if C denotes the copula belonging to F, we define the survival copula $\widetilde{C}$ by

$$\overline{F}(x_1,\ldots,x_n)=\widetilde{C}(\overline{F}_1(x_1),\overline{F}_2(x_2),\ldots,\overline{F}_n(x_n))\,. \tag{2.1.3}$$

Note that $\widetilde{C}$ is a copula, whereas $\overline{C}$ is not.

The following proposition states the relationship between $\tilde{C}$ and $\overline{C}$.

Proposition 2.1.1. *Let C be an n-copula, $\overline{C}$ its survival function, and $\widetilde{C}$ the corresponding survival copula. Then*

$$\widetilde{C}\left(u_1,\ldots,u_n\right)=1+\sum_{P\in\mathcal{P}^*}\left(-1\right)^{|P|}C\left(1_{1\in P}\left(1-u_1\right)+1_{1\notin P},\ldots,1_{n\in P}\left(1-u_n\right)+1_{n\notin P}\right) \tag{2.1.4}$$

and

$$\widetilde{C}\left(1-u\right)=\overline{C}\left(u\right) \tag{2.1.5}$$

for all $u=(u_1,...,u_n)\in[0,1]^n$. Here $\mathcal{P}^*$ is the set of nonempty subsets of $\{1,...,n\}$, i.e. $\mathcal{P}^*=\mathcal{P}\left(\{1,...,n\}\right)\setminus\{\emptyset\}$ and $|P|$ is the cardinal number of $P\in\mathcal{P}^*$.

Proof. Let F be a distribution function with marginals $F_1,...,F_n$. If C is the copula for F, the corresponding survival copula $\widetilde{C}$ fulfils

$$\overline{F}\left(x_1,...,x_n\right)=\widetilde{C}\left(\overline{F}_1\left(x_1\right),...,\overline{F}_n\left(x_n\right)\right)\ .$$

According to standard literature, $\overline{F}$ can be written as

$$\overline{F}\left(x_1,...,x_n\right)=1+\sum_{P\in\mathcal{P}^*}\left(-1\right)^{|P|}F_P\left(x_j,\ j\in P\right)\ .$$

Here F_P denotes the $|P|$-marginals of F belonging to the index sets in P. Writing $C\left(1_{1\in P}\left(u_1\right)+1_{1\notin P},\ldots,1_{n\in P}\left(u_n\right)+1_{n\notin P}\right)$ for the $|P|$-marginals of C we calculate

$$\begin{aligned}\overline{F}(x_1,...,x_n) &= 1+\textstyle\sum_{P\in\mathcal{P}^*}\left(-1\right)^{|P|}F_P\left(x_j,\ j\in P\right)\\ &= 1+\textstyle\sum_{P\in\mathcal{P}^*}(-1)^{|P|}C\left(1_{1\in P}(F_1(x_1))+1_{1\notin P},\ldots,1_{n\in P}(F_n(x_n))+1_{n\notin P}\right)\\ &= 1+\textstyle\sum_{P\in\mathcal{P}^*}(-1)^{|P|}C\left(1_{1\in P}(1-\overline{F}_1(x_1))+1_{1\notin P},\ldots,1_{n\in P}(1-\overline{F}_n(x_n))+1_{n\notin P}\right)\ .\end{aligned}$$

This proves (2.1.4). Equation (2.1.5) follows by

$$\begin{aligned}\overline{C}(u_1,...,u_n) &= 1+\textstyle\sum_{P\in\mathcal{P}^*}(-1)^{|p|}C(1_{1\in P}(u_1)+1_{1\notin P},\ldots,1_{n\in p}(u_n)+1_{n\notin P})=\\ &1+\textstyle\sum_{P\in\mathcal{P}^*}(-1)^{|p|}C(1_{1\in P}(1-(1-u_1))+1_{1\notin P},\ldots,1_{n\in P}(1-(1-u_n))+1_{n\notin P})\\ &= \widetilde{C}\left(1-u_1,...,1-u_n\right)\ .\end{aligned}$$

□

This implies the following result for copula densities.

Lemma 2.1.2. *Let C be an n-copula with density c, let $\widetilde{c}$ be the density belonging to the survival copula $\widetilde{C}$. Then for all $u\in[0,1]^n$ we have $\widetilde{c}(u)=c(1-u)$.*

Proof. Follows directly by equation (2.1.4) since

$$\widetilde{c}(u_1,\ldots,u_n)=\frac{\partial\widetilde{C}\left(U_1,\ldots,U_n\right)}{\partial U_1\cdots\partial U_n}\left(u_1,\ldots,u_n\right).$$

□

Let us approach Lemma 2.1.2 from a statistical point of view. Assume that we want to estimate a survival copula $\widetilde{C}_{\theta_s}$ depending on a parameter vector θ_s. In the light of the lemma, estimating θ_s by maximum likelihood (ML) is equivalent to ML fitting the original copula C_θ to the rotated dataset $1-u$ for data u. This is true since $\widehat{\theta} \equiv \widehat{\theta}_s$ because optimization within an ML routine is based solely on the density. Lemma (2.1.2) then yields $\widetilde{c}(u) = c(1-u)$.

Definition 2.1.2. *Let C be an n-copula. We call C* radially symmetric *if the boxes $[0,u_1]\times\ldots\times[0,u_n]$ and $[1-u_1,1]\times\ldots\times[1-u_n,1]$ have the same C-volume for all $u \in [0,1]^n$ and we call C* permutational symmetric *if $C(u_1,\ldots,u_n) = C\left(u_{\pi(1)},\ldots,u_{\pi(n)}\right)$ for all permutations π on $(1,\ldots,n)$ and $u=(u_1,\ldots,u_n)\in[0,1]^n$.*

So if rotating the data set does not change the volume under C, the copula and its survival copula are the same. (The following result is an n-dimensional combination of Theorem 2.7.2 and Theorem 2.7.3 in Nelsen [92].)

Lemma 2.1.3. *Let C be an n-copula, then C is radially symmetric iff $C=\widetilde{C}$.*

Proof. Let $(u_1,\ldots,u_n) \in [0,1]^n$ arbitrary. The C-volume over $[0,u_1]\times\ldots\times[0,u_n]$ is simply $C(u_1,...,u_n)$ what is assumed to be equal to the C-volume over $[1-u_1,1]\times\ldots\times[1-u_n,1]$. But the latter is just the survival probability of $(1-u_1,...,1-u_n)$ and hence given by $\overline{C}(1-u_1,...,1-u_n) = \widetilde{C}(u_1,...,u_n)$ due to equation (2.1.5). The opposite direction follows by going backwards through the proof. □

2.2 Dependence Concepts

The next two definitions recall global dependence concepts. Let $(X_1,\ldots,X_n)\sim F = C\circ(F_1,\ldots,F_n)$ where F has a density f and C is a copula.

Definition 2.2.1. *The copula C is called* positive upper orthant dependent *(short: PUOD) if*

$$\mathbf{P}(X_j > z_j, j=1,\ldots,n) \geq \prod_{k=1}^{n} \mathbf{P}(X_j > z_j)$$

for all z. If

$$\mathbf{P}(X_j \leq z_j, j=1,\ldots,n) \geq \prod_{k=1}^{n} \mathbf{P}(X_j \leq z_j),$$

C is called positive lower orthant dependent *(short: PLOD)*

Definition 2.2.2. *A non-negative function κ on $\mathbb{R}^n$ is called* multivariate totally positive of order 2 *(MTP2) if*

$$\kappa(\max\{x_1, y_1\}, \ldots, \max\{x_n, y_n\})\, \kappa(\min\{x_1, y_1\}, \ldots, \min\{x_n, y_n\}) \geq \kappa(x)\, \kappa(y)$$

for all $x, y \in \mathbb{R}^n$.

We say that a random vector X is MTP2 if its distribution function is MTP2.

We now assume that X and Y are distributed with the same copula and recall common definitions for stochastic order for random vectors.

Definition 2.2.3. *We say that random vectors X and Y are* stochastically ordered *if and only if $Ef(X) \leq Ef(Y)$ holds for all increasing functions f. We write for short $X \leq_{st} Y$. If the same relation holds for all convex f we call X and Y* convex ordered *and write $X \leq_{cx} Y$.*

By using stochastic ordering, we can define another form of positive dependence. We refer to Müller and Stoyan [90], p. 125 for the notation and the next 2 definitions. If X and Y are n-dimensional random vectors, X is said to be stochastically increasing in Y ($X \uparrow_{st} Y$ for short) if the conditional distribution of X given $Y = y$ is increasing in y with respect to the stochastic order $\leq_{st}$.

Definition 2.2.4. *A random vector $X = (X_1, \ldots, X_n)$ is called* conditionally increasing *(short: CI) if $X_i \uparrow_{st} (X_j,\, j \in J)$ for all $i \in \{1, \ldots, n\}$ and all subsets $J \subset \{1, \ldots, n\}$.*

So the random vector X is considered as positively dependent in the sense that the conditional distribution of a component given the knowledge of some other components, is stochastically increasing in the components that are known.

Remark 4. *For the bivariate case, the following relations can be shown.*

- *MTP2 implies PUOD. (see Joe [62] Theorem 2.3, p. 26)*
- *If a bivariate random vector X is MTP2, one can show that X is CI, too.*

Definition 2.2.5. *If C_1 and C_2 are copulas, we say C_1 is smaller than C_2 (or C_2 is larger than C_1), and write $C_1 \prec C_2$ (or $C_2 \succ C_1$) if $C_1(u, v) \leq C_2(u, v)$ for all $(u, v) \in [0,\, 1]^2$. This order is called* concordance order. *We call a parametric family of copulas* positively ordered *with respect to the parameter θ, if $\theta_1 \leq \theta_2$ implies $C_{\theta_1} \prec C_{\theta_2}$ for all θ_1, θ_2 and* negatively ordered, *if $C_{\theta_1} \succ C_{\theta_2}$ analogously.*

This order can help to compare the diversification of different portfolios.

2.2.1 Measures of Dependence

Let us turn towards an investigation how the dependence structure in F can be measured. In this subsection we will recall the definitions of the two most popular dependence measures (besides the correlation of course). In later chapters we will make extensive use of them.
Kendall's τ is an appropriate dependence measure for *monotonic* dependence.

Definition 2.2.6. *[Kendall's τ]Let the bivariate random vector $\left(\tilde{X}, \tilde{Y}\right)$ be an independent copy of (X, Y). Kendall's τ is defined as*

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

Since Kendall's τ is a rank correlation,

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

holds, i.e. Kendall's τ is completely determined by the copula of $(X, Y)'$ and so it depends only on the copula parameters of the distribution of $(X, Y)'$. Now let (x_i, y_i), (x_j, y_j) be realisations of a random vector (X, Y). The pairs (x_i, y_i) and (x_j, y_j) are *concordant* if ($x_i < x_j$ and $y_i < y_j$) or ($x_i > x_j$ and $y_i > y_j$) and they are *discordant* if ($x_i < x_j$ and $y_i > y_j$) or ($x_i > x_j$ and $y_i < y_j$).
So Kendall's τ can be interpreted as the probability for an observation of (X, Y) to be concordant minus the probability to be discordant. For a sample $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ of observations from a continuous (X, Y), a *sample version of Kendall's Tau* can be estimated due to

$$t = \frac{c - d}{c + d} = \frac{c - d}{\begin{pmatrix} n \\ 2 \end{pmatrix}} \tag{2.2.1}$$

where c is the number of concordant pairs and d the number of discordant pairs. Kendalls τ avoids some of the pitfalls known for the correlation in a non elliptical framework. Especially, we have the relation $\tau_{X,Y} = 1$ iff X, Y are comonotone, $\tau_{X,Y} = -1$ iff X, Y are countermonotone, is generally not true for the correlation measure. See, e.g., Embrechts et al. [42].

In addition to monotonic dependence, which is measured by rank correlation, financial data are likely to exhibit lower *tail* dependence, resulting in a high probability of extreme simultaneous losses. Tail-dependence definitions for random vectors are mostly related to their bivariate marginal distribution functions. Loosely speaking, tail dependence describes the limiting proportion of one margin

exceeding a certain threshold given that the other margin has already exceeded that threshold. The following approach (Joe [62], p. 33) represents one of many possible definitions of tail dependence. See Schmidt and Stadtmüller [106] for a generalization to multivariate tail dependence.

Definition 2.2.7. *Let $X = (X_1, X_2)$ be a 2-dimensional random vector with joint distribution function F and marginal distribution functions F_1 and F_2. We say that (X_1, X_2) is (bivariate) upper tail-dependent if*

$$\lambda_U := \lim_{v \to 1^-} \lambda_U(v) = \lim_{v \to 1^-} \mathbf{P}(X_1 > F_1^{\leftarrow}(v) \mid X_2 > F_2^{\leftarrow}(v)) > 0, \tag{2.2.2}$$

in case the limit exists. Here $F_1^{\leftarrow}$ and $F_2^{\leftarrow}$ denote the generalized inverse distribution functions of X_1, X_2. Consequently, we say (X_1, X_2) is upper tail-independent if λ_U equals 0. Further, we call λ_U the upper tail-dependence coefficient (TDC). Similarly, we define the lower tail-dependence *coefficient by*

$$\lambda_L := \lim_{v \to 0^+} \mathbf{P}(X_1 \leq F_1^{\leftarrow}(v) \mid X_2 \leq F_2^{\leftarrow}(v)). \tag{2.2.3}$$

For the sake of simplicity we write λ for the TDC whenever $\lambda_L = \lambda_U$. The tail-dependence concept can be embedded within the copula theory in the following sense. Thus many copula features transfer to the tail-dependence coefficient, for example the invariance under strictly increasing transformations of the marginals. If (X_1, X_2) is a continuous bivariate random vector, then

$$\lambda_U = \lim_{v \to 1^-} \frac{1 - 2v + C(v, v)}{1 - v}, \tag{2.2.4}$$

where C denotes the copula of (X_1, X_2). Analogously,

$$\lambda_L = \lim_{v \to 0^+} \frac{C(v, v)}{v} \tag{2.2.5}$$

holds for the lower tail-dependence coefficient.

Since tail dependence is defined by means of copula, beside Kendall's τ also λ_L and λ_U depend only on copula parameters.

2.3 Parametric Copula Classes

We now introduce two standard classes of copulas, which will be used later on.

We start with the class of *elliptical copulas* which are often used as benchmark models. Here, by elliptical copulas we simply mean the copulas of elliptical distributions, e.g. given by Equation (2.1.2) with F being elliptical. In particular, this includes the copula of the Student-t and the normal distribution function.

Example 1. *Examples of elliptical copulas are given by:*

(a) The t-copula C_t is given by

$$C_t(u,v) = T_{\nu,\rho}\left(T_\nu^{\leftarrow}(u), T_\nu^{\leftarrow}(v)\right), \tag{2.3.1}$$

where $T_{\nu,\rho}$ is the bivariate standardized Student-t distribution function with ν degrees of freedom and correlation ρ, while T_ν denotes the univariate standardized Student-t distribution function. The upper and lower tail dependence parameter λ for $\nu > 2$ is (see Embrechts et al. [41])

$$\lambda = 2\left(1 - T_{\nu+1}\left(\sqrt{\nu+1}\frac{\sqrt{1-\rho}}{\sqrt{1+\rho}}\right)\right). \tag{2.3.2}$$

Obviously, the tail dependence parameter depends on ρ, and if $\rho = 0$ the tail dependence is restricted by $\lambda < 0.5$ ($\nu \to 0$) and $\lambda < 0.1817$ ($\nu = 2$), respectively. (Note that the above expression for λ even holds in the case $0 < \nu \leq 2$ with a different interpretation of ρ.)

(b) For $\nu \to \infty$ the t-copula degenerates to the copula of the normal distribution

$$C_N(u,v) = N_\rho\left(N^{\leftarrow}(u), N^{\leftarrow}(v)\right), \tag{2.3.3}$$

where $N_\rho(\cdot)$ and $N(\cdot)$ denote the standard bivariate and the standard univariate normal distribution functions, respectively. From Equation (2.3.2), it is obvious that the normal copula reveals no tail dependence, i.e. $\lambda = 0$.

In general, elliptical copulas cannot be expressed in closed form and are restricted to permutational and radial symmetry. For a detailed discussion about limitations see Frahm et al. [50], and Chapter 3.

Secondly, we introduce the class of Archimedean copulas. These copulas can be described by a one-dimensional function called *generator*. We start with the two dimensional case and generalize it to arbitrary high dimensions afterwards in Chapter 4.

Proposition 2.3.1. *Let $\varphi : [0,\,1] \to [0,\,\infty]$ be continuous and strictly decreasing with $\varphi(1) = 0$. The function $C : [0,\,1]^2 \to [0,\,1]$ given by*

$$C(u,v) = \varphi^{\leftarrow}\left(\varphi(u) + \varphi(v)\right) \tag{2.3.4}$$

is a copula if and only if φ is convex.

Here $\varphi^{\leftarrow} : [0,\,\infty] \to [0,\,1]$ denotes the generalized inverse of φ. The copula constructed by (2.3.4) is called Archimedean. The function φ is called *generator* of the copula. A generator φ is called strict if $\varphi(0) = \infty$ and in this case it is invertible, i.e. $\varphi^{\leftarrow} = \varphi^{-1}$.

Example 2. *Examples of Archimedean copulas are given by:*

(a) The independence copula with generator $\varphi_\Pi(q) = -\ln q$ and

$$C_\Pi(u, v) = uv. \tag{2.3.5}$$

The independence copula exhibits neither lower nor upper tail dependence, i.e. $\lambda_L = \lambda_U = 0$.

(b) The Gumbel copula with generator $\varphi_G(q) = (-\ln q)^\delta$ where $\delta \in [1, \infty)$ and

$$C_G(u, v) = \exp\left(-\left[(-\ln u)^\delta + (-\ln v)^\delta\right]^{\frac{1}{\delta}}\right). \tag{2.3.6}$$

The Gumbel copula is upper tail dependent with $\lambda_U = 2 - 2^{\frac{1}{\delta}}$ but has no lower tail dependence.

(c) The Frank copula with generator $\varphi_F(q) = -\ln \frac{e^{-\vartheta \cdot q} - 1}{e^{-\vartheta} - 1}$ where $\vartheta \in (-\infty, \infty) \setminus \{0\}$ and

$$C_F(u, v) = -\frac{1}{\vartheta} \ln\left(1 + \frac{\left(e^{-\vartheta u} - 1\right)\left(e^{-\vartheta v} - 1\right)}{e^{-\vartheta} - 1}\right). \tag{2.3.7}$$

The Frank copula shows neither upper nor lower tail dependence

(d) The Clayton copula with generator $\varphi_C(q) = \frac{1}{\delta}\left(q^{-\delta} - 1\right)$ where $\delta \geq 0$ and

$$C_{Cl}(u, v) = \left(u^{-\delta} + v^{-\delta} - 1\right)^{-\frac{1}{\delta}}. \tag{2.3.8}$$

The Clayton copula has lower tail dependence with tail dependence parameter $\lambda_L = 2^{-\frac{1}{\delta}}$ and no upper tail dependence, $\lambda_U = 0$.
We would like to mention that the Clayton copula is sometimes called Cook-Johnson copula and is also mentioned in Joe [62], p. 141 as Family B4, introduced by Kimeldorf and Sampson (1975).

We will concentrate on Archimedean copulas and their properties in Chapter 4, where in Chapter 3 we will focus on elliptical copulas.

2.4 Pitfalls in Copula Theory

At the end of this chapter we will outline that the belief in copulas as a tool that can describe *any* dependence structure between random variables is, unfortunately, unjustified. Suppose e.g. Microsoft and a small enterprize testing the

prototype software for Microsoft. If the request for Microsoft products is high, there is a good chance for new releases and products developed by Microsoft, that have to be tested. If everyone switches to use Linux instead of Windows, there will be not enough work for the small enterprize to survive. So it is obvious that the fortune of the test company highly depends on the market conditions for Microsoft. Vice versa, the financial welfare of the test company has only little or in fact no influence on Microsoft at all. Economically, we can say that Microsoft is described by a real subset of the risk factors that can describe the test company. The resulting copula will be a linear combination of the total positive dependence copula (events that drive Microsoft and hence the small company, too) with the independence copula (events that drive the small company but has no effect on Microsoft).

Even if we see total dependence in an e.g. bivariate random vector (U, V), it is (economically) only ensured that one variable is a strict monotone function of the other. Mathematically, we have of course $V = F(U)$ what implies $U = F^{\leftarrow}(V)$ but to conclude that the dependence of U from V is as strong as the one of V from U is economically questionable. As an example we see in Figure 2.1 the

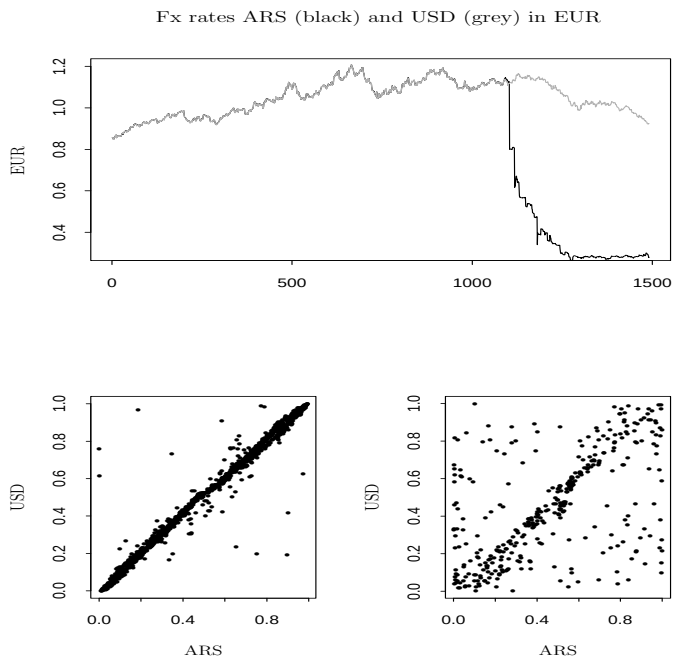

Figure 2.1: Fx rates ARS (black) and USD (grey) in EUR (upper plot), copula pseudo observations of the (ARS,USD) GARCH innovation portfolio during January 1999 - January 2002 (lower left plot), and during February 2002 - June 2003 (lower right plot), respectively.

scatterplot of the copula pseudo observations of the foreign exchange rate of the Argentinean peso (ARS) and the USD, both with respect to the EUR. For the mapping to uniformity we have used the empirical distribution functions of the GARCH innovations. For the left lower plot the observation period was January 1999 - January 2002, where the peso was by Argentinean law coupled to the USD. Due to economic pressure the coupling had to be resigned in January 2002. The right lower plot shows the observations from February 2002 - June 2003. As we can see from the left plot, either the Argentinean peso or the USD highly depend on the other currency. The problem is that from the copula solely we are not able to identify which one is the dominant factor. Of course we just know that the Argentinean peso was connected to the USD during the observation period. The upper plot shows the fx rate to the EUR of both currencies. It is obvious that the Argentinean crisis has not effected the USD in a significant way. In the right plot we are able to identify the dominance of the USD, because of the slightly observable asymmetry ($C(u, v) \neq C(v, u)$) for extremal loss events.
To summarize, whenever there is directional (or causal) dependence in-between the components of a random vector X in the sense that some X_i depends on X_j, $j \neq i$, in an other way than X_j on X_i, the copula may fail to incorporate the directional information. We are probably able to determine the existence of such dependence with copulas, but we are not able to derive the direction without further information that is not provided by the copula.

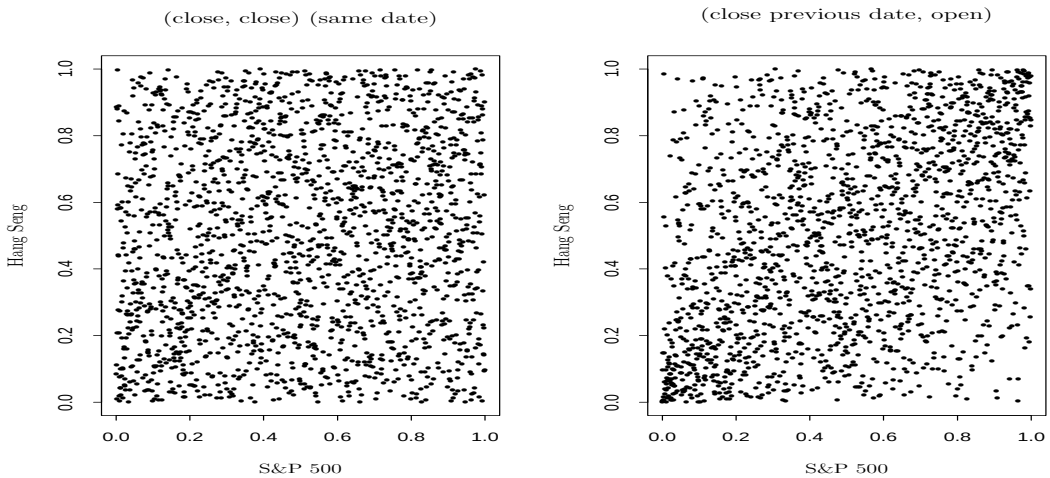

Figure 2.2: Pseudo copula observations of a (S&P 500, Hang Seng) GARCH innovation portfolio. Left plot shows scatterplot with observations created by (close, close) quotes of the same date, whereas in the right plot they are created by (close previous day, open) quotes.

Another pitfall is not to take caution of possible time lags in-between the

elements of multidimensional observations. For example, Figure 2.2 shows the scatterplots of the copula pseudo observations (po) of the GARCH innovations of an index-portfolio containing the S&P 500 and the Hang Seng stock indices, as quoted in New York and Hong Kong, respectively. The left plot shows the (close,close) pos of the same date. Since the New York stock exchange closes at 16:00 h (local time!) and the Hong Kong exchange at 15:30 h (local time!) the real time lag in-between the univariate observations is 13.5 hours. Coupling the close pos of S&P 500 of the previous day with the open pos of the Hang Seng leads to the right plot. With an opening time at 10:00 h (local time) of the Hong Kong exchange, this procedure minimizes the real time lag in-between the observations to 5 hours. Remark that New York and Hong Kong time is also separated by the international date line. However, it is obvious that there is a big difference between the two plots. Whereas the left one (big time lag) manifests only weak dependence, the right one (minimized time lag) shows evidence for significant positive dependence.

So copulas (especially their estimation) are very sensitive for time lags in-between the univariate elements of the multivariate observations. Hence, of course the estimation is also very sensitive for shift errors in data.

Chapter 3

Elliptical Copulas: Applicability and Limitations

In this chapter (based on Frahm et al. [50]), we study copulas generated by elliptical distributions. We show that their tail dependence can simply be computed with default routines on Student's t-distribution given Kendall's τ and the tail index. The copula family generated by the sub-Gaussian α-stable distribution is unable to cover the magnitude of tail dependence observed in financial data.

3.1 Introduction

Elliptical distribution families are widely applied in statistics and econometrics, especially in finance. For example, Owen and Rabinovitch [94] discuss the impact of elliptical distributions on portfolio theory, and Kim [72] studies the CAPM for this class of distributions. The normal distribution is the archetype of elliptical distributions in finance. Besides this distribution family, in recent time other elliptical distribution families became more important. Student's t-distribution is emphasized, e.g., by Blattberg and Gonedes [11]. The sub-Gaussian α-stable distribution is a special case of the extensive stable distribution class discussed in Rachev and Mittnik [99]. Portfolio selection with sub-Gaussian α-stable distributed returns is analyzed by Ortobelli et al. [93].

In this chapter we study the dependence structure generated by elliptical distributions. We focus on the multivariate Student's t-distribution and the multivariate sub-Gaussian α-stable distribution since these distribution families are frequently applied for modelling financial data. This subject is relevant with respect to risk analysis in finance. Here, the probability of joint extremal events has to be determined because the aggregate portfolio risk increases seriously in case of noticeable dependence of the particular assets in crash situations. Therefore, besides rank correlation measures, like Kendall's τ (see Definition 2.2.6), the

concept of tail dependence is of main interest. The copula concept is utilized to extract the dependence structure of a given distribution by the copula function. Thus, it is obtained a separation of the marginal distributions of the particular random components and the dependence structure that is contained in the copula. Recent developments on elliptical distributions concerning the dependence structure can be found in Hult and Lindskog [60], and Schmidt [104].

For copulas generated by elliptically distributed random vectors we discuss a general relationship between Kendall's τ, the tail index of the underlying elliptical distribution, and the tail dependence of the generated copula. Given Kendall's τ and the tail index ζ, the tail dependence can simply be computed with default routines on Student's t-distribution. Applying these results we find that sub-Gaussian α-stable copulas are not suitable for modelling the dependence structure of financial risk. Empirical investigations have shown that the tail dependence of stocks is smaller than those provided by the sub-Gaussian α-stable distribution assumption.

3.2 Copula and Marginal Parameters

An axiomatic definition of copulas is to be found in Joe [62] and Nelsen [92]. For a short introduction see Chapter 2.

In general, a multivariate distribution F contains parameters that do not affect the copula of F, whereas other parameters affect the copula and possibly the marginals. We want to call the latter type of parameters "copula parameters".

Let $\theta \in I\!R^n$ be a parameter vector and $F(\cdot\,;\theta)$ a continuous multivariate distribution with copula $C(\cdot\,;\theta)$. Let $I_C \subset I = \{1, ..., n\}$ be an index-set that contains all k for which at least one $u \in [0,1]^d$ exists, such that

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

So I_C contains all copula parameter indices.

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

$$F(x_1, ..., x_d; \theta) = F^*(h_1(x_1;\theta_1), ..., h_d(x_d;\theta_d); \theta_0), \tag{3.2.1}$$

with

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

Lemma 3.2.1. *Only the vector θ_0 contains copula parameters.*

Proof. The lemma follows from the fact that any copula is invariant under strictly monotone increasing transformations $h_1(\cdot\,;\theta_1), ..., h_d(\cdot\,;\theta_d)$. Thus also the parameters $\theta_1, ..., \theta_d$ can not affect the copula. □

So the parameters $\theta_1, ..., \theta_d$ are canceled down through copula separation and θ_0 still remains. We call the distribution $F^*(\cdot\,;\theta_0)$ the "underlying distribution" of $C(\cdot\,;\theta)$, and $C(\cdot\,;\theta)$ the "copula generated" by $F^*(\cdot\,;\theta_0)$. That is we use Equation (2.1.1) (Sklar's Theorem) to extract a copula from a known multidimensional distribution function. In particular, the copula generated by an elliptical distribution will be called "elliptical copula".

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

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

with given parameter vector θ_0, variable location parameters $\mu_1, ..., \mu_d$ and scale parameters $\sigma_1, ..., \sigma_d$. So this distribution family is generated by affine marginal transformations and the location and scale parameters are not copula parameters.

3.3 Elliptical Distributions

3.3.1 Characterizations

Definition 3.3.1. *The d-dimensional random vector X has an elliptical distribution if and only if the characteristic function $t \mapsto E(\exp(it'X))$ with $t \in \mathbb{R}^d$ has the representation*

$$t \longmapsto \phi_g(t; \mu, \Sigma, \vartheta) = \exp(it'\mu) \cdot g(t'\Sigma t; \vartheta).$$

Here $g(\cdot\,;\vartheta) : [0, \infty[\to \mathbb{R}$, $\vartheta \in \mathbb{R}^m$, $\mu \in \mathbb{R}^d$, and Σ is a symmetric positive semidefinite $d \times d$-matrix.

The parameter vector μ is a location parameter, and the matrix Σ determines the scale and the correlation of the random variables $X_1, ..., X_d$. The function $g(\cdot\,;\vartheta)$ constitutes the distribution family and is called "characteristic generator", whereas ϑ is a parameter vector that determines the shape, in particular the tail index of the distribution.

For example the characteristic function of the sub-Gaussian α-stable distribution (Ortobelli et al. [93]) is

$$t \longmapsto \exp\left(it'\mu\right) \cdot \exp\left(-\left(\frac{1}{2} \cdot t'\Sigma t\right)^{\alpha/2}\right), \qquad 0 < \alpha < 2,$$

so the characteristic generator is $x \mapsto \exp\left(-\left(\frac{1}{2} \cdot x\right)^{\alpha/2}\right)$ with shape parameter α.

Now let

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

with

$$\sigma_i := \sqrt{\sigma_{ii}}, \qquad i = 1, ..., d$$

and

$$\rho_{ij} := \frac{\sigma_{ij}}{\sigma_i \sigma_j}, \qquad i, j = 1, ..., d, \tag{3.3.1}$$

so that $\Sigma = \sigma\rho\sigma$ and $\phi_g\left(\cdot\,; \mu, \Sigma, \vartheta\right) \equiv \phi_g\left(\cdot\,; \mu, \sigma, \rho, \vartheta\right)$. Consider the multivariate elliptical distribution $F_g^*\left(\cdot\,; \rho, \vartheta\right)$ corresponding to the characteristic function $\phi_g\left(\cdot\,; \underline{0}, \underline{1}, \rho, \vartheta\right)$. The characteristic function corresponding to the multivariate distribution

$$(x_1, ..., x_d)' \longmapsto F_g^*\left(\frac{x_1 - \mu_1}{\sigma_1}, ..., \frac{x_d - \mu_d}{\sigma_d}; \rho, \vartheta\right)$$

with $\mu \in I\!R^d$ and $\sigma_1, ..., \sigma_d > 0$ is

$$\begin{aligned} t \longmapsto \exp\left(it'\mu\right) \cdot \phi_g\left(\sigma t; \underline{0}, \underline{1}, \rho, \vartheta\right) &= \exp\left(it'\mu\right) \cdot g\left(\left(\sigma t\right)' \rho\left(\sigma t\right); \vartheta\right) \\ &= \exp\left(it'\mu\right) \cdot g\left(t'\left(\sigma\rho\sigma\right) t; \vartheta\right) \\ &= \phi_g\left(t; \mu, \Sigma, \vartheta\right). \end{aligned}$$

Thus the location-scale-family generated by a standard elliptical distribution is elliptical, too. This is emphasized by a stochastic representation of the elliptical class which is stated by the following theorem:

Theorem 3.3.2 (Fang et al. [45]). *The d-dimensional random vector X is elliptically distributed with characteristic function $\phi_g\left(\cdot\,; \mu, \Sigma, \vartheta\right)$ and* $\operatorname{rank}\left(\Sigma\right) = r \leq d$ *if and only if there exists an r-dimensional random vector U, uniformly distributed on the unit sphere*

$$\left\{u \in [-1, 1]^r : \|u\| = 1\right\},$$

a nonnegative random variable $\mathcal{R}$ independent of U, and a $d \times r$-matrix $\sqrt{\Sigma}$ with $\sqrt{\Sigma}\sqrt{\Sigma}' = \Sigma$, such that

$$X \stackrel{d}{=} \mu + \mathcal{R}\sqrt{\Sigma}\, U. \tag{3.3.2}$$

Due to the transformation matrix $\sqrt{\Sigma}$ the uniform random vector U produces elliptically contoured density level surfaces, whereas the "generating random variable" $\mathcal{R}$ (Schmidt [104], p. 6) gives the distribution shape, in particular the tailedness of the distribution.

With the reparametrization (3.3.1) the equation $\sqrt{\Sigma} = \sigma\sqrt{\rho}$ holds, and so

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

3.3.2 Dependence Structure

The standard density of the d-dimensional sub-Gaussian α-stable distribution can be obtained through multivariate Fourier-transformation (see Appendix B.1) and is

$$\begin{aligned} f^*_{\alpha,\rho}(x) &= \frac{1}{(2\pi)^d} \cdot \int\limits_{\mathbb{R}^d} \phi_{\text{stable}}\left(t; \underline{0}, \underline{1}, \rho, \alpha\right) \cdot \exp\left(-it'x\right) dt' \\ &= \frac{1}{(2\pi)^d} \cdot \int\limits_{\mathbb{R}^d} \exp\left(-\left(\frac{1}{2} \cdot t'\rho t\right)^{\alpha/2}\right) \cdot \cos\left(t'x\right) dt', \qquad 0 < \alpha < 2. \end{aligned}$$

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

$$C_\alpha\left(u_1, \ldots, u_d\right) = F^*_{\alpha,\rho}\left(F^{*\leftarrow}_{\alpha,1}\left(u_1\right), \ldots, F^{*\leftarrow}_{\alpha,1}\left(u_d\right)\right),$$

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

$$F^*_{\alpha,\rho}(x) = \int\limits_{]-\infty,x]} f^*_{\alpha,\rho}(w)\, dw',$$

with $]-\infty, x] :=]-\infty, x_1] \times \cdots \times]-\infty, x_d]$, and $F^{*\leftarrow}_{\alpha,1}$ is the inverse of the univariate standard distribution function

$$F^*_{\alpha,1}(x) = \int\limits_{-\infty}^{x} f^*_{\alpha,1}(w)\, dw.$$

The standard density of the d-dimensional multivariate Student's t-distribution with ν degrees of freedom is

$$f^*_t(x) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \cdot \sqrt{\frac{1/|\rho|}{(\nu\pi)^d}} \cdot \exp\left(-\frac{1}{2} \cdot \log\left(1 + \frac{x'\rho^{-1}x}{\nu}\right)^{\nu+d}\right), \qquad \nu > 0.$$

For continuous elliptical distributions there is a straight link between Kendall's τ and Pearson's ρ:

Theorem 3.3.3 (Lindskog et al. [77]). *Let X be an elliptically distributed random vector with characteristic function $\phi_g(\cdot\,;\mu,\sigma,\rho,\vartheta)$. For two continuous components of X, say X_i and X_j, Kendall's τ is*

$$\tau(X_i, X_j) = \frac{2}{\pi} \cdot \arcsin(\rho_{ij}). \tag{3.3.3}$$

This is to say Kendall's τ depends only on ρ, and neither the characteristic generator nor the shape of the distribution affect the rank correlation (see also Fang et al. [44]).

We know that the specified distributions are heavy tailed, i.e. the marginal survival functions $\overline{F}_i$ exhibit a power law with tail index $\zeta > 0$:

$$\overline{F}_i(x) = \lambda_i(x) \cdot x^{-\zeta}, \qquad x > 0, \tag{3.3.4}$$

for $i = 1, ..., d$. Here $\lambda_1, ..., \lambda_d$ are slowly varying functions. Note that elliptically contoured distributions are symmetric (heavy) tailed, i.e. the tail indices for the upper tail $\overline{F}_i$ and the lower tail F_i coincide.

For the multivariate Student's t-distribution the tail index corresponds to the number of the degrees of freedom ν, and for the multivariate sub-Gaussian α-stable distribution the tail index equals α. Furthermore, in the elliptical framework the lower tail dependence is equal to the upper tail dependence. The following theorem connects the tail index with the tail dependence of elliptical distributions.

Theorem 3.3.4 (Schmidt [104]). *Let X be an elliptically distributed random vector with characteristic function $\phi_g(\cdot\,;\mu,\sigma,\rho,\vartheta)$ and tail index ζ. For two components of X, say X_i and X_j, the tail dependence is*

$$\lambda(X_i, X_j; \zeta, \rho_{ij}) = \frac{\int_0^{f(\rho_{ij})} \frac{u^\zeta}{\sqrt{u^2-1}}\, du}{\int_0^1 \frac{u^\zeta}{\sqrt{u^2-1}}\, du}, \qquad \zeta > 0,\ f(\rho_{ij}) = \sqrt{\frac{1+\rho_{ij}}{2}}. \tag{3.3.5}$$

So the tail dependence for elliptical copulas is a function $\rho_\zeta \longmapsto \lambda$, where the tail index ζ of the underlying elliptical distribution family results from the characteristic generator g and the shape parameter ϑ. Given the matrix ρ the tail dependence is a function $\zeta_\rho \longmapsto \lambda$, and due to Theorem 3.3.3 also the relation $\zeta_\tau \longmapsto \lambda$ holds for a given matrix of Kendall's τ.

Remark 5. *The tail index is a property of the elliptical distribution family from which the copula is extracted from, whereas the tail dependence concerns the copula itself. By Sklar's Theorem (Nelsen [92]), it is possible to construct new multivariate distributions with arbitrary marginals, providing a specific copula. In this*

case ζ is generally not the tail index of the proposed marginal distributions but still a copula parameter.

Substituting the integration variable u in equation (3.3.5) by $\cos(v)$ leads to the following equivalent representation (see B.2) of the tail dependence of two elliptically distributed random variables X_i and X_j (Hult and Lindskog [60]):

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

Due to relation, (3.3.3) we can substitute ρ_{ij} by $\sin\left(\tau_{ij} \cdot \frac{\pi}{2}\right)$, and thus get

$$\lambda(X_i, X_j; \zeta, \tau_{ij}) = \frac{\int_{h(\tau_{ij})}^{\pi/2} \cos^{\zeta}(v)\ dv}{\int_0^{\pi/2} \cos^{\zeta}(v)\ dv}, \qquad h(\tau_{ij}) = \frac{\pi}{2}\left(\frac{1-\tau_{ij}}{2}\right). \tag{3.3.6}$$

Hence for the limiting case $\zeta = 0$, the tail dependence is an affine function of Kendall's τ:

$$\lim_{\zeta \searrow 0} \lambda(X_i, X_j; \zeta, \tau_{ij}) = \frac{1+\tau_{ij}}{2}. \tag{3.3.7}$$

Remark 6. *Since every random variable $X_1, ..., X_d$ has, as a component of X, the same tail index ζ which is given by the generating random variable $\mathcal{R}$, the tail dependence λ_{ij} of each bivariate combination $(X_i, X_j)'$ is uniquely determined by τ_{ij}. Thus modelling the tail dependence structure of elliptical copulas especially for higher dimensions is restricted by the set $\{(\lambda, \tau) \in [0,1] \times [-1,1] : \lambda = \lambda(\zeta, \tau)\}$ given the tail index parameter ζ.*

The tail dependence of a bivariate t-distributed random vector $(X, Y)'$ with ν degrees of freedom can be obtained by computation of (2.2.3) and is

$$\begin{aligned} \lambda &= 2 \cdot \bar{t}_{\nu+1}\left(\sqrt{\nu+1} \cdot \sqrt{\frac{1-\rho}{1+\rho}}\right) \\ &= 2 \cdot \bar{t}_{\nu+1}\left(\sqrt{\nu+1} \cdot \sqrt{\frac{1-\sin\left(\tau \cdot \frac{\pi}{2}\right)}{1+\sin\left(\tau \cdot \frac{\pi}{2}\right)}}\right), \qquad \nu > 0, \end{aligned} \tag{3.3.8}$$

where $\bar{t}_{\nu+1}$ is the survival function of the univariate Student's t-distribution with $\nu + 1$ degrees of freedom (see, e.g., Embrechts et al. [41]).

Since Equation (3.3.8) holds for all $\nu > 0$, where ν corresponds to the tail index ζ of X and Y, and Theorem 3.3.4 states that the tail dependence of two elliptically distributed random variables depends only on ρ_{ij} and ζ, Equation

(3.3.5) can be replaced by

$$\begin{aligned}\lambda_{ij} &= 2 \cdot \bar{t}_{\zeta+1}\left(\sqrt{\zeta+1} \cdot \sqrt{\frac{1-\rho_{ij}}{1+\rho_{ij}}}\right) \\ &= 2 \cdot \bar{t}_{\zeta+1}\left(\sqrt{\zeta+1} \cdot \sqrt{\frac{1-\sin\left(\tau_{ij} \cdot \frac{\pi}{2}\right)}{1+\sin\left(\tau_{ij} \cdot \frac{\pi}{2}\right)}}\right), \qquad \zeta > 0.\end{aligned} \tag{3.3.9}$$

Student's t-distribution is a default routine in statistics software and is tabulated in many textbooks (Johnson et al. [61]). So it is more convenient to use formula (3.3.9) than (3.3.5). In Figure 3.1 we plot the barriers of tail dependence as a function of ρ for any elliptical copula allowing for $\zeta > 0$. The range of possible tail dependence in the special case $\zeta < 2$, which holds for the sub-Gaussian α-stable copula, is marked explicitly.

An investigation of several stocks from the German and the US market shows that the lower tail dependence ranges from 0 to 0.35, whereas Kendall's τ takes values in between 0 to 0.4, approximately (Junker [64]). With formula (3.3.6) we can plot the tail dependence barriers as a function of Kendall's τ, see Figure 3.2. Note that for the limit case $\zeta = 0$ the tail dependence is an affine function of τ, as stated by (3.3.7). We see that the sub-Gaussian α-stable copula restricts the scope of possible tail dependence too much. The dependence structure generated by the multivariate sub-Gaussian α-stable distribution is not suitable for modelling financial risk because the provided range of λ has only a small intersection with the empirical results.

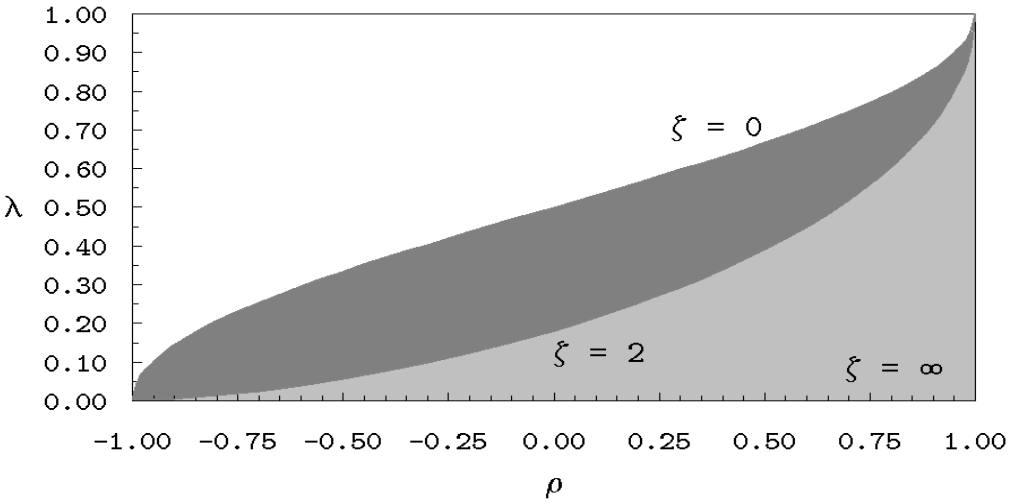

Figure 3.1: Range of possible tail dependence as a function of the correlation ρ.

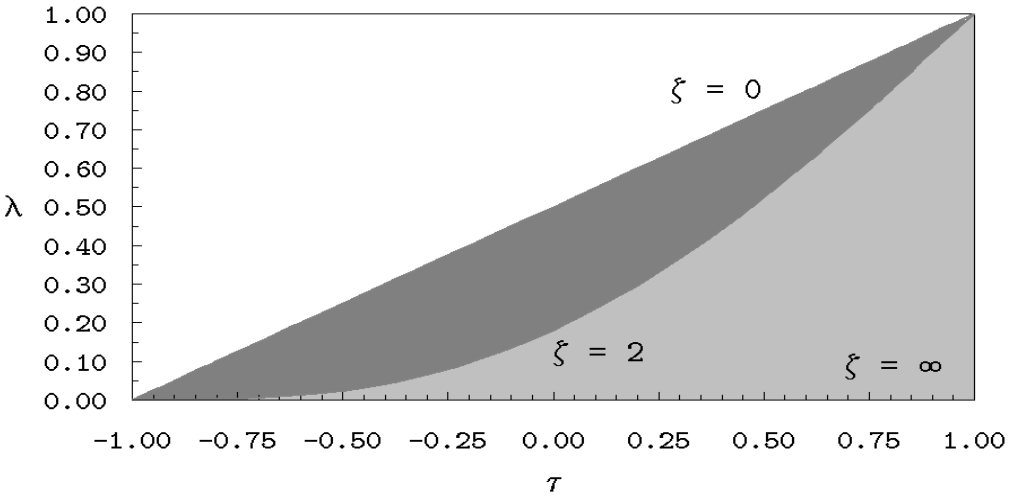

Figure 3.2: Range of possible tail dependence as a function of Kendall's τ.

Chapter 4

Archimedean Copulas- 2-dimensional Flexibility, Multidimensional Complexity

This chapter is partially taken out of Junker and May [67]. First we give a brief summary of well known results in the Archimedean class and show that this world is not closed with respect to convex linear combinations. We will further state a transformation rule for generators that extend the family of Archimedean copulas to new classes. Finally we come up with a special model that allows for different lower and upper tail dependence and contains the Frank and Gumbel copula as limit cases. The last section outlines the restrictions in constructing multidimensional models based on Archimedean copulas, and derives some of the resulting limitations.

4.1 Properties

Recall from Chapter 2 that a bivariate Archimedean copula C is completely described by a continuous, strictly decreasing and convex generator function $\varphi : [0, 1] \rightarrow [0, \infty]$ with $\varphi(1) = 0$ via the relation

$$C(u, v) = \varphi^{[-1]} \left(\varphi \left(u \right) + \varphi \left(v \right) \right). \tag{4.1.1}$$

The next proposition summarizes some of the properties of Archimedean copulas. A function g is m-monotone if all derivatives up to m exist and $(-1)^k \frac{d^k}{dt^k} g(t) \geq 0$ for all t and all $k = 0, 1, \ldots, m$. For $m = \infty$, all higher derivatives exists, and g is called completely monotone.

Proposition 4.1.1. *Let C be an Archimedean copula with generator φ. Then C is permutational symmetric, i.e. $C(u_1, u_2) = C(u_2, u_1)$ and associative, i.e.*

$C(C(u_1, u_2), u_3) = C(u_1, C(u_2, u_3))$. For any constant $\alpha > 0$, the scalar multiple $\alpha\varphi$ is a generator of C.

If the inverse generator $\varphi^{\leftarrow}$ is completely monotonic on $[0, \infty)$, then the density of C is totally positive of order 2, hence C is MTP2 implying positive quadrant dependence.

Since a completely monotonic inverse generator implies a totally positive density c of the copula C, a random vector U distributed via C has the property CI.

For the multivariate case $n \geq 2$ we define an Archimedean n-copula

$$C(u_1, \ldots, u_n) = \varphi^{[-1]}(\varphi(u_1) + \cdots + \varphi(u_n)) .$$

This yields a copula if and only if $\varphi^{\leftarrow}$ is completely monotone on $[0, \infty)$.

Even though the class of copulas is closed under taking the survival copula and under convex linear combinations, we cannot transfer this nice properties to the class of Archimedean copulas.

We recall the class of Frank copulas that first appeared in Frank [51] in an algebraic context. In the bivariate case, Frank [51] shows that the Frank copula is the only Archimedean copula that has an Archimedean survival copula. Furthermore, the Frank copula satisfies $\widetilde{C} = C$, i.e. it is the only radially symmetric one (Frank [51], Theorem 4.2, p. 218).

This enables us to show that neither the class of Archimedean copulas nor the Frank subfamily is closed under linear convex combinations. Further we can extend Frank's result to n dimensions.

Lemma 4.1.2. *The survival copula $\widetilde{C}$ of an Archimedean n-copula C is Archimedean if and only if C is the Frank copula and hence $\widetilde{C} = C$ (i.e. C is rotationally symmetric).*
The convex linear combination of two Archimedean copulas needs not to be Archimedean.

Proof. For the first statement we use the associativity of Archimedean copulas and carry out complete induction over the dimension n. For $n = 2$, Theorem 4.2 in Frank [51] proves the result. Now assume the statement to be true up to $n > 2$.

$$\begin{aligned} \widetilde{C_{n+1}}\left(u_1, \ldots, u_{n+1}\right) &\equiv C_{n+1}\left(u_1, \ldots, u_{n+1}\right) \\ \widetilde{C_2}\left(\widetilde{C_n}\left(u_1, \ldots, u_n\right), u_{n+1}\right) &\equiv C_2\left(C_n\left(u_1, \ldots, u_n\right), u_{n+1}\right) \end{aligned} \tag{4.1.2}$$

where C_i denotes the i-copula corresponding to C. The second relation is equivalent to the first by associativity and symmetry of Archimedean copulas (see Nelsen [92] p. 121). Equation 4.1.2 has to be fulfilled for all $u \in [0, 1]^{n+1}$ especially for all $u \in [0, 1]^n \times 1$. So from the second line we can conclude $\widetilde{C_n} \equiv C_n$.

By the induction hypothesis we have $\widetilde{C_n}$ to be an n-dimensional Frank copula. Furthermore 4.1.2 has to be fulfilled for all $u \in \underline{1}_n \times [0,1]$ so we can conclude $\widetilde{C_2} \equiv C_2$ and by Frank [51], Theorem 4.2 that $\widetilde{C_2}$ has to be the 2-dimensional Frank copula. So we have finally $\widetilde{C_{n+1}}$ to be an $n+1$-dimensional Frank copula.

For the second statement let C_{θ_1} and C_{θ_2} be copulas from the Frank family with $\theta_1 \neq \theta_2$. For $0 < \lambda < 1$ let $C^*(u,v) = \lambda C_{\theta_1}(u,v) + (1-\lambda)C_{\theta_2}(u,v)$. As members of the Frank family, C_{θ_1} and C_{θ_2} are radially symmetric so we know that $C_{\theta_j}(u,v) = \widetilde{C_{\theta_j}}(u,v) = u+v-1+C_{\theta_j}(1-u,1-v)$ for $j=1,2$. We use this relation to compute the survival copula

$$\begin{aligned}
\widetilde{C^*}(u,v) &= u+v-1+\lambda C_{\theta_1}(1-u,1-v)+(1-\lambda)C_{\theta_2}(1-u,1-v) \\
&= \lambda(u+v-1+C_{\theta_1}(1-u,1-v)) \\
&\quad +(1-\lambda)(u+v-1+C_{\theta_2}(1-u,1-v)) \\
&= \lambda\widetilde{C_{\theta_1}}(u,v)+(1-\lambda)\widetilde{C_{\theta_2}}(u,v) \\
&= C^*(u,v).
\end{aligned}$$

This implies that C^* must be a member of the Frank family, or else can not be Archimedean according to Frank [51].

To construct a contradiction we fix some parameter vector $(\theta_1, \theta_2, \lambda)$ and compute

$$C^*_{(\theta_1,\theta_2,\lambda)}(u,v) = \tag{4.1.3}$$
$$-\tfrac{\lambda}{\theta_1}\ln\left[1+\tfrac{(e^{-\theta_1 u}-1)\,(e^{-\theta_1 v}-1)}{e^{-\theta_1}+1}\right] - \tfrac{1-\lambda}{\theta_2}\ln\left[1+\tfrac{(e^{-\theta_2 u}-1)\,(e^{-\theta_2 v}-1)}{e^{-\theta_2}+1}\right].$$

Assume that C^* falls into the Frank family. Then in the light of Equation (4.1.3) there exists θ such that

$$C^*_{(\theta_1,\theta_2,\lambda)}(u,v) = -\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta u}-1)\,(e^{-\theta v}-1)}{(e^{-\theta}+1)}\right)$$

for all $(u,v) \in [0,1]^2$. Now take $\lambda = 0.5, \theta_1 = 1$ and $\theta_2 = 10$. For this choice of parameters we compute $\theta(u,v)$. This yields $\theta(0.1,0.1) \approx 4.74$ but $\theta(0.5,0.5) \approx 3.89$ so since θ depends on (u,v) the copula C^* cannot belong to the Frank family which contradicts the statement above. □

We conclude the section by stressing that associativity and the property $C(u,u) < u$ characterize Archimedean copulas (Nelsen [92], p. 93). Many arguments concerning Archimedean copulas are in fact based on associativity.

4.2 A Transformation Rule for Generators

We briefly summarize what we have learned about the theoretical properties of Archimedean copulas: They are completely determined by a one dimensional

generator, and they show flexibility in modelling the dependence structure. For bivariate cases, asymmetry in the tails is possible. Listing the drawbacks, the structural permutational symmetry of an Archimedean copula must be reflected in the data which is certainly the case in most financial data sets. When constructing new copulas from known ones, one must take into account that the subclass of Archimedean copulas is not closed under linear convex combinations (see 4.1.2). But the existence of a generator gives rise to a different approach that does not rely on this property. Here we introduce a generator based transformation rule for Archimedean copulas that enables us to construct new copulas from Archimedean ones.

Proposition 4.2.1. *Let φ be a generator and $g : [0, 1] \to [0, 1]$ a strictly increasing, concave function with $g(1) = 1$, then $\varphi \circ g$ is a generator.*

Let $f : [0, \infty] \to [0, \infty]$ be a strictly increasing, convex function with $f(0) = 0$, then $f \circ \varphi$ is a generator.

Proof. The composition of two strictly monotone functions—one increasing, the other decreasing—is strictly monotone decreasing. Furthermore, $\varphi \circ g$ and $f \circ \varphi$ are convex functions, since for all $\lambda \in (0, 1)$ and $x, y \in [0, 1]$ we compute

$$\begin{aligned} g(\lambda x + (1-\lambda) y) &\geq \lambda g(x) + (1-\lambda) g(y) \\ \Rightarrow \varphi \circ g(\lambda x + (1-\lambda) y) &\leq \varphi(\lambda g(x) + (1-\lambda) g(y)) \\ &\leq \lambda \varphi \circ g(x) + (1-\lambda) \varphi \circ g(y), \end{aligned}$$

and

$$\begin{aligned} \varphi(\lambda x + (1-\lambda) y) &\leq \lambda \varphi(x) + (1-\lambda) \varphi(y) \\ \Rightarrow f \circ \varphi(\lambda x + (1-\lambda) y) &\leq f(\lambda \varphi(x) + (1-\lambda) \varphi(y)) \\ &\leq \lambda f \circ \varphi(x) + (1-\lambda) f \circ \varphi(y). \end{aligned}$$

Together with $\varphi \circ g(1) = f \circ \varphi(1) = 0$, the proposition follows. □

Remark 7. *The proposition generalizes the transformation rule for the α and β family in Nelsen [92], where $g(t) = t^\nu$, $\nu \in (0, 1]$ and $f(\varphi) = \varphi^\delta$, $\delta \in [1, \infty)$. Since a sum of generators is a generator itself, and products of a generator with a strictly positive scalar lead to a generator of the same copula, every polynomial of generators*

$$\Phi(\varphi) = \sum_{i=1}^{n} a_i \varphi^i,$$

with $a_i \geq 0$ and at least one $a_i > 0$, is a generator.

It is easy to construct transformation functions g and f when using piecewise strictly increasing, concave and convex functions that analytically depend on a parameter. The parameter is then used to fulfill the conditions, i.e. transformation of support and range and normalization at 1. Examples are given by

(a) $g_1(t) = \frac{\sin(at)}{\sin(a)}, \quad a \in \left(0,\, \frac{\pi}{2}\right]$

(b) $g_2(t) = \frac{\ln(at+1)}{\ln(a+1)}, \quad a \in (0,\, \infty)$

(c) $g_3(t) = \frac{\lambda^{-\theta\cdot t}-1}{\lambda^{-\theta}-1}, \quad \lambda \in [1,\, \infty),\ \theta \in (-\infty,\, \infty) \setminus \{0\}$

(d) $f_4\,(\varphi) = a^{\varphi} - 1, \quad a \in (1,\, \infty)$

(e) $f_5\,(\varphi) = a^{-\varphi} - 1, \quad a \in (0,\, 1)$

(f) $f_6(\varphi) = \varphi^{\delta}, \quad \delta \in [1, \infty)$

Remark 8. *For the bivariate case we conclude that $f_6 \circ \Pi$ yields the generator of the Gumbel copula, whereas $\Pi \circ g_3$ is the generator of the Frank copula.*

The transformation rule $\varphi \circ g$ is closely related to the transformation rule γ introduced by Durrleman et al. [35]. In their case the tail dependence is preserved under such transformation. (Durrleman et al. [35], Theorem 12.) In the following we will show that this is not the case if we change the generator φ to $f \circ \varphi$.

4.2.1 A family covering the Gumbel and Frank families

Throughout this thesis we use the generator of the Frank copula

$$\varphi_{Frank}(t) = -\ln \frac{e^{-\theta\cdot t} - 1}{e^{-\theta} - 1}, \quad \theta \in (-\infty,\, \infty) \setminus \{0\} \tag{4.2.1}$$

and transform it by

$$\varphi_{\omega} = \left(\varphi_{Frank}\right)^{\delta}, \quad \omega = (\theta, \delta) \in (-\infty,\, \infty) \setminus \{0\} \times [1,\, \infty). \tag{4.2.2}$$

Remark 9. *Since the generator of the Frank copula is strict and twice differentiable, φ_{ω} is strict, i.e. $\varphi_{\omega}^{\leftarrow} = \varphi_{\omega}^{-1}$, and twice differentiable.*

If we set $\delta = 1$, we have the generator of the Frank copula and hence all the properties of a Frank copula are fulfilled.

The resulting Archimedean copula given by (2.3.4) and (4.2.2) is

$$C_{\omega}(u, v) = \varphi_{\omega}^{-1}\left(\varphi_{\omega}(u) + \varphi_{\omega}(v)\right) =$$
$$-\tfrac{1}{\theta} \ln \left[1 + \left(e^{-\theta} - 1\right) \exp \left[-\left(\left(-\ln\left[\tfrac{e^{-\theta\cdot u}-1}{e^{-\theta}-1}\right]\right)^{\delta} + \left(-\ln\left[\tfrac{e^{-\theta\cdot v}-1}{e^{-\theta}-1}\right]\right)^{\delta}\right)^{\frac{1}{\delta}}\right]\right] \tag{4.2.3}$$

with corresponding density

$$c_\omega(u,v) = -\frac{\varphi''_\omega(C_\omega(u,v))\varphi'_\omega(u)\varphi'_\omega(v)}{\left(\varphi'_\omega(C_\omega(u,v))\right)^3}.$$

The following result shows the relation between the transformed copula and the Gumbel copula.

Lemma 4.2.2. *Let C_ω be the transformed copula from (4.2.3) with parameters θ and δ, then for a fixed δ we have the following limit behavior*

$$\lim_{\theta\to 0} C_\omega = C_{\delta_{Gumbel}}$$

Proof. We prove the Lemma by showing that the generators of the two copulas coincide. This writes by using standard calculus

$$\begin{aligned}
\lim_{\theta\to 0}\left(-\ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)^\delta &= \left[-\ln\left(\lim_{\theta\to 0}\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)\right]^\delta \\
&= \left[-\ln\left(\lim_{\theta\to 0}\frac{-te^{-\theta t}}{e^{-\theta}-1}\right)\right]^\delta \\
&= [-\ln(t)]^\delta = \varphi_{\delta_{\text{Gumbel}}}
\end{aligned}$$

□

Remark 10. *In terms of generators, Lemma 4.2.2 yields if $\varphi_\theta \to -\ln t$, then ${\varphi_\theta}^\delta \to \varphi_{\delta Gumbel}$.*

We state that with the result of Lemma 4.2.2 the transformed copula C_ω includes two widely used copulas—the Frank and the Gumbel—as special cases and gives a common analytical form for them.

Lemma 4.2.3. *Let φ_1, φ_2 generators of Archimedean copulas C_1, and C_2 respectively. Let further $C_{1,\delta}$, $C_{2,\delta}$ be the Archimedean copulas generated by the transformed generators φ_1^δ, φ_2^δ with $\delta \geq 1$.*
a) If $\frac{\varphi_1}{\varphi_2}$ is nondecreasing on $(0,1)$, then $C_1 \prec C_2$ (Corollary 4.4.5 in Nelsen [92]) and further $C_{1,\delta} \prec C_{2,\delta}$.
b) If φ_1 and φ_2 are continuously differentiable on $(0,1)$, and if $\frac{\varphi_1'}{\varphi_2'}$ is nondecreasing on $(0,1)$, then $C_1 \prec C_2$ (Corollary 4.4.6 in Nelsen [92]) and further $C_{1,\delta} \prec C_{2,\delta}$.

Proof. If $\frac{\varphi_1}{\varphi_2}$ is nondecreasing on $(0,1)$, so is $\frac{\varphi_1^\delta}{\varphi_2^\delta} = \left(\frac{\varphi_1}{\varphi_2}\right)^\delta$. Part a) of the lemma follows now by Corollary 4.4.5 in Nelsen [92], p. 110. Part b) follows by Corollary 4.4.6 in Nelsen [92], p. 111 and the fact that $\frac{\left(\varphi_1^\delta\right)'}{\left(\varphi_2^\delta\right)'} = \left(\frac{\varphi_1'}{\varphi_2'}\right)^{\delta-1}$ is nondecreasing on $(0,1)$, since $\frac{\varphi_1}{\varphi_2}$ is nondecreasing on $(0,1)$ and $\delta \geq 1$. □

The following proposition summarizes some dependence properties of C_ω.

Proposition 4.2.4. *(a) C_ω has no lower tail dependence but is upper tail dependent with tail dependence parameter $\lambda_U = 2 - 2^{\frac{1}{\delta}}$.*

(b) Let $\theta > 0$, then φ_ω^{-1} is completely monotonic on $[0, \infty)$, and hence the density c_ω is totally positive of order 2.

(c) C_ω is positively ordered with respect to the concordance order and the parameters θ, δ, i.e. $\theta_1 \leq \theta_2 \Rightarrow C_{\theta_1}(u,v) \leq C_{\theta_2}(u,v)$ for all $(u,v) \in [0,\, 1]^2$.

Proof. (a) For the tail dependence we have to calculate the limits

$$\begin{aligned} \lambda_L &= \lim_{u\to 0} \tfrac{C_\omega(u,u)}{u} \\ \lambda_U &= \lim_{u\to 1} \tfrac{1-2u+C_\omega(u,u)}{1-u}. \end{aligned}$$

Using L' Hospital's rule we get $\lambda_L = 0$ and $\lambda_U = 2 - 2^{\frac{1}{\delta}}$.

(b) To show the complete monotonicity of φ_ω^{-1} we use Lemma 4.6.4 in Nelsen [92], p. 123, where it is shown that for a strict generator φ with a completely monotonic inverse on $[0, \infty)$, the transformation φ^δ, $\delta \geq 1$, leads to a generator with a completely monotonic inverse on $[0, \infty)$. For $\theta > 0$, the Frank generator φ_{Frank} is strict and its inverse is completely monotonic on $[0, \infty)$.

(c) The concordance ordering property follows directly from Lemma 4.2.3 and Theorem 4.5.2 in Nelsen [92], p. 114.

□

In the next section we will state a more general result.

Remark 11. *(a) Further transformations of the generator φ_ω yield new dependence structures. As an example we take the generator $\varphi_\omega(t^\nu)$, $\nu \in (0,\, 1]$ that allows for negative dependence in* one *tail (see Figure 4.1).*

The Gumbel copula which has the generator $\varphi(t) = (-\ln(t))^\delta$, $\delta \geq 1$, leads to the same upper tail dependence as C_ω. Nevertheless, C_ω is not an extreme value copula, i.e. $C(u^\kappa, v^\kappa) = C^\kappa(u,v)$ is not true for all $\kappa > 0$.

(b) Since the inverse generator φ_ω^{-1} is completely monotonic, the copula density is TP2, hence a C_ω distributed random vector is conditionally increasing.

We will now analyze some limit cases, and show that all boundary cases—totally positive and negative dependence as well as independence—are covered by C_ω. Recall that the Fréchet-Hoeffding bounds $W(u,v) = \max\{u+v-1, 0\}$

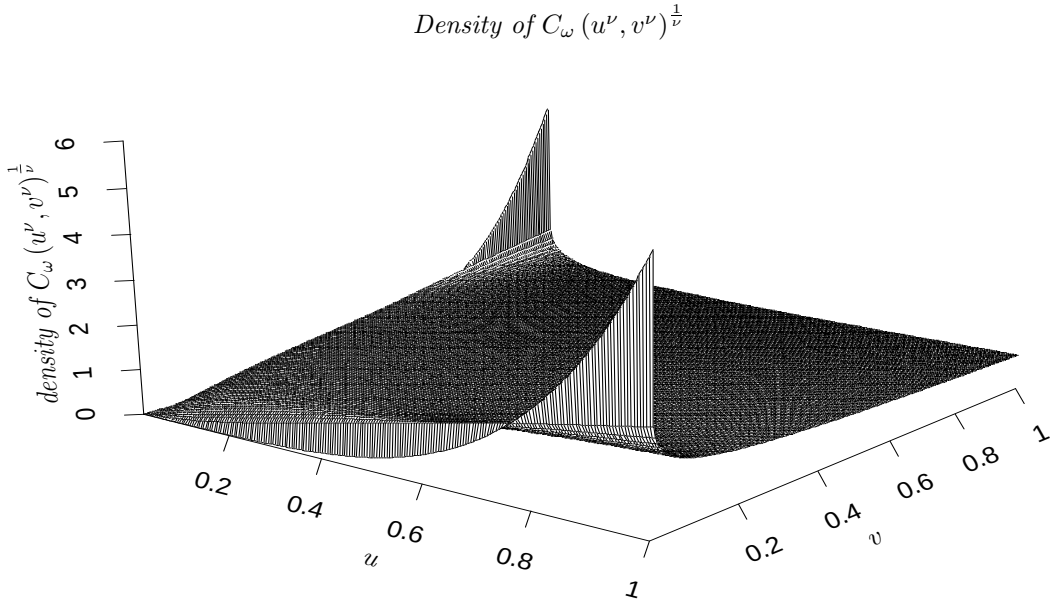

Figure 4.1: Density of $C_\omega\left(u^\nu, v^\nu\right)^{\frac{1}{\nu}}$ with $\theta = -5, \delta = 1, \nu = 0.2$.

and $M(u,v) = \min\{u,v\}$, representing totally negative and totally positive dependence respectively, are boundaries in the sense that for all copulas C it holds $W(u,v) \leq C(u,v) \leq M(u,v)$.

Proposition 4.2.5. *Let C_ω be the copula defined by Equation (4.2.3), with $\omega = (\theta, \delta)$, then*

$$\begin{aligned}
\lim_{\delta\to\infty} C_\omega &= M, \\
\lim_{\theta\to\infty} C_\omega &= M, \\
\lim_{\delta=1,\theta\to 0} C_\omega &= u \cdot v = \Pi, \\
\lim_{\delta=1,\theta\to -\infty} C_\omega &= W.
\end{aligned}$$

Proof. The first limit follows by Theorem 4.5.4 in Nelsen [92], p. 115 and the last two are just limit cases of the Frank copula.

To analyze $\lim_{\theta\to\infty} C_\omega$ we substitute $t = e^{-\theta} \Rightarrow \theta = -\ln(t)$ and get

$$\begin{aligned}
\lim_{\theta\to\infty} C_\omega(u,v) &= \lim_{t\to 0} \max\left\{ \frac{\ln\left[1-\exp\left[-\left((-\ln[1-t^u])^\delta + (-\ln[1-t^v])^\delta\right)^{\frac{1}{\delta}}\right]\right]}{\ln(t)}, 0 \right\} \\
&= \max\left\{ \lim_{t\to 0} \frac{\ln\left[1-\exp\left[-\left(-\ln\left[1-t^{\min\{u,v\}}\right]\right)\right]\right]}{\ln(t)}, 0 \right\} \\
&= \max\left\{ \lim_{t\to 0} \frac{\ln\left[t^{\min\{u,v\}}\right]}{\ln(t)}, 0 \right\} \\
&= \min\{u,v\} = M(u,v).
\end{aligned}$$

□

To catch asymmetry in the tails we take a convex linear combination of the proposed copula C_ω and a survival copula $\widetilde{C}_{\omega_s}$ out of the same family. Here $\omega_s = (\theta_s, \delta_s)$ is the parameter vector of the survival copula. Hence we define for $\alpha \in [0,\, 1]$ the copula

$$\begin{aligned} C_\Omega(u,v) &:= \alpha \cdot \widetilde{C}_{\omega_s}(u,v) + (1-\alpha)\cdot C_\omega(u,v) \\ &= \alpha \cdot (u+v-1+C_{\omega_s}(1-u,1-v)) + (1-\alpha)\cdot C_\omega(u,v). \end{aligned} \tag{4.2.4}$$

Proposition 4.2.6. *Let C_1, C_2 be 2-copulas with tail dependence parameters $\lambda_{L_1}, \lambda_{U_1}$ and $\lambda_{L_2}, \lambda_{U_2}$, respectively. Denote $\widetilde{\lambda}_L$, $\widetilde{\lambda}_U$ the lower and upper tail dependence of $\widetilde{C}$. Further let $C^\star = \alpha C_1 + (1-\alpha)C_2$, with $\alpha \in [0,1]$, be the copula resulting from the linear convex combination of C_1 and C_2. Then it holds*

(a) $\widetilde{\lambda}_L = \lambda_U$, $\widetilde{\lambda}_U = \lambda_L$

(b) $\lambda_L^\star = \alpha\lambda_{L_1} + (1-\alpha)\lambda_{L_2}$, *and the same holds for* $\lambda_U^\star$.

