An M-Estimator of Multivariate Tail Dependence

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit van Tilburg, op gezag van de rector magnificus, prof. dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op vrijdag 23 april 2010 om 10.15 uur door

Andrea Krajina

geboren op 22 maart 1981 te Tomislavgrad, Bosnië-Herzegovina.

Promotor: prof. dr. John Einmahl
Copromotor: dr. Johan Segers

Promotiecommissie: dr. Bojan Basrak
prof. dr. Anne-Laure Fougères
prof. dr. Laurens de Haan
dr. Roger Laeven
prof. dr. Bas Werker

Thomas Stieltjes Institute
for Mathematics

Acknowledgements

During the past four years that I worked on what is now presented in this thesis, I have learned a lot about Extreme Value Theory, but I have also learned a lot about myself. Some lessons were too hard to deal with alone and it is the people who helped me I wish to thank here.

First, I wish to thank my supervisors, prof. dr. John H.J. Einmahl and dr. Johan Segers. I am very thankful for having been given the opportunity to be their student. They have surpassed their role of great supervisors, which they were, by being kind and supportive, helping me overcome difficulties far bigger than the complicated proofs. Johan was always constructive in his comments, imaginative with the ideas, patient with his student when she was not as fast. The thoughtfulness, kindness, patience and consideration that are accompanying John's incredible knowledge and creativity make him a person that one can only aspire to be.

I warmly thank my committee members: dr. Bojan Basrak, prof. dr. Anne-Laure Fougères, prof. dr. Laurens de Haan, dr. Roger Laeven and prof. dr. Bas Werker. Also, I thank prof. dr. Claudia Klüppelberg for being my host at TUM, Munich, Germany in late 2007 and dr. Bojan Basrak for finding this Ph.D. project for me. Further, I thank Department of Mathematics at J.J. Strossmayer University of Osijek, Croatia; to prof. dr. Rudolf Scitovski, prof. dr. Dragan Jukić and prof. dr. Mirta Benšić in particular, for every way in which they supported me throughout these years.

I am grateful to Department of Econometrics and OR, FEB, Tilburg University for providing an inspiring work environment. I thank Hendri for his help with LaTeX and similar matters. To Sandra, Eva, Bauke, Baukelien and Eric I am grateful for the assistance with administrative matters that were often outside of their job description.

My close friends in Tilburg formed a family-like circle, providing help, support, opinions, food, humor, all with a twist of cynicism and irony, whenever it was needed (and sometimes when it was not). More gratitude than one can express in words goes to them: Ben, Cristina, Michele and Renata. Chance had made me share office with Amar and Maria and I am grateful for that, since having friends like them as office-mates made my working hours seem shorter. To all the old, new, temporary, and adjacent Lunch Group members, I am grateful for making my stay in Tilburg eventful, interesting and colorful, for many interesting conversations, the things they taught me, the times we had fun: Heejung, Owen, Willem, Edwin, Marcel, Bart, Gema, Benjamin, Linde, Anne, Anna, Akos, Aminah, DJ Mistoz, Momo, David, Hanka, Pedro B., Pedro R., Miguel, Nathanael, Khoa, Otilia, Karen, Verena, Amal.

To Ivana and Nenad I am grateful for being such good friends, no matter the distance.

Finally, I wish to express my deepest gratitude to the closest ones: to Chris, for everything; to my family, for believing in me and their unconditional love.

Contents

Chapter 1

Introduction

Extreme value theory is the part of probability and statistics that provides the theoretical background for modeling events that almost never happen. The interest in these events originates from their potentially large consequences, like in a case of a big flood or a stock market crash. Examples of extreme events come from fields such as meteorology (floods, wind storms, heavy rainfall, large scale forest fires), finance, non-life insurance and re-insurance, internet page ranking, athletics, etc., see de Haan (1990); Rootzén and Tajvidi (1995); Perrin, Rootzén, and Taesler (2006); Katz, Parlange, and Naveau (2002); Vrac and Naveau (2007); Ané and Kharoubi (2003); Embrechts, Klüppelberg, and Mikosch (1997); Volkovich (2009); Einmahl and Magnus (2008).

Over the past decades, extreme value theory, univariate as well as multivariate, has become widely used, known and studied. Some of the useful monographs that have facilitated that progress are: Resnick (1987); Embrechts, Klüppelberg, and Mikosch (1997); Coles (2001); Beirlant, Goegebeur, Segers, and Teugels (2004); Falk, Hüsler, and Reiss (2004); de Haan and Ferreira (2006); Resnick (2007).

In this introduction the core ideas in extreme value theory are presented, with the focus on the notions needed in the thesis. We conclude with an overview of the thesis.

1.1 Univariate extreme value theory

The theory of extreme values was inspired by the following question: "What happens with the limit distribution in the central limit theorem when the sequence of partial sums is replaced by the sequence of partial maxima?". Consider a random sample $X_1, \dots, X_n$ from a distribution function F. By a generalization of the central limit theorem we know that if there exists a sequence (s_n) of positive numbers and a sequence (m_n) of real numbers such that

$$\frac{\sum_{i=1}^n X_i - m_n}{s_n} \xrightarrow{d} Z, \quad \text{as} \quad n \to \infty,$$

then Z is a random variable with a stable distribution. Extreme value theory establishes under which conditions on F there exists a sequence (a_n) of positive numbers and a sequence (b_n) of real numbers such that as $n \to \infty$,

$$\frac{\max_{i=1,\dots,n} X_i - b_n}{a_n} \xrightarrow{d} Y,$$

where the distribution of Y is non-degenerate; and it describes the possible distributions of Y.

