

ulm university universität ulm

Time-Varying Copula Models for Financial Time Series

Magda Mroz

Universität Ulm
Fakultät für Mathematik und
Wirtschaftswissenschaften



ulm university universität **u**ulm

Time-Varying Copula Models for Financial Time Series

Dissertation zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
der Fakultät für Mathematik und Wirtschaftswissenschaften
an der Universität Ulm

vorgelegt von
Magda Mroz
aus Cieszyn (Polen)

im Juli 2012

Amtierender Dekan:	Prof. Dr. Paul Wentges
1. Gutachter:	Prof. Dr. Ulrich Stadtmüller
2. Gutachter:	Prof. Dr. Rüdiger Kiesel
Tag der Promotion:	19.10.2012

Ich versichere hiermit, dass ich die Arbeit selbständig angefertigt habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht habe.

Ulm, 30.07.2012 Magda Mroz

meinen Kindern

Acknowledgments

I first and foremost want to express my deepest gratitude to Professor Ulrich Stadtmüller for being my supervisor and for always providing me with great ideas and constant support during my doctorate. I very much enjoyed being a member of the Institute of Number Theory and Probability Theory, and in particular having the opportunity to participate in the student education.
I also thank Professor Rüdiger Kiesel for being the second advisor of this thesis and suggesting the empirical field of application for the theoretical results.
Furthermore, I am deeply indebted to Professor Karsten Urban who provided financial support for various conference visits and created a family-friendly environment for me, just to mention a few of the numerous benefits of being a scholar of the DFG-Research Training Group 1100.
For helping me with organizational matters, I want to thank Helga Lautenbach and Petra Hildebrand, the good souls of the Institute and of the Research Training Group.
Vito Baccelliere, Christian Hering, Christian Hirsch, Michael Weyhmüller and Marta Zampiceni were not only willing to thoroughly proofread this thesis, but also became good colleagues and companions during its completion. Our professional and personal discussions about copulas, commas, and all the other curiosities of life are most appreciated.
I have realized that family and true friends are the most important in life. Knowing your backing and support helped bridging the tough stages of this journey that is about to end. I have learned a lot during the last four years, but still stand in awe of the leaps that Joelle and Lucien have taken at the same time. Being a role model for you and watching you succeed is the greatest confirmation for me.

Ulm, July 2012 Magda Mroz

Contents

1 Introduction **1**

2 Mathematical Background **7**
2.1 Basic properties of copulas 8
2.1.1 Definitions 8
2.1.2 Fréchet-Hoeffding bounds and Sklar's Theorem 10
2.1.3 Measures of dependence 12
2.1.4 General sampling techniques 17
2.1.5 Some copula classes 18
2.2 Copula estimation 27
2.2.1 Empirical copulas 27
2.2.2 Maximum likelihood estimation 29
2.2.3 Other inference techniques 38
2.3 Extreme value theory 40
2.3.1 Basic theory 40
2.3.2 Norming constants 42
2.3.3 Dependent sequences 44

3 GARCH Time Series **49**
3.1 Univariate models 52
3.1.1 Stationarity and moment conditions 55
3.1.2 Finite-dimensional distribution 58
3.1.3 Quasi maximum likelihood estimation 60
3.2 Multivariate extensions 63
3.2.1 General approaches 64
3.2.2 SCOMDY model 66
3.3 Time-dynamic copula models 70

4 Homogeneity Tests **73**
4.1 Moving window estimators 73
4.2 Binomial test 77
4.2.1 Partial tests 77
4.2.2 Multiple test 82
4.3 Extremal test 89
4.3.1 Multivariate central limit theorem 90
4.3.2 Asymptotic extreme value theory 98
4.3.3 Local power 99

4.4 Test based on forecast regression . . . 102
4.5 Discussion . . . 108

5 Empirical Results **109**
5.1 Empirical size of the Binomial test . . . 109
5.1.1 Simulation . . . 110
5.1.2 Test and bootstrap . . . 112
5.1.3 Implementation . . . 116
5.2 Commodity contracts . . . 116
5.2.1 Data description . . . 117
5.2.2 SCOMDY modeling and testing . . . 119
5.2.3 Discussion . . . 125

A Examples of Copula Families Used **129**
A.1 Gaussian copula . . . 129
A.2 Student copula . . . 131
A.3 Marshall-Olkin copula . . . 132
A.4 Plackett copula . . . 134
A.5 Clayton copula . . . 135
A.6 Gumbel copula . . . 136
A.7 Frank copula . . . 137
A.8 Symmetrized Joe-Clayton copula . . . 139
A.9 Summary . . . 142

B Conditions for Asymptotic Properties in the SCOMDY Model **145**

C Conditions for Yaskov's Asymptotic Result **149**

Zusammenfassung **151**

Bibliography **155**

List of Figures **171**

List of Tables **173**

1 Introduction

"If I have seen further it is by standing on ye shoulders of Giants."
Isaac Newton, 1676

The initial motivation for the use of copulas in finance stems from the growing evidence that the dependence observed between the occurring quantities exhibits non-normal features. The term *non-normal* hereby stands for a type of dependence which is not compatible with the (multivariate) Gaussian distribution. A proper multivariate perspective on dependence modeling might be a challenging task. As Joe (1997, p. 3) observes, the "study of multivariate distributions is not easy because one cannot just write down a family of functions and expect it to satisfy the necessary conditions for multivariate cumulative distribution functions [...]." This is where the famous theorem due to Sklar (1959) comes in. Established about half a century ago, it laid the foundation for a separate consideration of marginal models and their (joint) dependency and made copulas so popular. Before then, modeling of multivariate distributions was indeed mainly confined to the multivariate normal distribution, or at most to the class of elliptical distributions, and dependence was virtually associated to Pearson's correlation coefficient. The two-stage, bottom-up approach propagated with Sklar's Theorem opens the door to flexible and comprehensive models. Furthermore, copulas allow for an alternative view on dependencies.
Owing to this flexibility, copulas have experienced an enormous jump in popularity in the last decade. Mikosch (2006) notes that a Google search for the word *copula* resulted in 10 000 responses in 2003, rising to 650 000 hits in 2005. Genest et al. (2009a) report on 750 000 Google hits in 2009, whereas we can state 1 820 000 responses as this thesis is written (see Google search). Being a very active research area, copula methodology as a means of representing multivariate distributions has entailed a plenty of profound mathematical theory with respect to probabilistic modeling (Sklar (1996) gives a compelling round-up on this) and statistical inference (the most important original contribution within this scope is perhaps Genest et al. (1995) that we will use extensively later on). Initially, copula theory was used and developed in medical and natural sciences, so that its field of application comprises for example hydrology (Genest and Favre, 2007 and Salvadori, 2007) and health science (Cui and Sun, 2004). However, it owes its rapid ascent mainly to the fact that from a statistician's point of view we have a very favorable data situation especially for economic problems: copula concepts prove useful and practicable in the fields of finance and insurance such as, for example, risk management for portfolios of financial derivatives (see McNeil et al.,

2005 or Cherubini et al., 2004), or dependence modeling in actuarial mathematics (see Frees and Valdez, 1998). Genest et al. (2009a) address the so-called "advent of copulas in finance" by surveying the related academic literature up to 2006, and they identify Embrechts et al. (2002) as being seminal for the spreading use of copulas for multivariate financial applications. Figure 1.1 (courtesy of M. Hofert) shows that its first appearance as technical report in 1998 initiated an enormous growth in the relevant literature.

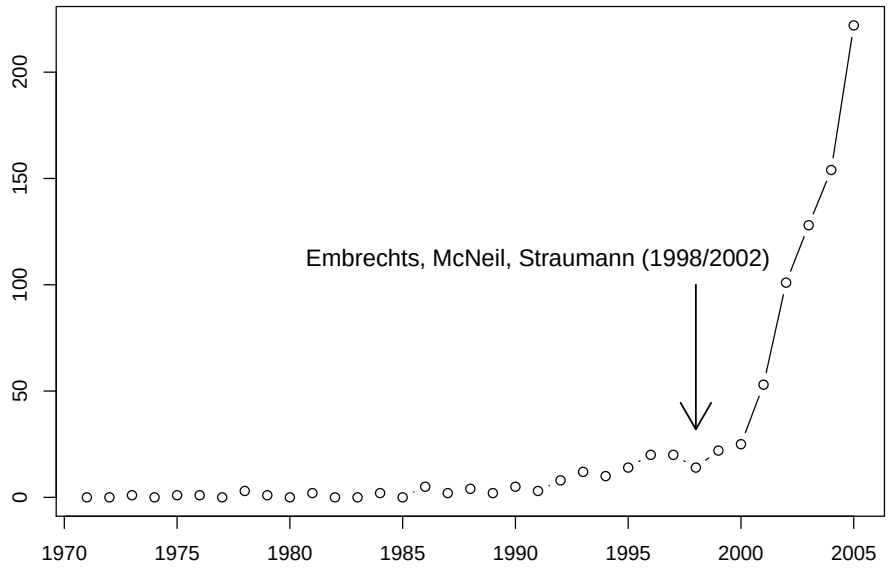


Figure 1.1: Annual number of publications on copula theory 1971–2005 (cf. Figure 1 in Genest et al., 2009a).

However, when we understand the copula approach as silver bullet concerning multivariate modeling, there unquestionably are downsides. Mikosch (2006) initiated a lively debate by pointing out that copulas might "not contribute to a better understanding of multivariate extremes" and "do not fit into the existing framework of stochastic processes and time series analysis [...]" (p. 19 ib.). Embrechts (2006) stated as a reply that by no means copulas yield a universal approach for global, dynamic multivariate questions, although they are "extremely useful in understanding the limitations caused by the data non-availability in financial and insurance risk management." (p. 46 ib.).
The main objective of this thesis is an analysis of the potential time-inhomogeneity in the dependence between multiple financial time series. With our research we want to contribute on the one hand to a better understanding of what really happens in real-world data and how we can draw conclusions on the basis of valid statistical inference. On the other hand, we want to fill the gap when copula theory is actually applied to financial modeling – so that certain well-accepted features, so-called stylized facts, of the data have to be taken into account – and tackle the question whether dependencies in such a case can be assumed constant throughout time or rather have to be modeled

in a time-inhomogeneous way. Patton (2012) provides a broad overview of copula theory used in economic time series models. In particular, he addresses the crucial role of appropriate dependence modeling in applications.
Note that we use the concepts of copula theory to capture the dependence between distinct series of data given for some joint points in time. Thus, we model the so-called *spatial dependence*, whereas the copula approach to multivariate modeling could basically also refer to the dependence structure of one specific variable at different points in time. Cherubini et al. (2011, p. 5) state that although copulas have mainly been intended in the latter sense in econometric applications, this is somewhat "curious" and "the flexibility of copulas immediately becomes a problem" as the usual restrictions of the so modeled processes to be for example Markovian or a martingale are not necessarily met. The spatial dependence can nevertheless exhibit a temporally changing structure. The components of a multivariate time series are themselves (univariate) time series, i.e. models to capture the temporal evolution of heteroscedastic financial data, and in our set-up their dependence is modeled by the multivariate distribution of the SCOMDY model's residuals, more precisely by the parametric copula of this joint distribution. At this point, the property of the residuals to be independent and identically distributed can be relaxed to the effect that the copula parameters are allowed to vary throughout time. This implies that through a changing copula parameter also the dependency is not constant, but may be stronger or weaker depending on the point of time. We will not assume any specific model for the evolution of the parameter, but test its homogeneity by a data-driven approach.

Outline

Our epigraph is part of a citation in Embrechts (2009) in support of a historical perspective when applying copulas. Keeping this in mind "copulas form a most useful concept for a lot of applied modeling [...]" (Embrechts, 2009, p. 13). In this spirit we develop our methodology based on some well-known and well-studied mathematical concepts while focusing on suitable inference techniques in the context of a special copula-based multivariate time series model.
In Chapter 2 we introduce the basic definitions and theorems related to copula theory. We start from an analytical point of view to proceed with Sklar's Theorem. It forms the basis for many further insights for example construction and sampling of copulas which we will briefly review. Then we focus on estimation techniques for multivariate data and within this scope on maximum likelihood related approaches. We conclude our account on the mathematical background with basic extreme value theory, particularly for weakly dependent sequences.
Chapter 3 provides a thorough presentation of heteroscedastic time series modeling. First, we review the common results on stationarity, marginal distributions and estimation in the univariate generalized autoregressive conditionally heteroscedastic (GARCH) model due to Engle (1982). Owing to its special structure, the semiparametric copula-based multivariate dynamic (SCOMDY) model due to Chan et al. (2009a) directly retains many findings presented in the univariate case. Although there are more general approaches

for multivariate heteroscedasticity, we decided in favor of the SCOMDY model due to its parsimonious parametrization and its statistical tractability. Anyway, we give a brief overview of these alternatives and review some of their properties. In particular, we are able to derive the tail behavior of the finite-dimensional distribution of the SCOMDY model by embedding it into a class of multivariate models for which this result was recently shown by Fernández and Muriel (2009). Finally, we give an overview of the literature related to time-dynamic copula models.
Chapter 4 contains our original contribution in pursuit of the objective to detect inhomogeneities in the dependence in multivariate time series. A possible time-varying dependence structure is captured by the parameter of the underlying copula family in the SCOMDY model.
We start by defining an estimator series of local estimators on overlapping subsamples of length $b = \mathcal{O}(T)$ and derive the basic joint distributional results as the overall observation horizon T tends to ∞ in Section 4.1. In Section 4.2 a Binomial test based on independent subsamples is constructed. We discuss the notion of statistical size and local power in detail, and show that the Binomial test has local power of order $\sqrt{b}$. The Extremal test presented in Section 4.3 is based on subsamples that overlap by a certain fixed amount of observations. The considerations concerning its local power of order $\sqrt{b/\log(T/b)}$ entail an extreme value limit result for asymptotically normal random variables. Finally, a third test that uses the whole of the moving window estimators and has local power of order $\sqrt{T}$ is presented in Section 4.4. A summary of the advantages and drawbacks of the different homogeneity tests concludes.
The first empirical analysis in Chapter 5 is a simulation study which shows that a bootstrap for assessing the estimation variance is unavoidable in finite samples. Secondly, we apply the aforementioned insight for the Binomial and the Extremal test to real-world financial data. We find that inhomogeneities in the dependence structure come along with unusual market events and can be remedied by careful univariate modeling. This is in accordance with the general perception that changes in the dependency are rather an exception than the rule, and stem from extreme economic events. In most cases a copula that is constant throughout time is sufficient to capture the dependence among the components of a multivariate time series model, and there is no gain in choosing a more complicated and statistically unhandy model with time-varying copula parameters. In Appendix A we give an overview of all the copula families used in this thesis.

Notation

We will use the following notational conventions throughout the text and give additional definitions if needed.

- Let $f\colon S\times\Theta \to B$, $f(x;\theta) = y$ be an arbitrary parametric function. By $\mathrm{dom}_\theta(f) := \{x \in S\colon \exists\, y \in B \text{ such that } f(x;\theta) = y\}$ we denote the function's *domain* for fixed θ, by $\mathrm{ran}_\theta(f) := \{y \in B\colon \exists\, x \in S \text{ such that } y = f(x;\theta)\}$ the function's *range* for fixed θ. The analogous definitions for nonparametric functions are $\mathrm{dom}(f)$ and $\mathrm{ran}(f)$. The short-hand notation for the *partial derivative* with respect to the

parameter (vector) θ is $f_\theta(x;\theta) := \frac{\partial}{\partial\theta} f(x;\theta)$. An analogous notation holds for higher order derivatives.

- The set of *real numbers* is denoted by $\mathbb{R}$, and its *closure* is given by $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. The subset of real numbers $0 \leq x \leq 1$ is the *unit interval* denoted by [0,1] and its d-variate analogue is the *unit cube* denoted by $[0,1]^d$.

- The *sign* of a real number $x \in \mathbb{R}$ is $\operatorname{sgn}(x) = -1$ if $x < 0$, $\operatorname{sgn}(x) = 1$ if $x > 0$ and $\operatorname{sgn}(0) = 0$. The *indicator function* on a set $A \subseteq \bar{\mathbb{R}}$ is $\mathbb{1}_A(x) = 1$ if $x \in A$ and $\mathbb{1}_A(x) = 0$ otherwise.

- For two real numbers $x, y \in \mathbb{R}$ we denote their *maximum* by $x \wedge y := \max\{x, y\}$ and their *minimum* by $x \vee y := \min\{x, y\}$.

- A *probability space* is given by $(\Omega, \mathcal{F}, \mathbb{P})$ with an arbitrary non-empty *sample space* Ω, a *σ-algebra* $\mathcal{F} \subseteq 2^\Omega$ and a *probability measure* $\mathbb{P}\colon \mathcal{F} \to [0,1]$. We can have *random variables* $X\colon \Omega \to \mathbb{R}$ or d-variate *random vectors* $\boldsymbol{X}\colon \Omega \to \mathbb{R}^d$ on this space. The symbol $\sim$ stands for *distributed according to* some distribution, $\xrightarrow{d}$ means *convergence in distribution*, and $\overset{d}{=}$ denotes *equality in distribution* of two random quantities possibly on different probability spaces.

- The (continuous) *uniform distribution* on the unit interval is denoted by $\mathcal{U}(0,1)$, the *binomial distribution* with size $n \geq 1$ and probability $p \in (0,1)$ by $\operatorname{Bin}(n,p)$, and the *exponential distribution* with mean $1/\lambda$, $\lambda > 0$ by $\operatorname{Exp}(\lambda)$. Furthermore we denote the *normal distribution* with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ by $\mathcal{N}(\mu, \sigma^2)$, so that $\mathcal{N}(0,1)$ corresponds to the *standard normal law* with distribution function $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2}\, dt$.

- Let $f = f(x)$ and $g = g(x) > 0$ be either real-valued functions of real-valued sequences for all $x \in \mathbb{R}$. Then, as $x \to a \in \bar{\mathbb{R}}$,
  - $f \in \mathcal{O}(g)$ means that $\limsup_{x\to a} \left|\frac{f(x)}{g(x)}\right| < \infty$, i.e., g is an *asymptotic upper bound* for f;
  - $f \in o(g)$ means that $\lim_{x\to a} \left|\frac{f(x)}{g(x)}\right| = 0$, i.e., f is *asymptotically negligible* compared to g;
  - $f \sim g$ means that $\lim_{x\to a} \frac{f(x)}{g(x)} = 1$, i.e., f and g are *asymptotically equal*;
  - $f \asymp g$ means that $0 < \liminf_{x\to a} |f(x)g(x)| < \infty \leq \limsup_{x\to a} |f(x)g(x)| < \infty$, i.e., f and g^{-1} are *asymptotically equivalent*.

- Let X be a random variable and $\{X_n\}_{n\geq 1}$ a sequence of random variables. Then, as $n \to \infty$
  - $X_n \xrightarrow{P} X$ means that $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$ for all $\varepsilon > 0$, i.e., X_n *converges in probability* to X;
  - $X_n \xrightarrow{a.s.} X$ means that $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$, i.e., X_n *converges almost surely* to X.

The symbols $\mathcal{O}$, o, $\sim$ and $\asymp$ can all be understood in the sense of these convergence notions.

2 Mathematical Background

This chapter is devoted to an introduction of the main mathematical building blocks of this thesis. In Section 2.1 the intuitive definition of a *copula*, namely

> "A d-dimensional copula is a d-dimensional distribution function on the unit cube $[0,1]^d$ with all one-dimensional marginal distributions being uniform on the unit interval $[0,1]$."

will be restated in a more analytical way.
Although the investigation of standardized distribution functions dates back to Hoeffding (1940, 1941), the link to stochastic dependency was not so intuitive in his work. It was not until Sklar's theorem (see Sklar, 1959) that laid the basis for the wide usability and rapidly growing popularity of copulas. In accordance to the original works and inspired by the pedagogical introduction in Nelsen (2006) the results will be given for the bivariate cases and, where a higher-dimensional extension is not straightforward, this will be stated.
To quantify dependence between random variables in a way as tangible as the well-known correlation coefficient by Pearson, some famous association measures will be introduced. These constitute for simple dependence coefficients ranging from -1, corresponding to perfect negative dependence (also called *countermonotonicity*), to 1, corresponding to perfect positive dependence (*comonotonicity*), and are related more closely to copulas. Within this scope we will see that the Fréchet-Hoeffding bounds are not only the lower and upper bounds for the possible magnitude of every copula, but also correspond to the two extremes of dependency. In order to illustrate these concepts further some examples of the most important copula classes as well as suitable simulation techniques will be given.
There will be an extensive account of different estimation techniques in Section 2.2. We will especially focus on maximum-likelihood-related parameter estimates and their asymptotic properties. Since there is rarely pure copula data at hand, the marginal distributions of a given sample of multivariate data have to be dealt with in some way. As it is the work of Genest et al. (1995) that forms the basis of all further steps, we will investigate *canonical maximum likelihood* estimators, their asymptotic variance and suitable variance estimates thoroughly.
Section 2.3 gives a brief overview of *extreme value theory* introducing the famous Fisher-Tippett-Theorem (see Fisher and Tippett, 1928), Fréchet, Weibull and Gumbel extremal distributions and maximum domains of attraction. Furthermore, norming constants for the extrema of normally distributed random variables will be given as well as conditions for weakly dependent normal sequences to belong to the maximum domain of attraction of the Gumbel distribution.

2.1 Basic properties of copulas

We opt to give an introduction to the basic concepts of copula theory from an analytical point of view along the lines of Nelsen (2006). Aiming at a self-contained exposition we give the corresponding proofs or at least a sketch of the principle ideas.

2.1.1 Definitions

The first step towards copulas is a multivariate analogue of a nondecreasing function of one variable. To this end, consider the following definitions.

Definition 2.1.1
Let $S_1, S_2 \subseteq \bar{\mathbb{R}}$ be nonempty subsets of the extended real line $\bar{\mathbb{R}} = [-\infty, \infty]$ and consider the real-valued function $H\colon S_1 \times S_2 \to \mathbb{R}$.

(1) For a rectangle $B = [x_1, x_2] \times [y_1, y_2]$ all of whose vertices lie in $\mathrm{dom}(H)$ the *H-volume of B* is defined as

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1).$$

$V_H(B)$ can be also understood as the *H-mass* of B.

(2) H is called 2-*increasing* if the H-volumes of all rectangles having vertices in $\mathrm{dom}(H)$ are non-negative. In this case, the H-volume of the rectangle B can be regarded as H-measure of B.

(3) If there exist elements $a_1 = \min S_1$ and $a_2 = \min S_2$ such that $H(a_1, y) = 0 = H(x, a_2)$ for all (x, y) in $S_1 \times S_2$, then H is called *grounded*.

We will now bundle these notions and restrict the domain to some subsets of the (bivariate) unit cube to define a *subcopula*.

Definition 2.1.2
A 2-dimensional *subcopula* C' is a real-valued function with the following properties

(i) $\mathrm{dom}(C') = S_1 \times S_2$ where $S_i \subseteq [0,1]$ and $0, 1 \in S_i$ for $i = 1, 2$,

(ii) C' is 2-increasing on $S_1 \times S_2$,

(iii) C' is grounded,

(iv) for every $u \in S_1$ and $v \in S_2$ it holds that $S(u, 1) = u$ and $S(1, v) = v$.

Some basic properties of this mathematical object can be derived right away. After defining a proper copula later on, we will see how these transfer.

Lemma 2.1.3
For a 2-dimensional subcopula C' and $(u, v) \in S_1 \times S_2$ it holds that

(1) C' is non-decreasing in each argument.

(2) $W(u,v) := \max\{u+v-1,0\} \leq C'(u,v) \leq \min\{u,v\} =: M(u,v)$.

(3) $|C'(u_2,v_2) - C'(u_1,v_1)| \leq |u_2-u_1| + |v_2-v_1|$.

Proof

(1) Property (ii) in Definition 2.1.2 implies that a subcopula is 2-increasing so that for all $u_1 \leq u_2 \in S_1$ and $v_1 \leq v_2 \in S_2$

$$C'(u_2,v_2) - C'(u_1,v_2) - C'(u_2,v_1) + C'(u_1,v_1) \geq 0. \qquad (\star)$$

Choose $v_1 = 0$ and fix v_2, then for $u_1 \leq u_2$

$$\begin{aligned} 0 &\leq C'(u_2,v_2) - C'(u_1,v_2) - C'(u_2,0) + C'(u_1,0) \\ &= C'(u_2,v_2) - C'(u_1,v_2) - 0 + 0 \\ &= C'(u_2,v_2) - C'(u_1,v_2), \end{aligned}$$

where the second equality is due to the groundedness of C' (property (iii) in Definition 2.1.2). This implies that C' is increasing in the first component and by symmetry the same is true for the second argument.

(2) Since $u, v \leq 1$ for all $(u,v) \in S_1 \times S_2$ and due to (1) above and (iv) in Definition 2.1.2 it holds that

$$C'(u,v) \leq C'(u,1) = u \quad \text{ as well as } \quad C'(u,v) \leq C'(1,v) = v.$$

Therefore, $C'(u,v) \leq \min\{u,v\}$.

Furthermore, by plugging in $(u_2,v_2) = (1,1)$ and $(u_1,v_1) = (u,v)$ into equation $(\star)$ we obtain

$$\begin{aligned} &C'(1,1) - C'(u,1) - C'(1,v) + C'(u,v) \geq 0 \\ \iff &1 - u - v + C'(u,v) \geq 0 \\ \iff &C'(u,v) \geq u + v - 1. \end{aligned}$$

Groundedness together with (1) ensures that $C'(u,v) \geq C'(0,v) = 0$ for every $v \in S_2$, such that

$$C'(u,v) \geq \max\{u+v-1,0\}.$$

(3) We will show the claim exemplarily for the case $u_1 \leq u_2 \in S_1$ and $v_1 \leq v_2 \in S_2$.

$$\begin{aligned} |C'(u_2,v_2) - C'(u_1,v_1)| &\leq |C'(u_2,v_2) - C'(u_2,v_1)| + |C'(u_2,v_1) - C'(u_1,v_1)| \\ &\overset{(1)}{\leq} |C'(1,v_2) - C'(1,v_1)| + |C'(u_2,1) - C'(u_1,1)| \\ &= |v_2 - v_1| + |u_2 - u_1|. \end{aligned}$$

For the other possible combinations analogous arguments hold.

□

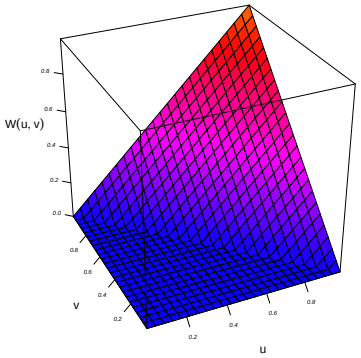
The final step towards the definition of copulas is merely a constraint on the domain.

Definition 2.1.4

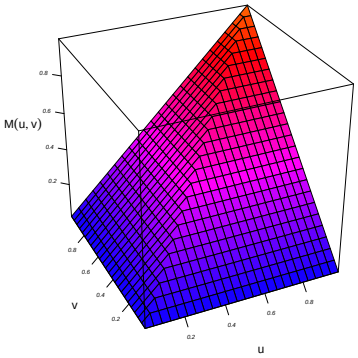
A 2-dimensional subcopula C is called *copula* if $\mathrm{dom}(C) = [0,1]^2$.

2.1.2 Fréchet-Hoeffding bounds and Sklar's Theorem

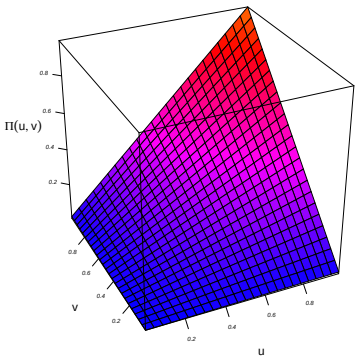
The inequalities in property (2) in Lemma 2.1.3 are known as *Fréchet-Hoeffding-bounds* and set limits for the magnitude of every subcopula. As every copula is a subcopula, the inequalities hold for every copula in every dimension $d \geq 2$. It's noteworthy that the *Fréchet-Hoeffding upper bound* M is itself a copula for every $d \geq 2$, whereas the *Fréchet-Hoeffding lower bound* W is a copula only in the case $d = 2$. Nevertheless, for every $\boldsymbol{u} \in [0,1]^d$ in an arbitrary dimension $d \geq 2$ there exists a copula C such that $C(\boldsymbol{u}) = W_d(\boldsymbol{u})$ where W_d is the d-variate Fréchet-Hoeffding lower bound (see Nelsen, 2006, Thm. 2.10.13).
Furthermore, property (3) in Lemma 2.1.3 establishes the uniform continuity of subcopulas, and hence of copulas, on their domains.



(a) Fréchet-Hoeffding lower bound W



(b) Fréchet-Hoeffding upper bound M



(c) Product copula Π

Figure 2.1: Graphs of the copulas W, M and Π in dimension $d = 2$

Another very important copula is the *independence* or *product copula* $\Pi(u,v) = uv$. The graphs of the bivariate copulas W, M and Π are depicted in Figure 2.1.
Although it is possible to define distribution and joint distribution functions in a general sense, i.e. not necessarily with the connection to random variables (see Nelsen, 2006, p. 17), we will focus on the probabilistic and statistical view from now on. To this end consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with sample space Ω, σ-algebra $\mathcal{F}$ and probability measure $\mathbb{P}$. The considered random variables are bivariate random vectors $(X,Y) : \Omega \to \mathbb{R}^2$ measurable on the defined probability space. In particular, the *joint distribution function* of the random vector is a function $H\colon \bar{\mathbb{R}}^2 \to [0,1]$ with $H(x,y) = \mathbb{P}(X \leq x, Y \leq y)$. We will simply use the term *distribution function* in the sequel. If any of the components x and y is either $-\infty$ or ∞ we set the distribution function to its respective limit.
We define $H(x,\infty) =: F(x) = \mathbb{P}(X \leq x)$ and $H(\infty,y) =: G(y) = \mathbb{P}(Y \leq y)$ as the *(univariate) marginal distribution functions* of X and Y. For matters of convention, we define the distribution functions to be *right-continuous*, i.e. $\lim_{h \downarrow 0} H(x+h, y+h) = H(x,y)$, and analogously for F and G.
Both groundedness of H in the sense of Definition 2.1.1(3) as well as H being 2-increasing as in Definition 2.1.1(2) are immediate consequences. The former is due to $H(-\infty, y) = 0 = H(x, -\infty)$ for all $x, y \in \bar{\mathbb{R}}$, the latter is implied by the fact that every non-empty subset in $\bar{\mathbb{R}}^2$ is assigned a non-negative mass by H. Therefore, the results derived so far directly apply to distribution functions.
Before turning to the most famous theorem in copula theory, recall the notion of *probability integral transforms* for continuous (univariate) random variables. For an absolutely continuous random variable X with distribution function F, it holds that

$$Y := F(X) \sim \mathcal{U}(0,1) \tag{2.1.1}$$

i.e. Y is uniformly distributed on $[0,1]$ with $\mathbb{P}(Y \leq y) = y$ for all $y \in [0,1]$. So, when applying the marginal distribution function to the components of a random vector, we are left with vectors in $[0,1]^2$, and this is where copulas come in.
The word *copula* itself is Latin for *link, tie* or *bond* connecting two different things (Marchant, 2011). In linguistics this term is used to name the part of a sentence that links the subject to the predicate. The specific part of speech depends on the language we consider, but for most cases copulas in this sense are verbs, also called *copular verbs* (for example in English or German *to be/sein* is the main copular verb), or have resemblance to pronouns (for example 是 in the meaning of *that* in Chinese). For further details see Moro (1997).
The copulas in our comprehension are also links – namely links between the univariate marginal distributions and the joint probability distribution of random vectors. Precisely, we have the following theorem.

Theorem 2.1.5 (Sklar, 1959)
Let H be a (2-dimensional) distribution function with (one-dimensional) margins F and G, then there exists a copula C such that for all $x, y \in \bar{\mathbb{R}}$

$$H(x,y) = C(F(x), G(y)). \tag{2.1.2}$$

If the distribution functions F and G are continuous, then C is unique; otherwise C is only uniquely defined on $\operatorname{ran}(F) \times \operatorname{ran}(G)$.
Conversely, given a copula C and two univariate distributions F and G, then $H(x, y)$ defined as in equation (2.1.2) is a bivariate distribution function with margins F and G.

Sketch of proof
As already mentioned above, all properties of (sub-)copulas confer to distribution functions. Especially, the application of the triangle inequality and the fact that H is 2-increasing yield

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)| \tag{2.1.3}$$

for all (x_1, x_2) and (y_1, y_2) in $\bar{\mathbb{R}}^2$. If $F(x_1) = F(x_2)$ and $G(y_1) = G(y_2)$, then obviously $H(x_1, y_1) = H(x_2, y_2)$. Therefore, the set of ordered pairs

$$\left\{ \left((F(x), G(y)), H(x, y) \right) : \ (x, y) \in \bar{\mathbb{R}}^2 \right\}$$

defines a real-valued function C' with domain $\operatorname{ran}(F) \times \operatorname{ran}(G)$. Furthermore, attributes such as being grounded, 2-increasing and having uniform margins carry over from H to C' with the result that C' is a subcopula.
If the marginal distributions $F, G \in \mathcal{C}(\mathbb{R})$, then $\operatorname{ran}(F) = \operatorname{ran}(G) = [0,1]$ and therefore C' is already a copula and also uniquely defined. Otherwise, a copula C as an extension of C' to $[0,1]^2$ can be constructed. This is done by a "bilinear interpolation" which is shown to satisfy the 2-increasing property (ii) in Definition 2.1.2. The constructive details of these steps can be found in Nelsen (2006, Lemma 2.3.5).
The converse case in the theorem is a matter of direct verification of the properties a probability distribution function must satisfy. □

2.1.3 Measures of dependence

Before proceeding to *measures of association*, consider the following remarks. Recall that the *generalized inverse* F^- of a univariate distribution function $F\colon \mathbb{R} \to [0,1]$, which is a (not necessarily strictly) monotone increasing function, is defined as

$$F^-(y) := \inf\{x \in \mathbb{R}\colon F(x) \geq y\} \text{ for all } y \in (0, 1] \tag{2.1.4}$$

with the convention that $\inf \emptyset = \infty$. For $y = 0$ we set $F^-(0) := \sup\{x \in \mathbb{R}\colon F(x) = 0\}$ with $\sup \emptyset = -\infty$. If the inverse in the classical sense exists it equals the generalized inverse. Furthermore, the generalized inverse is non-decreasing and left-continuous.
In Theorem 2.1.5 there is, on the one hand, the construction principle (2.1.2) which forms the basis for many sampling purposes; on the other hand, the following decomposition holds

$$C(u, v) = H(F^-(u), G^-(v)) \tag{2.1.5}$$

for any bivariate distribution H with margins F and G. Above equation (2.1.5) can be regarded as the removal of the margins from a multivariate distribution and allows us

to study dependence and other properties of random vectors independently from their marginal distributions. One can benefit from this separation of dependence structure and marginal distributions especially when it comes to comparing and ordering of random quantities.
In particular, it is noteworthy that the copula of (two) continuous random variables, denoted by $C_{X,Y}$, remains unchanged under strictly increasing transformations. Nelsen (2006, Thm. 2.4.3) shows that if α and β are strictly increasing on $\text{ran}(X)$ and $\text{ran}(Y)$ respectively, then

$$C_{X,Y} = C_{\alpha(X),\beta(Y)}. \tag{2.1.6}$$

For other strictly monotone mappings the resulting copula $C_{\alpha(X),\beta(Y)}$ can also be linked to the initial copula $C_{X,Y}$ of the untransformed random vector (X, Y) (see Nelsen, 2006, Thm. 2.4.4).
In the following paragraphs, various possible dependence measures will be presented. Note that there is a crucial difference between the bivariate concepts - these are the ones we have chosen to concentrate on - and the (generally non-unique) extensions to dimension higher than 2. First ideas for generalizations appeared some 30 years ago (see e.g. Wolff, 1980), but are still an active field of research (cf. Schmid and Schmidt, 2006, 2007a,b; Frahm, 2006). A concise and comprehensive account of copula-based association measures can be found in Schmid et al. (2010).
At first, we will introduce the widely used *correlation coefficient* whose name is ascribed to Pearson (1895), and some of its properties.

Definition 2.1.6
For two random variables X and Y such that $\mathbb{E}\,X^2$, $\mathbb{E}\,Y^2 < \infty$, *Pearson's correlation coefficient* is defined as

$$\rho_P := \rho_P(X,Y) := \frac{\text{Cov}\,(X,Y)}{\sqrt{\text{Var}\,(X)}\sqrt{\text{Var}\,(Y)}} = \frac{\mathbb{E}\,((X-\mathbb{E}\,(X))(Y-\mathbb{E}\,(Y)))}{\sqrt{\mathbb{E}\,((X-\mathbb{E}\,(X))^2)}\sqrt{\mathbb{E}\,((Y-\mathbb{E}\,(Y))^2)}}. \tag{2.1.7}$$

We will use the term *correlation* in the meaning of Pearson's correlation coefficient.

We summarize some well-known properties of the correlation coefficient given without proof. However, we comment on these properties subsequently.

Proposition 2.1.7
Let X and Y be random variables with finite second moments and correlation ρ_P. Then

(1) $-1 \leq \rho_P \leq 1$

(2) if X and Y are independent, then $\rho_P = 0$. The converse is generally not true. A special case in which uncorrelatedness is equivalent to independence is that of X and Y being jointly normally distributed.

(3) if $|\rho_P| = 1$, then $Y = c + mX$ almost surely, with constants $c, m \in \mathbb{R}$, $m \neq 0$.

(4) $\rho_P(c_X + m_X X, c_Y + m_Y Y) = \rho_P(X,Y)$ for $m_X, m_Y, c_X, c_Y \in \mathbb{R}$, $m_X, m_Y \neq 0$.

Property (4) implies that the correlation coefficient is invariant under strictly increasing *linear* transformations, but it is *not* under nonlinear ones. Take into account that in case of continuous marginal distributions F and G it holds that

$$\begin{aligned}\operatorname{Cov}(X,Y) &= \iint_{\mathbb{R}^2} (x - \mathbb{E}(X))(y - \mathbb{E}(Y))\, dH(x,y) \\ &= \iint_{[0,1]^2} (F^-(u) - \mathbb{E}(X))(G^-(v) - \mathbb{E}(Y))\, dC(u,v),\end{aligned}$$

where the latter equality is justified by the measure transformation dH on $\mathbb{R}^2$ to dC on $[0,1]^2$ implied by the mapping $(x,y)^\top \mapsto (F(x), G(y))^\top$. Then it becomes obvious that the correlation depends on both, the copula structure of the given random variables as well as on their marginal distributions.
This is a drawback to the effect that the actual range of ρ_P is a (usually strict) subset of the maximal range between -1 and 1 as stated in property (1). The attainable correlation again depends highly on the copula and the margins of the random vector (X,Y). Further pitfalls in the use of Pearson's correlation are described in McNeil et al. (2005, Sect. 2.5.1).
A better concept to measure dependence is that of *concordance*. In Joe (1997, p. 38) concordance measures as means of ordering bivariate distributions are introduced in an axiomatic way. We confine ourselves to mention the two most popular measures *Spearman's correlation coefficient* and *Kendall's tau* and their representation in terms of the copula of the random variables in the following proposition.

Proposition 2.1.8
For a given copula C and univariate distributions F and G, let (X,Y) and (X',Y') be independent random vectors distributed according to $H = C(F,G)$. Consider the definitions of *Kendall's tau*

$$\tau := \tau(X,Y) := \mathbb{E}\left(\operatorname{sgn}((X - X')(Y - Y'))\right), \tag{2.1.8}$$

and *Spearman's rho*

$$\rho := \rho(X,Y) := \rho_P(F(X), G(Y)). \tag{2.1.9}$$

If the marginal distributions F and G of the supposed bivariate distribution H are continuous, then the following representations in terms of the corresponding copula hold

$$\tau = 4 \iint_{[0,1]^2} C(u,v)\, dC(u,v) - 1 \tag{2.1.10}$$

and

$$\rho = 12 \iint_{[0,1]^2} C(u,v)\, du\, dv - 3. \tag{2.1.11}$$

In this case we will denote Kendall's tau and Spearman's rho as functions of the respective copula C as follows

$$\tau(X,Y) = \tau_C \quad \text{and} \quad \rho(X,Y) = \rho_C.$$

Proof

(1) From equation (2.1.8) above and the fact that $\operatorname{sgn}(a) = 1$ if $a > 0$ and $\operatorname{sgn}(a) = -1$ if $a < 0$ for any real number a, we have

$$\begin{aligned}\tau &= \mathbb{P}\left((X - X')(Y - Y') > 0\right) - \mathbb{P}\left((X - X')(Y - Y') < 0\right) \\ &= 2\,\mathbb{P}\left((X - X')(Y - Y') > 0\right) - 1\end{aligned}$$

Due to the interchangeability of (X, Y) and (X', Y'), it holds that

$$\mathbb{P}\left((X - X')(Y - Y') > 0\right) = 2\,\mathbb{P}\left(X > X', Y > Y'\right)$$

and therefore we continue the above chain of equalities

$$\begin{aligned}&= 4\,\mathbb{P}\left(X < X', Y < Y'\right) - 1 \\ &= 4 \iint_{\mathbb{R}^2} \mathbb{P}\left(X < x, Y < x\right)\, dH(x, y) - 1 \\ &= 4 \iint_{[0,1]^2} \mathbb{P}\left(X < F^-(u), Y < G^-(v)\right)\, dC(u, v) - 1 \\ &= 4 \iint_{[0,1]^2} C(u, v)\, dC(u, v) - 1,\end{aligned}$$

where the second equality can be regarded as a continuous version of the law of total probability and the third one is a consequence of substituting $x = F^-(u)$ and $y = G^-(v)$.

(2) Set $U := F(X), V := G(Y)$ and recall that both U and V are uniformly distributed on [0,1] due to the probability integral transform (2.1.1). Therefore, $\mathbb{E}(U) = \mathbb{E}(V) = \frac{1}{2}$ and $\operatorname{Var}(U) = \operatorname{Var}(V) = \frac{1}{12}$ and

$$\begin{aligned}\rho_S(X, Y) &= \rho(F(X), G(Y)) = \rho(U, V) \\ &= \frac{\operatorname{Cov}(U, V)}{\sqrt{\operatorname{Var}(U)}\sqrt{\operatorname{Var}(V)}} \\ &= 12 \iint_{[0,1]^2} \left(x - \tfrac{1}{2}\right)\left(y - \tfrac{1}{2}\right) dC(x, y)\end{aligned}$$

Using $\iint x\, dC(x, y) = \iint y\, dC(x, y) = \frac{1}{2}$ and $xy = \int_0^x \int_0^y 1\, du\, dv$ we obtain further

$$\begin{aligned}&= 12 \iint_{[0,1]^2} xy\, dC(x, y) + 12\left(-\tfrac{1}{4} - \tfrac{1}{4} + \tfrac{1}{4}\right) \\ &= 12 \iint_{[0,1]^2} \left(\iint_{[0,x]\times[0,y]} du\, dv \right) dC(x, y) - 3 \\ &= 12 \iint_{[0,1]^2} \left(\iint_{[u,1]\times[v,1]} dC(x, y) \right) du\, dv - 3 \\ &= 12 \iint_{[0,1]^2} C(u, v)\, du\, dv - 3.\end{aligned}$$

The second last equality is a consequence of interchanging the double integrals due to Tonelli's theorem, whereas in the last equality we can replace $1-u$ by u and $1-v$ by v respectively as the outer integration is over the whole unit cube $[0,1]^d$.

□

With this, we have now dependence measures at hand that (in case of continuous marginal distributions) only depend on the copula of a given random vector, i.e. on the inherent dependence structure which can easily be separated from the margins by Sklar's Theorem. Therefore, Kendall's tau and Spearman's rho inherit the property of being invariant under strictly increasing transformations (see equation (2.1.6)).
Generally, they are symmetric and take values in $[-1,1]$. Now we will see which copulas are implied by the extremes -1 and 1, and that it is possible to construct a bivariate distribution with any desired dependency in the range of τ or ρ.

Proposition 2.1.9
Let X and Y be continuous random variables with copula C and let κ be either $\tau(X,Y)$ or $\rho(X,Y)$. Then

(1) $\kappa = 1 \iff C = M \iff F(X) = G(Y)$ almost surely

(2) $\kappa = 0 \iff C = \Pi$

(3) $\kappa = -1 \iff C = W \iff F(X) = 1 - G(Y)$ almost surely

Proof
Showing that a special form of the copula implies τ and ρ attaining the given values is verified by simply calculating the integrals as given in Proposition 2.1.8.
For $\kappa = \rho$ the proof of the converse is a matter of straightforward verification of the claims taking advantage of definition by Pearson's correlation coefficient.
For $\kappa = \tau$ the converse is shown in Hofert (2010, Prop. 1.7.9). □

It is now clear that it makes sense to name the Fréchet-Hoeffding lower bound W the *countermonotonicity* and the upper bound M the *comonotonicity* copula. Since every convex combination of copulas is a copula itself (see Nelsen, 2006, Exercise 2.3), by

$$C(u,v) := \lambda W(u,v) + (1-\lambda)M(u,v)$$

we can define a copula $\forall \lambda \in [0,1]$ such that the respective association measures are

$$\tau_C = \rho_C = 1 - 2\lambda.$$

When considering extremal values of dependence, the notion of *tail dependence* must not be left unmentioned, i.e. the strength of dependence in the tails of a bivariate distribution.

Definition 2.1.10
Let X and Y be continuous random variables with marginal distributions F and G. In case the limits λ_l and λ_u exist

$$\lambda_l := \lim_{t \downarrow 0} \mathbb{P}\left(Y \leq G^-(t) \,|X \leq F^-(t)\right), \tag{2.1.12}$$

and

$$\lambda_u := \lim_{t\uparrow 1} \mathbb{P}\left(Y > G^-(t) \,|X > F^-(t)\right), \tag{2.1.13}$$

they are called the *coefficients of lower* and *upper tail dependence* respectively.
If $\lambda_l = 0$, then X and Y are called *lower tail-independent*; if $\lambda_u = 0$, then X and Y are *upper tail-independent*.

Using simple conditional probability, there is again a representation in terms of the copula C of the random variables (if the respective limits exist)

$$\lambda_l = \lim_{t\downarrow 0} \frac{C(t,t)}{t} \tag{2.1.14}$$

$$\lambda_u = 2 - \lim_{t\uparrow 1} \frac{1-C(t,t)}{1-t} \tag{2.1.15}$$

such that calculations are straightforward if the copula in question has a tractable closed form.
Furthermore, by Taylor expansion of $C(t,t)$ around $(0,0)$ it can be shown that in case all partial derivatives and the respective limits exist the following representations also hold true:

$$\lambda_l = \lim_{t\downarrow 0} \left(\frac{\partial}{\partial u} C(u,v) + \frac{\partial}{\partial v} C(u,v) \Big|_{u=v=t} \right) \tag{2.1.16}$$

$$\lambda_u = 2 - \lim_{t\uparrow 1} \left(\frac{\partial}{\partial u} C(u,v) + \frac{\partial}{\partial v} C(u,v) \Big|_{u=v=t} \right). \tag{2.1.17}$$

In Appendix A some famous examples of parametric copula families will be introduced. At that point, we will also give the respective tail dependence coefficients.

2.1.4 General sampling techniques

Let us now turn to simulation algorithms in order to sample random variates from a specified joint distribution. In case of a univariate distribution F, the most common procedure to obtain random variates distributed according to F relies on the inversion of the given distribution function and therefore it is called *inversion method*. A proof can be found in Hofert (2010, Prop. 1.1.6/1.1.7(2)).

Algorithm 2.1.11 (Inversion Method)
In order to generate a sample of the random variable $X \sim F$ proceed as follows

(1) Generate $u \sim \mathcal{U}(0,1)$

(2) $x := F^-(u)$ is the desired random variate

When a uniformly distributed sample is present, it suffices to transform it in the way described above to arrive at the desired (marginal) distribution. By virtue of Sklar's Theorem, we need only to generate a random pair (u,v) of uniform variates whose joint distribution is a copula C at first and then we attain a sample (x,y) of a random pair

$(X, Y) \sim H = C(F, G)$ by setting $(x, y) := (F^-(u), G^-(v))$. The most general procedure for the first step is the *conditional distribution method*. For a random pair $(U, V) \sim C$ consider the conditional distribution of V given $U = u$

$$c_u(v) := \mathbb{P}(V \leq v \,|U = u) = \lim_{t \downarrow 0} \frac{C(u+t, v) - C(u, v)}{t} = \frac{\partial}{\partial u} C(u, v). \tag{2.1.18}$$

Nelsen (2006, Thm. 2.2.7) shows that $\frac{\partial}{\partial u} C(u, v)$ exists and is nondecreasing almost everywhere on $[0,1]^2$. Hence the following approach will do the desired sampling.

Algorithm 2.1.12 (Conditional Distribution Method)
A sample of $(U, V) \sim C$ is obtained as follows

(1) Generate u and t uniform on [0,1] independently

(2) Set $v := c_u^-(t)$, where c_u^- is the generalized inverse of c_u as defined in (2.1.4)

(3) (u, v) is the desired pair

For higher dimensions $d > 2$ the connection of conditional copulas to partial derivatives holds only under the assumption of continuous partial derivatives (cf. Schmitz, 2003, Sect. 2.2) and allows for an analogous simulation algorithm (see Hofert, 2010, Thm. 1.8.1).

2.1.5 Some copula classes

In the following introductory review of the three most famous copula families of *elliptical*, *Archimedean* and *Marshall-Olkin* copulas we will focus on the definitions and on special sampling procedures. An overview of the copulas used in this thesis, especially regarding the analytic forms of copula and density, concordance and tail dependence measures and contour plots of samples, is given in Appendix A.
Elliptical copulas are examples of copulas derived from multivariate distributions and not having an explicit analytical form. First, we need to introduce *elliptical distributions*.

Definition 2.1.13
A d-dimensional random vector $\boldsymbol{X}$ has an *elliptical distribution* if there exists a *location vector* $\boldsymbol{\mu} \in \mathbb{R}^d$, a symmetric, positive semidefinite *dispersion matrix* $C \in \mathbb{R}^{d \times d}$ and a *characteristic generator* function $\phi\colon [0, \infty) \to \mathbb{C}$ such that

$$\mathbb{E}\left(\exp\left\{\mathrm{i}\, t^\top (\boldsymbol{X} - \boldsymbol{\mu})\right\}\right) = \phi(t^\top C t) \quad \text{for all } t \in \mathbb{R}^d,$$

i.e. the characteristic function of $\boldsymbol{X} - \boldsymbol{\mu}$ can be expressed in terms of the characteristic generator and the dispersion matrix. In this case we write $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, C, \phi)$.

By linearity of elliptical distributions (see Hofert, 2010, Prop. 1.9.3), for $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, C, \phi)$ it holds that $\boldsymbol{X} - \boldsymbol{\mu} \stackrel{d}{=} \boldsymbol{\mu} - \boldsymbol{X}$, i.e. every elliptical random vector is *radially symmetric* around its location vector $\boldsymbol{\mu}$. This results in elliptic contours as we will see in course of presenting some examples in Appendices A.1 and A.2. This fact coined the term *elliptical*.

Example 2.1.14 (Gaussian Copula)
For a covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$ let P be the corresponding correlation matrix. The d-variate *Gaussian copula* is defined for all $\boldsymbol{u} = (u_1, \dots, u_d) \in \mathbb{R}^d$ as

$$\begin{aligned} C_{\mathrm{Ga}}(\boldsymbol{u}; \Sigma) &= \Phi_{d,P}\left(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)\right) \\ &= \int_{-\infty}^{\Phi^{-1}(u_d)} \dots \int_{-\infty}^{\Phi^{-1}(u_1)} \frac{1}{\sqrt{(2\pi)^d \det(P)}} \exp\left\{-\frac{\boldsymbol{x}^\top P^{-1} \boldsymbol{x}}{2}\right\} dx_1 \dots dx_d \end{aligned} \tag{2.1.19}$$

where Φ^{-1} is the quantile function of the univariate $\mathcal{N}(0,1)$ distribution and $\Phi_{d,P}$ is the distribution function of the d-variate $\mathcal{N}(\mathbf{0}, P)$ with mean vector $\mathbf{0}$ and covariance matrix P which corresponds to $E_d(\mathbf{0}, P, \mathrm{e}^{-t/2})$ in terms of elliptical distributions.

As outlined above, we will apply Algorithm 2.1.11 on copula data obtained by Algorithm 2.1.12 as usually we are interested in sampling from a given joint distribution. In case of implicitly given copulas we are only able to sample directly from the multivariate distribution and then have to remove the marginal distributions by a probability integral transform to obtain pure copula data. We obtain the following algorithm.

Algorithm 2.1.15 (Gaussian Copula)
To sample from $\boldsymbol{U} \sim C_{\mathrm{Ga},\Sigma}$ consider the following:

(1) Determine the Cholesky square-root A of P, i.e. $P = AA^\top$

(2) Sample univariate independent $z_1, \dots, z_d \sim \mathcal{N}(0,1)$ and consider $\boldsymbol{z} = (z_1, \dots, z_d)$

(3) Set $\boldsymbol{x} := A\boldsymbol{z}^\top$

(4) $\boldsymbol{u} = (\Phi(x_1), \dots, \Phi(x_d))$ is the desired vector

In the bivariate case, it can be shown that $\lambda_l = \lambda_u = 0$ for any covariance structure Σ (see Appendix A.1). In our next example, the *Student copula*, at least this drawback can be removed.

Example 2.1.16 (Student Copula)
Let P be again the correlation matrix corresponding to a covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$. For all $\boldsymbol{u} = (u_1, \dots, u_d) \in \mathbb{R}^d$ the d-variate *Student* or t_ν*-copula* is defined as

$$\begin{aligned} C_{t_\nu}(\boldsymbol{u}; \Sigma) &= t_{\nu,P}\left(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d)\right) \\ &= \int_{-\infty}^{t_\nu^{-1}(u_d)} \dots \int_{-\infty}^{t_\nu^{-1}(u_1)} \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\nu\pi)^d \det(P)}} \left(1 + \frac{\boldsymbol{x}^\top P^{-1} \boldsymbol{x}}{\nu}\right)^{-\frac{\nu+d}{2}} dx_1 \dots dx_d \end{aligned} \tag{2.1.20}$$

where t_ν^{-1} is the quantile of the univariate t_ν-distribution with ν degrees of freedom and $t_{\nu,P}$ is the distribution function of the d-variate centered $t_{\nu,P}$-distribution with $\nu \in (0, \infty)$ degrees of freedom and correlation structure P. It corresponds to $E_d(\mathbf{0}, P, \mathbb{E}\left(\exp\left\{-Wt/2\right\}\right))$ where $W \sim \chi^2_\nu$.

The last comment gives rise to a stochastic representation of a multivariately t_ν-distributed random variable $\boldsymbol{Y}$ with mean μ in terms of a normally distributed $\boldsymbol{X} \sim \mathcal{N}(\mu, \Sigma)$ and a univariate $W \sim \chi^2_\nu$ independent of the former, namely

$$\boldsymbol{Y} \overset{d}{=} \frac{\boldsymbol{X}}{\sqrt{W/\nu}} \tag{2.1.21}$$

where the division is understood component-wise. With Algorithm 2.1.15 at hand we can easily set up the following.

Algorithm 2.1.17 (Student Copula)
A sample from $\boldsymbol{V} \sim C_{t_\nu,\Sigma}$ can be generated as follows:

(1) Sample $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, P)$

(2) Sample $w \sim \chi^2_\nu$ independent of $\boldsymbol{x}$ and set $s := \nu/w$

(3) Set $\boldsymbol{y} := \sqrt{s}\,\boldsymbol{x}$

(4) $\boldsymbol{u} = (t_\nu^{-1}(y_1), \ldots, t_\nu^{-1}(y_d))$ provides the desired sample

As shown by Demarta and McNeil (2004) the lower and upper tail dependence coefficients of the t_ν-copula are non-zero if the correlation of the components is non-zero. This and some illustrations are deferred to Appendix A.2.

In Marshall and Olkin (1967) a multivariate extension of the univariate exponential distribution was introduced. In the study of lifetimes, the so-called survival analysis, the exponential model is the most widely used parametric approach to survival data; this is mostly due to the fact that the exponential distribution is the only continuous distribution that exhibits the *lack of memory* property, i.e. the probability of surviving a certain time point $t + s$ given survival beyond time t equals the probability of surviving beyond s. The multivariate exponential distribution mentioned above is shown to be the only possible generalization to higher dimensions that remains memoryless in above sense.
Since we are aiming at a possible simulation algorithm in the end, we will introduce the resulting copula in a constructive way which also gives an intuitive interpretation right away. Consider a system of two components that are affected by (fatal) shocks following independent Poisson processes. The two shocks that hit the components individually are then $Z_1 \sim \text{Exp}(\lambda_1)$ and $Z_2 \sim \text{Exp}(\lambda_2)$ and the shock being fatal for both simultaneously is $Z_{12} \sim \text{Exp}(\lambda_{12})$ with $\lambda_1, \lambda_2, \lambda_{12} > 0$. For the joint survival probability of the individual lifetimes X and Y of the system we get

$$\begin{aligned}\bar{H}(s,t) &:= \mathbb{P}(X > s, Y > t) \\ &= \mathbb{P}(Z_1 > s, Z_2 > t, Z_{12} > \max\{s,t\}) \\ &= \exp\{-\lambda_1 s - \lambda_2 t - \lambda_{12}\max\{s,t\}\}.\end{aligned}$$

So far we have only considered the joint survival distribution of (X, Y). Notice that $X = \min\{Z_1, Z_{12}\}$ and $Y = \min\{Z_2, Z_{12}\}$ and since the exponential distribution is closed

under taking minima it is obvious that the marginal distributions of the components must be exponential, with parameters $\lambda_1 + \lambda_{12}$ and $\lambda_2 + \lambda_{12}$ respectively. Therefore, $\bar{F}(s) := \mathbb{P}(X > s) = \mathrm{e}^{-(\lambda_1+\lambda_{12})s}$ and $\bar{G}(t) := \mathbb{P}(Y > t) = \mathrm{e}^{-(\lambda_2+\lambda_{12})t}$ are the marginal survival distributions.
There is an analogue of Sklar's Theorem for multivariate survival functions (see e.g. McNeil et al., 2005 or Mai, 2010), which in the 2-dimensional case states that for every survival function $\bar{H}(s,t)$ of a random pair (X,Y) with marginal survival functions $\bar{F}(s)$ and $\bar{G}(t)$ there exists a proper copula $\check{C}$ such that

$$\bar{H}(s,t) = \check{C}(\bar{F}(s), \bar{G}(t)); \tag{2.1.22}$$

and that also the converse holds true, i.e. by $\check{C}(u,v) = \bar{H}(\bar{F}^-(u), \bar{G}^-(v))$ a copula $\check{C}$ is properly defined. $\check{C}$ is called the (in case of continuous margins unique) *survival copula* of the copula C and it is essential to stress that $\check{C}$ is a distribution function, namely that of $(1-U, 1-V)$ if $(U,V) \sim C$, and not a survival function.
We define $\alpha_i := \frac{\lambda_{12}}{\lambda_i+\lambda_{12}}$, $i = 1,2$ and as a consequence of Sklar's Theorem for survival distributions we obtain the following definition of the two-parameter family of *Marshall-Olkin copulas*.

Definition 2.1.18 (Marshall-Olkin Copula)
Let $\alpha_i \in [0,1]$, $i = 1,2$. A *Marshall-Olkin copula* is given for all $u, v \in [0,1]$ by

$$C_{\mathrm{MO}}(u,v;\alpha_1,\alpha_2) = \min\left\{vu^{1-\alpha_1}, uv^{1-\alpha_2}\right\}. \tag{2.1.23}$$

For $\alpha_1 = \alpha_2$ the resulting copula is called *Cuadras-Augé copula* (see Cuadras and Augé, 1981). Furthermore, $C_{\mathrm{MO};0\,\alpha_2} = C_{\mathrm{MO};\alpha_1 0} = \Pi$ and $C_{\mathrm{MO};11} = M$.

Note that Marshall-Olkin copulas possess both an absolutely continuous component for $u^{\alpha_1} \neq v^{\alpha_2}$ and a singular component with mass $\mathbb{P}(U^{\alpha_1} = V^{\alpha_2}) = \alpha_1\alpha_2/(\alpha_1+\alpha_2-\alpha_1\alpha_2)$ (the notions of absolutely continuous and singular components will be clarified in Section 2.2.2). Moreover, it can be shown that $\lambda_l = 0$ and $\lambda_u = \min\{\alpha_1, \alpha_2\}$ for $\alpha_1 \neq \alpha_2$, i.e. Marshall-Olkin copulas allow only for upper tail dependence.
For simulation purposes consider the following.

Algorithm 2.1.19 (Marshall-Olkin Copula)
We obtain a realization of $(U,V) \sim C_{\mathrm{MO};\alpha_1\alpha_2}$ with $\alpha_i = \frac{\lambda_{12}}{\lambda_i+\lambda_{12}}$ by

(1) Sample $z_j \sim \mathrm{Exp}(\lambda_j)$ for $j \in \{1,2,12\}$

(2) Set $x := \min\{z_1, z_{12}\}$ and $y := \min\{z_2, z_{12}\}$

(3) Set $u := \mathrm{e}^{-(\lambda_1+\lambda_{12})x}$ and $v := \mathrm{e}^{-(\lambda_2+\lambda_{12})y}$

(4) (u,v) is the desired sample

A recent account on Marshall-Olkin copulas can be found in the dissertation of Mai (2010) and numerous publications surrounding it (see e.g. Mai and Scherer, 2009). In Appendix A.3 we restrict ourselves to some illustrations of possible forms of the Marshall-Olkin copula.

Now, we proceed to the class of *Archimedean copulas* that is possibly the most famous copula class with the greatest range of applications. Archimedean copulas are popular mainly due to their easy construction principle (cf. Definition 2.1.20 below), the great variety of copula families that fall into this class and the possibility to capture many desirable properties that e.g. real-world data exhibit.

Definition 2.1.20 (Archimedean Copula)
A function $\varphi\colon [0,1] \to [0,\infty]$ is called *Archimedean generator* if $\varphi(1) = 0$ and φ is continuous and strictly decreasing on [0,1]. Set

$$\varphi^{[-1]}(y) = \begin{cases} \varphi^{-1}(y), & 0 \leq y \leq \varphi(0), \\ 0, & \varphi(0) \leq y \leq \infty. \end{cases} \tag{2.1.24}$$

as the *pseudo-inverse* of φ such that $\mathrm{dom}(\varphi^{[-1]}) = [0,\infty]$. If $\varphi(0) = \infty$, then the generator is called *strict* and the pseudo-inverse complies with the usual inverse.
A d-dimensional copula, $d \geq 2$, is an *Archimedean copula* if for every $(u_1,\ldots,u_d) \in \mathbb{R}^d$ the following representation with an Archimedean generator holds

$$C(u_1,\ldots,u_d) = \varphi^{[-1]}(\varphi(u_1) + \ldots + \varphi(u_d)). \tag{2.1.25}$$

Especially for higher dimensional models that are considered in more recent publications (see e.g. McNeil and Nešlehová, 2009 or Hofert et al., 2012), the roles of φ and $\varphi^{[-1]}$ are reversed. In this case, an Archimedean copula is defined to allow for the representation

$$C(u_1,\ldots,u_d) = \psi(\psi^{-1}(u_1) + \ldots + \psi^{-1}(u_d)) \tag{2.1.26}$$

with a continuous, strictly decreasing function $\psi\colon [0,\infty] \to [0,1]$ such that $\psi(0) = 1$ which then plays the role of the Archimedean generator. It is often for notational convenience that ψ is preferred over φ.
However, we will stick to the notation of the classical works such as Genest and MacKay (1986a) or Nelsen (2006) and translate the recent results to this language. Notice that in the definition above it is only stated that a copula with representation (2.1.25) is said to be Archimedean, but there is no statement about which generators φ actually generate proper copulas. We will approach a general answer to this question in three steps providing subsequently the following theorems.
The conditions for φ generating a 2-dimensional copula emerged in the course of studying *probabilistic metric spaces*, *associative functions* and *t-norms* and is due to Schweizer and Sklar. The theorem below corresponds to Theorem 4.1.4 in Nelsen (2006, p. 111) the proof of which is based on Alsina et al. (2006).