Proof. The first statement is proved through

$$\begin{aligned} \widetilde{\lambda_L} &= \lim_{u\to 0} \tfrac{\widetilde{C}(u,u)}{u} = \lim_{u'=(1-u)\to 1} \tfrac{1-2u'+C(u',u')}{1-u'} &= \lambda_U \\ \widetilde{\lambda_U} &= \lim_{u\to 1} \tfrac{1-2u+\widetilde{C}(u,u)}{1-u} = \lim_{u'=(1-u)\to 0} \tfrac{\widetilde{C}(u',u')}{u'} &= \lambda_L \end{aligned}$$

The second statement follows directly from a), since all tail dependence parameters are bounded, hence the corresponding limits exist and we can use the additivity of limits. □

Corollary 4.2.7. *C_Ω is lower tail dependent with $\lambda_L = \alpha\left(2-2^{\frac{1}{\delta_s}}\right)$ and upper tail dependent with $\lambda_U = (1-\alpha)\left(2-2^{\frac{1}{\delta}}\right)$. Therefore the tail dependence of C_Ω is restricted by $\lambda_L + \lambda_U \leq 1$.*

Proof. The corollary follows directly by Propositions 4.2.4 and 4.2.6 by setting $C_2 = \widetilde{C_1}$. □

Corollary 4.2.8. *Let C_Ω be the copula defined by Equation (4.2.4), then*

$$\begin{aligned} \lim_{\delta,\delta_s\to\infty} C_\Omega &= M, \\ \lim_{\theta,\theta_s\to\infty} C_\Omega &= M, \\ \lim_{\delta=\delta_s=1,\theta,\theta_s\to 0} C_\Omega &= u\cdot v = \Pi, \\ \lim_{\delta=\delta_s=1,\theta,\theta_s\to -\infty} C_\Omega &= W. \end{aligned}$$

Proof. We only want to sketch the ideas behind the calculations exemplarily for the first limit.

$$\begin{aligned}\lim_{\delta,\delta_s\to\infty} C_\Omega(u,v) &= \alpha \lim_{\delta_s\to\infty} \widetilde{C}_{\omega_s}(u,v) + (1-\alpha) \lim_{\delta\to\infty} C_\omega(u,v) \\ &= \alpha(u+v-1+\min\{1-u,1-v\}) + (1-\alpha)\min\{u,v\} \\ &= \min\{u,v\} = M(u,v).\end{aligned}$$

□

We now turn to the extension of Archimedean copulas to multidimensional models.

4.3 The multidimensional case

A multivariate Archimedean copula can be defined using the following proposition (see Theorem 4.6.2. in Nelsen [92], p. 122.).

Proposition 4.3.1. *Let φ be a continuous, strictly decreasing function from $[0,\,1]$ to $[0,\,\infty]$ such that $\varphi(0)=\infty$ and $\varphi(1)=0$, and let φ^{-1} denote the inverse of φ. The function $C^{[d]}$ from $[0,\,1]^d$ to $[0,\,1]$ given by*

$$C^{[d]}(\mathbf{u}) = \varphi^{-1}\left(\varphi\left(u_1\right) + \ldots + \varphi\left(u_d\right)\right) \tag{4.3.1}$$

is a d-copula for all $d \geq 2$ if and only if φ^{-1} is completely monotonic on $[0,\,\infty)$.

In Nelsen or other standard literature it is stated that the Gumbel and Clayton copula families are directly extendable to their multivariate version given by Equation 4.3.1, whereas the Frank family only has an extension if the parameter is restricted to $\theta \in (0,\infty]$ (see Example 4.22 in Nelsen [92]). One of the main disadvantages of Archimedean copulas is expressed by the following corollary (see e.g. Nelsen [92] Corollary 4.6.3).

Corollary 4.3.2. *If the inverse φ^{-1} of a strict generator φ of an Archimedean copula C is completely monotonic, then $C \succ \Pi$.*

In other words the corollary states that every multidimensional Archimedean copula can *only* describe positive dependence structures.
Since in most financial applications the dependence structure is either positive or negative, a possible solution to deal with the restriction given by Corollary 4.3.2 is sketched out in the following example. Suppose a portfolio of two stocks and a government bond. In a normal market it is assumed that the two univariate stochastic random variables driving the stock returns S_1, S_2 are positively dependent and both negatively dependent to the stochastic random variable driving the

bond returns B. So we end up with a 3-dimensional random vector (S_1, S_2, B). We can now model the dependence structure of $(S_1, S_2, -B)$ (that is by our assumptions a positive one) with a 3-dimensional Archimedean copula and run a Monte Carlo Simulation to achieve some realizations $(s_1, s_2, -b)$ that can easily be backtransformed to observations (s_1, s_2, b) of our portfolio innovations. Hence we are able to use Archimedean copulas and Monte Carlo Simulation methods to model multidimensional portfolios with negative and positive bivariate marginals. To check how the transformation rule 4.2.1 influences the multidimensional extensibility, and hence the completely monotonicity of the generator, we state the following proposition, also stated on p. 123 in Nelsen [92].

Proposition 4.3.1. *Let Ψ be an m-monotonic function, f a function with range in the domain of Ψ and g a function with domain in the range of Ψ.*

(a) If $(-1)^{j+1}\frac{d^j}{dt^j}f(t) \geq 0$ for all t and $j = 0, 1, \ldots, m$, then $\Psi \circ f$ is m-monotonic.

(b) If $\frac{d^j}{dt^j}g(t) \geq 0$ for all t and $j = 0, 1, \ldots, m$, then $g \circ \Psi$ is m-monotonic.

The first result were already given by Feller [46] and the second one by Widder [111]. Since the proof of the first result is a bit sloppy in Feller [46], we state here our own version of a proof, utilizing the following lemma.

Lemma 4.3.2. *Let n, k be integers with $k \geq n$. If a sequent of nonnegative integers $(a_i)_{i\in\mathbb{N}}$ fulfills $\sum_{i=1}^{\infty} a_i = n$ and $\sum_{i=1}^{\infty} i \cdot a_i = k$ then $a_i = 0$ for all $i > 1+k-n$.*

Proof. The proof is trivial for $n = 0$, so in the following we assume $n \geq 1$ and proof the lemma by contradiction.
Suppose there exists at least one $a_j \geq 1$, $j \geq 2 + k - n$. Now we can use the following minorizing

$$\begin{aligned} k &= \sum_{i\neq j,\, i=1}^{\infty} i \cdot a_i + j \cdot a_j \\ &\geq \sum_{i\neq j,\, i=1}^{\infty} i \cdot a_i + a_j + (j-1) \\ &\geq n + (1 + k - n) \end{aligned}$$

what lead to the contradiction $k \geq k + 1$. In the first inequality we used $a_j \geq 1$ and in the second $\sum_{i\neq j,\, i=1}^{\infty} i \cdot a_i + a_j \geq \sum_{i=1}^{\infty} a_i$ and $j \geq 2 + k - n$. □

Now we can proof Proposition 4.3.1.

Proof. (a) For m-monotonicity of $\Psi \circ f$ we have to show that $(-1)^k \frac{d^k}{dt^k}(\Psi \circ f) \geq 0$ for all $k = 1, \ldots, m$. For $k = 0$ we have $\Psi \circ f \geq 0$ due to the m-monotonicity of Ψ. Applying the multidimensional chain rule (see e.g. Bödewadt [12]) we achieve

$$\frac{d^k}{dt^k}(\Psi \circ f) = \sum_{n=1}^{k} \frac{d^n}{dt^n}\Psi \cdot \left[\sum_{(a)\neq 0} k! \prod_{i=1}^{1+k-n} \frac{1}{a_i!} \left(\frac{\frac{d^i}{dt^i} f}{i!} \right)^{a_i} \right],$$

where the sum runs over all nontrivial sequences fulfilling the conditions of Lemma 4.3.2. According to the condition on f given in Proposition 4.3.1 (a) and with use of Lemma 4.3.2 we derive

$$sgn\left(\sum_{(a)\neq 0} k! \prod_{i=1}^{1+k-n} \frac{1}{a_i!} \left(\frac{\frac{d^i}{dt^i} f}{i!} \right)^{a_i} \right) = (-1)^{\sum_{i=1}^{1+k-n}(i+1)a_i} = (-1)^{k+n}$$

Together with the m-monotonicity of Ψ we have

$$\frac{d^k}{dt^k}(\Psi \circ f) = \sum_{n=1}^{k} (-1)^n (-1)^{k+n} \cdot r_n,$$

where the r_n are some nonnegative real numbers, and hence

$$(-1)^k \frac{d^k}{dt^k}(\Psi \circ f) = \sum_{n=1}^{k} (-1)^k (-1)^n (-1)^{k+n} \cdot r_n = \sum_{n=1}^{k} (-1)^{2(k+n)} \cdot r_n \geq 0, \ \forall k, n$$

what finishes part (a).
(b) We have to show that $(-1)^k \frac{d^k}{dt^k}(g \circ \Psi) \geq 0$ for all $k = 1, \ldots, m$. For $k = 0$ we have $g \circ \Psi \geq 0$ due to $g \geq 0$. Analogously to part (a) we use the multidimensional chain rule and the m-monotonicity of Ψ to get

$$sgn\left(\sum_{(a)\neq 0} k! \prod_{i=1}^{1+k-n} \frac{1}{a_i!} \left(\frac{\frac{d^i}{dt^i} \Psi}{i!} \right)^{a_i} \right) = (1)^n (-1)^k$$

and hence

$$(-1)^k \frac{d^k}{dt^k}(g \circ \Psi) = \sum_{n=1}^{k} (-1)^k (-1)^k \cdot r_n = \sum_{n=1}^{k} (-1)^{2k} \cdot r_n \geq 0, \ \forall k, n$$

where the r_n are again some nonnegative real numbers. □

We are now able to come back to our transformation rules and provide in the following corollary further conditions on the transformation to guarantee multidimensional extensibility of the resulting Archimedean copulas.

Corollary 4.3.3. *Let φ be a generator of an Archimedean d-copula with $d > 2$. Let further g, f be functions as given in Proposition 4.2.1. Then the following holds:*
a) $\varphi \circ g$ ia a generator of an Archimedean d-copula iff $g^{-1} \circ \varphi^{-1}$ is completely monotonic, and the latter is true if $\frac{d^j}{dt^j} g \geq 0,\ \forall j = 0, 1, \ldots$
b) $f \circ \varphi$ is a generator of an Archimedean d-copula iff $\varphi^{-1} \circ f^{-1}$ is completely monotonic, and the latter is true if $(-1)^{j+1} \frac{d^j}{dt^j} f \geq 0,\ \forall j = 0, 1, \ldots$

Proof. The corollary follows directly by Theorem 4.6.2 in Nelsen [92] and Proposition 4.3.1. □

Corollary 4.3.3 generalizes Lemma 4.6.4 in Nelsen [92]. Remark that for the inverse of $f(t) = t^\delta$, $\delta \geq 1$ the condition of Corollary 4.3.3 is fulfilled since $\frac{d}{dt} f^{-1}(t) = \frac{1}{\delta} t^{\frac{1}{\delta}-1}$ and $\frac{1}{\delta} - 1 \leq 0$. Hence especially the generator $(\varphi_\theta)^\delta$ of C_ω defined by Equation (4.2.3) has a multidimensional extension according to (4.3.1) if and only if $\theta > 0$.
Despite the fact that the d-dimensional form (4.3.1) is restricted to positive dependence, it further restricts all k-marginals to have the same copula, and hence the same dependence structure. That means for example that in a portfolio every pair of assets has to have the same dependence. If dependence is e.g. measured by correlation, then the correlation matrix of the portfolio is assumed to have the same entries outside the diagonal. This does in no way appeal a very realistic assumption, and hence we conclude that (4.3.1) is not an appropriate choice to model higher dimensional financial problems. In the next subsection we state alternative multidimensional extension rules. For the sake of more flexibility we have to leave the Archimedean world and deal with highly complicated derived boundary conditions on the parameters.

4.3.1 More multidimensional extension rules

Equation (4.3.1) gives the general form for a multivariate Archimedean copula depending on *one* generator. More general types of dependence can be achieved when each bivariate margin is supposed to have its own generator, each of the form (2.3.4). Nesting them together, this leads for the trivariate case $d = 3$

$$\varphi_2^{-1}\left(\varphi_2 \circ \varphi_1^{-1}(\varphi_1(u_1) + \varphi_1(u_2)) + \varphi_2(u_3)\right) , \tag{4.3.2}$$

or in terms of copulas

$$C^{[3]}(u_1, u_2, u_3) = C_2(C_1(u_1, u_2), u_3) .$$

For higher dimensions d there are many possible nestings. Since our empirical analysis will deal with dimension $d = 4$, we restrict our discussion further on this

case. In 4 dimensions there are two distinct possibilities to generate a copula from its bivariate marginals, given by the following equations (see also Joe [62] p. 89)

$$\begin{aligned} C^{[4]}(u_1,\ldots,u_4) &= \varphi_3^{-1}\left[\varphi_3(\varphi_2^{-1}\left(\varphi_2\circ\varphi_1^{-1}(\varphi_1(u_1)+\varphi_1(u_2))+\varphi_2(u_3)\right),u_4)\right] \\ &= C_3\left(C_2\left(C_1\left(u_1,u_2\right),u_3\right),u_4\right) \end{aligned} \tag{4.3.3}$$

$$\begin{aligned} C^{[[4]]}(u_1,\ldots,u_4) &= \varphi_3^{-1}\left(\varphi_3(\varphi_1^{-1}[\varphi_1(u_1)+\varphi_1(u_2)])+\varphi_3(\varphi_2^{-1}[\varphi_2(u_3)+\varphi_2(u_4)]\right) \\ &= C_3\left(C_1\left(u_1,u_2\right),C_2\left(u_3,u_4\right)\right) \end{aligned} \tag{4.3.4}$$

We will refer to (4.3.3) by calling it *model a)* and to (4.3.4) by calling it *model b)*. For both models all trivariate marginals have the form (4.3.2) and all bivariate ones the Archimedean structure (2.3.4).
If all generators are the same in the above nesting constructions, we end up with the multidimensional Archimedean form given by (4.3.1) as a special case. Hence the first natural condition for the construction algorithm to generate copulas by nesting, is that all generators have to be strict with completely monotone inverse. For the the further conditions we introduce some notations about Laplace transforms. Let $\mathcal{L}_\infty = \{\psi\colon [0,\infty) \to [0,1] \mid \psi(0) = 1, \lim_{t\to\infty}\psi(t) = 0, (-1)^k\psi^{(k)} \geq 0, (\forall k \in I\!N)\}$ denote the set of Laplace transforms with value 0 at ∞. So Laplace transforms can be characterized as completely monotone functions on $I\!R^+$ with value 1 in 0. Hence the inverses of strict generators for copulas are Laplace transforms in $\mathcal{L}_\infty$ and vice versa. A related set is $\mathcal{L}_\infty^* = \{\varpi\colon [0,\infty) \to [0,\infty) \mid \varpi(0) = 0, \varpi(\lim_{t\to\infty}) = \infty, (-1)^{j-1}\varpi^{(j)} \geq 0, (\forall j \in I\!N)\}$, the class of infinitely differentiable increasing functions of $I\!R^+ \to I\!R^+$ with alternating sign for the derivatives.
The further condition that has to be fulfilled to guarantee the outputs of (4.3.2), (4.3.3), (4.3.4) and other nesting constructions for higher dimensions, being d-copulas, is $\varphi_{i+1}\circ\varphi_i^{-1} \in \mathcal{L}_\infty^*$ for all levels of nesting, i.e. for all $i = 1,\ldots,n-2$. Remark that for $\varphi_i = \varphi_j$ for all $i,j = 1,\ldots,n-1$ we end up again with the multidimensional Archimedean case (4.3.1) and $\varphi_{i+1}\circ\varphi_i^{-1}$ is the identity and hence trivially in $\mathcal{L}_\infty^*$. So we have the conditions:

Condition a):
Model a) (4.3.3) is a copula if and only if $\varphi_3\circ\varphi_2^{-1} \in \mathcal{L}_\infty^*$ and $\varphi_2\circ\varphi_1^{-1} \in \mathcal{L}_\infty^*$
Condition b):
Model b) (4.3.4) is a copula if and only if $\varphi_3\circ\varphi_2^{-1} \in \mathcal{L}_\infty^*$ and $\varphi_3\circ\varphi_1^{-1} \in \mathcal{L}_\infty^*$

Remark that the copulas created via (4.3.2), (4.3.3), (4.3.4) or higher dimensional nestings, are generally not permutational symmetric anymore and hence have left the Archimedean world. There is still some partial exchangeability in the variables, what yields one of the big restrictions of these types of copulas. More precisely, we formulate for the models a) and b) the following lemma.

Lemma 4.3.4. *Let $X = (X_1, X_2, X_3, X_4)$ be a 4-dimensional rv with copula $C^{[4]}$ (model a)), and $C^{[[4]]}$ (model b)), respectively. Then the following holds for copulas of the bivariate marginals under model a)*

$$C_{X_1,X_3} = C_{X_2,X_3} \ \wedge \ C_{X_1,X_4} = C_{X_2,X_4} = C_{X_3,X_4}$$

and under model b)

$$C_{X_1,X_3} = C_{X_1,X_4} = C_{X_2,X_3} = C_{X_2,X_4}.$$

Proof. The copula of (X_i, X_j) is just the projection of the copula of X in the (i, j) plane. The projection is carried out by setting every component of C_X to 1 except for i and j. Recalling that $C(u_1, \ldots, u_k = 1, \ldots, u_d) = C(u_1, \ldots, u_{k-1}, u_{k+1}, \ldots, u_d)$ for all d-copulas C finishes the proof. □

Remark 12. *The equality of the bivariate marginal copulas stated in Lemma 4.3.4 implies the equality of all dependence measures like Kendall's τ and the lower and upper tail dependence coefficient of these bivariate marginals. If we e.g. assume model b), we deduce directly $\lambda_{L,(i,j)} = \lambda_{L,(k,l)}$, $\lambda_{U,(i,j)} = \lambda_{U,(k,l)}$ and $\tau_{(i,j)} = \tau_{(k,l)}$ with $i, k = 1, 2,\ j, l = 3, 4$.*

So again we face the drawback of inflexibility. It improves (4.3.1), but in the sense of the previous example we have now assumed a correlation matrix with areas of same entries (if dependence measured in terms of correlation).

If we now check the parameter boundary conditions for some explicit models, the following well known theorem from the theory of Laplace transforms will help.

Theorem 4.3.5. *Let $\psi \in \mathcal{L}_\infty$ a Laplace transform. Then ψ^α is completely monotone for all $\alpha > 0$ if and only if $-\ln(\psi) \in \mathcal{L}^*_\infty$.*

Lemma 4.3.6. *(a) Let $\varphi_{\delta_1}, \varphi_{\delta_2}$ be generators of the Clayton type (see 2.3.8), then $\varphi_{\delta_2} \circ \varphi_{\delta_1}^{-1} \in \mathcal{L}^*_\infty$ if $\delta_2 < \delta_1$ for $\delta_{1,2} \geq 0$*

(b) Let $\varphi_{\delta_1}, \varphi_{\delta_2}$ be generators of the Gumbel type (see 2.3.6), then $\varphi_{\delta_2} \circ \varphi_{\delta_1}^{-1} \in \mathcal{L}^*_\infty$ if $\delta_2 < \delta_1$ for $\delta_{1,2} \geq 1$

(c) Let $\varphi_{\theta_1}, \varphi_{\theta_2}$ be generators of the Frank type (see 2.3.7), then $\varphi_{\theta_2} \circ \varphi_{\theta_1}^{-1} \in \mathcal{L}^*_\infty$ if $\theta_2 < \theta_1$ for $\theta_{1,2} > 0$

Proof. a) The generator for the Clayton copula is $\varphi_{Cl}(t) = t^{-\delta} - 1$ with $\delta > 0$. The functional inverse is given by $\varphi_{Cl}^{-1}(s) = (1+s)^{\frac{1}{\delta}}$, $\delta \geq 0$. Obviously, $\nu(s) = -\ln(\varphi_{Cl}^{-1}(s)) = \frac{1}{\delta}\ln(1+s)$ and $\nu \in \mathcal{L}_{\infty}^{*}$ for $\delta > 0$. We now assume that $\delta_1 > \delta_2$ and compute $\nu(s) = \varphi_{Cl,2}(\varphi_{Cl,1}^{-1}(s)) = \varphi_{Cl,2}((1+s)^{\frac{1}{\delta_1}}) = (1+s)^{\delta_2/\delta_1}$. Now $\frac{\delta_2}{\delta_1} - 1 < 0$ and hence $\nu \in \mathcal{L}_{\infty}^{*}$.
For the proof of b) and c) we refer to Joe [62], p. 375. □

Together with Conditions a) and b) the lemma leads directly to the boundary conditions on the parameters of several copula families as summarized in the next lemma.

Lemma 4.3.7. *(a) Let $\varphi_{\delta_1}, \varphi_{\delta_2}, \varphi_{\delta_3}$ be generators of the Clayton type (i.e. $\delta_{1,2,3} \geq 0$), then model a) (4.3.3) is a copula if $\delta_3 < \delta_2 < \delta_1$ and model b) (4.3.4) is a copula if $\delta_3 < \delta_1$ and $\delta_3 < \delta_1$.*

(b) Let $\varphi_{\delta_1}, \varphi_{\delta_2}, \varphi_{\delta_3}$ be generators of the Gumbel type (i.e. $\delta_{1,2,3} \geq 1$), then model a) (4.3.3) is a copula if $\delta_3 < \delta_2 < \delta_1$ and model b) (4.3.4) is a copula if $\delta_3 < \delta_1$ and $\delta_3 < \delta_1$.

(c) Let $\varphi_{\theta_1}, \varphi_{\theta_2}, \varphi_{\theta_3}$ be generators of the Frank type, then model a) (4.3.3) is a copula if $\theta_3 < \theta_2 < \theta_1$ and $\theta_{1,2,3} > 0$ and model b) (4.3.4) is a copula if $\theta_3 < \theta_1$ and $\theta_3 < \theta_1$ and $\theta_{1,2,3} > 0$

We now come back to the transformed Frank copula introduced in Subsection 4.2.1. We derive the following sufficient conditions via numerical simulation.

Lemma 4.3.8. *Let $C_{\omega}^{[4]}$ be a function following model a) with transformed Frank generators. If $\delta_3 < \delta_2 < \delta_1$ and $\theta_3 < \theta_2 < \theta_1$ for $\delta_j \geq 1$ and $\theta_j > 0$ for $j = 1, 2, 3$ then $C_{\omega}^{[4]}$ is a copula.*
Let $C_{\omega}^{[[4]]}$ be a function following model b) with transformed Frank generators. If $\delta_3 < \delta_2$ and $\delta_3 < \delta_1$ and $\theta_3 < \theta_2$ and $\theta_3 < \theta_1$ for $\delta_j \geq 1$ and $\theta_j > 0$ for $j = 1, 2, 3$ then $C_{\omega}^{[[4]]}$ is a copula.

We are not able to give a proof of Lemma 4.3.8, what will be subject of further research. The intuition is that we combine the parameter conditions of the two limit cases of C_{ω}, that is of the Gumbel and Frank copula, given by Lemma 4.3.7. For a numerical crosscheck we restricted the parameter space to be finite and sampled ten million parameter vectors according to the conditions of Lemma 4.3.8 with $\delta_{1,2,3} \in [1, 5]$ and $\theta_{1,2,3} \in (0, 10]$. As we will see in Chapter 8 these vectors broadly cover the parameter range of empirical interest. For every of the 10 million parameter vectors we had run a Monte Carlo Simulation over 1 million points in $[0, 1]^4$ and evaluated the supposed densities of $C_{\omega}^{[4]}$ and $C_{\omega}^{[[4]]}$ at these points. There were no negative values detected. So the extensive numerical check

gives no suspicion about the assumption that Lemma 4.3.8 is true. Remark that all other conditions for $C_\omega^{[4]}$ and $C_\omega^{[[4]]}$ being copulas (other than being d-increasing or equivalent having a positive density since we are grounded) are trivially fulfilled since the 2-dimensional C_ω is a copula and the construction algorithm is based on the bivariate form.

Because of the restrictions that we had sketched here for the 4-dimensional case in the Conditions a) and b) and Lemma 4.3.4, the model choice depends crucial on the ordering of the variables. Both the decision concerning an appropriate model and the choice of the variables $u_1, u_2, \ldots, u_d$ are responsible for the optimal fit of the model. In Chapter 8 we check for the 4-dimensional case if the restrictions are relevant for financial applications.

Chapter 5

An asymmetric Extreme Value Copula

This chapter is based on Junker [65].
All well known and widely used copula functions are permutational symmetric, like e.g. the members of the elliptical and Archimedean families. Here we introduce a new family of copulas allowing for asymmetry and upper tail dependence. It is shown that the Gumbel copula arises as a special case of this family.

5.1 An Asymmetric extreme value copula

Pickands [98] had developed an asymmetric extreme value distribution given by

$$F(x,y) = \exp(\mu(x,y)) = \exp\left(-\left(\frac{1}{x}+\frac{1}{y}\right)A\left(\frac{x}{x+y}\right)\right), \tag{5.1.1}$$

where μ is called the exponent measure and $A : [0,1] \to [0,1]$ is called the dependence function. Here we will restrict to the asymmetric logistic model as introduced in Klüppelberg and May [73], i.e.

$$A(t) = \left\{[\theta(1-t)]^{\delta} + [\phi t]^{\delta}\right\}^{\frac{1}{\delta}} + 1 - \theta(1-t) - \phi t, \tag{5.1.2}$$

where $0 \le \phi \le 1, 0 \le \theta \le 1, \delta \ge 1$.
We now can derive the copula of (5.1.1) with A given by (5.1.2) and call it the asymmetric logistic copula, defined by

$$
\begin{aligned}
C_{al}(u,v) &= \exp\left(-\mu\left(F_{\cdot}^{\leftarrow}(u), F_{\cdot}^{\leftarrow}(v)\right)\right) \\
&= \exp\left(\frac{\theta-1}{F_{\cdot}^{\leftarrow}(u)} + \frac{\phi-1}{F_{\cdot}^{\leftarrow}(u)} - \frac{\left\{\left[\theta F_{\cdot}^{\leftarrow}(u)\right]^{\delta} + \left[\phi F_{\cdot}^{\leftarrow}(v)\right]^{\delta}\right\}^{\frac{1}{\delta}}}{F_{\cdot}^{\leftarrow}(u)\, F_{\cdot}^{\leftarrow}(v)}\right) \\
&= u^{1-\theta} v^{1-\phi} \exp\left(-\left[\left(\frac{\theta}{F_{\cdot}^{\leftarrow}(u)}\right)^{\delta} + \left(\frac{\phi}{F_{\cdot}^{\leftarrow}(v)}\right)^{\delta}\right]^{\frac{1}{\delta}}\right),
\end{aligned}
$$

where $F_{\cdot} : \mathbb{R} \to [0,1]$ denotes an arbitrary univariate strictly monotone continuous distribution function. Assuming $F_{\cdot}$ being Fréchet distributed, i.e. $F_{\cdot} = F_F(t) = \exp\left(-\frac{1}{t}\right)$ with inverse $F_F^{\leftarrow} = F_F^{-1}(t) = -\frac{1}{\ln(t)}$, we get

$$
C_{alF}(u,v) = u^{1-\theta} v^{1-\phi} \exp\left(-\left\{(-\theta \ln(u))^{\delta} + (-\phi \ln(v))^{\delta}\right\}^{\frac{1}{\delta}}\right). \tag{5.1.3}
$$

For $\delta = 1$ we calculate $C_{alF}(u,v) = \Pi(u,v) = uv$, i.e., the random variables X and Y are independent.
In Figure 5.1 we have plotted 5000 realizations of C_{alF} with $\theta = 0.8, \phi = 0.2, \delta = 4$ (left picture) and $\theta = 0.2, \phi = 0.8, \delta = 4$ (right picture) respectively. As seen we have a "convex" asymmetry if $\theta > \phi$ and a "concave" one if $\theta < \phi$.

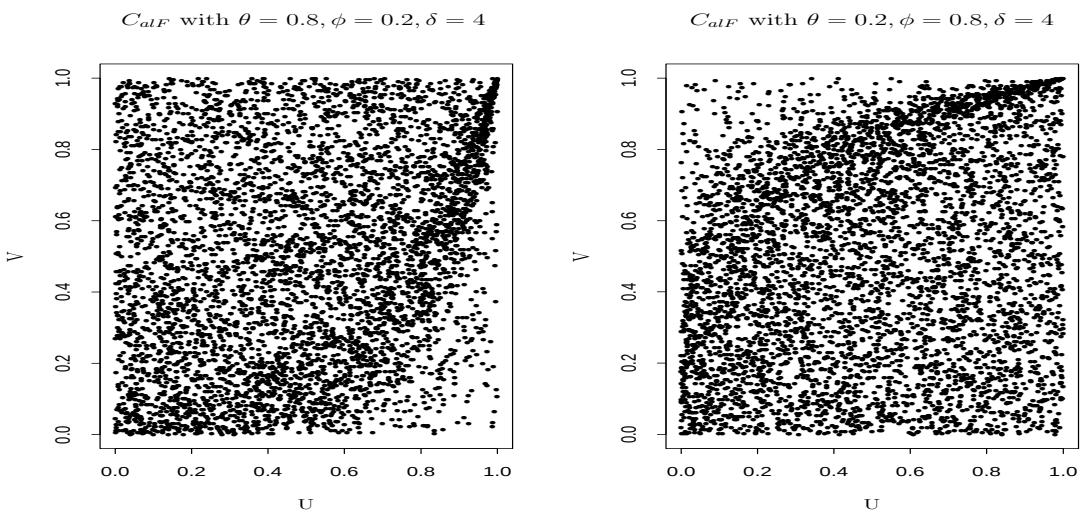

Figure 5.1: 5000 simulations of C_{alF} with $\theta = 0.8, \phi = 0.2, \delta = 4$ and $\theta = 0.8, \phi = 0.2, \delta = 4$ respectively

Lemma 5.1.1. *C_{alF} is an extreme value copula, i.e. $C_{alF}(u^{\nu}, v^{\nu}) = C_{alF}^{\nu}(u,v)$.*

Proof.

$$\begin{aligned}
C_{alF}(u^\nu, v^\nu) &= u^{\nu(1-\theta)} v^{\nu(1-\phi)} \exp\left(-\left\{(-\theta \ln(u^\nu))^\delta + (-\phi \ln(v^\nu))^\delta\right\}^{\frac{1}{\delta}}\right) \\
&= (u^{1-\theta} v^{1-\phi})^\nu \exp\left(-\left\{(-\theta\nu \ln(u))^\delta + (-\phi\nu \ln(v))^\delta\right\}^{\frac{1}{\delta}}\right) \\
&= (u^{1-\theta} v^{1-\phi})^\nu \exp\left(-\nu\left\{(-\theta \ln(u))^\delta + (-\phi \ln(v))^\delta\right\}^{\frac{1}{\delta}}\right) \\
&= C_{alF}^\nu(u, v)
\end{aligned}$$

□

If $\theta = \phi := \theta^\star$ we are in the symmetric case $C_{alF}^\star(u,v) = C_{alF}^\star(v,u)$ and can express $C_{alF}^\star$ in terms of Archimedean copulas. Therefore we define $\varphi^\star(t) := (-\theta^\star \ln(t))^\delta$. Since $t \mapsto -\ln(t)$ is the generator of the independence copula, the class of generators is closed under multiplication with positive constants, and the power of a generator with an exponent $\delta \geq 1$ is also still a generator, $\varphi^\star$ is the generator of an Archimedean copula $C_{\varphi^\star}$ (see Junker and May [67] or Chapter 4). Straight forward calculation yields

$$C_{alF}^\star(u,v) = \Pi(u,v)^{1-\theta^\star} C_{\varphi^\star}(u,v)^{1/\theta^\star}.$$

For $\theta = \phi = 1$ we achieve the well known Gumbel copula given by equation (2.3.6). The next proposition states the tail dependence properties of C_{alF}.

Proposition 5.1.2. *Let C_{alF} be a copula as in equation (5.1.3). Then C_{alF} has no lower tail dependence, but is upper tail dependent with*

$$\lambda_U = \theta + \varphi - \left(\theta^\delta + \varphi^\delta\right)^{\frac{1}{\delta}} < 1$$

Proof. We first calculate the upper tail dependence,

$$\begin{aligned}
\lambda_U &= \lim_{u\to 1} \frac{1 - 2u + C(u,u)}{1-u} \\
&= \lim_{u\to 1} \frac{1 - 2u + u^{2-\theta-\phi} \exp\left(-\left\{(-\theta \ln(u))^\delta + (-\phi \ln(u))^\delta\right\}^{\frac{1}{\delta}}\right)}{1-u} \\
&= \lim_{u\to 1} \frac{1 - 2u + u^{2-\theta-\phi} \exp\left(\ln(u)\left(\theta^\delta + \phi^\delta\right)^{\frac{1}{\delta}}\right)}{1-u} \\
&= \lim_{u\to 1} \frac{1 - 2u + u^{2-\theta-\phi} u^{\left(\theta^\delta + \phi^\delta\right)^{\frac{1}{\delta}}}}{1-u} \\
&= \theta + \phi - \left(\theta^\delta + \phi^\delta\right)^{\frac{1}{\delta}},
\end{aligned}$$

where we have applied l'Hospital's rule in the last step. Now $\theta + \phi - \left(\theta^\delta + \phi^\delta\right)^{\frac{1}{\delta}} < 1$, since

$$\theta + \phi - 1 \leq \min(\theta, \phi) < 2^{\frac{1}{\delta}} \min(\theta, \phi) \leq \left(\theta^\delta + \phi^\delta\right)^{\frac{1}{\delta}},$$

for all $\theta, \phi \in [0, 1]$ and $1 \leq \delta < \infty$. The lower tail dependence can be achieved again by the use of l'Hospital and $\lambda_U < 1$

$$\begin{aligned} \lambda_L &= \lim_{u \to 0} \frac{C(u,u)}{u} = u^{1-\theta-\phi+\left(\theta^\delta+\phi^\delta\right)^{\frac{1}{\delta}}} \\ &= \lim_{u \to 0} u^{1-\lambda_U} = 0 \end{aligned}$$

□

So for $\theta, \phi \to 0$ we have $\lambda_U \to 0$ and for $\theta = \phi = \theta^\star$ one calculate $\lambda_U = \theta^\star \left(2 - 2^{\frac{1}{\delta}}\right)$, i.e, the upper tail dependence is $\theta^\star$ times the upper tail dependence of the Gumbel copula.
The special form seen in Figure 5.1 may be of interest if modelling joint default probabilities. Suppose e.g. U representing the default probability (dp) (mapped to uniformity) of a small feeder plant for Daimler Chrysler, whose mapped dp is represented by V. We argue that the feeder plant is very sensible for a deteriorating dp of Daimler Chrysler, whereas Daimler is more robust against the dp changes of the feeder plant. We end up with a scenario represented in the left plot of Figure 5.1.

Chapter 6

Multidimensional Goodness-of-Fit Tests

This chapter, partially covered by Junker and May [67], gives an overview over different test procedures that allow to compare the in-sample and out of sample fit for parametric distribution models. We introduce a copula based mapping rule leading to joint uniform independence and as results of this mapping we present an easy method of multidimensional χ^2 testing and a new estimate for high dimensional parametric distribution functions. Further we summarize the multivariate test statistics of some well known tests like Anderson Darling and Aikaike and Bayesian Information Criterion, and give some finite sample refinements.

6.1 Multidimensional χ^2 tests

In the first step we recall the χ^2 test in a multivariate setting. We will give a special formulation that enables us to perform a goodness-of-fit test for any parametric distribution function without need of thinking how to divide the testspace in cells.

Let $X = (X_1, \ldots, X_d)$ be a d-dimensional random vector with marginales X_i, $i = 1, \ldots, d$, and absolutely continuous parametric distribution function $F_\theta \colon D \to [0,1]$, $D \subseteq I\!R^d$ with parameter vector θ. By $x^j = \left(x_1^{(j)}, ..., x_d^{(j)}\right)$ we denote the d-dimensional observations numbered with $j = 1, ..., m$. The empirical distribution function based on the m-sample of observations is then defined as follows

$$F_m\left(x_1, ..., x_d\right) = \frac{1}{m} \sum_{j=1}^{m} \prod_{i=1}^{d} 1_{\left\{x_i^{(j)} \leq x_i\right\}} \left(\forall x \in D\right). \tag{6.1.1}$$

We want to test the hypothesis $F_\theta = F_m$, i.e. we ask whether the sample comes from a population distributed with F_θ. As in the well-known univariate χ^2 test we have to divide the test space into regions (classes, cells) of known (mostly

equal) probability. The statistics is given by

$$\chi^2_{\mathrm{df}} = \sum_{l=1}^{k} \left(\frac{\mathbf{E}_{F_\theta}(c_l) - \#\left\{x^{(j)},\ j = 1, ..., m : x^{(j)} \in c_l\right\}}{\mathbf{E}_{F_\theta}(c_l)} \right)^2, \tag{6.1.2}$$

with $\mathbf{E}_{F_\theta}(c_l)$ denoting the number of expected observations in a cell c_l under the model F_θ. The statistics follows asymptotically a χ^2-distribution with $k-1-p$ degrees of freedom (df), where p is the length of θ, i.e. the number of parameters. In the next subsection we derive a way how the class separation can easily be done in a multivariate setting.

6.1.1 A mapping rule leading to joint uniform independence

We first transform our data into uniformly distributed ones. Let T be a transformation that maps the univariate marginals X_j onto uniformly distributed random variables on $[0,1]^d$. Transformations with these properties have been subject to investigations already in the thirties of the last century (e.g. Lévy [1930]). We refer to Rosenblatt [102] for a concise presentation of the idea behind.

For the univariate marginals X_j, $j = 1, \ldots, d$, let F_{X_j} denote the univariate distribution function of the j-th margin. Define the transformation $T : I\!R^d \to [0,1]^n$ with $T(x_1, \ldots, x_d) = (z_1, \ldots, z_d)$ via

$$\begin{aligned}
z_1 &= \mathbf{P}[X_1 \le x_1] = F_{X_1}(x_1) \\
z_2 &= \mathbf{P}[X_2 \le x_2 | X_1 = x_1] = F_{X_2|X_1}(x_2|x_1) \\
&\vdots \\
z_d &= \mathbf{P}[X_d \le x_d | X_1 = x_1, \ldots, X_{d-1} = x_{d-1}] \\
&= F_{X_d|X_1,\ldots,X_{d-1}}(x_d|x_1, \ldots, x_{d-1})\,.
\end{aligned}$$

Then the random variables $Z_j = T(X_j)$, $j = 1, \ldots, d$, are independent and the random vector $Z = (Z_1, \ldots, Z_d)$ is uniformly distributed on $[0,1]^d$. So testing whether $x^{(j)}$ is distributed with F is equivalent to testing on $Z \sim \Pi$ where $\Pi(z_1, \ldots, z_d) = \prod_{j=1}^{d} z_j$.

We now formulate things in terms of copulas. Let C be a copula such that

$$F(x_1, \ldots, x_d) = C(F_{X_1}(x_1), \ldots, F_{X_d}(x_d))).$$

The following Proposition 6.1.1 is a copula version based on the idea in Diebold et al. [29].

Proposition 6.1.1. *Let $U_1, \ldots, U_d$ be $U(0,1)$ random variables and $C \in \mathbf{C}^d$ a d-copula. If the joint distribution of $(U_1, \ldots, U_d)$ is C, the d-dimensional random vector*

$$(U_1, C_2(U_2 \mid U_1), \ldots, C_d(U_d \mid U_1, U_2, \ldots, U_{d-1})) \sim \Pi$$

is distributed like the d-dimensional independence copula Π. Here,

$$C_k(u_1, \ldots, u_k) = C(u_1, \ldots, u_k, 1, \ldots, 1), \quad k = 1, \ldots, d$$

denote the k-marginals of $C(u_1, \ldots, u_d)$ and the conditional distribution of U_k given the values of the first k-1 components of $(U_1, \ldots, U_d)$ is expressed by

$$\begin{aligned} C_k(u_k \mid u_1, \ldots, u_{k-1}) &= \mathbf{P}[U_k \leq u_k \mid U_1 = u_1, \ldots, U_{k-1} = u_{k-1}] \\ &= \frac{\partial^{k-1} C_k(u_1, \ldots, u_k)}{\partial u_1 \ldots \partial u_{k-1}} \Big/ \frac{\partial^{k-1} C_{k-1}(u_1, \ldots, u_{k-1})}{\partial u_1 \ldots \partial u_{k-1}} . \end{aligned}$$

Proof. By using the algorithm for random number generation from the copula C, we derive an alternative representation by

$$\begin{aligned} (U_1, \ldots, U_d) &\stackrel{d}{=} (V_1, C_2^{-1}(V_2 \mid V_1), \ldots, C_d^{-1}(V_d \mid V_1, C_2^{-1}(V_2 \mid V_1), \ldots)) \\ &\sim C \end{aligned} ,$$

where the $V_1, \ldots, V_d$ are iid $U(0,1)$ random variables. A componentwise comparison gives $U_1 \stackrel{d}{=} V_1$, $U_2 \stackrel{d}{=} C_2^{-1}(V_2 \mid U_1)$, ... , $U_d \stackrel{d}{=} C_d^{-1}(V_d \mid U_1, U_2, \ldots, U_{d-1})$. If we iteratively plug this in we get $U_1 \stackrel{d}{=} V_1$, $C_2(U_2 \mid U_1) = C_2(C_2^{-1}(V_2 \mid U_1) \mid U_1) = V_2$, ..., $C_d(U_d \mid U_1, U_2, \ldots, U_{d-1}) = C_d(C_d^{-1}(V_d \mid U_1, U_2, \ldots, U_{d-1}) \mid U_1, U_2, \ldots, U_{d-1}) = V_d$. So we have

$$\begin{aligned} (U_1, C_2(U_2 \mid U_1), \ldots, C_d(U_d \mid U_1, U_2, \ldots, U_{d-1})) &\stackrel{d}{=} (V_1, \ldots, V_d) \\ &\sim \Pi , \end{aligned} \tag{6.1.3}$$

and the proof is finished. □

6.1.2 χ^2 test with class separation

Let $(u_1, \ldots, u_d)^{(i)}$ for $i = 1, \ldots, m$ be m realizations of a d-dimensional uniform random vector. Due to Proposition 6.1.1, the following two hypotheses are equivalent:

Hypothesis A: The $(u_1, \ldots, u_d)^{(i)}$ are realizations of the copula C.

Hypothesis B: $(u_1, C_2(u_2 \mid u_1), \ldots, C_d(u_d \mid u_1, u_2, \ldots, u_{d-1}))^{(i)}$ are realizations of the d-dimensional independence copula Π.

Under B, the testspace $[0, 1]^d$ can easily be separated into d-dimensional cuboids, since the expected number of realizations in each cuboid is directly proportional to its volume.

Remark 13. *If the copula under one of the equivalent hypotheses A or B is Archimedean with a strict generator φ, the test becomes even more easy because of*

$$C_k(u_k \mid u_1, \ldots, u_{k-1}) = \frac{\varphi_{k-1}^{-1}\left(\varphi(u_1) + \ldots + \varphi(u_k)\right)}{\varphi_{k-1}^{-1}\left(\varphi(u_1) + \ldots + \varphi(u_{k-1})\right)},$$

where $\varphi_k^{-1} = \frac{\partial^k}{\partial u^k}\varphi^{-1}$.

Since any multidimensional continuous distribution function F can be written as $F(x_1, \ldots, x_d) = C(F_{X_1}(x_1), \ldots, F_{X_d}(x_d))$, where the F_{X_i} are the marginal distribution functions and C the copula representing the dependence structure of F, the Proposition 6.1.1 also provides an easy method for testing arbitrary multidimensional distributions.

Using Proposition 6.1.1 we can write the random variables Z_k for $2 \leq k \leq d$ using the conditional distributions of the k-copula C_k,

$$Z_k = C_k(F_{X_k}(X_k)|F_{X_1}(X_1), \ldots, F_{X_{k-1}}(X_{k-1}))\,. \tag{6.1.4}$$

Breymann et al. [14] now use the following idea.

Observe that $(F_{X_1}(X_1), F_{X_2}(X_2), \ldots, F_{X_d}(X_d)) \sim C$. Then $\Phi^{-1}(Z_k) \sim \mathrm{N}(0,1)$ iid for $i = 1, \ldots, d$. It follows that $\sum_{k=1}^{d}\left(\Phi^{-1}(Z_i)\right)^2$ has a χ^2 distribution with d degrees of freedom. The summation over all dimensions yields of course to a lack of multivariate information. Moreover —and a real drawback with respect to asymmetric dependent data (in the sense $C \neq \widetilde{C}$)— the test favors elliptical distributions (more precisely, generalized elliptical distributions as defined in Chapter 12) since there is no test on spherical uniformity of U in $\|Z\| = \sqrt{I}\cdot U$. We remark that this radial test coincides with a multidimensional test with further mapping to independent normal and spherical annuli (radial intervals) separation. The separation is done in a way that all annuli have equal mass under the hypothesis. Equivalently, the expected value for the radius of a point measured in L_2 norm is constant for all annuli. (Here we assume the biggest radius to be ∞.) This one-sided class separation can not test the sample for spherical uniformity.

Next we illustrate the power of the multivariate test with cubic separation after applying the mapping rule 6.1.1 compared to the above radial test for different scenarios.

The left picture in Figure 6.1 shows the average p-value for the test with cubic class separation (solid line) and radial test (Breymann et al., dotted line). The average is taken over the tests based on 100 simulated bivariate standard normal distributed samples with zero correlation and sample size 2000. The hypothesis is "$X \sim N(0, \rho)$" (x-axis), where we have taken $\rho \in [-0.5, 0.5]$ at 100 discretized equidistant points. The range for which the hypothesis can not be rejected on the 95% level, is $-0.15 \leq \rho \leq 0.12$ for cubic class separation and

$-0.33 \leq \rho \leq 0.32$ for radial separation. From maximum likelihood theory we deduce that the asymptotic 95% confidence interval for 2000 data points for an ML estimator of ρ is $[-0.09, 0.09]$. So both tests can not be viewed as optimal since the region of acceptance is too large compared to the included confidence interval.

The right plot in Figure 6.1 shows again the average p-values for 100 simulated samples of size 1000 from a survival copula of Gumbel type $\widetilde{C}_{G,\delta}$ where the parameter was chosen as $\delta = 3$. As hypothesis we test here

$$X \sim \frac{1}{2}\left(C_{G(\delta)} + \widetilde{C}_{G(\delta)}\right), \tag{6.1.5}$$

that is the linear combination of a Gumbel copula with parameter δ and its corresponding survival copula. Remark that under this hypothesis we have a symmetric model whereas the true model is highly asymmetric. The region where the hypothesis can not be rejected on the 95% level, is here computed as $1.31 \leq \delta \leq 1.52$ for the radial test whereas the hypothesis can not be accepted on any of the usual levels if the cubic class test is performed (maximal p value is below 0.0001.) The average ML estimated value under the hypothesis is $\overline{\widehat{\delta}} = 1.47$ and for the mean standard error we compute $\overline{\widehat{\sigma_\delta}} = 0.0343$. Thus the estimated asymptotic 95% confidence interval is $[1.41, 1.54]$, i.e. there is some evidence that a parameter estimated under the wrong hypothesis model falls into the region where the radial test can not reject the hypothesis.
For both plots we had run the tests over 100 cells.

Figure 6.2 shows the deviations from the theoretical expected value $\frac{2000}{100}$ for each of the 100 test cells in a sample from the survival Gumbel copula of size 2000 data points. Here we performed the mapping using the model $\frac{1}{2}\left(C_{G,\delta} + \widetilde{C}_{G,\delta}\right)$ where $\delta = 1.4$. The data were first mapped applying (6.1.5) onto a uniformly iid sample (left plot), in a second step onto bivariate standard normal distributed data (right plot). Bright colors show the tendency for underestimating, dark colors are a sign for overestimating.

As seen, the greyscale of the cubic separation (left plot) has a significant larger range compared to the radial separation (right plot), indicating the higher power of the cubic test.

Figure 6.3 shows the scatterplot representation of Figure 6.2. The right plot shows that the points are not likely to be (approximately) standard normally distributed. Despite of this fact, the radial test over 100 radial classes gives a p-value of 0.08. The plot shows every fifth disk segment, i.e. under the hypothesis there are 100 points between 2 radii. This is still satisfied under the wrong hypothesis (6.1.5).

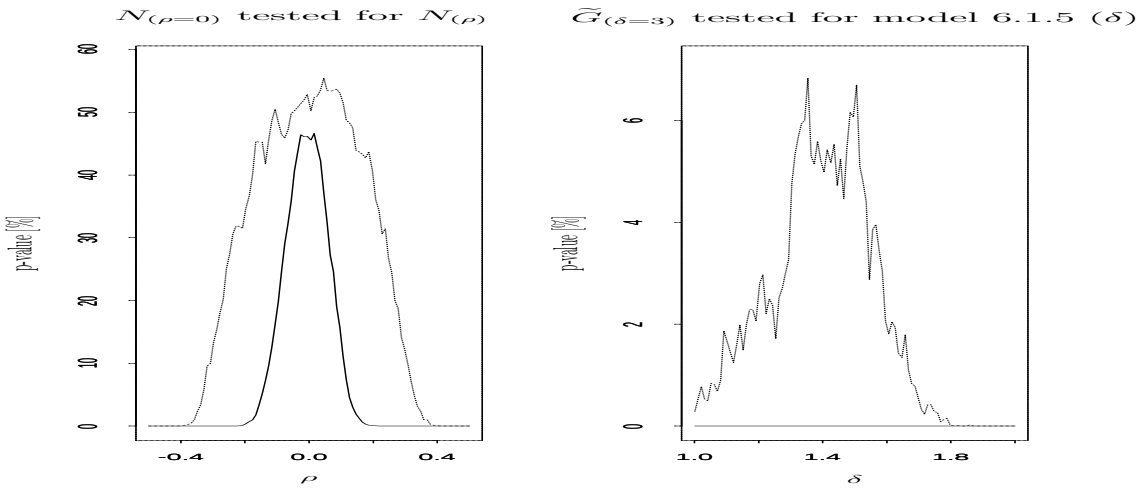

Figure 6.1: p-values for the tests with cubic (solid line) and radial (dotted line) testcells. To the left the hypothesis is "standard normal with different ρ" and we sampled standard normal data with $\rho = 0$. In the right plot the test runs on the symmetric hypothesis (6.1.5) with different δ, where we sampled survival Gumbel copulas with $\delta = 3$.

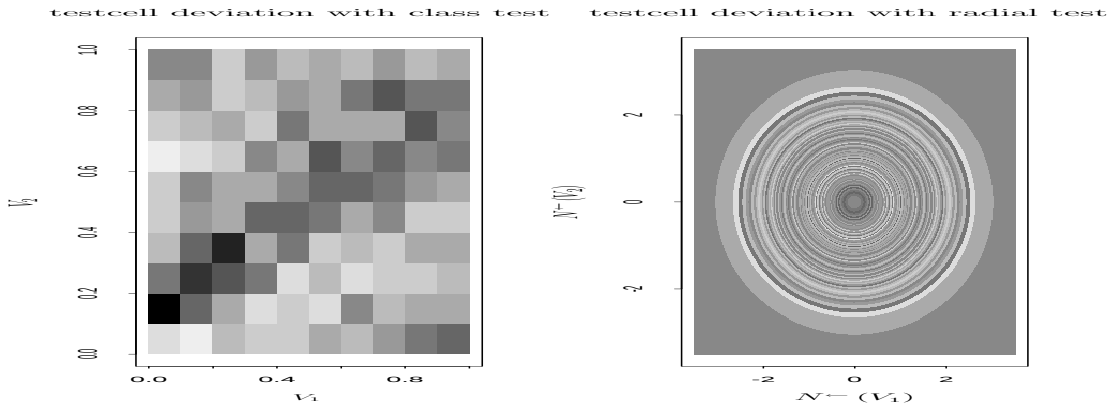

Figure 6.2: Deviation of the number of expected points in the testcells after applying the mapping rule from 6.1.1 on a survival Gumbel sample with hypothesis 6.1.5 and $\delta = 1.4$ for the class and radial test. Bright colors indicate underestimation, dark overestimation.

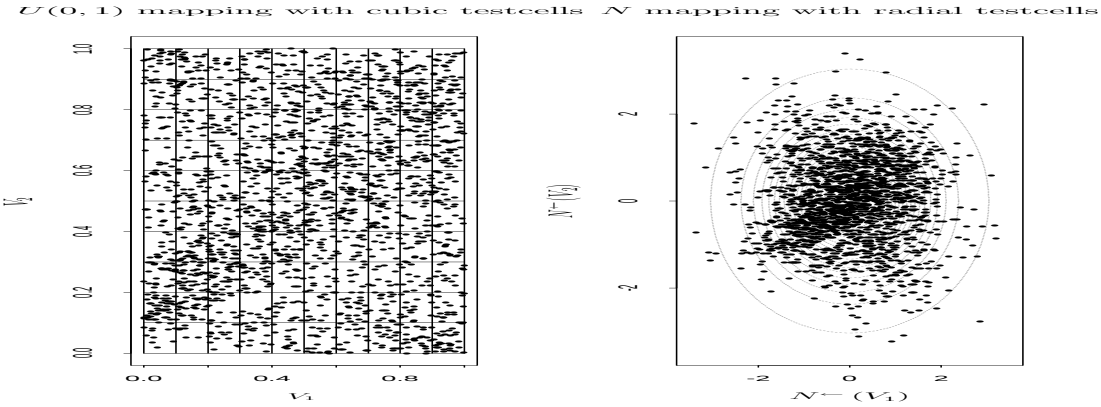

Figure 6.3: Standard uniform mapped sample with cubic testcells (left) and standard normal mapped sample with radial testcells (right). Both mapped under hypothesis (6.1.5).

The examples given showed clearly the better power of the cubic test in comparison to the radial test.

For high dimensions the cubic separation is in practise not possible any more, since we have at least 2^d classes what for large d grows fast to a greater number than the number of observations m. Changing to cuboid separation makes things a lot more complicated. A possible solution is the radial test since here we only have to separate the one dimensional radial space $\mathbb{R}^+$, but we face the weaker power as showed above. If the sample mapped to independent normal (under the hypothesis) is tested both for χ_d-distribution of the radius and for radial symmetry, or spherical uniformity of the scaled sample $X/\|X\|$ (see e.g. Mardia [85]), we will get extra information about the plausibility of the hypothesis. Unfortunately it is not clear how to combine both test results into one test.
The further mapping to independent normal may still be a benefit, since the separation in spherical annuli *and* spherical segments avoid the stated power problem by giving broader opportunities for class separation.
However, in the next subsection we will give an alternative and new approach that overcomes the drawback of separation preserving the multidimensional resolution.

6.1.3 Estimation and χ^2 testing using the distribution of differences

As before, we start with a d-dimensional random vector $(U_1, \ldots, U_d) \sim C$. The mapping described above yields a uniformly distributed

random vector $(\tilde{U}_1, \ldots, \tilde{U}_d) \sim \Pi$ with independent marginals. The transformed vector $Z = (\Phi^{-1}(\tilde{U}_1), \ldots, \Phi^{-1}(\tilde{U}_d))$ is multivariate normally distributed. We observe that $Z = \|Z\| \cdot V$ with $R \stackrel{def}{=} \|Z\| \sim \sqrt{\chi^2_n}$ and V uniformly distributed on the $n-1$ dimensional unit sphere S^{n-1}.

Lemma 6.1.2. *Let $Y = \|Z_i - Z_j\|^2$ for all i,j and independent normally distributed $Z_i, Z_j \sim N(0, I)$. Then the distribution function of Y writes as follows*

$$F_Y = \mathbf{E}_{(R_i,R_j)\sim(\sqrt{\chi^2_n},\sqrt{\chi^2_n})} \left[F_\beta \left(\frac{(R_i^2 + R_j^2 - \delta)^2}{4R_i^2 R_j^2}, \frac{1}{2}, \frac{n-1}{2} \right) \right].$$

Proof. Set without loss of generality $\delta \stackrel{def}{=} \|Z_1 - Z_2\|^2 = Y$, $R_1 \stackrel{def}{=} \|Z_1\|$, $R_2 \stackrel{def}{=} \|Z_2\|$ and discuss the triangle with wedges $\|Z_1\|, \|Z_2\|, \delta$. By trigonometric arguments, we compute $\delta = \|Z_1 - Z_2\|^2 = \|Z_1\|^2 + \|Z_2\|^2 - 2\|Z_1\| \, \|Z_2\| \cos\alpha$ where $\alpha = \alpha(\delta) = \arccos\left(\frac{-\delta + R_1^2 + R_2^2}{2R_1R_2}\right)$. Let $f_{(R_1,R_2,\alpha)}$ denote the density of the random vector (R_1, R_2, α). Write e.g. f_{R_1} for the corresponding univariate density and note that $f_{(R_1,R_2,\alpha)} = f_{R_1} f_{R_2} f_\alpha$ because the independence of R_1, R_2, α. We now apply the transformation rule to the conditional density $f_{\delta|(R_1,R_2)=(r_1,r_2)}$

$$\begin{aligned}
f_{\delta|(R_1,R_2)=(r_1,r_2)}(\delta) &= f_{R_1,R_2,\alpha}(r_1, r_2, \alpha) \cdot \left| \tfrac{\partial\alpha}{\partial\delta}(r_1, r_2) \right| \\
f(\delta) &= \int_0^\infty \int_0^\infty f_\alpha(\alpha) \cdot \left| \tfrac{\partial\alpha}{\partial\delta}(r_1, r_2) \right| \cdot f_{R_1}(r_1) \, f_{R_2}(r_2) \, dr_1 dr_2 \\
f(\delta) &= \mathbf{E}_{(R_1,R_2)\sim\left(\sqrt{\chi^2_n},\sqrt{\chi^2_n}\right)} \left[f_\alpha(\alpha) \cdot \left| \tfrac{\partial\alpha}{\partial\delta}(R_1, R_2) \right| \right]
\end{aligned}$$

In the next step we calculate the derivative and derive from Theorem 12.3.1, where it is stated that $\alpha \sim \frac{1}{2} + \frac{1}{2} \, sign\left(\alpha - \frac{\pi}{2}\right) \cdot F_\beta\left(\cos^2(\alpha), \frac{1}{2}, \frac{n-1}{2}\right)$, the density of α.

$$\begin{aligned}
f(\delta) &= \mathbf{E}_{(R_1,R_2)\sim\left(\sqrt{\chi^2_n},\sqrt{\chi^2_n}\right)} \left[\frac{2R_1R_2\Gamma\left(\frac{n}{2}\right)\sqrt{\frac{\left(R_1^2+R_2^2-\delta\right)^2}{R_1^2R_2^2}}\left(1-\frac{\left(R_1^2+R_2^2-\delta\right)^2}{4R_1^2R_2^2}\right)^{\frac{n-2}{2}}}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)\cdot\left(R_1^2+R_2^2-\delta\right)\cdot\left|2R_1R_2\sqrt{1-\frac{\left(R_1^2+R_2^2-\delta\right)^2}{4R_1^2R_2^2}}\right|} \right] \\
F(\delta) &= \mathbf{E}_{(R_1,R_2)\sim\left(\sqrt{\chi^2_n},\sqrt{\chi^2_n}\right)} \left[\int_{R_1^2+R_2^2-2R_1R_2}^{R_1^2+R_2^2} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} \frac{\left(1-\frac{\left(R_1^2+R_2^2-\delta\right)^2}{4R_1^2R_2^2}\right)^{\frac{n-3}{2}}}{\sqrt{R_1^2R_2^2}} \, d\delta \right] \\
&= \mathbf{E}_{(R_1,R_2)\sim\left(\sqrt{\chi^2_n},\sqrt{\chi^2_n}\right)} \left[\int_{\frac{\left(R_1^2+R_2^2-\delta\right)^2}{4R_1^2R_2^2}}^{0} -\frac{\Gamma\left(\frac{n-1}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} t^{\frac{1}{2}-1}(1-t)^{\frac{n-1}{2}-1} \, dt \right] \\
&= \mathbf{E}_{(R_1,R_2)\sim\left(\sqrt{\chi^2_n},\sqrt{\chi^2_n}\right)} \left[F_\beta\left(\frac{(R_1^2+R_2^2-\delta)^2}{4R_1^2R_2^2}, \frac{1}{2}, \frac{n-1}{2}\right) \right]
\end{aligned}$$

Here we have used in the second step the correspondence of $\delta \in [R_1^2 + R_2^2 - 2R_1R_2, R_1^2 + R_2^2]$ with $\alpha \in \left[0, \frac{\pi}{2}\right]$ and in the third the substitution $t = \frac{(R_1^2+R_2^2-\delta)^2}{4R_1^2R_2^2}$. $\square$

Let $x^{(i)}$ denote an observation in the n-dimensional m-sample. Applying the construction of Y from Lemma 6.1.2 we derive $\binom{m}{2}$ squared differences $\|z^i - z^j\|^2$ for $i, j = 1, \ldots, m$ with $j < i$. Note that these are not independent. Suppose e.g. $m = 3$, than given two differences $\sqrt{\delta_1}$ and $\sqrt{\delta_2}$ the third one has to fulfil $\sqrt{\delta_3} \in [\left|\sqrt{\delta_1} - \sqrt{\delta_2}\right|, \sqrt{\delta_1} + \sqrt{\delta_2}]$ and hence depends on δ_1, δ_2. One possibility to achieve independent distances is to fix a point, e.g. z^1, and take only the $m-1$ differences to this fixed point. That is $\|z^1 - z^j\|^2$ for $j = 2, \ldots, m$. Translate the origin to z^1 to show that we then have the same situation as in the radial test, and hence would lose multidimensional information. The following remark states an algorithm to overcome such problems.

Remark 14. *Draw m times without replacement from the pseudo observations $z^1, \ldots, z^m$ and denote the new series by $\widetilde{z}^1, \ldots, \widetilde{z}^m$. Now form the $\left\lfloor \frac{m}{2} \right\rfloor$ squared differences $y^k \stackrel{def}{=} \|\widetilde{z}^{2k-1} - \widetilde{z}^{2k}\|^2$, $k = 1, \ldots, \left\lfloor \frac{m}{2} \right\rfloor$, where $\lfloor\ \rfloor$ are the gaussian brackets. The hypothesis that the sample $x^{(i)}$ comes from F_θ-distributed data (where F_θ denotes the multivariate parametric distribution function as usual) can be tested by applying a one-dimensional χ^2 test to the transformed data y^k. In contrast to the above radial method the proposed approach makes use of the full multivariate information and is space covering.*

Furthermore, Lemma 6.1.2 yields a new ML estimation method for the parameters of F_θ based on the log-likelihood function $\sum_{k=1}^{\left\lfloor \frac{m}{2} \right\rfloor} \ln f(y^k)$. Remark therefore that the differences y^k depend on the parameter vector θ.

To summarize: We presented a procedure that is based on a univariate χ^2 test, i.e. no intersection in classes is necessary. By transforming the data, the volume of the data changes from m d-dimensional data to $\left\lfloor \frac{m}{2} \right\rfloor$ univariate independent differences whereas Breymann et al. can deal with m univariate radial data. An advantage of this the procedure is that it yields an explicit univariate distribution function and an ML estimation procedure for the parameters of the joint distribution function.
The properties of this ML estimate and the corresponding χ^2 test will be subject of forthcoming work. Here we also have to check if one can take benefit from the fact that there are $m!/2^{\left\lfloor \frac{m}{2} \right\rfloor}$ different admissible series y^k.

Another estimation method can be derived by taking the differences in the uniform space after applying the mapping 6.1.1. Suppose we have $X \sim F_\theta$ and

map $X = (X_1, \ldots, X_d)$ to $\widetilde{U} = \left(\widetilde{U}_1, \ldots, \widetilde{U}_d\right) \sim \Pi$. Let $\widetilde{u}^1, \ldots, \widetilde{u}^m$ be m independent realizations of $\widetilde{U}$. We can calculate the differences $\|\widetilde{u}^i - \widetilde{u}^j\|, i \neq j$ and denote them by Δ_j^i, $i = 1, \ldots, m,\ j = 1, \ldots, m-1$ where Δ_j^i is the distance between $\widetilde{u}^i$ and its jth next neighbor. We denote the expected value under the hypothesis (and hence independent of the supposed θ) with $\Delta_j \stackrel{def}{=} \mathbf{E}\left[\Delta_j^i\right]$. Note that the Δ_j only depend on j, d, m but not on i, since $\widetilde{U} \sim \Pi$.
Having observations $x^1, \ldots, x^m$ of X we can compute the $m(m-1)$ differences $\Delta_j^i(\theta)$ and estimate θ via the following least square equation

$$\widehat{\theta} \stackrel{def}{=} \arg\min_{\theta} \sum_{j=1}^{m-1} \sum_{i=1}^{m} \left(\Delta_j^i(\theta) - \Delta_j\right)^2$$

In the following we use this approach to achieve start estimators for the usual ML estimation methods stated in Subsection 7.3.1. The Δ_j are derived by a preciding Monte Carlo Simulation.

6.2 More goodness-of-fit tests

We now turn to different criteria and methods that allow for ordering different parametric models with respect to its goodness-of-fit. We briefly discuss the limitations when using the single methods.

Let us again assume that the data are d-dimensional, $X = (X_1, \ldots, X_d)$ and that $x^j = \left(x_1^j, ..., x_d^j\right)$ denotes the d-dimensional observations numbered by $j = 1, ..., m$.

Here, we fit parametric models that depend on p parameters, say, i.e. let θ denote the parameter vector of length p. We assume that the model is described by a parametric distribution function $F_\theta : S \to [0,1]$ with density f_θ and support S. We write MLf for the log likelihood function

$$\mathrm{ML}f\left(f_\theta; x^1, ..., x^m\right) = \sum_{j=1}^{m} \ln\left(f_\theta\left(x^j\right)\right) \tag{6.2.1}$$

and denote by $\widehat{\theta}$ the ML estimated parameter vector.

A good hint on the goodness-of-fit related to the number of model parameters is given by Akaike's Information Criterion (AIC) or the Bayesian Information Criterion (BIC),

$$\begin{aligned} AIC\left(f_{\widehat{\theta}}; x^1, ..., x^m\right) &= -2\mathrm{ML}f\left(f_{\widehat{\theta}}; x^1, ..., x^m\right) + 2p \\ BIC\left(f_{\widehat{\theta}}; x^1, ..., x^m\right) &= -2\mathrm{ML}f\left(f_{\widehat{\theta}}; x^1, ..., x^m\right) + p \ln m\,, \end{aligned} \tag{6.2.2}$$

Both criteria measure the probability of the *observations* for the different models with respect to the number of parameters used. So overparametrization is penalized by adding p to the ML-Function, and the BIC puts more penalty on the number of parameters (for $m > 7$) than the Akaike criterion. This is of special importance if $\widehat{\theta}$ is a ML estimator and the p degrees of freedom are then used to maximize (6.2.1)

The measures have their origin in the time series analysis where they are used to estimate the order of ARMA processes. In general they do not really have a clear interpretation but are still popular as a of thumb comparison. Lower values indicate a better fit but obviously the values do highly depend on the data. To be on the safe side we will further claim to have both criteria ordered in the same direction (i.e. both smaller or both larger), otherwise no deduction will be made.

The entropy is focused more on the model and hence gains more independence from the data sample. The definition is as follows

$$\begin{aligned} \text{Entropy}\left(f_{\widehat{\theta}}\right) &= \mathbf{E}\left[-\ln\left(f_{\widehat{\theta}}(X)\right)\right] = -\int_S \ln\left(f_{\widehat{\theta}}(x)\right)\, dF_{\widehat{\theta}}(x) \\ &= -\int_S \ln\left(f_{\widehat{\theta}}(x)\right) \cdot f_{\widehat{\theta}}(x)\, dx\,. \end{aligned} \tag{6.2.3}$$

Note that the last expression in equation (6.2.3) is only defined if $f_{\widehat{\theta}}$ exists. The entropy is based on the expected value of the negative log-likelihood function (6.2.1) with sample size $m = 1$. Throughout our considerations, we will use the integral form of the entropy whereas other authors take the mean over the pseudo observations $\ln\left(f_{\widehat{\theta}}(x^j)\right)$, $j = 1, \ldots, m$ only. So the information content of the different models is measured in means of the log likelihood function. This is to say that the entropy measures the probability that the *model* is the correct one. This contrasts to the AIC and BIC because due to the restriction to the expected value the outcome does *not* depend (directly) on the data.

The χ^2 test in contrast measures so to say differences in the densities whereas alternative goodness-of-fit tests like the Kolmogorov-Smirnov or the von Mises, rely on distances between the empirical and the parametric distribution function. (Recall that in the case $d > 2$ the Kolmogorov-Smirnov statistics is not the same for all continuous F; the same holds for the von Mises statistics.) Since convergence must coincide for the empirical and the theoretical distribution function in both tails, their representation of fit in the tails is weak and thus the tests are not as powerful as the χ^2 test. A test statistics which is superior in this respect dates back to Anderson and Darling [2]. The Anderson and Darling test uses relative instead of absolute deviations between the distribution functions and thereby gives a better representation of the fit in the tails. Mainly there are two types of Anderson Darling statistics in use, the supremum version SAD, and the

integrated version AD, respectively, both summarized in the following equations

$$\begin{aligned} SAD\left(F_{\widehat{\theta}}; x^1, ..., x^m\right) &= \sqrt{m} \sup_{x \in S} \frac{\left|F_m(x) - F_{\widehat{\theta}}(x)\right|}{\sqrt{F_{\widehat{\theta}}(x)\left(1 - F_{\widehat{\theta}}(x)\right)}} \\ AD\left(F_{\widehat{\theta}}; x^1, ..., x^m\right) &= m \int_S \frac{\left(F_m(x) - F_{\widehat{\theta}}(x)\right)^2}{F_{\widehat{\theta}}(x)\left(1 - F_{\widehat{\theta}}(x)\right)} \cdot dF_{\widehat{\theta}}(x) \\ &= m \int_S \frac{\left(F_m(x) - F_{\widehat{\theta}}(x)\right)^2}{F_{\widehat{\theta}}(x)\left(1 - F_{\widehat{\theta}}(x)\right)} \cdot f_{\widehat{\theta}}(x) dx \,, \end{aligned} \tag{6.2.4}$$

where the last line only holds if $f_{\widehat{\theta}}$ exists. Note that SAD is —on account of the supremum— sensitive to single "bad" observation whereas the integrated version is more robust because of the integration. In the following we relate on the AD results only.

In our empirical work we make use of a conditioned form of the AD test because we only take points that lie beyond a multidimensional quantile curve. This is due to the fact that for sampling we want to have $\sqrt{m}$ points at most to guarantee the convergence of $F_m \to F$ at the lower tail. So the integral in (6.2.4) is computed with respect to this condition only, and we will refer to that by calling AD the *conditional* Anderson Darling statistics.