If there exists a sequence (a_n) of positive numbers and (b_n) of real numbers such that

$$\mathbb{P}\left(\frac{\max_{i=1\dots,n} X_i - b_n}{a_n} \le x\right) = F^n\left(a_n x + b_n\right) \to G(x), \quad \text{as} \quad n \to \infty, \qquad (1.1.1)$$

for every continuity point of G, we say that F is in the *max-domain of attraction* of G, and call G an *extreme value distribution.* The class of extreme value distributions was first described in Fisher and Tippet (1928) and Gnedenko (1943). It holds that there exist $\gamma \in \mathbb{R}$, $a > 0$ and $b \in \mathbb{R}$ such that $G(x) = G_\gamma(ax + b)$, where

$$G_\gamma(x) = \begin{cases} \exp\{-(1+\gamma x)^{-1/\gamma}\}, & \text{if } 1 + \gamma x > 0 \text{ and } \gamma \in \mathbb{R} \setminus \{0\}, \\ \exp\{-e^{-x}\}, & \text{if } \gamma = 0. \end{cases}$$

The above parametrization is from von Mises (1936).

Up to location and scale, extreme value distributions form a single-parameter family. The parameter γ is called the *extreme value index*. We distinguish three different subclasses of extreme value distributions.

- If $\boldsymbol{\gamma > 0}$, the sequences (a_n) and (b_n) can be chosen such that G has a Fréchet$(1/\gamma)$ distribution, $G(x) = \exp\{-x^{-1/\gamma}\}$, for $x > 0$; the right end-point of F is ∞. Examples of distributions in the max-domain of attraction of such G are the Student and Pareto distributions.

- If $\boldsymbol{\gamma = 0}$, the sequences (a_n) and (b_n) can be chosen such that G has a Gumbel distribution, $G(x) = \exp\{-\exp\{-x\}\}$; the right end-point can be finite or infinite, and examples of distributions in this domain of attraction are the exponential and normal distributions.

- If $\boldsymbol{\gamma < 0}$, the sequences (a_n) and (b_n) can be chosen such that G has a reverse Weibull$(-1/\gamma)$ distribution, $G(x) = \exp\{-(-x)^{-1/\gamma}\}$, for $x < 0$; the right end-point of F is finite. The uniform or, in general, Beta distributions are examples of distributions in the max-domain of attraction of an extreme value distribution with a negative γ.

The max-domain of attraction condition (1.1.1) provides a (limit) model for the upper tail of a distribution and is used when dealing with the applications mentioned above. The estimation of γ is crucial, and several estimators of γ have been constructed. The most famous ones are the Hill (1975) estimator, the moment estimator as in Dekkers, Einmahl, and de Haan (1989), the Pickands (1975) estimator and the maximum likelihood estimator as in Smith (1987).

Inference is based on the top k (out of n) order statistics only, $k \in \{1, \ldots, n\}$, and the choice of this optimal sample fraction is a difficult issue in statistics of extremes. Several procedures for the choice of k have been suggested, see for example Dekkers and de Haan (1993); Drees and Kaufmann (1998); Danielsson, de Haan, Peng, and de Vries (2002); Gomes and Oliveira (2002).

1.2 Multivariate extreme value theory

The max-domain of attraction condition in the multivariate setting determines the limit distribution of the componentwise maxima of random vectors. Let $X_1, \dots, X_n$ be a random sample from a continuous d-variate distribution function F, $X_i = (X_{i1}, \dots, X_{id})$, $i = 1, \dots, n$. If there exist positive sequences $(a_{n,1}), \dots, (a_{n,d})$ and sequences $(b_{n,1}), \dots, (b_{n,d})$ of real numbers such that as $n \to \infty$,

$$F^n(a_{n,1}x_1 + b_{n,1}, \dots, a_{n,d}x_d + b_{n,d}) \to G(x) \tag{1.2.1}$$

for every continuity point $x := (x_1, \dots, x_d)$ of G, we say that F is in the max-domain of attraction of G, and we call G a (multivariate) extreme value distribution. Unlike in the univariate case, the family of multivariate extreme value distributions does not allow for a finite-dimensional parametrization.

Let $F_1, \dots, F_d$ be the marginals of F. The multivariate max-domain of attraction condition implies d univariate max-domain of attraction conditions for the marginal distributions, together with a max-domain of attraction condition for the dependence structure. Namely, (1.2.1) implies the existence of the limit

$$\lim_{t \downarrow 0} t^{-1} \mathbb{P} \left(1 - F_1(X_{11}) \le tx_1 \text{ or } \dots \text{ or } 1 - F_d(X_{1d}) \le tx_d\right) =: l(x), \tag{1.2.2}$$

for all $x \in [0, \infty)^d$. The function l is called the *stable tail dependence function*, and is one of the notions used to describe the tail dependence structure. The bivariate stable tail dependence function was introduced and studied in Huang (1992).

Every continuous distribution function F can be expressed in terms of its marginal distribution function and its dependence structure, for example its copula, C, as $F(x) = C(F_1(x), \dots, F_d(x))$. Using the copula, the stable tail dependence function can be written as

$$l(x) = \lim_{t \downarrow 0} t^{-1} \left(1 - C(1 - tx_1, \dots, 1 - tx_d)\right),$$

which presents the connection between the dependence structure C on one hand and the tail dependence structure described by l on the other hand. For instance, in the bivariate case, the coefficient of upper tail dependence is given by

$$\lim_{t\downarrow 0} t^{-1}\mathbb{P}\left(1 - F_1(X_{11}) \le t, 1 - F_2(X_{12}) \le t\right) = R(1,1) = 2 - l(1,1),$$

where $R(x_1, x_2) = x_1 + x_2 - l(x_1, x_2)$.