Theorem 2.1.21 (Schweizer and Sklar, 1963)
With an Archimedean generator φ as introduced in Definition 2.1.20, the two-place function

$$C(u,v) := \varphi^{[-1]}(\varphi(u) + \varphi(v)) \tag{2.1.27}$$

is a bivariate copula if, and only if, φ is convex.

The next theorem gives conditions on the Archimedean generator such that equation (2.1.25) defines a d-variate copula in any dimension $d \geq 2$.

Theorem 2.1.22 (Kimberling, 1974)
A function C as in (2.1.25) is a proper d-variate copula for any $d \geq 2$ if, and only if, for the Archimedean generator φ it holds that

(i) its pseudo-inverse $\varphi^{[-1]}$ is infinitely often differentiable on $(0, \infty)$

(ii) and $(-1)^k \frac{d^k}{dt^k} \varphi^{[-1]}(t) \geq 0$ for all $k = 0, 1, 2, \ldots$ and for all $t \in (0, \infty)$.

Then $\varphi^{[-1]}$ is called *completely monotone*.

This result can be generalized in the sense that the dimension d of the desired copula is fixed. It is the most recent and due to McNeil and Nešlehová.

Theorem 2.1.23 (McNeil and Nešlehová, 2009)
Let $d \geq 2$ be given and let C be a function via (2.1.25). C is a d-variate copula if, and only if, for the Archimedean generator φ it holds that

(i) its pseudo-inverse $\varphi^{[-1]}$ is $(d-2)$-times differentiable on $(0, \infty)$,

(ii) $(-1)^k \frac{d^k}{dt^k} \varphi^{[-1]}(t) \geq 0$ for all $k = 0, 1, \ldots, (d-2)$ and for all $t \in (0, \infty)$

(iii) and $(-1)^{(d-2)} \frac{d^{(d-2)}}{dt^{(d-2)}} \varphi^{[-1]}(t)$ is decreasing and convex.

Then $\varphi^{[-1]}$ is called *d-monotone*.

Most of the time, we will consider *parametric* copula families, i.e. the copulas are of the form $C(\ldots; \theta)$. Archimedean copulas in that sense are hence functions of parameters θ in a parameter space Θ through a parametric generator denoted by φ_θ. Some examples of parametric Archimedean copulas as well as closely related copulas are given in Appendices A.5 to A.8, many more together with their respective generators can be found in Nelsen (2006, Tab. 4.1).
Concerning sampling purposes various approaches were proposed, for concise overviews see e.g. Hofert (2008, 2011). Here, we will confine ourselves to presenting only three procedures for the Archimedean copulas introduced above. The used construction principle leads to multivariate copulas that are *exchangeable*. We refer to a copula C as exchangeable if it is the distribution function of an exchangeable random vector $(U_1, \ldots, U_d)$ of uniform variates with joint distribution C such that for all permutations $\pi \colon \{1, \ldots, d\} \to \{1, \ldots, d\}$ it holds that

$$(U_1, \ldots, U_d) \stackrel{d}{=} (U_{\pi(1)}, \ldots, U_{\pi(d)}).$$

In particular, for all such permutations it holds that $C(u_1, \ldots, u_d) = C(u_{\pi(1)}, \ldots, u_{\pi(d)})$. Exchangeability is often perceived as a severe drawback in modeling empirical data. Recently, there has been a lot of research towards more suitable dependence structures which still take advantage of the easy construction principle of Archimedean copulas.

Worth noting within this scope are *nested* or *hierarchical Archimedean copulas* (see McNeil, 2008; Okhrin et al., 2009; Hering et al., 2010) or *d-vine copulas* (see Kurowicka and Joe, 2011).
Let us start with an adaptation of the conditional distribution method in Algorithm 2.1.12 to suit the framework of Archimedean generators. In the bivariate setting recall equation (2.1.18) in terms of φ

$$c_u(v) = \frac{\partial}{\partial u} C(u,v) = \frac{\partial}{\partial u} \varphi^{[-1]}(\varphi(u) + \varphi(v)) = \frac{\varphi'(u)}{\varphi'(\varphi^{[-1]}(\varphi(u) + \varphi(v)))}. \tag{2.1.28}$$

We have avoided to differentiate $\varphi^{[-1]}$, but used the formula $(f^{-1})' = 1/f'(f^{-1})$ for derivatives of inverses. The sampling procedure that results herefrom has already been proposed in Genest and MacKay (1986a, Chap. 6). Let $\varphi'^{(-1)}$ denote the (generalized) inverse of the first derivative of the Archimedean generator φ.

Algorithm 2.1.24
In order to generate a sample from $(U,V) \sim C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ proceed as follows.

(1) Generate $u, t \sim \mathcal{U}(0,1)$ independently

(2) Set $w := \varphi'^{(-1)}(\varphi'(u)/t)$

(3) Set $v := \varphi^{[-1]}(\varphi(w) - \varphi(u))$

(4) (u,v) is the desired sample

Another possibility to obtain samples for Archimedean copulas is based on the following proposition due to Genest and Rivest (1993).

Proposition 2.1.25
Let $(U,V) \sim C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ and consider $K_C(t) := \mathbb{P}(C(U,V) \leq t)$, $t \in [0,1]$.

(1) Then $K_C(t) = t - \dfrac{\varphi(t)}{\varphi'(t^+)}$, where $t^+ = \lim\limits_{s \downarrow t} s$.

(2) The joint distribution of

$$S = \frac{\varphi(U)}{\varphi(U) + \varphi(V)} \quad \text{and} \quad T = C(U,V)$$

is $H(s,t) := \mathbb{P}(S \leq s, T \leq t) = s \cdot K_C(t)$.

Proof

(1) See Theorem 4.3.4 in Nelsen (2006, p. 127).

(2) See Theorem 4.3.7 in Nelsen (2006, p. 129).

□

Note that item (2) above implies that S is uniformly distributed on [0,1] and that S and T are independent. With $K_C^{(-1)}(t)$ the (generalized) inverse of K_C as given by (1) above the following algorithm results.

Algorithm 2.1.26
In order to generate a sample from $(U,V) \sim C(u,v) = \varphi^{[-1]}(\varphi(u)+\varphi(v))$ proceed as follows.

(1) Generate $s,t \sim \mathcal{U}(0,1)$ independently

(2) Set $w := K_C^{(-1)}(t)$

(3) Set $u := \varphi^{[-1]}(s\,\varphi(w))$ and $v := \varphi^{[-1]}((1-s)\varphi(w))$

(4) (u,v) is the desired pair

A remarkable result of Bernstein (1929) links completely monotone functions to Laplace-Stieltjes transforms of certain distributions. The following theorem corresponds to Theorem 1 in Feller (1971, p. 439) where also a proof can be found.

Theorem 2.1.27 (Bernstein, 1929)
A function $\psi\colon [0,\infty] \to [0,1]$ with $\psi(0) = 1$ is completely monotone if, and only if, there exists a non-negative random variable W on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution function F_W such that for all $t \geq 0$

$$\psi(t) = \mathbb{E}\,\mathrm{e}^{-tW} = \int_0^\infty \mathrm{e}^{-tw}\, dF_W(w), \tag{2.1.29}$$

i.e. ψ is the *Laplace-Stieltjes transform* of W.

By setting $\varphi^{[-1]} = \psi$, a simple link of Archimedean generators to the distribution of W can be established. Based on this fact, a sampling algorithm was introduced by Marshall and Olkin (1988). It is especially useful when we know the distribution function F_W for which a given generator inverse $\varphi^{[-1]}$ is the Laplace-Stieltjes transform.
It is noteworthy that sampling Archimedean copulas according to the following algorithm is very efficient and particularly suitable for multivariate purposes as the dimension only matters for the count of iid uniform samples.

Algorithm 2.1.28 (Marshall and Olkin, 1988)
A sample of $\boldsymbol{U} \sim C(u_1,\ldots,u_d) = \varphi^{[-1]}(\varphi(u_1)+\ldots+\varphi(u_d))$ can be generated as follows.

(1) Sample $w \sim F_W$ where $\varphi^{[-1]}(t) = \int_0^\infty \mathrm{e}^{-tw}\, dF_W(w)$

(2) Generate $x_1,\ldots,x_d \sim \mathcal{U}(0,1)$ independently and consider $\boldsymbol{x} = (x_1,\ldots,x_d)$

(3) Set $\boldsymbol{u} = \varphi^{[-1]}\left(-\log \boldsymbol{x}/w\right)$ to obtain the desired sample

For further sampling strategies see the overview in Hering (2011, Sect. 3.1.2) or Hofert (2010). They present procedures for the more sophisticated models of nested Archimedean copulas as well.

As a last point in this section we will show how some of the concordance measures and the tail dependence coefficients introduced in Section 2.1.3 can be expressed in terms of the Archimedean generator. This is often more convenient when it comes to computations.

Proposition 2.1.29
Let $C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ be a bivariate Archimedean copula with generator φ. Then Kendall's tau τ and the tail dependence coefficients λ_l and λ_u of C have the following representation.

(1) $\tau = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}dt$

(2) $\lambda_l = \lim_{x\to\infty} \frac{\varphi^{[-1]}(2x)}{\varphi^{[-1]}(x)}$ and $\lambda_u = 2 - \lim_{x\downarrow 0} \frac{1-\varphi^{[-1]}(2x)}{1-\varphi^{[-1]}(x)}$

Proof
Let U, V be uniform on [0,1], (U,V) be distributed according to the Archimedean copula C above and $C(U,V) \sim K_C(t)$.

(1) According to equation (2.1.10) we have

$$\begin{aligned}\tau &= 4\mathbb{E}\left(C(U,V)\right) - 1\\ &= 4\int_0^1 t dK_C(t) - 1\\ &\overset{(\star)}{=} 3 - 4\int_0^1 K_C(t)dt\\ &\overset{(\star\star)}{=} 3 - 4\int_0^1 \left(t - \frac{\varphi(t)}{\varphi'(t^+)}\right)dt = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}dt\end{aligned}$$

where $(\star)$ is a consequence of integration by parts and $(\star\star)$ is due to item (1) in Proposition 2.1.25.

(2) Use the representation in equation (2.1.14) to obtain

$$\lambda_l = \lim_{t\downarrow 0}\frac{C(t,t)}{t} = \lim_{t\downarrow 0}\frac{\varphi^{[-1]}(2\varphi(t))}{t} \overset{(\star)}{=} \lim_{x\to\infty}\frac{\varphi^{[-1]}(2x)}{\varphi^{[-1]}(x)}$$

respectively in equation (2.1.15) for

$$\lambda_u = 2 - \lim_{t\uparrow 1}\frac{1-C(t,t)}{1-t} = 2 - \lim_{t\uparrow 1}\frac{1-\varphi^{[-1]}(2\varphi(t))}{1-t} \overset{(\star)}{=} 2 - \lim_{x\downarrow 0}\frac{1-\varphi^{[-1]}(2x)}{1-\varphi^{[-1]}(x)}$$

where the equality in $(\star)$ is upon the monotone substituition $\varphi(t) = x$.

□

2.2 Copula estimation

This section is devoted to a general overview of inference techniques for copula-based models. We will focus on the estimation of the dependence structure of multivariate random vectors. In the introductory summary text on nonparametric estimation in Section 2.2.1 we will get acquainted with *empirical distribution functions*. After an in-depth treatment of the most classical parametric inference method of *maximum likelihood estimation*, we will conclude Section 2.2.2 with the semiparametric *canonical maximum likelihood* procedure. Other parametric approaches will be mentioned very briefly in Section 2.2.3.

Since we aim at multivariate time series models in the main part of this thesis the observation horizon is denoted by $T \in \mathbb{N}$ and we formulate the results in the d-variate setting for $d \geq 2$. Let $\boldsymbol{u}_1, \dots, \boldsymbol{u}_T$ be realizations of independent and identically distributed (iid) copies $\boldsymbol{U}_1, \dots, \boldsymbol{U}_T$ of the random vector $\boldsymbol{U} \sim C$ such that $\boldsymbol{u}_t = (u_{1t}, \dots, u_{dt}) \in [0,1]^d$ for all $t = 1, \dots, T$. Analogously, let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_T$ with $\boldsymbol{x}_t = (x_{1t}, \dots, x_{dt}) \in \mathbb{R}^d$, $t = 1, \dots, T$, be a sample of iid $\boldsymbol{X}_1, \dots, \boldsymbol{X}_T$ distributed as $\boldsymbol{X} \sim H = C(F_1, \dots, F_d)$ with copula C and univariate marginal distributions $F_1, \dots, F_d$.

2.2.1 Empirical copulas

With the aim of assessing the dependence structure of given data, the nonparametric estimation technique that will be presented in the following is the most general one in our context and works under the least assumptions concerning the actual class of the copula.
Although it is unrealistic to assume that the copula data itself is available in applications, we will present this theoretical case to get an impression of the resulting empirical estimators and processes. Based on $\boldsymbol{u}_1, \dots, \boldsymbol{u}_T$ consider the following estimator for the copula C

$$\bar{C}_T(\boldsymbol{u}) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}_{\{U_{1t} \leq u_{1t}, \dots, U_{dt} \leq u_{dt}\}}. \tag{2.2.1}$$

Since the data is iid, the summation is that of iid indicators that are binomially distributed with size 1 and probability $C(\boldsymbol{u})$ and therefore the Central Limit Theorem holds. For a fixed argument $\boldsymbol{u} \in [0,1]^d$ and $T \to \infty$ it follows that

$$\sqrt{T} \left(\bar{C}_T(\boldsymbol{u}) - C(\boldsymbol{u}) \right) \xrightarrow{d} \mathcal{N}\big(0, C(\boldsymbol{u})(1 - C(\boldsymbol{u}))\big). \tag{2.2.2}$$

Also a functional limit can be shown, namely that for $\boldsymbol{u} \in [0,1]^d$ and $T \to \infty$

$$\sqrt{T} \left(\bar{C}_T - C \right)(\boldsymbol{u}) \xrightarrow{\mathcal{D}} \mathcal{G}(\boldsymbol{u}) \tag{2.2.3}$$

where $\mathcal{G}$ is a centered Gaussian process with $\mathbb{E}\left(\mathcal{G}(\boldsymbol{u})\right) = 0$ and $\mathrm{Cov}\left(\mathcal{G}(\boldsymbol{u}), \mathcal{G}(\tilde{\boldsymbol{u}})\right) = \min\{C(\boldsymbol{u} \wedge \tilde{\boldsymbol{u}}) - C(\boldsymbol{u})C(\tilde{\boldsymbol{u}})\}$. Contrary to the result for empirical distribution functions

$\hat{F}_{jT}$ as in (2.2.4) below, where the limiting process is independent of the actual distribution F_j, in case of multivariate empirical distribution functions the variance of the limiting distribution of the empirical process is still a function of the copula C.
The usually considered empirical distribution function as estimator for the univariate distribution function F_j is defined as

$$\hat{F}_{jT}(x_j) := \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}_{\{X_{jt}\leq x_j\}} \tag{2.2.4}$$

for every $j = 1, \ldots, d$ and has a multivariate analogue for the joint distribution function H of the form

$$\hat{H}_T(\boldsymbol{x}) := \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}_{\{X_{1t}\leq x_1,\ldots,X_{dt}\leq x_d\}}. \tag{2.2.5}$$

Let the generalized inverses $\hat{F}_{jT}^{-}$ of $\hat{F}_{jT}$ be defined according to the definition in (2.1.4), to wit

$$\hat{F}_{jT}^{-}(y) = \begin{cases} X_{j(1)}, & 0 \leq y \leq \frac{1}{T} \\ X_{j(t)}, & \frac{t-1}{T} < y \leq \frac{t}{T},\ t = 2, \ldots, T. \end{cases} \tag{2.2.6}$$

Assuming continuous marginal distributions the application of Sklar's Theorem in the sense of equation (2.1.5) leads to the (unique) empirical copula

$$\hat{C}_T(\boldsymbol{u}) = \hat{H}_T\left(\hat{F}_{1T}^{-}(u_1), \ldots, \hat{F}_{dT}^{-}(u_d)\right) \tag{2.2.7}$$

and the empirical copula process

$$\hat{Z}_T(\boldsymbol{u}) = \sqrt{T}\left(\hat{C}_T - C\right)(\boldsymbol{u}). \tag{2.2.8}$$

The formulation in (2.2.7) is called *Bahadur representation* and is due to Bahadur (1966). In particular the following results concerning the empirical copula and the empirical copula process were derived. In Deheuvels (1979) the consistency of $\hat{C}_T$ was investigated, subsequently in Deheuvels (1980) the uniform convergence to C at rate $\mathcal{O}_{\text{a.s.}}\left(\sqrt{\log\log T}/\sqrt{T}\right)$ was established.
There exist interim results that consider the convergence of the empirical copula process, but it was not before Fermanian et al. (2004) who showed weak convergence of $\hat{Z}_T$ in $\mathbb{L}^\infty([0,1]^d)$ under the condition of continuous partial derivatives with the help of the functional delta method. Find the results condensed in Tsukahara (2005, Sect. 2). They argue that this is the most general possible result within this scope. But the covariance structure and the partial derivatives are usually of such a complicated form that the practical usability is restricted.
Anyhow, the empirical copula has another noteworthy field of application in dependence models. Whenever a measure of association has a representation in terms of copula or copula density this theoretical factors can be replaced by their empirical counterparts for inference tasks (see e.g. Schmid and Schmidt, 2007a,b). We will go into further detail of similar inference methods in Section 2.2.3.

2.2.2 Maximum likelihood estimation

The empirical copula and the empirical copula process in Section 2.2.1 might be the most omnibus techniques to determine dependence in terms of copulas, but the estimated object is only of limited use when it comes to modeling tasks. This is where parametric models show their strength. As already mentioned in the context of the Archimedean copula class in Section 2.1.5, a *parametric copula class* is a family $\{C_\theta : \theta \in \Theta\}$ where the copulas C_θ are functions of a parameter (vector) $\theta \in \Theta \subseteq \mathbb{R}^m$.
Although there are other possibilities to estimate the copula parameter θ, some of the recent developments are to be presented in Section 2.2.3, consider the following argument in Joe (1997, p. 297) that supports our decision to concentrate on maximum likelihood estimation:

> "[...] For almost all of the multivariate models [...] much of the classical statistical inference theory is not applicable. This includes exponential family results, sufficient statistics, ancillary statistics, minimum variance unbiased estimators, etc. [...] Practically the only theory that can be applied is the asymptotic maximum likelihood (ML) theory. [...] "

Even if this choice limits us to the parametric copula classes which possess a density, we can base our contribution on a 100-year history of mathematical research. Among a multitude of his publications, Fisher (1912, 1921, 1922) laid the foundation for the "most important development in 20th century statistics". A very nice paper following the making of maximum likelihood and the life of its founder Sir Ronald Aylmer Fisher between 1912 and 1921 is Aldrich (1997). Le Cam (1990) provides another valuable contribution to the discussion by highlighting some critical issues.
Before we start, we want to stress that we will apply maximum likelihood estimation only to multivariate distributions that possess an absolutely continuous copula (Ruiz-Rivas and Cuadras (1988) estimated the parameters of the Cuadras-Augé copula that posses a sigular component by decomposing the dominant measure into the respective parts). According to Nelsen (2006, p. 27) and if applicable, for every copula we can consider the decomposition $C(\boldsymbol{u}) = A_C(\boldsymbol{u}) + S_C(\boldsymbol{u})$ for $\boldsymbol{u} \in [0,1]^d$ where

$$A_C(\boldsymbol{u}) = \int_0^{u_1} \dots \int_0^{u_d} \frac{\partial^d}{\partial x_1 \dots \partial x_d} C(x_1, \dots, x_d)\, dx_d \dots dx_1 \quad \text{and} \quad S_C(\boldsymbol{u}) = C(\boldsymbol{u}) - A_C(\boldsymbol{u}).$$

If $C \equiv A_C$ on $[0,1]^d$, then the copula is called *absolutely continuous* and in parametric classes we denote its *copula density* by

$$c(u_1, \dots, u_d; \theta) = \frac{\partial^d}{\partial u_1, \dots, \partial u_d} C(u_1, \dots, u_d; \theta). \tag{2.2.9}$$

If $C \equiv S_C$, i.e. if $c(\boldsymbol{u}; \theta) = 0$ almost everywhere on $[0,1]^d$, then C is called *singular*. Otherwise, the copula C possesses an *absolutely continuous component* A_C and a *singular component* S_C which themselves are neither copulas as they do not have uniform margins. Note that distribution functions generally allow for a decomposition into three parts,

namely an absolutely continuous, a singular and a discrete component. In case of copulas – being distribution functions on the unit cube $[0,1]^d$ – the discrete part drops out as the marginal distribution are always continuously uniform.
Examples of copulas with both an absolutely continuous and a singular component are Marshall-Olkin copulas (see Appendix A.3) or Clayton copulas with $\theta \in [-1, 0)$ (see Appendix A.5). A purely singular copula is for example the Fréchet-Hoeffding upper bound M. A nonparametric absolutely continuous copula is the product copula Π; with few exceptions the copulas presented in Appendix A are parametric absolutely continuous copulas that have mixed derivatives of order d as their densities.
The maximum likelihood estimator for distributions with densities is defined as follows.

Definition 2.2.1
For a given density $h_\theta \colon \mathbb{R}^d \to (0, \infty)$ of a parametric class of distribution functions $H_\theta \colon D \to [0,1]$ with domain $D \subseteq \mathbb{R}^d$ and parameter $\theta \in \Theta \subseteq \mathbb{R}^m$ the *(theoretical) log-likelihood function* is defined as

$$l \colon D \times \Theta \to \mathbb{R}, \quad l(\boldsymbol{y}; \theta) = \log h(\boldsymbol{y}; \theta) \tag{2.2.10}$$

and the *joint log-likelihood function* is given by

$$L \colon D^T \times \Theta \to \mathbb{R}, \quad L(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T; \theta) = \sum_{t=1}^{T} l(\boldsymbol{y}_t; \theta). \tag{2.2.11}$$

For independent and identically distributed random variables $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T$ distributed according to H_θ the *maximum likelihood estimator* of the parameter θ is given by

$$\hat{\theta}(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T) := \underset{\theta \in \Theta}{\operatorname{argmax}}\, L(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T; \theta), \tag{2.2.12}$$

that is as that value where the joint log-likelihood function attains its (not necessarily unique) maximum with respect to θ. For specific iid samples $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$ we set $\hat{\theta} := \hat{\theta}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$ as the *sample maximum likelihood estimate* of θ.

In our context we plug in the copula density $c(\boldsymbol{u}; \theta)$ for h. Furthermore, consider the following partial derivatives of the log-likelihood function

$$l_\theta(\boldsymbol{u}; \theta) := \frac{\partial}{\partial \theta} l(\boldsymbol{u}; \theta) \quad \text{and} \quad l_j(u_1, \ldots, u_d; \theta) := \frac{\partial}{\partial u_j} l(u_1, \ldots, u_d; \theta) \, \forall \, j = 1, \ldots, d.$$

Note that the subscript is given together with the function argument and in this case means the respective partial derivative.
For a fundamental understanding of the prominent role of maximum likelihood estimators as a class of consistent and asymptotically normal and efficient estimators the following theorem is stated. It is a reformulation of Theorems 3.7 and 3.10 in Lehmann and Casella (2003, Sect. 2) in terms of multivariate densities when the parameter θ is not vector-valued, but actually a real number. A more general version suitable in the financial time series setting will be developed later on.
Recall that an estimator is called *efficient* if it is consistent and its asymptotic variance attains the Cramér-Rao lower bound that is the inverse of the Fisher information (for details see Theorem 2.2.2 below or the original references Cramér, 1946 and Rao, 1945).

Theorem 2.2.2
Let $\boldsymbol{Y}, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_d$ be independent and identically distributed according to a d-variate distribution H_θ with density h_θ. Let $\theta \in \Theta$ and $\boldsymbol{y} \in S := \operatorname{supp}(h_\theta) \subseteq \mathbb{R}^d$. Assume that the following conditions hold.

(i) The parameter space $\Theta \subseteq \mathbb{R}$ is an open interval.

(ii) The support S does not depend on θ.

(iii) $h(\boldsymbol{y}; \theta)$ is three times continuously differentiable with respect to θ.

(iv) $\int_S \frac{\partial}{\partial \theta} h(\boldsymbol{y}; \theta)\, d\boldsymbol{y} = \frac{\partial}{\partial \theta} \int_S h(\boldsymbol{y}; \theta)\, d\boldsymbol{y} = 0$.

(v) $\mathbb{E}\, l_\theta(\boldsymbol{Y}; \theta)^2 + \mathbb{E}\, l_{\theta\theta}(\boldsymbol{Y}; \theta) = 0$.

(vi) The *Fisher Information* $I(\theta) := \mathbb{E}\, l_\theta(\boldsymbol{Y}; \theta)^2 = -\mathbb{E}\, l_{\theta\theta}(\boldsymbol{Y}; \theta)$ is positive and finite.

(vii) For all $\theta_0 \in \Theta$ and $\theta \in U_\delta(\theta_0)$ there exists a measurable function M_{θ_0} with $\mathbb{E}_{\theta_0}(M(\boldsymbol{Y}; \theta_0)) < \infty$ such that $|l_{\theta\theta\theta}(\boldsymbol{y}; \theta)| \leq M(\boldsymbol{y}; \theta_0)$ for all $\boldsymbol{y} \in S$.

Let θ_0 be the true parameter. Then one has the following asymptotic properties.

(1) Under conditions (i) and (iii) and with probability tending to 1 as the sample size grows to infinity there exists $\hat{\theta}(T) = \hat{\theta}(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T)$ such that

$$\frac{\partial}{\partial \theta} L(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T; \hat{\theta}(T)) = 0 \tag{2.2.13}$$

and $\hat{\theta}(T) \xrightarrow[(T \to \infty)]{P} \theta_0$. A unique root of (2.2.13) yields *weak consistency* of $\hat{\theta}(T)$.

(2) Under conditions (i) – (vii) every $\hat{\theta}(T)$ that is weakly consistent in above sense, satisfies

$$\sqrt{T}\left(\hat{\theta}(T) - \theta_0\right) \xrightarrow[(T \to \infty)]{d} \mathcal{N}\Big(0, \frac{1}{I(\theta)}\Big),$$

i.e. the (unique) maximum likelihood estimator is asymptotically normal and efficient.

Proof

(1) See Theorem 2.3.7 in Lehmann and Casella (2003).

(2) See Theorem 2.3.10 in Lehmann and Casella (2003). Details of the proof will be stated later as we turn to a joint normality result in Section 4.1.

□

Note that sometimes conditions (iv) and (v) above are replaced by the more general assumption that

$$\frac{\partial^k}{\partial\theta^k}\int_S h(\boldsymbol{y};\theta)\,d\boldsymbol{y} = \int_S \frac{\partial^k}{\partial\theta^k} h(\boldsymbol{y};\theta)\,d\boldsymbol{y} \quad \text{for } k = 1, 2.$$

In the theorem above, the essence of maximum likelihood estimation is given in a very general formulation to suit our need to deal with different copula-based models. Depending on the present data situation, it can now be elaborated straightforwardly in terms of the respective densities. With the notion of an iid sample we indicate the realization of independent and identically distributed copies of a random vector distributed according to the distribution under consideration.

Copula maximum likelihood

It is merely of theoretical interest to consider an iid sample $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_T$ of pure copula data, i.e. realizations of C_θ with density c_θ and parameter space $\Theta \subseteq \mathbb{R}$, as this is rarely the case in empirical analyses. The corresponding log-likelihood function for $\boldsymbol{u} \in [0,1]^d$ is then

$$l(\boldsymbol{u};\theta) = \log c(\boldsymbol{u};\theta)$$

and the parameter estimate is given by

$$\hat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmax}} \sum_{t=1}^{T} l(\boldsymbol{U}_t;\theta).$$

Under the assumptions in Theorem 2.2.2 it is consistent and asymptotically normally distributed with mean 0 and variance $I(\theta)^{-1} = (\mathbb{E}\, l_\theta(\boldsymbol{U};\theta)^2)^{-1}$, and so asymptotically efficient.
To arrive at consistent estimates for this variance consider

$$\text{the } \textit{expected Fisher Information} \quad \hat{I} := I(\hat{\theta}) = \mathbb{E}\, l_\theta(\boldsymbol{U};\hat{\theta})^2 \tag{2.2.14}$$

$$\text{and the } \textit{observed Fisher Information} \quad \hat{J} := J(\hat{\theta}) = -\frac{1}{T}\sum_{t=1}^{T} l_{\theta\theta}(\boldsymbol{U}_t;\hat{\theta}). \tag{2.2.15}$$

Both are consistent as estimators for the Fisher information; the former due to the consistency of $\hat{\theta}$ for θ and the continuous mapping theorem, the latter as a consequence of the weak law of large numbers. This yields $\hat{I}^{-1}$ and $\hat{J}^{-1}$ as possible variance estimators in the above normal limit distribution. However, the observed Fisher information seems to be more suitable to assess the asymptotic estimation variance as its computation takes advantage of the present observations in two respects as opposed to the expected Fisher information. Apart from that, it is often computationally tedious or impossible to assess the expected Fisher information which implies evaluating d-dimensional integrals. Cox (1958) and Efron and Hinkley (1978) give a detailed comparison of the two approximations and argue for the appropriateness of the observed version as a variance estimate in likelihood estimation.

Full maximum likelihood

Let us now shed light on procedures to estimate the parameter θ of a copula given data from a multivariate distribution $H = C(F_1, \dots, F_d; \theta)$. It is again owing to Sklar's Theorem that this task can be accomplished nevertheless. In a fully parametrized model the copula C depends on a parameter $\theta \in \Theta$ and every marginal distribution F_j is a function of a (set of parameters) $\delta_j \in \Delta_j \subseteq \mathbb{R}^{p_j}$, $p_j \geq 1$, $j = 1, \dots, d$. The log-likelihood function can be directly derived with the help of the chain rule for mixed partial derivatives. For $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ we obtain

$$l(\boldsymbol{x}; \theta, \delta_1, \dots, \delta_d) = \log c(F_1(x_1; \delta_1), \dots, F_d(x_d; \delta_d); \theta) + \sum_{j=1}^{d} \log f_j(x_j; \delta_j). \tag{2.2.16}$$

Unlike the exclusion of singular copulas or copulas with a singular component, the marginal distributions are not restricted to absolutely continuous ones, but can also be discrete. As the case may be, we denote by f_j the density or the probability mass function corresponding to the absolutely continuous or discrete distribution function F_j for every margin $j = 1, \dots, d$.
The maximum likelihood estimate based on an iid sample $\boldsymbol{X}_1, \dots, \boldsymbol{X}_T$ of H is then

$$(\hat{\theta}, \hat{\delta}_1, \dots, \hat{\delta}_d) = \underset{\theta \in \Theta, \delta_1 \in \Delta_1, \dots, \delta_d \in \Delta_d}{\operatorname{argmax}} \sum_{t=1}^{T} l(\boldsymbol{X}_t; \theta, \delta_1, \dots, \delta_d) \tag{2.2.17}$$

by simultaneously maximizing the joint log-likelihood with respect to the parameters $(\theta, \delta_1, \dots, \delta_d) \in \Theta \times \Delta_1 \times \dots \times \Delta_d$ and therefore called *full maximum likelihood*. Under conditions similar to conditions (i) – (vii) in Theorem 2.2.2 imposed on $h(\boldsymbol{x}; \theta, \delta_1, \dots, \delta_d) = \exp\{l(\boldsymbol{x}; \theta, \delta_1, \dots, \delta_d)\}$, the joint consistency and joint asymptotic normality of the estimators $(\hat{\theta}, \hat{\delta}_1, \dots, \hat{\delta}_d)$ hold true. Textbook reference for the multiparameter case of maximum likelihood estimation is to be found for example in Theorem 5.1 in Lehmann and Casella (2003, Sect. 2). In particular we have (univariate) asymptotic normality of $\hat{\theta}$ with consistent variance estimate $J(\hat{\theta}, \hat{\delta}_1, \dots, \hat{\delta}_d)^{-1}$ where

$$J(\hat{\theta}, \hat{\delta}_1, \dots, \hat{\delta}_d) = -\frac{1}{T} \sum_{t=1}^{T} l_{\theta\theta}(F_1(X_{1t}; \hat{\delta}_1), \dots, F_d(X_{dt}; \hat{\delta}_d); \theta)|_{\theta=\hat{\theta}}.$$

Inference for margins

At a closer look at (2.2.16) we observe that only the first summand of the log-likelihood function in the fully parametrized model depends on the copula parameter θ. In Joe and Xu (1996) a computationally attractive alternative to the full maximum likelihood method was proposed. Instead of the simultaneous maximization over the dependence parameter θ and the marginal parameters δ_j of the multivariate model proceed as follows:

(1) Maximize every marginal log-likelihood with respect to the marginal parameter to obtain

$$\hat{\delta}_j = \underset{\delta_j \in \Delta_j}{\operatorname{argmax}} \sum_{t=1}^{T} \log f_j(X_{jt}; \delta_j). \tag{2.2.18}$$

(2) Plugging in the obtained marginal estimates into the first summand of (2.2.16) and maximizing over the copula parameter leads to

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{t=1}^{T} \log c(F_1(X_{1t}; \hat{\delta}_1), \ldots, F_d(X_{dt}; \hat{\delta}_d); \theta). \tag{2.2.19}$$

This technique is called *inference functions for margins* method as the inference functions are score functions of likelihoods of marginal distributions, and is applicable to wider classes of models as noted by Joe (1997, Sect. 10.1) and thoroughly investigated by Xu (1996).
Under the regularity conditions in Theorem 2.2.2 imposed on the marginal likelihoods $L_j(\delta_j) = \sum_{t=1}^{T} \log f_j(x_{jt}; \delta_j)$, $j = 1, \ldots, d$ and the copula likelihood $L(\theta) = \sum_{t=1}^{T} \log c(F_1(x_{1t}; \delta_1), \ldots, F_d(x_{dt}; \delta_d); \theta)$ and with the score function

$$s(\theta, \delta_1, \ldots, \delta_d) := \left(\frac{\partial}{\partial \theta} L(\theta), \frac{\partial}{\partial \delta_1} L_1(\delta_1), \ldots, \frac{\partial}{\partial \delta_d} L_d(\delta_d) \right)^\top$$

a joint normality result holds as $T \to \infty$

$$\sqrt{T} \left((\hat{\theta}, \hat{\delta}_1, \ldots, \hat{\delta}_d)^\top - (\theta, \delta_1, \ldots, \delta_d)^\top \right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \hat{G}^{-1}). \tag{2.2.20}$$

$\hat{G}$ denotes the estimator (by plugging in the parameter estimates $(\hat{\theta}, \hat{\delta}_1, \ldots, \hat{\delta}_d)$ for the parameter vector ϑ) of the *Godambe's information matrix* G which is defined as

$$G(\vartheta) = D^{-1}(\vartheta)\, V(\vartheta)\, D^{-1}(\vartheta)^\top, \quad D(\vartheta) = \mathbb{E}\left(\tfrac{\partial}{\partial \vartheta} s(\vartheta) \right), \quad V(\vartheta) = \mathbb{E}\left(s(\vartheta)\, s(\vartheta)^\top \right).$$

In Joe (1997, Sect. 10.1) it is further discussed how to construct jackknife estimators for the asymptotic variance of the inference for margins estimators to circumvent the analytically and numerically problematic Godambe matrix. By comparison to the full maximum likelihood estimation a lack of efficiency for the inference for margins method is demonstrated. But the loss in efficiency is so moderate, that one can take advantage of the significant reduction of numerical complexity: a numerical optimization with many parameters is usually much more time-consuming compared to a series of numerical optimizations, each with fewer parameters.

Canonical maximum likelihood

Similar to the inference procedure discussed above, the density decomposition for the log-likelihood function of multivariate data as in (2.2.16) motivates another two-stage estimation technique. It is the semiparametric *canonical maximum likelihood* method that works as follows:

(1) Estimate the marginal distributions nonparametrically through a rescaled version of the empirical distribution function (cf. (2.2.4)) based on an iid sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T$

$$\tilde{F}_{jT}(x_j) := \frac{T}{T+1} \hat{F}_{jT}(x_j) = \frac{1}{T+1} \sum_{s=1}^{T} \mathbb{1}_{\{X_{js} \leq x_j\}}. \tag{2.2.21}$$

(2) Based on $\tilde{F}_{1T}, \dots, \tilde{F}_{dT}$ maximize the copula likelihood with respect to the parameter θ

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \sum_{t=1}^{T} \log c(\tilde{F}_{1T}(X_{1t}), \dots, \tilde{F}_{dT}(X_{dt}); \theta). \tag{2.2.22}$$

As a result of step (1) we sometimes consider the rank-transformed observations

$$\tilde{\boldsymbol{U}}_t = \left(\tilde{U}_{1t}, \dots, \tilde{U}_{dt}\right) \quad \text{with} \quad \tilde{U}_{jt} = \tilde{F}_{jT}(X_{jt}) \tag{2.2.23}$$

for an iid sample $\boldsymbol{X}_1, \dots, \boldsymbol{X}_T$ with $\boldsymbol{X}_t = (X_{1t}, \dots, X_{dt})$ for all $t = 1, \dots, T$. Due to the probability integral transform (2.1.1) these quantities are univariately uniform on [0,1] and jointly distributed according to C_θ and therefore termed *pseudo-observations* of the copula. The transformation by the rescaled empirical distribution functions guarantees that the pseudo-observations remain bounded at the borders of the unit cube $[0,1]^d$. We need further regularity conditions that control the speed of divergence of the log-likelihood function at the boundaries (cf. conditions (C4), (N3) and (N6) in Appendix B) together with the prerequisites in Theorem 2.2.2 to derive consistency and asymptotic normality of the semiparametric estimator $\hat{\theta}$ above. It is the publications by Ruymgaart et al. (1972); Ruymgaart (1974); Rüschendorf (1976) where the asymptotic behavior of rank statistics is thoroughly investigated and where the additional assumptions stem from. Multivariate rank order statistics are essential in proving the following famous result that is due to Genest et al. (1995, Prop. 2.1).

Theorem 2.2.3
Based on an iid sample $\boldsymbol{X}_1, \dots, \boldsymbol{X}_T \sim C(F_1, \dots, F_d; \theta)$ let the semiparametric estimator $\hat{\theta}$ be given as in (2.2.22). Under regularity conditions (i) – (vii) as well as (C4), (N3) and (N6), $\hat{\theta}$ is strongly consistent and it holds that for $T \to \infty$

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \Sigma) \tag{2.2.24}$$

with

$$\Sigma = \frac{\operatorname{Var}\left(l_\theta(F_1(X_1), \dots, F_d(x_d); \theta) + \sum_{j=1}^{d} W_j(X_j)\right)}{\mathbb{E}\left(l_\theta(F_1(X_1), \dots, F_d(X_d); \theta)^2\right)^2} =: \frac{\sigma^2}{\beta^2} \tag{2.2.25}$$

where $W_j(X_j) = \int_{[0,1]^d} \mathbb{1}_{\{F_j(X_j) \le u_j\}} l_{\theta,j}(u_1, \dots, u_d; \theta) dC(u_1, \dots, u_d; \theta)$.

Proof
Take as a starting point that $\hat{\theta}$ solves

$$\frac{1}{T}\frac{\partial}{\partial\theta} L(\theta) = \frac{1}{T} \sum_{t=1}^{T} l_\theta(\tilde{F}_1(x_{1t}), \dots, \tilde{F}_d(x_{dt}); \theta) = 0$$

and expand the above expression into a Taylor series around the true parameter value θ. Then a first order approximation (denoted by $\doteq$) is given by

$$\frac{1}{T}\frac{\partial}{\partial\theta} L(\theta)|_{\theta=\hat{\theta}} = 0 \doteq A_T - (\hat{\theta} - \theta) B_T,$$

with $A_T = \frac{1}{T}\sum_{t=1}^{T} l_\theta(\tilde{F}_1(x_{1t}),\dots,\tilde{F}_d(x_{dt});\theta)$, $B_T = -\frac{1}{T}\sum_{t=1}^{T} l_{\theta\theta}(\tilde{F}_1(x_{1t}),\dots,\tilde{F}_d(x_{dt});\theta)$. Rearranging the equation leads to $\sqrt{T}(\hat{\theta}-\theta) \backsimeq \sqrt{T}\,A_T/B_T$. The limiting behavior of this expression can be deduced through multivariate rank order statistics by substitution of J in Proposition A.1(ii) in Genest et al. (1995) by l_θ to obtain

$$B_T \xrightarrow[T\to\infty]{} \beta$$

and by substituting $l_{\theta\theta}$ for J to arrive at

$$\sqrt{T}\,A_T \xrightarrow[(T\to\infty)]{d} N(0,\sigma^2),$$

with β and σ^2 as in (2.2.25). □

Remark 2.2.4
Note that in the proof of Proposition A.1(ii) in Genest et al. (1995) reference to Theorem 2.1 in Ruymgaart et al. (1972) is made. For the connection between the two results to become clear, we want to mention that the function J with domain $[0,1]^2$ of Genest et al. corresponds to the product $J \cdot K$ with domain $[0,1] \times [0,1]$ of Ruymgaart et al.. Although the reasoning remains similar, the margins have to be simultaneously considered, especially when applying the mean value theorem.

Let us now turn to a suitable variance estimate. Using integration by parts with respect to each $u_j, j = 1,\dots,d$ we obtain an alternative expression for $W_j(X_j)$, namely

$$\begin{aligned} W_j(X_j) &= \int_{[0,1]^d} \mathbb{1}_{\{F_j(X_j)\le u_j\}} l_{\theta,j}(u_1,\dots,u_d;\theta)\,dC(u_1,\dots,u_d;\theta) \\ &= -\int_{[0,1]^d} \mathbb{1}_{\{F_j(X_j)\le u_j\}} l_\theta(u_1,\dots,u_d;\theta) l_j(u_1,\dots,u_d;\theta)\,dC(u_1,\dots,u_d;\theta). \end{aligned}$$

Define $K(\boldsymbol{X}) := l_\theta(F_1(X_1),\dots,F_d(X_d);\theta)$ and $L(\boldsymbol{X}) = K(\boldsymbol{X}) + \sum_{j=1}^{d} W_j(X_j)$ and the rescaled version of the empirical copula

$$\tilde{C}_T(\boldsymbol{u}) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}_{\{\tilde{F}_1(X_1)\le u_1,\dots,\tilde{F}_d(X_d)\le u_d\}}. \tag{2.2.26}$$

With given iid observations $\boldsymbol{x}_1,\dots,\boldsymbol{x}_T$ of $\boldsymbol{X}_1,\dots,\boldsymbol{X}_T$ we can construct estimates $\hat{K}_t$ and $\hat{L}_t$ for the above quantities

$$\begin{aligned} \hat{K}_t &= l_\theta(\tilde{F}_1(x_{1t}),\dots,\tilde{F}_d(x_{dt});\hat{\theta}), \\ \hat{L}_t &= \hat{K}_t - \sum_{j=1}^{d} \int_{[0,1]^d} \mathbb{1}_{\{\tilde{F}_j(x_{jt})\le u_j\}}\, l_\theta(u_1,\dots,u_d;\hat{\theta})\, l_j(u_1,\dots,u_d;\hat{\theta})\, d\tilde{C}_T(u_1,\dots,u_d). \end{aligned}$$

To finally assess these quantities we will work in terms of ranked components now. Let us assume that the jth components of the vectors in the given sample are arranged in increasing order and let $S_{lt}^{(j)} = \#\{\tau : x_{l\tau} \le x_{lt}\}$ for every remaining component

$l \neq j$ be the rank of x_{jt} among the x_j's under the name ordering of the jth component. Especially it holds that $S_{jt}^{(j)} = t$. Then we can express $\hat{K}_t$ and $\hat{L}_t$ in terms of S as

$$\hat{K}_t = l_\theta\Big(\frac{t}{T+1}, \dots, \frac{S_{dt}^{(1)}}{T+1}; \hat{\theta}\Big),$$
$$\hat{L}_t = \hat{K}_t - \frac{1}{T}\sum_{j=1}^{d}\sum_{\tau=t}^{T} l_\theta\Big(\frac{S_{1\tau}^{(j)}}{T+1}, \dots, \frac{S_{d\tau}^{(j)}}{T+1}; \hat{\theta}\Big)\, l_j\Big(\frac{S_{1\tau}^{(j)}}{T+1}, \dots, \frac{S_{d\tau}^{(j)}}{T+1}; \hat{\theta}\Big).$$

Note that in the definition of $\hat{K}_t$ the choice of conditional ranks with respect to the first component, denoted by $S_{kt}^{(1)}$ for all $k > 2$, is quite arbitrary. The formula remains valid for any other choice of the reference component, too.
An estimator for β is then given by

$$\hat{\beta} = \frac{1}{T}\sum_{t=1}^{T} \hat{K}_t{}^2.$$

For σ^2 we can choose for example the (unbiased) sample variance

$$\hat{\sigma}^2 = \frac{1}{T-1}\sum_{t=1}^{T}\left(\hat{L}_t - \frac{1}{T}\sum_{t=1}^{T}\hat{L}_t\right)^2 = \frac{1}{T-1}\left(\sum_{t=1}^{T}\hat{L}_t{}^2 - T\left(\frac{1}{T}\sum_{t=1}^{T}\hat{L}_t\right)^2\right).$$

Both $\hat{\beta}$ and $\hat{\sigma}^2$ are consistent estimators and therefore by Slutsky's theorem we obtain a consistent estimator

$$\hat{\Sigma} = \hat{\sigma}^2/\hat{\beta}^2 \tag{2.2.27}$$

for the asymptotic variance Σ of the canonical maximum likelihood estimator.

Remark 2.2.5 (1) Note that the terms W_j in above theorem account for the fact that the asymptotic variance Σ for the canonical maximum likelihood estimator is greater or equal than the inverse of the Fisher information being the usual asymptotic variance in maximum likelihood estimation. Under further assumptions in Genest et al. (1995, Prop. 2.2) it is shown that these vanish at independence and that $\hat{\theta}$ is fully efficient then.

More recently, Genest and Werker (2001) discussed further conditions for efficient semiparametric copula estimators and proved that for normal copulas asymptotic efficiency is attained. This is actually a result already anticipated by Klaassen and Wellner (1997).

(2) From a practical point of view, let Kim et al. (2007) be mentioned who argue for canonical maximum likelihood being superior to full maximum likelihood and the inference for margins method in terms of mean squared error. They base their reasoning on an extensive simulation study and especially focus on the behavior of the estimators when the marginal distributions are misspecified. Under this aspect, it is not surprising that a nonparametric marginal estimation outperforms a parametric one under wrong assumptions.

2.2.3 Other inference techniques

For completeness a fairly perfunctory overview over other approaches to estimate dependence in terms of copulas will be given now.

Method of moments

The commonly known *method of moments* estimation is a parametric inference procedure for given iid realizations $y_1, \dots, y_n$ of a random variable $Y \sim F_\theta$. It consists in equating a convenient number of sample moments to the corresponding theoretical moments which are functions of the unknown parameter(s) θ, and subsequently inverting this functional relationship to arrive at parameter estimates. Formally, we set

$$\mu_r(\theta) := \mathbb{E}_\theta(Y^r) \stackrel{!}{=} \frac{1}{n}\sum_{i=1}^{n} y_i^r =: \hat{m}_r$$

and obtain $\hat{\theta} = \mu_r^{-1}(\hat{m}_r)$ by inversion. Such estimators are usually consistent and asymptotically normal, but in general not efficient. For textbook treatment of this topic see e.g. Cramér (1946, Sect. 33.1) or Serfling (2001, Sect. 4.3.1).
A similar approach for bivariate (copula) data is based on sample versions of association measures such as Kendall's τ and Spearman's ρ and the functional relation of these to copula parameters θ. Let $(x_1, y_1), \dots, (x_n, y_n)$ be a sample of size n of iid realizations of $(X, Y) \sim C(F, G; \theta)$, then we get the *sample Spearman's* ρ

$$\hat{\rho}_n = \frac{12}{n(n^2-1)}\sum_{i=1}^{n} \hat{F}_n(x_i)\hat{G}_n(y_i) - 3\frac{n+1}{n-1}, \tag{2.2.28}$$

where $\hat{F}_n$ and $\hat{G}_n$ are the usual empirical distribution functions for the margins, and the *sample Kendall's* τ

$$\hat{\tau}_n = \frac{4}{n(n-1)}\sum_{1\le i<j\le n} \operatorname{sgn}((x_i - x_j)(y_i - y_j)). \tag{2.2.29}$$

Often, it holds that $\rho(X, Y) = g_\rho(\theta)$ and $\tau(X, Y) = g_\tau(\theta)$ with some function $g_\cdot$ (for examples see Tables A.1 and A.2 in Appendix A) such that the *method of moments* analogue of a parameter estimator is given by $\hat{\theta} = g_\rho^{-1}(\hat{\rho}_n)$ and $\hat{\theta} = g_\tau^{-1}(\hat{\tau}_n)$ respectively. In the following we will restrict ourselves to some asymptotic results for the sample version of Kendall's τ. First of all, the statistic in (2.2.29) possesses a representation as U-statistic, namely as U-statistic with kernel $\kappa_\tau((x, y), (\tilde{x}, \tilde{y})) = \operatorname{sgn}((x - \tilde{x})(y - \tilde{y}))$ of order 2.

Definition 2.2.6
For iid random variables $Y_1, \dots, Y_n \in (\Omega, P)$ let $n \ge m \in \mathbb{N}$ and $\kappa\colon \Omega^m \to \mathbb{R}$ be a symmetric and measurable function. Then a *U-statistic with kernel of order m* is defined as

$$\hat{U}_n^m(\kappa) = \binom{n}{m}^{-1} \sum_{1\le i_1<\dots i_m\le n} \kappa(Y_{i_1}, \dots, Y_{i_m}). \tag{2.2.30}$$

There are general results concerning the consistency, moments and asymptotic normality of U-statistics (see e.g. Lee, 1990, or Serfling, 2001, Chap. 5), as well as recent results concerning Kendall's τ in particular (see Dengler, 2010).
Let us summarize the result for the moments in the following lemma given without proof.

Lemma 2.2.7
Let $\hat{\tau}_n$ be given as in (2.2.29). Then

(1) $\mathbb{E}(\hat{\tau}_n) = \tau$

(2) $\operatorname{Var}(\hat{\tau}_n) = \binom{n}{2}^{-1} \left(1 - \tau^2 + 2(n-2)\operatorname{Var}(\mathbb{E}(\operatorname{sgn}(X - \tilde{X})(Y - \tilde{Y})))\right)$

where (X, Y) and $\left(\tilde{X}, \tilde{Y}\right)$ are independent copies of the respective distribution. In particular, it holds that $\operatorname{Var}(\hat{\tau}_n) = \mathcal{O}(1/n)$.

Similar to comparison of semiparametric to parametric methods in Remark 2.2.5(2) different semiparametric inference procedures can be compared to each other. This is what Kojadinovic and Yan (2010a) do by contrasting canonical maximum likelihood estimators and parameter estimates based on the inversion of Kendall's τ and Spearman's ρ. They also conclude that the maximum-likelihood-based approach is superior to the method-of-moments estimation in terms of mean squared error.
Some recent publications dealing with method-of-moments approaches in copula estimation are for example Oh and Patton (2011) and Brahimi and Necir (2012).

Minimum distance estimation

Originally proposed by Wolfowitz (1957) in a non-copula context, the *minimum distance* method yields parameter estimators by minimizing the distance between the empirical distribution of given observations to the theoretical distribution function belonging to a parametric family. Putting this concept mathematically precise in terms of iid random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T$ according to a multivariate distribution $C(F_1, \ldots, F_d; \theta)$, consider their empirical copula $\hat{C}_T$ as in (2.2.7) and the corresponding empirical copula process $\hat{Z}_{T,\theta}$ as in (2.2.8). Then the minimum distance estimator of θ is defined as

$$\hat{\theta} := \operatorname*{argmin}_{\theta \in \Theta} \|\hat{C}_T - C_\theta\| = \operatorname*{argmin}_{\theta \in \Theta} \mathcal{D}(\hat{Z}_{T,\theta}), \tag{2.2.31}$$

where the norm $\|\,.\,\|$ measuring the distance between the empirical and the actual copula is arbitrary, but required to be equivalent to the uniform norm $\|\,.\,\|_\infty$, and can be expressed in terms of a distance statistic $\mathcal{D}(\,.\,)$ of the empirical copula process. Famous examples for $\mathcal{D}$ are the *Cramér-von-Mises* statistic

$$\mathcal{D}^{CvM} = \int_{[0,1]^d} \hat{Z}_{T,\theta}(\boldsymbol{u})^2 d\hat{C}_T(\boldsymbol{u}) \tag{2.2.32}$$

and the *Kolmogorov-Smirnov* distance

$$\mathcal{D}^{KS} = \sup_{\boldsymbol{u} \in [0,1]^d} |\hat{Z}_{T,\theta}(\boldsymbol{u})|. \tag{2.2.33}$$

These approaches originally appeared in testing goodness of fit for copula data (see Genest et al. (2009b) for an overview). Furthermore, the statistics above were proposed to be applied on other transforms than the empirical copula. In Genest et al. (2006) the asymptotic properties of empirical processes based on *Kendall's transform* were derived; Savu and Trede (2008) used it for goodness-of-fit tests in case of Archimedean copulas. *Rosenblatt's transform* was originally introduced in Rosenblatt (1952) and used for goodness-of-fit tests in Berg (2009).
When it concerns estimation, in Weiß (2011) a large scale simulation study is presented where minimum distance estimators (for many possible combinations of transforms and distances) are compared to maximum likelihood estimators. The author concludes that maximum likelihood yields smaller estimation biases at less computational effort than any of the minimum distance estimators. In Tsukahara (2005) a similar comparison was conducted with focus put on on the semiparametric canonical maximum likelihood. The existence, consistency and asymptotic normality of $\hat{\theta}$ under a set of conditions is shown ibidem.
Most recently, Hering (2011) and Hering and Stadtmüller (2012) exploit this technique for estimation and goodness-of-fit in certain high-dimensional copula classes.

2.3 Extreme value theory

In Chapter 4 we will derive global confidence regions for the time-varying parameter sequences of a copula. We will achieve this by taking maxima over these sequences. To be able to assess the distribution of the resulting quantities, we will need some results from *extreme value theory*. A brief outline of the main results will be given in the following sections.

2.3.1 Basic theory

Extreme value theory is a classical topic in probability theory and mathematical statistics and dates back to Fisher and Tippett (1928). Two subjects are central in this field: the main analytic tool is the theory of *regularly varying functions*, whereas the basic probabilistic tool is *point process* theory. For textbook treatment of extreme value theory, the interested reader may consult for example Leadbetter et al. (1983), Resnick (1987) or Galambos (1987).
A thorough treatment of the topic with many univariate applications to finance can be found in Embrechts et al. (2003, Chap. 3). For an extension to the multivariate case with a natural link to copulas see Joe (1997, Chap. 6). Our presentation is close in spirit to these.
We begin with the extremal types theorem which is *the* basic theorem for classical extreme value theory. It is usually known as *Fisher-Tippett-Theorem*.

Theorem 2.3.1
Let $X_1, \ldots, X_n$ be a sequence of iid random variables and $M_n = \max_{i=1,\ldots,n}\{X_i\}$ their maximum. If there exists a sequence $\{a_n\}_{n=1}^{\infty}$ of positive constants and a sequence

$\{d_n\}_{n=1}^{\infty}$ of real constants as well as a non-degenerate distribution function H such that

$$\frac{1}{a_n}(M_n - d_n) \xrightarrow[(n\to\infty)]{\mathrm{d}} H, \tag{2.3.1}$$

then H belongs to the type of one of the following three distribution functions:

$$\begin{aligned}
&(\textit{Fréchet}) && \Phi_\alpha(x) = \mathrm{e}^{-x^{-\alpha}} && \text{for } \alpha > 0 \text{ and } x > 0,\\
&(\textit{Weibull}) && \psi_\alpha(x) = \mathrm{e}^{-(-x)^{\alpha}} && \text{for } \alpha > 0 \text{ and } x < 0,\\
&(\textit{Gumbel}) && \Lambda(x) = \mathrm{e}^{-\mathrm{e}^{-x}} && \text{for all } x \in \mathbb{R}.
\end{aligned}$$

Proof
For a sketch of the proof see Theorem 3.2.3 in Embrechts et al. (2003), a complete proof may be found in Proposition 0.3 in Resnick (1987). □

In particular, H is a continuous distribution function on $\mathbb{R}$ and we call H an *extreme value distribution*. Briefly, Theorem 2.3.1 states that the extreme value distributions are the limit laws of maxima of iid random variables and that these distributions are the only possible limiting distributions for (normed and centered) maxima.
Although the three extreme value distribution types are very different for modeling purposes, they are closely linked under the mathematical point of view. Take a positive random variable $X > 0$, then the following equivalences hold

$$X \sim \Phi_\alpha \quad \Leftrightarrow \quad \log(X^\alpha) \sim \Lambda \quad \Leftrightarrow \quad -X^{-1} \sim \psi_\alpha.$$

Definition 2.3.2
Let $H \in \{\Phi_\alpha, \psi_\alpha, \Lambda\}$ and $X_1, \ldots, X_n \sim F$ be iid random variables distributed according to F. Set $M_n = \max_{i=1,\ldots,n}\{X_i\}$. We say that F belongs to the *maximum domain of attraction* of H if, and only if, there exists a positive sequence $\{a_n\}_{n\geq 1}$ and a real-valued sequence $\{d_n\}_{n\geq 1}$ such that

$$\frac{1}{a_n}(M_n - d_n) \xrightarrow[(n\to\infty)]{d} H.$$

In short, we denote this by $F \in \mathrm{MDA}(H)$.

Since the extreme value distributions are continuous on their domains, we see that the limiting law H is *max-stable*, i.e. for $Y, Y_1, \ldots, Y_n \sim H$ the following closure property holds

$$\max\{Y_1, \ldots, Y_n\} \overset{d}{=} \tilde{a}_n Y + \tilde{d}_n \tag{2.3.2}$$

for appropriate constants $\tilde{a}_n > 0, \tilde{d}_n \in \mathbb{R}$.
This fact yields the following characterization of maximum domains of attraction for H in terms of the survival function $\bar{F} = 1 - F$.

Proposition 2.3.3
A distribution function F belongs to the maximum domain of attraction of the extreme value distribution H with norming constants $a_n > 0$, $d_n \in \mathbb{R}$ if, and only if, for all $x \in \mathbb{R}$

$$\lim_{n\to\infty} n\bar{F}(a_n x + d_n) = -\log H(x).$$

For $H(x) = 0$ we interpret the limit as ∞.

The maximum domain of attraction for each of the three families of extreme value distributions can be characterized through the behavior of the tails $\bar{F}$ for all $F \in \mathrm{MDA}(H)$. Recall that a distribution is said to have *regularly varying* tails with index $-\alpha$ for some $\alpha \geq 0$ if

$$\lim_{x\to\infty} \frac{\bar{F}(xt)}{\bar{F}(x)} = t^{-\alpha}, \quad t > 0.$$

Furthermore, we say that two distribution functions F and G with respective right endpoints x_F and x_G are *tail equivalent* if

$$x_F = x_G \quad \text{and} \quad \lim_{x\to x_F^-} \frac{\bar{F}(x)}{\bar{G}(x)} = c \text{ for some constant } c \in \mathbb{R}_0^+.$$

It can be shown that every maximum domain of attraction is closed with respect to tail equivalence, i.e. if $F \overset{\text{tail}}{\sim} G$, then $F \in \mathrm{MDA}(H) \Leftrightarrow G \in \mathrm{MDA}(H)$ with the same norming constants $a_n > 0$ and $d_n \in \mathbb{R}$.

2.3.2 Norming constants

We are not so much interested in the full description of the maximum domains of attraction of the extreme value distributions, but want to determine the specific form of the norming constants for a given distribution F. Regular variation theory proves to be of great help for both tasks. In the latter case, norming constants for at least the families of the Fréchet and the Weibull distribution can be calculated straightforwardly. Consider $F \in \mathrm{MDA}(H)$ and $X_1, \ldots, X_n \sim F$ iid. Then

(1) in the case $H = \Phi_\alpha$ (Fréchet family) choose zero as centering constant, and the upper $(1 - 1/n)$-quantile of F as scaling constant:

$$d_n = 0 \quad \text{and} \quad a_n = F^-(1 - 1/n).$$

(2) in the case $H = \psi_\alpha$ (Weibull family) we center by the right endpoint, and scale by the lower $(1 - 1/n)$-quantile of F:

$$d_n = x_F \quad \text{and} \quad a_n = x_F - F^-(1 - 1/n).$$

Note that the quantiles are formed through the generalized inverses F^- of the underlying distribution.
In order to illustrate the Fréchet and Weibull maximum domains of attraction consider the following examples.

Example 2.3.4
The Pareto distribution with parameter $\gamma > 0$ has the distribution function $F_P(x) = 1 - x^{-1/\gamma}$ for $x > 1$. With norming constants $d_n = 0$ and $a_n = n^\gamma$, it holds that F_P belongs to the maximum domain of attraction of the Fréchet extremal distribution, since for $x > 0$ we have

$$F_P^n(a_n x + d_n) = \left(1 - (n^\gamma x)^{-1/\gamma}\right)^n = \left(1 - x^{-1/\gamma}/n\right)^n \xrightarrow[(n\to\infty)]{} \mathrm{e}^{-x^{-1/\gamma}} = \Phi_{1/\gamma}(x).$$

Example 2.3.5
For the uniform distribution on [0,1] it holds that $x_F = 1$ and $F_{\mathcal{U}(0,1)}(x) = x$ for $x \in [0,1]$. Choose as norming constants $d_n = 1$ and $a_n = n^{-1}$, then F_{unif} belongs to the maximum domain of attraction of the Weibull extremal distribution, since for $-n < x < 0$ we have

$$F_{unif}^n(a_n x + d_n) = (x/n + 1)^n = (1 - (-x)/n)^n \xrightarrow[(n\to\infty)]{} \mathrm{e}^{-(-x)^1} = \psi_1(x).$$

The uniform distribution is contained in the family of Beta(a,b)-distributions, $a, b > 0$, as special case with $a = b = 1$. It can be shown that in general Beta$(a,b) \in \mathrm{MDA}(\psi_\alpha)$.

The maximum domain of attraction of the Gumbel distribution consists of distributions with a great variety in the tail behavior. More advanced results are needed to characterize the norming constants in this case. As an introductory example consider the exponential distribution.

Example 2.3.6
The exponential distribution with parameter 1 has the distribution function $F_{\mathrm{Exp}(1)}(x) = 1 - \mathrm{e}^{-x}$ for $x > 0$. As norming constants we can choose $d_n = \log n$ and $a_n = 1$, such that F_{exp} belongs to the maximum domain of attraction of the Gumbel extremal distribution, since for $x \in \mathbb{R}$ we have

$$F_{exp}^n(a_n x + d_n) = \left(1 - \mathrm{e}^{-(x+\log n)}\right)^n = \left(1 - \mathrm{e}^{-x}/n\right)^n \xrightarrow[(n\to\infty)]{} \mathrm{e}^{-\,\mathrm{e}^{-x}} = \Lambda(x).$$

In order to apply extreme value theory to the maximum of a sequence of maximum likelihood estimators, we derive the norming constants for the normal distribution in the following.

Proposition 2.3.7
Let $X_1, \ldots, X_n$ be independent and identically distributed according to $\mathcal{N}(\mu, \sigma^2)$ and set $M_n = \max_{i=1,\ldots,n}\{X_i\}$ as the maximum of this iid sequence.