The drawback of the AD tests is some smoothing effect due the cumulative character of distribution functions, especially at the upper tail. This is in contrast to the χ^2 test discussed in the beginning of this chapter, that can be interpreted to measure differences in the densities (for sufficient large number of classes and observations) and hence does not have such a smoothing effect due to accumulation. Since we are interested in testing models that are designed for flexible fit even in the tails this is a severe drawback for our purpose.

We therefore present a related version of the AD statistics that is tailor made for our situation and is more precise with respect to the upper tail than (6.2.4). The idea is to reformulate the Anderson Darling setting in terms of survival functions. For completeness, Equation (6.2.5) lists both forms (i.e. the supremum and integrated version) but for further work we will again restrict on the integrated form only.

$$\begin{aligned} SADS\left(\overline{F}_{\widehat{\theta}}; x^1, ..., x^m\right) &= \sqrt{n} \sup_{x \in S} \frac{\left|\overline{F}_m(x) - \overline{F}_{\widehat{\theta}}(x)\right|}{\sqrt{\overline{F}_{\widehat{\theta}}(x)\left(1 - \overline{F}_{\widehat{\theta}}(x)\right)}} \\ ADS\left(\overline{F}_{\widehat{\theta}}; x^1, ..., x^m\right) &= n \int_S \frac{\left(\overline{F}_m(x) - \overline{F}_{\widehat{\theta}}(x)\right)^2}{\overline{F}_{\widehat{\theta}}(x)\left(1 - \overline{F}_{\widehat{\theta}}(x)\right)} \cdot dF_{\widehat{\theta}}(x) \,, \end{aligned} \tag{6.2.5}$$

where $\overline{F}_\theta$ is the survival function of F_θ and

$$\overline{F}_m(x_1, ..., x_d) = \frac{1}{m} \sum_{j=1}^{m} \prod_{i=1}^{d} 1_{\left\{x_i > x_i^j\right\}} \ \forall x \in S. \tag{6.2.6}$$

A copula based notation of these statistics can be achieved by replacing $F_\theta(x_1, ..., x_d) = C_{\theta_C}(F_{\theta_1,1}(x_1), ..., F_{\theta_d,d}(x_d))$ and $f_\theta(x_1, ..., x_d) = \prod_{i=1}^d f_{\theta_i,i}(x_i)\, c_{\theta_C}(x_1, ..., x_d)$, where $\theta = (\theta_C, \theta_1, \ldots, \theta_d)$ and $F_{\theta_i,i}$ are the marginal distribution functions, and $f_{\theta_i,i}$ the marginal density functions respectively. To compare different parametric copula models, we assume the estimated uniform realizations by using the univariate empirical distribution functions, to be true observations of a d-dimensional uniform random variable $U \sim C$, distributed with an unknown copula C. So we have $F_\theta = C_\theta$, $f_\theta = c_\theta$ and $S = [0,1]^d$ in the above equations. Note that this is different from the assumption that the uniform mapping was done with no mistakes, i.e. that we had used the true marginal distributions.

Finally we introduce a measure that is closely related to the AD test. We write PD for the relative probability deviation, and PDS for the relative probability deviation of the survival function, respectively. PD measures the relative deviation of the empirical distribution to the hypothetical distribution at a point representing a region that tightly covers the lower tail. Analogously, PDS measures the relative deviation of the empirical survival distribution to the hypothetical survival distribution at a point representing a region that close cover the upper tail. More precisely: For $0 \leq p \leq 1$ we fix a point on the "ridge" of the copula density of the hypothetical model, $q = q(p) = CR^{\leftarrow}_{\theta_C}(p) \in [0,1]^n$ where we denote by $CR^{\leftarrow}_{\theta_C}$ the inverse of the "ridge" section of the copula function C_{θ_C}, that is $p = CR_{\theta_C}(q_1, \ldots, q_n) = C_{\theta_C}(q_1, \ldots, q_n)$ with $(q_1, \ldots, q_n)$ is a point on the "ridge" of the copula C_{θ_C}. For permutational symmetric copulas (i.e. $C(u,v) = C(v,u)$) the "ridge" is just the diagonal and hence $q_1 = q_2 = \ldots = q_n$. Analogously, we define $qs = qs(p) = CRS^{\leftarrow}_{\theta_C}(p) \in [0,1]^n$ with $p = CRS_{\theta_C}(qs_1, \ldots, qs_n) = \overline{C}_{\theta_C}(qs_1, \ldots, qs_n)$ with $(qs_1, \ldots, qs_n)$ is a point on the "ridge" of the copula C_{θ_C}. We now define PD and PDS by the following equations

$$\begin{aligned} PD &= \tfrac{1}{p}\left(F_m\left(F^{\leftarrow}_{\theta_1,1}(q_1), \ldots, F^{\leftarrow}_{\theta_n,n}(q_n)\right) - p\right) \\ PDS &= \tfrac{1}{p}\left(\overline{F}_m\left(\overline{F}^{\leftarrow}_{\theta_1,1}(qs_1), \ldots, \overline{F}^{\leftarrow}_{\theta_n,n}(qs_n)\right) - p\right) \end{aligned} \tag{6.2.7}$$

Note that the way we defined here PD and PDS makes only sense when the hypothetical model has a positive dependence structure, otherwise it is not clear what means "ridge".

Chapter 7

Estimating the Tail-Dependence Coefficient

This chapter is mainly based on Frahm et al. [49]. The concept of tail dependence describes the amount of asymptotic dependence in the lower-left-quadrant tail or upper-right-quadrant tail of a bivariate distribution. An important measure for tail dependence is given by the so-called tail-dependence coefficient. In this chapter various parametric and nonparametric estimators for the tail-dependence coefficient are introduced and investigated. After presenting the statistical properties we provide a detailed simulation study which compares and illustrates the advantages and disadvantages of all estimators under consideration.

7.1 Introduction

During the last decade, dependencies between financial assets have increased due to globalization effects and relaxed market regulation. However, traditional dependence measures like the linear correlation-coefficient are not always suited for a proper understanding of dependency in financial markets (Embrechts et al. [42]). In particular dependencies of extreme events like, e.g., large credit-portfolio losses, cause the need of new dependence measures to support powerful asset-allocation strategies. Recently, several empirical surveys like Junker and May [66], Kiesel and Schmidt [71], Ané and Kharoubi [3], and Malevergne and Sornette [83] exhibited that the concept of tail dependence is a useful tool to describe extreme-dependence phenomena in financial data. Moreover, they showed that especially during volatile and bear markets tail dependence plays an increasing role. Therefore robust and efficient estimation methods are required to measure tail dependence. The tail-dependence coefficient (TDC) represents the most common measure of tail dependence.

The main difficulty in estimating the TDC arises in the limited availability

of extreme data. Therefore we suggest several parametric and nonparametric estimators ascribed either to the entire data set or to a subset of extreme data. Regarding the latter, the extreme value theory is the natural choice for inferences on extreme events of random vectors or the tail behavior of probability distributions. Usually one approximates the tail of a probability distribution by an appropriate extreme value distribution. In the one-dimensional setting the class of extreme value distributions has a solely parametric representation, so it suffices to apply parametric estimation methods. By contrast, multi-dimensional extreme-value distributions are characterized by a parametric and a nonparametric component. This leads to more complicated estimation methods. Parametric estimation methods have the advantage of being efficient given that the model is true whereas nonparametric estimation avoids model-misclassifications. Therefore we provide both types of estimators either based on extreme-value theory or based on other distributional assumptions like elliptically contoured laws.

First of all this workout serves as a summary of various estimators for the tail-dependence coefficient. Statistical methods testing for tail dependence or tail independence go beyond the scope of this work, for that we refer the reader to Ledford and Tawn [75] and Draisma et al. [30].

After providing the necessary basic definitions in Section 7.2 we present various estimation methods for the tail-dependence coefficient in Section 7.3. Thereby we distinguish between parametric and nonparametric estimators as well as estimation methods based on the entire data sample or based on extreme data. We present various statistical and empirical properties of the estimators and discuss their prospective application fields. Section 7.3.1 and 7.3.2 consider the important class of elliptically contoured distributions and their ability to model tail dependence. Finally, Section 7.4 presents an extensive simulation study where we analyze and discuss the introduced estimators regarding their applicability. Further, we illustrate some of the statistical properties established in the previous sections.

7.2 Basic definitions

For the definition of tail dependence see Chapter 2, especially Equations (2.2.3) and (2.2.2).

Recently, there is an increasing interest in the class of elliptically contoured distributions for financial and actuarial data modelling (Bingham and Kiesel [8], Embrechts et al. [40], Bingham et al. [9]), which inherit appealing estimation, simulation and modelling properties. Thus an investigation of the tail dependence property for this class is justified. For the definition of elliptically contoured distributions (briefly: elliptical distributions) see 3.3.1.

In this chapter we will use the following notations. Let $E_d(\mu, \Sigma, \Phi)$ denote the class of elliptically contoured distributions with parameters μ (location), Σ (dispersion), and Φ (characteristic generator). The density function, if existent, of an elliptically contoured distribution has the following form:

$$f(x) = |\Sigma|^{-1/2} g((x-\mu)^T \Sigma^{-1}(x-\mu)), \quad x \in I\!R^d, \tag{7.2.1}$$

for some function $g : I\!R_0^+ \to I\!R_0^+$, which we call the *density generator*. Observe that the name "elliptically contoured" distribution is related to the elliptical contours of the latter density. For a more detailed treatment of elliptical distributions see the monographs of Fang et al. [45] or Cambanis et al. [17].

Here we recall the stochastic representation of Section 3.3, that turns out to be useful. Let X be a d-dimensional elliptically distributed random vector, i.e. $X \in E_d(\mu, \Sigma, \Phi)$, with parameters μ and symmetric positive semi-definite matrix $\Sigma \in I\!R^{d \times d}$, $\operatorname{rank}(\Sigma) = m,\ m \leq d$. Then

$$X \stackrel{d}{=} \mu + \mathcal{R}_m A' U^{(m)}, \tag{7.2.2}$$

where $A'A = \Sigma$ and the univariate random variable $\mathcal{R}_m \geq 0$ is independent of the m-dimensional random vector $U^{(m)}$. The random vector $U^{(m)}$ is uniformly distributed on the unit hypersphere $\mathbb{S}_2^{m-1}$ in $I\!R^m$. In detail, $\mathcal{R}_m$ represents the radial part and $U^{(m)}$ represents an angle of the corresponding elliptical random vector X. We will call $\mathcal{R}_m$ the *generating variate* of X.

However, elliptical copulas are restricted to rotationally symmetry. So they never can be appropriate if $\lambda_L \neq \lambda_U$. It is empirically stated that sets of financial data exist where usually there is only one TDC significant greater than zero. Such an asymmetry is observed e.g. in Junker and May [67] for oil related products (see Chapter 8) or in Junker et al. [69] for USD yields (see Chapter 11). For this reason, we also use the class of Archimedean copulas (see Chapter 4), on which the threshold approach in Section 7.3.2 is based.

An example is given by the Frank copula that is neither lower nor upper tail dependent, i.e. we have: $\lambda_L = \lambda_U = 0$. Frank [51] has shown that the Frank copula is the only rotationally symmetric Archimedean copula. So there is no Archimedean copula with $\lambda_L = \lambda_U > 0$ and $C(u, v) = C(1-u, 1-v)$ what is an important special case that is often realized in financial data. Take e.g. foreign exchange rates, then the hypothesis of the elliptically t-copula dependency with $\lambda > 0$ can not be rejected (see Breymann et al. [14] or Junker and May [67], and Chapter 8, respectively).

Hence it is important to consider both —methods based on elliptical and on Archimedean dependency.

Some of the following results are characterized by regularly varying functions. For a definition see (1.2.1). Moreover, one may generalize regular variation and the (tail)-index for multivariate random vectors and their distribution functions (see Schmidt [104] for more details).

7.3 Estimation methods for the tail-dependence coefficient

According to Equation (2.2.5) the tail-dependence coefficients λ_L and λ_U are copula-properties, thus we may write $\lambda_L = \lambda_L(C)$ ($\lambda_U = \lambda_U(C)$) or $\lambda_L = \lambda_L(\theta)$ ($\lambda_U = \lambda_U(\theta)$) if C is a member of a parametric copula-family C_θ with parameter vector θ.

7.3.1 Estimation from the entire sample

In this section the lower and upper TDC are estimated from the entire distribution or sample size.

The (semi-)parametric method

One approach to model tail dependence is to choose a *parametric* copula family C_θ allowing for tail dependence and being appropriate to describe the data. Then the lower and upper TDC is estimated by some functional estimator $\widehat{\lambda} := \widehat{\lambda}_n = \lambda(\widehat{\theta}_n)$. Starting from a random sample $X^{(1)}, \ldots, X^{(n)}$ with common distribution function F, we may first estimate the marginal distribution functions $F_1, \ldots, F_d$ in order to obtain approximate realizations of the copula C_θ:

$$\widehat{U} = \left(\widehat{U}_1, ..., \widehat{U}_d\right) = \left(\widehat{F}_1\left(X_1\right), ..., \widehat{F}_d\left(X_d\right)\right).$$

Parametric and nonparametric procedures are distinguished to estimate the marginal distribution functions $F_1, \ldots, F_d$. Regarding the parametric procedure we assume that d parameter vectors $m_1, ..., m_d$ determine the marginals $F_1, \ldots, F_d$. In the semiparametric case all marginal distribution functions are substituted by their empirical counterpart.

Another approach to model tail dependence is to choose a semi-parametric copula allowing for tail dependence and to estimate the upper and lower TDC via the limiting relation (2.2.5) and (2.2.4), respectively.

Parametric estimation – One-step estimate. Within the one-step estimate the parameters $\theta, m_1, \ldots, m_d$ are estimated in *one* step by maximum likelihood (ML) methods. Under the usual regularity conditions for asymptotic

ML-theory (see e.g. Casella and Berger [18], p. 516) the functional estimator $\widehat{\lambda} = \lambda(\widehat{\theta})$ represents an ML estimate of λ. Hence, consistency and asymptotic normality are given, i.e.,

$$\sqrt{n}\left(\widehat{\lambda}_n - \lambda\right) \xrightarrow{d} Z \sim N\left(0, \sigma^2_{\widehat{\lambda}}\right) \quad \text{as } n \to \infty.$$

In case $\lambda(\cdot)$ is a continuous differentiable function in $\theta = (\theta_1, \ldots, \theta_m)$ the standard error $\sigma_{\hat{\lambda}}$ can be approximated by the linear part of the corresponding Taylor expansion, i.e.,

$$\widehat{\sigma}^2_{\widehat{\lambda}} = \sum_{i=1}^{m} \left(\frac{\partial \lambda}{\partial \theta_i}\left(\widehat{\theta}_i\right)\right)^2 \sigma^2_{\widehat{\theta}_i}, \tag{7.3.1}$$

where $\sigma_{\hat{\theta}_i},\ i = 1, \ldots, m$ denotes the standard error of the corresponding ML estimate of $\theta_i,\ i = 1, \ldots, m$.

Parametric estimation - Two-step estimate. One of the advantages of the copula concept is the preservation of good statistical estimation properties while estimating the parameters $\theta, m_1, \ldots, m_d$ in *two* steps. Here we assume that the copula parameter θ belongs only to the copula function C_θ and is independent of the marginal distribution functions. Thus we first estimate the marginal parameters $m_1, \ldots, m_d$ and fit the marginal distributions $F_1, \ldots, F_d$ to the observed realization $(x_j^{(1)}, \ldots, x_j^{(n)}),\ j = 1, \ldots, d$, of the random sample $X^{(1)}, \ldots, X^{(n)}$.

In the second step the copula parameter θ is estimated from the approximate realizations $(\hat{u}_j^{(1)}, \ldots, \hat{u}_j^{(n)}),\ j = 1, \ldots, d$, with $\hat{u}_j^{(i)} := F_{\hat{m}_j}(x_j^{(i)}),\ j = 1, \ldots, d$, $i = 1, \ldots, n$. Denote $\hat{u} := (\hat{u}^{(1)}, \ldots, \hat{u}^{(n)})$ as the matrix of normalized observations. Joe [62] Chapter 10 provides the asymptotic distribution and asymptotic covariance matrix. Note, this method is also referred to as pseudo maximum likelihood. Joe [62] also carried out some simulation studies that stated that there is not much difference in estimating the TDC parametric either in two or in one step. However, the two-step approach simplifies the numerics of the parameter estimation magnificently. Even for high-dimensional data sets the TDC estimation becomes a relief.

Semi-parametric estimation - Two-step estimate. To avoid misidentifications of the marginal distribution functions the marginals $F_1, \ldots, F_d$ are now estimated nonparametrically by their empirical distribution functions. In a second step the copula parameter θ is estimated from the normalized realizations $\hat{u}$ via an ML-estimator, i.e., $\hat{\theta} = \hat{\theta}(\hat{u})$. As shown e.g. by Genest et al. [53], the resulting estimator is consistent and asymptotically normal distributed, i.e., $\sqrt{n}(\theta - \widehat{\theta}_n(\widehat{u})) \xrightarrow{d} Z \sim N(0, \Sigma)$. For the covariance matrix and its estimate from a finite data sample, we refer to Genest et al. [53]. In case $\lambda(\cdot)$ is a smooth function the estimator $\hat{\lambda} = \lambda(\hat{\theta})$ turns out to be also consistent and asymptotic normal. In Section 7.4 our simulation study carries out that the relative efficiency, which

is defined as the ratio of mean-squared errors (MSE) of the semiparametric two-step ML estimator to the one-step parametric ML estimator, comes close to 1. Thus there is not much difference in estimating the TDC either with empirical marginals in two steps or purely parametric in one step. The latter semiparametric two-step estimator is favorably used in many recent empirical investigations, e.g., Ané and Kharoubi [3], and Junker and May [67].
On the other hand, marginal distributions are often better known than the dependence function. For instance for many financial models it is appropriate to assume Gaussian marginal innovations, whereas the dependence structure does not seem to be obvious at all. Therefore another semi-parametric method is provided by specifying the parametric marginal distributions in a first step and estimating the related copula via some nonparametric estimator in a second step. For instance we utilize an ML-estimator for the parameters $m_1, \ldots, m_d$ of the marginal distributions $F_1, \ldots, F_d$ at a first stage, and at a second stage the copula C is estimated via $\hat{C}_{\hat{m}_1,\ldots,\hat{m}_d} = F_n(F^{\leftarrow}_{\hat{m}_1}, \ldots, F^{\leftarrow}_{\hat{m}_d})$, where F_n denotes the multivariate empirical distribution function of F. Under suitable regularity conditions the process

$$\sqrt{n}(\hat{C}_{\hat{m}_1,\ldots,\hat{m}_d} - C) = \sqrt{n}(F_n(F^{\leftarrow}_{\hat{m}_1}, \ldots, F^{\leftarrow}_{\hat{m}_d}) - F(F^{\leftarrow}_1, \ldots, F^{\leftarrow}_d))$$

converges to a Gaussian limit (cf. Fermanian et al. [47]).

The parametric method for elliptically contoured distributions

In this section we assume that the observed data follow an elliptically-contoured random sample $X, X^{(1)}, \ldots, X^{(n)}$. If not stated otherwise we assume $X \in E_d(\mu, \Sigma, \Phi)$ with positive-definite dispersion matrix Σ. Then a nonsingular matrix $\sqrt{\Sigma}$ exists, i.e. the Cholesky root of Σ with inverse $\sqrt{\Sigma}^{-1}$. According to (7.2.2) the random vector X possesses the stochastic representation

$$X \stackrel{d}{=} \mu + \mathcal{R}_d \sqrt{\Sigma}' U^{(d)}. \tag{7.3.2}$$

Schmidt [104] shows essentially that the TDC of two components of an elliptical random vector X, say X_i and X_j, depends only on the tail index of the corresponding density-generator or generating distribution function, if they are regularly varying (see (1.2.1)), and the correlation-coefficient ρ_{ij} between X_i and X_j. Observe that the correlation-coefficient ρ_{ij} may not exist; in that case we refer to ρ_{ij} as some dependence parameter. However, if not stated otherwise ρ_{ij} is always denoted as the correlation-coefficient for notational convenience. For elliptical random vectors X a regularly varying density generator is equivalent to regular variation of X (Schmidt [104]). Thus in the following we always refer to

the tail index of X. Moreover (see Frahm et al. [50]), the TDC corresponds to

$$\lambda_{ij} = 2 \cdot \overline{t}_{\zeta+1}\left(\sqrt{\zeta+1} \cdot \sqrt{\frac{1-\rho_{ij}}{1+\rho_{ij}}}\right), \qquad \zeta > 0, \tag{7.3.3}$$

where $\overline{t}_{\zeta+1}$ denotes the survival function of the univariate Student's t-distribution with $\zeta + 1$ degrees of freedom. Thus given estimates for both, the tail index ζ and the correlation matrix ρ, the TDC between the random components X_i and X_j can be estimated by formula (7.3.3). Note that the tail index of the generating variate $\mathcal{R}_d$ corresponds to the tail index ζ of the elliptical random vector X (Schmidt [104]). In the following, a method for estimating both the distribution function of the generating variate $\mathcal{R}_d$ and the correlation matrix ρ will be presented. Once given an estimate for the generating distribution function, an appropriate model for $\mathcal{R}_d$ can be fitted.

Due to the stochastic representation (7.3.2) the inverse relation

$$\mathcal{R}_d U^{(d)} \stackrel{d}{=} \sqrt{\Sigma}'^{-1}(X-\mu)$$

holds. Hence

$$\delta\mathcal{O}\mathcal{R}_d U^{(d)} \stackrel{d}{=} \delta\mathcal{O}\sqrt{\Sigma}'^{-1}(X-\mu) = \Gamma(X-\mu), \qquad \delta > 0,\ \Gamma := \delta\mathcal{O}\sqrt{\Sigma}'^{-1}, \tag{7.3.4}$$

where δ denotes an arbitrary scaling constant and $\mathcal{O}$ may be any orthogonal matrix.

Given the location parameter vector μ and the inversion matrix Γ we can observe the realizations of the spherical random vector $\Gamma(X-\mu)$ and sieve the realizations of the rescaled generating variate $\delta\mathcal{R}$, since

$$\|\Gamma(X-\mu)\|_2 \stackrel{d}{=} \left\|\delta\mathcal{O}\mathcal{R}_d U^{(d)}\right\|_2 = \delta\mathcal{R}_d \cdot \left\|\mathcal{O}U^{(d)}\right\|_2 = \delta\mathcal{R}_d =: \mathcal{R}_d^*. \tag{7.3.5}$$

Lindskog [76] shows that the heavier the tails of the one-dimensional marginals of X the less is Pearson's standard covariance estimator suitable for estimating the dispersion matrix Σ and therefore also the correlation matrix ρ. The covariance-matrix of X may even be infinite if the tail index ζ of X is smaller than 2. In that case Pearson's covariance estimator is vulnerable to contaminated and outlier data. Moreover, even the sample-mean vector is also no appropriate estimator for μ in case $\zeta < 1$ as the mean does not exist. Hence robust estimation methods for both, the location parameter μ and the inversion matrix Γ are required. This task is tackled by the following.

Since elliptical distributions are symmetric, the median vector $\mu_{0.5}$ corresponds to the location μ. Further, under some technical condition given in Büning and Trenkler [16], p. 61, the sample-median $\overline{x}_{0.5}$ is a consistent estimator for μ.

So $\widehat{\mu} = \overline{x}_{0.5}$ is an appropriate estimator for μ.
For the estimation of Γ we will follow an approach introduced by Frahm and Junker [48] (see Chapter 12). If $\mu = 0$, then

$$\frac{\sqrt{\Sigma}U}{\left\|\sqrt{\Sigma}U\right\|_2} \stackrel{d}{=} \frac{\mathcal{R}_d\sqrt{\Sigma}U}{\left\|\mathcal{R}_d\sqrt{\Sigma}U\right\|_2} \stackrel{d}{=} \frac{X}{\|X\|_2}$$

Now Γ can be estimated via maximum likelihood, where the log-likelihood function is given by

$$\log\left(\mathcal{L}\left(\Gamma\,;s_{\cdot 1},...,s_{\cdot n}\right)\right) \propto n\cdot\log\left(\det\left(\Gamma'\Gamma\right)\right) - d\cdot\sum_{j=1}^{n}\log\left(s'_{\cdot j}\Gamma'\Gamma s_{\cdot j}\right), \qquad (7.3.6)$$

with

$$s_{\cdot j} := \frac{x_{\cdot j}}{\|x_{\cdot j}\|_2}, \qquad j = 1,...,n.$$

Since the log-likelihood function (7.3.6) is invariant under linear transformations of Γ, a constraint must be considered to get a unique solution for the maximum likelihood. Here we will use

$$\Gamma_{1,1} = \frac{1}{\prod_{i=2}^{d}\Gamma_{ii}},$$

so that the determinant of Γ is always 1. If furthermore the constraint $\Gamma_{ii} \neq 0$ $(i = 1,...,d)$ is met in each iteration, the existence and positive definiteness of $\Sigma^* = (\Gamma'\Gamma)^{-1}$ is guaranteed.
Let $\Gamma^\star$ be the ML-estimator of Γ, i.e. $\Gamma^\star := \arg\max\log\left(\mathcal{L}\left(\Gamma\,;\cdot\right)\right)$. The ML-estimator of the dispersion matrix Σ corresponds (up to scale) to

$$\widehat{\Sigma}_S := \left(\Gamma^{*\prime}\Gamma^*\right)^{-1} = \frac{1}{\delta^2}\hat{\Sigma}, \qquad (7.3.7)$$

where $\widehat{\Sigma}$ is the virtual estimator of Σ. Let the correlation matrix of X be defined as $\rho := \left[\Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}\,\right]$ even if the second moment of the generating variate $\mathcal{R}$ is infinite. Note that $\Sigma_{ii} > 0$ for every $i = 1,\ldots,d$ since Σ is positive definite. The correlation matrix can be estimated by

$$\hat{\rho}_S := \left[\frac{\widehat{\Sigma}_{S,ij}}{\sqrt{\widehat{\Sigma}_{S,ii}\cdot\widehat{\Sigma}_{S,jj}}}\right], \qquad (7.3.8)$$

robustly. We call both $\widehat{\Sigma}_S$ and $\widehat{\rho}_S$ 'spectral estimators', respectively.
For the theoretical details of this procedure we refer to Frahm and Junker [48].

There are a lot of other robust estimation methods for the correlation matrix of elliptical random vectors. For an overview see Lindskog [76], for instance. The minimum volume estimator, for instance, minimizes the volume of the ellipsoid generated by the Mahalanobis distance, i.e.

$$\left\{ x \in I\!R^d : \sqrt{(x-\widehat{\mu})' \widehat{\Sigma}^{-1} (x-\widehat{\mu})} \leq \mathcal{M} \right\},$$

covering a certain number of data points $m \leq n$. The problem is to find the "right" m (or equivalent the right positive "threshold" $\mathcal{M}$). If the number of covered data points is too large extreme values lead to high variance, and if m is too small the same results due to data reduction. Although a lot of robust estimation methods are based on the Mahalanobis distance this approach is very sensitive to outliers. But separating the correlation structure and the radial part of elliptically distributed random vectors like presented before enables us to eliminate extremes without reducing data and therefore to obtain more efficient correlation estimators. Notice that the latter correlation estimate is always positive definite; this is an important advantage as erroneous transformations can be avoided.

Due to equation (7.3.5) we are able to estimate the realizations of the rescaled generating variate $\mathcal{R}_d^* = \delta \mathcal{R}_d$,

$$\widehat{r}_i^* := \left\| \widehat{\Gamma} (x_i - \widehat{\mu}) \right\|_2, \qquad i = 1, ..., n.$$

Indeed the random variable $\mathcal{R}_d^*$ has not the same distribution as $\mathcal{R}_d$, but the tail index of $\mathcal{R}_d$ is invariant under scale transformations. So it is sufficient to fit an appropriate univariate model F_θ to the realizations $\widehat{r}_i^*$ $(i = 1, ..., n)$. Since the tail index is generally a function of θ, determined by the law F_θ, an appropriate estimate for ζ can be derived from $\zeta(\widehat{\theta})$.

7.3.2 Estimation from the distribution's tail or extreme sample

The threshold method

Parametric threshold methods. As in Section 7.3.1 we may choose a parametric copula family allowing for tail dependence and estimate the TDC λ via a function of the inferred distribution parameters. Of course we can distinguish here again between the use of empirical or parametric marginal models. However, we only believe the underlying model to be appropriate for the joint extremes and not for the entire data sample. Hence we choose a lower and upper 2-dimensional threshold and investigate only data below the threshold (thl) and

above the threshold (thu), respectively. The lower and upper TDC are achieved by the estimated parameter vector estimates $\widehat{\theta}_{thl}$ and $\widehat{\theta}_{thu}$ respectively.

The Gumbel copula as an extreme-value copula (see Joe [62]) seems to be one natural choice. The truncated Gumbel copula for the upper tail, and the truncated form for the lower tail respectively, are given by equation (7.3.9).

$$\begin{aligned} C_G^{thu}(u,v) &= \tfrac{C_G(u,v)-C_G(u,thu)-C_G(thu,v)+C_G(thu,thu)}{1-2thu+C_G(thu,thu)}, \quad \forall\,(u,v) \geq (thu,thu) \\ C_G^{thl}(u,v) &= \tfrac{C_G(1-u,1-v)-C_G(1-u,1-thl)-C_G(1-thl,1-v)+C_G(1-thl,1-thl)}{1-2(1-thl)+C_G(1-thl,1-thl)}, \\ & \quad \forall\,(u,v) \leq (thl,thl) \end{aligned} \tag{7.3.9}$$

A parametric threshold method for Archimedean copulas For bivariate Archimedean copulas, Juri and Wüthrich [70] have shown a useful limit behavior of the bivariate excesses. Denote by F_{th} the conditional distribution function

$$F_{th} := \mathcal{P}\left[U \leq u \mid U \leq th, V \leq th\right], \; 0 \leq u \leq 1.$$

Then the extreme tail dependence copula of a copula C, with threshold th is given by

$$C_{th}(u,v) = \mathcal{P}\left[U \leq F_{th}^{\leftarrow}(u), V \leq F_{th}^{\leftarrow}(v) \mid U \leq th, V \leq th\right]. \tag{7.3.10}$$

If C is now an Archimedean copula generated by a regularly varying differentiable generator $\varphi_C \in \mathcal{R}_{-\alpha}$, $0 < \alpha < \infty$, i.e. $\lim_{t\to 0^+} \frac{\varphi_C(kt)}{\varphi_C(kt)} = k^{-\alpha}$, then

$$\lim_{th\to 0^+} C_{th}(u,v) = C_{Cl}(u,v), \tag{7.3.11}$$

for all $0 \leq u,v \leq 1$, where C_{Cl} is the Clayton copula with generator $\varphi_{Cl} = \frac{1}{\alpha}(t^{-\alpha}-1)$, $\alpha > 0$ and hence given by

$$C_{Cl}(u,v) = \varphi_{Cl}^{\leftarrow}(\varphi_{Cl}(u) + \varphi_{Cl}(v)) = \left(u^{-\alpha} + v^{-\alpha}\right)^{-\frac{1}{\alpha}}.$$

Now we can achieve the lower and upper tail dependence coefficient by fitting a Clayton copula to the pseudo observations of $(F_{th=thl}(u), F_{th=thl}(v))$, $u,v \in [0,thl]^2$ and $(F_{th=1-thu}(1-u), F_{th=1-thu}(1-v))$, $u,v \in [thu,1]^2$ respectively. Let $\hat{\alpha}_{thl}$ and $\hat{\alpha}_{thu}$ be the estimates of the Clayton copula parameter, then

$$\begin{aligned} \hat{\lambda}_{thl}(\hat{\alpha}_{thl}) &= 2^{-\frac{1}{\hat{\alpha}_{thl}}} \\ \hat{\lambda}_{thu}(\hat{\alpha}_{thu}) &= 2^{-\frac{1}{\hat{\alpha}_{thu}}} \end{aligned} \tag{7.3.12}$$

Of course one can utilize the above approach, probably ignoring that C has to be Archimedean, and try to fit other parametric copulas allowing for lower tail

dependence to (7.3.10). A possible candidate is the survival Gumbel copula given by

$$\widetilde{C}_{G,\delta}(u,v) = u + v - 1 + \exp\left(-\left((-\log(u))^{\delta} + (-\log(u))^{\delta}\right)^{\frac{1}{\delta}}\right),\ \delta \geq 1.$$

The TDC estimates in analogy to (7.3.12) are

$$\begin{aligned} \hat{\lambda}_{thl}\left(\hat{\delta}_{thl}\right) &= 2 - 2^{\frac{1}{\hat{\delta}_{thl}}} \\ \hat{\lambda}_{thu}\left(\hat{\delta}_{thu}\right) &= 2 - 2^{\frac{1}{\hat{\delta}_{thu}}} \end{aligned} \tag{7.3.13}$$

In Section 7.4.1 we check the finite sample properties of the limit behavior (7.3.11), i.e. what happens if it is not possible to choose an arbitrary small threshold. Furthermore we check out the robustness of the result if applied to a non Archimedean dependence structured sample.

A test for TDC asymmetry. In case of an a priori assumption of radial symmetry and hence $\lambda_L = \lambda_U$, one only has to determine one threshold th and can use all data points below th and above $1-th$. For example we can mirror the lower and upper joint extreme-values of the copula by $(u_1, u_2) \to (1-u_1, 1-u_2)$ to the upper right and lower left corner and use the union set of original and mirrored data. Now we introduce a heuristic for TDC asymmetry, i.e. $\lambda_l \neq \lambda_u$, based on Akaike's (AIC) and the Bayesian (BIC) information criterion. Let (u_{thl}, v_{thl}) be the observations below thl and (u_{thu}, v_{thu}) the ones above thu. Further we define

$$\begin{aligned} L(\theta, n, u, v) &:= \sum_{i=1}^{n} \log\left(c_{C^{\star}(\theta)}\right) \\ L_{asy} &:= \frac{1}{2}\left(L\left(\hat{\theta}_{thl}, Nl, u_{thl}, v_{thl}\right) + L\left(\hat{\theta}_{thu}, Nu, 1-u_{thu}, 1-v_{thu}\right)\right) \\ L_{sym} &:= \frac{1}{4}\left(2L_{asy} + L\left(\hat{\theta}_{thu}, Nl, u_{thl}, v_{thl}\right) + L\left(\hat{\theta}_{thl}, Nu, 1-u_{thu}, 1-v_{thu}\right)\right), \end{aligned}$$

where Nl is the number of observations below thl, Nu is the number of observations above thu and $c_{C^{\star}(\theta)}$ is the density of the copula with parameter vector θ used to model the excesses or the limit copula (7.3.11) in the later approach.

Under the assumption of symmetric TDC's, $\hat{\theta}_{thl} = \hat{\theta}_{thu}$ should hold for $Nl \to \infty$, $Nu \to \infty$ and $thl \to 0$, $thu \to 0$. Hence $L_{sym} = L_{asy}$. For finite samples one still should have $L_{sym} \approx L_{asy}$ for sufficiently small (big) thresholds under symmetric TDC's. However, an asymmetric TDC assumption costs one more parameter for the description of $\lambda_l \neq \lambda_u$, what is the motivation for the

following AIC and BIC based definitions

$$\begin{aligned} AIC_{sym} &= -2L_{sym} + 2p \\ AIC_{asy} &= -2L_{asy} + 4p \\ BIC_{sym} &= -2L_{sym} + \frac{p}{2}\left(\log\left(Nl\right) + \log\left(Nu\right)\right) \\ BIC_{asy} &= -2L_{asy} + p\left(\log\left(Nl\right) + \log\left(Nu\right)\right), \end{aligned}$$

where p is the number of parameters of $C_{C^\star(\theta)}$. We will call a sample TDC asymmetric if $AIC_{asy} < AIC_{sym}$ and $BIC_{asy} < BIC_{sym}$. If a sample is not TDC asymmetric we call it TDC symmetric and can take

$$\hat{\lambda}_{thl,thu} = \frac{1}{2}\left(\hat{\lambda}_{thl} + \hat{\lambda}_{thu}\right)$$

for the reduction of the estimators volatility.

Non- and semiparametric threshold methods. At the end of Section 7.3.1 we outlined that it is possible to use parametric marginals and the empirical copula function to model multidimensional data. However, no matter how the estimates for the pseudo observations $\hat{u}$ were achieved, using the empirical copula to estimate λ_L leads to an approximation of the limit (2.2.5) by

$$\hat{\lambda}_{thl} = \frac{\hat{C}(thl, thl)}{thl}, \tag{7.3.14}$$

where $\hat{C}$ is the empirical copula. Of course we wish to choose thl as small as possible, but to guarantee $\hat{C}(thl, thl) \to C(thl, thl)$ there have to be enough data below thl. To estimate λ_U we mirror the data to $1-u$ and estimate $\widehat{\widetilde{C}}$. Now we set $thl = 1 - thu$ in Equation (7.3.14) and get the estimate $\hat{\lambda}_{thu} = \frac{\widehat{\widetilde{C}}(1-thu,1-thu)}{1-thu}$.

The parametric method for elliptical distributions

Recall from Section 7.3.1, that for regularly varying elliptical random vectors X the tail-dependence coefficient λ was given via a function of the tail index ζ which has been estimated from the entire data-sample. Alternatively, the tail index ζ is now estimated from extreme value sample realizations. The estimation reduces again to a one-dimensional problem as the random variables $\mathcal{R}_d$, $||X||_2$ possesses the same tail index ζ (Schmidt [104]). Hence we estimate the tail index of $\mathcal{R}_d$ with standard methods from univariate extreme value theory utilizing the extreme-value realizations of $||X||_2$.

In Section 7.3.1 a method for separating the correlation structure and the random variate $\mathcal{R}_d$ of an elliptically distributed random vector was introduced; to be precise, we obtained a rescaled random variate $\mathcal{R}_d^*$. Thus, alternatively,

the tail index ζ can be estimated from $\mathcal{R}_d^*$ as the rescaling has non influence on the tail index ζ. However, assuming a non trivial dispersion matrix (i.e. $\Sigma \neq \mathbf{I}$), $||X||_2$ can also be interpreted as a rescaled version of $\mathcal{R}_d$. But in contrast to $\mathcal{R}_d^*$ the scaling factor is not a constant, i.e. two independent realisations of $||X||_2$ have almost surely a different scale. This might have some influence regarding the finite sample estimation, what we have checked in Section 7.4.1.

Different methods for univariate tail index estimation are discussed e.g. in Embrechts et al. [38]. In contrast to the peaks over threshold and the block maxima approach, estimation methods based on upper order statistics (e.g. the Hill-estimator) do not require an estimate for the scale parameter of the generalized Pareto distribution or the generalized extreme value distribution, respectively. For our purpose a Hill-type estimator with optimal sample fraction proposed by Drees and Kaufmann [31] turns out to be satisfying.

The nonparametric methods

In this section we drop any assumption on the marginal distributions and the copula function, and provide two nonparametric estimators for the lower and the upper TDC, respectively. Further, results concerning the strong consistency and the asymptotic normality of the introduced estimators are provided. Nonparametric estimation methods prevent wrong specifications of the underlying distribution or copula and thus avoid misidentifications of the upper or lower TDC. Suppose again that $(X_1, X_2), (X_1^{(1)}, X_2^{(1)}), \ldots, (X_1^{(n)}, X_2^{(n)})$ are iid bivariate random vectors with distribution function F and copula C.

The following two nonparametric estimators for the lower and the upper TDC, λ_L and λ_U, are proposed. Let C_n be the empirical copula defined by

$$(u, v) \mapsto C_n(u, v) = F_n(F_{1n}^{\leftarrow}(u), F_{2n}^{\leftarrow}(v)), \tag{7.3.15}$$

with F_n, F_{1n} and F_{2n} denoting the empirical distribution functions corresponding to F, F_1 and F_2. Let $R_{1n}^{(j)}$ and $R_{2n}^{(j)}$ be the ranks of $X_1^{(j)}$ and $X_2^{(j)}$, $j = 1, \ldots, n$, respectively. The first set of estimators is motivated by the TDC-definition (2.2.2):

$$\begin{aligned} \hat{\lambda}_{U,n}(k) &= \frac{n}{k} \cdot C_n\Big(\Big(1 - \frac{k}{n}, 1\Big] \times \Big(1 - \frac{k}{n}, 1\Big]\Big) \\ &= \frac{1}{k} \cdot \sum_{j=1}^{n} \mathbf{1}_{\{R_{1n}^{(j)} > n-k \,\wedge\, R_{2n}^{(j)} > n-k\}}, \qquad 1 \leq k \leq n, \end{aligned} \tag{7.3.16}$$

where $C_n((a, b] \times (c, d])$ denotes the empirical copula-measure on the interval

$(a,b] \times (c,d]$, and thus the estimator writes

$$\begin{aligned} \hat{\lambda}_{L,n}(k) &= \frac{n}{k} \cdot C_n\Big(\frac{k}{n}, \frac{k}{n}\Big) \\ &= \frac{1}{k} \cdot \sum_{j=1}^{n} \mathbf{1}_{\{R_{1n}^{(j)} \le k \,\wedge\, R_{2n}^{(j)} \le k\}} \qquad 1 \le k \le n, \end{aligned} \tag{7.3.17}$$

with $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$.

The second set of estimators is motivated by the TDC-representation (2.2.4) and by techniques from extreme value theory. In particular, the relationship between the so-called dependence function of a bivariate extreme value distribution (see Huang [59] for a definition) and the tail-dependence coefficient is utilized. Notice that the limiting relation (2.2.4) can be rewritten in the form

$$\begin{aligned} \lambda_U &= \lim_{v \to 1^-} \tfrac{1-2v+C(v,v)}{1-v} = \lim_{v \to 1^-} \tfrac{2-2v}{1-v} + \lim_{v \to 1^-} \tfrac{-1+C(v,v)}{1-v} \\ &= 2 - \lim_{v \to 1^-} \tfrac{1-C(v,v)}{1-v}. \end{aligned} \tag{7.3.18}$$

Therefore we propose

$$\begin{aligned} \hat{\lambda}_{U,n}^{EVT}(k) &= 2 - \tfrac{n}{k} \cdot \Big(1 - C_n\Big(1 - \tfrac{k}{n}, 1 - \tfrac{k}{n}\Big)\Big) \\ &= 2 - \tfrac{1}{k} \cdot \textstyle\sum_{j=1}^{n} \mathbf{1}_{\{R_{1n}^{(j)} > n-k \,\vee\, R_{2n}^{(j)} > n-k\}}, \qquad 1 \le k \le n, \end{aligned} \tag{7.3.19}$$

with $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$. The estimator $\hat{\lambda}_{L,n}^{EVT}$ is defined analogously. The optimal choice of k is related to the usual variance-bias problem known from tail index estimations of regular varying distribution functions, and will be addressed in a forthcoming work. The following strong consistency and asymptotic normality properties of the introduced estimators are proven in Schmidt and Stadtmüller [106]. The proof of asymptotic normality is accomplished in two steps. In the first step the marginals F_1 and F_2 are assumed to be known and an asymptotic normality result is derived. In the second step the marginal distribution functions F_1 and F_2 are assumed to be unknown and the asymptotic result is proven by utilizing a particular version of the functional delta method. See also Schmidt [104] for related results in the case of elliptically contoured distributions and elliptical copulas. In this context the result of asymptotic normality is only given for the TDC-estimator $\hat{\lambda}_{L,n}$.

Let F be a bivariate distribution function with continuous marginal distribution functions F_1 and F_2. Under some technical conditions regarding the local behavior of the convergence (2.2.2) towards the TDC, like a second order condition (see Schmidt and Stadtmüller [106] for more details), and $\lambda_L > 0$ we obtain

$$\sqrt{k}\big\{\hat{\lambda}_{L,n} - \lambda_L\big\} \xrightarrow{d} \mathbb{G}_{\lambda_L},$$

with $\mathbb{G}_{\lambda_L}$ being centered and normally distributed, i.e. $\mathbb{G}_{\lambda_L} \sim N(0, \sigma_L^2)$ with

$$\begin{aligned}\sigma_L^2 = \lambda_L &+ \Big(\frac{\partial}{\partial x}\Lambda_L(1,1)\Big)^2 + \Big(\frac{\partial}{\partial y}\Lambda_L(1,1)\Big)^2 \\ &+ 2\lambda_L\Big(\Big(\frac{\partial}{\partial x}\Lambda_L(1,1) - 1\Big)\Big(\frac{\partial}{\partial y}\Lambda_L(1,1) - 1\Big) - 1\Big),\end{aligned}$$

where the function $(x, y) \mapsto \Lambda_L(x, y)$ is specified in Schmidt and Stadtmüller [106]. Further strong consistency holds if $k/\log\log n \to \infty$ as $n \to \infty$. Similar results hold for $\hat{\lambda}_{U,n}$ and $\hat{\lambda}_{U,n}^{EVT}$.

Another nonparametric estimator is motivated by the limiting relation (7.3.18). If the diagonal section $v \mapsto C(v, v)$ is differentiable in the interval $(1 - \varepsilon, 1)$ for some $\varepsilon > 0$, then

$$\lim_{v\to 1^-} \frac{1 - C(v,v)}{1 - v} = \lim_{v\to 1^-} \frac{dC(v,v)}{dv} = \lim_{v\to 1^-} \frac{\log C(v,v)}{\log v} \tag{7.3.20}$$

holds (see Coles et al. [22]). Therefore we can provide an estimator for the upper tail-dependence coefficient by

$$\begin{aligned}\widehat{\lambda}_{U,n}^{LOG}(k) &= 2 - \frac{\log C_n(\frac{n-k}{n}, \frac{n-k}{n})}{\log \frac{n-k}{n}} \\ &= 2 - \frac{\log\Big(\frac{1}{n}\cdot\sum_{j=1}^n \mathbf{1}_{\{R_{1n}^{(j)} \le n-k \,\wedge\, R_{2n}^{(j)} \le n-k\}}\Big)}{\log \frac{n-k}{n}}, \quad 1 \le k < n,\end{aligned} \tag{7.3.21}$$

with $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$. If the bivariate data is either comonotonic or stochastically independent this estimator is appropriate for all k, because in the former case $C(v, v) = v$ and thus

$$\lambda_U = 2 - \frac{\log C(v,v)}{\log v} = 2 - \frac{\log v}{\log v} = 1, \qquad \forall\, v \in (0,1), \tag{7.3.22}$$

whereas in the latter case $C(v, v) = v^2$ and hence

$$\lambda_U = 2 - \frac{\log C(v,v)}{\log v} = 2 - \frac{\log v^2}{\log v} = 2 - 2\cdot\frac{\log v}{\log v} = 0, \qquad \forall\, v \in (0,1). \tag{7.3.23}$$

That is to say the tail-dependence coefficient is given exactly at each point on the logarithm of the diagonal section if it follows a power law which indeed is the case for comonotonicity and independence. Since this is not true for the EVT-estimator defined by (7.3.19) we favor the log-estimator and compare it with the one defined by (7.3.16). Notice that if $C(v, v) = v^\alpha$, for all $v \in (1-\varepsilon, 1)$, $\alpha \in [1, 2]$ then $\lambda_U = 2 - \alpha$.

A similar result for the lower tail-dependence coefficient is motivated by Equation (2.2.5). It can be shown that

$$\lim_{v\to 0^+} \frac{dC(v,v)}{dv} = 2 - \lim_{v\to 0^+} \frac{\log\left(1-2v+C(v,v)\right)}{\log\left(1-v\right)}, \tag{7.3.24}$$

if the diagonal section $v \mapsto C(v,v)$ is differentiable in the interval $(0,\varepsilon)$ for some $\varepsilon > 0$. Therefore an estimator for the lower tail-dependence coefficient λ_L is provided by

$$\begin{aligned} \widehat{\lambda}_{L,n}^{LOG}(k) &= 2 - \frac{\log C_n\Big(\left(\frac{k}{n},1\right] \times \left(\frac{k}{n},1\right]\Big)}{\log \frac{n-k}{n}} \\ &= 2 - \frac{\log \frac{1}{n} \cdot \sum_{j=1}^{n} \mathbf{1}_{\{R_{1n}^{(j)}>k \,\wedge\, R_{2n}^{(j)}>k\}}}{\log \frac{n-k}{n}}, \quad 1 \le k < n, \end{aligned} \tag{7.3.25}$$

with $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$. Also if $1-2v+C(v,v) = (1-v)^\alpha$, for all $v \in (0,\varepsilon)$, $\alpha \in [1,2]$ then $\lambda_L = 2-\alpha$. Moreover, for the identity copula ($\alpha = 1$) and the product copula ($\alpha = 2$) the latter estimator fulfills the same stability properties as derived for the log-estimator (7.3.21).

In the pre-simulation Section 7.4.1 we have checked the finite sample properties of the direct estimation over the slope of the secant and the log-estimator. Regarding the statistical properties of the log-estimators (7.3.21) and (7.3.25), strong consistency holds if $k/\log\log n \to \infty$ as $n \to \infty$. This can be deduced from the strong consistency result of the nonparametric estimators $\hat{\lambda}_{U,n}$ and $\hat{\lambda}_{L,n}$. Asymptotic normality will be discussed in a forthcoming work.

7.3.3 A final remark about pitfalls

We would like to point out, that one can not be completely sure from a finite iid sample observation $\{x^1,\ldots,x^n\}$ of some d-dimensional random vector $X = (X_1,\ldots,X_d)'$ whether X is tail dependent or not. As for tail index estimation one can find certain model specifications which produce finite samples pretending fat tails, even for large sample sizes. For example Figure 7.1 shows the scatterplot of realizations from a copula corresponding to a mixture distribution of two different bivariate Gaussian distributions, namely:

$$NM = \frac{1}{2}N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix}\right) + \frac{1}{2}N\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1.44 & 0.72 \\ 0.72 & 1.44 \end{bmatrix}\right).$$

On the first glance the scatterplot reveals upper tail dependence although the NM distribution is tail-independent. The message is that one must be careful by

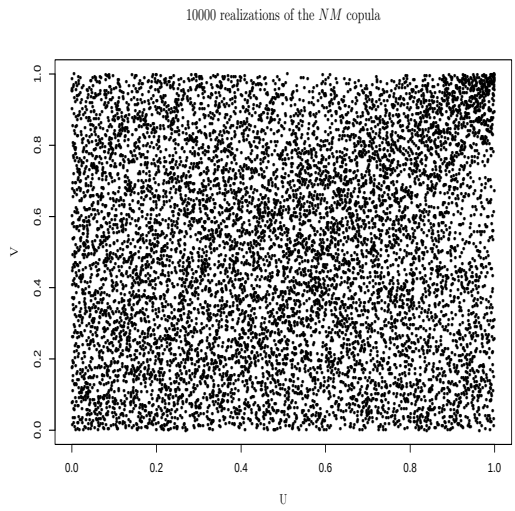

Figure 7.1: Scatterplot of simulated data from a normal mixture distribution NM.

inferring tail dependence from a finite random sample. The best way to protect against misidentifications is the application of several estimators, test or plots to the same data set.

The use of parametric marginal models instead of empirical ones contains the risk of misspecification and therefore the pitfall of wrong uniform mapping and wrong interpretation of the dependence structure. As example we had simulated 3000 realizations of a 2-dimensional random vector distributed like $H(x,y) = C_{G,\vartheta=2}(t_{\nu=3}(x), t_{\nu=3}(y))$, where t_ν denotes the univariate standard t-distribution with ν degrees of freedom. We then mapped the marginals with their empirical uniform distribution to achieve the 2-dimensional uniform sample U_{emp}. This one is compared with the 2-dimensional sample U_{norm} realised by fitting univariate normal distributions to the marginals, given by $U_{norm} = \left\{\left(\widehat{N}_X(x_i), \widehat{N}_Y(y_i)\right)\right\}$. Figure 7.2 clearly shows that U_{emp} is reflecting the simulated dependence of a Gumbel copula whereas U_{norm} has nearly lost all appearance for upper tail dependence and shows evidence for a sort of "linear" dependence. Seeing a scatterplot of U_{norm} one probably gets the feeling that a t-copula is an appropriate model. Remark that of course the hypothesis of uniform marginals (what is the constraint for calibrating a copula) would be rejected for U_{norm}. Here we have shown that not testing or ignoring the quality of the marginal fit can cause dramatic misinterpretation of the underlying dependence structure.

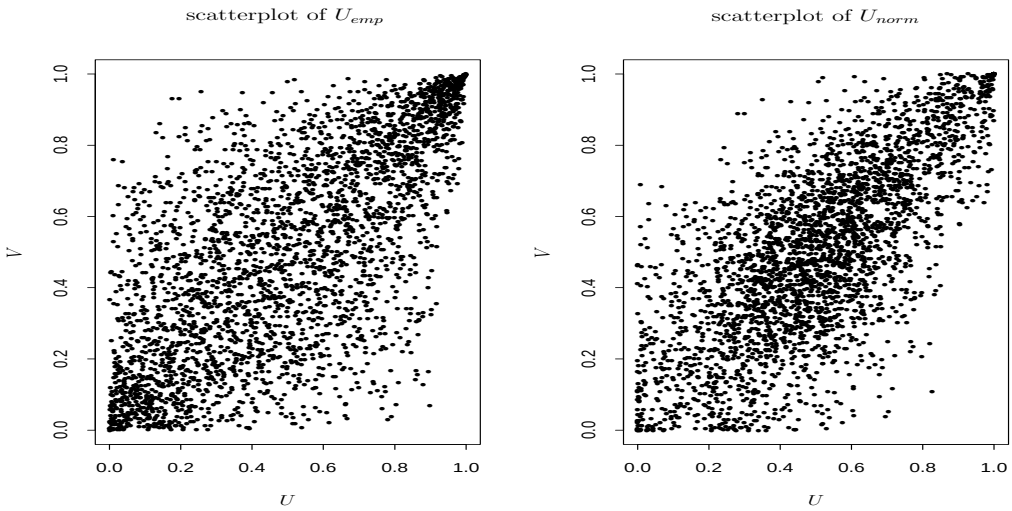

Figure 7.2: Comparison of empirical marginal mapping with misspecified marginal mapping.

7.4 Simulation Study

In order to compare the qualities and properties of the introduced tail-dependence estimators we run an extensive simulation study. Each simulated data set consists of 1000 independent copies of n realizations from a random sample $X^{(1)}, \ldots, X^{(n)}$ having one particular distribution out of four. Three different sample sizes n are considered for each data set: $n = 250, 1000, 5000$. The four different distributions under consideration are denoted by H, T, F and G. For example the data set S_H^{250} is defined by

$$S_H^{250} := \left\{ \left\{ X^{(i,j)}(\omega) : X^{(i,j)} \sim H(\lambda = 1, \alpha = 1, \rho = 0.5), j = 1, \ldots, 250 \right\} : i = 1, \ldots, 1000 \right\},$$

where $H(\lambda = 1, \alpha = 1, \rho = 0.5)$ denotes the standard bivariate symmetric generalized hyperbolic distribution with correlation coefficient $\rho = 0.5$ as defined in Hrycej et al. [58]. Distribution T corresponds to the Student's t-distribution with $\nu = 1.5$ degrees of freedom and $\rho = 0.5$. Another elliptical distribution F is characterized by the random variate $\mathcal{R}$ and scaling matrix

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix},$$

$$\mathcal{R} \sim \left(\frac{1}{2} \cdot \mathrm{LN}(\mu = 1, \sigma = 0.01) + \frac{1}{2} \cdot \mathrm{GPD}(\xi = \frac{2}{3}, \beta = 1) \right),$$

i.e., $\mathcal{R}$ is the mixture of a lognormal distribution and a generalized Pareto distribution.

The last distribution function G is characterized by

$$G(x_1, x_2) = C_{\vartheta,\gamma}\left(\Phi\left(x_1\right), \Phi\left(x_2\right)\right),$$

where Φ denotes the univariate standard normal distribution and $C_{\vartheta,\gamma}$ is an Archimedean copula achieved by a transformed Frank copula with generator

$$\varphi_{\vartheta,\gamma}(t) = \left(\varphi_{\text{Frank}}(t)\right)^{\gamma} = \left(-\log\frac{e^{-\vartheta t}-1}{e^{-\vartheta}-1}\right)^{\gamma}, \qquad \vartheta \neq 0, \gamma < 1,$$

introduced by Junker and May [67]. We choose $\vartheta = 1.56$ and $\gamma = -0.76$.

Note that H is elliptically distributed but has no tail dependence. The corresponding data sets are used to control the stability of the utilized estimation procedures under absence of tail dependence. In contrast, T and F are elliptical distributions with parameters $\rho = 0.5$ and tail index $\zeta = 1.5$ possessing tail dependence. The tail-dependence coefficient is obtained by the relation

$$\lambda = 2 \cdot \bar{t}_{\zeta+1}\left(\sqrt{\zeta+1} \cdot \sqrt{\frac{1-\rho}{1+\rho}}\right). \tag{7.4.1}$$

Observe that the copula $C_{\vartheta,\gamma}$ is lower tail-independent but upper tail-dependent, i.e., $\lambda_L^G = 0$ and $\lambda_U^G = 2 - 2^{\frac{1}{\gamma}}$. In order to provide comparability of the estimation results of the different models T, F, and G the parameterization is chosen such that $\lambda_L^T = \lambda_U^T = \lambda_L^F = \lambda_U^F = \lambda_U^G = 0.4406$ and $\tau^H = \tau^T = \tau^F = \tau^G = 0.\overline{3}$. Note that Kendall's τ for elliptical copulas can be obtained via the relation

$$\tau = \frac{2}{\pi} \cdot \arcsin\left(\rho\right). \tag{7.4.2}$$

Figure 7.3 illustrates the different tail-behavior of the H, T, F and G distribution by presenting the scatterplots of the respective simulated data-sample with sample size $n = 5000$ together with the corresponding empirical copula-density. For the copula-mapping we utilize the marginal empirical distribution functions.

7.4.1 Pre-simulation

The properties of the two-step estimate (empirical marginals and ML estimation of $\theta\left(\widehat{u}\right)$) in contrast to the one-step estimate (ML estimation of θ) is compared for the S_T^{1000} data set with parameters $\nu = 1.5$ and $\rho = 0.5$. We denote the parameters of the one-step ML estimator by $\widehat{\theta}_{T,ML} := \left(\widehat{\nu}_{ML}, \widehat{\rho}_{ML}\right)$ and the parameters of the two-step estimator by $\widehat{\theta}_{T,\widehat{u}} := \left(\widehat{\nu}, \widehat{\rho}\right)$. Comparing the empirical means

$$\overline{\widehat{\theta}_{T,ML}} = \left(1.5364, 0.5000\right), \overline{\widehat{\theta}_{T,\widehat{u}}} = \left(1.5388, 0.5004\right),$$

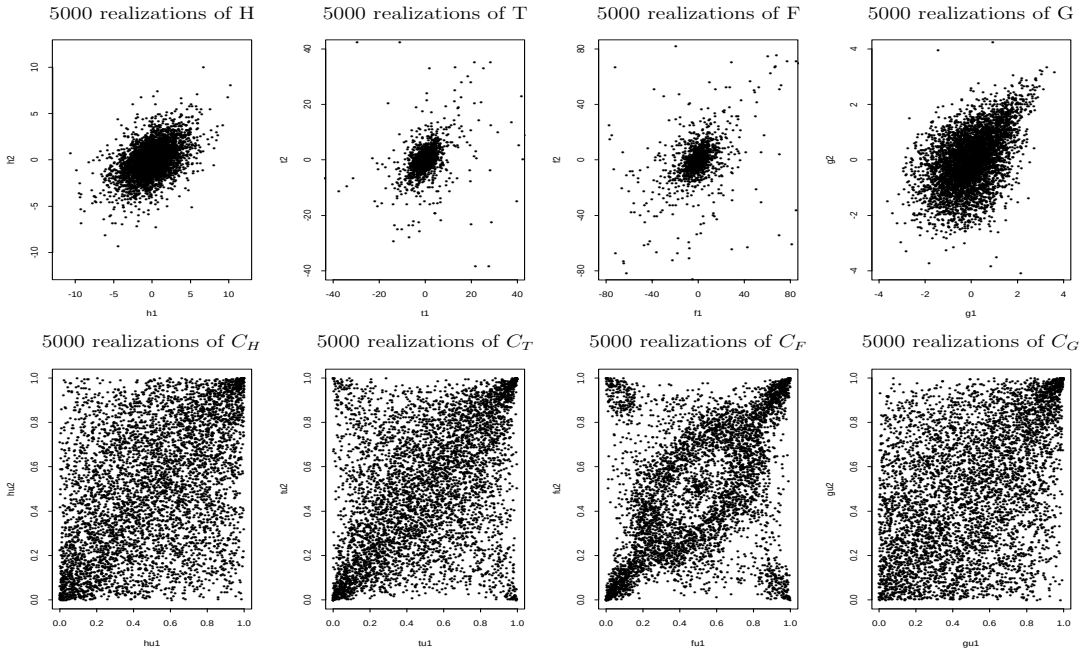

Figure 7.3: Scatterplots of the simulated distributions and corresponding empirical copula densities.

the medians $\mu_{0.5}\left(\widehat{\theta}_{T,ML}\right) = (1.5003, 0.5001)$, $\mu_{0.5}\left(\widehat{\theta}_{T,\widehat{u}}\right) = (1.5242, 0.5005)$ and the standard deviations $\sigma\left(\widehat{\theta}_{T,ML}\right) = (0.3562, 0.0293)$, $\sigma\left(\widehat{\theta}_{T,\widehat{u}}\right) = (0.1447, 0.0316)$ we conclude that there is almost no difference for estimating ρ. However, the standard deviations of $\widehat{\nu}_{ML}$ and $\widehat{\nu}$ differ significantly as a result of upper outliers of $\widehat{\nu}_{ML}$ (Even though there are no remarkable lower outliers they do not effect the standard deviation in the same way as upper outliers). Another statistical measure is provided by the ratio of the mean squared errors $MSE\left(\widehat{\theta}_{T,ML}\right)/MSE\left(\widehat{\theta}_{T,\widehat{u}}\right) = (5.71, 0.86)$. The ratio approves the worse estimation of $\widehat{\nu}_{ML}$ compared to $\widehat{\nu}$ and a similar estimation of the correlation coefficient. Therefore the two-step estimation is to recommend.

In what follows we compare the correlation-coefficient estimator for elliptical random vectors introduced in Section 7.3.1 (denoted by $\widehat{\rho}_s$) with the correlation-coefficient estimator introduced by Lindskog et al. [77] (denoted by $\widehat{\rho}_\tau$) which is based on Kendall's τ. This is of interest because the tail-dependence coefficient for elliptical distributions depends heavily on the correlation-coefficient estimate due to equation (7.3.3). In Figure 7.4 we see that the bias of both $\left(0.5 - \overline{\widehat{\rho}_\tau}, 0.5 - \overline{\widehat{\rho}_s}\right) = (0.0004, -0.0012)$ is negligible in the case of estimating the correlation coefficient from S_T^{1000}, and the standard deviations

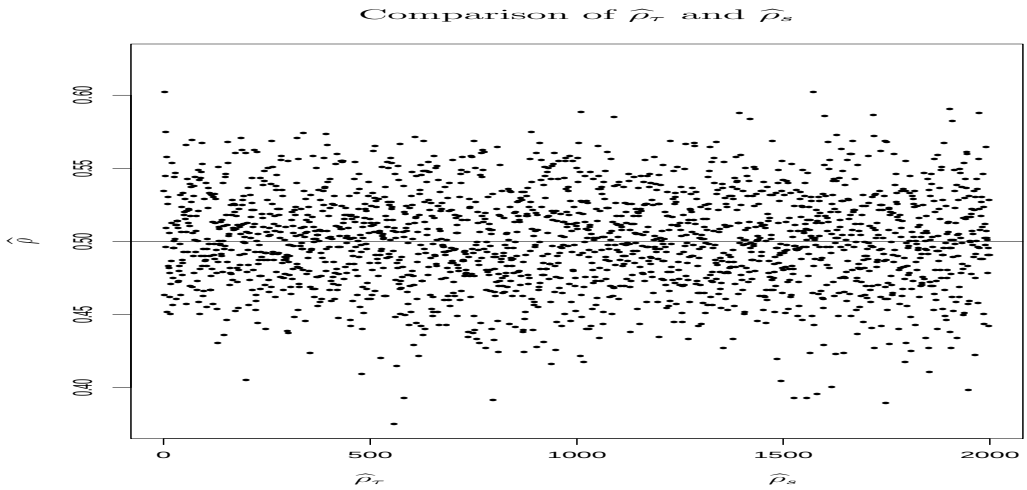

Figure 7.4: Comparison of $\widehat{\rho}_\tau$ and $\widehat{\rho}_s$. (The black line relates to the theoretical value.)

$(\sigma(\widehat{\rho}_\tau), \sigma(\widehat{\rho}_s)) = (0.0326, 0.0335)$ are of comparable size. The mean square error ratio $MSE(\widehat{\rho}_\tau)/MSE(\widehat{\rho}_s) = 0.95$ indicates a very small favor for $\widehat{\rho}_\tau$ as the result of its slightly lower volatility. However, note that the estimate $\widehat{\rho}_s$ yields immediately a correlation matrix for higher dimensions, whereas the estimator $\widehat{\rho}_\tau$ has to be adjusted to guarantee the resulting correlation matrix to be positive semi definite. Furthermore in high dimensions the calculation of the correlation matrix via Kendall's τ becomes a computational extensive task.

As mentioned above there are two possibilities that yield an estimate of the tail index of an elliptical distributed random vector X. We can use the realizations of $\|X\|_2$ as sketched in Section 7.3.2 or the estimates of $R_d^\star$ like outlined in Section 7.3.1. To check the possible influence on the tail index estimation with finite samples, we compare in Figure 7.5 the mean excess plots for the $\|X\|_2$ and the estimated $R_d^\star$ realizations of one element of S_T^{1000}. It turns out that the slope of a linear regression line for the points above a realistic chosen thresholds, is the same for both procedures. Hence the tail index will be estimated to nearly the same value. Nevertheless $\|X\|_2$ and $R_d^\star$ have of course different distribution functions what is expressed by Figure 7.6 showing the empirical survival distributions of the realizations.

In Section 7.3.2 we stated the limit result (7.3.11) of Juri and Wütherich [70]. A recent empirical study about the TDC's of high frequency data (Breymann et al. [14]) also investigated the modelling of the limit copula (7.3.10) with other parametric copula families than the Clayton one. They found that the best adaption over several thresholds measured by an AIC criterion is best for the Clayton

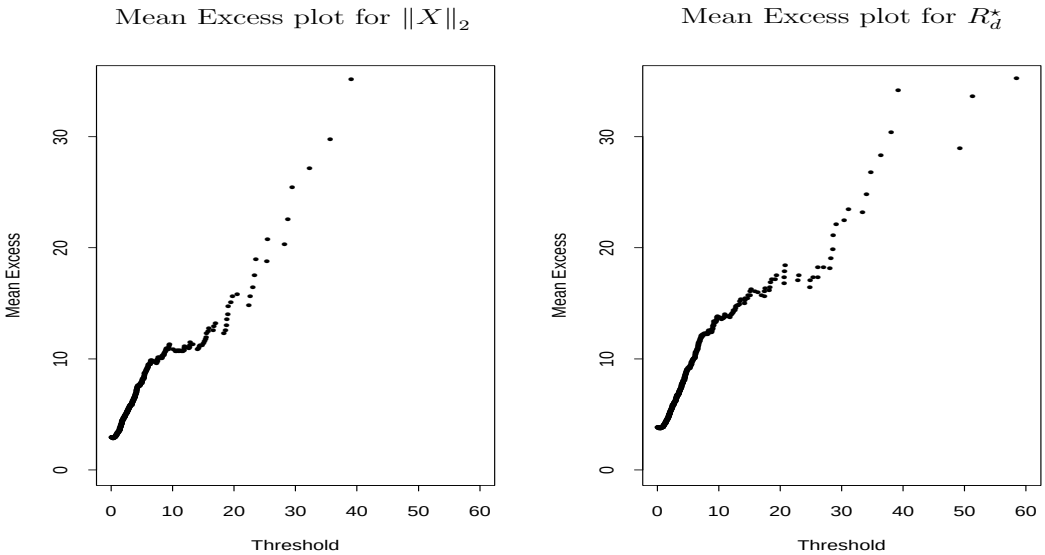

Figure 7.5: Mean Excess Plots for a realization of $\|X\|_2$ and $R_d^\star$

copula. However, here we are rather interested in a good estimation of the TDC than in a good fit of the limit copula (7.3.10). Of course we argue that there is a high coincidence between these goals. In Figure 7.7 we used the Clayton and the survival Gumbel copula to model the limit copula 7.3.10.

Plotted with c (Clayton) and g (Gumbel) are the means of the $\widehat{\lambda}_{thu}$ estimations over the 1000 elements of S_G^{1000} (left picture) and S_T^{1000} (right picture). Furthermore the lines represent the observed 95% confidence bands of the Clayton (solid) and survival Gumbel (dashed) $\widehat{\lambda}_{thu}$.

Not surprisingly the range of the estimates decreases for increasing thresholds. In the case of the Archimedean samples in S_G^{1000} the bias of the Clayton $\widehat{\lambda}_{thu}$ is rather small, whereas using the survival Gumbel causes in a significant underestimation. On the other side the confidence bands of the survival Gumbel are significantly smaller; so one has the old trade off between bias and variance. Under violation of the assumption that the observations are samples of an Archimedean dependence structure, like it is the case for the S_T^{1000} samples, both copula models especially the Clayton one, result in significant overestimation. Also here the variance of the survival Gumbel based estimator is slightly lower.

A remarkable result for both models and samples is the stability of the mean estimate over the thresholds. This stability gives us the justification of using kernel type estimates, e.g. taking the mean of estimates over a grid of a threshold interval. Such procedures can gain drastic volatility reduction.

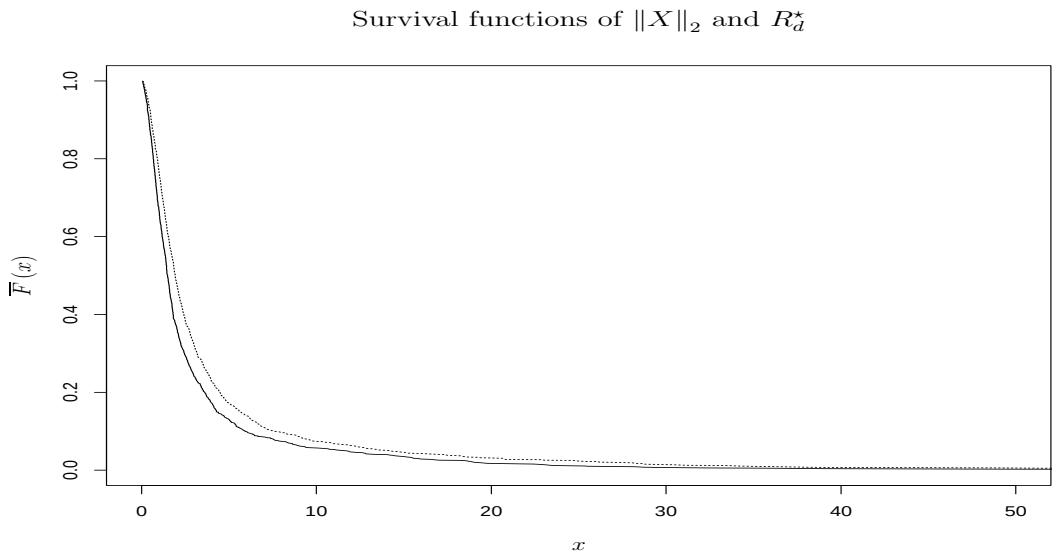

Figure 7.6: Survival function of $\|X\|_2$ (solid line) and $R_d^\star$ (dashed line).

