Statistical Inference for Copulas and Extremes

AXEL BÜCHER

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
an der Fakultät für Mathematik
der Ruhr-Universität Bochum

März 2011

Contents

Introduction

The modeling of dependence relations between random variables is one of the most important subjects in probability theory and statistics. Not surprisingly, a great variety of concepts for dependence structures has emerged, for an overview see the monograph [Joe, 1997]. The most popular approach is based on second moments of the underlying random variables, the covariance. Nevertheless, it is well known that only linear dependence can be captured by the covariance and that it is characterizing only for a few special classes of distributions, e.g. the multivariate normal distribution.

As a beneficial alternative, the concept of copulas going back to [Sklar, 1959] has drawn a lot of attention in the last decades. The copula $C : [0,1]^2 \to [0,1]$ of a random vector (X,Y) allows to separate the effect of dependence from the effects of the marginal distributions in the following manner:

$$\mathbb{P}[X \leq x, Y \leq y] = C\big(\mathbb{P}[X \leq x], \mathbb{P}[Y \leq y]\big).$$

While mathematical details will be given in the first chapter of this thesis, we summarize the main consequence of this identity: the copula completely characterizes the stochastic dependence between the random variables X and Y.

In the recent years, the theory of copulas has been introduced to many applied fields of science, e.g. mathematical finance, actuarial science or hydrology among others (see [McNeil et al., 2005; Frees and Valdez, 1998; Genest and Favre, 2007] and references therein). The increasing number of applications gave rise to a great demand for statistical methods. To name but a few accomplishments, semiparametric and nonparametric estimation of copulas was investigated in [Genest et al., 1995; Fermanian et al., 2004], while [Genest et al., 2009] gives an overview on testing the goodness-of-fit for copula models. In order to approximate the behavior of occurring test statistics within the framework of copulas several resampling procedures have been proposed, see e.g. [Fermanian et al., 2004] for the bootstrap based on resampling or [Genest and Rémillard, 2008] for the (semi-)parametric bootstrap.

Furthermore, [Rémillard and Scaillet, 2009] proposed a new multiplier bootstrap approach to approximate the empirical copula process. This stochastic process describes the behavior of the most popular nonparametric estimator for the copula, the empirical copula going back to [Deheuvels, 1979]. It is part of this thesis to extend and improve on these results.

Investigating the notion of copulas within the framework of multivariate extreme value theory leads to the so called *extreme value copulas*. These copulas can be characterized by the *Pickands dependence function A* going back to [Pickands, 1981]:

$$C(u,v) = \exp\left\{\log(uv)A\left(\frac{\log(v)}{\log(uv)}\right)\right\},$$

where $A : [0,1] \to [0,1]$ is convex and satisfies $\max\{t, 1-t\} \leq A(t) \leq 1$. The class of extreme value copulas forms a large and important class of copulas and its estimation in the setting where marginal distributions are unknown has recently been investigated in [Genest and Segers, 2009]. It is a further prime concern of this thesis to introduce new estimators for extreme value copulas and for its corresponding Pickands dependence function. The problem of testing for the hypothesis of being faced with an extreme value copula has drawn much less attention in the literature and is a concern of current research interest, see [Ghoudi et al., 1998] and [Kojadinovic and Yan, 2010]. The present thesis provides a new test for this hypothesis where the critical values of the test are obtained by the multiplier bootstrap.

Multivariate extreme value theory is concerned with the estimation of events outside the range of the data, see [de Haan and Ferreira, 2006]. For this purpose one is interested in the limiting distribution of some appropriate vector maximum of a given d-dimensional data sample. It is well known that the possible limit distributions can be characterized by different mathematical objects, for instance the exponent measure, the spectral measure or the stable tail dependence function (or, equivalently, the tail copula). Estimating these objects turns out to be rather difficult and the rate of convergence is usually smaller than $\sqrt{n}$, more precisely it is $\sqrt{k_n}$ where k_n tends to infinity with $k_n = o(n)$, see e.g. [Einmahl et al., 2001; Huang, 1992; Schmidt and Stadtmüller, 2006]. It is the third main concern of this thesis to revisit the estimators investigated in the latter two sources. The smoothness assumptions regarding the asymptotic theory of the corresponding estimators in [Huang, 1992; Schmidt and Stadtmüller, 2006] are so restrictive that there is no model at all satisfying them and therefore we show how the assumptions can be suitably weakened. Furthermore, it is a main goal of this thesis to develop a multiplier bootstrap procedure which yields an approximation for the estimators defined in [Huang, 1992; Schmidt and Stadtmüller, 2006].

This work is organized as follows: In Chapter 1 we state some basic definitions and results, which will be repeatedly used throughout this work. The concept of copulas will be introduced as well as the foundations of multivariate extreme value theory. A short survey on weak convergence in general metric spaces based on outer integrals finalizes the chapter.

Chapter 2 will deal with multiplier bootstrap approximations of the empirical copula process. We will give an overview on existing approaches and introduce two new procedures. All approaches are compared by a small simulation study concerning the finite sample properties.

Chapter 3 concerns the estimation of extreme value copulas. We will introduce a new class of minimum distance estimators for Pickands dependence function which easily allows for testing extreme value dependence, i.e. for the hypothesis that the underlying copula C is an extreme value copula. The critical values of the test are determined by the multiplier bootstrap approximation investigated in the preceding Chapter 2. All results are examined for their practical applicability by means of a simulation study.

In Chapter 4 we will investigate the empirical tail copula process under non-restrictive smoothness assumptions. We introduce two multiplier bootstrap procedures for the limiting variable and examine several applications, both theoretically and by means of simulation studies. The applications include a test for equality between two tail copulas, a new minimum distance estimate and a goodness-of-fit test for parametric tail copula models.

Finally, some technical details are deferred to the appendix.

* * *

I would like to express my deepest gratitude to Professor Holger Dette for his excellent guidance during the entire phase of my doctoral studies. His infecting enthusiasm in many fields of mathematics and statistics and his readiness to help were very encouraging.

Further, I extend my thanks to my colleagues at the Institute of Statistics of whom many contributed to this thesis in a way or another. In particular, a special thankyou goes to Dr. Stanislav Volgushev for many fruitful discussions and critical remarks.

I am also very grateful to Professor Johan Segers who provided me with several useful hints during his visit of Bochum.

Finally, my special gratitude belongs to my family and my friends for their encouragement throughout the last years.

Chapter 1

Mathematical Preliminaries

This chapter is devoted to the clarification of the mathematical concepts involved. In the first section we will introduce the copula concept and its relation to the notion of stochastic dependence. We proceed by an overview on some results in multivariate extreme value theory and end with a short summary of modern weak convergence in general metric spaces on the basis of outer integrals.

Some notation: Vectors in $\mathbb{R}^d$ will be denoted by $\mathbf{x} = (x_1, \ldots, x_d)^T$ and relations and operations are taken component-wise. For example, for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ the relation $\mathbf{x} < \mathbf{y}$ means $x_j < y_j$ for all $j = 1, \ldots, d$. For reals a, b the notation $a \vee b$ stands for the maximum of a and b, while $a \wedge b$ denotes the minimum. $\mathbf{x} \wedge \mathbf{y}$ is the component-wise minimum of $\mathbf{x}$ and $\mathbf{y}$. The origin of $\mathbb{R}^d$ is denoted by $\mathbf{0}$, rectangles in $\mathbb{R}^d$ are defined by $[\mathbf{x}, \mathbf{y}] = \{\mathbf{z} : \mathbf{x} \leq \mathbf{z} \leq \mathbf{y}\}$. The extended real line $\bar{\mathbb{R}}$ is the interval $[-\infty, \infty]$. The complement of a set A is denoted by A^c. For a distribution function F on the real line F^- denotes the generalized inverse of F and is formally defined by

$$F^-(p) := \begin{cases} \inf\{x \in \mathbb{R} \mid F(x) \geq p\}, & 0 < p \leq 1 \\ \sup\{x \in \mathbb{R} \mid F(x) = 0\}, & p = 0, \end{cases}$$

where $\inf \emptyset = \infty, \sup \emptyset = -\infty$. The range and the support of F are denoted by $\operatorname{ran} F$ and $\operatorname{supp} F$, respectively.

1.1 Modeling dependence: the copula concept

The basic idea underlying the modeling of stochastic dependence by copulas is the following: let $\mathbf{X} = (X_1, X_2)^T$ be a random vector with marginal distribution functions $F_1 = F_{X_1}$ and $F_2 = F_{X_2}$. For the moment assuming that these marginal distributions are continuous the probability integral transform applied to X_1 and X_2 defines two random variables $U_1 = F_1(X_1)$ and $U_2 = F_2(X_2)$ which are uniformly distributed on $[0, 1]$. If X_1 and X_2 were dependent so are U_1 and U_2 and since the transforms are invertible specifying the dependence between X_1 and X_2 is, in a way, the same as specifying the dependence between U_1 and U_2. The problem of investigating stochastic dependence has therefore been reduced to the problem of investigating bivariate distribution functions on the unit cube $[0,1]^2$ with uniform marginals, this is the copula.

Definition 1.1 (Copula)
A *(d-dimensional) copula* $C : [0,1]^d \to [0,1]$ is a multivariate cumulative distribution function on the unit cube with uniform marginals.

The motivation preceding this definition is summarized in the following well-known theorem of Sklar which underlies most applications of copulas. It turns out that relaxing the assumption of continuity of the marginals results in non-uniqueness of the associated copula.

Theorem 1.2 (Sklar's Theorem)
Let F be a d-dimensional distribution function with margins F_p for $p = 1, \ldots, d$. Then there exists a copula C such that for all $\mathbf{x} \in \bar{\mathbb{R}}^d$

$$F(\mathbf{x}) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{1.1}$$

If all marginals are continuous then C is uniquely determined by

$$C(\mathbf{u}) = F(F_1^-(u_1), \ldots, F_d^-(u_d)),$$

otherwise it is uniquely determined on $\operatorname{ran} F_1 \times \cdots \times \operatorname{ran} F_d$. Conversely, if C is a copula and $F_1, \ldots, F_d$ are distribution functions, then the function F defined by (1.1) is a joint distribution function with marginals $F_1, \ldots, F_d$.

The theorem first appeared in [Sklar, 1959], an interesting new proof has recently been published in [Rüschendorf, 2009]. Its importance for the notion of stochastic dependence can hardly be underestimated: every joint distribution function, which obviously contains all information about the dependence of a random vector distributed according to this function, can be split into a copula and into the marginal distributions. Since the marginals are not able to explain any dependence properties the dependence has to be completely characterized by the copula. This is underlined by the fact that usual measures of dependence like Kendall's-τ or Spearman's-ρ are functions of the copula, see the following remark where we collect some basic properties of copulas. For the sake of brevity we only state the results for $d = 2$.

Proposition 1.3 (Basic properties of copulas)
Let C be a 2-dimensional copula and suppose that $\mathbf{X} = (X_1, X_2)^T \sim F = C(F_1, F_2)$ is a random vector with marginals F_1, F_2 and copula C. Then the following results hold.

(i) **Fréchet-Hoeffding bounds.** *For all $\mathbf{u} \in [0,1]^2$*

$$W(\mathbf{u}) = \max\{u_1 + u_2 - 1, 0\} \le C(\mathbf{u}) \le \min\{u_1, u_2\} = M(\mathbf{u}).$$

The upper bound M corresponds to perfect positive dependence in the sense that X_1 is a strictly increasing function of X_2 almost surely. Analogously, W corresponds to perfect negative dependence.

(ii) **Independence.** *If F_1 and F_2 are continuous then X_1 and X_2 are independent if and only if $C(\mathbf{u}) = u_1 u_2 = \Pi(\mathbf{u})$. The latter is called the independence copula.*

(iii) **Lipschitz-continuity.** *C is Lipschitz-continuous with respect to the L^1-norm on $[0,1]^2$:*

$$|C(\mathbf{u}) - C(\mathbf{v})| \leq ||\mathbf{u} - \mathbf{v}||_1 = |u_1 - v_1| + |u_2 - v_2| \qquad \textit{for all } \mathbf{u}, \mathbf{v} \in [0,1]^2.$$

(iv) **Differentiability.** *For all $u_1 \in [0,1]$ it holds that the partial derivative $\partial_2 C(\mathbf{u})$ exists for λ_1-almost every u_2. Furthermore, $0 \leq \partial_2 C(\mathbf{u}) \leq 1$. The same is true for interchanged roles of u_1 and u_2.*

(v) **Invariance under increasing transformations.** *If F_1 and F_2 are continuous and α, β are strictly increasing, then the copula of $(\alpha \circ X_1, \beta \circ X_2)$ is C.*

(vi) **Kendall's-τ.** *If F_1 and F_2 are continuous then*

$$\tau = \tau_{X_1, X_2} = 4 \int_{[0,1]^2} C(\mathbf{u}) \, dC\mathbf{u} - 1.$$

For proofs of these results and more details regarding the theory of copulas we refer the reader to the monographs [Nelsen, 2006; Joe, 1997].

1.2 Modeling extremal dependence

In this section we will formally introduce the concept of tail dependence and embed it into the framework of copulas. Loosely spoken, bivariate tail dependence is the amount of dependence in the upper-quadrant tail or lower-quadrant tail of a bivariate distribution. The concept turns out to be deeply related to multivariate extreme value theory and we will clarify this relationship.

In the following let $\mathbf{X} = (X_1, X_2)^T$ be a two-dimensional random vector with joint distribution function F, marginal distribution functions F_1, F_2 and copula C. The following definition introduces the common scalar measure of tail dependence, the so-called coefficient of tail dependence, see [Joe, 1997].

Definition 1.4 (Tail dependence coefficient)
X_1 and X_2 are said to be *upper tail dependent* if the limit

$$\lambda_U = \lim_{u \to 1} \mathbb{P}(X_1 > F_1^-(u) \mid X_2 > F_2^-(u))$$

exists and is positive. X_1 and X_2 are said to be *upper tail independent* if $\lambda_U = 0$. λ_U is called the *upper tail-dependence coefficient*. Similarly, the *lower tail dependence coefficient* is defined as

$$\lambda_L = \lim_{u \to 0} \mathbb{P}(X_1 \leq F_1^-(u) \mid X_2 \leq F_2^-(u)),$$

if existent, and X_1 and X_2 are said to be *lower tail dependent (lower tail independent)* if $\lambda_L > 0$ (resp. $\lambda_L = 0$).

As one might expect, for continuous marginal distributions tail dependence is a property of the copula of $\mathbf{X}$. If $\bar{C}(\mathbf{u}) = u_1 + u_2 - 1 + C(1-u_1, 1-u_2)$ denotes the survival copula of $\mathbf{X}$ an easy calculation based on the definition of conditional probabilities justifies the following Proposition.

Proposition 1.5
Let $\mathbf{X}$ be continuous bivariate random vector, then

$$\lambda_L = \lim_{u\to 0} \frac{C(u,u)}{u} \quad \textit{and} \quad \lambda_U = \lim_{u\to 1} \frac{1-2u+C(u,u)}{1-u} = \lim_{u\to 0} \frac{\bar{C}(u,u)}{u}. \tag{1.2}$$

The limit relations in (1.2) show that tail dependence measured by the tail dependence coefficients is solely determined by the diagonal section of the copula. The stable tail dependence function (introduced by [Huang, 1992]), resp. the tail copula (see [Schmidt and Stadtmüller, 2006]) provides a natural tool generalizing the coefficients by taking into account the off-diagonal section of C. From now on let the marginals of $\mathbf{X}$ be continuous.

Definition 1.6 (Stable tail dependence function, tail copulas)
In case of existence of the following limits

$$\begin{aligned}\Lambda_L(\mathbf{x}) &= \lim_{t\to\infty} t\, C(x_1/t, x_2/t)\\ \Lambda_U(\mathbf{x}) &= \lim_{t\to\infty} t\, \bar{C}(x_1/t, x_2/t),\end{aligned}$$

where $\mathbf{x} = (x_1, x_2)^T \in \bar{\mathbb{R}}^2_+ := [0,\infty]^2 \setminus \{(\infty,\infty)\}$, the functions Λ_L and Λ_U are referred to as the *lower (resp. upper) tail copula* of $\mathbf{X}$ (or equivalently of F). The function

$$l(\mathbf{x}) = x_1 + x_2 - \Lambda_U(\mathbf{x}),$$

where $\mathbf{x} \in \mathbb{R}^2_+$, is called *stable tail-dependence function*.

Before we proceed a short note on the terminology: the concept and the name *stable tail dependence function* goes back to the phd-thesis [Huang, 1992]. The background of its definition stems from multivariate extreme value theory, for details see the forthcoming discussion. In the phd-thesis [Schmidt, 2003] and its attendant publication [Schmidt and Stadtmüller, 2006] the name *tail copula* was introduced, even though, at least the upper tail copula, provided no new concept. Nevertheless, since the lower tail copula does not have a clear and prominent expression in the literature up to the thesis [Schmidt, 2003], we use both names in the following.
Note that $\lambda_L = \Lambda_L(1,1)$ and $\lambda_U = \Lambda_U(1,1)$. For generalizations of tail copulas to higher dimension we refer the reader to the thesis [Schmidt, 2003]. Note that the tail copula is not a copula itself, the name is justified by the following properties, which are quite similar to the copula properties stated in Proposition 1.3.

Proposition 1.7 (Schmidt and Stadtmüller, 2006)
Let Λ denote either the lower or the upper tail copula. If Λ exists it has the following properties.

(i) **Groundedness.** $\Lambda(x, 0) = \Lambda(0, x) = 0$ *for all* $x \in [0, \infty]$ *and* $\Lambda(x, \infty) = \Lambda(\infty, x) = x$ *for all* $x \in [0, \infty)$.

(ii) **Homogeneity.** $\Lambda(t\mathbf{x}) = t\Lambda(\mathbf{x})$ *for all* $t > 0$ *and* $\mathbf{x} \in \bar{\mathbb{R}}^2_+$. *Hence,* Λ *is uniquely determined by its values on the unit circle.*

(iii) **Fréchet-Hoeffding bounds.** $0 \leq \Lambda(\mathbf{x}) \leq x_1 \wedge x_2$ *for all* $\mathbf{x} \in \bar{\mathbb{R}}^2_+$.

(iv) **Lipschitz-continuity.** Λ *is Lipschitz-continuous with respect to the maximum norm.*

(v) **2-increasing.** Λ *is* 2*-increasing.*

(vi) **Tail independence.** *If* Λ *is non-zero in a single point* $\mathbf{x} \in [0, \infty)^2$, *then it is non-zero everywhere on* $[0, \infty)^2$. *Hence,* $\Lambda|_{[0,\infty)^2} \equiv 0$ *in case of tail independence.*

These assertions are proved in [Schmidt and Stadtmüller, 2006].

In the following we will investigate the tail copulas (and the stable tail dependence function) within the framework of multivariate extreme value theory. It will turn out that the basis assumption of extreme value theory, namely that F lies in the maximum domain of attraction of an extreme value distribution G, is sufficient for the upper tail copula of F to exist everywhere on $\bar{\mathbb{R}}^2_+$. Even more, it suffices that the copula of F lies in some domain of attraction. Furthermore, the upper tail copula characterizes the copula of the limiting distribution G and therefore, in a way, entirely measures extremal dependence of F.

Definition 1.8 (Maximum domain of attraction, extreme value distribution, extreme value copulas)
Let F and G be d-dimensional distribution functions, where G is supposed to have non-degenerate marginals, and let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})^T$ be *i.i.d.* distributed according to F. If there exist sequences $a_{nj} > 0$ and b_{nj} real, $j = 1 \ldots, d$ such that

$$\lim_{n\to\infty} F^n(a_{n1}x_1 + b_{n1}, \ldots, a_{nd}x_d + b_{nd}) = G(\mathbf{x}) \tag{1.3}$$

for every continuity point $\mathbf{x} = (x_1, \ldots, x_d)^T$ of G then F is said to *lie in the maximum domain of attraction of* G. We denote this by $F \in D(G)$. Equivalently,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\max_{i=1}^n X_{i1} - b_{n1}}{a_{n1}} \leq x_1, \ldots, \frac{\max_{i=1}^n X_{id} - b_{nd}}{a_{nd}} \leq x_d\right) = G(\mathbf{x}).$$

The possible limit distributions are called *extreme value distributions* and for $d \geq 2$ their copulas are called *extreme value copulas*.

The well-known Fisher-Tippet Theorem completely characterizes the possible limit distributions in the one-dimensional setting. Surprisingly, all one-dimensional extreme value distributions can be parametrized by only three real-valued parameters.

Theorem 1.9 (Fisher and Tippet (1928))
Every one-dimensional extreme value distribution is of the form $G_{\gamma,a,b}(x) = G_\gamma(ax+b)$ with $a > 0$, $b \in \mathbb{R}$, where

$$G_\gamma(x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right), \qquad 1+\gamma x > 0,$$

with $\gamma \in \mathbb{R}$ and where for $\gamma = 0$ the right-hand side is interpreted as $\exp(-e^{-x})$.

Obviously, $F \in D(G_{\gamma,a,b})$ for some $a > 0, b \in \mathbb{R}$ is equivalent to $F \in D(G_\gamma)$. Section 1 in [de Haan and Ferreira, 2006] contains sufficient and necessary conditions on F to lie in the domain of attraction of some G_γ. From a statistical point of view, the finite-dimensionality of the family of extreme value distributions allows for parametric estimation methods which are usually more precise than nonparametric methods.

Unfortunately, in the multivariate setting the characterization of extreme value distribution turns out to be more complicated: they depend on a parametric and a nonparametric component. As one might expect the parametric component stems from the marginal behavior. We consider the two-dimensional case only and denote by F_1, F_2 (G_1, G_2) the marginals of F (resp. G). Since (1.3) implies convergence of the two marginal distributions we can choose the normalizing sequences $a_{n1}, a_{n2}, b_{n1}, b_{n2}$ in such a way that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\max_{i=1}^n X_{ij} - b_{nj}}{a_{nj}} \le x_j\right) = G_j(x_j) = \exp\left(-(1+\gamma_j x_j)^{-1/\gamma_j}\right)$$

for some $\gamma_j \in \mathbb{R}$ and $j = 1,2$. Unlike in the copula context, where marginals are standardized to uniform distributions, it is usual in multivariate EVT to standardize the marginals to Pareto or Fréchet distributions. This will yield a simplified limit relation. Using the notation $H^{-1}(x) = \inf\{y : H(y) \ge x\}$ for the left-continuous inverse of a nondecreasing function H, we set for $j = 1,2$

$$U_j(x) = \left(\frac{1}{1-F_j}\right)^{-1}(x) = F_j^{-1}\left(1-\frac{1}{x}\right), \qquad x > 1,$$
$$\psi_j(x) = \left(\frac{1}{-\log G_j}\right)^{-1}(x) = \frac{x^{\gamma_j}-1}{\gamma_j}, \qquad x > 0,$$

and transform F and G to F_0 and G_0 by

$$F_0(\mathbf{x}) = F(U_1(x_1), U_2(x_2)), \qquad G_0(\mathbf{x}) = G(\psi_1(x_1), \psi_2(x_2)).$$

In case F has continuous marginal distribution functions F_1 and F_2 the latter can be formulated as

$$F_0(\mathbf{x}) = \mathbb{P}\left(\frac{1}{1-F_1(X_1)} \le x_1, \frac{1}{1-F_2(X_2)} \le x_2\right) \tag{1.4}$$
$$G_0(\mathbf{x}) = \mathbb{P}\left(\frac{1}{-\log G_1(Z_1)} \le x_1, \frac{1}{-\log G_2(Z_2)} \le x_2\right), \tag{1.5}$$

where $\mathbf{X} = (X_1, X_2)^T \sim F$ and $\mathbf{Z} = (Z_1, Z_2)^T \sim G$. Bearing in mind that for a real random variable Y distributed according to some continuous distribution function H it holds $\mathbb{P}\left(1/(1 - H(Y)) \le y\right) = 1 - 1/y$ (the Pareto-distribution) and $\mathbb{P}\left(1/(-\log H(Y)) \le y\right) = \exp(-1/y)$ (the Fréchet-distribution), the identities (1.4) and (1.5) can be interpreted as a standardization of the marginals of F to Pareto-distributions and of the marginals of G to Fréchet-distributions.
Next, passing from F to F_0 and from G to G_0 yields the following simplified limit relation

$$\lim_{n\to\infty} F_0^n(n\mathbf{x}) = G_0(\mathbf{x}), \tag{1.6}$$

i.e. F_0 lies in the domain of attraction of G_0. This can be proved like following: Since $\lim_{n\to\infty} F_j^n(a_{nj}x_j + b_{nj}) = G_j(x_j)$ we obtain, by taking logarithms and exploiting the fact that $-\log z \sim 1 - z$ for $z \to 1$, that

$$\lim_{n\to\infty} n(1 - F_j(a_{nj}x_j + b_{nj})) = -\log G_j(x_j).$$

Passing to reciprocals and invoking an inversion formula (see Lemma 1.1.1 in [de Haan and Ferreira, 2006]) we obtain

$$\lim_{n\to\infty} \frac{U_j(nx_j) - b_{nj}}{a_{nj}} = \psi_j(x_j).$$

By continuity of G and monotonicity of F it follows that

$$F_0^n(n\mathbf{x}) = \mathbb{P}\left(\frac{\max X_{i1} - b_{n1}}{a_{n1}} \le \frac{U_1(nx_1) - b_{n1}}{a_{n1}}, \frac{\max X_{i2} - b_{n2}}{a_{n2}} \le \frac{U_2(nx_2) - b_{n2}}{a_{n2}}\right)$$

converges to $G(\psi_1(x_1), \psi_2(x_2)) = G_0(\mathbf{x})$ as asserted.

The following Proposition (see Proposition 5.15 in [Resnick, 1987]) shows that convergence of $F_0^n(n\mathbf{x})$ to $G_0(\mathbf{x})$, together with marginal convergence is also sufficient for $F \in D(G)$.

Proposition 1.10
$F \in D(G)$ if and only if $F_0 \in D(G_0)$, $F_1 \in D(G_1)$ and $F_2 \in D(G_2)$.

Therefore, in order to completely characterize bivariate extreme value distributions, it remains to characterize the possible limits G_0 having standard Fréchet-marginals. Since (1.6) implies

$$G_0(\mathbf{x}) = \lim_{n\to\infty} F_0^{nt}(nt\mathbf{x}) = \lim_{n\to\infty} F_0^n(nt\mathbf{x})^t = G_0(t\mathbf{x})^t \tag{1.7}$$

for every $t > 0$, we can conclude that $G_0^{1/t}(\mathbf{x}) = G_0(t\mathbf{x})$ is a distribution function. This property is called max-infinitely divisibility and allows for the definition of measures $\nu_n = nG_0^{1/n}$ on $\mathbb{E} = [0, \infty]^2 \setminus \{\mathbf{0}\}$, where lines through infinity are supposed to have measure 0. The topology of $\mathbb{E}$ is obtained by removing $\mathbf{0}$ from the compact set $[0,\infty]^2$ and may be metrized by

$$d(\mathbf{x},\mathbf{y}) = \left|\frac{1}{x_1 \vee x_2} - \frac{1}{y_1 \vee y_2}\right| + \left|\left|\frac{\mathbf{x}}{x_1 \vee x_2} - \frac{\mathbf{y}}{y_1 \vee y_2}\right|\right|_2,$$

see e.g. page 225 in [Resnick, 1987]. Note that subsets of $\mathbb{E}$ are compact if and only if they are closed and bounded away from $\mathbf{0}$. Since $-\log z \sim 1 - z$ for $z \to 1$ we obtain

$$\lim_{n\to\infty} \nu_n([\mathbf{0},\mathbf{x}]^C) = \lim_{n\to\infty} n(1 - G_0^{1/n}(\mathbf{x})) = -\log G_0(\mathbf{x}) \tag{1.8}$$

for every $\mathbf{x} \in \mathbb{E}$ and therefore

$$\sup_{n\in\mathbb{N}} \nu_n([\mathbf{0},\mathbf{x}]^C) < \infty.$$

Notice that $\mathbb{E} = \lim_{\mathbf{x}\to\mathbf{0}}[\mathbf{0},\mathbf{x}]^C$ and conclude that $\sup_{n\in\mathbb{N}} \nu_n(B) < \infty$ for every relatively compact subset of $\mathbb{E}$. By Proposition 3.16 in [Resnick, 1987] the sequence ν_n is vaguely relatively compact and by identity (1.8) all accumulation points of vaguely convergent subsequences are the same. Therefore, there exists a limit point $\nu = \lim_{n\to\infty} \nu_n$, which satisfies

$$G_0(\mathbf{x}) = \exp(-\nu([\mathbf{0},\mathbf{x}]^C)).$$

Relation (1.7) may be translated into a homogeneity property for ν

$$\nu([\mathbf{0},\mathbf{x}]^C) = t\nu([\mathbf{0},t\mathbf{x}]^C) = t\nu(t[\mathbf{0},\mathbf{x}]^C),$$

and the latter can be shown to hold for every Borel subset of $\mathbb{E}$. This yields the following Proposition.

Proposition 1.11
Suppose that (1.3) *holds. Then there exists an exponent measure ν defined on $\mathbb{E}$ such that the standardized limit distribution $G_0(\mathbf{x}) = G(\psi_1(x_1), \psi_2(x_2))$ satisfies*

$$G_0(\mathbf{x}) = \exp(-\nu([\mathbf{0},\mathbf{x}]^C))$$

for all $\mathbf{x} \in \mathbb{E}$. The measure ν is homogeneous in the sense that

$$\nu(B) = t\nu(tB) \qquad \text{for every Borel subset } B \subset \mathbb{E} \text{ and every } t > 0$$

and puts no mass on the lines through ∞.

The literature provides various estimators for ν, see for example [de Haan and Resnick, 1993]. The homogeneity property of ν enables us to estimate the probability of events that contain no data, i.e. of *rare events*.

The following proposition will clarify the relationship between the exponent measure of G_0 and the stable tail dependence function l or the upper tail copula Λ_U. Before its statement note that if $F = C(F_1, F_2)$ has continuous marginals F_1, F_2 then

$$F_0(\mathbf{x}) = F(F_1^-(1-1/x_1), F_2^-(1-1/x_2)) = C(1-1/x_1, 1-1/x_2) = C_0(\mathbf{x}).$$

Hence, by (1.6) if both $F \in D(G)$ and $C \in D(H)$ the corresponding exponent measures of G and H coincide.

Proposition 1.12
Suppose that $F = C(F_1, F_2)$ has continuous marginals F_1, F_2 and that the copula C lies in the domain of attraction of some extreme value distribution G. Then the stable tail dependence function l (and the upper tail copula Λ_U) exists everywhere on $\mathbb{R}_+^2$ and satisfies

$$l(\mathbf{x}) = x_1 + x_2 - \Lambda_U(\mathbf{x}) = \nu([\mathbf{0}, 1/\mathbf{x}]^C),$$

where $1/\mathbf{x} = (1/x_1, 1/x_2)^T$ and ν denotes the exponent measure of G_0.

Proof. Exploiting the fact that $-\log z \sim 1 - z$ for $z \to 1$ we obtain from (1.8) that

$$\lim_{n\to\infty} n(1 - C_0(n\mathbf{x})) = -\log G_0(\mathbf{x}) = \nu([\mathbf{0}, \mathbf{x}]^C).$$

This yields

$$\begin{aligned} l(\mathbf{x}) &= x_1 + x_2 - \lim_{n\to\infty} n\bar{C}(\mathbf{x}/n) = \lim_{n\to\infty} n(1 - C(1 - x_1/n, 1 - x_2/n)) \\ &= \lim_{n\to\infty} n(1 - C_0(n/x_1, n/x_2)) = \nu([\mathbf{0}, 1/\mathbf{x}]^C) \end{aligned}$$

as asserted. □

Remark 1.13
For proving the existence of the upper tail copula, we did not need the assumption that F itself lies in the domain of attraction of some extreme value distribution. It suffices that C does so, this is a crucial point for practical applications where marginals are often treated separately from the copula.

Not every measure with the requirements of Proposition 1.11 defines an extreme value distribution G_0. Hence, the exponent measure does not describe G_0 in a simple one-to-one manner. In the following we will fill this gap and exemplarily state three objects that completely characterize extreme value distributions.

First of all, the stable tail dependence function allows for a characterization of G_0 via convexity of its level sets, see p. 223 in [de Haan and Ferreira, 2006].

A second characterization is given by the *spectral measure*. The polar coordinate transformation

$$T : \mathbb{R}_+^2 \to \mathbb{R}_+ \times [0, \pi/2], \quad \mathbf{x} \mapsto (||\mathbf{x}||_2, \arctan(x_2/x_1))^T$$

translates the exponent measure ν into a product measure by utilizing the homogeneity property, see page 215 in [de Haan and Ferreira, 2006]. It yields the following Theorem.

Theorem 1.14
For any extreme value distribution G_0 with standard Fréchet marginals there exists a finite measure Ψ on the set $[0, \pi/2]$, called spectral measure, with the property that

$$G_0(\mathbf{x}) = \exp\left(-\int_0^{\pi/2} \left(\frac{\cos\theta}{x_1} \vee \frac{\sin\theta}{x_2}\right) \Psi(d\theta)\right). \tag{1.9}$$

The measure Ψ satisfies the side conditions

$$\int_0^{\pi/2} \cos\theta\, \Psi(d\theta) = \int_0^{\pi/2} \sin\theta\, \Psi(d\theta) = 1.$$

Conversely, any finite measure Ψ on $[0, \pi/2]$ fulfilling the side conditions gives rise to an extreme value distribution G_0 with standard Fréchet marginals via (1.9).

A third characterization is given by the *Pickands dependence function*. It is defined for $t \in [0, 1]$ by

$$A(t) = -\log G_0\left(\frac{1}{1-t}, \frac{1}{t}\right) = l(1-t, t).$$

By the homogeneity of $-\log G_0$ (which follows by the max-infinitely divisibility of G_0) the function A determines G_0:

$$G_0(\mathbf{x}) = \exp\left(-\frac{x_1 + x_2}{x_1 x_2} A\left\{\frac{x_1}{x_1 + x_2}\right\}\right).$$

It can be shown that A is convex and satisfies the boundary conditions $(1-t) \vee t \leq A(t) \leq 1$. Conversely, any function A with these properties defines an extreme value distribution, see [Pickands, 1981].

Note that for the characterization of bivariate extreme value distributions one can use any of the three objects mentioned above.

In the remaining part of this section we will draw the attention to extreme value copulas. Recall that C is an extreme value copula if and only if it is the copula of an extreme value distribution.

Proposition 1.15
The following conditions are equivalent.

(i) *C is an extreme value copula.*

(ii) *For all $t > 0$ and all $\mathbf{u} \in [0, 1]^2$ it holds* $C(\mathbf{u}^t) = C^t(\mathbf{u})$.

(iii) *There exists a convex function $A : [0, 1] \to [0, 1/2]$ satisfying $(1-t) \vee t \leq A(t) \leq 1$ for all $t \in [0, 1]$ such that*

$$C(\mathbf{u}) = \exp\left(\log(u_1 u_2) A\left\{\frac{\log u_2}{\log u_1 u_2}\right\}\right).$$

A is called Pickands dependence function.

Proof. Exploiting the fact that

$$C(\mathbf{u}) = G(G_1^-(u_1), G_2^-(u_2)) = G_0(-1/\log u_1, -1/\log u_2)$$

the homogeneity property of G_0 translates into

$$C(\mathbf{u}^t) = G_0\left(\frac{-1}{t\log u_1}, \frac{-1}{t\log u_2}\right) = G_0^t\left(\frac{-1}{\log u_1}, \frac{-1}{\log u_2}\right) = C^t(\mathbf{u}). \tag{1.10}$$

Therefore (i) implies (ii). For the converse, set $G_0(\mathbf{x}) = C(e^{-1/x_1}, e^{-1/x_2})$. A similar calculation as in (1.10) reveals that G_0 is homogeneous in the sense that $G_0^n(n\mathbf{x}) = G_0(\mathbf{x})$ for all $n \in \mathbb{N}$. Hence, $G_0 \in D(G_0)$ which yields the assertion. Moreover, an easy calculation shows that (iii) implies (ii). If C is an extreme value copula then $G_0(\mathbf{x}) = C(e^{-1/x_1}, e^{-1/x_2})$ is an extreme value distribution with standard Fréchet marginals. The characterization of those distributions by Pickands dependence functions yields (iii). $\square$

1.3 Weak convergence in metric spaces

Let $(\mathbb{D}, d)$ be a metric space and let $(P_n)_{n\in\mathbb{N}}$ and P be Borel probability measures on $(\mathbb{D}, \mathcal{D})$, where $\mathcal{D}$ denotes the Borel σ-field on $\mathbb{D}$. Weak convergence of P_n to P, which we write as $P_n \rightsquigarrow P$, is classically defined by the requirement that

$$\int_{\mathbb{D}} f\,dP_n \to \int_{\mathbb{D}} f\,dP \qquad \text{for all } f \in C_b(\mathbb{D}),$$

where $C_b(\mathbb{D})$ denotes the set of all bounded, continuous and real-valued functions on $\mathbb{D}$ (see e.g. [Billingsley, 1968]). For $\mathbb{D}$-valued random variables $(X_n)_{n\in\mathbb{N}}$ and X weak convergence is led back to their induced laws, so that $X_n \rightsquigarrow X$ if and only if

$$\mathbb{E}f(X_n) \to \mathbb{E}f(X) \qquad \text{for all } f \in C_b(\mathbb{D}). \tag{1.11}$$

This classical theory of weak convergence requires that all random variables involved are Borel measurable. While this condition usually holds for separable metric spaces such as $\mathbb{R}^d$ or $C[0,1]$, it easily fails when the metric spaces are nonseparable. The classical example is the càdlàg-space $D[0,1]$, containing all functions on the unit interval which are right-continuous and possess left-hand limits, equipped with the metric induced by the supremum norm. For *i.i.d.* random variables $X_1, \ldots, X_n$ on $[0,1]$ the empirical distribution function

$$F_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{X_i \le t\}, \qquad t \in [0,1],$$

as well as the empirical process

$$X_n(t) = \sqrt{n}(F_n(t) - F(t)), \qquad t \in [0,1], \tag{1.12}$$

seen as random variables in $D[0,1]$, are not Borel measurable if $D[0,1]$ is equipped with the supremum norm, see [Billingsley, 1968].
During the last decades several approaches to overcome this difficulty were suggested. In this section we will briefly summarize the most modern approach, going back to J. Hoffmann-Jørgensen and extensively investigated in [van der Vaart and Wellner, 1996; Kosorok, 2008].
The key idea is to drop the requirement of Borel measurability of each X_n, meanwhile upholding the requirement (1.11), where the expectations are replaced by outer expectations.

Definition 1.16 (Outer integral and outer probability)
Let T be an arbitrary map from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to the extended real line $\bar{\mathbb{R}}$. The *outer integral* of T with respect to $\mathbb{P}$ is defined as

$$\mathbb{E}^*T = \inf\left\{\mathbb{E}U : \ U \geq T, \ U : \Omega \to \bar{\mathbb{R}} \text{ measurable and } \mathbb{E}U \text{ exists}\right\}.$$

The *outer probability* of an arbitrary subset $B \subset \Omega$ is defined as

$$\mathbb{P}^*(B) = \inf\left\{\mathbb{P}(A) : \ A \supset B, \ A \in \mathcal{A}\right\}.$$

Inner integrals and *inner probabilities* are defined by $\mathbb{E}_*T = -\mathbb{E}^*(-T)$ and $\mathbb{P}_*(B) = 1 - \mathbb{P}^*(\Omega \setminus B)$.

The infima in the latter definitions are always achieved, see the following Lemma, which is proved in [van der Vaart and Wellner, 1996].

Lemma 1.17 (Measurable cover functions)
For any map $T : \Omega \to \bar{\mathbb{R}}$ there exists a measurable function $T^* : \Omega \to \bar{\mathbb{R}}$ with $T^* \geq T$ and $T^* \leq U$ a.s. for every measurable $U : \Omega \to \bar{\mathbb{R}}$ with $U \geq T$ a.s. For every such T^* it holds $\mathbb{E}^*T = \mathbb{E}T^*$, provided that $\mathbb{E}T^*$ exists. The latter is certainly true if $\mathbb{E}^*T < \infty$.

With Definition 1.16 and Lemma 1.17 at hand we can define weak convergence, outer almost sure convergence and convergence in outer probability for arbitrary, nonmeasurable maps.

Definition 1.18 (Convergence: Weak, outer almost surely and in outer probability)
Let $X_n : \Omega_n \to \mathbb{D}, X : \Omega \to \mathbb{D}$ be arbitrary maps defined on some probability spaces $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$, $(\Omega, \mathcal{A}, \mathbb{P})$.

(i) If X is Borel measurable we say that X_n *weakly converges to* X, written $X_n \rightsquigarrow X$, if and only if

$$\mathbb{E}^*f(X_n) \ \to \ \mathbb{E}f(X) \qquad \text{for all } f \in C_b(\mathbb{D}).$$

(ii) If X_n, X are defined on a common probability space we say that X_n *converges outer almost surely to* X if $d(X_n, X)^* \to 0$ almost surely for some version of $d(X_n, X)^*$. This is denoted by $X_n \overset{\text{as}*}{\to} X$.

(iii) If X_n, X are defined on a common probability space we say that X_n *converges in outer probability to* X if $d(X_n, X)^* \to 0$ in probability. This is equivalent to $\mathbb{P}^*(d(X_n, X) > \varepsilon) \to 0$ for every $\varepsilon > 0$ and is denoted by $X_n \overset{\mathbb{P}}{\to} X$.

With this definition much of the theory for non-measurable maps parallels the classical theory to a remarkable degree. For example, the parallels include a Portmanteau Theorem, continuous mapping results, a Prohorov Theorem and the "metrization" of weak convergence to separable limits by the bounded Lipschitz-metric. The latter is developed in Section 1.12 in [van der Vaart and Wellner, 1996] and states that $X_n \rightsquigarrow X$, where X is Borel measurable and separable if and only if

$$\sup_{f \in BL_1(\mathbb{D})} |\mathbb{E}^* f(X_n) - \mathbb{E} f(X)| \to 0, \tag{1.13}$$

where $BL_1(\mathbb{D})$ denotes the set of all real functions on $\mathbb{D}$ which are bounded by 1 and satisfy the Lipschitz condition $|f(x) - f(y)| \leq d(x, y)$ for all $x, y \in \mathbb{D}$.
Two further important properties for the investigation of stochastic convergence of non-measurable maps are summarized in the following definition.

Definition 1.19
Let $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ be a sequence of probability spaces and let $X_n : \Omega_n \to \mathbb{D}$ be arbitrary maps.

(i) $(X_n)_{n\in\mathbb{N}}$ is *asymptotically measurable* if $\mathbb{E}^* f(X_n) - \mathbb{E}_* f(X_n) \to 0$ for every $f \in C_b(\mathbb{D})$.

(ii) $(X_n)_{n\in\mathbb{N}}$ is *asymptotically tight* if for every $\varepsilon > 0$ there exists a compact set $K \subset \mathbb{D}$ with $\liminf_{n\to\infty} \mathbb{P}_*(X_n \in G) \geq 1 - \varepsilon$ for every open $G \supset K$.