(1) In the standard normal case with $\mu = 0$ and $\sigma^2 = 1$ choose

$$a_n = \sqrt{2\log n} \quad \text{and} \quad d_n = a_n - \frac{\log\log n + \log 4\pi}{2a_n} \tag{2.3.3}$$

as scaling and centering constants. Then $\lim_{n\to\infty} \mathbb{P}\left(M_n \le a_n x + d_n\right) = \Lambda(x) = \mathrm{e}^{-\,\mathrm{e}^{-x}}$ for all $x \in \mathbb{R}$.

(2) In the general case with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ the scaling and centering constants are

$$\tilde{a}_n = \sigma a_n \quad \text{and} \quad \tilde{d}_n = \mu + \sigma d_n. \tag{2.3.4}$$

Then $\lim_{n\to\infty} \mathbb{P}\left(M_n \leq \tilde{a}_n x + \tilde{d}_n\right) = \Lambda(x) = \mathrm{e}^{-\mathrm{e}^{-x}}$ for $x \in \mathbb{R}$.

Summarizing, we have that the normal distribution belong to the maximum domain of attraction of the Gumbel distribution with constants as above.

Proof

(1) See Example 3.3.29 in Embrechts et al. (2003, p. 145).

(2) A straightforward calculation yields that $\tilde{a}_n$ and $\tilde{d}_n$ can be determined from a_n and d_n through monotone transformations.

□

2.3.3 Dependent sequences

Classical extreme value theory is devoted to extremes in a sequence of independent and identically distributed random variables. The next question we want to address is whether the results can be extended to the non-iid case. Along the lines of Leadbetter et al. (1983, Chap. 3) and Galambos (1987, Chap. 3.7) we will outline that under certain (intertemporal) dependence restrictions there exists a natural extension of the theory at least for stationary sequences. Unless otherwise stated, we denote by *stationary* those sequences of random variables that are *strictly stationary*, i.e. sequences $\xi_1, \xi_2, \ldots$ of random variables with distribution function F such that

$$(\xi_{t_1}, \ldots, \xi_{t_m}) \stackrel{\mathrm{d}}{=} (\xi_{t_1+l}, \ldots, \xi_{t_m+l}) \tag{2.3.5}$$

for every $m, l \in \mathbb{N}$ and every index set $1 \leq t_1 < \ldots < t_m < \infty$.
Denote by $M_n = \max\{\xi_1, \ldots, \xi_n\}$ the maximum over a sample of size n of stationary random variables. An iid sequence $\{X_t\}$ with common distribution F is called *associated iid sequence* and we denote its maximum by $\tilde{M}_n = \max\{X_1, \ldots, X_n\}$. If $F \in \mathrm{MDA}(H)$ for any of the extreme value distributions H, we know by Definition 2.3.2 that there exist sequences $a_n > 0$ and $d_n \in \mathbb{R}$ such that $a_n^{-1}(\tilde{M}_n - d_n) \xrightarrow{\mathrm{d}} H$. Proposition 2.3.3 yields the following factorization property (see equation (4.37) in Embrechts et al., 2003)

$$\begin{aligned} \mathbb{P}\left(\tilde{M}_n \leq u_n\right) &= \mathbb{P}^n(X_1 \leq u_n) \\ &= \exp\{n \log(1 - \mathbb{P}\left(X_1 > u_n\right))\} \\ &\approx \exp\{-n\bar{F}(u_n)\} \end{aligned} \tag{2.3.6}$$

with a sequence $\{u_n\}$ to be further specified in the sequel. In particular, for any $0 \leq \tau < \infty$ the so-called *Poisson approximation* holds:

$$\mathbb{P}\left(\tilde{M}_n \leq u_n\right) \to \mathrm{e}^{-\tau} \iff n\bar{F}(u_n) \to \tau. \tag{2.3.7}$$

To extend this characterization to dependent stationary sequences we need a specific type of *asymptotic independence* stated below.

Condition $D(u_n)$
For any p, q and n let

$$1 \leq t_1 < \ldots < t_p < s_1 < \ldots < s_q \leq n$$

such that $s_1 - t_p \geq l$ and

$$|\mathbb{P}\left(M_{A_p \cup A_q} \leq u_n\right) - \mathbb{P}\left(M_{A_p} \leq u_n\right) \mathbb{P}\left(M_{A_q} \leq u_n\right)| \leq \alpha_{n,l} \tag{2.3.8}$$

where $A_p = \{t_1, \ldots, t_p\}$, $A_q = \{s_1, \ldots, s_q\}$, M_A is the maximum over the index set A and $\alpha_{n,l} \to 0$ is a decreasing function for some sequence $l = l(n) = o(n)$ as $n \to \infty$.

This is the necessary condition for the limiting maximal law to be an extreme value distribution, and condensed in the theorem below.

Theorem 2.3.8
If for appropriate constants $a_n > 0$, $d_n \in \mathbb{R}$

$$a_n^{-1}(M_n - d_n) \xrightarrow[(n\to\infty)]{\mathrm{d}} H$$

and $D(a_n x + d_n)$ holds for every $x \in \mathbb{R}$, then H is an extreme value distribution.

Moreover, another requirement has to be met by the threshold sequence $\{u_n\}$ to convey the Poisson approximation (2.3.7) and thus give sufficient conditions for convergence of the law of the maximum. This *anti-clustering condition* reads as follows:

Condition $D'(u_n)$

$$\lim_{k\to\infty} \limsup_{n\to\infty} n \sum_{t=2}^{[n/k]} \mathbb{P}\left(\xi_1 > u_n, \xi_t > u_n\right) = 0 \tag{2.3.9}$$

This yields the sufficient condition for existence of the limiting maximal law of a stationary sequence of random variables.

Theorem 2.3.9
Given a stationary sequence $\{\xi_t\}$ and a sequence $\{u_n\}$ satisfying $D(u_n)$ and $D'(u_n)$ and $\tau \in [0, \infty)$, the Poisson approximation

$$n\bar{F}(u_n) \to \tau \iff \mathbb{P}\left(M_n \leq u_n\right) \to \mathrm{e}^{-\tau} \tag{2.3.10}$$

for the maximum $M_n = \max\{\xi_1, \ldots, \xi_n\}$ holds. In particular

$$a_n^{-1}(M_n - d_n) \xrightarrow[(n\to\infty)]{\mathrm{d}} H \tag{2.3.11}$$

with $H \in \{\Phi_\alpha, \psi_\alpha, \Lambda\}$.

Summarizing, if a stationary sequence $\{\xi_t\}$ has a common distribution F in the maximum domain of attraction of an extreme value distribution H, i.e. there exist constants $a_n > 0$, $d_n \in \mathbb{R}$ such that

$$n\bar{F}(a_n x + d_n) \xrightarrow[(n\to\infty)]{} -\log H(x) \; \forall x \in \mathbb{R}$$

and the sequence $u_n = a_n x + d_n$ satisfies $D(u_n)$ and $D'(u_n)$, the maximum of $\{\xi_t\}$ exhibits exactly the same limit behavior with the same norming constants as the associated iid sequence. However, it is generally tedious to verify the asymptotic independence and especially the anti-clustering condition.
In case of a normal stationary sequence the only condition for $D(u_n)$ and $D'(u_n)$ to hold for $u_n = a_n x + d_n$ with norming constants as in Proposition 2.3.7 is the following. For the covariance sequence $\gamma_k = \mathrm{Cov}(\xi_t, \xi_{t+k})$, $k \geq 0$ it must apply that

$$\gamma_k \log k \to 0 \quad \text{as } k \to \infty. \tag{2.3.12}$$

This implication is thoroughly treated in (Leadbetter et al., 1983, Chap. 4.2/4.4) and (Galambos, 1987, Chap. 3.8) by means of the normal comparison lemma. The main concept there is that Gaussian multivariate distributions are close when their covariances are close. The decay condition (2.3.12) is similar to commonly known mixing conditions.

Remark 2.3.10

(1) The most simple dependence restriction is that of *k-dependence* where the random variables ξ_i and ξ_j are actually independent if $|i - j| > k$. The first basic results in this special case are due to Watson (1954) and Newell (1964). A less restrictive dependence assumption is that of *strong mixing* in which case the dependence between events generated by k-lagged random variables decays monotonically with increasing k. All of the aforementioned dependence restrictions fulfill conditions $D(u_n)$ and $D'(u_n)$ and therefore yield max-stable sequences.

(2) Apart from conditions on stationary sequences which guarantee maxima limit laws of the common type, other sequence types are also interesting. The case of non-stationary normal sequences can be regarded as adding a trend or seasonal component to a stationary normal sequence. By posing certain conditions on this additional terms, similar results as in the stationary case arise (see Leadbetter et al., 1983, Chap. 6.1). In case of strongly dependent normal sequences the weakening of the dependence decay conditions yields maximal distributions which differ from the simple three extremal types. If the condition $\gamma_k \log k \to \gamma > 0$ is fulfilled for the covariance sequence $\{\gamma_k\}$ of a stationary normal sequence, then the normalized maximum doesn't tend to the Gumbel distribution Λ, but to a convolution of Λ and a normal distribution. Further if $\gamma_k \log k \to \infty$ in a sufficiently smooth manner (i.e. $\gamma_k \to 0$ still), then a different normalization leads to distributional convergence to a normal distribution (for details see Leadbetter et al., 1983, Chapters 6.4–6.6, and Galambos, 1987, Theorems 3.8.1–3.8.4).

For our purposes it will suffice to show that a certain maximum likelihood estimator sequence is stationary, normal and k-dependent with some fixed k independent of the

sample size. As the observation horizon grows, that is the sample size goes to infinity, the Gumbel extremal distribution appears as limiting distribution. This will be exploited in Section 4.3 to construct asymptotical homogeneity tests.

3 GARCH Time Series

This chapter is devoted to suitable time series models in the context of financial data. We will especially focus on the statistical aspects concerning existence of *stationary* solutions, appropriate parameter *estimators* and the *finite-dimensional distributions* of the time series. We will limit our description to *autoregressive conditionally heteroscedastic* models as introduced by Engle (1982) and their *generalized* version due to Bollerslev (1986).
In Section 3.1 the basic concepts will be introduced in a univariate set-up. For an in-depth understanding of the said statistical implications we will confine ourselves to the most simple parametrization of the models, and formulate the respective results. In Section 3.2 a brief overview of possible extensions of the univariate model to higher dimensions will be given. With focus on suitable inference techniques, most of these generalizations are intractable and the verification of required assumptions can be tedious in many cases. The *semiparametric copula-based multivariate dynamic* model due to Chen and Fan (2006) turns out to be parsimonious in the parametrization and tractable when it comes to estimation. We will conclude with the distributional result by Chan et al. (2009a) concerning the resulting copula parameter estimates which forms the basis for the asymptotical tests in Chapter 4. In Section 3.3, we finally give an overview of time-dynamic copula models proposed in recent literature. These models are for the most part embedded in financial applications, however there the main focus is on the modeling of the copula dynamics. This section is meant to shed light on possible modeling approaches whenever the time-homogeneity of dependence through copulas should be rejected in favor of dynamic methods.

Before we go into the statistical details of modeling financial time series, let us first give the economical motivation for the need of special time series models.
Usually, we denote by $\{S_t\}_{t\geq 1}$ the raw time series of a financial price process, for example S_t could be the (daily) closing prices of an equity or of some financial option. Due to the inflation as time goes by, it can be observed that prices tend to have an exponential trend. Therefore, they are far from being stationary and will rarely be modeled directly with methods from time series analysis. To arrive at a stationary model we usually work with the compounded return of the underlying series. The so-called *logreturns* are defined to be

$$R_t := \log\left(\frac{S_t}{S_{t-1}}\right). \tag{3.0.1}$$

Note that $R_t = \log(S_t/S_{t-1}) \approx (S_t - S_{t-1})/S_{t-1}$ such that the logreturns correspond approximatively to one step ahead relative returns. As a result, they are independent of any monetary unit and different return series are comparable.

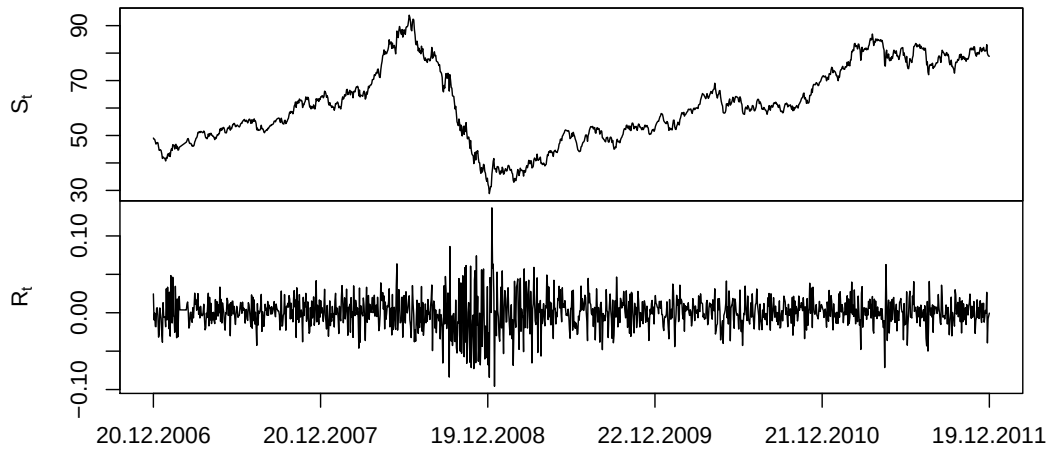


Figure 3.1: Raw financial time series and the corresponding logreturns (in €)

In the upper panel of Figure 3.1 we can see the raw time series S_t of the price process of one of the commodity futures that will be empirically analyzed in Chapter 5. In the lower panel the transformed logreturn series R_t is shown.
As working hypothesis, we can presume that logreturn series are at least mean-stationary (we will clarify the notion of *stationarity* in Definition 3.1.1). But on the other hand, R_t exhibits some features that require a different modeling as in classical linear approaches. At first glance, the volatility, as the variance of return series is usually called, is not constant over time. Besides this so-called *heteroscedasticity*, there are realizations that exceed a normal level. This hints at a finite dimensional distribution that has heavier tails than a normal distribution; a fact that can be observed when comparing a kernel density estimate in the upper left panel of Figure 3.2 to a corresponding normal density. Therefore, the distribution of returns is called *leptokurtic*. Furthermore, whenever such exceedances occur they seem to appear in clusters; this coined the term *volatility clusters* in financial time series. Finally, the upward and downward movements of the series do not appear at the same frequency, therefore the finite dimensional distributions are also *skewed* as opposed to symmetric normal distributions. Among others, the mentioned properties are summarized as *stylized facts* of financial returns. These qualitative properties of asset returns are observed empirically across a wide range of asset classes and are therefore assumed to be generally present. There have been many publications exploring the empirical properties of financial data, for an extensive overview including further stylized facts see Cont (2001) or Sewell (2011).
We can identify the changing volatility as the primary task in modeling and therefore set-up a time series model where the noise terms (also called innovations or residuals) come in in a multiplicative rather than in an additive way. Generally, we aim at the

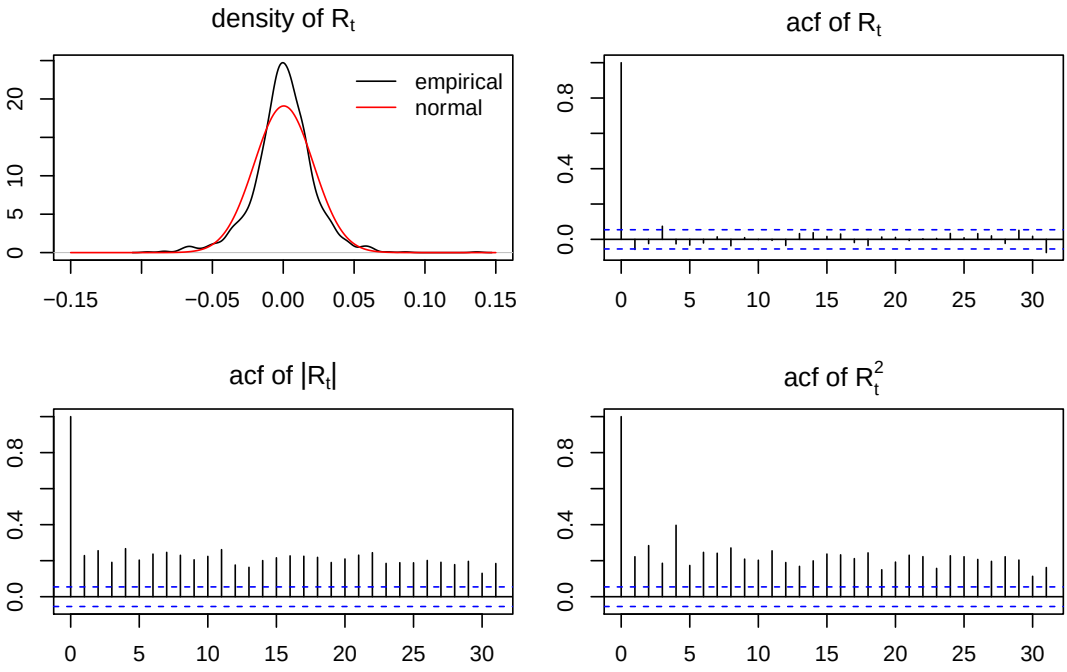


Figure 3.2: Smoothed density of logreturns and autocorrelation functions (acf)

following *conditional heteroscedasticity model*

$$\begin{aligned} R_t &= \mu_t + \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \mathrm{Var}\left(R_t \,|\mathcal{F}_{t-1}\right), \\ \mu_t &= \mathbb{E}\left(R_t\right), \end{aligned} \tag{3.0.2}$$

where $\varepsilon_t \sim \mathrm{WN}(0,1)$ is a process of uncorrelated random variables ε_t with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathrm{Var}(\varepsilon_t) = 1$ for all $t \geq 1$ (a so-called *white noise* process), and $\mathcal{F}_{t-1} = \sigma(R_{t-1}, R_{t-2}, \ldots)$ is the σ-algebra generated by the past observations up to R_{t-1}.
For statistical modeling the consideration of *autocorrelation* also indicates that the classical autoregression and moving average models are not suitable in capturing return properties. Autocorrelation of a stationary time series R_t means the correlation of R_t and R_{t+h} for a time lag $h > 0$; for a precise definition we refer to Definition 3.1.1(3). As illustrated in the upper right panel of Figure 3.2 the return series R_t itself does not seem to be autocorrelated, whereas $|R_t|$ and R_t^2 exhibit significant autocorrelation. Before the autocorrelation in the square returns will be exploited to specify (3.0.2) in the subsequent section, let us mention some consideration due to Shephard (1996) concerning two competing concepts in modeling changing volatility in the set-up $Y_t = \sigma_t \varepsilon_t$ (see above but with mean $\mu_t = 0$ for all t).

(1) On the one hand, the volatility can be specified in an *observation driven* way as

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \ldots + \alpha_q Y_{t-q}^2.$$

This leads to the ARCH-type models elaborated in Definition 3.1.3. The advantages of this approach are threefold: First, the one-step-ahead specification is very much in line with the actually present information structure in economic markets. Second, the same specification allows to obtain the likelihood functions through a prediction decomposition. Last, but not least, this modeling is parallel to the well-studied classical ARMA-type models so that certain techniques can be reused here.

(2) On the other hand, a *parameter driven* approach leads to

$$\log \sigma_t^2 = \gamma_0 + \gamma_1 \log \sigma_{t-1}^2 + \eta_t, \quad \eta_t \sim \mathrm{WN}(0, \sigma_\eta^2)$$

where σ_η^2 denotes the variance of the log-volatility $\log \sigma_t^2$. The major drawback of this model is that the log-volatility is not directly observable, but driven by some latent component. From a statistical point of view this implies that the one-step-ahead forecast densities are usually not available in an analytical way.

However, the need for tractability in a statistical sense gave rise to filtering techniques; the most famous being perhaps the *Kalman filter* (see Kálmán, 1960). Furthermore, the parameter driven models are sometimes easier to understand and to generalize to higher dimensions from an econometric point of view. Finally they posses simpler continuous-time analogues which is an important feature in modern financial theory where diffusions are preferably used.

Filtering applications are still an active field of research, see for example the dissertation of Rahe (2012) where the market price of risk is estimated by using the unscented Kalman filter in the Heston stochastic volatility model.

3.1 Univariate models

After the motivation in the introduction above, we will define the needed concepts more precisely in the following. We will concentrate on the model for the stochastic volatility, that is we will only consider the time series $Y_t = \sigma_t \varepsilon_t, t \geq 1$. Recall that a univariate *stochastic process* is a family of random variables $\{Y_t\}_{t \in I}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ measurable with respect to the product σ-algebra $\otimes_{t \in I} \mathcal{F}$ where I is an arbitrary index set for the times t. A *realization* or *path* of the stochastic process is the family $\{Y_t(\omega^*)\}_{t \in I}$ for a fixed $\omega^* \in \Omega$. Instead of specifying ω^* for a path, the realizations are often given in lower case letters. In our considerations I is finite and the observations are equidistant, therefore a path will be denoted by $y_1, \ldots, y_T$ with $T = |I|$.
The most important index sets are $I \subseteq \mathbb{R}$ (in which case the stochastic process is called *time-continuous*) and $I \subseteq \mathbb{Z}$ (then the process is called *time-discrete*). Time-discrete processes are usually named *time series*. Further definitions are given in the following.

Definition 3.1.1
Let $\{Y_t\}_{t\in I}$ be a stochastic process with $\mathbb{E}\left(Y_t^2\right) < \infty$ for all $t \in I$.

(1) The real-valued function $\mu\colon I \to \mathbb{R}$ with $\mu(t) = \mathbb{E}\left(Y_t\right)$, $t \in I$ is called the *mean function* of the stochastic process $\{Y_t\}_{t\in I}$.

(2) The real-valued function $\gamma\colon I \times I \to \mathbb{R}$ defined by $\gamma(s,t) = \mathrm{Cov}\left(Y_s, Y_t\right)$, $s,t \in I$ is called the *autocovariance function* of $\{Y_t\}_{t\in I}$.

(3) By $\rho\colon I \times I \to \mathbb{R}$ with $\rho(s,t) = \mathrm{Corr}\left(Y_s, Y_t\right) = \frac{\gamma(s,t)}{\sqrt{\mathrm{Var}(Y_s)}\sqrt{\mathrm{Var}(Y_t)}}$ we define the *autocorrelation function* of $\{Y_t\}_{t\in I}$.

(4) A stochastic process is called *(weakly) stationary* if

(i) $\mu(t) \equiv \mu$ for all $t \in I$,

(ii) $\gamma(s,t) = \gamma(s+r, t+r)$ for all $r,s,t \in I$ such that $s+r, t+r \in I$.

Especially, in this case the autocovariance function can then be written as $\gamma(h) = \gamma(0,h)$, $h \in I$, i.e. as a function of merely the time lag $h = t - s$.

(5) Denote by $F_{Y_{t_1},\dots,Y_{t_n}}(y_1,\dots,y_n) = \mathbb{P}\left(Y_{t_1} \le y_1, \dots, Y_{t_n} \le y_n\right)$ the finite-dimensional distribution of $Y_{t_1},\dots,Y_{t_n}$ for any $n \in \mathbb{N}$, $y_i \in \mathbb{R}$, $i = 1,\dots,n$. Then the stochastic process is called *strictly stationary* if for all $h \in I$

$$F_{Y_{t_1},\dots,Y_{t_n}} = F_{Y_{t_1+h},\dots,Y_{t_n+h}},$$

i.e. the finite-dimensional distributions are invariant under joint shifts of length h.

A class of models for stationary time series are *linear processes*, i.e. stochastic processes $X_t = \sum_{\nu\in\mathbb{Z}} \alpha_\nu \varepsilon_{t-\nu}$, $t \in \mathbb{Z}$ with a real sequence $\{\alpha_\nu\} \in \ell_1$ and a white noise process $\varepsilon_t \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$. The ARMA$(p,q)$ models are well-studied specifications within this class and defined as the solution of

$$X_t - \sum_{\nu=1}^{p} \alpha_\nu X_{t-\nu} = \varepsilon_t + \sum_{\mu=1}^{q} \beta_\mu \varepsilon_{t-\mu} \tag{3.1.1}$$

with iid white noise ε_t, $p,q \ge 0$ and $\alpha_\nu, \beta_\mu \in \mathbb{R}$. Define $\Phi(z) := 1 - \sum_{\nu=1}^{p} \alpha_\nu z^\nu$ and $\Psi(z) := 1 + \sum_{\mu=1}^{q} \beta_\mu z^\mu$, then for a stationary solution it is necessary and sufficient that Φ has no complex roots in $|z| \le 1$, if Φ and Ψ have no zeros in common. In this case, X_t can be expressed as linear process $X_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$ where $\{c_j\}$ is the sequence of the (absolutely summable) Taylor coefficients of $\Psi(z)/\Phi(z)$ at $z_0 = 0$. For further details on linear and ARMA processes consult for example Brockwell and Davis (2006).

Definition 3.1.2
For a linear process $\{X_t\}_{t\in\mathbb{Z}}$ let $\hat{X}_{k+1} = \sum_{j=1}^{k} \phi_{kj} X_{k+1-j}$ be the best linear prediction of X_{k+1} by $X_1,\dots,X_k$. The sequence $\{\phi(k)\}_{k\in\mathbb{N}}$ defined by $\phi(k) = \phi_{kk}$ is called the *partial autocorrelation function* of the process.

Generally speaking, $\phi(k+1)$ gives the correlation between X_{k+1} minus its linear prediction based on $X_1, \dots, X_k$ and X_0 minus its linear prediction based on $X_1, \dots, X_k$.
A first step towards the identification of the orders p and q given a data set is to consider the empirical counterparts of the autocorrelation and the partial autocorrelation function (see Figure 3.2 for the empirical autocorrelation). Whenever $\phi(k)$ disappears abruptly at lag $k = p^*$ and $\gamma(k)$ exhibits exponential decay, this hints at an ARMA$(p^*, 0)$ model, also called *autoregression* of order p^*. Whereas, if $\phi(k)$ exhibits exponential decay and $\gamma(k)$ disappears abruptly at lag $k = q^*$, this is an indicator for an ARMA$(0, q^*)$ model, also denoted as *moving average* of order q^*. Time series possessing actually both components are not that clearly identifiable anymore as the decay and the disappearance superpose each other.
Based on this identification strategy and due to the fact that the squares of financial returns exhibit a similar behavior as autoregressive processes, Engle (1982) specified the conditional variance in (3.0.2) as autoregression, and Bollerslev (1986) generalized this idea further. Let us consider a precise definition.

Definition 3.1.3
Let $p, q \geq 0$ be given. A real-valued time series $\{Y_t\}_{t\in\mathbb{Z}}$ defined by $Y_t = \sigma_t \varepsilon_t$ with $\varepsilon_t \sim$ WN(0,1) iid and $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ a stochastic process with coefficients $\alpha_0, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q \geq 0$, $\alpha_p, \beta_q \neq 0$ satisfying the difference equation

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2 \tag{3.1.2}$$

is called GARCH(p, q) process (stemming from *Generalized AutoRegressive Conditionally Heteroscedastic*). $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ is called the *volatility process*. If $q = 0$, $\{Y_t\}_{t\in\mathbb{Z}}$ is called an ARCH(p) process.

The specification in (3.1.2) is sufficient to account for variance inhomogeneity, volatility clusters and heavy-tailed finite-dimensional distributions (as we will see shortly). However, as outlined in the introduction, sometimes it is required to model skewness and asymmetries. This can for example be accomplished by the approach of exponential GARCH; this specification will be introduced in the empirical part in Chapter 5. There are many more possibilities to specify the variance evolution; for a very broad overview on the flavors of GARCH see Bollerslev (2008).
In the three subsequent sections we will give the respective stationary, consistency and tail results in case of a GARCH(1,1) time series Y_t. With white noise $\varepsilon_t \sim$ WN(0,1) and parameters $\alpha_0, \beta_1 \geq 0, \alpha_1 > 0$ we consider

$$Y_t = \sigma_t \varepsilon_t, \tag{3.1.3}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{3.1.4}$$

On the one hand, this serves the notational clarity and the possibility to sketch the proofs. On the other hand, the GARCH(1,1) specification proved to be adequate for most applications and the gain in accuracy of higher-order models is disproportional to the rising complexity of the parametrization (see Hansen and Lunde, 2001). For completeness,

we will shortly indicate how to generalize the respective results to GARCH(p,q) and provide the corresponding references.

3.1.1 Stationarity and moment conditions

Due to the definition in (3.1.2) and the white noise property of the residuals ε_t we have that σ_t^2 only depends on ε_s, $s<t$ and therefore that σ_t^2 and ε_t are independent. Furthermore, σ_t^2 is measurable with respect to the σ-algebra $\mathcal{F}_{t-1}=\sigma(\varepsilon_s\colon s\leq t-1)$. Consider $\eta_t=Y_t^2-\sigma_t^2$. Then by conditioning on $\mathcal{F}_{t-1}$ it can be shown that

$$\mathbb{E}\left(\eta_t\right)=\mathbb{E}\left(\sigma_t^2(\varepsilon_t^2-1)\right)=\mathbb{E}\left(\sigma_t^2\mathbb{E}\left(\varepsilon_t^2-1\,|\mathcal{F}_{t-1}\right)\right)=\mathbb{E}\left(\sigma_t^2\cdot 0\right)=0$$

and for $s\leq t$ that

$$\mathbb{E}\left(\eta_s\eta_t\right)=\begin{cases}\mathbb{E}\left(\sigma_s^2(\varepsilon_s^2-1)\sigma_t^2\,\mathbb{E}\left(\varepsilon_t^2-1\,|\mathcal{F}_{t-1}\right)\right)=0, & s<t\\ \mathbb{E}\left(\sigma_t^4\,\mathbb{E}\left(\left(\varepsilon_t^2-1\right)\right)^2\right)<\infty, & s=t.\end{cases}$$

For the validity of the second case it is required that $\mathbb{E}\left(\sigma_t^4\right)<\infty$ and $\mathbb{E}\left(\varepsilon_t^4\right)<\infty$. Then $\{\eta_t\}$ fulfills the white noise property of zero mean, finite variance and uncorrelatedness. By straightforward calculations it can be shown that Y_t^2 has a representation as ARMA(1,1) process, namely

$$Y_t^2=\alpha_0+(\alpha_1+\beta_1)Y_{t-1}^2-\beta_1\eta_{t+1}+\eta_t$$

with $\eta_t\sim$ WN(0,1). Furthermore, we have $\mathbb{E}\left(Y_t\right)=0\ \forall t$ and $\mathbb{E}\left(Y_sY_t\right)=\mathbb{E}\left(\sigma_t^2\right)\mathbb{1}_{\{s=t\}}$ $\forall s\leq t$. This implies that if $\mathbb{E}\left(\sigma_t^2\right)<\infty$ then Y_t is uncorrelated, and if $\mathbb{E}\left(\sigma_t^2\right)\equiv const$ then Y_t is weakly stationary.
Deriving permissible parameterizations under which σ_t^2 is strictly stationary and possesses moments up to a certain order is less trivial. These conditions will then imply strict stationarity and existence of moments of Y_t, as well.
The following result is due to Nelson (1990). Besides referencing the corresponding theorems therein, we will sketch the proof to get an idea about the required conditions.

Theorem 3.1.4
Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a GARCH(1,1) process and $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ the respective volatility process.

(1) Then both (3.1.3) and (3.1.4) have a strictly stationary solution if, and only if,

$$\mathbb{E}\left(\log(\alpha_1\varepsilon_1^2+\beta_1)\right)<0. \tag{3.1.5}$$

In this case, $\sigma_t^2=\alpha_0(1+\sum_{j=1}^{\infty}\prod_{\nu=1}^{j}(\alpha_1\varepsilon_{t-\nu}^2+\beta_1))$, hence σ_t^2 is measurable with respect to $\mathcal{F}_{t-1}$, and the unique strictly stationary solution of (3.1.4).

In particular, $\alpha_1+\beta_1<1$ is sufficient for condition (3.1.5), such that in this case

$$\mathbb{E}\,X_t^2=\mathbb{E}\,\sigma_t^2=\frac{\alpha_0}{1-\alpha_1-\beta_1}.$$

(2) If additionally $\alpha_0 > 0$ and $\mathbb{E}\,|\varepsilon_t|^{2q} < \infty$ for some $q > 0$, then there exists a p with $0 < p < q$ and $\mathbb{E}\left((\alpha_1\varepsilon_1^2 + \beta_1)^p\right) < 1$. If, moreover, we have that

$$\mathbb{E}\left((\alpha_1\varepsilon_1^2 + \beta_1)^r\right) < 1 \tag{3.1.6}$$

for $0 < r < q$, then there exists $\delta > 0$ such that $\mathbb{E}\left((\alpha_1\varepsilon_1^2 + \beta_1)^{r+\delta}\right) < 1$. Especially, this implies that Y_t possesses moments of order $2r$.

Proof

(1) The sufficiency of (3.1.5) under the condition that $\alpha_1 + \beta_1 < 1$ is shown right away through Jensen's inequality

$$\mathbb{E}\left(\log(\alpha_1\varepsilon_1^2 + \beta_1)\right) \leq \log(\alpha_1\mathbb{E}\,\varepsilon_1^2 + \beta_1) = \log(\alpha_1 + \beta_1) < 0.$$

By application of Tonelli's theorem we see that

$$\begin{aligned}\mathbb{E}\,X_t^2 = \mathbb{E}\,\sigma_t^2 &= \mathbb{E}\left(\alpha_0(1 + \sum_{j=1}^{\infty}\prod_{\nu=1}^{j}(\alpha_1\varepsilon_{t-\nu}^2 + \beta_1))\right) \\ &\leq \alpha_0(1 + \sum_{j=1}^{\infty}\prod_{\nu=1}^{j}(\alpha_1\mathbb{E}\,\varepsilon_{t-\nu}^2 + \beta_1)) = \alpha_0(\sum_{j=0}^{\infty}(\alpha_1 + \beta_1)^j) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}.\end{aligned}$$

For the necessity assume that $\mathbb{E}\left(\log(\alpha_1\varepsilon_1^2 + \beta_1)\right) < 0$ holds true. The GARCH equations can be written in differential form, i.e. for $t \geq 1$

$$Y_t = (\alpha_1\varepsilon_{t-1}^2 + \beta_1)Y_{t-1}, \tag{3.1.7}$$

$$\sigma_t^2 = \alpha_0 + (\alpha_1\varepsilon_{t-1}^2 + \beta_1)\sigma_{t-1}^2. \tag{3.1.8}$$

From the general theory of difference equations (or perpetuities as they are sometimes called) we see that (3.1.7) is in homogeneous form and in particular possesses the solution $Y_t = \prod_{j=1}^{t}(\alpha_1\varepsilon_{t-j}^2 + \beta_1)Y_0$ for an arbitrary Y_0. Moreover, $Y_t \to 0$ almost surely as $t \to \infty$ since the following strong law of large numbers holds. There exists $\rho > 1$ such that

$$\frac{1}{t}\sum_{j=1}^{t}[\log\rho + \log(\alpha_1\varepsilon_{t-j}^2 + \beta_1)] \xrightarrow[(t\to\infty)]{\text{a.s.}} \log\rho + \mathbb{E}\left(\log(\alpha_1\varepsilon_1^2 + \beta_1)\right).$$

This is equivalent to $\prod_{j=1}^{t}(\alpha_1\varepsilon_{t-j}^2 + \beta_1)) \xrightarrow{\text{a.s.}} 0$ with geometric speed of convergence as $t \to \infty$, i.e. $\prod_{j=1}^{t}(\alpha_1\varepsilon_{t-j}^2 + \beta_1)) \leq c\rho^{-t}$ with a constant $c \in \mathbb{R}$ and $\rho > 1$. Thus, the infinite sum in the formula for σ_t^2 converges, and therefore σ_t^2 is well-defined. Moreover, it solves (3.1.4) which can be checked by plugging it in. As the ε_t are iid, it is also a strictly stationary solution and also unique.

A k-fold iteration of (3.1.4) leads to

$$\sigma_t^2 \geq \alpha_0(1 + \sum_{j=1}^{k} \prod_{\nu=1}^{j} (\alpha_1 \varepsilon_{t-\nu} + \beta_1)). \tag{3.1.9}$$

Now let us assume that (3.1.5) is violated: Either $\mathbb{E}\left(\log(\alpha_1 \varepsilon_1^2 + \beta_1)\right) > 0$ which implies that (3.1.9) tends to $+\infty$ as $t \to \infty$ by similar arguments as in the application of the strong law above. Or it holds that $\mathbb{E}\left(\log(\alpha_1 \varepsilon_1^2 + \beta_1)\right) = 0$ leading to the same contradiction as before as can be shown with the help of a more detailed analysis using random walks (see the proof of the third case in Theorem 1 in Nelson, 1990). In both cases $\sigma_t^2 = \infty$ almost surely.

(2) The first two claims correspond to Theorem 4 in Nelson (1990). The latter is a consequence of (29) in the Corollary to Theorem 3 in Nelson (1990). The proofs are based on Minkowski-type inequalities for Stieltjes integrals for the most part.

□

Note that the condition $\alpha_1 + \beta_1 < 1$ is not necessary for strict stationarity. There exist parameterizations of (3.1.4) that yield strictly stationary solutions and an infinite variance of Y_t at the same time. The condition for covariance stationarity is similar for GARCH(p, q) models, namely that $\sum_{i=1}^{p} \alpha_i + \sum_{k=1}^{q} \beta_k < 1$ implies $\mathbb{E}\, Y_t^2 = \alpha_0/(1 - \alpha_1 - \ldots - \alpha_p - \beta_1 - \ldots - \beta_q) < \infty$ (see Bollerslev, 1986).
Concerning strict stationarity in general GARCH(p, q) models, Bougerol and Picard (1992a) derive that strict negativity of the top Lyapunov exponent associated with matrices in a perpetuity representation of Y_t similar to (3.1.7) is a necessary and sufficient condition. The reasoning is based on their general result for generalized autoregressive processes in Bougerol and Picard (1992b).
The stationarity condition (3.1.5) and the moment condition (3.1.6) are at best tedious to verify, usually even impractical for deriving permissible parameter regions. However, in case of standard normally and Cauchy distributed residuals ε_t the boundary conditions on the GARCH(1,1) parameters can be given in analytical form. The specific forms in terms of various types of hypergeometric functions among others are elaborated in Nelson (1990). We will restrict ourselves to reproducing an illustrative display of the respective parameter regions given therein for residuals $\varepsilon_t \sim \mathcal{N}(0,1)$ iid. In Figure 3.3 we consider $(\alpha_1, \beta_1) \in (0, \infty) \times [0, \infty)$ under the condition $\alpha_0 > 0$.
For α_1 and β_1 in the rightmost grey-shaded region, condition (3.1.5) is violated, i.e. the GARCH(1,1) equation (3.1.4) has no (strictly) stationary solution. In the adjacent region of vertical blue lines, we have strict stationarity of σ_t^2 and hence of Y_t, but Y_t does not possess first moments. For parameters in the blue checked region, we have strong stationarity and first moments, but still infinite variance of Y_t. The classical, covariance stationary GARCH(1,1) models have parameters in the checks-with-diagonals area of Figure 3.3. Note that the boundaries of the different regions are attributed to the respective righthand area.
For $\beta_1 = 0$ the GARCH(1,1) model reduces to an ARCH(1) model with $Y_t = \sigma_t \varepsilon_t, \sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2$, $\alpha_0 > 0, \alpha_1 \geq 0$. The condition for strict stationarity is then $\log(2\alpha_1) +$

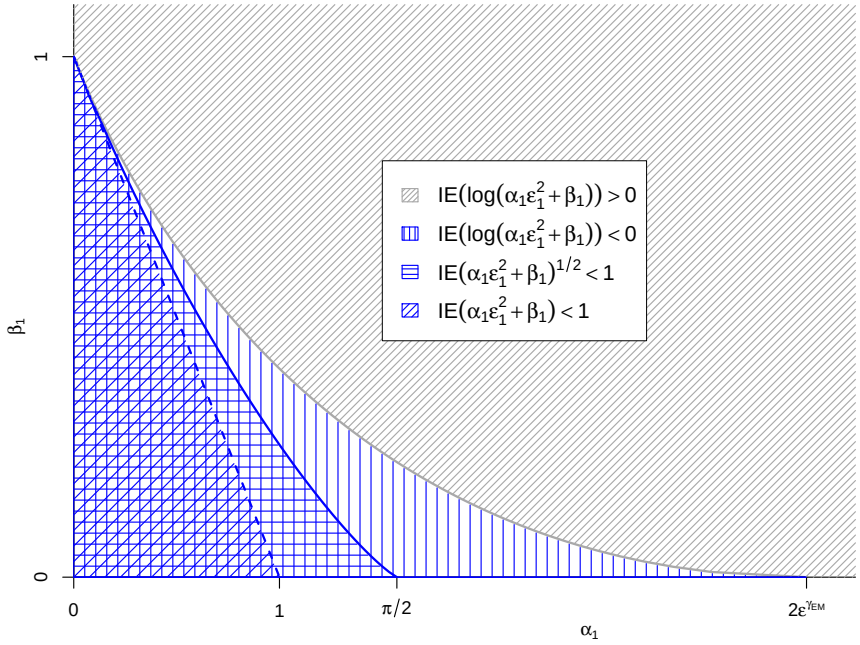


Figure 3.3: Parameter regions of $\alpha_1 > 0, \beta_1 \geq 0$ in GARCH(1,1) , $\varepsilon_t \sim \mathcal{N}(0,1)$ iid

$\psi(1/2) < 0$ where $\psi(\,.\,)$ is the Euler psi function with $\psi(1/2) = -2\log 2 - \gamma_{EM}$ and $\gamma_{EM} \approx 0.5772$ (the Euler-Mascheroni constant). This results in non-stationarity for $\alpha_1 > 2\,\mathrm{e}^{\gamma_{EM}} \approx 3.562$, and strict stationarity otherwise. In this case, first moments of Y_t exist if $\sqrt{2\alpha_1}/\sqrt{\pi}\Gamma(1) < 1$ where $\Gamma(\,.\,)$ is the Gamma function with $\Gamma(1) = 1$. This condition is equivalent to $\alpha_1 < \pi/2 \approx 1.571$ for $\mathbb{E}\,|Y_t| < \infty$. A covariance stationary solution exists if $0 < \alpha_1 < 1$. These results are also derived in Embrechts et al. (2003, Thm. 8.4.9).

3.1.2 Finite-dimensional distribution

When introducing conditional heteroscedasticity models in the original publications (see Engle, 1982; Bollerslev, 1986) the distribution of the residuals $\{\varepsilon_t\}$ was presumed to be standard normal. The normal distribution is probably the most prominent representative of *light-tailed* distributions in the sense of exponentially decaying tails. To be more precise, if there exists $s > 0$ such that $\mathrm{e}^{sx}\,\bar{F}(x) \to 0$ for $x \to \infty$ and $\mathrm{e}^{-sx}\,F(x) \to 0$ for $x \to -\infty$ for a distribution function F, then $X \sim F$ is said to have *light tails*. This especially implies that X possesses moments of every order $h_0 \leq \infty$. As for *heavy-tailed* distributions this notion is often associated with regular variation of the respective tail distribution function. If for $X \sim F$ there exists $\alpha > 0$ such that $lim_{x\to\infty}\bar{F}(tx)/\bar{F}(x) = t^{-\alpha}$ for all $t > 0$, then $\bar{F}$ is called *regularly varying with index* $-\alpha$. Then X is said to be regularly varying with index α and thus possesses

heavy tails. Recently, the dissertation of Janßen (2010) deals with some aspects of the extremal behavior of light tailed and heavy tailed distributions and provides an intuitive introduction to regular variation and its generalization to higher dimensions.
We will now focus on the tail of the marginal distribution of a GARCH(1,1) process and clarify what the statement "light-tailed residuals cause heavy-tailed marginal distributions" within the scope of heteroscedastic models means. To this end, set $A_t = \alpha_1 \varepsilon_{t-1}^2 + \beta_1$ and $B_t = \alpha_0$ and consider

$$\sigma_t^2 = B_t + A_t \sigma_{t-1}^2, \tag{3.1.10}$$

which is a representation equivalent to (3.1.8). Since the ε_t are iid, $A_t \stackrel{d}{=} A := \alpha_1 \varepsilon_1^2 + \beta_1$ holds for all $t \in \mathbb{Z}$.
The next theorem as well as its proof are along the lines of Theorem 2.1 in Mikosch and Starica (2000).

Theorem 3.1.5
Assume that the law of $\log A$ is non-arithmetic, $\mathbb{E} \log A < 0$ (this is condition (3.1.5) in Theorem 3.1.4), $\mathbb{P}(A > 1) > 0$, and there exists $h_0 \leq \infty$ such that $\mathbb{E} A^h < \infty$ for all $h < h_0$ and $\mathbb{E} A^{h_0} = \infty$.

(1) The equation

$$\mathbb{E} A^{\kappa/2} = 1 \tag{3.1.11}$$

has a unique solution $\kappa > 0$.

(2) If $\alpha_0 > 0$ and κ solves (3.1.11), then there exist strictly stationary solutions $\sigma_t^2 \stackrel{d}{=} \sigma^2$ and $Y \stackrel{d}{=} Y_t$ of the GARCH(1,1) equations (3.1.8) and (3.1.7). For independent A and σ, there exists a constant

$$c = \frac{\mathbb{E}\left((\alpha_0 + A\sigma^2)^{\kappa/2} - (A\sigma^2)^{\kappa/2}\right)}{\kappa/2 \mathbb{E}\left(A^{\kappa/2} \log A\right)} \tag{3.1.12}$$

such that for $x \to \infty$

$$\mathbb{P}(\sigma > x) \sim c x^{-\kappa} \tag{3.1.13}$$

and

$$\mathbb{P}(|Y| > x) \sim \mathbb{E}|\varepsilon_1|^{\kappa} \, \mathbb{P}(\sigma > x). \tag{3.1.14}$$

The expression $f(x) \sim g(x)$ as $x \to \infty$ denotes asymptotic equivalence of f and g in the sense that $\lim_{x \to \infty} f(x)/g(x) = 1$.

Proof

(1) Consider the function $h(u) = \mathbb{E} A^u$ which can be shown to be convex in u. $\mathbb{E} \log A < 0$ implies that h attains values smaller than 1 in the neighborhood of $u = 0$. Furthermore, $\mathbb{E} A^u \geq 1$ for sufficiently large u since $\mathbb{P}(A > 1) > 0$ and $\mathbb{E} A^{h_0} = \infty$. Therefore, $\mathbb{E} A^{\kappa/2} = 1$ has a unique positive solution.

(2) (3.1.13) is basically a consequence of Kesten (1973, Thm. 5) elaborated in Goldie (1991, Thm. 4.1) for the univariate perpetuities in case of GARCH(1,1) presented here. Our conditions ensure that $\mathbb{E}\, A^{\kappa/2+\delta} < \infty$ for small $\delta > 0$ and therefore all needed prerequisites to apply the mentioned theorem are fulfilled.

(3.1.14) is a consequence of the following result: Let ξ and ζ be two independent non-negative random variables such that $\mathbb{P}(\xi > x) = L(x)x^{-\kappa}$ with some regularly varying function L with index 0 (L is then also called *slowly varying*, and is extremely heavy-tailed), and $\mathbb{E}\,\zeta^{\kappa+\delta} < \infty$ for some $\delta > 0$. Then we have $\mathbb{P}(\zeta\xi > x) = \mathbb{E}\,\zeta^{\kappa}\mathbb{P}(\xi > x)$ as $x \to \infty$.

□

Not restricting ourselves to standard normal residuals, we formulated the result for residuals having moments of at least order h_0. The case of $\varepsilon_t \sim \mathcal{N}(0,1)$ and $\beta_1 = 0$ (i.e. an ARCH(1) model) is one of the rare situations where it is possible to determine the constant c explicitly (see Embrechts et al., 2003, Thm. 8.4.12). There also the proof of Pareto tails of Y_t (the tail behavior in (3.1.13) and (3.1.14) is named as such) is elaborated in more detail.
For a similar result in case of the general GARCH(p, q) model see Basrak et al. (2002). Here the GARCH equations are reformulated in terms of difference equations (perpetuities) and the claim of regularly varying tails is established with the general renewal theory for random difference equations due to Kesten (1973).

3.1.3 Quasi maximum likelihood estimation

In this section we plan to introduce an estimation procedure for the parameters in a GARCH(p, q) model. The first step towards this, is to specify conditions that ensure the identifiability of the true parameter set $\gamma = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$.

Theorem 3.1.6
Let $Y_t = \sigma_t\varepsilon_t$ be given according to (3.1.2) with iid residuals $\varepsilon_t \sim$ WN(0,1) and parameters $\alpha_0 > 0$, $\alpha_1, \ldots, \alpha_p \geq 0$ and $\beta_1, \ldots, \beta_q \geq 0$ such that a strictly stationary solution exists. It is sufficient for a unique parametrization of the model

(i) that the distribution function of ε_t^2 is not concentrated at one point, i.e. ε_t^2 is a non-degenerate random variable for all $t \in \mathbb{Z}$,

(ii) that $\alpha_p\beta_q > 0$, i.e. that the order of the model is actually p and q, and

(iii) that the polynomials $A(z) = \sum_{\nu=1}^{p} \alpha_\nu z^\nu$ and $B(z) = 1 - \sum_{\mu=1}^{q} \beta_\mu z^\mu$ have no joint zeros for all $z \in \mathbb{R}$.

Proof
Assume $\gamma = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$ and $\tilde{\gamma} = (\tilde{\alpha}_0, \tilde{\alpha}_1, \ldots, \tilde{\alpha}_p, \tilde{\beta}_1, \ldots, \tilde{\beta}_q)$ being two different parameterizations. Bougerol and Picard (1992a) showed that condition (3.1.5), necessary and sufficient for strict stationarity, yields $\sum_{k=1}^{q} \beta_k < 1$ and $\sum_{k=1}^{q} \tilde{\beta}_k < 1$ respectively. This in turn implies that $B(z) \neq 0$ and analogously $\tilde{B}(z) \neq 0$ for $|z| \leq 1$.

Hence $1/B(z)$ and $1/\tilde{B}(z)$ are holomorphic functions in $|z| \leq R$ with some (joint) $R > 1$ and can be expanded into a Taylor series around $z_0 = 0$. According to the classical theory for linear processes, this allows for two representations

$$\sigma_t^2 = c_0 + \sum_{j=1}^{\infty} c_j Y_{t-j}^2 = \tilde{c}_0 + \sum_{j=1}^{\infty} \tilde{c}_j Y_{t-j}^2$$

in terms of c_j and $\tilde{c}_j$ being the Taylor coefficients of $A(z)/B(z)$ and $\tilde{A}(z)/\tilde{B}(z)$ respectively. Whenever $i_0 \in \mathbb{Z}$ is the smallest index such that $c_{i_0} \neq \tilde{c}_{i_0}$, this implies that ε_{t-i_0} is constant, which contradicts the white noise property. Therefore, it has to be $A(z)/B(z) = \tilde{A}(z)/\tilde{B}(z)$. As the respective numerators and denominators do not have common zeros, $A(z)$ and $\tilde{A}(z)$ as well as $B(z)$ and $\tilde{B}(z)$ must have the same zeros. Due to (ii) this implies $A(z) = \tilde{A}(z)$ and $B(z) = \tilde{B}(z)$, and hence $\gamma = \tilde{\gamma}$. □

We will state consistency and asymptotic normality of a maximum-likelihood-type parameter estimator $\hat{\gamma}$ due to Berkes et al. (2003). Along the lines of their publication consider the following quantities under the assumption that $q \geq p$ (in the opposite case, the roles of p and q have to be reversed)

$$\begin{aligned}
d_0(\gamma) &= \frac{\alpha_0}{1 - \sum_{k=1}^{q} \beta_k} \\
d_1(\gamma) &= \alpha_1 \\
d_2(\gamma) &= \alpha_2 + \beta_1 d_1(\gamma) \\
&\vdots \\
d_p(\gamma) &= \alpha_p + \beta_1 d_{p-1}(\gamma) + \ldots + \beta_{p-1} d_1(\gamma) \\
d_{p+1}(\gamma) &= \beta_1 d_p(\gamma) + \ldots + \beta_p d_1(\gamma) \\
&\vdots \\
d_q(\gamma) &= \beta_1 d_{q-1}(\gamma) + \ldots + \beta_{q-1} d_1(\gamma)
\end{aligned}$$

and for $r > \max\{p, q\}$

$$d_r \gamma = \beta_1 d_{r-1}(\gamma) + \ldots + \beta_q d_{r-q}(\gamma).$$

Moreover, let the admissible parameter set be given as

$$\begin{aligned}
\Gamma := \Big\{ & \gamma = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q) \colon \alpha_i, \beta_j > 0, \\
& \sum_{j=1}^{q} \beta_j \leq \Delta_0^* < 1, 0 < \Delta_1^* \leq \min\{\gamma\} \leq \max\{\gamma\} \leq \Delta_2^*, q\Delta_1^* < \Delta_0^* \Big\}.
\end{aligned} \tag{3.1.15}$$

Then under the assumption that $\varepsilon_t \sim \mathcal{N}(0,1)$ iid for all $t \in \mathbb{Z}$, the distribution of Y_t/σ_t conditionally on the information up to time point $t-1$ given by $\sigma(\varepsilon_k \colon -\infty < k \leq t-1)$ is also standard normal. This can be used to derive a suitable log-likelihood function and to define the following estimator.

Definition 3.1.7
Set $w_t(\gamma) = d_0(\gamma) + \sum_{r=1}^{\infty} d_r(\gamma) Y_{t-r}^2$ for a GARCH(p,q) process $\{Y_t\}$ with parameter set $\gamma = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$. Then the *quasi maximum likelihood estimator* for γ is defined as

$$\hat{\gamma} = \underset{\gamma \in \Gamma}{\operatorname{argmax}} \; -\frac{1}{2} \sum_{t=1}^{T} \left(\log(w_t(\gamma)) + \frac{Y_t^2}{w_t(\gamma)} \right). \tag{3.1.16}$$

Even if the residuals are not standard normally distributed, the quasi maximum likelihood estimator can be shown to be strongly consistent and asymptotically normal. In the following theorem these results are given under the most general conditions known at present.

Theorem 3.1.8
Under the identifiability conditions in Theorem 3.1.6 and assuming that the true parameters lie in the interior of Γ, let $\hat{\gamma}$ be given as in (3.1.16).

(1) If $\mathbb{E}\,|\varepsilon_1^2|^{1+\delta} < \infty$ for some $\delta > 0$, $\lim_{t \to 0} t^{-\mu} \mathbb{P}\left(\varepsilon_1^2 \leq t\right) = 0$ for some $\mu > 0$ and $\mathbb{E}\,\varepsilon_1^2 = 1$, then as $T \to \infty$

$$\hat{\gamma} \xrightarrow{\text{a.s.}} \gamma. \tag{3.1.17}$$

(2) If additionally $\mathbb{E}\,|\varepsilon_1^2|^{2+\delta} < \infty$ for some $\delta > 0$, then as $T \to \infty$

$$\hat{\gamma} - \gamma = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{2} (\varepsilon_1^2 - 1) B^{-1} \frac{w_t'(\gamma)}{w_t(\gamma)} + o_P(T^{-1/2}), \tag{3.1.18}$$

where $w_t'(\gamma) = \left(\frac{\partial}{\partial \alpha_0} w_t(\gamma), \frac{\partial}{\partial \alpha_1} w_t(\gamma), \ldots, \frac{\partial}{\partial \alpha_p} w_t(\gamma), \frac{\partial}{\partial \beta_1} w_t(\gamma), \ldots, \frac{\partial}{\partial \beta_q} w_t(\gamma) \right)$ and

$$B = \mathbb{E}\left(\frac{w_1'(\gamma)}{w_1(\gamma)} \left(\frac{w_1'(\gamma)}{w_1(\gamma)} \right)^{\top} \right).$$

That is, we have the following distributional convergence

$$\sqrt{T}(\hat{\gamma} - \gamma) \xrightarrow{d} \mathcal{N}(\mathbf{0}, (\mathbb{E}\,{\varepsilon_1}^4 - 1) B^{-1}). \tag{3.1.19}$$

Proof

(1) See Theorem 4.1 in Berkes et al. (2003).

(2) See Theorem 4.2 in Berkes et al. (2003).

□

In practice, only realizations of $Y_1, \ldots, Y_T$ with finite T are available and the evaluation of the log-likelihood function in (3.1.16) will not be possible. Therefore we substitute

the infinite sums in the definition of the quasi maximum likelihood estimator by finite sums to arrive at a modified version of the estimate

$$\tilde{\gamma} = \underset{\gamma \in \Gamma}{\operatorname{argmax}} \; -\frac{1}{2} \sum_{t=1}^{T} \left(\log(\tilde{w}_k(\gamma)) + \frac{Y_t^2}{\tilde{w}_k(\gamma)} \right) \tag{3.1.20}$$

with $\tilde{w}_k(\gamma) = d_0(\gamma) + \sum_{r=1}^{k-1} d_r(\gamma) Y_{k-r}^2$. Using this approach we also arrive at estimators $\tilde{\gamma}$ that are strongly consistent (see Theorem 4.3 in Berkes et al., 2003) and asymptotically normal at the same rate and with the same limiting variance (see Theorem 4.4 in Berkes et al., 2003).
Under identifiability it holds that $\sigma_t^2 = w_t(\gamma)$. As shown in Lemma 5.8 in Berkes et al. (2003) under the conditions of Theorem 3.1.8(1) there exists $\delta > 1$ such that

$$\delta^k |w_t(\gamma) - \tilde{w}_t(\gamma)| = \mathcal{O}_P(1),$$

i.e. the approximation $\tilde{\sigma}_t^2 = \tilde{w}_t(\gamma)$ converges in probability to the true, but usually unknown, conditional variance σ_t^2 exponentially fast. This can now be used to assess another quantity that is not directly observable, namely the residuals ε_t, through the technique of *deGARCHing*. Set $\tilde{w}_1(\gamma) = 1$ and the empirical conditional variance $\tilde{\sigma}_t^2 = \tilde{w}_t(\hat{\gamma})$ with a quasi maximum likelihood estimator $\hat{\gamma}$ for γ. Then the *empirical residuals* for all $t \geq 1$ are defined to be

$$\tilde{\varepsilon}_t := Y_t / \sqrt{\tilde{\sigma}_t^2}. \tag{3.1.21}$$

By plugging in the specific realizations of Y_t, we arrive at so-called *pseudo-data* from the (univariate) distribution of the residuals. When generalizing GARCH models to higher dimensions in Section 3.2 we will use this transformation on multiple marginal GARCH(p, q) series to arrive at pseudo-data from a copula-based multivariate distribution.

3.2 Multivariate extensions

Now, we will consider multivariate generalizations of the univariate GARCH models presented in Section 3.1. The major objective is to model conditional heteroscedasticity in the marginal univariate time series and, at the same time, to capture possible interdependence between the series, thus finally to obtain a multivariate view on multiple financial time series. For most of the time, dependence in multivariate GARCH time series is modeled through, be it constant or dynamic, conditional covariance or correlation matrices or both. Bauwens et al. (2006) as well as Silvennoinen and Teräsvirta (2009) provide comprehensive accounts on multivariate GARCH models. The short overview in Section 3.2.1 concerning proposed extensions from univariate to multivariate models for financial time series follows their lines. By pointing out some critical issues such as over-parameterization entailing intractability for inference procedures as well as tedious stationarity and moment conditions, we intend to motivate a different approach. In Section 3.2.2 the *semiparametric copula-based multivariate dynamic* model will be

presented. Its main building blocks are the dependence modeling through a copula-based multivariate distribution of the residuals as well as univariate GARCH models for the individual processes of returns. Furthermore, a suitable estimation technique that is intrinsically a part of the model will be presented together with its asymptotic properties.

3.2.1 General approaches

In complete analogy to condition (3.0.2) in the univariate case, we now seek a suitable model for a vector-valued process $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$ with $\boldsymbol{Y}_t \in \mathbb{R}^d$, $d \geq 2$ such that with $\mathcal{F}_{t-1} = \sigma(\boldsymbol{Y}_{t-1}, \boldsymbol{Y}_{t-2}, \ldots)$, representing the information set up to time $t-1$, we have

$$\begin{aligned} \boldsymbol{Y}_t &= \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\Sigma}_t &= \operatorname{Cov}(\boldsymbol{Y}_t \,|\mathcal{F}_{t-1}), \end{aligned} \tag{3.2.1}$$

where $\{\boldsymbol{\varepsilon}_t\}$ is a uncorrelated zero-mean, identity-covariance d-variate white noise process now, also denoted by $\boldsymbol{\varepsilon}_t \sim \mathrm{WN}(\mathbf{0}, \mathbf{I}_d)$. Note that $\boldsymbol{\varepsilon}_t$ are uncorrelated both in their components as well as among each other, i.e. throughout time. For $\boldsymbol{\Sigma}_t^{1/2}$ to be the square-root matrix of $\boldsymbol{\Sigma}_t$ for every $t \in \mathbb{Z}$ (for example the Cholesky square-root is such that $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\Sigma}_t^{1/2})^\top$), the conditional covariance matrix $\boldsymbol{\Sigma}_t$ is not an arbitrary $(d \times d)$-matrix, but has to be positive semi-definite. This restriction has to be taken into consideration in every possible extension of the univariate evolution equation.
Definition 3.1.3 was extended by Engle and Kroner (1995) to d dimensions in the most general way by means of the so-called vec-parametrization. The bijective *vec transformation* is the mapping $\operatorname{vec}\colon \mathbb{R}^{d_1\times d_2} \to \mathbb{R}^{d_1 d_2}$ that maps every $(d_1 \times d_2)$-matrix to a $d_1 d_2$-dimensional vector by stacking the columns of the matrix from the leftmost to the rightmost one below one another. In Definition 3.2.1 we will also state very general conditions for $\boldsymbol{\Sigma}_t$ to be positive semi-definite in terms of the involved matrices.

Definition 3.2.1
Let $p, q \geq 0$ be given. Consider the vector-valued time series $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$, $\boldsymbol{Y}_t \in \mathbb{R}^d$ defined by $\boldsymbol{Y}_t = \boldsymbol{\Sigma}_t \boldsymbol{\varepsilon}_t$ with $\boldsymbol{\varepsilon}_t \sim \mathrm{WN}(\mathbf{0}, \mathbf{I}_d)$ iid and $\{\boldsymbol{\Sigma}_t\}_{t\in\mathbb{Z}}$ a process of matrices $\boldsymbol{\Sigma}_t \in \mathbb{R}^{d\times d}$ such that

$$\operatorname{vec}(\boldsymbol{\Sigma}_t) = \operatorname{vec}(C) + \sum_{i=1}^{p} \tilde{A}_i \operatorname{vec}(\boldsymbol{Y}_{t-i}\boldsymbol{Y}_{t-i}^\top) + \sum_{k=1}^{q} \tilde{B}_k \operatorname{vec}(\boldsymbol{\Sigma}_{t-k}). \tag{3.2.2}$$

If $C \in \mathbb{R}^{d\times d}$ is positive semi-definite and $\tilde{A}_i, \tilde{B}_k \in \mathbb{R}^{d^2\times d^2}$, $i = 1, \ldots, p, k = 1, \ldots, q$ are matrices mapping the vectorized positive semi-definite matrices into themselves, then $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$ is called a *multivariate* GARCH(p, q) process in vec-parametrization. $\{\boldsymbol{\Sigma}_t\}_{t\in\mathbb{Z}}$ is the process of *volatility matrices*, and for $q = 1$ we refer to $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$ as *multivariate* ARCH(p) process.