The differences in the finite sample properties of the secant and log-estimator introduced in Section 7.3.2 are illustrated in Figure 7.8. The upper picture shows the upper TDC estimates for different k of an independent normal distributed sample. The lower one is a t-distributed sample out of S_T^{1000}. The black lines indicate the true values. Note that the smoothing effect of the logarithm causes a decreasing bias even in the case of independence, i.e. in a special case of $\lambda_U = 0$. The log-estimator is much more robust against the choice of k. So one can take a benefit from the homogenous property of the log-estimator by using kernel type estimates over k to reduce the estimators variance.

7.4.2 Simulation and estimation

For comparison of the different estimation methods we calculate the mean $\overline{\widehat{\lambda}}_n$, the finite sample standard deviation $\sigma\left(\widehat{\lambda}_n\right)$ and the approximate asymptotic standard deviation of $\sqrt{n}\left(\widehat{\lambda}_n - \lambda\right)$, if possible. The root mean square-error

$$\mathrm{RMSE}\left(\widehat{\lambda}_n\right) := \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(\widehat{\lambda}_{n,i} - \lambda\right)^2} \tag{7.4.3}$$

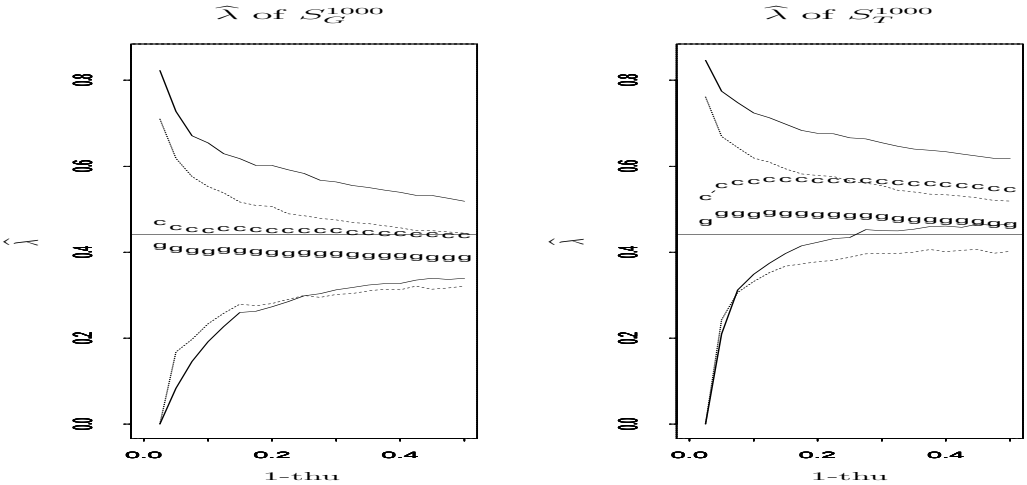

Figure 7.7: TDC mean estimates for the S_G^{1000} (left plot) and the S_T^{1000} samples (right plot) over several thresholds. The Clayton copula is marked with c (solid line 95% confidence interval), the survival Gumbel with g (dashed line 95% confidence interval).

allows to observe the bias-variance trade-off for the different sample sizes and estimation procedures. We introduce another statistical quantity

$$\mathrm{MESE}\left(\widehat{\lambda}_n\right) := \frac{\mathrm{RMSE}\left(\widehat{\lambda}_n\right)}{\sigma\left(\widehat{\lambda}_n\right)} - 1, \tag{7.4.4}$$

which measures the distance of the standard error $\sigma\left(\widehat{\lambda}_n\right)$ to the bias due to model misidentification. We call MESE the mean error of the standard error. $\mathrm{MESE}\left(\widehat{\lambda}_n\right) \to_{n\to\infty} 0$ if the estimator of λ is unbiased, since in that case $\mathrm{RMSE}\left(\widehat{\lambda}_n\right) \to_{n\to\infty} \sigma(\widehat{\lambda}_n)$. For every biased estimation the RMSE is bounded away from zero and therefore the latter MESE-limit diverges to infinity.

Estimation from the entire sample

Specifying a parametric copula

This section relates to Section 7.3.1. For the elliptical data sets S_H, S_T, and S_F we fit a t-copula, since this seems to be a realistic choice by looking only at the scatterplots in Figure 7.3. However, the empirical copula density of these data sets indicate that the t-copula may not be suitable for S_F. Although not fitting the true copula-model we estimate the tail-dependence coefficient via the relation (7.4.1). In Figure C.1 the 3×1000 consecutive estimates corresponding to S_T are presented.

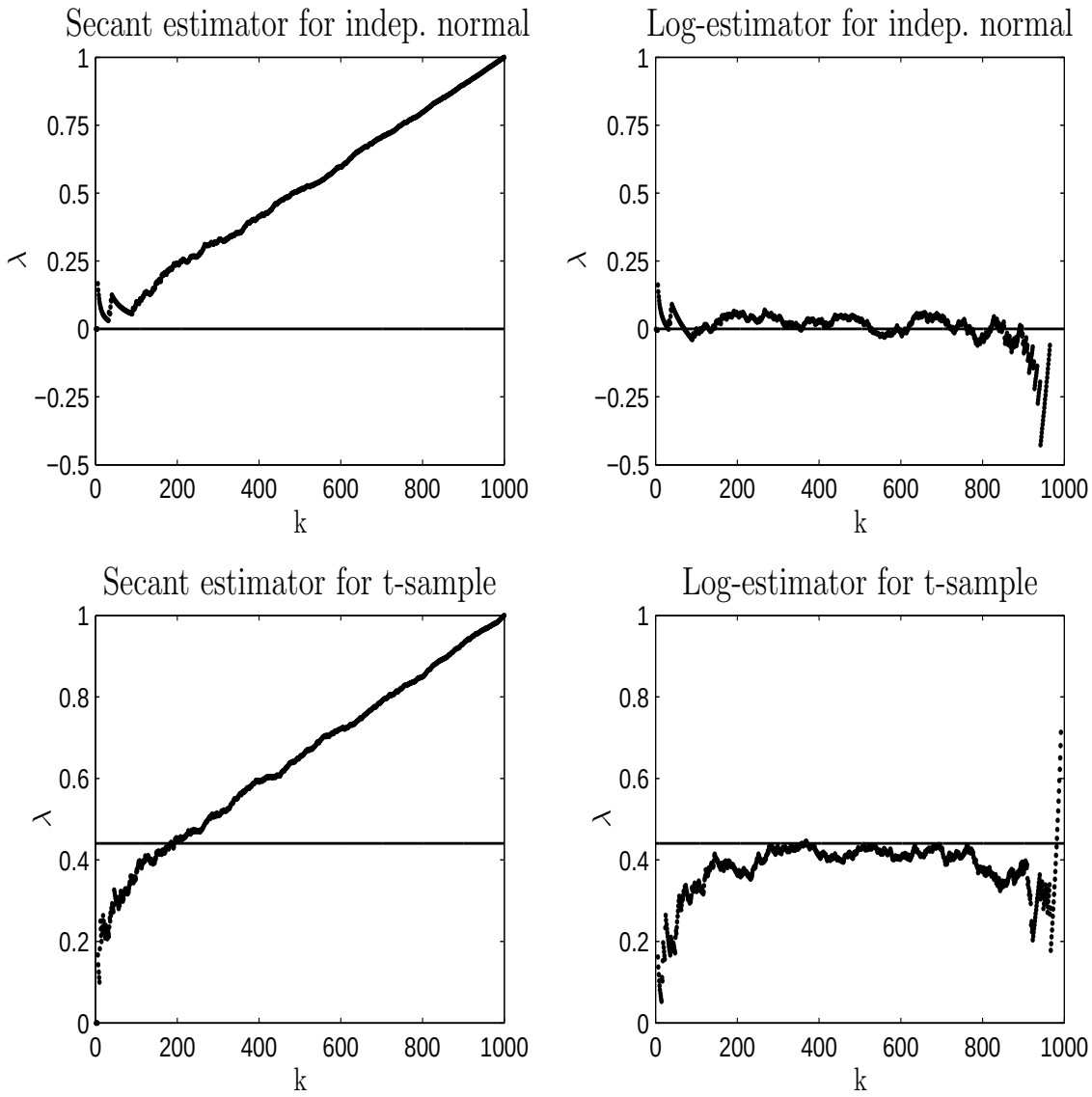

Figure 7.8: Secant estimator and log-estimator for an independent normal sample (upper picture) and for a t-sample out of S_T^{1000} (lower picture).

In the following we will plot the estimates for the data set S_T under different estimation procedures and summarize the figures in Appendix C. For easy comparison we fix the range of the plots. Regarding the data set S_G, we fit a Gumbel-copula which seems to be the true underlying model according to Figure 7.3 and estimate the upper tail-dependence coefficient via $\widehat{\lambda}_U^G = 2 - 2^{\frac{1}{\widehat{\gamma}}}$.

The estimation results for all data sets are summarized in Table 7.1.

Estimation from extreme samples

The threshold method

In the pre-simulation Section 7.4.1 we have seen that the use of the Clayton copula for the threshold approach as introduced in Section 7.3.2, can cause drastic overestimation in the case of violation of the Archimedean dependence assumption. Since we assume a threshold procedure should require as weak assumptions as possible, we use the survival Gumbel copula that is not best in the case of Archimedean dependence, but also not worst if the archimedean assumption does not hold.
As outlined in Section 7.4.1 we can use kernel type estimates to reduce the variance. Here we just take the mean of estimations over an equidistant grid of 10 thresholds in the interval $\left[\frac{1}{\sqrt{n}}, \frac{1}{n^{1/4}}\right]$ and $\left[1 - \frac{1}{n^{1/4}}, 1 - \frac{1}{\sqrt{n}}\right]$ respectively. So $\lambda_{thl} = \frac{1}{10}\sum_{i=1}^{10} \lambda_{thl_i}$ and λ_{thu} analogously.
Whenever the asymmetry test introduced in Section 7.3.2 rejects asymmetry for more than half of the pairs $(\lambda_{thl_i}, \lambda_{thu_i})$, we state $\lambda_{thl,thu}$, else λ_{thu}.

The statistical figures corresponding to the data sets S_H, S_T,S_F and S_G are provided in Table 7.1. Figure C.4 gives a graphical survey, again for the set S_T.

The parametric method for elliptical distributions

For the data sets S_H, S_T and S_F we estimated the correlation matrix and the tail index by the method presented in Section 7.3.1 and provide the estimates for the tail-dependence coefficient via relation (7.4.1). Additionally, for the data set S_T the tail index is estimated via $\|X\|_2$ and the correlation matrix via Kendall's τ and the results are summarized in Figure C.2. In Figure C.3 the estimates of both the tail index and the correlation matrix are derived utilizing the separation algorithm described in Section 7.3.1.

Obviously the data in S_G is not elliptical, so here it is not possible to make use of this method. For the numerical evaluations see again the table at the end of this chapter.

The nonparametric approach

For the data sets S_H, S_T,S_F and S_G the statistical figures of the nonparametric estimator (7.3.16) and the log-estimator (7.3.21) introduced in Section 7.3.2 are provided in Table 7.2. Again for the data set S_T, Figure C.5 and Figure C.6 give a graphical view of both procedures.

The plateaus are chosen by a heuristic plateau finding algorithm that first smoothes the plateau-plot by a simple box kernel with "bandwidth" bd, i.e. the mean of $bd+1$ successive points of the plateau-plot vector $(\lambda_1, ..., \lambda_n)$ leads to a new point of the smoothed plateau $\Lambda = (\lambda_1, ..., \lambda_{n-bd})$.

Here we have taken $bd = \lfloor 0.005 \cdot n \rfloor$. In a second step, a plateau of length m, here $m = \left\lfloor \sqrt{n-bd} \right\rfloor$, is defined as a vector $(\lambda_i, \lambda_{i+1}, ..., \lambda_{i+m-1})$, $i \in \{1, ..., n-bd\}$ belonging to elements of Λ. The algorithm takes the first plateau k for $k = 1, ..., n-bd-m+1$ whose elements fulfil the condition $\sum_{i=k}^{k+m-1} |\lambda_k - \lambda_i| \leq 2\sigma(\Lambda)$. If such a plateau does not exist, the TDC is set to zero, otherwise let $\{\lambda_{k^\star}, ..., \lambda_{k^\star+m-1}\}$ be the appointed plateau, and set $\widehat{\lambda} := \frac{1}{m} \sum_{i=0}^{m-1} \lambda_{k^\star+i}$.
As seen from the pre-simulation Section 7.4.1 we may choose a greater bandwidth bd for the log-estimator to reduce the variance of the estimation. However, for better comparison between the two types, we do not take this opportunity here.

7.4.3 Discussion of the results

The parametric method 7.3.1 has mostly the lowest standard error for the used data sets (see Table 7.1). This is because here we have the strongest model assumptions and hence the lowest degree of freedom within the compared approaches. On the other side the observed bias is not such bad, surprisingly even in the case of misspecified dependence models. Together this leads to the lowest root mean square error for the most samples. However, if the chosen model was not the true one (i.e. for all except the S_T samples), the mean error of the standard error is significantly high, often the worst. So the standard error is too low to be a reliable measure for the estimation error.

If there is a well founded suspicion of the sample being generated by an elliptical distribution, we recommend the use of one of the elliptical methods introduced in Section 7.3.1. One benefit is that there are well known methods to test the pseudo observations of $\frac{X}{\|X\|_2}$ for uniformity on the sphere (see e.g. Mardia [85]) or the observations of X for elliptical symmetry (see e.g. Manzotti et al. [84]). In Table 7.1 we see that the bias for both subtypes is of comparable small size, where the standard errors are also comparable and with exception of the S_H data of "realistic error size" as indicated by the very low MESE. Here it is hard to

indicate a favor, either for the $\|X\|_2$ or the $\mathcal{R}_d^*$ approach. As mentioned in Sections 7.3.1 and 7.4.1, for higher dimensions the use of Kendall's τ to achieve the correlation matrix is complicated if positive definiteness is to be guaranteed. So for multivariate tasks the estimation via $\mathcal{R}_d^*$ is in some sense the favorable method.

As seen in the pre-simulation Section 7.4.1, the threshold method in Section 7.3.2 with use of the Clayton copula in (7.3.11) is a really good choice, if there is a strong evidence for the assumption of being in an Archimedean world. This is strengthend if using a kernel type estimator over different thresholds for the reduction of noise. In Table 7.2 the results for the more robust use (against non Archimedean dependence) of the Gumbel copula in 7.3.11 are stated. The lower index gives the number (in percentage of the 1000 data sets) that are found to be TDC asymmetric as defined in Section 7.3.2. The introduced test seemed to work quite well. In all TDC symmetric data (S_H, S_T, S_F) there were only a few sets declared to have different TDC's, whereas for the S_G data the number of detected asymmetric TDC samples increases rapidly with the increase of sample length. The upper index gives the mean number of observations that was used for the estimation. For the elliptical data the TDC was significantly overestimated, whereas for the Archimedean samples in S_G we observe significant underestimation, what reflects the bias observed in the pre-simulation 7.4.1. The underestimation may probably also be caused by the fact that we take the average whenever the hypothesis of symmetry could not be rejected. The hope using not the entire data set is always to yield a better bias, mostly with the cost of higher variance. Here the bias *and* the variance are mostly the worst over all methods. So of course is the RMSE. So as an approach that should not make any a priori assumptions about the dependence structure, it has not much to recommend on. Remark that the use of the Clayton copula will even worsen the findings for the elliptical data, as it is seen in the pre-simulation. Again, like stated above, in an Archimedean world everything is fine if the Clayton copula is used.

The nonparametric methods in Section 7.3.2 lead to the estimates given in Table 7.2. These procedures have no assumption beside the one that the investigated sample has a TDC greater zero. So for the S_H data sets where this assumption is violated, the bias is worst with respect to the other methods. This can be fixed by testing for strict positive TDC before running the estimation. For a description of such tests see Ledford and Tawn [75]. The TDC's of other data were captured quite well. Of course the variance is here bigger than under the methods requiring distributional assumptions, and the bias can not be as low as in the data where given restrictions are fulfilled. But for example the S_G data are estimated here with the lowest bias. Also for the case S_F the bias can partly beat the one of

the semiparametric method of Section 7.3.1. The comparison of both submethods established a small benefit of using the proposed log-estimator, since here the variances are overall slightly smaller, as expected from the pre-simulation Section 7.4.1, even though we have taken the same bandwidth for the kernel procedure introduced in Section 7.4.2. This causes in smaller RMSE's.

7.5 Conclusion

We found that there is not *the one* TDC estimation method. The semiparametric method 7.3.1, that is specifying a parametric copula model, is quite acceptable. Even in the case of misidentification the results are not too badly biased, but one has to be careful not to misinterpret the very low standard error as a sign of excellent estimation.
On the other side we have the completely nonparametric approach 7.3.2 that has the typical bias variance trade-off compared to the former method. Taking the logarithm of the diagonal section causes in slightly better root mean square errors than observing the diagonal section directly. Utilizing the observed homogeneity of the log-estimator by taking broader bandwidth in the kernel estimation will lead to further variance reduction.
A good compromise seems to be the methods with weak assumptions about the underlying dependence structure. For the elliptical case we observed low bias and low variance for the methods that separate the estimation of the correlation matrix and the estimation of the tail index of the radial random variable (see Section 7.3.1). In the case of Archimedean dependency, the threshold method of Section 7.3.2 with the fit of a Clayton copula to the extreme value copula achieves nearly unbiased results, where the variance can be reduced by kernel type estimators over several thresholds.

Method	Data set	$\overline{\widehat{\lambda}_U}$	BIAS$\left(\widehat{\lambda}_U\right)$	$\sigma\left(\widehat{\lambda}_U\right)$	RMSE$\left(\widehat{\lambda}_U\right)$	MESE
t/Gumbel copula and	S_H^{250}	0.1618	0.1618^+	0.0817	0.1812^+	1.2184^+
empirical marginals	S_H^{1000}	0.1698	0.1698	0.0413	0.1747	3.2315
cf. Section 7.3.1	S_H^{5000}	0.1739	0.1739	0.0187^+	0.1749	8.3676^-
and 7.4.2	S_T^{250}	0.4281	-0.0125	0.0403^+	0.0422^+	0.0463
	S_T^{1000}	0.4374	-0.0032	0.0204^+	0.0206^+	0.0098
	S_T^{5000}	0.4400	-0.0006^+	0.0092^+	0.0092^+	0.0000^+
	S_F^{250}	0.4873	0.0467	0.0316^+	0.0564^+	0.7831^-
	S_F^{1000}	0.4940	0.0534	0.0150^+	0.0555^+	2.6862^-
	S_F^{5000}	0.4946	0.0540	0.0065^+	0.0544^+	7.3265^-
	S_G^{250}	0.3905	-0.0501	0.0437^+	0.0664^+	0.5214^-
	S_G^{1000}	0.3922	-0.0484	0.0212^+	0.0529^+	1.4914^-
	S_G^{5000}	0.3919	-0.0487^-	0.0097^+	0.0497	4.1237^-
Elliptical approach	S_H^{250}	0.2031	0.2031	0.0588^+	0.2114	2.5959
via $\|X\|_2$	S_H^{1000}	0.1815	0.1815	0.0377^+	0.1854	3.9171
cf. Section 7.3.2	S_H^{5000}	0.1575	0.1575	0.0220	0.1590	6.2286
and 7.4.2	S_T^{250}	0.4379	-0.0027^+	0.0465	0.0466	0.0012^+
	S_T^{1000}	0.4432	0.0026^+	0.0242	0.0243	0.0054^+
	S_T^{5000}	0.4437	0.0031	0.0109	0.0113	0.0398
	S_F^{250}	0.4712	0.0306	0.1228	0.1265	0.0302
	S_F^{1000}	0.4122	-0.0284^+	0.0955^-	0.0996	0.0428^+
	S_F^{5000}	0.4298	-0.0108	0.0450	0.0463	0.0277
Elliptical approach	S_H^{250}	0.1865	0.1865	0.0660	0.1978	1.9970
via $\mathcal{R}_d^*$	S_H^{1000}	0.1675	0.1675^+	0.0388	0.1719^+	3.4304
cf. Section 7.3.2	S_H^{5000}	0.1424	0.1424^+	0.0235	0.1444^+	5.1447
and 7.4.2	S_T^{250}	0.4374	-0.0032	0.0491	0.0492	0.0020
	S_T^{1000}	0.4435	0.0029	0.0235	0.0237	0.0085
	S_T^{5000}	0.4440	0.0034	0.0113	0.0118	0.0442
	S_F^{250}	0.4483	0.0077^+	0.1412	0.1413	0.0004^+
	S_F^{1000}	0.4033	-0.0373	0.0916	0.0988	0.0786
	S_F^{5000}	0.4394	-0.0012^+	0.0327	0.0327^+	0.0000^+

Table 7.1: Statistical figures for various parametric, semi-parametric and non-parametric estimation methods. Best values over the different methods are ticked with a plus, worst ones with a minus sign behind.

Method	Data set	$\overline{\widehat{\lambda}_U}$	BIAS$\left(\widehat{\lambda}_U\right)$	$\sigma\left(\widehat{\lambda}_U\right)$	RMSE$\left(\widehat{\lambda}_U\right)$	MESE
Threshold method	S_H^{250}	$0.2462^{31}_{6.8}$	0.2462	0.1177^-	0.2729	1.3186
cf. Section 7.3.2	S_H^{1000}	0.2040^{76}_{2}	0.2040	0.0789^-	0.2187	1.7719^+
and 7.4.2	S_H^{5000}	$0.1885^{214}_{1.5}$	0.1885	0.0510	0.1953	2.8294
	S_T^{250}	0.4715^{36}_{12}	0.0309	0.1159^-	0.1199^-	0.0345
	S_T^{1000}	$0.4834^{97}_{4.4}$	0.0428^-	0.0659^-	0.0786	0.1927
	S_T^{5000}	$0.4900^{302}_{1.7}$	0.0494^-	0.0346	0.0603^-	0.7428^-
	S_F^{250}	$0.5376^{39}_{13.9}$	0.0970^-	0.1013^-	0.1403^-	0.3850
	S_F^{1000}	$0.5218^{111}_{4.9}$	0.0812	0.0559	0.0986	0.7639
	S_F^{5000}	$0.5126^{347}_{3.1}$	0.0720^-	0.0321	0.0788	1.4548
	S_G^{250}	$0.3436^{29}_{29.7}$	-0.0970^-	0.1512^-	0.1796^-	0.1878
	S_G^{1000}	$0.3594^{63}_{54.9}$	-0.0812^-	0.1169^-	0.1422^-	0.2164
	S_G^{5000}	$0.3925^{162}_{89.6}$	-0.0481	0.0634^-	0.0796^-	0.2555
Nonpara. method	S_H^{250}	0.3636	0.3636^-	0.1016	0.3775^-	2.7158
emp. copula est.	S_H^{1000}	0.3056	0.3056^-	0.0717	0.3139	3.3779
cf. Section 7.3.2	S_H^{5000}	0.2390	0.2390	0.0932^-	0.2565^-	1.7525^+
and 7.4.2	S_T^{250}	0.4681	0.0275	0.0800	0.0845	0.0574
	S_T^{1000}	0.4587	0.0181	0.0513	0.0545	0.0606
	S_T^{5000}	0.4463	0.0057	0.0431^-	0.0435	0.0087
	S_F^{250}	0.5326	0.0920	0.0758	0.1192	0.5718
	S_F^{1000}	0.5326	0.0920^-	0.0554	0.1074^-	0.9386
	S_F^{5000}	0.4880	0.0474	0.0694^-	0.0841^-	0.2113
	S_G^{250}	0.4841	0.0435	0.0796	0.0907	0.1397
	S_G^{1000}	0.4650	0.0244^+	0.0482	0.0541	0.1208^+
	S_G^{5000}	0.4453	0.0047^+	0.0603	0.0605	0.0030^+
Nonpara. method	S_H^{250}	0.3144	0.3144	0.0828	0.3251	2.9263^-
log estimator	S_H^{1000}	0.2893	0.2893	0.0539	0.2943	4.4601^-
cf. Section 7.3.2	S_H^{5000}	0.2567	0.2567^-	0.0377	0.2595^-	5.8833
and 7.4.2	S_T^{250}	0.3951	-0.0455^-	0.0727	0.0857	0.1788^-
	S_T^{1000}	0.4132	-0.0274	0.0491	0.0562	0.1446^-
	S_T^{5000}	0.4240	-0.0166	0.0352	0.0389	0.1051
	S_F^{250}	0.4228	-0.0178	0.0913	0.0930	0.0186
	S_F^{1000}	0.4932	0.0526	0.0612	0.0807	0.3186
	S_F^{5000}	0.4802	0.0396	0.0406	0.0567	0.3966
	S_G^{250}	0.4016	-0.0390^+	0.0719	0.0818	0.1377^+
	S_G^{1000}	0.4098	-0.0308	0.0448	0.0544	0.2121
	S_G^{5000}	0.4229	-0.0177	0.0233	0.0293^+	0.2575

Table 7.2: Statistical figures for various parametric, semi-parametric and non-parametric estimation methods. Best values over the different methods are ticked with a plus, worst ones with a minus sign behind.

Chapter 8

Copula Estimations for Risk Factor Portfolios

This chapter is based on the empirical results of Junker and May [67]. We observe the copulas of different risk factor portfolios, namely stock indices, fx rates, oil related commodities and MLE metals. We will discuss the bivariate portfolios as well as the corresponding multivariate portfolios up to dimension 4. The modelling is based on the new copula family combining the Frank and Gumbel one, as presented in Subsection 4.2.1. For running the goodness-of-fit tests we utilize the mapping rule to independent uniformity and the resulting χ^2 test as introduced in Subsection 6.1.2. We further use some other test diagnostics as outlined in Section 6.2.

8.1 Pre-considerations

Recent studies which apply copula functions in finance such as for example Ané and Kharoubi [3] and Scaillet [103] indicate that the concept appeals in modelling complex dependence structures. Malevergne and Sornette [83] argue that the hypothesis of the normal copula cannot be rejected for a variety of financial returns, including stock and exchange rate returns. However, this finding may relate to the amount of data available and to issues of power of the testing procedures (They used the radial class test with the drawbacks presented in Subsection 6.1.2.). Indeed, the authors also find that alternative copula models cannot be rejected either. Breymann et al. [14] fitted different copula models to high frequency fx data, where they used several other diagnostics beside the radial class test. Within this literature, there is an ongoing debate which copula models are most appropriate. In the next subsection we will discuss our model choices.

8.1.1 Pre-selection of the models

There are no hints concerning an asymmetric dependence structure with respect to permutational symmetry in most financial data sets. So the class of Archimedean copulas seems to be a natural choice for our setting in the sense that the restriction to permutational symmetric models seemed to be uncritical. Moreover, this class contains many tractable parametric families and has a one dimensional generator. One important special case in financial dependence structure is the rotational symmetric case like it is assumed e.g. in a normal or more general in an elliptical world. As mentioned in Chapter 4 the only rotational symmetric Archimedean copula family is the Frank family. Furthermore from the survey of Ané and Kharoubi [3] one can deduce from the tables, that the Frank copula fits the data (here stock indices) better than the normal copula. On the other hand tail dependence is an important stylized fact of financial data (or at least a concentration of the dependence in the tail region, if one does not believe in the asymptotic concept of tail dependence). Ané and Kharoubi [3] found the Clayton family to be best to cover lower tail dependence. Despite the fact that they missed to compare there results with the survival Gumbel copula, other authors considering the survival Gumbel as benchmark model like e.g. Breymann et al. [14], came to the same conclusion. So a model that covers the Frank and Clayton families would be optimal. Unfortunately, it was not possible to join both families with the transformation rule introduced by Proposition 4.2.1. As seen in Subsection 4.2.1 we are able to combine the Frank and the Gumbel families, whereas the latter one is mostly stated as the second best choice for describing tail dependence. From the pre-simulation (see Subsection 7.4.1) for the threshold method for TDC estimation, we also see that the Gumbel copula is probably more robust against different underlying true models. However, the convex linear combination of the transformed Frank copula (4.2.3) with a survival transformed Frank copula, as given by equation (4.2.4) will be our favored model in the next chapters.

As benchmark models we take the convex linear combination of a Clayton copula with a survival Clayton copula and denote it by C_C

$$C_C(u,v) = \alpha \cdot C_{Cl_\delta}(u,v) + (1-\alpha) \cdot \widetilde{C_{Cl_{\delta_s}}}(u,v) \text{ with } 0 \leq \alpha \leq 1\,, \tag{8.1.1}$$

where $\delta, \delta_s \geq 0$. Due to Corollary 4.2.7, C_C is lower tail dependent with parameter $\lambda_L = \alpha \cdot 2^{-\frac{1}{\delta}}$ and upper tail dependent with $\lambda_U = (1-\alpha) \cdot 2^{-\frac{1}{\delta_s}}$. The big disadvantage in using the Clayton copula is that in the absence of tail dependency, i.e. $\delta \to 0$, we in the limit end up with the independence copula C_Π. Hence, the Clayton model and with it C_C, can not adapt any dependency in the absence of

tail dependency.
Further we investigate the t-copula C_t, given by

$$C_t(u_1, \ldots, u_n) = T_{\nu,\rho}({t_\nu}^{-1}(u_1), \ldots, {t_\nu}^{-1}(u_n)) \,, \tag{8.1.2}$$

where $T_{\nu,\rho}$ is the multivariate standard Student's t-distribution with ν degrees of freedom and correlation matrix ρ, and t_ν is the univariate standard Student's t-distribution. C_t is upper and lower tail dependent with TDC given by equation (2.3.2).
So we have for 2-dimensions the 4 parameter model C_Ω, the 3 parameter model C_C and the 2 parameter model C_t. Remark that we have covered the Frank, Gumbel, survival Gumbel, Clayton, survival Clayton and normal Copula as special cases of our three models.

For the 4-dimensional case we have two possibilities to construct the multidimensional extensions of C_Ω and C_C that we have called model a) (4.3.3) and model b) (4.3.4). In both cases, the modelling is done on 3 planes, where the third plane in model b)—in contrast to model a)—is very restrictive. We stress that model a) is in a sense a more probable (and flexible) choice since less joint equality is proposed (3 identical pairs vs. 4 identical pairs), see also Lemma 4.3.4 for that. In terms of entropy, model a) has intuitively the higher entropy and so is capable for the higher mixture of states. When taking the expected value for the (negative) entropy over the admissible parameter space (see Lemma 4.3.7) we compute e.g. for the Clayton type copulas $C_{Cl}^{[4]}$ ($\delta_3 < \delta_2 < \delta_1$) and $C_{Cl}^{[[4]]}$ ($\delta_3 < \delta_1$ and $\delta_3 < \delta_1$)

$$\int_0^\infty \int_{\delta_3}^{\delta_1} \int_0^\infty \int_{[0,1]^4} (-\log c_{Cl}^{[4]}(u_1, u_2, u_3, u_4)) dc_{Cl}^{[4]}(u_1, u_2, u_3, u_4) df(\delta_1, \delta_2, \delta_3) \tag{8.1.3}$$

$$\int_0^\infty \int_0^{\delta_1} \int_0^{\delta_1} \int_{[0,1]^4} (-\log c_{Cl}^{[[4]]}(u_1, u_2, u_3, u_4)) dc_{Cl}^{[[4]]}(u_1, u_2, u_3, u_4) df(\delta_1, \delta_2, \delta_3), \tag{8.1.4}$$

where f is the density of the distribution of the parameter vectors. Because of $\tilde{c}(u) = c(1-u)$, the entropy of the survival copula coincides with the entropy for the copula. In general, the above equations read $\mathbf{E}_{\text{param}}(\mathbf{E}_u(-logc_a))$ and $\mathbf{E}_{\text{param}}(\mathbf{E}_u(-logc_b))$. Of course we can in the same way derive the entropy for the t-copula model. If now a researcher has an intuition about what parameter vectors are likely to be realized under the different choices of models, he or she can specify the distribution density of the parameter vectors, calculate the negative model entropies and take the model with the lowest one. Here we do not have

such an intuition and therefore provide a second way how to choose the model.

The above model choice over the model entropy is only possible if one feels inclined to choose a priori a distribution of the parameter vectors. Despite this fact we further need to choose the order of the variables for the multidimensional "Archimedean" case, see Subsection 4.3.1. When mapping the risk factors (assets) to the variables, the order in which this is done makes a difference in the quality of the multivariate fit. In the following we show how bivariate results may help for the right model and order choice. First we summarize the parameter conditions for the $C_{\Omega}^{[4]}$, $C_{\Omega}^{[[4]]}$ and $C_{C}^{[4]}$, $C_{C}^{[[4]]}$ models according to Lemmas 4.3.4 and 4.3.7. Therefore we remark that in 4 dimensions we have 6 bivariate marginals (projections) and we will denote them by the indices (i,j), $i<j$, $i=1,2,3$, $j=2,3,4$. So if we e.g. model the (i,j) bivariate margin with a Clayton copula, i.e. we suppose the copula of (X_i, X_j) to be a Clayton copula, we will achieve the parameter $\delta_{(i,j)}$ and have the bivariate projection to be lower tail dependent with $\lambda_{L_{(i,j)}} = 2^{-\frac{1}{\delta_{(i,j)}}}$. Further recall from Subsection 4.3.1 that for e.g. $C_{\Omega}^{[4]}$ and $C_{\Omega}^{[[4]]}$ all bivariate marginals have the form C_{Ω}.

Lemma 8.1.1. *The parameters of $C_{C}^{[4]}$ (model a)) have to fulfill all of the following conditions.*

- $\alpha \equiv \alpha_{(1,2)} = \alpha_{(1,3)} = \alpha_{(1,4)} = \alpha_{(2,3)} = \alpha_{(2,4)} = \alpha_{(3,4)} \in [0,1]$,
- $\delta_3 \equiv \delta_{(1,4)} = \delta_{(2,4)} = \delta_{(3,4)} < \delta_2 \equiv \delta_{(1,3)} = \delta_{(2,3)} < \delta_1 \equiv \delta_{(1,2)}$,
- $\delta_{3_s} \equiv \delta_{(1,4)_s} = \delta_{(2,4)_s} = \delta_{(3,4)_s} < \delta_{2_s} \equiv \delta_{(1,3)_s} = \delta_{(2,3)_s} < \delta_{1_s} \equiv \delta_{(1,2)_s}$.

For $C_{\Omega}^{[4]}$ we have additionally

- $\theta_3 \equiv \theta_{(1,4)} = \theta_{(2,4)} = \theta_{(3,4)} < \theta_2 \equiv \theta_{(1,3)} = \theta_{(2,3)} < \theta_1 \equiv \theta_{(1,2)}$.

The parameters of $C_{C}^{[[4]]}$ (model b)) have to fulfill,

- $\alpha \equiv \alpha_{(1,2)} = \alpha_{(1,3)} = \alpha_{(1,4)} = \alpha_{(2,3)} = \alpha_{(2,4)} = \alpha_{(3,4)} \in [0,1]$,
- $\delta_3 \equiv \delta_{(1,3)} = \delta_{(1,4)} = \delta_{(2,3)} = \delta_{(2,4)} < \delta_1 \equiv \delta_{(1,2)} \quad \wedge \quad \delta_3 < \delta_2 \equiv \delta_{(3,4)}$,
- $\delta_{3_s} \equiv \delta_{(1,3)_s} = \delta_{(1,4)_s} = \delta_{(2,3)_s} = \delta_{(2,4)_s} < \delta_{1_s} \equiv \delta_{(1,2)_s} \quad \wedge \quad \delta_{3_s} < \delta_{2_s} \equiv \delta_{(3,4)_s}$.

For $C_{\Omega}^{[[4]]}$ we have additionally

- $\theta_3 \equiv \theta_{(1,3)} = \theta_{(1,4)} = \theta_{(2,3)} = \theta_{(2,4)} < \theta_1 \equiv \theta_{(1,2)} \quad \wedge \quad \theta_3 < \theta_2 \equiv \theta_{(3,4)}$.

So one can estimate the parameters of the 2-dimensional models of all (here 6) bivariate marginals (i, j) in a first step and then check whether the conditions of model a) or model b) are better satisfied. As a thumb rule we can say, when one of the bivariate pairs is dominant (in the sense of positive dependence), model a) is a good choice, whereas model b) fits well when 2 dominant bivariate pairs are present. Here we are mainly interested in a good tail adaption and since for C_C and C_Ω the tail dependence coefficients are functions of α, δ and α, δ_s, respectively, we will couple their impact in terms of the TDCs. The boundary regions of the TDCs are summarized by the next lemma and follow directly from Lemma 8.1.1 and Corollary **??**.

Lemma 8.1.2. *For a copula in dimension 4 the following restrictions must hold.*

(a) The tail dependence coefficients of the bivariate projections of the models of type a) $C_\Omega^{[4]}$ and $C_C^{[4]}$, are restricted by:

a1) $\lambda_{L_{(1,2)}} + \lambda_{U_{(1,2)}} \leq 1$,

a2) $\lambda_{L_3} \equiv \lambda_{L_{(i,4)}} < \lambda_{L_2} \equiv \lambda_{L_{(j,3)}} < \lambda_{L_1} \equiv \lambda_{L_{(1,2)}}$ *for* $i = 1, 2, 3$ *and* $j = 1, 2$, $\lambda_{U_3} \equiv \lambda_{U_{(i,4)}} < \lambda_{U_2} \equiv \lambda_{U_{(j,3)}} < \lambda_{U_1} \equiv \lambda_{U_{(1,2)}}$ *for* $i = 1, 2, 3$ *and* $j = 1, 2$, $\lambda_{(L,U)_{(i,4)}} = \lambda_{(L,U)_{(k,4)}}$, *for* $i, k = 1, 2, 3, 4$, *and* $\lambda_{(L,U)_{(j,3)}} = \lambda_{(L,U)_{(\ell,3)}}$, *for* $j, \ell = 1, 2$

(b) The tail dependence coefficients of the bivariate projections of the models of type b) $C_\Omega^{[[4]]}$ and $C_C^{[[4]]}$, are restricted by:

b1) $\lambda_{L_{(1,2)}} + \lambda_{U_{(1,2)}} \leq 1$ *and* $\lambda_{L_{(3,4)}} + \lambda_{U_{(3,4)}} \leq 1$,

b2) $\lambda_{L_3} \equiv \lambda_{L_{(i,j)}} < \lambda_{L_1} \equiv \lambda_{L_{(1,2)}}$, $\lambda_{U_3} \equiv \lambda_{U_{(i,j)}} < \lambda_{U_1} \equiv \lambda_{U_{(1,2)}}$ *and* $\lambda_{L_3} \equiv \lambda_{L_{(i,j)}} < \lambda_{L_2} \equiv \lambda_{L_{(3,4)}}$, $\lambda_{U_3} \equiv \lambda_{U_{(i,j)}} < \lambda_{U_2} \equiv \lambda_{U_{(3,4)}}$, *for* $i = 1, 2$ *and* $j = 3, 4$,
$\lambda_{(L,U)_{(i,j)}} = \lambda_{(L,U)_{(k,\ell)}}$, *for* $i, k = 1, 2$ *and* $j, \ell = 3, 4$

From a1) and a2) and from b1) and b2) we deduce $\lambda_{L,k} + \lambda_{U,k} \leq 1$ *for* $k = 1, \ldots, 4$.

Remark that the first condition is naturally fulfilled due to the restrictions of the convex linear combination construction of the bivariate C_C and C_Ω, as stated in Corollary 4.2.7. Remark further, that the $\widehat{\theta}_{(i,j)}$ should not systematically interfere with the estimates of $\widehat{\alpha}_{(i,j)}$, since we had set $\theta = \theta_s$ because of the rotational symmetric impact of the "Frank parameter" θ.
Later on we will use the bivariate information in terms of the estimated parameters $\theta_{(i,j)}$ and the estimated TDCs $\lambda_{(L,U)_{(i,j)}}$ to choose the model and to enumerate the assets in an optimal way.

For the limitations of the 4-dimensional t-copula see Chapter 3, for the restrictions of possible TDCs of the bivariate projections see Remark 6.
Regardless of wether we choose model a) or model b) we will end up for our 4-dimensional models with a 10 parametric C_Ω type, a 7 parametric C_C type and the 7 parametric t-copula.

8.1.2 Estimation method

As already mentioned in Section 7.3.1 the copula concept allows for the splitting of the estimation of the marginal parameters and the copula parameters. Here we will follow such a two step estimation, where we model the univariate marginals by their empirical distribution functions and use the estimated pseudo observations $\hat{u}$ to estimate the copula parameters by a maximum likelihood procedure. So we end up with the "Semi-parametric estimation - Two-step estimate" of Section 7.3.1. Copula densities often show numerical critical behavior in the lower and upper corners, sometimes, e.g. for the Gumbel copula, the densities have here singularities. To avoid numerical problems we transfer the ML estimation to a unimodal density with no singularities. This is done by fitting a distribution with standard normal marginals and the copula model to be estimated, to the pseudo observations $(\Phi^{\leftarrow}(\hat{u}_1), \ldots, \Phi^{\leftarrow}(\hat{u}_n))$, where $\Phi^{\leftarrow}$ is the inverse of the univariate standard normal distribution. Since the ML functions of this and the "Semi-parametric estimation - Two-step estimate" are only differing by a constant not depending on the parameters of the copula model, both estimates are theoretical the same. This procedure also leads to a better graphical interpretation of copula scatterplots. As seen in Figure 8.1 the asymmetric tail behavior is much better to figure out in the right plot, representing the pseudo observations of a distribution with the copula of Nasdaq - S&P 500 and standard normal marginals, than in the left one representing the scatterplot of the pseudo copula observations directly.

To achieve a start estimator we used the Least Square method introduced at the end of Subsection 6.1.3, where the number of runs of the Monte Carlo Simulation to determine the expected distances were set to 10,000. Furthermore we restricted the estimation to the deviations from the next neighbor distances to the corresponding expected value. Remark that one can adapt the number of runs to the length of the used data set, since the variance of the distances decreases with the number of data m.

8.1.3 Diagnostics

To validate and to compare our estimated models we use several goodness-of-fit tests as presented in Chapter 6. These are the χ^2 test (with cubic class separation

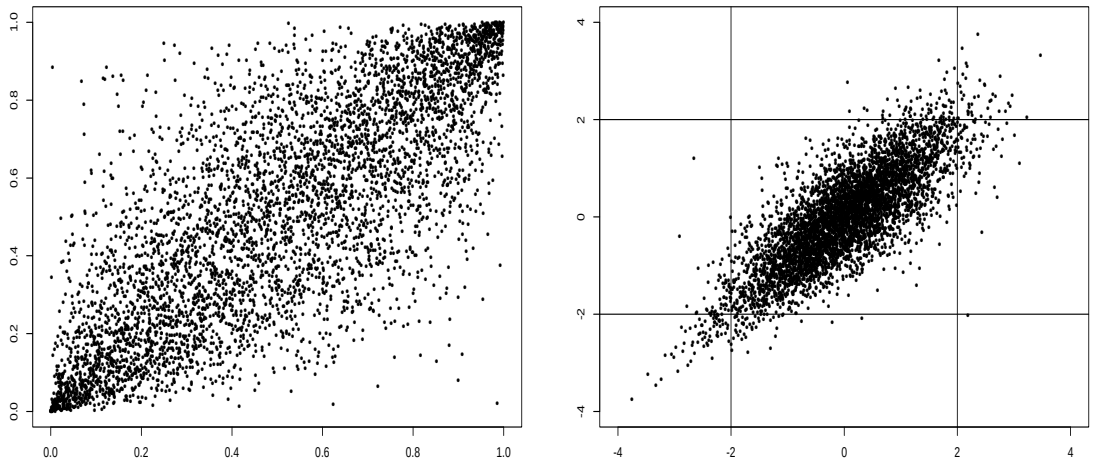

Figure 8.1: Scatterplot of the Nasdaq - S&P 500 copula pseudo observations (left plot) and scatterplot of the pseudo observations of a distribution with standard normal marginals and the copula of Nasdaq - S&P 500 (right plot).

according to Subsection 6.1.2), the AIC, BIC criterions (see equations (6.2.2), the Entropy statistics (6.2.3), the conditional integrated Anderson Darling statistics and the corresponding survival form (6.2.4) and (6.2.5) and finally the probability deviation measure PD and the survival form PDS given by equation 6.2.7. Here the χ^2 test runs over 81 cells and the integrals for the Entropy and Anderson Darling statistics in Equations (6.2.3), (6.2.4) and (6.2.5) are solved numerically by Monte Carlo integration over 100,000 points. For an overview of Monte Carlo integration we refer to Sobol' [108]. The PD and PDS statistics were calculated for $p = 0.01$, i.e. under the hypothesis we have 1% of the data set in the lower tail covering region and 1% in the upper tail covering region, respectively.

From Section 7.3.1 we recall that the used two step approach does not lead to real ML estimates. We stated that Genest et al. [53] has shown that the estimates are still asymptotically normal distributed, with a VCV matrix given by a mixture of the ML copula VCV matrix with the VCV matrix representing the fluctuations of the empirical univariate uniform estimates. For details we refer to Genest et al. [53]. In this thesis we stated the standard errors and resulting t-values calculated by the approach of Genest et al.. For the TDC estimates as functions of the parameter estimates we face the problem that the distribution of the TDC estimate can not concluded to be normal, since the parameter estimates are not true ML estimates. The standard errors are here calculated by a Taylor expansion

of first order. To achieve some significance statements for the TDC estimates we use a Monte Carlo simulation over the estimated asymptotic normal distribution of the parameter estimates with 1 million runs, and achieve 1 million pseudo realizations of lower and upper TDC estimates. Instead of a t-value we state now the empirical probability of the TDC Monte Carlo observations, to be greater than zero. We denote this probability by p_L for the estimated lower TDC and by p_U for the upper, respectively. If $0 \leq p_{L,U} \ll 1$ there is evidence for $\lambda_{L,U} = 0$, contrary $\frac{1}{2} \ll p_{L,U} \leq 1$ indicates a significant lower (upper) tail dependence. To check the observed portfolios for TDC asymmetry we run like above a Monte Carlo simulation to generate pseudo observations of $\lambda_{\text{diff}} \stackrel{def}{=} \lambda_L - \lambda_U$. We now define $p_{asy} \stackrel{def}{=} 2\left|\mathbf{P}(\lambda_{\text{diff}} > 0) - \frac{1}{2}\right|$. Remark that p_{asy} is only the probability of TDC asymmetry, but not the probability of TDC asymmetry *in the estimated direction*. If p_{asy} is close to 0 there is no significant TDC asymmetry, whereas for values close to 1 we conclude evidence for TDC asymmetry. We further define $p_{asy.real} \stackrel{def}{=} p_L p_U p_{asy}$, what measures the probability of asymmetric significant TDCs. Whenever $p_{asy.real}$ is small compared to p_{asy} we can argue that probably the convex linear combination in the models C_Ω, C_C is not necessary and either C_ω, C_{Cl} or their survival forms would have been also proper models to describe the tail dependence behavior.
In the multivariate context we derive the significance of the chosen model — a) or b)— by observing e.g. $\delta_3 - \delta_2$. Remark that our estimates of δ_3, δ_2 are asymptotically normal distributed according to Genest et al. [53], hence their difference is too. So we can indicate the significance of the estimated model by checking the conditions of Lemma 8.1.1 due the calculated t-values of e.g. $\delta_3 - \delta_2$. If they and the corresponding differences are all significant lower than zero, then we call the model significant. If the differences are only significant for certain planes or parameters, we call it partially significant in the corresponding plane and parameter (e.g. if only $\delta_3 - \delta_2$ is significant lower zero for $C_\Omega^{[4]}$, the model is δ significant in the (1,2,3) - 4 planes).

8.1.4 Data sets, data preparation and filtering

The Anderson Darling and probability deviation statistics does not allow for penalizing an overparameterized model. It is natural that the AD, ADS, PD and PDS statistics will be better the more parameters the hypothesis model has, if the test runs over the same data as the parameters were estimated from. So at least these statistics should be calculated *out of sample* on a so called test set. Also the statistics having a correction term for in sample tests, are more powerful if derived out of sample. For this reason we divided in our investigations the data sets in a trainings (estimation) set and a testset. However the above statement

is of course only true, if trainings and testset are samples of the same stochastic process, i.e. there were no change points during the observation time. Unfortunately this is not the case for all of our data sets. Whenever a remarkable high increase of the e.g. χ^2 statistic can be observed for all models when changing from the trainings to the testset, one should be alarmed that the iid assumption of the multivariate innovations should be rejected. For such cases the only reliable conclusions can be made with the help of the χ^2, AIC and BIC statistics. As an example Figure 8.2 shows the copula pseudo observations of the innovations of the HKD-YEN portfolio on the trainings set (left) and on the test set (right). Obviously there has been a dependence change point from weak to significant high positive dependence.

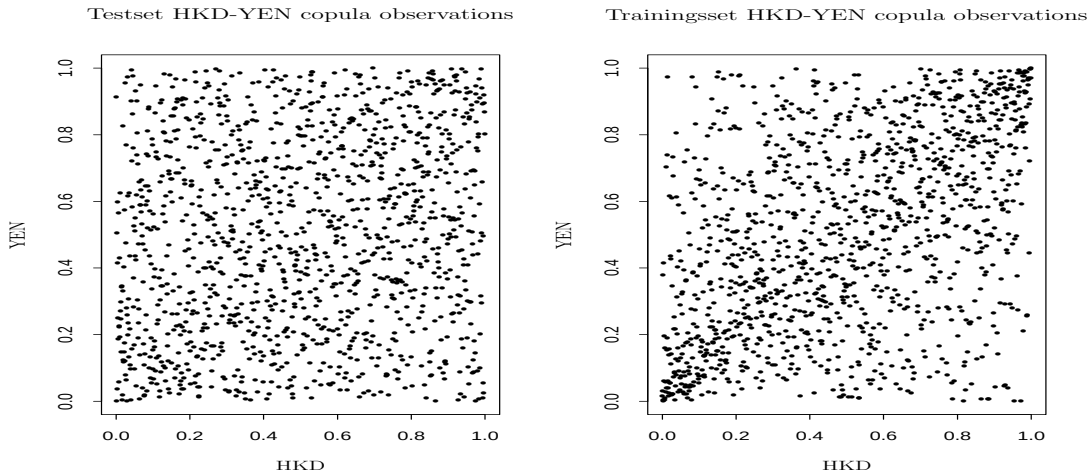

Figure 8.2: Testset and trainingsset of the HKD-YEN copula observations.

Time interdependence of the univariate time series (the dependence structure is assumed to be constant in time) is detected via plotting the autocorrelation function (ACF) of the log-returns against the time lags to check for autocorrelation, and by plotting the ACF for the squared log-returns to check for heteroscedasticity. Weekend effects were neglected. In the following we look at the pseudo-observations resulting from a time-interdependence filter (e.g. AR(1)-GARCH(1,1)) as innovations. It is common practise to handle these innovations as true observations, i.e. to ignore estimation errors of the filter. Using this approach we further neglect the error influences on the test statistics and estimation procedures of the joint innovations distribution.
All used data are daily quoted and adjusted for non trading days. Here for every portfolio a non trading day is defined as a day where at least one of the assets in the portfolio was not traded. We now come to the data itself.

The stock indices are adjusted with respect to dividend payments and splits. The Nasdaq and S&P 500 are observed at the New York stock exchange, the Nikkei quotes are from Tokyo and the Hang-Seng quotes from Hong Kong. According to Section 2.4 we minimized the time gap in-between the components of the multidimensional observations by joining the close quotes in New York of the previous day with the open quotes in Asia. All indices show heteroscedasticity but no autocorrelation. Hence the log-returns were filtered by a GARCH(1,1) process. The testset covers the time 01.01.1993 - 30.09.1996 what gives, depending on the country, 870-950 points. The trainingsset covers the time period 01.10.1996 - 30.06.2000 with 870-950 points (depending on the country). The data are provided by Thompson Financial data stream.
The foreign exchange (fx) rates of the Australian dollar (AUD), the Singapore dollar (SGD), the Hong Kong dollar (HKD) and the Yen (YEN) are measured in EUR and were downloaded from www.markt-daten.de/daten. They show heteroscedasticity but no autocorrelation, hence we used a GARCH(1,1) filter again. Testset: 01.01.1993 - 31.12.1997 (1289 points); trainingsset: 01.01.1998 - 31.12.2002 (1279 points).
The metal (cash (spot)) prices of aluminium (Al), copper (Cu), lead (Pb) and zinc (Zn) are stated in USD per tonne and were quoted at the London Metal Exchange (LME). They were downloaded from www.lme.co.uk before the relaunch of the webside (now these data are not available in the used form anymore). The data show heteroscedasticity and autocorrelation. To filter the autocorrelation we used an autoregressive process (AR) of order one (the order were chosen by an AIC criterium) and for the heteroscedasticity we utilized a GARCH(1,1) filter. Testset: 01.01.1990 - 31.12.1995 (1513 points); trainingsset: 01.01.1996 - 31.12.2002 (1764 points).
The (cash (spot)) prices of the oil related commodities, kerosine, diesel, heatingoil and gasoline, are quoted in US cent per gallon in New York. All data show heteroscedasticity (again we use a GARCH(1,1) filter). In addition, diesel and gasoline follow an AR(1) process, where the heatingoil log-return series follows an AR(4) process. Testset: 01.01.1996 - 30.06.1999 (874 points); trainingsset: 01.07.1999 - 31.12.2002 (873 points)

8.2 Estimation results and discussion

The computational estimation and testing routines were written in S-Plus version 5.1. In the next two subsections we state our empirical findings for the described bivariate and 4-dimensional portfolios.

8.2.1 2-dimensional data

Table D.1 to D.7 in the Appendix show the empirical results for the bivariate portfolios. For each of the 4 classes of portfolios with different risk factors we show in a first table the estimated parameters with standard error and t-value (estimated on the trainingsset). A second table shows the results of the different goodness-of-fit tests (tested over the trainingsset and testset). The best values are bold typed.
All sums $\lambda_L + \lambda_U$ were well below 1, even if we add as a puffer 2 times the standard error we only achieve $\max\{\lambda_L + \lambda_U + 2 \cdot s.e.\} = 1.0128 \approx 1$ for the C_C model and the (heatingoil,diesel) portfolio. We conclude that the restriction from Corollary 4.2.7 for the application of the convex linear combination, seems not to be of empirical evidence.
All observed bivariate dependence structures show positive dependence, at least independence. So we have no doubt in using the PD and PDS statistics as defined in Chapter 6. Especially there is no a priori objection against the direct use of the multidimensional "Archimedean" models.
Table D.14 summarizes the minimum and maximum values of p_{asy} and $p_{asy.real}$. We use this information when discussing now the TDC and TDC asymmetry properties of the different risk factors.

For the stock indices portfolios we have not surprisingly the strongest dependence for the only one country portfolio Nasdaq - S&P 500 (see Table D.1). Here we also detect the highest upper and especially lower tail dependence. This pair reveals significant asymmetric TDCs under the C_Ω model, beside the Nasdaq - Hang Seng and S&P 500 - Hang Seng pairs that show evidence for TDC asymmetry (see Table D.14). Surprisingly the later two (USA - Asia) pairs show stronger dependence than the Asian market pair of Hang Seng - Nikkei.
The TDC asymmetry is generally smaller under the C_C model, causing in the lag of mass in the middle of the distribution. However, all "Frank contributions" of the C_Ω models were estimated quite low for all portfolios, indicating that the dependence is mainly driven by the tails. Observing the test statistics in Table D.2, none of the models can really be rejected. For the TDC asymmetric Nasdaq - S&P 500 there is a favor for the C_Ω model, whereas for the other portfolios the results are rather mixed. Here the smaller asymmetries do not justify the two more parameters of C_Ω compared to C_t. The degrees of freedom of the estimated t-copulas are (despite of the Nasdaq - S&P 500 pair) quite large, indicating that the portfolios are not too far away from normal dependence, what confirms the results of Malevergne and Sornette [83].

The analysis of the portfolio with fx rates shows hardly TDC asymmetry in the data (see Table D.14). Only for the pair HKD - SGD there is some evidence for TDC asymmetry. The survey of Malevergne and Sornette [83] concluded that fx rates can be modelled by the normal copula (using the radial test as diagnostic). In contrast to them we have to neglect the normal hypothesis since all degrees of freedom of the estimated t-copulas are quite low and significant. Furthermore under all models the TDCs are significantly above zero (see Table D.3). Following the diagnostics of Table D.4 all models can cover the dependence structure. Hence the C_Ω model here has the disadvantage of too many parameters, that are consequently not or weakly significant according to the observed t-values. Only for the HKD-SGD and SGD-YEN portfolios we have a significant "Frank contribution" of the C_Ω models, that in the HKD-SGD case leads to a rejected C_C model (on a 95% level) with respect to the χ^2-test. Unfortunately, there seemed to be a dependence change point between trainings and testset, inferring that it is not possible to derive reliable out-of-sample statistics. At least we can say for the above two portfolios that the in sample statistics (especially PD and PDS) do not contradict the hypothesis of better tail adaption of C_Ω.
Overall we conclude again that the t-copula is an appropriate choice for modelling fx risk factor dependence. These findings coincide with the results of Breymann et al. [14] for a daily tick.

For bivariate MLE metal portfolios Malevergne and Sornette [83] were able to show that the normal copula is not an appropriate model to describe the dependence structure. In our survey we observed from the diagnostics of Table D.6 that also the t-copula is not a perfect choice. Since the estimated degrees of freedom are quite large (hence C_t is close to the normal copula) and the calculated TDCs are of moderate size under the C_t and C_Ω model, this is in line with the findings of Malevergne and Sornette [83] (see Table D.5). Overall the C_Ω model seemed to be the favorite. The observed TDC asymmetry is of minor, mostly negligible size (see Table D.14). So the C_Ω model displays here over parametrization in the tails, as seen by the low t-values of the estimates of δ_s and δ. That it is even the best choice, may lie in the significant high estimated "Frank contribution", represented by the parameter θ. This is a good example that even for non asymmetric data the behavior of the data need not necessarily be elliptical, as already mentioned in Subsection 8.1.1. The C_C model is missing such a term describing the "middle region" dependence and is therefore penalized in the goodness-of-fit tests.

For the portfolio with oil products the data show very clear lower tail dependence under all models, where the upper tail dependence is only significant for the C_C model and of course due symmetry for the t-copula (see Table D.7). For the

C_Ω model we see that all δ are estimated to the lower boundary of 1 and hence all upper TDCs λ_U are estimated zero. So probably the reduced 2 parameter model of the survival transformed Frank copula $\widetilde{C}_\omega$ would be sufficient. Consequently, we see under the C_Ω model in Table D.14 strong TDC asymmetry indicated in $p_{asy} \approx 1$, whereas the $p_{asy.real}$ values are not significant. The asymmetry is still present under the C_C model, but here of smaller size. The reason is that the θ's of C_Ω and the correlations of C_t are estimated quite large, indicating a high positive "global" dependence. The C_C model is missing a parameter describing such dependence and can only capture it with the joint increase of δ_s and δ, what has the effect of a joint increase of λ_L and λ_U. In the case of the e.g. kerosine - diesel portfolio this can even lead to a turn around of the TDC asymmetry direction.
However, the diagnostics in Table D.8 elect the C_Ω model as the winner. The t-copula is penalized for the missing possibility of asymmetric tails, the C_C copula for the missing "global" dependence structure.

8.2.2 4-dimensional data

We now turn to the discussion of the 4-dimensional portfolios. Before starting the estimation we show first how to derive the model choice —a) or b)— from the bivariate findings, and how to map the risk factors (assets) to the variables.

Model choice

Under the models a) and b) we have different boundary conditions for the parameters as outlined by Lemma 8.1.1. As mentioned in Subsection 8.1.1 we will use the bivariate information about the estimated $\theta_{(i,j)}$ and TDCs $\lambda_{(L,U)_{(i,j)}}$ to choose our model and to achieve the order in which the assets contribute to the model.

For the portfolio with stock indices we choose

$u_1 \;:\; S\&P500 \qquad u_2 \;:\; Nasdaq$

$u_3 \;:\; Hang-Seng \quad u_4 \;:\; Nikkei\,.$

Model b) is inappropriate because of the dominance of the pair Nasdaq - S&P 500. The order is the more or less best possible for the $C_\Omega^{[4]}$ and $C_C^{[4]}$ model, but still it is a problem to match the conditions in the (1,2,3) - 4 planes. However, from the bivariate findings we are not able to conclude a better arrangement.

For the portfolio with fx rates we choose

$u_1 \;:\; HKD \quad u_2 \;:\; SGD$

$u_3 \;:\; YEN \quad u_4 \;:\; AUD\,.$

The application of model b)—even with different ordering of the variables—fails because of the dominance of Hong Kong and Singapure Dollar (see Table D.3). For the above order, model b) is inhibited by the fact that $\lambda_{L_{(1,4)}}$ is very large compared to $\lambda_{L_{(3,4)}}$ and $\lambda_{L_{(2,4)}} >> \lambda_{L_{(3,4)}}$, respectively. So model a) is chosen for both, the C_Ω copula and the C_C copula.

For the portfolio with metals we choose

$u_1 \;:\; Al \quad u_2 \;:\; Cu$

$u_3 \;:\; Zn \quad u_4 \;:\; Pb\,.$

For the C_C copula this order turns out to be best. For the C_Ω copula the above order is good with respect to λ but is suboptimal with respect to θ where equality turns out to be the problem in the condition. One could try the order Cu, Zn, Al, Pb instead with the result that now the condition is perfect for θ but bad in λ. So it is not clear what the optimal choice for the C_Ω copula should be. For model b) other choices are worse with respect to θ, but still this order has a problem with the matching of Zn. We will continue to work with model a) and the above order for both models.

For the portfolio with oil products we choose

$u_1 \;:\; heatingoil \quad u_2 \;:\; diesel$

$u_3 \;:\; kerosine \quad u_4 \;:\; gasoline\,.$

Here model a) fulfils the conditions for both θ and λ very well, so the order will be applied for the C_Ω copula. The above order is applicable for the C_C copula as well. Model b) fails because of the dominance of the pair (heating oil, diesel).

The discussion shows that in general model a) is the best choice for our data.

Estimation

The empirical results are listed in the Appendix, Table D.9 to D.12. The test statistics are summarized in Table D.13 and Table D.15 state the values p_{asy} and $p_{asy.real}$ for the bivariate projections. The interpretation of the results for the different risk factors is as follows:

As seen from the model checks, the models $C_\Omega^{[4]}$ and $C_C^{[4]}$ are only partially significant in the (1,2)-3 plane, earlier one is also there missing model significance in the θ parameter (see Table D.9). The diagnostics consequently reject both the C_Ω and the C_C copula (see Table D.13). Only the t copula is clearly accepted by our diagnostics. Also when looking at the results for PD, PDS the restriction of the "archimedean" models are clearly shown. Having a look at the TDCs of the bivariate projections we see that under the models $C_\Omega^{[4]}$ and $C_C^{[4]}$ the 1-2 plane is nearly unchanged (from theory we have here unrestricted choice of the TDCs). In the (1,2)-3 plane especially the lower TDCs of the projections are estimated

significant smaller than under the bivariate models.

For the portfolio with fx rates, none of the proposed models is accepted by the diagnostics (see Table D.13). Again the t-copula is the best model with respect to the gof tests. This is in line with the bivariate findings, where we could not find strong evidence for TDC asymmetry and had no objection against the choice of the bivariate t-copula. The TDC estimates of the bivariate projections of the estimated t-copula show mostly reasonable small deviations compared to the bivariate survey. Only for $\lambda_{(2,4)}$, the projection on AUD-SGD, we detect a significant deviation (see Table D.10). In contrast to that the TDCs of the bivariate projections of $C_{\Omega}^{[4]}$ and $C_{C}^{[4]}$ differ remarkable from the bivariate findings. The models turned out to have significantly higher upper TDCs and smaller lower TDCs compared to the bivariate findings. Surprisingly, the somehow unrestricted $\lambda_{(L,U)_{(1,2)}}$ corresponding to the HKD-SGD plane, exhibit the largest deviations. Model $C_{C}^{[4]}$ reveals significance, whereas $C_{\Omega}^{[4]}$ is only significant in δ.

Based on the diagnostics, the t-copula approach leads the best fit for the metal portfolio among the three models (see Table D.13). Despite that fact also the other models perform much better than in the former two cases. Unfortunately there is a dependence change point between traings and testset, otherwise probably the $C_{\Omega}^{[4]}$ model could had score points in the Anderson Darling and PD, PDS statistics as argued by the in-sample tests. Corresponding to the bivariate findings the projections do not show any significant TDC asymmetry. Hence $C_{\Omega}^{[4]}$ can only achieve weak model significance, since the freedom of $\delta_s \neq \delta$ is unnecessarily. $C_{C}^{[4]}$ is model significant (the "missing" θ leads to more significant δ_s, δ). Compared to the size of the standard errors the TDCs of the bivariate projections of all models do not significantly differ from the corresponding bivariate estimates.

For the oil product portfolio no model can be accepted, but from the statistics $C_{\Omega}^{[4]}$ performs clearly best (see Table D.13). As already seen in the bivariate survey, the data exhibit strong tail asymmetry, that can not be captured by the t-copula. Here the impact is such strong that the statistics for the estimated C_t model is significantly the worst. Even though the flexibility of the C_{Ω} model is needed here with respect to asymmetry in the tails, there is no statistical fit of the model to the data. So in contrast to the bivariate case, the flexibility to fit the model to data with real asymmetry does not yield to an accepted model for higher dimension. The $C_{C}^{[4]}$ model is significant, whereas $C_{\Omega}^{[4]}$ is not significant in δ (no significant upper tail dependence). So analogously to the bivariate case a reduced 6 parameter $\widetilde{C^{[4]}}_{\omega}$ model would probably be sufficient. Under the $C_{\Omega}^{[4]}$

model the TDCs of the projections are, within the region of their standard errors, equivalent to the bivariate TDC estimates. For the $C_C^{[4]}$ model there are smaller TDC estimate deviations, for the C_t model significant ones.

8.3 Conclusion

In a 2-dimensional framework we were able to prove the existence of tail dependence asymmetry in stock indices portfolios, fx-rate and especially in oil related commodity baskets. For MLE metal portfolios we have seen that even in a rotational symmetric world, the Archimedean shape of a "Frank contribution" may be a better choice than the use of elliptical models. Despite these facts the t-copula, taken all aspects into account, is in most cases an appropriate model. The flexibility of the proposed C_Ω copula is only needed (in terms of gof-tests) in highly asymmetric data like the oil related commodities. Later on in Chapter 11 we will see that also the asymmetry of interest rate data justify the use of C_ω.
When modelling more than 2 dimensions working with Archimedean copulas shows some difficulties that are hard to overcome. The models turn out to be very restrictive, and a lot of knowledge about the correct choice of the order of the assets is necessary. The t-copula is the clear favorite for not too asymmetric multidimensional data. Despite its easy multidimensional extension, the restriction to rotationally symmetry seemed not to be as relevant as the restrictions of the "Archimedean" type copulas $C_\Omega^{[4]}$, $C_C^{[4]}$ as seen from several goodness-of-fit tests. So in contrast to the bivariate case, the use of such complex models is hard to justify. We stress that for data with real asymmetry in the tails (as the oil commodities) the diagnostics still allow for the conclusion that there is a better fit in this region but the overall goodness-of-fit test does not accept any model here.
In general we end up with the conclusion that the flexibility of Archimedean models is important to model bivariate asymmetric complex dependence structures, but their complexity and restrictions in multidimensional settings make them hard to handle, if not inappropriate to model multidimensional data. To capture asymmetry also for high dimensional data a generalized form of elliptical distributions, as later introduced in Chapter 12, is probably an alternative.

Chapter 9

Measurement of Aggregate Risk with Copulas

In this chapter, based on Junker and May [66], we apply the copula concept to Risk Management and quantify financial risk by standard risk measures like Value at Risk. For this purposes we show the typical procedure exemplarily for a portfolio of two stocks, namely Hoechst and Volkswagen.

9.1 Copula specification

The observed time period for our two stocks (Hoechst and Volkswagen) ranges from Oct. 23, 1989 to Sep. 29, 2000. This yields 2701 data points. Figure 9.1 shows a scatterplot representation for the GARCH innovations of the two stock portfolio, with it's empirical copula and standard normal margins. The plot indicate a possible TDC asymmetry, since joint losses occurred more often than joint gains.

We now use the two step approach with empirical margins (see Subsection 8.1.2) to achieve the copula estimates of our introduced standard models, i.e. of the linear convex combined transformed Frank copula (equation (4.2.4)), the linear convex combined Clayton copula (equation (8.1.1)) and the t-copula (equation (8.1.2)). The copula parameter estimates and χ^2 tests over 196 cells are summarized in Table 9.1 and approve our suspicion of lower tail dependence. The χ^2 test indicate the C_Ω model to be a good choice, but it hasn't enough power to reject the other models. However, in Figure 9.2 we see the absolute errors, i.e. the model probability to be in a test cell minus the empirical probability mass of this cell, over 196 cells for the estimated Hoechst-Volkswagen $\widehat{C}_\Omega$ and $\widehat{C}_t$ copula. The left errorplot ($\widehat{C}_\Omega$) indicates that there is no specific model error, whereas for the symmetric $\widehat{C}_t$ copula (right plot) the overestimated upper tail is remarkable visible. Another graphical method to examine the quality of the fit is given by

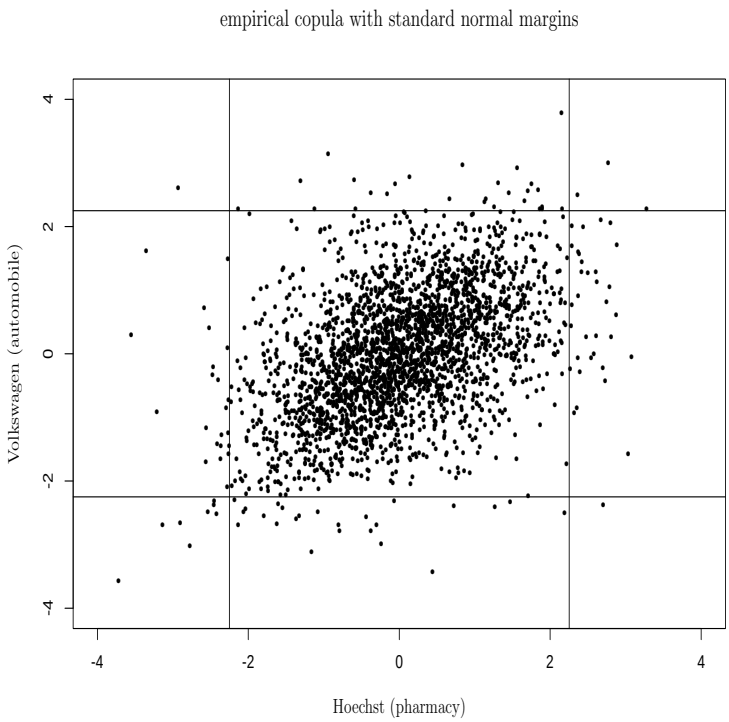

Figure 9.1: A scatterplot representation of the empirical Hoechst-Volkswagen copula with standard normal margins.

copula	$\widehat{\alpha}$	$\widehat{\delta_s}$	$\widehat{\theta}$	$\widehat{\delta}$	$\widehat{\lambda}_L$	$\widehat{\lambda}_U$	χ^2_{191}
C_Ω	0.4924	1.6841	1.5573	1.0000	0.2417	0	176.74
s.e. / p-value	(0.0876)	(0.1183)	(0.3258)	(0.0395)	(0.0457)	(0.0194)	(76.25)
t-value	(5.62)	(5.78)	(4.78)	-	(5.29)	-	
	$\widehat{\alpha}$	$\widehat{\delta}$		$\widehat{\delta_s}$	$\widehat{\lambda}_L$	$\widehat{\lambda}_U$	χ^2_{192}
C_C	0.5546	1.3048		0.5282	0.3260	0.1199	197.49
s.e. / p-value	(0.0442)	(0.1431)		(0.1008)	(0.0322)	(0.0323)	(37.77)
t-value	(12.54)	(9.12)		(5.24)	(10.13)	(3.71)	
		df	*cor*		$\widehat{\lambda}_L = \widehat{\lambda}_U$		χ^2_{193}
C_t		5.9776	0.4805		0.1618		208.67
s.e. / p-value		(0.7510)	(0.0157)		(0.0184)		(20.88)
t-value		(6.63)	(30.68)		(8.82)		

Table 9.1: Parameter estimates for the different copula models (fitted to the hoechst volkswagen sample), with resultant tail dependence parameters with standard errors (s.e.) and t-values in parantheses, and the χ^2-statistics with the appendant p-values (in parantheses) in percentage.