By Lemma 1.3.8 in [van der Vaart and Wellner, 1996] weak convergence of $(X_n)_n$ to a tight limit implies both asymptotic tightness and asymptotic measurability.
In the remainder of this section we consider the important special case $\mathbb{D} = l^\infty(T)$, where $l^\infty(T)$ is the set of all uniformly bounded, real-valued functions on a given fixed set T. The metric d is the metric induced by the supremum norm $||z||_t = \sup_{t\in T} |z(t)|$. In this case weak convergence of the finite dimensional distributions (i.e. of all vectors $(X_n(t_1), \ldots, X_n(t_k)) \in \mathbb{R}^k$ with $t_i \in T$ and $k \in \mathbb{N}$) and asymptotic tightness are sufficient for weak convergence in $l^\infty(T)$.
The most important example of a sequence of maps in a space of the latter form is the empirical process. Given a sample $X_1, \ldots, X_n$ of *i.i.d.* random variables with distribution P on an arbitrary sample space $\mathcal{X}$ we define the *empirical measure* as $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$,

where δ_x is the dirac measure at x. For $f \in \mathcal{F} = \{f : \mathcal{X} \to \mathbb{R}\}$ we set $\mathbb{P}_n f = n^{-1} \sum_{i=1}^n f(X_i)$ and define the *empirical process* as

$$\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - P)f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i) - Pf,$$

which can be seen as an element of $l^\infty(\mathcal{F})$ provided

$$\sup_{f \in \mathcal{F}} |f(x) - Pf| < \infty \qquad \text{for every } x.$$

Chapter 2 in [van der Vaart and Wellner, 1996] assembles conditions under which

$$\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P) \rightsquigarrow \mathbb{G} \qquad \text{in } l^\infty(\mathcal{F}),$$

where the limit $\mathbb{G}$ is a tight Borel measurable element in $l^\infty(\mathcal{F})$. Considering marginal convergence one can conclude that $\mathbb{G}$ must be a centered Gaussian process with covariance function $\mathbb{E}[\mathbb{G}f\mathbb{G}g] = Pfg - PfPg$. A class $\mathcal{F}$ with the property that the empirical process weakly converges to a tight limit is called *P-Donsker*.
Since $D[0,1] \subset l^\infty[0,1]$ the example of the empirical process $X_n(t)$ as defined in (1.12) can be embedded in the preceding context by identifying $l^\infty[0,1]$ with $l^\infty(\mathcal{F})$ where $\mathcal{F} = \{\mathbb{I}_{[0,t]} : t \in [0,1]\}$. More generally the same identification can be made for any subset of $\mathbb{R}^d$ which will be frequently considered in the subsequent sections.
Since the limit process $\mathbb{G}$ of the empirical process depends on the (unknown) distribution P various bootstrap procedures have been proposed in order to approximate the distribution of $\mathbb{G}$. The usual resampling bootstrap is based on the bootstrap empirical measure $\hat{\mathbb{P}}_n f = n^{-1} \sum_{i=1}^n W_{ni} f(X_i)$, where $W_n = (W_{n1}, \ldots, W_{nn})$ is a multinomial random vector with parameter n and success probabilities $(1/n, \ldots, 1/n)$, independent of the data $X_1, \ldots, X_n$. The conditional distribution of $\hat{\mathbb{G}}_n = \sqrt{n}(\hat{\mathbb{P}}_n - \mathbb{P}_n)$ given $X_1, X_2, \ldots$ can be shown to converge weakly to the law of $\mathbb{G}$ in the following sense, see [Kosorok, 2008].

Definition 1.20
Let $\hat{Z}_n = \hat{Z}_n(X_1, \ldots, X_n, M_1, \ldots, M_n) : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{D}$ be a bootstrapped process depending on some random variables $X_1, \ldots, X_n$ and some random weights M. If Z is a tight process in $\mathbb{D}$ then $\hat{Z}_n$ *weakly converges to* Z *given* $X_1, X_2, \ldots$ *in probability* if and only if

(i) $\sup_{f \in BL_1(\mathbb{D})} \left| \mathbb{E}_M f(\hat{Z}_n) - \mathbb{E} f(Z) \right| \xrightarrow{\mathbb{P}} 0,$

(ii) $\mathbb{E}_M f(\hat{Z}_n)^* - \mathbb{E}_M f(\hat{Z}_n)_* \xrightarrow{\mathbb{P}} 0$ for all $f \in BL_1(\mathbb{D})$,

where the subscript M indicates conditional expectation over the weights M given the remaining data and where $h(\hat{Z}_n)^*$ and $h(\hat{Z}_n)_*$ denote measurable majorants and minorants with respect to the joint data. We denote this convergence by $\hat{Z}_n \overset{\mathbb{P}}{\underset{M}{\rightsquigarrow}} Z$.

Note that this definition is motivated by the "metrization" of weak convergence by the bounded Lipschitz metric, see (1.13), and by the asymptotic measurability of weakly convergent sequences. Its usefulness is explained in [Giné and Zinn, 1990] where it is demonstrated that the concept of weak convergence conditional on the data in probability allows for the construction of asymptotic confidence regions for P.
For further results we refer the reader to the monographs [van der Vaart and Wellner, 1996; Kosorok, 2008].

Chapter 2

Bootstrap approximations for the empirical copula process

2.1 Introduction

The empirical copula C_n is the most famous and easiest nonparametric estimator for the copula C of a random vector. It is well known that the standardized process $\sqrt{n}(C_n - C)$ converges weakly towards a Gaussian field $\mathbb{G}_C$ with covariance structure depending on the unknown copula and its derivatives; see e.g. [Fermanian et al., 2004]. Because these quantities are usually difficult to estimate several authors have suggested approximating the limit distribution using bootstrap procedures. [Fermanian et al., 2004] proposed a bootstrap procedure based on resampling and proved its consistency. A wild bootstrap approach based on the multiplier method was recently proposed by [Rémillard and Scaillet, 2009] and applied to the problem of testing the equality between two copulas. Recently [Kojadinovic and Yan, 2010] and [Kojadinovic et al., 2010] used the same method to construct a goodness-of-fit test for the parametric form of a copula.
The present chapter of this work has three purposes. First of all, [Rémillard and Scaillet, 2009] only investigate the multiplier bootstrap approach unconditionally, while usually bootstrap results are stated conditionally given the observed data. Clearly, the latter concept fits better to statistical applications since, loosely spoken, it deletes any randomness being descended from the data. Usually observing only one data set in practical applications this is a desirable demand on bootstrap results.
Secondly, we propose two modifications of the multiplier bootstrap approach developed in [Rémillard and Scaillet, 2009]. Note that the latter approach requires the estimation of the partial derivatives of the unknown copula. Bearing in mind that estimating derivatives usually is a difficult task we propose a modification, which avoids the estimation problem. Moreover, the results of [Rémillard and Scaillet, 2009] are extended to the case of non-i.i.d. multinomial multipliers.
Finally, we investigate the finite sample properties of the proposed bootstrap procedures. In particular it is demonstrated that despite the fact that the modified multiplier method and the bootstrap based on resampling avoid the problem of estimating derivatives, the

procedure proposed in [Rémillard and Scaillet, 2009] and its extension to multinomial multipliers yield the best approximations in most cases.

2.2 The empirical copula process and four bootstrap approximations

For the sake of brevity, we restrict ourselves to the case of a bivariate copula, but all results can easily be transferred to higher dimensions. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent identically distributed bivariate random vectors with cumulative distribution function (cdf) F, continuous marginal distribution functions F_1 and F_2 and copula C. Due to Sklar's Theorem 1.2 there is the relationship

$$C(\mathbf{u}) = F(F_1^-(u_1), F_2^-(u_2)), \tag{2.1}$$

where $\mathbf{u} = (u_1, u_2)^T \in [0,1]^2$ and $H^-(u) = \max\{\inf\{t \in \mathbb{R} | H(t) \geq u\}, \inf \operatorname{supp} H\}$ denotes the generalized inverse of a real distribution function H. The empirical copula as the simplest nonparametric estimator for C (going back to [Deheuvels, 1979]) simply replaces the unknown terms in equation (2.1) by their empirical counterparts, that is

$$C_n(\mathbf{u}) = F_n(F_{n1}^-(u_1), F_{n2}^-(u_2)) \tag{2.2}$$

where

$$F_n(\mathbf{x}) = F_n(x_1, x_2) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{X_{i1} \leq x_1, X_{i2} \leq x_2\},$$

$$F_{np}(x_p) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{X_{ip} \leq x_p\}, \ \ p = 1, 2$$

denote the corresponding empirical distribution functions. The asymptotic behavior of C_n was studied in several papers, including [Gänssler and Stute, 1987; Ghoudi and Rémillard, 2004; Tsukahara, 2005; Fermanian et al., 2004] among others. It is remarkable that all the results in these papers can be substantially weakened regarding the first order properties of the copula: instead of requiring continuous partial derivatives of C on the whole unit square we only require $\partial_1 C$ being continuous on $(0,1) \times [0,1]$ and $\partial_2 C$ being continuous on $[0,1] \times (0,1)$, see [Segers, 2010]. The importance of this simplification can hardly be underestimated since most of the common copula families have discontinuous partial derivatives in some boundary points. For example, the only extreme value copula possessing continuous partial derivatives on the whole unit square is the independence copula, see [Segers, 2010].
Throughout this chapter $\rightsquigarrow$ denotes weak convergence in the metric space $l^\infty([0,1]^2)$ of all uniformly bounded functions on the unit square $[0,1]^2$ equipped with the metric induced by the supremum norm. Weak convergence is understood in the sense of Hoffmann-Jørgensen, see Section 1.3. The following Theorem is proved in [Segers, 2010] and an alternative proof based on the functional delta method is given in Section 2.3.

Theorem 2.1 (Segers, 2010)
If the Copula C satisfies the following first order property

$$\partial_p C(\mathbf{u}) \text{ exists and in continuous on } \{\mathbf{u} \in [0,1]^2 \mid u_p \in (0,1)\} \tag{2.3}$$

$(p = 1,2)$, *then the empirical copula process* $\sqrt{n}(C_n - C)$ *weakly converges towards a Gaussian field* $\mathbb{G}_C$,

$$\alpha_n = \sqrt{n}(C_n - C) \quad \rightsquigarrow \quad \mathbb{G}_C \quad in \quad l^\infty([0,1]^2).$$

The limiting process can be represented as

$$\mathbb{G}_C(\mathbf{u}) = \mathbb{B}_C(\mathbf{u}) - \partial_1 C(\mathbf{u})\mathbb{B}_C(u_1,1) - \partial_2 C(\mathbf{u})\mathbb{B}_C(1,u_2), \tag{2.4}$$

where $\partial_p C$, $p = 1,2$ *is defined as* 0 *on the set* $\{\mathbf{u} \in [0,1]^2 \mid u_p \in \{0,1\}\}$ *and* $\mathbb{B}_C$ *denotes a centered Gaussian field with covariance structure*

$$\tilde{r}(\mathbf{u},\mathbf{v}) = \mathrm{Cov}(\mathbb{B}_C(\mathbf{u}),\mathbb{B}_C(\mathbf{v})) = C(\mathbf{u} \wedge \mathbf{v}) - C(\mathbf{u})C(\mathbf{v}).$$

The minimum in the last expression is understood component-wise.

The literature provides several similar nonparametric estimators for the copula. For example, [Genest et al., 1995] studied the rank-based estimator

$$\bar{C}_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{F_{n1}(X_{i1}) \le u_1, F_{n2}(X_{i2}) \le u_2\}.$$

In the latter expression the marginal edfs F_{np} are often replaced by their rescaled counterparts $\hat{F}_{np} = \frac{n}{n+1}F_{np}$. Both modifications do not affect the asymptotic behavior, see [Fermanian et al., 2004]. See also [Chen and Huang, 2007; Omelka et al., 2009] for a smoothed version of this process.

The limiting Gaussian variable $\mathbb{G}_C$ depends on the unknown copula C and for this reason it is not directly applicable for statistical inference. In the following discussion we will present two known and two new bootstrap approximations for the distribution of the limiting process. We begin with the usual bootstrap based on resampling , which was proposed in [Fermanian et al., 2004]. To be precise, let $W_n = (W_{n1}, \dots, W_{nn})$ be multinomial distributed random vectors with parameter n and success probabilities $(1/n, \dots, 1/n)$ and set

$$C_n^{W,W}(\mathbf{u}) = F_n^W(F_{n1}^{W-}(u_1), F_{n2}^{W-}(u_2)),$$

where

$$F_n^W(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} W_{ni}\mathbb{I}\{X_{i1} \le x_1, X_{i2} \le x_2\},$$

$$F_{np}^W(x_p) = \frac{1}{n}\sum_{i=1}^{n} W_{ni}\mathbb{I}\{X_{ip} \le x_p\}, \qquad p = 1,2.$$

The double superscript W, W indicates that we use multinomial weights W both in the joint and the marginal ecdfs. Finally define

$$\alpha_n^{res} = \sqrt{n}(C_n^{W,W} - C_n)$$

as the bootstrap process based on resampling. For a precise statement of the asymptotic properties of this process we denote by $\overset{\mathbb{P}}{\underset{W}{\rightsquigarrow}}$ *weak convergence conditional on the data in probability* as defined in Definition 1.20. The following result has been established by [Fermanian et al., 2004], the proof follows along similar lines as the proof of Theorem 2.4 below and is given in Section 2.3.

Theorem 2.2
Under the preceding notations and assumptions the bootstrap approximation $C_n^{W,W}$ of the empirical copula yields a valid approximation of the limit variable $\mathbb{G}_C$ in the sense that

$$\alpha_n^{res} = \sqrt{n}(C_n^{W,W} - C_n) \overset{\mathbb{P}}{\underset{W}{\rightsquigarrow}} \mathbb{G}_C \quad in \quad l^\infty([0,1]^2).$$

In a recent paper [Rémillard and Scaillet, 2009] considered the problem of testing the equality between two copulas (see also [Scaillet, 2005]) and proposed a multiplier bootstrap approach to approximate the distribution of the limiting process $\mathbb{G}_C$. To be precise let $Z_1, \ldots, Z_n$ be independent identically distributed centered random variables with variance one, independent of the data $\mathbf{X}_1, \ldots, \mathbf{X}_n$, which satisfy

$$||Z||_{2,1} = \int_0^\infty \sqrt{P(|Z| > x)}\, dx < \infty.$$

Note that the latter condition is implied by a finite moment of any order $r > 2$. [Rémillard and Scaillet, 2009] defined the bootstrap process

$$C_n^*(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n Z_i \, \mathbb{I}\{F_{n1}(X_{i1}) \leq u_1, F_{n2}(X_{i2}) \leq u_2\}$$

and showed that C_n^* approximates the Gaussian field $\mathbb{B}_C$ in the sense that

$$\left(\sqrt{n}(\tilde{C}_n - C),\ \sqrt{n}(C_n^* - \bar{Z}_n C_n)\right) \quad \rightsquigarrow \quad (\mathbb{B}_C, \mathbb{B}'_C) \quad \text{in} \quad l^\infty([0,1]^2)^2,$$

where $\tilde{C}_n(\mathbf{u}) = n^{-1} \sum_{i=1}^n \mathbb{I}\{F_1(X_{i1}) \leq u_1, F_2(X_{i2}) \leq u_2\}$, $\bar{Z}_n = n^{-1} \sum_{i=1}^n Z_i$ and $\mathbb{B}'_C$ is an independent copy of $\mathbb{B}_C$. Since one is interested in an approximation of $\mathbb{G}_C$ one is able to utilize identity (2.4) by estimating the partial derivatives of the copula C. As proposed by [Rémillard and Scaillet, 2009] and more precisely by [Segers, 2010] we use

$$\widehat{\partial_1 C}(\mathbf{u}) := \begin{cases} \frac{C_n(u_1, u_2+h) - C_n(u_1, u_2-h)}{2h} & \text{if } u_1 \in [h, 1-h] \\ \frac{C_n(u_1, 2h)}{2h} & \text{if } u_1 \in [0, h) \\ \frac{u_1 - C_n(u_1, 1-2h)}{2h} & \text{if } u_1 \in (1-h, 1] \end{cases} \tag{2.5}$$

$$\widehat{\partial_2 C}(\mathbf{u}) := \begin{cases} \frac{C_n(u_1+h, u_2) - C_n(u_1-h, u_2)}{2h} & \text{if } u_2 \in [h, 1-h] \\ \frac{C_n(2h, u_2)}{2h} & \text{if } u_2 \in [0, h) \\ \frac{u_2 - C_n(1-2h, u_2)}{2h} & \text{if } u_2 \in (1-h, 1] \end{cases} \tag{2.6}$$

where $h = n^{-1/2} \to 0$ (for a smooth version of these estimators see [Scaillet, 2005]). Under strong continuity assumptions [Rémillard and Scaillet, 2009] showed that these estimates are uniformly consistent, while [Segers, 2010] was able to relax this assertion and its assumptions in a suitable manner. In order to approximate the limiting process $\mathbb{G}_C$ the authors set

$$\tilde{\alpha}_n^{pdm}(\mathbf{u}) = \tilde{\beta}_n(\mathbf{u}) - \widehat{\partial_1 C}(\mathbf{u})\tilde{\beta}_n(u_1, 1) - \widehat{\partial_2 C \mathbf{u}})\tilde{\beta}_n(1, u_2), \tag{2.7}$$

where the process $\tilde{\beta}_n$ is defined by $\tilde{\beta}_n = \sqrt{n}(C_n^* - \bar{Z}_n C_n)$. The upper index pdm in (2.7) denotes the fact that estimates of the partial derivatives and a multiplier concept are used. Proposition 4.2 in [Segers, 2010] states that under the assumptions of Theroem 2.1

$$\left(\sqrt{n}(C_n - C),\, \tilde{\alpha}_n^{pdm}\right) \quad \rightsquigarrow \quad (\mathbb{G}_C, \mathbb{G}_C') \quad \text{in} \quad l^\infty([0,1]^2)^2,$$

where $\mathbb{G}_C'$ is an independent copy of $\mathbb{G}_C$, see also [Rémillard and Scaillet, 2009] for an analogous result under stronger smoothness assumptions. To conclude, $\tilde{\alpha}_n^{pdm}$ approximates the limit distribution unconditionally. As argued in the introduction we are rather interested in conditional weak convergence given the observed data. This can be done by the following slightly modified version of the multiplier approach.

Let $\xi_1, \ldots, \xi_n$ denote independent identically distributed nonnegative random variables, independent of the data $\mathbf{X}_1, \ldots, \mathbf{X}_n$, with expectation μ and finite variance $\tau^2 > 0$ such that $||\xi||_{2,1} < \infty$. We define $\bar{\xi}_n = n^{-1} \sum_{i=1}^n \xi_i$ as the mean of $\xi_1, \ldots, \xi_n$ and consider the multiplier statistics

$$C_n^{\xi,\cdot}(\mathbf{u}) = F_n^{\xi}(F_{n1}^-(u_1), F_{n2}^-(u_2)),$$

where

$$F_n^{\xi}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{\xi_i}{\bar{\xi}_n} \mathbb{I}\{X_{i1} \le x_1, X_{i2} \le x_2\}.$$

If we standardize the ξ_i to $Z_i = (\xi_i - \mu)\tau^{-1}$ we observe that both multiplier approaches are indeed closely related by

$$\sqrt{n}\,\frac{\mu}{\tau}\,(C_n^{\xi,\cdot} - C_n) \approx \sqrt{n}\,\frac{\mu}{\bar{\xi}_n}(C_n^* - \bar{Z}_n C_n).$$

The next theorem states that the slightly modified multiplier approach $C_n^{\xi,\cdot}$ approximates the Gaussian field $\mathbb{B}_C$ conditionally on the data. Using the notations

$$\beta_n(\mathbf{u}) = \sqrt{n}\,\frac{\mu}{\tau}\,(C_n^{\xi,\cdot}(\mathbf{u}) - C_n(\mathbf{u}))$$

and

$$\alpha_n^{pdm}(\mathbf{u}) = \beta_n(\mathbf{u}) - \widehat{\partial_1 C}(\mathbf{u})\beta_n(u_1, 1) - \widehat{\partial_2 C}(\mathbf{u})\beta_n(1, u_2). \tag{2.8}$$

we obtain the following theorem, which is proved in the following section.

Theorem 2.3
Under the preceding notations and assumptions we have

$$\beta_n \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{B}_C \quad in \quad l^\infty([0,1]^2) \qquad and \qquad \alpha_n^{pdm} \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_C \quad in \quad l^\infty([0,1]^2).$$

The next resampling concept considered in this section is new and combines both approaches in order to avoid the estimation of the derivatives. On the one hand it makes use of multipliers and on the other hand it is also based on identity (2.1) and the functional delta method. To be precise we consider i.i.d. multipliers $\xi_1, \ldots, \xi_n$ as defined above and define the statistic

$$C_n^{\xi,\xi}(\mathbf{u}) = F_n^\xi(F_{n1}^{\xi-}(u_1), F_{n2}^{\xi-}(u_2)),$$

where

$$F_n^\xi(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n \frac{\xi_i}{\bar{\xi}_n}\mathbb{I}\{X_{i1} \le x_1, X_{i2} \le x_2\},$$

$$F_{np}^\xi(x_p) = \frac{1}{n}\sum_{i=1}^n \frac{\xi_i}{\bar{\xi}_n}\mathbb{I}\{X_{ip} \le x_p\}, \qquad p = 1,2.$$

As before set

$$\alpha_n^{dm} = \sqrt{n}\,\frac{\mu}{\tau}\,(C_n^{\xi,\xi} - C_n).$$

We call this bootstrap the direct multiplier method, which is reflected by the the superscript *dm* in its definition. The following result shows that the process α_n^{dm} yields a consistent bootstrap approximation of the empirical copula process conditional on the observed data. Just as the resampling bootstrap note that this approach avoids the estimation of the partial derivatives of the copula.

Theorem 2.4
Under the preceding notations and assumptions we have

$$\alpha_n^{dm} = \sqrt{n}\,\frac{\mu}{\tau}(C_n^{\xi,\xi} - C_n) \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_C \quad in \quad l^\infty([0,1]^2).$$

The last new resampling concept combines the existing approaches in a different manner. It is similar in kind to the *pdm*-bootstrap but we use multinomial weights in the joint ecdf instead of i.i.d. multipliers; we therefore call it the *pdr*-method. More precisely, we set

$$C_n^{W,\cdot}(\mathbf{u}) = F_n^W(F_{n1}^-(u_1), F_{n2}^-(u_2)),$$
$$\gamma_n(\mathbf{u}) = \sqrt{n}\,(C_n^{W,\cdot}(\mathbf{u}) - C_n(\mathbf{u})).$$

It turns out, that γ_n only approximates $\mathbb{B}_C$ and we therefore have to estimate the partial derivatives of the copula as in (2.5) and (2.6) and use

$$\alpha_n^{pdr}(\mathbf{u}) = \gamma_n(\mathbf{u}) - \widehat{\partial_1 C}(\mathbf{u})\gamma_n(u_1, 1) - \widehat{\partial_2 C}(\mathbf{u})\gamma_n(1, u_2)$$

as an approximation for $\mathbb{G}_C$. In analogy to Theorem 2.3 we obtain the following result.

Theorem 2.5
Under the preceding notations and assumptions we have

$$\gamma_n \overset{\mathbb{P}}{\underset{W}{\rightsquigarrow}} \mathbb{B}_C \quad in \quad l^\infty([0,1]^2) \qquad and \qquad \alpha_n^{pdr} \overset{\mathbb{P}}{\underset{W}{\rightsquigarrow}} \mathbb{G}_C \quad in \quad l^\infty([0,1]^2).$$

2.3 Proofs

In this section we give the proofs for Theorem 2.1 - Theorem 2.5. First note that it is sufficient to consider only the case of independent identically distributed random vectors with $\mathcal{U}[0,1]$-marginals and copula C. Indeed, let $\mathcal{U}_1,\dots,\mathcal{U}_n$ be independent identically distributed random vectors with cdf C and set

$$G_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{\mathcal{U}_{i1} \le x, \mathcal{U}_{i2} \le x_2\},$$
$$G_{np}(x_p) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{\mathcal{U}_{ip} \le x_p\}\,, \qquad p = 1,2.$$

Clearly,

$$F_n(\mathbf{x}) \overset{\mathcal{D}}{=} G_n(F_1(x_1), F_2(x_2)),$$
$$F_{np}(x_p) \overset{\mathcal{D}}{=} G_{np}(F_p(x_p))\,, \qquad p = 1,2$$

and from the definition of the generalized inverse we conclude

$$F_{np}^-(u_p) \overset{\mathcal{D}}{=} F_p^-(G_{np}^-(u_p))\,, \qquad p = 1,2,$$

so that

$$C_n(\mathbf{u}) \overset{\mathcal{D}}{=} G_n(G_{n1}^-(u_1), G_{n2}^-(u_2))$$

as asserted. An analogue result holds for $C_n^{W,W}, C_n^{\xi,\cdot}, C_n^{\xi,\xi}$ and $C_n^{W,\cdot}$ and for this reasoning we may assume in the following that $\mathbf{X}_1,\dots,\mathbf{X}_n$ are independent identically distributed distributed according to the cdf C.
Now consider the mapping

$$\Phi : \begin{cases} \mathbb{D}_\Phi & \rightarrow \quad l^\infty[0,1]^2 \\ H & \mapsto \quad H(H_1^-, H_2^-), \end{cases}$$

where $\mathbb{D}_\Phi$ denote the set of all distribution functions H on $[0,1]^2$ whose marginal cdfs $H_1 = H(\cdot,1)$ and $H_2 = H(1,\cdot)$ satisfy $H_1(0) = H_2(0) = 0$. The following Lemma is an extension of Lemma 2 in [Fermanian et al., 2004] showing Hadamard-differentiability of Φ under essentially weaker assumptions.

Lemma 2.6
Let C be a copula whose partial derivatives satisfy the following first order properties

$$\partial_p C(\mathbf{u}) \text{ exists and in continuous on } \{\mathbf{u} \in [0,1]^2 \mid u_p \in (0,1)\}.$$

Then Φ is Hadamard-differentiable at C tangentially to

$$\mathbb{D}_0 = \left\{ F \in C[0,1]^2 \mid F(0,x) = F(x,0) = 0 \text{ for all } x \in [0,1],\ F(1,1) = 0 \right\}.$$

Its derivative at C in $\alpha \in \mathbb{D}_0$ is given by

$$\left(\Phi'_C(\alpha)\right)(\mathbf{x}) = \alpha(\mathbf{x}) - \partial_1 C(\mathbf{x})\alpha(x_1, 1) - \partial_2 C(\mathbf{x})\alpha(1, x_2),$$

where $\partial_p C$, $p = 1,2$ is defined as 0 on the set $\{\mathbf{u} \in [0,1]^2 \mid u_p \in \{0,1\}\}$.

Proof. Let $\mathbb{E}$ denote the set of distribution functions F on $[0,1]$ with $F(0) = 0$ and define $\mathbb{E}^-$ as the set of all generalized inverse functions F^- with $F \in \mathbb{E}$. Now decompose $\Phi = \Phi_3 \circ \Phi_2 \circ \Phi_1$, where

$$\begin{aligned}
\Phi_1 &: \begin{cases} \mathbb{D}_\Phi & \rightarrow \quad \mathbb{D}_{\Phi_2} = \mathbb{D}_\Phi \times \mathbb{E} \times \mathbb{E} \\ H & \mapsto \quad (H, H_1, H_2) \end{cases} \\
\Phi_2 &: \begin{cases} \mathbb{D}_{\Phi_2} & \rightarrow \quad \mathbb{D}_{\Phi_3} = \mathbb{D}_\Phi \times \mathbb{E}^- \times \mathbb{E}^- \\ (H, F, G) & \mapsto \quad (H, F^-, G^-) \end{cases} \\
\Phi_3 &: \begin{cases} \mathbb{D}_{\Phi_3} & \rightarrow \quad l^\infty[0,1]^2 \\ (H, F, G) & \mapsto \quad H \circ (F, G). \end{cases}
\end{aligned} \tag{2.9}$$

The first mapping Φ_1 is Hadamard-differentiable at C with derivative $\Phi'_{1,C} = \Phi_1$ since it is linear and continuous.
Considering the second mapping let $U \in \mathbb{E}$ be the identity, then $\Psi : F \mapsto F^-$ with $F \in \mathbb{E}$ is Hadamard-differentiable at U tangentially to the set $\mathbb{E}_0 = \{F \in C[0,1] \,:\, F(0) = F(1) = 0\}$ with derivative $\Psi'_U(h) = -h$. To see this let $t_n \to 0$, $h_n \in l^\infty[0,1]$ with $h_n \to h \in \mathbb{E}_0$ and $U + t_n h_n \in \mathbb{E}$. We have to show that

$$\sup_{x\in[0,1]} \left| \frac{\Psi(U + t_n h_n) - \Psi(U)}{t_n} - \Psi'_U(h) \right| (x) = \sup_{x\in[0,1]} \left| \frac{(U + t_n h_n)^-(x) - x}{t_n} + h(x) \right| \to 0. \tag{2.10}$$

We deal with the case $x = 0$ first and show that

$$A_n = (U + t_n h_n)^-(0) = \sup\{x \in [0,1] \,:\, x + t_n h_n(x) = 0\} = o(t_n).$$

Suppose A_n is non-zero, eventually. Choose a sequence $x_n \in [A_n/2, A_n]$ satisfying $x_n + t_n h_n(x_n) = 0$, i.e. $h_n(x_n) = -x_n/t_n$. Since h_n is bounded we obtain $x_n = O(t_n) = o(1)$. By uniform convergence of h_n and continuity of h this implies $-x_n/t_n = h_n(x_n) \to h(0) = 0$, that is $x_n = o(t_n)$ and therefore $A_n = o(t_n)$.

Now consider the case $x > 0$. Since $U + t_n h_n$ is a distribution function with $(U + t_n h_n)(0) = 0$ it follows that $\xi_n(x) = (U + t_n h_n)^-(x) \in (0,1]$ for all $x \in (0,1]$. Set $\varepsilon(x) = t_n^2 \wedge \xi_n(x) > 0$, then

$$(U + t_n h_n)(\xi_n(x) - \varepsilon(x)) \leq x \leq (U + t_n h_n)(\xi_n(x))$$

for all $x \in (0,1]$ by the definition of the generalized inverse function. This implies

$$-t_n h_n(\xi_n(x)) \leq \xi_n(x) - x \leq -t_n h_n(\xi_n(x) - \varepsilon(x)) + t_n^2.$$

Since h_n is uniformly bounded we obtain $\xi_n(x) \to x$ uniformly in $x \in (0,1]$. Divide the last equation by t_n and use uniform convergence of h_n and continuity of h to conclude that $(\xi_n(x) - x)/t_n \to -h(x)$ uniformly in $x \in (0,1]$. Together with the case $x = 0$ this yields (2.10).
It now easily follows that Φ_2 is Hadamard-differentiable at (C, U, U) tangentially to $\mathbb{D}_0 \times \mathbb{E}_0 \times \mathbb{E}_0$ with derivative

$$\Phi'_{2,(C,U,U)}(\alpha, h_1, h_2) = (\alpha, -h_1, -h_2).$$

Last but not least consider Φ_3. We assert that Φ_3 is Hadamard-differentiable at (C, U, U) tangentially to $\mathbb{D}_0 \times \mathbb{E}_0 \times \mathbb{E}_0$ with derivative

$$\Phi'_{3,(C,U,U)}(\alpha, f, g)(\mathbf{x}) = \alpha(\mathbf{x}) + \partial_1 C(\mathbf{x}) f(x_1) + \partial_2 C(\mathbf{x}) g(x_2).$$

To see this let $t_n \to 0$ and $(\alpha_n, f_n, g_n) \in l^\infty[0,1]^2 \times l^\infty[0,1] \times l^\infty[0,1]$ with $(\alpha_n, f_n, g_n) \to (\alpha, f, g) \in \mathbb{D}_0 \times \mathbb{E}_0 \times \mathbb{E}_0$ such that $(C + t_n\alpha_n, U + t_n f_n, U + t_n g_n) \in \mathbb{D}_\Phi \times \mathbb{E}^- \times \mathbb{E}^-$. Now decompose

$$t_n^{-1}\{\Phi_3(C + t_n\alpha_n, U + t_n f_n, U + t_n g_n) - \Phi_3(C, U, U)\} = L_{n1} + L_{n2}$$

where

$$\begin{aligned} L_{n1} &= t_n^{-1}\{C \circ (U + t_n f_n, U + t_n g_n) - C\} \\ L_{n2} &= \alpha_n \circ (U + t_n f_n, U + t_n g_n) \end{aligned}$$

and consider both terms separately. Exploiting the facts that α_n and (f_n, g_n) converge uniformly and that α is uniformly continuous one can conclude that $||L_{n2} - \alpha||_\infty = o(1)$. Concerning L_{n1} we have to deal with nine different cases. If $\mathbf{x} \in (0,1)^2$ a series expansion of C at $\mathbf{x}$ yields

$$L_{n1}(\mathbf{x}) = \partial_1 C(\mathbf{x}) f_n(x_1) + \partial_2 C(\mathbf{x}) g_n(x_2) + r_n(\mathbf{x}),$$

where the error term r_n can be written as

$$r_n(\mathbf{x}) = \big(\partial_1 C(\mathbf{y}) - \partial_1 C(\mathbf{x})\big) f_n(x_1) + \big(\partial_2 C(\mathbf{y}) - \partial_2 C(\mathbf{x})\big) g_n(x_2)$$

with some intermediate point $\mathbf{y} = \mathbf{y}(n)$ between $\mathbf{x}$ and $(U + t_n f_n, U + t_n g_n)(\mathbf{x})$. The main term uniformly converges to $\partial_1 C(\mathbf{x}) f(x_1) + \partial_2 C(\mathbf{x}) g(x_2)$ as asserted, hence it remains to show that the error term converges to 0 uniformly in $\mathbf{x}$. To see this, let $\varepsilon > 0$. Using uniform convergence of f_n, uniform continuity of f and the fact that $f(0) = f(1) = 0$ one can conclude that there exists $\delta > 0$ such that $|f_n(x_1)| \leq \varepsilon/2$ for all $x_1 < \delta$ and $x_1 > 1 - \delta$. Since partial derivatives of copulas are bounded by 1 the first half of the error term is uniformly small for $x_1 < \delta$ and $x_1 > 1 - \delta$. On the quadrangle $[\delta, 1 - \delta] \times [0, 1]$ the partial derivative $\partial_1 C$ is uniformly continuous which yields the desired convergence under consideration of $\mathbf{y}(n) \to \mathbf{x}$ and boundedness of f. The same arguments apply for the partial derivative in the second direction and the case $\mathbf{x} \in (0,1)^2$ is finished.
Now consider $\mathbf{x} \in (0,1) \times \{1\}$. Decompose $L_{n1}(x_1, 1) = L_{n1}^{(1)}(x_1) + L_{n1}^{(2)}(x_1)$ where

$$\begin{aligned} L_{n1}^{(1)}(x_1) &= t_n^{-1}\{C(x_1 + t_n f_n(x_1), 1) - C(x_1, 1)\} \\ L_{n1}^{(2)}(x_1) &= t_n^{-1}\{C(x_1 + t_n f_n(x_1), 1 + t_n g_n(1)) - C(x_1 + t_n f_n(x_1), 1)\} \end{aligned}$$

By the same arguments as in the previous case a Taylor expansion in the first coordinate yields

$$L_{n1}^{(1)}(x_1) = f(x_1) + o(1) = \partial_1 C(x_1, 1) f(x_1) + \partial_2 C(x_1, 1) g(1) + o(1)$$

uniformly in x_1 (note that $\partial_1 C(x_1, 1) = 1$ and $g(1) = 0$). Lipschitz-continuity of C implies $|L_{n1}^{(2)}(x_1)| \leq |g_n(1)| \to |g(1)| = 0$ and the second case is finished.
If $\mathbf{x} \in (0,1) \times \{0\}$ use $C(\cdot, 0) \equiv 0$ and Lipschitz-continuity of C to estimate

$$|L_{n1}(x_1, 0)| = t_n^{-1} \left| C(x_1 + t_n f_n(x_1), t_n g_n(0)) - C(x_1 + t_n f_n(x_1), 0) \right| \leq |g_n(0)| \to 0.$$

Since $\partial_1 C(x_1, 0) f(x_1) + \partial_2 C(x_1, 0) g(0) = 0$ this yields the assertion.
The cases $\mathbf{x} \in \{0,1\} \times (0,1)$ and $\mathbf{x} \in \{0,1\} \times \{0,1\}$ follow along similar lines and are therefore omitted. To conclude, Φ_3 is Hadamard-differentiable as asserted.
Now apply the chain rule, Lemma 3.9.3 in [van der Vaart and Wellner, 1996], to $\Phi = \Phi_3 \circ \Phi_2 \circ \Phi_1$ to conclude the assertion of the Lemma. □

Next, note that Theorem 2.6 in [Kosorok, 2008] yields

$$\begin{aligned} \sqrt{n}(F_n - C) &\rightsquigarrow \mathbb{B}_C, \\ \sqrt{n}\, c\, (F_n^M - F_n) &\underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{B}_C, \end{aligned}$$

where $M \in \{\xi, W\}$ and $c = \mu\tau^{-1}$ (resp. 1) if $M = \xi$ (resp. W).
Since $C_n^{M,M} = \Phi(F_n^M)$, $C_n = \Phi(F_n)$ and $C = \Phi(C)$, the functional delta method, see [Kosorok, 2008], yields

$$\begin{aligned} \sqrt{n}\, c\, (C_n - C) = \sqrt{n}\, c\, (\Phi(F_n) - \Phi(C)) &\rightsquigarrow \Phi_C'(\mathbb{B}_C) = \mathbb{G}_C, \\ \sqrt{n}\, c\, (C_n^{M,M} - C_n) = \sqrt{n}\, c\, (\Phi(F_n^M) - \Phi(F_n)) &\underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} \Phi_C'(\mathbb{B}_C) = \mathbb{G}_C. \end{aligned}$$

This proves Theorem 2.1, Theorem 2.2 and Theorem 2.4.

Next consider the mapping

$$\Psi : (G, H) \mapsto G(H_1^-, H_2^-),$$

defined for two distribution functions $(G, H) \in \mathbb{D}_\Phi \times \mathbb{D}_\Phi$. Note that

$$C_n^{\xi,\cdot} = \Psi(F_n^\xi, F_n), \quad C_n^{W,\cdot} = \Psi(F_n^W, F_n), \quad C_n = \Psi(F_n, F_n), \quad C = \Psi(C, C).$$

The mapping Ψ is Hadamard-differentiable at (C, C) tangentially to $C([0,1]^2)^2$ with derivative

$$\Psi'_{(C,C)}(G, H)(\mathbf{u}) = G(\mathbf{u}) - \partial_1 C(\mathbf{u}) H(u_1, 1) - \partial_2 C(\mathbf{u}) H(1, u_2).$$

To see this decompose

$$\Psi = \Phi_3 \circ \Phi_2 \circ \Psi_1,$$

where Φ_2 and Φ_3 are defined in (2.9) and Ψ_1 is given by

$$\Psi_1 : \begin{cases} \mathbb{D}_\Phi \times \mathbb{D}_\Phi & \rightarrow \quad \mathbb{D}_{\Phi_2} = \mathbb{D}_\Phi \times \mathbb{E} \times \mathbb{E} \\ (G, H) & \mapsto \quad (G, H_1, H_2). \end{cases}$$

Ψ_1 is Hadamard-differentiable since it is linear and continuous, Φ_2 and Φ_3 are dealt with in the proof of Lemma 2.6. Apply the chain rule, Lemma 3.9.3 in [van der Vaart and Wellner, 1996] to obtain the desired Hadamard-differentiability of Ψ.

Next note that, unlike in the proof of Theorem 2.2 and 2.4 above, we do not have weak convergence (resp. weak conditional convergence) of

$$\begin{aligned} \sqrt{n}\big((F_n, F_n) - (C, C)\big) &\rightsquigarrow (\mathbb{B}_C, \mathbb{B}_C), \\ \sqrt{n}\, c\, \big((F_n^M, F_n) - (F_n, F_n)\big) &\underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} (\mathbb{B}_C, 0) \end{aligned}$$

towards the same limiting field, which would be necessary for an application of the functional delta method. Nevertheless, we can mimic certain steps in the proof of Theorem 12.1 in [Kosorok, 2008].

First note that, unconditionally, $U_n^M = \sqrt{n}\big((F_n^M, F_n) - (F_n, F_n)\big) \rightsquigarrow (c^{-1}\mathbb{B}_C, 0) =: c^{-1}\mathbb{U}_1$. To see this let $h \in BL_1((l^\infty[0,1]^2)^2)$ and use Fubini's Theorem (Lemma 6.14 in [Kosorok, 2008]) and the triangle inequality to estimate

$$|\mathbb{E}^* h(U_n^M) - \mathbb{E} h(\mathbb{U}_1)| \leq |\mathbb{E}_X \mathbb{E}_M h(U_n^M)^* - \mathbb{E}_X^* \mathbb{E}_M h(U_n^M)| + \mathbb{E}_X^* |\mathbb{E}_M h(U_n^M) - \mathbb{E} h(\mathbb{U}_1)|.$$

Since $\left|\mathbb{E}_M h(U_n^M) - \mathbb{E} h(\mathbb{U}_1)\right|^*$ converges to 0 in probability and is uniformly integrable with respect to $\mathbf{X}_1, \ldots, \mathbf{X}_n$ the second term converges to 0. The first term is bounded by $\mathbb{E}_X \mathbb{E}_M h(U_n^M)^* - \mathbb{E}_X \mathbb{E}_M h(U_n^M)_*$ which also converges to 0 since

$$\mathbb{E}_M h(U_n^M)^* - \mathbb{E}_M h(U_n^M)_* = o_{\mathbb{P}}(1)$$

is uniformly integrable. Theorem 1.12.2 in [van der Vaart and Wellner, 1996] now yields the weak convergence as asserted.
Next, by successive application of Lemma 1.4.3 and Lemma 1.4.4 in [van der Vaart and Wellner, 1996], the vector (U_n^M, U_n) is asymptotically tight and asymptotically measurable. It is easy to see that asymptotic tightness and asymptotic measurability are pertained under continuous functionals, and therefore the same is true for $U_n^M + U_n$.
Now, if $\mathbb{U}_2$ is an independent copy of $\mathbb{U}_1$ we have for $h \in BL_1((l^\infty[0,1]^2)^2)$

$$\begin{aligned}
\left|\mathbb{E}^* h(U_n^M + U_n, U_n) - \mathbb{E} h(c^{-1}\mathbb{U}_1 + \mathbb{U}_2, \mathbb{U}_2)\right| \\
\le \left|\mathbb{E}_X \mathbb{E}_M h(U_n^M + U_n, U_n)^* - \mathbb{E}^* \mathbb{E}_M h(U_n^M + U_n, U_n)\right| \\
+ \mathbb{E}^* \left|\mathbb{E}_M h(U_n^M + U_n, U_n)^* - \mathbb{E}_{\mathbb{U}_1} h(c^{-1}\mathbb{U}_1 + U_n, U_n)\right| \\
+ \left|\mathbb{E}^* \mathbb{E}_{\mathbb{U}_1} h(c^{-1}\mathbb{U}_1 + U_n, U_n)^* - \mathbb{E}_{\mathbb{U}_2} \mathbb{E}_{\mathbb{U}_1} h(c^{-1}\mathbb{U}_1 + \mathbb{U}_2, \mathbb{U}_2)\right|.
\end{aligned}$$

The first term on the right goes to 0 by asymptotic measurability of $(U_n^M + U_n, U_n)$ (use Lemma 1.4.4 in [van der Vaart and Wellner, 1996]). The second term goes to 0 by the fact that $U_n^M \underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} c^{-1}\mathbb{U}_1$ and the fact that the map $x \mapsto h(x + U_n, U_n)$ is bounded Lipschitz. The third term goes to 0 since $(U_n, U_n) \rightsquigarrow (\mathbb{U}_2, \mathbb{U}_2)$ and since the map $x \mapsto \mathbb{E}_{\mathbb{U}_1} h(c^{-1}\mathbb{U}_1 + x, x)$ is also bounded Lipschitz. Using Theorem 1.12.2 in [van der Vaart and Wellner, 1996] one can conclude that, unconditionally,

$$\sqrt{n}\begin{pmatrix} F_n^M - C \\ F_n - C \end{pmatrix} \rightsquigarrow \begin{pmatrix} c^{-1}\mathbb{U}_1 + \mathbb{U}_2 \\ \mathbb{U}_2 \end{pmatrix}.$$

Hadamard-differentiability of the mapping $(\beta, \gamma) \mapsto (\Psi(\beta, \gamma), \Psi(\gamma, \gamma), (\beta, \gamma), (\gamma, \gamma))$ and the functional delta method, see Theorem 2.8 in [Kosorok, 2008], yields

$$\sqrt{n}\begin{pmatrix} \Psi(F_n^M, F_n) - \Psi(C, C) \\ \Psi(F_n, F_n) - \Psi(C, C) \\ (F_n^M, F_n) - (C, C) \\ (F_n, F_n) - (C, C) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \Psi'_{(C,C)}(c^{-1}\mathbb{U}_1 + \mathbb{U}_2, \mathbb{U}_2) \\ \Psi'_{(C,C)}(\mathbb{U}_2, \mathbb{U}_2) \\ (c^{-1}\mathbb{U}_1 + \mathbb{U}_2, \mathbb{U}_2) \\ (\mathbb{U}_2, \mathbb{U}_2) \end{pmatrix}.$$

Observing that $\Psi'_{(C,C)}$ is linear we can conclude that

$$c\sqrt{n}\begin{pmatrix} \Psi(F_n^M, F_n) - \Psi(F_n, F_n) \\ (F_n^M, F_n) - (F_n, F_n) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \Psi'_{(C,C)}(\mathbb{U}_1, 0) \\ (\mathbb{U}_1, 0) \end{pmatrix} = \begin{pmatrix} \mathbb{U}_1 \\ (\mathbb{U}_1, 0) \end{pmatrix}.$$

Continuity of the map $(\alpha, (\beta, \gamma)) \mapsto ||\alpha - \beta||$ yields

$$||c\sqrt{n}\left(\Psi(F_n^M, F_n) - \Psi(F_n, F_n)\right) - c\sqrt{n}(F_n^M - F_n)|| \longrightarrow 0$$

in outer probability. Hence, we can use Lemma A.1 and the weak conditional convergence $c\sqrt{n}(F_n^M - F_n) \underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{B}_C$ to conclude that $c\sqrt{n}\left(\Psi(F_n^M, F_n) - \Psi(F_n, F_n)\right) \underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{B}_C$, or in other words, both $\beta_n \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{B}_C$ and $\gamma_n \underset{W}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{B}_C$ in $l^\infty([0,1]^2)$. Regarding the assertions concerning α_n^{pdm} and α_n^{pdr} we set

$$\begin{aligned}\bar{\alpha}_n^{pdm}(\mathbf{u}) &= \beta_n(\mathbf{u}) - \partial_1 C(\mathbf{u})\beta_n(u_1,1) - \partial_2 C(\mathbf{u})\beta_n(1,u_2),\\ \bar{\alpha}_n^{pdr}(\mathbf{u}) &= \gamma_n(\mathbf{u}) - \partial_1 C(\mathbf{u})\gamma_n(u_1,1) - \partial_2 C(\mathbf{u})\gamma_n(1,u_2).\end{aligned}$$

Observing Lemma 4.1 in [Segers, 2010] and exploiting the fact that β_n and γ_n are asymptotically tight we can proceed as in the proof of Proposition 4.2 in [Segers, 2010] to conclude that

$$||\alpha_n^{pdm} - \bar{\alpha}_n^{pdm}|| \xrightarrow{\mathbb{P}} 0, \qquad ||\alpha_n^{pdr} - \bar{\alpha}_n^{pdr}|| \xrightarrow{\mathbb{P}} 0.$$

By Lemma A.1 this is sufficient for the assertions of Theorem 2.3 and 2.5. □

2.4 Finite sample properties

In this section we present a small comparison of the finite sample properties of the four bootstrap approximations given in the previous section. For the sake of brevity we only consider one scenario, namely the Clayton copula with parameter $\theta = 1$, corresponding to Kendall's-$\tau = 1/3$, and note that other scenarios lead to similar results. The sample size in our study is either $n = 100$ or $n = 200$.