The matrices $\tilde{A}_i, \tilde{B}_k$ can be regarded as linear operators that are subject to restrictions to ensure that $\boldsymbol{\Sigma}_t$ are proper covariance matrices. These constraints brought about a different parametrization that guarantees positive semi-definiteness per se. Engle and Kroner (1995) also introduced the BEKK-parametrization (the acronym BEKK is an

appreciation of the works of Baba, Engle, Kraft and Kroner that laid the foundation for this multivariate model). The volatility matrices $\boldsymbol{\Sigma}_t$ are defined to be

$$\boldsymbol{\Sigma}_t = C + \sum_{i=1}^{p} \sum_{\nu=1}^{n_i} \bar{A}_{i\nu} \boldsymbol{Y}_{t-i} \boldsymbol{Y}_{t-i}^\top \bar{A}_{i\nu}^\top + \sum_{k=1}^{q} \sum_{\mu=1}^{m_k} \bar{B}_{k\mu} \boldsymbol{\Sigma}_{t-k} \bar{B}_{k\mu}^\top, \tag{3.2.3}$$

in this case with arbitrary matrices $\bar{A}_{i\nu}, \bar{B}_{k\mu} \in \mathbb{R}^{d\times d}$ for $i = 1,\dots,p, \nu = 1,\dots,n_i$ and $k = 1,\dots,q, \mu = 1,\dots,m_k$. The equivalence to the vec-parametrization can be established through $\tilde{A}_i = \sum_{\nu=1}^{n_i} \bar{A}_{i\nu} \otimes \bar{A}_{i\nu}$ and $\tilde{B}_k = \sum_{\mu=1}^{m_k} \bar{B}_{k\mu} \otimes \bar{B}_{t\mu}$ where $\otimes$ denotes the Kronecker matrix product. Note that the parametrization is not unique whenever $n_i, m_k > 1$, i.e. an identifiability problem of the model may arise in this case.
If one takes into consideration that covariance matrices are symmetric, the number of parameters still is $(p+q)(d(d+1)/2)^2 + d(d+1)/2 = \mathcal{O}(d^4)$ with the vec-parametrization and up to $(p+q)Kd^2 + d(d+1)/2 = \mathcal{O}(d^2)$ in the BEKK model with $K = \max_{i,k}\{n_i, m_k\}$. Apart from models where the involved matrices are restricted to be diagonal (for example diagonal vec was used by Bollerslev et al., 1988) and thus reducing the number of parameters significantly, a different, in a way indirect approach was proposed. It consists in specifying the conditional covariance matrix $\boldsymbol{\Sigma}_t$ in equation (3.2.1) using a decomposition into conditional standard deviations and correlations. This reads as follows

$$\boldsymbol{\Sigma}_t = D_t P D_t, \tag{3.2.4}$$

with $D_t = \operatorname{diag}(\sigma_{1t},\dots,\sigma_{dt})$ and $P = (\rho_{ij})_{i,j=1,\dots,d}$ a positive definite $(d \times d)$-matrix where $\rho_{ii} = 1$ for all $i = 1,\dots,d$. This representation allows to model σ_{jt}^2, $j = 1,\dots,d$ as volatility processes in univariate GARCH(p,q) according to (3.1.2). The vector $\boldsymbol{\sigma}_t^2 = (\sigma_{1t}^2,\dots,\sigma_{dt}^2)$ has the following form in this case

$$\boldsymbol{\sigma}_t^2 = \boldsymbol{\omega} + \sum_{i=1}^{p} A_i \boldsymbol{Y}_{t-i}^{(2)} + \sum_{k=1}^{q} B_k \boldsymbol{\sigma}_{t-k}^2 \tag{3.2.5}$$

where $\boldsymbol{\omega} \in \mathbb{R}^d$ has positive components, $A_i, B_k \in \mathbb{R}^{d\times d}$ are diagonal matrices with positive entries and $\boldsymbol{Y}_t^{(2)} = \boldsymbol{Y}_t \odot \boldsymbol{Y}_t$, $\odot$ denotes the Hadamard matrix product. The off-diagonal elements of $\boldsymbol{\Sigma}_t$ are $\rho_{ij}\sigma_{it}\sigma_{jt}$ for $i \neq j$. Note that if P is diagonal, then the conditional covariances of the components $Y_{1t},\dots,Y_{dt}$ only depend on the evolution in the respective component itself, i.e. $\boldsymbol{\Sigma}_t$ is also diagonal, otherwise they possibly depend on the evolution of the whole vector $\boldsymbol{\sigma}_t$. This is usually meant by *volatility spillovers*. In the following definition we will give certain specifications for P leading to some conditional correlation models that are well-known in financial applications.

Definition 3.2.2
Let $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$ be a vector-valued process such that $\boldsymbol{Y}_t = \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\varepsilon}_t$ with a vector-valued white noise process $\boldsymbol{\varepsilon}_t \sim \mathrm{WN}(\mathbf{0}, \mathbf{I}_d)$ and the conditional covariance matrices $\boldsymbol{\Sigma}_t$ as in (3.2.4) and (3.2.5) under the respective conditions. According to whether

(1) $P \equiv P_0$ is constant and diagonal, $\boldsymbol{Y}_t$ is called a *constant conditional correlation* (CCC) GARCH(p,q) process (introduced by Bollerslev, 1990), if P is not diagonal,

extended constant conditional correlation (ECCC) GARCH(p, q) process (due to Jeantheau, 1998),

(2) or $P = P_t$ for every $t \in \mathbb{Z}$ such that

$$P_t = (\mathrm{I}_d \odot Q_t)^{-1/2} Q_t (\mathrm{I}_d \odot Q_t)^{-1/2} \tag{3.2.6}$$

$$Q_t = (1 - a - b)S + a\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top + bQ_{t-1} \tag{3.2.7}$$

with $a > 0, b \geq 0, a + b < 1$, $S = \mathrm{Cov}(\boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_1^\top)$ and a positive definite $Q_0 \in \mathbb{R}^{d \times d}$ as starting value, then $\boldsymbol{Y}_t$ is called a *dynamic conditional correlation* (DCC) GARCH(p, q) process (introduced by Engle, 2002).

Many different specifications in a way ranging from Definition 3.2.1 to Definition 3.2.2 have been proposed. The former can be seen as the most general and flexible parametrization, but a very cumbersome one for inference tasks, the latter is fairly restrictive with regards to flexibility in modeling, but suitable for applying well-known and well-studied estimation procedures. Bauwens et al. (2006) focus on a broad presentation of many possible models within the context of multivariate GARCH processes, whereas Silvennoinen and Teräsvirta (2009) in particular shed light on the statistical issues. We want to complement the discussion in their Chapter 3 by highlighting some of the latest publications, especially regarding the analogies to our presentation in the univariate case:

- Boussama et al. (2011) state conditions on the matrices and the residual distribution in the BEKK GARCH(p, q) model that are necessary and sufficient for *strict stationarity*. They show the implication with the help of algebraic geometry. However, their conditions imply covariance stationarity at the same time. The extension to strictly stationary solutions with infinite variance "seems not possible at the moment" (Boussama et al., 2011, p. 2336).

- Generalizing the conditions as in Theorem 3.1.5 on the matrices and the residual distribution to the multivariate CCC GARCH(p, q) model, Fernández and Muriel (2009) show joint (multivariate) *regular variation* of $\boldsymbol{Y}_t$ and $\boldsymbol{\sigma}_t^2$.

- Engle and Sheppard (2001) provide conditions for the positive definiteness of $\boldsymbol{\Sigma}_t$ in DCC GARCH(p, q) as well as for *consistency* and *asymptotic normality* of a proposed two-step quasi maximum likelihood estimator for the involved parameters.

3.2.2 SCOMDY model

We have reached the point at which we can finally bring together multivariate financial time series with copula theory in order to obtain a statistically tractable model. As we have approached this aim from the line of thought of univariate conditional heteroscedasticity models, these will naturally be the main ingredient besides multivariate copula-based distributions.

Copulas have become common and enjoy a growing popularity in financial applications (see Genest et al., 2009a). Although they are often applied to capture the dependence in time series data, some of their inherent features are often disregarded and the data itself is assumed to be iid. It is essential to distinguish between input data for time series models and data of copula-based distributions. We now define a multivariate time series model where from both the conceptual and the statistical point of view this requirement is fulfilled.
The model was first introduced by Chen and Fan (2006), and subsequently used for further statistical developments such as value-at-risk calculations in financial portfolios in Härdle et al. (2008) or local change-point tests of the copula parameter in Giacomini et al. (2009) among others.

Definition 3.2.3
Let $\{\boldsymbol{Y}_t\}_{t\geq 1}$ be a vector-valued process such that $\boldsymbol{Y}_t = \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\varepsilon}_t$ with vector-valued residuals $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{dt})$ and conditional covariance matrices $\boldsymbol{\Sigma}_t = (\sigma_{ij}^2)_{i,j=1,\ldots,d}$.

(1) Let $\boldsymbol{\Sigma}_t = \operatorname{diag}(\sigma_{1t}^2, \ldots, \sigma_{dt}^2)$ be such that for every $j = 1, \ldots, d$ a univariate GARCH(p, q) specification holds

$$\sigma_{jt}^2 = \alpha_{j0} + \sum_{i=1}^{p} \alpha_{ji} Y_{j,t-i}^2 + \sum_{k=1}^{q} \beta_{jk} \sigma_{j,t-k}^2 \tag{3.2.8}$$

with $\gamma_j = (\alpha_{j0}, \alpha_{j1}, \ldots, \alpha_{jp}, \beta_{j1}, \ldots, \beta_{jq}) \in (0, \infty) \times [0, \infty)^{p+q}$. Furthermore, let $\boldsymbol{\varepsilon}_t$, $t \geq 1$, be independent and identically distributed according to a d-variate distribution with distribution function $H = C(F_1, \ldots, F_d; \theta)$ where $F_1, \ldots, F_d$ are arbitrary marginal distributions and $C(\,.\,;\theta)$ belongs to a parametric copula family with parameter $\theta \in \Theta \subseteq \mathbb{R}$. Let $\mathbb{E}\,\boldsymbol{\varepsilon}_t = \mathbf{0}$ and $\operatorname{Cov}\boldsymbol{\varepsilon}_t = R = (r_{ij})_{i,j=1,\ldots,d}$ such that for all $i = 1, \ldots, d$

$$r_{ii} = \operatorname{Var}\varepsilon_{it} = 1 \tag{3.2.9}$$

and $r_{ij} = \operatorname{Cov}(\varepsilon_{it}, \varepsilon_{jt})$ for $i \neq j$.

(2) Let $\tilde{\gamma}_j$ be the quasi maximum likelihood estimator defined in equation (3.1.20) and let $\tilde{\varepsilon}_{jt} = Y_{jt}/\tilde{\sigma}_{jt}$ be the empirical residuals as in equation (3.1.21) in every component $j = 1, \ldots, d$. With $\nu = \nu(T)$ being an offset, set

$$\tilde{U}_{jt} = \frac{1}{T - \nu + 1} \sum_{s=\nu}^{T} \mathbb{1}_{\{\tilde{\varepsilon}_{js} \leq \tilde{\varepsilon}_{jt}\}} \tag{3.2.10}$$

and consider $\tilde{\boldsymbol{U}}_t = (\tilde{U}_{1t}, \ldots, \tilde{U}_{dt})$ for $t = \nu, \ldots, T$ as a pseudo-sample from the copula. As an estimator for the copula parameter θ consider

$$\hat{\theta} = \operatorname*{argmax}_{\theta \in \Theta} \frac{1}{T - \nu + 1} \sum_{t=\nu}^{T} \log c(\tilde{U}_{1t}, \ldots, \tilde{U}_{dt}; \theta), \tag{3.2.11}$$

a canonical maximum likelihood estimator similar to (2.2.22) but offset by ν.

The process $\{\boldsymbol{Y}_t\}_{t\geq 1}$ with the time series specification in (1) and the estimation procedure described in (2) is called SCOMDY model (stemming from *Semiparametric COpula-based Multivariate DYnamic* model).

Remark 3.2.4
Note that the residuals $\boldsymbol{\varepsilon}_t$ are no longer white noise with identity covariance matrix, but iid random vectors with mean zero and covariance matrix R which corresponds to the correlation matrix due to (3.2.9). Apart from a different covariance matrix, all white noise properties are preserved, so that we denote the residual process by $\boldsymbol{\varepsilon}_t \sim \mathrm{WN}(\mathbf{0}, R)$. Condition (3.2.9) is required for identifiability of the parameters in the univariate GARCH(p,q) models and can be implemented into the SCOMDY model by rescaling the components by the respective variances. For this consider $\boldsymbol{\varepsilon}_t^* = (\varepsilon_{1t}^*, \ldots, \varepsilon_{dt}^*) \sim H^* = C^*(F_1^*, \ldots, F_d^*)$ with $\mathrm{Var}(\varepsilon_{jt}^*) = \mathrm{v}_j$ and set $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}^*/\mathrm{v}_1, \ldots, \varepsilon_{dt}^*/\mathrm{v}_d)$. Then the joint distribution for all $(x_1, \ldots, x_d) \in \mathbb{R}^d$ is

$$H(x_1, \ldots, x_d) = H^*\big(F_1^*(\mathrm{v}_1 x_1), \ldots, F_d^*(\mathrm{v}_d x_d)\big), \tag{3.2.12}$$

and the marginal distributions and their (generalized) inverses for all $x_j, y_j \in \mathbb{R}, j = 1, \ldots, d$ are

$$F(x_j) = F_j^*(\mathrm{v}_j x_j) \quad \text{and} \quad F_j^-(y_j) = F_j^{*-}(y_j)/\mathrm{v}_j. \tag{3.2.13}$$

For the copula C of the distribution of $\boldsymbol{\varepsilon}_t$, plug in (3.2.12) and (3.2.13) into the second part of Sklar's theorem such that for all $(u_1, \ldots, u_d) \in [0,1]^d$ we have

$$\begin{aligned} C(u_1, \ldots, u_d) &= H\big(F_1^-(u_1), \ldots, F_d^-(u_d)\big) \\ &= H^*\big(\mathrm{v}_1 F_1^{*-}(u_1)/\mathrm{v}_1, \ldots, \mathrm{v}_d F_d^{*-}(u_d)/\mathrm{v}_d\big) \\ &= H^*\big(F_1^{*-}(u_1), \ldots, F_d^{*-}(u_d)\big) \\ &= C^*(u_1, \ldots, u_d). \end{aligned}$$

This implies that restricting the matrix R to have a unit diagonal is merely a condition on the marginal distributions (which can easily be accomplished by rescaling), and does not impinge on the freedom of choice as to the desired copula.

The formulations of (1) and (2) are very specific for the SCOMDY model we use as the basis for our upcoming analysis. Evidently, other specifications are feasible in this context, for example Hafner and Reznikova (2010) estimate the copula parameter also based on empirical residuals obtained by deGARCHing, but by semiparametric kernel estimates.
However, in our case a consistency and an asymptotic normality result due to Chan et al. (2009a) can be shown. We restate the respective theorems in Theorem 3.2.5 below, the mentioned prerequisites together with the needed notation can be found in Appendix B. Furthermore, for $\boldsymbol{x} = (x_1, \ldots, x_d)$ and $\boldsymbol{y} = (y_1, \ldots, y_d)$ let the real-valued Gaussian process $U(\boldsymbol{x})$ with $\mathbb{E}\,U(\boldsymbol{x}) = 0$ and $\mathbb{E}\,(U(\boldsymbol{x})U(\boldsymbol{y})) = \prod\{x_i \wedge y_i\} - \prod x_i y_i$ be given.

Theorem 3.2.5
Let $\{\boldsymbol{Y}_t\}_{t\geq 1}$ and $\hat{\theta}$ be given as in Definition 3.2.3. The following holds true.

(1) Let conditions (A1), (C1) – (C4) as in Appendix B be satisfied, and suppose that $\nu(T)/T \to 0$, $T \to \infty$, then as $T \to \infty$ we have weak consistency of $\hat{\theta}$, i.e.

$$\hat{\theta} \xrightarrow{P} \theta. \tag{3.2.14}$$

(2) Suppose that (A1)–(A2),(C1)–(C4) and (N1)–(N6) in Appendix B hold. If as $T \to \infty$ we have $\nu(T)/\log T \to \infty$ and $\nu(T)/T \to 0$, then $\sqrt{T}(\hat{\theta} - \theta)$ converges in distribution to Z distributed as

$$\begin{aligned} &- \Sigma^{-1}(\theta)\Big\{ \int \delta(T_d^-(x_1, \dots, x_d); \theta) dU(x_1, \dots, x_d) \\ &+ \sum_{j=1}^{d} \int \delta_j(x_1, \dots, x_d; \theta) U((1, \dots, 1, x_j, 1, \dots, 1)^\top) c(x_1, \dots, x_d; \theta) dx_1 \dots dx_d \Big\}, \end{aligned} \tag{3.2.15}$$

where $\Sigma(\theta)$ is given by (B.0.3), δ_j by (B.0.1) and T_d by (B.0.2).

The authors note that the distributional result in (2) means that the estimation of the copula parameter is independent of deGARCHing, in particular of the estimation of the GARCH parameters via quasi maximum likelihood, and refer to their previous paper Chen and Fan (2006) where a similar result is stated in a more heuristic version. In Proposition 3.2(2) therein it is stated that $\sqrt{T}(\hat{\theta} - \theta)$ converges to a centered normal distribution with variance corresponding to the asymptotic variance at canonical maximum likelihood estimation, i.e. to Σ as in (2.2.25) as if the residuals ε_t were observable. The proof references to Chen and Fan (2004), a manuscript that is not publicly available. Therefore the heuristic result could neither be verified, nor reconstructed.
Furthermore, they comment that "it's hard to see whether the limit in the above theorem is the same as in Chen and Fan (2006)" (Chan et al., 2009a, p. 62). We guess that it is probably not (see also our comment in Appendix B) as the proposed parametric bootstrap procedure to assess the asymptotic estimator variance did not work very well and we had to switch to a nonparametric bootstrap (for further details see the empirical results in Chapter 5).
Despite the necessity to fall back on bootstrapping as an auxiliary variance estimation, the SCOMDY model is our model of choice and we assume for all further considerations that we work on the residual vectors $\tilde{\varepsilon}_t$. The model is not only numerically tractable in the sense that quasi maximum likelihood for the GARCH parameters and canonical maximum likelihood for the copula parameters are well established and available in many statistical software implementations. It also fits into the theoretical framework presented in Section 3.1. As the evolution in the single components only depends on the respective component itself, the stationarity conditions for univariate GARCH(p, q) models as outlined in Section 3.1.1 confer directly to the marginal conditional variance models in SCOMDY. The same is true for the asymptotic results with respect to the quasi maximum likelihood estimators for the respective parameter sets. Here the results of Section 3.1.3 carry over. As to the finite-dimensional distributions, the SCOMDY model

embeds itself in the class of constant conditional correlation GARCH(p, q) models (cf. Definition 3.2.2(1)), and therefore the heavy-tailedness result by Fernández and Muriel (2009) transfers.
In the subsequent section, possible extensions to capture time-dynamic dependencies in copula-based models will be presented. As the main focus of this thesis lies on testing whether the dependence in multivariate time series changes over time, this overview should be considered as a supplement to our chief objective. In case the homogeneity assumption is rejected, modeling the time-dynamic copula parameters through one of the following models might be suitable.

3.3 Time-dynamic copula models

It has become a stylized fact that another dependence measure, namely Pearson's correlation (see Definition 2.1.9) measuring linear dependency between asset returns, is not constant throughout time, a finding documented long before the advent of copulas in finance, for example by Erb et al. (1994), Longin and Solnik (1995) or Engle (2002). Time-varying correlations can not only be captured in the DCC GARCH model (see Definition 3.2.2), but also in the stochastic volatility model with stochastic correlations by Yu and Meyer (2006) or the regime switching model for dynamic correlations by Pelletier (2006). A semiparametric model for the correlation dynamics was further proposed by Hafner et al. (2005).
In the same spirit, dependencies modeled through copulas (which at least allows for more flexible and more realistic non-normal dependence structures) can also be treated as time-dynamic. The first research concerning time-varying copula-based dependence modeling was conducted in Patton (2006). There Sklar's Theorem was extended to conditional distributions and a parametric model to describe the copula parameter evolution was set up. Further in Dias and Embrechts (2004) a statistical test for structural breaks in the parameter of the copula over time was proposed for the first time.
Subsequently, a large amount of research was carried out concerning the application, modeling and testing of time-varying copulas. Some recent rather econometric contributions can be found for example in Chollete et al. (2009); van den Goorbergh et al. (2005); Guegan and Zhang (2009); Jondeau and Rockinger (2006). In the following, we will shortly summarize the survey paper of Reznikova and Manner (2009) to demonstrate the two main approaches to the topic.

- The *observation driven* copula models of Patton (2006) and Creal et al. (2008) are based on the assumption that the copula parameter is a parametric function f of lagged rescaled observations and an autoregressive term. To determine the copula parameter θ_t for every time point, the copula likelihood is assumed to have the parameters of the function f as arguments and is maximized with respect to these. Hereby, the evolution function for θ_t serves as a constraint.

- This constrained maximum likelihood estimation can also be applied when the parameter dynamics are modeled by an adapted *DCC GARCH* approach (the

original model of Engle (2002) can easily be carried over to copula dependence). After fitting univariate GARCH specifications to the margins first, the copula parameter is a parametric function of the cross product of standardized residuals and again an autoregressive term.

- In contrast to the observation based autoregression, Hafner and Manner (2008) propose the copula parameter to be driven by an independent stochastic process λ_t which they call *stochastic autoregression* as λ_t is a stationary autoregressive process of order one with normal residuals. To determine the time-varying copula parameter, the unobservable process λ_t is integrated out of the joint likelihood via Monte Carlo simulation. The simulated likelihood is then maximized with respect to the parameters λ.

- In Hafner and Reznikova (2010) the copula parameter is modeled as a smooth function of time. The time-variation in the parameter is assessed by locally weighting the copula likelihood by a nonparametric kernel function and then maximizing it for every point in time. This procedure confers the name of *semiparametric dynamic copula* modeling on this approach. The choice of an appropriate bandwidth in the kernel guarantees optimality in the sense of the mean squared error.

- A statistical test for *structural breaks* in the parameter specification is proposed by Dias and Embrechts (2004). Assuming that the dependence can be captured appropriately by a fixed parametric copula family, the parameter is allowed to change after a break point. The location of this break is determined by a likelihood ratio statistic. In this approach the concept of local homogeneity appears for the first time as the parameter before the break is assumed to be constant and equal to θ_1, and after the break constant as well, but equal to $\theta_2 \neq \theta_1$.

- An *adaptive estimation* procedure, namely the local parametric estimation on adaptively chosen subsamples, was then proposed by Giacomini et al. (2009). Here again the dependence is captured by a fixed parametric copula family and time-variation by time-varying parameters which are assumed to be locally constant. The data driven procedure of local change point analysis (LCPA) leads to a local maximum likelihood estimator sequence for the parameters.

The local homogeneity assumption is the main feature which distinguishes the latter two approaches from the former ones. It also accounts for the estimation technique of maximum likelihood on subsamples. Moreover, it allows a specification-free method to obtain the dynamic behavior of an estimator time series. In the following, we restrict ourselves to the moving window maximum likelihood estimation (see e.g. Franke et al., 2008, Chap. 17.3.3), the most simple specification of local parametric fitting. This procedure was originally proposed by Giacomini et al. (2009) in analogy to the RiskMetrics methodology (see J.P.Morgan/Reuters, 1996). In RiskMetrics the conditional distribution of logreturns from financial time series is assumed to be multivariately

normal and the respective covariance matrix is estimated by an exponentially weighted moving average model. The results of both moving window and RiskMetrics estimation were compared to the more sophisticated LCPA in this publication.
Besides the presented parametric modeling of time-variation, there are several non-parametric approaches to model and estimate time-varying copulas. Among others, Fermanian and Scaillet (2003) derive asymptotic properties of kernel estimators for copulas in the context of stationary strong mixing processes. For test purposes, Busetti and Harvey (2008) developed stationary tests of copula constancy applied to nonparametrically estimated indicators for bivariate series which were then extended by Harvey (2008, 2010) to time-varying histograms and quantiles, respectively.
In the subsequent chapter we will give a precise definition of the moving window estimation and derive homogeneity tests.

4 Homogeneity Tests

Assume that $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T$, $T \geq 1$ is a finite sample coming from a SCOMDY process $\{\boldsymbol{Y}_t\}_{t\geq 1}$ as given in Definition 3.2.3(1). According to Definition 3.2.3(2) we obtain the empirical residuals $\tilde{\boldsymbol{\varepsilon}}_1, \ldots, \tilde{\boldsymbol{\varepsilon}}_T$ as pseudo-realizations of the residual distribution H, and pseudo-realizations $\tilde{\boldsymbol{U}}_1, \ldots, \tilde{\boldsymbol{U}}_T$ of the copula C_θ through a probability integral transform in every component of the vectors. As we aim at testing for homogeneity of the copula parameter throughout time, consider for the moment

$$\tilde{\boldsymbol{U}}_t \sim C_{\theta_t} \tag{4.0.1}$$

where θ_t is the present parameter for every point in time $t = 1, \ldots, T$. For an arbitrary $\theta_0 \in \Theta$, the pair of hypotheses we want to test is then

$$\mathbf{H}_0^*\colon \{\theta_t \equiv \theta_0 \,\forall\, t\} \quad \text{vs.} \quad \mathbf{H}_A^*\colon \{\exists t\colon \theta_t \neq \theta_0\}. \tag{4.0.2}$$

We address this task nonparametrically through a purely data-driven approach in Section 4.1. Based on the resulting estimators and their asymptotic distribution three different tests will be constructed and analyzed with respect to their ability to detect actual deviations from the homogeneity hypotheses. In Section 4.2 we will present a Binomial multiple test based on independent subsamples of the given data, in Section 4.3 we will examine whether the extremes of a parameter series based on overlapping subsamples exceed certain confidence levels. Finally, in Section 4.4 a fluctuation test using each of the moving window estimators is presented.

4.1 Moving window estimators

In order to assess θ_t for $1 \leq t \leq T$ we define maximum likelihood estimators on certain subsamples of a bandwidth b in the following.

Definition 4.1.1
Let $\tilde{\boldsymbol{U}}_1, \ldots, \tilde{\boldsymbol{U}}_T$, $T \geq 1$ be pseudo-realizations of a parametric copula C_θ according to Definition 3.2.3(2). With a bandwidth $1 \leq b = b(T) \leq T$ the *local* or *moving window maximum likelihood estimator* $\hat{\theta}_t$ of θ for every $t = b, \ldots, T$ is defined by

$$\hat{\theta}_t = \underset{\theta\in\Theta}{\operatorname{argmax}} \sum_{s=t-b+1}^{t} \log c(\tilde{U}_{1s}, \ldots, \tilde{U}_{ds}; \theta). \tag{4.1.1}$$

In Proposition 4.1.2 the consistency and asymptotical normality result for the copula estimators in the SCOMDY model are transferred to the local estimators $\hat{\theta}_t$ in terms of the bandwidth $b = b(T)$.

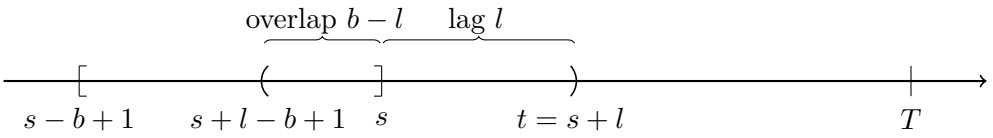
Proposition 4.1.2
Under the conditions of Theorem 3.2.5 and under the assumption that $b = b(T)$ is chosen such that $b \to \infty$ and $b/T \to 0$ for $T \to \infty$, in shorthand notation denoted by $b = \mathcal{O}(T)$ $(T \to \infty)$, we have the following for $T \to \infty$.

(1) $\hat{\theta}_t \xrightarrow{P} \theta$.

(2) $\sqrt{b}(\hat{\theta}_t - \theta) \xrightarrow{d} \mathcal{N}(0, \Sigma)$, where Σ is the asymptotic variance in canonical maximum likelihood estimation as in (2.2.25).

Proof
(1) and (2) correspond to Theorem 3.2.5(1) and (2) respectively. By posing the condition $b = \mathcal{O}(T)$ $(T \to \infty)$ on the bandwidth b, the respective offset conditions on ν are fulfilled. □

Now, we consider the joint limiting distribution of (two) moving window estimators and show that the covariance structure is non-degenerate if the overlap of the subsamples, which the respective estimators are based on, is asymptotically constant. For points in time $1 \leq s < t \leq T$ consider $\hat{\theta}_t$ and $\hat{\theta}_s$ based on subsamples of bandwidth b according to (4.1.1). Their *lag* is defined as $l := |s - t|$, whereas the *overlap* of the subjacent subsamples is $b - l$. The situation for $t = s + l$ can be visualized on a time scale as follows.



Proposition 4.1.3
Let $1 \leq s, t \leq T$ and let $\hat{\theta}_t$ and $\hat{\theta}_s$ be the respective moving window estimators. Let $1 \leq l := |s-t| < b(T)$ be such that $l/b(T) \to \mu_l \in (0,1)$ as $T \to \infty$. Under the conditions of Theorem 3.2.5 the following joint limiting law holds

$$\sqrt{b}\left((\hat{\theta}_s, \hat{\theta}_t)^\top - (\theta, \theta)^\top\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma\bar{M}) \tag{4.1.2}$$

with matrix $\bar{M} = \begin{pmatrix} 1 & 1-\mu_l \\ 1-\mu_l & 1 \end{pmatrix}$ and Σ the scalar variance as in (2.2.25).

Proof
For clarity of presentation we will assume at first that we act on an iid sample of pure copula data. In this case, the involved variance term is the inverse Fisher information. Secondly, we will give the arguments in case of canonical maximum likelihood on general multivariate data; the asymptotic variance is Σ in this case.

(i) Let $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_T$ be independent and identically distributed realization of $\boldsymbol{U} \sim C_\theta$. Recall the log-likelihood function $l(\boldsymbol{u}; \theta) = \log c(\boldsymbol{u}; \theta)$ for $\boldsymbol{u} \in [0,1]^d$, and its

partial derivatives $l_\theta(\boldsymbol{u};\theta) = \frac{\partial}{\partial\theta} l(\boldsymbol{u};\theta)$ and $l_{\theta\theta}(\boldsymbol{u};\theta) = \frac{\partial^2}{\partial\theta^2} l(\boldsymbol{u};\theta)$ with respect to θ. Furthermore, assume that $s < t$. Then $\hat{\theta}_s$ and $\hat{\theta}_t$ allow for the following representation

$$\hat{\theta}_\tau = \underset{\theta}{\operatorname{argmax}} \sum_{i=\tau-b+1}^{\tau} l(\boldsymbol{U}_i;\theta) = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{b} l(\boldsymbol{U}_{\tau-i+1};\theta), \tag{4.1.3}$$

with $\tau \in \{s,t\}$. For the function $f(\theta,t) := \frac{1}{b}\sum_{i=1}^{b} l_\theta(\boldsymbol{U}_{t-i};\theta)$ the maximum likelihood approach then yields that

$$f(\hat{\theta}_s, s) = f(\hat{\theta}_t, t) = 0. \tag{4.1.4}$$

This implies that the first term in the first order Taylor expansion of f for $\tau \in \{s,t\}$ around $\hat{\theta}_\tau$ cancels out. With $\zeta = \theta$ as intermediate point, we obtain

$$\frac{1}{\sqrt{b}} \sum_{i=1}^{b} l_\theta(\boldsymbol{U}_{\tau-i};\theta) = -\frac{1}{b} \sum_{i=1}^{b} l_{\theta\theta}(\boldsymbol{U}_{\tau-i};\theta)\sqrt{b}(\hat{\theta}_\tau - \theta). \tag{4.1.5}$$

By the strong law of large numbers and under condition (vi) in Theorem 2.2.2 the first term on the right-hand side converges almost surely to the Fisher information $I(\theta)$ as the bandwidth b goes to infinity with $T \to \infty$, precisely

$$-\frac{1}{b}\sum_{i=1}^{b} l_{\theta\theta}(\boldsymbol{U}_{\tau-i};\theta) \xrightarrow{a.s.} -\mathbb{E}\, l_{\theta\theta}(\boldsymbol{U};\theta) = I(\theta). \tag{4.1.6}$$

The left-hand side of (4.1.5) converges to a univariate normal distribution due the the classical central limit theorem. Under condition (iv) in Theorem 2.2.2 the limiting distribution has mean zero, and under condition (vi) the asymptotic variance is the Fisher information, i.e. as $b(T) \to \infty$ ($T \to \infty$) we have

$$Z_b^{(\tau)} := \frac{1}{\sqrt{b}} \sum_{i=1}^{b} l_\theta(\boldsymbol{U}_{\tau-i};\theta) \xrightarrow{d} Z^{(\tau)} \sim \mathcal{N}(0, I(\theta)). \tag{4.1.7}$$

Consider now $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ and $\boldsymbol{Z}_b = \left(Z_b^{(s)}, Z_b^{(t)}\right)$ and their product

$$\begin{aligned}
\boldsymbol{\lambda}\boldsymbol{Z}_b^\top &= b^{-1/2}\left(\lambda_1 \sum_{i=1}^{b} l_\theta(\boldsymbol{U}_{s-i+1};\theta) + \lambda_2 \sum_{j=1}^{b} l_\theta(\boldsymbol{U}_{t-j+1};\theta)\right) \\
&= b^{-1/2}\left(\lambda_1 \sum_{i=b-l+1}^{b} l_\theta(\boldsymbol{U}_{s-i+1};\theta)\right. \\
&\qquad \left. +(\lambda_1+\lambda_2)\sum_{i=1}^{b-l} l_\theta(\boldsymbol{U}_{s-i+1};\theta) + \lambda_2 \sum_{i=1}^{l-1} l_\theta(\boldsymbol{U}_{s+l-i+1};\theta)\right)
\end{aligned}$$

which results from rearranging the sums and an index shift. Since $\mathbb{E}\, l_\theta(\boldsymbol{U};\theta) = 0$, we have $\mathbb{E}\,\boldsymbol{\lambda}\boldsymbol{Z}_b^\top = 0$ and furthermore

$$\begin{aligned}
\operatorname{Var}\boldsymbol{\lambda}\boldsymbol{Z}_b^\top &= b^{-1}\left(\lambda_1^2 \sum_{i=b-l+1}^{b} \mathbb{E}\, l_\theta^2(\boldsymbol{U}_{s-i+1};\theta)\right. \\
&\qquad \left. +(\lambda_1+\lambda_2)^2 \sum_{i=1}^{b-l} \mathbb{E}\, l_\theta^2(\boldsymbol{U}_{s-i+1};\theta) + \lambda_2^2 \sum_{i=1}^{l-1} \mathbb{E}\, l_\theta^2(\boldsymbol{U}_{s+l-i+1};\theta)\right) \\
&= b^{-1}(l\lambda_1^2 + (b-l)(\lambda_1+\lambda_2)^2 + l\lambda_2^2)\mathbb{E}\, l_\theta^2(\boldsymbol{U};\theta) \\
&= (\lambda_1^2 + 2(1-l/b)\lambda_1\lambda_2 + \lambda_2^2) I(\theta) \\
&= (\lambda_1,\lambda_2)\begin{pmatrix} 1 & 1-l/b \\ 1-l/b & 1 \end{pmatrix} I(\theta) \begin{pmatrix}\lambda_1 \\ \lambda_2\end{pmatrix} =: \boldsymbol{\lambda} I(\theta)\bar{M}_b\boldsymbol{\lambda}^\top.
\end{aligned}$$

Note that under the assumption that $l/b \to \mu_l$ for $b \to \infty$, we have $\lim_{b\to\infty} \bar{M}_b = \bar{M}$. The terms are grouped in such a way that they are iid with the result that the central limit theorem is applicable. Finally, we have shown that as $b(T) \to \infty$ ($T \to \infty$) we have $\boldsymbol{\lambda}\boldsymbol{Z}_b^\top \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\lambda} I(\theta)\bar{M}\boldsymbol{\lambda}^\top)$, and by a Cramér-Wold type argument (see Cramér and Wold, 1936) we conclude that

$$\boldsymbol{Z}_b \xrightarrow{d} \mathcal{N}(\mathbf{0}, I(\theta)\bar{M}). \tag{4.1.8}$$

The joint asymptotic normality of $\hat{\theta}_s$ and $\hat{\theta}_t$ is now a result of (4.1.6), (4.1.8) and an application of Slutky's theorem, and reads as follows

$$\sqrt{b}\begin{pmatrix}\hat{\theta}_s - \theta \\ \hat{\theta}_t - \theta\end{pmatrix} \xrightarrow{d} \frac{1}{I(\theta)}\mathcal{N}(\mathbf{0}, I(\theta)\bar{M}) = \mathcal{N}(\mathbf{0}, \bar{M}/I(\theta)). \tag{4.1.9}$$

In particular, the asymptotic covariance of two moving window maximum likelihood estimators lagged by l observations $\operatorname{Cov}(\hat{\theta}_t, \hat{\theta}_s) = (1-\mu_l)/I(\theta) =: \gamma_l$ is a function of this lag only.

(ii) The reasoning for moving window estimators based on multivariate data stemming from a multivariate distribution or being the residuals of a SCOMDY model is similar to the Taylor approximation technique used above. For the joint asymptotic distribution of $\hat{\theta}_s$ and $\hat{\theta}_t$, we have additionally to use the theory for multivariate rank order statistics as in Ruymgaart et al. (1972). There the asymptotic behavior of the quantities in question is also reduced to sums over certain expressions depending on the given sample. Similarly to equation (3.4) in the proof of their Theorem 2.1 numerous terms arise that finally vanish asymptotically, and only three of the terms account for the asymptotic normality in the end. The reordering of the respective sums can be performed analogously, so that the initial overlapping of the subsamples is remedied and the correlation can be factored out into the matrix M.

□

The moving window estimators $\hat{\theta}_t$ are defined for all $t = b, \ldots, T$ and a given realization $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T$ of a SCOMDY model. For the subsequent sections we will regard the estimators as an estimator time series, and proceed in testing the null hypothesis (4.0.2) depending on the index set for the times t. To this end, consider the following definitions.

Definition 4.1.4
Let a SCOMDY sample of size $T \geq 1$ and the respective moving window estimators $\hat{\theta}_t$ with bandwidth $b = b(T)$ and for all $t = b, \ldots, T$ be given. Consider the *thinning constant* $c \in [1/b, 1]$ and the *series length* $n^{(c)} = \lfloor \frac{T-b}{cb} \rfloor$, where $\lfloor x \rfloor := \max\{k \in \mathbb{Z} \colon k \leq x\}$ for all $x \in \mathbb{R}$. Set the *index set* as $I^{(c)} := \{t_k = \lfloor T - (n^{(c)} - k)cb \rfloor \colon k = 1, \ldots, n^{(c)}\}$ and consider the *estimator series* $\{\hat{\theta}_t\}_{t \in I^{(c)}}$.

(1) For $c = 1$, $\{\hat{\theta}_t\}_{t \in I^{(c)}}$ is called *independent* estimator series. In particular, we set $n := n^{(1)} = \lfloor \frac{T-b}{b} \rfloor$ and $\hat{\theta}_k := \hat{\theta}_{t_k}$ for all $t_k = T - (n-k)b$, $k = 1, \ldots, n$.

(2) For fixed c with $1/b < c < 1$, the estimator series $\{\hat{\theta}_t\}_{t \in I^{(c)}}$ is called *thinned*.

(3) For $c = c(b) = 1/b$, we refer to $\{\hat{\theta}_t\}_{t \in I^{(c)}}$ as *full* estimator series. In this case, we stress the dependence on the bandwidth by writing $n(b) := n^{(1/b)} = T - b$ and $t_k(b) = t_k^{(1/b)} = k + b$.

Along with this three-step definition we will proceed by presenting a suitable approach to test for parameter homogeneity in the independent estimator series in Section 4.2.

4.2 Binomial test

Let $\hat{\theta}_1, \ldots, \hat{\theta}_n$ be an independent estimator series as in Definition 4.1.4(1). The general null hypothesis is replaced by a first approximation in order to take advantage of the independence of the individual estimators $\hat{\theta}_k$. We aim at testing

$$\mathbf{H}_0 \colon \{\theta_{t_1} = \ldots = \theta_{t_n} = \theta_0 \colon t_k = T - (n-k)b\} \quad \text{vs.} \quad \mathbf{H}_A \colon \{\exists t_k \colon \theta_{t_k} \neq \theta_0\}. \tag{4.2.1}$$

We will clarify the notation and the statistical concepts for one individual estimator in Section 4.2.1. In Section 4.2.2 we will perform a multiple simultaneous test of the partial null hypotheses to arrive at the hypothesis given in (4.2.1).

4.2.1 Partial tests

Consider an arbitrary $k = 1, \ldots, n$ and set $\hat{\theta} = \hat{\theta}_k$ as well as $\theta = \theta_{t_k}$ in the following. Furthermore, consider Σ as the asymptotic variance in canonical maximum likelihood estimation given in equation (2.2.25). Note that the copula parameter estimator $\hat{\theta}$ is determined locally on subsamples preceding the point in time t_k. The following asymptotically normal test statistic for $b(T) \to \infty$ ($T \to \infty$) can be constructed

$$T_b := \frac{\sqrt{b}(\hat{\theta} - \theta_0)}{\sqrt{\Sigma}} \xrightarrow{d} \mathcal{N}(0,1) \tag{4.2.2}$$

to test the partial null hypothesis $\mathrm{H}_0\colon \{\theta = \theta_0\}$.
In order to quantify a statistical test's ability to detect deviations that hint at the null hypothesis being false, the notion of *statistical power* is generally accepted. Consider the following formal definitions.

Definition 4.2.1
Let the random variables $X_1, \ldots, X_b \sim F$ be identically distributed according to the distribution function $F \in \Delta$ belonging to a family of distributions $\Delta = \Delta_0 \cup \Delta_A$.

(1) We call $\mathrm{H}_0\colon \{F \in \Delta_0\}$ the *null hypothesis*, and $\mathrm{H}_A\colon \{F \in \Delta_A\}$ the *alternative hypothesis*.

(2) A subset $K \subseteq \mathbb{R}^b$ is called *region of rejection* or *critical region* if for the sample function $\varphi\colon \mathbb{R}^b \to [0,1]$ we have

$$\varphi(x_1, \ldots, x_b) = \mathbb{1}_{\{(x_1,\ldots,x_b)\in K\}}. \tag{4.2.3}$$

(3) For fixed $\alpha \in (0,1)$ we call φ a *test to the level* α if

$$\mathbb{P}_F(\varphi(X_1, \ldots, X_b) = 1) \leq \alpha \quad \text{for all } F \in \Delta_0. \tag{4.2.4}$$

A test is called *unbiased* if furthermore

$$\mathbb{P}_F(\varphi(X_1, \ldots, X_b) = 1) \geq \alpha \quad \text{for all } F \in \Delta_A. \tag{4.2.5}$$

If the inequalities in (4.2.4) and (4.2.5) hold asymptotically for $b \to \infty$, then the test is *asymptotically* to the level α and *asymptotically* unbiased respectively. In any case, α is called the *size* or *significance level* of the test.

(4) The function $G\colon \Delta \to [0,1]$ with $G(F) = \mathbb{P}_F(\varphi(X_1, \ldots, X_b) = 1)$ is called power function, and the restriction $\gamma = G|_{\Delta_A}$ the *statistical power* of the test.

(5) If $\Delta_A = \Delta_A(b)$ is such that $\Delta_A \to \Delta_0$ for growing sample size $b \to \infty$, then the alternative hypothesis is called *local*. If the convergence of the alternative to the null hypothesis as $b \to \infty$ is as fast as $\beta_b^{-1} \to 0$ for a null sequence $\{\beta_b^{-1}\}$, and the test is unbiased, then the test is said to have *local power of order* β_b. In shorthand notation we denote the speed of convergence by $\mathrm{H}_A - \mathrm{H}_0 \asymp \beta_b$ as $b \to \infty$.

As noted in Lehmann and Romano (2005, p. 433), local power considerations for a given test have to be conducted in every special case of test statistic, asymptotic distribution and alternative hypothesis. Subsequently, we elaborate the definitions in terms of the test statistic T_b given in (4.2.2) and the partial null hypothesis $\mathrm{H}_0\colon \{\theta = \theta_0\}$. For the power calculations we consider a fixed and a local alternative hypothesis. On the one hand, we have $\mathrm{H}_A(\eta)\colon \{\theta = \theta(\eta) = \theta_0 + \eta\}$ as the fixed alternative, on the other hand, $\bar{\mathrm{H}}_A(\eta)\colon \{\theta = \theta_b(\eta) = \theta_0 + \eta/\sqrt{b}\}$ as the local alternative. Both can be regarded as *one-sided* if $\eta > 0$ or $\eta < 0$ respectively, and as *two-sided* if $\eta \neq 0$.
For the one-sided alternatives, H_0 is rejected to the significance level $\alpha \in (0,1)$ if $T_b > q_\alpha$ or $T_b < q_{-\alpha}$, whereas for the two-sided alternative, rejection occurs if $|T_b| > q_{\alpha/2}$.

Here we denote by q_α the $(1-\alpha)$-quantile of the standard normal distribution, i.e. $\mathbb{P}(Z \leq q_\alpha) = 1-\alpha$ for $Z \sim \mathcal{N}(0,1)$. For the critical region K this implies

$$K(\eta) = \begin{cases} (q_\alpha, \infty), & \text{for } \eta > 0 \\ (-\infty, -q_\alpha), & \text{for } \eta < 0 \\ (-\infty, -q_{\alpha/2}) \cup (q_{\alpha/2}, \infty), & \text{for } \eta \neq 0. \end{cases}$$

Depending on whether the alternative is local or fixed, we will derive the power of both the one- and the two-sided tests in the following lemma. Without loss of generality we assume that $\theta_0 = 0$ and $\Sigma = 1$ for notational ease. The results remain unchanged under general $\theta_0 \in \Theta$ and $\Sigma > 0$.

Lemma 4.2.2
By setting the null hypothesis as $\mathrm{H}_0 \colon \{\theta = 0\}$ we obtain the fixed alternative $\mathrm{H}_A(\eta) \colon \{\theta = \eta\}$ and the local alternative $\bar{\mathrm{H}}_A(\eta) \colon \{\theta = \eta/\sqrt{b}\}$. Then we have the following.

(1) Let the sample size $b \geq 1$ be fixed and $\eta \in \mathbb{R}$. Then for $\gamma_b(\eta) = \mathbb{P}_{\mathrm{H}_A(\eta)}(T_b \in K(\eta))$ we have

$$\mathbb{P}_{\mathrm{H}_A(\eta)}(T_b \in K(\eta)) = (1 + o(1)) \cdot \begin{cases} 1 - \Phi(q_\alpha - \sqrt{b}\eta), & \eta > 0 \\ \Phi(-q_\alpha - \sqrt{b}\eta), & \eta < 0 \\ 1 - \Phi(q_{\alpha/2} - \sqrt{b}\eta) + \Phi(-q_{\alpha/2} - \sqrt{b}\eta), & \eta \neq 0. \end{cases}$$

Note that the asymptotic equality of the left-hand and the right-hand side is with respect to $b \to \infty$.

(2) For $\eta \in \mathbb{R}$ we have asymptotically

$$\bar{\gamma}(\eta) = \lim_{b \to \infty} \mathbb{P}_{\bar{\mathrm{H}}_A(\eta)}(T_b \in K(\eta)) = \begin{cases} 1 - \Phi(q_\alpha - \eta), & \eta > 0 \\ \Phi(-q_\alpha - \eta), & \eta < 0 \\ 1 - \Phi(q_{\alpha/2} - \eta) + \Phi(-q_{\alpha/2} - \eta), & \eta \neq 0. \end{cases}$$

Φ is the distribution function of the standard normal law and $\mathbb{P}_{\mathrm{H}_A}$ denotes the probability under the (alternative) hypothesis H_A.

Note that for the fixed alternative we consider γ_b, i.e. the power of the test depending on the sample size b. For $b \to \infty$ we have $\gamma_b(\eta) \to 1$ for any η in any case of either one-sided or two-sided testing. The advantage of local alternatives becomes apparent here as the power $\bar{\gamma}$ is independent of the sample size and the limit for $b \to \infty$ can be established right away.

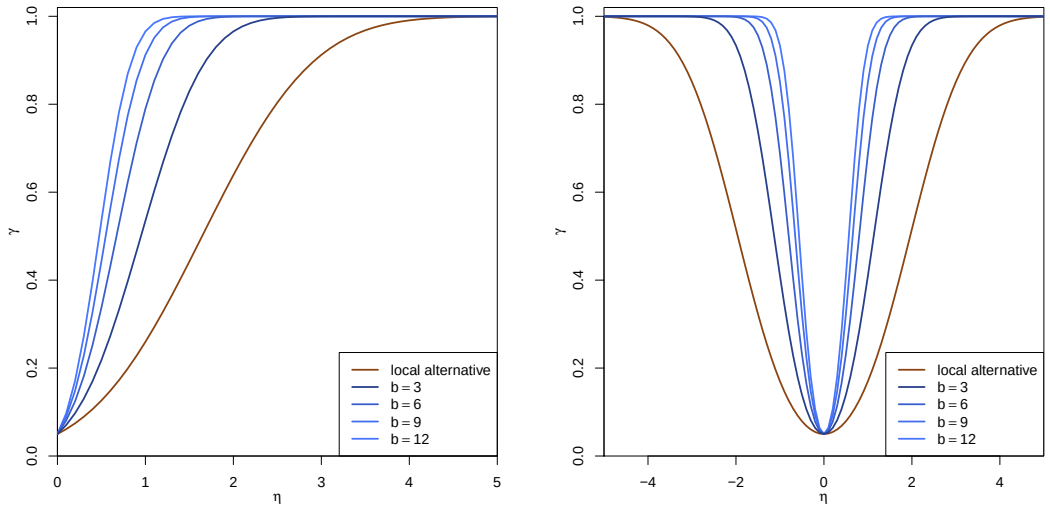
Proof
For a shorthand notation of a standard normally distributed random variable, we will consider $Z \sim \mathcal{N}(0,1)$ in the following. Exemplarily, we will now establish the statistical power of the partial test under the two-sided fixed alternative.

For $\eta \neq 0$ and with $\mathbb{P} = \mathbb{P}_{\mathrm{H}_A(\eta)}$ the probability under $\mathrm{H}_A(\eta)$ we compute

$$\begin{aligned}\gamma_b(\eta) &= \mathbb{P}\left(|T_b| > q_{\alpha/2}\right) = \mathbb{P}\left(T_b > q_{\alpha/2}\right) + \mathbb{P}\left(T_b < -q_{\alpha/2}\right) \\ &= \mathbb{P}\left(\sqrt{b}(\hat{\theta} - \theta(\eta)) > q_{\alpha/2} - \sqrt{b}\theta(\eta)\right) + \mathbb{P}\left(\sqrt{b}(\hat{\theta} - \theta(\eta)) < -q_{\alpha/2} - \sqrt{b}\theta(\eta)\right) \\ &\approx \mathbb{P}\left(Z > q_{\alpha/2} - \sqrt{b}\eta\right) + \mathbb{P}\left(Z < -q_{\alpha/2} - \sqrt{b}\eta\right) \\ &= 1 - \Phi(q_{\alpha/2} - \sqrt{b}\eta) + \Phi(-q_{\alpha/2} - \sqrt{b}\eta),\end{aligned}$$

where $x_b \approx y_b$ is a short-hand notation for $x_b = (1 + \mathcal{O}(1))\, y_b$ as $b \to \infty$. The asymptotic equality is due the the distributional convergence of $\sqrt{b}(\hat{\theta} - \theta(\eta))$ to the standard normal law under the alternative. For the remaining cases analogous considerations prove the claims. □

In Figure 4.1, the power curves $\gamma(\eta)$ of the one-sided test with $\eta > 0$ and of the two-sided test as functions of η are depicted, exemplarily for $\alpha = 5\%$. As expected, both the fixed and the local power curves converge to 1 as η goes to infinity in modulus, i.e. the test $T_b \in K(\eta)$ is suitable in order to detect deviations from the null hypothesis $\mathrm{H}_0 \colon \{\theta = \theta_0\}$. However, in case of the fixed alternative the detection ability becomes worse the greater the sample size b gets.



(a) One-sided partial test

(b) Two-sided partial test

Figure 4.1: Local power $\gamma = \bar{\gamma}$ (brown) and fixed power $\gamma = \gamma_b$ (blue) of partial tests

Furthermore, we see that $\gamma(0) = \alpha$ and $\gamma(\eta) > \alpha$ if $\eta \neq 0$. The following theorem formally shows that the test is unbiased, which in case of the local alternative can be used for a result concerning the local power and its order.

Theorem 4.2.3

Let $\gamma_b(\eta)$ and $\bar{\gamma}(\eta)$ be given as in Lemma 4.2.2(1) and (2), i.e. the fixed and local

power of the test to the level $\alpha \in (0,1)$ of $\mathrm{H}_0\colon \{\theta = 0\}$ against $\mathrm{H}_A(\eta)\colon \{\theta = \eta\}$ and $\bar{\mathrm{H}}_A(\eta)\colon \{\theta = \eta/\sqrt{b}\}$, respectively. Then we have for $\eta \neq 0$

$$\gamma_b(\eta) > \alpha \quad \text{and} \quad \bar{\gamma}(\eta) > \alpha, \tag{4.2.6}$$

i.e. the test is asymptotically unbiased against both alternatives H_A and $\bar{\mathrm{H}}_A$. Furthermore, the test against $\bar{\mathrm{H}}_A$ has local power of order $\sqrt{b}$.

Proof

We will again confine ourselves to showing $\gamma_b(\eta) > \alpha$ for $\eta \neq 0$, i.e. the case of the two-sided test with a fixed alternative. In case of a local alternative the main ideas are left unchanged. The bound for the one-sided tests are even simpler to prove. Here all we have to consider is the sum of an integral over the positive normal density and α, which is naturally greater or equal to α.

We will directly act under the asymptotic form of the power function of the two-sided partial test with fixed alternative and denote this by $\approx$. Add $\mp\Phi(q_{\alpha/2})$ twice to obtain

$$\begin{aligned}
\gamma_b(\eta) &\approx 1 - \Phi(q_{\alpha/2} - \sqrt{b}\eta) + \Phi(-q_{\alpha/2} - \sqrt{b}\eta) \\
&= 1 \mp \Phi(q_{\alpha/2}) - \Phi(q_{\alpha/2} - \sqrt{b}\eta) - (\mp\Phi(-q_{\alpha/2}) - \Phi(-q_{\alpha/2} - \sqrt{b}\eta)) \\
&= 1 - (1 - \alpha/2) + \int_{q_{\alpha/2}-\sqrt{b}\eta}^{q_{\alpha/2}} \phi(x)dx + \alpha/2 - \int_{-q_{\alpha/2}-\sqrt{b}\eta}^{-q_{\alpha/2}} \phi(x)dx \\
&= \alpha + \int_{q_{\alpha/2}-\sqrt{b}\eta}^{q_{\alpha/2}} \phi(x)dx - \int_{-q_{\alpha/2}-\sqrt{b}\eta}^{-q_{\alpha/2}} \phi(x)dx \\
&=: \alpha + \psi(\alpha, \eta, b)
\end{aligned}$$

where $\phi(x)$ is the standard normal density. In this notation, we assume that $\alpha < 1/2$ and therefore $q_{\alpha/2} < 0$. Otherwise the limits in the integrals have to be reversed, but similar results hold. Due to the standard normal density's symmetry around 0, we have that $\psi(\alpha, 0, b) = \phi(q_{\alpha/2}) - \phi(-q_{\alpha/2}) = 0$. Furthermore, let b be fixed and consider the partial derivative of ψ with respect to $\eta > 0$ taking into account the Leibniz integral rule

$$\begin{aligned}
\frac{\partial}{\partial\eta}\psi(\alpha, \eta, b) &= 0 + 0 - \phi(q_{\alpha/2} - \sqrt{b}\eta)(-\sqrt{b}) - \Big(0 + 0 - \phi(-q_{\alpha/2} - \sqrt{b}\eta)(-\sqrt{b})\Big) \\
&= \sqrt{b}\Big(\phi(q_{\alpha/2} - \sqrt{b}\eta) - \phi(-q_{\alpha/2} - \sqrt{b}\eta)\Big) \\
&= \frac{\sqrt{b}}{\sqrt{2\pi}}\left(\exp\left\{-\frac{1}{2}(q_{\alpha/2} - \sqrt{b}\eta)^2\right\} - \exp\left\{-\frac{1}{2}(-q_{\alpha/2} - \sqrt{b}\eta)^2\right\}\right) \\
&= \frac{\sqrt{b}}{\sqrt{2\pi}}\left(\exp\left\{-\frac{1}{2}(q_{\alpha/2} - \sqrt{b}\eta)^2\right\} - \exp\left\{-\frac{1}{2}(+q_{\alpha/2} + \sqrt{b}\eta)^2\right\}\right) \\
&> \frac{\sqrt{b}}{\sqrt{2\pi}}\left(\exp\left\{-\frac{1}{2}(q_{\alpha/2} - \sqrt{b}\eta)^2\right\} - \exp\left\{-\frac{1}{2}(q_{\alpha/2} - \sqrt{b}\eta)^2\right\}\right) = 0,
\end{aligned}$$

where the inequality in the last line is due to $q_{\alpha/2} < 0$. This yields that the function ψ is strictly increasing in $\eta > 0$.

Finally, we know that $\psi(\alpha, \eta, b) > 0$ for $\eta > 0$. The same is true for $\eta < 0$. Therefore, $\gamma_b(0) = \alpha$ and $\gamma_b(\eta) > \alpha$ for $\eta \neq 0$. By definition the test is asymptotically unbiased against the alternative H_A.
As $\bar{\gamma}(\eta) > \alpha$ for $\bar{\mathrm{H}}_A \colon \{\theta = \eta/\sqrt{b}\}$ and $\eta/\sqrt{b} \to 0$, $(b \to \infty)$, for every fixed $\eta \neq 0$, we have $\bar{\mathrm{H}}_A - \mathrm{H}_0 \asymp \sqrt{b}$. This fact yields local power of order $\sqrt{b}$. □

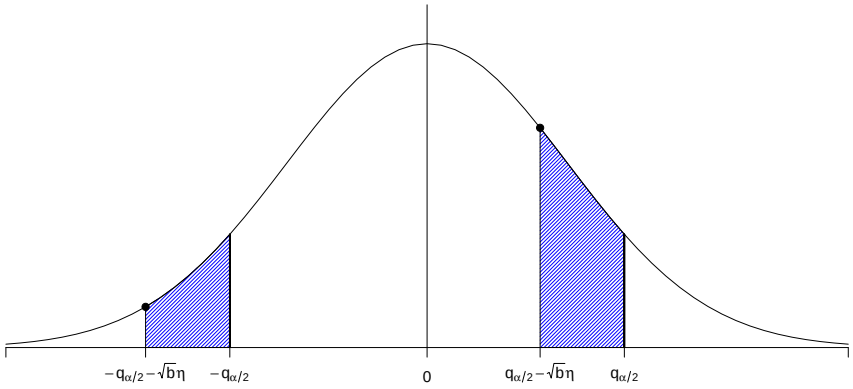


Figure 4.2: Density of the standard normal distribution

In Figure 4.2 the blue-shaded area on the right-hand side corresponds to the first integral in the definition of $\psi(\alpha, \eta, b)$, whereas the blue-shaded area on the left-hand side corresponds to the second integral and is subtracted. Apart from that, the values of $\phi(q_{\alpha/2} - \sqrt{b}\eta)$ and $\phi(-q_{\alpha/2} - \sqrt{b}\eta)$ respectively are indicated as dots in the picture. It becomes intuitively clear, that both the difference between the areas (equal to $\psi(\alpha, \eta, b)$) as well as the difference between the points (equal to $\frac{\partial}{\partial \eta}\psi(\alpha, \eta, b)$ divided by $\sqrt{b}$) are positive.
We have shown that the asymptotically normal test statistic T_b based on an estimator $\hat{\theta}$ for the copula parameter θ on a subsample of size b is suitable to test $\mathrm{H}_0 \colon \{\theta = \theta_0\}$. The test against the local alternative $\bar{\mathrm{H}}_A(\eta) \colon \{\theta = \theta_b(\eta) = \theta_0 + \eta/\sqrt{b}\}$ has been shown to have local power of order $\sqrt{b}$. We will now take advantage of the fact that the estimators $\hat{\theta}_1, \ldots, \hat{\theta}_n$ as in Definition 4.1.4(1) are established on subsamples that do not overlap, i.e. that are independent, to construct a multiple test.

4.2.2 Multiple test

Our aim is to simultaneously test the following pairs of hypotheses

$$\mathrm{H}_{0k} \colon \{\theta_{t_k} = \theta_0\} \quad \text{vs.} \quad \bar{\mathrm{H}}_{Ak}(\eta) \colon \{\theta_{t_k} = \theta_b(\eta)\} \tag{4.2.7}$$

for all $t_k = T - (n-k)b$, $n = \lfloor \frac{T-b}{b} \rfloor$. This notation also implies that we limit the following considerations to the local alternatives $\theta_b(t) = \theta_0 + \eta/\sqrt{b}$, $\eta \neq 0$. For every

$k = 1, \ldots, n$ construct the partial test statistics

$$T_{bk} := \frac{\sqrt{b}(\hat{\theta}_k - \theta_0)}{\sqrt{\Sigma}}. \tag{4.2.8}$$

Consider the random variables $Y_k = \mathbb{1}_{\{T_{bk} \in K(\eta)\}}$ that indicate whether the k-th partial test rejects the null hypothesis H_{0k} to the level $\alpha \in (0,1)$. Summing them up, we obtain

$$S_n = \sum_{k=1}^{n} Y_k \sim \mathrm{Bin}(n, \alpha) \tag{4.2.9}$$

under the null hypothesis of parameter constancy. Note that since the partial tests are asymptotical tests, the size of the above Binomial distribution is only approximatively α. S_n is now suitable to test the simultaneous null hypothesis $\mathbf{H}_0 = \bigcap_{k=1}^{n} \mathrm{H}_{0k}$ of all the partial null hypotheses being true. The alternatives might range from one false hypothesis up to all hypotheses being false. Formally, we specify the alternative hypothesis for $1 \leq m \leq n$ by $\mathbf{H}_{A,m}(\eta) = \bigcap_{k=1}^{m} \bar{\mathrm{H}}_{Ak}(\eta) \cap \bigcap_{k=m+1}^{n} \mathrm{H}_{0k}$, that is

$$\mathbf{H}_{A,m}(\eta)\colon \{\theta_{t_k} = \theta_b(\eta) \text{ for } k \leq m,\ \theta_{t_k} = \theta_0 \text{ for } k > m\}. \tag{4.2.10}$$

Note that in this set-up all the partial alternatives have to be of the same absolute value with respect to η.
The null hypothesis $\mathbf{H}_0$ is rejected to the significance level $\alpha^* \in (0,1)$ if $S_n > c_{\alpha^*}$ where c_{α^*} is the $(1-\alpha^*)$-quantile of the Binomial distribution with size n and probability α. The critical region has the form $K = \{c_{\alpha^*} + 1, c_{\alpha^*} + 2, \ldots, n\}$. Note that with fixed $\alpha, \alpha^* \in (0,1)$ the rejection probability under the null hypothesis $\mathbf{H}_0$ is in general

$$\mathbb{P}_{\mathbf{H}_0}(S_n > c_{\alpha^*}) = \sum_{j=c_{\alpha^*}+1}^{n} \binom{n}{j} \alpha^j (1-\alpha)^{n-j} \leq \alpha^*, \tag{4.2.11}$$

i.e. the size of the test is less than the desired significance level α^* for most of the time. This is usually the case when the distribution of the test statistic is discrete. In order to actually attain the proper size, Binomial tests of such kind have to be *randomized*. The existence of p in the following modification of the sample function φ is due to the *fundamental lemma of Neyman and Pearson* (see Theorem 3.2.1 in Lehmann and Romano, 2005). The lemma basically states that for every given significance level $\alpha \in (0,1)$, a test that exactly attains this level can be constructed.