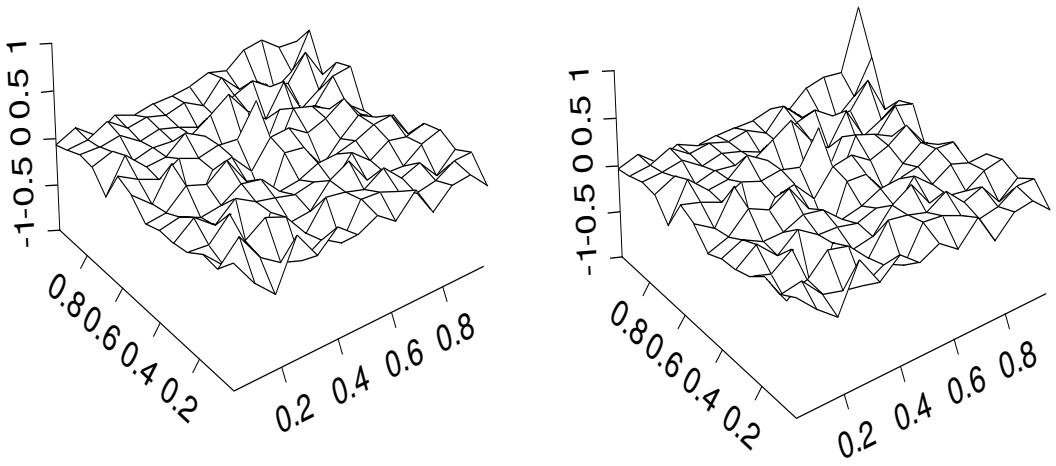

Figure 9.2: Absolute error of the estimated Hoechst-Volkswagen copula ($\widehat{C}_\Omega$ left plot, $\widehat{C}_t$ right plot) in percentage.

Figure 9.3, where the contourlines of the empirical copula are compared with the ones of the estimated $\widehat{C}_\Omega$, $\widehat{C}_C$ and t-copula. In contrast to the $\widehat{C}_\Omega$ and $\widehat{C}_t$ copulas which do not show a significant divergence to the empirical contourlines, the spaced dotted line of the $\widehat{C}_C$ copula clearly points out the problems of the missing flexibility in modelling uniform dependencies, leading to overestimated $(0,1)$ and $(1,0)$ corners.

In Figure 9.4, the estimated density of the $\widehat{C}_\Omega$ copula of the Hoechst-Volkswagen portfolio, clearly marks out the asymmetry of the tail dependencies.

9.2 Value at Risk calculation under different distribution assumptions

In this section we quantify the risk of a portfolio under different joint distribution assumptions. As risk measures we choose the Value at Risk (VaR) and Expected Shortfall, both widely used in practice. For a numerical example, we will come back to our equally weighted stock portfolio Hoechst-Volkswagen, i.e. the portfolio is assumed to contain the same amount of money in each risk factor (stock).

We start by presenting the profit and loss (P&L) distribution of a portfolio in a copula setting.

The portfolio wealth at time $t+1$ is given by

$$V(t+1) = V(t)\left(1 + \sum_{i=1}^{m} \alpha_i R_i(t+1)\right),$$

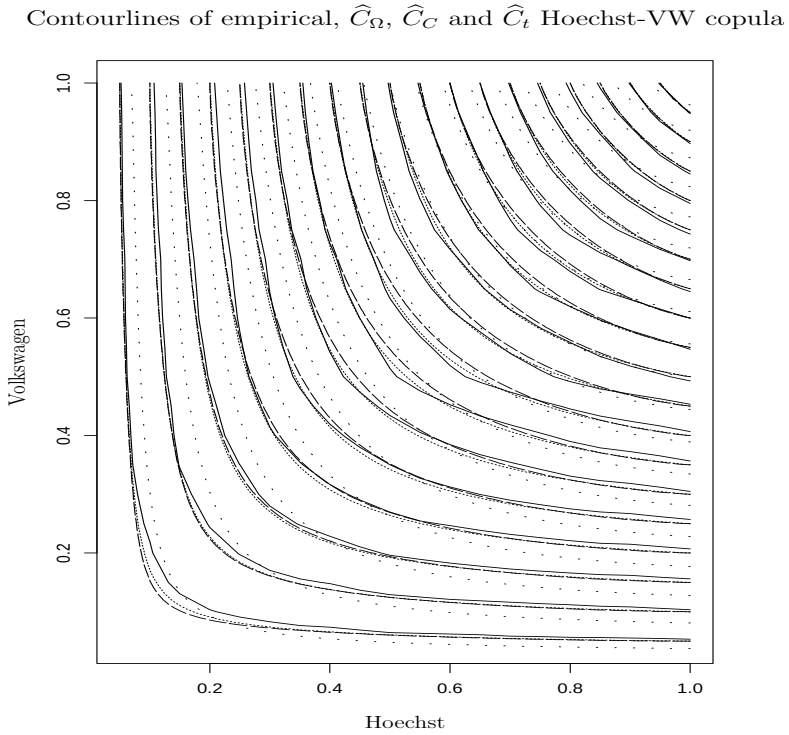

Figure 9.3: Contourlines of the Hoechst-VW empirical copula (solid line) compared with the contourlines of the estimated $\widehat{C}_\Omega$ copula (dotted line), $\widehat{C}_C$ copula (spaced dotted) and the $\widehat{C}_t$ (dashed line).

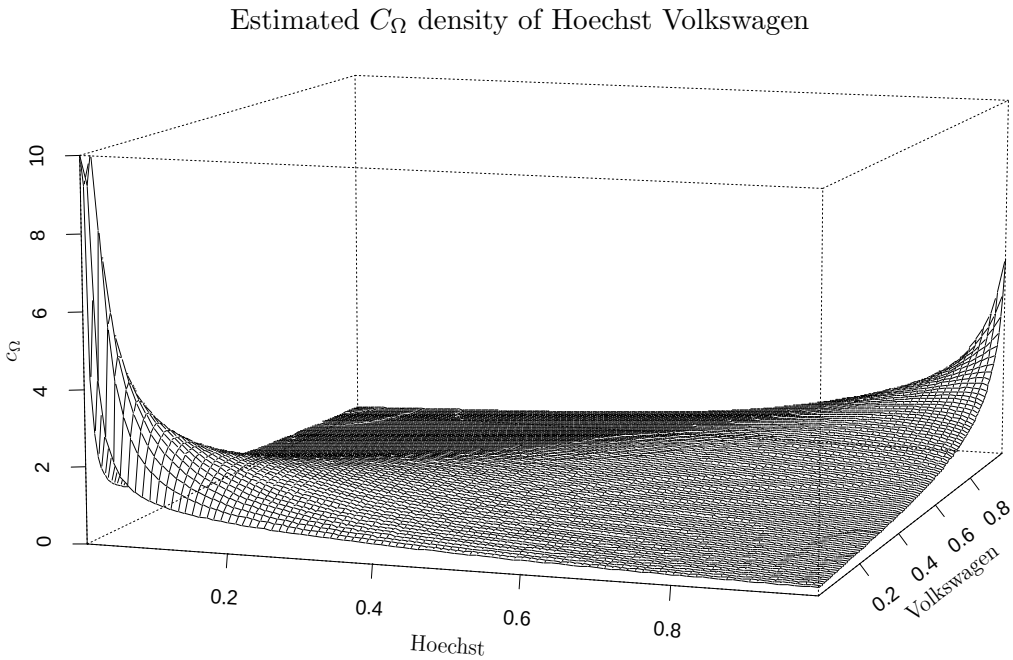

Figure 9.4: Estimated $\widehat{C}_\Omega$ density of the Hoechst Volkswagen portfolio.

where m is the number of assets (risk factors) in the portfolio, α_i stands for the percentage of asset i in the portfolio, and R_i are the (log)returns of asset i. With our model assumption, this leads to $R_i(t+1) = \mu_i(t, R_i(t)) + \sigma_i(t, R_i(t))\varepsilon_{i,t+1}$. The portfolio return R_P is now given by

$$R_P(t+1) = \frac{V(t+1)}{V(t)} - 1 = \sum_{i=1}^{m} \alpha_i R_i(t+1).$$

As a useful result, we show that the scenarios leading to a given R_P can be expressed explicitly.

Remark 15. *Let R_P be fixed. Then the respective scenarios can be represented as points on an $m-1$ dimensional hyperplane $\mathcal{H}$ in $[0,\ 1]^m$.*

Let $\{(u_1, .., u_{m-1}, u_m(\Theta, R_P, u_1, ..., u_{m-1})) \mid (u_1, ..., u_{m-1}) \in [0,\ 1]^{m-1}\} = \mathcal{H}$, where Θ contains all parameters of the marginals. Then $\mathcal{H}$ is determinded by the following formulae:

$$\begin{aligned}
& \tfrac{R_P}{\alpha_m} - \sum_{i=1}^{m-1} \tfrac{\alpha_i}{\alpha_m} R_i = R_m = \mu_m + \sigma_m \varepsilon_m \\
\Rightarrow\ & \varepsilon_m = \tfrac{1}{\sigma_m}\left(\tfrac{R_P}{\alpha_m} - \sum_{i=1}^{m-1} \tfrac{\alpha_i}{\alpha_m}\left(\mu_i + \sigma_i \varepsilon_i\right) - \mu_m\right) \\
\Rightarrow\ & u_m = F_{\varepsilon_m}\left[\tfrac{1}{\sigma_m}\left(\tfrac{R_P}{\alpha_m} - \sum_{i=1}^{m-1} \tfrac{\alpha_i}{\alpha_m}\left(\mu_i + \sigma_i F_{\varepsilon_i}^{\leftarrow}(u_i)\right) - \mu_m\right)\right]
\end{aligned}$$

The percentile of R_P is the volume of the copula density over the m dimensional plane

$$H_m(\Theta, R_P) := cl\big\{(u_1, .., u_m) \in [0,\ 1]^{m-1} \times [0,\ u_m^{\star}(\Theta, R_P, u_1, ..., u_{m-1})]\big\},$$

so the P&L distribution is given by

$$PL(R_P) = \int\limits_{H_m(\Theta, R_P)} c(u_1, ..., u_m)\, du_1 ... du_m\,. \tag{9.2.1}$$

The VaR can now be calculated by solving 9.2.1 numerically.

With $S = \big\{(u_1, ..., u_m) \mid \sum_{i=1}^{m} \alpha_i F_{R_i}^{\leftarrow}(u_i) \leq \mathrm{VaR}\big\}$, the Expected Shortfall reads

$$ES = \mu_P + \int\limits_{S} \left(\sum_{i=1}^{m} \alpha_i \sigma_i F_{\varepsilon_i}^{\leftarrow}(u_i)\right) \prod_{i=1}^{m} f_{\varepsilon_i}\left(F_{\varepsilon_i}^{\leftarrow}(u_i)\right)\, c(u_1, ..., u_m)\, du_1 ... du_m\,, \tag{9.2.2}$$

where $\mu_P = \sum_{i=1}^{m} \alpha_i \mu_i$ is the expected portfolio return.

When the R_i are log-returns, we have to transform the equations (9.2.1) and (9.2.2), respectively to

$$PL(\exp(R_P) - 1) = \int\limits_{H_m(\Theta, R_P)} c(u_1, ..., u_m)\, du_1 ... du_m\,. \tag{9.2.3}$$

and

$$ES = \mu_P + \int\limits_S \left(\sum_{i=1}^{m} \left(\exp\left(\alpha_i \sigma_i F_{\varepsilon_i}^{\leftarrow}(u_i) \right) - 1 \right) \right) \prod_{i=1}^{m} f_{\varepsilon_i}\left(F_{\varepsilon_i}^{\leftarrow}(u_i) \right) c(u_1, ..., u_m)\, du_1 ... du_m, \tag{9.2.4}$$

where this time $S = \left\{ (u_1, ..., u_m) \mid \sum_{i=1}^{m} \alpha_i F_{R_i}^{\leftarrow}(u_i) \leq VaR_{LR} \right\}$, and $VaR_{LR} = \ln(VaR) + 1$

If it is easy to generate random numbers from the assumed joint distribution of the portfolio, then it can be more efficient to use a Monte Carlo Method instead of numerically solving (9.2.1) and (9.2.2) respectively, to obtain VaR and Expected Shortfall.

We show some numerical results for a portfolio with equally weighted Hoechst and Volkswagen shares. In Table 9.2 we summarize the mean adjusted conditional VaR and conditional ES under different marginal distribution and copula assumptions. The stated empirical result means that we use the empirical distribution function to describe the innovations distribution and state the conditional "empirical" VaR and ES in that sense. Furthermore, the χ^2 statistics are given. Junker and Szimayer [68] propose an asymmetric univariate distribution function allowing for heavy tails. We used this distribution function F—obtained by a linear combination of transformed generalized pareto distributions—for the marginal fit. For the definition see 1.2.3, where here we have used the 6 parameter version. For more details see Chapter 1.

For an interpretation of Table 9.2, we first have a look at the χ^2 statistics under the different joint models. Independently from the marginal assumption, the proposed C_Ω copula shows the best results, whereas the normal copula C_N is worst. As a result of the joint parameter estimation for marginals and copula, the use of F_θ or Student's t-distribution yields a better fit than the empirical marginals. This states the crucial role of heavy tailed marginals that improve the quality of the fit even though the copula choice is not the optimal one. If one compares different copulas, the statistics worsen if we do not allow for asymmetric tail dependency, i.e. C_t instead of C_Ω, or tail dependency itself, i.e. C_N instead of C_t. So it seems to be as important to allow for tail dependency, as for asymmetric tail dependency. To give a brief summary: The joint fit can be improved by the use of marginal models that allow for heavy tails. In the sense of goodness of the fit, the possibility of asymmetric tails in the marginals has only minor influence.

Recall from Remark 11 that C_Ω is conditionally increasing. Then the following Theorem holds for C_Ω distributed random vectors. Müller and Scarsini [89]: If $X_i \leq_{cx} Y_i$ then the affine combination reads $\sum_{i=1}^{n} \alpha_i X_i \leq_{cx} \sum_{i=1}^{n} \alpha_i Y_i$ for all non-negative coefficients $\alpha_1, \ldots, \alpha_n$. In a portfolio setting, this means that for assets that have a common C_Ω copula the inequality for the single assets is transferred

Marginals	Copula	χ^2-Statistic (p-value)	99% & 1% VaR	99% & 1% ES
empirical model		-	-3.33% 3.01%	-4.47% 3.73%
empirical	C_Ω	176.74 (0.76)	-3.28% 3.11%	-4.48% 3.75%
	C_t	208.67 (0.21)	-3.33% 3.33%	-4.56% 4.12%
	C_C	197.49 (0.38)	-3.31% 3.12%	-4.67% 3.81%
	C_N	247.13 (0.01)	-3.30% 3.24%	-4.41% 3.94%
pareto-like (F_θ)	C_Ω	179.53 (0.47)	-3.30% 3.17%	-4.62% 3.81%
	C_t	203.84 (0.12)	-3.28% 3.50%	-4.37% 4.43%
	C_C	204.30 (0.10)	-3.50% 3.23%	-4.93% 3.98%
	C_N	217.03 (0.04)	-3.19% 3.35%	-4.20% 4.11%
t-distribution	C_Ω	188.77 (0.41)	-3.25% 3.22%	-4.15% 4.03%
	C_t	212.39 (0.10)	-3.32% 3.43%	-4.28% 4.50%
	C_C	211.98 (0.09)	-3.41% 3.32%	-4.43% 4.38%
	C_N	218.04 (0.07)	-3.20% 3.31%	-4.00% 4.21%
normal	C_Ω	286.44 (0.00)	-3.05% 3.00%	-3.53% 3.43%
	C_t	328.47 (0.00)	-3.10% 3.19%	-3.58% 3.72%
	C_C	369.45 (0.00)	-3.19% 3.10%	-3.73% 3.59%
	C_N	392.81 (0.00)	-2.92% 3.00%	-3.32% 3.44%

Table 9.2: Mean adjusted VaR and ES under different marginal distribution and copula assumptions for the 99% and 1% level respectively.

on a portfolio level.

We now turn to the question of accuracy in VaR estimation. The empirical VaR will be taken as a benchmark model. As a general result, we obtain from Table 9.2 that the tail dependence is the driving factor for the parameter estimates. Here the lower tail shows more mass and is hence fitted in a more accurate way whereas in contrary the less dependent (here: upper) tail shows greater deviation. This effect is most drastically shown when using parametric models in the marginals because some extremes are taken away by a cut in the empirical marginals.

In coincidence with our expectation, the normal copula overestimates the correlation and subsequently the win tail. The t-copula improves the fit in the loss tail but widely overestimates the win tail. The Clayton model overestimates both tails and provides no fit in the middle of the distribution.

When talking about Expected Shortfall as a more sensible risk measure, we also notice a not negligible effect of asymmetric tails in the margins if we compare e.g. the pairs (F_θ, C_Ω) and (t_ν, C_Ω). When talking about estimation error, we end up with the same result than in the VaR case. For example, under the naive normal assumption (Φ, C_N) the 99% ES is 25.7%, and the 1% ES is 7.8% underestimated, while under the more sophisticated model (F_θ, C_Ω) the 99% ES

is 3.7%, and the 1% ES is 2.1% overestimated. The overestimated tail dependencies under the C_C model (see also Chapter 8 on this subject) lead to significantly overestimated VaRs and Expected Shortfalls. This effect can compensate a misspecification in the marginals.

Note that we have stated the VaR and Expected Shortfall in the return world. In the log-return world, the symmetric models constructed by the C_t and C_N copula with either normal or t-distributed marginals would lead to identically absolute values of the 99% and 1% VaR, and Expected Shortfall respectively.

In Figure 9.5 we have plotted the contourlines of the estimated (F_θ, C_Ω) joint GARCH innovations density on the scatterplot of the Hoechst Volkswagen GARCH Residuals. One can clearly identify the impacts of fat lower tails in the marginals and the lower tail dependence.

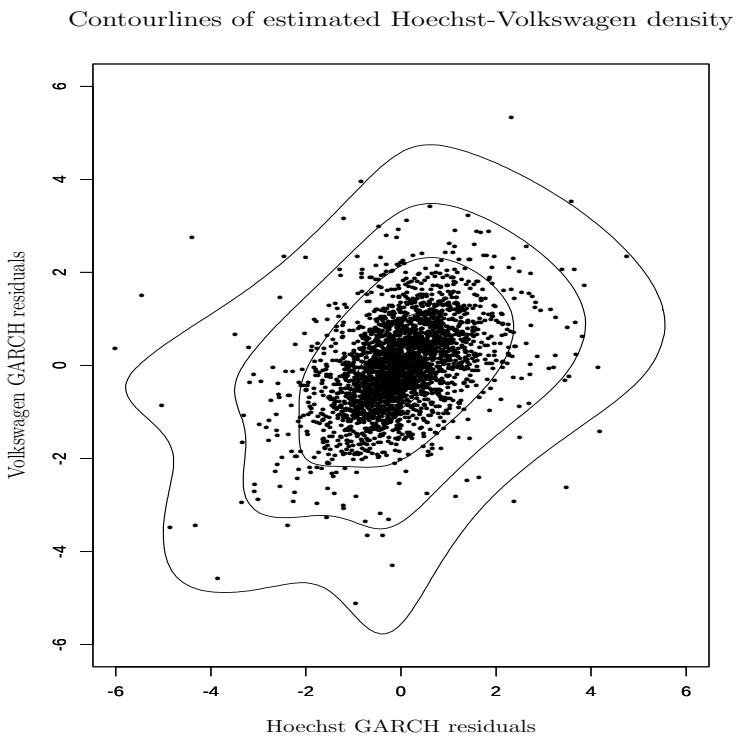

Figure 9.5: Contourlines of the estimated (F_θ, C_Ω) joint GARCH innovations density on the scatterplot of the Hoechst Volkswagen GARCH Residuals.

We follow equation (9.2.1) and compute the P&L-distribution under the different models. Figure 9.6 shows QQ-plots for some jointly modeled P&L's against the empirical P&L. For better comparison of the different tail adaption we stay here in the log-return world. Here, the (F_θ, C_Ω) model indicates only minor deviations, whereas the case (F_θ, C_C) is clearly influenced by the overestimated

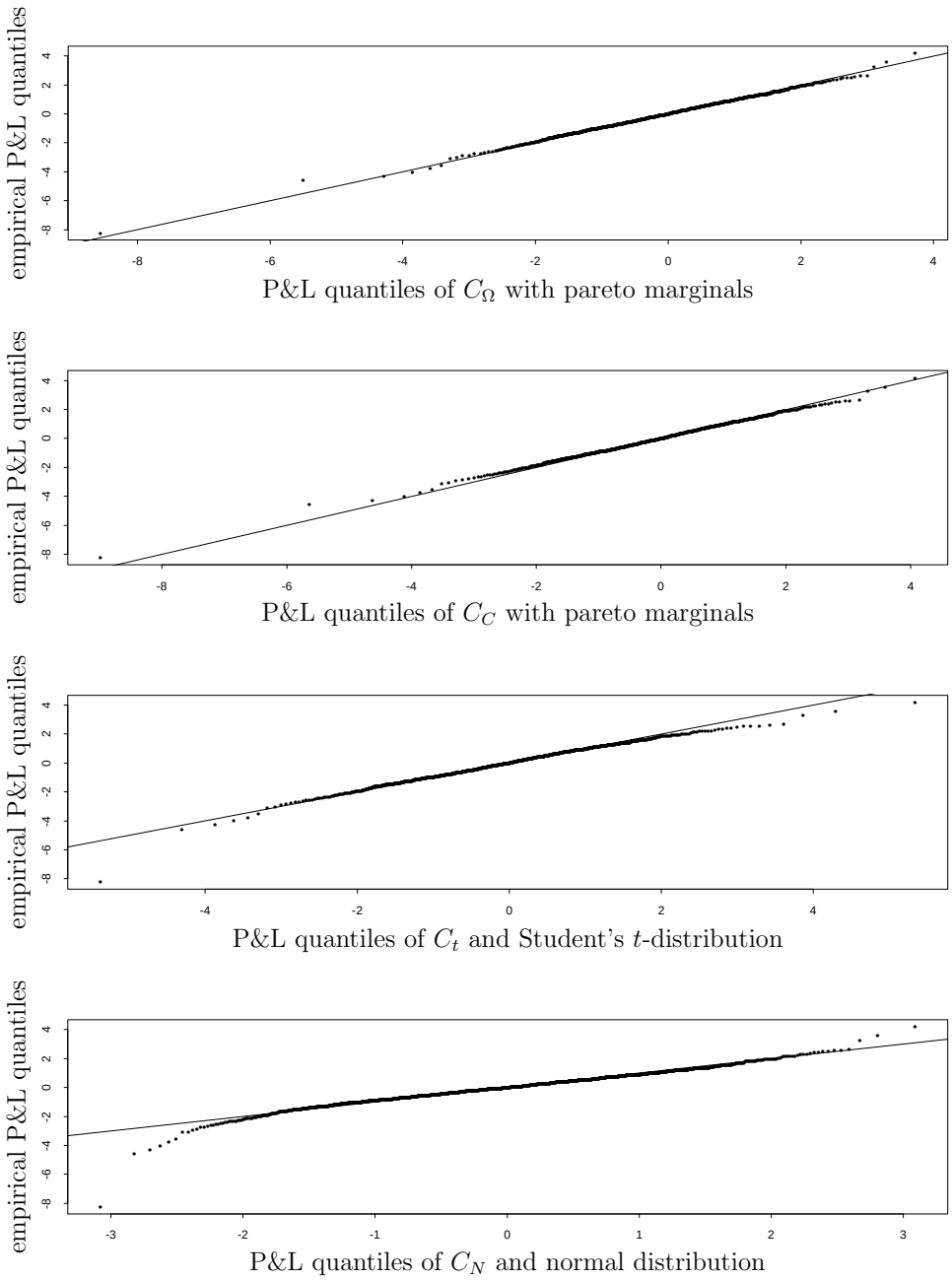

Figure 9.6: QQ-plots in the log-return world for the portfolio P& L's; empirical vs. different joint models.

tails.

The assumption of a Student's t-distribution leads to a significant overestimation in the win tail, since the lower tail dependence drives the parameter estimation and the symmetry of the model. The normal model lacks heavy tails in the margins, and together with the missing tail dependency this leads to a significant underestimation of the loss tail, whereas the win tail is adapted quite well. Here it seems that the overestimated correlation just covers the lag of slightly heavy win tails in the margins.

Chapter 10

European Real Estate Firms in Crash Situations

In this chapter, that is a copy of Junker [64], we analyse the crash behaviour of major European real estate firms in comparison to blue chips. The single asset risk characteristics in terms of volatility, heavy tailedness and Value-at-Risk are investigated via an extreme value theory approach. First we filter the data with a GARCH model to capture heteroscedasticity effects, then we measure the tail fatness of the residuals by adjusting a generalised Pareto distribution. The diversification effects of the admixture of real estate firms to stock portfolios are observed by correlation, Kendall's τ and tail dependence. To obtain an estimate for the tail dependence we fit a transformed Frank copula. We conclude that real estate firms generally show lighter tails than stocks and that their admixture in portfolios can gain a high diversification for daily returns, that even does not break down in crash situations.

10.1 Introduction

During the last years, incorporated real estate firms (REFs) in Europe have attracted growing attention. One reason is the breakdown of inter-generation contract based retirement pay, demanding state-aided, private retirement insurance, e.g., the Riester Rente in Germany, and giving banks and financial service companies the opportunity to establish large pension funds. Here the requirement of crash stability plays a leading role. Furthermore, insurance companies may have the same interest in order to obtain crash resistant investment strategies for their reserves. Whenever the focus is on portfolio diversification and crash stability, real estate based stocks are of increasing interest, since there is hope of gaining the liquidity and tractability of stocks combined with diversification and stability effects of real estates. Recent work has been done to analyse if this aspiration holds.

In Maurer and Sebastian [86] a portfolio of German REFs is compared to

the German stock index DAX, the German bond index REXP and a portfolio of German real estate funds. They observed an only slightly lower volatility of the REFs with respect to the DAX in contrast to a significantly lower volatility of the real estate funds. Furthermore, they found a significantly high correlation of the REFs with the DAX whereas there was no correlation with the real estate funds. These findings correspond to empirical studies done for US equity real estate trusts (EREITs), that are probably the best analogue to the European REF. Giliberto [55] also observed a high correlation between EREITs and stocks and no correlation of EREITs compared to real estates. By using the residuals of a linear regression he removed the market effects of stock and bond returns on the EREIT returns and showed that there is a significant positive correlation between the regression residuals and real estate returns. This enabled him to infer the existence of a common factor moving EREIT and real estate returns. Myer and Webb [91] extended this approach by using a vector autoregressive model and running a Granger causality test with the outcome that EREIT returns can Granger cause real estate returns. Furthermore, they analysed the stylized facts of some single EREIT returns and ran different tests of normality. It turned out that the normal hypothesis can not be discarded under most tests and only few data showed significant skewness and kurtosis.

In this paper, we focus on the behaviour of REFs in crash situations and compare them to blue chips. We investigate the single asset attributes as well as the effect of REF admixture to common stock portfolios. It is well known and examined that stock returns have the stylized fact of heavy tails, especially at the loss end. See, e.g., Danielsson and De Vries [26], Frey and McNeil [52] and Longin [79]. The findings by Myer and Webb [91] give a first hint that this is not true for EREIT returns. We use Extreme Value Theory (EVT) to proof that intuition. As guidelines to EVT, we refer to Resnick [100] and Embrechts et al. [39].

The unexpected high correlation with common stocks and the absence of correlation with real estates found in the references above are a drawback in the effort of using REFs as a diversification tool in common stock portfolios. However, we will not find such high correlation in our studies and the common factor driving the EREIT and real estate returns mentioned above is an indicator that there should be a diversification effect. Since we are mainly interested in the crash behaviour, we emphasize the dependence of extreme events and measure it using tail dependence. For example Ané and Kharoubi [3] and Junker and May [66] have observed that portfolios of common stocks tend to be lower tail dependent; i.e., a possible existing diversification effect breaks down if extreme losses occurs. For benchmark reasons we also investigate pure blue chip and REF portfolios.

The chapter is organized as follows: In Section 10.2 we present the mathematical concepts needed and summarize facts and definitions about EVT and copulas. Section 10.3 is devoted to the analysis of financial data and the question of parameter estimation. For single assets, we focus on log-returns of blue chips and European REFs and apply a GARCH-type model to capture conditional heteroskedasticity effects. For each time series, the loss tails of the innovations are fitted by a generalised Pareto distribution (GPD) to obtain a tail index and to get an accurate Value-at-Risk (VaR) and Expected Shortfall estimator as measures of risk. On the aggregated (portfolio) level, the joint distribution function is established along the lines of Junker and May [66] by a copula function based on a transformed Frank copula. Here we operate on the innovations of the univariate time-series, modeled by their empirical distribution function and yield an estimator for the tail dependence. The quality of the estimation is examined by performing a χ^2 goodness-of-fit (gof) test.

10.2 Mathematical Framework

For an overview about univariate extreme value theory we refer to Embrechts et al. [39]. Basic definitions can be found in Chapter 1, e.g. the definition of the generalized Pareto distribution (equation (1.2.3)) or the mean excess function (equation (1.2)).

The GPD is defined on the positive half axis. Often, we need to shift the distribution to some new starting point u that is called threshold. In general, the GPD might only reflect the tail behaviour of a given random variable. In this case we can determine the threshold u by graphical data analysis. We choose u such that the empirical mean excess function

$$\widehat{e}(x) = \frac{1}{N(x)} \sum_{i=1}^{N} x_i \mathbf{1}_{\{x_i \geq x\}}$$

of the observed sample $\{x_1, \ldots, x_N\}$, with $N(x) = |\{x_i \mid x_i \geq x,\ i = 1, \ldots, N\}|$, is approximately linear for $x \geq u$, see Embrechts et al. [39], pp.352. An estimator for a p%-quantile $\widehat{x}_p > u$ is attained by inverting the GPD

$$\widehat{x}_p = u + \frac{\widehat{\beta}}{\widehat{\xi}} \left(\left(\frac{N}{N(u)} (1-p) \right)^{-\widehat{\xi}} - 1 \right). \tag{10.2.1}$$

Since we are particularly interested in extreme values and their joint occurrence, we will use the asymptotic measure for tail dependence as defined by

equations (2.2.5) and (2.2.4). For the calculation we will fit the copula C_Ω as introduced in Chapter 4 to the data sets. For further details we refer to Junker and May [66] or Chapter 4.

10.3 Empirical studies

We now turn to the question of empirical evidence for a different crash behaviour for REFs with respect to common stocks. For this purpose we investigate the 15 biggest European REFs ranked by market capitalisation (see Table 10.1) and compare them with 15 blue chips (see Table 10.2), primarily taken from the European market. The data analysed here are daily log-returns in an observed

company	market capitalisation [in Mill. EUR]	number of zero-returns [in % of samplesize]
Land Securities, GB	6.9	2.57
Canary Wharf Finance, GB	4.5	14.73
British Land Company Plc, GB	4.1	2.63
Rodamco CE, NL	3.2	10.48
Unibail, F	2.7	8.26
Hammerson Plc, GB	2.3	11.59
Slough Estates, GB	2.2	11.49
Simco, F	2.1	10.01
Liberty, GB	2.0	23.79
Gecina, F	1.8	15.43
Corio, NL	1.7	17.40
IVG, D	1.4	9.04
Klépierre, F	1.4	12.91
Vallehermoso, E	1.1	4.73
Drott, S	1.0	13.70

Table 10.1: European REFs ranked by their market capitalisation.

time period ranging from Jan. 1997 to Jan. 2002 for the REFs, and from Oct. 1989 to Oct. 2000 for the blue chips respectively. As a lower bound measure for liquidity for the REFs, we give in Table 10.1 the number of observed zero log-returns in percentage of the sample size.

All datasets used show heteroskedasticity and some turn out to be autocorrelated. To deal with these effects, we describe the mean by an AR(1) model, and the volatility of the log-returns by a GARCH(1,1) model, i.e. we model the

log-returns denoted by R_t, via

$$\begin{aligned} R_t &= \mu_t + \sigma_t \varepsilon_t \\ \mu_t &= \mu + \alpha R_{t-1} \\ \sigma_t^2 &= \omega + \beta \varepsilon_{t-1}^2 \, \sigma_{t-1}^2 + \gamma \sigma_{t-1}^2, \end{aligned}$$

where $\alpha = 0$ for the log-returns that do not show autocorrelation. For a related model approach we refer to Frey and McNeil [52].

10.3.1 Single asset studies

In the following we want to investigate some attributes, like volatility and heavy tailedness, of the REF log-returns. Since we want to compare them with common stocks we only use data here in the joint observed time period ranging from Jan. 1997 to Oct. 2000.

We use a maximum likelihood estimator to get the parameters of a GPD describing the loss tail of the innovations. To examine the quality of the fit we perform a χ^2 goodness-of-fit test based on the data below the chosen thresholds and the GPD parameter estimates. In Table 10.2 the results for the blue chips are summarized. The tail index ξ measures the innovation risk, the unconditional volatility $\sigma^\diamond = \sqrt{\frac{\omega}{1-\beta-\gamma}}$ of the GARCH(1,1) process gives the volatility risk, and the 99% VaR and 99% Expected Shortfall (ES), both calculated with the fitted GPD, quantify the total single asset risk. Here the Expected Shortfall is obtained by a Monte Carlo Simulation with ten thousand runs. The unconditional mean return $\mu^\diamond = \frac{\mu}{1-\alpha}$ contains the payed risk premium. Here we will state the risk and performance measures with respect to the log-returns rather than the real returns. However, this chapter focuses more on qualitative and comparative analysis than on exact quantitative results. Furthermore, the transformation ($R = \exp(LR) - 1$) from the log-returns LR to the real returns R is strictly monotone (increasing). Hence we have no doubt about using the following results. There are 10 of the 15 blue chips heavy tailed with 80% significance and 3 with 95% significance. There is no significant light tailed blue chip. The p-values[1] of the χ^2 goodness-of-fit tests against the hypothesis of GPD distributed data over the chosen thresholds indicates satisfying loss tail approximations with exception of BP and Aventis; but even here the hypothesis can not be rejected on an 99% level. Table 10.3 contains the results for the REFs. Here it seemed to be less usual to observe heavy tailed losses, since only 5 (3) of 15 have a positive estimated tail index ξ with a significance of 80% (95%) and the real estate firm Drott is the only asset with a light tail in our studies. Also the mean of the estimated tail indices is

[1] The hypothesis of the test can not be rejected for levels higher than $1 - p$.

companies	$\widehat{\xi}$	$\widehat{\mu^\diamond}$ [in %]	$\widehat{\sigma^\diamond}$ [in %]	99% VaR [in %]	99% ES [in %]	$\frac{\widehat{\mu^\diamond}}{\widehat{\sigma^\diamond}}$ [in %]	p-value [in %]
BP, GB	0.0667	0.0609	1.9809*	4.56	5.40	3.08	1.77
Reuters, GB	0.1784*	0.0915	4.3202*	10.86	14.61	2.12	49.50
Lloyds, GB	0.0352	0.0841	2.5226°	5.92	7.28	3.33	83.01
Aventis, F	0.4128°	0.1275°	2.6231*	6.25	8.62	4.86°	2.21
Totalfina, F	0.2832°	0.0951°	2.4501*	5.95	7.54	3.88°	13.55
Allianz, D	0.2894*	0.0721	2.4899*	6.22	9.04	2.90	89.57
BASF, D	0.3002°	0.0567	2.0187*	4.88	6.47	2.81	96.68
Deutsche Telekom, D	0.1965°	0.1178°	3.5110*	8.54	11.11	3.35	21.92
Hoechst, D	0.0592	0.0381	5.1417*	13.74	17.62	0.74	42.99
VW, D	0.1001°	-0.0134	2.4111*	6.43	8.41	-0.56	82.88
Nestlé, CH	-0.0039	0.1027*	1.5222*	4.13	5.08	6.75°	18.48
Exxon, US	0.2037°	0.0634	1.7731°	4.08	5.13	3.57	95.98
IBM, US	0.2626*	0.1333°	2.4300*	6.29	8.95	5.49°	61.53
Microsoft, US	0.2019°	0.1097°	2.5458*	6.49	8.77	4.31°	91.73
SUN, US	-0.0233	0.4250*	3.3762*	8.53	9.86	12.59*	84.91

Table 10.2: Estimates for the blue chips with the p-values for the fitted GPD model. Values marked with a $\circ$, ($\star$), are significant on a 80%-level (95%-level).

with 0.0733 for the REFs less than half as big than the 0.1708 mean of the blue chips. Furthermore, the amount of zero-returns in Table 10.1 indicates a potential illiquidity risk for the REFs. In contrast none of the blue chips has more than 3% zero-returns during the observed time period. Usually, such illiquidity of moderate size results in more heavy tailed returns, so the conclusion that REFs tend to have less fat tailed innovation distributions than common stocks, is not affected.

The mean unconditional daily volatility of the REFs (1.59%) is 72% lower as for the blue chips (2.74%), even the estimated volatility of illiquid assets is usually higher than the true one. Since we do not cover any illiquidity risk in our Value-at-Risk or Expected Shortfall calculation, the lower volatility and less heavy tailed innovations of the REFs results directly to a lower 99%- average VaR (4.16%) and Expected Shortfall (5.41%) with respect to the blue chips (6.86% VaR and 8.93% Expected Shortfall). Comparing the estimated quotients $\frac{\mu^\diamond}{\sigma^\diamond}$ we have a mean value of 1.83% for the REFs and 3.95% for the blue chips and additionally only 3 of the 15 single REFs can beat the equally weighted blue chip portfolio in that sense. This indicates that there is a certain risk premium payed for the common stocks. Hence the market realizes the higher innovation and volatility risk for the blue chips and prices it higher than an eventual illiquidity risk for the REFs. This

companies	$\widehat{\xi}$	$\widehat{\mu^{\diamond}}$ [in %]	$\widehat{\sigma^{\diamond}}$ [in %]	99% VaR [in %]	99% ES [in %]	$\frac{\widehat{\mu^{\diamond}}}{\widehat{\sigma^{\diamond}}}$ [in %]	p-value [in %]
Land Securities, GB	0.2348*	-0.0044	1.4920*	3.75	5.11	-0.29	22.96
Canary Wharf Finance, GB	0.1982°	0.0526	2.4925*	6.85	9.84	2.11	73.14
British Land Company Plc, GB	-0.0745	0.0195	2.0411*	5.18	6.18	0.96	19.13
Rodamco CE, NL	0.0145	-0.0480	1.3571*	3.66	4.67	-3.53	12.30
Unibail, F	0.0954	0.0652°	1.5000*	3.98	5.15	4.35°	63.97
Hammerson Plc, GB	-0.0348	0.0385	1.2709*	3.35	4.12	3.03	48.04
Slough Estates, GB	-0.0133	0.0150	1.1701*	2.98	3.68	1.28	51.08
Simco, F	-0.0106	0.0038	1.4155*	3.65	4.53	0.27	39.40
Liberty, GB	-0.0427	0.0085	1.1271*	3.00	3.72	0.76	1.89
Gecina, F	0.0947°	0.0345	1.1625*	3.28	4.33	2.97	10.07
Corio, NL	0.3159*	0.0009	0.9466*	2.82	4.52	0.10	40.68
IVG, D	0.1138	0.0357	1.9163*	4.73	6.17	1.86	87.39
Klépierre, F	0.2373*	0.0913*	1.5196*	4.37	6.39	6.01*	33.12
Vallehermoso, E	0.1411	0.0566	2.0265*	4.56	5.58	2.80	94.25
Drott, S	-0.1702*	0.1131°	2.3612*	6.23	7.19	4.79°	56.10

Table 10.3: Estimates for the REFs and the p-values for the fitted GPD model. Values marked with a $\circ$, ($\star$), are significant on a 80%-level (95%-level).

observation is in line with Glascock and Davidson III [56], who found for the US market that, on average, real estate firm returns are lower than a benchmark return based on common stocks. They concluded that REFs underperform the market, even on a Sharpe and Treynor risk adjusted basis. However, a Sharpe and Treynor risk adjustment does not cover the observed higher innovation risk for common stocks. So an investment in REFs may still be fair.

10.3.2 Portfolio investigations

There were no obvious change points in the analyzed REFs and blue chips. So the GARCH residuals are iid samples and hence it is possible to compare the innovation distributions, even if they are generated from different time intervals. This enables us to use the maximal time interval for each investigated portfolio, e.g., Oct. 1989 to Oct. 2000 for the blue chips, Jan. 1997 to Oct. 2000 for the mixed portfolios and Jan. 1997 to Jan. 2002 for the pure REF portfolios.

The lower tail dependencies λ_L, as measures of crash diversification of the investigated portfolios, are obtained by fitting the copula given by Equation (2.5)

and applying Corollary 4.2.7. Furthermore, the estimated correlation ρ and the sample version of Kendall's τ, are stated as general diversification measures, where one should remember the pitfalls of using correlation mentioned above and in, e.g., Embrechts et al. [42]. Table 10.4 summarises the results for the blue chip portfolios as benchmark portfolios. We concentrate on country portfolios. All of

portfolio		$\widehat{\lambda}_L$	$\widehat{\rho}$	$\widehat{\tau}$	p-value [in %]
BP	Lloyds	$0.0706^{\star\star}$	0.2131	0.1661	73.50
	Reuters	$0.1093^{\star\star}$	0.2663	0.1808	92.23
Lloyds	Reuters	$0.1074^{\star\star}$	0.2228	0.1731	59.26
Aventis	Totalfina	$0.1522^{\star\star}$	0.2978	0.1966	98.58
Allianz	BASF	$0.2615^{\star\star}$	0.4983	0.3664	41.39
	D. Telek.	$0.1972^{\star\star}$	0.3659	0.2593	76.01
	Hoechst	$0.3071^{\star\star}$	0.4686	0.3546	90.40
	VW	$0.1766^{\star\star}$	0.4688	0.3527	29.85
BASF	D. Telek.	$0.1013^{\star}$	0.2876	0.2001	54.53
	Hoechst	$0.3310^{\star\star}$	0.6584	0.5241	27.88
	VW	$0.2809^{\star\star}$	0.5331	0.3980	11.46
D. Telek.	Hoechst	$0.2166^{\star\star}$	0.2923	0.2058	11.28
	VW	$0.2063^{\star\star}$	0.2579	0.1772	52.23
Hoechst	VW	$0.2417^{\star\star}$	0.4752	0.3592	76.25
IBM	Microsoft	$0.1733^{\star\star}$	0.3691	0.2545	76.85
	Sun	$0.0507^{\star\star}$	0.3573	0.2451	52.98
Microsoft	Sun	$0.1652^{\star\star}$	0.4196	0.2914	94.74
Allianz	BP	$0.0607^{\star\star}$	0.1262	0.0959	73.11
	Nestlé	$0.1091^{\star\star}$	0.2959	0.2283	28.19
	Microsoft	$0.0731^{\star\star}$	0.0925	0.0513	84.44

Table 10.4: Estimated lower tail dependency $\widehat{\lambda}_L$, correlation $\widehat{\rho}$ and samples Kendall's Tau $\widehat{\tau}$ for the blue chip portfolios, and the p-values of the fitted copula model with respect to a χ^2 goodness-of-fit test. All $\widehat{\lambda}_L$ values marked with a $\star\star$ are significant on a 99% level.

the 20 observed portfolios, even the few international ones, show lower tail dependency on a 99% significance level in addition to BASF-Deutsche Telekom where it is the 95% level. All levels of the TDC estimates were achieved by the delta method, i.e. by a first order Taylor approximation of the distribution of λ as a function of the copula parameter estimates (which are random variables). This indicates the tendency of common stock portfolios to have a minor diversification effect in extreme loss situations, as probably expected from their correlation or

Kendall's τ. The mean size of the lower tail dependency is $\overline{\widehat{\lambda}_L} = 0.17$. These findings are in line with the studies of Ané and Kharoubi [3] and Junker and May [66]. The p-values of the fitted copula model with respect to a performed χ^2 goodness-of-fit test indicate that the data are well adapted. Table 10.5 contains the results for the mixed portfolios, i.e. containing a common stock and a REF.

Here we also concentrate on country portfolios. There is only one portfolio (British Land-Reuters) estimated with a moderate high and significant lower tail dependence of $\widehat{\lambda}_L = 0.12$. All the other 28 investigated mixed portfolios have lower tail dependencies well below 0.1. So the mean size of the lower tail dependency is with $\overline{\widehat{\lambda}_L} = 0.03$ on a 99% level significantly lower than for the blue chip portfolios. The computed correlations and Kendall's τ's are for almost all mixed portfolios lower than for the blue chip portfolios, and hence the mean correlation and Kendall's τ are with $\overline{\widehat{\rho}} = 0.10$ and $\overline{\widehat{\tau}} = 0.07$ 71% and 72% lower than the corresponding values $\overline{\widehat{\rho}} = 0.35$ and $\overline{\widehat{\tau}} = 0.25$ for the common stock portfolios. So there is a high diversification effect of REFs compared to stocks. This is in contrast to Giliberto [55] and Maurer and Sebastian [86] who found high correlation of EREITs and REFs, respectively, with common stocks. Both did not use any heteroscedasticity filter what can cause an overestimated correlation in the absence of homoscedasticity. Furthermore, they studied monthly returns, whereas we here investigate daily log-returns. Together with the result of Liu and Mei [78] who found that EREIT returns show a high predictability compared with stocks and bonds, our differing findings may indicate a time shifted co-movement of the REFs with the common stocks or the existence of a systematic dependence that is suppressed by a dominating white noise dependence for high frequency data. Hence the observed diversification effects may be of no benefit for a buy and hold strategy. Therefore we compute the correlation and Kendall's τ for the mixed portfolios with lags up to 60 days (1 quarter), i.e., we calculate $corr(X_{[1,\ldots,N-k+1]}, Y_{[k,\ldots,N]})$ where X, Y are the innovation time series of the portfolio assets, N the series length and $k = 1, \ldots 60$ the lags. We could not find any remarkable and significant lagged correlation or Kendall's τ, see, e.g., the correlation plot of Land Securities versus British Petrolium in Figure 10.1. So we can neglect a time shifted co-movement. To investigate if there is a systematic dependence, we in Table 10.6 give the correlation and Kendall's τ for the mixed portfolios with a $1, \ldots, 4$ week log-return frequency. We can not extend this survey to the tail dependence, since the observed time horizon is not long enough to guarantee a satisfactory estimator $\widehat{\lambda}_L$. For all portfolios, except for Land Securities-Reuters, the maximum of the log-return correlations and Kendall's τs occur for middle frequency, mostly for the 3 and 4 week frequency. The scale of the maxima is of the same size as the results of Giliberto

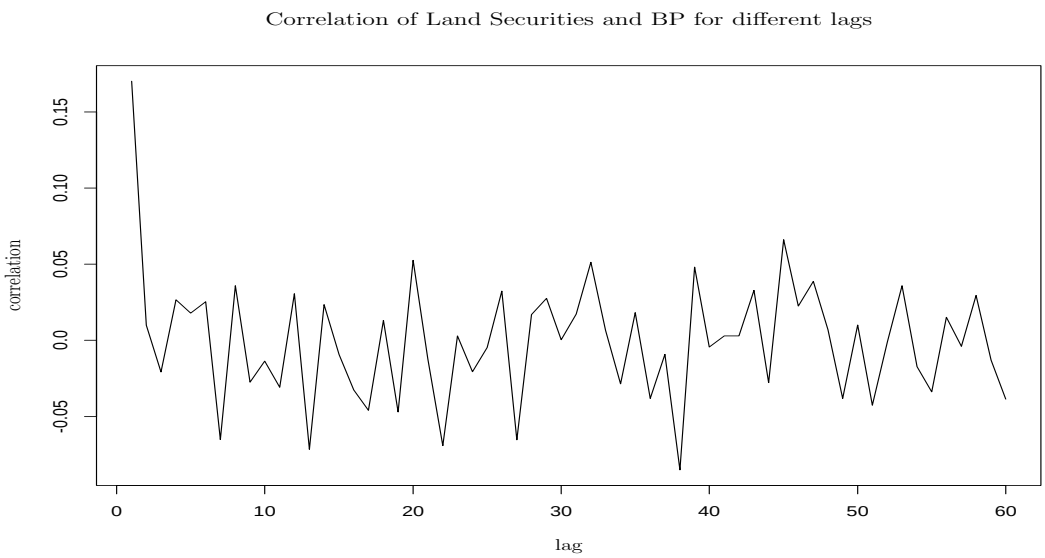

Figure 10.1: Correlation of Land Securities with British Petrolium for different lags.

[55] and Maurer and Sebastian [86]. We can not conclude a long term and systematic dependency, since for half of the investigated portfolios the correlation and Kendall's τ are already decreasing for the 4 week frequency. A long term study should be subject of further research to become certainty in this point. However, at this point we can come up with the result of a high diversification effect of REFs for daily frequency, which, in contrast to common stock portfolios, even in crash situations does not break down.

For completeness we have a look on pure REF portfolios in Table 10.7. The results are here somehow mixed. In the main country portfolios, the mean correlation and Kendall's τ are with $\overline{\widehat{\rho}} = 0.18$ and $\overline{\widehat{\tau}} = 0.10$ in-between the range of common stock and mixed portfolios. There are 3 of the 17 portfolios with a significant and moderate high lower tail dependence, whereas the others have an estimated $\widehat{\lambda}_L$ well below 0.1. With $\overline{\widehat{\lambda}_L} = 0.06$ the tail dependence parameters are very low compared to the blue chip portfolios. These findings are remarkable, since at least all REFs are in the same business line of real estates.

10.4 Conclusion

We use extreme value theory to examine the crash behaviour in means of heavy tailedness of single asset REFs. In comparison to blue chips they turn out to have less heavy tailed GARCH residuals and hence a lower innovation risk. With an observed lower unconditional volatility, this results in a lower Value-at-risk and

Expected Shortfall. A criticism is that we have stated the risk and performance measures with respect to the log-returns rather than the real returns. However, this chapter focused more on qualitative and comparative analysis than on exact quantitative results.

The concept of lower tail dependence allows us to survey dependence effects in crash situations. We conclude that in the sense of correlation and Kendall's τ, REFs can gain high diversification effects as admixture in common stock portfolios. In contrast to pure blue chip portfolios this diversification even does not break down in crash situations, indicated by a very small, in fact not present, lower tail dependence. These findings are for a daily frequency of the observed log-returns, and hence give a benefit for a buy and sell strategy and a one day risk management improvement. To come up with a statement for a buy and hold framework, further research should be carried out to examine the long term dependence structure of REFs with common stocks.

portfolio		$\widehat{\lambda}_L$	$\widehat{\rho}$	$\widehat{\tau}$	p-value [in %]
Land Securities	BP	0.0137**	0.1698	0.1202	97.90
	Lloyds	0.0000	0.1681	0.1092	97.39
	Reuters	0.0244**	0.1727	0.1363	43.46
British Land	BP	0.0537**	0.1821	0.1241	31.81
	Lloyds	0.0137	0.1309	0.0946	96.62
	Reuters	0.1221**	0.1083	0.0942	36.73
Hammerson	BP	0.0306	0.0789	0.0517	27.24
	Lloyds	0.0137	0.0762	0.0468	39.67
	Reuters	0.0710	0.1004	0.0740	92.31
Slough Estates	BP	0.0449	0.0822	0.0411	68.20
	Lloyds	0.0325°	0.0702	0.0319	19.29
	Reuters	0.0229°	0.0398	0.0227	48.11
Liberty	BP	0.0000	0.0923	0.0553	86.91
	Lloyds	0.0241**	-0.0005	-0.0073	82.93
	Reuters	0.0136**	0.0470	0.0483	96.08
Unibail	Aventis	0.0135	0.0304	0.0325	86.24
	Totalfina	0.0531**	0.0640	0.0614	89.90
Simco	Aventis	0.0136	0.0537	0.0584	66.30
	Totalfina	0.0137	0.0102	0.0347	94.84
Klépierre	Aventis	0.0008	0.0379	0.0222	99.82
	Totalfina	0.0515**	0.0181	0.0116	35.10
IVG	Allianz	0.0611*	0.1420	0.0911	91.11
	BASF	0.0000	0.1621	0.1099	45.30
	D. Telek.	0.0332*	0.1498	0.1000	97.32
	Hoechst	0.0000	0.1412	0.0949	79.20
	VW	0.0000	0.1552	0.1041	97.60
IVG	BP	0.0434°	0.1226	0.0819	37.82
	Nestlé	0.0649	0.1417	0.0957	98.94
	Microsoft	0.0240	0.0445	0.0293	95.33

Table 10.5: Estimated lower tail dependency $\widehat{\lambda}_L$, correlation $\widehat{\rho}$ and samples Kendall's Tau $\widehat{\tau}$ for the mixed portfolios, and the p-values of the fitted copula model with respect to a χ^2 goodness-of-fit test. $\widehat{\lambda}_L$ values marked with a $\circ$, ($\star$), ($\star\star$) are significant on a 80%, (95%), (99%) level.

portfolio		$\widehat{\rho_1}$	$\widehat{t_1}$	$\widehat{\rho_2}$	$\widehat{t_2}$	$\widehat{\rho_3}$	$\widehat{t_3}$	$\widehat{\rho_4}$	$\widehat{t_4}$
Land Securities	BP	0.1446	0.1119	0.2080	0.1301	**0.3388**	0.2115	0.3312	**0.2192**
	Lloyds	0.2483	0.1793	0.3115	0.2253	0.3954	**0.2737**	**0.3955**	0.2483
	Reuters	0.1086	0.1112	0.0333	0.0119	0.0682	0.0445	-0.1744	-0.1249
British Land	BP	0.1455	0.1275	0.3090	0.2359	0.4517	0.3159	**0.4800**	**0.3228**
	Lloyds	0.3030	0.2266	0.02511	0.1828	**0.5065**	**0.3744**	0.2771	0.2092
	Reuters	0.1187	0.0811	0.1344	0.0803	**0.1589**	**0.0957**	-0.0236	-0.0305
Hammerson	BP	0.1620	0.0834	0.3977	0.2437	0.4514	0.2965	**0.5725**	**0.3765**
	Lloyds	0.2143	0.1737	0.3125	0.1985	**0.4715**	**0.3072**	0.3318	0.2126
	Reuters	0.1494	**0.1242**	**0.1802**	0.1000	0.1566	0.0558	-0.0254	-0.0583
Slough Estates	BP	0.0483	0.0129	0.3038	0.1869	0.3006	**0.2227**	**0.3181**	0.1693
	Lloyds	0.1809	0.0991	0.1509	0.0864	**0.3304**	**0.2159**	0.2803	0.1497
	Reuters	0.0854	0.0536	0.1163	**0.0798**	**0.1609**	0.0374	-0.0096	0.0638
Liberty	BP	0.2443	0.1576	0.4041	0.2546	**0.4305**	**0.3333**	0.4138	0.2821
	Lloyds	0.1597	0.1017	0.3240	0.2459	0.3235	0.2438	**0.4045**	**0.2993**
	Reuters	-0.0207	0.0221	0.0665	0.0343	**0.1527**	**0.0865**	-0.1372	-0.1397
Unibail	Aventis	0.0829	0.0639	**0.1634**	**0.1149**	0.0690	0.0544	0.0821	0.0846
	Totalfina	0.1440	0.0611	0.1089	0.0831	**0.1990**	**0.1315**	0.0777	0.0021
Simco	Aventis	0.1164	0.0831	0.1508	0.0914	**0.2276**	**0.1584**	0.1726	0.0655
	Totalfina	**0.1956**	0.1062	0.1878	0.1170	0.1919	0.1303	0.1766	**0.1924**
Klépierre	Aventis	0.0564	0.0550	0.0419	0.04911	0.1112	0.0883	**0.1907**	**0.1163**
	Totalfina	-0.0539	-0.0822	-0.0953	-0.1259	**0.0472**	-0.0029	-0.0212	0.0085
IVG	Allianz	0.2943	0.1572	**0.3279**	**0.1791**	0.1488	0.0650	0.1898	0.0731
	BASF	0.2444	0.1548	0.2817	**0.1801**	0.2485	0.1572	**0.3397**	0.1230
	D. Telek.	0.1665	0.0348	**0.1741**	**0.1238**	0.1112	0.0353	0.1292	0.1082
	Hoechst	0.1515	0.1020	0.0978	0.0158	0.2010	0.1459	**0.2782**	**0.1915**
	VW	0.3216	0.1467	0.2968	0.2249	0.1217	0.1183	**0.3399**	**0.2488**
IVG	BP	0.2667	0.2058	0.1984	0.1459	0.2743	0.1867	**0.3800**	**0.2734**
	Nestlé	0.2695	0.1757	0.2377	0.1216	0.2230	0.2110	**0.3714**	**0.2537**
	Microsoft	0.1204	0.0599	0.0618	0.0417	0.1285	0.1164	**0.2686**	**0.1919**

Table 10.6: Correlation and Kendall's τ for $1, \ldots, 4$ week log-return frequencys, where the relativ maximal value is in bold style.

portfolio		$\widehat{\lambda}_L$	$\widehat{\rho}$	$\widehat{\tau}$	p-value [in %]
Land Securities	British Land	$0.1048^{\star\star}$	0.4125	0.2785	19.39
	Hammerson	0.0279	0.1686	0.1038	54.77
	Slough Estates	0.0000	0.2029	0.1151	69.37
	Liberty	0.0000	0.0863	0.0651	57.28
British Land	Hammerson	0.0331	0.2224	0.1322	98.02
	Slough Estates	0.0330	0.2196	0.1160	96.85
	Liberty	0.0325	0.1308	0.0827	62.24
Hammerson	Slough Estates	0.0611	0.3573	0.2236	44.74
	Liberty	$0.1416^{\star\star}$	0.1803	0.0972	85.44
Slough Estates	Liberty	$0.1339^{\star\star}$	0.1549	0.0774	77.44
Unibail	Simco	$0.0611^{\star}$	0.1386	0.0606	61.85
	Klépierre	0.0669	0.1263	0.0734	84.59
Simco	Klépierre	0.0820	0.1387	0.0608	79.18
IVG	Land Securities	0.0501°	0.0536	0.0355	51.14
	Corio	0.0001	0.0981	0.0455	82.25
	Unibail, F	0.0320	0.0757	0.0383	64.74
	Vallehermoso	0.0887	0.1759	0.0753	87.29

Table 10.7: Estimated lower tail dependency $\widehat{\lambda}_L$, correlation $\widehat{\rho}$ and samples Kendall's τ $\widehat{\tau}$ for the REF portfolios, and the p-values of the fitted copula model with respect to a χ^2 goodness-of-fit test. $\widehat{\lambda}_L$ values marked with a $\circ$, ($\star$), ($\star\star$) are significant on a 80%, (95%), (99%) level.

Chapter 11

Nonlinear Term Structure Dependence

This chapter, based on Junker et al. [69][1], documents nonlinear cross-sectional dependence in the term structure of U.S. Treasury yields and points out risk management implications. The analysis is based on a Kalman filter estimation of a two-factor affine model which specifies the yield curve dynamics. We then apply a broad class of copula functions for modelling dependence in factors spanning the yield curve. Our sample of monthly yields in the 1982 to 2001 period provides evidence of upper tail dependence in yield innovations; i.e., large positive interest rate shocks tend to occur under increased dependence. In contrast, the best fitting copula model coincides with zero lower tail dependence. This asymmetry has substantial risk management implications. We give an example in estimating bond portfolio loss quantiles and report the biases which result form an application of the normal dependence model.

11.1 Introduction

The class of affine term structure models (ATSMs) as proposed by Duffie and Kan [33] and further characterized by Dai and Singleton [24], has recently become a benchmark in modelling the term structure of default-free interest rates. Within the model class, the term structure is characterized by the current realizations as well as the dynamics of a set of state variables. Logarithmic bond prices are then affine functions of these state variables. The class appeals by its analytical tractability and contains the well-known models by Vasicek [109], Cox et al. [23], Chen and Scott [21], and Longstaff and Schwartz [80], for example. However, recent empirical evidence indicates that term structure data do not fully confirm the ATSM class. A series of articles document distinct nonlinearities e.g. in the

[1] A first draft of that paper was written and the dataset was obtained while the third author Niklas Wagner was visiting Stanford GSB. Additional data were kindly provided by David Tien.

drift and volatility function of the short-rate, particularly implying that mean-reversion in the short-rate depends on its level; see for example Aït-Sahalia [1] and Stanton [107]. Ang and Bekaert [4] focus on these findings and develop a Markovian switching-model which captures such nonlinearities. Also, empirical results on one- and two-factor ATSMs by Duan and Simonato [32] indicate a rejection of the affine model assumption when tested against local alternatives.[2] In general, findings of nonlinearity in the term structure of interest rates are important for at least three reasons. First, only an exact assessment of state variable dynamics and their dependence allows for an accurate modelling of the term structure. Second, derivatives pricing is frequently based on assumptions imposed by the class of ATSMs. And lastly, effective bond portfolio risk management builds upon models which give reliable risk implications.

While previous empirical studies have focused on time-series nonlinearities and discontinuities in the process dynamics, this paper analyzes nonlinear cross-sectional dependence between factors that span the yield curve. We show that the dependence structure of the long and short end of the yield curve exhibits nonlinearity which can be characterized under a particular focus on extremal dependence. The starting point of our model is the benchmark-class of ATSMs. Based on this theory, the term structure dynamics in our study are given by a Gaussian two-factor generalized Vasicek model. This model was applied for example by Babbs and Nowman [6] who find that the two-factor approach provides a good description of the yield curves for a broad sample of mature bond markets. Formulating a discrete time model in state-space representation allows for parameter estimation. We then focus on cross-sectional dependence in the term structure by modelling general forms of dependence in discrete factor innovations by a broad choice of copula functions. While a Gaussian factor model allows for correlated factors only, copula functions as outlined in Joe [62] and Nelsen [92], generalize the dependence concept by separating the treatment of dependence and marginal behavior. Based on the model, it is possible to characterize dependence in the center of the distribution independently from dependence in the distribution tails. Hence, we can impose various combinations of symmetric as well as asymmetric tail dependence on the factor innovations. Recent studies which apply copula functions in finance such as for example Ané and Kharoubi [3] and Scaillet [103] as discussed in Chapter 8, indicate that the concept is useful in

[2]Other specifications of the term structure include Ahn et al. [5] as well as alternative formulations of the short rate e.g. by Chan et al. [20] and Aït-Sahalia [1]. For an extense survey of models see also Dai and Singleton [25]. Besides the linear structure, the distributional assumptions imposed by ATSMs is critical: Björk et al. [10] extend the diffusion driven ATSMs by allowing for jumps. Eberlein and Raible [37] study term structure models driven by general Lévy processes.

modelling complex dependence structures. Conditional copula functions are studied by Patton [95] who models conditional dependence in U.S.-Dollar exchange rate returns and by Rockinger and Jondeau [101] who examine conditional dependence in international stock market returns. To our knowledge, no evidence for the dependence structure within the term structure of interest rates has yet been provided.

In our empirical investigation we focus on the term structure of U.S.-Treasuries which represent the largest government bond market worldwide. We use a sample of monthly yield curve observations as in the empirical studies for example by Ang and Bekaert [4] and De Jong [27]. The sample covers the 20-year period from October 1982 to December 2001. We form two 10-year subsamples in order to check for the robustness of the empirical results. The empirical investigation in the paper is then organized in two steps. In the first step, we use the class of affine term structure models to specify the yield curve dynamics. In particular, we choose a two-factor generalized Vasicek model characterized by a jointly normal bivariate factor process. We then extract factors representing yields, namely the interest rates on zero-coupon bonds with one year and five years to maturity. The model parameters are estimated by Kalman filtering as supported by maximum likelihood arguments; this was also done in previous studies such as for example Lund [81], Duan and Simonato [32], and Dewachter et al. [28]. In the second step, we model the dependence structure within the yield curve. We thereby focus on the dependence relation between short-term and long-term interest rates as represented by the two yield factors. To this aim, a broad set of different copula functions is used.

Based on our empirical findings, we show that the class of rotationally symmetric elliptical copula functions —including copulas such as the normal and the Student t— has characteristics which violate the observed complex dependence structure. Hence, the yield factor dependence cannot be characterized by a correlation coefficient as in the normal model nor with a rotationally symmetric Student t-model. While the copula function of the normal distribution does not allow for dependence in the tails, the Student t-copula does not allow for asymmetric tail dependence. However, dependence models contained in the class of Archimedean copulas can indeed capture dependence in the yield curve which is characterized by distinct asymmetry and upper tail dependence. The Gumbel as well as a transformed Frank copula turn out to be more suitable choices than the Student t-copula. Considering all candidate models used in our study, we find the transformed Frank copula to be the most appropriate model. Moreover, the goodness-of-fit tests for the two subsamples indicate that our main conclusions are robust within the observation period.

Given our empirical findings, we demonstrate the risk management implications in a bond portfolio setting. Based on the affine model of factor dynamics and the alternative copula models of factor dependence, we study the pricing effects of nonlinear dependence in the yield factors. Particularly, we use the ATSM implication that bond prices are exponential affine functions of the state variables. By sampling from one year and five year yield factors under the fitted copula functions we then estimate loss quantiles for bond portfolios with alternative durations. Our analysis highlights that the normal copula function —which is implied by the assumption of linear dependence in affine term structures— yields a substantial bias in the assessment of portfolio risk. When compared to the transformed Frank copula which captures the asymmetric dependence in the data, we report a bias structure in the upper and lower bond portfolio loss quantiles which yields values as high as 6 percent as compared to the normal model.

The remainder of this paper is organized as follows. In the next section, we outline the model used in the analysis. Term structure time series dynamics are given with the class of ATSMs. Cross-sectional dependence in bivariate term structure innovations is modeled by two candidate classes of copula functions. The empirical investigation and the estimation results are given in Section 11.3. The application to bond portfolio risk management which points out risk implications of nonlinear factor dependence is given in Section 11.4. Section 11.5 concludes.

11.2 The Term Structure Model

The starting point of our model is the class of benchmark ATSMs. We model the term structure time series dynamics within a continuous time two-factor generalized Vasicek model. A state-space representation allows for observational noise and prepares estimation based on a discrete time vector autoregressive version of the model. We then focus on cross-sectional dependence in the term structure by modelling dependence in factor innovations by copula functions. The functions stem from two wide classes of copula functions. Based on the copula model, we can impose various combinations of symmetric as well as asymmetric tail dependence on the factor innovations.

11.2.1 Term Structure Dynamics

Affine Term Structure Models

The affine term structure model is a class of models in which the yields to maturity are affine functions of some state variable vector. The state vector X is assumed

to obey the following dynamics

$$dX(t) = \kappa(\theta - X(t))\,dt + \Sigma\sqrt{S(t)}\,dW(t), \tag{11.2.1}$$

where W is a d-dimensional standard Brownian motion, θ is a d-vector, κ and Σ are $d \times d$ matrices, and $S(t)$ is a $d \times d$ diagonal matrix with diagonal elements which are affine functions of the state vector X. Provided that a parameterization is admissible, the price of a zero bond $P(t,\tau)$ in time t with time to maturity τ can be expressed as

$$P(t,\tau) = \exp\left(A(\tau) + B(\tau)^\top X(t)\right), \tag{11.2.2}$$

where A is a scalar function, and B is a d-dimensional vector function. The instantaneous interest rate is, as usual, defined as

$$r(t) = -\lim_{\tau \searrow 0} \frac{\ln P(t,\tau)}{\tau}. \tag{11.2.3}$$

Duffie and Kan [33] show that $P(\cdot,\cdot)$ is generically exponential affine, i.e. in the form of equation (11.2.2), if and only if the mean and variance in equation (11.2.1), and the short rate r are affine functions in the state variable X. Moreover, A and B in equation (11.2.2) are obtained as solutions to ordinary differential equations. Let $R(t,\tau)$ denote the time-t continuously compounded yield on a zero bond with maturity τ. The yield to maturity of this bond is

$$R(t,\tau) = -\frac{\ln P(t,\tau)}{\tau}. \tag{11.2.4}$$

The Two-Factor Affine Model

The special case of the two-factor generalized Vasicek model is given by

$$\begin{aligned} r(t) &= R_0 + X_1(t) + X_2(t), \\ dX(t) &= -\kappa\, X(t)\,dt + \Sigma\, dW(t), \end{aligned} \tag{11.2.5}$$

where W is a 2-dimensional standard Brownian motion, and

$$\kappa = \begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix}, \text{ and } \Sigma = \begin{pmatrix} \sigma_1 & 0 \\ \rho\,\sigma_2 & \sqrt{1-\rho^2}\,\sigma_2 \end{pmatrix}.$$

The parameter R_0 is the mean level of the instantaneous rate r, the state processes X_1 and X_2 fluctuate around zero with mean reversion rates κ_1, κ_2, and diffusion coefficients σ_1, σ_2, and correlation ρ. Details on the functions A and B describing the term structure implied by the two-factor model are given in Duffie and Kan [33] and Babbs and Nowman [6]; see also Appendix E.1.

State-Space Representation

Estimation of the above term structure model can be carried out via transformation to state-space representation; see for example Babbs and Nowman [6], [7] and Duan and Simonato [32] for term structure estimation applications and Harvey [57] for a general treatment of state-space models.

Assume that the yields for different maturities are observed with error. After the addition of measurement error, the yield to maturity, using the bond pricing formula (11.2.2), can be written as

$$R(t,\tau) = \frac{-A(\tau)}{\tau} + \frac{-B(\tau)^\top X(t)}{\tau} + \varepsilon(t,\tau), \tag{11.2.6}$$

where $\varepsilon(t,\tau)$ is assumed to be a normally distributed error term with mean zero and standard deviation $\sigma_{\varepsilon_\tau}$. Hence, given that N bond yields for different maturities are observed, the N corresponding yields have the following representation:

$$\begin{bmatrix} R(t,\tau_1) \\ \vdots \\ R(t,\tau_N) \end{bmatrix} = \begin{bmatrix} \frac{-A(\tau_1)}{\tau_1} \\ \vdots \\ \frac{-A(\tau_N)}{\tau_N} \end{bmatrix} + \begin{bmatrix} \frac{-B(\tau_1)^\top}{\tau_1} \\ \vdots \\ \frac{-B(\tau_N)^\top}{\tau_N} \end{bmatrix} X(t) + \begin{bmatrix} \varepsilon(t,\tau_1) \\ \vdots \\ \varepsilon(t,\tau_N) \end{bmatrix}. \tag{11.2.7}$$

In terms of the state-space model, this equation is referred to as the measurement equation. To obtain the transition equation for the state-space model, the expressions for the conditional mean and variance for the unobserved state variable process over a discrete time interval of length h have to be derived. Define $m(X(t);h) = \mathbf{E}\{X(t+h)\,|X(t)\}$ and $\Phi(X(t);h) = \mathrm{Var}(X(t+h)\,|X(t)\,)$, then the transition equation reads

$$X(t+h) = m(X(t);h) + \Phi(X(t);h)^{1/2}\,\eta(t,h), \tag{11.2.8}$$

where $\eta(t,h)$ is a d-vector of Gaussian white noise with $\Phi(X(t);h)^{1/2}$ denoting the Cholesky decomposition of $\Phi(X(t);h)$.