In our first example we show a comparison of the different resampling methods studying their covariances. We chose four points $\left\{\left(\frac{i}{3}, \frac{j}{3}\right), i,j = 1,2\right\}$ in the unit square and show in the first row of Table 2.1 the true covariances of the limiting process. The second row in the two table shows the simulated covariances of the the process $\sqrt{n}(C_n - C)$ on the basis of 10^6 simulation runs (note that this distribution cannot be used in applications because the "true" copula is usually not known). We observe a rather good approximation of the covariances of the limiting process by the empirical copula process α_n. Rows 3 - 6 of Table 2.1 show the covariances obtained by the bootstrap approximation. These covariances are based on the average of 1000 simulation runs, where in each run the covariance is estimated on the basis of $B = 1000$ bootstrap replications. The corresponding results for the mean squared errors are shown in Table 2.3. The multipliers for the partial derivative and the direct multiplier bootstrap are simulated from two-point distributions with mean and variance 1. We have also investigated other multipliers but it turns out that the two-point distributions with variance 1 yield the best results (the other results are not presented here for the sake of brevity).

The results of Table 2.1 - 2.4 show that the partial derivative multiplier methods (*pdm* and *pdr*) yield the best approximations in almost all cases, despite the fact that they requires the estimation of the partial derivatives of the copula. The advantages of these approach are particularly visible in the estimation of the variances. The approximations based on the resampling bootstrap and the direct multiplier bootstrap are similar but less accurate than the results obtained by the partial derivative method. It is also worthwhile

to mention that the *pdm*- and *pdr*-method need slightly more computational time to simulate a bootstrap sample, since they require evaluation of the multiplier processes in the boundary-points.
In our second example we investigate the approximation of the 90% and 95% quantile of the Kolmogorov-Smirnov statistic

$$K_n = \sup_{x \in [0,1]^2} |f_n(x)| \tag{2.11}$$

and the Crámer-von Mises statistic

$$L_n = \int_{[0,1]^2} f_n^2(x)\, dx. \tag{2.12}$$

The corresponding results are presented in Table 2.5 where the first and sixth row show the quantiles of the "true" process $f_n = \alpha_n$, which are calculated by 10^6 simulation runs. For the bootstrap methods the quantiles are estimated by 1000 simulation runs with $B = 1000$ Bootstrap-replications in each scenario. We observe again that the partial derivatives methods *pdm* and *pdr* yield the best approximation of the quantiles, while the resampling bootstrap and the direct multiplier bootstrap usually give too large quantiles, in particular for sample size $n = 100$. A similar observation for the partial multiplier derivative and the resampling method has been made by [Scaillet, 2005] in the context of testing hypothesis regarding the copula.
On the basis of the results presented in this study we conclude our investigation with the statement that, despite the fact that the partial derivatives bootstrap requires the estimation of the partial derivatives, it outperforms the resampling and the direct multiplier bootstrap.

		(1/3,1/3)	(1/3,2/3)	(2/3,1/3)	(2/3,2/3)
True	(1/3,1/3)	0.0486	0.0202	0.0202	0.0100
	(1/3,2/3)		0.0338	0.0093	0.0185
	(2/3,1/3)			0.0338	0.0185
	(2/3,2/3)				0.0508
α_n	(1/3,1/3)	0.0489	0.0198	0.0198	0.0097
	(1/3,2/3)		0.0334	0.0089	0.0181
	(2/3,1/3)			0.0333	0.0180
	(2/3,2/3)				0.0510
α_n^{pdr}	(1/3,1/3)	0.0510	0.0200	0.1999	0.0092
	(1/3,2/3)		0.0351	0.0091	0.0182
	(2/3,1/3)			0.0349	0.0184
	(2/3,2/3)				0.0541
α_n^{pdm}	(1/3,1/3)	0.0527	0.0205	0.0205	0.0093
	(1/3,2/3)		0.0361	0.0092	0.0188
	(2/3,1/3)			0.0360	0.0188
	(2/3,2/3)				0.0554
α_n^{res}	(1/3,1/3)	0.0619	0.0244	0.0236	0.0094
	(1/3,2/3)		0.0460	0.0091	0.0211
	(2/3,1/3)			0.0450	0.0208
	(2/3,2/3)				0.0694
α_n^{dm}	(1/3,1/3)	0.0627	0.0251	0.0248	0.0112
	(1/3,2/3)		0.0456	0.0119	0.0213
	(2/3,1/3)			0.0451	0.0233
	(2/3,2/3)				0.0711

Table 2.1: *Sample covariances for the Clayton Copula with $\theta = 1$ and sample size $n = 100$. The first and second rows show the true covariances and the covariances of the empirical copula process, while rows 3 - 5 show the corresponding results for the bootstrap approximations.*

		(1/3,1/3)	(1/3,2/3)	(2/3,1/3)	(2/3,2/3)
True	(1/3,1/3)	0.0486	0.0202	0.0202	0.0100
	(1/3,2/3)		0.0338	0.0093	0.0185
	(2/3,1/3)			0.0338	0.0185
	(2/3,2/3)				0.0508
α_n	(1/3,1/3)	0.0492	0.0203	0.0203	0.0100
	(1/3,2/3)		0.0339	0.0093	0.0185
	(2/3,1/3)			0.0339	0.0185
	(2/3,2/3)				0.0508
α_n^{pdr}	(1/3,1/3)	0.0506	0.0199	0.0199	0.0094
	(1/3,2/3)		0.0351	0.0086	0.0182
	(2/3,1/3)			0.0350	0.0182
	(2/3,2/3)				0.0530
α_n^{pdm}	(1/3,1/3)	0.0513	0.0203	0.0201	0.0092
	(1/3,2/3)		0.0356	0.0087	0.0184
	(2/3,1/3)			0.0355	0.0185
	(2/3,2/3)				0.0537
α_n^{res}	(1/3,1/3)	0.0583	0.0228	0.0228	0.0098
	(1/3,2/3)		0.0413	0.0092	0.0199
	(2/3,1/3)			0.0417	0.0202
	(2/3,2/3)				0.0609
α_n^{dm}	(1/3,1/3)	0.0577	0.0226	0.0227	0.0104
	(1/3,2/3)		0.0408	0.0103	0.0210
	(2/3,1/3)			0.0412	0.0213
	(2/3,2/3)				0.0634

Table 2.2: *Sample covariances for the Clayton Copula with $\theta = 1$ and sample size $n = 200$. The first and second rows show the true covariances and the covariances of the empirical copula process, while rows 3 - 5 show the corresponding results for the bootstrap approximations.*

		(1/3,1/3)	(1/3,2/3)	(2/3,1/3)	(2/3,2/3)
α_n^{pdr}	(1/3,1/3)	0.7244	0.5058	0.4722	0.3321
	(1/3,2/3)		0.9067	0.1821	0.2731
	(2/3,1/3)			0.8285	0.2551
	(2/3,2/3)				0.5612
α_n^{pdm}	(1/3,1/3)	0.8887	0.5210	0.5222	0.3716
	(1/3,2/3)		1.0112	0.1799	0.2988
	(2/3,1/3)			0.9899	0.2818
	(2/3,2/3)				0.6250
α_n^{res}	(1/3,1/3)	2.2612	0.6640	0.5424	0.3447
	(1/3,2/3)		2.3702	0.1781	0.3554
	(2/3,1/3)			2.1336	0.3554
	(2/3,2/3)				3.9469
α_n^{dm}	(1/3,1/3)	2.6734	0.7566	0.7067	0.3037
	(1/3,2/3)		2.3636	0.2461	0.5189
	(2/3,1/3)			2.2544	0.5324
	(2/3,2/3)				4.6142

Table 2.3: *Mean squared error (multiplied with* 10^4*) of the different estimates for the covariance. The underlying copula is the Clayton copula with* $\theta = 1$ *and the sample size is* $n = 100$.

		(1/3,1/3)	(1/3,2/3)	(2/3,1/3)	(2/3,2/3)
α_n^{pdr}	(1/3,1/3)	0.4307	0.2423	0.2634	0.1601
	(1/3,2/3)		0.4601	0.0992	0.1528
	(2/3,1/3)			0.4922	0.1659
	(2/3,2/3)				0.2672
α_n^{pdm}	(1/3,1/3)	0.4595	0.2673	0.2798	0.1961
	(1/3,2/3)		0.5211	0.1069	0.1577
	(2/3,1/3)			0.5092	0.1681
	(2/3,2/3)				0.2992
α_n^{res}	(1/3,1/3)	1.3820	0.3476	0.3715	0.2102
	(1/3,2/3)		1.0414	0.1133	0.1940
	(2/3,1/3)			1.2112	0.1993
	(2/3,2/3)				1.614
α_n^{dm}	(1/3,1/3)	1.2682	0.3602	0.3471	0.2083
	(1/3,2/3)		1.0394	0.1101	0.2484
	(2/3,1/3)			1.0544	0.2642
	(2/3,2/3)				1.9483

Table 2.4: *Mean squared error (multiplied with* 10^4*) of the different estimates for the covariance. The underlying copula is the Clayton copula with* $\theta = 1$ *and the sample size is* $n = 200$.

n	f_n	90% (L^2)	95% (L^2)	90% (KS)	95% (KS)
100	α_n	0.04593	0.05722	0.59254	0.65000
	α_n^{pdr}	0.04882	0.06113	0.60000	0.66500
	α_n^{pdm}	0.04870	0.06086	0.62042	0.68611
	α_n^{res}	0.07060	0.08700	0.80000	0.80000
	α_n^{dm}	0.07402	0.09241	0.76154	0.83721
200	α_n	0.04544	0.05660	0.58925	0.64829
	α_n^{pdr}	0.04823	0.06010	0.60140	0.66242
	α_n^{pdm}	0.04715	0.05867	0.61236	0.67528
	α_n^{res}	0.06030	0.07425	0.70711	0.77782
	α_n^{dm}	0.06066	0.07507	0.70192	0.77030

Table 2.5: *Sample quantiles of the Crámer van Mises statistic (2.12) and the Kolmogorov-Smirnov statistic (2.11) for the Clayton copula with parameter $\theta = 1$.*

Chapter 3

Extreme value copulas: New estimators and tests

3.1 Introduction

Recall Proposition 1.15: A bivariate copula C is an extreme value copula if and only if it has a representation of the form

$$C(\mathbf{u}) = \exp\Big(\log(u_1u_2)A\Big(\frac{\log u_2}{\log(u_1u_2)}\Big)\Big),$$

where $A : [0,1] \to [1/2,1]$ is a convex function satisfying $\max\{t, 1-t\} \le A(t) \le 1$, which is called Pickands dependence function. Using the transformation $\mathbf{u} = (y^{1-t}, y^t)$ it is easy to see that this is equivalent to

$$C(y^{1-t}, y^t) = y^{A(t)} \quad \forall\, y, t \in [0,1]. \tag{3.1}$$

Extreme value copulas arise naturally as the possible limits of copulas of component-wise maxima of independent, identically distributed or strongly mixing stationary sequences (see [Deheuvels, 1984; Hsing, 1989] or the discussion in Chapter 1). These copulas provide flexible tools for modeling joint extremes in risk management. An important application of extreme value copulas appears in the modeling of data with positive dependence, and in contrast to the more popular class of Archimedean copulas they are not symmetric (see [Tawn, 1988; Ghoudi et al., 1998]). Further applications can be found in [Coles et al., 1999; Cebrian et al., 2003] among others.
The representation of (3.1) of the extreme value copula C depends only on the one-dimensional function A and statistical inference on a bivariate extreme value copula C may now be reduced to inference on its Pickands dependence function A.
The problem of estimating Pickands dependence function nonparametrically has found considerable attention in the literature. Roughly speaking, there exist two classes of estimators. The classical nonparametric estimator is that of [Pickands, 1981] (see [Deheuvels, 1991] for its asymptotic properties) and several variants have been discussed. Alternative estimators have been proposed and investigated in the papers by [Capéraà et al.,

1997; Rojo Jiménez et al., 2001; Hall and Tajvidi, 2000; Segers, 2007; Zhang et al., 2008], where the last-named authors also discussed the multivariate case. In most references the estimators of Pickands dependence function are constructed assuming knowledge of the marginal distributions. Recently [Genest and Segers, 2009] proposed rank-based versions of the estimators of [Pickands, 1981] and [Capéraà et al., 1997], which do not require knowledge of the marginal distributions. In general all of these estimators are neither convex nor do they satisfy the boundary restriction $\max\{t, 1-t\} \leq A(t) \leq 1$, in particular the endpoint constrains $A(0) = A(1) = 1$. However, the estimators can be modified without changing their asymptotic properties in such a way that these constraints are satisfied, see e.g. [Fils-Villetard et al., 2008].

Before the specific model of an extreme value copula is selected it is necessary to check this assumption by a statistical test, that is a test for the hypotheses

$$H_0 : C \in \mathcal{C} \quad \text{vs.} \quad H_1 : C \notin \mathcal{C}, \tag{3.2}$$

where $\mathcal{C}$ denotes the class of all copulas satisfying (3.1). Throughout this chapter we call (3.2) the hypothesis of extreme value dependence. The problem of testing this hypothesis has found much less attention in the literature. To our best knowledge, only two tests of extremeness are currently available in the literature. The first one was proposed by [Ghoudi et al., 1998]. It exploits the fact that for a random vector $\mathbf{X} \sim F = C(F_1, F_2)$ with an extreme value copula C the random variable $W = F(\mathbf{X}) = C(F_1(X_1), F_2(X_2))$ satisfies the identity

$$-1 + 8\mathbb{E}[W] - 9\mathbb{E}[W^2] = 0. \tag{3.3}$$

The properties of this test have been studied by [Ben Ghorbal et al., 2009], who determined the finite- and large-sample variance of the test statistic. In particular, the test proposed by [Ghoudi et al., 1998] is not consistent against alternatives satisfying (3.3). The second class of tests was recently introduced by [Kojadinovic and Yan, 2010] who proposed to compare the empirical copula and a copula estimator which is constructed from the estimators proposed by [Genest and Segers, 2009] under the assumption of an extreme value copula. These tests are only consistent against alternatives that are left tail decreasing in both arguments and satisfy strong smoothness assumptions on the copula and convexity assumptions on an analogue of Pickands dependence function, which are hard to verify analytically.

The present chapter of this thesis has two purposes. The first is the development of some alternative estimators of Pickands dependence function using the principle of minimum distance estimation. We propose to consider the best approximation of the logarithm of the empirical copula C_n evaluated in the point (y^{1-t}, y^t), i.e. $\log C_n(y^{1-t}, y^t)$, by functions of the form

$$\log(y)A(t) \tag{3.4}$$

with respect to a weighted L^2-distance. It turns out that the minimal distance and the corresponding optimal function can be determined explicitly. On the basis of this result, and by choosing various weight functions in the L^2-distance we obtain an infinite dimensional class of estimators for the function A. Our approach is closely related to the theory of Z-estimation and in Section 3.3 we indicate how this point of view provides several interesting relationships between the different concepts for constructing estimates of Pickands dependence function.

The second purpose of this chapter is to present a new test for the hypothesis of extreme value dependence, which is consistent against a much broader class of alternatives than the tests which have been proposed so far. Here our approach is based on an estimator of a weighted minimum L^2-distance between the true copula and the class of functions satisfying (3.4) and the corresponding tests are consistent with respect to all positive quadrant dependent alternatives satisfying weak differentiability assumptions of first order. To our best knowledge, this method provides the first test in this context which is consistent against such a general class of alternatives. Moreover, in contrast to [Ghoudi et al., 1998] and [Kojadinovic and Yan, 2010] we also provide a weak convergence result under fixed alternative which can be used for studying the power of the test.

The remaining part of the chapter is organized as follows. In Section 3.2 we consider the approximation problem from a theoretical point of view. In particular, we derive explicit representations for the minimal L^2-distance between the logarithm of the copula and its best approximation by a function of the form (3.4), which will be the basis for all statistical applications in this chapter. The new estimators, say $\hat{A}_n$, are defined in Section 3.3, where we also prove weak convergence of the process $\{\sqrt{n}(\hat{A}_n(t) - A(t))\}_{t\in[0,1]}$ in the space of uniformly bounded functions on the interval $[0,1]$ under appropriate assumptions on the weight function used in the L^2-distance. Furthermore, we give a theoretical and empirical comparison of the new estimators with the estimators proposed in [Genest and Segers, 2009]. We will also determine "optimal" estimators in the proposed class by minimizing the asymptotic MSE with respect to the choice of the weight function used in the L^2-distance. In particular, we demonstrate that some of the new estimators have a substantially smaller asymptotic variance than the estimators proposed by the last-named authors. We also provide a simulation study in order to investigate the finite sample properties of the different estimates. In Section 3.4 we introduce and investigate the new test of extreme value dependence. In particular, we derive the asymptotic distribution of the test statistic under the null hypothesis as well as under the alternative. In order to approximate the critical values of the test we introduce a multiplier bootstrap procedure, prove its consistency and study its finite sample properties by means of a simulation study. Finally, the proofs of the results in this chapter are deferred to Section 3.5.

3.2 A measure of extreme value dependence

Let $\mathcal{A}$ denote the set of all functions $A : [0,1] \to [1/2, 1]$ and let Π be the copula corresponding to independent random variables, i.e. $\Pi(\mathbf{u}) = u_1 u_2$, see Proposition 1.3 (ii). Throughout this chapter we assume that the copula C satisfies $C \geq \Pi$ which holds for any extreme value copula due to the lower bound for its Pickands dependence function A. As pointed out by [Scaillet, 2005] this property is equivalent to the concept of positive quadrant dependence, that is

$$\mathbb{P}(\mathbf{X} \leq \mathbf{x}) \geq \mathbb{P}(X_1 \leq x_1)\mathbb{P}(X_2 \leq x_2) \quad \forall\, \mathbf{x} = (x_1, x_2)^T \in \mathbb{R}^2.$$

For a copula with this property we define the weighted L^2-distance

$$M_h(C, A) = \int_{(0,1)^2} \Big(\log C(y^{1-t}, y^t) - \log(y) A(t)\Big)^2 h(y)\, d(y,t) \tag{3.5}$$

where $h : [0,1] \to \bar{\mathbb{R}}^+$ is a continuous weight function.
The following result is essential for our approach and provides an explicit expression for the best L^2-approximation of the logarithm of the copula by the logarithm of a function of the form (3.1) and as a by-product characterizes the function A^* minimizing $M_h(C, A)$.

Theorem 3.1
Assume that the given copula satisfies $C \geq \Pi^\kappa$ for some $\kappa \geq 1$ and that the weight function h satisfies $\int_0^1 (\log y)^2 h(y) dy < \infty$. Then the function

$$A^* = \arg\min\{M_h(C, A) \mid A \in \mathcal{A}\}$$

is unique and given by

$$A^*(t) = B_h^{-1} \int_0^1 \frac{\log C(y^{1-t}, y^t)}{\log y} h^*(y)\, dy\,, \tag{3.6}$$

where the associated weight function h^* is defined by

$$h^*(y) = \log^2(y) h(y), \quad y \in (0,1), \tag{3.7}$$

and

$$B_h = \int_0^1 (\log y)^2 h(y)\, dy = \int_0^1 h^*(y)\, dy.$$

Moreover, the minimal L^2-distance between the logarithms of the given copula and the class of functions of the form (3.4) *is given by*

$$M_h(C, A^*) = \int_{(0,1)^2} \left(\frac{\log C(y^{1-t}, y^t)}{\log y}\right)^2 h^*(y)\, d(y,t) - B_h \int_0^1 (A^*(t))^2\, dt. \tag{3.8}$$

Proof. Since $C \geq \Pi^\kappa$, we get $0 \geq \log C(y^{1-t}, y^t) \geq \kappa \log y$ and thus $|\log C(y^{1-t}, y^t)| \leq \kappa |\log y|$ and all integrals exist. Rewriting the L^2 distance in (3.5) gives

$$M_h(C, A) = \int_0^1 \int_0^1 \Big(\frac{\log C(y^{1-t}, y^t)}{\log y} - A(t)\Big)^2 (\log y)^2 h(y)\, dy\, dt$$

and the assertion is now obvious. □

Note that $A^*(t) = A(t)$ if C is an extreme value copula of the form (3.1) with Pickands dependence function A. Furthermore, the following Lemma shows that the minimizing function A^* defined in (3.6) satisfies the boundary conditions of Pickands dependence functions.

Lemma 3.2
Assume that C is a copula satisfying $C \geq \Pi$. Then the function A^* defined in (3.6) has the following properties

(i) $A^*(0) = A^*(1) = 1$

(ii) $A^*(t) \geq t \vee (1-t)$

(iii) $A^*(t) \leq 1$.

Proof. Assertion (i) is obvious. For a proof of (ii) one uses the Fréchet-Hoeffding bound $C(\mathbf{u}) \leq u_1 \wedge u_2$ (see e.g. [Nelsen, 2006]) and obtains the assertion by a direct calculation. Similarly assertion (iii) follows from the inequality $C \geq \Pi$. □

Unfortunately, the function A^* is in general not convex for every copula satisfying $C \geq \Pi$. A counterexample can be derived from Theorem 3.2.2 in [Nelsen, 2006] and is given by the following shuffle of the copula $u \wedge v$

$$C(\mathbf{u}) = \begin{cases} \min\{u_1, u_2, 1/2\}, & \mathbf{u} \in [0, \sqrt{1/2}]^2 \\ \min\{u_1, u_2 + 1/2 - \sqrt{1/2}\}, & \mathbf{u} \in [0, \sqrt{1/2}] \times [\sqrt{1/2}, 1] \\ \min\{u_1 + 1/2 - \sqrt{1/2}, u_2\}, & \mathbf{u} \in [\sqrt{1/2}, 1] \times [0, \sqrt{1/2}] \\ \min\{u_1, u_2, u_1 + u_2 + 1/2 - 2\sqrt{1/2}\}, & \mathbf{u} \in [\sqrt{1/2}, 1]^2, \end{cases}$$

for which an easy calculation shows that the mapping $t \mapsto -\log C(1/2^{1-t}, 1/2^t)$ is not convex. Consequently, one can find a weight function h such that the corresponding best approximating function A^* is not convex.

With the notation

$$f_y(t) = C(y^{1-t}, y^t),$$

the function A^* is convex (for every weight function h) if and only if the function $g_y(t) = -\log f_y(t)$ is convex for every $y \in (0,1)$. The following Lemma is now obvious.

Lemma 3.3
If the function $t \to f_y(t) = C(y^{1-t}, y^t)$ is twice differentiable and the inequality

$$[f_y'(t)]^2 \geq f_y''(t) f_y(t)$$

holds for every $(y, t) \in (0,1)^2$, then the best approximation A^* defined by (3.6) *is convex.*

It is worthwhile to mention that the function A^* is convex for some frequently considered classes of copulas, one of which will be illustrated in the following example.

Example 3.4
Consider the Clayton copula

$$C_{\text{Clayton}}(\mathbf{u};\theta) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}, \quad \theta > 0. \tag{3.9}$$

Then a tedious calculation yields

$$\begin{aligned}[f_y'(t)]^2 - f_y''(t) f_y(t) &= \theta \log^2(y) \left\{C_{\text{Clayton}}(y^{1-t}, y^t; \theta)\right\}^{2+2\theta} \left(4y^{-\theta} - y^{-\theta t} - y^{-\theta(1-t)}\right) \\ &\geq \theta \log^2(y) \left\{C_{\text{Clayton}}(y^{1-t}, y^t; \theta)\right\}^{2+2\theta} \left(3y^{-\theta} - 1\right) \geq 0,\end{aligned}$$

where the inequalities follow observing that $m(t) = y^{-\theta t} + y^{-\theta(1-t)} \leq m(0) = 1 + y^{-\theta}$ and $y^{-\theta} \geq 1$. Therefore we obtain from Lemma 3.3 that the best approximation A^* is convex and corresponds to an extreme value copula.

Example 3.5
In the following we discuss the weight function $h_k(y) = -y^k / \log y$ $(k \geq 0)$ with associated function $h_k^*(y) = -y^k \log y$, which will be used later for the construction of the new estimators of Pickands dependence function. One the one hand this choice is made for mathematical convenience, because it allows for easy explicit calculations of the asymptotic variance in specific examples. On the other hand estimates constructed on the basis of this weight function turn out to have good asymptotic and finite sample properties (see the discussion in Section 3.3.7). It follows that

$$B_{h_k} = -\int_0^1 y^k \log y \, dy = (k+1)^{-2}$$

and

$$A^*(t) = -(k+1)^2 \int_0^1 \log C(y^{1-t}, y^t)\, y^k \, dy,$$

which simplifies in the case $k = 0$ to the representation

$$A^*(t) = -\int_0^1 \log\ C(y^{1-t}, y^t) dy.$$

Example 3.6
In the following we calculate the minimal distance $M_h(C, A^*)$ and its corresponding best approximation A^* for two copula families and the associated weight function $h_1^*(y) = -y \log y$ from Example 3.5. First we investigate the Gaussian copula defined by

$$C_{\text{Gauß}}(\mathbf{u};\rho) = \Phi_2(\Phi^-(u_1), \Phi^-(u_2), \rho),$$

where Φ is the standard normal distribution function and $\Phi_2(\cdot,\cdot,\rho)$ is the distribution function of a bivariate normal random variable with standard normally distributed margins and correlation $\rho \in [0,1]$. For the limiting cases $\rho = 0$ and $\rho = 1$ we obtain the independence and perfect dependence copula, respectively, while for $\rho \in (0,1)$ the copula $C_{\text{Gauß}}(\,\cdot\,;\rho)$ is not an extreme value copula. The minimal distances are plotted as a function of ρ in the left part of the first line of Figure 3.1. In the right part we show some functions A^* corresponding to the best approximation of the logarithm of the Gaussian copula by a function of the form (3.4). We note that all functions A^* are convex although $C_{\text{Gauß}}(\,\cdot\,;\rho)$ is only an extreme value copula in the case $\rho = 0$.
In the second example we consider a convex combination of a Gumbel copula with parameter $\theta_1 = \log 2/\log 1.5$ (corresponding to a coefficient of tail dependence of 0.5) and a Clayton copula with parameter $\theta_2 = 2$, i.e.

$$C_\alpha(\mathbf{u}) = \alpha\, C_{\text{Clayton}}(\mathbf{u};\theta_2) + (1-\alpha)\, C_{\text{Gumbel}}(\mathbf{u};\theta_1), \quad \alpha \in [0,1],$$

where the the Clayton copula is given in (3.9) and the Gumbel copula is defined by

$$C_{\text{Gumbel}}(\mathbf{u};\theta) = \exp\left(-\left\{(-\log u_1)^\theta + (-\log u_2)^\theta)\right\}^{1/\theta}\right), \quad \theta > 1.$$

Note that only the Gumbel copula is an extreme value copula and obtained for $\alpha = 0$. The minimal distances are depicted in the left part of the lower panel of Figure 3.1 as a function of α. In the right part we show the functions A^* corresponding to the best approximation of the logarithm of C_α by a function of the form (3.4). Again all approximations are convex, which means that A^* corresponds in fact to an extreme value copula.

3.3 A class of minimum distance estimators

3.3.1 Pickands and the CFG estimator

From now on, let $\mathbf{X}_1,\ldots,\mathbf{X}_n$ (with $\mathbf{X}_i = (X_{i1},X_{i2})^T$) denote independent identically distributed bivariate random variables with cumulative distribution function F, copula C and marginals F_1 and F_2. Most of the estimates which have been proposed in the literature so far are based on the fact that the random variable

$$\xi(t) = \frac{-\log F_1(X_1)}{1-t} \wedge \frac{-\log F_2(X_2)}{t}$$

is exponentially distributed with parameter $A(t)$. In particular we have $E[\xi(t)] = 1/A(t)$. If the marginal distributions would be known, an estimate of $A(t)$ could be obtained by the method of moments. In the case of unknown marginals [Genest and Segers, 2009] proposed to replace F_1 and F_2 by their empirical counterparts and obtained

$$\hat{A}^P_{n,r}(t) = \Big(\frac{1}{n}\sum_{i=1}^n \hat{\xi}_i(t)\Big)^{-1}$$

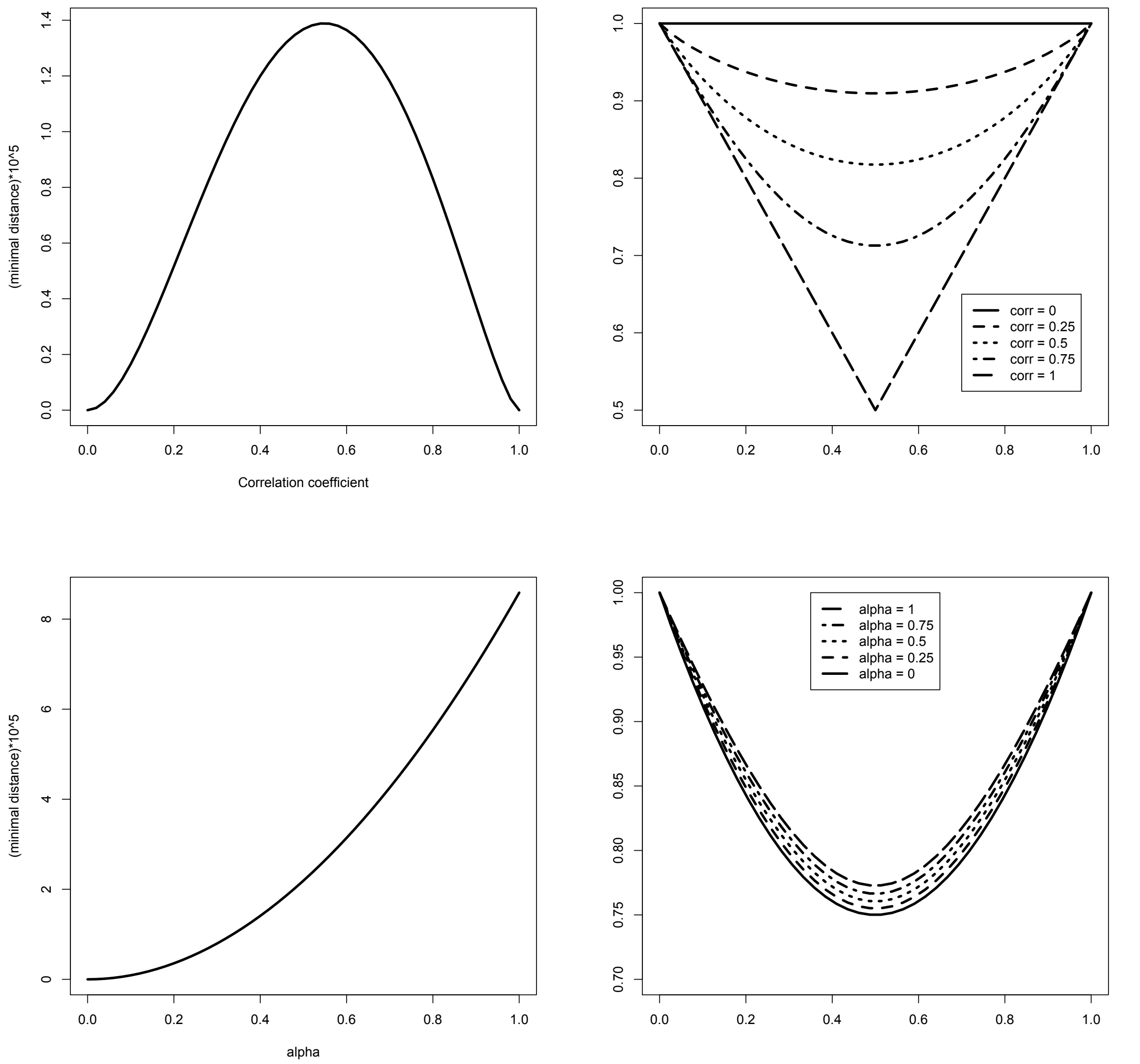


Figure 3.1: *Left: Minimal distances $M_h(C, A^*) \times 10^5$ for the Gaussian copula (as a function of its correlation coefficient) and for the convex combination of a Gumbel and a Clayton copula (as a function of the parameter α in the convex combination). Right: the functions A^* corresponding to the best approximations by functions of the form* (3.4).

as a rank-based version of Pickands estimate, where

$$\hat{\xi}_i(t) = \frac{-\log F_{n1}(X_{i1})}{1-t} \wedge \frac{-\log F_{n2}(X_{i2})}{t} \qquad i = 1, \dots, n,$$

and

$$F_{np}(X_{ip}) = \frac{1}{n+1} \sum_{j=1}^{n} I\{X_{jp} \le X_{ip}\}, \quad p = 1, 2,$$

denotes the (slightly modified) empirical distribution function of the sample $\{X_{jp}\}_{j=1}^n$ at the point X_{ip}. Similarly, observing the identity $E[\log \xi(t)] = -\log A(t) - \gamma$ (here $\gamma = -\int_0^\infty \log x \; e^{-x} dx$ denotes Euler's constant), they obtained a rank-based version of the estimate proposed by [Capéraà et al., 1997], that is

$$\hat{A}_{n,r}^{CFG}(t) = \exp\Big(-\gamma - \frac{1}{n} \sum_{i=1}^{n} \log \hat{\xi}_i(t)\Big).$$

For illustrative purposes we finally recall two integral representations for the rank-based version of Pickands and CFG estimate, which we use in Section 3.3.6 to put all estimates considered in this chapter in a general context, i.e.

$$\frac{1}{\hat{A}_{n,r}^{P}(t)} = \int_0^1 \frac{C_n(y^{1-t}, y^t)}{y} dy, \tag{3.10}$$

$$\gamma + \log \hat{A}_{n,r}^{CFG}(t) = \int_0^1 \frac{C_n(y^{1-t}, y^t) - I\{y > e^{-1}\}}{\log y} dy, \tag{3.11}$$

where

$$C_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^{n} I\{F_{n1}(X_{i1}) \le u_1,\; F_{n2}(X_{i2}) \le u_2\}, \tag{3.12}$$

denotes the empirical copula (see [Genest and Segers, 2009] for more details). Note that the definition of C_n slightly differs from the one in (2.2), but the asymptotic behavior is not affected.

3.3.2 New estimators and weak convergence

Theorem 3.1 suggests to define a class of new estimators for Pickands dependence function by replacing the unknown copula in (3.6) through the empirical copula defined in (3.12). The asymptotic properties of the corresponding estimators will be investigated in this section. For technical reasons we require that the argument in the logarithm in the representation (3.6) is positive and propose to use the estimator

$$\tilde{C}_n = C_n \vee n^{-\gamma}, \tag{3.13}$$

where the constant γ satisfies $\gamma > 1/2$ and the empirical copula C_n is defined in (3.12). Under this assumption, the process $\sqrt{n}(\tilde{C}_n - C)$ shows the same limiting behavior as the empirical copula process $\sqrt{n}(C_n - C)$, see Theorem 2.1. Note that the condition (2.3) in Theorem 2.1 can be shown to hold for any extreme value copula with continuously differentiable Pickands function A, see [Segers, 2010].
Now, observing the representation (3.6) we obtain the estimator

$$\hat{A}_{n,h}(t) = B_h^{-1} \int_0^1 \frac{\log \tilde{C}_n(y^{1-t}, y^t)}{\log y} h^*(y) dy \tag{3.14}$$

for Pickands dependence function, where $\tilde{C}_n$ is defined in (3.13). Note that this relation specifies an infinite dimensional class of estimators indexed by the set of all admissible weight functions. The following results specify the asymptotic properties of these estimators. We begin with a slightly more general statement, which shows weak convergence for the weighted integrated process

$$\sqrt{n}\mathbb{W}_{n,w}(t) = \sqrt{n} \int_0^1 \log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} w(y,t)\, dy,$$

where the weight function $w : [0,1]^2 \to \bar{\mathbb{R}}$ depends on y and t. The result (and some arguments in its proof) are also needed in Section 3.4.

Theorem 3.7
Assume that for the weight function $w : [0,1]^2 \to \bar{\mathbb{R}}$ there exists a function $\bar{w} : [0,1] \to \bar{\mathbb{R}}_0^+$ such that

$$\forall\, (y,t) \in [0,1]^2 : |w(y,t)| \le \bar{w}(y) \tag{3.15}$$

$$\forall\, \varepsilon > 0 : \sup_{y \in [\varepsilon, 1]} \bar{w}(y) < \infty \tag{3.16}$$

$$\int_0^1 \bar{w}(y) y^{-\lambda} dy < \infty \tag{3.17}$$

for some $\lambda > 1$. If the copula C satisfies (2.3) and $C \ge \Pi$, then we have for any $\gamma \in (1/2, \lambda/2)$ as $n \to \infty$

$$\sqrt{n}\mathbb{W}_{n,w}(t) = \sqrt{n} \int_0^1 \log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} w(y,t)\, dy \rightsquigarrow \mathbb{W}_{C,w}(t) = \int_0^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} w(y,t)\, dy$$

in $l^\infty[0,1]$.

The following result is now an immediate consequence of Theorem 3.7 using $w(y,t) := -B_h^{-1} h^*(y)$ (recall the definition of the associated weight function h^* in (3.7)) and yields the weak convergence of the process $\sqrt{n}(\hat{A}_{n,h} - A^*)$ for a broad class of weight functions.

Theorem 3.8
If the copula $C \geq \Pi$ satisfies condition (2.3) and the weight function h satisfies the conditions

$$\text{for all } \varepsilon > 0 : \sup_{y \in [\varepsilon,1]} \left| \frac{h^*(y)}{\log y} \right| < \infty \tag{3.18}$$

$$\int_0^1 h^*(y)(-\log y)^{-1} y^{-\lambda}\, dy < \infty \tag{3.19}$$

for some $\lambda > 1$, then we have for any $\gamma \in (1/2, \lambda/2)$ as $n \to \infty$

$$\mathbb{A}_{n,h} = \sqrt{n}(\hat{A}_{n,h} - A^*) \rightsquigarrow \mathbb{A}_{C,h} \quad \text{in} \quad l^\infty[0,1],$$

where the process $\mathbb{A}_{C,h}$ is given by

$$\mathbb{A}_{C,h}(t) = B_h^{-1} \int_0^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} \frac{h^*(y)}{\log y} dy. \tag{3.20}$$

Remark 3.9
(a) Conditions (3.18) and (3.19) restrict the behavior of the function h^* near the boundary of the interval $[0,1]$. A simple sufficient condition for (3.18) and (3.19) is given by

$$\sup_{x \in [0,1]} \left| \frac{h^*(x)}{x^\alpha (1-x)^\beta} \right| < \infty$$

for some $\alpha > 0, \beta \geq 1$. In this case λ can be chosen as $1 + \alpha/2$.

(b) In the construction discussed so far, it is also possible to use weight functions that depend on t, i.e. functions of the form $\tilde{h}^*(y,t)$. As long as $\tilde{h}^*(y,t) > 0$ for $(y,t) \in (0,1) \times [0,1]$, the corresponding best approximation A^* will still be well defined and correspond to the Pickands dependence function if C is an extreme value copula. Theorem 3.7 provides the asymptotic properties of the corresponding estimator A if we set $w(y,t) := \tilde{h}^*(y,t)/(-\log y)$ and assume that $\int_0^1 \tilde{h}^*(y,t)dy = 1$ for all t. However, for the sake of a clear presentation, we will only use weight functions that do not depend on t.

Note that Theorem 3.8 is also correct if the given copula is not an extreme value copula. In other words: it establishes weak convergence of the process $\sqrt{n}(\hat{A}_{n,h} - A^*)$ to a centered Gaussian process, where A^* denotes the function corresponding to the best approximation of the logarithm of the copula C by a function of the form (3.4). If A^* is convex, it corresponds to an extreme value copula and coincides with Pickands dependence function. Note also that Theorem 3.8 excludes the case $h_0^*(y) = -\log y$, because condition (3.19) is not satisfied for this weight function. Nevertheless, under the additional assumption that C is an extreme value copula with twice continuously differentiable Pickands dependence function A, the assertion of the preceding theorem is still valid.

Theorem 3.10
Assume that C is an extreme value copula with twice continuously differentiable Pickands dependence function A. For the weight function $h_0^*(y) = -\log y$ we have for any $\gamma \in (1/2, 3/4)$ as $n \to \infty$

$$\mathbb{A}_{n,h_0}(t) = \sqrt{n}(\hat{A}_{n,h_0} - A)(t) \rightsquigarrow \mathbb{A}_{C,h_0}(t) = -\int_0^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\, dy$$

in $l^\infty[0,1]$, where $\hat{A}_{n,h_0}(t) = -\int_0^1 \log \tilde{C}_n(y^{1-t}, y^t)\, dy$.

Remark 3.11
(a) If the marginals of (X, Y) are independent the distribution of the random variable $\mathbb{A}_{\Pi,h_0}$ coincides with the distribution of the random variable

$$\mathbb{A}_r^P(t) = -\int_0^1 \mathbb{G}_\Pi(y^{1-t}, y^t) y^{-1}\, dy,$$

which appears as the weak limit of the appropriately standardized Pickands estimator, see [Genest and Segers, 2009]. In fact, a much more general statement is true: by using weight functions $\tilde{h}^*(y,t)$ depending on t it is possible to obtain for any extreme value copula estimators of the form (3.14) which show the same limiting behavior as the estimators proposed by [Genest and Segers, 2009]. This already indicates that for any extreme value copula it is possible to find weight functions which will make the new minimum distance estimators asymptotically at least as efficient (in fact better, as will be shown in Section 3.3.4) as the estimators introduced by [Genest and Segers, 2009].

(b) A careful inspection of the proof of Theorem 3.7 reveals that the condition $C \geq \Pi$ can be relaxed to $C \geq \Pi^\kappa$ for some $\kappa > 1$, if one imposes stronger conditions on the weight function.

(c) The estimator depends on the parameter γ which is used for the construction of the statistic $\tilde{C}_n = C_n \vee n^{-\gamma}$. This modification is only made for technical purposes and from a practical point of view the behavior of the estimators does not change substantially provided that γ is chosen larger than $2/3$.

Remark 3.12
The new estimators can be alternatively motivated observing that the identity (3.1) yields the representation $A(t) = \log C(y^{1-t}, y^t)/\log y$ for any $y \in (0,1)$. This leads to a simple class of estimators, i.e.

$$\tilde{A}_{n,\delta_y}(t) = \frac{\log \tilde{C}_n(y^{1-t}, y^t)}{\log y}\,; \quad y \in (0,1),$$

where δ_y is the Dirac measure at the point y and $\tilde{C}_n$ is defined in (3.13). By averaging these estimators with respect to a distribution, say π, we obtain estimators of the form

$$\tilde{A}_{n,\pi}(t) = \int_0^1 \frac{\log \tilde{C}_n(y^{1-t}, y^t)}{\log y} \pi(dy),$$

which coincide with the estimators obtained by the concept of best L^2-approximation.