Definition 4.2.4
Let a pair of hypotheses H_0 vs. H_A and the significance level $\alpha \in (0,1)$ be given. Denote by ∂K the boundary of the respective critical region K and by $\overline{K}$ its closure, and set

$$\tilde{\varphi}(x_1, \ldots, x_b) = \begin{cases} 1, & (x_1, \ldots, x_b) \in K, \\ R, & (x_1, \ldots, x_b) \in \partial K, \\ 0, & (x_1, \ldots, x_b) \notin \overline{K}, \end{cases} \tag{4.2.12}$$

where $R \sim \mathrm{Bin}(1, p)$ and the value $p \in (0,1)$ can be chosen such that $\mathbb{P}_{\mathrm{H}_0}(\tilde{\varphi} = 1) = \alpha$ and $\mathbb{P}_{\mathrm{H}_A}(\tilde{\varphi} = 0) \geq \alpha$. This test is called a *randomized test* to the level α.

In case of the Binomial test statistic S_n in (4.2.9) we set

$$\tilde{\varphi}(s) = \begin{cases} 1, & c_{\alpha^*} \leq s-1, \\ R, & s-1 < c_{\alpha^*} \leq s, \\ 0, & s < c_{\alpha^*}, \end{cases} \tag{4.2.13}$$

for a realization $0 \leq s \leq n$ of $S \sim \text{Bin}(n, \alpha)$, and choose $p \in (0,1)$ such that

$$1 \cdot \mathbb{P}\left(S \geq c_{\alpha^*} + 1\right) + p \cdot \mathbb{P}\left(S = c_{\alpha^*}\right) = \alpha^*, \tag{4.2.14}$$

i.e. $p = \left(\alpha^* - \mathbb{P}\left(S \geq c_{\alpha^*} + 1\right)\right) / \mathbb{P}\left(S = c_{\alpha^*}\right)$. Practically, we sample r from $R \sim \text{Bin}(1, p)$ whenever the specific realization s of S_n is equal c_{α^*}, and reject $\mathbf{H}_0$ if $s + r = c_{\alpha^*} + 1$. It is highly recommended to avoid randomized tests as the outcome for a specific sample with a fixed significance level is not deterministic, and hence not reproducible (see Lehmann and Romano, 2005, p. 75). Since we sum over integer-valued j in (4.2.11), it is possible to fix $\kappa := c_{\alpha^*} + 1 \in \mathbb{N}$ and then solve

$$\sum_{j=\kappa}^{n} \binom{n}{j} \alpha^j (1-\alpha)^{n-j} = \alpha^* \tag{4.2.15}$$

for the partial level $\alpha \in (0,1)$. With this α, it holds that $\kappa - 1 = c_{\alpha^*}$ and that

$$\mathbb{P}_{\mathbf{H}_0}(S_n > c_{\alpha^*}) = \alpha^*. \tag{4.2.16}$$

There cannot be given a clear advice how to choose κ. Its choice depends very much on the number n of partial hypotheses and on the alternative. We will discuss this question at the end of the current section. Formula (4.2.15) is also given in Teriokhin et al. (2007) (note that there is an error in the lower bound of the second sum due to a misprint of 1 for $l = n - m$). They compare the Binomial multiple test to other testing procedures for multiple comparisons such as the Simes (1986) or the Fisher (1970) test.
Now, we will turn to the local power $\mathbb{P}_{\mathbf{H}_{A,m}(\eta)}(S_n > c_{\alpha^*})$ of the multiple test and show that the local power of order $\sqrt{b}$ confers to the multiple test regardless of the number m of false partial alternatives. First, we establish an exact formula for the power function depending on the partial local power $\gamma = \bar{\gamma}(\eta)$.

Lemma 4.2.5
Let $\gamma \in [\alpha, 1]$ be the (local) power of the partial tests. Let $\Gamma(\eta) := \mathbb{P}_{\mathbf{H}_{A,m}(\eta)}(S_n > c_{\alpha^*})$ be the (local) power of the multiple test under the alternative hypothesis $\mathbf{H}_{A,m}(\eta)$ given in (4.2.10). Then we have

$$\Gamma(\eta) = \sum_{j=c_{\alpha^*}+1}^{n} \sum_{i=0 \vee j-(n-m)}^{j \wedge m} \binom{m}{i} \gamma^i (1-\gamma)^{m-i} \binom{n-m}{j-i} \alpha^{j-i} (1-\alpha)^{n-m-(j-i)} \tag{4.2.17}$$

for $\eta \in \mathbb{R}$, $m = 0, \ldots, n$ and with $x \vee y = \max\{x, y\}$ and $x \wedge y = \min\{x, y\}$.

Proof
Consider the following quantities that are Binomially distributed under $\mathbf{H}_{A,m}(\eta)$ with sizes being approximatively γ and α respectively

$$S_m^{(\gamma)} := \sum_{k=1}^{m} Y_k \sim \mathrm{Bin}(m, \gamma),$$
$$S_{n-m}^{(\alpha)} := \sum_{k=m+1}^{n} Y_k \sim \mathrm{Bin}(n-m, \alpha).$$

Note that $S_n = S_m^{(\gamma)} + S_{n-m}^{(\alpha)}$, and $S_m^{(\gamma)}$ and $S_{n-m}^{(\alpha)}$ are independent random variables. In the following let $\mathbb{P} = \mathbb{P}_{\mathbf{H}_{A,m}(\eta)}$ be the probability under $\mathbf{H}_{A,m}(\eta)$. Then the local power of the multiple test is

$$\begin{aligned}
\mathbb{P}\left(S_n > c_{\alpha^*}\right) &= \sum_{j=c_{\alpha^*}+1}^{n} \mathbb{P}\left(S_n = j\right) \\
&= \sum_{j=c_{\alpha^*}+1}^{n} \mathbb{P}\left(S_m^{(\gamma)} + S_{n-m}^{(\alpha)} = j\right) \\
&= \sum_{j=c_{\alpha^*}+1}^{n} \sum_{i=0\vee j-(n-m)}^{j\wedge m} \mathbb{P}\left(S_m^{(\gamma)} = i\right)\, \mathbb{P}\left(S_{n-m}^{(\alpha)} = j-i\right) \\
&= \sum_{j=c_{\alpha^*}+1}^{n} \sum_{i=0\vee j-(n-m)}^{j\wedge m} \binom{m}{i} \gamma^i (1-\gamma)^{m-i} \binom{n-m}{j-i} \alpha^{j-i} (1-\alpha)^{n-m-(j-i)}.
\end{aligned}$$

□

Under $\mathbf{H}_0$, we obviously have $\gamma(0) = \alpha$, $S_n \sim \mathrm{Bin}(n, \alpha)$, and therefore $\Gamma(0) = \alpha^*$ due to (4.2.16). To prove that $\Gamma(\eta) > \alpha^*$ for $\eta \neq 0$, we can interpret S_n as the sum of two independent, binomially distributed random variables under the null hypothesis $\mathbf{H}_0$, i.e. as $S_n = S_m^{(\alpha)} + S_{n-m}^{(\alpha)}$. We will show that for $\eta \neq 0$ we have

$$\alpha^* = \mathbb{P}_{\mathbf{H}_0}(S_n > c_{\alpha^*}) \;<\; \mathbb{P}_{\mathbf{H}_{A,m}(\eta)}(S_n > c_{\alpha^*}) = \Gamma(\eta). \tag{4.2.18}$$

For this we need the definitions of *stochastic orderings* and *stochastic coupling*.

Definition 4.2.6 (1) Let μ and ν be two probability laws. Then μ is *stochastically smaller* than ν, denoted by $\mu \leq_{st} \nu$, if for all $x \in \mathbb{R}$

$$\mu([x, \infty)) \leq \nu([x, \infty)). \tag{4.2.19}$$

Equivalently, for random variables $X \sim \mathbb{P}_\mu$ and $Y \sim \mathbb{P}_\nu$,

$$X \leq_{st} Y \quad \Longleftrightarrow \quad \mathbb{P}_\mu(X > x) \leq \mathbb{P}_\nu(Y > x) \tag{4.2.20}$$

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

(2) For a collection X_i, $i \in I$ of random variables generally on different probability spaces and with an arbitrary index set I, a corresponding *coupling* is a collection of random variables $\tilde{X}_i$ on a common probability space, such that $X_i \overset{d}{=} \tilde{X}_i$ for all $i \in I$. We stress the fact that the random variables $\tilde{X}_i$ are defined on a joint probability space by denoting the collection as $\{\tilde{X}_i\}_{i \in I}$ in this case.

The technique of coupling is useful whenever properties of stochastic laws are to be shown as well as when the joint distribution of some random variables is of interest. Note that even if the original variables X_i do not have a joint distribution, it is possible to construct the coupling $\tilde{X}_i$ with a specified joint law. In the proof of Proposition 4.2.7 we used so-called *quantile coupling*. For a deeper account on coupling see Thorisson (2000).

Proposition 4.2.7
Let $n \in \mathbb{N}$ and $0 \leq p_1 < p_2 \leq 1$. Then we have

$$\text{Bin}(n, p_1) \leq_{st} \text{Bin}(n, p_2). \tag{4.2.21}$$

Proof
Let $U_1, \ldots, U_n \in [0,1]$ be independent and uniformly distributed random variables on a joint probability space. Define $N_i = \sum_{k=1}^{n} \mathbb{1}_{\{U_k \leq p_i\}}$ for $i = 1, 2$. Since $\{U_k \leq p_1\} \subseteq \{U_k \leq p_2\}$, we have

$$N_1 = \sum_{k=1}^{n} \mathbb{1}_{\{U_k \leq p_1\}} \overset{a.s.}{\leq} \sum_{k=1}^{n} \mathbb{1}_{\{U_k \leq p_2\}} = N_2.$$

The pointwise inequality carries over to the respective laws, $N_1 \sim \text{Bin}(n, p_1)$ and $N_2 \sim \text{Bin}(n, p_2)$, such that $\text{Bin}(n, p_1) \leq_{st} \text{Bin}(n, p_2)$. □

It can be argued with coupling as well that stochastic orderings are preserved under convolution.

Proposition 4.2.8
Let $X_1 \leq_{st} X_1'$ and $X_2 \leq_{st} X_2'$ such that X_1 is independent of X_2 and X_1' is independent of X_2'. Then we have

$$X_1 + X_2 \leq_{st} X_1' + X_2'. \tag{4.2.22}$$

Proof
If $X_1 \leq_{st} X_1'$, then there exists a coupling $(\tilde{X}_1, \tilde{X}'_1)$ of X_1 and X_1', such that $\tilde{X}_1 \leq \tilde{X}'_1$. The same is true for $(\tilde{X}_2, \tilde{X}'_2)$, a coupling of X_2 and X_2', such that $\tilde{X}_2 \leq \tilde{X}'_2$. Let $(\tilde{X}_1, \tilde{X}'_1)$ and $(\tilde{X}_2, \tilde{X}'_2)$ be independent bivariate random vectors.
Then $(\tilde{X}_1 + \tilde{X}_2, \tilde{X}'_1 + \tilde{X}'_2)$ is a coupling of $X_1 + X_2$ and $X_1' + X_2'$ such that

$$\tilde{X}_1 + \tilde{X}_2 \leq \tilde{X}'_1 + \tilde{X}'_2.$$

The pointwise domination carries over to the original random variables and hence results in stochastic order. □

Theorem 4.2.9
The Binomial test $S_n > c_{\alpha^*}$ against the alternative $\mathbf{H}_{A,m}(\eta)$ is asymptotically unbiased for $\eta \neq 0$, and thus has local power of order $\sqrt{b}$ for all $m = 1, \ldots, n$.

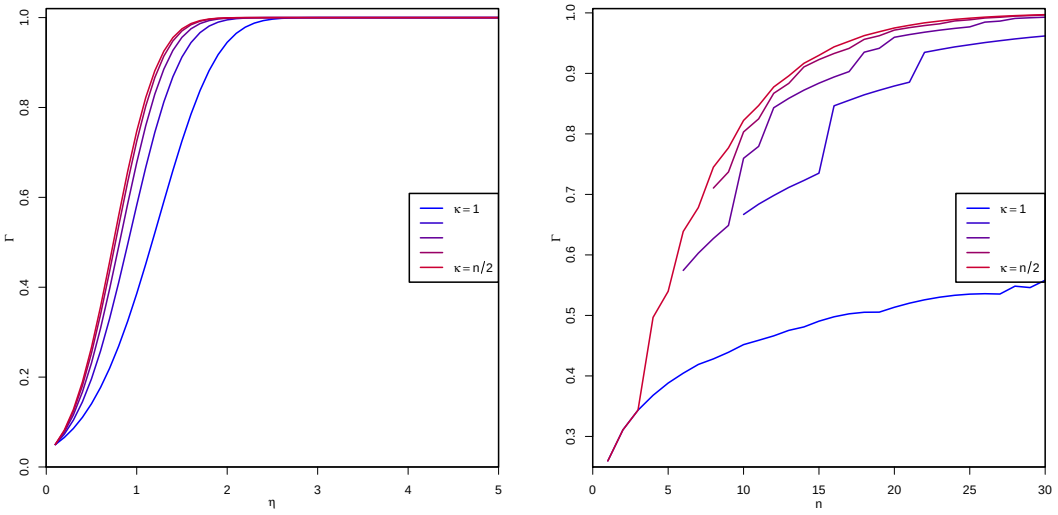
Proof
As shown in Section 4.2.1, the power $\gamma(\eta)$ of the partial tests is always greater than its nominal level α for $\eta \neq 0$, and according to Proposition 4.2.7 we have $S_m^{(\alpha)} \leq_{st} S_m^{(\gamma)}$. Trivially, it also is $S_{n-m}^{(\alpha)} \leq_{st} S_{n-m}^{(\alpha)}$. Since $S_m^{(.)}$ and $S_{n-m}^{(.)}$ are independent in any case, Proposition 4.2.8 can be applied to conclude

$$S_n|_{\mathbf{H}_0} = S_m^{(\alpha)} + S_{n-m}^{(\alpha)} \leq_{st} S_m^{(\gamma)} + S_{n-m}^{(\alpha)} = S_n|_{\mathbf{H}_{A,m}(\eta)}. \tag{4.2.23}$$

By definition this implies (4.2.18), i.e. $\Gamma(\eta) > \alpha^*$ for $\eta \neq 0$, and therefore unbiasedness of the Binomial test. Analogously to the convergence of the hypothesis pair for the partial test, we have that $\mathbf{H}_{A,m}(\eta) - \mathbf{H}_0 \asymp \sqrt{b}$ as $b \to \infty$, so that the test has local power of the desired order. □

Optimal parameter choice

Considering the local power of the Binomial multiple test $\Gamma(\eta)$ as a function of η and of the number n of partial tests, we want to figure out how to choose κ in (4.2.15) properly. We want to test the null hypothesis $\mathbf{H}_0$ against the multiple alternative $\mathbf{H}_{A,m}(\eta)$, $1 \leq m \leq n$, which is composed of m "false" partial hypotheses $\mathrm{H}_{Ak}(\eta)\colon \{\theta_{t_k} = \theta(\eta) = \theta_0 + \eta/\sqrt{b}\}$ for $k = 1, \ldots, m$ and $n - m$ "true" partial hypotheses $\mathrm{H}_{0k}\colon \{\theta_{t_k} = \theta_0\}$ for $k = m+1, \ldots, n$. Depending on m in the global alternative, a different choice of κ is suitable.

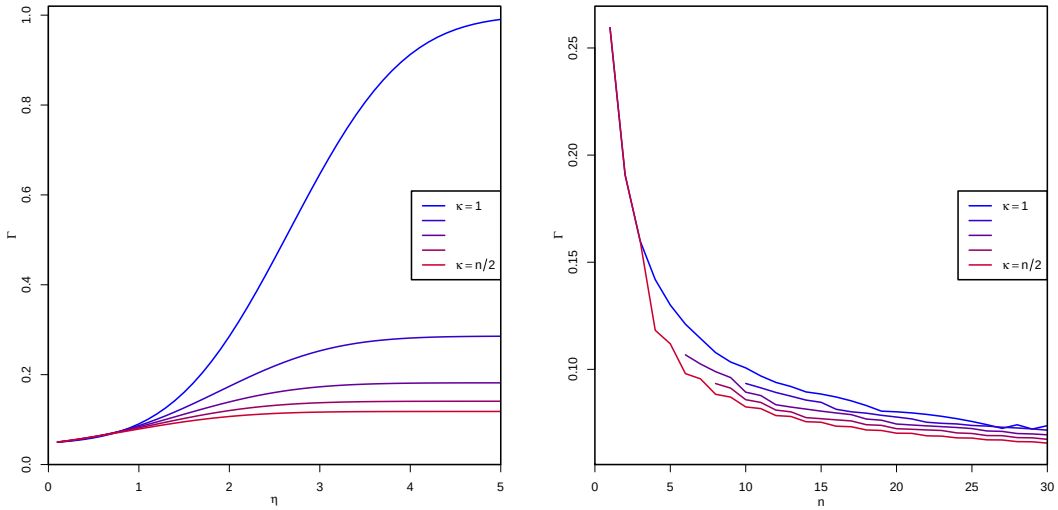


(a) $n = 10$ fixed, η varies

(b) $\eta = 1$ fixed, n varies

Figure 4.3: $\Gamma(\eta)$ under $\mathbf{H}_{A,n}(\eta)$, $\eta > 0$

It is easy to see that if all partial hypotheses are false, i.e. $m = n$, then the test statistic S_n is $\mathrm{Bin}(n, \gamma)$-distributed. For $\kappa \to n$, the local level $\alpha \to 1$ and simultaneously $\gamma \to 1$. Let $S_n(\kappa)$ denote the test statistic where the local level α is determined depending on κ, then Proposition 4.2.7 yields that for $\kappa_1 \leq \kappa_2$ we have $\mathbb{P}(S_n(\kappa_1) > c) \leq \mathbb{P}(S_n(\kappa_2) > c)$ for all $c = 0, \ldots, n$. This fact implies increasing power $\Gamma(\eta)$ with increasing κ. Nevertheless, we cap κ by $n/2$ for practical reasons. The situation for $n = 10$ in case of one-sided partial tests is depicted in Figure 4.3(a). The same is true if the partial tests are two-sided.
For a fixed $\gamma = \gamma(\eta)$ and with increasing n the same order of the power curves persists. The case for $\eta = 1$ is depicted in Figure 4.3(b). The lines in the colors between red and blue indicate some of the intermediate set-ups between $\kappa = 1$ and $\kappa = \lfloor n/2 \rfloor$.

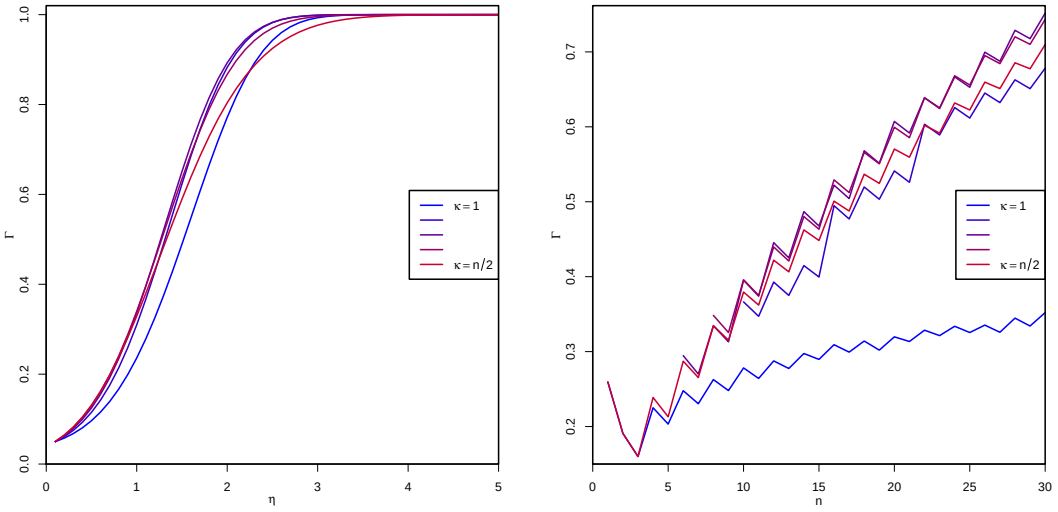


(a) $n = 10$ fixed, η varies

(b) $\eta = 1$ fixed, n varies

Figure 4.4: $\Gamma(\eta)$ under $\mathbf{H}_{A,1}(\eta)$, $\eta > 0$

This relation is just reversed if the alternative to test against consists of only one false partial hypothesis. This is due to the fact that the tails of distribution functions are monotonically non-increasing, in particular those of convolutions of Binomial distributions (see proof of Theorem 4.2.9). Therefore, $\mathbb{P}(S_n > c_1) \geq \mathbb{P}(S_n > c_2)$ for $c_1 \leq c_2$, and thus the power is maximized for κ as small as possible. Note that the contribution of a changing κ on α does not outweigh the contribution of κ on γ. This finding is depicted in Figure 4.4 for both cases of fixed $n = 10$ and varying η as well as for fixed $\eta = 1$ and increasing n.
In all other cases $1 < m < n$ the ordering of the power curves $\Gamma(\eta)$ is not monotone anymore. That means that the contributions of κ on α and γ are not balanced in the sense that the dependence of $\Gamma(\eta)$ on κ remains monotonic. Depending on η (for fixed n) there are some intersections. This can be seen in Figure 4.5(a). The same happens for fixed η and a growing number n of partial tests. The intersections are depicted in Figure 4.5(b).



(a) $n = 10$ fixed, t varies (b) $t = 1$ fixed, n varies

Figure 4.5: $\Gamma(\eta)$ under $\mathbf{H}_{A,n/2}(\eta)$, $\eta > 0$

Concerning the optimal parameter choice for κ, a definitive advice can only be given if the alternative to be tested against is in a sense extreme, i.e. if it is either the falsity of all partial hypotheses or the falsity of only one. In all other cases, the power results corresponding to certain values of κ depend on the number n of partial tests.
Summarizing, we have constructed a multiple Binomial test with proper size and local power of order $\sqrt{b}$. The multiple null hypothesis is composed of n independent partial null hypotheses and is rejected when there are at least κ partial rejections. We are free to choose $1 \leq \kappa \leq \lfloor n/2 \rfloor$, and arrive at a test that attains the significance level α^* under the null hypothesis. In order to maximize the local power, we should take into account the number of false partial hypotheses in our choice of κ.
In the subsequent section we turn to another approximation of the global null hypothesis given in (4.0.2) that allows to test against a more general alternative. The test we will construct makes use of a greater portion of the estimator series $\{\hat{\theta}_t\}$ by allowing the underlying subsamples to overlap.

4.3 Extremal test

The null hypothesis $\mathbf{H}_0^*$: $\{\theta_t \equiv \theta_0 \; \forall t\}$ means that the copula parameter is constant for all points in time t. A consequence of this is that in particular $\max_t \theta_t = \theta_0$. Thus, we will subsequently test the following pair of hypotheses

$$\mathbf{H}_0' \colon \{\max_t \theta_t = \theta_0\} \quad \text{vs.} \quad \mathbf{H}_A' \colon \{\max_t \theta_t > \theta_0\} \tag{4.3.1}$$

for an arbitrary $\theta_0 \in \Theta$. As basis for our test statistic, consider a fixed constant c with $1/b < c < 1$ and the thinned estimator series $\{\hat{\theta}_{t_k}\}$ for $t_k = T - (n^{(c)} - k)cb$,

$k = 1, \ldots, n^{(c)} = \lfloor (T-b)/cb \rfloor$ (cf. Definition 4.1.4(2)).
For notational convenience we use a centered and scaled version of $\hat{\theta}_{t_k}$ for all $k = 1, \ldots, n^{(c)}$, such that under $\mathbf{H}_0^*$ we have

$$\xi_k := \sqrt{b}(\hat{\theta}_{t_k} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma) \tag{4.3.2}$$

as $b \to \infty$ due to Proposition 4.1.2(2).
In order to test the pair of hypotheses in (4.3.1), a test statistic based on the vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_{n^{(c)}})$ is suitable. For assessing the statistic's distribution, we first need the joint distribution of the vector $\boldsymbol{\xi}$, and secondly, extreme value theory for asymptotically normal variables. This will be elaborated in Sections 4.3.1 and 4.3.2. Subsequently, these results will be used to construct a test and establish its local power in Section 4.3.3.

4.3.1 Multivariate central limit theorem

We will trace back the structure of $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_{n^{(c)}})$ to the summands in the likelihood estimation, and then establish its asymptotic distribution. To this end we will work on a random quantity $\boldsymbol{S}_b$ exhibiting a special structure that we need to obtain our results. Although the subsample size b grows with the overall observation horizon T, the overlap proportion remains fixed. A reordering of the original summation enables us to apply a recent result on convergence rates in the multivariate central limit theorem that depends on the involved dimension (due to Bentkus, 2004). Together with bounds on the elements in the square-root inverse of the occurring covariance matrices (a result by Demko et al., 1984), this finally leads to conditions on b and T that guarantee distributional convergence of $\boldsymbol{S}_b$ to a Gaussian distribution.

Let $X_1, \ldots, X_T$ be a sample of length T of independent and identically distributed random variables with $\mathbb{E}\, X_1 = 0$, $\mathbb{E}\, X_1^2 = \Sigma < \infty$ and $\mathbb{E}\,|X_1|^3 < \infty$.
Furthermore, choose a bandwidth $b < T$ such that b tends to infinity as $T \to \infty$, but at the same time $b/T \to 0$. Denote this restriction by $b = b(T) = \mathcal{O}(T^\delta)$, $\delta \in (0,1)$. For a constant $c \in (1/b, 1)$, consider now sums over subsamples of length b

$$S_k := X_{t_k - b + 1} + \ldots + X_{t_k} \tag{4.3.3}$$

starting at $t_k = \lfloor T - (n^{(c)} - k)cb \rfloor$ for all $k = 1, \ldots, n^{(c)} = \lfloor \frac{T-b}{cb} \rfloor$. The respective subsamples overlap by $\lfloor (1-c)b \rfloor$ summands. For notational convenience, assume that $c = 1/\mathcal{K}$, $\mathcal{K} \in \mathbb{N}$, and that $cb \in \mathbb{N}$. Furthermore, choose T and b in such a way that $t_1 = 1$ and $t_{n^{(c)}} = T - b + 1$, i.e. the whole sample is used.
The vector $\boldsymbol{S} = (S_1, \ldots, S_{n^{(c)}})$ of the subsample sums will be normalized by $1/\sqrt{b}$ and

denoted by $\boldsymbol{S}_b$ in the following. To be precise, we have

$$\boldsymbol{S}_b = \frac{1}{\sqrt{b}} \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_{n^{(c)}} \end{pmatrix} = \frac{1}{\sqrt{b}} \begin{pmatrix} X_1 & + & X_2 & + & \dots & + & X_b \\ X_{cb+1} & + & X_{cb+2} & + & \dots & + & X_{cb+b} \\ X_{2cb+1} & + & X_{2cb+2} & + & \dots & + & X_{2cb+b} \\ \vdots & & \vdots & & & & \vdots \\ \underbrace{X_{n^{(c)}cb+1}}_{=:\boldsymbol{X}^{(1)}} & + & \underbrace{X_{n^{(c)}cb+2}}_{=:\boldsymbol{X}^{(2)}} & + & \dots & + & \underbrace{X_{n^{(c)}cb+b}}_{=:\boldsymbol{X}^{(b)}} \end{pmatrix}.$$

Hence, our target variable $\boldsymbol{S}_b = (\boldsymbol{X}^{(1)} + \ldots + \boldsymbol{X}^{(b)})/\sqrt{b}$ can be viewed as a sum of $n^{(c)}$-dimensional vectors. For all $k = 1, \ldots, b$, the vectors $\boldsymbol{X}^{(k)}$ have mean zero and Σ times the identity matrix as covariance matrix, i.e.

$$\mathbb{E}\, \boldsymbol{X}^{(k)} = \boldsymbol{0},\ \operatorname{Cov} \boldsymbol{X}^{(k)} = \Sigma \mathbf{I}_{n^{(c)}}.$$

Although they are identically distributed as X_i are iid, they exhibit a very unfavorable dependence structure. This is due to the fact that for all $i = 1, \ldots, cb$ the vectors $\boldsymbol{X}^{(\nu cb+i)}$ and $\boldsymbol{X}^{(\mu cb+i)}$ have $n^{(c)} - (\mu - \nu)$ identical components as long as $0 \le \nu \le \mu \le \mathcal{K} - 1$. To clarify this, the case for $n^{(c)} = 7$, $\mathcal{K} = 4$ and $i = 1$ is schematically depicted in Figure 4.6. The arrows indicate which of the components coincide.
At a closer look, this scheme provides a remedy for the strong dependence between the vectors $\boldsymbol{X}^{(k)}$. For every $i = 1, \ldots, cb$ define the $n^{(c)}$-dimensional vector $\boldsymbol{Y}^{(i)}$ by summing up the vectors $\boldsymbol{X}^{(\nu cb+i)}$ that share common components and are therefore highly dependent

$$\boldsymbol{Y}^{(i)} = \boldsymbol{X}^{(i)} + \boldsymbol{X}^{(cb+i)} + \ldots + \boldsymbol{X}^{((\mathcal{K}-1)cb+i)}. \tag{4.3.4}$$

Then $\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(cb)}$ are iid random vectors with mean zero and covariance matrix $\operatorname{Cov} \boldsymbol{Y}^{(1)} = \Sigma M$ where

$$M = \begin{pmatrix} \mathcal{K} & \mathcal{K}-1 & \dots & 1 & & & & \\ \mathcal{K}-1 & \ddots & \ddots & & \ddots & & 0 & \\ \vdots & \ddots & \ddots & \ddots & & \ddots & & \\ 1 & & \ddots & \ddots & \ddots & & \ddots & \\ & \ddots & & \ddots & \ddots & \ddots & & 1 \\ & & \ddots & & \ddots & \ddots & \ddots & \vdots \\ & 0 & & \ddots & & \ddots & \ddots & \mathcal{K}-1 \\ & & & & 1 & \dots & \mathcal{K}-1 & \mathcal{K} \end{pmatrix}. \tag{4.3.5}$$

M is not only symmetric and positive definite as a covariance matrix, but also $(\mathcal{K}-1)$-banded and Toeplitz with off-diagonal decaying elements.
A matrix $M = (m_{ij})$ is called *k-banded* if $m_{ij} = 0$ for $|i - j| > k$, that is it has k non-trivial (i.e. non-zero) secondary diagonals. A diagonal matrix is therefore 0-banded.

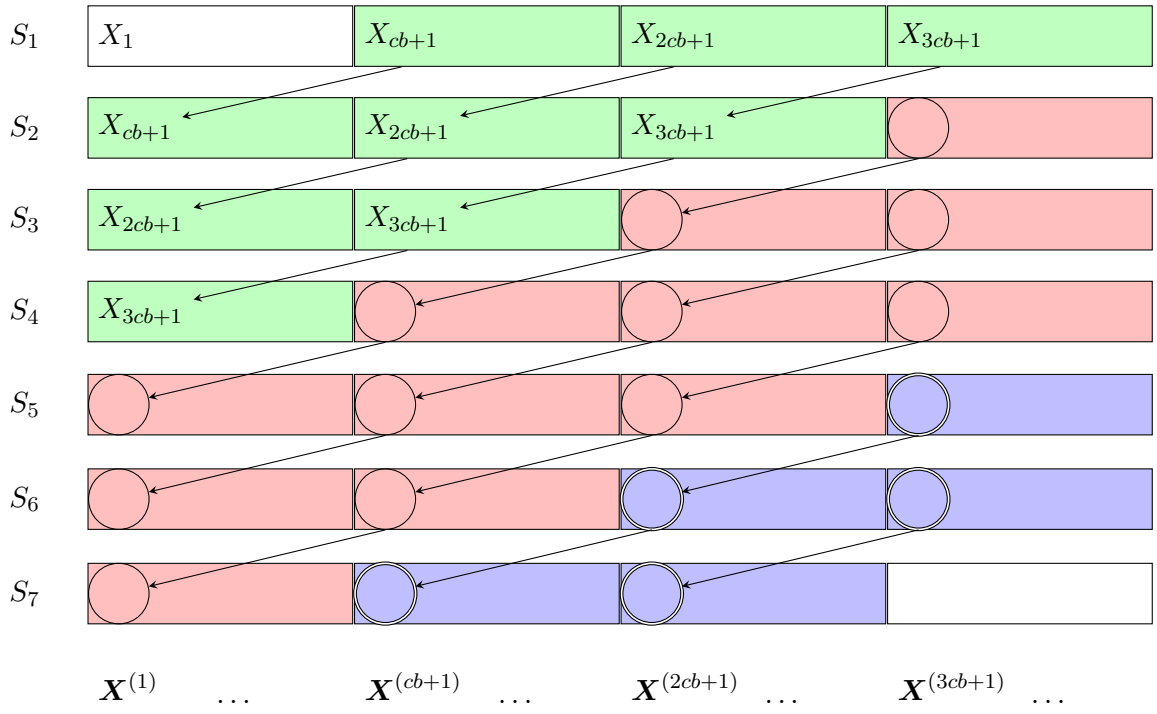


Figure 4.6: Summation Scheme

Symmetric matrices having k non-trivial bands above (and via symmetry also below) the main diagonal are also called k-banded, rather than $2k$-banded.
A *Toeplitz* or *diagonal-constant* matrix has constant entries in each descending diagonal. For symmetric Toeplitz matrices it suffices to give the generating sequence $\boldsymbol{a} = (a_0, a_1, \ldots, a_d)$ in order to know that $m_{ij} = a_{|i-j|}$. In this case, off-diagonal decaying elements are characterized by $a_0 \geq a_1 \geq \ldots \geq a_d$.
It is generally true that the Cholesky decomposition of a symmetric positive definite band matrix preserves the band structure (see e.g. Schwarz and Köckler, 2006, p. 62). Moreover, it has been recently shown that these special Toeplitz matrices have a Cholesky square-root that exhibits an off-diagonal decay in a certain wider sense. Let's sum up these two results in the following theorem.

Theorem 4.3.1
Let M_d be a symmetric k-banded Toeplitz matrix of dimension d generated by a sequence

$\boldsymbol{a} = (a_0, a_1, \ldots, a_k)$, $k < d$. That is, M_d has the following structure

$$M_d = \begin{pmatrix} a_0 & a_1 & \ldots & a_k & & 0 \\ & \ddots & \ddots & & \ddots & \\ & & \ddots & \ddots & & a_k \\ & & & \ddots & \ddots & \vdots \\ & \text{symm.} & & & \ddots & a_1 \\ & & & & & a_0 \end{pmatrix}.$$

Moreover, let $\boldsymbol{a}$ be positive, monotone and convex, i.e.

i) $a_0 \geq a_1 \geq \ldots \geq a_k \geq 0$ and

ii) $a_0 - a_1 \geq a_1 - a_2 \geq \ldots \geq a_{k-1} - a_k \geq 0$.

Then the Cholesky decomposition $M_d = G_d G_d^\top$ with $G_d = (g_{ij})_{1 \leq j \leq i \leq d}$ satisfies

i) $g_{ij} = 0$ for all $i > k + 1$,

ii) $g_{ij} \geq 0$ for all $i \leq k + 1$,

iii) $g_{ij} \geq g_{i+1,j}$.

Proof

See Theorem 1 in Berenhaut and Bandyopadhyay (2005). □

Theorem 4.3.1 states that G_d is a lower triangular, k-banded matrix with positive and monotone entries:

$$G_d = \begin{pmatrix} g_{11} & & & & \\ \vdots & \ddots & & 0 & \\ g_{k+1,1} & & \ddots & & \\ & \ddots & & \ddots & \\ 0 & & g_{d,k+1} & \cdots & g_{dd} \end{pmatrix}.$$

It is straightforwardly shown that the diagonal elements of G_d are bounded. Let $M_d = (m_{ij})_{1 \leq i,j \leq d}$. Since G_d is lower triangular, m_{ii} can be written as

$$m_{ii} = \sum_{k=1}^{i} g_{ik}^2 \geq g_{ii}^2.$$

Positive definiteness of M_d ensures that $g_{ii} > 0$. Therefore,

$$g_{ii} \leq \sqrt{m_{ii}} = \sqrt{a_0} \quad \text{for all } i = 1, \ldots, d.$$

As long as the band k remains fixed, we can consider the limits $M = \lim_{d \to \infty} M_d$ and $G = \lim_{d \to \infty} G_d$ and show their boundedness on $\ell^2(\mathbb{Z})$. To this end recall the notion of

norms $\|\,.\,\|$ on a possibly infinite-dimensional vector space V. By the *vector norm* $\|\boldsymbol{x}\|$ of $\boldsymbol{x} \in V$, $\boldsymbol{x} = (x_1, x_2, \ldots)$, we mean either the ℓ^p-norm for arbitrary $1 \leq p < \infty$, i.e.

$$\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p},$$

or the supremum norm, i.e. $\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_\infty = \sup_{i \in \mathbb{N}} |x_i|$. We additionally assume that the norms are *equivalent*, i.e. that there exist a positive constant $\rfloor$ such that $\|\boldsymbol{x}\|_1 \leq c\|\boldsymbol{x}\|_2$ for all $\boldsymbol{x} \in V$ and any of the norms $\|\,.\,\|_i$, $i = 1, 2$. The induced *operator norm* $\|A\|_{op}$ for a linear operator $A \colon V \to V$ is then

$$\|A\|_{op} = \max_{\boldsymbol{x} \neq 0} \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|$$

with the corresponding vector norm $\|\,.\,\|$.

Proposition 4.3.2
The linear operators

$$M = \lim_{d \to \infty} M_d \quad \text{and} \quad G = \lim_{d \to \infty} G_d$$

are bounded operators on $\ell^2(\mathbb{Z})$.

Proof
For all $l = 0, \ldots, k$, let $E_l = (e_{ij}^{(l)})_{i,j \in \mathbb{Z}}$ have the following entries

$$e_{ij}^{(l)} = \begin{cases} 1, & |i - j| = l, \\ 0, & \text{elsewhere.} \end{cases}$$

That is, every E_l is an infinitely dimensional matrix having one's on the l-th diagonals and zeros elsewhere. Its operator norm is finite, as

$$\|E_l\|_{op} = \sup \frac{\|E_l x\|}{\|x\|} = \sup_{\|x\|=1} \|E_l x\| < \infty \quad \text{for all } \; l,$$

so that E_l is a bounded operator. We can now set $M = \sum_{l=0}^{k} a_l E_l$. Since $a_0 < \infty$, M is also bounded as a finite weighted sum of bounded operators.
For the Cholesky square-root we have

$$\|G\|_{op} \leq \|\widetilde{G}\|_{op}, \quad \text{where we define } \widetilde{G} = \sum_{l=0}^{k} \sqrt{a_0} E_l = \sqrt{a_0} \sum_{l=0}^{k} E_l.$$

The same argument as for M yields boundedness of G. □

As for now, we have seen that G is a bounded operator and, since $g_{ii} > 0 \; \forall \, i$ and G is lower triangular, its inverse G^{-1} also exists. The Bounded Inverse Theorem (see e.g. Renardy and Rogers, 2004, p. 241) guarantees that $\|G^{-1}\|_{op} < \infty$. Therefore, G is said to be boundedly invertible.

The next theorem is due to Demko et al. (1984) and gives exponential bounds on the off-diagonal decay of banded operators that have a matrix representation. Note that the *condition* of an operator is defined as

$$\operatorname{cond}(A) = \frac{\max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|}{\min_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|}$$

with a suitable vector norm $\|\,.\,\|$, and in case of an invertible operator can be expressed as $\operatorname{cond}(A) = \|A\|_{op}\|A^{-1}\|_{op}$.

Theorem 4.3.3
Let G and its inverse $G^{-1} = (g_{ij}^-)$ be bounded operators on $\ell^2(\mathbb{Z})$. If G is k-banded and quasi-centered, i.e. the main diagonal is contained in the non-trivial matrix bands, then for all $i, j \in \mathbb{Z}$

$$|g_{ij}^-| \leq C_1 \lambda_1^{|i-j|}$$

where

$$\lambda_1 = \left(\frac{\operatorname{cond}(G)-1}{\operatorname{cond}(G)+1}\right)^{1/k} < 1 \quad \text{and} \quad C_1 = \frac{k+1}{\lambda_1^k}\|G^{-1}\|_{op} \max\left\{1, \frac{1}{2}\left(\frac{1+\operatorname{cond}(G)}{\operatorname{cond}(G)}\right)^2\right\}.$$

Proof
See Theorem 2.4 in Demko et al. (1984). □

Note that the inversion of band matrices usually destroys the band structure. In this sense, the exponential decay of the off-diagonal elements is the best we could expect. Later on, we will need the absolute summability of the entries of the square-root inverse of the covariance matrix. From this point of view, exponential decay is also desirable.
Now we can proceed to the distributional convergence of $\boldsymbol{S}_b$ as the sample size increases to ∞. In a recent publication of Bentkus (2004), Lyapunov-type Berry-Esseen bounds in the multivariate central limit theorem are given inheriting a dimensional dependence. A version of this result suitable for our purposes is stated below.

Theorem 4.3.4
Let $\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(n)} \in \mathbb{R}^d$ be independent random vectors with $\mathbb{E}\,\boldsymbol{Y}^{(k)} = \boldsymbol{0}$ for all k. Let $\boldsymbol{S} = \boldsymbol{Y}^{(1)} + \ldots + \boldsymbol{Y}^{(n)}$ and let $\mathbf{C}^2 = \operatorname{Cov} \boldsymbol{S}$. Denote by $\Phi_{d,\mathbf{C}^2}$ the distribution function of the d-variate centered Gaussian distribution with covariance matrix $\mathbf{C}^2$.
Then there exists an absolute and positive constant C_2 such that

$$\Delta := \sup_{\boldsymbol{x}\in\mathbb{R}^d} |F_{\boldsymbol{S}}(\boldsymbol{x}) - \Phi_{d,\mathbf{C}^2}(\boldsymbol{x})| \leq C_2 \cdot d^{1/4} \cdot \beta, \tag{4.3.6}$$

where $F_{\boldsymbol{S}}$ is the d-variate distribution function of $\boldsymbol{S}$ and $\beta = \beta_1 + \ldots + \beta_n$ with $\beta_k = \mathbb{E}\,|\mathbf{C}^{-1}\boldsymbol{Y}^{(k)}|^3$.

Proof
See Theorem 1.1 in Bentkus (2004). □

Besides this convergence result, the special structure of the covariance matrix will now be exploited to derive the limiting distribution of the target variable $\boldsymbol{S}_b$ as well as a convergence rate depending on the sample sizes T and b.
$\boldsymbol{S}_b$ can now be seen as the sum of the iid vectors $\boldsymbol{Y}^{(k)}$ as in equation (4.3.4)

$$\boldsymbol{S}_b = \frac{1}{\sqrt{b}} \left(\boldsymbol{Y}^{(1)} + \ldots + \boldsymbol{Y}^{(cb)} \right) \tag{4.3.7}$$

with covariance matrix

$$\mathbf{C}^2 = \operatorname{Cov} \boldsymbol{S}_b = c\, \Sigma M \tag{4.3.8}$$

where c is the thinning constant. Theorem 4.3.3 and Theorem 4.3.4 are the principal building blocks in the proof of the following theorem. The first main result of this thesis is essential in deriving the maximal limiting law of the components of $\boldsymbol{S}_b$ later in Section 4.3.3.

Theorem 4.3.5
For $b = T^\delta$, $\delta \in (0,1)$ and $n^{(c)} = \mathcal{O}(T/b)$, let Δ_b be defined in analogy to equation (4.3.6), namely as

$$\Delta_b = \sup_{\boldsymbol{x} \in \mathbb{R}^{n^{(c)}}} |F_{\boldsymbol{S}_b}(\boldsymbol{x}) - \Phi_{n^{(c)}, \mathbf{C}^2}(\boldsymbol{x})|.$$

If $\delta > 7/9$, then $\Delta_b \to 0$ as $T \to \infty$.

Proof
According to equation (4.3.4) the vectors $\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(cb)} \in \mathbb{R}^{n^{(c)}}$ are iid with $\mathbb{E}\, \boldsymbol{Y}^{(k)} = \mathbf{0}$ and $\operatorname{Cov} \boldsymbol{Y}^{(k)} = \Sigma M$ as in equation (4.3.5).
For $\boldsymbol{S}_b = \frac{1}{\sqrt{b}} \left(\boldsymbol{Y}^{(1)} + \ldots + \boldsymbol{Y}^{(cb)} \right)$ the covariance matrix $\mathbf{C}^2 = \operatorname{Cov} \boldsymbol{S}_b$ is given by

$$\mathbf{C}^2 = c\, \Sigma M.$$

Let $\mathbf{C}^{-1} = (c_{ij}^-)$ denote its square-root inverse, i.e. $\mathbf{C}^{-1} = (\sqrt{\mathbf{C}^2})^{-1} = G^{-1}/\sqrt{c\,\Sigma}$ where $M = G\, G^\top$ is the Cholesky decomposition of M, and let $G^{-1} = (g_{ij}^-)$.
For $n = cb$ and $d = n^{(c)}$, Theorem 4.3.4 yields

$$\begin{aligned} \Delta_b &= \sup_{\boldsymbol{x} \in \mathbb{R}^{n^{(c)}}} |F_{\boldsymbol{S}_b}(\boldsymbol{x}) - \Phi_{n^{(c)}, \mathbf{C}^2}(\boldsymbol{x})| \\ &\leq C_2 \cdot {n^{(c)}}^{1/4} \cdot \beta \end{aligned}$$

where $\beta = \sum_{k=1}^{cb} \beta_k = cb \cdot \beta_1$. The last equality holds due to the fact that $\boldsymbol{Y}^{(k)}$ are iid. Let $\boldsymbol{Y} \overset{d}{=} \boldsymbol{Y}^{(k)}$, then $\beta_1 = \mathbb{E}\, |\mathbf{C}^{-1} \boldsymbol{Y}|^3$.

For a vector $\boldsymbol{x} = (x_1, \ldots, x_{n^{(c)}}) \in \mathbb{R}^{n^{(c)}}$ let $|\boldsymbol{x}|^2 = \sum_{i=1}^{n^{(c)}} x_i^2$, then we have

$$\begin{aligned}
\mathbb{E}\,|\mathbf{C}^{-1}\boldsymbol{Y}|^3 &= \mathbb{E}\left(\left| \left(\sum_{j=1}^{n^{(c)}} c_{ij}^{-} Y_j \right)_{i=1,\ldots,n^{(c)}} \right|^2 \right)^{3/2} \\
&= \mathbb{E}\left(\sum_{i=1}^{n^{(c)}} \left(\sum_{j=1}^{n^{(c)}} c_{ij}^{-} Y_j \right)^2 \right)^{3/2} \\
&= \mathbb{E}\left(\sum_{i=1}^{n^{(c)}} \tilde{X}_i{}^2 \right)^{3/2},
\end{aligned}$$

where we define the random variables $\tilde{X}_i := \sum_{j=1}^{n^{(c)}} c_{ij}^{-} Y_j$ for all i. Let us now take advantage of $\tilde{X}_i{}^2 = |\tilde{X}_i{}^2 \cdot 1|$ and apply the Hölder inequality with $p = 3/2$ and $q = 3$ to further conclude

$$\begin{aligned}
\mathbb{E}\,|\mathbf{C}^{-1}\boldsymbol{Y}|^3 &\leq \mathbb{E}\left(\left(\sum_{i=1}^{n^{(c)}} |\tilde{X}_i{}^2|^{3/2} \right)^{2/3} \cdot \left(\sum_{i=1}^{n^{(c)}} 1^3 \right)^{1/3} \right)^{3/2} \\
&= {n^{(c)}}^{1/2} \cdot \mathbb{E}\left(\sum_{i=1}^{n^{(c)}} |\tilde{X}_i|^3 \right) \\
&\leq {n^{(c)}}^{3/2} \cdot \mathbb{E}\,|\tilde{X}_{(\max)}|^3
\end{aligned}$$

where $\tilde{X}_{(\max)}$ is such that $\mathbb{E}\,|\tilde{X}_i|^3 \leq \mathbb{E}\,|\tilde{X}_{(\max)}|^3$ for all $i = 1, \ldots, n^{(c)}$. For now, we have the following upper bound for β

$$\beta = cb \cdot \beta_1 \leq cb \cdot {n^{(c)}}^{3/2} \cdot \mathbb{E}\,|\tilde{X}_{(\max)}|^3.$$

Taking a closer look at $\mathbb{E}\,|\tilde{X}_i|^3$ for an arbitrary $i = 1, \ldots, n^{(c)}$ we obtain

$$\begin{aligned}
\mathbb{E}\,|\tilde{X}_i|^3 &= \mathbb{E}\left| \sum_{j=1}^{n^{(c)}} c_{ij}^{-} Y_j \right|^3 \\
&\leq \mathbb{E}\left(\sum_{j=1}^{n^{(c)}} \sqrt{c_{ij}^{-}}\, \sqrt{c_{ij}^{-}}\, |Y_j| \right)^3.
\end{aligned}$$

Again by applying Hölder's inequality, this can be further bounded from above so that

$$\begin{aligned}
\mathbb{E}\,|\tilde{X}_i|^3 &\leq \mathbb{E}\left(\left(\sum_{j=1}^{n^{(c)}} |c_{ij}^{-}|^{3/4} \right)^{2/3} \cdot \left(\sum_{j=1}^{n^{(c)}} |c_{ij}^{-}|^{3/2} |Y_j|^3 \right)^{1/3} \right)^3 \\
&= \left(\sum_{j=1}^{n^{(c)}} |c_{ij}^{-}|^{3/4} \right)^2 \sum_{j=1}^{n^{(c)}} |c_{ij}^{-}|^{3/2}\, \mathbb{E}\,|Y_1|^3.
\end{aligned}$$

The last equality stems from the fact that the vector $\boldsymbol{Y}$ has identically distributed components.

In Theorem 4.3.3 we have shown that the components $|g_{ij}^-|$ of the Cholesky square-root of M are exponentially decaying as $|i-j| \to \infty$ and, since $|c_{ij}^-| = |g_{ij}^-|/\sqrt{c\,\Sigma}$, the same is true for $|c_{ij}^-|$. For $n^{(c)} \to \infty$, the above sums thus converge to some positive well-defined constants $\tilde{C}_i$, and we have $\tilde{C} = \max \tilde{C}_i$.
As $\boldsymbol{S}_b = \boldsymbol{Y}^{(1)}/\sqrt{b} + \ldots + \boldsymbol{Y}^{(cb)}/\sqrt{b}$, we have to put $Y_1 = \sum_{\nu=1}^{\mathcal{K}} X_\nu/\sqrt{b}$ with X_ν iid and therefore

$$\mathbb{E}\,|Y_1|^3 \leq \frac{1}{b^{3/2}c^3}\mathbb{E}\,|X_1|^3.$$

It has been assumed from the beginning that $\mathbb{E}\,|X_1|^3 < \infty$, so that we may now conclude the upper bound

$$\mathbb{E}\,|\tilde{X}_{(max)}|^3 \leq \tilde{C} \cdot \frac{1}{b^{3/2}}.$$

For β this implies

$$\beta \leq \tilde{C} \cdot b \cdot {n^{(c)}}^{3/2} \cdot \frac{1}{b^{3/2}} = \tilde{C} \cdot \frac{{n^{(c)}}^{3/2}}{b^{1/2}}.$$

Altogether, we get convergence to zero

$$\begin{aligned}\Delta_b &\leq \tilde{C} \cdot {n^{(c)}}^{1/4} \cdot \beta \\ &\leq \tilde{C} \cdot {n^{(c)}}^{1/4} \cdot {n^{(c)}}^{3/2}/b^{1/2} \\ &\leq \tilde{C} \cdot T^{7/4}/b^{9/4} \xrightarrow[(T\to\infty)]{} 0\end{aligned}$$

if, and only if, $b > T^{7/9}$. This is equivalent to the claim that $\delta > 7/9$ and concludes the proof. □

Note that this specific ratio of the subsample size b to the overall sample size T ensures that the vector of normalized subsample sums is asymptotically jointly normal, even if the samples overlap by a certain fixed amount of b.

4.3.2 Asymptotic extreme value theory

This section is devoted to bringing together the previous part and extreme value theory as outlined in Section 2.3. Finally, we show that the maximal component of the asymptotically Gaussian target vector $\boldsymbol{S}_b$ converges in law. To this end, recall that the k-th component of $\boldsymbol{S}_b$ is

$$S_{bk} = \frac{1}{\sqrt{b}}S_k = \frac{1}{\sqrt{b}}\sum_{\tau=t_k-b+1}^{t_k} X_\tau.$$

Furthermore, the covariance matrix is denoted by $\mathbf{C}^2 = \mathrm{Cov}(S_{b1}, \ldots, S_{bN}) = \mathrm{Cov}\,\boldsymbol{S}_b = c\,\Sigma M$ as in equation (4.3.8). Then the following theorem holds.

Theorem 4.3.6
If $b > T^{7/9}$ and $n^{(c)} = \frac{T-b}{cb}$ then

$$\mathbb{P}\left(\max_{k=1,\dots,n^{(c)}} S_{bk} \leq z_{n^{(c)}}\right) \to \mathrm{e}^{-\mathrm{e}^{-z}} \ (T \to \infty)$$

with $z_{n^{(c)}} = a_{n^{(c)}} z + d_{n^{(c)}}$ for any $z \in \mathbb{R}$ as in Proposition 2.3.7(2).

Proof
Let $n^{(c)} = n^{(c)}(T) = \frac{T-b}{cb} = \mathcal{O}(T/b)$ and recall that $c = \frac{1}{\mathcal{K}}$, $\mathcal{K} \in \mathbb{N}$.
For $\mathbf{C}^2 = (c_{ij}^2)$ it holds that

$$c_{ij}^2 = \gamma_{|i-j|} = \begin{cases} \Sigma\left(1 - \frac{|i-j|}{\mathcal{K}}\right), & 0 \leq |i-j| \leq \mathcal{K}, \\ 0, & |i-j| > \mathcal{K}. \end{cases}$$

This means that $\{S_{bk}\}_{k \geq 1}$ is a $\mathcal{K}$-dependent sequence. In particular, we have

$$\lim_{h \to \infty} \gamma_h \log h = 0.$$

Moreover it is stationary, i.e. $S_{bk} \sim F_b$ for all k, since the summation extends over iid variables X_i. Define $\boldsymbol{z}_{n^{(c)}} = (z_{n^{(c)}}, \dots, z_{n^{(c)}})$ as the $n^{(c)}$-dimensional vector with all components being $z_{n^{(c)}}$ as above. According to Theorem 4.3.4 it holds that

$$\begin{aligned} &\mathbb{P}\left(\max_{k=1,\dots,n^{(c)}} S_{bk} \leq z_{n^{(c)}}\right) \\ &= \mathbb{P}\left(S_{b1} \leq z_{n^{(c)}}, \dots, S_{bn^{(c)}} \leq z_{n^{(c)}}\right) \\ &= F_{\boldsymbol{S}_b}(\boldsymbol{z}_{n^{(c)}}) \\ &\leq \Phi_{n^{(c)}, \mathbf{C}^2}(\boldsymbol{z}_{n^{(c)}}) + \Delta_b \\ &= \mathbb{P}\left(Z_1 \leq z_{n^{(c)}}, \dots, Z_{n^{(c)}} \leq z_{n^{(c)}}\right) + \Delta_b, \end{aligned}$$

where the vector $(Z_1, \dots, Z_{n^{(c)}})$ is $n^{(c)}$-variate Gaussian with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{C}^2$.
Theorem 4.3.5 in Section 4.3.1 and the assumption that $b > T^{7/9}$ ensure that the second summand Δ_b converges to zero as $T \to \infty$. For the first summand, Theorem 2.3.9 can be applied. Since $n^{(c)}$ tends to infinity as $T \to \infty$ and due to the appropriate covariance structure (see equation 4.3.8) it holds

$$\mathbb{P}\left(Z_1 \leq z_{n^{(c)}}, \dots, Z_{n^{(c)}} \leq z_{n^{(c)}}\right) = \mathbb{P}\left(M_{n^{(c)}} \leq a_{n^{(c)}} z + d_{n^{(c)}}\right) \xrightarrow[(T \to \infty)]{} \mathrm{e}^{-\mathrm{e}^{-z}}$$

This concludes the proof. □

4.3.3 Local power

Now, we are ready to state an asymptotic distributional result for the extremes of ξ_k as an application of Theorem 4.3.6.

Theorem 4.3.7
Let $a_{n^{(c)}} = \sqrt{2 \log n^{(c)}}$ and $d_{n^{(c)}} = \Sigma(a_{n^{(c)}} - (\log\log n^{(c)} + \log 4\pi)/2a_{n^{(c)}})$. If $b > T^{7/9}$, we have

$$\mathbb{P}\Big(\max_{k=1,\dots,n^{(c)}} \xi_k \leq a_{n^{(c)}} z + d_{n^{(c)}}\Big) \to \mathrm{e}^{-\mathrm{e}^{-z}} \tag{4.3.9}$$

for all $z \in \mathbb{R}$ as $T \to \infty$.

Proof
Proposition 4.1.3 in Section 4.1 can be extended beyond the bivariate case to show that the $n^{(c)}$-variate vector $(\xi_1, \dots, \xi_{n^{(c)}})$ is asymptotically normally distributed with mean $\mathbf{0}$ and covariance matrix $c\,\Sigma M$. To be precise, we know that

$$(\xi_1, \dots, \xi_{n^{(c)}}) \xrightarrow{d} (Z_1, \dots, Z_{n^{(c)}}) \sim \mathcal{N}(\mathbf{0}, c\,\Sigma M).$$

By assumption, it holds that $\delta > 7/9$, so by setting $S_{bk} = \xi_k$, Theorem 4.3.6 is directly applicable to establish (4.3.9). □

This result can now be exploited to construct a global test of parameter constancy over the observation horizon T. Define $M_{n^{(c)}} := \max_{k=1,\dots,n^{(c)}} \xi_k$ and consider the test statistic

$$T'_{n^{(c)}} = (M_{n^{(c)}} - d_{n^{(c)}})/a_{n^{(c)}} \xrightarrow{d} \Lambda \tag{4.3.10}$$

as $n^{(c)} \to \infty$ where Λ denotes the Gumbel extremal distribution. The Extremal test rejects the null hypothesis $\mathbf{H}'_0$ against the alternative that the maximal value of the thinned parameter series exceeds a fixed value $\theta_0 \in \Theta$ to the confidence level $\alpha \in (0,1)$ if $T_{n^{(c)}} > \lambda_\alpha$ where λ_α denotes the $(1-\alpha)$-quantile of the Gumbel extremal distribution. The critical region of the test is then $K' = (\lambda_\alpha, \infty)$. We are again interested in the local power of the test, and thus consider the following local hypothesis for $\eta > 0$

$$\mathbf{H}'_A(\eta)\colon \Big\{\theta_{t_k} = \theta'(\eta) = \theta_0 + \eta/\sqrt{b/\log n^{(c)}}\Big\}. \tag{4.3.11}$$

The Extremal test's properties with respect to its statistical power are condensed in the following theorem.

Theorem 4.3.8
Let $T'_{n^{(c)}}$ be given as in (4.3.10), i.e. the centered and normalized maximum of the thinned and rescaled estimator series $\{\xi_k\}_{k=1,\dots,n^{(c)}}$. Under the conditions of Theorem 4.3.7, especially under the assumption that $b > T^{7/9}$, and under the local alternative $\mathbf{H}'_A(\eta)$ for $\eta \geq 0$, consider the extremal test to the level $\alpha \in (0,1)$ as $T'_{n^{(c)}} > \lambda_\alpha$.

(1) As $n^{(c)} \to \infty$ with $b \to \infty$, the asymptotic local power of the extremal test is

$$\mathbb{P}_{\mathbf{H}'_A(\eta)}(T'_{n^{(c)}} > \lambda_\alpha) \to \Gamma'(\eta) = 1 - \Lambda(\lambda_\alpha - \eta/\sqrt{2}) \tag{4.3.12}$$

where $\Lambda(\,.\,)$ is the distribution function of the Gumbel extremal distribution.

(2) For $\eta = 0$, we have $\Gamma'(\eta) = \alpha$, and for $\eta > 0$, we have $\Gamma'(\eta) > \alpha$. Thus, the extremal test has local power of order $\sqrt{b/\log n^{(c)}}$.

Proof
For notational convenience we set $\mathbb{P} = \mathbb{P}_{\mathbf{H}'_A}$, $n = n^{(c)}$ and $\theta_0 = 0$.

(1) Consider the rejection probability under the alternative

$$\begin{aligned}
&\mathbb{P}\left(T'_n > \lambda_\alpha\right) = \mathbb{P}\left((M_n - d_n)/a_n > \lambda_\alpha\right)\\
=&\mathbb{P}\left(M_n > a_n\lambda_\alpha + d_n\right) = \mathbb{P}\left(\xi_1 > a_n\lambda_\alpha + d_n, \ldots, \xi_n > a_n\lambda_\alpha + d_n\right)\\
=&\mathbb{P}\left(\sqrt{b}(\hat{\theta}_{t_1} - \theta'(\eta)) > a_n\lambda_\alpha + d_n - \sqrt{b}\theta'(\eta), \ldots, \sqrt{b}(\hat{\theta}_{t_n} - \theta'(\eta)) > a_n\lambda_\alpha + d_n - \sqrt{b}\theta'(\eta)\right)\\
=&\mathbb{P}\left(\tilde{\xi}_1 > a_n\lambda_\alpha + d_n - \eta\sqrt{\log n}, \ldots, \tilde{\xi}_n > a_n\lambda_\alpha + d_n - \eta\sqrt{\log n}\right)\\
=&\mathbb{P}\left(\tilde{M}_n > a_n\lambda_\alpha + d_n - \eta\sqrt{\log n}\right)\\
=&\mathbb{P}\left((\tilde{M}_n - d_n)/a_n > \lambda_\alpha - \eta\sqrt{\log n}/a_n\right)\\
=&\mathbb{P}\left((\tilde{M}_n - d_n)/a_n > \lambda_\alpha - \eta/\sqrt{2}\right) \to 1 - \Lambda(\lambda_\alpha - \eta/\sqrt{2}),
\end{aligned}$$

since under the alternative $\mathbf{H}'_A(\eta)$ the quantities $\tilde{\xi} \xrightarrow{d} \mathcal{N}(0,1)$ and $\tilde{M}_n = \max \tilde{\xi}_k$ satisfy the conditions in Theorem 4.3.7.

(2) Obviously, $\Gamma'(0) = 1 - \Lambda(\lambda_\alpha) = 1 - (1 - \alpha) = \alpha$, and for $\eta > 0$ we have $\lambda_\alpha - \eta/\sqrt{2} < \lambda_\alpha$ so that $\Gamma(\eta) > \alpha$ as tail distribution functions are decreasing. This yields asymptotic unbiasedness of the test. We have $\mathbf{H}'_A(\eta) - \mathbf{H}'_0 \asymp \sqrt{b/\log n^{(c)}}$ as $b \to \infty$ so that, together with unbiasedness, the local power has the same order, as claimed.

□

Note that for any sequence $\{z_n\}_{n\in\mathbb{N}}$ we have that $\min\{z_n\} = -\max\{-z_n\}$, and furthermore we know that for $X \sim \mathcal{N}(0,1)$ it holds that $X \stackrel{d}{=} -X$. Therefore a similar result for the minimal limiting law of the estimator series $\{\xi_k\}$ holds true, namely for all $z \in \mathbb{R}$ we have

$$\mathbb{P}\Big(\min_{k=1,\ldots,n^{(c)}} \xi_k > -a_{n^{(c)}} z - d_{n^{(c)}}\Big) \to \mathrm{e}^{-\mathrm{e}^{-z}} \tag{4.3.13}$$

with $a_{n^{(c)}}$ and $d_{n^{(c)}}$ as in Theorem 4.3.7. This can be used to test $\mathbf{H}'_0$ against the alternative $\mathbf{H}''_A$: $\{\min_t \theta_t < \theta_0\}$ with the test statistic $T''_{n^{(c)}} = -(m_{n^{(c)}} + d_{n^{(c)}})/a_{n^{(c)}}$ where $m_{n^{(c)}} = \min_{k=1,\ldots,n^{(c)}} \xi_k$. Rejecting the null hypothesis upon $T''_{n^{(c)}} < \lambda_{1-\alpha}$ yields an asymptotically unbiased test with local power of the same order as in the maximum case.
Furthermore, the sequence $\{\xi_k\}_{k=1,\ldots,n^{(c)}}$ is $\mathcal{K}$-dependent and thus satisfies condition $D(u)$ given in equation (2.3.8) with either $u = a_{n^{(c)}} z + d_{n^{(c)}}$ or $u = -a_{n^{(c)}} z - d_{n^{(c)}}$. This yields asymptotic independence of $M_{n^{(c)}}$ and $m_{n^{(c)}}$, so that we can also test against the alternative $\mathbf{H}'''_A$: $\{\min_t \theta_t < \theta_0 \text{ or } \max_t \theta_t > \theta_0\}$, i.e. that the extrema of the parameter series exceed some confidence band. The null hypothesis is then rejected if either $T'_{n^{(c)}} > \lambda_{\alpha/2}$ or $T''_{n^{(c)}} < \lambda_{1-\alpha/2}$. This testing procedure results in a symmetric simultaneous test with proper asymptotic size and local power of the same order as in the preceding cases.