The function l is a convex function that takes values between $\max\{x_1, \ldots, x_d\}$ and $x_1+\cdots+x_d$, where the maximum of the coordinates corresponds to complete tail dependence, and the sum of the coordinates to tail independence. Also, l is homogeneous of order 1, that is $l(tx) = t\, l(x)$, for all $t > 0$ and all $x \in [0, \infty)^d$.

Another way of writing the function l is in terms of the spectral measure H by

$$l(x) = \int_{\Delta_{d-1}} \max_{j=1,\ldots,d} \{w_j x_j\} H(\mathrm{d}w),$$

where $\Delta_{d-1} = \{w \in \mathbb{R}^d : w_j \ge 0,\ w_1 + \cdots + w_d = 1\}$ is the unit simplex in $\mathbb{R}^d$ on which the measure H is defined. The spectral measure also describes the tail dependence structure and can be used interchangeably with l. Any finite Borel measure H on Δ_{d-1} that satisfies the moment conditions

$$\int_{\Delta_{d-1}} w_j H(\mathrm{d}w) = 1, \quad j = 1, \ldots, d,$$

is a spectral measure. Often it is more natural to use the spectral measure H than the function l. An example is the factor model, which will be studied in detail in Chapter 4. Here we will only introduce the "canonical" d-dimensional r-factor model $X = (X_1, \ldots, X_d)$,

$$X_j = \max_{i=1,\ldots,r} \{b_{ij} Z_i\}, \quad j = 1, \ldots, d, \tag{1.2.3}$$

where the Z_i are independent standard Fréchet random variables, and b_{ij} are nonnegative numbers such that $\sum_{i=1}^r b_{ij} = 1$ for all $j = 1, \ldots, d$. Factor models are rather general models that are used in many different areas such as psychology or, especially, finance, see for example Fama and French (1993); Malevergne and Sornette (2004) and Geluk, de Haan, and de Vries (2007), and the references therein. The spectral measure of the factor model above is discrete: it assigns mass to r atoms, and is zero everywhere else. The function l of the factor model in (1.2.3) is given by

$$l(x) = \sum_{i=1}^{r} \max_{j=1,\dots,d} \{b_{ij} x_j\},$$

and it is not differentiable. As we will show later, this lack of smoothness calls for a new estimation procedure.

In the bivariate case a nonparametric estimator of l is

$$\hat{l}_n(x_1, x_2) := \frac{1}{k} \sum_{i=1}^{n} \mathbf{1} \left\{ R_i^1 \geq n + \frac{1}{2} - kx_1 \text{ or } R_i^2 \geq n + \frac{1}{2} - kx_2 \right\}, \tag{1.2.4}$$

where $k \in \{1, \dots, n\}$ and where R_i^1 and R_i^2 are the ranks of X_{i1} and X_{i2} in the two marginal samples respectively, for $i = 1, \dots, n$. If $k = k_n$ is an intermediate sequence, $k \to \infty$ and $k/n \to 0$, then the estimator in (1.2.4) is consistent and, under some second-order conditions, asymptotically normal, see Huang (1992); Drees and Huang (1998); Einmahl, de Haan, and Li (2006).

Just like the definition of the function l itself, the definition of the above estimator is easily translated to general dimension d. However, estimation of a function in higher dimensions is not easy. In addition, the fraction of the data that can be used to estimate l is small. Therefore, it might be helpful to impose a parametric model for l. Estimation of l then reduces to estimation of the parameter vector. This is the approach followed in the thesis.

1.3 Outline of the thesis

Chapter 2. A Method of Moments Estimator of Bivariate Tail Dependence. A new estimator for the two-dimensional stable tail dependence function is introduced in Chapter 2, which corresponds to the paper Einmahl, Krajina, and Segers (2008). Assuming that the stable tail dependence function belongs to some parametric family with an unknown parameter θ from a parameter space $\Theta \subseteq \mathbb{R}^p$, we define an estimator $\hat{\theta}_n$ of θ as the solution of

$$\int_{[0,1]^2} g(x)\hat{l}_n(x)\mathrm{d}x = \int_{[0,1]^2} g(x)l(x; \hat{\theta}_n)\mathrm{d}x,$$

where $\hat{l}_n$ is the nonparametric estimator of l from (1.2.4), and $g\colon [0,1]^2 \to \mathbb{R}^p$ is an auxiliary function that we choose. Note that we do not require that F_1, F_2 or C are parametric, that is we consider a large semiparametric model.

We prove that, under mild conditions, the method of moments estimator $\hat{\theta}_n$ is consistent and asymptotically normal. These results do not rely on the continuity, not even the existence of the partial derivatives of l (with respect to x), which is the standard requirement for asymptotic normality of all other estimators of l, the nonparametric one, as well as the maximum likelihood estimators, see Coles and Tawn (1991); Joe, Smith, and Weissman (1992); Smith (1994); Ledford and Tawn (1996); de Haan, Neves, and Peng (2008). The absence of the differentiability assumption enables the estimation of tail dependence in a wider class of models. For example, we estimate the discrete two-point spectral measure corresponding to the bivariate two-factor model.

Chapter 3. A Method of Moments Estimator of Tail Dependence in Elliptical Copula Models. Elliptical distributions form a family of models that are widely used in finance and insurance, see Embrechts, McNeil, and Straumann (2002); Landsman and Valdez (2003); Kaynar, Birbil, and Frenk (2007). Bivariate elliptical distributions yield an explicit form of the function l, or equivalently, of the function R given by

$$R(x_1, x_2) = \frac{\int_{-\arcsin\rho}^{\pi/2} \min\left\{x_1(\cos\phi)^\nu, x_2(\sin(\phi + \arcsin\rho))^\nu\right\} \mathrm{d}\phi}{\int_{-\pi/2}^{\pi/2} (\cos\phi)^\nu \mathrm{d}\phi}.$$

The same function R is obtained for every distribution with the same copula. Such distributions form the class of the *elliptical copula models*, which are also known as the meta-elliptical distributions, as introduced in Fang, Fang, and Kotz (2002).