The two-factor model (11.2.5) defines the state variables as Gauss-Markov processes and thus the conditional mean and the conditional variance are:

$$m(x;h) = (m^h_{i,j})_{2\times 2} \cdot x = \begin{pmatrix} e^{-\kappa_1 h} & 0 \\ 0 & e^{-\kappa_2 h} \end{pmatrix} x, \tag{11.2.9}$$

$$\Phi(x;h) = (\Phi^h_{i,j})_{2\times 2} = \begin{pmatrix} \frac{\sigma_1^2}{2\kappa_1}(1-e^{-2\kappa_1 h}) & \frac{\rho\sigma_1\sigma_2}{\kappa_1+\kappa_2}(1-e^{-(\kappa_1+\kappa_2)h}) \\ \frac{\rho\sigma_1\sigma_2}{\kappa_1+\kappa_2}(1-e^{-(\kappa_1+\kappa_2)h}) & \frac{\sigma_2^2}{2\kappa_2}(1-e^{-2\kappa_2 h}) \end{pmatrix}. \tag{11.2.10}$$

Given observations of the yield vector in (11.2.6) and under a discrete sampling scheme with interval h, the exact likelihood function can be established based on the Kalman filter estimate of the unobservable state variable process X.

11.2.2 Nonlinear Term Structure Dependence

The Discrete-Time Factor Process

Section above outlined the two-factor affine term structure model which we apply in our study for capturing the term structure dynamics. The generalized Vasicek model (11.2.5) is based on continuous time factor dynamics dX driven by two-dimensional Brownian motion.

The factor process given by transition equation (11.2.8) is linear in the drift and non-stochastic in the diffusion coefficient. Hence, given (11.2.9) and (11.2.10), a discrete-time sample of X under $h = 1$, dropping h superscripts, is given by a vector autoregressive process of order one

$$\begin{aligned} X_{1,t} &= m_{1,1}\, X_{1,t-1} + (\Phi^{1/2})_{1,1}\eta_{1,t} + (\Phi^{1/2})_{1,2}\eta_{2,t}, \\ X_{2,t} &= m_{2,2}\, X_{2,t-1} + (\Phi^{1/2})_{2,1}\eta_{1,t} + (\Phi^{1/2})_{2,2}\eta_{2,t}, \end{aligned} \tag{11.2.11}$$

with $t = 0,\ 1,\ ...,\ T$. The variables $\eta_{i,t}$, $i = 1, 2$ are uncorrelated iid standard normal innovations. In this setting, factor dependence is completely characterized by the correlation coefficient ρ. Generalizing the above model, we now rewrite the discrete-time factor dynamics as

$$\begin{aligned} X_{1,t} &= m_{1,1}\, X_{1,t-1} + Z_{1,t}, \\ X_{2,t} &= m_{2,2}\, X_{2,t-1} + Z_{2,t}, \end{aligned} \tag{11.2.12}$$

and assume that the innovations $(Z_{1,t}, Z_{2,t})$ are iid vectors with common joint distribution function $H(z_1, z_2)$. This relaxes the assumption of joint normality as imposed by the class of ATSMs.

Copula Functions

Based on (11.2.12), copula functions allow us to treat general versions of factor dependence in the two-factor generalized Vasicek model (11.2.5). For the present application, we restrict the exposition to the two-dimensional case, and will model the joint distribution of Z_1 and Z_2.

Let F_{Z_1} and F_{Z_2} denote the continuous marginal distribution functions of Z_1 and Z_2, i.e. $H(z_1, \infty)$ and $H(\infty, z_2)$, respectively. By transformation we obtain uniform random variables as $U = F_{Z_1}(Z_1)$ and $V = F_{Z_2}(Z_2)$. The copula function $C : [0, 1]^2 \to [0, 1]$ for the bivariate random vector (Z_1, Z_2) is defined as the joint distribution function of the uniform random vector $(U, V) = (F_{Z_1}(Z_1), F_{Z_2}(Z_2))$, that is, $C(u, v) = P[U \leq u, V \leq v]$. Hence, it follows

$$H(z_1, z_2) = C(F_{Z_1}(z_1), F_{Z_2}(z_2)), \tag{11.2.13}$$

which is known as Sklar's Theorem, see also Chapter 2.

In the following we will use elliptical copulas as outlined by Chapter 3, namely the t-copula (see equation (2.3.1)) and the normal copula (see equation (2.3.3)), and Archimedean copulas, namely the Gumbel copula (see equation (2.3.6)) and the transformed Frank copula C_ω as introduced in Chapter 4 (see equation (4.2.3)).

11.3 Empirical Analysis of Nonlinear Term Structure Dependence

Our program for the empirical analysis is as follows. First we briefly introduce the zero-coupon yield dataset. We then estimate the term structure parameters of the two-factor model based on Kalman filter approach as outlined in Section 11.2.2. The empirical analysis of dependence between unpredictable innovations in the long end and the short end of the yield curve is based on an examination of our different parametric copula functions (normal-copula, t-copula, Gumbel copula and transformed Frank copula). We argue that the theoretical properties of the copula functions given above, together with careful empirical testing, allow us to identify a suitable model which is consistent with the dependence in the yield structure.

11.3.1 The Sample

As pointed out in the introduction, our empirical analysis of the U.S.-Treasury term structure is based on a sample of monthly zero-coupon yields. The yield observations are obtained from the refined Fama-Bliss zero-coupon yield dataset as introduced in Fama and Bliss [43]. The maturities range from one to five years. The sample covers the period October 1982 to December 2001 with 231 monthly observations. Of course, the amount of sample information comes with a trade-off concerning stationarity. We did therefore not extend the sample back to periods in the early eighties where much different economic and economic policy regimes prevailed. Still, with the given dataset covering nearly twenty years, a check of robustness of the empirical results with respect to sample choice is important. We hence form two subsamples covering the October 1982 to December 1991 and the January 1992 to December 2001 period, which yields 111 monthly observations and 120 monthly observations, respectively.

In the following, we consider the monthly zero-coupon yields $R(t, \tau_i)$, $t = 0, \ldots, 230$ with τ_i denoting the i-year to maturity bond with $i = 1, 2, 3, 4, 5$. All yields are given on an annualized continuously compounded basis. We express

all maturities in months and the parameter estimates can interpreted as monthly values; (annualized values can be calculated from these given for reasons of comparison). The length of the discrete sampling interval, h, equals 1 month. Table 1 reports summary statistics for the entire sample period as well as for the two subperiods.

October 1982 to December 2001			
Maturity	Mean	Std. Dev.	Autocorr.
1 year	0.0638	0.0201	0.9852
2 years	0.0675	0.0206	0.9859
3 years	0.0701	0.0204	0.9854
4 years	0.0721	0.0205	0.9851
5 years	0.0732	0.0205	0.9858
October 1982 to December 1991			
Maturity	Mean	Std. Dev.	Autocorr.
1 year	0.0794	0.0157	0.9641
2 years	0.0837	0.0165	0.9707
3 years	0.0863	0.0165	0.9709
4 years	0.0886	0.0166	0.9696
5 years	0.0897	0.0167	0.9708
January 1992 to December 2001			
Maturity	Mean	Std. Dev.	Autocorr.
1 year	0.0494	0.0109	0.9712
2 years	0.0526	0.0098	0.9581
3 years	0.0550	0.0088	0.9447
4 years	0.0569	0.0082	0.9371
5 years	0.0579	0.0080	0.9387

Table 11.1: Summary statistics (sample mean, sample standard deviation, and first order sample autocorrelation) for the monthly U.S. Treasury zero-coupon yield data. Sample period 1982 to 2001 and subsamples 1982 to 1991 and 1992 to 2001.

The statistics in Table 11.1 exhibit an on average increasing yield curve. The sample autocorrelation coefficients indicate typical first order linear dependence in monthly bond yields. Comparing the results for the subsamples with those for the entire period indicates lower levels of interest rates together with lower levels of volatility in yield changes for the second subsample.

11.3.2 Term Structure Model Estimation

This section presents the term structure estimation results based on Kalman filtering as outlined in Section 11.2.2. In general, the application of the Kalman filter requires the state process X to have normally distributed innovations which is typically violated in estimation applications. This means that the parameter estimation approach is based on a quasi-likelihood function.[3]

Based on the Kalman filter approach, maximum likelihood (ML) estimation yields an estimate of the parameter vector $\psi = (R_0, \kappa_1, \kappa_2, \gamma_1, \gamma_2, \sigma_1, \sigma_2, \rho)$ of the two-factor generalized Vasicek model (11.2.5). As additional parameters to those previously introduced, we estimate γ_1 and γ_2 which denote the risk premiums for the first and the second factor, respectively (see Appendix E.1). We assume a diagonal covariance structure of the measurement errors $\varepsilon(t, \tau)$ in (11.2.7) where the diagonal elements are denoted by $\sigma^2_{\varepsilon_i}$. The estimation results are given in Table 11.2. All parameter estimates contained in ψ turn out to be significantly different from zero at usual confidence levels. The estimates of γ_1 are negative throughout and significantly different from zero for the second subsample, indicating a negative premium in yield curve shifts. The significant estimates of γ_2 indicate a positive risk premium on shifts in yield curve steepness. The estimates of the correlation coefficient ρ reveal a typical negative linear dependence between the two yield factors. The estimated standard errors $\widehat{\sigma}_{\varepsilon_i}$ are relatively homogeneous for all maturities with a slight tendency for larger measurement error variability for the 1-year maturity yields.

11.3.3 Derivation of Term Structure Innovations

Given the estimate of ψ in Table 11.2, we can derive unpredictable innovations for our term structure sample. By choosing two observable yields, namely the short end τ_s-year and the long end τ_l-year maturity yield, $R(t, \tau_s)$ and $R(t, \tau_l)$, the dynamics of the two-dimensional yield factor X can be expressed in terms of the estimated term structure parameters; see Appendix E.2 for details on the

[3]Inference based on the Kalman iteration and likelihood maximization faces two specification issues. Firstly, the Kalman Filter estimates of X_t do not exactly correspond to the conditional expectations given the observed yields since the filter relies on a linear projection. Secondly, in a non-Gaussian model, the filtering errors –the differences between X_t and the linear projections– are not normally distributed. Brandt and He [15] discuss the first-order approximation for non-normalities introduced by affine factor dynamics. Duan and Simonato [32] discuss the estimation of square-root models by Kalman filtering and show in a simulation study that the biases are small, and, as also the results in Lund [81] indicate, economically insignificant.

Kalman Filter Estimates			
Parameters	Oct. 82 to Dec. 01	Oct. 82 to Dec. 91	Jan. 92 to Dec. 01
R_0	0.0589	0.0637	0.0286
	(0.0291)	(0.0254)	(0.0176)
κ_1	0.0691	0.1225	0.1918
	(0.0227)	(0.0285)	(0.0143)
σ_1	0.0203	0.0260	0.2402
	(0.0034)	(0.0048)	(0.0151)
γ_1	-0.1850	-0.0465	-0.4359
	(0.1400)	(0.1696)	(0.1465)
κ_2	0.3719	0.4954	0.2131
	(0.0413)	(0.0610)	(0.0162)
σ_2	0.0188	0.0230	0.2385
	(0.0033)	(0.0047)	(0.0158)
γ_2	1.3358	1.4057	1.5395
	(0.1838)	(0.2509)	(0.2310)
ρ	-0.7807	-0.8199	-0.9991
	(0.0797)	(0.0774)	(0.0001)
σ_{ε_1}	0.0014	0.0008	0.0014
	(0.0001)	(0.0002)	(0.0001)
σ_{ε_2}	0.0004	0.0006	0.0002
	(0.0001)	(0.0001)	(0.0001)
σ_{ε_3}	0.0006	0.0007	0.0003
	(0.0001)	(0.0001)	(0.0001)
σ_{ε_4}	0.0006	0.0008	0.0002
	(0.0001)	(0.0001)	(0.0001)
σ_{ε_5}	0.0005	0.0006	0.0005
	(0.0001)	(0.0001)	(0.0001)

Table 11.2: Estimation results for the two-factor generalized Vasicek model with monthly observations on 1, 2, 3, 4 and 5 year maturity yields. Sample period 1982 to 2001 and subsamples 1982 to 1991 and 1992 to 2001. Kalman filter recursions are initialized with the values of the stationary mean and variance of the unobserved state variables. Maximization of the log-likelihood function is based on a sequential quadratic programming algorithm. Heteroskedasticity-consistent standard errors of the parameter estimates given in parenthesis.

yield factor representation. From the affine structure, it follows

$$\begin{pmatrix} R(t,\tau_s) \\ R(t,\tau_l) \end{pmatrix} = \mathcal{A} + \mathcal{B}X_t = \begin{pmatrix} -A(\tau_s)/\tau_s \\ -A(\tau_l)/\tau_l \end{pmatrix} + \begin{pmatrix} -B(\tau_s)^\top/\tau_s \\ -B(\tau_l)^\top/\tau_l \end{pmatrix} X_t.$$

Due to the autoregressive structure of X in (11.2.12) this results in

$$\begin{pmatrix} R(t,\tau_s) \\ R(t,\tau_l) \end{pmatrix} = (\mathbf{I}_2 - \mathcal{B}\,A\,\mathcal{B}^{-1})\,\mathcal{A} + \mathcal{B}\,A\,\mathcal{B}^{-1} \begin{pmatrix} R(t-1,\tau_s) \\ R(t-1,\tau_l) \end{pmatrix} + \mathcal{B} \begin{pmatrix} Z_{\tau_s,t} \\ Z_{\tau_l,t} \end{pmatrix}.$$

In short, this can be rewritten as

$$\begin{pmatrix} R(t,\tau_s) \\ R(t,\tau_l) \end{pmatrix} = \mu_R + A_R \begin{pmatrix} R(t-1,\tau_s) \\ R(t-1,\tau_l) \end{pmatrix} + \begin{pmatrix} \epsilon_{\tau_s,t} \\ \epsilon_{\tau_l,t} \end{pmatrix}. \tag{11.3.1}$$

Note that the common joint distribution function $G(\epsilon_{\tau_s}, \epsilon_{\tau_l})$ of the innovations is completely determined by $H(z_{\tau_s}, z_{\tau_l})$ and $\mathcal{B}$, i.e. the distribution of $(\epsilon_{\tau_s,t}, \epsilon_{\tau_l,t})$ follows from that of $(Z_{\tau_s,t}, Z_{\tau_l,t})$ via scaling with a constant.

In the following we choose the one-year and the five-year maturity yield, $R(t,1)$ and $R(t,5)$, to represent the short and the long yield factor, respectively. Based on the Kalman filter estimate $\widehat{\psi}$, we can derive the estimates $\widehat{\mu}_R$ and $\widehat{A}_R$. The time $t-1$ conditional expectation $\mathbf{E}_{t-1,\widehat{\psi}}$ is defined by equation (11.3.1). We then obtain the sequence of bivariate empirical yield innovations as

$$\begin{pmatrix} \widehat{\epsilon}_{1,t} \\ \widehat{\epsilon}_{5,t} \end{pmatrix} = \begin{pmatrix} R(t,1) - \mathbf{E}_{t-1,\widehat{\psi}}\{R(t,1)\} \\ R(t,5) - \mathbf{E}_{t-1,\widehat{\psi}}\{R(t,5)\} \end{pmatrix}, \quad t = 1, ..., T, \tag{11.3.2}$$

with $T = 230$.

11.3.4 Analysis of the Term Structure Innovations

Estimation of the copula parameter vector based on sample innovations is widely used in empirical research. Given the parameter estimates for the term structure dynamics above, the empirical marginal distribution functions are determined for each component of the bivariate yield innovation series (11.3.2). We thereby check for the joint normality assumption which is imposed by the ATSM. In the second step, parametric estimation of the copula functions is carried out. The copula functions introduced in Section 11.2.2 are our respective candidate dependence models for the bivariate yield innovation series.

Distributional Properties of the Term Structure Innovations

The Kalman filtering estimates of Section 11.3.2 result in the bivariate yield innovation series $(\widehat{\epsilon}_{1,t}, \widehat{\epsilon}_{5,t})_{t=1,...,T}$ as defined by equation (11.3.2).

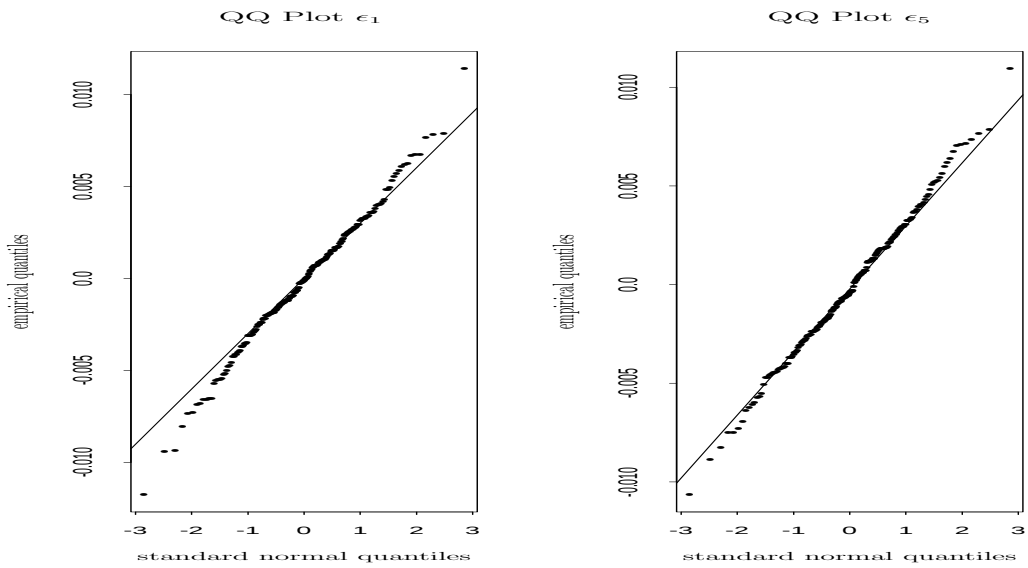

Figure 11.1: QQ-plots of the marginal distributions of $\widehat{\epsilon}_1$ and $\widehat{\epsilon}_5$ each against the standard normal distribution. Sample period 1982 to 2001.

a) Univariate Properties

Closer inspection of the series' univariate distributional properties reveals that the assumption of uncorrelated normally distributed innovations appears to be a suitable approximation.

The empirical marginal (univariate) distribution functions $F_{\cdot,T}$ are determined for each component separately. They are given as

$$F_{\cdot,T}(x) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{\{\cdot_t \leq x\}}. \tag{11.3.3}$$

Figure 11.1 shows QQ-plots of the marginal distributions of the innovations where the quantiles of the empirical distributions are plotted against those of the standard normal distribution. The two plots indicate a reasonably good approximation by the normal distribution, where the fit in the lower tail is better for the long maturity factor innovations. Additional results from the univariate χ^2-test over 17 cells indicate that fitted normality models (with mean zero) cannot be rejected with p-values of 0.40 for $\widehat{\epsilon}_1$ and 0.15 for $\widehat{\epsilon}_5$. We next consider the innovations' time-series properties. Figure 11.2 shows the sample autocorrelation functions for the univariate and the squared univariate series with lags up to order 23. The estimated autocorrelations for the raw innovations stay within the 95% confidence intervals with one exception in the $\widehat{\epsilon}_1$-plot which is an expected violation under the given confidence level. For the squared innovations we find six coefficients to exceed the 95% confidence interval in both plots. These irregular exceedances are evidence of some heteroskedasticity in both series, though they

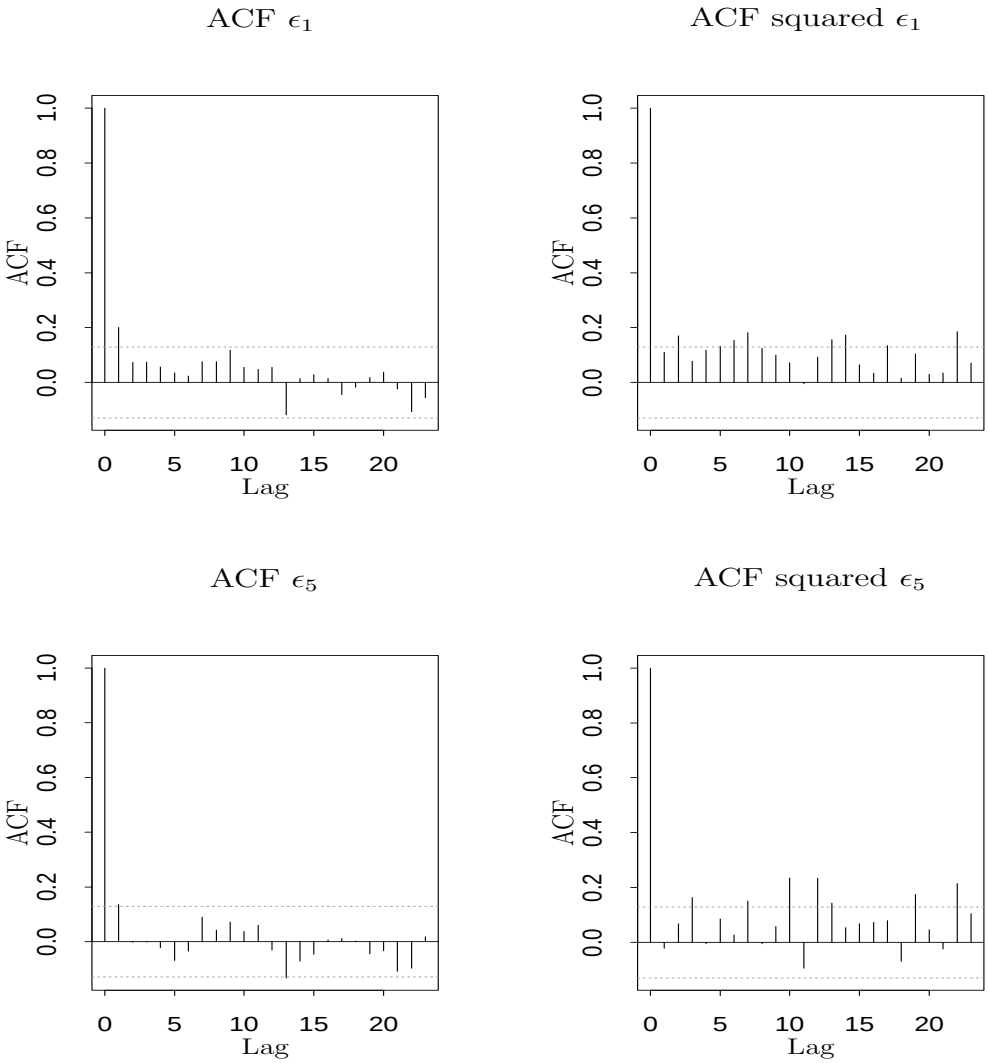

Figure 11.2: Plots of the autocorrelations of the innovations $\widehat{\epsilon}_1$ and the squared innovations $\widehat{\epsilon}_5^2$. The dotted straight line indicates the 95% confidence interval under the Null of uncorrelatedness. Sample period 1982 to 2001.

are weak as compared to what is otherwise typically observed for financial return series.

b) Multivariate Properties

While the above results support the marginal model assumptions of the affine model, a brief graphical analysis of the joint distribution of the factor innovations in Figure 11.3 casts doubt on the assumption of joint normality of the factor innovations.

Figure 11.3 gives a scatterplot representation of the joint density of the yield innovations for the overall sample and the two subsamples. Apart from standard scatterplot representations of $(\widehat{\epsilon}_{1,t}, \widehat{\epsilon}_{5,t})$, the figure contains plots of the mapped observations

$$(\widehat{\epsilon}_{1,t}, \widehat{\epsilon}_{5,t})_{t=1,\ldots,T} \longmapsto (\widehat{u}_t, \widehat{v}_t)_{t=1,\ldots,T} = \left(F_{\epsilon_1,T}\left(\widehat{\epsilon}_{1,t}\right), F_{\epsilon_5,T}\left(\widehat{\epsilon}_{5,t}\right)\right)_{t=1,\ldots,T}. \tag{11.3.4}$$

As an empirical application of Sklar's theorem (11.2.13), the mapped observations are defined on the uniform space $[0,1] \times [0,1] = [0,1]^2$. The $(\widehat{\epsilon}_{1,t}, \widehat{\epsilon}_{5,t})$-plots in the figure reveal a somewhat stronger uniform (e.g. linear) factor dependence during the second subsample period. All three $(\widehat{u}_t, \widehat{v}_t)$-plots reveal some clustering of observations in the upper right-hand corner of $[0,1]^2$ which indicates upper tail dependence. In contrast, graphical inspection gives little evidence of lower tail dependence as would be indicated by clustering in the lower left-hand corners of the plots. In the next three paragraphs, we estimate our parametric copula functions and analyze which models may capture the observed factor dependence.

Copula Estimation Methodology

Given the above results supporting the univariate model assumptions of the affine model, we assume in the following that the marginal distributions of the yield innovations are normal with mean zero and variances $\sigma^2_{1,\cdot}$ and $\sigma^2_{5,\cdot}$, respectively. Hence, we can write Sklar's Theorem (11.2.13) as

$$G_{\cdot}\left(\epsilon_1, \epsilon_5\right) = C_{\cdot}\left(N\left(\epsilon_1/\sigma_{1,\cdot}\right),\ N\left(\epsilon_5/\sigma_{5,\cdot}\right)\right), \tag{11.3.5}$$

where the notation ' $\cdot$ ' indicates the choice of one of the respective copula functions C_t, C_N, C_ω and C_G. For the different copula functions, the parameter vectors are given as $\omega_t = (\rho, \nu)$, $\omega_N = (\rho)$, $\omega_\omega = (\vartheta, \delta)$, and $\omega_G = (\delta)$. We use ML-estimation to obtain simultaneous estimates of the parameters $(\omega_{\cdot}, \sigma_{1,\cdot}, \sigma_{5,\cdot})$ of the joint distribution function $G_{\cdot}$. Note that these estimates are optimal for the overall joint distributional assumption imposed by $G_{\cdot}$, which includes the

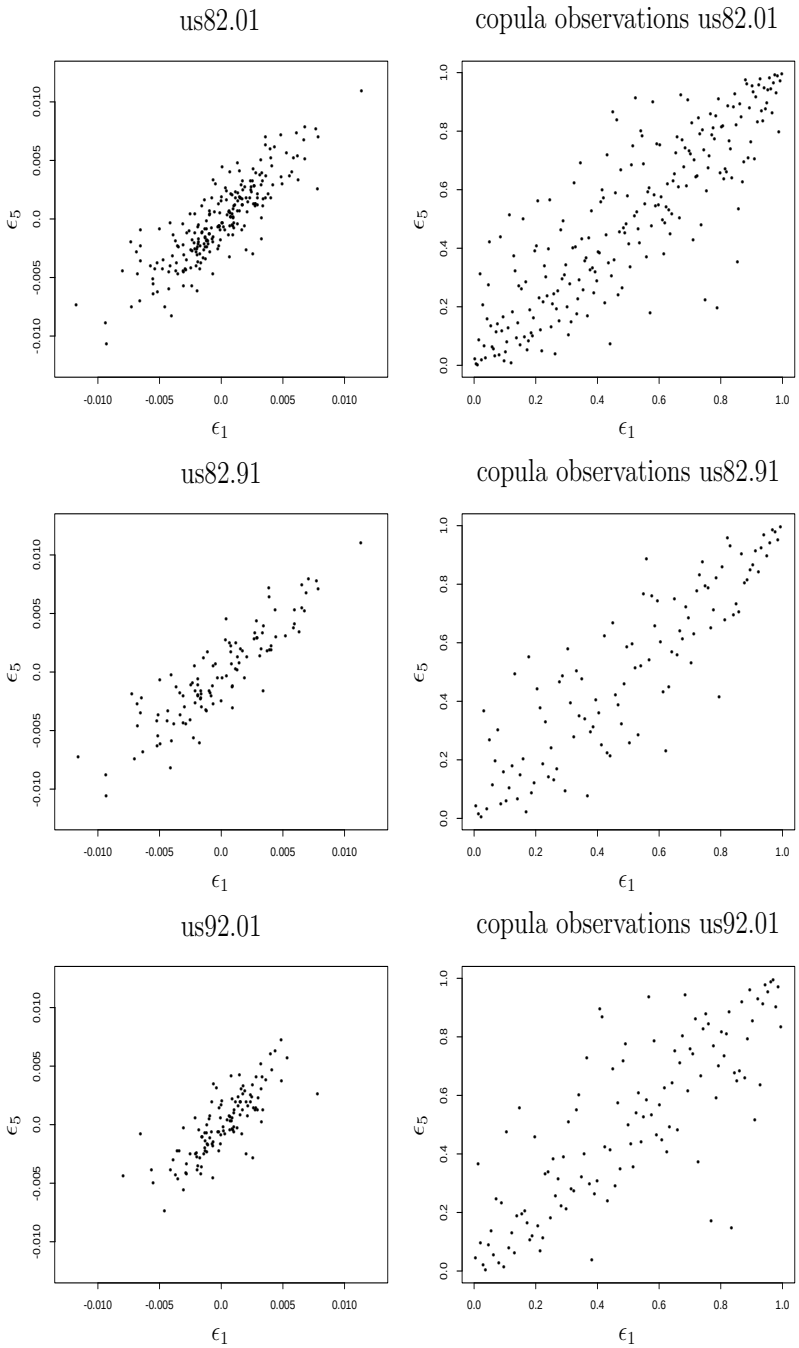

Figure 11.3: Scatterplot representation of the yield innovations and the yield innovation copula densities. Sample period 1982 to 2001 and subsamples 1982 to 1991 and 1992 to 2001.

marginal distributions as well as the dependence structure. With the joint density function $g.(\epsilon_1, \epsilon_5)$ derived from (11.3.5) the log likelihood function reads

$$\ln L.\left(\omega., \sigma_{1,\cdot}, \sigma_{5,\cdot};\ \epsilon_{1,t}, \epsilon_{5,t}\right) =$$
$$= \sum_{t=1}^{T} \ln \left[\frac{1}{\sigma_{1,\cdot}\sigma_{5,\cdot}}\ c.\left(N\left(\epsilon_{1,t}/\sigma_{1,\cdot}\right),\ N\left(\epsilon_{5,t}/\sigma_{5,\cdot}\right)\right)\ N^{'}\left(\epsilon_{1,t}/\sigma_{1,\cdot}\right)\ N^{'}\left(\epsilon_{5,t}/\sigma_{5,\cdot}\right) \right], \tag{11.3.6}$$

where $N^{'}$ denotes the density of the standard normal distribution and $c.$ is one of our respective copula densities. As the estimates of the copula parameters $\widehat{\omega}_T$ will have ML-properties (ignoring the time interdependence filtering impacts), the estimates of the tail dependence parameters, $\widehat{\lambda}_T = \lambda(\widehat{\omega}_T)$ will be consistent and normally distributed with

$$\sqrt{T}(\widehat{\lambda}_T - \lambda) \stackrel{d}{=} N\left(0, \sigma_\lambda^2\right). \tag{11.3.7}$$

Given that $\lambda\left(\cdot\right)$ are suitably smooth functions, the variance σ_λ^2 above can be approximated by a first order Taylor series expansion of the form

$$\sigma_{\widehat{\lambda}}^2 = \sum_{i=1}^{p} \left(\frac{\partial \lambda(\widehat{\omega}_T)}{\partial \omega_i} \right)^2 \sigma_{\widehat{\omega}_i}^2,$$

where $p = \#\{\omega\}$ denotes the number of parameters of the copula model and $\sigma_{\widehat{\omega}_i}$ denotes the ML standard errors of the respective copula parameter estimates.

Diagnostics for the Estimated Copula Models

Based on our estimates of the parametric copulas, we compare the in-sample model fit based on a set of different goodness-of-fit test procedures. These include seven statistics which are given as follows.

A general goodness-of-fit test as introduced in Chapter 6 given by equation (6.1.2) and determined by the mapping rule (6.1.1). Based on a grid of $6 \times 6 = 36$ cells of identical size for the overall sample, this implies 32 degrees of freedom for the normal and the Gumbel copula and 31 degrees of freedom for the t-copula and the transformed Frank copula. For the two subsamples, we use a grid of $4 \times 4 = 16$ cells of identical size and adjust the degrees of freedom accordingly.

Three additional tests of the overall model fit are based on the maximized log-likelihood function $\ln L.$. These include the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the Entropy EN, all introduced in Chapter 6. To calculate the latter we had run a Monte Carlo simulation.

Furthermore we use the integrated version of an Anderson and Darling test (see Anderson and Darling [2] and Chapter 6).

Our last goodness-of-fit diagnostics particularly focuses on the fit in the distribution tails. The AD statistics, due to its use of the cumulative distribution function, has the drawback of a smoothing effect particularly present in the upper tail. When considering model fit in the tails, we therefore apply a diagnostics which is based not on the overall probability deviations, but on the probability deviations at a particular quantile of the joint distribution function. Let C_T denote the empirical copula function and $\overline{C}_T$ the empirical survival copula function, as defined in Chapter 6 by equations (6.1.1) and (6.2.6). With these empirical copulas we measure deviations at the upper and the lower tail independently. The relative lower tail probability deviation PD_p is defined as the deviation of the model probabilities from the empirical probabilities measured at a point (q, q) in the lower corner of the set $[0, 1]^2$. Here, $q = C_T^{\leftarrow}(p)$, $0 \leq p \leq 1$ and $C_T^{\leftarrow}$ is the inverse of the diagonal section of the empirical copula function C_T. The relative lower tail probability deviation PD_p is given as

$$PD_p = \frac{G_{\cdot}\left(F_{\epsilon_1,T}^{\leftarrow}\left(C_T^{\leftarrow}(p)\right),\ F_{\epsilon_5,T}^{\leftarrow}\left(C_T^{\leftarrow}(p)\right)\right) - p}{p}. \tag{11.3.8}$$

Based on survival functions, the upper tail relative probability deviation is defined by the survival probability deviation PDS_p at the point (q, q) in the upper right corner of $[0, 1]^2$. It is given as

$$PDS_p = \frac{\overline{G}_{\cdot}\left(\overline{F}_{\epsilon_1,T}^{\leftarrow}\left(\overline{C}_T^{\leftarrow}(p)\right),\ \overline{F}_{\epsilon_5,T}^{\leftarrow}\left(\overline{C}_T^{\leftarrow}(p)\right)\right) - p}{p}. \tag{11.3.9}$$

Setting p equal to a small positive value, the probability deviations PD_p and PDS_p allow us to measure deviations in the tails. Note that pT observations are available for the calculation of the empirical distribution function and —given that pT is sufficiently large— ensure convergence towards the true distribution function.

Note that here we have used a different definition of PD, and PDS respectively, as in Chapter 6. Here p is the empirical probability for a point to be in the lower and upper corner, respectively. In equations (6.2.7) it is the probability under the hypothesis, i.e. under the assumed distribution model. However, here we are forced to follow such a data adaptive approach since the lack of enough data. Using the definitions (11.3.8) and (11.3.9) instead, enables us to force to have an absolute minimum of data for the calculation of the empirical distribution functions.

October 1982 to December 2001						
$T = 230$	$\widehat{\sigma}_{1,\cdot}$	$\widehat{\sigma}_{2,\cdot}$	$\widehat{\rho}$	$\widehat{\nu}$	$\widehat{\lambda}_L$	$\widehat{\lambda}_U$
C_t	3.4291E-3	3.4730E-3	0.8556	10.2957	0.3681	0.3681
s.e.	(0.1806E-3)	(0.1701E-3)	(0.0196)	(7.6051)	(0.1716)	(0.1716)
t-value	(18.99)	(20.42)	(43.76)	(1.35)	(2.14)	(2.14)
C_N	3.4517E-3	3.4510E-3	0.8537	-	0	0
s.e.	(0.1845E-3)	(0.1679E-3)	(0.0190)	-	-	-
t-value	(18.71)	(20.55)	(44.97)	-	-	-
	$\widehat{\sigma}_{1,\cdot}$	$\widehat{\sigma}_{2,\cdot}$	$\widehat{\vartheta}$	$\widehat{\delta}$	$\widehat{\lambda}_L$	$\widehat{\lambda}_U$
C_ω	3.4152E-3	3.4738E-3	4.1759	1.8101	0	0.5334
s.e.	(0.17785E-3)	(0.1665E-3)	(1.1523)	(0.2463)	-	(0.0764)
t-value	(19.20)	(20.86)	(3.62)	(3.29)	-	(6.98)
C_G	3.5237E-3	3.5433E-3	-	2.8805	0	0.7279
s.e.	(0.18045E-3)	(0.16912E-3)	-	(0.2047)	-	(0.0218)
t-value	(19.53)	(20.95)	-	(9.19)	-	(33.46)

Table 11.3: Parametric ML-estimates of the joint distribution function G under the alternative copula models. Standard errors and t-values of the parameter estimates given in parenthesis. Sample period 1982 to 2001.

Copula Estimation Results

Tables 11.3, 11.4 and 11.5 give the estimation results for the overall sample 1982-2001 and the two subsamples 1982-1991 and 1992-2001, respectively. We give the standard deviation estimates for the marginal distributions as well as the estimates of the parameters in the copula parameter vectors ω_t, ω_N, ω_ω, and ω_G. In Table 11.6 we compare the goodness-of-fit for the competing copula models for the overall sample as well as for the two subsamples. For the evaluation we use the seven statistics χ^2, AIC, BIC, EN, AD, PD and PDS as defined above.

Table 11.3 summarizes the estimation results for the joint distribution functions in the overall sample period. The normal copula yields an estimate of the correlation coefficient of 0.85 which indicates a quite strong positive linear dependence in the yield factors. The subsample results in Tables 11.4 and 11.5 indicate comparable linear dependence, with an estimate of 0.89 in the first and 0.79 in the second subsample. The results also show that the estimates of the standard deviations for the marginal distributions vary somewhat depending on the copula model, where the Gumbel copula assigns the largest standard deviations to the marginals. Considering the subsamples, makes obvious that the standard deviations in the yield factor realizations were higher in earlier subsample period 1982-1991 than in the later 1992-2001 period. This is also notable in the plots of Figure 11.3. A graphical illustration of the estimated 1982-2001 joint distribution

October 1982 to December 1991						
$T = 110$	$\widehat{\sigma}_{1,\cdot}$	$\widehat{\sigma}_{2,\cdot}$	$\widehat{\rho}$	$\widehat{\nu}$	$\widehat{\lambda}_L$	$\widehat{\lambda}_U$
C_t	4.2012E-3	4.0293E-3	0.8871	27.6529	0.2008	0.2008
s.e.	(0.2740E-3)	(0.2684E-3)	(0.0206)	(77.8816)	(0.6241)	(0.6241)
t-value	(15.33)	(15.01)	(43.13)	(0.36)	(0.32)	(0.32)
C_N	4.2051E-3	4.0202E-3	0.8868	-	0	0
s.e.	(0.2746E-3)	(0.2627E-3)	(0.0202)	-	-	-
t-value	(15.31)	(15.31)	(43.79)	-	-	-
	$\widehat{\sigma}_{1,\cdot}$	$\widehat{\sigma}_{2,\cdot}$	$\widehat{\vartheta}$	$\widehat{\delta}$	$\widehat{\lambda}_L$	$\widehat{\lambda}_U$
C_ω	4.1678E-3	4.0233E-3	1.9274	2.6545	0	0.7016
s.e.	(0.2619E-3)	(0.2596E-3)	(1.8253)	(0.6329)	-	(0.0808)
t-value	(15.91)	(15.50)	(1.06)	(2.61)	-	(8.68)
C_G	4.2200E-3	4.0745E-3	-	3.2772	0	0.7645
s.e.	(0.2507E-3)	(0.2526E-3)	-	(0.3146)	-	(0.0251)
t-value	(16.83)	(16.13)	-	(7.24)	-	(30.48)

Table 11.4: Parametric ML-estimates of the joint distribution function G under the alternative copula models. Standard errors and t-values of the parameter estimates given in parenthesis. Sample period 1982 to 1991.

January 1992 to December 2001						
$T = 119$	$\widehat{\sigma}_{1,\cdot}$	$\widehat{\sigma}_{2,\cdot}$	$\widehat{\rho}$	$\widehat{\nu}$	$\widehat{\lambda}_L$	$\widehat{\lambda}_U$
C_t	2.5580E-3	2.9075E-3	0.8148	6.2034	0.4188	0.4188
s.e.	(1.8802E-3)	(0.1980E-3)	(0.0342)	(3.1757)	(0.1185)	(0.1185)
t-value	(13.60)	(14.68)	(23.83)	(1.95)	(3.53)	(3.53)
C_N	2.5634E-3	2.7780E-3	0.7886	-	0	0
s.e.	(0.1864E-3)	(0.1686E-3)	(0.0366)	-	-	-
t-value	(13.75)	(16.48)	(21.52)	-	-	-
	$\widehat{\sigma}_{1,\cdot}$	$\widehat{\sigma}_{2,\cdot}$	$\widehat{\vartheta}$	$\widehat{\delta}$	$\widehat{\lambda}_L$	$\widehat{\lambda}_U$
C_ω	2.5364E-3	2.8771E-3	5.8523	1.3362	0	0.3201
s.e.	(0.1847E-3)	(0.1764E-3)	(2.7736)	(0.3693)	-	(0.2408)
t-value	(13.73)	(16.31)	(2.11)	(0.91)	-	(1.33)
C_G	2.5802E-3	2.9629E-3	-	2.4207	0	0.6684
s.e.	(0.1895E-3)	(0.1963E-3)	-	(0.2360)	-	(0.0372)
t-value	(13.62)	(15.09)	-	(6.02)	-	(17.98)

Table 11.5: Parametric ML-estimates of the joint distribution function G under the alternative copula models. Standard errors and t-values of the parameter estimates given in parenthesis. Sample period 1992 to 2001.

functions under the different copula functions is given in Figure 11.4. The plotted joint density contour lines visualize the dependence implications of the different copulas, while the plotted yield innovations allow for a first visual inspection of model fit.

We next turn to the goodness-of-fit tests in Table 11.6. Given the size of the data set we have to point out that it is impossible to strictly reject any of the copula models. However, it turns out that the transformed Frank copula shows best overall fit and never obtains one of the worst diagnostics results in one of the two subsamples.

Starting with the overall sample, our results clearly indicate that the transformed Frank copula is the superior dependence model. All seven statistics including those which penalize the number of model parameters (χ^2, AIC and BIC) happen to favour the C_ω-model where the second best model follows with some diagnostic distance. Considering the symmetric models, the Student's t-model shows advantages in the χ^2 and the entropy statistics, but not for AIC and BIC. It is remarkable that the AD test always shows very high deviation values for the Student's t-model as compared to the other models. Considering Figure 11.4, an explanation for this finding is that the contours of the Student's t-model gather quickliest in the overall region of the lower left quadrant $[-\infty, 0]\times[-\infty, 0]$ of the joint distribution function causing large relative deviations in the empirical versus theoretical distributions for moderate negative values. Additionally, due to the sample size, the AD statistics should be interpreted with some caution; the number of 230 observations may not fully guarantee convergence of the empirical distribution functions, which is a requirement for the AD statistics. For the probability deviations in the tails, PD and PDS, we choose $p = 0.05$ which yields 11 observations for the calculation of the marginal distribution function. The PD and PDS results indicate that the transformed Frank copula has lowest deviations from the empirical observations in the upper as well as in the lower tail. The symmetric models tend to overestimate the probability of observations in the lower tail, which is demonstrated by large positive values for PD. The Gumbel copula shows a tendency to overestimate the probability of observations in the upper tail showing a large deviation PDS; note that the Gumbel copula models overall dependence and upper tail dependence jointly via the δ parameter which implies strong upper tail dependence under strong uniform dependence, and vice versa.

We next turn to the subsamples, i.e. the 1982-1991 and 1992-2001 subsample results. Note that the PD and PDS statistics are now based on $p = 0.1$ which under a subsamples' size of roughly $T/2$, implies a number of tail observations roughly equal to these for the full sample diagnostics. For the subsamples, the

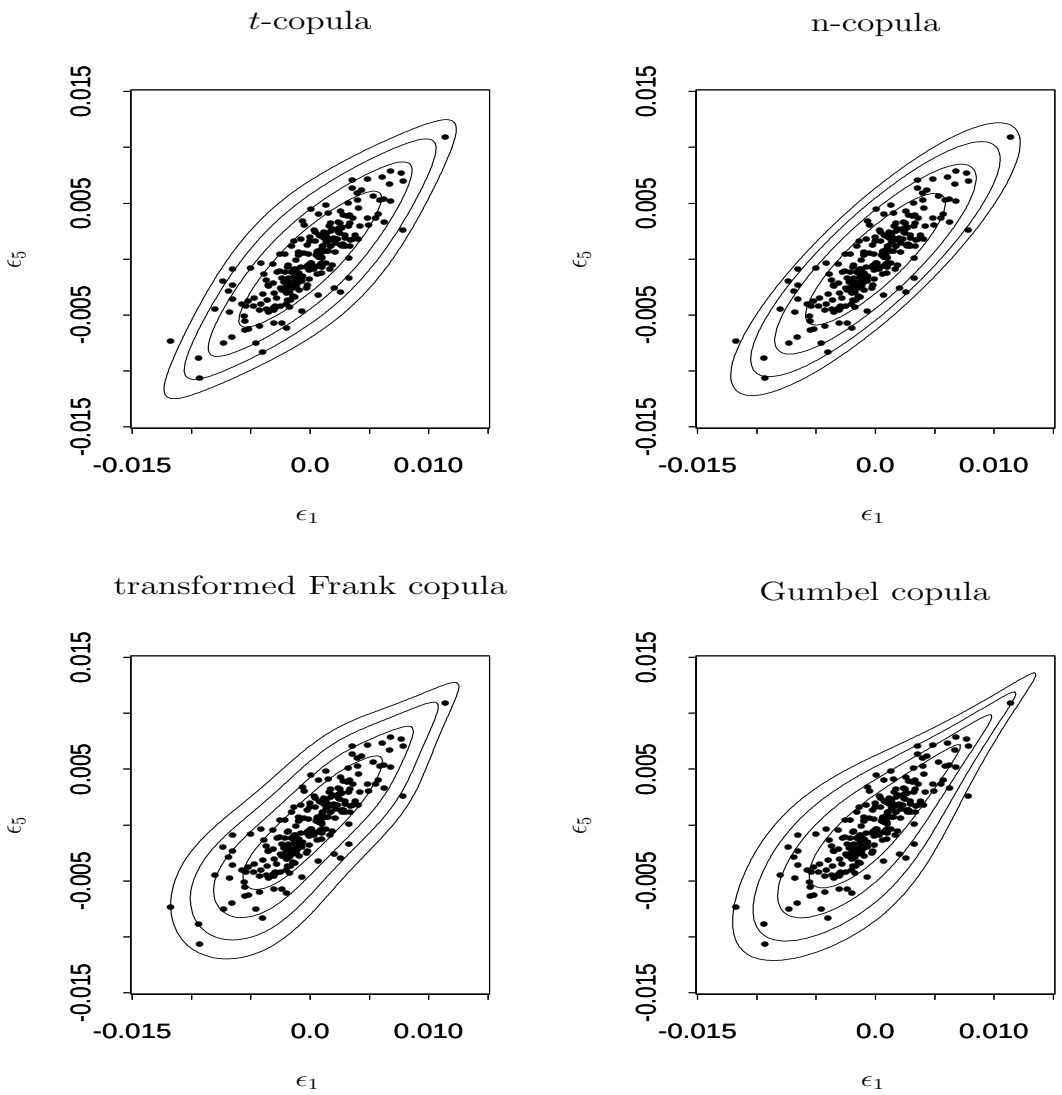

Figure 11.4: Contourlines of the estimated densities under the different copula assumptions with normal marginals. Sample period 1982 to 2001.

October 1982 to December 2001				
$T = 230$	C_t	C_N	C_ω	C_G
$\chi^2_{31,32}$ (p-value)	0.34	0.25^-	0.90^+	0.30
AIC	-4204.30	-4204.24	-4215.34^+	-4203.88^-
BIC	-4188.54^-	-4191.93	-4199.58^+	-4191.57
EN	-9.15	-9.15	-9.19^+	-9.14^-
AD	21.61^-	0.69	0.52^+	0.63
$PD_{p=0.05}$	$22.58\%^-$	20.26%	$6.19\%^+$	9.67%
$PDS_{p=0.05}$	-4.08%	-7.21%	$-3.67\%^+$	$21.76\%^-$
October 1982 to December 1991				
$T = 110$	C_t	C_N	C_ω	C_G
$\chi^2_{11,12}$ (p-value)	0.65	0.73^+	0.47	0.37^-
AIC	-1954.91^-	-1956.80	-1960.44	-1961.27^+
BIC	-1942.11^-	-1946.70	-1947.64	-1951.17^+
EN	-8.92	-8.91^-	-8.96^+	-8.94
AD	14.6768^-	0.6751	0.6134^+	0.6489
$PD_{p=0.1}$	$7.25\%^-$	6.36%	$-0.91\%^+$	-1.86%
$PDS_{p=0.1}$	0.57%	$0.40\%^+$	7.66%	$14.29\%^-$
January 1992 to December 2001				
$T = 119$	C_t	C_N	C_ω	C_G
$\chi^2_{11,12}$ (p-value)	0.86^+	0.78	0.76	0.12^-
AIC	-2267.39^+	-2255.23	-2265.22	-2254.10^-
BIC	-2244.28	-2244.89	-2252.10^+	-2243.76^-
EN	-9.49	-9.51	-9.56^+	-9.47^-
AD	6.0714^-	0.2756^+	0.3732	0.2796
$PD_{p=0.1}$	$0.72\%^+$	-9.81%	-7.35%	$-12.81\%^-$
$PDS_{p=0.1}$	4.35%	-5.68%	$0.41\%^+$	$16.22\%^-$

Table 11.6: Goodness-of-fit statistics for the ML-estimates of the joint distribution function G under the alternative copula models. $^+$indicates best model fit for a given statistics, $^-$indicates worst model fit for a given statistics. Sample period 1982 to 2001 and subsamples 1982 to 1991 and 1992 to 2001.

assignment of the best goodness-of-fit statistics in Table 11.6 shows notable variations across the models. This is due to the substantial decrease in sample size, which makes the interpretation of the results less conclusive than for the overall sample. However, the transformed Frank copula still obtains the best results when evaluated by the number of best fit-results among all models. Also, the C_ω-copula function is never assigned one of the worst-fit results, what does not hold for the other models. Given a smaller data set, the statistics χ^2, AIC and BIC considerably penalize the two-parameter copula functions. At the same time, the C_t parameter estimate of ν in Table 11.4 as well as the C_ω parameter estimate of δ in Table 11.5, exhibit high standard errors with corresponding low respective t-values. A relatively stable pattern in Table 11.6 is provided by the PDS statistics; high deviations for the overall sample as well as for the subsamples point out that a drawback of the Gumbel copula is its tendency to overestimate the upper tail.

Turning to the first subsample, 1982-1991, C_ω shows the best fit according to the EN, AD and PD statistics. The one-parameter copula models C_N and C_G also perform relatively well. The normal model even gives best fit according to the chi-square test statistics. In the second subsample, 1992-2001, as mentioned above, we report lower estimates of the volatility in the marginal distributions. However, our results do not indicate that the dependence structure is much different in the two subsamples. The C_ω-copula again yields results better than for the other models with best fit as measured by the BIC, the EN and the upper tail fit PDS. The Student's t-model has the second worst BIC and worst AD statistics; still it has the best fit according to the χ^2 and AIC measures. The normal copula performs reasonably well having the best AD statistics and the worst BIC. To summarize the subsample comparison results, we can state that —given a high variability in the statistics— the Gumbel copula provides a second best fit in the first subperiod while the normal copula provides a second best fit in the second subperiod. In both subperiods however, the statistics indicate that the transformed Frank copula has best overall fit.

11.4 Application: Measuring Bond Portfolio Risk

Based on the affine term structure model of Section 11.2.1, the term structure of interest rates is completely described by two risk factors. Clearly, the dependence characteristics of the joint distribution of the 1-year yield and the 5-year yield influences the risk measurement of portfolios. In this section we analyze the different distributional specifications' impact on risk management decisions.

We start with a graphical illustration of the copula function estimation results of Section 11.3. Figure 11.5 contains four plots of the fitted conditional densities of

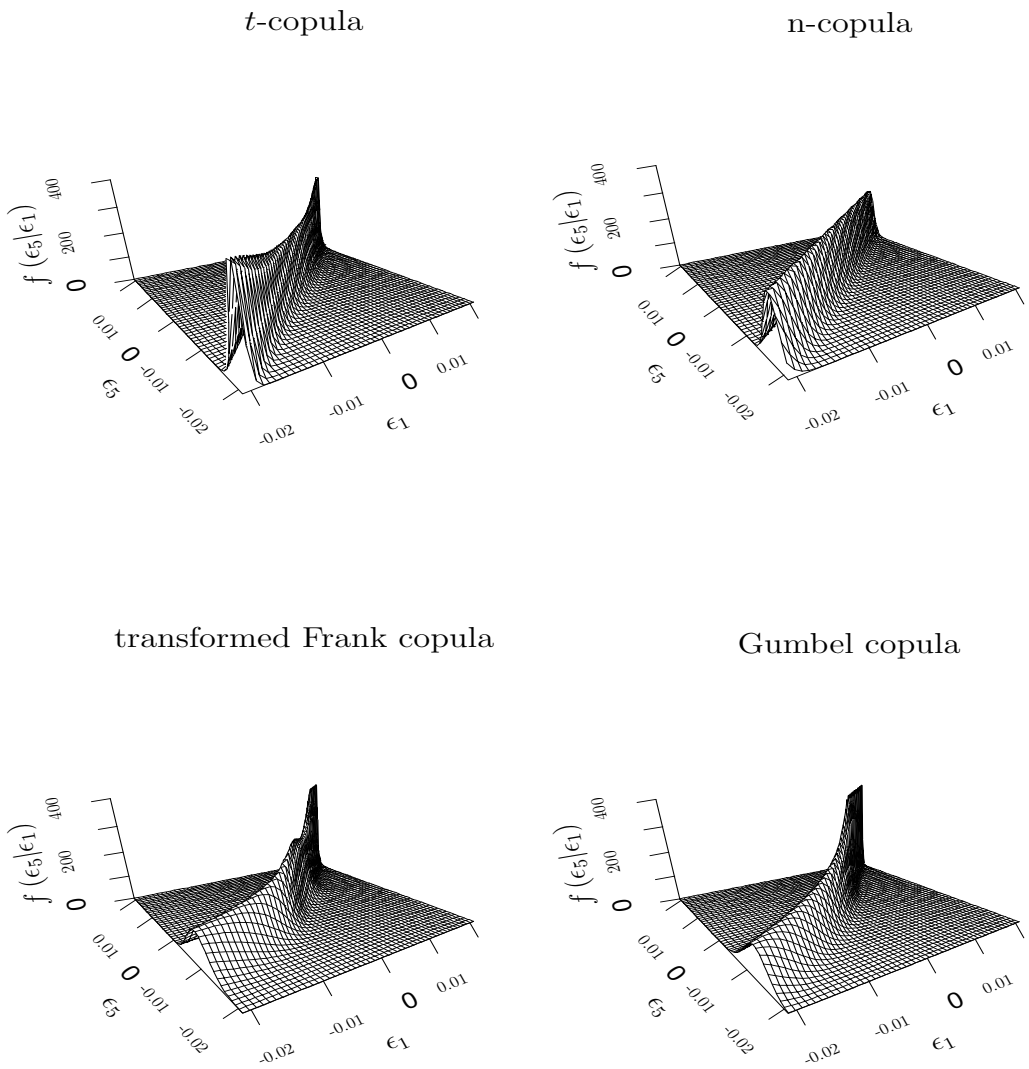

Figure 11.5: Conditional densities under the different fitted copula functions. Sample period 1982 to 2001.

the 5-year yield given a fixed realization of the 1-year yield. Each plot represents one of the four different copula models. As can be seen, the conditional densities show large structural differences especially including the probability of joint upper or lower tail events. For example, given a negative shock to the short rate of –0.02 in the transformed Frank model, the conditional density for the five year yield has high variance while, given a positive shock, the conditional density has low variance. As illustrated, the structure looks much different for the symmetric models.

We next apply the estimated dependence structures to quantify the risk of different bond portfolios and compare the results. We consider portfolios which invest in the 1-year zero bond at price $P(t,1)$ and the 5-year zero bond at price $P(t,5)$ and then study the return of this investment after 1 month of time. Denote by R_Λ the return of the portfolio which has initial portfolio duration Λ. As a risk measure ϱ we utilize Value-at-Risk (VaR), i.e. the quantile of the profit-and-loss distribution of the portfolio. When adjusted for the expected portfolio return VaR is

$$\varrho = F^{\leftarrow}_{R_\Lambda}(\alpha) - \mathbf{E}\{R_\Lambda\},$$

where α is the confidence level and F_{R_Λ} is the cumulative distribution function of R_Λ. We introduce the superscript '+' to the risk measure ϱ when measuring the risk of a long position, and the superscript '−' for measuring the risk of a short position, respectively. Additionally, the subscript at the risk measure ϱ indicates the copula applied for defining the dependence structure. The confidence levels we discuss are $\alpha = 99\%$ and $\alpha = 99.9\%$ for which the VaR numbers ϱ^+_N, ϱ^+_t, ϱ^+_G, ϱ^+_ω, and ϱ^-_N, ϱ^-_t, ϱ^-_G, ϱ^-_ω are calculated.

We compare the risk measures ϱ by fixing the risk measure induced by the normal copula which is the standard risk measure, and calculate the relative deviations from this measure. The relative deviations are

$$\Delta_{\cdot} = \frac{\varrho_{\cdot} - \varrho_N}{\varrho_N},$$

where' $\cdot$ ' indicates the choice of one of the three copula functions C_t, C_ω, and C_G. The relative deviations $\Delta_{\cdot}$ for long and short bond portfolio holdings as a function of the initial duration Λ are plotted in Figure 11.6.

The results for holding a long position in the interest rate portfolio $\Delta^+_{\cdot}$ are displayed in Figure 11.6 (upper plots). The results for the t-copula model Δ^+_t are indicated by the solid line. As can be seen in the upper left plot, the t-copula produces VaR numbers which are close to the normal copula model for the 99% confidence level. When increasing the confidence level to 99.9% in the upper right plot, the maximum relative deviation increases from 0.2% to 3% which reflects the property of the t-copula to adopt to the (upper) tail dependence

existing in our data set. A similar pattern is observed for the model given by transformed Frank copula (see the dashed line). The relative deviation Δ_ω^+ takes a maximum value of approximately 4% for the 99.9% confidence level, and the VaR is persistently larger than the numbers based on the t-copula model. The Gumbel copula (dashed-dotted line) generates the highest VaR. The maximum relative deviation Δ_G^+ is 2.5% for the 99% confidence level, and around 4.5% for the 99.9% confidence level, respectively.

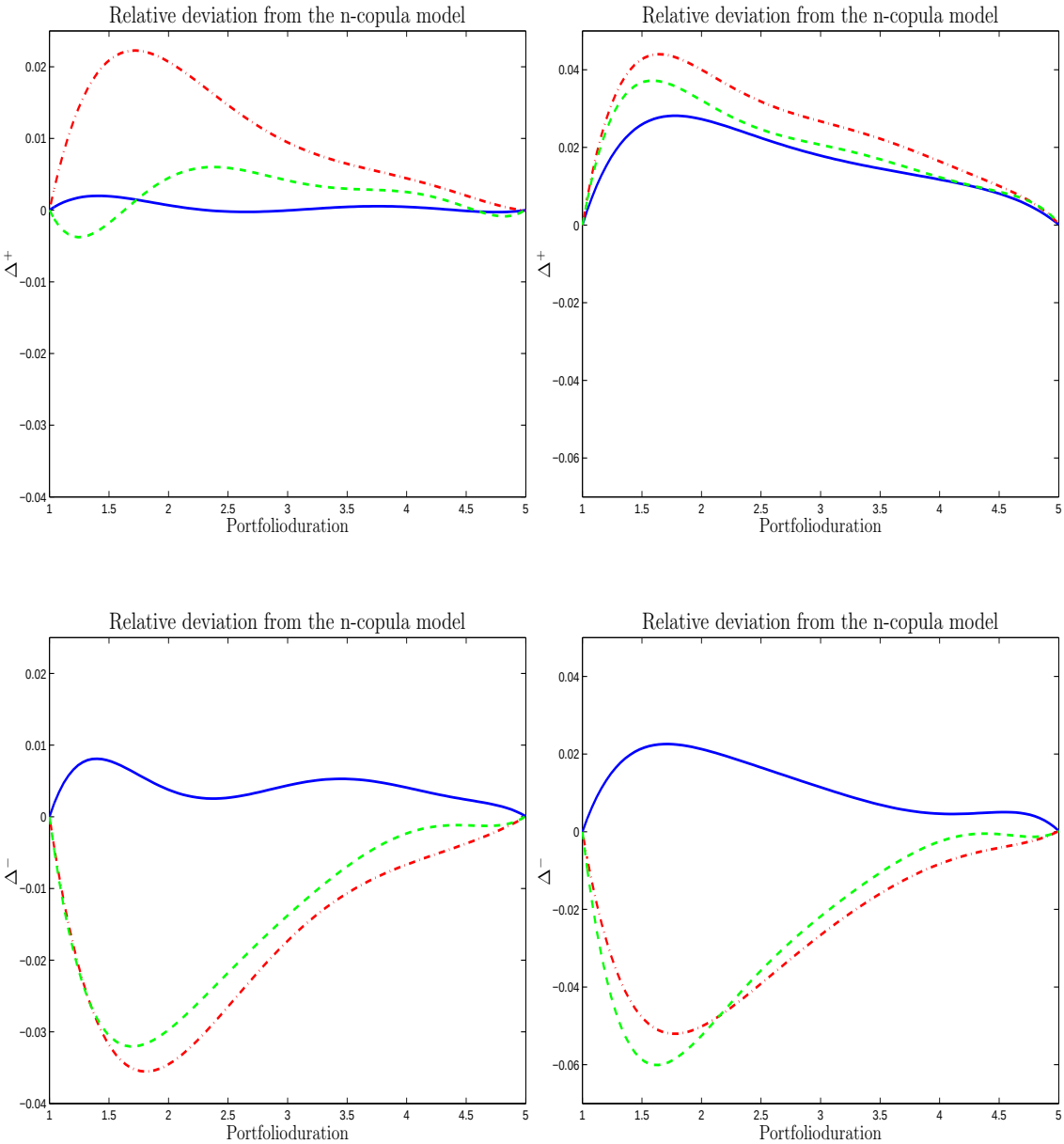

Figure 11.6: Relative deviation of the risk measures from the normal copula model. Upper pictures Δ^+ for the $\alpha = 99\%$ quantile (left), and $\alpha = 99.9\%$

quantile (right) respectively. Lower pictures Δ^- for the $\alpha = 99\%$ quantile (left), and $\alpha = 99.9\%$ quantile (right) respectively. The solid line belongs to the t-copula, the dashed-dotted line to the Gumbel copula, and the dashed line to the transformed Frank copula. Sample period 1982 to 2001.

Figure 11.6 (lower plots) show the results $\Delta^-_{\cdot}$ for a short position in bond portfolios with initial duration Λ. The relative deviation of the t-copula model Δ^-_t has similar characteristics as in the case of the long position. The VaR turns out to be relatively close to the VaR given under the normal copula, where the positive deviations tend to become overall larger when the confidence level is increased from 99% to 99.9%. The t-copula quantiles exceed the normal ones because the t-copula features lower tail dependence which is not present in the data. In contrast to the t-copula, the transformed Frank copula and the Gumbel copula both produce negative relative deviations of the VaR measures when compared to the normal copula. The maximum relative deviation is around 3% for the 99% confidence level and around 6% for the 99.9% confidence level.

The above findings can be interpreted as follows. In Section 11.3.4, the transformed Frank copula proved to be the dependence model which reflects the observed dependence structure in the most appropriate way. Assuming that the data are generated by a joint distribution with normal marginals and a transformed Frank copula then implies that the normal copula produces a systematic bias in measured VaR.

For long bond portfolio positions, the normal copula tends to underestimate VaR where the lack in risk capital may approximately amount to up to 4% in our example. Clearly, the negative bias in VaR produced by the normal copula is related to the upper tail dependence which is present in the data but not a characteristic of the normal dependence model. The t-copula results in VaR numbers which are much closer to the transformed Frank numbers than the normal numbers with a maximum deviation of approximately 1%. This finding is due to the upper tail dependence which is incorporated in the t-copula. The Gumbel copula features characteristics which are present in the analyzed data set: upper tail dependence and asymmetry. The VaR numbers are of reasonable quality especially for the high confidence levels of 99.9%. For the 99% confidence level, the Gumbel copula produces the maximum relative deviation to transformed Frank numbers. For this particular case, the Gumbel copula performs poorly compared to the alternative dependence models (Figure 11.6 upper left plot).

For short bond portfolio positions, the normal copula overestimates VaR. The maximal relative deviation takes a value around 6%. Though the data were not found to show lower tail dependence in Section 11.3.4, which is in accordance with the normal copula, bias is again present but with opposite sign. The explanation for this finding is reasonably simple. As the upper and lower tail of the normal

copula are estimated simultaneously, the realized estimate is a balanced result of both shortcomings of the normal copula, namely its lack of tail dependence and its symmetric characteristic. Also, due to its symmetric structure, the absolute biases generated by the t-copula are high when compared to the transformed Frank model. Hence, the t-copula turns out to produce overestimated VaR numbers for short positions. As it turns out, even for moderate confidence levels of 99%, the copula functions' ability to reproduce a complex observed dependence structure becomes important: The Gumbel copula is a parsimoniously parameterized model which captures upper tail dependence. The relative deviations for the best-fitting transformed Frank model indicate that the normal model can produce VaR biases of up to 6% in the given example.

11.5 Conclusion

As is well-known, the concept of linear dependence breaks down under non-normality. However, as the present investigation documents, statistical theory offers more flexible models of dependence which are relevant to financial modelling.

Based on the benchmark model given by the affine class of term structures which assumes joint normality in yield innovations, this paper analyses cross-sectional nonlinearity in the term structure of U.S.-Treasury yields. The nonlinearities documented in the data represent a profound statistical characteristic which is shown to be of economic significance as well. Deviations from linear dependence have implications on risk management when financial risk is for example measured by the commonly used VaR methodology. Most strikingly, we conclude that the normal copula as a benchmark model of dependence imposes two main problems, namely absence of tail dependence and symmetry, which prevent accurate risk measurement. Our findings are not limited to bond pricing and bond portfolio VaR applications. The model bias due to the normality assumption should be even more pronounced when the pricing implications for nonlinear contracts, i.e. for interest rate derivatives, are considered.

Chapter 12

Generalized Elliptical Distributions - Models and Estimation

This chapter, covered by Frahm and Junker [48], generalizes the class of multivariate elliptical distributions, allowing for asymmetry but preserving all the ordinary components, i.e. the generating variate, the location vector, and the dispersion matrix. Furthermore, some of the limitations mentioned in the former chapters can be avoided. A positive definite ML-estimator for the dispersion matrix which does not require any assumption about the generating distribution function is derived. Even though the estimator recognizes all the available data points it is robust within the class of generalized elliptical distributions. Given robust estimates both for the location and the scatter it is shown how the generating distribution function can robustly be estimated, too. Finally, some generalized elliptical distribution models are presented.
Here we will only give a theoretical outlook. Further results and empirics about the proposed approaches will be given in the forthcoming doctoral thesis by Gabriel Frahm.

12.1 Motivation

Every d-dimensional elliptical random vector X can be represented by $X \stackrel{d}{=} \mu + \mathcal{R}\sqrt{\Sigma}\,U$, where $\mu \in I\!R^d$, $\sqrt{\Sigma} \in I\!R^{d\times k}$, $\sqrt{\Sigma}\sqrt{\Sigma}' =: \Sigma$, U is a k-dimensional random vector uniformly distributed on the unit hypersphere, and $\mathcal{R}$ is a nonnegative random variable independent of U (Cambanis et al. [17]). The covariance matrix of X corresponds to $\mathbf{E}\left(\mathcal{R}^2\right)/k \cdot \Sigma$, provided the second moment of $\mathcal{R}$ is finite (Fang et al. [45]). But estimating the covariance matrix of elliptical random vectors via the method of moments, especially the correlation matrix by Pearson's correlation coefficient is dangerous when the underlying distribution is not normal (Lindskog [76]). This is because Pearson's correlation coefficient is very sensitive to outliers and the smaller the tail index of $\mathcal{R}$ is the higher is the estimator's variance. Indeed, there are a lot of robust techniques to insulate its estimation from the 'bad influence' of outliers (Peña and Prieto [96]). All these

methods principally fall back on the Mahalanobis distance of the observed data points and one either has to detect the outliers so as to eliminate them or to specify another more subtle penalty function applying to extreme realizations. But from the viewpoint of extreme value theory this has the annoying effect of neglecting useful information contained in the extremes. In particular, estimating the tail index of $\mathcal{R}$ is impossible without outliers.

In this article the class of multivariate elliptical distributions is generalized to allow for asymmetry. Nevertheless, all the ordinary components of elliptical distributions, i.e. the generating variate $\mathcal{R}$, the location vector μ and the dispersion matrix Σ remain for this new class of 'generalized elliptical distributions'. We introduce an ML-estimator for Σ, recognizing all the available data points. Positive definiteness is guaranteed and the estimator is robust since it does not require any assumption about the cumulative density function (c.d.f.) of $\mathcal{R}$, i.e. the generating distribution function of the generalized elliptical random vector. Hence it is not disturbed by outliers. Furthermore, given the estimates of location and scatter the empirical generating distribution function can be extracted even preserving the outliers. Hence the tail index of the generating variate $\mathcal{R}$ can be estimated using standard methods from extreme value theory (Embrechts et al. [38]). For the special case of an elliptical random vector X also the tail-dependence coefficient (Joe [62]) of two arbitrary components of X can be estimated given the estimates for the tail index and the dispersion matrix (Frahm et al. [50]).

12.2 Generalized Elliptical Distributions

12.2.1 Definition

Definition 12.2.1. *The d-dimensional random vector X is said to be 'generalized elliptically distributed' if and only if there exist a k-dimensional random vector U, uniformly distributed on the unit hypersphere*

$$\mathcal{S}^{k-1} := \left\{ x \in I\!R^k : \ \|x\|_2 = 1 \right\},$$

a nonnegative random variable $\mathcal{R}$, a vector $\mu \in I\!R^d$, and a matrix $\sqrt{\Sigma} \in I\!R^{d\times k}$, such that

$$X \stackrel{d}{=} \mu + \mathcal{R}\sqrt{\Sigma}\, U. \tag{12.2.1}$$

In contrast to elliptical distributions the generating variate $\mathcal{R}$ may depend stochastically on the direction determined by U. Hence the dependence structure of $\mathcal{R}$ and U constitutes the multivariate c.d.f. of X essentially. In particular, X does not need to be symmetrically distributed anymore, and its covariance matrix is not necessarily equal to $\mathbf{E}\left(\mathcal{R}^2\right)/k \cdot \Sigma$.