In the subsequent section we will discuss a further testing procedure that takes into account the whole estimator series $\{\hat{\theta}_t\}_{t=b,\dots,T}$. Although, this test has the advantage that its local power is of order $\sqrt{T}$, the alternative to the null hypothesis of parameter constancy is that of a finite number of change points. Furthermore, the location of the change points has to be known a priori.

4.4 Test based on forecast regression

The following concepts were originally developed for the task of comparing the quality of forecasts from competing models for a given time series. We present them here as a motivating introduction for our third test. West (1996) introduced this approach which was brought forth for example by West and McCracken (1998), White (2000), Clark and McCracken (2001) and Corradi and Swanson (2005) thereupon. All these publications consider series that exhibit a stationary structure. Yaskov (2010b) extends the above method to data that changes its structure finitely many times, and we adopt his findings to develop a homogeneity test with optimal local power.
Consider the original set-up due to West and McCracken (1998). Under the assumption of a parametric model, the parameter value up to time t can either be estimated *recursively* (based on the first to t-th observation), or on a *fixed* (of length $R < T$) or a *rolling* subsample. The latter exactly corresponds to our moving window estimators based on the b observations preceding the actual point in time t where the estimation is required. Therefore, we can think of $\hat{\theta}_t$ in the following as the local maximum likelihood estimator of the copula parameter in the SCOMDY model as given in (4.1.1). In particular, $\hat{\theta}_t$ is in compliance with the more general requirement that under the null hypothesis of $\theta_t = \theta_0$ for all $t \leq T$ we have the factorization

$$\hat{\theta}_t - \theta_0 = Q_t^{-1} H_t \tag{4.4.1}$$

such that $Q_t \to Q$ almost surely as $T \to \infty$ and $H_t = \sum_{s=t-b+1}^{t} h_s$ with $\mathbb{E}\, h_s = 0$. Furthermore, consider the real-valued *prediction error function* $v_{t+1} = v_{t+1}(\theta)$ and its empirical counterpart $\hat{v}_{t+1} = v_{t+1}(\hat{\theta}_t)$, as well as the possibly vector-valued *objective function* $g_{t+1} = g_{t+1}(\theta)$ and its empirical version $\hat{g}_{t+1} = g_{t+1}(\hat{\theta}_t)$. We want to stress that v_{t+1} and g_{t+1} themselves are functions of the given data. Therefore, $\hat{v}_{t+1}$ and $\hat{g}_{t+1}$ depend doubly on the data: On the one hand, a parameter estimate $\hat{\theta}_t$ based on (part of) the observations is plugged in; on the other hand, the functions themselves are defined to depend on the data. The pair of hypotheses we want to test is

$$\mathrm{H}_0 \colon \{\mathbb{E}\, v_{t+1} g_{t+1} = 0 \text{ for all } t\} \quad \text{vs.} \quad \mathrm{H}_A \colon \{\mathbb{E}\, v_{t+1} g_{t+1} \neq 0 \text{ for some } t\}. \tag{4.4.2}$$

Depending on the specific question, v_{t+1} and g_{t+1} have to be defined appropriately by the statistician. If we want to test for example a time series model of the form $y_t = s(x_t; \theta)$ with respect to the mean prediction error, we could choose $g_{t+1} \equiv 1$ and $v_{t+1} = y_t - s(x_t; \theta_t)$, and then test for $\mathbb{E}\, v_{t+1} = 0$. For further examples see West and McCracken (1998, p. 820).

Now, we specify the *objective* or *forecast regression* equation

$$\hat{v}_{t+1} = \hat{g}_{t+1}^{\top}\beta + \epsilon_{t+1} \tag{4.4.3}$$

with $\epsilon_t \sim \mathrm{WN}(0,1)$. By regressing the (empirical) prediction error $\hat{v}_{t+1}$ on the (empirical) objective function $\hat{g}_{t+1}$, we can equivalently test the null hypothesis $\tilde{\mathrm{H}}_0\colon \{\beta = 0\}$. Assuming the relationship in (4.4.3), the falsity of H_0, that we actually want to detect, implies necessarily that $\tilde{\mathrm{H}}_0$ is rejected. This auxiliary construction thus provides certainty with respect to the violation of the null hypothesis H_0. For the test consider the least squares estimator for β

$$\hat{\beta} = \Big(\sum_{t=b}^{T-1} \hat{g}_{t+1}\hat{g}_{t+1}^{\top}\Big)^{-1} \sum_{t=b}^{T-1} \hat{g}_{t+1}\hat{v}_{t+1}. \tag{4.4.4}$$

For the limiting distribution of $\hat{\beta}$ as $T \to \infty$ consider further notations

$$f_{t+1} = g_{t+1}v_{t+1}, \quad f'_{t+1} = \frac{\partial}{\partial\theta} f_{t+1}, \quad Q_{f'} = \mathbb{E}\, f'_{t+1}, \quad \begin{pmatrix} Q_{ff} & Q_{fh} \\ Q_{fh}^{\top} & Q_{hh} \end{pmatrix} = \lim_{T\to\infty} \mathrm{Cov}\,(f_{t+1}, h_t)\,.$$

Then we have the following theorem due to West and McCracken (1998).

Theorem 4.4.1
Let $\hat{\beta}$ be given as in (4.4.4) and assume that $(T-b)/b \to \varpi \geq 1$ as $T \to \infty$. Then we have

$$\sqrt{T-b}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}) \tag{4.4.5}$$

where $\mathbf{V} = (\mathbb{E}\, g_t g_t^{\top})^{-1}\Omega(\varpi)(\mathbb{E}\, g_t g_t^{\top})$ and $\Omega(\varpi) = Q_{ff} + (1 - 1/2\varpi)(Q_{f'}Q^{-1}{Q_{fh}}^{\top} + Q_{fh}{Q^{-1}}^{\top}Q_{f'}^{\top}) + (1 - 1/3\varpi)Q_{f'}Q^{-1}Q_{hh}{Q^{-1}}^{\top}Q_{f'}^{\top}$.

Proof
In Lemmas 4.1 – 4.3 and Theorem 4.1 in West and McCracken (1998) the results for a general estimation scheme (i.e. either recursive, fixed or rolling) are given. A crucial step in the proof is the following Taylor approximation

$$\sum \hat{g}_{t+1}\hat{v}_{t+1} = \sum g_{t+1}v_{t+1} + Q_{f'}Q^{-1}\sum H_t + o_P(1). \tag{4.4.6}$$

Subsequently an application of the law of large numbers, the central limit theorem and Slutsky's theorem yield the general result. Upon replacing $\lambda_{fh} = 1 - 1/2\varpi$ and $\lambda_{hh} = 1 - 1/3\varpi$ we arrive at our claim. □

The second and third summands in the expression $\Omega(\varpi)$ are terms that arise additionally in the variance of a least squares estimator of regression coefficients. They are due to the fact that we work with the empirical quantities $\hat{g}_{t+1}$ and $\hat{v}_{t+1}$ instead of the theoretical g_{t+1} and v_{t+1}, and also stem from the Taylor approximation in (4.4.6).
The notation introduced hints at possible generalizations of the forecast regression. We proceed along the lines of Yaskov (2010b) who introduces modifications of the initial model in two respects: On the one hand, we directly test the pair of hypotheses

$$\mathrm{H}'_0\colon \{\mathbb{E}\, f_t = 0\} \quad \text{vs.} \quad \mathrm{H}'_A\colon \{\mathbb{E}\, f_t \neq 0\} \tag{4.4.7}$$

with a suitable function f_t. As the objective regression in (4.4.3) was originally introduced to discriminate two competing forecast (time series) models, f_t can be understood as the difference or distance of the prediction losses $L_1(\boldsymbol{x}_t;\theta_1)$ and $L_2(\boldsymbol{x}_t;\theta_2)$ of two models at choice. The losses here also depend doubly on the data $\boldsymbol{x}_1,\dots,\boldsymbol{x}_T$. We have $f_t(\theta) = L_1(\boldsymbol{x}_t;\theta_1) - L_2(\boldsymbol{x}_t;\theta_2)$, $\theta = (\theta_1,\theta_2)$. For the prediction loss we can for example consider the empirical residuals of a fitted model. When comparing a reference model to more than one alternative, f_t is set as the maximum over the relative losses. The limiting law is then the maximum of Gaussian variables and can be assessed via bootstrap procedures.
On the other hand, Yaskov allows the data that forms the basis for the estimation of θ to change structure finitely many times throughout the observation horizon T, i.e. the parameter series $\{\theta_t\}$ exhibits l different regimes.
Taking this twofold modification as a starting point, we consider the following test statistic

$$S'_{Tb} = \frac{1}{\sqrt{T}} \sum_{t=b}^{T-1} (f_{t+1}(\hat{\theta}_t) - \mathbb{E}\, f_{t+1}). \tag{4.4.8}$$

According to Theorem 4.4.1 we have asymptotic normality of this quantity, namely $S'_{Tb} \xrightarrow{d} \mathcal{N}(0,\Omega(\varpi))$ under the condition that $(T-b)/b \to \varpi \geq 1$ as $T \to \infty$. As West and McCracken (1998) note, the assessment of ϖ in finite samples might give rise to problems such as considerable size distortions in the respective tests.
The main result in Yaskov (2010b) can be used to circumvent these difficulties in the testing of H'_0. Besides establishing the bias of S'_{Tb}, that possibly accounts for finite sample distortions, the author expresses this mean and the asymptotic variance in terms of T and b and their relative magnitude. The conditions we refer to as well as the specific expressions for μ_{Tb} and σ^2_{Tb} in the following theorem are postponed to Appendix C. Note that the original reference is in Russian, and the following statements rely on our own translation and personal communication with the author.

Theorem 4.4.2
Under the conditions (Y1) – (Y7) in Appendix C, and if $T = \mathcal{O}(b^\delta)$, $\delta < 3$, it holds that for $T, b \to \infty$

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\frac{S'_{Tb} - \mu_{Tb}}{\sigma_{Tb}} \leq x \right) - \Phi(x) \right| = \mathcal{o}(1) \tag{4.4.9}$$

with μ_{Tb} and σ^2_{Tb} given in (C.0.1) and (C.0.2). Φ is the distribution function of the standard normal law.

Proof
The result corresponds to Theorem 1 in Yaskov (2010b). □

The expressions for μ_{Tb} and σ^2_{Tb} as given in (C.0.1) and (C.0.2) allow for an estimation based on pseudo-observations of the involved quantities. As these pseudo-samples are established with consistent estimates for θ_t, the estimation of μ_{Tb} and σ^2_{Tb} is consistent as well. Yaskov (2010b) argues in favor of autocorrelation and heteroscedasticity consistent (so-called HAC) estimators $\hat{\mu}_{Tb}$ and $\hat{\sigma}^2_{Tb}$. Such type of estimators are often kernel-based

and smoothed versions of the usual sample estimates and account for possible violations of the iid assumption in the underlying data. An introduction to HAC estimation can for example be found in Andrews (1991).
In order to test the null hypothesis $\mathrm{H}_0'\colon \{\mathbb{E}\, f_t = 0\}$ against the alternative $\mathrm{H}_A'\colon \{\mathbb{E}\, f_t \neq 0\}$ consider $\hat{S}_{Tb} = \sum_{t=b}^{T-1} f_{t+1}(\hat{\theta}_t)/\sqrt{T}$. H_0' is rejected to the (asymptotical) level $\alpha \in (0,1)$ if

$$\left| \frac{\hat{S}_{Tb} - \hat{\mu}_{Tb}}{\hat{\sigma}_{Tb}} \right| > q_{\alpha/2} \tag{4.4.10}$$

where $q_{\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. Here the symmetric two-sided test is presented, but entirely similar results hold in the one-sided case (cf. Section 4.2.1).
It is furthermore argued that the following modification of S'_{Tb} is appropriate to account for regime changes in θ:

$$\bar{S}_{Tb} = S'_{Tb} - \frac{1}{\sqrt{T}} \sum_{t \in U_{Tb}} (f_{t+1}(\hat{\theta}_t) - \mathbb{E}\, f_{t+1}) \tag{4.4.11}$$

where $U_{Tb} \subseteq [b, T]$, $|U_{Tb}| = o(T)$ is the union of all change point environments. If there are finitely many change points, i.e.

$$\bar{S}_{Tb} = S'_{Tb} + o_P(1),$$

the test statistic is robust against changes in the parameters, and its asymptotic normality holds even under weaker prerequisites.
Our primary concern is to construct a test that enables us to detect possible change points in a given parameter series, rather than give a test statistic that is insensitive to disturbances in the parameter structure. However, Yaskov's results are very general and lend themselves to construct a test that meets our needs nevertheless. We will see that our situation is even simpler, and the distributional limit can be established directly. We confine ourselves to the alternative of one change point τ throughout the observation time $t = 1, \ldots, T$. However, as long as the number of change points is finite, the ideas confer. The pair of hypotheses under consideration is then

$$\tilde{\mathbf{H}}_0\colon \{\theta_t \equiv \theta_0\ \forall t\} \quad \text{vs.} \quad \tilde{\mathbf{H}}_A(\eta)\colon \{\theta_t = \theta_0 \text{ for } t \leq \tau, \theta_t = \theta_T(\eta) = \theta_0 + \eta/\sqrt{T} \text{ for } t > \tau\} \tag{4.4.12}$$

where $\tilde{\mathbf{H}}_A(\eta) - \tilde{\mathbf{H}}_0 \asymp \sqrt{T}$ for $\eta \neq 0$, i.e. we want to test against a local alternative of order $\sqrt{T}$ this time. The objective function to test the above pair of hypotheses is

$$f(\theta_t) = \begin{cases} (\mathbb{1}_{\{t \leq \tau\}} - \mathbb{1}_{\{\tau < t\}})(\theta_t - \theta_0), & t \geq b, \\ 0, & t < b. \end{cases} \tag{4.4.13}$$

By plugging in the respective estimates for θ_t we obtain the test statistic

$$S_{Tb} = \frac{1}{\sqrt{T}} \sum_{t=b}^{\tau} (\hat{\theta}_t - \hat{\theta}_{t+\tau}) - 2(T/2 - \tau)\theta_0. \tag{4.4.14}$$

We directly show the asymptotic normality of $\tilde{S}_{Tb}$ in the following.

Theorem 4.4.3
If $b = \mathcal{O}(T)$ as $T \to \infty$ and under the null hypothesis $\tilde{\mathbf{H}}_0 \colon \{\theta_t \equiv \theta_0 \; \forall t\}$ we have

$$S_{Tb} \xrightarrow{d} \mathcal{N}(0, \Sigma), \tag{4.4.15}$$

where $\Sigma = \Sigma(\theta_0)$ is the limiting variance in canonical maximum likelihood estimation of copula parameters given in (2.2.25).

Proof
For notational convenience we assume that the change point divides the sample into two equal halves, i.e. $\tau = T/2$. The test statistic then reads as $S_{Tb} = \sum_{t=b}^{T/2}(\hat{\theta}_t - \hat{\theta}_{t+T/2})/\sqrt{T}$. Define the coefficients $\alpha_t = 1$ for $t \leq T/2$ and $\alpha_t = -1$ for $t > T/2$, then we have

$$S_{Tb} = \frac{1}{\sqrt{T}} \sum_{t=b}^{T} \alpha_t (\hat{\theta}_t - \theta_0).$$

Analogously to the proof of Theorem 2.2.3 we obtain a first order approximation (denoted by $\doteq$) with $(\hat{\theta}_t - \theta_0) \doteq A_b / B_b$ where

$$A_b = \frac{1}{b} \sum_{\nu=t-b+1}^{t} l_\theta(\tilde{\boldsymbol{U}}_\nu; \theta_0) \;\text{ and }\; B_b = -\frac{1}{b} \sum_{\nu=t-b+1}^{t} l_{\theta\theta}(\tilde{\boldsymbol{U}}_\nu; \theta_0)$$

and $\tilde{\boldsymbol{U}}_\nu$ are the vectors of the rescaled rank-transformed theoretical residuals, i.e. with components $\tilde{U}_{j\nu} = \tilde{F}_{jT}(\varepsilon_{j\nu})$ for all $j = 1, \ldots, d$. For $b \to \infty$ we have the convergence of B_b to the Fisher information at the true parameter value θ_0 denoted by $B_b \to I(\theta_0)$. The test statistic S_{Tb} can thus be approximated by

$$\begin{aligned} S_{Tb} &\doteq \frac{1}{\sqrt{T}} \sum_{t=b}^{T} \alpha_t \sum_{\nu=t-b+1}^{t} l_\theta(\tilde{\boldsymbol{U}}_\nu; \theta_0)/(bI(\theta_0)) \\ &= \frac{1}{\sqrt{T}} \sum_{\nu=1}^{T} l_\theta(\tilde{\boldsymbol{U}}_\nu; \theta_0) \sum_{t=\nu}^{\nu+b-1} \alpha_t/(bI(\theta_0)) \\ &= \frac{1}{\sqrt{T}} \sum_{\nu=1}^{T} \tilde{\beta}_\nu \, l_\theta(\tilde{\boldsymbol{U}}_\nu; \theta_0), \end{aligned}$$

with weights

$$\tilde{\beta}_\nu = \frac{1}{I(\theta_0)} \begin{cases} 1, & \nu \leq T/2 - b, \\ (T - b + 1 - 2\nu)/b, & T/2 - b < \nu \leq T/2, \\ -1, & T/2 < \nu. \end{cases}$$

The number of the terms in the second case is $\sqrt{b}$. Because of $b = \mathcal{O}(T)$, we have $\sqrt{b}/\sqrt{T} \to 0$ and therefore this portion of the summation is asymptotically negligible, i.e. as $T \to \infty$ we have

$$S_{Tb} = \frac{1}{\sqrt{T}} \sum_{\nu=1}^{T} \tilde{\beta}_\nu \, l_\theta(\tilde{\boldsymbol{U}}_\nu; \theta_0) \sim \frac{1}{\sqrt{T}} \sum_{\nu=1}^{T} \beta_\nu \, l_\theta(\tilde{\boldsymbol{U}}_\nu; \theta_0)$$

with $\beta_\nu = 1/I(\theta_0)$ for $\nu \leq T/2$ and $\beta_\nu = -1/I(\theta_0)$ for $\nu > T/2$. Furthermore, $\sqrt{T}A_T \xrightarrow{d} \mathcal{N}(0,\sigma^2)$ as in equation (2.2.25). Analogously to the proof of Theorem 2.2.3 we conclude that $S_{Tb} \xrightarrow{d} \mathcal{N}\big(0,\sigma^2/I^2(\theta_0)\big) = \mathcal{N}(0,\Sigma)$. □

Similar considerations as for the local power of the partial tests (see Theorem 4.2.3) yield that the change point test has a local power of order $\sqrt{T}$. Define the empirical test statistic $\tilde{S}_{Tb} = \sum_{t=b\vee(\tau-T/2)}^{\tau\wedge T/2}(\hat{\theta}_t - \hat{\theta}_{t+\tau})/\sqrt{T}$ and reject $\tilde{\mathbf{H}}_0$ as in (4.4.12) to the asymptotical level $\alpha \in (0,1)$ if

$$|\hat{\Sigma}^{-1/2}\tilde{S}_{Tb}| > q_{\alpha/2}$$

with $\hat{\Sigma}$ the variance estimate as in (2.2.27).

Theorem 4.4.4
For an arbitrary $\alpha \in (0,1)$ and $\eta \neq 0$ we have that

$$\lim_{T\to\infty} \mathbb{P}_{\tilde{\mathbf{H}}_A(\eta)}\left(|S_{Tb}/\sqrt{\Sigma}| > q_{\alpha/2}\right) = \tilde{\Gamma}(\eta) > \alpha. \tag{4.4.16}$$

The change point test is thus asymptotically unbiased and has local power of order $\sqrt{T}$.

Proof
Let $\mathbb{P} = \mathbb{P}_{\mathrm{H}_A(\eta)}$ be the probability under the alternative hypothesis $\tilde{\mathbf{H}}_A(\eta)$. For clarity of notation we assume that $\theta_0 = 0$ and $\tau = T/2$. By denoting the test statistic under the alternative by $S_{Tb}^A(\eta)$, we have

$$\begin{aligned}
&\mathbb{P}\left(|S_{Tb}/\sqrt{\Sigma}| > q_{\alpha/2}\right)\\
=&\mathbb{P}\left(\left|\sum_{t=b}^{T/2}(\hat{\theta}_t - \hat{\theta}_{t+T/2})/\sqrt{T} \mp \sqrt{T}\theta_T(\eta)/2\right| > q_{\alpha/2}\sqrt{\Sigma}\right)\\
=&\mathbb{P}\left(\left|\sum_{t=b}^{T/2}(\hat{\theta}_t - (\hat{\theta}_{t+T/2} + \eta/\sqrt{T}))/\sqrt{T} + \frac{\eta}{2}\right| > q_{\alpha/2}\sqrt{\Sigma}\right)\\
=&\mathbb{P}\left(\left|S_{Tb}^A(\eta) + \frac{\eta}{2}\right|/\sqrt{\Sigma} > q_{\alpha/2}\right) \xrightarrow[(T\to\infty)]{} \tilde{\Gamma}(\eta) > \alpha.
\end{aligned}$$

The last assertion is due to the fact that $S_{Tb}^A(\eta)/\sqrt{\Sigma} \xrightarrow{d} \mathcal{N}(0,1)$ under $\tilde{\mathbf{H}}_A(\eta)$ as $T \to \infty$. Furthermore, for $Z \sim \mathcal{N}(0,1)$ the probability $\mathbb{P}(|Z+\eta| > q)$ is increasing if $\eta \neq 0$ is increasing in modulus. This yields asymptotic unbiasedness. Moreover, we have $\tilde{\mathbf{H}}_A(\eta) - \tilde{\mathbf{H}}_0 \asymp \sqrt{T}$ which yields local power of order $\sqrt{T}$ of the change point test. □

Note, however, that the specific form of the test statistic heavily relies on the a priori knowledge of the change points. If the actual locations deviate from the supposed locations, the test is no longer capable of detecting the change points and will not lead to the rejection of the homogeneity hypothesis in this case.
We will conclude with a discussion of the three presented tests and their respective local power in detecting deviations from a time-homogeneous parameter structure.

4.5 Discussion

Conclusively, we discuss the three types of tests from the previous sections with respect to their ability to detect whether and of which type there are deviations from the null hypothesis of parameter homogeneity.

The Extremal test presented in Section 4.3 has local power of order $\sqrt{b/n^{(c)}}$ where $n^{(c)} = \lfloor \frac{T-b}{cb} \rfloor$ and $b = \mathcal{O}(T^{7/9})$ with respect to the overall observation horizon T. This local power is the least among the considered tests. Furthermore, the applicability of the procedure is limited in finite samples due to the restriction $b = \mathcal{O}(T^{7/9})$. This bandwidth condition is practically only fulfilled in samples over a long observation horizon, say for T in the order of decades if we consider daily price data. Nevertheless the test is suitable to test against the most general alternative in our set-up, namely whether there is an exceedance from a long term constant parameter value at all, and if the deviation is in a sense extreme, i.e. the maximal (or minimal) parameter value peaks high above (or below) an assumed average level. Motivated by the empirical observations in real-world data where the estimated parameter series sometimes exhibit high peaks (mostly above the average parameter level), this test was developed as an omnibus test that allows to judge if peaks in the parameter value lie within a predefined confidence region. In the course of the data description and empirical analyses in Chapter 5, we will also present what the rejection of the extremal null hypothesis $\mathbf{H}'_A$ could be used for after that.

In Section 4.2 we presented a Binomial test with local power of order $\sqrt{b}$ where $b = \mathcal{O}(T)$ as $T \to \infty$. The facts that the test does not achieve the usual local power of order $\sqrt{T}$, as might be expected when a sample of size T is available, and that the local estimates are restricted to be independent, may seem to be a disadvantage at first sight. At the same time, take into account that using moving window estimation we implicitly presume a locally homogeneous structure in the copula parameter anyway. Under this perspective, a rejection of one (or more) of the partial hypotheses hints at one (or more) regime changes in the parameter series. We again refer to Chapter 5 for an empirical application of the Binomial test on real-world data.

Finally, some recent results regarding forecast regression for discriminating competing forecast models led to the construction of a change point test in Section 4.4 that actually has local power of order $\sqrt{T}$. However, a priori knowledge of the change point locations is necessary as the alternative is directly incorporated into the test statistic.

Summarizing, each of the presented tests has certain pros and cons, and on its own is only able to detect the special alternatives that were outlined in the respective sections. We can though combine the testing of the respective null hypotheses in the same order as they were presented in this discussion: First, the Extremal test lends itself to decide whether there is an extremal deviation of an average copula parameter, second, the Binomial test provides evidence for moderate deviation and/or the location of possible change points, finally, a suspected regime change can be verified through the forecast regression based test.

In the subsequent chapter, we will demonstrate the first two steps applied on special financial data, and give an interpretation of the resulting findings.

5 Empirical Results

In this chapter we will demonstrate some practical applications of the modeling and testing theory presented so far.
Section 5.1 is dedicated to the description and evaluation of a simulation study. Under the null hypothesis of parameter constancy we simulate copula data of different dimensions and with different parameter values. Then the Binomial test is run with various types of base data in order to investigate the test's empirical size. Within this scope, the proper determination of the involved estimation variance proves vital for a correct testing procedure. The main outcome is that bootstrapping the variance cannot be avoided in our finite-sample set-up notwithstanding the theoretical results.
With this insight we turn to the analysis of certain commodity price data in Section 5.2. After describing the data, we set up a SCOMDY model appropriate for this special kind of financial data and run the Binomial and the Extremal test on the respective estimators. A discussion of the economic interpretation of the test decisions concludes this section.

5.1 Empirical size of the Binomial test

As we will see in the subsequent section where real-world data is investigated, a goodness-of-fit procedure is performed and the family of Clayton copulas (see Appendix A.5) emerges as the dependence model that fits the empirical data best. For this reason and for computational tractability, we restrict ourselves to the case of residuals in the SCOMDY model having a multivariate distribution with a Clayton copula only allowing for positive dependence. Recall that for $\theta \in (0, \infty)$ and $d \geq 2$, a Clayton copula has the analytical form

$$C_{\mathrm{C}}(u_1, \ldots, u_d; \theta) = \left(\sum_{j=1}^{d} u_j^{-\theta} - (d-1) \right)^{-1/\theta} \tag{5.1.1}$$

for all $\boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d$. Restricting the copula parameter to be positive yields a dependence that is positive in the sense of comonotonicity. However, it is possible to extend the parameter range to $\theta \geq -1/(d-1)$ and still obtain a proper copula for arbitrary dimensions $d \geq 2$ (see Example 2.3 in McNeil and Nešlehová, 2009).
In a finite sample set-up and with a thinning constant $c = 1/4$, the Extremal test based on data sampled under parameter constancy hardly ever produces rejections of the null hypothesis. This is mainly due to the fact that the extreme value results exhibit a rather slow speed of convergence, and thus in our finite sample set-up we are not sufficiently close to the respective limits, yet. Therefore we will not repost those results in detail

here, but restrict ourselves to the Binomial test in the following. As the main focus in this section is on assessing the appropriate variance estimates, these considerations will be sufficient.

5.1.1 Simulation

Our Monte Carlo simulation consists of generating $M = 1000$ random samples of $T = 500$ data points, i.e. T is the observation horizon in a time series setting. The bandwidth for the local estimators of the copula parameter θ is set to $b = T/10 = 50$, which results in $n = 9$ independent moving window estimators $\hat{\theta}_k$, $k = 1, \dots, n$, representing the parameter value at the point in time t_k. As significance level we choose the usual $\alpha^* = 5\%$ where α^* is the nominal level of the multiple Binomial test as given in (4.2.11). The model dimension is $d \in \{2, 5, 10\}$ and the Clayton copula parameter is $\theta_0 \in \{0.1, 0.5, 1.5, 6\}$. Let us summarize the values of the input variables in tabular form.

number of Monte Carlo runs	$M = 1000$
dimension	$d \in \{2, 5, 10\}$
copula parameter	$\theta_0 \in \{0.1, 0.5, 1.5, 6\}$
observation horizon	$T = 500$
bandwidth for moving windows	$b = 50$
number of independent estimators	$n = 9$
significance level	$\alpha^* = 5\%$

Table 5.1: Input parameters for Monte Carlo simulation study

We consider three different data set-ups in order to analyze the respective variance estimates:

(I) First, we use a sample of *pure copula data* of length T. We simulate $\boldsymbol{u}_1, \dots, \boldsymbol{u}_T$ independent and identically distributed according to $C_{\mathrm{C};\theta_0}$ so that for all $t = 1, \dots, T$ we have $\boldsymbol{u}_t = (u_{1t}, \dots, u_{dt}) \in [0,1]^d$. The moving window estimator is determined through copula maximum likelihood estimation on a subsample of length b so that for t_k, $k = 1, \dots, n$ it is

$$\hat{\theta}_k = \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{t=t_k-b+1}^{t_k} \log c_{\mathrm{C}}(u_{1t}, \dots, u_{dt}; \theta). \tag{5.1.2}$$

The variance in the limiting distribution for $\sqrt{b}(\hat{\theta}_k - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{v})$ corresponds to the inverse Fisher information at the true parameter value θ_0, i.e.

$$\mathrm{v} = I(\theta_0)^{-1}. \tag{5.1.3}$$

As an appropriate variance estimate we have chosen the inverse of the observed Fisher information as in (2.2.15) at the local parameter estimate, i.e.

$$\hat{\mathrm{v}} = J(\hat{\theta}_k)^{-1}. \tag{5.1.4}$$

(II) With $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T$ simulated in (I) and arbitrary univariate distribution functions $F_1, \ldots, F_d$ at hand, we use the univariate inversion method (see Algorithm 2.1.11) in every component of the copula vectors to obtain

$$\boldsymbol{x}_t = (x_{1t}, \ldots, x_{dt}) = \left(F_1^-(u_{1t}), \ldots, F_d^-(u_{dt})\right), \tag{5.1.5}$$

and thus a sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$ of length T of independent and identically distributed realizations of the *copula-based multivariate distribution* $H = C_{\mathrm{C}}(F_1, \ldots, F_d; \theta)$. For the simulation we chose the standard normal distribution in every component, i.e. $F_1 = \ldots = F_d = \Phi$, and for the local estimation of the copula parameter we applied canonical maximum likelihood estimators on subsamples of length b. Formally, for every t_k, $k = 1, \ldots, n$ we have

$$\hat{\theta}_k = \operatorname*{argmax}_{\theta \in \Theta} \sum_{t=t_k-b+1}^{t_k} \log c_{\mathrm{C}}(\tilde{u}_{1t}, \ldots, \tilde{u}_{dt}; \theta) \tag{5.1.6}$$

where $\tilde{u}_{jt} = \tilde{F}_{jT}(x_{jt})$ are the rescaled componentwise ranks of the observations obtained by a transformation using the version $\tilde{F}_{jT}$ of the empirical distribution function as given in (2.2.21). In this case, the theoretical limiting variance in $\sqrt{b}(\hat{\theta}_k - \theta_0) \to \mathcal{N}(0, \mathrm{v})$ corresponds to the variance due to Genest et al. (1995) at the true parameter value θ_0, i.e.

$$\mathrm{v} = \Sigma(\theta_0) \tag{5.1.7}$$

where Σ is given in (2.2.25). The estimate is chosen as in (2.2.27) at the local parameter estimate, i.e.

$$\hat{\mathrm{v}} = \hat{\Sigma}(\hat{\theta}_k). \tag{5.1.8}$$

(III) Through (5.1.5) we obtain independent and identically distributed realizations of the multivariate distribution $H = C_{\mathrm{C}}(F_1, \ldots, F_d; \theta)$ that serve as residuals when simulating a path of a SCOMDY model. We choose the most parsimonious parametrization, namely that of GARCH(1,1) models in every component with the parameter set being identical. We set $\alpha_{j0} = \alpha_0 = 10^{-6}$, $\alpha_{j1} = \alpha_1 = 0.1$ and $\beta_{j1} = \beta_1 = 0.8$ for all $j = 1, \ldots, d$. These numerical values represent a typical parametrization observed in real-world financial data; note that the values lie near the boundary of the parameter set that ensures covariance stationarity of the

resulting time series (cf. Figure 3.3). Analogously to (3.1.3) and (3.1.4), we obtain realizations $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{dt})$ of the specified SCOMDY model for all $t = 1, \ldots, T$:

$$\begin{aligned} y_{jt} &= \sigma_{jt} x_{jt}, \\ \sigma_{jt}^2 &= \alpha_0 + \alpha_1 y_{j,t-1}^2 + \beta_1 \sigma_{j,t-1}^2, \end{aligned}$$

with starting values $y_{j0} = 0$ and $\sigma_{j0} = \alpha_0/(1-\alpha_1-\beta_1)$ and simulated iid residuals $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$.

With quasi maximum likelihood estimates $\hat{\gamma}_j = (\hat{\alpha}_{j0}, \hat{\alpha}_{j1}, \hat{\beta}_{j1})$ in every univariate GARCH(1,1) model and the empirical conditional variance $\tilde{\sigma}_{jt}^2 = \tilde{w}_t(\hat{\gamma}_j)$ (see Section 3.1.3), the empirical residuals $\tilde{\boldsymbol{\varepsilon}}_1, \ldots, \tilde{\boldsymbol{\varepsilon}}_T$ are given by

$$\tilde{\varepsilon}_{jt} = y_{jt}/\sqrt{\tilde{\sigma}^2},\ j = 1, \ldots, d, \tag{5.1.9}$$

in analogy to (3.1.21). The copula parameter is again estimated locally for all t_k, $k = 1, \ldots, n$ by canonical maximum likelihood estimation on subsamples, i.e.

$$\hat{\theta}_k = \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{t=t_k-b+1}^{t_k} \log c_{\mathrm{C}}(\tilde{\tilde{u}}_{1t}, \ldots, \tilde{\tilde{u}}_{dt}; \theta) \tag{5.1.10}$$

where $\tilde{\tilde{u}}_{jt} = \tilde{F}_{jT}(\tilde{\varepsilon}_{jt})$ are the rescaled componentwise ranks of the empirical residuals (cf. (II)). Under the conditions of Theorem 3.2.5 we have for $\sqrt{b}(\hat{\theta}_k - \theta_0) \to \mathcal{N}(0, \mathrm{v})$ that $\mathrm{v} = \Sigma(\theta_0)$ and $\hat{\mathrm{v}} = \hat{\Sigma}(\hat{\theta}_k)$ are the respective asymptotic variance and an allegedly suitable variance estimate.

In the subsequent section we adapt the test statistics for the respective data set-ups. Within this scope we especially provide a detailed discussion on the variance estimation in (III) and the bootstrapping procedure that is required to attain the empirical size of the Binomial test.

5.1.2 Test and bootstrap

Let us recall the distributional limit that forms the basis of all tests presented in Chapter 4. For $b = o(T)$ as $T \to \infty$ we have

$$\sqrt{b}(\hat{\theta}_k - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{v}) \tag{5.1.11}$$

with limiting variance v according to the underlying data situation (see the previous Section 5.1.1). Therefore the theoretical form of the test statistic of the partial test (see Section 4.2.1) is

$$T_k = \frac{\sqrt{b}(\hat{\theta}_k - \theta_0)}{\sqrt{\mathrm{v}}} \xrightarrow{d} \mathcal{N}(0,1). \tag{5.1.12}$$

It is common to plug in a suitable variance estimate $\hat{\mathrm{v}}$ for the asymptotic variance v, and in case of a consistent estimate the distributional limit still holds true. Unlike the

situation when data according to a prespecified copula with a known parameter value $\theta_0 \in \Theta$ is generated, we usually have no knowledge about the true parameter value. However, for testing the null hypothesis of parameter constancy this is not necessarily required. Especially, if each of the local parameter values θ_t is assumed to be constant and equal to θ_0, a global parameter will attain the same value. Such a global value can be estimated analogously, but this time on the whole sample of size T, i.e. we define the global parameter estimate as

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{t=1}^{T} \log c_{\mathrm{C}}(\boldsymbol{z}_t; \theta) \tag{5.1.13}$$

where $\boldsymbol{z}_t$ is either pure copula data $\boldsymbol{u}_t$, rank-transformed realizations $\tilde{\boldsymbol{u}}_t$ of a multivariate copula-based distribution, or the rank-transformed empirical residuals $\tilde{\hat{\boldsymbol{u}}}_t$ in a SCOMDY model. Since $b = o(T)$, i.e. that both b and T tend to infinity, but T faster than b, we have the empirical test statistic

$$\hat{T}_k = \frac{\sqrt{b}(\hat{\theta}_k - \hat{\theta})}{\sqrt{\hat{\mathrm{v}}}} \xrightarrow{d} \mathcal{N}(0,1), \tag{5.1.14}$$

that is asymptotically standard normally distributed. Given a partial significance level $\alpha \in (0,1)$, we have for $q_{\alpha/2}$ being the $(1-\alpha/2)$-quantile of the standard normal distribution

$$\hat{S}_n = \sum_{k=1}^{n} \mathbb{1}_{\{|\hat{T}_k| > q_{\alpha/2}\}} \tag{5.1.15}$$

the empirical version of the multiple test statistic S_n, and $\hat{S}_n \xrightarrow{d} \mathrm{Bin}(n, \alpha)$ still holds as $T \to \infty$. With c_{α^*} being the $(1-\alpha^*)$-quantile of the $\mathrm{Bin}(n, \alpha)$-distribution, we have $\lim_{T\to\infty} \mathbb{P}(\hat{S}_n > c_{\alpha^*}) \leq \alpha^*$. In our set-up, we have $n = 9$, $\alpha = \alpha^* = 5\%$ and $c_{\alpha^*} = 2$, so that $\mathbb{P}(\hat{S}_9 > 2) = 0.84\%$ which is notably less than the desired level of 5%. As discussed in Section 4.2.2 there is the possibility to keep the rejection tolerance fixed, i.e. to choose an integer $\kappa = c_{\alpha^*} + 1$, and then continuously adjust the partial significance level $\alpha \in (0,1)$. On the other hand, a test with a discrete statistic can usually be randomized in order to attain exactly the desired level. This is what we opt to do in the present case as will be explained next.

Recall Definition 4.2.4 and the subsequent specification of randomization in case of an (asymptotic) test with a binomially distributed test statistic. Taking equations (4.2.12) and (4.2.14) into account, p is the probability to reject the null hypothesis if the value of the test statistic happens to lie on the boundary of the critical region. That means that if $\hat{S}_n \geq c_{\alpha^*} + 1$, the null hypothesis is rejected, whereas if $\hat{S}_n < c_{\alpha^*}$, the null hypothesis is not rejected. In case $c_{\alpha^*} \leq \hat{S}_n < c_{\alpha^*} + 1$, we generate a random number r as realization of $R \sim \mathrm{Bin}(1, p)$ and reject the null hypothesis if $r = 1$. In particular, with $n = 9$, $\alpha = \alpha^* = 5\%$ and $c_{\alpha^*} = 2$ we have

$$p = \frac{\alpha^* - \mathbb{P}\left(S_n \geq c_{\alpha^*} + 1\right)}{\mathbb{P}\left(S_n = c_{\alpha^*}\right)} \approx \frac{5\% - 0.84\%}{6.29\%} = 0.6625. \tag{5.1.16}$$

Let us come back to the appropriate variance estimate $\hat{v}$ in the definition of the partial test statistic in (5.1.14). In Section 5.1.1 the details for the cases (I) of pure copula data and (II) of multivariate data with a copula-based distribution were already discussed. Thus, we choose as variance estimate the inverse of the observed Fisher information (cf. (5.1.4)) and the Genest variance estimate (cf. (5.1.8)) evaluated at the corresponding local estimate for θ_{t_k}, respectively. The columns marked (I) and (II) in Table 5.2 show the empirical size of the randomized Binomial test when based on the respective data.

		(I)	(II)	(III)(1)	(III)(2)	(III)(3)
$d = 2$	$\theta = 0.1$	0.066	0.076	0.113	0.130	0.076
	0.5	0.059	0.058	0.080	0.066	0.058
	1.5	0.059	0.065	0.070	0.074	0.065
	6	0.047	0.056	0.077	0.060	0.056
$d = 5$	$\theta = 0.1$	0.032	0.038	0.048	0.072	0.030
	0.5	0.055	0.047	0.073	0.052	0.043
	1.5	0.050	0.052	0.072	0.062	0.050
	6	0.051	0.063	0.095	0.074	0.056
$d = 10$	$\theta = 0.1$	0.049	0.042	0.068	0.074	0.062
	0.5	0.037	0.047	0.071	0.072	0.060
	1.5	0.053	0.052	0.080	0.056	0.049
	6	0.043	0.074	0.094	0.108	0.064

Table 5.2: Empirical size of randomized Binomial tests under different data set-ups (nominal level $\alpha^* = 5\%$)

The proper estimation of the moving window estimator variance in case of SCOMDY residuals proves to be a more difficult task. We have to take three different approaches to finally arrive near the desired size:

(1) Unlike the canonical maximum likelihood on realizations of a copula-based multivariate distribution, the Binomial test with a statistic where the Genest variance estimator was used, does not attain the nominal level, but exceeds $\alpha^* = 5\%$ considerably in some cases. The specific values of the empirical size are tabulated in column (III)(1). This hints at the fact that the theoretical variance, even when taking the locality of the parameter estimates into account, is too small for the finite sample set-up.

(2) Relying on the empirical studies reported in Chan et al. (2009a), we continue our efforts towards a test with proper size by parametrically bootstrapping the needed variance. As the authors argue, and we have reported in Section 3.2.2 and Appendix B, their distributional result implies that the copula parameter estimation in the SCOMDY model is independent of the deGARCHing step. Our

parametric bootstrap procedure consists in re-sampling copula data with the estimated global parameter value and re-estimating the parameters.

Analogously to the outer Monte Carlo simulation, we choose sample lengths of $T = 500$, and repeat the iterations $B = 200$ times, i.e. our bootstrap sample is of size B. For each of the M outer Monte Carlo iterations, the procedure for a bootstrap variance estimate is exactly the same and independent of the other runs. So consider $\hat{\theta}$ as the global copula parameter estimate, and for all $i = 1, \dots, B$ sample $\boldsymbol{u}_{1i}^*, \dots, \boldsymbol{u}_{Ti}^*$ according to $C_{\mathrm{C};\hat{\theta}}$. Estimate $\hat{\theta}_i^*$ via canonical maximum likelihood on the whole sample for all i, and then consider the sample variance

$$\mathrm{v}^* = \frac{1}{B-1} \sum_{i=1}^{B} \left(\hat{\theta}_i^* - \bar{\theta}^*\right)^2 \tag{5.1.17}$$

as an estimator for the variance of the global estimate in the bootstrap sample, where $\bar{\theta}^* = \sum_{i=1}^{B} \hat{\theta}_i^*/B$ denotes the sample mean. However, plugging in v^* for the variance in the partial test statistic, does not lead to substantially different results than in (1), i.e. the Binomial test still exceeds its theoretical empirical size under the null hypothesis. The respective values are tabulated in column (III)(2). Unfortunately, neither shorter or longer bootstrap samples (ranging from $T/10$ to 1000 observations), nor more bootstrap repetitions (up to $B = 1000$) improve upon this size distortions.

(3) Finally, we apply a nonparametric bootstrap on the empirical residuals from the SCOMDY model. For this, we sample B times a series of length T of independent and identically distributed realizations of the joint empirical distribution function of $\tilde{\boldsymbol{\varepsilon}}_1, \dots, \tilde{\boldsymbol{\varepsilon}}_T$. This is the classical understanding of bootstrapping (see e.g. Efron and Tibshirani, 1993), and is practically performed by "drawing with replacement" from a given sample of interest. Provided we have a series of empirical residuals $\tilde{\boldsymbol{\varepsilon}}_1, \dots, \tilde{\boldsymbol{\varepsilon}}_T$ in every outer Monte Carlo run, we sample $\boldsymbol{\varepsilon}_{1i}^*, \dots, \boldsymbol{\varepsilon}_{Ti}^*$ for every $i = 1, \dots, B$ by randomly and possibly repeatedly choosing observations from the given series. Based on every bootstrap sample, a global parameter estimate $\hat{\theta}_i^{**}$ is determined via canonical maximum likelihood, and then the bootstrap variance $\mathrm{v}^{**} = \sum_{i=1}^{B} (\hat{\theta}_i^{**} - \bar{\theta}^{**})^2/(B-1)$ is computed analogously to (5.1.17). Finally, the Binomial test based on the test statistic $\hat{T}_k$ with the variance replaced by v^{**} has attained the proper statistical size of $\alpha^* = 5\%$ in our Monte Carlo simulations. The specific numbers are provided in column (III)(3) of Table 5.2.

Usually, when there are theoretical (asymptotic) distributional assertions about a quantity of interest, one seeks to derive a statistical model-based inference procedure, and preferably avoids purely data-driven auxiliary constructions. The need for bootstrapping hints at a poor approximation quality under a finite sample set-up. Nevertheless, the simulation results presented so far give us at least the confirmation that applying a similar procedure on data whose distribution is not known a priori - as will be the case for the real-world empirical data in the next section - produces reliable test decisions with respect to errors of the first kind.

5.1.3 Implementation

The numerical simulations and the estimations and testing based on this synthetical data were done with the help of the statistical programming language `R` in its 2.11 version (see R Development Core Team, 2010).
For the copula related computations such as generating random numbers and estimating the copula parameters we used the `copula` package (see e.g. Yan, 2007 or Kojadinovic and Yan, 2010b). The computations concerning the simple GARCH(p, q) models were done with the `fGarch` package (see Wuertz et al., 2010); for estimation we directly used the provided `garchFit()` function, for simulation we adapted the `garchSim()` function to our need of a copula-based multivariate residual distribution. For more sophisticated GARCH models (used for the empirical data in Section 5.2) we used the `rugarch` package (see Ghalanos, 2012). For the quantiles of the Gumbel extremal distribution we used the `evd` package (see Stephenson, 2002).
The computational runs were performed on the resources provided by the bwGRiD project (see bwGRiD), a distributed aggregation of computer nodes operated by eight Baden-Württemberg universities. In total the bwGRiD provides 11 000 computing cores, but we in particular used one of the nodes of the cluster of the Ulm University for our numerical experiments. In the following description (see bwGRiD Cluster Ulm) the nodes are denoted by blades in accordance to the technical specification.

> "The resources of Ulm and Konstanz are jointly administered at Ulm. This allows to offer 20 IBM Bladecenter Type H with 14 IBM HS21XM Blades each, summing up to 280 Blades to users in Ulm and Konstanz. Every blade comprehends two 4-Core Intel Harpertown CPUs with 2.83 GHz and 16 GByte RAM."

Besides a usual Gigabit Ethernet connection, an additional InfiniBand interface (see InfiniBand® Trade Association) is provided and allows for high-speed communication and data exchange between the single nodes, for example via the Message-Passing-Interface (MPI) standard (see MPI Forum, 2009). However, we restricted ourselves to parallelizing our `R`-scripts merely through multicore-distributed computations. This was achieved by using the `foreach`-loop provided by the `foreach` package (see Revolution Analytics, 2011b) with the multicore routines in the `doMC` package (see Revolution Analytics, 2011a). For further information on parallel computing with `R`, especially on the pitfalls concerning distributed random number generation through MPI, see e.g. McCallum and Weston (2011) and Eugster et al. (2011).

5.2 Commodity contracts

We apply our modeling and testing approach on real-world financial data now. First, we describe the specific type of forward contract whose logreturns are assumed to evolve according to a SCOMDY model. We decided to work with commodity contracts, which is perhaps not a typical choice in the present literature. In many publications dealing with statistical models for financial data, popular data examples are equity prices of

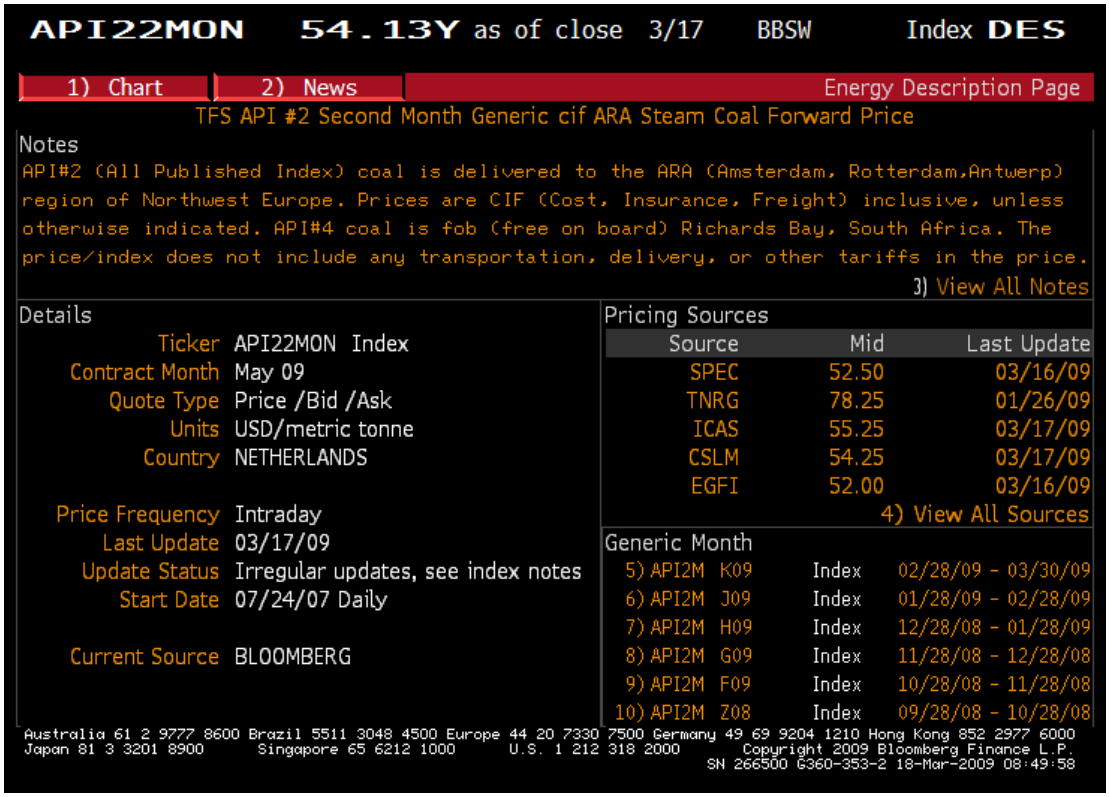


Figure 5.1: `API22MON Index` description page from Bloomberg

big global companies or the evolution of (stock price) indices, such as the German DAX or the American S&P 500 index. However, we want to stress that our procedure for homogeneity testing can be applied in rather general settings and is by no means restricted to the special kind of data described in the following.

5.2.1 Data description

We have chosen to work on so-called second front month forward prices for different types of commodities such as coal, natural gas, crude oil and electricity.
Along the lines of Hull (2002) we consider the following underlying concepts. A *forward* contract is defined to be "an agreement to buy or sell an asset at a certain future time for a certain price. [...] contrasted with a spot contract, which is an agreement to buy or sell an asset today. A forward contract is traded in the over-the-counter market [...]." In an *over-the-counter* market the trading parties set up the respective contracts on their individual terms with regards for example to trading volume, settlement agreements or maturity dates. In contrast, an *exchange-traded* market is a market where standardized contracts, with standards on volume, quality, currency, maturity and the like defined by the exchange, are traded by individuals. Often the exchange institution serves as a *clearing house* for the parties involved, i.e. as an intermediary providing settlement

services and in this role reducing the exposure to default risk for the treaty partners. Forward contracts that are exchange-traded are called *future contracts* or shortly *futures*. Besides the standardization of the contracts' framework, a *collateral* deposit, also know as *margin* deposit, is required, i.e. both parties have to provide a margin amount of (usually) money on an account as a security. These sums are then used for a regular *marking to market*, i.e. the contracted agreements are compared to the actual market prices and the respective accounts are balanced on a daily basis. As the profits and losses are already realized through the interim partial payments during the *tenor*, i.e. the contract's duration, there is usually a spot price settlement at the delivery date.
For both forward and future contracts the delivery price of the underlying asset, i.e. the price at which the settlement takes place at maturity, is called *forward price*. A *second front month* forward price quoted today is then the price for a certain amount of the underlying that is to be paid at the delivery date exactly two months as from today. The prices we use for our empirical analyses are determined as *rolling*, i.e. as an average over all quoted delivery prices of the available contracts maturing in two months. They can be regarded as a generic price index, and it has to be taken into account that they are only updated if contracts with the corresponding maturity are available, i.e. possibly not on a daily basis. We have decided to consider the second front month as these are the contracts that are traded most liquidly. Yet, the eleven different underlyings presented below are only a subset of a larger number of available index prices. The price series that we omitted exhibit long periods without any updates and therefore are not suitable for our estimation purposes. Usually the front month prices are based on over-the-counter forwards. Whenever exchange-traded futures are the calculation base, this will be indicated in the description of the respective commodities below.
The historical time series data for our analysis is retrieved from Bloomberg's business database (see Bloomberg Professional®, the electronic financial data provision service of the Bloomberg L.P. corporation) through access provided by the LBBW Trading Room operated by the Institute of Financial Mathematics at Ulm University (see LBBW Trading Room). Each security of interest, especially the generic front month prices described above, is assigned a unique identifier in the Bloomberg database, the so-called *ticker*. Together with an identifier for the respective market sector (such as `Index` for index securities or `Comdty` for commodity-related underlyings) access to for example historical data, standard options and descriptions of the underlying is provided. We now give a list of the tickers and a short description of how the prices are calculated. The description page of the first coal forward (`API2MON Index`) is provided as an example of a typical Bloomberg screen in Figure 5.1.
Note that all prices are either noted in Euro or are converted from the original currency.

`API2MON Index` coal, price per metric ton (based on **A**ll **P**ublished **I**ndices), delivered to the Amsterdam, Rotterdam and Antwerp region (ARA) in Northwest Europe.

`API4MON Index` coal, same as above, but the delivery takes place to Richards Bay, South Africa. The pricing source was not Bloomberg, but Credit Suisse.

`TTFG2MON Index` natural gas, price per megawatt hour, based on the virtual gas hub

covering all entry and exit points in the Netherlands.

NBPG2MON Index natural gas, price per megawatt hour, based on the market movements at the virtual gas hub covering mainland Britain.

CO2 Comdty Brent crude oil, price per barrel, supplied at Sullom Voe on Shetland, Scotland.

BRSWMO2 Index Brent crude oil, price per barrel, calculated by Bloomberg based on the Intercontinental Exchange (ICE) futures prices.

EL{U,G,B}B2MON Index price per megawatt hour of baseload electricity in the **U**nited Kingdom, **G**ermany and **B**elgium respectively.

EL{U,G}P2MON Index price per megawatt hour of peakload electricity in the **U**nited Kingdom and **G**ermany respectively.

Note that baseload electricity refers to contracts that have a 24-hour delivery period (23:00–23:00), whereas peakload electricity has a delivery period of 12 hours (07:00–19:00).
We retrieve the historical quotes of the raw second front month forward prices, i.e. the series $\{S_t^{(j)}\}_{t\geq 1}$ for $j = 1, \ldots, 11$ representing the different commodities. There are global non-trading days, such as weekends, that lead to nil values at the respective days across all considered time series. After removing them from the series, there are still nil values in the data, for example at local non-trading days due to national holidays or to missing price updates at certain days where no corresponding contracts were available. The over-all liquidity of the considered commodity contracts guarantees that the missing values are isolated in the single time series and therefore linearly interpolating them is unproblematic.
After this step, the transformed series of logreturns

$$R_t^{(j)} = \log\left(\frac{S_t^{(j)}}{S_{t-1}^{(j)}}\right)$$

for all $j = 1, \ldots, 11$, $t \geq 1$ forms the basis of all further considerations.
In the subsequent section we will describe the univariate GARCH models for the logreturn series, how we proceeded to decide on goodness-of-fit of the different copula classes and the test results of the Binomial and the Extremal test.

5.2.2 SCOMDY modeling and testing

After the data retrieval and preparation, the focus is on fitting appropriate time series models and finding a proper copula family to capture the dependence structure.
In a first analysis we use one year of historical data in the period from 18-03-2008 to 18-03-2009 and consider all of the possible pairwise combinations resulting in 55 bivariate series vectors. Subsequently we fit several of the copula families presented in

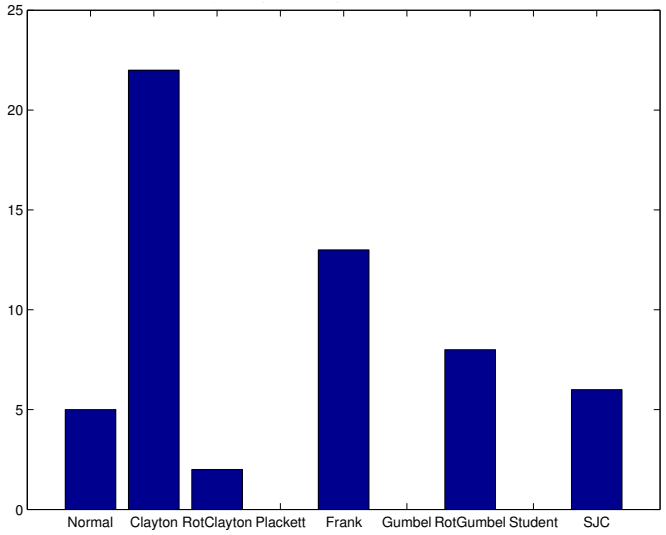


Figure 5.2: Histogram of the goodness-of-fit results

Appendix A to the rank-transformed logreturns, i.e. we do not deGARCH the series first, but directly investigate the present pairwise dependence structure. As opposed to the successive analyses, this first data examination is conducted in `matlab` with Patton's copula toolbox (see Patton, 2008). According to the maximum likelihood logic, we choose the minimal value of the negative log-likelihood function of the whole series for $t = 1, \ldots, T$ as the goodness-of-fit criterion, i.e. we select a copula family as "best fit" for a logreturn pair when the respective negative log-likelihood function is minimal compared to the other families. The result is that the Clayton copula family with a positive parameter (see A.5) emerges most frequently as "best fit". A histogram of the frequencies is given in Figure 5.2; on the x-axis the different copula families are also given.
The second step is taken with regard to suitable univariate GARCH models. As already argued in Section 3.1 and with reference to Hansen and Lunde (2001), it suffices in most cases to set up the parsimonious GARCH(1,1) model. Therefore, we assume that for the logreturns we have

$$\begin{aligned} R_t^{(j)} &= \mu_j + Y_{jt} \\ Y_{jt} &= \sigma_{jt}\varepsilon_{jt} \\ \sigma_{jt}^2 &= \omega_j + \alpha_j Y_{j,t-1}^2 + \beta_j \sigma_{j,t-1}^2 \end{aligned} \tag{5.2.1}$$

with parameters $\mu_j, \omega_j, \alpha_j, \beta_j$ and univariate white noise residuals $\varepsilon_{jt} \sim \text{WN}(0,1)$ with respect to t for all $j = 1, \ldots, 11$. This time, we take as basis for the estimation two years' time series data from 19-11-2007 to 19-11-2009 and obtain the following parameter estimates in the univariate models. In Table 5.3 the respective estimates are given together with the sum $\hat{\alpha}_j + \hat{\beta}_j$. A value greater than or equal to 1 indicates that the

respective model is not covariance-stationary; note however, that this does not affect the consistency of the quasi maximum likelihood estimates.

j		$\hat{\mu}_j$	$\hat{\omega}_j$	$\hat{\alpha}_j$	$\hat{\beta}_j$	$\hat{\alpha}_j + \hat{\beta}_j$
1	API2	8.7409e-04	6.4357e-06	0.146726	0.86039960	1.00713
2	API4	1.8548e-03	1.3157e-05	0.350214	0.70303320	1.05325
3	TTFG	5.7206e-04	1.8513e-05	0.135967	0.87672740	1.01269
4	NBPG	7.1524e-04	2.9631e-05	0.077660	0.91306180	0.99072
5	CO2	9.5303e-04	8.7841e-06	0.071987	0.91476680	0.98675
6	BRSW	9.2096e-04	8.2616e-06	0.063581	0.92285210	0.98643
7	ELUB	2.1332e-04	8.9599e-05	0.321522	0.62147820	0.94300
8	ELGB	3.5195e-05	8.0567e-04	0.211854	0.00000001	0.21185
9	ELBB	2.4977e-04	1.0554e-03	0.160923	0.03653276	0.19746
10	ELUP	-8.6258e-05	1.3132e-04	0.231266	0.63050610	0.86177
11	ELGP	2.1356e-04	8.9710e-04	0.358092	0.00000001	0.35809

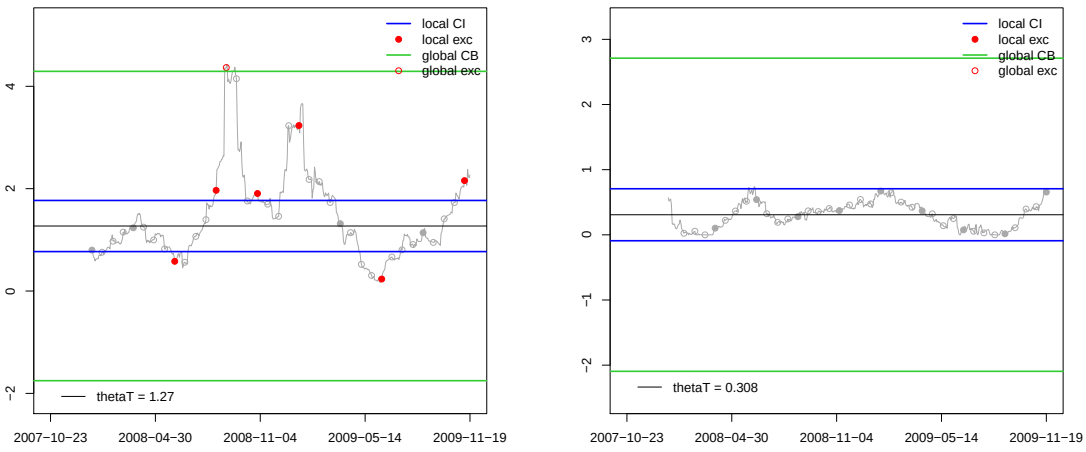
Table 5.3: Estimated parameters for univariate commodity GARCH(1,1) models

Note that we abbreviated the tickers. After deGARCHing the original logreturn series, we obtain 11 univariate series of empirical residuals which are rank-transformed before the further investigation. All bivariate combinations result in $\binom{11}{2} = 55$ possible pairs. The pairwise dependence is captured by a bivariate Clayton copula. In Table 5.4 the global parameter estimates $\hat{\theta}$ are given in the lower left triangle, and the respective variance estimates $\hat{\Sigma}$ obtained by the bootstrap procedure described in 5.1.2 are given in the upper right triangle.
The two considered trading years sum up to the observation horizon $T = 539$, the bandwidth is chosen as one tenth of T, namely $b = 54$ (which accounts for merely $n = 9$ independent estimators), the thinning constant is $c = 1/4$ (which accounts

	3.4427	1.5737	1.5563	2.2915	2.3264	1.4272	1.6389	1.4685	1.3076	1.5001	API2
1.2703		1.5303	1.4775	2.0483	2.1947	1.6920	1.7098	1.7401	1.5123	1.5528	API4
0.1407	0.1635		8.2404	1.6686	1.6423	2.2104	2.1753	1.6393	2.0779	2.4668	TTFG
0.1881	0.1702	1.6905		1.8589	1.7635	2.4607	2.8032	1.8740	1.9904	2.5159	NBPG
0.3345	0.3159	0.1743	0.1895		78.7038	1.8754	1.8699	1.5593	1.6921	1.5389	CO2
0.3095	0.3081	0.1151	0.1514	6.0296		1.7605	1.8124	1.5718	1.5494	1.5377	BRSW
0.2659	0.1791	0.6536	0.8289	0.2813	0.2536		3.0064	2.0581	2.4142	2.4250	ELUB
0.3125	0.2288	0.4261	0.4527	0.1774	0.1587	0.6175		4.4794	3.9574	7.2225	ELGB
0.2395	0.1776	0.4166	0.3651	0.2070	0.1873	0.5471	1.2602		2.4444	2.5953	ELBB
0.1943	0.1685	0.6918	0.7331	0.2387	0.2068	1.6195	0.7405	0.7813		4.6296	ELUP
0.1739	0.1172	0.4409	0.4247	0.1221	0.1088	0.6480	1.6319	1.1894	0.9103		ELGP
API2	API4	TTFG	NBPG	CO2	BRSW	ELUB	ELGB	ELBB	ELUP	ELGP	

Table 5.4: Estimated global parameter (lower left) and bootstrapped variance estimate (upper right) for bivariate SCOMDY models

for $N = 37$ weakly dependent estimators in the thinned series). With a significance level of $\alpha = \alpha^* = 5\%$ and the number of local exceedances less than $q_{\alpha^*} = 2$ the null hypothesis of parameter constancy cannot be rejected. On the other hand if the number of exceedances is greater or equal 3, then the null hypothesis is rejected. For exactly two local exceedances a randomization has to be conducted, i.e. by generating r as a realization of a Bernoulli random variable with size p as in (5.1.16), the null is rejected if $r = 1$ and cannot be rejected if $r = 0$. Furthermore, the Extremal test is conducted in its min-max version, i.e. we test for both deviations of the maximum above the upper global confidence level, as well as deviations of the minimum below the lower global confidence level. The test results can be visualized in a very intuitive way (see Figure 5.3); by giving two examples in case of the two-year data we demonstrate on the one hand so-called global exceedances which are deviations of the extremal values of the estimator series, and that they possibly may come along with extreme real-world events. On the other hand we will explain the different occurring elements of the graphics.