3.3.3 A special class of weight functions

In this subsection we illustrate the results investigating Example 3.5 discussed at the end of Section 3.2. For the associated weight function $h_k^*(x) = -y^k \log y$ with $k \geq 0$ we obtain

$$\hat{A}_{n,h_k}(t) = -(k+1)^2 \int_0^1 \log \tilde{C}_n(y^{1-t}, y^t)\, y^k\, dy.$$

The process $\{\mathbb{A}_{n,h_k}(t)\}_{t\in[0,1]}$ converge weakly in $l^\infty[0,1]$ to the process $\{\mathbb{A}_{C,h_k}\}_{t\in[0,1]}$, which is given by

$$\mathbb{A}_{C,h_k}(t) = -(k+1)^2 \int_0^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\, y^k\, dy.$$

Consequently, for $C \in \mathcal{C}$, the asymptotic variance of $\hat{A}_{n,h_k}$ is obtained as

$$\mathrm{Var}(\mathbb{A}_{C,h_k}(t)) = (k+1)^4 \int_0^1 \int_0^1 \sigma(u,v;t)(uv)^{k-A(t)}\, du\, dv, \tag{3.21}$$

where the function σ is given by

$$\sigma(u,v;t) = \mathrm{Cov}\left(\mathbb{G}_C(u^{1-t}, u^t), \mathbb{G}_C(v^{1-t}, v^t)\right).$$

In order to find an explicit expression for these variances we assume that the function A is differentiable and introduce the notation

$$\mu(t) = A(t) - tA'(t), \quad \nu(t) = A(t) + (1-t)A'(t),$$

where A' denotes the derivative of A. The following result can be shown by similar arguments as given in [Genest and Segers, 2009], the proof is given in Section 3.5.

Proposition 3.13
For $t \in [0,1]$ let $\bar{\mu}(t) = 1 - \mu(t)$ and $\bar{\nu}(t) = 1 - \nu(t)$. If C is an extreme value copula with Pickands dependence function A, then the variance of the random variable $\mathbb{A}_{C,h_k}(t)$ is given by

$$\begin{aligned}
(k+1)^2 \Big\{ & \frac{2(k+1)}{2k+2-A(t)} - (\mu(t)+\nu(t)-1)^2 - \frac{2\mu(t)\bar{\mu}(t)(k+1)}{2k+1+t} - \frac{2\nu(t)\bar{\nu}(t)(k+1)}{2k+2-t} \\
& + 2\mu(t)\nu(t)\frac{(k+1)^2}{(1-t)t} \int_0^1 \left(A(s) + (k+1)\left(\frac{1-s}{1-t} + \frac{s}{t}\right) - 1\right)^{-2} ds \\
& - 2\mu(t)\frac{(k+1)^2}{(1-t)t} \int_0^t \left(A(s) + (k+t)\frac{1-s}{1-t} + (k+1-A(t))\frac{s}{t}\right)^{-2} ds \\
& - 2\nu(t)\frac{(k+1)^2}{(1-t)t} \int_t^1 \left(A(s) + (k+1-A(t))\frac{1-s}{1-t} + (k+1-t)\frac{s}{t}\right)^{-2} ds \Big\}.
\end{aligned}$$

Note that the limiting process in (3.20) is a centered Gaussian process. This means that, asymptotically, the quality of the new estimators (as well as of the estimators of [Genest and Segers, 2009], which show a similar limiting behavior) is determined by the variance. Based on these observations, we will now provide an asymptotic comparison of the new estimators $\hat{A}_{n,h_k}(t)$ with the estimators investigated by [Genest and Segers, 2009]. Some finite sample results will be presented in the following section for various families of copulas. For the sake of brevity we restrict ourselves to the independence copula Π, for which $A(t) \equiv 1$. In the case $k = 0$ we obtain from Proposition 3.13 the same variance as for the rank-based version of Pickands estimator, that is

$$\mathrm{Var}(\mathbb{A}_{\Pi,h_0}) = \frac{3t(1-t)}{(2-t)(1+t)} = \mathrm{Var}(\mathbb{A}_r^P),$$

see Corollary 3.4 in [Genest and Segers, 2009], while the case $k > 0$ yields

$$\mathrm{Var}(\mathbb{A}_{\Pi,h_k}) = \frac{(3+4k)(k+1)^2}{2k+1} \frac{t(1-t)}{(2k+2-t)(2k+1+t)}.$$

Investigating the derivative in k, it is easy to see that $\mathrm{Var}(\mathbb{A}_{\Pi,h_k})$ is strictly decreasing in k with

$$\lim_{k\to\infty} \mathrm{Var}(\mathbb{A}_{\Pi,h_k}) = \frac{t(1-t)}{2}.$$

Therefore, we have

$$\mathrm{Var}(\mathbb{A}_r^P) = \mathrm{Var}(\mathbb{A}_{\Pi,h_0}) \geq \mathrm{Var}(\mathbb{A}_{\Pi,h_k})$$

for all $k \geq 0$ with strict inequality for all $k > 0$. This means that for the independence copula all estimators obtained by our approach with associated weight function $h_k^*(y) = -y^k \log y$, $k > 0$ have a smaller asymptotic variance than the rank-based version of Pickands estimator. On the other hand a comparison with the CFG estimator proposed by [Genest and Segers, 2009] does not provide a clear picture about the superiority of one estimator and we defer this comparison to the following section, where optimal weight functions for the new estimates $\hat{A}_{n,h}$ are introduced.

3.3.4 Optimal weight functions

In this section, we discuss asymptotically optimal weight functions corresponding to the class of estimates introduced in Section 3.3.2. As pointed out in the previous section, from an asymptotic point of view the mean squared error of the estimates is dominated by the variance and therefore we concentrate on weight functions minimizing the asymptotic variance of the estimate $\hat{A}_{n,h}$. The finite sample properties of the mean squared error of the various estimates will be investigated by means of a simulation study in Section 3.3.7.

Note that an optimal weight function depends on the point t where Pickands dependence function has to be estimated and on the unknown copula. Therefore an estimator with an optimal weight function cannot be implemented in concrete applications without preliminary knowledge about the copula. However, it can serve as a benchmark for user-specified weight functions. To be precise, observe that by Theorem 3.8 the variance of the limiting process $\mathbb{A}_{C,h}$ is of the form

$$V(\xi) = \int_0^1 \int_0^1 k_t(x,y) d\xi(x) d\xi(y), \tag{3.22}$$

where ξ denotes a probability measure on the interval $[0,1]$ defined by $d\xi(x) = B_h^{-1} h^*(x) dx$ and the kernel $k_t(x,y)$ is given by

$$k_t(x,y) = \mathbb{E}\Big[\frac{\mathbb{G}_C(x^{1-t}, x^t)}{C(x^{1-t}, x^t)\log(x)} \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)\log(y)}\Big]$$

It is easy to see that V defines a convex function on the space of all probability measures on the interval $[0,1]$ and the existence of a minimizing measure follows if the kernel k_t is continuous on $[0,1]^2$. The following result characterizes the minimizer of V and is proved in Section 3.5.

Theorem 3.14
A probability measure η on the interval $[0,1]$ minimizes V if and only if the inequality

$$\int_0^1 k_t(x,y) d\eta(y) \geq \int_0^1 \int_0^1 k_t(x,y) d\eta(x) d\eta(y) \tag{3.23}$$

is satisfied for all $x \in [0,1]$.

Theorem 3.14 can be used to check the optimality of a given weight function. For example, if the copula C is given by the independence copula Π we have

$$k_t(x,y) = \frac{(x^t \wedge y^t - (xy)^t)(x^{1-t} \wedge y^{1-t} - (xy)^{1-t})}{x \log(x)\, y \log(y)},$$

and it is easy to see that none of the associated weight functions $h_k^*(y) = -y^k \log y$ with $k \geq 0$ is optimal in the sense that it minimizes the asymptotic variance of the estimate $\hat{A}_{n,h}$ with respect to the choice of the weight function. On the other hand the result is less useful for an explicit computation of optimal weight functions. Deriving an analytical expression for the optimal weight function seems to be impossible, even for the simple case of the independence copula.
However, approximations of the optimal weight function can easily be computed numerically. To be precise we approximate the double integral appearing in the representation of $\mathrm{Var}(\mathbb{A}_{C,h}(t))$ by the finite sum

$$V(\xi) \approx \sum_{i=1}^N \sum_{j=1}^N \xi_{i,N} \xi_{j,N} k_t(i/N, j/N) = \Xi^T K_t \Xi \tag{3.24}$$

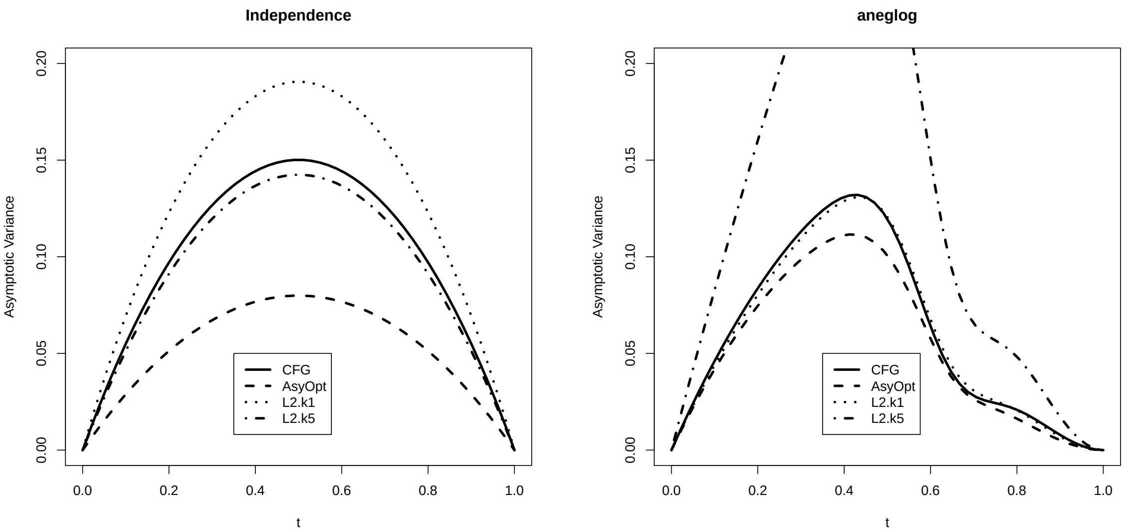


Figure 3.2: *Asymptotic variances of various estimators of the Pickands dependence function. Left panel: independence copula; right panel: asymmetric negative logistic model.*

where $N \in \mathbb{N}$, $K_t = (k_t(i/N, j/N))_{i,j=1}^N$ denotes an $N \times N$ matrix, $\Xi = (\xi_{i,N})_{i=1}^N$ is an vector of length N and $\xi_{i,N} = \xi((i-1)/N, i/N]$ represents the mass of ξ allocated to the interval $((i-1)/N, i/N]$ $(i = 1, \ldots, N)$. Minimizing the right hand side of the above equation with respect to Ξ under the constrains $\xi_{i,N} \geq 0$, $\sum_{i=1}^N \xi_{i,N} = 1$ is a quadratic (convex) optimization problem which can be solved by standard methods; see for example [Nocedal and Wright, 2006] and approximations of the optimal weight function can be calculated with arbitrary precision by increasing N.

In the remaining part of this section we will compare the asymptotic variance of the Pickands-, the CFG-estimator proposed by [Genest and Segers, 2009] and the new estimates, where the new estimators are based on the weight functions h_k^* discussed in Section 3.3.3 for two values of k as well as on the optimal weights minimizing the right hand side of (3.24), where we set $N = 100$. In order to compute the solution Ξ_{opt}, we used the routine *ipop* from the R-package *kernlab* by [Karatzoglou et al., 2004]. In the left part of Figure 3.2 we show the asymptotic variances of the different estimators for the independence copula. We observe that Pickands estimator has the largest asymptotic variances (this curve is not displayed in the figure), while the CFG estimator of [Genest and Segers, 2009] yields smaller variances than the estimator $\hat{A}_{n,h_1}$, but larger asymptotic variances than the estimators $\hat{A}_{n,h_5}$. On the other hand the estimate $\hat{A}_{n,h_{opt}}$ corresponding to the numerically determined optimal weight function yields a substantially smaller variance than all other estimates under consideration. In the right part of Figure 3.2 we display the corresponding results for the asymmetric negative logistic model [see [Joe, 1990]]

$$A(t) = 1 - \left\{ (\psi_1(1-t))^{-\theta} + (\psi_2 t)^{-\theta} \right\}^{-1/\theta} \tag{3.25}$$

with parameters $\psi_1 = 1, \psi_2 = 2/3$ and $\theta \in (0, \infty)$ chosen such that the coefficient of tail dependence is 0.6. We observe that the estimate $\hat{A}_{n,h_5}$ yields the largest asymptotic variance. The CFG estimate proposed by [Genest and Segers, 2009] and the estimate $\hat{A}_{n,h_1}$ show a similar behavior (with minor advantages for the latter), while the best results are obtained for the new estimate corresponding to the optimal weight function.
We conclude this section with the remark that we have presented a comparison of the different estimators based on the asymptotic variance which determines the mean squared error asymptotically. For finite samples, minimizing only the variance might increase the bias and therefore the asymptotic results can not directly be transferred to applications. In the finite sample study presented in Section 3.3.7 we will demonstrate that not all of the asymptotic results yield good predictions for the finite-sample behavior of the corresponding estimators.

3.3.5 Convex estimates and endpoint corrections

In general, all of the estimates discussed so far (including those proposed by [Genest and Segers, 2009]) will neither be convex, nor will they satisfy the other characterizing properties of Pickands dependence functions. However, the literature provides many proposals on how to enforce these conditions. Various endpoint corrections have been proposed by [Deheuvels, 1991; Segers, 2007; Hall and Tajvidi, 2000] among others. [Fils-Villetard et al., 2008] proposed an L^2-projection of the estimate of Pickands dependence function on a space of partially linear functions which is arbitrarily close to the space of all convex functions in $\mathcal{A}$ satisfying the conditions of Lemma 3.2. They also showed that this transformation decreases the L^2-distance between the "true" dependence function and the estimate. An alternative concept of constructing convex estimators is based on the greatest convex minorant, which yields a decrease in the sup-norm, that is

$$\sup_{0<t<1} |\hat{A}_n^{\mathrm{gcm}}(t) - A(t)| \leq \sup_{0<t<1} |\hat{A}_n(t) - A(t)| \,,$$

where $\hat{A}_n$ is any initial estimate of Pickands dependence function and $\hat{A}_n^{\mathrm{gcm}}$ its greatest convex minorant (see e.g. [Marshall, 1970; Wang, 1986; Robertson et al., 1996] among others). It is also possible to combine this concept with an endpoint correction calculating the greatest convex minorant of the function

$$t \longrightarrow (\hat{A}_n(t) \wedge 1) \vee t \vee (1-t)$$

(see [Genest and Segers, 2009] who also proposed alternative special endpoint corrections for their estimators). All these methods can be used to produce an estimate of A which has the characterizing properties of a Pickands dependence function.

3.3.6 M- and Z-estimates

As mentioned in the introduction a broader class of estimates could be obtained by minimizing more general distances between the given copula and the class of functions defined by (3.1) and in this paragraph we briefly indicate this principle. Consider the best approximation of the copula C by functions of the form (3.1) with respect to the distance

$$D_w(C,A) = \int_0^1 \int_0^1 \Phi\Big(C(y^{1-t},y^t),y^{A(t)}\Big)w(y,t)dydt \tag{3.26}$$

where $\Phi : [0,1]\times[0,1] \to \mathbb{R}_0^+$ denotes a “distance” and w is a given weight function. Note that the minimization in (3.26) can be carried out by separately minimizing the inner integral for every value of t. Consequently, the problem reduces to a one-dimensional minimization problem and assuming differentiability it follows that for fixed t the optimal value $A^*(t)$ minimizing the interior integral in (3.26) is obtained as a solution of the equation

$$\frac{\partial}{\partial a}\int_0^1 \Phi\Big(C(y^{1-t},y^t),y^a\Big)w(y,t)dy\Big|_{a=A^*(t)} \;=\; 0.$$

Under suitable assumptions integration and differentiation can be exchanged and we have

$$\int_0^1 \Psi\Big(C(y^{1-t},y^t),y^a\Big)(\log y)y^a w(y,t)\Big|_{a=A^*(t)}dy \;=\; 0 \tag{3.27}$$

where $\Psi = \partial_2\Phi$ denotes the derivative of Φ with respect to the second argument. In general the solution of (3.27) is only defined implicitly as a functional of the copula C. Therefore, if C is replaced through the empirical copula the analysis of the stochastic properties of the corresponding process turns out to be extremely difficult because in many cases one has to control improper integrals (see the proofs of Theorem 3.7 and 3.10 in Section 3.5). This is beyond the scope of this thesis.
Nevertheless, equation (3.27) yields a different view on the estimation problem of Pickands dependence function. Note that the estimate introduced in Section 3.3.2 is obtained by the choice $w(y,t) = h(y)B_h^{-1}$ and

$$\Phi(z_1,z_2) = (\log z_1 - \log z_2)^2\,;\;\; \Psi(z_1,z_2) = -2(\log z_1 - \log z_2)/z_2$$

in (3.27). This estimate corresponds to a minimum distance estimate. Similarly, an estimate corresponding to the classical L^2-distance is obtained for the choice

$$\Phi(z_1,z_2) = (z_1-z_2)^2\,;\;\; \Psi(z_1,z_2) = -2(z_1-z_2).$$

This yields for (3.27) the equation

$$\int_0^1 \Big(C(y^{1-t},y^t) - y^a\Big)(\log y)^2 y^a h(-\log y)\Big|_{a=A^*(t)}dy \;=\; 0,$$

which cannot be solved analytically. The rank-based versions of Pickands and the CFG estimator proposed by [Genest and Segers, 2009] do not correspond to M-estimates, but could be considered as Z-estimates obtained from (3.27) for the function

$$\Psi(z_1, z_2) = (z_1 - z_2)/z_2$$

with $w_{\mu,\nu}(y) = y^{\mu-1}/(-\log y)^{2+\nu}$ with $\mu = \nu = 0$ and $\mu = 0, \nu = 1$ respectively. In fact this choice leads to a general class of estimators which relates the Pickands and the CFG estimate in an interesting way. To be precise, note that for $\nu \in [0, 1)$ equation (3.27) yields

$$\begin{aligned} \int_0^1 \frac{(C(y^{1-t}, y^t) - I\{y > e^{-1}\})y^{\mu-1}}{(-\log y)^\nu} dy &= \int_0^1 \frac{(y^{A(t)} - I\{y > e^{-1}\})y^{\mu-1}}{(-\log y)^\nu} dy \\ &= \frac{\Gamma(1-\nu)}{(A(t)+\mu)^{1-\nu}} - \int_0^1 \frac{e^{-\mu x}}{x^\nu} dx \end{aligned} \tag{3.28}$$

Here the case $\nu = 1$ has to be interpreted as the limit $\nu \to 1$, which yields a generalization of the defining equation for the CFG estimate, that is

$$-\log \mu - \int_\mu^\infty \frac{e^{-t}}{t} dt + \log(A(t) + \mu) = \int_0^1 \frac{(C(y^{1-t}, y^t) - I\{y > e^{-1}\})y^{\mu-1}}{\log y} dy.$$

Observing the relation

$$\lim_{\mu \to 0} \log \mu + \int_\mu^\infty \frac{e^{-t}}{t} dt = -\gamma$$

we obtain the defining equation for the estimate proposed by [Genest and Segers, 2009] (see equation (3.11)). Similarly, if $\nu \in [0, 1)$ it follows from (3.28)

$$\int_0^1 \frac{C(y^{1-t}, y^t)y^{\mu-1}}{(-\log y)^\nu} dy = \frac{\Gamma(1-\nu)}{(A(t)+\mu)^{1-\nu}} \tag{3.29}$$

and we obtain a defining equation for a generalization of the Pickands estimate. The classical case is obtained for $\mu = \nu = 0$ (see [Genest and Segers, 2009] or equation (3.10)), but (3.29) defines many other estimates of this type. Therefore, the Pickands and the CFG estimate correspond to the extreme cases in the class $\{w_{\mu,\nu} | \ \mu \geq 0, \nu \in [0, 1]\}$.
We finally note that there are numerous other functions Ψ, which could be used for the construction of alternative Z-estimates, but most of them do not lead to an explicit solution for $A^*(t)$. In this sense the CFG-estimator, Pickands-estimator and the estimates proposed in this chapter could be considered as attractive special cases, which can be explicitly represented in terms of an integral of the empirical copula.

3.3.7 Finite sample properties

In this subsection we investigate the small sample properties of the new estimators by means of a simulation study. Especially, we compare the new estimators with the rank-based estimators suggested by [Genest and Segers, 2009], which are most similar in spirit with the method proposed in this chapter. We study the finite sample behavior of the greatest convex minorants of the endpoint corrected versions of the various estimators. The new estimators are corrected in a first step by

$$\hat{A}_{n,h}^{corr}(t) := \max(t, 1-t, \min(\hat{A}_{n,h}, 1)) \tag{3.30}$$

and in a second step the greatest convex minorant of $\hat{A}_{n,h}^{corr}$ is calculated. For the rank-based CFG and Pickands estimators we first used the endpoint corrections proposed in [Genest and Segers, 2009], then applied (3.30) and finally calculated the greatest convex minorant. Hereby, we compare the performance of the different statistical procedures which will be used in concrete applications and apply the corrections, that are most favorable for the respective estimators. The greatest convex minorants are computed using the routine *gcmlcm* from the package *fdrtool* by [Strimmer, 2009]. All results presented here are based on 5000 simulation runs and the sample size is $n = 100$.
As estimators we consider the statistics defined in (3.14) with the weight function h_k and the optimal weight function determined in Section 3.3.4. An important question is the choice of the parameter k for the statistic $\hat{A}_{n,h_k}$ in order to achieve a balance between bias and variance. For this purpose, we first study the performance of the estimator $\hat{A}_{n,h_k}$ with respect to different choices for the parameter k and consider the asymmetric negative logistic model defined in (3.25) and the symmetric mixed model (see [Tawn, 1988]) defined by

$$A(t) = 1 - \theta t + \theta t^2, \quad \theta \in [0,1]. \tag{3.31}$$

The results for other copula models are similar and are omitted for the sake of brevity. For the Pickands dependence function (3.25) we used the parameters $\psi_1 = 1$ and $\psi_2 = 2/3$ such that the coefficient of tail dependence is given by $\rho = 2\,(3^\theta + 2^\theta)^{-1/\theta}$ and varies in the interval $(0, 2/3)$, while the parameter $\theta \in [0,1]$ used in (3.31) yields $\rho = \theta/2 \in [0, 1/2]$.
The quality of an estimator $\hat{A}$ is measured with respect to mean integrated squared error

$$\mathrm{MISE}(\hat{A}) = \mathbb{E}\left[\int_0^1 (\hat{A}(t) - A(t))^2\, dt\right],$$

which was computed by taking the average over 5000 simulated samples. The new estimators turned out to be rather robust with respect to the choice of the parameter γ in the definition of the process $\tilde{C}_n = C_n \vee n^{-\gamma}$ provided that $\gamma \geq 2/3$. For this reason we use $\gamma = 0.95$ throughout this section. Analyzing the impact of choosing different values for k, in Figure 3.3 we display the simulated curves

$$k \mapsto \frac{\mathrm{MISE}(\hat{A}_{n,h_k})}{\min_{\ell \geq 0} \mathrm{MISE}(\hat{A}_{n,h_\ell})} \tag{3.32}$$

for the asymmetric negative logistic and the mixed models with different coefficients of tail dependence ρ, as well as the maximum over such curves for different values of ρ

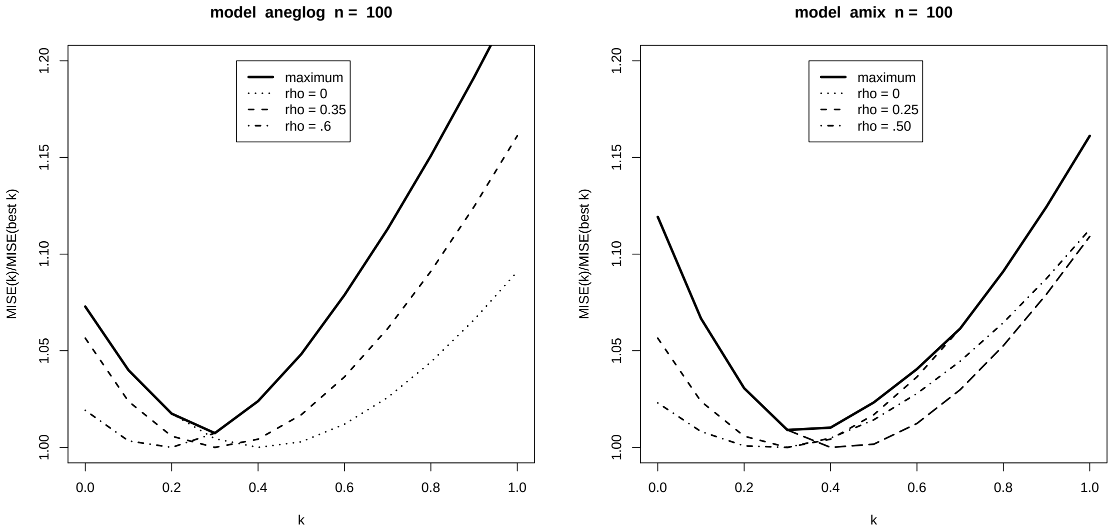


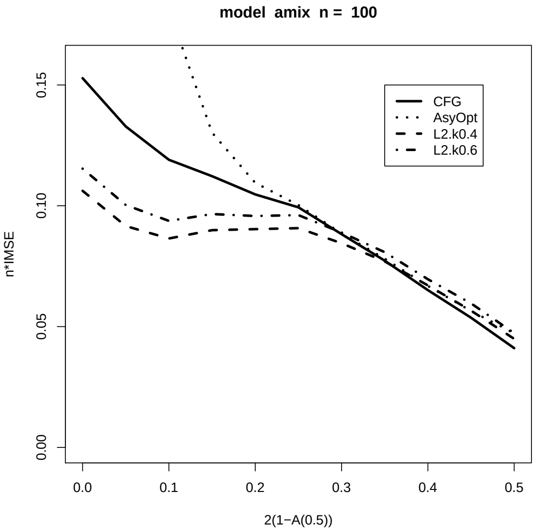
Figure 3.3: *The function defined in* (3.32) *for various models and coefficients of tail dependence. The minimum corresponds to the optimal value of k in the weight function h_k. The solid curve corresponds to the worst case defined by* (3.33). *The sample size is $n = 100$ and the MISE is calculated by* 5000 *simulation runs. Left panel: asymmetric negative logistic model. Right panel: mixed model.*

(solid curves), that is

$$k \mapsto \max_{\rho} \frac{MISE_\rho(\hat{A}_{n,h_k})}{\min_{\ell \geq 0} MISE_\rho(\hat{A}_{n,h_\ell})}, \tag{3.33}$$

where by $MISE_\rho$ we denote the MISE for the tail dependence coefficient ρ. The curves in (3.32) attain their minima in the optimal k for the respective ρ, and their shapes provide information about the performance of the estimators for non-optimal values of k. The solid curve gives an impression about the "worst case" scenario (with respect to ρ) in every model. The simulations indicate, that for $n = 100$ the optimal values of k for different models and tail dependence coefficients lie in the interval $[0.2, 0.6]$. Moreover, for values of k in this interval the quality of the estimators remains very stable. For $n = 200$, $n = 500$ and additional models the picture remains quite similar and these results are not depicted for the sake of brevity. We thus recommend using $k = 0.4$ in practical applications. Note that the asymptotic analysis in Section 3.3.4 suggests that the asymptotically optimal k should differ substantially for various models. However, this effect is not visible for sample size up to $n = 500$. In these cases the optimal values for k usually varies in the interval $[0.2, 0.8]$.

Next we compare the new estimators with rank-based versions of Pickands and the CFG estimator proposed by [Genest and Segers, 2009]. In Figure 3.4, the normalized MISE is plotted as a function of the tail dependence parameter ρ for the asymmetric negative



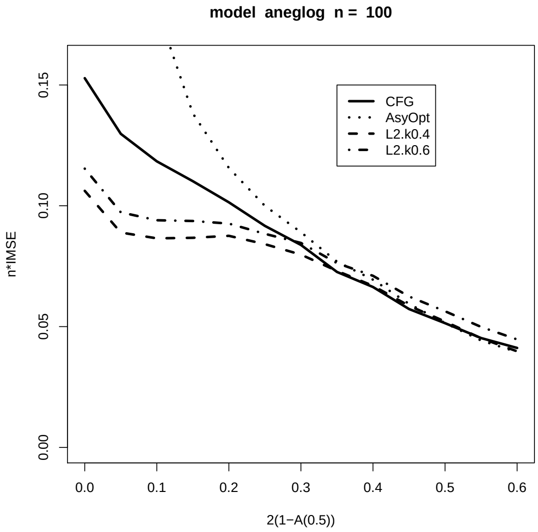


Figure 3.4: $100\times$*MISE for various estimators, models and coefficients of tail dependence, based on* 5000 *samples of size* $n = 100$.

logistic and the mixed model, where the parameter θ is chosen in such a way, that the coefficient of tail dependence $\rho = 2(1 - A(0.5))$ varies over the specific range of the corresponding model. For each sample we computed the rank-based versions of Pickands estimator, the CFG estimator (see [Genest and Segers, 2009]) and two of the new estimators $\hat{A}_{n,h_k}$ ($k = 0.4, 0.6$). In this comparison we also include the estimator $\hat{A}_{n,h_{opt}}$ which uses the optimal weight function determined in Section 3.3.4.

Summarizing the results one can conclude that in general the best performance is obtained for our new estimator based on the weight function h_k with $k = 0.4$ and $k = 0.6$, in particular if the coefficient of tail dependence is small. A comparison of the two estimators $\hat{A}_{n,h_{0.4}}$ and $\hat{A}_{n,h_{0.6}}$ shows that the choice $k = 0.4$ performs slightly better than the choice $k = 0.6$ in both models. In both settings, the MISE obtained by $\hat{A}_{n,h_{0.4}}$ and $\hat{A}_{n,h_{0.6}}$ is smaller than the MISE of the CFG estimator proposed in [Genest and Segers, 2009] if the coefficient of tail dependence is small. On the other hand the latter estimators yield sightly better results for a large coefficient of tail dependence. The results for rank-based version of the Pickands estimator are not depicted, because this estimator yields a uniformly larger MISE. Simulations of other scenarios show similar results and are also not displayed for the sake of brevity. It is remarkable that the optimal weight function usually yields an estimator with a substantially larger MISE than all other estimates if the coefficient of tail dependence is small. Similar results can be observed for the sample size $n = 500$ (these results are not depicted). This indicates that the advantages of the asymptotically optimal weight function only start to play a role for rather large sample sizes.

3.4 A test for an extreme value dependence

3.4.1 The test statistic and its weak convergence

From the definition of the functional $M_h(C, A)$ in (3.5) it is easy to see that, for a strictly positive weight function h with $h^* \in L^1(0,1)$, a copula function C is an extreme value copula if and only if

$$\min_{A \in \mathcal{A}} M_h(C, A) = M_h(C, A^*) = 0,$$

where A^* denotes the best approximation defined in (3.6). This suggests to use $M_h(\tilde{C}_n, \hat{A}_{n,h})$ as a test statistic for the hypothesis (3.2), i.e.

$$H_0 : C \quad \text{is an extreme value copula.}$$

Recalling the representation (3.8)

$$M_h(C, A^*) = \int_0^1 \int_0^1 \bar{C}^2(y,t) h^*(y)\, dy\, dt - B_h \int_0^1 \big(A^*(t)\big)^2\, dt$$

with $\bar{C}(y,t) = -\log C(y^{1-t}, y^t)$ and defining $\bar{C}_n(y,t) := -\log \tilde{C}_n(y^{1-t}, y^t)$ we obtain the decomposition

$$\begin{aligned}
& M_h(\tilde{C}_n, \hat{A}_{n,h}) - M_h(C, A^*) \\
&= \int_0^1 \int_0^1 \Big(\bar{C}_n^2(y,t) - \bar{C}^2(y,t)\Big) \frac{h^*(y)}{(\log y)^2}\, dy\, dt - B_h \int_0^1 \hat{A}_{n,h}^2(t) - \big(A^*(t)\big)^2\, dt \\
&= 2 \int_0^1 \int_0^1 \Big(\bar{C}_n(y,t) - \bar{C}(y,t)\Big) \bar{C}(y,t) \frac{h^*(y)}{(\log y)^2}\, dy\, dt - 2B_h \int_0^1 (\hat{A}_{n,h}(t) - A^*(t)) A^*(t)\, dt \\
&\quad + \int_0^1 \int_0^1 \Big(\bar{C}_n(y,t) - \bar{C}(y,t)\Big)^2 \frac{h^*(y)}{(\log y)^2}\, dy\, dt - B_h \int_0^1 (\hat{A}_{n,h}(t) - A^*(t))^2\, dt \\
&= 2 \int_0^1 \int_0^1 \Big(\bar{C}_n(y,t) - \bar{C}(y,t)\Big) \Big(\bar{C}(y,t) - A^*(t)(-\log y)\Big) \frac{h^*(y)}{(\log y)^2}\, dy\, dt \\
&\quad + \int_0^1 \int_0^1 \Big(\bar{C}_n(y,t) - \bar{C}(y,t)\Big)^2 \frac{h^*(y)}{(\log y)^2}\, dy\, dt - B_h \int_0^1 (\hat{A}_{n,h}(t) - A^*(t))^2\, dt \\
&=: S_1 + S_2 + S_3,
\end{aligned} \tag{3.34}$$

where the last identity defines the terms S_1, S_2 and S_3 in an obvious manner. Note that under the null hypothesis of extreme value dependence we have $A^* = A$ and thus $\bar{C}(y,t) = A^*(t)(-\log y)$. This means that under H_0 the term S_1 will vanish and the asymptotic distribution will be determined by the large sample properties of the random variable $S_2 + S_3$. Under the alternative the equality $\bar{C}(y,t) = A^*(t)(-\log y)$ will not hold anymore and it turns out that in this case the statistic is asymptotically dominated by the random variable S_1. With the following results we will derive the limiting distribution of the proposed test statistic under the null hypothesis and the alternative.

Theorem 3.15
Assume that the given copula C satisfies condition (2.3) *and is an extreme value copula with Pickands dependence function A^*. If the function $\bar{w}(y) := h^*(y)/(\log y)^2$ fulfills conditions* (3.16) - (3.17) *for some $\lambda > 2$ and the weight function h is strictly positive and satisfies assumptions (3.18) - (3.19) for $\tilde{\lambda} := \lambda/2 > 1$, then we have for any $\gamma \in (1/2, \lambda/4)$ and $n \to \infty$*

$$nM_h(\tilde{C}_n, \hat{A}_{n,h}) \rightsquigarrow Z_0,$$

where the random variable Z_0 is defined by

$$Z_0 := \int_0^1 \int_0^1 \Big(\frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\Big)^2 \bar{w}(y)\, dy\, dt - B_h \int_0^1 \mathbb{A}^2_{C,h}(t)\, dt$$

with $B_h = \int_0^1 h^*(y)dy$ and the process $\{\mathbb{A}_{C,h}(t)\}_{t\in[0,1]}$ is defined in Theorem 3.8.

The next theorem gives the distribution of the test statistic $M_h(\tilde{C}_n, \hat{A}_{n,h})$ under the alternative. Note that in this case we have $M_h(C, A^*) > 0$.

Theorem 3.16
Assume that the given copula C satisfies $C \geq \Pi$, condition (2.3) *and that $M_h(C, A^*) > 0$. If additionally the weight function h is strictly positive and h and the function $\bar{w}(y) := h^*(y)/(\log y)^2$ satisfy the assumptions* (3.18) - (3.19) *and* (3.16) - (3.17) *for some $\lambda > 1$, respectively, then we have for any $\gamma \in (1/2, (1+\lambda)/4 \wedge \lambda/2)$ and $n \to \infty$*

$$\sqrt{n}(M_h(\tilde{C}_n, \hat{A}) - M_h(C, A^*)) \rightsquigarrow Z_1,$$

where the random variable Z_1 is defined as

$$Z_1 = 2\int_0^1 \int_0^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} v(y,t)\, dy\, dt,$$

with

$$v(y,t) = (\log C(y^{1-t}, y^t) - \log(y)A^*(t))\frac{h^*(y)}{(\log y)^2}.$$

Remark 3.17
(a) Note that the weight functions $h_k^*(y) = -y^k \log y$ satisfy the assumptions of Theorem 4.1 and 4.2 for $k > 1$ and $k > 0$, respectively.

(b) The preceding two theorems yield a consistent asymptotic level α test for the hypothesis of extreme value dependence by rejecting the null hypothesis H_0 if

$$n\, M_h(\tilde{C}_n, \hat{A}_{n,h}) > z_{1-\alpha}, \tag{3.35}$$

where $z_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the distribution of the random variable Z_0.

(c) By Theorem 3.16 the power of the test (3.35) is approximately given by

$$\mathbb{P}\left(n\,M_h(\tilde{C}_n,\hat{A}_{n,h}) > z_{1-\alpha}\right) \approx 1-\Phi\left(\frac{z_{1-\alpha}}{\sqrt{n}\,\sigma} - \sqrt{n}\,\frac{M_h(C,A^*)}{\sigma}\right) \approx \Phi\left(\sqrt{n}\,\frac{M_h(C,A^*)}{\sigma}\right),$$

where the function A^* is defined in (3.6) corresponding to the best approximation of the logarithm of the copula C by a function of the form (3.4), σ is the standard deviation of the distribution of the random variable Z_1 and Φ is the standard normal distribution function. Thus the power of the test (3.35) is an increasing function of the quantity $M_h(C,A^*)\,\sigma^{-1}$.

3.4.2 Multiplier bootstrap

In general the distribution of the random variable Z_0 can not be determined explicitly, because of its complicated dependence on the (unknown) copula C. We hence propose to determine the quantiles by the multiplier bootstrap approach as described in Chapter 2. For the sake of brevity we only consider the *partial derivative multiplier bootstrap* α_n^{pdm} defined in (2.8). As a consequence of Theorem 2.3 we obtain the following bootstrap approximation for Z_0.

Theorem 3.18
If condition (2.3) *is satisfied, the weight function h satisfies the conditions of Theorem 3.15 and $h^*(y)(y\log y)^{-2}$ is uniformly bounded then*

$$\hat{Z}_0^* = \int_0^1\int_0^1 \Big(\frac{\alpha_n^{pdm}(y^{1-t},y^t)}{\tilde{C}_n(y^{1-t},y^t)}\Big)^2 \frac{h^*(y)}{(\log y)^2}\,dy\,dt$$
$$- B_h^{-1}\int_0^1\Big(\int_0^1 \frac{\hat{\alpha}_n^{pdm}(y^{1-t},y^t)}{\tilde{C}_n(y^{1-t},y^t)}\,\frac{h^*(y)}{\log y}\,dy\Big)^2 dt.$$

converges weakly to Z_0 conditional on the data in the sense of Definition 1.20, i.e. $\hat{Z}_0^* \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} Z_0$.

By Theorem 3.18 $\hat{Z}_0^*$ is a valid bootstrap approximation for the distribution of Z_0. Consequently, repeating the procedure B times yields a sample $\hat{Z}_0^*(1),\ldots,\hat{Z}_0^*(B)$ that is approximately distributed according to Z_0 and we can use the empirical $(1-\alpha)$-quantile of this sample, say $z_{1-\alpha}^*$, as an approximation for $z_{1-\alpha}$. Therefore rejecting the null hypothesis if

$$n\,M_h(\tilde{C}_n,\hat{A}_{n,h}) > z_{1-\alpha}^* \tag{3.36}$$

yields a consistent asymptotic level α test for extreme value dependence.
Note that the condition on the boundedness of the function $h^*(y)(y\log y)^2$ is not satisfied for any member of the class $h_k^*(y) = -y^k/\log(y)$ from Example 3.5. Nevertheless, mimicking the procedure from [Kojadinovic and Yan, 2010] and using $h_k^*(y)\mathbb{I}_{[\varepsilon,1-\varepsilon]}(y)$ instead of $h_k^*(y)$ is sufficient for the boundedness. Since this is the procedure being usually performed in practical applications, Theorem 3.18 is still valuable for the weight functions investigated in this thesis.

3.4.3 Finite sample properties

In this subsection we investigate the finite sample properties of the test for extreme value dependence. We consider the asymmetric negative logistic model (3.25), the symmetric mixed model (3.31) and additionally the symmetric model of Gumbel

$$A(t) = \left(t^\theta + (1-t)^\theta\right)^{1/\theta} \tag{3.37}$$

with parameter $\theta \in [1, \infty)$ (see [Gumbel, 1960]) and the model of Hüsler and Reiss

$$A(t) = (1-t)\Phi\left(\theta + \frac{1}{2\theta}\log\frac{1-t}{t}\right) + t\Phi\left(\theta + \frac{1}{2\theta}\log\frac{t}{1-t}\right), \tag{3.38}$$

where $\theta \in (0, \infty)$ and Φ is the standard normal distribution function (see [Hüsler and Reiss, 1989]). The coefficient of tail dependence in (3.38) is given by $\rho = 2(1 - \Phi(\theta))$, i.e. independence is obtained for $\theta \to \infty$ and complete dependence for $\theta \to 0$. For the Gumbel model (3.37) complete dependence is obtained in the limit as θ approaches infinity while independence corresponds to $\theta = 1$. The coefficient of tail dependence $\rho = 2(1 - A(0.5))$ is given by $\rho = 2 - 2^{1/\theta}$.

We generated 1000 random samples of sample size $n = 200$ from various copula models and calculated the probability of rejecting the null hypothesis. Under the null hypothesis we chose the model parameters in such a way that the coefficient of tail dependence ρ varies over the specific range of the corresponding model. Under the alternative the coefficient of tail dependence does not need to exist and we therefore chose the model parameters, such that Kendall's τ is an element of the set $\{1/4, 1/2, 3/4\}$. The weight function is chosen as $h_{0.4}(y) = -y^{0.4}/\log(y)$ and the critical values are determined by the multiplier bootstrap approach as described in Section 3.4.2 with $B = 200$ Bootstrap replications. The results are stated in Table 3.1.

We observe from the left part of Table 3.1 that the level of test is accurately approximated for most of the models, if the tail dependence is not too strong. For a large tail dependence coefficient the bootstrap test is conservative. This phenomenon can be explained by the fact that for the limiting case of random variables distributed according to the upper Fréchet-Hoeffding bound the empirical copula C_n does not converge weakly to a non-degenerate process at a rate $1/\sqrt{n}$, rather in this case it follows that $||C_n - C|| = O(1/n)$. Consequently, the approximations proposed in this thesis, which are based on the weak convergence of $\sqrt{n}(C_n - C)$ to a non-degenerate process, are not appropriate for small samples, if the tail dependence coefficient is large. Considering the alternative we observe reasonably good power for the Frank and Clayton copulas, while for the Gaussian or t-copula deviations from an extreme value copula are not detected well with a sample size $n = 200$. In some cases the power of the test (3.36) is close to the nominal level. This observation again can be explained by the closeness to the upper Fréchet-Hoeffding bound.

Indeed, we can use the minimal distance $M_h(C, A^*)$ as a measure of deviation from an extreme value copula. Calculating the minimal distance $M_h(C, A^*)$ (with Kendall's $\tau = 0.5$

H_0-model	ρ	0.05	0.1
Independence	0	0.031	0.075
Gumbel	0.25	0.045	0.098
	0.5	0.029	0.066
	0.75	0.025	0.065
Mixed model	0.25	0.043	0.09
	0.5	0.047	0.10
Asy. Neg. Log.	0.25	0.041	0.09
	0.5	0.038	0.077
Hüsler-Reiß	0.25	0.04	0.091
	0.5	0.045	0.089
	0.75	0.009	0.053

H_1-model	τ	0.05	0.1
Clayton	0.25	0.874	0.916
	0.5	1	1
	0.75	0.999	1
Frank	0.25	0.291	0.396
	0.5	0.73	0.822
	0.75	0.783	0.898
Gaussian	0.25	0.168	0.240
	0.5	0.237	0.336
	0.75	0.084	0.156
t_4	0.25	0.105	0.187
	0.5	0.158	0.263
	0.75	0.046	0.092

Table 3.1: *Simulated rejection probabilities of the test* (3.36) *for the null hypothesis of an extreme value copula for various models. The first four columns deal with models under the null hypothesis, while the last four are from the alternative.*

and $h = h_{0.4}$) we observe that the minimal distances are about ten times smaller for the Gaussian and t_4 than for the Frank and Clayton copula, i.e.

$$M_h(C, A^*_{\text{Clayton}}) = 1.65 \times 10^{-3}, \qquad M_h(C, A^*_{\text{Frank}}) = 5.87 \times 10^{-4},$$
$$M_h(C, A^*_{\text{Gaussian}}) = 2.08 \times 10^{-4}, \qquad M_h(C, A^*_{t_4}) = 1.18 \times 10^{-4}.$$

Moreover, as explained in Remark 4.3 (b) the power of the tests (3.35) and (3.36) is an increasing function of the quantity $p(\text{copula}) = M_h(C, A^*)\, \sigma^{-1}$. For the four copulas considered in the simulation study (with $\tau = 0.5$) the corresponding ratios are approximately given by

$$p(\text{Clayton}) = 0.230, \quad p(\text{Frank}) = 0.134, \quad p(\text{Gaussian}) = 0.083, \quad p(t_4) = 0.064,$$

which provides a theoretical explanation of the findings presented in Table 3.1. Loosely speaking, if the value $M_h(C, A^*)\, \sigma^{-1}$ is very small a larger sample size is required to detect a deviation from an extreme value copula. This statement is confirmed by further simulations results. For example, for the Gaussian and t_4 copula (with Kendall's $\tau = 0.75$) we obtain for the sample size $n = 500$ the rejection probabilities 0.766 (0.629) and 0.40 (0.544) for the bootstrap test with level 5% (10%), respectively.