Remark 16. *A d-dimensional random vector X is elliptically distributed if and only if there exist a vector $\mu \in I\!R^d$, a positive semidefinite matrix $\Sigma \in I\!R^{d\times d}$, and a function ϕ such that the characteristic function of $X-\mu$ can be represented by $t \mapsto \phi(t'\Sigma t)$ (see, e.g., Cambanis et al. [17]). The function ϕ is called the 'characteristic generator' of X (Fang et al. [45]). Note that in Definition 12.2.1 it is not presumed that* $\operatorname{rank}(\Sigma) = k$. *Nevertheless, if $\mathcal{R}$ and U are stochastically independent, the characteristic function of $X-\mu$ corresponds to $t \mapsto \phi((t'\sqrt{\Sigma})(t'\sqrt{\Sigma})') = \phi(t'\Sigma t)$, where $t \mapsto \phi(t't)$ is the characteristic function of $\mathcal{R}U$. Conversely, if a random vector X is elliptically distributed then X is always representable as in Definition 12.2.1, with $\mathcal{R}$ and U being independent (Cambanis et al. [17]). Hence the class of generalized elliptical distributions includes the class of elliptical distributions.*

Remark 17. *From Remark 16 we see that the dispersion of an elliptically distributed random vector is uniquely defined via the matrix Σ, i.e. the particular matrix decomposition is irrelevant. Intuitively speaking, this is because (under some conditions concerning $\sqrt{\Sigma}$) the linear transformation $u \mapsto \sqrt{\Sigma}\,u$ ($u \in \mathcal{S}^{k-1}$) generates an ellipsoid which is invariant under different decompositions of Σ, and $\mathcal{R}$ does not depend on the realization of U. But for generalized elliptical distributions this is not true and $\sqrt{\Sigma}$ must be —in general— specified.*

Now we will give some remarks on the numerical simulation of generalized elliptical random vectors.

Remark 18. *The k-dimensional random vector $X \sim N(\underline{0}, I_k)$ has the stochastic representation $X \overset{d}{=} \sqrt{\chi_k^2}\,U$, where U is uniformly distributed on $\mathcal{S}^{k-1}$ and stochastically independent of $\sqrt{\chi_k^2}$. Since the relation*

$$U \overset{a.s.}{=} \frac{U}{\|U\|_2} \overset{a.s.}{=} \frac{\mathcal{R}U}{\|\mathcal{R}U\|_2}$$

holds for any positive random variable $\mathcal{R}$, we can always use $X/\|X\|_2$ with $X \sim N(\underline{0}, I_k)$ to simulate the random vector U.

Remark 19. *A d-dimensional generalized elliptically distributed random vector X can be simulated after specifying a location vector $\mu \in I\!R^d$, a matrix $\sqrt{\Sigma} \in I\!R^{d\times k}$, and the conditional distribution functions $r \mapsto F_{\mathcal{R}|U=u}(r) := P(\mathcal{R} \le r \,|\, U = u)$ for every $u \in \mathcal{S}^{k-1}$. Using the conditional quantile functions*

$$p \longmapsto F^{\leftarrow}_{\mathcal{R}|U=u}(p) := \inf\left\{r : F_{\mathcal{R}|U=u}(r) \ge p\right\}, \qquad 0 < p < 1,$$

the random vector X results from

$$X := \mu + F^{\leftarrow}_{\mathcal{R}|U=\tilde{U}}(Z)\,\sqrt{\Sigma}\,\tilde{U},$$

where $\tilde{U}$ is uniformly distributed on $\mathcal{S}^{k-1}$ and can be simulated as described in Remark 18, and $Z \sim U(0,1)$ being stochastically independent of $\tilde{U}$.

Theorem 12.2.2. *Let X be a d-dimensional generalized elliptical random vector with $\mu \in I\!R^d$ and $\sqrt{\Sigma} \in I\!R^{d\times d}$ with $\det(\sqrt{\Sigma}\,) \neq 0$. Further, let the joint c.d.f. of $\mathcal{R}$ and U be absolutely continuous. Then the probability density function (p.d.f.) of X is given by*

$$x \longmapsto f_X(x) = \sqrt{\det(\Sigma^{-1})} \cdot g\left((x-\mu)'\,\Sigma^{-1}(x-\mu)\,;u\right), \qquad x \neq \mu,$$

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

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

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

Proof. Since the joint c.d.f. of $\mathcal{R}$ and U is absolutely continuous the joint p.d.f. $(r,u) \mapsto f_{(\mathcal{R},U)}(r,u)$ exists. Consider the conditional density function of $\mathcal{R}$, i.e. $r \mapsto f_{\mathcal{R}|U=u}(r) := f_{(\mathcal{R},U)}(r,u)\,/f_U(u)$ $(r>0)$, where $f_U(u) = \Gamma\left(\frac{d}{2}\right)/(2\pi^{d/2})$ is the uniform density on the unit hypersphere $\mathcal{S}^{d-1}$ (Fang et al. [45]). Thus the joint p.d.f. of $\mathcal{R}$ and U corresponds to

$$(r,u) \longmapsto f_{(\mathcal{R},U)}(r,u) = \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^{d/2}} \cdot f_{\mathcal{R}|U=u}(r), \qquad r > 0.$$

To get the density of $\mathcal{R}U =: Y$ we define the transformation $h:\ ]0,\infty[\,\times \mathcal{S}^{d-1} \to I\!R^d \setminus \{\underline{0}\}$, $(r,u) \mapsto ru =: y$. Note that h is injective and the p.d.f. of Y is given by

$$y \longmapsto f_Y(y) = f_{(\mathcal{R},U)}\left(h^{-1}(y)\right) \cdot |J_h|^{-1}, \qquad y \neq \underline{0},$$

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

$$\mathcal{S}_r^{d-1} := \left\{x \in I\!R^d :\ \|x\|_2 = r > 0\right\}$$

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

$$J_h = \det\left(\begin{bmatrix} 1 & \underline{0}' \\ \underline{0} & r\cdot I_{d-1} \end{bmatrix}\right) = r^{d-1} = \|y\|_2^{d-1}, \qquad y \neq \underline{0}.$$

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

$$\begin{aligned} y \longmapsto f_Y(y) &= f_{(\mathcal{R},U)}\left(\|y\|_2\,, y/\,\|y\|_2\right) \cdot \|y\|_2^{-(d-1)} \\ &= \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^{d/2}} \cdot \|y\|_2^{-(d-1)} \cdot f_{\mathcal{R}|U=u}\left(\|y\|_2\right), \qquad y \neq \underline{0}, \end{aligned}$$

where $u = y/\|y\|_2$. Now we define the transformation $q\colon \mathbb{R}^d \setminus \{\underline{0}\} \to \mathbb{R}^d \setminus \{\mu\}$, $y \mapsto \mu + \sqrt{\Sigma}\, y =: x$. The Jacobian determinant of $\partial(\mu + \sqrt{\Sigma}\, y)/\partial y'$ corresponds to $J_q = \det(\sqrt{\Sigma}\,)$, and thus the p.d.f. of $X := \mu + \sqrt{\Sigma}\, Y = \mu + \mathcal{R}\sqrt{\Sigma}\, U$ is given by

$$\begin{aligned} x \longmapsto f_X(x) &= f_Y\left(q^{-1}(x)\right) \cdot |J_q|^{-1} \\ &= f_Y\left(\sqrt{\Sigma}^{-1}(x-\mu)\right) \cdot |\det(\sqrt{\Sigma}\,)|^{-1}, \qquad x \neq \mu. \end{aligned}$$

Since $|\det(\sqrt{\Sigma}\,)| = \sqrt{\det(\sqrt{\Sigma}\sqrt{\Sigma}')} = \sqrt{\det(\Sigma)}$ and hence $|\det(\sqrt{\Sigma}\,)|^{-1} = \sqrt{\det(\Sigma^{-1})}$, the p.d.f. of X becomes

$$x \mapsto f_X(x) = \sqrt{\det(\Sigma^{-1})} \cdot \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^{d/2}} \cdot \|\sqrt{\Sigma}^{-1}(x-\mu)\|_2^{-(d-1)} \cdot f_{\mathcal{R}|U=u}\left(\|\sqrt{\Sigma}^{-1}(x-\mu)\|_2\right),$$

with

$$u = \frac{\sqrt{\Sigma}^{-1}(x-\mu)}{\|\sqrt{\Sigma}^{-1}(x-\mu)\|_2},$$

and $x \neq \mu$. This is equivalent to the formula given in Theorem 12.2.2. □

12.2.2 Spectral Density

Definition 12.2.3. *Let $\sqrt{\Sigma} \in \mathbb{R}^{d\times k}$ with $\det(\sqrt{\Sigma}\sqrt{\Sigma}') = \det(\Sigma) \neq 0$. We call the random vector*

$$S_{\sqrt{\Sigma}} := \frac{\sqrt{\Sigma}\, U}{\|\sqrt{\Sigma}\, U\|_2} \tag{12.2.2}$$

the 'unit random vector generated by $\sqrt{\Sigma}$'.

Remark 20. *Let X be a d-dimensional generalized elliptically distributed random vector with $\mu = \underline{0}$, $\sqrt{\Sigma}$ defined as in Definition 12.2.3, and let the generating variate $\mathcal{R}$ be positive. Then*

$$S_{\sqrt{\Sigma}} \overset{a.s.}{=} \frac{\mathcal{R}\sqrt{\Sigma}\, U}{\|\mathcal{R}\sqrt{\Sigma}\, U\|_2} \overset{d}{=} \frac{X}{\|X\|_2}, \tag{12.2.3}$$

which neither depends on the particular c.d.f. of $\mathcal{R}$ nor on the dependence structure of $\mathcal{R}$ and U.

Note that $S_{\sqrt{\Sigma}}$ does not have a density in the Lebesgue sense but a spectral density on the unit hypersphere.

Theorem 12.2.4. *The spectral density function of the d-dimensional unit random vector generated by $\sqrt{\Sigma}$ corresponds to*

$$s \longmapsto \psi(s) = \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^{d/2}} \cdot \sqrt{\det(\Sigma^{-1})} \cdot \sqrt{s'\Sigma^{-1}s}^{\,-d}, \qquad s \in \mathcal{S}^{d-1}. \tag{12.2.4}$$

Proof. Due to the invariance property described in Remark 20 it can be assumed w.l.o.g. that $X \stackrel{d}{=} \sqrt{\chi_k^2}\sqrt{\Sigma}\,U \sim N(\underline{0}, \Sigma)$ (see also Remark 18). The p.d.f. of X under the condition $\|X\|_2 = r$ is

$$x \longmapsto \varphi_r(x) := c_r^{-1} \cdot \varphi_X(x), \qquad x \in \mathcal{S}_r^{d-1},$$

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

$$S_{\sqrt{\Sigma}} \stackrel{d}{=} \frac{X}{\|X\|_2}, \qquad X \sim N(\underline{0}, \Sigma),$$

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

$$s \longmapsto \psi_r(s) = \varphi_r\left(h^{-1}(s)\right) \cdot |J_{h^{-1}}| = c_r^{-1} \cdot \varphi_X(rs) \cdot |J_{h^{-1}}|, \qquad s \in \mathcal{S}^{d-1}.$$

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

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

Since the tangent plane on the hypersphere $\mathcal{S}_r^{d-1}$ has only $d-1$ dimensions (see also the proof of Theorem 12.2.2) $J_{h^{-1}}$ is given by

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

Now the p.d.f. of $S_{\sqrt{\Sigma}}$ is

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

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

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

□

Remark 21. *If* $\Sigma = I_d$ *then* $\psi(s) = \Gamma(d/2) / \left(2\pi^{d/2}\right)$ *for every* $s \in \mathcal{S}^{d-1}$. *Thus the unit random vector generated by* I_d *is uniformly distributed on* $\mathcal{S}^{d-1}$ *and* $2\pi^{d/2}/\Gamma(d/2)$ *is the surface area of* $\mathcal{S}^{d-1}$.

Remark 22. *The spectral density function* ψ *is invariant under linear transformations* $\Sigma \mapsto \sigma\Sigma$, $\sigma > 0$, *since*

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

and

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

Remark 23. *The unit random vector* $S_{\sqrt{\Sigma}}$ *is generalized elliptically distributed, since* $S_{\sqrt{\Sigma}} \overset{d}{=} \mu + \mathcal{R}\sqrt{\Sigma}\,U$, *with* $\mu = \underline{0}$ *and* $\mathcal{R} = \|\sqrt{\Sigma}\,U\|_2^{-1}$. *Anyhow, the distribution of* $S_{\sqrt{\Sigma}}$ *does not depend on the particular matrix decomposition* $\Sigma = \sqrt{\Sigma}\sqrt{\Sigma}'$ *(see Remark 17).*

12.2.3 Estimation

In the following we will derive some robust estimation methods for the location vector μ, the dispersion matrix Σ, and the generating distribution function $r \mapsto F_{\mathcal{R}}(r) := P(\mathcal{R} \le r)$, assuming that the d-dimensional random vector X is generalized elliptically distributed. The sample realizations of n independent copies of X are denoted by

$$x := \begin{bmatrix} x_{\cdot 1} & x_{\cdot 2} & \cdots & x_{\cdot n} \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1n} \\ x_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{d1} & \cdots & \cdots & x_{dn} \end{bmatrix}.$$

Remark 24. *Let* $h_1, \ldots, h_d$ *be some continuous and strictly monotone increasing functions. If* $Y \sim F$, *where* F *is some multivariate generalized elliptical c.d.f., then* $x \mapsto G(x) := F(h_1(x_1), \ldots, h_d(x_d))$ *is the multivariate c.d.f. of the random vector* $X := (h_1^{\leftarrow}(Y_1), \ldots, h_d^{\leftarrow}(Y_d))$. *This can be used for modelling new distribution functions based on the class of generalized elliptical distributions. Conversly, if* X *is not generalized elliptically distributed, one possibly could nevertheless find some strictly monotone increasing functions* $h_1, \ldots, h_d$ *such that the random vector* $h(X) := (h_1(X_1), \ldots, h_d(X_d))$ *is generalized elliptically distributed. Then the multivariate c.d.f. of* X *is given by* $x \mapsto G(x) = F(h(x))$, *where* F *is the multivariate generalized elliptical c.d.f. of* $h(X)$. *This can be used for pre-processing, i.e. the estimation methods for generalized elliptical distributions are to be applied on the transformed data* $h(x_{\cdot j})$ $(j = 1, \ldots, n)$ *so as to obtain* $\widehat{F}$ *and thus* $\widehat{G} = \widehat{F} \circ h$.

Hence the following estimation methods can be applied even if the observed data are not generalized elliptically distributed provided the transformations $h_1, \ldots, h_d$ are known. Suppose that both the marginal distributions $G_1, \ldots, G_d$ of X and the marginal distributions $F_1, \ldots, F_d$ of Y are continuous and that the copula C of X is generated by Y, i.e. $P(G_1(X_1) \leq u_1, \ldots, G_d(X_d) \leq u_d) = P(F_1(Y_1) \leq u_1, \ldots, F_d(Y_d) \leq u_d)$ for all $(u_1, \ldots, u_d) \in [0,1]^d$. Then the transformations $h_1, \ldots, h_d$ are given by $h_i = F_i^{\leftarrow} \circ G_i$ $(i = 1, \ldots, d)$ since the multivariate c.d.f. of X corresponds to

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

where $F_1^{\leftarrow}, \ldots, F_d^{\leftarrow}$ are the marginal quantile functions of F, i.e. $p \mapsto F_i^{\leftarrow}(p) := \inf\{x_i : F_i(x_i) \geq p\}$, $p \in]0,1[$ $(i = 1, \ldots, d)$. The c.d.f. G is said to be a 'meta-elliptical distribution function' (Fang et al. [44]).

Estimating the dispersion matrix Σ

First, we assume that the assumptions of Remark 20 are fulfilled. Then the dispersion matrix Σ of X can be estimated up to a linear transfomation by maximizing the log-likelihood function

$$\log\left(\mathcal{L}\left(\tilde{\Sigma}; s_{\cdot 1}, ..., s_{\cdot n}\right)\right) \propto n \cdot \log\left(\det(\tilde{\Sigma}^{-1})\right) - d \cdot \sum_{j=1}^{n} \log\left(s_{\cdot j}' \tilde{\Sigma}^{-1} s_{\cdot j}\right), \quad (12.2.5)$$

where

$$s_{\cdot j} := \frac{x_{\cdot j}}{\|x_{\cdot j}\|_2}, \qquad j = 1, ..., n.$$

Even though this is a true ML-procedure there is no need for information about the generating distribution function. In particular, the estimator does not depend of the moments of X being finite and not even on their existence. This is due to the separation of the radial and the angular part of X caused by relation (12.2.3).

Proposition 12.2.5. *If* $\Sigma = \sqrt{\Sigma}\sqrt{\Sigma}'$, *then* $\det(\Sigma) \neq 0$ *if and only if* Σ *is positive definite.*

Proof. If $\det(\Sigma) \neq 0$ then $\operatorname{rank}(\sqrt{\Sigma}') = d$, implying that $\|\sqrt{\Sigma}'x\|_2 > 0$ for every $x \neq \underline{0}$. Hence $0 < (\sqrt{\Sigma}'x')(\sqrt{\Sigma}'x) = x'\sqrt{\Sigma}\sqrt{\Sigma}'x = x'\Sigma x$ $(\forall\, x \neq \underline{0})$. The 'if' is obvious. $\square$

Since $\det(\Sigma) \neq 0$ the matrix $\tilde{\Sigma}$ must be positive definite and we can apply the Cholesky decomposition $\tilde{\Sigma} = LL'$, where L is a real-valued lower triangular matrix. Thus $\tilde{\Sigma}^{-1} = (LL')^{-1} = (L^{-1})' L^{-1} = \Gamma'\Gamma$, with $\Gamma := L^{-1}$ being a lower triangular matrix, too. Instead of the log-likelihood function (12.2.5) the reparametrized form

$$\log(\mathcal{L}(\Gamma; s_{\cdot 1}, ..., s_{\cdot n})) \propto n \cdot \log(\det(\Gamma'\Gamma)) - d \cdot \sum_{j=1}^{n} \log\left(s'_{\cdot j}\Gamma'\Gamma s_{\cdot j}\right) \qquad (12.2.6)$$

is proposed for two reasons. Firstly, if the function (12.2.6) is maximized numerically the algorithm does not have to calculate the inverse of $\tilde{\Sigma}$ anymore which prevents suboptimal solutions due to $\tilde{\Sigma}$ being close to singularity. Secondly, it is easy to control the positive definiteness of $\tilde{\Sigma}$ by means of Γ which is shown by the following proposition.

Proposition 12.2.6. *Let $\Gamma \in I\!R^{d\times d}$ be a lower triangular matrix. The matrix $(\Gamma'\Gamma)^{-1}$ is positive definite if and only if $\Gamma_{ii} \neq 0$ for all $i = 1, ..., d$.*

Proof. Since the determinant of any triangular matrix corresponds to the product of the entries on its main diagonal, $\det(\Gamma) \neq 0$ if and only if $\Gamma_{ii} \neq 0$ for all $i = 1, ..., d$. This is equivalent to the existence of Γ^{-1} and thus of $(\Gamma'\Gamma)^{-1} = \Gamma^{-1}\Gamma^{-1\prime}$. Since Γ^{-1} is nonsingular the positive definiteness follows directly from Proposition 12.2.5. □

Remark 25. *When the constraint $\Gamma_{ii} \neq 0$ $(i = 1, ..., d)$ is fulfilled in each iteration, the existence and positive definiteness of $\tilde{\Sigma} = (\Gamma'\Gamma)^{-1}$ is guaranteed. Conversely, since every admissible $\tilde{\Sigma}$ is positive definite there is one and only one lower triangular matrix L such that $LL' = \tilde{\Sigma}$ and $\det(\tilde{\Sigma}) = \det^2(L) > 0$, i.e. the matrix L must be nonsingular. Then the inverse L^{-1} exists and corresponds to Γ. That is to say each matrix $\tilde{\Sigma}$ has a unique representation in Γ, and vice versa. So the reparametrized log-likelihood function (12.2.6) must lead to the same maximum as the original one (12.2.5).*

Remark 26. *The log-likelihood function (12.2.6) is invariant under linear transformations of Γ (see also Remark 22). Thus another constraint must be considered to get a unique solution for the maximum likelihood. A simple straight forward one is given by $\Gamma_{1,1} = 1$. An alternative constraint is given by*

$$\Gamma_{1,1} = \frac{1}{\prod_{i=2}^{d} \Gamma_{ii}},$$

so that the determinant of Γ is always 1 what is equivalent to $\det(\tilde{\Sigma}) = 1$.

Let Γ^* be the ML-estimator of Γ, i.e. $\Gamma^* := \arg\max \log\left(\mathcal{L}\left(\Gamma\,;\cdot\right)\right)$. Due to the equivariance property of Γ^* discussed in Remark 25 the ML-estimator of the dispersion matrix Σ corresponds (up to scale) to

$$\widehat{\Sigma}_S := \left(\Gamma^{*\prime}\Gamma^*\right)^{-1} = \sigma\hat{\Sigma}, \tag{12.2.7}$$

where $\widehat{\Sigma}$ is the virtual estimator of Σ and $\sigma > 0$ is an unknown scaling constant. Let the correlation matrix of X be defined as $\rho := \left[\Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}\,\right]$ even if the second moment of the generating variate $\mathcal{R}$ is infinite. Note that $\Sigma_{ii} > 0$ for every $i = 1, \ldots, d$ since Σ is positive definite. The correlation matrix can be estimated by

$$\hat{\rho}_S := \left[\frac{\widehat{\Sigma}_{S,ij}}{\sqrt{\widehat{\Sigma}_{S,ii}\cdot\widehat{\Sigma}_{S,jj}}}\right], \tag{12.2.8}$$

robustly. We call both $\widehat{\Sigma}_S$ and $\widehat{\rho}_S$ 'spectral estimators', respectively.

Estimating the location vector μ

Now let μ be an arbitrary real-valued location vector, i.e. $\mu \in I\!R^d$. If μ is known, we can apply the spectral estimator simply on

$$s_{\cdot j} := \frac{x_{\cdot j} - \mu}{\left\|x_{\cdot j} - \mu\right\|_2}, \qquad j = 1, ..., n.$$

Otherwise, the location vector must be substituted by an adequate estimate $\widehat{\mu}$.

Definition 12.2.7. *Let $F_i^{\leftarrow}$ be the i-th marginal quantile function (for the definition see Remark 24) of an arbitrary d-dimensional random vector X ($i = 1, \ldots, d$). The median of X is defined as the vector*

$$x_{0.5} := \left[\frac{1}{2}\cdot\left(F_i^{\leftarrow}\left(0.5^-\right) + F_i^{\leftarrow}\left(0.5^+\right)\right)\right].$$

If $F_i^{\leftarrow}\left(0.5^-\right) = F_i^{\leftarrow}\left(0.5^+\right)$ for all $i = 1, \ldots, d$, then we say that X has a 'strict median'.

Proposition 12.2.8. *Let the random vector X be generalized elliptically distributed with location μ. If X has a strict median, then μ corresponds to the median of X.*

Proof. Because the median of X is strict, it is sufficient to prove that

$$P\left(X_i \le \mu_i\right) \ge 0.5 \quad \wedge \quad P\left(X_i < \mu_i\right) \le 0.5, \qquad i = 1, \ldots, d. \tag{12.2.9}$$

Let $V := \sqrt{\Sigma}\, U$ and V_i be the i-th component of V. Since $X_i \stackrel{d}{=} \mu_i + \mathcal{R}V_i$, assertion (12.2.9) is equivalent to

$$P\left(\mathcal{R}V_i \leq 0\right) \geq 0.5 \quad \wedge \quad P\left(\mathcal{R}V_i < 0\right) \leq 0.5, \qquad i = 1, \ldots, d.$$

Obviously, $P\left(V_i \leq 0\right) = 0.5$. Furthermore,

$$\begin{aligned} P\left(\mathcal{R}V_i \leq 0\right) &= \int_{-\infty}^{\infty} P\left(\mathcal{R}v_i \leq 0 \,|\, V_i = v_i\right) dF_{V_i}\left(v_i\right) \\ &\geq \int_{-\infty}^{0} P\left(\mathcal{R}v_i \leq 0 \,|\, V_i = v_i\right) dF_{V_i}\left(v_i\right) = 0.5, \qquad i = 1, \ldots, d, \end{aligned}$$

since $P\left(\mathcal{R}v_i \leq 0 \,|\, V_i = v_i\right) = 1$ for P-almost all $\omega \in \Omega$ with $V_i\left(\omega\right) = v_i \leq 0$, $i = 1, \ldots, d$. Otherwise

$$\begin{aligned} P\left(\mathcal{R}V_i < 0\right) &= \int_{-\infty}^{\infty} P\left(\mathcal{R}v_i < 0 \,|\, V_i = v_i\right) dF_{V_i}\left(v_i\right) \\ &= \int_{-\infty}^{0} P\left(\mathcal{R}v_i < 0 \,|\, V_i = v_i\right) dF_{V_i}\left(v_i\right) \leq 0.5, \qquad i = 1, \ldots, d, \end{aligned}$$

since $P\left(\mathcal{R}v_i < 0 \,|\, V_i = v_i\right) = 0$ for P-almost all $\omega \in \Omega$ with $V_i\left(\omega\right) = v_i > 0$, $i = 1, \ldots, d$. $\square$

Let $x_{i1}^*, \ldots, x_{in}^*$ be a permutation of the i-th component realizations $x_{i1}, \ldots, x_{in}$ such that $x_{i1}^* \leq \ldots \leq x_{in}^*$, $i = 1, \ldots, d$. Further, let $x_{\cdot j}^*$ be the j-th column vector of the matrix $x^* := \left[x_{ij}^*\right]$. The sample median

$$\widehat{x}_{0.5,n} := \begin{cases} \frac{1}{2} \cdot \left(x_{\cdot,0.5\cdot n}^* + x_{\cdot,0.5\cdot n+1}^*\right), & n \text{ even}, \\ x_{\cdot,0.5\cdot(n+1)}^*, & n \text{ odd}, \end{cases}$$

is proposed as a robust estimator for μ due to its persistence against non existing higher moments. That is to say the sample median is not disturbed by outliers.

Remark 27. *Since we presume that X has a strict median, the sample median converges strongly to the theoretical one (Pestman, [97]). Due to Proposition 12.2.8 this is equivalent to $\widehat{x}_{0.5,n} \rightarrow^{a.s.} \mu$ as $n \rightarrow \infty$.*

Now the spectral estimator can be applied on

$$s_{\cdot j} := \frac{x_{\cdot j} - \widehat{x}_{0.5}}{\left\| x_{\cdot j} - \widehat{x}_{0.5} \right\|_2}, \qquad j = 1, ..., n,$$

where $\widehat{x}_{0.5} := \widehat{x}_{0.5,n}$ for a given sample size n. Hence the spectral estimator becomes pseudo-ML.

Estimating the Generating Distribution Function

For estimating the generating distribution function we require additionally that $k = d$, i.e. $\sqrt{\Sigma} \in I\!R^{d\times d}$ being nonsingular. Then, due to the stochastic representation (12.2.1) of generalized elliptical random vectors the radial variable $\mathcal{R}$ is given by

$$\mathcal{R} \overset{a.s.}{=} \|\mathcal{R}U\|_2 \overset{d}{=} \|\sqrt{\Sigma}^{-1}(X-\mu)\|_2.$$

This is equivalent to the Mahalanobis distance of X because

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

Thus, having the spectral estimate $\widehat{\Sigma}_S$ of the dispersion matrix Σ, the realizations of $\mathcal{R}_1, \ldots, \mathcal{R}_n$ can be estimated by

$$\widehat{\mathcal{R}}_{S,j} := \sqrt{(x_{\cdot j} - \widehat{x}_{0.5})'\,\widehat{\Sigma}_S^{-1}(x_{\cdot j} - \widehat{x}_{0.5})} = \frac{\widehat{\mathcal{R}}_j}{\sqrt{\sigma}}, \qquad j = 1, \ldots, n, \tag{12.2.10}$$

up to an unknown scale parameter $\sigma > 0$, too. Here $\widehat{\mathcal{R}}_j$ symbolizes the virtually estimated realization of $\mathcal{R}_j$, i.e.

$$\widehat{\mathcal{R}}_j := \sqrt{(x_{\cdot j} - \widehat{x}_{0.5})'\,\widehat{\Sigma}^{-1}(x_{\cdot j} - \widehat{x}_{0.5})}, \qquad j = 1, \ldots, n.$$

Obviously, $\widehat{\Sigma}_S^{-1}$ in Eq. (12.2.10) can be replaced by $\Gamma^{*\prime}\Gamma^*$ from Eq. (12.2.7), i.e.

$$\widehat{\mathcal{R}}_{S,j} = \sqrt{(x_{\cdot j} - \widehat{x}_{0.5})'\,\Gamma^{*\prime}\Gamma^*(x_{\cdot j} - \widehat{x}_{0.5})} = \|\Gamma^*(x_{\cdot j} - \widehat{x}_{0.5})\|_2, \qquad j = 1, \ldots, n.$$

Remark 28. *Suppose that the survival function of $\mathcal{R}$ is regularly varying with tail index $\alpha > 0$ (Resnick [100]), i.e.*

$$\overline{F}_{\mathcal{R}}(x) = \mathcal{L}(x) \cdot x^{-\alpha}, \qquad x > 0,$$

where $\mathcal{L}$ is a positive, Lebesgue measurable function on $]0, \infty[$ with $\mathcal{L}(tx)/\mathcal{L}(x) \to 1$, $x \to \infty$, for every $t > 0$. Then the survival function of $\mathcal{R}/\sqrt{\sigma}$ is regularly varying with the same tail index. Hence for estimating the tail index it is sufficient to observe $\mathcal{R}$ merely up to scale.

Remark 29. *Suppose that the random vector X is elliptically distributed with $\mathcal{R}$ possessing a regularly varying survival function with tail index $\alpha > 0$. The tail-dependence coefficient (Joe [62]) of two arbitrary components of X, say X_i and X_j, corresponds to*

$$\lambda_{ij} = \lim_{u \searrow 0} \frac{C(u,u)}{u} = 2 \cdot \overline{t}_{\alpha+1}\left(\sqrt{\alpha+1} \cdot \sqrt{\frac{1-\rho_{ij}}{1+\rho_{ij}}}\right), \tag{12.2.11}$$

where C is the copula of (X_i, X_j) and $\bar{t}_{\alpha+1}$ is the survival function of Student's univariate t-distribution with $\alpha+1$ degrees of freedom (Frahm et al. [50] in conjunction with Schmidt [104]). Let $\widehat{\alpha}_S$ be an estimate of the tail index based on $\widehat{\mathcal{R}}_{S,1}, \widehat{\mathcal{R}}_{S,2}, \ldots, \widehat{\mathcal{R}}_{S,n}$ given by Eq. (12.2.10). Further, let $\widehat{\rho}_{S,ij}$ be the spectral estimate of the correlation coefficient of X_i and X_j. Then the tail-dependence coefficient can be estimated coherently by substituting α and ρ_{ij} in Eq. (12.2.11) by their corresponding spectral estimates $\widehat{\alpha}_S$ and $\widehat{\rho}_{S,ij}$.

12.3 Models

If the radial variable $\mathcal{R}$ is stochastically independent of the direction vector U, then the well investigated traditional class of multivariate elliptically contoured distributions emerges (see Cambanis et al. [17], as well as Fang et al. [45]). This class contains a lot of well-known multivariate distributions like, e.g., the Gaussian distribution, the sub-Gaussian α-stable distribution, the symmetric generalized hyperbolic distribution, Student's t-distribution, etc. Note that every elliptical distribution is symmetric. To allow for asymmetry $\mathcal{R}$ must depend in some way on U. In the following we will develop a feasible method for the modelling of asymmetric generalized elliptical distributions.

Let $v_1, \ldots, v_m$ be some fixed 'reference vectors' on the unit hypersphere. We assume that the conditional c.d.f. of $\mathcal{R}$ is a function of some 'distances' $\delta(u, v_1), \ldots, \delta(u, v_m)$ between u and the reference vectors $v_1, \ldots, v_m$, i.e.

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

where $H(\cdot, \delta_1, \ldots, \delta_m)$ is a c.d.f. for all $(\delta_1, \ldots, \delta_m)$, with $H(r, \delta_1, \ldots, \delta_m) = 0$ for every $r \leq 0$.

For an adequate definition of the reference vectors $v_1, \ldots, v_m$ we diagonalize the dispersion matrix Σ, i.e. $\Sigma = V \Lambda V'$, where V is the orthonormal basis of the eigenvectors and Λ is the diagonal matrix of the eigenvalues of Σ. Hence we obtain the diagonal root $\sqrt{\Sigma} = V\sqrt{\Lambda}$ of Σ. Let $Y := \sqrt{\Lambda}\,\mathcal{R}U$, then $X = \mu + VY$. Now we can interpret the components of $Y = (Y_1, \ldots, Y_d)$ as the 'driving factors' of X, each factor having another variance determined by the corresponding eigenvalue, and direction determined by its eigenvector. Note that the eigenvectors are determined only up to sign. That is to say if w is an eigenvector of Σ then w can be substituted by its negative conjugate $-w$. Now we define both the eigenvectors $v_1^+, \ldots, v_d^+$ and their negative conjugates $v_1^-, \ldots, v_d^-$ as reference vectors.

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

Theorem 12.3.1. *Let the d-dimensional random vector U be uniformly distributed on the unit hypersphere. The c.d.f. of the angle $\measuredangle(U,v)$ between U and a given reference vector $v \in \mathcal{S}^{d-1}$ is*

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

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

$$F_{\text{Beta}}(x\,;\beta_1,\beta_2) = \int_0^x \frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\,\Gamma(\beta_2)} \cdot t^{\beta_1-1}(1-t)^{\beta_2-1}\,dt, \qquad 0 \le x \le 1,$$

is the Beta distribution function with parameters $\beta_1, \beta_2 > 0$.

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

$$P\left(\measuredangle(U,v) \le a\right) = P\left(U_1 \le \cos(\pi - a)\right) = P\left(U_1 \le -\cos(a)\right).$$

The p.d.f. of U_1 corresponds to

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

see Fang et al. [45]. If $0 \le a < \pi/2$, after substituting u by $-\sqrt{t}$ we get

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

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

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

□

Now we define

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

and propose it as a distance measure taking the number of dimensions adequately into consideration. Note that $\delta(u,v) \in [0,1]$ is the area of the spherical cap on $\mathcal{S}^{d-1}$ spanned by u and v divided by the surface area of $\mathcal{S}^{d-1}$ (see Remark 21). For $d = 2$ the distance $\delta(u,v)$ becomes simply $\arccos\langle u,v\rangle/\pi$, for instance. So one can interpret δ as a probabilistic generalization of the radian measure for d dimensions.

There is an interesting relationship between the tail-dependence coefficient λ of elliptical random pairs and the distance measure δ stated as follows.

Corollary 12.3.2. *Consider Student's univariate t-distribution with ν degrees of freedom, i.e.*

$$t_\nu(x) = \int_{-\infty}^{x} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{1}{\sqrt{\nu}} \cdot \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt, \qquad \nu > 0. \tag{12.3.1}$$

For $x \le 0$ the univariate t-distribution can be represented by

$$t_\nu(x) = \frac{1}{2} \cdot \overline{F}_{\text{Beta}}\left(\frac{x^2}{x^2+\nu}; \frac{1}{2}, \frac{\nu}{2}\right) = P\left(\measuredangle(U,v) \le \arccos\sqrt{\frac{x^2}{x^2+\nu}}\right),$$

where U is uniformly distributed on the unit hypersphere S^ν, $\nu > 0$. Further, let the random vector X be elliptically distributed satisfying the conditions of Remark 29. Then the tail-dependence coefficient of two arbitrary components of X corresponds to

$$\lambda_{ij} = 2 \cdot P\left(\measuredangle(U,v) \le \arccos\sqrt{\frac{1-\rho_{ij}}{2}}\right), \qquad \rho_{ij} \in [-1,1],$$

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

Proof. Substituting t in Eq. (12.3.1) by $-\sqrt{\nu\left((1-u)^{-1}-1\right)}$ $(0 \le u < 1)$ leads to

$$\begin{aligned}t_\nu(x) &= \int_{1}^{l_\nu(x)} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{1}{\sqrt{\nu}} \cdot (1-u)^{\frac{\nu+1}{2}} \cdot \left(-\frac{1}{2}\right) \cdot \sqrt{\frac{\nu}{(1-u)^{-1}-1}} \cdot (1-u)^{-2}\, du \\ &= \frac{1}{2} \cdot \int_{l_\nu(x)}^{1} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)} \cdot u^{\frac{1}{2}-1} \cdot (1-u)^{\frac{\nu}{2}-1}\, du \\ &= \frac{1}{2} \cdot \overline{F}_{\text{Beta}}\left(l_\nu(x); \frac{1}{2}, \frac{\nu}{2}\right) = P\left(\measuredangle(U,v) \le \arccos\sqrt{l_\nu(x)}\right),\end{aligned}$$

where

$$l_\nu(x) := 1 - \frac{1}{1+\frac{x^2}{\nu}} = \frac{x^2}{x^2+\nu}, \qquad x \le 0,$$

and U is uniformly distributed on $\mathcal{S}^\nu$, $\nu > 0$. Further, the tail-dependence coefficient (see Remark 29) corresponds to

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

where U is uniformly distributed on $\mathcal{S}^{\alpha+1}$, and

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

such that

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

Note that if $\rho_{ij} = -1$ the tail-dependence coefficient always corresponds to 0 (Frahm et al. [50]). □

With the definitions above we now can give some examples for generalized elliptical distributions.

Example 1. *Let the conditional c.d.f. of $\mathcal{R}$ be $r \mapsto P(\mathcal{R} \le r \,|\, U = u) = P(\phi(u) \cdot \mathcal{R}^* \le r)$, where*

$$\phi(u) = \phi_0 + \sum_{i=1}^{d} \alpha_i \delta^{\vartheta_i}\left(\frac{\sqrt{\Sigma}\,u}{\|\sqrt{\Sigma}\,u\|_2}, v_i^+\right) + \sum_{i=1}^{d} \beta_i \delta^{\theta_i}\left(\frac{\sqrt{\Sigma}\,u}{\|\sqrt{\Sigma}\,u\|_2}, v_i^-\right), \qquad (12.3.2)$$

with $\phi_0 > 0$, $\alpha_1, \ldots, \alpha_d, \beta_1, \ldots, \beta_d \ge 0$, $\vartheta_1, \ldots, \vartheta_d, \theta_1, \ldots, \theta_d > 0$. Further, $\mathcal{R}^*$ is a positive random variable possessing a p.d.f. and being stochastically independent of U. Hence $f_{\mathcal{R}|U=u}(\cdot) = f_{\mathcal{R}^*}(\cdot/\phi(u))/\phi(u)$ and due to Theorem 12.2.2 the multivariate p.d.f. of the random vector X is given by

$$f_X(x) = \sqrt{\det(\Sigma^{-1})} \cdot \sigma^{-d}(x) \cdot g\left(\frac{(x-\mu)'\,\Sigma^{-1}(x-\mu)}{\sigma^2(x)}\right), \qquad x \ne \mu,$$

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

$$\sigma(x) := \phi\left(\sqrt{\Sigma}^{-1}(x-\mu)\right).$$

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

Example 2. *If $\mathcal{R}^2/d \sim F_{d,\nu}$, where $F_{d,\nu}$ is the F-distribution with d and ν degrees of freedom, then $\mathcal{R}$ is the generating variate of Student's d-variate*

t-distribution with ν degrees of freedom (Hult and Lindskog [60]). Let the conditional c.d.f. of $\mathcal{R}$ be $r \mapsto P(\mathcal{R} \leq r \mid U = u) = P(\mathcal{R}^2/d \leq r^2/d \mid U = u) = F_{d,\phi(u)}(r^2/d)$, where ϕ is the positive scaling function (12.3.2) defined in Example 1. Similarly to the p.d.f. of Student's multivariate t-distribution the p.d.f. of X is given by

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

where $x \neq \mu$ and $\nu(x) \equiv \sigma(x)$. For the degenerate case $\alpha_1, \ldots, \alpha_d, \beta_1, \ldots, \beta_d = 0$ we get Student's d-variate t-distribution with location μ, dispersion Σ and ϕ_0 degrees of freedom. Moreover, for $\phi_0 \to \infty$ the d-variate Gaussian distribution $N(\mu, \Sigma)$ appears.

In Figure 12.1 we see some density contour lines of Example 1 (left hand) and of Example 2 (right hand) where $d = 2$, $\mu = \underline{0}$, $\Sigma_{11} = \Sigma_{22} = 1$, and $\Sigma_{12} = \Sigma_{21} = 0.5$. The density generator of Example 1 corresponds to the density generator of Student's bivariate t-distribution with 100 degrees of freedom. For each example there is only one reference vector, more precisely $v_1^+ = (\cos(\pi/4), \sin(\pi/4))$ for Example 1 and $v_1^- = (-\cos(\pi/4), -\sin(\pi/4))$ for Example 2. The parameters are $\phi_0 = 1$, $\alpha_1 = 0.25$, and $\vartheta_1 = 1$, as well as $\phi_0 = 2$, $\beta_1 = 98$, and $\theta_1 = 2$, respectively. All remaining parameters are set to zero, i.e. $\beta_1 = 0$ in Example 1 and $\alpha_1 = 0$ in Example 2, and especially $\alpha_2 = \beta_2 = 0$ in both examples. The dashed contour lines symbolize the density of Student's bivariate t-distribution with 100 degrees of freedom with the location and dispersion given above. This corresponds to the degenerate cases $\alpha_1 = 0$, and $\phi_0 = 100$, $\beta_1 = 0$, respectively.

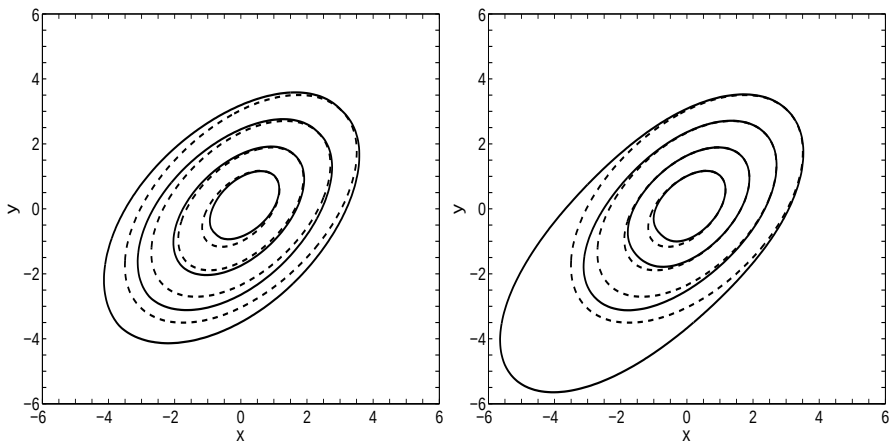

Figure 12.1: Density contours of Example 1 (left hand) and of Example 2 (right hand). The degenerate cases are represented by the dashed contour lines.

Conclusion

There was a lot to do.
We have done a lot.

Now there is a lot to do!

Bibliography

[1] Aït-Sahalia, Y. (1996), Testing Continuous-Time Models of the Spot Interest Rate, *Review of Financial Studies* **9**, 385-426

[2] Anderson, T. W., Darling, D. A. (1952), Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes, *Annals of Mathematical Statistics* **23**, 193-212

[3] Ané, T., Kharoubi, C. (2001), Dependence Structure and Risk Measure, *The Journal of Business*, forthcoming

[4] Ang, A., Bekaert, G. (2000), Short Rate Nonlinearities and Regime Switches, *Journal of Economic Dynamics and Control*, forthcoming

[5] Ahn, D. H., Dittmar, R. F., Gallant A. R. (2002), Quadratic Term Structure Models: Theory and Evidence, *Review of Financial Studies* **15**, 243-288

[6] Babbs, S. H., Nowman, K. B. (1998), Econometric Analysis of a Continuous Time Multi-Factor Generalized Vasicek Term Structure Model: International Evidence, *Asia-Pacific Financial Markets* **5**, 159-183

[7] Babbs, S. H., Nowman, K. B. (1999), Kalman Filtering of Generalized Vasicek Term Structure Models, *Journal of Financial and Quantitative Analysis* **34**, 115-130

[8] Bingham, N. H., Kiesel, R. (2002), Semi-parametric modelling in finance: theoretical foundation, *Quantitative Finance* **2**, 241-250

[9] Bingham, N. H., Kiesel, R., Schmidt, R. (2002), Semi-parametric models in finance: Econometric applications, *www.mathematik.uni-ulm.de/finmath/*

[10] Björk, T., Kabanov, Y., Runggaldier, W. (1997), Bond Market Structure in the Presence of Marked Point Processes, *Mathematical Finance* **7**, 211-239

[11] Blattberg, R.C., Gonedes, N. (1974), A comparison of the stable and student distributions as models for stock prices, *Journal of Business* **47**, 244-280

[12] Bödewadt, U. F. (1942), Die Kettenregel für höhere Ableitungen, *Mathematische Zeitschriften* **48**, 735-746

[13] Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity. In: *Journal of Econometrics* **31**, 307-327

[14] Breymann W., Dias, A., Embrechts, P. (2003), Dependence Structures for Multivariate High-Frequency Data in Finance, *Quantitative Finance*, forthcoming

[15] Brandt, M. W., He, P. (2002), Simulated Likelihood Estimation of Affine Term Structure Models from Panel Data, working paper, University of Pennsylvania

[16] Büning, H., Trenkler, G. (1978), Nichtparametrische statistische Methoden, *Walter de Gruyter Verlag*, Berlin, New York

[17] Cambanis, S., Huang, S., Simons, G. (1981), On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis* **11**, 368-385

[18] Casella, G., Berger, R. L. (2002), Statistical Inference, *Duxbury Press Pacific Grove* 2. edn.

[19] Cassidy, C., Gizycki, M. (1997), Measuring traded market risk: Value-at-Risk and Backtesting Techniques. Discussion Paper, *Reserve Bank of Australia*

[20] Chan, K. C., Karolyi, G. A., Longstaff, F. A., Sanders, A. B. (1992), An Empirical Comparison of Alternative Models of the Short-Term Interest Rate, *Journal of Finance* **47**, 1209-1227

[21] Chen, R., Scott, L. (1992), Pricing Interest Rate Options in a Two-Factor Cox-Ingersoll-Ross Model of the Term Structure, *Review of Financial Studies* **5**, 613-636

[22] Coles, S., Heffernan, J., Tawn, J. (1999), Dependence Measures for Extreme Value Analyses, *Extremes* **2**, 339-365

[23] Cox, J. C., Ingersoll, J. E., Ross, S. A. (1985), A Theory of the Term Structure of Interest Rates, *Econometrica* **53**, 385-407

[24] Dai, Q., Singleton, K. J. (2000), Specification Analysis of Affine Term Structure Models, *Journal of Finance* **55**, 1943-1978

[25] Dai, Q., Singleton, K. J. (2002), Fixed Income Pricing, Mimeo, *NYU and Stanford University*

[26] Danielsson, J., De Vries, C.G. (1997), Tail Index and Quantile Estimation with Very High Frequency Data. In: *Journal of Empirical Finance* **4**, 241-257.

[27] De Jong, F. (2000), Time-Series and Cross-Section Information in Affine Term-Structure Models, *Journal of Business and Economic Statistics* **18**, 300-314

[28] Dewachter, H., Lyrio, M., Maes, K. (2002), The Effect of Monetary Union on German Bond Markets, working paper, *University of Leuven*

[29] Diebold, F. X., Gunther, T. A., Tay, A. S. (1997), Evaluating Density Forecasts, working paper, *Federal Reserve Bank of Philadelphia*

[30] Draisma, G., Drees, H., Ferreira, A., de Haan, L. (2001), Tail dependence in independence, *Eurandom* preprint

[31] Drees, H., Kaufmann, E. (1998), Selecting the optimal sample fraction in univariate extreme value estimation, *Stochastic Processes and their Applications* **75**, 149-172

[32] Duan, J. C., Simonato, J. G. (1999), Estimating Exponential-Affine Term Structure Models by Kalman Filter, *Review of Quantitative Finance and Accounting* **13**, 111-135

[33] Duffie, D., Kan, R. (1996), A Yield-Factor Model of Interest Rates, *Mathematical Finance* **6**, 379-406

[34] Duffie, D., Pan, J. (1997), An Overview of Value-at-Risk. In: *Journal of Derivatives* **4**, 7-49

[35] Durrleman, V., Nikeghbali A., Roncalli T. (2000), A simple transformation of copulas, *Crédit Lyonnais*, working paper

[36] Eberlein, E., Keller, U. (1995), Hyperbolic distributions in finance. *Bernoulli* **1**, 281–299

[37] Eberlein, E., Raible, S. (1999), Term Structure Models driven by General Lévy Processes, *Mathematical Finance* **9**, 31-53

[38] Embrechts, P., Klüppelberg, C., Mikosch, T. (2003), Modelling extremal events (for insurance and finance), Corrected 4th printing, *Springer*, Berlin

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

[40] Embrechts, P., Lindskog, F. , McNeil, A. (2001), Modelling Dependence with Copulas and Applications to Risk Management, *ETH Zürich* preprint

[41] Embrechts, P., McNeil, A., Straumann, D. (2002), Correlation and dependence in risk management: properties and pitfalls, in: M. Dempster, ed., *Risk management: value-at-risk and beyond (Cambridge University Press)*

[42] Embrechts P, McNeil A., Strautmann, D. (1999), Correlation and Dependence in Risk Management: Properties and Pitfalls, *ETH Zürich* preprint

[43] Fama, E. F., Bliss, R. R. (1987), The Information in Long-Maturity Forward Rates, *American Economic Review* **77**, 680-692

[44] Fang, H.B., Fang, K.T., Kotz, S. (2002), The meta-elliptical distributions with given marginals, *Journal of Multivariate Analysis* **82**, 1-16

[45] Fang, K.T., Kotz, S., Ng, K.W. (1990), Symmetric Multivariate and Related Distributions, *Chapman and Hall*, London

[46] Feller, W. (1971), An Introduction to Probability Theory and Its Applications Volume II, *Wiley series in probability and mathematical statistics*, second edition, *Wiley*, New York and other

[47] Fermanian, J-D., Radulović, D., Wegkamp, M. (2002), Weak convergence of empirical copula processes, preprint

[48] Frahm, G., Junker, M. (2003), Generalized Elliptical Distributions: Models and Estimation, *caesar* working paper

[49] Frahm, G., Junker, M., Schmidt, R. (2003), Estimating the tail-dependence coefficient Properties and Pitfalls, working paper

[50] Frahm, G., Junker, M., Szimayer, A. (2003), Elliptical copulas: applicability and limitations, *Statistics and Probability Letters* **63**, Issue 3, 275-286

[51] Frank, M. J. (1979), On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$, *Aequationes Math.* **19**, 194-226

[52] Frey, R., McNeil, A. (2000), Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach, *Journal of Empirical Finance* **7**, 271-300

[53] Genest, C., Ghoudi, K., Rivest, L-P. (1995), A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions, *Biometrika* **82**, 543-552.

[54] Giles, D. E. A. (2000), A Saddlepoint Approximation to the Distribution Function of the Anderson-Darling Test Statistic, *University of Victoria* Econometrics working paper

[55] Giliberto, S. M. (1990), Equity Real Estate Investment Trust and Real Estate Returns *The Journal of Real Estate Research* **5**, 259-263

[56] Glascock, J. L., Davidson III, W. N. (1995), Performance Measures of Real Estate Firm Common Stock Returns, *Alternative Ideas in Real Estate Investment*, Boston, 143-156

[57] Harvey, A. C. (1989), Structural Time Series Models and the Kalman Filter, *Cambridge University Press*, Cambridge

[58] Hrycej, T., Schmidt, R., Stützle, E. (2002), Multidimensional data modelling with generalized hyperbolic distributions, *www.mathematik.uni-ulm.de/finmath/*

[59] Huang, X. (1992), Statistics of Bivariate Extreme Values, *Thesis Publishers and Tinbergen Institute*

[60] Hult, H., Lindskog, F. (2001), Multivariate extremes, aggregation and dependence in elliptical distributions, *Risklab* (Switzerland) working paper

[61] Johnson, N.L., Kotz, S., Balakrishnan, N. (1995), *Continuous Univariate Distributions, Volume 2*, 2nd ed., *Wiley*, Canada

[62] Joe, H. (1997), Multivariate Models and Dependence Concepts *Monographs on Statistics and Applied Probability* **73**, *Chapmann & Hall*, London

[63] Jorion, P. (1997), Value-at-Risk, *Irvin*, Chicago

[64] Junker, M. (2002), European real estate firms in crash situations, *caesar* preprint

[65] Junker, M. (2003), An asymmetric extreme value Copula, *caesar* working paper

[66] Junker, M., May, A. (2002), Measurement of Agregate Risk with Copulas, *caesar* preprint

[67] Junker, M., May, A. (2003), Measurement of Agregate Risk with Copulas, *caesar* working paper

[68] Junker, M., Szimayer, A. (2002), A probability density applying to different types of heavy tail behaviors, *caesar* preprint

[69] Junker, M., Szimayer, A., Wagner, N., Nonlinear Term Structure Dependence: Copula Functions, Empirics, and Risk Implications, *www.ssrn.com*

[70] Juri, A., Wühterich, M. V. (2002), Copula convergence theorems for tail events, *Insurance Math. Econom.*, **24**, 139-148

[71] Kiesel, R., Schmidt, R. (2003), Extreme dependency in financial data, *www.mathematik.uni-ulm.de/finmath/*

[72] Kim, C. (1994), Stochastic dominance, Pareto optimality, and equilibrium asset pricing, *EconWPA* working paper

[73] Klüppelberg, C., May, A. (1998), The dependence function for bivariate extreme value distributions - a systematic approach, *Munich University of Technology* preprint

[74] Kupiec, P. H. (1995), Techniques for Verifying the Accuracy of Risk Measurement Models, *The Journal of Derivatives* **3**, 73-84

[75] Ledford, A.W., Tawn J.A. (1996), Statistics for near independence in multivariate extreme values, *Biometrika* **83**, 169-187

[76] Lindskog, F. (2000), Linear correlation estimation, *Risklab*, preprint

[77] Lindskog, F., McNeil, A., Schmock, U. (2001), Kendall's tau for elliptical distributions, *Risklab* (Switzerland) working paper

[78] Liu, C. H., Mei, J. (1992), The Predictability of Returns on Equity REITs and Their Co-Movement with Other Assets, *Journal of Real Estate Finance and Economics* **5**, 401-418

[79] Longin, F. M. (1999), From Value at Risk to Stress Testing: the Extreme Value Approach, *Journal of Banking and Finance* **7**, 1097-1130

[80] Longstaff, F. A., Schwartz, E. S. (1992), Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model, *Journal of Finance* **47**, 1259-1282

[81] Lund, J. (1997), Non-Linear Kalman Filtering Techniques for Term-Structure Models, *Aarhus School of Business* working paper

[82] Malevergne, Y., Sornette, D. (2002), Tail dependence of factor models, *University of Nice* preprint

[83] Malevergne, Y., Sornette, D. (2002), Testing the Gaussian Copula Hypothesis for Financial Assets Dependences, *UCLA* working paper

[84] Manzotti, A., Perez, F. J., Quiroz, A. J. (2002), A statistic for testing the null hypothesis of elliptical symmetry, *Journal of Multivariate Analysis* **81**, 274-285

[85] Mardia, K. V. (1972), Statistics of Directional Data, *Academic Press* London and New York

[86] Maurer, R., Sebastian, S. (1998), Immobilienfonds und Immobilienaktiengesellschaften als finanzwirtschaftliche Substitute für Immobiliendirektanlagen *University of Mannheim SFB 504* working paper

[87] May, A., Szimayer, A. (2001), Testing for Conditional Heteroscedasticity in Financial Time-Series, *Technische Universität Darmstadt* preprint

[88] Moore, D.S. (1986), Tests of Chi-squared Type, *Goodness-of-Fit Techniques*, D'Agostino, R.B. and M.A. Stevens (Eds.), *Dekker*, New York.

[89] Müller, A., Scarsini, M. (2001), Stochastic comparison of random vectors with a common copula, *Mathematics of Operations Research* **26**, 723-740

[90] Müller, A., Stoyan, D. (2002), Comparison Methods for Stochastic Models and Risks, *Wiley Series in Probability and Statistics*, *Wiley*, Chichester

[91] Myer, F. C. N., Webb, J. R. (1993), Return Properties of Equity REITs, Common Stocks, and Commercial Real Estate: A Comparison, *The Journal of Real Estate Research* **8**, 87-106

[92] Nelson, R. B. (1999), An Introduction to Copulas, *Lecture Notes in Statistics* **139** *Springer*, New York

[93] Ortobelli, S., Huber, I., Schwartz, E. (2002), Portfolio selection with stable distributed returns, *Mathematical Methods of Operations Research* **55**, 265-300.

[94] Owen, J., Rabinovitch, R. (1983), On the class of elliptical distributions and their applications to the theory of portfolio choice, *Journal of Finance* **38**, 745-752

[95] Patton, A. J. (2001), Modelling Time-Varying Exchange Rate Dependence Using the Conditional Copula, *U.C. San Diego* working paper

[96] Peña, D., Prieto, F.J. (2001), Multivariate outlier detection and robust covariance matrix estimation, *Technometrics* **43**, 286-301

[97] Pestman, W.R. (1998), Mathematical statistics, *De Gruyter* Berlin

[98] Pickands, J. (1981), Multivariate extreme value distributions, *Proceedings of the 43rd session og the ISI Buenos Aires* **49**, 859-878

[99] Rachev, S., Mittnik, S. (2000), Stable Paretian Models in Finance, *Wiley*, Chichester

[100] Resnick, S. (1987), Extreme Values, Regular Variation, and Point Processes, *Springer*, New York

[101] Rockinger, M., Jondeau, E. (2001), Conditional Dependency of Financial Series: An Application of Copulas, *CEPR and HEC* working paper

[102] Rosenblatt, M. (1952), Remarks on a multivariate transformation, *Annals Mathematical Statistics* **23**, 470-472

[103] Scaillet, O. (2002), A Nonparametric Analysis of Stock Index Return Dependence through Bivariate Copulas, *European Investment Review* **1**, 7-16

[104] Schmidt, R. (2002), Tail dependence for elliptically contoured distributions, *Mathematical Methods of Operations Research* **55**, 301-327

[105] Schmidt, R. (2003), Credit risk modelling and estimation via elliptical copulae, Credit Risk: Measurement, Evaluation and Management, ed. G. Bohl, G. Nakhaeizadeh, S.T. Rachev, T. Ridder and K.H. Vollmer, *Physica-Verlag* Heidelberg

[106] Schmidt, R., Stadtmüller, U. (2002), Nonparametric estimation of tail dependence, *submitted, www.mathematik.uni-ulm.de/finmath/*

[107] Stanton, R. (1997), A Nonparametric Model of the Term Structure Dynamics and the Market Price of Interest Rate Risk, *Journal of Finance* **52**, 1973-2002

[108] Sobol', I. M. (1994), A primer for the Monte Carlo method, *CRC Press*, Boca Raton

[109] Vasicek, O. A. (1977), An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics* **5**, 177-188

[110] White, H. (1982), Maximum Likelihood Estimation of Misspecified Models, *Econometrica* **50**, 1-25

[111] Widder, D. V. (1941), The Laplace Transform, *Princeton University Press*, Princeton

Appendix A

A Probability Density Applying to Asymmetric Pareto Type Tails

A.1 Proof of Proposition 1.2.4:

To show (a) we need to calculate the first derivative of $h_{\delta,s,\beta}$. For $x \in \mathbb{R}$ we have

$$h'_{\delta,s,\beta}(x) = \frac{\beta[1+\exp(-\delta\cdot\frac{x}{\beta})]\cdot[1+\frac{x}{\sqrt{s^2+x^2}}]+\delta\exp(-\delta\cdot\frac{x}{\beta})\cdot[x+\sqrt{s^2+x^2}]}{2\cdot[1+\exp(-\delta\cdot\frac{x}{\beta})]^2}.$$

From $\delta, s, \beta > 0$ we conclude $|x| < \sqrt{s^2+x^2}$, $x+\sqrt{s^2+x^2} > 0$ and $1+\frac{x}{\sqrt{s^2+x^2}} > 0$. Hence $h'_{\delta,s,\beta} > 0$, i.e. $h_{\delta,s,\beta}$ is strictly increasing. So the functions F_θ and $F^\star_\theta$ are increasing and continuous, since $h_{\delta,s,\beta}$, $h^\star_{\delta,s,\beta,\xi}$ and $G_{\xi,\beta}$ are increasing and continuous on their support. F_θ is a df as a convex linear combination of two dfs.

To show (b), let $\xi_l, \xi_r \geq 0$, then $F_\theta \in \mathbf{C}^\infty(\mathbb{R},[0,\,1])$, since $h_{\delta,s,\beta}$ and G_ξ are compositions of $\mathbf{C}^\infty$-functions. In particular, the density of F_θ is given by

$$f_\theta(x) = \alpha f^l_\theta(x) + (1-\alpha) f^r_\theta(x), \tag{A.1.1}$$

where

$$f^l_\theta(x) = h'_{\delta_l,s_l,\beta_l}(\rho - x)\cdot g_{\xi_l}(h_{\delta_l,s_l,\beta_l}(\rho - x)) \tag{A.1.2}$$

$$f^r_\theta(x) = h'_{\delta_r,s_r,\beta_r}(x - \rho)\cdot g_{\xi_r}(h_{\delta_r,s_r,\beta_r}(x - \rho)), \tag{A.1.3}$$

and $g_{\xi,\beta}$ denotes the density of G_ξ.

If $\xi_l < 0$ or $\xi_r < 0$, then the preceding differentiability statement holds except for x_l and x_r, respectively. In this cases, we find for the density of $F_\theta^\star$

$$f_\theta^\star(x) \;=\; \begin{cases} 1_{(x_l,\,\infty)}(x)\,\alpha f_\theta^l(x) + (1-\alpha) f_\theta^r(x), & \xi_l < 0, \xi_r \geq 0 \\ \alpha f_\theta^l(x) + 1_{(-\infty,\,x_r)}(x)\,(1-\alpha) f_\theta^r(x), & \xi_l \geq 0, \xi_r < 0 \\ 1_{(x_l,\,\infty)}(x)\,\alpha f_\theta^l(x) + 1_{(-\infty,\,x_r)}(x)\,(1-\alpha) f_\theta^r(x), & \xi_l, \xi_r < 0. \end{cases} \tag{A.1.4}$$

Note that due to the conditions on s_l, s_r, we have $x_l < \rho < x_r$. Furthermore, $f_\theta^\star$ is not defined in x_l and x_r. □

A.2 Lemma A.2.1

Lemma A.2.1. *Let $g \in \mathcal{R}_{-\gamma}^+$ and $h \in \mathcal{R}_{-\vartheta}^+$, where $\gamma, \vartheta > 0$. Then $\frac{g}{h} \in \mathcal{R}_{-\gamma+\vartheta}^+$. If $\gamma = \infty$ and $\vartheta > 0$, then $\frac{g}{h} \in \mathcal{R}_{-\infty}^+$. Moreover, if $\gamma > \vartheta$ or $\gamma = \infty$ then $\lim_{x\to+\infty} \frac{g(x)}{h(x)} = 0$.*

Proof. For $\gamma < \infty$ we find

$$\begin{aligned} \lim_{x\to+\infty} \frac{g(tx)/h(tx)}{g(x)/h(x)} &= \lim_{x\to+\infty} \frac{g(tx)}{g(x)} \lim_{x\to+\infty} \frac{h(x)}{h(tx)} \\ &= t^{-\gamma}\, t^{\vartheta} \quad = \quad t^{-\gamma+\vartheta} \end{aligned}$$

hence $\frac{g}{h} \in \mathcal{R}_{-\gamma+\vartheta}^+$. If $\gamma = \infty$, the result follows in an analogous way.

For $\vartheta < \gamma < \infty$, we have $\frac{g}{h} \in \mathcal{R}_{-\gamma+\vartheta}^+$, where $-\gamma+\vartheta < 0$. By Corollary A3.4 in Embrechts et al. [39] we find $\lim_{x\to+\infty} \frac{g(x)}{h(x)} = 0$. If $\gamma = \infty$, we have $\frac{g}{h} \in \mathcal{R}_{-\infty}^+$. By application of Theorem A3.12 (b) in Embrechts et al. [39] the result is shown. □

Appendix B

Elliptical Copulas: Applicability and Limitations

B.1 Multidimensional Sub-Gaussian Standard Density

The standard characteristic function of the d-dimensional sub-Gaussian α-stable distribution is

$$\phi_{\text{stable}}\left(t;\underline{0},\underline{1},\rho,\alpha\right)=\exp\left(-\left(\frac{1}{2}\cdot t\rho t'\right)^{\alpha/2}\right),\qquad t=\left(t_1,...,t_d\right). \tag{B.1.1}$$

Hence ϕ_{stable} is an even function due to the quadratic form contained in (B.1.1). Furthermore, the integral

$$\int\limits_{\mathbb{R}^d}\exp\left(-\left(\frac{1}{2}\cdot t\rho t'\right)^{\alpha/2}\right)dt$$

converges absolutely, that is to say the Fourier-transform

$$f^*_{\alpha,\rho}\left(x\right)=\frac{1}{\left(2\pi\right)^d}\cdot\int\limits_{\mathbb{R}^d}\exp\left(-\left(\frac{1}{2}\cdot t\rho t'\right)^{\alpha/2}\right)\cdot\exp\left(-itx'\right)dt \tag{B.1.2}$$

exists.

At first, we consider the bivariate case. Since

$$\begin{aligned}\exp\left(-itx'\right)&=\exp\left(-it_1x_1\right)\cdot\exp\left(-it_2x_2\right)\\&=\left(\cos\left(t_1x_1\right)-i\sin\left(t_1x_1\right)\right)\left(\cos\left(t_2x_2\right)-i\sin\left(t_2x_2\right)\right)\\&=\left(\cos\left(t_1x_1\right)\cos\left(t_2x_2\right)-\sin\left(t_1x_1\right)\sin\left(t_2x_2\right)\right)\\&\qquad-i\left(\cos\left(t_1x_1\right)\sin\left(t_2x_2\right)+\sin\left(t_1x_1\right)\cos\left(t_2x_2\right)\right),\end{aligned}$$

the real part of $\exp(-itx')$ is an even function whereas its imaginary part is an odd function of $t = (t_1, t_2)$. Because the characteristic function is even also the real part of the integrand

$$\exp\left(-\left(\frac{1}{2}\cdot t\rho t'\right)^{\alpha/2}\right)\cdot\exp(-itx')$$

is even, and the imaginary part remains odd, too. Hence the imaginary part can be eliminated from the Fourier-transform (B.1.2), and with the addition formula

$$\cos(\varphi\pm\theta) = \cos(\varphi)\cos(\theta)\mp\sin(\varphi)\sin(\theta). \tag{B.1.3}$$

we get

$$f^*_{\alpha,\rho}(x) = \frac{1}{(2\pi)^2}\cdot\int\limits_{I\!R^2}\exp\left(-\left(\frac{1}{2}\cdot t\rho t'\right)^{\alpha/2}\right)\cdot\cos(tx')\,dt.$$

This can be extended analogously for $d > 2$ dimensions.

B.2 Equivalent Representation of Equation (3.3.5)

We recall Equation (3.3.5) is

$$\lambda(X_i, X_j; \zeta, \rho_{ij}) = \frac{\int_0^{f(\rho_{ij})}\frac{u^\zeta}{\sqrt{u^2-1}}\,du}{\int_0^1\frac{u^\zeta}{\sqrt{u^2-1}}\,du}, \qquad \zeta > 0,\ f(\rho_{ij}) = \sqrt{\frac{1+\rho_{ij}}{2}}.$$

Substituting u by $\cos(v)$ leads to

$$\begin{aligned}\lambda(X_i, X_j; \zeta, \rho_{ij}) &= \frac{\int_{\arccos(0)}^{\arccos(f(\rho_{ij}))}\frac{\cos^\zeta(v)}{\sqrt{\cos^2(v)-1}}\cdot(-\sin(v))\,dv}{\int_{\arccos(0)}^{\arccos(1)}\frac{\cos^\zeta(v)}{\sqrt{\cos^2(v)-1}}\cdot(-\sin(v))\,dv}\\ &= \frac{\int_{\pi/2}^{\arccos(f(\rho_{ij}))}\frac{\cos^\zeta(v)}{\sqrt{-1}\cdot\sqrt{1-\cos^2(v)}}\cdot(-\sin(v))\,dv}{\int_{\pi/2}^{0}\frac{\cos^\zeta(v)}{\sqrt{-1}\cdot\sqrt{1-\cos^2(v)}}\cdot(-\sin(v))\,dv}.\end{aligned}$$

Since $\sqrt{1-\cos^2(v)} = \sin(v)$,

$$\lambda(X_i, X_j; \zeta, \rho_{ij}) = \frac{-\frac{1}{i}\int_{\pi/2}^{\arccos(f(\rho_{ij}))}\cos^\zeta(v)\,dv}{-\frac{1}{i}\int_{\pi/2}^{0}\cos^\zeta(v)\,dv} = \frac{\int_{\arccos(f(\rho_{ij}))}^{\pi/2}\cos^\zeta(v)\,dv}{\int_0^{\pi/2}\cos^\zeta(v)\,dv}.$$

Now ρ_{ij} is substituted by $\sin\left(\tau_{ij}\cdot\frac{\pi}{2}\right)$ which leads to the lower limit

$$\arccos\left(f\left(\rho_{ij}\right)\right)=\arccos\left(f\left(\sin\left(\tau_{ij}\cdot\frac{\pi}{2}\right)\right)\right)=\arccos\left(\sqrt{\frac{1+\sin\left(\tau_{ij}\cdot\frac{\pi}{2}\right)}{2}}\right).$$

We consider the addition formula

$$\sin\left(\varphi\pm\theta\right)=\sin\left(\varphi\right)\cos\left(\theta\right)\pm\cos\left(\varphi\right)\sin\left(\theta\right). \tag{B.2.1}$$

Since $\sin\left(\tau\cdot\frac{\pi}{2}\right)=\sin\left(\frac{\pi}{2}-\left(1-\tau\right)\frac{\pi}{2}\right)$ it follows from (B.2.1) that

$$\sin\left(\tau\cdot\frac{\pi}{2}\right)=\sin\left(\frac{\pi}{2}\right)\cos\left(\left(1-\tau\right)\frac{\pi}{2}\right)-\cos\left(\frac{\pi}{2}\right)\sin\left(\left(1-\tau\right)\frac{\pi}{2}\right)=\cos\left(\left(1-\tau\right)\frac{\pi}{2}\right).$$

Now we are looking for a function h that fulfills

$$h\left(\tau\right)=\arccos\left(\sqrt{\frac{1+\sin\left(\tau\cdot\frac{\pi}{2}\right)}{2}}\right)=\arccos\left(\sqrt{\frac{1+\cos\left(\left(1-\tau\right)\frac{\pi}{2}\right)}{2}}\right),$$

but without trigonometric components. That is to say

$$\begin{aligned}\cos\left(\left(1-\tau\right)\frac{\pi}{2}\right)&=2\cos^2\left(h\left(\tau\right)\right)-1=\cos^2\left(h\left(\tau\right)\right)-\left(1-\cos^2\left(h\left(\tau\right)\right)\right)\\&=\cos^2\left(h\left(\tau\right)\right)-\sin^2\left(h\left(\tau\right)\right),\qquad\forall\,\tau\in\left[-1,1\right].\end{aligned}$$

For $\varphi=\theta$ formula (B.1.3) becomes $\cos\left(2\varphi\right)=\cos^2\left(\varphi\right)-\sin^2\left(\varphi\right)$, and thus

$$\cos\left(\left(1-\tau\right)\frac{\pi}{2}\right)=\cos\left(2\cdot h\left(\tau\right)\right).$$

Hence

$$h:\ \tau\longmapsto\frac{\pi}{2}\left(\frac{1-\tau}{2}\right),$$

and thus

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

Appendix C

Estimating the Tail-Dependence Coefficient

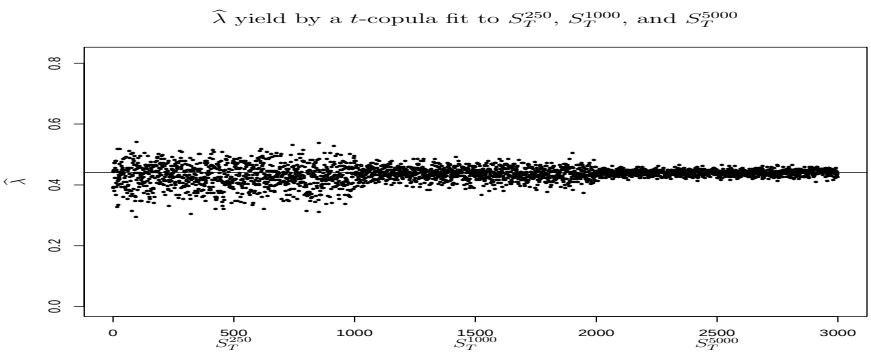

Figure C.1: Estimated λ from S_T^{250}, S_T^{1000}, and S_T^{5000} utilizing a t-copula and empirical marginals, cf. Section 7.3.1. (The black line indicates the true value.)