The two parameters ρ and ν in the above expression have different meanings and properties, and are therefore treated differently in the estimation procedure. Since the correlation parameter ρ depends on the whole copula, we estimate it using the whole sample. Next we plug it into the expression for R and estimate the tail parameter ν using the method of moments presented in Chapter 2.

This application is of interest because of the wide use of elliptical models. However, the implementation and the asymptotic properties are not straightforward. The estimation procedure, the asymptotic results for the estimator of (ν, ρ) and a simulation study are presented in Chapter 3, which corresponds to Krajina (2009).

Chapter 4. An M-Estimator of Tail Dependence in Arbitrary Dimensions. In Chapter 4, which is based on Einmahl, Krajina, and Segers (2009), we assume that the function l is parametric and we extend the estimator of its parameter vector θ from Chapter 2 in two directions. First, we allow for estimation in arbitrary dimension d, $d \geq 2$. Second, we use M-estimation instead of the method of moments approach. The first extension addresses an important issue in statistics of extremes, where the estimation of the tail dependence in high(er) dimensions is a challenge, since the existing estimators are either likelihood based or nonparametric. The maximum likelihood estimators can be notoriously difficult to compute due to the untractable form of the likelihood; moreover, the assumptions on the model are restrictive, including higher-order differentiability of l.

We choose a parametric approach as this enables us to impose some structure on the form of dependence and potentially reduce the estimation error. As in Chapter 2, it is important to note that the parametric assumptions are made only at the level of the tail dependence structure, resulting in a large semiparametric model. No assumptions on the marginal distributions of F were made; we do not even assume a parametric model for the copula.

Again without the requirement of differentiability, we define an M-estimator of θ. We prove consistency and asymptotic normality and present a test for a submodel within a chosen parametric model. The simulation study for some examples, including two different factor models, shows that the estimators perform well in dimensions higher than two.

Chapter 2

A Method of Moments Estimator of Bivariate Tail Dependence

[Based on joint work with J.H.J. Einmahl and J. Segers, *A Method of Moments Estimator of Tail Dependence*, Bernoulli 14(2008), 1003–1026.]

Abstract. In the world of multivariate extremes, estimation of the dependence structure presents a challenge and an interesting problem. A procedure for the bivariate case is presented that opens the road to a similar way of handling the problem in a truly multivariate setting. We consider a semiparametric model in which the stable tail dependence function is parametrically modeled. Given a random sample from a bivariate distribution function, the problem is to estimate the unknown parameter. A method of moments estimator is proposed where a certain integral of a nonparametric, rank-based estimator of the stable tail dependence function is matched with the corresponding parametric version. Under very weak conditions, the estimator is shown to be consistent and asymptotically normal. Moreover, a comparison between the parametric and nonparametric estimators leads to a goodness-of-fit test for the semiparametric model. The performance of the estimator is illustrated for a discrete spectral measure that arises in a factor-type model and for which likelihood based methods break down. A second example is that of a family of stable tail dependence functions of certain meta-elliptical distributions.

2.1 Introduction

A bivariate distribution function F with continuous marginal distribution functions F_1 and F_2 is said to have a *stable tail dependence function* l if for all $x \geq 0$ and $y \geq 0$ the following limit exists:

$$\lim_{t \to 0} t^{-1} \mathbb{P} \left(1 - F_1(X) \leq tx \text{ or } 1 - F_2(Y) \leq ty\right) = l(x, y); \tag{2.1.1}$$

see Huang (1992); Drees and Huang (1998). Here (X, Y) is a bivariate random vector with distribution F.

The relevance of condition (2.1.1) comes from multivariate extreme value theory: if F_1 and F_2 are in the max-domains of attraction of extreme value distributions G_1 and G_2 and if (2.1.1) holds, then F is in the max-domain of attraction of an extreme value distribution G with marginals G_1 and G_2 and with copula determined by l; see Section 2.2 for more details.

Inference problems on multivariate extremes therefore generally fall apart into two parts. The first one concerns the marginal distributions and is simplified by the fact that univariate extreme value distributions constitute a parametric family. The second one concerns the dependence structure in the tail of F and forms the subject of this chapter. In particular, we are interested in the estimation of the function l. The marginals will not be assumed to be known and will be estimated nonparametrically. As a consequence, the new inference procedures are rank-based and therefore invariant with respect to the marginal distribution, in accordance with (2.1.1).

The class of stable tail dependence functions does not constitute a finite-dimensional family. This is an argument for nonparametric, model-free approaches. However, the accuracy of these nonparametric approaches is often poor in higher dimensions. Moreover, stable tail dependence functions satisfy a number of shape constraints (bounds, homogeneity, convexity; see Section 2.2), which are typically not satisfied by nonparametric estimators.

The other approach is the semiparametric one, i.e. we model l parametrically. At the price of an additional model risk, parametric methods yield estimates that are always proper stable tail dependence functions. Moreover, they do not suffer from the curse of dimensionality. A large number of models have been proposed in the literature, allowing for various degrees of dependence and asymmetry, and new models continue to be invented; see Beirlant, Goegebeur, Segers, and Teugels (2004); Kotz and Nadarajah (2000) for an overview of the most common ones.