3.5 Proofs

Proof of Theorem 3.7. Fix $\lambda > 1$ as in (3.17) and $\gamma \in (1/2, \lambda/2)$. Due to Lemma 1.10.2 (i) in [van der Vaart and Wellner, 1996], the process $\sqrt{n}(\tilde{C}_n - C)$ will have the same weak limit (with respect to the $\rightsquigarrow$ convergence) as $\sqrt{n}(C_n - C)$.
For $i = 2, 3, \ldots$ we consider the following random functions in $l^\infty[0,1]$

$$\begin{aligned} W_n(t) &= \int_0^1 \sqrt{n}\big(\log \tilde{C}_n(y^{1-t}, y^t) - \log C(y^{1-t}, y^t)\big) w(y,t)\, dy, \\ W_{i,n}(t) &= \int_{1/i}^1 \sqrt{n}\big(\log \tilde{C}_n(y^{1-t}, y^t) - \log C(y^{1-t}, y^t)\big) w(y,t)\, dy, \\ W(t) &= \int_0^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} w(y,t)\, dy, \\ W_i(t) &= \int_{1/i}^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} w(y,t)\, dy. \end{aligned}$$

We prove the theorem by an application of Theorem 4.2 in [Billingsley, 1968], adapted to the concept of weak convergence in the sense of Hoffmann-Jørgensen, see e.g. [van der Vaart and Wellner, 1996]. More precisely, we will show in Lemma A.3 in the Appendix that the weak convergence $W_n \rightsquigarrow W$ in $l^\infty[0,1]$ follows from the following three assertions

$$\begin{aligned} &\text{(i) For every } i \geq 2: \; W_{i,n} \rightsquigarrow W_i \;\text{ for } n \to \infty \;\text{ in } l^\infty[0,1], \\ &\text{(ii) } W_i \rightsquigarrow W \;\text{ for } i \to \infty \;\text{ in } l^\infty[0,1], \\ &\text{(iii) For every } \varepsilon > 0: \; \lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*\big(\sup_{t\in[0,1]} |W_{i,n}(t) - W_n(t)| > \varepsilon\big) = 0. \end{aligned} \tag{3.39}$$

The main part of the proof now consists in the verification assertion (iii).
We begin by proving assertion (i). For this purpose set $T_i = [1/i, 1]^2$ and consider the mapping

$$\Phi_1 : \begin{cases} \mathbb{D}_{\Phi_1} \to l^\infty(T_i) \\ f \mapsto \log \circ f \end{cases}$$

where its domain $\mathbb{D}_{\Phi_1}$ is defined by $\mathbb{D}_{\Phi_1} = \{f \in l^\infty(T_i) : \inf_{x\in T_i} |f(x)| > 0\} \subset l^\infty(T_i)$. By Lemma 12.2 in [Kosorok, 2008] it follows that Φ_1 is Hadamard-differentiable at C, tangentially to $l^\infty(T_i)$, with derivative $\Phi'_{1,C}(f) = f/C$. Since $\tilde{C}_n \geq n^{-\gamma}$ and $C \geq \Pi$ we have $\tilde{C}_n, C \in \mathbb{D}_{\Phi_1}$ and the functional delta method (see Theorem 2.8 in [Kosorok, 2008]) yields

$$\sqrt{n}(\log \tilde{C}_n - \log C) \rightsquigarrow \mathbb{G}_C / C$$

in $l^\infty(T_i)$. Next we consider the operator

$$\Phi_2 : \begin{cases} l^\infty(T_i) \to l^\infty([1/i, 1] \times [0,1]) \\ f \mapsto f \circ \varphi, \end{cases}$$

where the mapping $\varphi : [1/i, 1] \times [0, 1] \to T_i$ is defined by $\varphi(y, t) = (y^{1-t}, y^t)$. Observing

$$\sup_{(y,t)\in[1/i,1]\times[0,1]} |f \circ \varphi(y, t) - g \circ \varphi(y, t)| \le \sup_{\mathbf{x}\in T_i} |f(\mathbf{x}) - g(\mathbf{x})|$$

we can conclude that Φ_2 is Lipschitz-continuous. By the continuous mapping theorem (see e.g. Theorem 7.7 in [Kosorok, 2008]) and conditions (3.15) and (3.16) we immediately obtain

$$\sqrt{n}\big(\log \tilde{C}_n(y^{1-t}, y^t) - \log C(y^{1-t}, y^t)\big)w(y, t) \rightsquigarrow \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} w(y, t)$$

in $l^\infty([1/i, 1] \times [0, 1])$. The assertion in (i) now follows by continuity of integration with respect to the variable y.
For the proof of assertion (ii) we simply note that $\mathbb{G}_C$ is bounded on $[0, 1]^2$ and that

$$K(y, t) = \frac{w(y, t)}{C(y^{1-t}, y^t)}$$

is uniformly bounded with respect to $t \in [0, 1]$ by the integrable function $\bar{K}(y) = \bar{w}(y)\, y^{-1}$.
For the proof of assertion (iii) choose some $\alpha \in (0, 1/2)$ such that $\lambda\alpha > \gamma$ and consider the decomposition

$$\begin{aligned} W_n(t) - W_{i,n}(t) &= \int_0^{1/i} \sqrt{n}\big(\log \tilde{C}_n(y^{1-t}, y^t) - \log C(y^{1-t}, y^t)\big)w(y, t)\, dy \\ &= B_i^{(1)}(t) + B_i^{(2)}(t), \end{aligned} \tag{3.40}$$

where

$$B_i^{(j)}(t) = \int_{I_{B_i^{(j)}(t)}} \sqrt{n} \log \frac{\tilde{C}_n}{C}(y^{1-t}, y^t)w(y, t)\, dy,\ j = 1, 2$$

and

$$I_{B_i^{(1)}(t)} = \left\{0 < y < 1/i \,|\, C(y^{1-t}, y^t) > n^{-\alpha}\right\},\ I_{B_i^{(2)}(t)} = (0, 1) \setminus I_{B_i^{(1)}(t)}. \tag{3.41}$$

The usual estimate

$$\mathbb{P}^*\big(\sup_{t\in[0,1]} |W_{i,n}(t) - W_n(t)| > \varepsilon\big) \le \mathbb{P}^*\big(\sup_{t\in[0,1]} |B_i^{(1)}(t)| > \varepsilon/2\big) + \mathbb{P}^*\big(\sup_{t\in[0,1]} |B_i^{(2)}(t)| > \varepsilon/2\big) \tag{3.42}$$

allows for individual investigation of both terms, and we begin with $\sup_{t\in[0,1]} |B_i^{(1)}(t)|$.
By the mean value theorem applied to the logarithm we have

$$\log \frac{\tilde{C}_n}{C}(y^{1-t}, y^t) = \log \tilde{C}_n(y^{1-t}, y^t) - \log C(y^{1-t}, y^t) = (\tilde{C}_n - C)(y^{1-t}, y^t)\, \frac{1}{C^*(y, t)}, \tag{3.43}$$

where $C^*(y,t)$ is some intermediate point satisfying

$$|C^*(y,t) - C(y^{1-t}, y^t)| \le |\tilde{C}_n(y^{1-t}, y^t) - C(y^{1-t}, y^t)|.$$

Especially, observing $C \ge \Pi$ we have

$$C^*(y,t) \ge (C \wedge \tilde{C}_n)(y^{1-t}, y^t) \ge y \wedge \left(y \, \frac{\tilde{C}_n}{C}(y^{1-t}, y^t) \right)$$

and therefore

$$\begin{aligned} \sup_{t\in[0,1]} |B_i^{(1)}(t)| &\le \sup_{t\in[0,1]} \int_{I_{B_i^{(1)}(t)}} \sqrt{n} \Big| (\tilde{C}_n - C)(y^{1-t}, y^t) \Big| \times \Big| 1 \vee \frac{C}{\tilde{C}_n}(y^{1-t}, y^t) \Big| \, w(y,t) \, y^{-1} \, dy \\ &\le \sup_{\mathbf{x}\in[0,1]^2} \sqrt{n} |\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})| \times \Big(1 \vee \sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})>n^{-\alpha}} \Big| \frac{C}{\tilde{C}_n}(\mathbf{x}) \Big| \Big) \times \psi(i), \end{aligned}$$

with $\psi(i) = \int_0^{1/i} \bar{w}(y) \, y^{-1} \, dy = o(1)$ for $i \to \infty$. This yields for the first term on the right hand side of (3.42)

$$\begin{aligned} &\mathbb{P}^*(\sup_{t\in[0,1]} |B_i^{(1)}(t)| > \varepsilon) \\ \le &\mathbb{P}^*\Big(\sup_{\mathbf{x}\in[0,1]^2} \sqrt{n} |\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})| > \sqrt{\frac{\varepsilon}{\psi(i)}} \Big) + \mathbb{P}^*\Big(1 \vee \sup_{C(\mathbf{x})>n^{-\alpha}} \Big| \frac{C}{\tilde{C}_n}(\mathbf{x}) \Big| > \sqrt{\frac{\varepsilon}{\psi(i)}} \Big). \end{aligned} \quad (3.44)$$

Since $\sup_{\mathbf{x}\in[0,1]^2} \sqrt{n} |\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})|$ is asymptotically tight we immediately obtain

$$\lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*\Big(\sup_{\mathbf{x}\in[0,1]^2} \sqrt{n} |\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})| > \sqrt{\frac{\varepsilon}{\psi(i)}} \Big) = 0. \quad (3.45)$$

For the estimation of the second term in equation (3.44) we note that

$$\sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})>n^{-\alpha}} \Big| \frac{\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})}{C(\mathbf{x})} \Big| < n^{\alpha} \sup_{\mathbf{x}\in[0,1]^2} |\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})| \xrightarrow{\mathbb{P}} 0,$$

which in turn implies

$$\begin{aligned} &\sup_{C(\mathbf{x})>n^{-\alpha}} \Big| \frac{C}{\tilde{C}_n}(\mathbf{x}) \Big| = \sup_{C(\mathbf{x})>n^{-\alpha}} \Big| 1 + \frac{\tilde{C}_n - C}{C}(\mathbf{x}) \Big|^{-1} \\ \le &\Big(1 - \sup_{C(\mathbf{x})>n^{-\alpha}} \Big| \frac{\tilde{C}_n - C}{C}(\mathbf{x}) \Big| \Big)^{-1} \mathbb{I}_{A_n} + \Big(\sup_{C(\mathbf{x})>n^{-\alpha}} \Big| 1 + \frac{\tilde{C}_n - C}{C}(\mathbf{x}) \Big|^{-1} \Big) \mathbb{I}_{\Omega\setminus A_n} \xrightarrow{\mathbb{P}} 1, \end{aligned} \quad (3.46)$$

where $A_n = \Big\{ \sup_{C(\mathbf{x})>n^{-\alpha}} \Big| \frac{\tilde{C}_n - C}{C}(\mathbf{x}) \Big| < 1/2 \Big\}$. Thus the function

$$\max \Big\{ 1, \sup_{C(\mathbf{x})>n^{-\alpha}} \Big| \frac{C}{\tilde{C}_n}(\mathbf{x}) \Big| \Big\}$$

can be bounded by a function that converges to one in outer probability, which implies

$$\lim_{i\to\infty}\limsup_{n\to\infty}\mathbb{P}^*\Big(1\vee\sup_{C(\mathbf{x})>n^{-\alpha}}\Big|\frac{C}{\tilde{C}_n}(\mathbf{x})\Big|>\sqrt{\frac{\varepsilon}{\psi(i)}}\Big)=0.$$

Observing (3.44) and (3.45) it remains to estimate the second term on the right hand side of (3.42). We make use of the mean value theorem again, see equation (3.43), but use the estimate

$$C^*(y,t)\geq (C\wedge\tilde{C}_n)(y^{1-t},y^t)\geq y^\lambda\wedge y^\lambda\,\frac{\tilde{C}_n}{C^\lambda}(y^{1-t},y^t) \tag{3.47}$$

(recall that $\lambda>1$ by assumption (3.17)). This yields

$$\begin{aligned}\sup_{t\in[0,1]}|B_i^{(2)}(t)| &\leq \sup_{t\in[0,1]}\int_{I_{B_i^{(2)}}(t)}\sqrt{n}\Big|(\tilde{C}_n-C)(y^{1-t},y^t)\Big|\times\Big|1\vee\frac{C^\lambda}{\tilde{C}_n}(y^{1-t},y^t)\Big|\,w(y,t)\,y^{-\lambda}\,dy\\ &\leq \sup_{\mathbf{x}\in[0,1]^2}\sqrt{n}|\tilde{C}_n(\mathbf{x})-C(\mathbf{x})|\times\Big(1\vee\sup_{\mathbf{x}\in[0,1]^2:C(\mathbf{x})\leq n^{-\alpha}}\Big|\frac{C^\lambda}{\tilde{C}_n}(\mathbf{x})\Big|\Big)\times\phi(i),\end{aligned}$$

where $\phi(i)=\int_0^{1/i}\bar{w}(y)y^{-\lambda}\,dy=o(1)$ for $i\to\infty$ by condition (3.17). Using analogous arguments as for the estimation of $\sup_{t\in[0,1]}|B_i^{(1)}(t)|$ the assertion follows from

$$\sup_{\mathbf{x}\in[0,1]^2:C(\mathbf{x})\leq n^{-\alpha}}\Big|\frac{C^\lambda}{\tilde{C}_n}(\mathbf{x})\Big|\leq\sup_{\mathbf{x}\in[0,1]^2:C(\mathbf{x})\leq n^{-\alpha}}\Big|n^\gamma C^\lambda(\mathbf{x})\Big|\leq n^{\gamma-\lambda\alpha}=o(1)$$

due to the choice of γ and α. □

Proof of Theorem 3.10. The proof will also be based on Lemma A.3 in the Appendix verifying conditions (i) - (iii) in (3.39). A careful inspection of the previous proof shows that the verification of condition (i) in (3.39) remains valid. Regarding condition (ii) we have to show that the process $\frac{\mathbb{G}_C}{C}(y^{1-t},y^t)$ is integrable on the interval $(0,1)$. For this purpose we write

$$\mathbb{G}_C(\mathbf{x})=\mathbb{B}_C(\mathbf{x})-\partial_1C(\mathbf{x})\mathbb{B}_C(x_1,1)-\partial_2C(\mathbf{x})\mathbb{B}_C(1,x_2)$$

and consider each term separately. From Theorem G.1 in [Genest and Segers, 2009] we know that for any $\omega\in(0,1/2)$ the process

$$\tilde{\mathbb{B}}_C(\mathbf{x})=\begin{cases}\frac{\mathbb{B}_C(\mathbf{x})}{(x_1\wedge x_2)^\omega(1-x_1\wedge x_2)^\omega} & \text{, if } \quad x_1\wedge x_2\in(0,1)\\ 0 & \text{, if } \quad x_1=0 \text{ or } x_2=0 \text{ or } \mathbf{x}=(1,1),\end{cases}$$

has continuous sample paths on $[0,1]^2$. Considering $C(y^{1-t},y^t)\geq y$ and using the notation

$$\begin{aligned}K_1(y,t)&=q_\omega(y^{1-t}\wedge y^t)y^{-1}\\ K_2(y,t)&=\partial_1C(y^{1-t},y^t)q_\omega(y^{1-t})y^{-1}\\ K_3(y,t)&=\partial_2C(y^{1-t},y^t)q_\omega(y^t)y^{-1}\end{aligned} \tag{3.48}$$

with $q_\omega(t) = t^\omega(1-t)^\omega$ it remains to show that there exist integrable functions $K_j^*(y)$ with $K_j(y,t) \le K_j^*(y)$ for all $t \in [0,1]$ $(j = 1,2,3)$. For K_1 this is immediate because $K_1(y,t) \le (y^{1-t} \wedge y^t)^\omega y^{-1} \le y^{\omega/2-1}$. For K_2, note that $\partial_1 C(y^{1-t}, y^t) = \mu(t)\, y^{A(t)-(1-t)}$, with $\mu(t) = A(t) - tA'(t)$. Therefore

$$K_2(y,t) \le \mu(t)\, y^{A(t)-(1-\omega)(1-t)-1} \le \mu(t)\, y^{\omega/2-1} \le 2\, y^{\omega/2-1}, \tag{3.49}$$

where the second estimate follows from the inequality $t \vee (1-t) \le A(t) \le 1$ and holds for $\omega \in (0,2)$. A similar argument works for the term K_3.
For the verification of condition (iii) we proceed along similar lines as in the previous proof. We begin by choosing some $\beta \in (1, 9/8), \omega \in (1/4, 1/2)$ and some $\alpha \in (4/9, \gamma \wedge (2-\omega)^{-1})$ in such a way that $\gamma < \beta\alpha$. First note that $y \le 1/(n+2)^2$ implies $\tilde{C}_n(y^{1-t}, y^t) = n^{-\gamma}$ for all $t \in [0,1]$. This yields

$$\int_0^{(n+2)^{-2}} \sqrt{n}(\log \tilde{C}_n - \log C)(y^{1-t}, y^t)\, dy = O\Big(\frac{\log n}{n^{3/2}}\Big)$$

uniformly with respect to $t \in [0,1]$, and therefore it is sufficient to consider the decomposition in (3.40) with the sets

$$I_{B_i^{(1)}(t)} = \{1/(n+2)^2 < y < 1/i \,|\, C(y^{1-t}, y^t) > n^{-\alpha}\}, \;\; I_{B_i^{(2)}(t)} = (1/(n+2)^2, 1/i) \setminus I_{B_i^{(1)}(t)}.$$

We can estimate the term $B_i^{(1)}(t)$ analogously to the previous proof by

$$|B_i^{(1)}(t)| \le \int_{I_{B_i^{(1)}(t)}} \sqrt{n}\Big|(\tilde{C}_n - C)(y^{1-t}, y^t)\Big| \times \Big|1 \vee \frac{C}{\tilde{C}_n}(y^{1-t}, y^t)\Big|\, y^{-1}\, dy.$$

Let H_n denote the empirical distribution function of the sample $\{F_1(X_{i1}), F_2(X_{i2})\}_{i=1}^n$. By the results in [Segers, 2010], Section 5, we can decompose $\sqrt{n}(\tilde{C}_n - C) = \sqrt{n}(C_n \vee n^{-\gamma} - C)$ as follows

$$\begin{aligned}\sqrt{n}(\tilde{C}_n - C)(\mathbf{x}) &= \sqrt{n}(C_n - C)(\mathbf{x}) + \sqrt{n}(\tilde{C}_n - C_n)(\mathbf{x}) \\ &= \alpha_n(\mathbf{x}) - \partial_1 C(\mathbf{x})\alpha_n(x_1, 1) - \partial_2 C(\mathbf{x})\alpha_n(1, x_2) + \tilde{R}_n(\mathbf{x}),\end{aligned} \tag{3.50}$$

where $\alpha_n(\mathbf{x}) = \sqrt{n}(H_n - C)(\mathbf{x})$ and the remainder satisfies

$$\sup_{\mathbf{x}\in[0,1]^2} |\tilde{R}_n(\mathbf{x})| = O(n^{1/2-\gamma} + n^{-1/4}(\log n)^{1/2}(\log\log n)^{3/4}) \quad \text{a.s.} \tag{3.51}$$

Note that the estimate of (3.51) requires validity of condition 5.1 in [Segers, 2010]. This condition is satisfied provided that the function A is assumed to be twice continuously differentiable, see Example 6.3 in [Segers, 2010]. With (3.50) we can estimate the term

$|B_i^{(1)}(t)|$ analogously to decomposition (3.40) by $B_{i,1}^{(1)}(t) + \cdots + B_{i,4}^{(1)}(t)$, where

$$B_{i,1}^{(1)}(t) = \int_{I_{B_i^{(1)}(t)}} \left|\alpha_n(y^{1-t}, y^t)\right| \left|1 \vee \frac{C}{\tilde{C}_n}(y^{1-t}, y^t)\right| y^{-1}\, dy,$$

$$B_{i,2}^{(1)}(t) = \int_{I_{B_i^{(1)}(t)}} \partial_1 C(y^{1-t}, y^t) \left|\alpha_n(y^{1-t}, 1)\right| \left|1 \vee \frac{C}{\tilde{C}_n}(y^{1-t}, y^t)\right| y^{-1}\, dy,$$

$$B_{i,3}^{(1)}(t) = \int_{I_{B_i^{(1)}(t)}} \partial_2 C(y^{1-t}, y^t) \left|\alpha_n(1, y^t)\right| \left|1 \vee \frac{C}{\tilde{C}_n}(y^{1-t}, y^t)\right| y^{-1}\, dy,$$

$$B_{i,4}^{(1)}(t) = \int_{I_{B_i^{(1)}(t)}} \left|\tilde{R}_n(y^{1-t}, y^t)\right| \left|1 \vee \frac{C}{\tilde{C}_n}(y^{1-t}, y^t)\right| y^{-1}\, dy.$$

The decomposition in (3.50), Theorem G.1 in [Genest and Segers, 2009] and the inequality $\alpha < \gamma \wedge (2-\omega)^{-1}$ may be used to conclude

$$\sup_{(y,t):\, C(y^{1-t}, y^t) > n^{-\alpha}} \left| \frac{\tilde{C}_n - C}{C}(y^{1-t}, y^t) \right| = o_{\mathbb{P}}(1),$$

which in turn implies

$$1 \vee \sup_{(y,t):\, C(y^{1-t}, y^t) > n^{-\alpha}} \left| \frac{C}{\tilde{C}_n}(y^{1-t}, y^t) \right| = O_{\mathbb{P}}(1) \tag{3.52}$$

analogously to (3.46). Together with (3.51) and the inequality $\int_{(n+2)^{-2}}^{1/i} y^{-1}\, dy \le 2\log(n+2)$ we obtain, for $n \to \infty$

$$\sup_{t\in[0,1]} B_{i,4}^{(1)}(t) = O_{\mathbb{P}}(n^{1/2-\gamma}\log n + n^{-1/4}(\log n)^{3/2}(\log\log n)^{3/4}) = o_{\mathbb{P}}(1),$$

which implies

$$\lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*\Big(\sup_{t\in[0,1]} B_{i,4}^{(1)}(t) > \varepsilon/4 \Big) = 0.$$

Observing that $q_\omega(y^{1-t} \wedge y^t) \le y^{\omega/2}$ the first term $B_{i,1}^{(1)}(t)$ can be estimated by

$$\sup_{t\in[0,1]} B_{i,1}^{(1)}(t) \le \sup_{\mathbf{x}\in[0,1]^2} \frac{|\alpha_n(\mathbf{x})|}{q_\omega(x_1 \wedge x_2)} \times \Big(1 \vee \sup_{(y,t):\, C(y^{1-t}, y^t) > n^{-\alpha}} \left| \frac{C}{\tilde{C}_n}(y^{1-t}, y^t) \right| \Big) \times \psi(i),$$

where $\psi(i) = \int_0^{1/i} y^{-1+\omega/2}\, dy = o(1)$ for $i \to \infty$. Using analogous arguments as in the previous proof we can conclude, under consideration of (3.52) and Theorem G.1 in

[Genest and Segers, 2009], that $\lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*(\sup_{t\in[0,1]} B_{i,1}^{(1)}(t) > \varepsilon/4) = 0$. For the second summand we note that

$$\sup_{t\in[0,1]} B_{i,2}^{(1)}(t) \le \sup_{x_1\in[0,1]} \frac{|\alpha_n(x_1,1)|}{q_\omega(x_1)} \times \Big(1 \vee \sup_{(y,t):\, C(y^{1-t},y^t)>n^{-\alpha}} \Big|\frac{C}{\tilde{C}_n}(y^{1-t},y^t)\Big|\Big) \times \sup_{t\in[0,1]} \int_0^{1/i} K_2(y,t)\,dy,$$

where $K_2(y,t)$ is defined in (3.48). From (3.49), we have $\lim_{i\to\infty} \sup_{t\in[0,1]} \int_0^{1/i} K_2(y,t)\,dy = 0$. Again, under consideration of (3.52) and Theorem G.1 in [Genest and Segers, 2009], we obtain $\lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*\Big(\sup_{t\in[0,1]} B_{i,2}^{(1)}(t) > \varepsilon/4\Big) = 0$. A similar argument works for $B_{i,3}^{(1)}$ and from the estimates for the different terms the assertion

$$\lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*(\sup_{t\in[0,1]} |B_i^{(1)}(t)| > \varepsilon) = 0$$

follows. Considering the term $\sup_{t\in[0,1]} |B_i^{(2)}(t)|$ we proceed along similar lines as in the proof of Theorem 3.7. For the sake of brevity we only state the important differences: in estimation (3.47) replace λ by β, then make use of decomposition (3.50), calculations similar to (3.49), and Theorem G.1 in [Genest and Segers, 2009] again and for the estimation of the remainder note that $\int_{1/(n+2)^2}^{1/i} y^{-\beta}\,dy = O(n^{2(\beta-1)})$. □

Proof of Proposition 3.13. The proof follows by similar lines as given in the proof of Proposition 3.3 in [Genest and Segers, 2009]. Observing that $\partial_1 C(u^{1-t},u^t) = u^{A(t)+t-1}\mu(t)$ and $\partial_2 C(u^{1-t},u^t) = u^{A(t)-t}\nu(t)$ we can decompose $\sigma(u,v;t)$ into

$$\sigma(u,v;t) = \sigma_0(u,v;t) + (uv)^{A(t)}\Big\{\sum_{l=1}^4 \sigma_l(u,v;t) - \sum_{l=5}^8 \sigma_l(u,v;t)\Big\},$$

where

$$\begin{aligned}
\sigma_0(u,v;t) &= (u\wedge v)^{A(t)} - (uv)^{A(t)} \\
\sigma_1(u,v;t) &= (u^{t-1}\wedge v^{t-1} - 1)\mu^2(t) \\
\sigma_2(u,v;t) &= (u^{-t}\wedge v^{-t} - 1)\nu^2(t) \\
\sigma_3(u,v;t) &= (u^{t-1}v^{-t}C(u^{1-t},v^t) - 1)\mu(t)\nu(t) \\
\sigma_4(u,v;t) &= (u^{-t}v^{1-t}C(v^{1-t},u^t) - 1)\mu(t)\nu(t) \\
\sigma_5(u,v;t) &= (u^{-A(t)}v^{t-1}C(u^{1-t}\wedge v^{1-t},u^t) - 1)\mu(t) \\
\sigma_6(u,v;t) &= (u^{t-1}v^{-A(t)}C(u^{1-t}\wedge v^{1-t},v^t) - 1)\mu(t) \\
\sigma_7(u,v;t) &= (u^{-A(t)}v^{-t}C(u^{1-t},u^t\wedge v^t) - 1)\nu(t) \\
\sigma_8(u,v;t) &= (u^{-t}v^{-A(t)}C(v^{1-t},u^t\wedge v^t) - 1)\nu(t).
\end{aligned}$$

In view of (3.21) we need to evaluate the integrals $\int_0^1 \int_0^1 \sigma_0(u,v;t)(uv)^{k-A(t)}\,du\,dv$ and $\int_0^1 \int_0^1 \sigma_l(u,v;t)(uv)^k\,du\,dv$ for $l = 1,\ldots,8$. By symmetry, some of these integrals coincide, that is

$$\int_0^1 \int_0^1 \sigma_l(u,v;t)(uv)^k\,du\,dv = \int_0^1 \int_0^1 \sigma_{l+1}(u,v;t)(uv)^k\,du\,dv, \qquad l = 3,5,7.$$

Considering the remaining integrals straightforward calculations yield

$$\begin{aligned}
\int_0^1 \int_0^1 \sigma_0(u,v;t)(uv)^{k-A(t)}\,du\,dv &= \frac{2}{(k+1)(2k+2-A(t))} - \frac{1}{(k+1)^2}, \\
\int_0^1 \int_0^1 \sigma_1(u,v;t)(uv)^k\,du\,dv &= \Big(\frac{2}{(k+1)(2k+1+t)} - \frac{1}{(k+1)^2}\Big)\mu^2(t), \\
\int_0^1 \int_0^1 \sigma_2(u,v;t)(uv)^k\,du\,dv &= \Big(\frac{2}{(k+1)(2k+2-t)} - \frac{1}{(k+1)^2}\Big)\nu^2(t).
\end{aligned}$$

Regarding the integral with respect to σ_3 we need to evaluate

$$H_1(t) = \int_0^1 \int_0^1 u^{k+t-1}v^{k-t}C(u^{1-t},v^t)\,du\,dv = \frac{1}{t(1-t)}\int_0^1 \int_0^1 C(x,y)x^{\frac{k+1}{1-t}-2}y^{\frac{k+1}{t}-2}\,dx\,dy,$$

where we have used the substitution $u^{1-t} = x$ and $v^t = y$. Next substitute $x = w^{1-s}$ and $y = w^s$, then $w = xy \in (0,1]$ and $s = \frac{\log y}{\log xy} \in [0,1]$, while the Jacobian of the transformation is given by $-\log w$. One obtains

$$H_1(t) = \frac{1}{t(1-t)}\int_0^1 \Big(A(s) + (k+1)\Big(\frac{1-s}{1-t} + \frac{s}{t}\Big) - 1\Big)^{-2} ds,$$

where the last equality follows by integration by parts. In consequence,

$$\begin{aligned}
&\int_0^1 \int_0^1 \sigma_3(u,v;t)(uv)^k\,du\,dv \\
&= \Big\{\frac{1}{t(1-t)}\int_0^1 \Big(A(s) + (k+1)\Big(\frac{1-s}{1-t} + \frac{s}{t}\Big) - 1\Big)^{-2} ds - \frac{1}{(k+1)^2}\Big\}\mu(t)\nu(t).
\end{aligned}$$

Regarding the integral of σ_5 we decompose

$$\begin{aligned}
&\int_0^1 \int_0^1 u^{k-A(t)}v^{k+t-1}C(u^{1-t}\wedge v^{1-t},u^t)\,du\,dv \\
&= \int_0^1 \int_0^v u^{k-A(t)}v^{k+t-1}C(u^{1-t},u^t)\,du\,dv + \int_0^1 \int_v^1 u^{k-A(t)}v^{k+t-1}C(v^{1-t},u^t)\,du\,dv.
\end{aligned}$$

Straightforward calculations show that the first integral equals $\big((k+1)(2k+1+t)\big)^{-1}$. For the second integral we substitute $v^{1-t} = x$ and $u^t = y$ to obtain

$$\frac{1}{t(1-t)}\int_0^1 \int_0^{y^{(1-t)/t}} y^{\frac{k+1-A(t)}{t}-1}x^{\frac{k+1}{1-t}-2}C(x,y)\,dx\,dy.$$

We proceed by the same transformation as for σ_3, namely $x = w^{1-s}$ and $y = w^s$. The inequality $x < y^{(1-t)/t}$ transforms to $t > s$ and in consequence the latter integral equals

$$
\begin{aligned}
&-\frac{1}{t(1-t)}\int_0^t\int_0^1 w^{s\left(\frac{k+1-A(t)}{t}-1\right)+(1-s)\left(\frac{k+1}{1-t}-2\right)+A(s)}\log w\,dw\,ds \\
&=\frac{1}{t(1-t)}\int_0^t\left(A(s)+(k+t)\frac{1-s}{1-t}+(k+1-A(t))\,\frac{s}{t}\right)^{-2}ds,
\end{aligned}
$$

where the last equality follows by integration by parts. Combining all terms for σ_5 we obtain

$$
\begin{aligned}
\int_0^1\int_0^1 \sigma_5(u,v;t)(uv)^k\,du\,dv = \mu(t)\bigg\{&\frac{1}{(k+1)(2k+1+t)} \\
&+\frac{1}{t(1-t)}\int_0^t\left(A(s)+(k+t)\frac{1-s}{1-t}+(k+1-A(t))\,\frac{s}{t}\right)^{-2}ds-\frac{1}{(k+1)^2}\bigg\}.
\end{aligned}
$$

For the integrals with respect to σ_7 similar calculations yield

$$
\begin{aligned}
\int_0^1\int_0^1 \sigma_7(u,v;t)(uv)^k\,du\,dv = \nu(t)\bigg\{&\frac{1}{(k+1)(2k+2-t)} \\
&+\frac{1}{t(1-t)}\int_t^1\left(A(s)+(k+1-A(t))\,\frac{1-s}{1-t}+(k+1-t)\frac{s}{t}\right)^{-2}ds-\frac{1}{(k+1)^2}\bigg\}
\end{aligned}
$$

and the conclusion finally follows by assembling all terms. □

Proof of Theorem 3.14. Let η denote a probability measure minimizing the functional V defined in (3.22). Note that V is convex and define for $\alpha \in [0,1]$ and a further probability measure ξ on $[0,1]$ the function

$$g(\alpha) = V(\alpha\xi + (1-\alpha)\eta).$$

Because V is convex it follows that η is optimal if and only if the directional derivative of η in the direction $\xi - \eta$ satisfies

$$
\begin{aligned}
0 \le g'(0+) &= \lim_{\alpha\to 0+}\frac{g(\alpha)-g(0)}{\alpha} \\
&= 2\int_0^1\int_0^1 k_t(x,y)d\xi(x)d\eta(y) - 2\int_0^1\int_0^1 k_t(x,y)d\eta(x)d\eta(y)
\end{aligned}
$$

for all probability measures ξ. Using Dirac measures for ξ yields that this inequality is equivalent to (3.23), which proves Theorem 3.14. □

Proof of Theorem 3.15. Since the integration mapping is continuous, it suffices to establish the weak convergence $W_n(t) \rightsquigarrow W(t)$ in $l^\infty[0,1]$ where we define

$$
\begin{aligned}
W_n(t) &= \int_0^1 n\Big(\log\frac{\tilde{C}_n(y^{1-t},y^t)}{C(y^{1-t},y^t)}\Big)^2\bar{w}(y)\,dy - nB_h(\hat{A}_{n,h}(t)-A^*(t))^2, \\
W(t) &= \int_0^1\Big(\frac{\mathbb{G}_C(y^{1-t},y^t)}{C(y^{1-t},y^t)}\Big)^2\bar{w}(y)\,dy - B_h\mathbb{A}^2_{C,h}(t).
\end{aligned}
$$

We prove this assertion along similar lines as in the proof of Theorem 3.7. For $i \geq 2$ we recall the notation $\bar{w}(y) = h^*(y)/(\log y)^2$ and consider the following random functions in $l^\infty[0,1]$

$$W_{i,n}(t) = \int_{1/i}^{1} n\Big(\log \frac{\tilde{C}_n(y^{1-t},y^t)}{C(y^{1-t},y^t)}\Big)^2 \bar{w}(y)\,dy - B_h^{-1}\Big(\int_{1/i}^{1} \sqrt{n}\Big(\log \frac{\tilde{C}_n(y^{1-t},y^t)}{C(y^{1-t},y^t)}\Big)\frac{h^*(y)}{\log y}\,dy\Big)^2$$

$$W_i(t) = \int_{1/i}^{1} \Big(\frac{\mathbb{G}_C(y^{1-t},y^t)}{C(y^{1-t},y^t)}\Big)^2 \bar{w}(y)\,dy - B_h^{-1}\Big(\int_{1/i}^{1} \frac{\mathbb{G}_C(y^{1-t},y^t)}{C(y^{1-t},y^t)}\frac{h^*(y)}{\log y}\,dy\Big)^2.$$

By an application of Lemma A.3 in the Appendix, it suffices to show the conditions listed in (3.39). By arguments similar to those in the proof of Theorem 3.7 we obtain

$$\sqrt{n}\log\frac{\tilde{C}_n(y^{1-t},y^t)}{C(y^{1-t},y^t)} \rightsquigarrow \frac{\mathbb{G}_C(y^{1-t},y^t)}{C(y^{1-t},y^t)}$$

in $l^\infty([1/i,1]\times[0,1])$. Assertion (i) now follows immediately by the boundedness of the functions $\bar{w}(y)$ and $h^*(y)(-\log y)^{-1}$ on $[1/i,1]$ (see conditions (3.15), (3.16) and (3.18)) and the continuous mapping theorem.
For the proof of assertion (ii) we simply note that $\mathbb{G}_C^2$ and $\mathbb{G}_C$ are bounded on $[0,1]^2$ and $K_1(y,t) = \frac{\bar{w}(y)}{C^2(y^{1-t},y^t)}$ and $K_2(y,t) = \frac{h^*(y)}{C(y^{1-t},y^t)}$ are bounded uniformly with respect to $t \in [0,1]$ by the integrable functions $\bar{K}_1(y) = \bar{w}(y)\,y^{-2}$ and $\bar{K}_2(y) = h^*(y)(-\log y)^{-1}\,y^{-1}$.

For the proof of assertion (iii) we fix some $\alpha \in (0,1/2)$ such that $\lambda\alpha > 2\gamma$ and consider the decomposition

$$W_n(t) - W_{i,n}(t) = B_i^{(1)}(t) + B_i^{(2)}(t) + B_i^{(3)}(t),$$

where

$$B_i^{(1)}(t) = \int_{I_{B_i^{(1)}}(t)} n\Big(\log\frac{\tilde{C}_n(y^{1-t},y^t)}{C(y^{1-t},y^t)}\Big)^2 \bar{w}(y)\,dy,$$

$$B_i^{(2)}(t) = \int_{I_{B_i^{(2)}}(t)} n\Big(\log\frac{\tilde{C}_n(y^{1-t},y^t)}{C(y^{1-t},y^t)}\Big)^2 \bar{w}(y)\,dy,$$

$$B_i^{(3)}(t) = -B_h^{-1} I(t,1/i)(2I(t,1) - I(t,1/i)),$$

$I_{B_i^{(1)}}(t)$ and $I_{B_i^{(2)}}(t)$ are defined in (3.41) and

$$I(t,a) = \sqrt{n}\int_0^a \Big(\log\frac{\tilde{C}_n(y^{1-t},y^t)}{C(y^{1-t},y^t)}\Big)\frac{h^*(y)}{\log y}\,dy.$$

By the same arguments as in the proof of Theorem 3.7 we have for every $\varepsilon > 0$

$$\lim_{i\to\infty}\limsup_{n\to\infty}\mathbb{P}^*\big(\sup_{t\in[0,1]}|I(t,1/i)| > \varepsilon\big) = 0,$$

and $\sup_{t\in[0,1]} |I(t,1)| = O_{\mathbb{P}^*}(1)$, which yields $\lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*(\sup_{t\in[0,1]} |B_i^{(3)}(t)| > \varepsilon) = 0$. For $B_i^{(1)}(t)$ we obtain the estimate

$$\begin{aligned}\sup_{t\in[0,1]} |B_i^{(1)}(t)| &\le \sup_{t\in[0,1]} \int_{I_{B_i^{(1)}(t)}} n\Big|(\tilde{C}_n - C)(y^{1-t}, y^t)\Big|^2 \Big|1 \vee \frac{C^2}{\tilde{C}_n^2}(y^{1-t}, y^t)\Big|\, \bar{w}(y)\, y^{-2}\, dy \\ &\le \sup_{\mathbf{x}\in[0,1]^2} n|\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})|^2 \times \Big(1 \vee \sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})>n^{-\alpha}} \Big|\frac{C^2}{\tilde{C}_n^2}(\mathbf{x})\Big|\Big) \times \psi(i),\end{aligned}$$

where $\psi(i) := \int_0^{1/i} \bar{w}(y) y^{-2} dy$, which can be handled by the same arguments as in the proof of Theorem 3.7. Finally, the term $B_i^{(2)}(t)$ can be estimated by

$$\begin{aligned}\sup_{t\in[0,1]} |B_i^{(2)}(t)| &\le \sup_{t\in[0,1]} \int_{I_{B_i^{(2)}(t)}} n\Big|(\tilde{C}_n - C)(y^{1-t}, y^t)\Big|^2 \Big|1 \vee \frac{C^\lambda}{\tilde{C}_n^2}(y^{1-t}, y^t)\Big|\, \bar{w}(y)\, y^{-\lambda}\, dy \\ &\le \sup_{\mathbf{x}\in[0,1]^2} n|\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})|^2 \times \Big(1 \vee \sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})\le n^{-\alpha}} \Big|\frac{C^\lambda}{\tilde{C}_n^2}(\mathbf{x})\Big|\Big) \times \phi(i),\end{aligned}$$

where $\phi(i) = \int_0^{1/i} \bar{w}(y) y^{-\lambda}\, dy = o(1)$ for $i \to \infty$ by condition (3.17). Mimicking the arguments from the proof of Theorem 3.7 completes the proof. □

Proof of Theorem 3.16. Recall the decomposition $M_h(\tilde{C}_n, \hat{A}_{n,h}) - M_h(C, A^*) = S_1 + S_2 + S_3$ where S_1, S_2 and S_3 are defined in (3.34). With the notation $\bar{v}(y) := 2\, h^*(y)/(-\log y)$ it follows that $|v(y,t)| \le \bar{v}(y)$ and the assumptions on h yield the validity of (3.15)-(3.17) for $v(y,t)$. This allows for an application of Theorem 3.7 and together with the continuous mapping theorem we obtain $\sqrt{n} S_1 \rightsquigarrow Z_1$, where Z_1 is the limiting process defined in (3.16). Thus it remains to verify the negligibility of $S_2 + S_3$. For S_3 we note that by Theorem 3.8 and the continuous mapping theorem we have $S_3 = O_{\mathbb{P}}(1/n)$ and it remains to consider S_2. To this end we fix some $\alpha \in (0, 1/2)$ such that $(1 + (\lambda - 1)/2)\alpha > \gamma$ and consider the decomposition

$$\begin{aligned}&\int_0^1 \log^2 \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} \frac{h^*(y)}{(\log y)^2}\, dy \\ &= \int_{I_{B_1^{(1)}(t)}} \log^2 \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} \frac{h^*(y)}{(\log y)^2}\, dy + \int_{I_{B_1^{(2)}(t)}} \log^2 \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} \frac{h^*(y)}{(\log y)^2}\, dy \\ &=: T_1(t,n) + T_2(t,n)\end{aligned}$$

where the sets $I_{B_1^{(j)}(t)}, j = 1, 2$ are defined in (3.41). On the set $I_{B_1^{(1)}(t)}$ we use the estimate

$$\begin{aligned}\log^2 \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} &\le \frac{|\tilde{C}_n - C|^2}{(C^*)^2}(y^{1-t}, y^t) \le \frac{|\tilde{C}_n - C|^2}{C^*}(y^{1-t}, y^t) \frac{1}{n^{-\alpha}(1 \wedge \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)})} \\ &\le n^\alpha \frac{|\tilde{C}_n - C|^2}{C^*}(y^{1-t}, y^t)\Big(1 \vee \sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})>n^{-\alpha}} \frac{C(\mathbf{x})}{\tilde{C}_n(\mathbf{x})}\Big)\end{aligned}$$

where $|C^*(y,t) - C(y^{1-t}, y^t)| \leq |\tilde{C}_n(y^{1-t}, y^t) - C(y^{1-t}, y^t)|$. By arguments similar to those used in the proof of Theorem 3.7, it is now easy to see that

$$\sqrt{n} \sup_t |T_1(t,n)| \leq \sup_{\mathbf{x}\in[0,1]^2} n^{\alpha+1/2} |\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})|^2 \times \left(1 \vee \sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})>n^{-\alpha}} \left|\frac{C}{\tilde{C}_n}(\mathbf{x})\right|\right)^2 \times K$$

converges to 0 in outer probability, where $K := \int_0^1 \bar{w}(y)\, y^{-1}\, dy < \infty$ denotes a finite constant (see condition (3.17)). Now set $\beta := (\lambda - 1)/2 > 0$. From the estimate

$$C^*(y,t) \geq y^{1+\beta} \left(1 \wedge \frac{\tilde{C}_n}{C^{1+\beta}}(y^{1-t}, y^t)\right) = y^{-\beta} y^{\lambda} \left(1 \wedge \frac{\tilde{C}_n}{C^{1+\beta}}(y^{1-t}, y^t)\right)$$

we obtain by similar arguments as in the proof of the negligibility of $|B_i^{(2)}(t)|$ in the proof of Theorem 3.7 (note that on $I_{B_1^{(2)}(t)}$ we have $y \leq C(y^{1-t}, y^t) \leq n^{-\alpha}$)

$$\begin{aligned}\sup_{t\in[0,1]} |T_2(t,n)| \leq \log(n) n^{-\beta\alpha} \sup_{\mathbf{x}\in[0,1]^2} \sqrt{n} |\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})| \times \\ \left(1 \vee \sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})\leq n^{-\alpha}} \left|\frac{C^{1+\beta}}{\tilde{C}_n}(\mathbf{x})\right|\right) \times \tilde{K}\end{aligned}$$

where $\tilde{K} := \gamma \int_0^1 (1 - \log y) \frac{h^*(y)}{(\log y)^2} y^{-\lambda}\, dy$ denotes a finite constant (see conditions (3.17) and (3.19)) and we used the estimate

$$\begin{aligned}\left|\log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right|^2 &\leq (\gamma \log n - \log y) \left|\log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right| \\ &\leq \gamma \log(n)(1 - \log y) \left|\log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right|,\end{aligned}$$

which holds for sufficiently large n. Finally, we observe that

$$\sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})\leq n^{-\alpha}} \left|\frac{C^{1+\beta}}{\tilde{C}_n}(\mathbf{x})\right| \leq \sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})\leq n^{-\alpha}} \left|n^{\gamma} C^{1+\beta}(\mathbf{x})\right| \leq n^{\gamma-(1+\beta)\alpha} = o(1).$$

Now the proof is complete. □

Proof of Theorem 3.18 The conditions on the weight function imply that all integrals in the definition of Z_0 are proper and therefore the mapping $(\mathbb{G}_C, C) \mapsto Z_0(\mathbb{G}_C, C)$ is continuous. Hence, the result follows by Theorem 2.3 and the continuous mapping theorem for the bootstrap, see e.g. Theorem 10.8 in [Kosorok, 2008]. □

Chapter 4

Multiplier Bootstrap approximations for tail copulas

4.1 Introduction

Let $\mathbf{X}$ be a 2-dimensional random vector distributed according to some continuous bivariate cumulative distribution function F with associated copula $C = F(F_1^-, F_2^-)$, where $F_p(x_p) = \mathbb{P}(X_p \leq x_p)$, $p = 1, 2$, denote the marginal cumulative distribution functions of $\mathbf{X}$. Recall Definition 1.6: Provided the limits

$$\begin{aligned} \Lambda_L(\mathbf{x}) &= \lim_{t \to \infty} t\, C(x_1/t, x_2/t) \\ \Lambda_U(\mathbf{x}) &= \lim_{t \to \infty} t\, \bar{C}(x_1/t, x_2/t), \end{aligned}$$

exist, the functions Λ_L and Λ_U are referred to as the lower (resp. upper) tail copula of $\mathbf{X}$. Here $\mathbf{x} = (x_1, x_2)^T \in \bar{\mathbb{R}}_+^2 := [0, \infty]^2 \setminus \{(\infty, \infty)\}$ and $\bar{C}(\mathbf{u}) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2)$ denotes the survival copula of $\mathbf{X}$.