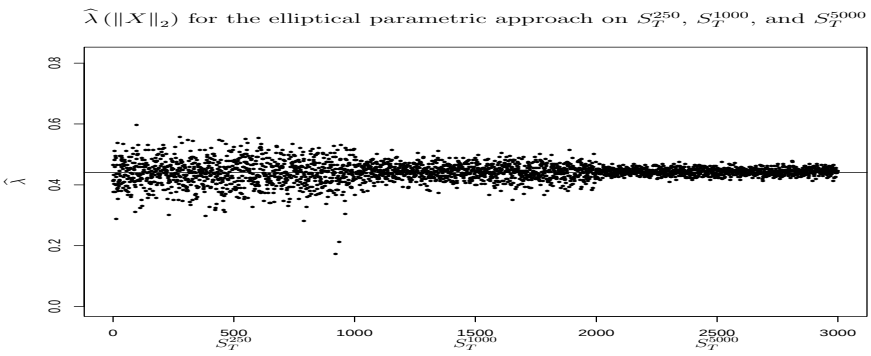

Figure C.2: Estimated λ from S_T^{250}, S_T^{1000}, and S_T^{5000}, where the tail index is estimated via $\|X\|_2$ and the correlation is estimated via Kendall's τ. (The black line indicates the true value.)

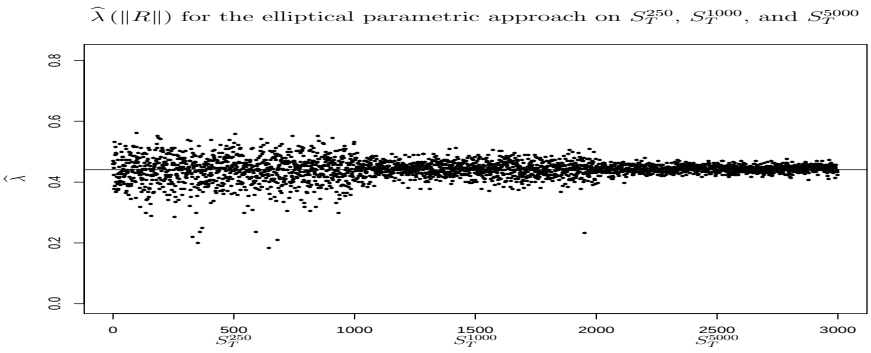

Figure C.3: Estimated λ from S_T^{250}, S_T^{1000}, and S_T^{5000} with the elliptical parametric approach of Section 7.3.1, where the tail index is estimated via $\mathcal{R}_d^*$. (The black line indicates the true value.)

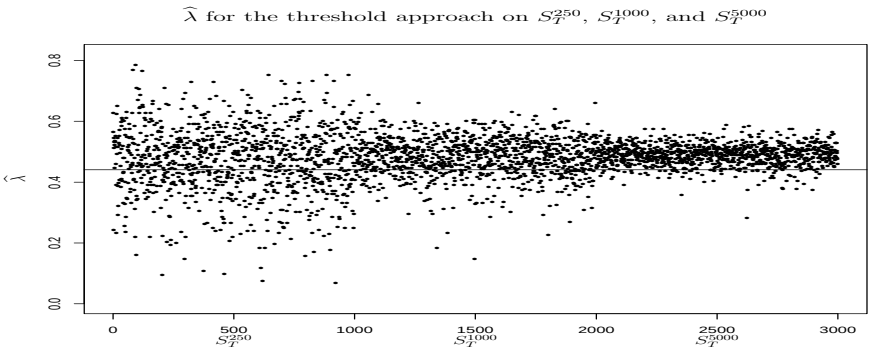

Figure C.4: Estimated λ from S_T^{250}, S_T^{1000}, and S_T^{5000} with the threshold approach, cf. Section 7.3.2. (The black line indicates the true value.)

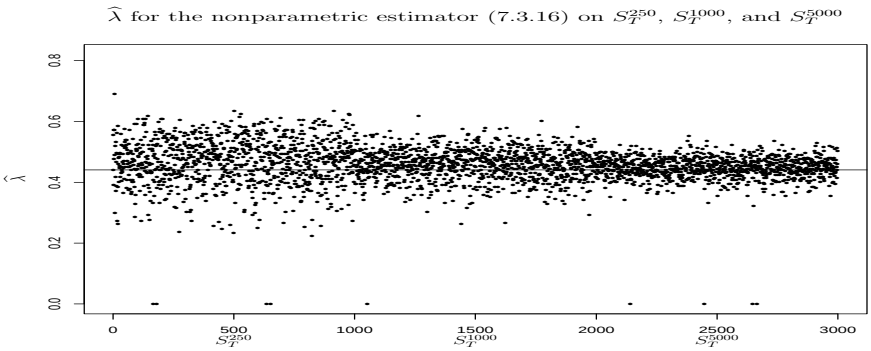

Figure C.5: Estimated λ from S_T^{250}, S_T^{1000}, and S_T^{5000} with the nonparametric estimator (7.3.16). (The black line indicates the true value.)

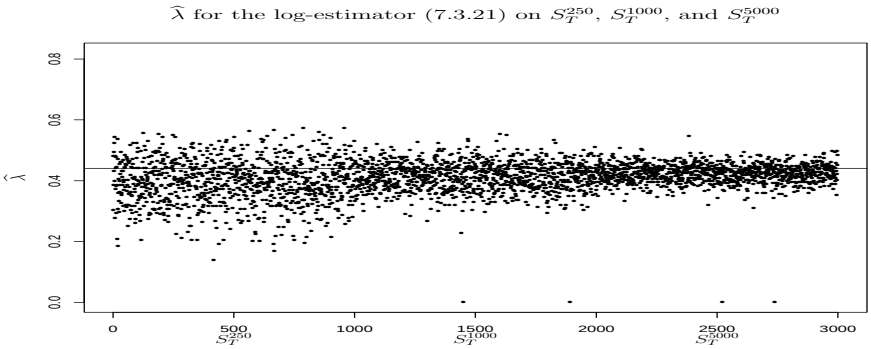

Figure C.6: Estimated λ from S_T^{250}, S_T^{1000}, and S_T^{5000} with the log-estimator (7.3.21). (The black line indicates the true value.)

Appendix D

Copula estimations for risk factor portfolios

On the next pages, we summarize the estimation and test results for the bivariate and 4-dimensional portfolios of our risk factors. For all estimates we state the standard error and the corresponding t-value. For the TDC estimates we give instead of a t-value (since asymptotic normality can not be guaranteed) the p_L and p_U probabilities as described in Subsection 8.1.3. In the test statistic tables the best values under the trainings and testset over the different models are marked bold for each portfolio. Under the χ^2 column we stated the p-value of the test in percentage. The PD and PDS values are also given in percentage.
Table D.14 present the bivariate portfolios under which the models C_Ω and C_C have the maximal and minimal value of p_{asy} (as defined in Subsection 8.1.3). Here we used the abbreviations Nasdaq (NAS), S&P 500 (SP), Nikkei (NIK), kerosine (ker), gasoline (gas), diesel (die) and heating oil (hea). Analogously Table D.15 states the p_{asy} and $p_{asy.real}$ values for the bivariate projections of the 4-dimensional portfolios.

NAS-SP						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.8043	2.1597	1.7065	1.9135	0.4999	0.1103
s.e.	0.1290	0.3067	1.1630	0.5155	0.0948	0.0777
t-value	6.2300	3.7800	1.4700	1.7700	0.9999	0.9058
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.6057	2.8137		2.6986	0.4734	0.3050
s.e.	0.0532	0.2531		0.3267	0.0429	0.0422
t-value	11.3900	11.1200		8.2600	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		7.2720	0.8156		0.3853	
s.e.		2.0427	0.0118		0.0642	
t-value		3.5600	68.9500		1.0000	

SP-NIK						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.1363	1.3313	0.8760	1.0247	0.0432	0.0286
s.e.	0.3656	0.9602	0.4735	0.0461	0.1444	0.0531
t-value	0.3700	0.3400	1.8500	0.5300	0.2839	0.6953
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.3365	0.6101		0.1875	0.1081	0.0165
s.e.	0.1879	0.3926		0.0802	0.0994	0.0264
t-value	1.7900	1.5500		2.3400	0.8961	0.9808
t-Copula		df	ρ		λ	
est.		24.7419	0.2191		0.0004	
s.e.		21.8524	0.0328		0.0023	
t-value		1.1300	6.6900		1.0000	

NAS-NIK						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.2855	1.1348	0.8841	1.0223	0.0451	0.0214
s.e.	1.2302	0.6440	0.4905	0.0738	0.2665	0.0782
t-value	0.2300	0.2100	1.8000	0.3000	0.1778	0.3653
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.3429	0.5239		0.2046	0.0913	0.0222
s.e.	0.1967	0.3361		0.0897	0.0936	0.0336
t-value	1.7400	1.5600		2.2800	0.8920	0.9790
t-Copula		df	ρ		λ	
est.		47.4151	0.2190		0	
s.e.		81.5157	0.0322		0	
t-value		0.5800	6.7900		0.8132	

SP-Hang						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.8987	1.2243	1.0448	3.3670	0.2143	0.0781
s.e.	0.0695	0.0879	0.6268	1.5879	0.0664	0.0550
t-value	12.9200	2.5500	1.6700	1.4900	0.9947	0.8600
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.6134	0.8722		0.9806	0.2771	0.1907
s.e.	0.0844	0.1576		0.2400	0.0551	0.0531
t-value	7.2700	5.5300		4.0900	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		8.5274	0.4736		0.0963	
s.e.		3.2116	0.0288		0.0599	
t-value		2.6600	16.4300		1.0000	

NAS-Hang						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.7105	1.3732	0.2267	1.2710	0.2440	0.0795
s.e.	0.2236	0.1790	0.7582	0.3123	0.1091	0.0909
t-value	3.1800	2.0900	0.3000	0.8700	0.9808	0.7109
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.5882	0.8135		0.5708	0.2509	0.1223
s.e.	0.1113	0.1741		0.1769	0.0659	0.0566
t-value	5.2900	4.6700		3.2300	1.0000	0.9989
t-Copula		df	ρ		λ	
est.		13.7396	0.4215		0.0273	
s.e.		8.5623	0.0307		0.0465	
t-value		1.6000	13.7300		1.0000	

Hang-NIK						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	1	1.0560	0.9299	–	0.0721	0
s.e.	–	0.3848	0.5122	–	–	-
t-value	–	0.1500	1.8200	–	–	-
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.7780	0.2765		0.4539	0.0634	0.0482
s.e.	0.1942	0.0925		0.4586	0.0555	0.0855
t-value	4.0100	2.9900		0.9900	0.9971	0.7005
t-Copula		df	ρ		λ	
est.		23.1097	0.2253		0.0007	
s.e.		20.3950	0.0329		0.0036	
t-value		1.1300	6.8500		1.0000	

Table D.1: Estimation results for the bivariate stock index portfolios.

NAS-SP								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**99.33**	-1043.53	-1024.12	-53.18	**0.0628**	**0.0774**	**-1.72**	**-7.78**
C_C.est	57.18	-981.03	-966.48	-47.98	0.2290	0.3398	-10.11	-17.67
t-Copula.est	89.44	-1040.32	-1030.61	**-55.45**	0.0936	0.1165	-6.43	20.94
Ω.test	**0.00**	-631.00	-611.59	-53.18	1.4313	1.3671	14.49	**-3.66**
C_C.test	**0.00**	-611.03	-596.47	-47.98	**0.7682**	**0.7228**	**6.78**	-8.29
t-Copula.test	**0.00**	-626.88	-617.18	**-55.45**	1.2576	1.4086	9.76	85.87

SP-NIK								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	80.96	-36.50	-17.33	-2.17	**0.2058**	0.2144	-13.84	**10.02**
C_C.est	**95.67**	-37.09	-22.71	-2.46	0.2213	**0.2019**	**-5.37**	12.64
t-Copula.est	93.29	**-40.56**	**-30.98**	**-2.60**	0.2727	0.2415	-11.82	39.76
Ω.test	49.10	-3.88	15.33	-2.17	0.6695	0.6426	-13.38	**25.01**
C_C.test	**62.81**	-6.75	7.66	-2.46	**0.4629**	0.4430	**-3.77**	28.62
t-Copula.test	61.33	**-11.47**	**-1.86**	**-2.60**	0.6473	**0.4261**	-7.76	28.60

NAS-NIK								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	54.95	-34.04	-14.87	-2.69	0.2352	0.3697	-15.18	**3.76**
C_C.est	**69.58**	-36.13	-21.75	**-2.79**	0.2605	**0.3414**	**-4.93**	10.58
t-Copula.est	54.30	**-39.89**	**-30.30**	-2.23	**0.2227**	0.3512	-8.87	10.88
Ω.test	36.18	2.17	21.38	-2.69	0.6704	0.6868	**18.52**	14.34
C_C.test	**59.35**	0.79	15.19	**-2.79**	**0.5111**	**0.5425**	23.49	14.08
t-Copula.test	52.64	**-1.90**	**7.70**	-2.23	0.6120	0.6522	19.81	**13.60**

SP-Hang								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	91.74	-235.40	-216.20	-13.93	**0.0985**	0.1250	**-0.6**	**-6.81**
C_C.est	64.50	-233.40	-219.00	**-14.06**	0.1083	**0.1098**	3.64	12.65
t-Copula.est	**96.83**	-230.08	-220.48	-13.87	0.2319	0.2021	-9.16	18.38
Ω.test	10.20	-108.33	-89.09	-13.93	1.0314	1.0975	16.3	**14.13**
C_C.test	**29.74**	-114.78	-100.35	**-14.06**	0.8554	**0.9976**	17.31	26.25
t-Copula.test	20.38	**-122.18**	**-112.56**	-13.87	**0.6742**	1.5568	**12.52**	48.11

NAS-Hang								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**92.50**	-176.16	-156.96	**-10.64**	**0.0952**	**0.1146**	**-5.78**	**3.06**
C_C.est	88.83	-175.19	-160.79	-9.57	0.1236	0.1301	-12.67	3.15
t-Copula.est	31.76	**-177.78**	**-168.18**	-10.07	0.2271	0.1590	-32.01	22.84
Ω.test	67.55	-64.71	-45.46	**-10.64**	0.8526	0.7871	40.04	7.57
C_C.test	69.93	-67.59	-53.16	-9.57	**0.6457**	**0.5665**	38.22	**7.26**
t-Copula.test	**86.28**	**-74.41**	**-60.78**	-10.07	0.8115	0.9861	**14.39**	33.17

Hang-NIK								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	99.14	-38.00	-18.92	-2.88	**0.1249**	0.1893	-16.93	12.63
C_C.est	**99.98**	-39.54	-25.23	-2.61	0.1679	**0.1804**	**-2.23**	**-5.98**
t-Copula.est	99.41	**-41.75**	**-32.21**	**-2.90**	0.1889	0.3075	-19.09	25.02
Ω.test	94.93	2.40	21.53	-2.88	0.7232	0.8222	-2.72	**37.69**
C_C.test	**97.17**	-5.49	8.86	-2.61	**0.5115**	**0.5384**	6.66	38.01
t-Copula.test	85.21	-4.35	5.22	**-2.90**	0.6706	0.7542	**-2.50**	45.30

Table D.2: Test results for the bivariate stock index portfolios.

AUD-HKD						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.3225	1.4604	0.7120	1.2750	0.1266	0.1882
s.e.	0.1632	0.2586	0.6884	0.1233	0.0775	0.0763
t-value	1.9800	1.7800	1.0300	2.2300	0.9386	0.9871
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.4361	0.8493		0.8415	0.1928	0.2475
s.e.	0.0679	0.1663		0.1281	0.0430	0.0430
t-value	6.4300	5.1100		6.5700	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		7.8943	0.4605		0.1037	
s.e.		2.3388	0.0245		0.0488	
t-value		3.3800	18.8300		1.0000	

AUD-SGD						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.6193	1.2771	0.3201	1.7850	0.1729	0.2000
s.e.	0.1069	0.1236	0.7350	0.2224	0.0634	0.0624
t-value	5.7900	2.2400	0.4400	3.5300	0.9876	0.9996
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.5746	0.7533		1.2847	0.2290	0.2480
s.e.	0.0600	0.1082		0.2113	0.0386	0.0413
t-value	9.5700	6.9600		6.0800	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		8.5732	0.4864		0.1003	
s.e.		2.5016	0.0239		0.0475	
t-value		3.4300	20.3800		1.0000	

AUD-YEN						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.3304	1.4892	0.2999	1.1932	0.1346	0.1422
s.e.	0.1640	0.2483	0.6226	0.0893	0.0783	0.0626
t-value	2.0100	1.9700	0.4800	2.1600	0.9536	0.9847
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.6181	0.3777		1.1193	0.0987	0.2056
s.e.	0.0817	0.0762		0.2513	0.0388	0.0524
t-value	7.5700	4.9600		4.4500	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		6.1886	0.3659		0.1094	
s.e.		1.3377	0.0269		0.0364	
t-value		4.6300	13.5900		1.0000	

HKD-SGD						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.4946	4.0933	2.2987	1.5392	0.4033	0.2179
s.e.	0.0578	0.4409	0.7291	0.1439	0.0483	0.0417
t-value	8.5600	7.0200	3.1500	3.7500	1.0000	0.9999
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.4232	4.1743		2.9919	0.3584	0.4575
s.e.	0.0369	0.3035		0.1682	0.0315	0.0299
t-value	11.4700	13.7500		17.7900	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		2.4037	0.8452		0.6261	
s.e.		0.3295	0.0108		0.0229	
t-value		7.2900	77.9100		1.0000	

HKD-YEN						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.3835	2.3951	0.0417	1.2059	0.2548	0.1377
s.e.	0.0769	0.2519	0.6972	0.0881	0.0534	0.0491
t-value	4.9800	5.5400	0.0600	2.3400	1.0000	0.9902
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.3637	2.1206		0.5503	0.2623	0.1806
s.e.	0.0543	0.3798		0.0750	0.0420	0.0346
t-value	6.7000	5.5800		7.3400	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		4.1532	0.4809		0.2351	
s.e.		0.6814	0.0248		0.0363	
t-value		6.1000	19.3800		1.0000	

SGD-YEN						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.4138	2.0716	2.0376	1.1968	0.2494	0.1263
s.e.	0.1182	0.2358	0.6866	0.0965	0.0746	0.0551
t-value	3.5000	4.5400	2.9700	2.0400	0.9998	0.9794
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.5033	1.6568		1.4872	0.3312	0.3117
s.e.	0.0453	0.1580		0.1437	0.0326	0.0317
t-value	11.1200	10.4900		10.3500	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		4.9698	0.6430		0.2984	
s.e.		0.9441	0.0181		0.0435	
t-value		5.2600	35.4500		1.0000	

Table D.3: Estimation results for the bivariate fx portfolios.

AUD-HKD								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**99.08**	-315.18	-294.57	-12.45	**0.0938**	**0.1073**	**20.87**	**-0.96**
C_C.est	95.89	-315.64	-300.18	-12.27	0.1299	0.1299	23.79	6.35
t-Copula.est	98.85	**-317.56**	**-307.26**	**-12.61**	0.1276	0.1272	27.16	-2.19
Ω.test	84.09	-210.48	-189.83	-12.45	0.9965	0.7356	**8.03**	-11.90
C_C.test	**87.28**	-213.10	-197.61	-12.27	**0.7316**	**0.5392**	21.27	**-4.76**
t-Copula.test	68.77	-208.26	-197.93	**-12.61**	1.1529	0.6739	37.84	-13.18

AUD-SGD								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	96.82	-358.81	-338.20	**-14.11**	0.1024	**0.1191**	8.60	**2.00**
C_C.est	**99.42**	**-361.02**	**-345.56**	-13.81	**0.0982**	0.1233	5.19	5.77
t-Copula.est	95.76	-353.26	-342.95	-13.76	0.1020	0.1312	**2.30**	5.96
Ω.test	0.02	-116.62	-95.97	**-14.11**	2.8173	2.0736	56.41	-31.05
C_C.test	0.01	-106.49	-91.01	-13.81	2.7395	**2.0297**	55.55	-16.51
t-Copula.test	**0.03**	**-126.36**	**-116.03**	-13.76	**2.5660**	2.1426	**49.64**	**-14.82**

AUD-YEN								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	37.67	-193.30	-172.69	**-8.01**	0.3481	**0.2945**	**-3.66**	-7.79
C_C.est	34.43	-189.50	-174.04	-7.52	0.3693	0.3222	-7.02	**-6.07**
t-Copula.est	**41.16**	**-206.75**	**-196.44**	-8.00	**0.3305**	0.3202	4.14	-6.58
Ω.test	0.00	11.86	32.51	**-8.01**	3.1345	3.2811	24.75	90.07
C_C.test	**0.02**	18.04	33.52	-7.52	**2.7664**	2.8488	**22.59**	102.69
t-Copula.test	0.01	**-7.55**	**2.77**	-8.00	3.0464	**2.8282**	34.63	**86.42**

HKD-SGD								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	50.85	-1648.75	-1628.13	-65.85	**0.1818**	**0.1541**	**-3.62**	-7.60
C_C.est	1.47	-1582.27	-1566.81	-60.21	0.3483	0.2510	-17.89	-9.23
t-Copula.est	**57.29**	**-1656.66**	**-1646.35**	**-70.17**	0.2212	0.2532	11.66	**5.31**
Ω.test	**0.00**	-2201.76	-2181.11	-65.85	**0.9046**	**0.7741**	-20.65	-27.62
C_C.test	**0.00**	-2039.12	-2023.64	-60.21	1.9574	1.4018	-36.68	-28.71
t-Copula.test	**0.00**	-2193.42	-2183.10	**-70.17**	0.9138	0.8061	**-3.65**	**-9.28**

HKD-YEN								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**86.89**	-378.17	-357.55	-14.28	**0.1703**	**0.1107**	-10.94	8.80
C_C.est	72.70	-365.79	-350.33	-14.52	0.1932	0.1354	-13.16	**-8.37**
t-Copula.est	54.43	-375.15	-364.85	**-15.97**	0.2458	0.1550	**-4.07**	8.83
Ω.test	**0.00**	159.54	180.19	-14.28	8.8686	9.0277	207.90	248.87
C_C.test	**0.00**	173.91	189.39	-14.52	**8.4361**	8.6109	206.07	**232.60**
t-Copula.test	**0.00**	**147.75**	**158.08**	**-15.97**	9.3842	**8.3506**	**205.85**	370.39

SGD-YEN								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**81.72**	-697.53	-676.91	**-28.08**	**0.1096**	**0.1146**	**4.40**	**-8.13**
C_C.est	80.37	-681.78	-666.32	-26.44	0.1913	0.1541	11.40	-8.52
t-Copula.est	73.39	-695.21	-684.90	-27.32	0.2348	0.2444	13.64	13.49
Ω.test	**0.00**	330.72	351.37	**-28.08**	16.9586	16.3207	**157.85**	**287.12**
C_C.test	**0.00**	346.40	361.89	-26.44	**15.0874**	**14.3958**	187.09	394.03
t-Copula.test	**0.00**	**314.21**	**324.53**	-27.32	15.8523	14.6341	226.01	430.60

Table D.4: Test results for the bivariate fx portfolios.

Al-Cu						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.6687	1.4268	2.4111	1.3284	0.2505	0.1043
s.e.	0.1629	0.1678	0.6029	0.2268	0.0871	0.0715
t-value	4.1000	2.5400	4.0000	1.4500	0.9945	0.9051
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.5125	1.6777		1.5205	0.3390	0.3090
s.e.	0.0426	0.1593		0.1434	0.0312	0.0301
t-value	12.0300	10.5300		10.6000	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		12.0079	0.6542		0.1231	
s.e.		3.4060	0.0137		0.0516	
t-value		3.5300	47.7800		1.0000	

Al-Pb						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.9277	1.1158	1.6738	1.2980	0.1288	0.0213
s.e.	0.2772	0.0646	0.4073	1.3249	0.0731	0.1057
t-value	3.3500	1.7900	4.1100	0.2200	0.9632	0.2097
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.6111	0.6797		0.7336	0.2204	0.1512
s.e.	0.0725	0.1028		0.1690	0.0429	0.0433
t-value	8.4300	6.6100		4.3400	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		17.2522	0.4043		0.0122	
s.e.		7.1999	0.0202		0.0176	
t-value		2.4000	20.0300		1.0000	

Al-Zn						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.4708	1.2222	2.4401	1.1302	0.1115	0.0813
s.e.	0.4050	0.2390	0.4548	0.1438	0.1329	0.0984
t-value	1.1600	0.9300	5.3700	0.9100	0.7012	0.7217
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.4629	1.2840		0.8405	0.2698	0.2355
s.e.	0.0538	0.1652		0.0959	0.0365	0.0324
t-value	8.6000	7.7700		8.7600	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		23.8716	0.5286		0.0105	
s.e.		11.7977	0.0167		0.0187	
t-value		2.0200	31.6500		1.0000	

Cu-Pb						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.5989	1.1794	2.0042	1.0000	0.1198	0.0000
s.e.	0.5340	0.1883	0.4169	0.0807	0.1471	0.0449
t-value	1.1200	0.9500	4.8100	0.0000	0.6987	0.4053
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.5816	0.7463		0.6935	0.2297	0.1540
s.e.	0.0694	0.1109		0.1396	0.0419	0.0402
t-value	8.3800	6.7300		4.9700	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		16.0341	0.4148		0.0167	
s.e.		6.8586	0.0199		0.0226	
t-value		2.3400	20.8800		1.0000	

Cu-Zn						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.4081	1.1829	2.7461	1.1536	0.0830	0.1044
s.e.	0.5518	0.3154	0.4445	0.1738	0.1603	0.1379
t-value	0.7400	0.5800	6.1800	0.8800	0.4895	0.6696
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.4805	1.1281		1.1418	0.2600	0.2831
s.e.	0.0524	0.1386		0.1214	0.0345	0.0339
t-value	9.1600	8.1400		9.4100	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		27.3156	0.5565		0.0083	
s.e.		8.5599	0.0160		0.0100	
t-value		3.1900	34.8600		1.0000	

Pb-Zn						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.7362	1.1443	2.3837	1.2470	0.1232	0.0677
s.e.	0.3476	0.1046	0.4378	0.4029	0.0947	0.1216
t-value	2.1200	1.3800	5.4500	0.6100	0.8990	0.5074
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.5498	0.8963		1.1306	0.2537	0.2439
s.e.	0.0534	0.1019		0.1517	0.0332	0.0352
t-value	10.3000	8.8000		7.4500	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		24.0897	0.5209		0.0094	
s.e.		13.3425	0.0171		0.0194	
t-value		1.8100	30.4200		1.0000	

Table D.5: Estimation results for the bivariate metal portfolios.

Al-Cu								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**93.65**	**-989.08**	**-967.17**	**-28.34**	**0.1049**	**0.1208**	-7.38	**-4.15**
C_C.est	18.79	-955.69	-939.26	-27.74	0.2011	0.2797	0.29	11.67
t-Copula.est	87.48	-977.82	-966.87	-27.30	0.1583	0.2493	**0.19**	12.75
Ω.test	**0.00**	-191.95	-170.66	**-28.34**	5.5882	4.4046	**79.64**	**53.93**
C_C.test	**0.00**	**-213.66**	**-197.70**	-27.74	**4.2047**	**3.3834**	93.73	59.51
t-Copula.test	**0.00**	-186.50	-175.86	-27.30	5.0847	3.6564	89.24	60.33

Al-Pb								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**93.76**	-315.76	-293.86	-8.40	**0.1470**	**0.1835**	7.19	-64.94
C_C.est	62.82	-306.50	-290.07	-8.63	0.2691	0.2931	28.09	**-2.52**
t-Copula.est	85.67	-308.33	-297.38	**-9.46**	0.2622	0.4160	**3.66**	4.09
Ω.test	95.19	-177.22	-155.93	-8.40	0.7587	0.5237	**8.73**	**-9.69**
C_C.test	91.40	-169.17	-153.20	-8.63	0.5661	**0.3161**	24.11	25.66
t-Copula.test	**97.45**	**-187.75**	**-177.11**	**-9.46**	**0.4379**	0.5437	8.84	32.46

Al-Zn								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**89.20**	-573.02	-551.12	**-17.04**	**0.1258**	**0.1195**	**20.93**	-8.9
C_C.est	28.48	-539.97	-523.54	-14.77	0.4408	0.3916	38.96	6.90
t-Copula.est	71.91	-570.14	-559.19	-16.55	0.2481	0.3161	35.69	**6.39**
Ω.test	0.47	**-196.97**	**-175.68**	**-17.04**	2.5721	1.8695	**12.11**	**33.8**
C_C.test	**23.92**	-186.13	-170.17	-14.77	**1.6595**	**1.2309**	35.39	101.38
t-Copula.test	0.41	-176.16	-165.52	-16.55	2.2213	1.4268	18.94	96.64

Cu-Pb								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**61.17**	**-340.82**	**-318.92**	-9.05	**0.0941**	**0.1132**	-2.22	-97.28
C_C.est	43.61	-318.77	-382.34	**-10.48**	0.4439	0.4699	17.33	43.90
t-Copula.est	52.98	-324.85	-313.90	-9.65	0.2819	0.5292	**1.12**	**12.16**
Ω.test	**35.26**	-208.05	-186.76	-9.05	0.6899	0.4714	**11.19**	-15.51
C_C.test	23.92	-189.96	-173.99	**-10.48**	**0.5443**	0.3837	69.68	**12.42**
t-Copula.test	10.48	-206.71	-196.07	-9.65	0.5535	**0.2402**	50.23	14.11

Cu-Zn								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	26.75	-647.17	-625.27	**-18.95**	**0.1334**	**0.1176**	**-6.16**	**-2.71**
C_C.est	0.45	-603.77	-587.35	-18.47	0.4782	0.4457	17.65	19.80
t-Copula.est	**30.50**	-644.58	-633.63	-18.01	0.3145	0.3503	15.18	-20.17
Ω.test	5.76	**-259.88**	**-238.59**	**-18.95**	1.8050	1.5993	**42.70**	**57.98**
C_C.test	**17.78**	-250.72	-234.76	-18.47	**1.1832**	**1.1866**	54.12	121.21
t-Copula.test	2.65	-220.17	-209.52	-18.01	1.8905	1.2066	65.26	104.91

Pb-Zn								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**93.98**	-558.38	-536.48	-16.44	**0.1446**	**0.1510**	-10.02	**3.95**
C_C.est	22.28	-527.77	-511.35	-16.19	0.5069	0.6442	14.39	22.99
t-Copula.est	70.84	-551.32	-540.37	**-16.95**	0.3082	0.4851	**2.14**	16.15
Ω.test	9.86	-317.77	-296.48	-16.44	0.9750	0.8186	-23.25	**9.64**
C_C.test	**40.56**	-315.54	-299.58	-16.19	**0.4409**	**0.4286**	**-7.93**	36.41
t-Copula.test	4.67	-304.55	-293.90	**-16.95**	0.8556	0.6247	-16.87	20.59

Table D.6: Test results for the bivariate metal portfolios.

kerosine-heating oil						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.7708	4.2218	6.3789	1.0000	0.6333	0.0000
s.e.	0.0406	0.6945	1.3868	0.1866	0.0414	0.0593
t-value	18.9900	4.6400	4.6000	0.0000	1.0000	0.4995
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.4672	3.6750		14.103	0.3869	0.5073
s.e.	0.0378	0.2138		1.186	0.0316	0.0361
t-value	12.3500	17.1900		11.890	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		1.5220	0.9449		0.8095	
s.e.		0.2046	0.0054		0.0129	
t-value		7.4400	176.0900		1.0000	

kerosine-gasoline						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.7109	1.9211	1.9210	1.0000	0.4020	0.0000
s.e.	0.1157	0.1776	0.7663	0.1071	0.0737	0.0429
t-value	6.1400	5.1900	2.5100	0.0000	1.0000	0.4995
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.6149	1.9342		1.2280	0.4297	0.2190
s.e.	0.0526	0.1789		0.1633	0.0394	0.0341
t-value	11.6900	10.8100		7.5200	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		4.1705	0.6573		0.3470	
s.e.		0.7146	0.0199		0.0396	
t-value		5.8400	33.0300		1.0000	

kerosine-diesel						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.7696	3.6289	5.0746	1.0000	0.6076	0.0000
s.e.	0.0338	0.4532	1.0740	0.1272	0.0347	0.0406
t-value	22.7600	5.8000	4.7200	0.0000	1.0000	0.4999
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.4633	3.5216		8.2662	0.3806	0.4935
s.e.	0.0382	0.2131		0.6918	0.0317	0.0353
t-value	12.1300	16.5200		11.9500	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		1.7959	0.9170		0.7524	
s.e.		0.2195	0.0069		0.0150	
t-value		8.1800	132.5100		1.0000	

heating oil-gasoline						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.7765	1.7234	2.2748	1.0000	0.3920	0.0000
s.e.	0.1423	0.1740	0.7530	0.1542	0.0859	0.0478
t-value	5.4600	4.1600	3.0200	0.0000	1.0000	0.4984
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.6495	1.8080		1.5211	0.4427	0.2222
s.e.	0.0550	0.1665		0.2207	0.0406	0.0378
t-value	11.8100	10.8600		6.8900	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		5.6490	0.6738		0.2943	
s.e.		1.3387	0.0188		0.0544	
t-value		4.2200	35.7800		1.0000	

heating oil-diesel						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.8544	4.607	5.1644	1.0000	0.7157	0.0000
s.e.	0.0242	0.727	1.3861	0.1503	0.0311	0.0303
t-value	35.2600	4.960	3.7300	0.0000	1.0000	0.5001
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.5079	5.9320		11.1621	0.4519	0.4625
s.e.	0.0379	0.3358		0.7329	0.0339	0.0357
t-value	13.4000	17.6600		15.2300	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		1.5086	0.9553		0.8290	
s.e.		0.2167	0.0047		0.0124	
t-value		6.9600	205.3300		1.0000	

gasoline-diesel						
Ω	α	δ_s	θ	δ	λ_L	λ_U
est.	0.6879	1.9538	2.3030	1.0000	0.3949	0.0000
s.e.	0.1137	0.2056	0.7759	0.1071	0.0748	0.0463
t-value	6.0500	4.6400	2.9700	0.0000	1.0000	0.5002
C_C	α	δ		δ_s	λ_L	λ_U
est.	0.6648	2.0740		1.4023	0.4759	0.2045
s.e.	0.0503	0.1911		0.2166	0.0389	0.0344
t-value	13.2000	10.8500		6.4800	1.0000	1.0000
t-Copula		df	ρ		λ	
est.		4.8127	0.6804		0.3348	
s.e.		1.1179	0.0192		0.0528	
t-value		4.3100	35.5000		1.0000	

Table D.7: Estimation results for the bivariate oil product portfolios.

kerosine-heating oil								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**4.33**	**-2048.52**	**-2029.44**	-119.77	**0.1106**	**0.1043**	**5.17**	**-13.64**
C_C.est	0.0	-1881.04	-1866.72	-106.94	0.2969	0.4028	-7.35	-17.80
t-Copula.est	0.01	-1996.15	-1986.61	**-123.13**	0.1205	0.1231	14.20	14.31
Ω.test	**0.0**	**-1257.74**	**-1238.65**	-119.77	0.9637	0.9425	33.99	17.34
C_C.test	**0.0**	-1114.91	-1100.59	-106.94	**0.5379**	**0.5139**	**3.36**	**-1.37**
t-Copula.test	**0.0**	-1239.67	-1230.12	**-123.13**	0.8570	0.8206	43.17	45.35

kerosine-gasoline								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**77.04**	**-509.35**	**-490.26**	**-31.66**	**0.1199**	**0.1317**	-9.51	**3.59**
C_C.est	44.87	-473.21	-458.89	-27.18	0.2083	0.2404	**-8.47**	-3.60
t-Copula.est	66.99	-487.11	-477.57	-31.64	0.2501	0.4147	-11.13	30.17
Ω.test	**92.18**	-464.63	-445.54	**-31.66**	**0.0798**	**0.1153**	61.05	6.99
C_C.test	40.12	-450.37	-436.05	-27.18	0.1227	0.2050	68.78	**-0.81**
t-Copula.test	89.19	-463.60	-454.06	-31.64	0.1254	0.2635	**52.66**	79.00

kerosine-diesel								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**56.65**	**-1671.71**	**-1652.63**	-97.36	**0.0863**	**0.0993**	**5.10**	-11.63
C_C.est	0.00	-1525.50	-1511.18	-87.23	0.2149	0.2494	-15.20	-16.38
t-Copula.est	4.98	-1626.04	-1615.50	**-100.71**	0.1588	0.1485	19.51	**5.41**
Ω.test	0.17	**-1291.33**	**-1272.24**	-97.36	0.4047	0.3996	19.34	8.43
C_C.test	**4.70**	-1217.00	-1202.69	-87.23	**0.2054**	**0.2130**	**-8.02**	**7.36**
t-Copula.test	0.00	-1271.86	-1262.32	**-100.71**	0.4062	0.3791	36.89	33.92

heating oil-gasoline								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**99.94**	**-540.18**	**-521.10**	**-33.09**	**0.0620**	**0.0913**	**-17.77**	-6.01
C_C.est	83.04	-514.58	-500.26	-30.59	0.1596	0.2115	-37.10	**-5.83**
t-Copula.est	83.86	-516.83	-507.28	-30.34	0.2409	0.3780	-18.95	27.27
Ω.test	**92.69**	-552.20	-533.11	**-33.09**	**0.0873**	**0.1246**	16.80	11.08
C_C.test	15.99	-539.81	-525.49	-30.59	0.2007	0.3324	21.55	**10.09**
t-Copula.test	92.32	**-558.55**	**-549.01**	-30.34	0.1610	0.2641	**9.83**	40.98

heating oil-diesel								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	56.04	**-2232.84**	**-2213.75**	-131.27	**0.0709**	**0.0682**	**-5.84**	**-9.20**
C_C.est	0.00	-2054.98	-2040.66	-113.36	0.2413	0.2866	-26.17	-16.70
t-Copula.est	**66.99**	-2193.92	-2184.38	**-133.71**	0.1059	0.1147	6.42	19.02
Ω.test	3.02	-1955.50	-1936.41	-131.27	0.1944	0.1987	**-0.27**	**8.25**
C_C.test	0.66	-1874.18	-1859.86	-113.36	0.3279	0.4175	-17.78	-11.05
t-Copula.test	**25.17**	**-1968.03**	**-1958.48**	**-133.71**	**0.1565**	**0.1600**	7.23	35.38

gasoline-diesel								
	χ^2	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	60.87	**-564.29**	**-545.21**	-32.66	**0.1037**	0.1538	-13.56	7.72
C_C.est	**82.62**	-541.22	-526.90	**-33.84**	0.1180	**0.1512**	**-3.50**	**-6.46**
t-Copula.est	14.71	-528.81	-519.27	-32.81	0.3059	0.5327	-14.14	34.27
Ω.test	55.88	-530.14	-511.04	-32.66	**0.0996**	**0.1209**	21.17	3.39
C_C.test	3.21	-509.60	-495.28	**-33.84**	0.1558	0.2679	38.25	**1.13**
t-Copula.test	**91.81**	**-546.08**	**-536.54**	-32.81	0.1509	0.2275	**19.66**	49.04

Table D.8: Test results for the bivariate oil product portfolios.

stock portfolio $C_{\Omega}^{[4]}$										
parameter estimates	α	δ_{1_s}	θ_1	δ_1	δ_{2_s}	θ_2	δ_2	δ_{3_s}	θ_3	δ_3
est.	0.6396	2.3447	1.4574	1.9540	1.1125	0.6203	1.2270	1.1125	0.6203	1.2269
s.e.	0.1107	0.3614	1.3857	0.4357	0.0597	0.5504	0.1837	0.0541	0.3657	0.1154
t-value	5.78	3.72	1.05	2.19	1.88	1.13	1.24	2.08	1.70	1.97
model check		$\delta_{2_s} - \delta_{1_s}$	$\theta_2 - \theta_1$	$\delta_2 - \delta_1$				$\delta_{3_s} - \delta_{2_s}$	$\theta_3 - \theta_2$	$\delta_3 - \delta_2$
est.		-1.2322	-0.8371	-0.7270				0.0000	0.0000	0.0000
s.e.		0.3663	1.4910	0.4729				0.0806	0.6608	0.2169
t-value		-3.36	-0.56	-1.54				0.00	0.00	0.00
λ estimates		λ_{L_1}		λ_{U_1}	λ_{L_2}		λ_{U_2}	λ_{L_3}		λ_{U_3}
est.		0.4196		0.2070	0.0866		0.0867	0.0866		0.0867
s.e.		0.0825		0.0754	0.0426		0.0599	0.0391		0.0429
$p_{L,U}$		0.9999		0.9938	0.9658		0.9628	0.9812		0.9748

stock portfolio $C_{C}^{[4]}$										
parameter estimates	α	δ_1		δ_{1_s}	δ_2		δ_{2_s}	δ_3		δ_{3_s}
est.	0.5926	2.7766		2.3740	0.4796		0.3473	0.4796		0.3473
s.e.	0.0545	0.3300		0.3411	0.1276		0.1539	0.0686		0.0858
t-value	10.87	8.41		6.96	3.76		2.26	6.99		4.05
model check		$\delta_2 - \delta_1$		$\delta_{2_s} - \delta_{1_s}$				$\delta_3 - \delta_2$		$\delta_{3_s} - \delta_{2_s}$
est.		-2.2970		-2.0268				0.0000		0.0000
s.e.		0.3538		0.3742				0.1449		0.1761
t-value		-6.49		-5.42				0.00		0.00
λ estimates		λ_{L_1}		λ_{U_1}	λ_{L_2}		λ_{U_2}	λ_{L_3}		λ_{U_3}
est.		0.4617		0.3042	0.1397		0.0554	0.1397		0.0553
s.e.		0.0446		0.0427	0.0552		0.0495	0.0316		0.0283
$p_{L,U}$		1.0000		1.0000	1.0000		0.9991	1.0000		0.9999

stock portfolio C_t										
parameter estimates		df	$\rho_{(1,2)}$	$\rho_{(1,3)}$	$\rho_{(1,4)}$	$\rho_{(2,3)}$	$\rho_{(2,4)}$	$\rho_{(3,4)}$		
est.		12.0597	0.8196	0.2100	0.4144	0.2054	0.4689	0.2259		
s.e.		2.4495	0.0113	0.0360	0.0332	0.0346	0.0288	0.0351		
t-value		4.92	72.35	5.83	12.47	5.93	16.27	6.44		
λ estimates			$\lambda_{(1,2)}$	$\lambda_{(1,3)}$	$\lambda_{(1,4)}$	$\lambda_{(2,3)}$	$\lambda_{(2,4)}$	$\lambda_{(3,4)}$		
est.			0.2756	0.0119	0.0368	0.0116	0.0488	0.0130		
s.e.			0.0493	0.0086	0.0194	0.0084	0.0231	0.0092		
$p_{L,U}$			1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		

Table D.9: Estimation results for 4-dimensional stock indices portfolios.

fx-rate portfolio $C_{\Omega}^{[4]}$										
parameter estimates	α	δ_{1_s}	θ_1	δ_1	δ_{2_s}	θ_2	δ_2	δ_{3_s}	θ_3	δ_3
est.	0.3270	1.0334	3.5593	3.0581	1.0301	1.4486	1.7064	1.0300	1.3869	1.2745
s.e.	0.0392	0.1396	0.9350	0.2830	0.1004	0.5854	0.1079	0.0612	0.3643	0.0582
t-value	8.34	0.24	3.81	7.27	0.30	2.47	6.55	0.49	3.81	4.72
model check		$\delta_{2_s} - \delta_{1_s}$	$\theta_2 - \theta_1$	$\delta_2 - \delta_1$				$\delta_{3_s} - \delta_{2_s}$	$\theta_3 - \theta_2$	$\delta_3 - \delta_2$
est.		-0.0033	-2.1106	-1.3517				0.0000	-0.0617	-0.4320
s.e.		0.1719	1.1032	0.3028				0.1176	0.6895	0.1226
t-value		-0.02	-1.91	-4.46				0.00	-0.09	-3.52
λ estimates		λ_{L_1}		λ_{U_1}	λ_{L_2}		λ_{U_2}	λ_{L_3}		λ_{U_3}
est.		0.0145		0.5018	0.0131		0.3358	0.0131		0.1866
s.e.		0.0580		0.0342	0.0421		0.0325	0.0257		0.0308
$p_{L,U}$		0.6588		1.0000	0.7003		1.0000	0.6886		1

fx-rate portfolio $C_{C}^{[4]}$										
parameter estimates	α	δ_1		δ_{1_s}	δ_2		δ_{2_s}	δ_3		δ_{3_s}
est.	0.4462	4.1378		2.7416	1.2984		0.9825	0.7571		0.6394
s.e.	0.0350	0.4144		0.2031	0.1695		0.1043	0.1055		0.0600
t-value	12.74	9.99		13.50	7.66		9.42	7.18		10.66
model check		$\delta_2 - \delta_1$		$\delta_{2_s} - \delta_{1_s}$				$\delta_3 - \delta_2$		$\delta_{3_s} - \delta_{2_s}$
est.		-2.8394		-1.7591				-0.5412		-0.3430
s.e.		0.4477		0.2283				0.1996		0.1204
t-value		-6.34		-7.70				-2.71		-2.85
λ estimates		λ_{L_1}		λ_{U_1}	λ_{L_2}		λ_{U_2}	λ_{L_3}		λ_{U_3}
est.		0.3774		0.4301	0.2616		0.2735	0.1786		0.1873
s.e.		0.0303		0.0284	0.0275		0.0268	0.0268		0.0224
$p_{L,U}$		1.0000		1.0000	1.0000		1.0000	1.0000		1.0000

fx-rate portfolio C_t										
parameter estimates		df	$\rho_{(1,2)}$	$\rho_{(1,3)}$	$\rho_{(1,4)}$	$\rho_{(2,3)}$	$\rho_{(2,4)}$	$\rho_{(3,4)}$		
est.		5.0402	0.8507	0.5098	0.4650	0.6463	0.4734	0.3370		
s.e.		0.4860	0.0074	0.0234	0.0258	0.0172	0.0257	0.0284		
t-value		10.37	114.94	21.75	18.00	37.60	18.43	11.87		
λ estimates			$\lambda_{(1,2)}$	$\lambda_{(1,3)}$	$\lambda_{(1,4)}$	$\lambda_{(2,3)}$	$\lambda_{(2,4)}$	$\lambda_{(3,4)}$		
est.			0.5112	0.2106	0.1877	0.2978	0.1918	0.1339		
s.e.			0.0214	0.0235	0.0229	0.0247	0.0231	0.0198		
$p_{L,U}$			1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		

Table D.10: Estimation results for 4-dimensional fx-rate portfolios.

metal basket $C_{\Omega}^{[4]}$										
parameter estimates	α	δ_{1_s}	θ_1	δ_1	δ_{2_s}	θ_2	δ_2	δ_{3_s}	θ_3	δ_3
est.	0.4492	1.3669	2.9168	1.2838	1.1896	2.5220	1.1427	1.1265	2.2452	1.0805
s.e.	0.0680	0.1613	0.6851	0.1123	0.0972	0.4498	0.0663	0.0631	0.2957	0.0435
t-value	6.61	2.27	4.26	2.53	1.95	5.61	2.15	2.00	7.59	1.85
model check		$\delta_{2_s} - \delta_{1_s}$	$\theta_2 - \theta_1$	$\delta_2 - \delta_1$				$\delta_{3_s} - \delta_{2_s}$	$\theta_3 - \theta_2$	$\delta_3 - \delta_2$
est.		-0.1733	-0.3948	-0.1410				-0.0631	-0.2768	-0.0622
s.e.		0.1883	0.8195	0.1305				0.1159	0.5382	0.0793
t-value		-0.94	-0.48	-1.08				-0.54	-0.51	-0.78
λ estimates		λ_{L_1}		λ_{U_1}	λ_{L_2}		λ_{U_2}	λ_{L_3}		λ_{U_3}
est.		0.1525		0.1565	0.0940		0.0914	0.0673		0.0554
s.e.		0.0502		0.0487	0.0409		0.0373	0.0304		0.0278
$p_{L,U}$		0.9983		0.9990	0.9959		0.9927	0.9776		0.9680

metal basket $C_{C}^{[4]}$										
parameter estimates	α	δ_1		δ_{1_s}	δ_2		δ_{2_s}	δ_3		δ_{3_s}
est.	0.5473	1.5673		1.6477	0.9401		1.0255	0.7255		0.8053
s.e.	0.0310	0.1599		0.1852	0.0990		0.1190	0.0564		0.0714
t-value	17.64	9.80		8.90	9.50		8.62	12.86		11.28
model check		$\delta_2 - \delta_1$		$\delta_{2_s} - \delta_{1_s}$				$\delta_3 - \delta_2$		$\delta_{3_s} - \delta_{2_s}$
est.		-0.6272		-0.6222				-0.2147		-0.2202
s.e.		0.1880		0.2202				0.1139		0.1388
t-value		-3.34		-2.83				-1.88		-1.59
λ estimates		λ_{L_1}		λ_{U_1}	λ_{L_2}		λ_{U_2}	λ_{L_3}		λ_{U_3}
est.		0.3517		0.2972	0.2618		0.2303	0.2105		0.1914
s.e.		0.0255		0.0248	0.0252		0.0240	0.0197		0.0196
$p_{L,U}$		1.0000		1.0000	1.0000		1.0000	1.0000		1.0000

metal basket C_t										
parameter estimates		df	$\rho_{(1,2)}$	$\rho_{(1,3)}$	$\rho_{(1,4)}$	$\rho_{(2,3)}$	$\rho_{(2,4)}$	$\rho_{(3,4)}$		
est.		20.9773	0.6545	0.5272	0.4045	0.5581	0.4162	0.5233		
s.e.		3.9844	0.0132	0.0171	0.0205	0.0159	0.0199	0.0171		
t-value		5.26	49.42	30.79	19.74	35.10	20.89	30.60		
λ estimates			$\lambda_{(1,2)}$	$\lambda_{(1,3)}$	$\lambda_{(1,4)}$	$\lambda_{(2,3)}$	$\lambda_{(2,4)}$	$\lambda_{(3,4)}$		
est.			0.435	0.0161	0.0058	0.0205	0.0064	0.0156		
s.e.			0.0200	0.0101	0.0047	0.0120	0.0050	0.0098		
$p_{L,U}$			1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		

Table D.11: Estimation results for 4-dimensional metal baskets.

oil related commodity basket $C_{\Omega}^{[4]}$										
parameter estimates	α	δ_{1_s}	θ_1	δ_1	δ_{2_s}	θ_2	δ_2	δ_{3_s}	θ_3	δ_3
est.	0.7948	4.9276	5.2819	1.1841	3.6752	5.2819	1.0000	1.6289	2.4552	1.0000
s.e.	0.0232	0.6468	1.4847	0.1930	0.5720	1.4618	0.1807	0.1349	0.6599	0.0877
t-value	34.26	6.07	3.56	0.95	4.68	3.61	0.00	4.66	3.72	0.00
model check		$\delta_{2_s} - \delta_{1_s}$	$\theta_2 - \theta_1$	$\delta_2 - \delta_1$				$\delta_{3_s} - \delta_{2_s}$	$\theta_3 - \theta_2$	$\delta_3 - \delta_2$
est.		-1.2524	0.0000	-0.1841				-2.0463	-2.8267	0.0000
s.e.		0.8634	2.0835	0.2644				0.5876	1.6038	0.2008
t-value		-1.45	0.00	-0.70				-3.48	-1.76	0.00
λ estimates		λ_{L_1}		λ_{U_1}	λ_{L_2}		λ_{U_2}	λ_{L_3}		λ_{U_3}
est.		0.6747		0.0419	0.6298		0.0000	0.3732		0.0000
s.e.		0.0259		0.0355	0.0336		0.0514	0.0442		0.0250
$p_{L,U}$		1.0000		0.8830	1.0000		0.4995	1.0000		0.5005

oil related commodity basket $C_{C}^{[4]}$										
parameter estimates	α	δ_1		δ_{1_s}	δ_2		δ_{2_s}	δ_3		δ_{3_s}
est.	0.5083	5.3700		11.4860	3.7669		8.9926	1.6357		1.2933
s.e.	0.0353	0.5322		0.8882	0.4457		0.5301	0.1749		0.1154
t-value	14.40	10.09		12.93	8.45		16.96	9.35		11.21
model check		$\delta_2 - \delta_1$		$\delta_{2_s} - \delta_{1_s}$				$\delta_3 - \delta_2$		$\delta_{3_s} - \delta_{2_s}$
est.		-1.6031		-2.4933				-2.1312		-7.6993
s.e.		0.6941		1.0344				0.4788		0.5425
t-value		-2.31		-2.41				-4.45		-14.19
λ estimates		λ_{L_1}		λ_{U_1}	λ_{L_2}		λ_{U_2}	λ_{L_3}		λ_{U_3}
est.		0.4467		0.4629	0.4229		0.4552	0.3327		0.2877
s.e.		0.0315		0.0333	0.0308		0.0327	0.0276		0.0248
$p_{L,U}$		1.0000		1.0000	1.0000		1.0000	1.0000		1.0000

oil related commodity basket C_t										
parameter estimates		df	$\rho_{(1,2)}$	$\rho_{(1,3)}$	$\rho_{(1,4)}$	$\rho_{(2,3)}$	$\rho_{(2,4)}$	$\rho_{(3,4)}$		
est.		4.9457	0.7905	0.7552	0.3945	0.6628	0.4155	0.3652		
s.e.		0.5827	0.0263	0.0296	0.0603	0.0460	0.0598	0.0577		
t-value		8.49	30.09	25.50	6.54	14.39	6.94	6.33		
λ estimates			$\lambda_{(1,2)}$	$\lambda_{(1,3)}$	$\lambda_{(1,4)}$	$\lambda_{(2,3)}$	$\lambda_{(2,4)}$	$\lambda_{(3,4)}$		
est.			0.4365	0.3979	0.1597	0.3146	0.1686	0.1479		
s.e.			0.0390	0.0395	0.0334	0.0445	0.0345	0.0311		
$p_{L,U}$			1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		

Table D.12: Estimation results for 4-dimensional oil related commodity basket.

stock indices								
	χ^2-statistic (p-value[%])	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	115.04 (0.06)	-578.74	-521.05	-0.7068	0.7028	0.6956	-35.00	13.52
C_C.est	126.20 (0.01)	-557.38	-516.99	-0.6046	0.9865	1.1934	7.71	-12.89
t-Copula.est	**78.40 (31.16)**	**-614.03**	**-573.64**	**-0.7337**	**0.5019**	**0.3507**	**-7.53**	**2.69**
Ω.test	137.71 (0.00)	-329.63	-271.80	-0.7068	2.2417	2.7507	32.21	125.30
C_C.test	101.39 (1.56)	-336.18	-295.70	-0.6046	171.20	**1.3999**	**30.61**	**37.92**
t-Copula.test	**92.76 (5.92)**	**-367.63**	**-327.14**	**-0.7337**	**1.9155**	2.9954	33.69	108.01

fx-rates								
	χ^2-statistic (p-value[%])	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	224.03 (0.00)	-1288.18	-1226.65	-1.0441	0.6668	0.5123	-22.82	6.98
C_C.est	229.35 (0.00)	-1228.33	-1185.25	-0.9565	0.9468	1.0297	**-4.52**	-16.37
t-Copula.est	**109.91 (0.34)**	**-1354.72**	**-1311.64**	**-1.1236**	**0.3932**	**0.3640**	15.27	**-5.53**
Ω.test	**437.14 (0.00)**	**-981.81**	**-920.20**	-1.0441	11.2573	12.8212	**29.14**	186.54
C_C.test	474.46 (0.00)	-926.29	-883.16	-0.9565	371.8833	**8.0041**	40.99	**87.62**
t-Copula.test	459.26 (0.00)	-960.93	-917.80	**-1.1236**	**10.4183**	11.8035	65.12	146.11

metal basket								
	χ^2-statistic (p-value[%])	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	95.33 (2.38)	-1166.42	-1101.66	-0.6666	**0.3696**	**0.3579**	**4.69**	**-1.35**
C_C.est	108.46 (0.44)	-1112.02	-1066.70	-0.6589	1.9696	2.4132	30.09	20.67
t-Copula.est	**59.38 (69.53)**	**-1180.05**	**-1134.72**	**-0.6763**	0.5582	1.0503	22.24	20.79
Ω.test	124.19 (0.01)	**-492.17**	**-428.95**	-0.6666	6.7884	4.1303	**29.31**	**18.10**
C_C.test	**124.61 (0.02)**	-469.00	-424.75	-0.6589	325.9473	**2.1571**	95.85	20.34
t-Copula.test	123.97 (0.01)	-466.20	-421.94	**-0.6763**	**5.8653**	3.0023	63.93	20.91

oil related commodity basket								
	χ^2-statistic (p-value[%])	AIC	BIC	Entropy	AD	ADS	PD	PDS
Ω.est	**122.95 (0.01)**	**-2415.60**	**-2357.88**	**-2.8604**	**0.1458**	**0.2043**	**-22.52**	-8.52
C_C.est	178.52 (0.01)	-2201.11	-2160.71	-2.4864	1.1695	1.1501	-27.69	**-3.86**
t-Copula.est	1357.79 (0.00)	-1740.87	-1700.47	-1.1017	30.5945	28.9588	-57.86	-47.67
Ω.test	**102.07 (0.74)**	**-1947.45**	**-1889.72**	**-2.8604**	**0.4662**	**0.5282**	32.35	-5.31
C_C.test	108.57 (0.44)	-1824.40	-1783.99	-2.4864	106.42	0.6348	**-0.99**	**-0.17**
t-Copula.test	667.78 (0.00)	-1643.13	-1602.72	-1.1017	26.3588	24.8209	-53.43	-48.85

Table D.13: Test results for the 4-dimensional portfolios.

	p_{asy}	portfolio	$p_{asy.real}$	portfolio
stock indices				
max C_Ω	0.9791	NAS-SP	0.8868	NAS-SP
max C_C	0.9675	NAS-SP	0.9675	NAS-SP
min C_Ω	0.6205	SP-NIK	0.0479	NAS-NIK
min C_C	0.3529	NAS-NIK	0.3082	NAS-NIK
fx rates				
max C_Ω	0.9792	HKD-SGD	0.9791	HKD-SGD
max C_C	0.8952	HKD-SGD	0.8952	HKD-SGD
min C_Ω	0.1611	AUD-YEN	0.1513	AUD-YEN
min C_C	0.2380	AUD-SGD	0.2380	AUD-SGD
metal baskets				
max C_Ω	0.8760	Al-Pb	0.7147	Al-Cu
max C_C	0.8751	Cu-Pb	0.8751	Cu-Pb
min C_Ω	0.0823	Al-Zn	0.0416	Al-Zn
min C_C	0.1700	Pb-Zn	0.1700	Pb-Zn
oil related commodities				
max C_Ω	1.0000	ker-hea, ker-die, hea-die	0.5001	hea-die
max C_C	0.9999	gas-die	0.9999	gas-die
min C_Ω	0.9966	hea-gas	0.4967	gas-die
min C_C	0.1220	hea-die	0.1220	hea-die

Table D.14: Bivariate portfolios with the maximal and minimal values of p_{asy} under the C_Ω and C_C model, together with the corresponding values of p_{asy} and $p_{asy.real}$.

portfolio	model	plane 1-2		plane (1,2)-3		plane (1,2,3)-4	
		p_{asy}	$p_{asy.real}$	p_{asy}	$p_{asy.real}$	p_{asy}	$p_{asy.real}$
stock indices	C_Ω	0.8703	0.8648	0.0141	0.0131	0.0975	0.0933
	C_C	0.9544	0.9544	0.7884	0.7877	0.9431	0.9430
fx-rates	C_Ω	1.0000	0.6588	1.0000	0.7003	1.0000	0.6886
	C_C	0.6384	0.2082	0.1911	0.6384	0.2082	0.1911
metal basket	C_Ω	0.0352	0.0351	0.0458	0.0453	0.2567	0.2429
	C_C	0.5392	0.5328	0.0768	0.5392	0.5328	0.0768
oil related	C_Ω	1.0000	0.8830	1.0000	0.4995	1.0000	0.5005
commodity basket	C_C	0.5896	0.7376	0.7939	0.5896	0.7376	0.7939

Table D.15: p_{asy} and $p_{asy.real}$ values for the bivariate projections of the 4-dimensional portfolios.

Appendix E

Nonlinear Term Structure Dependence

E.1 Affine Models

We briefly review the concept of ATSMs following Duffie and Kan [33], and Dai and Singleton [24]. In contrast to the considerations above, we start with an equivalent martingale measure Q, and later on we work out the link to the real world under the so-called physical measure P. In the absence of arbitrage opportunities, the price at time t of a zero-coupon bond that matures at time $t+\tau$, denoted $P(t,\tau)$, is given by

$$P(t,\tau) = \mathbf{E}_t^Q \left\{ \exp\left(-\int_t^{t+\tau} R(s)\, ds \right) \right\} , \tag{E.1.1}$$

where $\mathbf{E}_t^Q\{\,\cdot\,\}$ denotes the conditional expectation under the risk neutral measure Q. A d-factor affine term structure model is obtained under the assumption that the instantaneous short rate R is an affine function of a d-dimensional vector process of state variables $X = (X_1, ..., X_d)$:

$$R(t) = R_0 + a_1 X_1(t) + ... + a_d X_d(t) = R_0 + a^\top X(t)\,, \tag{E.1.2}$$

and that X follows an affine diffusion:

$$dX(t) = \kappa_Q(\theta_Q - X(t))\, dt + \Sigma\sqrt{S(t)}\, dW_Q(t)\,, \tag{E.1.3}$$

where W_Q is a d-dimensional standard Brownian motion under the measure Q, θ_Q is a d-vector, κ_Q and Σ are $d \times d$ matrices, and $S(t)$ is a $d \times d$ diagonal matrix

with the ith diagonal element:

$$\{S(t)\}_{ii} = \alpha_{S\,i} + \beta_{S\,i}{}^{\top} X(t). \tag{E.1.4}$$

Provided that a parameterization is admissible, we know from Duffie and Kan [33]:

$$P(t,\tau) = \exp\left(A(\tau) + B(\tau)^{\top} X(t)\right) , \tag{E.1.5}$$

where A and B satisfy the ordinary differential equations:

$$\frac{dA(\tau)}{d\tau} = \theta_Q^{\top}\kappa_Q^{\top} B(\tau) + \frac{1}{2}\sum_{i=1}^{d}\left[\Sigma^{\top} B(\tau)\right]_i^2 \alpha_{S\,i} - R_0 , \tag{E.1.6}$$

$$\frac{dB(\tau)}{d\tau} = -\kappa_Q^{\top} B(\tau) + \frac{1}{2}\sum_{i=1}^{d}\left[\Sigma^{\top} B(\tau)\right]_i^2 \beta_{S\,i} - a . \tag{E.1.7}$$

The particular specification of the Q-dynamics of X in Equation (E.1.3) and the definition of R in Equation (E.1.2) allow the exponential affine representation of the bond prices in Equation (E.1.5). It is well known that the measure Q is generated by a change of measure with respect to the empirical/physical measure P that describes the behavior of the stochastic factors in the "real world". To obtain an affine structure for X under both measures P and Q, we restrict the measure change in terms of the market price of risk Λ to:

$$\Lambda(t) = \sqrt{S(t)}\lambda , \tag{E.1.8}$$

where λ is a d-vector of constants. Thus, the P-dynamics of the state process X is:

$$dX(t) = \kappa(\theta - X(t))\, dt + \Sigma\sqrt{S(t)}\, dW(t) , \tag{E.1.9}$$

where W is a d-dimensional standard Brownian motion under P and:

$$\kappa = \kappa_Q - \Sigma\Psi , \tag{E.1.10}$$

$$\theta = \kappa^{-1}\left(\kappa_Q\theta_Q + \Sigma\phi\right) . \tag{E.1.11}$$

The ith row of the $d \times d$ matrix Ψ is filled in by $\lambda_i \beta_{S\,i}{}^{\top}$ and ϕ is a d-vector whose ith element is given by $\lambda_i \alpha_{S\,i}$. The functions A and B describing the two-factor

generalized Vasicek model of Babbs and Nowman [6] are given by the parameters $R_0 \in \mathbb{R}$, $\kappa_1, \kappa_2, \sigma_1, \sigma_2 \in \mathbb{R}^+$, $\rho \in]-1, 1[$, and $\lambda_1, \lambda_2 \in \mathbb{R}$ in the following way:

$$A(\tau) = \sum_{i=1,2} \left[-\gamma_i \left(B_i(\tau) + \tau \right) - \frac{\sigma_i^2}{4\,\kappa_i} B_i(\tau)^2 \right] - R_0 \tau + A_\rho(\tau)\,, \quad \text{(E.1.12)}$$

$$B(\tau) = \left(B_1(\tau), B_2(\tau) \right)^\top = \left(\frac{e^{-\kappa_1 \tau} - 1}{\kappa_1}, \frac{e^{-\kappa_2 \tau} - 1}{\kappa_2} \right)^\top , \quad \text{(E.1.13)}$$

where $\gamma_1 = \frac{\sigma_1 \lambda_1}{\kappa_1} - \frac{\sigma_1^2}{2\,\kappa_1^2}$, $\gamma_2 = \frac{\sigma_2 \left(\rho \lambda_1 + \sqrt{1-\rho^2}\, \lambda_2 \right)}{\kappa_2} - \frac{\sigma_2^2}{2\,\kappa_2^2}$, and:

$$A_\rho(\tau) = \frac{\rho\, \sigma_1\, \sigma_2}{\kappa_1 + \kappa_2} \left[\frac{1}{\kappa_1} (B_1(\tau) + \tau) + \frac{1}{\kappa_2} (B_2(\tau) + \tau) - B_1(\tau)\, B_2(\tau) \right]. \quad \text{(E.1.14)}$$

E.2 Yield-Factor Representation

A distinct feature of ATSM framework is that the latent state variables can be transferred to an appropriate set of yields, see Duffie and Kan [33]. Moreover, the affine structure of the latent variables is preserved for the yields, and the yields can be viewed at as a new set of state variables provided some technical conditions hold. Given a d-factor ATSM with state variable $X = (X_1, ..., X_d)^\top$. For a set of maturities $(\tau_1, ..., \tau_d)$ the corresponding yields $Y = (Y_1, ..., Y_d)^\top$ are given by Equation (E.1.5):

$$Y(t) = A + B\, X(t), \quad \text{(E.2.1)}$$

where

$$A = \begin{pmatrix} -\frac{A(\tau_1)}{\tau_1} \\ \vdots \\ -\frac{A(\tau_d)}{\tau_d} \end{pmatrix}, \quad B = \begin{pmatrix} -\frac{B_1(\tau_1)}{\tau_1} & \dots & -\frac{B_d(\tau_1)}{\tau_1} \\ \vdots & \ddots & \vdots \\ -\frac{B_1(\tau_d)}{\tau_d} & \dots & -\frac{B_d(\tau_d)}{\tau_d} \end{pmatrix}. \quad \text{(E.2.2)}$$

Provided B is non-singular, we can state the analogue of Equation (E.1.9), i.e. the state equation for the yield vector Y:

$$dY(t) = \tilde{\kappa}(\tilde{\theta} - Y(t))\, dt + \widetilde{\Sigma} \sqrt{\widetilde{S}(t)}\, dW(t)\,, \quad \text{(E.2.3)}$$

where

$$\tilde{\kappa} = B\, \kappa\, B^{-1}\,, \quad \tilde{\theta} = B\, \theta + A\,, \quad \widetilde{\Sigma} = B\, \Sigma\,, \text{ and } \widetilde{S}(t) = \{\tilde{\alpha}_i + \tilde{\beta}_i^\top Y(t)\}\,, \quad \text{(E.2.4)}$$

and $\tilde{\alpha}_i = \alpha_{S\,i} - {\beta_{S\,i}}^\top B^{-1} A$, and $\tilde{\beta}_i = {B^{-1}}^\top \beta_{S\,i}$. We briefly discuss the yield dynamics implied by the two-factor generalized Gaussian model. Here, we find $\theta = 0$, and $S(t) = \mathbf{I}_d$ what results into $\tilde{\kappa} = B\,\kappa\,B^{-1}$, $\tilde{\theta} = A$, $\widetilde{\Sigma} = B\,\Sigma$, and $\widetilde{S}(t) = \mathbf{I}_d$:

$$dY(t) = B\,\kappa\,B^{-1}(A - Y(t))\,dt + B\,\Sigma\,dW(t)\,, \tag{E.2.5}$$

where $B\,\kappa\,B^{-1}$ describes the mean reversion including cross-dependencies between Y_1 and Y_2, and the covariance is given by $B\,\Sigma\,\Sigma^\top\,B^\top$.

The Author

Markus Junker was born 1973 in Marburg an der Lahn. After his Abitur 1993 at the Gauss Gymnasium in Worms he studied mathematics and physics at the University of Kaiserslautern. He received his diploma in mathematics in 1999 with a focus on financial time series analysis, and the "1. Staatsexamen" in mathematics and physics in the year 2000. Both degrees were certified by the University of Kaiserslautern.
In the year 2000 he joined the Financial Engineering group of the center of advanced european studies and research (caesar) in Bonn, where he worked in several projects with the industrial partners of the group. Since 2002 he is the deputy head of the Financial Engineering group and is currently working as a specialist for multidimensional dependence structures.

Markus Junker wurde am 30.11.1973 in Marburg an der Lahn geboren. Nach seinem Abitur 1993 am Gauss Gymnasium Worms studierte er Mathematik und Physik an der Universität Kaiserslautern. Sein Diplom in Mathematik (mit Schwerpunkt in Finanzzeitreihen) wurde ihm 1999, das 1. Staatsexamen in Mathematik und Physik für das höhere Lehramt im Jahr 2000 von der Universität Kaiserslautern verliehen.
Herr Junker began 2000 seine Mitarbeit in der Gruppe Financial Engineering am center of advanced european studies and research (caesar) in Bonn und hat an verschiedenen Projekten der Gruppe und ihrer Industriepartner mitgearbeitet. Seit 2002 ist er stellvertretender Arbeitsgruppenleiter und arbeitet zur Zeit als Spezialist für mehrdimensionale Abhängigkeitsstrukturen.

Contact

Markus Junker	Ludwig-Erhard-Allee 2	e-mail: junker@caesar.de
caesar research center	53175 Bonn, Germany	http://www.caesar.de/fen