In this chapter, we propose an estimator based on the method of moments: given a parametric family $\{l(\cdot;\theta) : \theta \in \Theta\}$ with $\Theta \subseteq \mathbb{R}^p$ and a function $g : [0,1]^2 \to \mathbb{R}^p$, the moment estimator $\hat{\theta}_n$ is defined as the solution to the system of equations

$$\iint_{[0,1]^2} g(x,y) l(x,y;\hat{\theta}_n) \,\mathrm{d}x\,\mathrm{d}y = \iint_{[0,1]^2} g(x,y) \hat{l}_n(x,y) \,\mathrm{d}x\,\mathrm{d}y.$$

Here $\hat{l}_n$ is the nonparametric estimator of l. Moreover, a comparison of the parametric and nonparametric estimators yields a goodness-of-fit test for the postulated model.

The method of moments estimator is to be contrasted with the maximum likelihood estimator in point process models for extremes Coles and Tawn (1991); Joe, Smith, and Weissman (1992) or the censored likelihood approach proposed in Smith (1994); Ledford and Tawn (1996) and studied for single-parameter families in de Haan, Neves, and Peng (2008). In parametric models, moment estimators yield consistent estimators but often with a lower efficiency than the maximum likelihood estimator. However, as we shall see, the set of conditions required for the moment estimator is smaller, the conditions that remain to be imposed are much simpler, and most importantly, there are no restrictions whatsoever on the smoothness (not even on the existence) of the partial derivatives of l. Even for nonparametric estimators of l, theorems on asymptotic normality require l to be differentiable Huang (1992); Drees and Huang (1998); Einmahl, de Haan, and Li (2006).

Such a degree of generality is needed for instance if the spectral measure underlying l is discrete. In this case, there is no likelihood at all, so the maximum likelihood method breaks down. An example is the linear factor model $\boldsymbol{X} = \boldsymbol{\beta}\boldsymbol{F} + \boldsymbol{\varepsilon}$, where $\boldsymbol{X}$ and $\boldsymbol{\varepsilon}$ are 2×1 random vectors, $\boldsymbol{F}$ is a $r \times 1$ random vector of factor variables and $\boldsymbol{\beta}$ is a constant $2 \times r$ matrix of factor loadings. If the r factor variables are mutually independent and if their common marginal tail is of Pareto-type and heavier than the ones of the noise variables $\varepsilon_1, \varepsilon_2$, then the spectral measure of the distribution of $\boldsymbol{X}$ is discrete with point masses determined by $\boldsymbol{\beta}$ and the tail index of the factor variables. The heuristic is that if $\boldsymbol{X}$ is far from the origin, then with high probability it will be dominated by a single component of $\boldsymbol{F}$. Therefore, in the limit, there are only a finite number of directions for extreme outcomes of $\boldsymbol{X}$. Section 2.5 deals with a two-factor model of the above type, which gives rise to a discrete spectral measure concentrated only on two atoms. For more examples of factor models and further references see Geluk, de Haan, and de Vries (2007).

The outline of the chapter is as follows. Basic properties of stable tail dependence functions and spectral measures are reviewed in Section 2.2. The estimator and goodness-of-fit test statistic are defined in Section 2.3. Section 2.4 states the main results on the large-sample properties of the new procedures. In Section 2.5, the example of a spectral measure with two atoms is worked out, and the finite-sample performance of the moment estimator is evaluated through simulations. Section 2.6 follows the same program for the stable tail dependence functions of elliptical distributions. All proofs are deferred to Section 2.7.

2.2 Tail dependence

Let $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ be independent random vectors in $\mathbb{R}^2$ with common continuous distribution function F and marginal distribution functions F_1 and F_2. The central assumption in this chapter is the existence for all $(x, y) \in [0, \infty)^2$ of the limit l in (2.1.1). Obviously, by the probability integral transform and the inclusion-exclusion formula, (2.1.1) is equivalent to the existence for all $(x, y) \in [0, \infty]^2 \setminus \{(\infty, \infty)\}$ of the limit

$$\lim_{t \to 0} t^{-1} \mathbb{P}\left(1 - F_1(X) \le tx, 1 - F_2(Y) \le ty\right) = R(x, y), \tag{2.2.1}$$

so $R(x, \infty) = R(\infty, x) = x$. The functions l and R are related by $R(x, y) = x + y - l(x, y)$, for $(x, y) \in [0, \infty)^2$. Note that $R(1, 1)$ is the upper tail dependence coefficient.

If C denotes the copula of F, that is, if $F(x, y) = C\{F_1(x), F_2(y)\}$, then (2.1.1) is equivalent to

$$\lim_{t \to 0} t^{-1} (1 - C(1 - tx, 1 - ty)) = l(x, y) \tag{2.2.2}$$

for all $x, y \geq 0$, and also to

$$\lim_{n \to \infty} C^n(u^{1/n}, v^{1/n}) = \exp\{-l(-\log u, -\log v)\} =: C_\infty(u, v)$$

for all $(u, v) \in (0, 1]^2$. The left-hand side in the previous display is the copula of the pair of componentwise maxima $(\max_{i=1,\ldots,n} X_i, \max_{i=1,\ldots,n} Y_i)$ and the right-hand side is the copula of a bivariate max-stable distribution. If in addition the marginal distribution functions F_1 and F_2 are in the max-domains of attraction of extreme value distributions G_1 and G_2, that is, if there exist positive sequences (a_n), (c_n), and sequences $(b_n) \in \mathbb{R}$ and $(d_n) \in \mathbb{R}$ such that $F_1^n(a_n x + b_n) \xrightarrow{d} G_1(x)$ and $F_2^n(c_n y + d_n) \xrightarrow{d} G_2(y)$, then actually

$$F^n(a_n x + b_n, c_n y + d_n) \xrightarrow{d} G(x, y) = C_\infty (G_1(x), G_2(y)),$$

that is, F is in the max-domain of attraction of a bivariate extreme value distribution G with marginals G_1 and G_2 and copula C_∞. However, in this chapter we shall make no assumptions on the marginal distributions F_1 and F_2 whatsoever except for continuity.