Since its introduction various parametric and nonparametric estimates of the tail copulas and of the stable tail dependence function have been proposed in the literature (regarding the relationship between tail copulas and the stable tail dependence function see Section 1.2). Several authors assume that the dependence function belongs to some parametric family. [Coles and Tawn, 1994; Tiago de Oliveira, 1980; Einmahl et al., 1993] imposed restrictions on the marginal distributions to estimate multivariate extreme value distributions. Nonparametric estimates of the stable tail dependence function have been investigated by [Huang, 1992; Qi, 1997; Drees and Huang, 1998], while corresponding estimates for tail copulas have been discussed by [Schmidt and Stadtmüller, 2006]. More recent work on inference on the stable tail dependence function can be found in [Einmahl et al., 2008] and [Einmahl et al., 2006], who investigated moment estimators of tail dependence and weighted approximations of tail copula processes, respectively.
The present chapter of this thesis has two main purposes. First we clarify some curiosities in the literature on tail copula estimation, which stem from the fact that most authors

assume the existence of continuous partial derivatives of the tail copula (see e.g. [Huang, 1992; Drees and Huang, 1998; Schmidt and Stadtmüller, 2006; de Haan and Ferreira, 2006; Peng and Qi, 2008; de Haan et al., 2008] among others). However, the tail copula corresponding to asymptotic independence is the only tail copula with this property, because the partial derivatives of a (lower or upper) tail copula satisfy

$$\partial_1 \Lambda(0,x) = \begin{cases} \lim_{t\to\infty} \Lambda(1,t) & \text{if } x > 0 \\ 0 & \text{if } x = 0. \end{cases} \tag{4.1}$$

As a consequence we provide a result regarding the weak convergence of the empirical tail copula process (and thus also of the empirical stable tail dependence function) under weak smoothness assumptions (see Theorem 4.2 in the following section). The smoothness conditions are nonrestrictive in the sense, that in the case where they are not satisfied, the candidate limiting process does not have continuous trajectories.

The second objective of this chapter is devoted to the approximation of the distribution of estimators for the tail copulas by new bootstrap methods. In contrast to the problem of estimation of the stable dependence function and tail copulas, this problem has found much less attention in the literature. Recently, [Peng and Qi, 2008] considered the tail empirical distribution function and showed the consistency of the bootstrap based on resampling (again under the assumption of continuous partial derivatives). These results were used to construct confidence bands for the tail dependence function. While these authors considered the naive bootstrap, the present chapter of this thesis is devoted to multiplier bootstrap procedures for tail copula estimation, see also Chapter 2. Note that the parametric bootstrap, which is commonly applied in goodness-of-fit testing problems (see [de Haan et al., 2008]), has very high computational costs, because it heavily relies on random number generation and estimation (see also [Kojadinovic and Yan, 2010] and [Kojadinovic et al., 2010] for a more detailed discussion of the computational efficiency of the multiplier bootstrap). Moreover, it was pointed out in Chapter 2 in the context of nonparametric copula estimation that multiplier bootstrap procedures lead to more reliable approximations than the bootstrap based on resampling.

In Section 4.2 we briefly review the nonparametric estimates of the tail copula and discuss their main properties. In particular we establish weak convergence of the empirical tail copula process under nonrestrictive smoothness assumptions, which are satisfied for many commonly used models. In Section 4.3 we introduce the multiplier bootstrap for the empirical tail copula and prove its consistency. The procedure itself as well as the method of proof are similar in kind to the multiplier bootstrap procedures for the empirical copula investigated in Section 2 of this thesis. More precisely, we consider a *partial derivatives multiplier bootstrap* which utilizes the structure of the limiting field of the empirical tail copula process (see also Theorem 2.3). As a consequence, this approach requires the estimation of the partial derivatives of the tail copula. Secondly, a *direct multiplier bootstrap* is proposed which uses multipliers in the two-dimensional empirical distribution function and in the estimates of the marginal distributions. This method is comparable to the one investigated in Theorem 2.4. Finally, in Section 4.4 we discuss several statistical applications of the multiplier bootstrap. In particular, we investigate the problem of testing for equality between two tail copulas and we discuss the bootstrap approximations in the context of testing parametric assumptions for the tail copula. Finally, the proofs and some of the technical details are deferred to Section 4.5.

4.2 Empirical tail copulas

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ denote an i.i.d. sample of random variables distributed according to $F = C(F_1, F_2)$ and denote the empirical distribution functions of F, F_1 and F_2 by F_n, F_{n1} and F_{n2}, respectively. Following [Schmidt and Stadtmüller, 2006] we consider the estimators

$$\hat{\Lambda}_L(\mathbf{x}) = \frac{n}{k} C_n\left(\frac{kx_1}{n}, \frac{kx_2}{n}\right), \tag{4.2}$$

$$\hat{\Lambda}_U(\mathbf{x}) = \frac{n}{k} \bar{C}_n\left(\frac{kx_1}{n}, \frac{kx_2}{n}\right) \tag{4.3}$$

for the lower and upper tail copula, respectively, where $k \to \infty$ such that $k = o(n)$, and C_n and $\bar{C}_n$ denote the empirical copula and the empirical survival copula, respectively. The latter is defined as

$$\bar{C}_n(\mathbf{u}) = \bar{F}_n(\bar{F}_{n1}^{-}(u_1), \bar{F}_{n2}^{-}(u_2)),$$

where $\bar{F}_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{\mathbf{X}_i > \mathbf{x}\}$ and $\bar{F}_{np}(x_p) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{X_{ip} > x_p\}$, $p = 1, 2$. It is easy to see that the estimators in (4.2) and (4.3) are asymptotically equivalent to the estimates

$$\hat{\Lambda}_L(\mathbf{x}) \approx \frac{1}{k}\sum_{i=1}^n \mathbb{I}\{R(X_{i1}) \leq kx_1, R(X_{i2}) \leq kx_2\}, \tag{4.4}$$

$$\hat{\Lambda}_U(\mathbf{x}) \approx \frac{1}{k}\sum_{i=1}^n \mathbb{I}\{R(X_{i1}) > n - kx_1, R(X_{i2}) > n - kx_2\}, \tag{4.5}$$

where $R(X_{ip}) = nF_{np}(X_{ip})$ denotes the rank of X_{ip} among $X_{1p}, \ldots X_{np}$ ($p = 1, 2$). Therefore we introduce analogs of (4.4) and (4.5) where the marginals F_1 and F_2 are assumed to be known, that is

$$\tilde{\Lambda}_L(\mathbf{x}) = \frac{1}{k}\sum_{i=1}^n \mathbb{I}\{F_1(X_{i1}) \leq \frac{kx_1}{n}, F_2(X_{i2}) \leq \frac{kx_2}{n}\}, \tag{4.6}$$

$$\tilde{\Lambda}_U(\mathbf{x}) = \frac{1}{k}\sum_{i=1}^n \mathbb{I}\{F_1(X_{i1}) > 1 - \frac{kx_1}{n}, F_2(X_{i2}) > 1 - \frac{kx_2}{n}\}. \tag{4.7}$$

For the sake of brevity we restrict our investigations to the case of lower tail copulas and we assume that this function is non-zero in a single point $\mathbf{x} \in \mathbb{R}_+^2$ (and as a consequence non-zero everywhere on $\mathbb{R}_+^2$, see Proposition 1.7.

Let $\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$ denote the space of all functions $f : \bar{\mathbb{R}}_+^2 \to \mathbb{R}$, which are locally uniformly bounded on every compact subset of $\bar{\mathbb{R}}_+^2$, with metric

$$d(f_1, f_2) = \sum_{i=1}^\infty 2^{-i}(||f_1 - f_2||_{T_i} \wedge 1),$$

where the sets T_i are defined recursively by $T_{3i} = T_{3i-1} \cup [0,i]^2, T_{3i-1} = T_{3i-2} \cup ([0,i] \times \{\infty\}), T_{3i-2} = T_{3(i-1)} \cup (\{\infty\} \times [0,i])$ and $T_0 = \emptyset$. Note that with this metric the set $\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$ is a complete metric space and that a sequence f_n in $\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$ converges with respect to d if and only if it converges uniformly on every T_i, see [van der Vaart and Wellner, 1996].
[Schmidt and Stadtmüller, 2006] assumed that the lower tail copula Λ_L satisfies the second-order condition

$$\lim_{t\to\infty} \frac{\Lambda_L(\mathbf{x}) - tC(x_1/t, x_2/t)}{A(t)} = g(\mathbf{x}) \tag{4.8}$$

locally uniformly for $\mathbf{x} = (x_1, x_2)^T \in \bar{\mathbb{R}}_+^2$, where g is a non-constant function and the function $A : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies $\lim_{t\to\infty} A(t) = 0$. Under this and the additional assumptions $\Lambda_L \neq 0$, $\sqrt{k}A(n/k) \to 0$, $k = k(n) \to \infty$, $k = o(n)$, they showed that the lower tail copula process with known marginals defined in (4.6) converges weakly in $\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$, that is

$$\sqrt{k}\left(\tilde{\Lambda}_L(\mathbf{x}) - \Lambda_L(\mathbf{x})\right) \rightsquigarrow \mathbb{G}_{\tilde{\Lambda}_L}(\mathbf{x}), \tag{4.9}$$

where $\mathbb{G}_{\tilde{\Lambda}_L}$ is a centered Gaussian field with covariance structure given by

$$\mathbb{E}\, \mathbb{G}_{\tilde{\Lambda}_L}(\mathbf{x})\mathbb{G}_{\tilde{\Lambda}_L}(\mathbf{y}) = \Lambda_L(x_1 \wedge y_1, x_2 \wedge y_2). \tag{4.10}$$

For the empirical tail copula $\hat{\Lambda}_L(\mathbf{x})$ they established the weak convergence

$$\alpha_n(\mathbf{x}) = \sqrt{k}\left(\hat{\Lambda}_L(\mathbf{x}) - \Lambda_L(\mathbf{x})\right) \rightsquigarrow \mathbb{G}_{\hat{\Lambda}_L}(\mathbf{x}) \tag{4.11}$$

in $\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$, provided that the tail copula has continuous partial derivatives. Here the limiting process $\mathbb{G}_{\hat{\Lambda}_L}$ has the representation

$$\mathbb{G}_{\hat{\Lambda}_L}(\mathbf{x}) = \mathbb{G}_{\tilde{\Lambda}_L}(\mathbf{x}) - \partial_1\Lambda_L(\mathbf{x})\mathbb{G}_{\tilde{\Lambda}_L}(x_1, \infty) - \partial_2\Lambda_L(\mathbf{x})\mathbb{G}_{\tilde{\Lambda}_L}(\infty, x_2). \tag{4.12}$$

The assumption of continuous partial derivatives is made in the whole literature on estimation of stable tail dependence functions and tail copulas. However, as demonstrated in (4.1) there does not exist any tail copula $\Lambda_L \neq 0$ with continuous partial derivatives at the origin $(0,0)^T$. With our first result we will fill this gap and prove weak convergence of the empirical tail copula process under weaker smoothness assumptions. For this purpose we will use a similar approach as in [Schmidt and Stadtmüller, 2006] since this turns out to be also useful for a proof of consistency of the multiplier bootstrap. First we consider the case of known marginals. Due to the second order condition (4.8) the proof of (4.9) can be given by showing weak convergence of the centered statistic

$$\tilde{\alpha}_n(\mathbf{x}) := \sqrt{k}\left(\tilde{\Lambda}_L(\mathbf{x}) - \frac{n}{k}C(x_1k/n, x_2k/n)\right). \tag{4.13}$$

Lemma 4.1
If $\Lambda_L \neq 0$ and the second order condition (4.8) *holds with* $\sqrt{k}A(n/k) \to 0$*, where* $k = k(n) \to \infty$ *and* $k = o(n)$*, then we have, as n tends to infinity*

$$\tilde{\alpha}_n(\mathbf{x}) = \sqrt{k}\left(\tilde{\Lambda}_L(\mathbf{x}) - \frac{n}{k}C(x_1k/n, x_2k/n)\right) \rightsquigarrow \mathbb{G}_{\tilde{\Lambda}_L}(\mathbf{x}) \tag{4.14}$$

in $\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$*, where* $\mathbb{G}_{\tilde{\Lambda}_L}$ *is a tight centered Gaussian field concentrated on* $\mathcal{C}_\rho(\bar{\mathbb{R}}^2_+)$ *with covariance structure given in (4.10), where* ρ *is a pseudometric on the space* $\bar{\mathbb{R}}^2_+$ *defined by*

$$\rho(\mathbf{x},\mathbf{y}) = \mathbb{E}\left[(\mathbb{G}_{\tilde{\Lambda}_L}(\mathbf{x}) - \mathbb{G}_{\tilde{\Lambda}_L}(\mathbf{y}))^2\right]^{1/2} = (\Lambda_L(\mathbf{x}) - 2\Lambda_L(\mathbf{x}\wedge\mathbf{y}) + \Lambda_L(\mathbf{x}))^{1/2},$$

$\mathbf{x} = (x_1, x_2)^T$, $\mathbf{y} = (y_1, y_2)^T$, $\mathbf{x}\wedge\mathbf{y} = (x_1\wedge y_1, x_2\wedge y_2)^T$ *and* $\mathcal{C}_\rho(\bar{\mathbb{R}}^2_+) \subset \mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$ *denotes the subset of all functions that are uniformly* ρ*-continuous on every* T_i*.*

This assertion is proved in [Schmidt and Stadtmüller, 2006] by showing convergence of the finite dimensional distributions and tightness. The proof of consistency of the bootstrap procedures proposed in the following section follows in part by arguments from an alternative proof of (4.14) based on Donsker classes which will be accomplished in the appendix.
For a proof of a corresponding result for the empirical tail copula process with estimated marginals in (4.11) we will use the functional delta method in (4.9) with some suitable functional.

Theorem 4.2
Let $\Lambda_L \neq 0$ *be a lower tail copula whose partial derivatives satisfy the following first order properties*

$$\partial_p\,\Lambda_L \text{ exists on } \{\mathbf{x}\in\bar{\mathbb{R}}^2_+ \mid x_p < \infty\} \text{ and is continuous on } \{\mathbf{x}\in\bar{\mathbb{R}}^2_+ \mid 0 < x_p < \infty\} \tag{4.15}$$

for $p = 1, 2$*. If additionally the assumptions of Lemma 4.1 are satisfied then we have*

$$\alpha_n(\mathbf{x}) = \sqrt{k}\left(\hat{\Lambda}_L(\mathbf{x}) - \Lambda_L(\mathbf{x})\right) \rightsquigarrow \mathbb{G}_{\hat{\Lambda}_L}(\mathbf{x})$$

in $\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$*, where the process* $\mathbb{G}_{\hat{\Lambda}_L}$ *is defined in* (4.12) *and* $\partial_p\,\Lambda_L$*,* $p = 1, 2$ *is defined as 0 on the set* $\{\mathbf{x}\in\bar{\mathbb{R}}^2_+ \mid x_p = \infty\}$*.*

Theorem 4.2 has been proved by [Schmidt and Stadtmüller, 2006] under the additional assumption that the tail copula has continuous partial derivatives. As pointed out in the previous paragraphs there does not exist any tail copula $\Lambda_L \neq 0$ with this property.

4.3 Multiplier bootstrap approximation

4.3.1 Asymptotic theory

In this section we will construct multiplier bootstrap approximations of the Gaussian limit distributions $\mathbb{G}_{\tilde{\Lambda}_L}$ and $\mathbb{G}_{\hat{\Lambda}_L}$ specified in (4.9) and (4.11), respectively. We proceed as in Chapter 2: let ξ_i be i.i.d. positive random variables, independent of the $\mathbf{X}_i$, with mean μ in $(0, \infty)$ and finite variance τ^2, which additionally satisfy $||\xi||_{2,1} := \int_0^\infty \sqrt{P(|\xi| > x)}\, dx < \infty$. We will first deal with the case of known marginals and define a multiplier bootstrap analogue of (4.6) by

$$\tilde{\Lambda}_L^{\xi}(\mathbf{x}) = \frac{1}{k} \sum_{i=1}^{n} \frac{\xi_i}{\bar{\xi}_n} \mathbb{I}\{F_1(X_{i1}) \leq \frac{kx_1}{n}, F_2(X_{i2}) \leq \frac{kx_2}{n}\} \tag{4.16}$$

where $\bar{\xi}_n = n^{-1} \sum_{i=1}^n \xi_i$ denotes the mean of $\xi_1, \ldots, \xi_n$. We have

$$\tilde{\alpha}_n^m(\mathbf{x}) = \frac{\mu}{\tau} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\xi_i}{\bar{\xi}_n} - 1 \right) f_{n,\mathbf{x}}(U_i) = \frac{\mu}{\tau} \sqrt{k} \big(\tilde{\Lambda}_L^{\xi} - \tilde{\Lambda}_L \big), \tag{4.17}$$

where the function $f_{n,\mathbf{x}}(U_i)$ is defined by

$$f_{n,\mathbf{x}}(\mathbf{U}_i) = \sqrt{\frac{n}{k}}\, \mathbb{I}\left\{ U_{i1} \leq kx_1/n, U_{i2} \leq kx_2/n \right\}, \tag{4.18}$$

and

$$\mathbf{U}_i = (U_{i1}, U_{i2}); \qquad U_{ip} = F_p(X_{ip}) \quad \text{for } p = 1, 2.$$

The following result shows that the process (4.17) provides a valid bootstrap approximation of the process (4.13).

Theorem 4.3
If $\Lambda_L \neq 0$ and the second order condition (4.8) *holds with $\sqrt{k}A(n/k) \to 0$, $k = k(n) \to \infty$ and $k = o(n)$ we have, as n tends to infinity,*

$$\tilde{\alpha}_n^m = \frac{\mu}{\tau} \sqrt{k} \big(\tilde{\Lambda}_L^{\xi} - \tilde{\Lambda}_L \big) \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_{\tilde{\Lambda}_L}$$

in the metric space $\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$.

Since Theorem 4.3 states that we have weak convergence of $\tilde{\alpha}_n^m$ to $\mathbb{G}_{\tilde{\Lambda}_L}$ conditional on the data U_i, it provides a bootstrap approximation of the empirical tail copula process in the case where the marginal distributions are known. To be precise, consider $B \in \mathbb{N}$ independent replications of the random variables $\xi_1, \ldots, \xi_n$ and denote them by $\xi_{1,b}, \ldots, \xi_{n,b}$. Compute the statistics $\tilde{\alpha}_{n,b}^m = \tilde{\alpha}_n^m(\xi_{1,b}, \ldots, \xi_{n,b})$ $(b = 1, \ldots, B)$ and use the empirical distribution of $\tilde{\alpha}_{n,1}^m, \ldots, \tilde{\alpha}_{n,B}^m$ as an approximation for the limiting distribution of $G_{\tilde{\Lambda}_L}$.
Because in most cases of practical interest there will be no information about the marginals one cannot use Theorem 4.3 in many statistical applications. We will now develop two consistent bootstrap approximation for the limiting distribution of the process (4.11) which do not require knowledge of the marginals. Intuitively, it is natural to replace the unknown marginal distributions in (4.16) by its empirical counterparts, that is

$$\hat{\Lambda}_L^{\xi,\cdot}(\mathbf{x}) = \frac{1}{k} \sum_{i=1}^{n} \frac{\xi_i}{\bar{\xi}_n} \mathbb{I}\{X_{i1} \leq F_{n1}^{-1}(kx_1/n), X_{i2} \leq F_{n2}^{-1}(kx_2/n)\}$$

which yields the process

$$\beta_n(\mathbf{x}) = \frac{\mu}{\tau}\sqrt{k}\left(\hat{\Lambda}_L^{\xi,\cdot} - \hat{\Lambda}_L\right) = \frac{\mu}{\tau}\frac{1}{\sqrt{k}}\sum_{i=1}^{n}\left(\frac{\xi_i}{\bar{\xi}_n} - 1\right)\mathbb{I}\{X_{i1} \le F_{n1}^{-1}(kx_1/n), X_{i2} \le F_{n2}^{-1}(kx_2/n)\}.$$

Similar to the multiplier approximations considered in Chapter 2 this intuitive approach does not yield an approximation for the distribution of the process $\mathbb{G}_{\hat{\Lambda}_L}$, but of $\mathbb{G}_{\tilde{\Lambda}_L}$.

Theorem 4.4
Suppose that the assumptions of Theorem 4.2 hold. Then we have, as n tends to infinity

$$\beta_n = \frac{\mu}{\tau}\sqrt{k}(\hat{\Lambda}_L^{\xi,\cdot} - \hat{\Lambda}_L) \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_{\tilde{\Lambda}_L}$$

in the metric space $\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$.

Proceeding as in Chapter 2 we can make use of Theorem 4.4 together with the representation (4.12) for $\mathbb{G}_{\hat{\Lambda}_L}$ to get an approximation of $\mathbb{G}_{\hat{\Lambda}_L}$. We begin by estimating the derivatives of the tail copula by

$$\widehat{\partial_p\Lambda_L}(\mathbf{x}) := \begin{cases} \frac{\hat{\Lambda}_L(\mathbf{x}+h\mathbf{e}_p)-\hat{\Lambda}_L(\mathbf{x}-h\mathbf{e}_p)}{2h} & ,\infty > x_p \ge h \\ \widehat{\partial_p\Lambda_L}(\mathbf{x}+(h-x_p)\mathbf{e}_p) = \frac{\hat{\Lambda}_L(\mathbf{x}+2h\mathbf{e}_p)-\hat{\Lambda}_L(\mathbf{x}-x_p\mathbf{e}_p)}{2h} & ,x_p < h \\ 0 & ,x_p = \infty \end{cases}$$

where $\mathbf{e}_1 = (1,0), \mathbf{e}_2 = (0,1)$ and $h \sim k^{-1/2}$ tends to 0 with increasing sample size. We will show in the Appendix (see the proof of the following Theorem in Section 4.5) that these estimates are consistent, and consequently we define the process

$$\alpha_n^{pdm}(\mathbf{x}) = \beta_n(\mathbf{x}) - \widehat{\partial_1\Lambda_L}(\mathbf{x})\beta_n(x_1,\infty) - \widehat{\partial_2\Lambda_L}(\mathbf{x})\beta_n(\infty,x_2). \tag{4.19}$$

Note that α_n^{pdm} only depends on the data and the multipliers $\xi_1,\ldots,\xi_n$. As a consequence, a bootstrap sample can easily be generated as described in the previous paragraph and as in Chapter 2 we call this method *partial derivatives multiplier bootstrap (pdm-bootstrap)* in the following discussion. Our next result shows that the *pdm*-bootstrap provides a valid approximation for the distribution of the process $\mathbb{G}_{\hat{\Lambda}_L}$.

Theorem 4.5
Under the assumptions of Theorem 4.2 we have

$$\alpha_n^{pdm} \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_{\hat{\Lambda}_L}$$

in the metric space $\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$.

Similarly as in Chapter 2 it turns out that there is an alternative valid multiplier bootstrap procedure in the case of unknown marginal distributions, which is attractive because it avoids the problem of estimating the partial derivatives of the lower tail copula. This method not only introduces multiplier random variables in the two-dimensional distribution function but also in the inner estimators of the marginals. To be precise define

$$\begin{aligned}
F_n^{\xi}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i}{\bar{\xi}_n}\mathbb{I}\{X_{i1}\le x_1, X_{i2}\le x_2\}\\
F_{nj}^{\xi}(x_j) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i}{\bar{\xi}_n}\mathbb{I}\{X_{ij}\le x_j\},\ \ j=1,2\\
C_n^{\xi,\xi}(\mathbf{u}) &= F_n^{\xi}(F_{n1}^{\xi-}(u_1), F_{n2}^{\xi-}(u_2)).
\end{aligned}$$

and consider the process

$$\begin{aligned}
\hat{\Lambda}_L^{\xi,\xi}(\mathbf{x}) := \frac{n}{k}C_n^{\xi,\xi}\left(\frac{k}{n}\mathbf{x}\right) &= \frac{1}{k}\sum_{i=1}^{n}\frac{\xi_i}{\bar{\xi}_n}\mathbb{I}\{X_{i1}\le F_{n1}^{\xi-}(kx_1/n), X_{i2}\le F_{n2}^{\xi-}(kx_2/n)\} \qquad (4.20)\\
&\approx \frac{1}{k}\sum_{i=1}^{n}\frac{\xi_i}{\bar{\xi}_n}\mathbb{I}\{F_{n1}^{\xi}(X_{i1})\le kx_1/n, F_{n2}^{\xi}(X_{i2})\le kx_2/n\}
\end{aligned}$$

As before we will call this bootstrap method the *direct multiplier bootstrap (dm-bootstrap)*.

Theorem 4.6
Under the assumptions of Theorem 4.2 we have

$$\alpha_n^{dm}(\mathbf{x}) = \frac{\mu}{\tau}\sqrt{k}\left(\hat{\Lambda}_L^{\xi,\xi}(\mathbf{x}) - \hat{\Lambda}_L(\mathbf{x})\right) \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_{\hat{\Lambda}_L} \quad in\ \mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)). \qquad (4.21)$$

4.3.2 Finite sample results

In this section we will present a small comparison of the finite sample properties of the two bootstrap approximations given in this section. For the sake of brevity we only consider data generated from the Clayton copula with a coefficient of lower tail dependence $\lambda_L = 0.25$. The Clayton copula,

$$C_{\text{Clayton}}(\mathbf{u};\theta) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta},\ \ \theta > 0, \qquad (4.22)$$

is a widely used copula family for the modeling of negative tail dependent data. Its lower tail copula is given by

$$\Lambda_L(\mathbf{x}) = \left(x_1^{-\theta} + x_2^{-\theta}\right)^{-1/\theta}.$$

In Tables 4.1 - 4.3 we investigate the accuracy of the bootstrap approximation of the covariances of the limiting variable $\mathbb{G}_{\hat{\Lambda}_L}$, see also Section 2.4 for an analogous investigation for the approximations of the empirical copula process. More precisely, we chose

True	$\frac{\pi}{8}$	$2\frac{\pi}{8}$	$3\frac{\pi}{8}$	α_n	$\frac{\pi}{8}$	$2\frac{\pi}{8}$	$3\frac{\pi}{8}$
$\frac{\pi}{8}$	0.0874	0.0754	0.0516	$\frac{\pi}{8}$	0.0889	0.0737	0.0476
$2\frac{\pi}{8}$		0.1160	0.0754	$2\frac{\pi}{8}$		0.1218	0.0741
$3\frac{\pi}{8}$			0.0874	$3\frac{\pi}{8}$			0.0892

Table 4.1: *Left part: True covariances of $\mathbb{G}_{\hat{\Lambda}_L}$ for the Clayton Copula with $\lambda_L = 0.25$. Right part: sample covariances of the empirical tail copula process α_n with sample size $n = 1000$ and parameter $k = 50$.*

α_n^{pdm}	$\frac{\pi}{8}$	$2\frac{\pi}{8}$	$3\frac{\pi}{8}$	α_n^{dm}	$\frac{\pi}{8}$	$2\frac{\pi}{8}$	$3\frac{\pi}{8}$	α_n^{res}	$\frac{\pi}{8}$	$2\frac{\pi}{8}$	$3\frac{\pi}{8}$
$\frac{\pi}{8}$	0.094	0.072	0.046	$\frac{\pi}{8}$	0.100	0.071	0.045	$\frac{\pi}{8}$	0.100	0.070	0.043
$2\frac{\pi}{8}$		0.130	0.072	$2\frac{\pi}{8}$		0.136	0.707	$2\frac{\pi}{8}$		0.136	0.070
$3\frac{\pi}{8}$			0.094	$3\frac{\pi}{8}$			0.099	$3\frac{\pi}{8}$			0.099

Table 4.2: *Averaged sample covariances of the Bootstrap approximations α_n^{pdm}, α_n^{dm} and α_n^{res} of $\mathbb{G}_{\hat{\Lambda}_L}$ under the conditions of Table 4.1.*

α_n^{pdm}	$\frac{\pi}{8}$	$2\frac{\pi}{8}$	$3\frac{\pi}{8}$	α_n^{dm}	$\frac{\pi}{8}$	$2\frac{\pi}{8}$	$3\frac{\pi}{8}$	α_n^{res}	$\frac{\pi}{8}$	$2\frac{\pi}{8}$	$3\frac{\pi}{8}$
$\frac{\pi}{8}$	3.676	4.688	3.650	$\frac{\pi}{8}$	3.869	3.496	2.724	$\frac{\pi}{8}$	4.217	3.851	3.219
$2\frac{\pi}{8}$		8.110	4.877	$2\frac{\pi}{8}$		8.892	3.259	$2\frac{\pi}{8}$		8.731	3.640
$3\frac{\pi}{8}$			3.706	$3\frac{\pi}{8}$			3.777	$3\frac{\pi}{8}$			3.900

Table 4.3: *Mean squared error $\times 10^4$ of the different estimates for the covariances in Table 4.2.*

three points on the unit circle $\{e^{i\varphi}, \varphi = k\pi/8 \text{ with } k = 1, 2, 3\}$ and show in the first four columns of Table 4.1 the true covariances of the limiting process $\mathbb{G}_{\hat{\Lambda}_L}$. The remaining columns show the simulated covariances of the process α_n on the basis of $5 \cdot 10^5$ simulation runs, where the sample size is $n = 1000$ and the parameter k is chosen as 50. This table is the benchmark for the bootstrap approximations of the covariances stated in Table 4.2. For the sake of completeness we also investigated the resampling bootstrap considered in [Peng and Qi, 2008] (which is hereafter denoted by α_n^{res}). The covariances are based on the average of 1000 simulation runs, where in each run the covariance is estimated on the basis of $B = 500$ bootstrap replications. In Table 4.3 we present the corresponding mean squared error.

As one can see all bootstrap procedures yield approximations of quite comparable magnitude. Considering only the bias in Table 4.2 the *pdm*-bootstrap has slight advantages in all cases, while there are basically no differences between the *dm*- and the resampling bootstrap. A comparison of the mean squared error in Table 4.3 shows that the *pdm*-bootstrap has the best performance on the diagonal. On the other hand, it yields a less accurate approximation in case of approximating off-diagonal covariances. In this case, the *dm*-bootstrap yields the best results.

4.4 Statistical applications

In this section we investigate several statistical applications of the multiplier bootstrap. In particular we discuss the problem of comparing lower tail copulas from different samples, the problem of constructing confidence intervals and the problem of testing for a parametric form of the lower tail copula.

4.4.1 Testing for equality between two tail copulas

Let $\mathbf{X}_1, \ldots, \mathbf{X}_{n_1}$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_{n_2}$ denote two independent samples of i.i.d. random variables (we will relax the assumption of independence between the samples later on) with cumulative distribution function $F = C(F_1, F_2)$ and $H = D(H_1, H_2)$, respectively. We assume that the marginal distributions F_1, F_2 and H_1, H_2 of F and H are continuous and that for both distributions the corresponding lower tail copulas, say $\Lambda_{L,X}$ and $\Lambda_{L,Y}$, exist and do not vanish. We are interested in a test for the hypothesis

$$\mathcal{H}_0 : \Lambda_{L,X} \equiv \Lambda_{L,Y} \quad \text{vs.} \quad \mathcal{H}_1 : \Lambda_{L,X} \neq \Lambda_{L,Y}. \tag{4.23}$$

Due to the homogeneity of tail copulas we have $\Lambda_L(t\mathbf{x}) = t\Lambda_L(\mathbf{x}) \quad \forall t > 0, \mathbf{x} \in \mathbb{R}_+^2$, and the hypotheses are equivalent to

$$\mathcal{H}_0 : \varrho(\Lambda_{L,X}, \Lambda_{L,Y}) = 0 \quad \text{vs} \quad \mathcal{H}_1 : \varrho(\Lambda_{L,X}, \Lambda_{L,Y}) > 0,$$

where the distance ϱ is defined by

$$\begin{aligned} \varrho(\Lambda_{L,X}, \Lambda_{L,Y}) &:= \int_0^{\pi/2} \left(\Lambda_{L,X}(\cos\varphi, \sin\varphi) - \Lambda_{L,Y}(\cos\varphi, \sin\varphi)\right)^2 d\varphi \\ &= \int_0^{\pi/2} \left(\Lambda^{\angle}_{L,X}(\varphi) - \Lambda^{\angle}_{L,Y}(\varphi)\right)^2 d\varphi \end{aligned} \tag{4.24}$$

and we have used the notation

$$\Lambda^{\angle}_{L,X}(\varphi) = \Lambda_{L,X}(\cos\varphi, \sin\varphi), \ \Lambda^{\angle}_{L,Y}(\varphi) = \Lambda_{L,Y}(\cos\varphi, \sin\varphi).$$

We propose to base the test for the hypothesis (4.23) on the distance between the empirical tail copulas and define

$$\mathcal{S}_n = \frac{k_1 k_2}{k_1 + k_2} \varrho(\hat{\Lambda}_{L,X}, \hat{\Lambda}_{L,Y}) = \frac{k_1 k_2}{k_1 + k_2} \int_0^{\pi/2} (\hat{\Lambda}^{\angle}_{L,X}(\varphi) - \hat{\Lambda}^{\angle}_{L,Y}(\varphi))^2 \, d\varphi,$$

where $\hat{\Lambda}^{\angle}_{L,X}(\varphi) = \hat{\Lambda}_{L,X}(\cos(\varphi), \sin(\varphi)), \hat{\Lambda}^{\angle}_{L,Y} = \hat{\Lambda}_{L,Y}(\cos(\varphi), \sin(\varphi))$ denote the empirical tail copulas $\hat{\Lambda}_{L,X}$ and $\hat{\Lambda}_{L,Y}$ with corresponding parameters k_1 and k_2, which satisfy the (second order) conditions of Lemma 4.1 and

$$k_1/(k_1 + k_2) \to \lambda \in (0, 1).$$

Note that $\mathcal{S}_n$ can easily be computed from the ranks of X_{ip} and Y_{ip} in their respective samples, $X_{1p}, \dots, X_{n_1 p}$ and $Y_{1p}, \dots, Y_{n_2 p}$ that is

$$\begin{aligned}\mathcal{S}_n = \frac{k_1 k_2}{k_1 + k_2} \Big\{ &\frac{1}{k_1^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \big[(\arccos(S_{i1}) \wedge \arccos(S_{j1})) - (\arcsin(S_{i2}) \vee \arcsin(S_{j2})) \big]^+ \\ &- \frac{2}{k_1 k_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \big[(\arccos(S_{i1}) \wedge \arccos(T_{j1})) - (\arcsin(S_{i2}) \vee \arcsin(T_{j2})) \big]^+ \\ &+ \frac{1}{k_2^2} \sum_{i=1}^{n_2} \sum_{j=1}^{n_2} \big[(\arccos(T_{i1}) \wedge \arccos(T_{j1})) - (\arcsin(T_{i2}) \vee \arcsin(T_{j2})) \big]^+ .\end{aligned}$$

Here we used the notation $S_{ip} = \frac{R(X_{ip})}{k_1} \wedge 1$, $T_{ip} = \frac{R(Y_{ip})}{k_1} \wedge 1$ $(p = 1, 2)$ and $[f]^+$ denotes the positive part of the function f.
Under the null hypothesis (4.23) of equality between the tail copulas we have $\mathcal{S}_n = \mathcal{T}_n$ with

$$\mathcal{T}_n = \int_0^{\pi/2} \mathcal{E}_n^2(\cos\varphi, \sin\varphi)\, d\varphi,$$

where

$$\mathcal{E}_n(\mathbf{x}) = \sqrt{\frac{k_2}{k_1 + k_2}} \sqrt{k_1} (\hat{\Lambda}_{L,X}(\mathbf{x}) - \Lambda_{L,X}(\mathbf{x})) - \sqrt{\frac{k_1}{k_1 + k_2}} \sqrt{k_2} (\hat{\Lambda}_{L,Y}(\mathbf{x}) - \Lambda_{L,Y}(\mathbf{x})).$$

Since the two samples X and Y are independent we obtain independently of the hypotheses that

$$\mathcal{E}_n \rightsquigarrow \sqrt{1-\lambda}\, \mathbb{G}_{\hat{\Lambda}_{L,X}} - \sqrt{\lambda}\, \mathbb{G}_{\hat{\Lambda}_{L,Y}} =: \mathcal{E}.$$

in the metric space $\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$, where the two-dimensional centered Gaussian fields $\mathbb{G}_{\hat{\Lambda}_{L,X}}$ and $\mathbb{G}_{\hat{\Lambda}_{L,Y}}$ are defined in (4.12). This yields by the continuous mapping theorem

$$\mathcal{T}_n \rightsquigarrow \int_0^{\pi/2} \mathcal{E}^2(\cos\varphi, \sin\varphi)\, d\varphi =: \mathcal{T}$$

under both the null hypothesis and the alternative. Note that

$$\varrho(\hat{\Lambda}_{L,X}, \hat{\Lambda}_{L,Y}) \xrightarrow{\mathbb{P}} \varrho(\Lambda_{L,X}, \Lambda_{L,Y}),$$

which vanishes if and only if the null hypothesis (4.23) is satisfied. Therefore we can conclude that

$$\mathcal{S}_n \rightsquigarrow_{\mathcal{H}_0} \mathcal{T}, \quad \mathcal{S}_n \xrightarrow{\mathbb{P}}_{\mathcal{H}_1} \infty, \tag{4.25}$$

which shows that a test, which rejects the null hypothesis (4.23) for large values of $\mathcal{T}_n$ is consistent.

In order to determine critical values for the test we approximate the limiting distribution $\mathcal{T}$ by the multiplier bootstrap proposed in Section 4.3. For this purpose we exemplarily consider the *pdm*-bootstrap (the extension to the *dm*-bootstrap defined in (4.21) is straightforward) using the definition in equation (4.19) and denote for any $b \in \{1, \dots, B\}$ by $\xi_{1,b}, \dots, \xi_{n_1,b}, \zeta_{1,b}, \dots, \zeta_{n_2,b}$ independent and identically distributed non-negative random variables with mean μ_1 (resp. μ_2) and variance τ_1^2 (resp. τ_2^2). We compute for each b and both samples the bootstrap statistics as given in (4.19), i.e.

$$\alpha^{pdm}_{X,n_1,b}(\mathbf{x}) = \beta_{X,n_1,b}(\mathbf{x}) - \widehat{\partial_1 \Lambda_{L,X}}(\mathbf{x})\beta_{X,n_1,b}(x_1,\infty) - \widehat{\partial_2 \Lambda_{L,X}}(\mathbf{x})\beta_{X,n_1,b}(\infty,x_2),$$
$$\alpha^{pdm}_{Y,n_2,b}(\mathbf{x}) = \beta_{Y,n_2,b}(\mathbf{x}) - \widehat{\partial_1 \Lambda_{L,Y}}(\mathbf{x})\beta_{Y,n_2,b}(x_1,\infty) - \widehat{\partial_2 \Lambda_{L,Y}}(\mathbf{x})\beta_{Y,n_2,b}(\infty,x_2),$$

where

$$\beta_{X,n_1,b}(\mathbf{x}) = \frac{\mu_1}{\tau_1}\frac{1}{\sqrt{k_1}}\sum_{i=1}^{n_1}\left(\frac{\xi_{i,b}}{\bar{\xi}_{\cdot,b_{n_1}}} - 1\right)\mathbb{I}\{F_{n_1 1}(X_{i1}) \le k_1 x_1/n_1, F_{n_1 2}(X_{i2}) \le k_1 x_2/n_1\},$$
$$\beta_{Y,n_2,b}(\mathbf{x}) = \frac{\mu_2}{\tau_2}\frac{1}{\sqrt{k_2}}\sum_{i=1}^{n_2}\left(\frac{\zeta_{i,b}}{\bar{\zeta}_{\cdot,b_{n_2}}} - 1\right)\mathbb{I}\{H_{n_2 1}(Y_{i1}) \le k_2 x_1/n_2, H_{n_2 2}(Y_{i2}) \le k_2 x_2/n_2\},$$

and $\widehat{\partial_p \Lambda_{L,X}}$ and $\widehat{\partial_p \Lambda_{L,Y}}$ are the corresponding estimates of the partial derivatives ($p = 1, 2$). For all $\mathbf{x} \in \bar{\mathbb{R}}^2_+$ and all $b \in \{1, \dots, B\}$ define

$$\hat{\mathcal{E}}_n^{(pdm,b)}(\mathbf{x}) := \sqrt{\frac{k_2}{k_1+k_2}}\alpha^{pdm}_{X,n_1,b}(\mathbf{x}) - \sqrt{\frac{k_1}{k_1+k_2}}\alpha^{pdm}_{Y,n_2,b}(\mathbf{x}),$$
$$\hat{\mathcal{T}}_n^{(pdm,b)} := \int_0^{\pi/2} \left\{ \hat{\mathcal{E}}_n^{(pdm,b)}(\cos\varphi, \sin\varphi) \right\}^2 d\varphi.$$

By Theorem 4.5 and Theorem 10.8 in [Kosorok, 2008], it follows for every $b \in \{1, \dots, B\}$

$$\hat{\mathcal{T}}_n^{(pdm,b)} \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{T}^{(b)},$$

where $\mathcal{T}^{(b)}$ is an independent copy of $\mathcal{T}$ (note that we consider the processes $\hat{\mathcal{E}}_n^{(\beta,b)}$ and $\hat{\mathcal{E}}_n^{(\gamma,b)}$ in the Banachspace $l^\infty([0,1]^2)$). Similarly, we have $\hat{\mathcal{T}}_n^{(dm,b)} \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{T}^{(b)}$. From (4.25) we therefore obtain a consistent asymptotic level α test for the null hypothesis (4.23) by rejecting $\mathcal{H}_0$ for large values of $\mathcal{S}_n$, that is

$$\mathcal{S}_n > q^m_{1-\alpha} \tag{4.26}$$

where $q^m_{1-\alpha}$ denotes the $(1-\alpha)$ quantile of the empirical distribution function

$$K_n^m(s) = \frac{1}{B}\sum_{b=1}^{B}\mathbb{I}\{\hat{\mathcal{T}}_n^{(m,b)} \le s\}$$

$\lambda_{L,X}$	$\lambda_{L,Y}$	$\alpha = 0.15$	$\alpha = 0.1$	$\alpha = 0.05$
0.25	0.25	0.143	0.098	0.054
0.5	0.5	0.140	0.099	0.047
0.75	0.75	0.117	0.078	0.029
0.25	0.5	0.764	0.706	0.605
0.5	0.75	0.896	0.856	0.783
0.25	0.75	1	1	1

Table 4.4: *Simulated rejection probabilities of the pdm bootstrap test* (4.26) *for the hypothesis* (4.23) .