(a) `API22MON` vs. `API42MON` (b) `API42MON` vs. `BRSWM02`

Figure 5.3: Global exceedance and absence of exceedances

Note that the order of the independent and the thinned sequence is not the "backward time" type as stated in Definition 4.1.4, but starts with the first estimator $\hat{\theta}_b$. The horizontal black lines depict the global parameter estimate $\hat{\theta}$ (the respective values are given as `thetaT` in the legend), the whole time-varying parameter series $\{\hat{\theta}_t\}$ is displayed in grey. The solid circles mark the independent estimators $\hat{\theta}_k$, whereas the non-solid circles mark the thinned series $\hat{\theta}_{t_k}$. The marked estimators are grey whenever they lie within the local (blue) respectively the global (green) bands, otherwise they are colored in red. The observation horizon is given on the time axis, the parameter values on the y-axis. In particular, we see the following:

`API22MON` vs. `API42MON` The two coal contracts exhibit a very unstable dependence structure. We observe 6 local exceedances in the series of independent estimators,

in the thinned series there are even more local and one global exceedance. This leads to the rejection of parameter constancy in both the local and the global test. Let us summarize the special features of this pair which might help to understand this behavior. First, coal is the underlying of both contracts, solely with different delivery places. The global dependence which stems from a parameter value of $\hat{\theta} = 1.27$ is rather high compared to the other results. Last, the so called sub-prime mortgage crisis in September 2008 falls into the observation horizon, and precisely in this period the dependence exhibits a peak which exceeds the global bands.

API42MON vs. BRSWMO2 The coal contract versus the oil contract depicted here is an example for a parameter series which fluctuates around the global parameter in a very quiet way. Adding the local and global bands confirms the hypothesis of parameter constancy since there are no exceedances at all.

The numerical results that have been described so far represent the first experiments and a preliminary data survey that we conducted to get a sense for the specific features of the data. Now we report on a comprehensive numerical analysis that was conducted conclusively. We use five years of historical data from 19-12-2006 to 19-12-2011. After a preliminary data preparation (we have removed nil values, for instance), we have chosen to apply more sophisticated univariate GARCH models for the logreturn series. We proceed along the lines of Chevallier (2011b); in Chapter 6 therein a series of carbon prices, i.e. the logreturns of CO_2 allowance contracts, is modeled according to an exponential GARCH(1,1) model due to Nelson (1991). We add an autoregressive part and arrive at

$$\begin{aligned} R_t^{(j)} &= \mu_j + \rho_j R_{t-1}^{(j)} + Y_{jt} \\ Y_{jt} &= \sigma_{jt}\varepsilon_{jt} \\ \log(\sigma_{jt}^2) &= \omega_j + \alpha_j \varepsilon_{j,t-1} + \gamma_j(|\varepsilon_{j,t-1}| - \mathbb{E}\,(\varepsilon_{j,t-1})) + \beta_j \log(\sigma_{j,t-1}^2) \end{aligned} \tag{5.2.2}$$

with real parameters $\mu_j, \rho_j, \omega_j, \alpha_j, \gamma_j, \beta_j$ and univariate white noise residuals $\varepsilon_{jt} \sim \mathrm{WN}(0,1)$ with respect to t for all $j = 1, \ldots, 11$. For the estimation of these models the `rugarch` package described in Section 5.1.3 proves to be of great help. Furthermore, in accordance with the ideas presented in Chevallier (2011a) again for carbon prices, univariate structural break analyses are performed. For this task the fluctuation tests due to Bai and Perron (1998b) are used. The original application of their theory is to test for structural changes of the parameter in linear regression models at unknown dates based on least squares estimates. Furthermore, the asymptotic properties of estimators of the break dates are derived. In a companion paper (see Bai and Perron, 1998a) the practical issues of the empirical application of the procedure above are addressed, in particular the implementation of the global minimizers of the sum of squared residuals based on dynamic programming approach in the GAUSS programming language (see Aptech Systems) is provided.
Although Bai and Perron (2003) provide improved critical values in order to obtain the significance level of their test in finite samples, Zeileis and Kleiber (2005) report on further numerical difficulties when trying to reproduce these results and suppose that

this is due to the optimization routines inherent in the software. The attempt to validate the implementation was repeated with `R` and lead to an alternative implementation of the test now provided by the package `strucchange` (see Zeileis et al., 2010), which we use for our numerical experiments.
However, the reports on these analyses are rather extensive, and moreover, neither the comparison of a simple GARCH(1,1) versus the more sophisticated exponential GARCH(1,1) finally used, nor the structural break analyses brought about additional insight into the univariate data structures. Let us merely note that the asymmetric time series model yields a slightly better fit to the data than the pure GARCH approach, especially with respect to filtering the heteroscedasticity effects from the given series. Moreover the structural change test due to Bai and Perron indicates such a high number of structural changes in the univariate models that there is no meaningful economic explanation for the break dates anymore. Therefore, we refrain from a detailed report of these results and refer to the additional material provided with this thesis.
The empirical residuals of the AR(1)-eGARCH(1,1) model given in (5.2.2) are subsequently the basis of a pairwise dependence analysis, i.e. the bivariate vectors of empirical residuals are assumed to be distributed according to a copula-based multivariate distribution. We condense the results we obtained for the parameter estimate of bivariate Clayton copulas in Table 5.5.

	$\hat{\theta}$	$\hat{\Sigma}$	$\hat{\Sigma}^{(BS)}$	$\hat{\tau} = \hat{\theta}/(\hat{\theta}+2)$	$\hat{\lambda}_{\mathrm{L}} = 2^{-1/\hat{\theta}}$
∅	0.41403	0.00336	0.07040	0.17151	0.18747
min	0.00998	0.00067	0.00405	0.00497	0.00000
max	7.30908	0.10148	2.71164	0.78516	0.90952

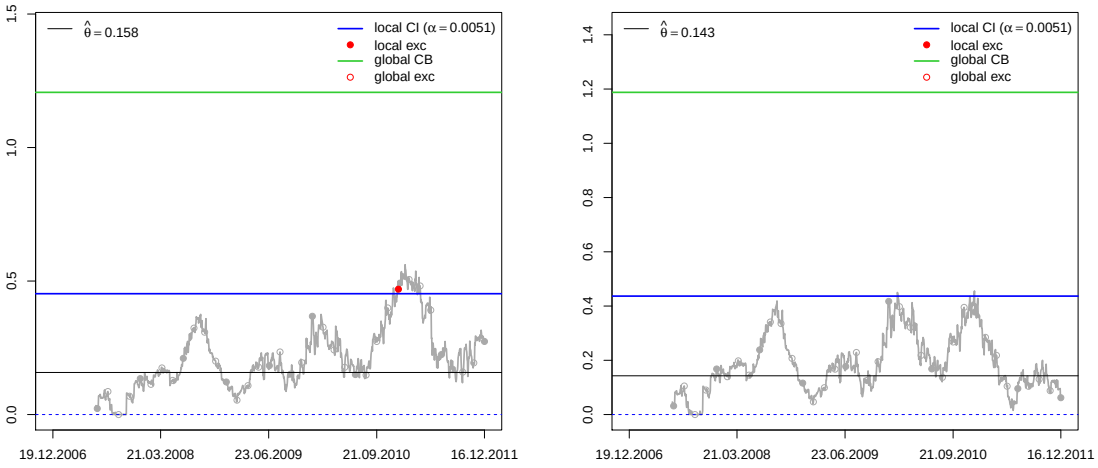
Table 5.5: Summary statistics for Clayton copula estimation with 5-year commodity data

Here, we only give the average (∅), the minimal (min) and the maximal (max) value of the global parameter estimate $\hat{\theta}$ based on the whole sample of about $T = 1300$ observations, as well as of the variance estimate $\hat{\Sigma}$ stemming from the canonical maximum likelihood estimation and the bootstrapped variance estimate $\hat{\Sigma}^{(BS)}$. The bootstrapped version of the variance estimator was obtained nonparametrically as described in Section 5.1.2 based on $B = 200$ estimations on a bootstrap sample of the same length as the bandwidth of the moving window estimates, i.e. $b = T/10$. Furthermore, we give the respective range of the estimator of Kendall's tau and the lower tail dependence. The estimation of $\hat{\theta}$ does actually work for 47 of the 55 possible pairs, so that we can conduct the Binomial and the Extremal test for these commodity pairs. This results in 25 local rejections (53%), i.e. rejections of the null hypothesis $\mathbf{H}_0$ as in (4.2.1) based on the Binomial test, and 3 global rejections (6%), i.e. rejections of $\mathbf{H}_0'$ as in (4.3.1) based on the Extremal test for the maximum of the parameter series. Note that we test $\mathbf{H}_0$ against the alternative that one of the partial hypotheses is false, i.e. against $\mathbf{H}_{A,1}$, and indeed use the level

adjustment given in equation (4.2.15) to obtain the exact size of $\alpha^* = 5\%$. For this, we choose $\kappa = 1$ and arrive at $\alpha = 0.51\%$ as the significance level in the partial tests. This is the method described in Section 4.2.2 to avoid the randomization of the Binomial test used in the simulation study in the preceeding Section 5.1.
In Figure 5.4, the graphical display of an exemplary test result on the 5-year historical data similar to Figure 5.3 is given. In subfigure 5.4(a) the estimator series $\hat{\theta}_t$ together with the local confidence regions due to the Binomial test and the global confidence bands due to the Extremal test are depicted. The empirical residuals that form the basis for the copula estimation stem from a SCOMDY model over the whole observation horizon $t = 1, \ldots, T = 1296$.
The Binomial test leads to the rejection of the null hypothesis of parameter constancy as there is one exceedance beyond the local confidence band at the point in time $\tau = 1037$, which corresponds to the date 22-12-2010. With this knowledge, we re-estimate the univariate GARCH models with two different regimes: the first is for $t = 1, \ldots, \tau$ and the second for $t = \tau + 1, \ldots, T$. Then the empirical residuals are concatenated and again used as pseudo-realizations of the residual copula-based distribution. Canonical maximum likelihood estimation of the global copula parameter $\hat{\theta}$ and the estimator series $\hat{\theta}_t$, as well as bootstrapping a variance estimate $\hat{\Sigma}^{(BS)}$ based on this concatenated series then results in a Binomial test that does not reject the null hypothesis, as subfigure 5.4(b) shows.
This approach is advantageous compared to another attempt to remedy time-variation in the copula parameter: the respective regimes are detected separately in the univariate models (with the methods provided by `strucchange`) in order to re-estimate the GARCH parameters on different regimes. With the latter approach, we are not able to arrive at time-homogeneous copula parameters, whereas interpreting the local or global exceedances in the Binomial or Extremal test respectively as change point indicators leads to regime-changed GARCH components with a constant dependence structure.
We conclude this section with an additional discussion of the possible economic interpretation of our results.

5.2.3 Discussion

The assessment of the empirical size of the Binomial test proved to be a non-trivial task and gave rise to finite-sample considerations especially for suitable variance estimates. We found that the theoretical result due to Chan et al. (2009a) cannot be unconditionally transferred to our simulated nor to the used real-world data respectively. Without accounting for the finiteness of the samples and the estimation error that naturally arises, the size of the test could not be reached. Neither through different univariate GARCH models nor by adapting the sample sizes (i.e. longer overall horizons T or relatively longer window bandwidths b respectively) this could be remedied. A considerable upward size distortion hinted at the original variance estimate being too small. Finally, applying a nonparametric bootstrap based on the empirical residuals provided a suitable variance estimate, and also the insight that the variance estimate due to Genest et al. (1995) is not apt to capture the variability in SCOMDY models.



(a) before change point handling

(b) after re-estimation of univariate models

Figure 5.4: Binomial test with proper size for **API2MON** vs. **ELGP2MON**

Concerning the empirical data, let us note the following: In shorter time series we observed fewer exceedances in sum, but also some extreme exceedances that finally gave the motivation for the Extremal test. This happened especially with pairs which have the same underlying commodity. In the 5-year time series data, the Binomial test gave more overall rejections, but we had fewer rejections when applying the Extremal test.
In our subsequent analyses, we found that there are cases when exceedances hint at some regime changes in the univariate models. Taking times of local (or global) exceedances as change points in the GARCH parameters, re-estimating the univariate models, and using the concatenated empirical residuals as basis for a repeated canonical maximum likelihood estimation of the global and local copula parameters, we found that the dependency represented by the copula parameter series exhibited a time-homogeneous structure at least in some of the investigated cases.
However, there were also cases where regime-changed GARCH models did not have any effect on the heterogeneity in the copula parameter. With a merely graphical valuation of the results, the estimator series seemed to have either different regimes themselves, or exhibited an autoregressive structure. The change point test given in Section 4.4 was though not applied for this specific data set, i.e. the change point hypothesis was not tested formally. The hypothesis that the estimator series evolves according to an autoregression was neither tested, in particular since we did not develop any such testing approach.
From the economic point of view, our observations and test results are in line with similar econometric findings. For example, Gaisser et al. (2009) measure a portfolio's dependence structure through Spearman's ρ and find that significant deviations from the hypothesis of constancy come along with unusual economic events (their procedure actually detects the volatility peaks that the financial markets exhibited in 2002 and

the beginning of 2004).
A very recent example regarding the assessment of time-dynamic dependencies with the help of copulas is Stöber and Czado (2012). There a pair copula construction, also called vine copula, is used to model the empirical residuals of high-dimensinal financial data. By applying Bayesian estimation techniques to determine Markov-switching parameters, the authors find that different regimes in the dependency parameters are often implied by exceptional economic events (such as the 2008 subprime crisis).
Moreover, Almeida et al. (2012) investigate the dependence between the DAX 30 constituents with vine copulas and incorporate time-dynamics by the stochastic autoregression due to Hafner and Manner (2008). For their empirical study they use a two-step estimation procedure, namely first a deGARCHing of financial time series data and second a Bayesian estimation of all the occuring dependence parameters. Besides their finding that according to the BIC model selection criterion non-dynamic models outperform the time-varying ones in most of the levels of the vine construction, they note in their concluding section that "the model uncertainty introduced by assuming the marginal parameter estimates as true ones in the two-step approach has [still] be to assessed." (see Almeida et al., 2012, p. 23).
The empirical application of the Binomial and the Extremal test that were introduced and theoretically examined in Chapter 4 showed that a suitable variance estimation (through a bootstrapping procedure) is essential for their use in finite samples. Taking this possible pitfall into account, the tests prove useful in detecting deviations from time-homogeneous copula parameters. These deviations can be seen as hints at regime-changed underlying models and as starting point for an improved univariate analysis.

A Examples of Copula Families Used

This Appendix is devoted to the illustration of the copula families used in this thesis. These are some of the most famous representatives of the copula classes presented in Section 2.1.5. As we aim at calculating concordance and tail dependence measures as well as at a graphical presentation of these numbers it is appropriate to consider the bivariate case. The contour plots represent the densities of multivariate distributions with the respective copulas and standard normal marginal distributions.
In the following let $(u, v) \in [0,1]^2$. To stress the functional dependence of the copulas on a parameter θ we will write $C(u, v; \theta)$ whenever we are interested in the arguments (u, v) and C_θ otherwise. This notation will be analogously used for the densities, but not for τ, ρ, λ_l and λ_u. Archimedean generators will always be written as φ_θ in accordance with the usual notation.
Furthermore, we will denote by C_{θ_0} the limit $\lim_{\theta\to\theta_0} C_\theta$ for certain values $\theta_0 \in \Theta$ to demonstrate hereby special limiting cases in the copula families. Most often these will be the comonotonicity copula M, the independence copula Π and the countermonotonicity copula W.

A.1 Gaussian copula

Equation (2.1.19) is formulated in the bivariate case now. It then becomes apparent that it is Pearson's correlation coefficient that mainly captures the dependence in Gaussian models.
Let $\rho \in [-1, 1]$ be a given correlation and consider the following covariance matrix which is at the same time the corresponding correlation matrix

$$\Sigma = P = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \tag{A.1.1}$$

The Gaussian copula is then only a function of ρ and will correspondingly be denoted by $C_{\mathrm{Ga}}(u, v; \rho)$. Recall the density of the univariate standard normal distribution $\phi(x) = (2\pi)^{-1/2}\,\mathrm{e}^{-x^2/2}$ and its distribution function $\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$ as well as the density function of the centered bivariate normal distribution with covariance matrix P

$$\phi_\rho(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\}.$$

For all $(u, v) \in [0,1]^2$ we hence have a distribution function, the Gaussian copula

$$C_{\mathrm{Ga}}(u, v; \rho) = \int_{-\infty}^{\Phi^{-1}(v)} \int_{-\infty}^{\Phi^{-1}(u)} \phi_\rho(x, y)dydx \tag{A.1.2}$$

and the corresponding copula density in implicit form

$$c_{\mathrm{Ga}}(u, v; \rho) = \frac{\phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\,\phi(\Phi^{-1}(v))}. \tag{A.1.3}$$

Kendall's concordance coefficient can now be expressed in terms of the correlation

$$\tau = \frac{2}{\pi} \arcsin \rho. \tag{A.1.4}$$

A proof of this result can be found in Fang et al. (2002). The most remarkable fact about this representation is that it holds for all elliptical, even for the generalization to meta-elliptical distributions; in particular it is hence true for the Student copula in Section A.2 below.

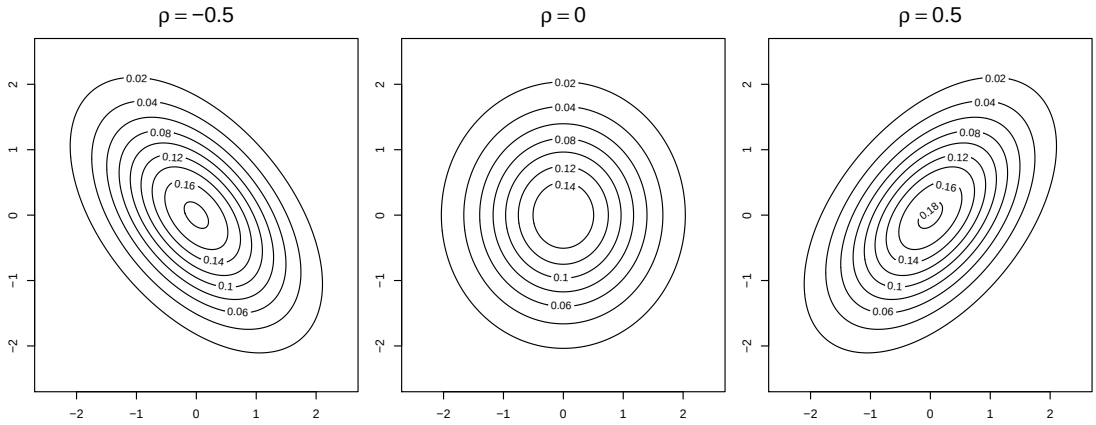


Figure A.1: Gaussian copula $C_{\mathrm{Ga};\rho}$ with standard normal margins

Use the representation in equation (2.1.17) for the upper tail dependence coefficient λ_u and the fact that for $(X, Y) \sim \mathcal{N}(\mathbf{0}, P)$ the conditional distribution of Y given $X = t$ is the univariate normal $\mathcal{N}(\rho t, 1 - \rho^2)$, to conclude

$$\begin{aligned}
\lambda_u &= 2 - \lim_{t \uparrow 1} \left(\frac{\partial}{\partial u} C(u, v) + \frac{\partial}{\partial v} C(u, v) \Big|_{u=v=t} \right) \\
&= 2 - \lim_{t \uparrow 1} \left(\mathbb{P}\left(V \leq t \,|U = t\right) + \mathbb{P}\left(U \leq t \,|V = t\right) \right) \\
&= 2 - 2 \lim_{t \uparrow 1} \mathbb{P}\left(V \leq t \,|U = t\right) \\
&\overset{(\star)}{=} 2 - 2 \lim_{t \to \infty} \mathbb{P}\left(Y \leq t \,|X = t\right) \\
&= 2 - 2 \lim_{t \to \infty} \Phi\left(\frac{t(1-\rho)}{\sqrt{1-\rho^2}} \right) = 2 - 2 \cdot 1 = 0,
\end{aligned}$$

where $(\star)$ denotes the substitution $X = \Phi^{-1}(U)$ and $Y = \Phi^{-1}(V)$. The lack of upper tail dependence transfers to the lower tail due to the radial symmetry of Gaussian copulas. In Hofert (2010, p. 48) the arguments are similar to derive $\lambda_l = 0$ directly.
This tail independence as well as the elliptic contours are apparent in Figure A.1. There are depicted the densities of multivariate distributions with bivariate Gaussian copulas and standard normal univariate margins. Pearson's correlation ranges from $\rho = -0.5$ (negative dependence) over $\rho = 0$ (independence) to $\rho = 0.5$ (positive dependence).

A.2 Student copula

With the same correlation matrix P as in (A.1.1) above we now give the bivariate form of the Student copula with ν degrees of freedom and correlation matrix P. To this end recall the distribution function of the univariate Student distribution with ν degrees of freedom

$$t_\nu(x) = \int_{-\infty}^{x} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{s^2}{\nu}\right)^{-\frac{\nu+1}{2}} ds \tag{A.2.1}$$

as well as the distribution function of the bivariate Student distribution with ν degrees of freedom and correlation matrix P

$$t_{\nu,\rho}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2})\nu\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 - 2\rho st + t^2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} dsdt \tag{A.2.2}$$

In accordance to the defining equation (2.1.20) the bivariate Student copula is given in fully analogous form

$$C_{t_\nu}(u,v;\rho) = \int_{-\infty}^{t_\nu^{-1}(v)}\int_{-\infty}^{t_\nu^{-1}(u)} \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2})\nu\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 - 2\rho st + t^2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} dsdt \tag{A.2.3}$$

and its density can be simplified to

$$c_{t_\nu}(u,v;\rho) = \rho^{-1/2}\frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})^2} \left(\frac{1 + \frac{(t_\nu^{-1}(u))^2 + (t_\nu^{-1}(v))^2 - 2\rho t_\nu^{-1}(u)t_\nu^{-1}(v)}{\nu(1-\rho^2)}}{\left(1 + \frac{(t_\nu^{-1}(u))^2}{\nu}\right)\left(1 + \frac{(t_\nu^{-1}(v))^2}{\nu}\right)}\right)^{-\frac{\nu+2}{2}}. \tag{A.2.4}$$

In Demarta and McNeil (2004) it is shown that $\lambda_l = \lambda_u = 2t_{\nu+1}(-\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho})$. The arguments are the same as in the Gaussian case above utilizing the fact that for a bivariately t_ν-distributed random pair (X,Y) with correlation ρ, under the condition that $X = t$, it holds that

$$\sqrt{\frac{\nu+1}{\nu+t^2}}\frac{Y - \rho t}{\sqrt{1-\rho^2}} \sim t_{\nu+1}.$$

Thus, with Student copulas it is possible to model tail dependence, but necessarily the lower and upper tail dependence coefficients are equal due to the radial symmetry of elliptical distributions.

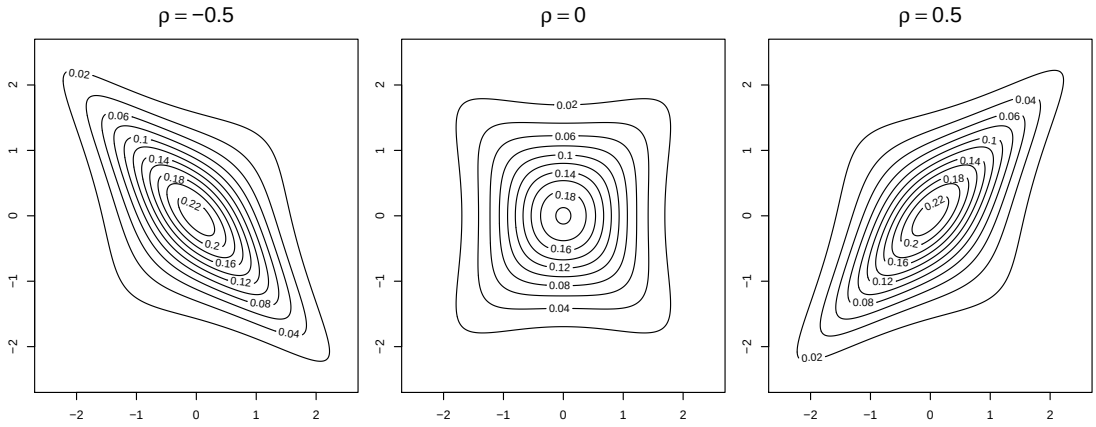


Figure A.2: Student copula $C_{t_\nu;\rho}$ with $\nu = 2$ degrees of freedom and standard normal margins

In Figure A.2 there are the contour plots of the densities of multivariate distributions with Student copulas with $\nu = 2$ degrees of freedom, $\rho \in \{-0.5, 0, 0.5\}$ and univariate standard normal margins.

A.3 Marshall-Olkin copula

Let $\lambda_1, \lambda_2, \lambda_{12} > 0$ be the intensities of the underlying Poisson processes driving the shocks that affect the first, the second or both components in the given survival system respectively. Define $\alpha_i := \frac{\lambda_{12}}{\lambda_i + \lambda_{12}}$, $i = 1, 2$ such that $\alpha_1, \alpha_2 \in (0,1)$. Recall the definition of the bivariate Marshall-Olkin copula (see Definition 2.1.18) for all $(u, v) \in [0,1]^2$

$$C_{\mathrm{MO}}(u, v; \alpha_1, \alpha_2) = \min\left\{vu^{1-\alpha_1}, uv^{1-\alpha_2}\right\}. \tag{A.3.1}$$

First, let us take a look at the density of the Marshall-Olkin copula where it is defined, i.e. beyond the singular component

$$c_{\mathrm{MO}}(u, v; \alpha_1, \alpha_2) = \begin{cases} (1-\alpha_1)u^{-\alpha_1}, & u^{\alpha_1} < v^{\alpha_2} \\ (1-\alpha_2)v^{-\alpha_2}, & u^{\alpha_1} > v^{\alpha_2}. \end{cases} \tag{A.3.2}$$

Kendall's tau for Marshall-Olkin copulas can be derived taking into account the fact that for every copula $C\colon [0,1]^2 \to [0,1]$ it holds that (see Nelsen, 2006, Thm. 5.1.5)

$$\iint_{[0,1]^2} C(u, v)\, dC(u, v) = \frac{1}{2} - \iint_{[0,1]^2} \frac{\partial}{\partial u} C(u, v) \frac{\partial}{\partial v} C(u, v) du dv.$$

Especially we then have

$$\iint_{[0,1]^2} C_{\mathrm{MO}}(u, v; \alpha_1, \alpha_2)\, dC_{\mathrm{MO}}(u, v; \alpha_1, \alpha_2) = \frac{1}{4}\left(1 - \frac{\alpha_1\alpha_2}{\alpha_1 - \alpha_1\alpha_2 + \alpha_2}\right)$$

and as a consequence of equation (2.1.10)

$$\tau = \frac{\alpha_1\alpha_2}{\alpha_1 - \alpha_1\alpha_2 + \alpha_2}. \tag{A.3.3}$$

Straightforward application of equation (2.1.11) leads to

$$\rho = \frac{3\alpha_1\alpha_2}{2\alpha_1 - \alpha_1\alpha_2 + 2\alpha_2}. \tag{A.3.4}$$

Furthermore, the tail dependence coefficients are derived in the following. Note that $\alpha_1, \alpha_2 \neq 1$ as $\lambda_i > 0$ for all $i \in \{1, 2, 12\}$. For λ_l we use equation (2.1.14) to conclude that

$$\lambda_l = \lim_{t\downarrow 0} \frac{C_{\mathrm{MO}}(t,t;\alpha_1,\alpha_2)}{t} = \lim_{t\downarrow 0} \left(t^{1-\alpha_1} \wedge t^{1-\alpha_2}\right) = 0$$

and for λ_u we apply equation (2.1.17) to see that

$$\begin{aligned}
\lambda_u &= 2 - \lim_{t\uparrow 1} \left(((1-\alpha_1)t^{1-\alpha_1} \wedge t^{1-\alpha_2}) + (t^{1-\alpha_2} \wedge (1-\alpha_2)t^{1-\alpha_2})\right) \\
&= \begin{cases} 2 - (1 - \alpha_1 + 1) = \alpha_1, & \text{if } \alpha_1 < \alpha_2 \\ 2 - (1 + 1 - \alpha_2) = \alpha_2, & \text{if } \alpha_2 < \alpha_1 \end{cases} \\
&= \min\{\alpha_1, \alpha_2\}
\end{aligned}$$

where $a \wedge b$ denotes the minimum of a and b.
The upper tail dependence as well as the singular component $\{u^{\alpha_1} = v^{\alpha_2}\}$ become visible in Figure A.3. In the three unit cubes $[0,1]^2$ there are scatterplots of random samples of size 1000 of pairs (u, v) with $\lambda_{12} = 1$ and λ_1 and λ_2 as given above the figures (the respective values for α_1 and α_2 are given below the figures).

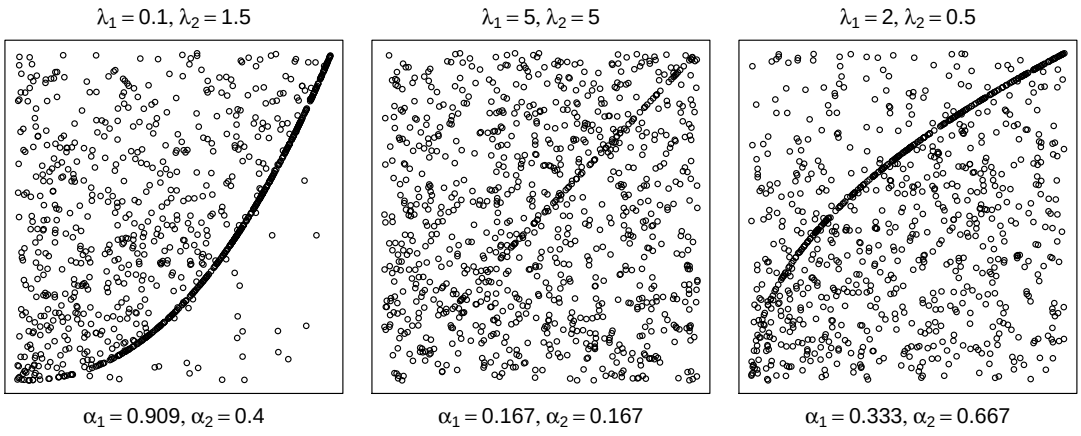


Figure A.3: Marshall-Olkin copula $C_{\mathrm{MO};\alpha_1,\alpha_2}$ with constant $\lambda_{12} = 1$, $n = 1000$

A.4 Plackett copula

The family of Plackett copulas was first introduced in Plackett (1965) and is a family of absolutely continuous copulas. The originally bivariate model can be extended to higher dimensions, see Molenberghs and Lesaffre (1994). With the parameter θ being element of $\Theta = (0, \infty)\backslash\{1\}$, the functional form of the bivariate copula is

$$C_{\mathrm{P}}(u, v; \theta) = \frac{\left(1 + (\theta - 1)(u + v) - \sqrt{(1 + (\theta - 1)(u + v))^2 - 4\theta(\theta - 1)uv}\right)}{2(\theta - 1)}. \quad \text{(A.4.1)}$$

The corresponding density is

$$c_{\mathrm{P}}(u, v; \theta) = \frac{\theta\,(1 + (\theta - 1)(u + v - 2uv))}{\sqrt{((1 + (\theta - 1)(u + v))^2 - 4\theta(\theta - 1)uv)^3}}. \quad \text{(A.4.2)}$$

The Plackett copula family is *comprehensive*, i.e. with (the closure of) the given parameter range it is possible to attain the Fréchet-Hoeffding bounds as well as the independence copula. Here we have $C_{\mathrm{P};0} = W$ over $C_{\mathrm{P};1} = \Pi$ to $C_{\mathrm{P};\infty} = M$.
We can obtain a functional connection of Spearman's rho to the copula parameter, namely

$$\rho(C_{\mathrm{P}}) = \frac{\theta + 1}{\theta - 1} - \frac{2\theta \log(\theta)}{(\theta - 1)^2}, \quad \text{(A.4.3)}$$

but no closed-form expression for Kendall's tau is known.
However, it can be shown that $\lambda_l = \lambda_u = 0$, i.e. with Plackett copulas we are not able to capture any tail dependence. This also becomes visible in Figure A.4 with θ such that U and V are almost countermonotonic ($\theta = 0.1$), independent ($\theta \to 1$) and almost comonotonic ($\theta = 10$). The margins are transformed to standard normal.

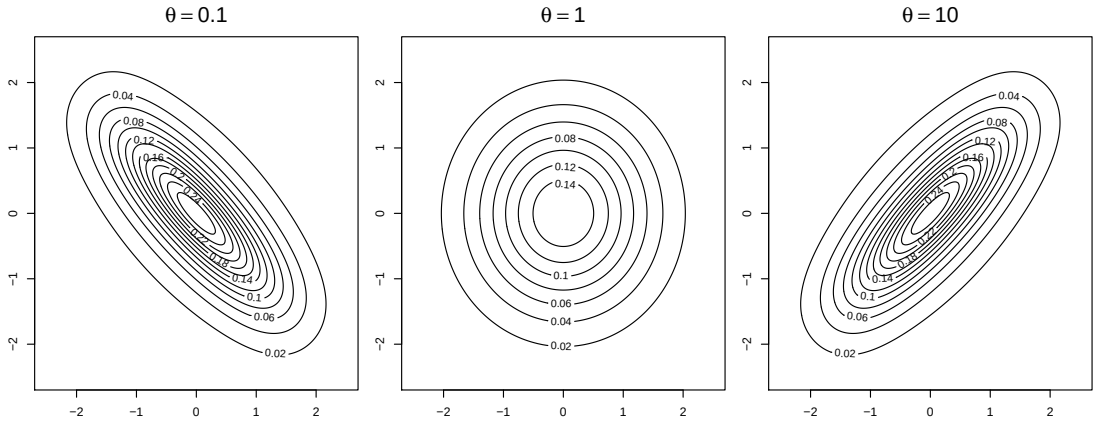


Figure A.4: Plackett copula $C_{\mathrm{P};\theta}$ with standard normal margins

A.5 Clayton copula

The Clayton copula family appeared for the first time in Clayton (1978), was discussed in several publications afterwards (see e.g. Oakes, 1982, 1986; Cox and Oakes, 1984; Cook and Johnson, 1981, 1986; Genest and MacKay, 1986b; Hutchinson and Lai, 1990), but it was Genest and Rivest (1993) that coined the wording used most often today. The Clayton copulas form a family of Archimedean copulas with generator

$$\varphi_\theta(t) = \frac{1}{\theta}(t^{-\theta} - 1), \quad \theta \in \Theta = [-1, \infty)\backslash\{0\}.$$

For $\theta > 0$ the generator is strict, for $\theta \in [-1, 0)$ the copula consists of both an absolutely continuous and a singular component. The singular component is visible in the scatterplots in Figure A.5 where samples of 1000 random pairs (u, v) on the unit cube $[0,1]^2$ are depicted.
In the bivariate case the copula has the following form

$$C_{\mathrm{C}}(u, v; \theta) = \left(\max\{u^{-\theta} + v^{-\theta} - 1, 0\}\right)^{-1/\theta}. \tag{A.5.1}$$

The corresponding density for $\theta > 0$ is

$$c_{\mathrm{C}}(u, v; \theta) = (1 + \theta)(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta - 2}(uv)^{-\theta - 1}. \tag{A.5.2}$$

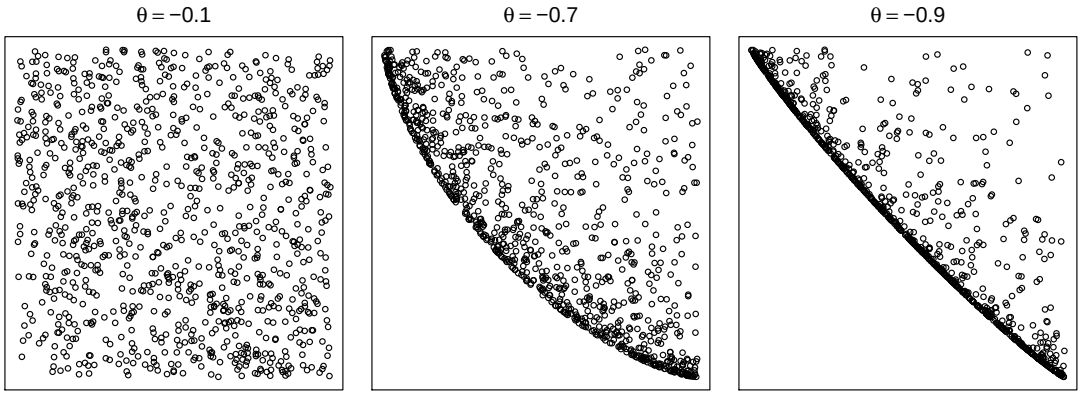


Figure A.5: Clayton copula $C_{\mathrm{C};\theta}$ with $\theta < 0$, $n = 1000$

Furthermore, the Clayton family is also comprehensive. In particular, it holds that $C_{\mathrm{C};-1} = W$, $C_{\mathrm{C};0} = \Pi$ and $C_{\mathrm{C};\infty} = M$. Contour plots in case of (almost) independence and comonotonicity (again of densities for a bivariate distribution with Clayton copula and standard normal margins) are depicted in the leftmost and the rightmost subfigures of Figure A.6 respectively.

We apply Proposition 2.1.29 to derive Kendall's tau and the tail dependence coefficients. Firtsly, we need $\varphi'_\theta(t) = -t^{-\theta-1}$ in order to obtain

$$\tau = 1 + 4\int_0^1 \frac{\varphi_\theta(t)}{\varphi'_\theta(t)} = \frac{\theta}{\theta+2}$$

and secondly, we used $\varphi_\theta^{[-1]}(t) = (1+\theta t)-1/\theta$ in order to obtain

$$\lambda_l = \lim_{x\to\infty} \frac{\varphi_\theta^{[-1]}(2x)}{\varphi_\theta^{[-1]}(x)} = 2^{-1/\theta} \quad \text{and} \quad \lambda_u = 2 - \lim_{x\downarrow 0} \frac{1-\varphi_\theta^{[-1]}(2x)}{1-\varphi_\theta^{[-1]}(x)} = 0.$$

We see that random pairs (U, V) distributed according to a Clayton copula exhibit only lower tail dependence, in particular tail asymmetry. At the same time, the pair $(1-U, 1-V)$ is said to be distributed according to the rotated Clayton copula C_{rC} which then allows for upper tail dependence of the same magnitude. Denote C_{rC} by

$$C_{\mathrm{rC}}(u, v; \theta) = C_{\mathrm{C}}(1-u, 1-v; \theta). \tag{A.5.3}$$

This copula is mentioned here for completeness and can be considered also as survival copula belonging to the Clayton distribution. It is not an Archimedean copula in the sense of Definition 2.1.20.

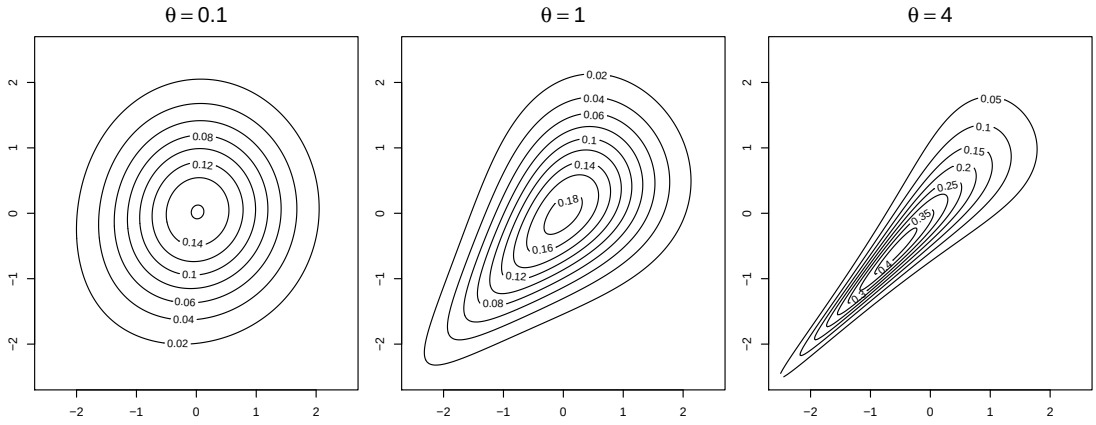


Figure A.6: Clayton copula $C_{\mathrm{C};\theta}$ with standard normal margins

A.6 Gumbel copula

The Gumbel copula family was first discussed in Gumbel (1960). But it also appeared in Hougaard (1986) and Hutchinson and Lai (1990) afterwards and is therefore sometimes referred to as Gumbel-Hougaard copula family (see e.g. copula number 4 in Nelsen, 2006, Tab. 4.1).

Gumbel copulas are Archimedean copulas and for $\theta \in \Theta = [1, \infty)$ they are generated by the Archimedean generator

$$\varphi_\theta(t) = (-\log t)^\theta.$$

They possess the following bivariate form

$$C_{\mathrm{G}}(u, v; \theta) = \exp\left\{-\left((-\log u)^\theta + (-\log v)^\theta\right)^{1/\theta}\right\}. \tag{A.6.1}$$

The corresponding density is

$$c_{\mathrm{G}}(u, v; \theta) = C_{\mathrm{G}}(u, v; \theta) \frac{((-\log u)(-\log v))^{\theta-1}}{((-\log u)^\theta + (-\log v)^\theta)^{2-1/\theta}} \left(((-\log u)^\theta + (-\log v)^\theta) + \theta - 1\right). \tag{A.6.2}$$

Kendall's tau and the tail dependence coefficients are again calculated as given in Proposition 2.1.29. We need $\varphi'_\theta(t) = -\frac{\theta}{t}(-\log t)^{\theta-1}$ in order to obtain

$$\tau = 1 + 4\int_0^1 \frac{\varphi_\theta(t)}{\varphi'_\theta(t)} dt = \frac{\theta - 1}{\theta}$$

and $\varphi_\theta^{[-1]}(t) = \mathrm{e}^{-t^{1/\theta}}$ to derive

$$\lambda_l = \lim_{x \to \infty} \frac{\varphi_\theta^{[-1]}(2x)}{\varphi_\theta^{[-1]}(x)} = 0 \quad \text{and} \quad \lambda_u = 2 - \lim_{x \downarrow 0} \frac{1 - \varphi_\theta^{[-1]}(2x)}{1 - \varphi_\theta^{[-1]}(x)} = 2 - 2^{1/\theta}.$$

If interested in modeling upper tail dependence as in Figure A.7 the Gumbel copula may be chosen. However, it should be noted that Gumbel copulas are only able to capture positive dependence. Especially, they merely interpolate between independence and comonotonicity as $C_{\mathrm{G};1} = \Pi$ and $C_{\mathrm{G};\infty} = M$.
In analogy to the Clayton copula in Section A.5 let the rotated Gumbel copula C_{rG} be mentioned as a possibility to model lower tail dependence. Formally, we have

$$C_{\mathrm{rG}}(u, v; \theta) = C_{\mathrm{G}}(1 - u, 1 - v; \theta). \tag{A.6.3}$$

Again, it is not an Archimedean copula but corresponds to the copula of the random pair $(1 - U, 1 - V)$ in case (U, V) is distributed according to a Gumbel copula.

A.7 Frank copula

The Frank family of parametric copulas was introduced in a non-statistical context in Frank (1979) and thereafter discussed in Nelsen (1986) and Genest (1987). For $\theta \in \mathbb{R}\backslash\{0\}$ the Archimedean generator of this copula family is given by

$$\varphi_\theta(t) = -\log\left(\frac{\mathrm{e}^{-\theta t} - 1}{\mathrm{e}^{-\theta} - 1}\right)$$

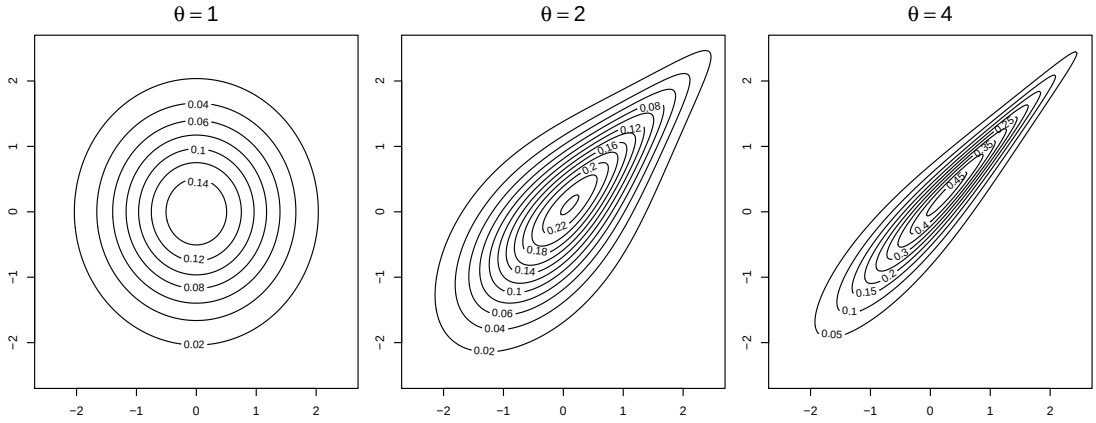


Figure A.7: Gumbel copula $C_{\mathrm{G};\theta}$ with standard normal margins

such that for bivariate random vectors with uniform margins the following distribution function results

$$C_{\mathrm{F}}(u,v;\theta) = -\frac{1}{\theta}\log\left(1+\frac{(\mathrm{e}^{-\theta u}-1)(\mathrm{e}^{-\theta v}-1)}{\mathrm{e}^{-\theta}-1}\right). \tag{A.7.1}$$

The respective copula density has the following form

$$c_{\mathrm{F}}(u,v;\theta) = \frac{\theta(\mathrm{e}^{-\theta}-1)\,\mathrm{e}^{-\theta(u+v)}}{((\mathrm{e}^{-\theta}-1)-(\mathrm{e}^{-\theta u}-1)(\mathrm{e}^{-\theta v}-1))^2}. \tag{A.7.2}$$

The most remarkable property, already derived in the original publication by Frank (1979), is that of radial symmetry. Frank copulas are indeed the only Archimedean copulas such that

$$C(u,v) = \check{C}(u,v)$$

where $\check{C}(u,v) = u+v-1+C(1-u,1-v)$ is the survival copula corresponding to C, i.e. the distribution function of $(1-U,1-V)$ in case (U,V) is distributed according to C. Moreover, the Frank copula family is comprehensive as it holds that $C_{\mathrm{F};-\infty} = W$, $C_{\mathrm{F};0} = \Pi$ and $C_{\mathrm{F};\infty} = M$.
In the given contour plot not only the radial symmetry is visible, but also that Frank copulas do not exhibit tail dependence. Formally, using $\varphi_\theta^{[-1]}(t) = -\frac{1}{\theta}\log\left(\mathrm{e}^{-t}(\mathrm{e}^{-\theta}-1)+1\right)$ in Proposition 2.1.29(2) yields

$$\lambda_l = \lim_{x\to\infty}\frac{\varphi_\theta^{[-1]}(2x)}{\varphi_\theta^{[-1]}(x)} = 0 \quad \text{and} \quad \lambda_u = 2-\lim_{x\downarrow 0}\frac{1-\varphi_\theta^{[-1]}(2x)}{1-\varphi_\theta^{[-1]}(x)} = 0.$$

For Kendall's tau we use $\varphi'_\theta(t) = \frac{\theta\,\mathrm{e}^{-\theta t}}{\mathrm{e}^{-\theta t}-1}$ in Proposition 2.1.29(1) and obtain

$$\tau = 1+4\int_0^1 \frac{\varphi_\theta(t)}{\varphi'_\theta(t)}dt = 1+\frac{4}{\theta}(D_1(\theta)-1)$$

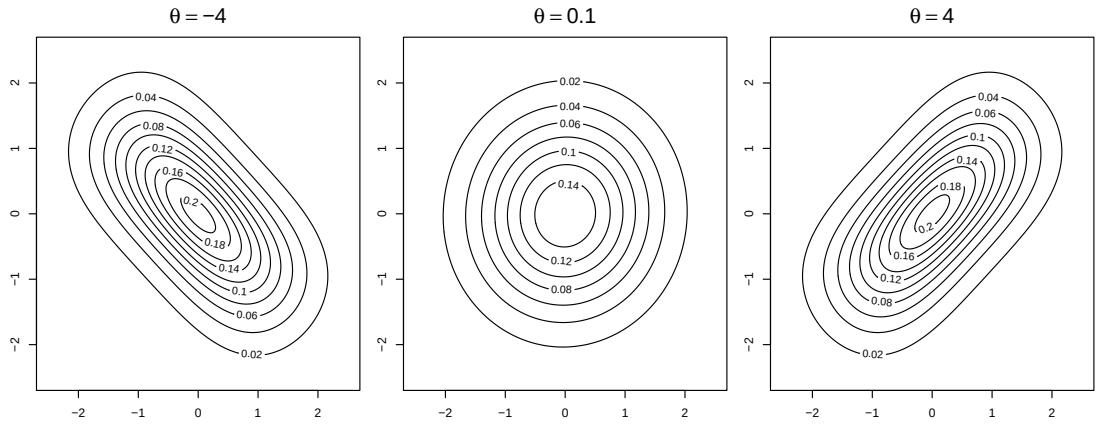


Figure A.8: Frank copula $C_{\mathrm{F};\theta}$ with standard normal margins

with $D_1(\theta) := \frac{1}{\theta}\int_0^\theta \frac{t}{\mathrm{e}^t-1}dt$ the so-called *Debye function of order* 1.

A.8 Symmetrized Joe-Clayton copula

The family of symmetrized Joe-Clayton copulas is a two-parameter copula family introduced by Patton (2006). It is a slight modification of the BB7 family in Joe (1997, p. 153). A BB7 copula (also called the Joe-Clayton copula) is a two-parameter Archimedean copula with generator

$$\varphi_{\theta,\delta}(t) = \Big(1-(1-t)^\theta\Big)^{-\delta} - 1$$

and has the following bivariate form

$$C_{\mathrm{JC}}(u,v;\lambda_l,\lambda_u) = 1 - \left(\Big((1-(1-u)^\theta)^{-\delta} + (1-(1-v)^\theta)^{-\delta}\Big)^{-1/\delta}\right)^{1/\theta} \tag{A.8.1}$$

where the parameters $\theta = 1/\log_2(2-\lambda_u)$ and $\delta = -1/\log_2(\lambda_l)$ are functions of the lower and upper tail dependence coefficients λ_l, $\lambda_u \in (0,1)$ constrained to the open unit interval.
The major drawback of the Joe-Clayton copula is that even if the tail dependence coefficients are equal, i.e. $\lambda_l = \lambda_u$, the copula exhibits asymmetry. This can be remedied by a linear combination of the copula C_{JC} and its survival copula $\check{C}_{\mathrm{JC}}$, the resulting symmetrized Joe-Clayton copula has the following bivariate form

$$C_{\mathrm{SJC}}(u,v;\lambda_l,\lambda_u) = \frac{1}{2}\Big(C_{\mathrm{JC}}(u,v;\lambda_l,\lambda_u) + \check{C}_{\mathrm{JC}}(u,v;\lambda_u,\lambda_l)\Big). \tag{A.8.2}$$

This construction does not mainly aim at symmetric dependency structures, but at a consistent model in which the tail dependence measures completely determine the

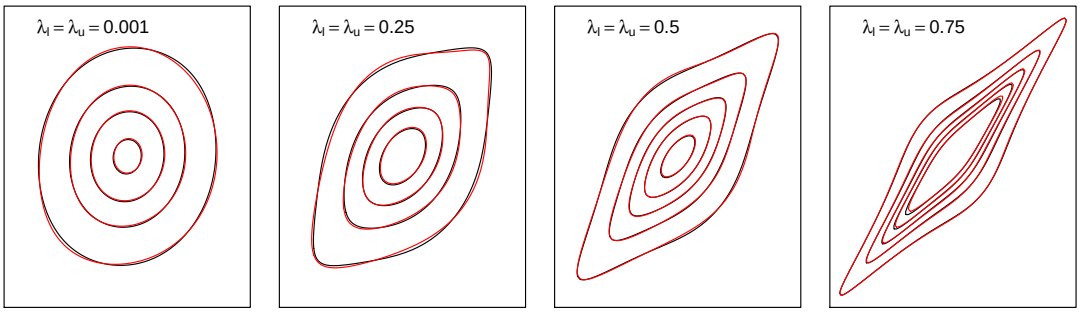


Figure A.9: Joe-Clayton copula $C_{\mathrm{JC};\lambda_l,\lambda_u}$ (red) vs. symmetrized Joe-Clayton copula $C_{\mathrm{SJC};\lambda_l,\lambda_u}$ (black) with standard normal margins

presence or absence of asymmetries. The lack of symmetry for Joe-Clayton copulas as opposed to symmetrized Joe-Clayton copulas becomes visible in contrasting the respective contour plots. In Figure A.9 there are contour plots of the densities of bivariate distributions with C_{JC} and C_{SJC}, respectively, where we have chosen $\lambda_l = \lambda_u \in \{0.001, 0.25, 0.5, 0.75\}$ and again standard normal margins. The overlay demonstrates that the contours of the Joe-Clayton based densities (red lines) are slightly asymmetric compared to the densities based on the symmetrized Joe-Clayton copula (black lines). The tail dependence coefficients can also be given in terms of $\theta \geq 1$ and $\delta > 0$ by straightforward transformations as $\lambda_l = 2^{-1/\delta}$ and $\lambda_u = 2 - 2^{1/\theta}$. As limiting cases for $\theta, \delta \to \infty$ we have $C_{\mathrm{SJC};\lambda_l,1} = C_{\mathrm{SJC};1,\lambda_u} = M$.
The copula density of C_{SJC} is

$$c_{\mathrm{SJC}}(u, v; \lambda_l, \lambda_u) = \frac{1}{2}\left(c_{\mathrm{JC}}(u, v; \lambda_l, \lambda_u) + c_{\mathrm{JC}}(1-u, 1-v; \lambda_u, \lambda_l)\right). \tag{A.8.3}$$

Above it is given as a linear combination of the densities of Joe-Clayton copulas. These have the following functional form (see also the online supplement Patton, 2003)

$$\begin{aligned} c_{\mathrm{JC}}(u, v) =& \frac{A}{B} \qquad (A.8.4)\\ A =& (1-\eta_u)^{\delta-1}\frac{\eta_u}{1-u} \\ & (-1+\theta(\delta(-1+(-1+(1-\eta_u)^{-\delta}+(1-\eta_u)^{-\delta})^{1/\delta}) \\ & +(-1+(1-\eta_u)^{-\delta}+(1-\eta_v)^{-\delta})^{1/\delta}) \\ & (1-(-1+(1-\eta_u)^{-\delta}+(1-\eta_v)^{-\delta})^{1/\delta})^{1/\theta} \\ & (1-\eta_v)^{\delta-1}\frac{\eta_v}{1-v}) \\ B =& ((-1+(-1+(1-\eta_u)^{-\delta}+(1-\eta_v)^{-\delta})^{1/\delta})^2 \\ & ((1-\eta_u)^{\delta}+(1-\eta_v)^{\delta}-(1-\eta_u)^{\delta}(1-\eta_v)^{\delta})^2 \end{aligned}$$

where $\eta_u = (1-u)^\theta$ and $\eta_v = (1-v)^\theta$. Note that the density of the survival Joe-Clayton copula was replaced by that of the Joe-Clayton copula but with u and v replaced by $1-u$ and $1-v$.

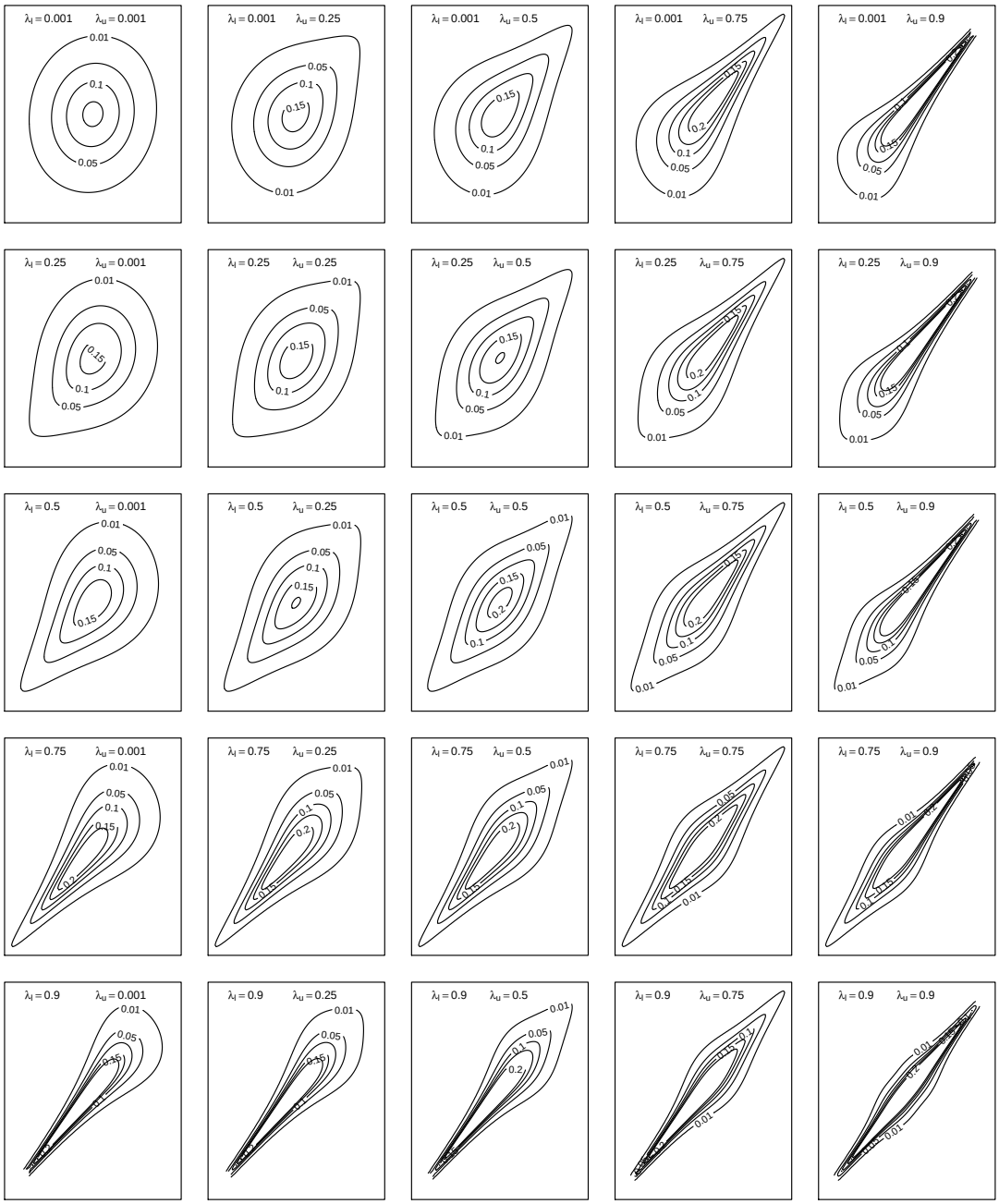


Figure A.10: Symmetrized Joe-Clayton copula $C_{\mathrm{SJC};\lambda_l,\lambda_u}$ with standard normal margins

In Figure A.10 the contour plots of the densities of bivariate distributions with symmetrized Joe-Clayton copulas and standard normal margins are depicted. The range of tail dependence is $\Lambda = \{0.01, 0.25, 0.5, 0.75, 0.9\}$ with all possible combinations $(\lambda_l, \lambda_u) \in \Lambda \times \Lambda$. For clarity of the illustrations we refrained from displaying the ranges of the x- and y-axis. These were $[-3, 3]$ respectively.

A.9 Summary

Let us summarize the characteristics derived above for the presented bivariate copula families.
In Table A.1 the parameter range, the respective concordance measures (τ or ρ) and the tail dependence coefficients for the introduced non-Archimedean copula families such as Gaussian, Student, Plackett and Marshall-Olkin copulas are given.
In Table A.2 we sum up what we know about the Archimedean copula families; here the parametric Archimedean generator φ_θ and its pseudo-inverse $\varphi_\theta^{[-1]}$, the parameter space Θ, Kendall's concordance coefficient τ as well as the lower and upper tail dependence coefficients λ_l and λ_u are tabulated.