Directly from the definition of l it follows that $x \vee y \leq l(x, y) \leq x + y$ for all $(x, y) \in [0, \infty)^2$. Similarly, $0 \leq R(x, y) \leq x \wedge y$ for $(x, y) \in [0, \infty)^2$. Moreover, the functions l and R are homogenous of order one: for all $(x, y) \in [0, \infty)^2$ and all $t > 0$,

$$l(tx, ty) = tl(x, y),$$

$$R(tx, ty) = tR(x, y).$$

In addition, l is convex and R is concave. It can be shown that these requirements on l (or, equivalently, R) are necessary and sufficient for l to be a stable tail dependence function.

The following representation of the tail dependence functions will be highly useful: there exists a finite Borel measure H on $[0,1]$, called *spectral* or *angular measure*, such that for all $(x,y) \in [0,\infty)^2$

$$\begin{aligned} l(x,y) &= \int_{[0,1]} \max\{wx,(1-w)y\} H(dw), \\ R(x,y) &= \int_{[0,1]} \min\{wx,(1-w)y\} H(dw). \end{aligned} \tag{2.2.3}$$

The identities $l(x,0) = l(0,x) = x$ for all $x \geq 0$ imply the following moment constraints for H :

$$\int_{[0,1]} wH(dw) = \int_{[0,1]} (1-w)H(dw) = 1. \tag{2.2.4}$$

Again, equation (2.2.4) constitutes a necessary and sufficient condition for l in (2.2.3) to be a stable tail dependence function. For more details on multivariate extreme value theory, see for instance Beirlant, Goegebeur, Segers, and Teugels (2004); Coles (2001); Falk, Hüsler, and Reiss (2004); Galambos (1987); de Haan and Ferreira (2006); Resnick (1987).

2.3 Estimation and testing

Let R_i^X and R_i^Y be the rank of X_i among $X_1,\dots,X_n$ and the rank of Y_i among $Y_1,\dots,Y_n$ respectively, where $i = 1,\dots,n$. Replacing $\mathbb{P}, F_1, F_2$ on the left-hand side of (2.1.1) by their empirical counterparts, we obtain a nonparametric estimator for l. Estimators obtained in this way are

$$\begin{aligned} \hat{L}_n^1(x,y) &:= \frac{1}{k}\sum_{i=1}^n \mathbf{1}\left\{R_i^X > n+1-kx \text{ or } R_i^Y > n+1-ky\right\}, \\ \hat{L}_n^2(x,y) &:= \frac{1}{k}\sum_{i=1}^n \mathbf{1}\left\{R_i^X \geq n+1-kx \text{ or } R_i^Y \geq n+1-ky\right\}, \end{aligned}$$

defined in Einmahl, de Haan, and Li (2006) and Drees and Huang (1998); Huang (1992), respectively; here $k \in \{1,\dots,n\}$. The estimator we will use here is similar to those above and is defined by

$$\hat{l}_n(x,y) := \frac{1}{k}\sum_{i=1}^{n} \mathbf{1}\left\{R_i^X > n + \frac{1}{2} - kx \text{ or } R_i^Y > n + \frac{1}{2} - ky\right\}.$$

For finite samples, simulation experiments show that the latter estimator usually performs slightly better. The large sample behaviors of the three estimators coincide however, since $\hat{L}_n^1 \leq \hat{L}_n^2 \leq \hat{l}_n$ and as $n \to \infty$

$$\sup_{0 \leq x, y \leq 1} \left| \sqrt{k} \left(\hat{l}_n(x, y) - \hat{L}_n^1(x, y) \right) \right| \leq \frac{2}{\sqrt{k}} \to 0, \tag{2.3.1}$$

where $k = k_n$ is an intermediate sequence, that is, $k \to \infty$ and $k/n \to 0$.

Assume that the stable tail dependence function l belongs to some parametric family $\{l(\,\cdot\,, \cdot\,; \theta) : \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^p$, $p \geq 1$. Observe that this does not mean that C (or F) belongs to a parametric family, i.e. we have constructed a semiparametric model. Let $g : [0, 1]^2 \to \mathbb{R}^p$ be an integrable function such that $\varphi : \Theta \to \mathbb{R}^p$ defined by

$$\varphi(\theta) := \iint_{[0,1]^2} g(x, y) l(x, y; \theta) \, \mathrm{d}x \, \mathrm{d}y \tag{2.3.2}$$

is a homeomorphism between Θ^o, the interior of the parameter space Θ, and its image $\varphi(\Theta^o)$. For examples of the function φ see Section 2.5 and Section 2.6. Let θ_0 denote the true parameter value and assume $\theta_0 \in \Theta^o$.

The method of moments estimator $\hat{\theta}_n$ of θ_0 is defined as the solution of

$$\iint_{[0,1]^2} g(x, y) \hat{l}_n(x, y) \, \mathrm{d}x \, \mathrm{d}y = \iint_{[0,1]^2} g(x, y) l(x, y; \hat{\theta}_n) \mathrm{d}x \mathrm{d}y = \varphi(\hat{\theta}_n),$$

that is,

$$\hat{\theta}_n := \varphi^{-1} \left(\iint_{[0,1]^2} g(x, y) \hat{l}_n(x, y) \, \mathrm{d}x \, \mathrm{d}y \right), \tag{2.3.3}$$

whenever the right-hand side is defined. For definiteness, if $\iint g \hat{l}_n \notin \varphi(\Theta^o)$, let $\hat{\theta}_n$ be some arbitrary, fixed value in Θ.