(here m is either *pdm* or *dm* corresponding to partial derivative or direct multiplier bootstrap).
The discussion up till now holds true for two independent populations $\mathbf{X}_i$ and $\mathbf{Y}_i$. Nevertheless it is easy to check that the methodology of the previous sections also applies if we are faced with paired observations, i.e. $\mathbf{X}_i$ is not independent of $\mathbf{Y}_i$, but $n_1 = n_2 = n$. In that case we have to set $\zeta_{i,b} = \xi_{i,b}$ for all $i = 1, \ldots, n$ and $b = 1, \ldots, B$. To see this, set $\mathbf{Z}_i = (X_{i1}, X_{i2}, Y_{i1}, Y_{i2})^T$ and denote the (empirical) copula of $\mathbf{Z}_i$ by $(\mathcal{C}_n)$ $\mathcal{C}$. Clearly,

$$C(u_1, u_2) = \mathcal{C}(u_1, u_2, 1, 1), \qquad D(v_1, v_2) = \mathcal{C}(1, 1, v_1, v_2),$$
$$C_n(u_1, u_2) = \mathcal{C}_n(u_1, u_2, 1, 1), \qquad D_n(v_1, v_2) = \mathcal{C}_n(1, 1, v_1, v_2).$$

If we set

$$\Lambda_{L,Z}(\mathbf{x}, \mathbf{y}) = \lim_{t\to\infty} t\, \mathcal{C}(\mathbf{x}/t, \mathbf{y}/t),$$
$$\hat{\Lambda}_{L,Z}(\mathbf{x}, \mathbf{y}) = \frac{n}{k} \mathcal{C}_n(\frac{n\mathbf{x}}{k}, \frac{n\mathbf{y}}{k}),$$

we obtain

$$\Lambda_{L,X}(\mathbf{x}) = \Lambda_{L,Z}(\mathbf{x}, \infty, \infty), \;\; \Lambda_{L,Y}(\mathbf{y}) = \Lambda_{L,Z}(\infty, \infty, \mathbf{y}),$$
$$\hat{\Lambda}_{L,X}(\mathbf{x}) = \hat{\Lambda}_{L,Z}(\mathbf{x}, \infty, \infty), \;\; \hat{\Lambda}_{L,Y}(\mathbf{y}) = \hat{\Lambda}_{L,Z}(\infty, \infty, \mathbf{y}).$$

Similar relations for the multiplier approximations are straightforward and the result follows along similar lines as in the previous sections.
For an investigation of the finite sample property we consider two independent samples of i.i.d. distributed random variables according to the Clayton copula, see (4.22), with a coefficient of lower tail dependence λ_L varying in the set $\{0.25, 0.5, 0.75\}$.
In Table 4.4 and 4.5 we show the simulated rejection probabilities of the *pdm* and *dm* bootstrap test defined in (4.26) for various nominal levels on the basis of 1000 simulation runs. The sample size is $n = 1000$, $k = 50$ and $B = 500$ bootstrap replications with $Laplacian(0, 2)$ multipliers have been used.
We observe that the nominal level is well approximated by the *pdm* bootstrap if the coefficient of tail dependence is not too large. For a larger coefficient the test is conservative.

$\lambda_{L,X}$	$\lambda_{L,Y}$	$\alpha = 0.15$	$\alpha = 0.1$	$\alpha = 0.05$
0.25	0.25	0.125	0.091	0.052
0.5	0.5	0.108	0.069	0.036
0.75	0.75	0.068	0.051	0.023
0.25	0.5	0.713	0.643	0.529
0.5	0.75	0.869	0.822	0.713
0.25	0.75	0.999	0.999	0.997

Table 4.5: *Simulated rejection probabilities of the dm bootstrap test* (4.26) *for the hypothesis* (4.23)

On the other hand, the *dm* bootstrap test is slightly more conservative and this effect is increasing with the coefficient of tail dependence. The alternative of different lower tail copulas is detected with reasonable power where both tests yield rather similar results with slight advantages for the *pdm*-bootstrap.

4.4.2 Bootstrap approximation of a minimum distance estimate and a computationally efficient goodness-of-fit test

In this section we are interested in estimating the tail copula of $\mathbf{X}$ under the additional assumption that the tail copula lies in some parametric class, say

$$\mathcal{L} = \{\Lambda_L(\cdot;\theta) \mid \theta \in \Theta\}. \tag{4.27}$$

Recently, estimates for parametric classes of tail copulas and stable tail dependence functions have been investigated by [de Haan et al., 2008] and [Einmahl et al., 2008] who proposed a censored likelihood and a moment based estimator, respectively. In the present section we investigate a further estimate, which is based on the minimum distance method. To be precise let Λ_L denote an arbitrary lower tail copula and $\Lambda_L(\,\cdot\,;\theta)$ an element in the parametric class $\mathcal{L}$ and define

$$\Lambda_L^{\angle}(\varphi) = \Lambda_L(\cos\varphi, \sin\varphi),\ \Lambda_L^{\angle}(\varphi;\theta) = \Lambda_L(\cos\varphi, \sin\varphi;\theta).$$

We consider the parameter corresponding to the best approximation by the distance ϱ defined in (4.24)

$$\theta_B = T(\Lambda_L) = \arg\min_{\theta\in\Theta} \varrho(\Lambda_L, \Lambda_L(\,\cdot\,;\theta)),$$

and call $\hat{\theta}_n^{MD} = T(\hat{\Lambda}_L)$ a minimum distance estimator for θ, where $\hat{\Lambda}_L$ is the empirical lower tail copula defined in (4.2) and $\hat{\Lambda}_L^{\angle} = \hat{\Lambda}_L(\cos\varphi, \sin\varphi)$. Note that θ_B is the "true" parameter if the null hypothesis is satisfied.

Throughout this section let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ denote i.i.d. bivariate random variables with cumulative distribution function $F = C(F_1, F_2)$ and existing lower tail copula Λ_L. Furthermore, we introduce the notations

$$Q(\theta) = \varrho(\Lambda_L, \Lambda_L(\cdot;\theta)), \quad \psi(\theta) = \partial_\theta\, Q(\theta),$$
$$Q_n(\theta) = \varrho(\hat{\Lambda}_L, \Lambda_L(\cdot;\theta)), \quad \psi_n(\theta) = \partial_\theta\, Q_n(\theta),$$

and assume that the following regularity conditions are satisfied.

(B.1) The parameter space Θ has non-empty interior, say Θ^0, and the parameter $\theta_B = T(\Lambda_L) \in \Theta^0$ corresponding to the best approximation of the lower tail copula by the parametric class $\mathcal{L}$ exists and is unique.

(B.2) $\Lambda_L(\cdot;\theta)$ is continuously differentiable with respect to $\theta \in \Theta^0$ with $\delta_\theta(\mathbf{x}) = \partial_\theta\, \Lambda_L(\mathbf{x};\theta)$ and the mapping $\mathbf{x} \mapsto \sup_{\theta\in\Theta} ||\delta_\theta(\mathbf{x})||$ is integrable on $K_+ = \{(\cos\varphi, \sin\varphi)^T : \varphi \in [0, \pi/2]\}$.

(B.3) For every $\varepsilon > 0$

$$\inf_{\theta: ||\theta-\theta_B|| \geq \varepsilon} ||\psi(\theta)|| > 0 = ||\psi(\theta_B)||.$$

(B.4) $\partial_\theta\, \delta_\theta(\mathbf{x})$ exists for every $\mathbf{x} \in K_+$ and is continuous in θ_B.

(B.5) The matrix

$$A_{\theta_B} := \int_0^{\pi/2} \delta^{\angle}_{\theta_B}(\varphi)\delta^{\angle}_{\theta_B}(\varphi)^T + \partial_\theta\, \delta^{\angle}_{\theta_B}(\varphi)(\Lambda^{\angle}_L(\varphi;\theta_B) - \Lambda^{\angle}_L(\varphi))\, d\varphi$$

exists and is non-singular, where $\delta^{\angle}_\theta(\varphi) = \delta_\theta(\cos\varphi, \sin\varphi)$.

Theorem 4.7
If the assumptions (B.1) - (B.3) *hold and the true tail copula* Λ_L *satisfies the first order condition* (4.15) *of Theorem 4.2, then the minimum distance estimator* $\hat{\theta}^{MD}_n$ *is consistent for the parameter* θ_B *corresponding to the best approximation with respect to the distance* ϱ. *If additionally the conditions* (B.4) *and* (B.5) *also hold, then the minimum distance estimator* $\hat{\theta}^{MD}_n$ *is asymptotically normally distributed, more precisely*

$$\Theta^{MD}_n := \sqrt{k}(\hat{\theta}^{MD}_n - \theta_B) = \sqrt{k}\int_0^{\pi/2} \gamma_{\theta_B}(\varphi)\left(\hat{\Lambda}^{\angle}_L(\varphi) - \Lambda^{\angle}_L(\varphi)\right) d\varphi + o_{\mathbb{P}^*}(1)$$
$$\rightsquigarrow \int_0^{\pi/2} \gamma_{\theta_B}(\varphi)\mathbb{G}^{\angle}_{\hat{\Lambda}_L}(\varphi)\, d\varphi =: \Theta^{MD},$$

where $\gamma_{\theta_B}(\varphi) = A^{-1}_{\theta_B}\delta^{\angle}_{\theta_B}(\varphi)$ *and* $\mathbb{G}^{\angle}_{\hat{\Lambda}_L}(\varphi) = \mathbb{G}_{\hat{\Lambda}_L}(\cos\varphi, \sin\varphi)$. *The limiting variable* Θ^{MD} *is centered normally distributed with variance*

$$\sigma^2 = \int_{[0,\pi/2]^2} \gamma_{\theta_B}(\varphi)\gamma_{\theta_B}(\varphi')r(\cos\varphi, \sin\varphi, \cos\varphi', \sin\varphi')\, d(\varphi, \varphi'),$$

where r denotes the covariance functional of the process $\mathbb{G}_{\hat{\Lambda}_L}$ *defined in (4.12).*

n	λ_L	90%	95%	n	λ_L	90%	95%
1000	0.25	0.895	0.955	3000	0.25	0.833	0.914
	0.5	0.893	0.936		0.5	0.900	0.941
	0.75	0.838	0.887		0.75	0.890	0.935

Table 4.6: *Simulated coverage probability of the confidence intervals based on the pdm-bootstrap,* $n = 1000$ $(k = 50)$ *and* $n = 3000$ $(k = 100)$

In order to make use of the latter result in statistical applications one needs the quantiles of the limiting distribution. We propose to use the multiplier bootstrap discussed in the previous section. The following theorem shows that the *pdm* and *dm* bootstrap yield a valid approximation of the distribution of the random variable Θ^{MD}.

Theorem 4.8
If the assumptions of the Theorems 4.7, 4.5 and 4.6 hold and Γ_n *denotes either the process* α_n^{pdm} *(Theorem 4.5) or* α_n^{dm} *(see Theorem 4.6) obtained by the pdm- or dm-bootstrap, respectively, then*

$$\Theta_n^{MD,m} := \int_0^{\pi/2} \gamma_{\hat{\theta}_n^{MD}}(\varphi)\Gamma_n^{\angle}(\varphi)\, d\varphi \ \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}}\ \Theta^{MD},$$

where $\Gamma_n^{\angle}(\varphi) = \Gamma_n(\cos\varphi, \sin\varphi)$.

On the basis of this result it is possible to construct asymptotic confidence regions for the parameter θ as well as to test point hypotheses regarding the parameter. In Table 4.6 we present a small simulation study regarding the finite sample coverage probabilities of some confidence intervals for the parameter of a Clayton tail copula. The sample size is $n = 1000$ or $n = 3000$ and we used $B = 500$ bootstrap replications and 1000 simulation runs to calculate the coverage probabilities. The parameter of the Clayton tail copula is chosen in such a way that the tail dependence coefficient varies in the set $\{1/4, 2/4, 3/4\}$. As one can see we get accurate results in all cases. For small sample sizes the approximation works better for weak tail dependence while for rather large sample sizes strong tail dependence yields slightly more accurate results.
It is also notable that the *dm*- and *pdm*-bootstrap can be used to construct a consistent approximation of the asymptotic distribution of the censored likelihood and moment estimator investigated in [de Haan et al., 2008] and [Einmahl et al., 2008].
In the following we will use the multiplier bootstrap to construct a computationally efficient goodness-of-fit test for the hypothesis that the lower tail copula has a specific parametric form, i.e.

$$\mathcal{H}_0 : \Lambda_L \in \mathcal{L}, \quad \mathcal{H}_1 : \Lambda_L \notin \mathcal{L}. \tag{4.28}$$

This problem has also been discussed in [de Haan et al., 2008] and [Einmahl et al., 2008] who proposed a comparison between a nonparametric and a parametric estimate of the

lower tail copula by an L^2-distance. In both cases the limiting distribution of the corresponding test statistic under the null hypothesis depends in a complicated way on the process $\mathbb{G}_{\hat{\Lambda}_L}$ and the unknown true parameter θ_B. While [Einmahl et al., 2008] does not propose any bootstrap approximation, [de Haan et al., 2008] proposed to use the parametric bootstrap. However, it was pointed out by [Kojadinovic and Yan, 2010] or [Kojadinovic et al., 2010] that for copula models, approximations based on multiplier bootstraps are computationally more efficient, especially for large sample sizes. We will now illustrate how the multiplier bootstrap can be successfully applied in the problem of testing the hypothesis (4.28).
To be precise, we propose to compare a parametric (using the minimum distance estimate $\hat{\theta}_n^{MD}$) and a nonparametric estimate of the tail copula and to reject the null hypothesis (4.28) for large values of the statistic

$$GOF_n := k\,\varrho(\hat{\Lambda}_L, \Lambda_L(\cdot;\hat{\theta}_n^{MD})) = k\int_0^{\pi/2} \left(\hat{\Lambda}_L^{\angle}(\varphi) - \Lambda_L^{\angle}(\varphi;\hat{\theta}_n^{MD})\right)^2 d\varphi,$$

where $\hat{\theta}_n^{MD}$ denotes the minimum distance estimate. If the assumptions (B.1) - (B.5) are satisfied we obtain for the process $H_n = \sqrt{k}\,(\hat{\Lambda}_L - \Lambda_L(\cdot;\hat{\theta}_n^{MD}))$ under the null hypothesis $\mathcal{H}_0 : \Lambda_L = \Lambda_L(\cdot;\theta_B)$

$$\begin{aligned}
H_n &= \sqrt{k}\left(\hat{\Lambda}_L - \Lambda_L - (\hat{\Lambda}_L(\cdot,\hat{\theta}_n^{MD}) - \Lambda_L(\cdot,\theta))\right)\\
&= \sqrt{k}\left(\hat{\Lambda}_L - \Lambda_L - \delta_\theta(\hat{\theta}_n^{MD} - \theta)\right) + o_{\mathbb{P}}(1)\\
&= \sqrt{k}\left(\hat{\Lambda}_L - \Lambda_L - \delta_\theta \int_0^{\pi/2} \gamma_\theta(\varphi)(\hat{\Lambda}_L^{\angle}(\varphi) - \Lambda_L^{\angle}(\varphi)\,d\varphi\right) + o_{\mathbb{P}}(1)\\
&\rightsquigarrow \mathbb{G}_{\hat{\Lambda}_L} - \delta_\theta \int_0^{\pi/2} \gamma_\theta(\varphi)\mathbb{G}_{\hat{\Lambda}_L}^{\angle}(\varphi)\,d\varphi = \mathbb{G}_{\hat{\Lambda}_L} - \delta_\theta\,\Theta^{MD}.
\end{aligned}$$

Under the alternative hypothesis we get an additional summand

$$H_n = \sqrt{k}\left(\hat{\Lambda}_L - \Lambda_L - \delta_\theta(\hat{\theta}_n^{MD} - \theta) - (\Lambda_L(\cdot;\theta_B) - \Lambda_L)\right) + o_{\mathbb{P}}(1),$$

which converges to either plus or minus infinity whenever $\Lambda_L(\mathbf{x},\theta_B) \neq \Lambda_L(\mathbf{x})$. The continuous mapping theorem yields the following result.

Theorem 4.9
Assume that assumptions of Theorem 4.7 are satisfied. If the null hypothesis is valid then

$$GOF_n = \int_0^{\pi/2} \{H_n^{\angle}(\varphi)\}^2\,d\varphi \rightsquigarrow \int_0^{\pi/2} \left(\mathbb{G}_{\hat{\Lambda}_L}^{\angle}(\varphi) - \delta_\theta^{\angle}(\varphi)\,\Theta^{MD}\right)^2 d\varphi,$$

while under the alternative

$$GOF_n = \int_0^{\pi/2} \{H_n^{\angle}(\varphi)\}^2\,d\varphi \xrightarrow{\mathbb{P}} \infty.$$

n	λ_L	$\alpha = 0.15$	$\alpha = 0.1$	$\alpha = 0.05$	n	λ_L	$\alpha = 0.15$	$\alpha = 0.1$	$\alpha = 0.05$
1000	0.25	0.124	0.087	0.037	3000	0.25	0.129	0.090	0.048
	0.5	0.097	0.068	0.032		0.5	0.105	0.069	0.031
	0.75	0.091	0.048	0.018		0.75	0.084	0.056	0.026

Table 4.7: *Simulated rejection probabilities of the pdm-bootstrap test* (4.29) *for the hypothesis* (4.27) *under the null hypothesis;* $n = 1000$ $(k = 50)$, $n = 3000$ $(k = 200)$.

The critical values of the test, which rejects the null hypothesis for large values of GOF_n can be calculated on the basis of the following theorem. The proof is similar to the proof of Theorem 4.8 in Section 4.5 and is therefore omitted.

Theorem 4.10
If the assumptions of the Theorems 4.7, 4.5 and 4.6 hold and Γ_n *denotes either the process* α_n^{pdm} *(Theorem 4.5) or* α_n^{dm} *(Theorem 4.6) obtained by the pdm- or dm-bootstrap, respectively, then it holds independently of the hypotheses that*

$$H_n^m := \Gamma_n - \delta_{\hat{\theta}_n^{MD}} \int_0^{\pi/2} \gamma_{\hat{\theta}_n^{MD}}(\varphi)\, \Gamma_n^{\angle}(\varphi)\, d\varphi \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_{\hat{\Lambda}_L} - \delta_{\theta_B}\, \Theta^{MD}.$$

Therefore

$$GOF_n^m = \int_0^{\pi/2} \{H_n^{m\angle}(\varphi)\}^2\, d\varphi \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \int_0^{\pi/2} \left(\mathbb{G}_{\hat{\Lambda}_L}^{\angle}(\varphi) - \delta_{\theta}^{\angle}(\varphi)\, \Theta^{MD}\right)^2 d\varphi.$$

In order to investigate the finite sample properties of a goodness-of-fit test on the basis of the multiplier bootstrap we show in Table 4.7 the simulated level of the *pdm*-bootstrap test

$$GOF_n > q_{1-\alpha}^{(pdm)} \tag{4.29}$$

where $q_{1-\alpha}^{(pdm)}$ denotes the $(1-\alpha)$ quantile of the bootstrap distribution. For the null hypothesis we considered as the parametric class the family of Clayton tail copulas. In particular we investigated three scenarios corresponding to a coefficient of tail dependence varying in $\{0.25, 0.5, 0.75\}$. The results are based on 1000 simulation runs, while the sample size is either $n = 1000$ and $k = 50$ or $n = 3000$ and $k = 200$. For each test $B = 500$ bootstrap replications with $Laplacian(0, 2)$ multipliers have been performed. We observe a reasonable power and approximation of the nominal level. Note that for the sample size $n = 1000$ the *pdm*-bootstrap test is conservative and this effect is increasing with the level of tail dependence.

n	λ_L	$\alpha = 0.15$	$\alpha = 0.1$	$\alpha = 0.05$	n	λ_L	$\alpha = 0.15$	$\alpha = 0.1$	$\alpha = 0.05$
1000	1/12	0.095	0.052	0.017	3000	1/12	0.217	0.155	0.092
	2/12	0.124	0.066	0.029		2/12	0.374	0.292	0.176
	3/12	0.298	0.200	0.088		3/12	0.868	0.819	0.696

Table 4.8: *Simulated rejection probabilities of the pdm-bootstrap test* (4.29) *for the hypothesis* (4.27) *under the alternative* $C = 1/3\ C_{Clayton} + 2/3\ \Pi$*, where* Π *denotes the independence copula;* $n = 1000$ $(k = 50)$*,* $n = 3000$ $(k = 200)$*.*

4.5 Proofs

Proof of Lemma 4.1. Due to Theorem 1.6.1 in [van der Vaart and Wellner, 1996] the proof of weak convergence of $\tilde{\alpha}_n$ in $\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$ can be given for each $l^\infty(T_i)$ separately. To this end we note that every T_i can be written in the form $T = [0, M_1] \times \{\infty\} \cup \{\infty\} \times [0, M_2] \cup [0, M_3]^2$, where $M_1, M_2, M_3 \in \mathbb{N}$, and show weak convergence in $l^\infty(T)$. Recalling the notation of $f_{n,\mathbf{x}}(\mathbf{U}_i)$ in (4.18) we can express $\tilde{\alpha}_n$ as

$$\tilde{\alpha}_n(\mathbf{x}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f_{n,\mathbf{x}}(\mathbf{U}_i) - \mathbb{E} f_{n,\mathbf{x}}(\mathbf{U}_i)),$$

and the assertion now follows by an application of Theorem 11.20 in [Kosorok, 2008]. For this purpose we show that the assumptions for this result are satisfied. Let $\mathcal{F}_n = \{f_{n,\mathbf{x}} : \mathbf{x} \in T\}$ be a class of functions changing with n and denote by

$$F_n(\mathbf{u}) = \sqrt{\frac{n}{k}}\, \mathbb{I}\left\{u_1 \le kM/n \ \text{ or } \ u_2 \le kM/n\right\},$$

$M = M_1 \vee M_2 \vee M_3$ a corresponding sequence of envelopes of $\mathcal{F}_n$. We have to prove that

(*i*) $(\mathcal{F}_n, F_n)$ satisfies the bounded uniform entropy integral condition

$$\limsup_{n\to\infty} \sup_{Q} \int_0^1 \sqrt{\log N(\varepsilon ||F_n||_{Q,2}, \mathcal{F}_n, L_2(Q))}\, d\varepsilon < \infty, \tag{4.30}$$

where for each n the supremum ranges over all probability measures Q with finite support and $||F_n||_{Q,2} = \left(\int F_n(x)^2\, dQ(x)\right)^{1/2} > 0$.

(*ii*) The limit $H(\mathbf{x}, \mathbf{y}) = \lim_{n\to\infty} \mathbb{E}[\tilde{\alpha}_n(\mathbf{x})\tilde{\alpha}_n(\mathbf{y})]$ exists for every $\mathbf{x}$ and $\mathbf{y}$ in T .

(*iii*) $\limsup_{n\to\infty} \mathbb{E}F_n^2(\mathbf{U}_1) < \infty$

(*iv*) $\lim_{n\to\infty} \mathbb{E}F_n^2(\mathbf{U}_1)\mathbb{I}\{F_n(\mathbf{U}_1) > \varepsilon\sqrt{n}\} = 0$ for all $\varepsilon > 0$.

(v) $\lim_{n\to\infty} \rho_n(\mathbf{x},\mathbf{y}) = \rho(\mathbf{x},\mathbf{y})$ for all $\mathbf{x},\mathbf{y} \in \bar{R}_+^2$, where

$$\rho_n(\mathbf{x},\mathbf{y}) = \left(\mathbb{E}(f_{n,\mathbf{x}}(U_1)) - f_{n,\mathbf{y}}(U_1))^2\right)^{1/2}.$$

Furthermore, for all sequences $(\mathbf{x}_n)_n, (\mathbf{y}_n)_n$ in T the convergence $\rho_n(\mathbf{x}_n,\mathbf{y}_n) \to 0$ holds, provided $\rho(\mathbf{x}_n,\mathbf{y}_n) \to 0$.

(vi) The sequence $\mathcal{F}_n$ of classes is almost measurable Suslin (AMS), i.e. for all $n \geq 1$ there exists a Suslin topological space $T_n \subset T$ with Borel sets $\mathcal{B}_n$ such that

(a) $\mathbb{P}^*(\sup_{\mathbf{x}\in T} \inf_{\mathbf{y}\in T_n} |f_{n,\mathbf{x}}(\mathbf{U}_1) - f_{n,\mathbf{y}}(\mathbf{U}_1)| > 0) = 0$,

(b) $f_{n,\cdot} : [0,1]^2 \times T_n \to \mathbb{R}$ is $\mathcal{B}|_{[0,1]^2} \times \mathcal{B}_n$-measurable for $i = 1,\dots,n$.

In order to prove the bounded uniform entropy integral condition (i) we decompose $\mathcal{F}_n = \bigcup_{i=1}^3 \mathcal{F}_n^{(i)}$ with $\mathcal{F}_n^{(i)} = \{f_{n,\mathbf{x}}^{(i)}, \mathbf{x} \in T\}$ and

$$\begin{aligned}
f_{n,\mathbf{x}}^{(1)}(\mathbf{U}_i) &= \sqrt{\frac{n}{k}}\, \mathbb{I}\{U_{i1} \leq kx_1/n\}\, \mathbb{I}\{x_2 = \infty\}, \\
f_{n,\mathbf{x}}^{(2)}(\mathbf{U}_i) &= \sqrt{\frac{n}{k}}\, \mathbb{I}\{U_{i2} \leq kx_2/n\}\, \mathbb{I}\{x_1 = \infty\}, \\
f_{n,\mathbf{x}}^{(3)}(\mathbf{U}_i) &= \sqrt{\frac{n}{k}}\, \mathbb{I}\{U_{i1} \leq kx_1/n, U_{i2} \leq kx_2/n\}\, \mathbb{I}\{x_1 < \infty, x_2 < \infty\}.
\end{aligned}$$

The corresponding envelopes of the classes $\mathcal{F}_n^{(i)}$ are given by

$$\begin{aligned}
F_n^{(1)}(\mathbf{U}_i) &= \sqrt{\frac{n}{k}}\, \mathbb{I}(U_{i1} \leq kM/n), \quad F_n^{(2)}(\mathbf{U}_i) = \sqrt{\frac{n}{k}}\, \mathbb{I}(U_{i2} \leq kM/n), \\
F_n^{(3)}(\mathbf{U}_i) &= \sqrt{\frac{n}{k}}\, \mathbb{I}(U_{i1} \leq kM/n, U_{i2} \leq kM/n),
\end{aligned}$$

so that $F_n(\mathbf{U}_i) = \max_{i=1}^3 \{F_n^{(i)}(\mathbf{U}_i)\}$. If we prove that the sequences $(\mathcal{F}_n^{(i)}, F_n^{(i)})$ satisfy the bounded uniform integral entropy condition given in (4.30), then the condition holds also for $(\mathcal{F}_n, F_n)$ by Lemma A.4 in the appendix and thus the assertion in (i) is proved. We only consider the (hardest) case of $\mathcal{F}_n^{(3)}$. Note that $\mathcal{F}_n^{(3)} = \{f_{n,\mathbf{x}}, \mathbf{x} \in [0, M_3]^2\} = \mathcal{G}_n^{(1)} \cdot \mathcal{G}_n^{(2)}$, where

$$\begin{aligned}
f_{n,\mathbf{x}} &= (n/k)^{1/2}\, \mathbb{I}\{U_{i1} \leq kx_1/n, U_{i2} \leq kx_2/n\}, \\
\mathcal{G}_n^{(j)} &= \{g_{n,t} = (n/k)^{1/4} \mathbb{I}\{U_{ij} \leq kt/n\} \mid t \in [0, M_3]\}
\end{aligned}$$

for $j = 1,2$. Since the functions $g_{n,t}$ are increasing in t the $G_n^{(j)}$ are VC-classes with VC-index 2. Thus by Lemma 11.21 in [Kosorok, 2008] both classes satisfy the bounded uniform integral entropy condition (4.30). Proposition 11.22 in [Kosorok, 2008] shows that $\mathcal{F}_n^{(3)}$ has the same property and by the discussion at the beginning of this paragraph (i) is satisfied.

For the proof of (ii) note that $\mathbb{E}[\tilde{\alpha}_n(\mathbf{x})\tilde{\alpha}_n(\mathbf{y})] = n/k\left(C(\frac{(\mathbf{x}\wedge\mathbf{y})k}{n}) - C(\frac{\mathbf{x}k}{n})C(\frac{\mathbf{y}k}{n})\right)$, which converges to $\Lambda_L(\mathbf{x}\wedge\mathbf{y}) =: H(\mathbf{x},\mathbf{y})$, since $\frac{n}{k}C(\frac{\mathbf{x}k}{n})C(\frac{\mathbf{y}k}{n}) \to 0$.
Regarding (iii) and (iv) we note that $\mathbb{E}F_n(\mathbf{U}_1)^2 = 2M - \frac{n}{k}C(Mk/n, Mk/n)$, which converges to $2M - \Lambda_L(M,M)$. Further,

$$\begin{aligned}\mathbb{E}F_n^2(\mathbf{U}_1)\mathbb{I}\{F_n(\mathbf{U}_1) > \varepsilon\sqrt{n}\} &= \int_{\{F_n(\mathbf{U}_1)>\varepsilon\sqrt{n}\}} F_n^2(\mathbf{U}_1)\,d\mathbb{P} \\ &\le \frac{n}{k}\mathbb{P}(\frac{1}{k}\mathbb{I}\{U_{11} \le kM/n \text{ or } U_{12} \le kM/n\} > \varepsilon) = 0\end{aligned}$$

for sufficiently large n, such that $k > 1/\varepsilon$. For (v) we note that

$$\begin{aligned}\rho_n(\mathbf{x},\mathbf{y}) = \left(\mathbb{E}(f_{n,\mathbf{x}}(\mathbf{U}_1) - f_{n,\mathbf{y}}(\mathbf{U}_1))^2\right)^{1/2} &= \sqrt{\frac{n}{k}}\left(C(\mathbf{x}k/n) - 2C((\mathbf{x}\wedge\mathbf{y})k/n) + C(\mathbf{y}k/n)\right)^{1/2} \\ &\to \left(\Lambda_L(\mathbf{x}) - 2\Lambda_L(\mathbf{x}\wedge\mathbf{y}) + \Lambda_L(\mathbf{y})\right)^{1/2} =: \rho(\mathbf{x},\mathbf{y}).\end{aligned}$$

Due to Theorem 1 in [Schmidt and Stadtmüller, 2006] we have locally uniform convergence in the latter expression, which yields the second condition stated in (v).
For the proof of condition (vi) we use Lemma 11.15 and the discussion on page 224 in [Kosorok, 2008] and show separability of $\mathcal{F}_n$, i.e. for every $n \ge 1$ there exists a countable subset $T_n \subset T$ such that

$$\mathbb{P}^*\left(\sup_{\mathbf{x}\in T}\inf_{\mathbf{y}\in T_n}|f_{n,\mathbf{y}}(\mathbf{U}_1) - f_{n,\mathbf{x}}(\mathbf{U}_1)| > 0\right) = 0.$$

Choose $T_n = (\mathbb{Q}\cap[0,M_1]\times\{\infty\}) \cup (\{\infty\}\times\mathbb{Q}\cap[0,M_2]) \cup (\mathbb{Q}^2\cap[0,M_3]^2)$, then we have (note that the functions $f_{n,\mathbf{x}}$ are built by indicators) that for every ω and every $\mathbf{x}\in T$ there is an $\mathbf{y}\in T_n$ with $|f_{n,\mathbf{x}}(\mathbf{U}_1(\omega)) - f_{n,\mathbf{y}}(\mathbf{U}_1(\omega))| = 0$. This yields the assertion and thus the proof of Lemma 4.1 is finished. □

Proof of Theorem 4.2. Let $\mathcal{B}_\infty(\mathbb{R}_+)$ denote the set of functions $f : \mathbb{R}_+ \to \mathbb{R}$ (where $\mathbb{R}_+ = [0,\infty)$) that are uniformly bounded on compact sets (equipped with the topology of uniform convergence on compact sets) and define $\mathcal{B}_\infty^{I,0}(\mathbb{R}_+)$ as the subset of all non-decreasing functions $f : \mathbb{R}_+ \to \mathbb{R}_+$ which satisfy $f(0+) = 0$ and for which $\sup\operatorname{ran} f < \infty$ implies that there exists a x_0 with $f(x_0) = \sup\operatorname{ran} f$. The latter condition implies that the adjusted generalized inverse function defined by

$$f^-(z) = \begin{cases}\sup\{x\in\mathbb{R}_+ \mid f(x) = 0\}, & z = 0\\ \inf\{x\in\mathbb{R}_+ \mid f(x) \ge z\}, & 0 < z < \sup\operatorname{ran} f\\ \inf\{x\in\mathbb{R}_+ \mid f(x) = \sup\operatorname{ran} f\}, & z \ge \sup\operatorname{ran} f\end{cases}$$

stays in $\mathcal{B}_\infty(\mathbb{R}_+)$ for every $f \in \mathcal{B}_\infty^{I,0}(\mathbb{R}_+)$. Further set

$$\mathcal{B}_\infty^{I,0}(\bar{\mathbb{R}}_+^2) := \left\{\gamma \in \mathcal{B}_\infty(\bar{\mathbb{R}}_+^2) \mid \gamma(\cdot,\infty) \in \mathcal{B}_\infty^{I,0}(\mathbb{R}_+), \gamma(\infty,\cdot) \in \mathcal{B}_\infty^{I,0}(\mathbb{R}_+)\right\}$$

and now define a map $\Phi : \mathcal{B}_\infty^{I,0}(\bar{\mathbb{R}}_+^2) \to \mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$ by

$$\gamma \longmapsto \Phi(\gamma) = \begin{cases} \gamma(\gamma^-(x,\infty), \gamma^-(\infty, y)) & \text{, if} \quad x, y \neq \infty \\ \gamma(\gamma^-(x,\infty), \infty) & \text{, if} \quad y = \infty \\ \gamma(\infty, \gamma^-(\infty, y)) & \text{, if} \quad x = \infty, \end{cases}$$

see also [Schmidt and Stadtmüller, 2006]. Observing that $\tilde{\Lambda}_L \in \mathcal{B}_\infty^{I,0}(\bar{\mathbb{R}}_+^2)$ and that the adjusted generalized inverse of $\tilde{\Lambda}_L(x,\infty)$ is given by $\frac{n}{k} F_1(F_{n1}^-(kx/n))$, one can conclude that $\Phi(\Lambda_L) = \Lambda_L$ and $\Phi(\tilde{\Lambda}_L) = \hat{\Lambda}_L$ ($\mathbb{P}$-almost surely) and the proof of Theorem 4.2 follows from the functional delta method (Theorem 3.9.4 in [van der Vaart and Wellner, 1996]) and the following Lemma, which is an extension of the result in the proof of Theorem 5 in [Schmidt and Stadtmüller, 2006].

Lemma 4.11
Let Λ_L be a lower tail copula whose partial derivatives satisfy the following first order properties

$$\partial_p \Lambda_L \textit{ exists on } \{\mathbf{x} \in \bar{\mathbb{R}}_+^2 \mid x_p < \infty\} \textit{ and is continuous on } \{\mathbf{x} \in \bar{\mathbb{R}}_+^2 \mid 0 < x_p < \infty\}$$

for $p = 1, 2$. Then Φ is Hadamard-differentiable at Λ_L tangentially to the set

$$\mathcal{C}^0(\bar{\mathbb{R}}_+^2) = \left\{\gamma \in \mathcal{B}_\infty(\bar{\mathbb{R}}_+^2) \mid \gamma \textit{ continuous with } \gamma(\cdot, 0) = \gamma(0, \cdot) = 0\right\}.$$

Its derivative at Λ_L in $\gamma \in \mathcal{C}^0(\bar{\mathbb{R}}_+^2)$ is given by

$$\Phi'_{\Lambda_L}(\gamma)(\mathbf{x}) = \gamma(\mathbf{x}) - \partial_1 \Lambda_L(\mathbf{x}) \gamma(x_1, \infty) - \partial_2 \Lambda_L(\mathbf{x}) \gamma(\infty, x_2) \tag{4.31}$$

where $\partial_p \Lambda_L$, $p = 1, 2$ is defined as 0 on the set $\{\mathbf{x} \in \bar{\mathbb{R}}_+^2 \mid x_p = \infty\}$.

Proof. Decompose $\Phi = \Phi_3 \circ \Phi_2 \circ \Phi_1$ where

$$\begin{aligned}
\Phi_1 :& \mathcal{B}_\infty^{I,0}(\bar{\mathbb{R}}_+^2) \to \mathcal{B}_\infty^{I,0}(\bar{\mathbb{R}}_+^2) \times \mathcal{B}_\infty^{I,0}(\mathbb{R}_+) \times \mathcal{B}_\infty^{I,0}(\mathbb{R}_+) \\
& \gamma \longmapsto (\gamma, \gamma(\cdot, \infty), \gamma(\infty, \cdot)) \\
\Phi_2 :& \mathcal{B}_\infty^{I,0}(\bar{\mathbb{R}}_+^2) \times \mathcal{B}_\infty^{I,0}(\mathbb{R}_+) \times \mathcal{B}_\infty^{I,0}(\mathbb{R}_+) \to \mathcal{B}_\infty^{I,0}(\bar{\mathbb{R}}_+^2) \times \mathcal{B}_\infty^{I,0,-}(\mathbb{R}_+) \times \mathcal{B}_\infty^{I,0,-}(\mathbb{R}_+) \\
& (\gamma, f, g) \longmapsto (\gamma, f^-, g^-)) \\
\Phi_3 :& \mathcal{B}_\infty^{I,0}(\bar{\mathbb{R}}_+^2) \times \mathcal{B}_\infty^{I,0,-}(\mathbb{R}_+) \times \mathcal{B}_\infty^{I,0,-}(\mathbb{R}_+) \to \mathcal{B}_\infty(\bar{\mathbb{R}}_+^2) \\
& (\gamma, f, g) \longmapsto \begin{cases} \gamma(f(x), g(y)) & \text{, if} \quad x, y \neq \infty \\ \gamma(f(x), \infty) & \text{, if} \quad y = \infty \\ \gamma(\infty, g(y)) & \text{, if} \quad x = \infty, \end{cases}
\end{aligned}$$

where $\mathcal{B}_\infty^{I,0,-}(\mathbb{R}_+)$ denotes the set of all adjusted generalized inverse functions f^- with $f \in \mathcal{B}_\infty^{I,0}(\mathbb{R}_+)$. Now Φ_1 is Hadamard-differentiable at Λ_L tangentially to $\mathcal{C}^0(\bar{\mathbb{R}}_+^2)$ since it is linear and continuous. The second map Φ_2 is Hadamard-differentiable at $(\Lambda_L, \mathrm{id}_{\mathbb{R}_+}, \mathrm{id}_{\mathbb{R}_+})$ tangentially to $\mathcal{C}^0(\bar{\mathbb{R}}_+^2) \times \mathcal{C}^0(\mathbb{R}_+) \times \mathcal{C}^0(\mathbb{R}_+)$ where $\mathcal{C}^0(\mathbb{R}_+)$ consists of all continuous functions f on $\mathbb{R}_+$ with $f(0) = 0$ and its derivative at $(\Lambda_L, \mathrm{id}_{\mathbb{R}_+}, \mathrm{id}_{\mathbb{R}_+})$ in (γ, f, g) is given by $\Phi'_{2,(\Lambda_L,\mathrm{id}_{\mathbb{R}_+},\mathrm{id}_{\mathbb{R}_+})}(\gamma,f,g) = (\gamma,-f,-g)$. The proof follows along similar as the one of Lemma 2.6 and is therefor omitted, we just note that $(\mathrm{id}_{\mathbb{R}_+} + t_n f_n)^-(x) > 0$ for all $x > 0$ is implied by the additional assumption of continuity in 0 for functions in the set $\mathcal{B}^{I,0}(\mathbb{R}_+)$. Some more efforts are necessary to show that Φ_3 is Hadamard-differentiable at $(\Lambda_L, \mathrm{id}_{\mathbb{R}_+}, \mathrm{id}_{\mathbb{R}_+})$ tangentially to $\mathcal{C}^0(\bar{\mathbb{R}}^2_+) \times \mathcal{C}^0(\mathbb{R}_+) \times \mathcal{C}^0(\mathbb{R}_+)$ with derivative

$$\Phi'_{3,(\Lambda_L,\mathrm{id}_{\mathbb{R}_+},\mathrm{id}_{\mathbb{R}_+})}(\gamma,f,g)(\mathbf{x}) = \gamma(\mathbf{x}) + \partial_1\Lambda_L(\mathbf{x})f(x_1) + \partial_2\Lambda_L(\mathbf{x})g(x_2).$$

To see this let $t_n \to 0$, $(\gamma_n, f_n, g_n) \in \mathcal{B}_\infty(\bar{\mathbb{R}}^2_+) \times \mathcal{B}_\infty(\mathbb{R}_+) \times \mathcal{B}_\infty(\mathbb{R}_+)$ with $(\gamma_n, f_n, g_n) \to (\gamma, f, g) \in \mathcal{C}^0(\bar{\mathbb{R}}^2_+) \times \mathcal{C}^0(\mathbb{R}_+) \times \mathcal{C}^0(\mathbb{R}_+)$ such that $(\Lambda_L + t_n\gamma_n, \mathrm{id}_{\mathbb{R}_+} + t_n f_n, \mathrm{id}_{\mathbb{R}_+} + t_n g_n) \in \mathcal{B}^{I,0}_\infty(\bar{\mathbb{R}}^2_+) \times \mathcal{B}^{I,0,-}_\infty(\mathbb{R}_+) \times \mathcal{B}^{I,0,-}_\infty(\mathbb{R}_+)$. Now Φ_3 is linear in its first argument and we introduce the decomposition

$$t_n^{-1}\left\{\Phi_3(\Lambda_L + t_n\gamma_n, \mathrm{id}_{\mathbb{R}_+} + t_n f_n, \mathrm{id}_{\mathbb{R}_+} + t_n g_n) - \Phi_3(\Lambda_L, \mathrm{id}_{\mathbb{R}_+}, \mathrm{id}_{\mathbb{R}_+})\right\} = L_{n1} + L_{n2},$$

where

$$\begin{aligned} L_{n1} &= t_n^{-1}\left\{\Phi_3(\Lambda_L, \mathrm{id}_{\mathbb{R}_+} + t_n f_n, \mathrm{id}_{\mathbb{R}_+} + t_n g_n) - \Phi_3(\Lambda_L, \mathrm{id}_{\mathbb{R}_+}, \mathrm{id}_{\mathbb{R}_+})\right\} \\ L_{n2} &= \Phi_3(\gamma_n, \mathrm{id}_{\mathbb{R}_+} + t_n f_n, \mathrm{id}_{\mathbb{R}_+} + t_n g_n). \end{aligned}$$