Copula class	parameter range	concordance measure	tail dependence
Gaussian $C_{\mathrm{Ga};\rho}$	$\rho \in (-1,1)$	$\tau = 2/\pi \arcsin \rho$	$\lambda_l = \lambda_u = 0$
Student $C_{t_\nu;\rho}$	$\rho \in (-1,1), \nu > 2$	$\tau = 2/\pi \arcsin \rho$	$\lambda_l = \lambda_u = 2t_{\nu+1}(-\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho})$
Plackett $C_{\mathrm{P};\theta}$	$\theta \in (0,\infty)$	$\rho = \frac{\theta+1}{\theta-1} - \frac{2\theta\log(\theta)}{(\theta-1)^2}$	$\lambda_l = \lambda_u = 0$
Marshall-Olkin $C_{\mathrm{MO};\alpha_1,\alpha_2}$	$\alpha_i \in [0,1]$, $i = 1,2$	$\tau = \frac{\alpha_1\alpha_2}{\alpha_1-\alpha_1\alpha_2+\alpha_2}$, $\rho = \frac{3\alpha_1\alpha_2}{2\alpha_1-\alpha_1\alpha_2+2\alpha_2}$	$\lambda_l = 0$, $\lambda_u = \alpha_1 \wedge \alpha_2$

Table A.1: Characteristics of non-Archimedean copula families

Name	$\varphi_\theta(t)$	$\varphi_\theta^{[-1]}(t)$	Θ	τ	λ_l	λ_u
Clayton	$(t^{-\theta}-1)/\theta$	$(1+\theta t)^{-1/\theta}$	$[-1,\infty)\backslash\{0\}$	$\theta/(\theta+2)$	$2^{-1/\theta}$	0
Gumbel	$(-\log t)^\theta$	$\mathrm{e}^{-t^{1/\theta}}$	$[1,\infty)$	$(\theta-1)/\theta$	0	$2-2^{1/\theta}$
Frank[a]	$-\log\left(\frac{\mathrm{e}^{-\theta t}-1}{\mathrm{e}^{-\theta}-1}\right)$	$-\log(\mathrm{e}^{-t}(\mathrm{e}^{-\theta}-1)+1)/\theta$	$\mathbb{R}\backslash\{0\}$	$1+4(D_1(\theta)-1)/\theta$	0	0
Joe-Clayton[b]	$\left(1-(1-t)^\theta\right)^{-\delta}-1$	$1-(1-(t+1)^{-1/\delta})^{1/\theta}$	$[1,\infty)\times(0,\infty)$	$1-\frac{2}{\delta(2-\theta)}+\frac{4}{\theta^2\delta}B\left(\frac{2-\theta}{\theta},\delta+2\right)$	$2^{-1/\theta}$	$2-2^{1/\theta}$

Table A.2: Characteristics of Archimedean copula families

[a] $D_1(\theta) = \int_0^\theta t/(\mathrm{e}^t-1)dt/\theta$ (Debye function of order 1)
[b] $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$ (Beta function/Euler integral of the first kind)

B Conditions for Asymptotic Properties in the SCOMDY Model

Following the lines of Chan et al. (2009a) a concise list of conditions being sufficient for consistency and asymptotic normality of the maximum likelihood parameter estimators introduced in Sections 2.2.2 and 3.2.2 will be given.
First, we need conditions on the residuals ε_{tj} in the GARCH models for every component j. In particular, these are necessary for studying the empirical and the weighted empirical process of the estimated residuals in the considered time series models.

(A1) For $j = 1, \dots, d$, $\gamma_j \in \Gamma_j$ (see equation (3.1.15)), $\mathbb{E}\,\varepsilon_j^4 < \infty$ and there exists $\mu > 0$ such that

$$\lim_{t \to 0} t^{-\mu} \mathbb{P}\left(\varepsilon_j^2 \leq t\right) = 0.$$

(A2) For $j = 1, \dots, d$ the support of ε_{jt} is $(-\infty, \infty)$, F_j has a continuous density F_j', and there exist $\beta_3 \in (0, 1/4)$ and $\Delta_3 > 0$ such that

$$\sup_s \sup_{|x-1| \leq \Delta_3} \frac{s F_j'(sx)}{(F_j(x)(1 - F_j(x)))^{\beta_3}} < \infty$$

for all $j = 1, \dots, d$.

The next set of assumptions concerns the likelihood function for estimating the copula parameter θ and will ensure consistency of the resulting estimates.

(C1) The log-likelihood function $l(u_1, \dots, u_d; \theta) = \log c(u_1, \dots, u_d; \theta)$ is a continuous function of θ for all $(u_1, \dots, u_d) \in [0,1]^d$.

(C2) The parameter space Θ is a compact subspace of $\mathbb{R}^m, m \geq 1$.

(C3) $\mathbb{E} \sup_{\theta \in \Theta} |\, l(F_1(\varepsilon_1), \dots, F_d(\varepsilon_d); \theta)| < \infty$.

(C4) For any $\Delta_0 \in (0, 1/2)$ and $\Delta_1 \in (1/2, 1)$, there exist $\beta_0 \in [0,1]$, $M_0 > 0$, $\beta_1 > 0$ and $M_1 > 0$ such that for $\boldsymbol{u} = (u_1, \dots, u_d)$, $\boldsymbol{v} = (v_1, \dots, v_d)$

$$\sup_{\theta \in \Theta} |\, l(\boldsymbol{u}; \theta)| \leq M_0 (\min_j u_j)^{-\beta_0}$$

for $\min_j u_j \leq \Delta_0$,

$$\sup_{\theta \in \Theta} |\, l(\boldsymbol{u}; \theta)| \leq M_0 (1 - \max_j u_j)^{-\beta_0}$$

for $\max_j u_j \geq \Delta_1$, and

$$\sup_{\theta\in\Theta} |\, l(\boldsymbol{u};\theta) - l(\boldsymbol{v};\theta)| \leq M_1 \sum_{j=1}^{d} |u_j - v_j|^{-\beta_1}$$

for $\Delta_0/2 < \min_j u_j \leq \max_j u_j < \Delta_1 + (1-\Delta_1)/2$ and $\Delta_0/2 < \min_j v_j \leq \max_j v_j < \Delta_1 + (1-\Delta_1)/2$.

Note that conditions (C1) to (C3) are the standard conditions for consistency of maximum likelihood estimators.
Condition (C4) is similar to that in Genest et al. (1995); it controls the speed of divergence of the log-likelihood function at the boundaries, and is satisfied by all the copula densities presented in Appendix A.

Before we state the conditions for asymptotic normality of the copula parameter estimators we will introduce some additional notations. Consider the following gradient and partial derivatives

$$\nabla_\theta\, c(\,.\,;\theta) = \left(\frac{\partial}{\partial\theta_1} c(\,.\,;\theta), \ldots, \frac{\partial}{\partial\theta_m} c(\,.\,;\theta)\right),\ \delta(\,.\,;\theta) = \frac{\nabla_\theta\, c(\,.\,;\theta)}{c(\,.\,;\theta)},$$

and for all $j = 1, \ldots, d$

$$\delta_j(u_1, \ldots, u_d;\theta) = \frac{\partial}{\partial u_j}\, c(u_1, \ldots, u_d;\theta). \tag{B.0.1}$$

Furthermore, for $j = 2, \ldots, d$ consider the following conditional marginal distributions

$$\begin{aligned} C_1(u_1) &= u_1, \\ C(u_j|u_1, \ldots, u_{j-1}) &= \mathbb{P}\left(F_j(\varepsilon_j) \leq u_j \,|F_1(\varepsilon_1) \leq u_1, \ldots, F_{j-1}(\varepsilon_{j-1}) \leq u_{j-1}\right) \end{aligned}$$

and the vector

$$T_j(u_1, \ldots, u_j) = (C_1(u_1), C_2(u_2|u_1), \ldots, C(u_j|u_1, \ldots, u_{j-1})) \tag{B.0.2}$$

as well as the partial derivatives for $i, j = 1, \ldots, m$

$$l_{ij}(\,.\,;\theta) = \frac{\partial^2}{\partial\theta_i\,\partial\theta_j} l(\,.\,;\theta).$$

Finally, define the Fisher information matrix $\Sigma(\theta)$ in the fully parametric model as

$$\Sigma(\theta) = \left(\mathbb{E}\left(\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log c(F_1(\varepsilon_1), \ldots, F_d(\varepsilon_d);\theta)\right)\right)_{i,j=1,\ldots,m}. \tag{B.0.3}$$

Then the conditions for asymptotic normality are

(N1) For all $j = 2, \ldots, d$ the function $C_j(u_j|T_{j-1}^-(u_1, \ldots, u_{j-1}))$ is differentiable with respect to $u_1, \ldots, u_{j-1}$ in the interior of $[0,1]^{j-1}$, and

$$\sum_{i=1}^{j-1} \int_{[0,1]^{j-1}} \left| \frac{\partial}{\partial u_i} C_j(u_j|T_{j-1}^-(u_1, \ldots, u_{j-1})) \right| du_1 \ldots du_{j-1} \leq M_2 \in (0, \infty).$$

(N2) There exists $\beta_2 \in (0, 1/2)$ such that for $\boldsymbol{u} = (u_1, \ldots, u_d)$

$$\sup_{\boldsymbol{u} \in [0,1]^d} \prod_{j=1}^{d} u_j^{\beta_2} (1 - u_j)^{\beta_2} |T(\boldsymbol{u}; \theta_0)| < \infty,$$

$$\int_{[0,1]^d} \prod_{j=1}^{d} u_j^{\beta_2} (1 - u_j)^{\beta_2} |dT(\boldsymbol{u}; \theta_0)| < \infty.$$

(N3) For any $\Delta_4 \in (0, 1/2)$ and $\Delta_5 \in (1/2, 1)$, there exist $\beta_4 \in (0, \beta_2)$, $M_4 > 0$, $\beta_5 > 0$ and $M_5 > 0$ such that for $\boldsymbol{u} = (u_1, \ldots, u_d)$, $\boldsymbol{v} = (v_1, \ldots, v_d)$

$$|\delta_j(\boldsymbol{u}; \theta_0)| \leq M_4 u_j (\min_i u_i)^{-\beta_4}$$

for $\min_i u_i \leq \Delta_4$,

$$|\delta_j(\boldsymbol{u}; \theta_0)| \leq M_4 (1 - u_j)(1 - \min_i u_i)^{-\beta_4}$$

for $\max_i u_i \geq \Delta_5$, and

$$|\delta(\boldsymbol{u}; \theta_0) - \delta(\boldsymbol{v}; \theta_0)| \leq M_5 \sum_{j=1}^{d} |u_j - v_j|^{\beta_5}$$

for $\Delta_4/2 < \min_j u_j \leq \max_j u_j < \Delta_5 + (1 - \Delta_5)/2$ and $\Delta_4/2 < \min_j v_j \leq \max_j v_j < \Delta_5 + (1 - \Delta_5)/2$.

(N4) For $i, j = 1, \ldots, m$ the partial derivatives $l_{ij}(\,.\,; \theta)$ are continuous functions of θ in an open neighborhood of θ_0 for all $\boldsymbol{u} \in [0,1]^d$.

(N5) There exists an open neighborhood Θ_0 of θ_0 such that, for $i, j = 1, \ldots, m$

$$\mathbb{E} \sup_{\theta \in \Theta_0} |l_{ij}(F_1(\varepsilon_1), \ldots, F_d(\varepsilon_d); \theta)| < \infty.$$

(N6) For any $\Delta_6 \in (0, 1/2)$ and $\Delta_7 \in (1/2, 1)$, there exist $\beta_6 \in [0,1]$, $M_6 > 0$, $\beta_7 > 0$ and $M_7 > 0$ such that for $\boldsymbol{u} = (u_1, \ldots, u_d)$, $\boldsymbol{v} = (v_1, \ldots, v_d)$

$$\sup_{\theta \in \Theta_0} |\, l_{ij}(\boldsymbol{u}; \theta)| \leq M_6 (\min_j u_j)^{-\beta_6}$$

for $\min_j u_j \leq \Delta_6$,

$$\sup_{\theta \in \Theta_0} |\, l_{ij}(\boldsymbol{u}; \theta)| \leq M_6 (1 - \max_j u_j)^{-\beta_6}$$

for $\max_j u_j \geq \Delta_7$, and

$$\sup_{\theta \in \Theta_0} |\, l_{ij}(\boldsymbol{u}; \theta) - l_{ij}(\boldsymbol{v}; \theta)| \leq M_7 \sum_{j=1}^{d} |u_j - v_j|^{-\beta_1}$$

for $\Delta_6/2 < \min_j u_j \leq \max_j u_j < \Delta_7 + (1 - \Delta_7)/2$ and $\Delta_6/2 < \min_j v_j \leq \max_j v_j < \Delta_7 + (1 - \Delta_7)/2$.

Again the standard conditions for asymptotic normality are contained as Condition (N4) and (N5) above. Textbook references for all standard (C) and (N) conditions are for example Casella and Berger (2001, Chap.6.3) and Lehmann and Casella (2003, Thm.2.6). Conditions (N1) and (N2) are introduced by Csörgő and Révész (1975) for multivariate empirical processes. Conditions (N3) and (N6) are similar to those in Genest et al. (1995) imposed again to control the speed of divergence at the boundaries.

Based on these assumptions, three results concerning the asymptotic properties of the estimators in copula-based multivariate GARCH models can be derived. The exact formulations can be found in Theorems 2.1, 2.2 and 2.3 in Chan et al. (2009a), we will only give heuristic explanations of the results. A supplementary personal comment will be given at the end.

In Theorem 2.1 it is shown that under (A1) and (A2) the empirical process of the estimated residuals $\hat{F}_j(x) - F(x)$ converges to a Kiefer process at the rate $\mathcal{O}(\sqrt{T})$ and a certain variance disturbance has to be taken into account.

The consistency result corresponds to Theorem 2.2 under (A1) and (C1)-(C4) which states that the maximum likelihood estimator $\hat{\theta}$ based on a pseudo-sample of residuals converges in probability to the true parameter value θ_0 as the sample size grows to infinity.

Finally, in their Theorem 2.3 it is established under (A1)-(A2), (C1)-(C4) and (N1)-(N6) that $\sqrt{T}(\hat{\theta} - \theta_0)$ converges in distribution to a normal distribution with a covariance structure as in Genest and Rivest (1993) that is in particular independent of the GARCH filtering.

Unfortunately, we were neither able to verify the details in the theoretical proof given in the online supplement Chan et al. (2009b) nor to reproduce the simulation results as stated therein. The main problem was that the variance that we observed in the empirical analyses was greater than it should have been according to the theoretical result. At best, this is a problem due to the finite sample set-up that empirical studies bear. However, simulations with different sample sizes as well as the results of Berkes and Horváth (2003) similar to those in Theorem 2.1 do not support their argument. More likely, it is an inherent problem of the GARCH estimation that cannot be remedied.

C Conditions for Yaskov's Asymptotic Result

Yaskov (2010b) is part of the author's dissertation (see Section 2.3 in Yaskov, 2010a). We work under the assumption that two competing models are compared, i.e. f_t is defined as

$$f_t(\theta) = L_1(x_t;\theta_1) - L_2(x_t;\theta_2),\ \theta = (\theta_1,\theta_2) \in \mathbb{R}^2,$$

and furthermore, that for $\theta_t = (\theta_{1t},\theta_{2t})$ the estimators are determined as the solution to an extremal problem, i.e. $\hat{\theta} = (\hat{\theta}_{1t},\hat{\theta}_{2t})$ with

$$\hat{\theta}_{it} = \underset{\theta_i\in\mathbb{R}}{\operatorname{argmax}} \sum_{s=t-b+1}^{t} \phi_i(x_s;\theta_i)$$

for an estimation function ϕ_i, $i = 1,2$ and iid data $x_1,\dots,x_T$. For S'_{Tb} as given in (4.4.14) consider the following set of conditions.

(Y1) $\theta = (\theta_1,\theta_2)$ lies in a compact set $\Theta \subseteq \mathbb{R}^2$.

(Y2) ϕ_1, ϕ_2, f_t are three times differentiable almost everywhere on $\Theta \subseteq \mathbb{R}^2$.

(Y3) The absolute moments of order p of

$$\sup_\theta |f_t(\theta)|^{1/p},\ f_t(\theta_t),\ \frac{\partial^k}{\partial\theta_i^k} f_t(\theta_t),\ \sup_\theta \left|\frac{\partial^3}{\partial\theta_i^3} f_t(\theta_t)\right|,\ \sup_{\theta_i} \left|\frac{\partial^k}{\partial\theta_i^k}\phi_i(x_t;\theta_i)\right|$$

are uniformly bounded in t for all $i = 1,2$, $k = 0,1,2$ and some $p > 3$.

(Y4) For sufficiently small $\epsilon > 0$ it holds for all t and i

$$\sup_{\|\bar{\theta}_i - \theta_{it}\|>\epsilon} \left(\mathbb{E}\,\phi_i(x_t;\bar{\theta}_i) - \mathbb{E}\,\phi_i(x_t;\theta_{it})\right) > 0.$$

(Y5) The function

$$\psi_i(\epsilon) = \max_{k=0,1} \sup_{t,q_i} \left| \mathbb{E} \sup_{\|\theta_i - q_i\|\le\epsilon} \frac{\partial^k}{\partial\theta_i^k}\phi_i(x_t;\theta_i) - \mathbb{E} \inf_{\|\theta_i - q_i\|\le\epsilon} \frac{\partial^k}{\partial\theta_i^k}\phi_i(x_t;\theta_i) \right|$$

decreases to 0 as $\epsilon \to \infty$.

(Y6) The number of different time intervals of the form $[a, b]$ on which $\{x_t\}_{t=1}^T$ is strictly stationary, namely $l = l(T)$, and b satisfy the relation

$$l = o(\sqrt{T}/b) \quad \text{as} \quad T \to \infty.$$

(Y7) $\{x_t\}_{t=1}^\infty$ satisfies the strong mixing condition and has mixing coefficients α_T of order $\mathcal{O}(T^{3p/(2p-6)})$.

The strict stationarity of $\{x_t\}_{t=1}^T$ in condition (Y6) can be relaxed by assuming covariance stationarity of the vectors of quantities of the form $h(x_t; \theta_t)$ where h is taken as f, $\frac{\partial f}{\partial \theta}$, ϕ_i, $\frac{\partial \phi_i}{\partial \theta}$.

With further notation

$$g_t = g_t(\theta_t), \quad \partial g = \frac{\partial g}{\partial \theta}, \quad \partial^1(g_1, g_2)^\top = \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} \\ \frac{\partial g_2}{\partial \theta_2} \end{pmatrix}, \quad \partial^i(g_1, g_2)^\top = \begin{pmatrix} \frac{\partial^i g_1}{\partial \theta_1^i} & 0 \\ 0 & \frac{\partial^i g_2}{\partial \theta_2^i} \end{pmatrix}, \; i = 2, 3,$$

$$\phi_t(\theta) = \begin{pmatrix} \phi_1(x_t; \theta_1) \\ \phi_2(x_t; \theta_2) \end{pmatrix}, \quad \begin{pmatrix} a \\ b \end{pmatrix}^* = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \quad u_t = [\mathbb{E}\, \partial^2 \phi_t]^{-1} \partial^1 \phi_t, \quad w_t = \mathbb{E}\, \partial^2 f_{t+1} \cdot u_t,$$

$$v_t = \mathbb{E}\, \partial^2 f_{t+1} [\mathbb{E}\, \partial^2 \phi_t]^{-1} \partial^2 \phi_t, \quad z_t = \mathbb{E}\, \partial^2 f_{t+1} [\mathbb{E}\, \partial^2 \phi_t]^{-1} u_t^* \mathbb{E}\, \partial^3 \phi_t, \quad h_t = \mathbb{E}\, \partial f_{t+1} \cdot u_t$$

we can finally establish the expression for the mean and the variance term in Theorem 4.4.2 as

$$\begin{aligned} \mu_{Tb} = \frac{1}{b\sqrt{T}} \sum_{k=-b}^{b} \left(1 - \frac{|k|}{b}\right) \sum_{t=b \vee (b-k)}^{(T-1) \wedge (T+k-1)} \left(\operatorname{Cov}(v_t, u_{t-k}^\top) + \frac{\operatorname{Cov}(w_t^\top, u_{t-k}^\top) - \operatorname{Cov}(z_t, u_{t-k}^\top)}{2} \right) \\ - \frac{1}{b\sqrt{T}} \sum_{k=0}^{b} \sum_{t=b \vee (b-k)}^{(T-1) \wedge (T+k-1)} \operatorname{Cov}(\partial f_{t+1}, u_{t-k}^\top), \end{aligned} \tag{C.0.1}$$

and

$$\begin{aligned} \sigma_{Tb}^2 = \frac{1}{T} \sum_{s,t=b}^{T-1} \operatorname{Cov}(f_{t+1} - h_t, f_{s+1} - h_s) + \frac{1}{T} \sum_{s,t=1}^{b-1} \frac{ts}{b^2} \operatorname{Cov}(h_t, h_s) \\ + \frac{1}{T} \sum_{s,t=1}^{b-1} \frac{ts}{b^2} \operatorname{Cov}(h_{T-t}, h_{T-s}) - \frac{1}{T} \sum_{s,t=1}^{b-1} \frac{s}{b} \operatorname{Cov}(f_{T-t+1} - h_{T-t}, h_{T-s}). \end{aligned} \tag{C.0.2}$$

We assume the well-definedness of the occurring expression such that $\liminf_{T,b} \sigma_{Tb}^2 > 0$.

Zusammenfassung

Das Hauptziel der vorliegenden Arbeit ist es, zu untersuchen, ob sich die Abhängigkeitsbeziehungen zwischen den einzelnen Komponenten mehrdimensionaler Zeitreihen von Finanzmarktdaten im Zeitverlauf verändern oder ob diese als zeithomogen angesehen werden können.
Zunächst einmal ist die Modellierung multivariater Verteilungen keine triviale Aufgabe. Es wird selten der Fall sein, dass eine ad hoc definierte Familie von Funktionen alle Voraussetzungen an multivariate Verteilungsfunktionen erfüllt (siehe Joe, 1997, S. 3), mag sie die Abhängigkeiten noch so gut abbilden. Dieses Problem tritt bei der Verwendung einer wohlbekannten Verteilungsfamilie wie z. B. der multivariaten Normalverteilung nicht auf; dass diese die geforderten Grenzwert-, Monotonie- und Stetigkeitseigenschaften (siehe Joe, 1997, Abschnitt 1.4.2) erfüllt, lässt sich direkt nachprüfen. Es ist eher die Tatsache, dass die Daten, die man typischerweise an Finanzmärkten beobachtet, in diesem Sinne *nicht-normale* Eigenschaften aufweisen, die der Verwendung der Normalverteilung entgegenstehen. So werden beispielsweise gemeinsame Kursverluste im Allgemeinen häufiger beobachtet als gemeinsame Kursgewinne, was auf eine asymmetrische Abhängigkeitsstruktur hindeutet, die mithilfe der elliptischen, also radialsymmetrischen, Normalverteilung nicht modelliert werden kann.
Beschränkt man sich bei der Quantifizierung von Abhängigkeiten mittels einer reellen Kennzahl auf den Korrelationskoeffizienten nach Pearson, so misst man zum einen nur die Stärke der linearen Korrelation; diese Betrachtungsweise ist oftmals nicht ausreichend, um multivariate Phänomene zu erklären. Andererseits deckt die Pearson-Korrelation unter Umständen nicht ihren gesamten Wertebereich $[-1,1]$, sondern nur eine viel kleinere Bandbreite an möglichen Abhängigkeiten ab; dies ist unter anderem darauf zurückzuführen, dass diese Abhängigkeitsmaßzahl nicht nur von der gemeinsamen Verteilung, sondern auch von den entsprechenden Randverteilungen abhängt.
An dieser Stelle erweist sich das wohl bekannteste Resultat moderner Copula-Theorie, nämlich Sklars Theorem, als äußerst hilfreich: Sklar (1959) hat gezeigt, dass sich jede d-dimensionale Verteilungsfunktion H folgendermaßen zerlegen lässt

$$H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)),$$

wobei $(x_1, \ldots, x_d) \in \mathbb{R}^d$, $F_j \colon \mathbb{R} \to [0,1]$ eine univariate Verteilungsfunktion für jedes $j = 1, \ldots, d$ und $C \colon [0,1]^d \to [0,1]$ eine Copula sind. Eine Copula ist hierbei selbst eine d-variate Verteilungsfunktion auf dem Einheitskubus $[0,1]^d$, deren univariate Ränder alle stetige Gleichverteilungen auf $[0,1]$ sind. Der darauf basierende, zweistufige Ansatz, dass aus beliebigen gegebenen Randverteilungen und einer gegebenen Copula eine zulässige gemeinsame Verteilung konstruiert werden kann, führt zu einer enormen Flexibilisierung multivariater Modelle. Durch die getrennte Betrachtung von Randverteilungen und

der Abhängigkeitsstruktur, die in der entsprechenden Copula abgebildet wird, können alternative Maßzahlen für die Abhängigkeit hergeleitet werden. Diese sind oftmals nicht nur unabhängig von den Rändern und somit nicht eingeschränkt, was ihren Wertebereich angeht, sondern ermöglichen es auch beispielsweise die Wahrscheinlichkeit für gemeinsame Extrema zu quantifizieren.

Es ist vor allem dieser Flexibilität geschuldet, dass Copulas im vergangenen Jahrzehnt einen enormen Popularitätszuwachs erfahren haben. Mikosch (2006) berichtet, dass eine Google-Anfrage nach dem Wort *copula* im Jahre 2003 gerade einmal 10.000 Treffer lieferte. 2005 ergab dieselbe Suche schon 650.000 Treffer, und 2009 berichteten Genest et al. (2009a) von 750.000 Suchergebnissen. Während diese Arbeit entsteht, ergibt eine ähnliche Suche (siehe Google search) schon über 1,8 Millionen Einträge.

Die sehr aktive Forschung im Bereich der Copula-Methodik für die Modellierung multivariater Fragestellungen hat grundlegende mathematische Ergebnisse im Bereich der Wahrscheinlichkeitstheorie und statistischen Inferenz hervorgebracht. Für erstere findet sich bei Sklar (1996) ein schöner Überblick, zu letzterer haben Genest et al. (1995) einen wichtigen Originalbeitrag geliefert.

Ursprünglich stammen die behandelten Fragestellungen aus den Naturwissenschaften, weswegen die Anwendungsgebiete von Copulas beispielsweise Hydrologie (siehe Genest and Favre, 2007 oder Salvadori, 2007) und Genetik (siehe Cui and Sun, 2004) umfassen. Ihren rasanten Aufstieg verdanken sie jedoch der Tatsache, dass in Hinblick auf statistische Auswertungen vor allem bei Wirtschaftsfragen eine sehr günstige Datensituation vorliegt, d.h. für Anwendungen in diesem Bereich gibt es eine ausreichend umfangreiche Datenbasis. Copulas lassen sich somit auf finanz- und versicherungsmathematische Probleme z. B. beim Risikomanagement von Portfolios (siehe McNeil et al., 2005 oder Cherubini et al., 2004) oder bei der Abhängigkeitsmodellierung von aktuariellen Größen (siehe Frees and Valdez, 1998) anwenden. Genest et al. (2009a) untersuchen, wie Copulas im Finance-Bereich so schnell an Bedeutung gewinnen konnten, indem sie die zugehörige akademische Literatur sichten. Dabei identifizieren sie den ersten Entwurf von Embrechts et al. (2002) als Diskussionspapier im Jahre 1998 als den Ausgangspunkt für die Anwendung von Copulas auf multivariate finanzmathematische Fragestellungen.

Trotz allem darf Copula-Theorie nicht als Königsweg für mehrdimensionale Modellierung angesehen werden. Hierbei tauchen fraglos auch Nachteile auf. Mikosch (2006) hat mit dem Hinweis auf eventuelle Schwierigkeiten bei multivariater Extremwert- und Zeitreihenmodellierung mithilfe von Copulas eine lebhafte Debatte angestoßen. Embrechts (2006) erwiderte daraufhin, dass Copulas mit Sicherheit keinen universellen Ansatz für ein globales, dynamisches multivariates Modell liefern, sich aber als äußerst nützlich erweisen, wenn wegen nicht direkt beobachtbarer Abhängigkeitsstrukturen an die (Modell-)Grenzen stößt (S. 46 ebd.).

Genau an dieser Stelle setzt der vorliegende Forschungsbeitrag an. Auf der einen Seite soll diese Arbeit zu einem besseren Verständnis der Abhängigkeiten zwischen den Komponenten in realen Datenreihen beitragen. Zum anderen soll untersucht werden, welche Implikationen für die verwendeten Methoden daraus resultieren, dass auf multivariaten Zeitreihendaten, anstatt auf reinen Copula-Daten gerechnet wird. Vor allem die Inhomogenität von Copula-Parametern über die Zeit soll analysiert und getestet

werden. Welche Möglichkeiten es für zeitdynamische Modelle gibt, ist in Patton (2012) dargestellt.
Das Konzept der Copulas wird in der vorliegenden Dissertation in der sogenannten räumlichen Dimension angewendet, d. h. um die Abhängigkeit zwischen einzelnen Zeitreihen, deren Beobachtungen an gemeinsamen Zeitpunkten vorliegen, zu modellieren. Die Verwendung in der zeitlichen Dimension, d. h. um die Abhängigkeit einer einzelnen Zeitreihe an verschiedenen Zeitpunkten zu erfassen, ist prinzipiell auch möglich (siehe Patton, 2006), wird jedoch als „merkwürdig“ empfunden (siehe Cherubini et al., 2011). Dennoch kann sich die so modellierte räumliche Abhängigkeitsstruktur im Zeitverlauf ändern. Die Komponenten einer multivariaten Zeitreihe sind selbst (univariate) Zeitreihen, d.h. sie erfassen die zeitliche Entwicklung heteroskedastischer Finanzdaten, und im vorliegenden Fall wird ihre Abhängigkeit durch die multivariate Verteilung der Residuen im SCOMDY-Modell erfasst, insbesondere durch die parametrische Copula dieser gemeinsamen Verteilung. An dieser Stelle wird die Forderung an die multivariaten Residuen, für jeden Zeitpunkt unabhängig und identisch verteilt zu sein, abgeschwächt, indem man zulässt, dass die Copula-Parameter zeitdynamisch sind. Dies hat zur Folge, dass sich durch den zeitvariablen Parameter auch die Abhängigkeit ändert und manchmal stärker, manchmal schwächer ist, je nachdem an welchem Zeitpunkt man diese misst. Für die folgenden Analysen wird kein spezifisches Modell für die zeitliche Entwicklung der Copula-Parameter zugrunde gelegt, sondern es werden Homogenitätstests durchgeführt, die auf einem rein datengetriebenen Ansatz beruhen.
Die Arbeit ist wie folgt gegliedert. In Kapitel 2 werden die grundlegenden Definitionen und Theoreme im Zusammenhang mit der Copula-Theorie unter analytischen Gesichtspunkten vorgestellt. Sklars Theorem bildet dann die Grundlage für viele weitere Erkenntnisse, z. B. was Konstruktions- und Simulationstechniken angeht, die kurz besprochen werden. Das Hauptaugenmerk liegt jedoch auf geeigneten Schätzverfahren auf Basis multivariater Daten, hierbei wird insbesondere auf Methoden der Maximum-Likelihood-Schätzung eingegangen. Dieses einführende Kapitel wird mit einer kurzen Übersicht über Extremwerttheorie, insbesondere solche für schwach abhängige Folgen, abgeschlossen.
In Kapitel 3 werden heteroskedastische Zeitreihenmodelle ausführlich diskutiert. Zunächst werden die bekannten Resultate zu Stationarität, endlich-dimensionalen Verteilungen und geeigneten Parameterschätzern im univariaten GARCH-Modell präsentiert. Aufgrund seiner speziellen Struktur übertragen sich viele der univariaten Ergebnisse auf das SCOMDY-Modell von Chan et al. (2009a), das den weiteren Betrachtungen zugrunde liegt. Durch den Vergleich mit anderen multivariaten Modellen für Heteroskedastizität gelingt es im diesem Fall, die Tails der endlich-dimensionalen Verteilungen herzuleiten. Eine Diskussion zeitdynamischer Copula-Modelle schließt dieses Kapitel ab.
Kapitel 4 enthält den Originalbeitrag der Arbeit, der dem Ziel dient, zeitinhomogene Abhängigkeiten in multivariaten Zeitreihen aufzudecken. Eine eventuelle Zeitdynamik wird dabei als dynamischer Parameter der zugrunde liegenden Copula-Familie im SCOMDY-Modell modelliert. Eine Schätzerfolge von lokalen Schätzern, die auf sich überlappenden Teilstichproben bestimmt werden, bildet die Basis der weiteren Untersuchungen. Die gemeinsame asymptotische Verteilung der lokalen Schätzer wird dann genutzt, um einen

Binomialtest, einen Extremaltest sowie einen Test zu konstruieren, der die gesamte Schätzerfolge berücksichtigt. Alle drei Tests werden hinsichtlich ihrer lokalen Power und der Bandbreite an detektierbaren Alternativen untersucht. Die Vor- und Nachteile, die sich dabei herausstellen, werden abschließend diskutiert.
Die erste empirische Analyse in Kapitel 5 stellt eine Simulationsstudie dar, die vor allem dazu dient, die empirische Size des Binomialtests festzustellen. Dabei zeigt sich, dass ein Bootstrap zur Bestimmung der Schätzervarianz bei den endlichen Stichprobenumfängen unumgänglich ist. Mit dieser Erkenntnis wird der Binomial- und der Extremaltest auf reale Finanzdaten angewendet. Dabei stellt sich heraus, dass sich die Abhängigkeiten zwischen den einzelnen Zeitreihen immer dann zu ändern scheinen, wenn ein ungewöhnliches reales Ereignis an den Finanzmärkten auftritt. Diese Parameterinhomogenitäten können jedoch in einigen Fällen behoben werden, indem man in den univariaten Modellen unterschiedliche Regimes zulässt. Dies stimmt mit der allgemeinen Wahrnehmung überein, dass sich die Abhängigkeiten eher zeitkonstant verhalten und Änderungen nur über sehr kurze, abgegrenzte Zeiträume zu beobachten sind, in denen extreme wirtschaftliche Ereignisse auftreten. In den meisten Fällen genügt es, eine zeitkonstante Copula zu wählen, und man gewinnt nichts Wesentliches hinzu, wenn ein komplizierteres und statistisch unhandliches Modell mit zeitdynamischen Parametern gewählt wird.
Obwohl das probabilistische Konzept von Copulas noch relativ jung ist, steht es momentan im Zentrum der Aufmerksamkeit einer ganzen Forschergeneration. Im Geiste des berühmten Ausspruchs „Wenn ich weiter gesehen habe als andere, so deshalb, weil ich auf den Schultern von Riesen stehe“, der in dieser Form Isaac Newton zugesprochen wird, ist diese Dissertation ein kleiner Beitrag, um mit dem Wissen um so profunde Theorie wie Maximum-Likelihood-Schätzung, Extremwerttheorie und statistische Inferenztechniken weiter zu sehen.

Bibliography

Aldrich, J. (1997). R. A. Fisher and the Making of Maximum Likelihood 1912 – 1922. *Statistical Science*, 12(3):162–176.

Almeida, C., Czado, C., and Manner, H. (2012). Modeling high dimensional time-varying dependence using D-vine SCAR models. http://arxiv.org/pdf/1202.2008v1. [Online; accessed 13-07-2012].

Alsina, C., Schweizer, B., and Frank, M. (2006). *Associative functions: Triangular norms and copulas*. World Scientific.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858.

Aptech Systems (http://www.aptech.com/). The GAUSS programming language. [Online; accessed 07-07-2012].

Bahadur, R. R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics*, 37(3):577–580.

Bai, J. and Perron, P. (1998a). Computation and analysis of multiple structural-change models. Cahier 9807, Université de Montréal.

Bai, J. and Perron, P. (1998b). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.

Bai, J. and Perron, P. (2003). Critical values for multiple structural change tests. *Econometrics Journal*, 6:72–78.

Basrak, B., Davis, R. A., and Mikosch, T. (2002). Regular variation of GARCH processes. *Stochastic Processes and their Applications*, 99(1):95–115.

Bauwens, L., Laurent, S., and Rombouts, J. V. K. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21:79–109.

Bentkus, V. (2004). A Lyapunov-type bound in $\mathbb{R}^d$. *Theory of Probability and its Applications*, 49(2):311–323.

Berenhaut, K. and Bandyopadhyay, D. (2005). Monotone convex sequences and Cholesky decomposition of symmetric Toeplitz matrices. *Linear Algebra and Its Applications*, 403:75–85.

Berg, D. (2009). Copula goodness-of-fit testing: An overview and power comparison. *The European Journal of Finance*, 15(7-8):675–701.

Berkes, I. and Horváth, L. (2003). Limit results for the empirical process of squared residuals in GARCH models. *Stochastic Processes and their Applications*, 105:271–298.

Berkes, I., Horváth, L., and Kokoszka, P. (2003). GARCH processes: Structure and estimation. *Bernoulli*, 9(2):201–227.

Bernstein, S. (1929). Sur les fonctions absolument monotones. *Acta Mathematica*, 52(1):1–66.

Bloomberg Professional® (http://www.bloomberg.com/professional). [Online; accessed 04-07-2012].

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, 72(3):498–505.

Bollerslev, T. (2008). Glossary to ARCH (GARCH). CREATES Research Paper 2008-49, School of Economics and Management, University of Aarhus, http://ssrn.com/abstract=1263250. [Online; accessed 04-01-2011].

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131.

Bougerol, P. and Picard, N. (1992a). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, 52(1–2):115–127.

Bougerol, P. and Picard, N. (1992b). Strict stationarity of generalized autoregressive processes. *The Annals of Probability*, 20(4):1714–1730.

Boussama, F., Fuchs, F., and Stelzer, R. (2011). Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and their Applications*, 121(10):2331–2360.

Brahimi, B. and Necir, A. (2012). A semiparametric estimation of copula models based on the method of moments. *Statistical Methodology*, 9(4):467–477.

Brockwell, P. J. and Davis, R. A. (2006). *Time series: Theory and methods*. Springer, 2nd edition.

Busetti, F. and Harvey, A. C. (2008). When is a copula constant? A test for changing relationships. Cambridge working papers in economics, Faculty of Economics, University of Cambridge.

bwGRiD (http://www.bw-grid.de). Member of the German D-Grid initiative, funded by the Ministry for Education and Research and the Ministry for Science, Research and Arts Baden-Württemberg. [Online; accessed 02-07-2012].

bwGRiD Cluster Ulm (http://www.uni-ulm.de/index.php?id=13040). [Online; accessed 02-07-2012].

Casella, G. and Berger, R. L. (2001). *Statistical Inference*. Duxbury Press, 2nd edition.

Chan, N.-H., Chen, J., Chen, X., Fan, Y., and Peng, L. (2009a). Statistical inference for multivariate residual copula for GARCH models. *Statistica Sinica*, 19(1):53–70.

Chan, N.-H., Chen, J., Chen, X., Fan, Y., and Peng, L. (2009b). Statistical inference for multivariate residual copula for GARCH models: Online supplement. http://www3.stat.sinica.edu.tw/statistica/password.asp?vol=19&num=1&art=2b&sup=s. [Online; accessed 10-11-2010].

Chen, X. and Fan, Y. (2004). Asymptotic properties of several estimators for semiparametric copula-based multivariate dynamic models. Manuscript, New York University and Vanderbilt University.

Chen, X. and Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, 135(1-2):125–154.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance*. Wiley.

Cherubini, U., Mulinacci, S., Gobbi, F., and Romagnoli, S. (2011). *Dynamic Copula Methods in Finance*. Wiley.

Chevallier, J. (2011a). Detecting instability in the volatility of carbon prices. *Energy Economics*, 33(1):99–110.

Chevallier, J. (2011b). *Econometric Analysis of Carbon Markets: The European Union Emissions Trading Scheme and the Clean Development Mechanism*. Springer.

Chollete, L., Heinen, A., and Valdesogo, A. (2009). Modeling international financial returns with a multivariate regime-switching copula. *Journal of Financial Econometrics*, 7(4):437–480.

Clark, T. E. and McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105(1):85–110.

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1:223–236.

Cook, R. D. and Johnson, M. E. (1981). A family of distributions for modelling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society*, 43(2):210–218.

Cook, R. D. and Johnson, M. E. (1986). Generalized Burr-Pareto-Logistic distributions with applications to a Uranium exploration data set. *Technometrics*, 28(2):123–131.

Corradi, V. and Swanson, N. R. (2005). Predictive density evaluation. In Granger, C., Elliot, G., and Timmerman, A., editors, *Handbook Of Economic Forecasting*. Elsevier.

Cox, D. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman & Hall.

Cox, D. R. (1958). Some problems connected with statistical inference. *Annals of Mathematical Statistics*, 29(2):357–372.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

Cramér, H. and Wold, H. (1936). Some theorems on distribution functions. *Journal of the London Mathematical Society*, 11(4):290–294.

Creal, D., Koopman, S. J., and Lucas, A. (2008). A general framework for observation driven time-varying parameter models. Technical report, Tinbergen Institute Discussion Paper No. 08-108/4.

Csörgő, M. and Révész, P. (1975). A strong approximation of the multivariate empirical process. *Studia Scientiarum Mathematicarum Hungarica*, 10:427–434.

Cuadras, C. M. and Augé, J. (1981). A continuous general multivariate distribution and its properties. *Communications in Statistics - Theory and Methods*, 10(4):339–353.

Cui, S. and Sun, Y. (2004). Checking for the gamma frailty distribution under the marginal proportional hazards frailty model. *Statistica Sinica*, 14:249–267.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. *Académie Royale de Belgique-Bulletin de la Classe des Sciences-5e Série*, 65(5):274–292.

Deheuvels, P. (1980). Non parametric tests of independence. In Raoult, J.-P., editor, *Statistique non Paramétrique Asymptotique*, volume 821 of *Lecture Notes in Mathematics*, pages 95–107. Springer.

Demarta, S. and McNeil, A. J. (2004). The t copula and related copulas. http://www.math.ethz.ch/~mcneil/ftp/tCopula.pdf. [Online; accessed 02-04-2011].

Demko, S., Moss, W. F., and Smith, P. W. (1984). Decay rates for inverses of band matrices. *Mathematics of Computation*, 43:491–491.

Dengler, B. (2010). *On the Asymptotic Behaviour of the Estimator of Kendall's Tau*. PhD thesis, Vienna University of Technology.

Dias, A. and Embrechts, P. (2004). Change-point analysis for dependence structures in finance and insurance. In *Risk Measures for the 21st Century*, pages 321–335.

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, 65(3):457–487.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall.

Embrechts, P. (2006). Discussion of "copulas: Tales and facts", by thomas mikosch. *Extremes*, 9:45–47.

Embrechts, P. (2009). Copulas: A personal view. *Journal of Risk and Insurance*, 76:639–650.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2003). *Modelling Extremal Events in Insurance and Finance*. Springer.

Embrechts, P., McNeil, A. J., and Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. In Dempster, M. A. H., editor, *Risk Management: Value at Risk and Beyond*, chapter 7. Cambrige University Press.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica*, 50:987–1008.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, 20(3):339–350.

Engle, R. and Kroner, K. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150.

Engle, R. and Sheppard, K. (2001). Theoretical and empirical properties of Dynamic Conditional Correlation multivariate GARCH. Economics working paper series, University of California at San Diego, Department of Economics, http://www.escholarship.org/uc/item/5s2218dp.pdf. [Online; accessed 21-05-2012].

Erb, C. B., Harvey, C. R., and Viskanta, T. E. (1994). Forecasting international equity correlations. *Financial Analysts Journal*, 50:32–45.

Eugster, M. J. A., Knaus, J., Porzelius, C., Schmidberger, M., and Vicedo, E. (2011). Hands-on tutorial for parallel computing with R. *Journal of Computational Statistics*, 26:219–239.

Fang, H.-B., Fang, K.-T., and Kotz, S. (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis*, 82(1):1–16.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications, Volume 2.* Wiley New York, 2nd edition.

Fermanian, J.-D., Radulović, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10:847–860.

Fermanian, J.-D. and Scaillet, O. (2003). Nonparametric estimation of copulas for time series. FAME Research Paper Series, International Center for Financial Asset Management and Engineering.

Fernández, B. and Muriel, N. (2009). Regular variation and related results for the multivariate GARCH(p, q) model with constant conditional correlations. *Journal of Multivariate Analysis*, 100(7):1538–1550.

Fisher, R. A. (1912). On an absolute criterion for fitting frequency curves. *Messenger of Mathematics*, 41:155–160.

Fisher, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, 1:3–32.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222:309–368.

Fisher, R. A. (1970). *Statistical Methods for Research Workers.* Oliver and Boyd.

Fisher, R. A. and Tippett, L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24:180–190.

Frahm, G. (2006). On the extremal dependence coefficient of multivariate distributions. *Statistics & Probability Letters*, 76(14):1470–1481.

Frank, M. (1979). On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Mathematicae*, 19:194–226.

Franke, J., Härdle, W. K., and Hafner, C. M. (2008). *Statistics of Financial Markets.* Springer.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2:1–25.

Gaisser, S., Memmel, C., Schmidt, R., and Wehn, C. (2009). Time dynamic and hierarchical dependence modelling of an aggregated portfolio of trading books: a multivariate nonparametric approach. Discussion Paper Series 2: Banking and financial studies 2009,07, Frankfurt am Main.

Galambos, J. (1987). *The asymptotic theory of extreme order statistics.* Krieger.

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika*, 74(3):549–555.

Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12:347–368.

Genest, C., Gendron, M., and Bourdeau-Brien, M. (2009a). The advent of copulas in finance. *European Journal of Finance*, 15(7-8):609–618.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.

Genest, C. and MacKay, R. J. (1986a). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canadian Journal of Statistics*, 14(2):145–159.

Genest, C. and MacKay, R. J. (1986b). The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, 40(4):280–283.

Genest, C., Quessy, J.-F., and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33(2):337–366.

Genest, C., Rémillard, B., and Beaudoin, D. (2009b). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44:199–213.

Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88(423):1034–1043.

Genest, C. and Werker, B. (2001). Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. Technical report, Tilburg University, http://ideas.repec.org/p/ner/tilburg/urnnbnnlui12-86725.html. [Online; accessed 30-04-12].

Ghalanos, A. (2012). *rugarch: Univariate GARCH models.* R package version 1.0-9.

Giacomini, E., Härdle, W., and Spokoiny, V. (2009). Inhomogeneous dependency modelling with time varying copulae. *Journal of Business and Economic Statistics*, 27:224–234.

Goldie, C. M. (1991). Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability*, 1(1):126–166.

Google search (http://tinyurl.com/GoogleCopula). [Online; accessed 17-07-2012].

Guegan, D. and Zhang, J. (2009). Change analysis of dynamic copulas for measuring dependence in multivariate financial data. *Quantitative Finance*, 10(4):421–430.

Gumbel, E. (1960). Distributions des valeurs extrêmes en plusiers dimensions. *Publications de l'Institut de Statistique de l'Université de Paris*, 9:171–173.

Hafner, C., Dijk, D. v., and Franses, P. (2005). Semi-parametric modelling of correlation dynamics. Technical Report EI 2005-26, Erasmus University Rotterdam, Econometric Institute.

Hafner, C. M. and Manner, H. (2008). Dynamic stochastic copula models: Estimation, inference and applications. Research Memoranda 043, METEOR, Maastricht.

Hafner, C. M. and Reznikova, O. (2010). Efficient estimation of a semiparametric dynamic copula model. *Computational Statistics & Data Analysis*, 54(11):2609–2627.

Hansen, P. R. and Lunde, A. (2001). A comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7):873–889.

Härdle, W., Okhrin, O., and Okhrin, Y. (2008). Modeling dependencies with copulae. In Härdle, W., Kleinow, T., and Stahl, G., editors, *Applied Quantative Finance*, pages 3–26. Springer, 2nd edition.

Harvey, A. (2008). Dynamic distributions and changing copulas. Cambridge Working Papers in Economics 0839, Faculty of Economics, University of Cambridge.

Harvey, A. (2010). Tracking a changing copula. *Journal of Empirical Finance*, 17(3):485–500.

Hering, C. (2011). *Estimation Techniques and Goodness-of-fit Tests for Certain Copula Classes in Large Dimensions*. PhD thesis, Ulm University.

Hering, C., Hofert, M., Mai, J.-F., and Scherer, M. (2010). Constructing hierarchical Archimedean copulas with Lévy subordinators. *Journal of Multivariate Analysis*, 101(6):1428–1433.

Hering, C. and Stadtmüller, U. (2012). Estimating Archimedean copulas in high dimensions. *Scandinavian Journal of Statistics*, Forthcoming.

Hoeffding, W. (1940). Maßstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 5:179–233.

Hoeffding, W. (1941). Maßstabinvariante Korrelationsmaße für diskontinuierliche Verteilungen. *Archiv für die mathematische Wirtschafts- und Sozialforschung*, 7:49–70.

Hofert, J. M. (2010). *Sampling nested Archimedean copulas with applications to CDO pricing*. PhD thesis, Ulm University.

Hofert, M. (2008). Sampling Archimedean copulas. *Computational Statistics & Data Analysis*, 52(12):5163–5174.

Hofert, M. (2011). Efficiently sampling nested Archimedean copulas. *Computational Statistics & Data Analysis*, 55(1):57–70.

Hofert, M., Mächler, M., and McNeil, A. J. (2012). Likelihood inference for Archimedean copulas. http://arxiv.org/abs/1108.6032. [Online; accessed 02-04-2012].

Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika*, 73(3):671–678.

Hull, J. C. (2002). *Options, Futures, and other Derivatives*. Prentice Hall, 5th edition.

Hutchinson, T. and Lai, C. (1990). *Continuous Bivariate Distributions, Emphasising Applications*. Rumsby Scientific Publishing.

InfiniBand® Trade Association (http://www.infinibandta.org). [Online; accessed 05-07-2012].

Janßen, A. (2010). *On Some Connections between Light Tails, Regular Variation and Extremes*. PhD thesis, Georg-August-Universität Göttingen.

Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, 14(1):70–86.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall.

Joe, H. and Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models. Technical Report 166, Department of Statistics, University of British Columbia.

Jondeau, E. and Rockinger, M. (2006). The copula-GARCH model of conditional dependencies: An international stock market application. *Journal of International Money and Finance*, 25:827–853.

J.P.Morgan/Reuters (1996). RiskMetrics - Technical Document. http://www.riskmetrics.com/rmcovv.html. [Online; accessed 03-05-2010].

Kálmán, R. E. (1960). A new approach to linear filtering and prediction problems. *Transaction of the ASME, Journal of Basic Engineering*, 82(D):35–45.

Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, CXXXI:207–248.

Kim, G., Silvapulle, M. J., and Silvapulle, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis*, 51(6):2836–2850.

Kimberling, C. H. (1974). A probabilistic interpretation of complete monotonicity. *Aequationes Mathematicae*, 10(2-3):152–164.

Klaassen, C. A. J. and Wellner, J. A. (1997). Efficient estimation in the bivariate normal copula model: Normal margins are least favourable. *Bernoulli*, 3(1):55–77.

Kojadinovic, I. and Yan, J. (2010a). Comparison of three semiparametric methods for estimating dependence parameters in copula models. *Insurance: Mathematics and Economics*, 47(1):52–63.

Kojadinovic, I. and Yan, J. (2010b). Modeling multivariate distributions with continuous margins using the `copula R` package. *Journal of Statistical Software*, 34:1–20.

Kurowicka, D. and Joe, H. (2011). *Dependence Modeling: Vine Copula Handbook*. World Scientific.

LBBW Trading Room (http://www.uni-ulm.de/index.php?id=10666). [Online; accessed 04-07-2012].

Le Cam, L. (1990). Maximum likelihood: An introduction. *International Statistical Review*, 58:153–171.

Leadbetter, M. R., Lindgren, G., and Rootzen, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer.

Lee, A. (1990). *U-Statistics: Theory and Practice*. Statistics, textbooks and monographs. M. Dekker.

Lehmann, E. L. and Casella, G. (2003). *Theory of Point Estimation*. Springer, 2nd edition.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer, 3rd edition.

Longin, F. and Solnik, B. (1995). Is the correlation in international equity returns constant: 1960–1990? *Journal of International Money and Finance*, 14:3–26.

Mai, J.-F. (2010). *Extendibility of Marshall-Olkin distributions via Lévy subordinators and an application to portfolio credit risk*. PhD thesis, Technische Universität München.

Mai, J.-F. and Scherer, M. (2009). Lévy-frailty copulas. *Journal of Multivariate Analysis*, 100(7):1567–1585.

Marchant, J. (2011). *Cassell's Latin Dictionary: Latin-English and English-Latin*. BiblioBazaar.

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):30–44.

Marshall, A. W. and Olkin, I. (1988). Families of multivariate distributions. *Journal of the American Statistical Association*, 83(403):834–841.

McCallum, Q. E. and Weston, S. (2011). *Parallel R*. O'Reilly.

McNeil, A. J. (2008). Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, 78(6):567–581.

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

McNeil, A. J. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. *Annals of Statistics*, 37(5B):3059–3097.

Mikosch, T. (2006). Copulas: Tales and facts. *Extremes*, 9.

Mikosch, T. and Starica, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *The Annals of Statistics*, 28(5):1427–1451.

Molenberghs, G. and Lesaffre, E. (1994). Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association*, 89(426):633–644.

Moro, A. (1997). *The Raising of Predicates: Predicative Noun Phrases and the Theory of Clause Structure*. Cambridge Studies in Linguistics. Cambridge University Press.

MPI Forum (2009). MPI: A Message-Passing Interface Standard. Version 2.2. http://www.mpi-forum.org. [Online; accessed 03-07-2012].

Nelsen, R. B. (1986). Properties of a one-parameter family of bivariate distributions with specified marginals. *Communications in Statistics - Theory and Methods*, 15(11):3277–3285.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, 2nd edition.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6(03):318–334.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59:347–370.

Newell, G. F. (1964). Asymptotic extremes for m-dependent random variables. *The Annals of Mathematical Statistics*, 35:1322–1325.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society*, 44(3):414–422.

Oakes, D. (1986). Semiparametric inference in a model for association in bivariate survival data. *Biometrika*, 73(2):353–361.

Oh, C. H. and Patton, A. (2011). Simulated method of moments estimation for copula-based multivariate models. Technical report, Duke University, http://public.econ.duke.edu/~ap172/Oh_Patton_SMM_copulas_nov11.pdf. [Online; accessed 01-05-2012].

Okhrin, O., Okhrin, Y., and Schmid, W. (2009). Properties of hierarchical Archimedean copulas. Discussion Paper 2009-014, Sonderforschungsbereich 649, Humboldt Universität zu Berlin, http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2009-014.pdf. [Online; accessed 10-03-2010].

Patton, A. (2003). The Joe-Clayton and symmetrised Joe-Clayton density functions. http://public.econ.duke.edu/~ap172/sym_jc_density_function.pdf. [Online; accessed 12-04-12].

Patton, A. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47:527–556.

Patton, A. (2008). Copula toolbox for `matlab`, version 1.07. http://public.econ.duke.edu/~ap172/Patton_copula_toolbox.zip. [Online; accessed 05-07-2012].

Patton, A. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, forthcoming.

Pearson, K. (1895). Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society*, 58:240–242.

Pelletier, D. (2006). Regime switching for dynamic correlations. *Journal of Econometrics*, 131(1-2):445–473.

Plackett, R. L. (1965). A class of bivariate distributions. *Journal of the American Statistical Association*, 60(310):516–522.

R Development Core Team (2010). *`R`: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rahe, F. (2012). *Option pricing under time-varying risk aversion with applications in risk management*. PhD thesis, Ulm University.

Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37:81–89.

Renardy, M. and Rogers, R. (2004). *An introduction to partial differential equations*. Texts in applied mathematics. Springer.

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

Revolution Analytics (2011a). *`doMC`: Foreach parallel adaptor for the `multicore` package*. `R` package version 1.2.3.

Revolution Analytics (2011b). *`foreach`: Foreach looping construct for R*. `R` package version 1.3.2.

Reznikova, O. and Manner, H. (2009). Time varying copulas: A survey. Institut de Statistique, Université Catholique de Louvain.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472.

Ruiz-Rivas, C. and Cuadras, C. M. (1988). Inference properties of a one-parameter curved exponential family of distributions with given marginals. *Journal of Multivariate Analysis*, 27(2):447–456.

Ruymgaart, F. H. (1974). Asymptotic normality of nonparametric tests for independence. *The Annals of Statistics*, 2(5):892–910.

Ruymgaart, F. H., Shorack, G. R., and van Zwet, W. R. (1972). Asymptotic normality of nonparametric tests for independence. *The Annals of Mathematical Statistics*, 43(4):1122–1135.

Rüschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics. *The Annals of Statistics*, 4:912–923.

Salvadori, G. (2007). *Extremes in Nature: An Approach Using Copulas*. Water Science and Technology Library. Springer.

Savu, C. and Trede, M. (2008). Goodness-of-fit tests for parametric families of Archimedean copulas. *Quantitative Finance*, 8(2):109–116.

Schmid, F. and Schmidt, R. (2006). Bootstrapping Spearman's multivariate rho. In Rizzi, A. and Vichi, M., editors, *COMPSTAT: Proceedings in computational statistics: 17th symposium held in Rome, Italy*. Springer.

Schmid, F. and Schmidt, R. (2007a). Multivariate conditional versions of Spearman's rho and related measures of tail dependence. *Journal of Multivariate Analysis*, 98(6):1123–1140.

Schmid, F. and Schmidt, R. (2007b). Nonparametric inference on multivariate versions of Blomqvist's beta and related measures of tail dependence. *Metrika*, 66(3):323–354.

Schmid, F., Schmidt, R., Blumentritt, T., Gaißer, S., and Ruppert, M. (2010). Copula-based measures of multivariate association. In Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., editors, *Copula Theory and Its Applications*, volume 198 of *Lecture Notes in Statistics*, pages 209–236. Springer.

Schmitz, V. (2003). *Copulas and Stochastic Processes*. PhD thesis, Rheinisch-Westfälische Technische Hochschule Aachen.

Schwarz, H. and Köckler, N. (2006). *Numerische Mathematik*. Teubner.

Schweizer, B. and Sklar, A. (1963). Associative functions and abstract semigroups. *Publicationes Mathematicae Debrecen*, 10:69–81.

Serfling, R. (2001). *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics. Wiley.

Sewell, M. (2011). Characterization of financial time series. Research note rn/11/01, UCL Department of Computer Science, http://www.cs.ucl.ac.uk/fileadmin/UCL-CS/images/Research_Student_Information/RN_11_01.pdf. [Online; accessed 12-05-2012].

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D.R. Cox, D. V. H. and Barndorff-Nielsen, O. E., editors, *Time Series Models in Econometrics, Finance and Other Fields*, pages 1–67. Chapman & Hall.

Silvennoinen, A. and Teräsvirta, T. (2009). Multivariate GARCH models. In Andersen, T. G., Davis, R. A., Kreiss, J.-P., and Mikosch, T., editors, *Handbook of Financial Time Series*, pages 201–229. Springer.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *L'Institut de Statistique de l'Université de Paris*, 8:229–231.

Sklar, A. (1996). Random variables, distribution functions, and copulas – a personal look backward and forward. In *Distributions with Fixed Marginals and Related Topics*, volume 28, pages 1–14.

Stephenson, A. G. (2002). `evd`: Extreme value distributions. *R News*, 2(2):31–32.

Stöber, J. and Czado, C. (2012). Detecting regime switches in the dependence structure of high dimensional financial data. http://arxiv.org/pdf/1202.2009v1. [Online; accessed 13-07-2012].

Teriokhin, A., de Meeûs, T., and Guégan, J. (2007). On the power of some binomial modifications of the Bonferroni multiple test. *Journal of General Biology*, 68(5):332–340.

Thorisson, H. (2000). *Coupling, Stationarity, and Regeneration*. Springer.

Tsukahara, H. (2005). Semiparametric estimation in copula models. *The Canadian Journal of Statistics*, 33(3):357–375.

van den Goorbergh, R. W., Genest, C., and Werker, B. J. (2005). Bivariate option pricing using dynamic copula models. *Insurance: Mathematics and Economics*, 37:101–114.

Watson, G. S. (1954). Extreme values in samples from m-dependent stationary stochastic processes. *Annals of Mathematical Statistics*, 25:798–800.

Weiß, G. (2011). Copula parameter estimation by maximum-likelihood and minimum-distance estimators: A simulation study. *Computational Statistics*, 26:31–54.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64(5):1067–84.

West, K. D. and McCracken, M. W. (1998). Regression-based tests of predictive ability. *International Economic Review*, 39(4):817–840.

White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5):1097–1126.

Wolff, E. F. (1980). N-dimensional measures of dependence. *Stochastica*, IV(3):175–188.

Wolfowitz, J. (1957). The minimum distance method. *The Annals of Mathematical Statistics*, 28(1):75–88.

Wuertz, D., with contribution from Michal Miklovic, Y. C., Boudt, C., Chausse, P., and others (2010). *`fGarch`: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling*. `R` package version 2110.81/r4996.

Xu, J. J. (1996). *Statistical modelling and inference for multivariate and longitudinal discrete response data*. PhD thesis, University of British Columbia, Department of Statistics.

Yan, J. (2007). Enjoy the joy of copulas: With a package `copula`. *Journal of Statistical Software*, 21(4):1–21.

Yaskov, P. (2010a). *Law of large numbers in modern stochastic models [In Russian]*. PhD thesis, Lomonosov Moscow State University, Department of Mechanics and Mathematics.

Yaskov, P. (2010b). Testing for predictive ability in the presence of structural breaks [in Russian]. *Quantile*, 8:127–135.

Yu, J. and Meyer, R. (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews*, 25:361–384.

Zeileis, A. and Kleiber, C. (2005). Validation of multiple structural change models - a case study. *Journal of Applied Econometrics*, 20:685–690.

Zeileis, A., Leisch, F., Hornik, K., and Kleiber, C. (2010). `strucchange`: An `R` package for testing for structural change in linear regression models. *Journal of Statistical Software*, 7(2):1–38.

List of Figures

1.1 Annual number of publications on copula theory 1971–2005 (cf. Figure 1 in Genest et al., 2009a). 2

2.1 Graphs of the copulas W, M and Π in dimension $d = 2$ 10

3.1 Raw financial time series and the corresponding logreturns (in €) 50
3.2 Smoothed density of logreturns and autocorrelation functions (acf) . . . 51
3.3 Parameter regions of $\alpha_1 > 0, \beta_1 \geq 0$ in GARCH(1,1) , $\varepsilon_t \sim \mathcal{N}(0,1)$ iid . . 58

4.1 Power of Partial Tests 80
4.2 Density of the standard normal distribution 82
4.3 "all false" alternative 87
4.4 "one false" alternative 88
4.5 "half false" alternative 89
4.6 Summation Scheme 92

5.1 `API22MON Index` description page from Bloomberg 117
5.2 Histogram of the goodness-of-fit results 120
5.3 Global exceedance and absence of exceedances 122
5.4 Binomial test with proper size for **`API22MON`** vs. **`ELGP2MON`** 126

A.1 Gaussian copula $C_{\mathrm{Ga};\rho}$ with standard normal margins 130
A.2 Student copula $C_{t_\nu;\rho}$ with $\nu = 2$ degrees of freedom and standard normal margins 132
A.3 Marshall-Olkin copula $C_{\mathrm{MO};\alpha_1,\alpha_2}$ with constant $\lambda_{12} = 1$, $n = 1000$ 133
A.4 Plackett copula $C_{\mathrm{P};\theta}$ with standard normal margins 134
A.5 Clayton copula $C_{\mathrm{C};\theta}$ with $\theta < 0$, $n = 1000$ 135
A.6 Clayton copula $C_{\mathrm{C};\theta}$ with standard normal margins 136
A.7 Gumbel copula $C_{\mathrm{G};\theta}$ with standard normal margins 138
A.8 Frank copula $C_{\mathrm{F};\theta}$ with standard normal margins 139
A.9 Joe-Clayton copula $C_{\mathrm{JC};\lambda_l,\lambda_u}$ (red) vs. symmetrized Joe-Clayton copula $C_{\mathrm{SJC};\lambda_l,\lambda_u}$ (black) with standard normal margins 140
A.10 Symmetrized Joe-Clayton copula $C_{\mathrm{SJC};\lambda_l,\lambda_u}$ with standard normal margins 141

List of Tables

5.1 Input parameters for Monte Carlo simulation study 110
5.2 Empirical size of randomized Binomial tests under different data set-ups (nominal level $\alpha^* = 5\%$) 114
5.3 Estimated parameters for univariate commodity GARCH(1,1) models . 121
5.4 Estimated global parameter (lower left) and bootstrapped variance estimate (upper right) for bivariate SCOMDY models 121
5.5 Summary statistics for Clayton copula estimation with 5-year commodity data 124

A.1 Characteristics of non-Archimedean copula families 143
A.2 Characteristics of Archimedean copula families 143

Der Lebenslauf wurde aus Gründen des Datenschutzes in der elektronischen Version der Arbeit entfernt.