Consider the goodness-of-fit testing problem, $\mathcal{H}_0 : \; l \in \{l(\,\cdot\,, \cdot\,; \theta) : \theta \in \Theta\}$ against $\mathcal{H}_a : \; l \notin \{l(\,\cdot\,, \cdot\,; \theta) : \theta \in \Theta\}$. We propose the test statistic

$$\iint_{[0,1]^2} \left(\hat{l}_n(x, y) - l(x, y; \hat{\theta}_n) \right)^2 \mathrm{d}x \, \mathrm{d}y, \tag{2.3.4}$$

with $\hat{\theta}_n$ as in (2.3.3). The null hypothesis is rejected for large values of the test statistic.

2.4 Results

The method of moments estimator is consistent for every intermediate sequence $k = k_n$ under minimal conditions on the model and the function g.

Theorem 2.4.1 (Consistency)**.** *Let $g : [0,1]^2 \to \mathbb{R}^p$ be integrable. If φ in* (2.3.2) *is a homeomorphism between Θ^o and $\varphi(\Theta^o)$ and if $\theta_0 \in \Theta^o$, then as $n \to \infty$, $k \to \infty$ and $k/n \to 0$, the right-hand side of* (2.3.3) *is well-defined with probability tending to one and $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$.*

Denote by W a mean-zero Wiener process on $[0,\infty]^2 \setminus \{(\infty,\infty)\}$ with covariance function

$$\mathbb{E}W(x_1,y_1)W(x_2,y_2) = R(x_1 \wedge x_2, y_1 \wedge y_2),$$

and for $x, y \in [0,\infty)$ denote

$$W_1(x) := W(x,\infty), \qquad W_2(y) := W(\infty,y).$$

Further, for $(x,y) \in [0,\infty)^2$ let $R_1(x,y)$ and $R_2(x,y)$ be the right-hand partial derivatives of R at the point (x,y) with respect to the first and second coordinate, respectively. Since R is concave, R_1 and R_2 defined in this way always exist, although they are discontinuous at points where $\frac{\partial}{\partial x}R(x,y)$ or $\frac{\partial}{\partial y}R(x,y)$ do not exist.

Finally, define the stochastic process B on $[0,\infty)^2$ and the p-variate random vector $\tilde{B}$ by

$$\begin{aligned} B(x,y) &= W(x,y) - R_1(x,y)W_1(x) - R_2(x,y)W_2(y), \\ \tilde{B} &= \iint_{[0,1]^2} g(x,y)B(x,y)\,\mathrm{d}x\,\mathrm{d}y. \end{aligned}$$

Theorem 2.4.2 (Asymptotic Normality)**.** *In addition to the conditions in Theorem 2.4.1, assume:*

(C1) The function φ is continuously differentiable in some neighborhood of θ_0 and its derivative matrix $D_\varphi(\theta_0)$ is invertible.

(C2) There exists $\alpha > 0$ such that as $t \to 0$,

$$t^{-1}\mathbb{P}\left(1 - F_1(X) \leq tx, 1 - F_2(Y) \leq ty\right) - R(x, y) = O(t^\alpha),$$

uniformly on the set $\{(x, y) : x + y = 1, x \geq 0, y \geq 0\}$.

(C3) $k = k_n \to \infty$ and $k = o(n^{2\alpha/(1+2\alpha)})$ as $n \to \infty$.

Then

$$\sqrt{k}(\hat{\theta}_n - \theta_0) \xrightarrow{d} D_\varphi(\theta_0)^{-1}\tilde{B}. \tag{2.4.1}$$

Note that condition (C2) is a second-order condition quantifying the speed of convergence in (2.2.1). Condition (C3) gives an upper bound on the speed with which k can grow to infinity. This upper bound is related to the speed of convergence in (C2) and ensures that $\hat{\theta}_n$ is asymptotically unbiased.

The limiting distribution in (2.4.1) depends on the model and on the auxiliary function g. The optimal g would be the one minimizing the asymptotic variance, but this minimization problem is typically difficult to solve. In the examples in Section 2.5 and Section 2.6 the functions g were chosen so as to simplify the calculations.

From the definition of the process B it follows that the distribution of $\tilde{B}$ is p-variate normal with mean zero and covariance matrix

$$\Sigma(\theta_0) = \mathrm{Var}(\tilde{B}) = \iiiint_{[0,1]^4} g(x, y)g(u, v)^\top \sigma(x, y, u, v; \theta_0)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}u\, \mathrm{d}v, \tag{2.4.2}$$

where σ is the covariance function of the process B, that is, for $\theta \in \Theta$,

$$\begin{aligned}\sigma(x, y, u, v; \theta) &= \mathbb{E}B(x, y)B(u, v) \\ &= R(x \wedge u, y \wedge v; \theta) + R_1(x, y; \theta)R_1(u, v; \theta)(x \wedge u) \\ &\quad + R_2(x, y; \theta)R_2(u, v; \theta)(y \wedge v) - 2R_1(u, v; \theta)R(x \wedge u, y; \theta) \\ &\quad - 2R_2(u, v; \theta)R(x, y \wedge v; \theta) + 2R_1(x, y; \theta)R_2(u, v; \theta)R(x, v; \theta). \end{aligned} \tag{2.4.3}$$

Denote by H_θ the spectral measure corresponding to $l(\cdot,\cdot;\theta)$. The following corollary allows the construction of confidence regions.