By the definition of d it suffices to show uniform convergence on sets T of the form $T = [0, M_1] \times \{\infty\} \cup \{\infty\} \times [0, M_2] \cup [0, M_3]^2$, where $M_1, M_2, M_3 \in \mathbb{N}$. Since $T \subset \bar{\mathbb{R}}^2_+$ is compact (f_n, g_n) converges uniformly and γ is uniformly continuous; hence L_{n2} uniformly converges to γ.
Considering L_{n1} we split the investigation into six different cases. First, let $\mathbf{x} \in (0, M_3]^2$. A series expansion at $\mathbf{x}$ yields

$$L_{n1} = \partial_1\Lambda_L(\mathbf{x})f_n(x_1) + \partial_2\Lambda_L(\mathbf{x})g_n(x_2) + r_n(\mathbf{x}),$$

where the error term r_n can be written as

$$r_n(\mathbf{x}) = \big(\partial_1\Lambda_L(\mathbf{y}) - \partial_1\Lambda_L(\mathbf{x})\big)f_n(x_1) + \big(\partial_2\Lambda_L(\mathbf{y}) - \partial_2\Lambda_L(\mathbf{x})\big)g_n(x_2)$$

with some intermediate point $\mathbf{y} = \mathbf{y}(n)$ between $\mathbf{x}$ and $(x_1 + t_n f_n(x_1), x_2 + t_n f_n(x_2))$. The dominating term converges uniformly to $\partial_1\Lambda_L(\mathbf{x})f(x_1) + \partial_2\Lambda_L(\mathbf{x})g(x_2)$, hence it remains to show that $r_n(\mathbf{x})$ converges to 0 uniformly in $\mathbf{x}$. For a given $\varepsilon > 0$ uniform convergence of f_n and uniform continuity of f on $[0, M_3]$ as well as the fact that $f(0) = 0$ allows to choose a $\delta > 0$ such that $|f_n(x_1)| < \varepsilon$ for all $x_1 < \delta$. Since partial derivatives of tail copulas are bounded by 1, the first term of $r_n(\mathbf{x})$ is uniformly small for $x_1 < \delta$. On the quadrangle $[\delta, M_3] \times (0, M_3]$ the partial derivative $\partial_1\Lambda_L$ is uniformly continuous which yields the desired convergence under consideration of $\mathbf{y}(n) \to \mathbf{x}$ and boundedness of f. The same arguments apply for the second derivative and the case $\mathbf{x} \in (0, M_3]^2$ is finished. Now consider the case $\mathbf{x} \in (0, M_3] \times \{0\}$. By Lipschitz-continuity of Λ_L on $\mathbb{R}^2_+$ we get

$$\begin{aligned} |L_{n1}(x_1, 0)| &= t_n^{-1}|\Lambda_L(x_1 + t_n f_n(x_1), t_n g_n(0))| \\ &= t_n^{-1}|\Lambda_L(x_1 + t_n f_n(x_1), t_n g_n(0)) - \Lambda_L(x_1 + t_n f_n(x_1), 0)| \\ &\le |g_n(0)| \to g(0) = 0. \end{aligned}$$

Since $\partial_1\Lambda_L(x_1,0)f(x_1) + \partial_2\Lambda_L(x_1,0)g(0) = 0$ this yields the assertion. For the cases $\mathbf{x} = (0,0)^T$ and $\mathbf{x} \in \{0\} \times (0, M_3]$ the arguments are similar and we proceed with $\mathbf{x} \in [0, M_1] \times \{\infty\}$ (and analogously $\mathbf{x} \in \{\infty\} \times [0, M_2]$)

$$L_{n1}(x_1,\infty) = t_n^{-1}(\Lambda_L(x_1 + t_n f_n(x_1),\infty) - \Lambda_L(x_1,\infty)) = f_n(x_1) \to f(x_1).$$

By $\partial_1\Lambda_L(x_1,\infty) = 1$ and $\partial_2\Lambda_L(x_1,\infty) = 0$ this yields the assertion. To conclude, Φ_3 is Hadamard-differentiable as asserted.
An application of the chain rule (see Lemma 3.9.3 in [van der Vaart and Wellner, 1996]) completes the proof of the Lemma. □

Proof of Theorem 4.3. Due to Lemma A.5 in the Appendix B we can proceed as in the proof of Lemma 4.1 and consider just convergence in $l^\infty(T)$ with $T = [0, M_1] \times \{\infty\} \cup \{\infty\} \times [0, M_2] \cup [0, M_3]^2$, where M_1, M_2 and M_3 are arbitrary constants in $\mathbb{N}$. The assertion now follows from Theorem 11.23 in [Kosorok, 2008], because the corresponding sufficient conditions have already been established in the proof of Lemma 4.1. □

Proof of Theorem 4.6. For technical reasons we give a proof of Theorem 4.6 in advance of Theorem 4.4 and 4.5. The proof is essentially a consequence of a bootstrap version of the functional delta method, see Theorem 12.1 in [Kosorok, 2008]. Since this result only holds for Banach space valued stochastic processes some adjustments have to be made. Note that the space $\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$ is a complete topological vector space with a metric d and some care is necessary whenever technical results depending on the norm are used.
Due to Lemma 4.1 and Theorem 4.3 we have

$$\sqrt{k}(\tilde{\Lambda}_L - \Lambda_L) \rightsquigarrow \mathbb{G}_{\tilde{\Lambda}_L}, \; \sqrt{k}\,\frac{\mu}{\tau}(\tilde{\Lambda}_L^\xi - \tilde{\Lambda}_L) \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_{\tilde{\Lambda}_L}$$

in $\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$. Observing that the generalized inverses of $\tilde{\Lambda}_L(x,\infty)$ and $\tilde{\Lambda}_L^\xi(x,\infty)$ are given by $\frac{n}{k}F_1(F_{n1}^{-}(kx/n))$ and $\frac{n}{k}F_1(F_{n1}^{\xi-}(kx/n))$, respectively, one can conclude that $\Phi(\Lambda_L) = \Lambda_L, \Phi(\tilde{\Lambda}_L) = \hat{\Lambda}_L$ and $\Phi(\tilde{\Lambda}_L^\xi) = \hat{\Lambda}_L^{\xi,\xi}$ ($\mathbb{P}$-almost surely). By Lemma 4.11 Φ is Hadamard-differentiable on $\mathcal{B}_\infty^{I,0}(\bar{\mathbb{R}}^2_+)$ at $\gamma_0 = \Lambda_L$ tangentially to $\mathcal{C}^0(\bar{\mathbb{R}}^2_+) \subset \mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$. Therefore it remains to argue why Theorem 12.1 in [Kosorok, 2008] can be applied in the present context.
A careful inspection of the proof of Theorem 12.1 in [Kosorok, 2008] shows that properties going beyond our specific assumptions (i.e. the complete topological vector space $(\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+), d)$) are used only three times. First of all the mapping Φ'_{Λ_L} needs to be extended to the whole space $\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$, which is possible using equation (4.31) as the defining identity. Secondly, the proof of Theorem 12.1 in [Kosorok, 2008] uses the usual functional delta method as stated in Theorem 2.8 in the same reference, but this result can be replaced by Theorem 3.9.4 in [van der Vaart and Wellner, 1996], which provides a functional delta method holding in general metrizable topological vector spaces. Finally, the proof of Theorem 12.1 in [Kosorok, 2008] makes use of a bootstrap continuous mapping theorem, see Theorem 10.8 in [Kosorok, 2008], which would yield that

$$\sqrt{k}\frac{\mu}{\tau}(\tilde{\Lambda}_L^\xi - \tilde{\Lambda}_L) \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_{\tilde{\Lambda}_L} \;\Rightarrow\; \Phi'_{\Lambda_L}(\sqrt{k}\frac{\mu}{\tau}(\tilde{\Lambda}_L^\xi - \tilde{\Lambda}_L)) \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \Phi'_{\Lambda_L}(\mathbb{G}_{\tilde{\Lambda}_L}).$$

In our specific context this statement follows immediately from the Lipschitz continuity of the derivative Φ'_{Λ_L} and an application of Lemma A.2 in the Appendix. $\square$

Proof of Theorem 4.4. Consider the mapping $\Psi : \mathcal{B}_\infty^{I,0}(\bar{\mathbb{R}}_+^2) \times \mathcal{B}_\infty^{I,0}(\bar{\mathbb{R}}_+^2) \longrightarrow \mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$ defined by $\Psi = \Phi_3 \circ \Phi_2 \circ \Psi_1$, where Φ_3 and Φ_2 are defined in the proof of Lemma 4.11 and Ψ_1 is given by

$$\begin{aligned}\Psi_1 : &\mathcal{B}_\infty^{I,0}(\bar{\mathbb{R}}_+^2) \times \mathcal{B}_\infty^{I,0}(\bar{\mathbb{R}}_+^2) \to \mathcal{B}_\infty^{I,0}(\bar{\mathbb{R}}_+^2) \times \mathcal{B}_\infty^{I,0}(\mathbb{R}_+) \times \mathcal{B}_\infty^{I,0}(\mathbb{R}_+)\\ &(\beta, \gamma) \longmapsto (\beta, \gamma(\cdot, \infty), \gamma(\infty, \cdot)).\end{aligned}$$

We obtain the representations $\Psi(\Lambda_L, \Lambda_L) = \Lambda_L$, $\Psi(\tilde{\Lambda}_L, \tilde{\Lambda}_L) = \hat{\Lambda}_L$ and $\Psi(\tilde{\Lambda}_L^\xi, \tilde{\Lambda}_L) = \hat{\Lambda}_L^{\xi,\cdot}$ ($\mathbb{P}$-almost surely) . Clearly, Ψ_1 is Hadamard-differentiable at (Λ_L, Λ_L) since it is linear and continuous. Φ_2 and Φ_3 are Hadamard-differentiable tangentially to suitable subspaces as well, see the proof of Lemma 4.11. By an application of the chain rule, see Lemma 3.9.3 in [van der Vaart and Wellner, 1996], we can conclude that Ψ is Hadamard-differentiable (Λ_L, Λ_L) tangentially to $\mathcal{C}^0(\bar{\mathbb{R}}_+^2) \times \mathcal{C}^0(\bar{\mathbb{R}}_+^2)$ with derivative

$$\Psi'_{(\Lambda_L, \Lambda_L)}(\beta, \gamma)(\mathbf{x}) = \beta(\mathbf{x}) - \partial_1 \Lambda_L(\mathbf{x}) \gamma(x_1, \infty) - \partial_2 \Lambda_L(\mathbf{x}) \gamma(\infty, x_2).$$

Note that, unlike in the previous proof, we do not have weak convergence (resp. weak conditional convergence) of $\sqrt{k}\left((\tilde{\Lambda}_L, \tilde{\Lambda}_L) - (\Lambda_L, \Lambda_L)\right)$ and $\frac{\mu}{\tau}\sqrt{k}\left((\tilde{\Lambda}_L^\xi, \tilde{\Lambda}_L) - (\tilde{\Lambda}_L, \tilde{\Lambda}_L)\right)$ towards the same limiting field, which would be necessary for an application of the functional delta method for the bootstrap (see for example Theorem 12.1 in [Kosorok, 2008]). Nevertheless, we can mimic certain steps in the proof of this theorem to conclude the result. To be precise, note that we obtain by analogous arguments as on page 236 of [Kosorok, 2008] that

$$\sqrt{k}\begin{pmatrix} \tilde{\Lambda}_L^\xi - \Lambda_L \\ \tilde{\Lambda}_L - \Lambda_L \end{pmatrix} \rightsquigarrow \begin{pmatrix} c^{-1}\mathbb{G}_1 + \mathbb{G}_2 \\ \mathbb{G}_2 \end{pmatrix},$$

unconditionally, where $\mathbb{G}_1$ and $\mathbb{G}_2$ denote independent copies of $\mathbb{G}_{\tilde{\Lambda}_L}$ and $c = \mu\tau^{-1}$. Hadamard-differentiability of the mapping $(\beta, \gamma) \mapsto (\Psi(\beta, \gamma), \Psi(\gamma, \gamma), (\beta, \gamma), (\gamma, \gamma))$ and the usual functional delta method (Theorem 3.9.4 in [van der Vaart and Wellner, 1996]) yields

$$\sqrt{k}\begin{pmatrix} \Psi(\tilde{\Lambda}_L^\xi, \tilde{\Lambda}_L) - \Psi(\Lambda_L, \Lambda_L) \\ \Psi(\tilde{\Lambda}_L, \tilde{\Lambda}_L) - \psi(\Lambda_L, \Lambda_L) \\ (\tilde{\Lambda}_L^\xi, \tilde{\Lambda}_L) - (\Lambda_L, \Lambda_L) \\ (\tilde{\Lambda}_L, \tilde{\Lambda}_L) - (\Lambda_L, \Lambda_L) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \Psi'_{(\Lambda_L, \Lambda_L)}(c^{-1}\mathbb{G}_1 + \mathbb{G}_2, \mathbb{G}_2) \\ \Psi'_{(\Lambda_L, \Lambda_L)}(\mathbb{G}_2, \mathbb{G}_2) \\ (c^{-1}\mathbb{G}_1 + \mathbb{G}_2, \mathbb{G}_2) \\ (\mathbb{G}_2, \mathbb{G}_2) \end{pmatrix}.$$

Observing that $\Psi'_{(\Lambda_L, \Lambda_L)}$ is linear we can conclude that

$$c\sqrt{k}\begin{pmatrix} \Psi(\tilde{\Lambda}_L^\xi, \tilde{\Lambda}_L) - \Psi(\tilde{\Lambda}_L, \tilde{\Lambda}_L) \\ (\tilde{\Lambda}_L^\xi, \tilde{\Lambda}_L) - (\tilde{\Lambda}_L, \tilde{\Lambda}_L) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \Psi'_{(\Lambda_L, \Lambda_L)}(\mathbb{G}_1, 0) \\ (\mathbb{G}_1, 0) \end{pmatrix} = \begin{pmatrix} \mathbb{G}_1 \\ (\mathbb{G}_1, 0) \end{pmatrix}.$$

Continuity of the map $(\alpha, (\beta, \gamma)) \mapsto d(\alpha, \beta)$ yields

$$d\left(c\sqrt{k}\left(\Psi(\tilde{\Lambda}_L^{\xi}, \tilde{\Lambda}_L) - \Psi(\tilde{\Lambda}_L, \tilde{\Lambda}_L)\right), c\sqrt{k}\left(\tilde{\Lambda}_L^{\xi} - \tilde{\Lambda}_L\right)\right) \longrightarrow 0$$

in outer probability. Since $c\sqrt{k}(\tilde{\Lambda}_L^{\xi} - \tilde{\Lambda}_L) \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_1$ we obtain the assertion by an application of Lemma A.1. □

Proof of Theorem 4.5. Let T be a set of the form $T = [0, M_1] \times \{\infty\} \cup \{\infty\} \times [0, M_2] \cup [0, M_3]^2$, see also the beginning of the proof of Lemma 4.1. We start the proof with an assertion regarding consistency of $\widehat{\partial_p \Lambda_L}$ and claim that for any $\delta \in (0, 1)$

$$\sup_{\mathbf{x} \in T : x_p \geq \delta} \left| \widehat{\partial_p \Lambda_L}(\mathbf{x}) - \partial_p \Lambda_L(\mathbf{x}) \right| \longrightarrow 0 \tag{4.32}$$

in outer probability. For a proof of (4.32) split T into three subsets as indicated by its definition and then proceed similar as in the proof of Lemma 4.1 in [Segers, 2010]. The details are omitted. Regarding the assertion of the Theorem we set

$$\bar{\alpha}_n^{pdm}(\mathbf{x}) = \beta_n(\mathbf{x}) - \partial_1 \Lambda_L(\mathbf{x}) \beta_n(x_1, \infty) - \partial_2 \Lambda_L(\mathbf{x}) \beta_n(\infty, x_2).$$

Under consideration of Lemma A.1 it suffices to prove that $d(\alpha_n^{pdm}, \bar{\alpha}_n^{pdm})$ converges to 0 in outer probability. By the definition of d we have to show uniform convergence on the set T. Since $|\alpha_n^{pdm} - \bar{\alpha}_n^{pdm}| \leq D_{n1} + D_{n2}$, where

$$D_{n1} = \left| \widehat{\partial_1 \Lambda_L} - \partial_1 \Lambda_L \right| |\beta_n(\cdot, \infty)|, \qquad D_{n2} = \left| \widehat{\partial_2 \Lambda_L} - \partial_2 \Lambda_L \right| |\beta_n(\infty, \cdot)|$$

we can consider both summands D_{np} separately and deal with D_{n1} exemplarily. First consider the case $\mathbf{x} \in [0, M_3]^2$, then for arbitrary $\varepsilon > 0$ and $\delta \in (0, 1)$

$$\begin{aligned} \mathbb{P}^*\left(\sup_{\mathbf{x} \in [0, M_3]^2} D_{n1}(\mathbf{x}) > \varepsilon\right) \leq\; & \mathbb{P}^*\left(\sup_{\mathbf{x} \in [0, M_3]^2, x_1 \geq \delta} D_{n1}(\mathbf{x}) > \varepsilon/2\right) \\ & + \mathbb{P}^*\left(\sup_{\mathbf{x} \in [0, M_3]^2, x_1 < \delta} D_{n1}(\mathbf{x}) > \varepsilon/2\right). \end{aligned} \tag{4.33}$$

Since $\widehat{\partial_1 \Lambda_L}$ is uniformly consistent on $\{\mathbf{x} \in [0, M_3]^2 \,|\, x_1 \geq \delta\}$ and since β_n is asymptotically tight in $l^\infty(T)$ (β_n converges unconditionally by the results in Chapter 10 of [Kosorok, 2008]) the first probability on the right-hand side converges to zero.
Regarding the second summand note that $F_{n1}^{-}(kx/n) = X_{\lceil kx \rceil : n, 1}$ (where $\lceil x \rceil = \min\{k \in \mathbb{Z} \,|\, k \geq x\}$) so that

$$\sup_{\mathbf{x} \in [0, M_3]^2} \left| \widehat{\partial_1 \Lambda_L}(\mathbf{x}) \right| \leq \sup_{\mathbf{x} \in [0, M_3]^2, x_1 \geq h} \frac{\lceil k(x_1 + h) \rceil - \lceil k(x_1 - h) \rceil}{2h} \leq 1 + \frac{M_3}{2kh} \leq 2$$

for sufficiently large n. Hence the right-hand side of equation (4.33) is bounded by

$$\mathbb{P}^*\left(\sup_{\mathbf{x}\in[0,M_3]^2, x_1<\delta} |\beta_n(\mathbf{x})| > \varepsilon/4\right),$$

eventually. As $\beta_n \rightsquigarrow \mathbb{G}_{\tilde{\Lambda}_L}$ (unconditionally) the lim sup of this outer probability is bounded by

$$\mathbb{P}\left(\sup_{\mathbf{x}\in[0,M_3]^2, x_1<\delta} |\mathbb{G}_{\tilde{\Lambda}_L}(\mathbf{x})| > \varepsilon/4\right).$$

Since $\mathbb{G}_{\tilde{\Lambda}_L}$ has continuous trajectories and $\mathbb{G}_{\tilde{\Lambda}_L}(0, x_2) = 0$ (almost surely) this probability can be made arbitrary small by choosing δ sufficiently small. The case $\mathbf{x} \in [0, M_3]^2$ is finished. For $\mathbf{x} \in [0, M_1] \times \{\infty\}$ the arguments are similar, while for $\mathbf{x} \in \{\infty\} \times [0, M_2]$ we have $D_{n1} = 0$ and nothing has to be shown. To conclude, $\sup_{\mathbf{x}\in T} D_{n1}(\mathbf{x})$ converges to zero in outer probability and because the term $\sup_{\mathbf{x}\in T} D_{n_2}$ can be treated similarly the proof is finished. □

Proof of Theorem 4.7. The consistency follows by Theorem 5.9 in [van der Vaart, 1998] observing the inequality

$$\sup_{\theta\in\Theta} ||\psi_n(\theta) - \psi(\theta)|| \le 2\int \sup_{\theta\in\Theta} ||\delta_\theta^{\angle}(\varphi)||\, |\hat{\Lambda}_L^{\angle}(\varphi) - \Lambda_L^{\angle}(\varphi)|\, d\varphi = o_{\mathbb{P}}(1)$$

and the consistency of the empirical tail copula.
Regarding the asymptotic normality note that by a Taylor expansion we have

$$0 = \psi_n(\hat{\theta}_n^{MD}) = \psi_n(\theta_b) + \partial_\theta\, \psi_n(\bar{\theta})(\hat{\theta}_n^{MD} - \theta_B),$$

where $\| \bar{\theta} - \theta_B \| \le \| \theta_B - \hat{\theta}_n^{MD} \|$. Due to consistency of both the empirical tail copula $\hat{\Lambda}_L$ and the MD-estimator $\hat{\theta}_n^{MD}$ we can conclude (note that the functions $\Lambda_L(\cdot\,;\theta)$, δ_θ and $\partial_\theta\, \delta_\theta$ are continuous in θ_B) that

$$\partial_\theta\, \psi_n(\bar{\theta}) = 2\int \delta_{\bar{\theta}}^{\angle}(\varphi)\delta_{\bar{\theta}}^{\angle}(\varphi)^T + \partial_\theta\, \delta_{\bar{\theta}}^{\angle}(\varphi)(\Lambda_L^{\angle}(\varphi;\bar{\theta}) - \hat{\Lambda}_L^{\angle}(\varphi))\, d\varphi \xrightarrow{\mathbb{P}} 2\, A_{\theta_B}.$$

Since $0 = \psi(\theta_B) = 2\int \delta_{\theta_B}^{\angle}(\varphi)(\Lambda_L^{\angle}(\varphi, \theta_B) - \Lambda_L^{\angle}(\varphi))\, d\varphi$ we obtain

$$2\int \delta_{\theta_B}^{\angle}(\varphi)(\hat{\Lambda}_L^{\angle}(\varphi) - \Lambda_L^{\angle}(\varphi))\, d\varphi = -\psi_n(\theta_B) = 2\,(A_{\theta_B} + o_{\mathbb{P}}(1))(\hat{\theta}_n^{MD} - \theta_B).$$

The probability that $(2\, A_{\theta_B} + o_{\mathbb{P}}(1))$ is invertible, converges to one, which yields the assertion by multiplying the last equality with $\sqrt{k}\, 1/2(A_{\theta_B} + o_{\mathbb{P}}(1))^{-1}$. □

Proof of Theorem 4.8. Since $\Gamma_n \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_{\hat{\Lambda}_L}$ and $\gamma_{\hat{\theta}_n^{MD}} \xrightarrow{\mathbb{P}} \gamma_{\theta_B}$ (consistency of $\hat{\theta}_n^{MD}$ for θ_B) it is easy to see that

$$\sup_{\mathbf{x}\in[0,1]^2} |\gamma_{\hat{\theta}_n^{MD}}\Gamma_n(\mathbf{x}) - \gamma_{\theta_B}\Gamma_n(\mathbf{x})|$$

converges to zero in outer probability. Hence, by Lemma A.1, $\gamma_{\hat{\theta}_n^{MD}}\Gamma_n \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_{\hat{\Lambda}_L}$ in $l^\infty[0,1]^2$ and the assertion follows invoking the Lipschitz-continuous mapping Theorem for the bootstrap, Lemma A.2. $\square$

Appendix A

Auxiliary Results

Lemma A.1
Let $Y_n = Y_n(X_1, \ldots, X_n, \xi_1, \ldots, \xi_n)$ and $Z_n = Z_n(X_1, \ldots, X_n, \xi_1, \ldots, \xi_n)$ be two (bootstrap) statistics in a metric space $(\mathbb{D}, d)$, depending on the data $X_1, \ldots, X_n$ and on some multipliers $\xi_1, \ldots, \xi_n$. If $Y_n \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} Y$ in $\mathbb{D}$, where Y is tight, and $d(Y_n, Z_n) \overset{\mathbb{P}}{\to} 0$, then also $Z_n \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} Y$ in $\mathbb{D}$.

Proof. We only prove (i) in Definition 1.20, the assertion about the asymptotic measurability in (ii) follows along similar lines. Observing the estimate

$$\sup_{h \in BL_1(\mathbb{D})} \left| \mathbb{E}_\xi h(Z_n) - \mathbb{E} h(Y) \right| \leq \mathbb{E}_\xi \left[d(Y_n, Z_n)^* \wedge 2 \right] + \sup_{h \in BL_1(\mathbb{D})} \left| \mathbb{E}_\xi h(Y_n) - \mathbb{E} h(Y) \right|$$

it suffices to show that $\mathbb{E}_\xi \left[d(Y_n, Z_n)^* \wedge 2 \right]$ converges to 0 in outer probability. Now the random variable $d(Y_n, Z_n)^* \wedge 2$ is uniformly integrable and converges in probability by assumption, hence it also converges in L^1. We finally use Markov's inequality to obtain $\mathbb{E}_\xi [d(Y_n, Z_n)^* \wedge 2] \overset{\mathbb{P}}{\to} 0$, which proves the assertion. □

The following Lemma extends a bootstrap continuous mapping theorem for Lipschitz-continuous functions between Banach spaces, see Proposition 10.7 in [Kosorok, 2008], to the case of metrized topological vector spaces.

Lemma A.2
Suppose that $g : \mathbb{D}_1 \longrightarrow \mathbb{D}_2$ is a Lipschitz-continuous map between metrized topological vector spaces. If $G_n = G_n(X_1, \ldots, X_n, \xi_1, \ldots, \xi_n) \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} G$ in $\mathbb{D}_1$, where G is tight, then $g(G_n) \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} g(G)$ in $\mathbb{D}_2$.

Proof. Let $c_0 \geq 1$ denote a Lipschitz constant for g. For arbitrary $h \in BL_1(\mathbb{D}_2)$ we have $|hg(\alpha) - hg(\beta)| \leq d(g\alpha, g\beta) \leq c_0 d(\alpha, \beta)$, which yields that the mapping $h' := c_0^{-1} hg$ lies in $BL_1(\mathbb{D}_1)$ and therefore

$$\sup_{h \in BL_1(\mathbb{D}_2)} |\mathbb{E}_\xi h(g(G_n)) - \mathbb{E} h(g(G))| \leq c_0 \sup_{h \in BL_1(\mathbb{D}_1)} |\mathbb{E}_\xi h(G_n) - \mathbb{E} h(G)| \overset{\mathbb{P}}{\to} 0.$$

The asymptotic measurability follows along similar lines. □

Lemma A.3
Let $X_n, X_{i,n} : \Omega \to \mathbb{D}$ for $i, n \in \mathbb{N}$ be arbitrary maps with values in the metric space $(\mathbb{D}, d)$ and $X_i, X : \Omega \to \mathbb{D}$ be Borel-measurable. Suppose that

$$\begin{aligned} &\text{(i) For every } i \in \mathbb{N}: \; X_{i,n} \rightsquigarrow X_i \text{ for } n \to \infty, \\ &\text{(ii) } X_i \rightsquigarrow X \text{ for } i \to \infty, \\ &\text{(iii) For every } \varepsilon > 0: \; \lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*(d(X_{i,n}, X_n) > \varepsilon) = 0. \end{aligned}$$

Then $X_n \rightsquigarrow X$ for $n \to \infty$.

Proof. Let $F \subset \mathbb{D}$ be closed and fix $\varepsilon > 0$. If $F^\varepsilon = \{x \in \mathbb{D} : d(x, F) \leq \varepsilon)$ denotes the ε-enlargement of F we obtain

$$\mathbb{P}^*(X_n \in F) \leq \mathbb{P}^*(X_{i,n} \in F^\varepsilon) + \mathbb{P}^*(d(X_{i,n}, X_n) > \varepsilon).$$

By hypothesis (i) and the Portmanteau-Theorem [see [van der Vaart and Wellner, 1996]]

$$\limsup_{n\to\infty} \mathbb{P}^*(X_n \in F) \leq \mathbb{P}(X_i \in F^\varepsilon) + \limsup_{n\to\infty} \mathbb{P}^*(d(X_{i,n}, X_n) > \varepsilon).$$

By conditions (ii) and (iii) $\limsup_{n\to\infty} \mathbb{P}^*(X_n \in F) \leq P(X \in F^\varepsilon)$ and since $F^\varepsilon \downarrow F$ for $\varepsilon \downarrow 0$ and closed F the result follows by the Portmanteau-Theorem. □

Lemma A.4
Suppose $\mathcal{G}_n$ and $\mathcal{H}_n$ are sequences of measurable functions with envelopes G_n and H_n, so that $(\mathcal{G}_n, G_n)$ and $(\mathcal{H}_n, H_n)$ satisfy the bounded uniform integral entopry condition as stated in (4.30). Then the bounded uniform entropy integral condition (4.30) holds also for $\mathcal{F}_n = \mathcal{G}_n \cup \mathcal{H}_n$, with envelopes $F_n = G_n \vee H_n$.

Proof. Note that

$$\begin{aligned} N &:= N(\varepsilon||F_n||_{Q,2}, \mathcal{F}_n, L_2(Q)) \\ &\leq N(\varepsilon||F_n||_{Q,2}, \mathcal{G}_n, L_2(Q)) + N(\varepsilon||F_n||_{Q,2}, \mathcal{H}_n, L_2(Q)) \leq N_1 + N_2, \end{aligned}$$

where

$$N_1 = N(\varepsilon||G_n||_{Q,2}, \mathcal{G}_n, L_2(Q)) \quad \text{and} \quad N_2 = N(\varepsilon||H_n||_{Q,2}, \mathcal{H}_n, L_2(Q)).$$

By monotonicity and subadditivity of $\log(n)$ and $\sqrt{n}$ for $n \geq 2$ we obtain the inequality

$$\sqrt{\log N} \leq \sqrt{\log N_1} + \sqrt{\log N_2},$$

which yields the assertion. □

Lemma A.5
Suppose $G_n = G_n(\mathbf{X}_1, \ldots \mathbf{X}_n, \xi_1, \ldots \xi_n)$ is some statistic taking values in $\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$. Then a conditional version of Theorem 1.6.1 in [van der Vaart and Wellner, 1996] holds , namely $G_n \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} G$ in $\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$ is equivalent to $G_n \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} G$ in $l^\infty(T_i)$ for every $i \in \mathbb{N}$.

Proof. We first show that

$$a_n = \sup_{h \in BL_1(\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+))} |\mathbb{E}_\xi h(G_n) - \mathbb{E}h(G)| \overset{\mathbb{P}}{\to} 0 \tag{A.1}$$

is equivalent to

$$a_n(T_i) = \sup_{h \in BL_1(l^\infty(T_i))} |\mathbb{E}_\xi h(G_n|_{T_i}) - \mathbb{E}h(G|_{T_i})| \overset{\mathbb{P}}{\to} 0 \tag{A.2}$$

for all $i \in \mathbb{N}$.
Suppose first that (A.1) holds. For arbitrary $h \in BL_1(l^\infty(T_i))$ we consider the mapping

$$h' : \mathcal{B}_\infty(\bar{\mathbb{R}}^2_+) \to \mathbb{R}, \alpha \mapsto h(\alpha|_{T_i}).$$

Then

$$\begin{aligned}|h'(\alpha) - h'(\beta)| &\leq \sup_{\mathbf{x} \in T_i} |\alpha(\mathbf{x}) - \beta(\mathbf{x})| \wedge 2 \leq 2 \sum_{j=1}^{i} \left(\sup_{\mathbf{x} \in T_j} |\alpha(\mathbf{x}) - \beta(\mathbf{x})| \wedge 1 \right) \\ &\leq 2^{i+1} \sum_{j=1}^{i} 2^{-j} \left(\sup_{\mathbf{x} \in T_j} |\alpha(\mathbf{x}) - \beta(\mathbf{x})| \wedge 1 \right) \leq 2^{i+1} d(\alpha, \beta),\end{aligned}$$

and therefore the mapping $h'' := 2^{-i-1}h'$ is an element of $BL_1(\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+))$. Observing

$$\begin{aligned}|\mathbb{E}_\xi h(G_n|_{T_i}) - \mathbb{E}h(G|_{T_i})| &= 2^{i+1}|\mathbb{E}_\xi h''(G_n) - \mathbb{E}h''(G)| \\ &\leq 2^{i+1} \sup_{h \in BL_1(\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+))} |\mathbb{E}_\xi h(G_n) - \mathbb{E}h(G)|,\end{aligned}$$

this yields the assertion.
Now suppose that (A.2) holds for all $i \in \mathbb{N}$. For $h \in BL_1(\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+))$ define $h_i(\alpha) = h(\alpha\, \mathbb{I}_{T_i})$, then $h_i \in BL_1(l^\infty(T_i))$ by the following reasoning: Obviously $||h_i|| \leq 1$ and

$$\begin{aligned}|h_i(\alpha) - h_i(\beta)| = |h(\alpha\, \mathbb{I}_{T_i}) - h(\beta\, \mathbb{I}_{T_i})| &\leq \sum_{j=1}^{\infty} 2^{-j} \left(\sup_{\mathbf{x} \in T_j} \left|(\alpha\, \mathbb{I}_{T_i})(\mathbf{x}) - (\beta\, \mathbb{I}_{T_i})(\mathbf{x})\right| \wedge 1 \right) \\ &= \sum_{j=1}^{i} 2^{-j} \left(\sup_{\mathbf{x} \in T_j} \left|\alpha(\mathbf{x}) - \beta(\mathbf{x})\right| \wedge 1 \right) + \left(\sup_{\mathbf{x} \in T_i} \left|\alpha(\mathbf{x}) - \beta(\mathbf{x})\right| \wedge 1 \right) \sum_{j=i+1}^{\infty} 2^{-j} \\ &\leq \left(\sup_{\mathbf{x} \in T_i} \left|\alpha(\mathbf{x}) - \beta(\mathbf{x})\right| \wedge 1 \right) \sum_{j=1}^{\infty} 2^{-j} \\ &\leq \sup_{\mathbf{x} \in T_i} \left|\alpha(\mathbf{x}) - \beta(\mathbf{x})\right|.\end{aligned}$$

Now we choose for $\varepsilon > 0$ an i_0 with $\sum_{j=i_0+1}^{\infty} 2^{-j} < \varepsilon$, then for any $i \geq i_0$

$$|h(\alpha) - h_i(\alpha)| = |h(\alpha) - h(\alpha\, \mathbb{I}_{T_i})| \leq d(\alpha, \alpha\, \mathbb{I}_{T_i}) = \sum_{j=1}^{\infty} 2^{-j} \sup_{\mathbf{x} \in T_j} \left|\alpha(\mathbf{x}) - (\alpha\, \mathbb{I}_{T_i})(\mathbf{x})\right| \wedge 1$$
$$= \sum_{j=i+1}^{\infty} 2^{-j} \left(\sup_{\mathbf{x} \in T_j} \left|\alpha(\mathbf{x})\right| \wedge 1 \right) = \sum_{j=i+1}^{\infty} 2^{-j} < \varepsilon.$$

This yields, for any $i \geq i_0$

$$\begin{aligned} |\mathbb{E}_\xi h(G_n) - \mathbb{E} h(G)| &\leq |\mathbb{E}_\xi h(G_n) - \mathbb{E}_\xi h_i(G_n)| + |\mathbb{E}_\xi h_i(G_n) - \mathbb{E} h_i(G)| + |\mathbb{E} h_i(G) - \mathbb{E} h(G)| \\ &\leq 2\varepsilon + |\mathbb{E}_\xi h_i(G_n) - \mathbb{E} h_i(G)| \\ &\leq 2\varepsilon + \sup_{h \in BL_1(l^\infty(T_i))} |\mathbb{E}_\xi h(G_n) - \mathbb{E} h(G)| \end{aligned}$$

and the latter summand converges to 0 in outer probability by assumption. Since $\varepsilon > 0$ was arbitrary the assertion follows.
It remains to show that $\mathbb{E}_\xi h(G_n)^* - \mathbb{E}_\xi h(G)_* \xrightarrow{\mathbb{P}} 0$ for all $h \in BL_1(\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2))$ if and only if $\mathbb{E}_\xi h(G_n|_{T_i})^* - \mathbb{E}_\xi h(G_n|_{T_i})_* \xrightarrow{\mathbb{P}} 0$ for every $h \in BL_1(l^\infty(T_i))$ and every $i \in \mathbb{N}$. This assertion can be proved along similar lines as in the proof of the first equivalence given above. The necessity follows from

$$\mathbb{E}_\xi h(G_n|_{T_i})^* - \mathbb{E}_\xi h(G_n|_{T_i})_* = 2^{i+1} \mathbb{E}_\xi h''(G_n)^* - \mathbb{E}_\xi h''(G_n)_* \xrightarrow{\mathbb{P}} 0,$$

while sufficiency can be concluded from

$$\begin{aligned} &\mathbb{E}_\xi h(G_n)^* - \mathbb{E}_\xi h(G_n)_* \\ &\leq \mathbb{E}_\xi |h(G_n)^* - h_i(G_n)^*| + \mathbb{E}_\xi h_i(G_n)^* - \mathbb{E}_\xi h_i(G_n)_* + \mathbb{E}_\xi |h_i(G_n)_* - h(G_n)_*| \\ &\leq 2\varepsilon + \mathbb{E}_\xi h_i(G_n)^* - \mathbb{E}_\xi h_i(G_n)_* \xrightarrow{\mathbb{P}} 2\varepsilon, \end{aligned}$$

where we estimated the first and the last summand in the second line under consideration of Lemma 1.2.2 in [van der Vaart and Wellner, 1996]. □

Remark A.6
With a similar argument one can conclude that $G_n \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} G$ in $\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$ implies $G_n \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} G$ in $l^\infty([0,1]^2)$, which is the space needed in most applications.

List of Abbreviations and Symbols

Notation that is largely confined to sections or chapters is mostly excluded from the list below.

Modes of convergence

$\xrightarrow{\text{as}*}$	convergence outer almost surely	13
$\underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}}$	conditional weak convergence in probability	14
$\xrightarrow{\mathbb{P}}$	convergence in (outer) probability	13
$\rightsquigarrow$	weak convergence	12

Miscellaneous

$\mathbb{B}_C$	Gaussian field on $[0,1]^2$	19
C	copula	2
C_n	empirical copula	18
$D(G)$	the maximum domain of attraction of G	5
F^-	generalized inverse of F	1
G_γ	extreme value distribution function	6
$\mathbb{G}_C$	Gaussian field on $[0,1]^2$	19
l	stable tail dependence function	4
$l^\infty(T)$	the space of all uniformly bounded functions on T	13
λ_L	coefficient of lower tail dependence	3
Λ_L	lower tail copula	4
λ_U	coefficient of upper tail dependence	3
Λ_U	upper tail copula	4

Bibliography

Ben Ghorbal, N., Genest, C., and Nešlehová, J. (2009). On the Ghoudi, Khoudraji, and Rivest test for extreme-value dependence. *Canad. J. Statist.*, 37(4):534–552.

Billingsley, P. (1968). *Convergence of probability measures*. John Wiley & Sons Inc., New York.

Capéraà, P., Fougères, A. L., and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3):567–577.

Cebrian, A. C., Denuit, M., and Lambert, P. (2003). Analysis of bivariate tail dependence using extreme values copulas: An application to the SOA medical large claims database. *Belgian Actuarial Journal*, 3:33–41.

Chen, S. X. and Huang, T.-M. (2007). Nonparametric estimation of copula functions for dependence modelling. *Canad. J. Statist.*, 35(2):265–282.

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2:339–365.

Coles, S. G. and Tawn, J. A. (1994). Statistical methods for multivariate extremes: An application to structural design. *Journal of Applied Statistics*, 43:1–48.

de Haan, L. and Ferreira, A. (2006). *Extreme value theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York.

de Haan, L., Neves, C., and Peng, L. (2008). Parametric tail copula estimation and model testing. *J. Multivariate Anal.*, 99(6):1260–1275.

de Haan, L. and Resnick, S. I. (1993). Estimating the limit distribution of multivariate extremes. *Comm. Statist. Stochastic Models*, 9(2):275–309.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. (5)*, 65(6):274–292.

Deheuvels, P. (1984). Probabilistic aspects of multivariate extremes. In de Oliveira, J. T., editor, *Statistical Extremes and Applications*. Reidel, Dordrecht.

Deheuvels, P. (1991). On the limiting behavior of the pickands estimator for bivariate extreme-value distributions. *Statistics and Probability Letters*, 12:429–439.

Drees, H. and Huang, X. (1998). Best attainable rates of convergence for estimates of the stable tail dependence functions. *Journal of Multivariate Analysis*, 64:25–47.

Einmahl, H. J., de Haan, L., and Li, D. (2006). Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *Annals of Statistics*, 34:1987–2014.

Einmahl, H. J., de Haan, L., and Xin, H. (1993). Estimating a multidimensional extreme-value distribution. *Journal of Multivariate Analysis*, 47:35–47.

Einmahl, H. J., Krajina, A., and Segers, J. (2008). A method of moments estimator of tail dependence. *Bernoulli*, 14:1003–1026.

Einmahl, J. H. J., de Haan, L., and Piterbarg, V. I. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statist.*, 29(5):1401–1423.

Fermanian, J.-D., Radulović, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847–860.

Fils-Villetard, A., Guillou, A., and Segers, J. (2008). Projection estimators of Pickands dependence functions. *Canad. J. Statist.*, 36(3):369–382.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *N. Am. Actuar. J.*, 2(1):1–25.

Gänssler, P. and Stute, W. (1987). *Seminar on empirical processes*, volume 9 of *DMV Seminar*. Birkhäuser Verlag, Basel.

Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Engineering*, 12:347–368.

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 Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann. Inst. H. Poincaré. Probabilités et Statistiques*, 44:1096–1127.

Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit test for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44:199–213.

Genest, C. and Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *Ann. Statist.*, 37(5B):2990–3022.

Ghoudi, K., Khoudraji, A., and Rivest, L.-P. (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *Canad. J. Statist.*, 26(1):187–197.

Ghoudi, K. and Rémillard, B. (2004). Empirical processes based on pseudo-observations. II. The multivariate case. In *Asymptotic methods in stochastics*, volume 44 of *Fields Inst. Commun.*, pages 381–406. Amer. Math. Soc., Providence, RI.

Giné, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.*, 18(2):851–869.

Gumbel, É. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statist. Univ. Paris*, 9:171–173.

Hall, P. and Tajvidi, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli*, 6(5):835–844.

Hsing, T. (1989). Extreme value theory for multivariate stationary sequences. *J. Multivariate Anal.*, 29(2):274–291.

Huang, X. (1992). *Statistics of bivariate extreme values.* PhD thesis, Tinbergen Institute Research Series, Netherlands.

Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: between independence and complete dependence. *Statist. Probab. Lett.*, 7(4):283–286.

Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statist. Probab. Lett.*, 9(1):75–81.

Joe, H. (1997). *Multivariate models and dependence concepts*, volume 73 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A. (2004). Kernlab – an S4 package for kernel methods in R. *Journal of Statistical Software*, 11(9):1–20.

Kojadinovic, I. and Yan, J. (2010). A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing*. in press.

Kojadinovic, I., Yan, J., and Holmes, M. (2010). Fast large sample goodness-of-fit tests for copulas. *Statistica Sinica*. to appear.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics, New York.

Marshall, A. W. (1970). Discussion of Barlow and van Zwets papers. In Pudi, M. L., editor, *Nonparametric Techniques in Statistical Inference*, pages 175–176. Cambridge University Press, London.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management*. Princeton Series in Finance. Princeton University Press, Princeton, NJ. Concepts, techniques and tools.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition.

Omelka, M., Gijbels, I., and Veraverbeke, N. (2009). Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *Ann. Statist.*, 37(5B):3023–3058.

Peng, L. and Qi, Y. (2008). Bootstrap approximation of tail dependence function. *Journal of Multivariate Analysis*, 99:1807–1824.

Pickands, J. (1981). Multivariate extreme value distributions. *Bull. Inst. Internat. Statist.*, 49(2):859–878, 894–902.

Qi, Y. (1997). Almost sure convergence of the stable tail empirical dependence function in multivariate extreme statistics. *Acta Math. Appl. Sinica (English Ser.)*, 13:167–175.

Rémillard, B. and Scaillet, O. (2009). Testing for equality between two copulas. *J. Multivariate Anal.*, 100(3):377–386.

Resnick, S. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer Series in Operations Research and Financial Engineering, New York.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1996). *Ordered Restricted Statistical Inference*. Wiley, New York.

Rojo Jiménez, J., Villa-Diharce, E., and Flores, M. (2001). Nonparametric estimation of the dependence function in bivariate extreme value distributions. *J. Multivariate Anal.*, 76(2):159–191.

Rüschendorf, L. (2009). On the distributional transform, Sklar's theorem, and the empirical copula process. *J. Statist. Plann. Inference*, 139(11):3921–3927.

Scaillet, O. (2005). A Kolmogorov-Smirnov type test for positive quadrant dependence. *Canad. J. Statist.*, 33(3):415–427.

Schmidt, R. (2003). *Dependencies of extreme events in finance*. PhD thesis, Universität Ulm, Germany.

Schmidt, R. and Stadtmüller, U. (2006). Non-parametric estimation of tail dependence. *Scand. J. Statist.*, 33(2):307–335.

Segers, J. (2007). Nonparametric inference for bivariate extreme-value copulas. In Ahsanullah, M. and Kirmani, S. N. U. A., editors, *Topics in Extreme Values*. Nova Science Publishers, New York.

Segers, J. (2010). Weak convergence of empirical copula processes under nonrestrictive smoothness assumptions. *arXiv:1012.2133v*.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.

Strimmer, K. (2009). *fdrtool: Estimation and Control of (Local) False Discovery Rates*. R package version 1.2.6.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415.

Tiago de Oliveira, J. (1980). Bivariate extremes: Foundations and statistics. In Krishnaiah, P. R., editor, *Multivariate Analysis 5*, pages 349–366. North-Holland, Amsterdam.

Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canad. J. Statist.*, 33(3):357–375.

van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York.

Wang, J. L. (1986). Asymptotically minimax estimators for distributions with increasing failure rate. *Annals of Statistics*, 44:1113–1131.

Zhang, D., Wells, M. T., and Peng, L. (2008). Nonparametric estimation of the dependence function for a multivariate extreme value distribution. *J. Multivariate Anal.*, 99(4):577–588.