Corollary 2.4.3. *Under the assumptions of Theorem 2.4.2, if the map $\theta \mapsto H_\theta$ is weakly continuous at θ_0 and if $\Sigma(\theta_0)$ is non-singular, then as $n \to \infty$,*

$$k(\hat{\theta}_n - \theta_0)^\top D_\varphi(\hat{\theta}_n)^\top \Sigma(\hat{\theta}_n)^{-1} D_\varphi(\hat{\theta}_n)(\hat{\theta}_n - \theta_0) \xrightarrow{d} \chi^2_p.$$

Finally, we derive the limit distribution of the test statistic in (2.3.4).

Theorem 2.4.4 (Test)**.** *Assume that the null hypothesis $\mathcal{H}_0$ holds and let $\theta_{\mathcal{H}_0}$ denote the true parameter. If*

1) *for all $\theta_0 \in \Theta$ the conditions of Theorem 2.4.2 are satisfied (and hence Θ is open);*

2) *on Θ, the mapping $\theta \mapsto l(x, y; \theta)$ is differentiable for all $(x, y) \in [0, 1]^2$, and its gradient is bounded in $(x, y) \in [0, 1]^2$,*

then

$$\begin{aligned}
&\iint_{[0,1]^2} k\left(\hat{l}_n(x, y) - l(x, y; \hat{\theta}_n)\right)^2 dx\, dy \\
&\xrightarrow{d} \iint_{[0,1]^2} \left(B(x, y) - D_{l(x,y;\theta)}(\theta_{\mathcal{H}_0}) D_\varphi(\theta_{\mathcal{H}_0})^{-1} \tilde{B}\right)^2 dx\, dy
\end{aligned}$$

as $n \to \infty$, where $D_{l(x,y;\theta)}(\theta_{\mathcal{H}_0})$ is the gradient of $\theta \mapsto l(x, y; \theta)$ at $\theta_{\mathcal{H}_0}$.

2.5 Example 1: Two-point spectral measure

The *two-point spectral measure* is a spectral measure H that is concentrated on only two points in $(0, 1) \setminus \{1/2\}$, call them a and $1 - b$. The moment conditions (2.2.4) imply that one of those points is less than $1/2$ and the other one is greater than $1/2$, and the masses on those points are determined by their locations. For definiteness, let $a \in (0, 1/2)$ and $1 - b \in (1/2, 1)$, so the parameter vector

$\theta = (a, b)$ takes values in the square $\Theta = (0, 1/2)^2$. The masses assigned to a and $1 - b$ are

$$q := H(\{a\}) = \frac{1 - 2b}{1 - a - b} \text{ and } 2 - q = H(\{1 - b\}) = \frac{1 - 2a}{1 - a - b}.$$

This model is also known as the *natural model* and was first described in Tiago de Oliveira (1980, 1989).

By (2.2.3), the corresponding stable tail dependence function is

$$l(x, y; a, b) = q \max\{ax, (1 - a)y\} + (2 - q) \max\{(1 - b)x, by\}.$$

The partial derivatives of l with respect to x and y are

$$\frac{\partial l(x, y; a, b)}{\partial x} = \begin{cases} 1 & \text{if } y < \frac{a}{1-a}x, \\ (1 - b)(2 - q) & \text{if } \frac{a}{1-a}x < y < \frac{1-b}{b}x, \\ 0 & \text{if } y > \frac{1-b}{b}x, \end{cases}$$

and $(\partial/\partial y)l(x, y; a, b) = (\partial/\partial y)l(y, x; b, a)$. Note that the partial derivatives do not exist on the lines $y = \frac{a}{1-a}x$ and $y = \frac{1-b}{b}x$. The same is true for the partial derivatives of R. As a consequence, the maximum likelihood method is not applicable and the asymptotic normality of the nonparametric estimator breaks down. However, the method of moments estimator can still be used since in Theorem 2.4.2 no smoothness assumptions are made on l whatsoever.

As explained in the introduction, discrete spectral measures arise whenever extremes are determined by a finite number of independent, heavy tailed factors. Specifically, let the random vector (X, Y) be given by

$$(X, Y) = (\alpha Z_1 + (1 - \alpha)Z_2 + \varepsilon_1, (1 - \beta)Z_1 + \beta Z_2 + \varepsilon_2), \tag{2.5.1}$$

where $0 < \alpha < 1$ and $0 < \beta < 1$ are coefficients and where Z_1, Z_2, ε_1 and ε_2 are independent random variables satisfying the following conditions: there exist $\nu > 0$ and a slowly varying function L such that $\mathbb{P}(Z_i > z) = z^{-\nu}L(z)$ for some $\nu > 0$, $i = 1, 2$; $\mathbb{P}(\varepsilon_j > z)/\mathbb{P}(Z_1 > z) \to 0$ as $z \to \infty$, $j = 1, 2$. (Recall that a positive, measurable function L defined in a neighborhood of infinity is called slowly varying if $L(yz)/L(z) \to 1$ as $z \to \infty$ for all $y > 0$.) Straightforward but lengthy computations show that the spectral measure of the random vector (X, Y) is a two-point spectral measure having masses q and $2 - q$ at the points a and $1 - b$, where

$$\begin{aligned} q &:= \frac{(1-\alpha)^\nu}{\alpha^\nu + (1-\alpha)^\nu} + \frac{\beta^\nu}{\beta^\nu + (1-\beta)^\nu} \\ a &:= \frac{(1-\alpha)^\nu}{\alpha^\nu + (1-\alpha)^\nu} q^{-1}, \\ 1-b &:= \frac{\alpha^\nu}{\alpha^\nu + (1-\alpha)^\nu} (2-q)^{-1}. \end{aligned}$$

Write $\Delta = \{(x, y) \in [0, 1]^2 : x + y \leq 1\}$ and let $\mathbf{1}_\Delta$ be its indicator function. The function $g_\Delta : [0, 1]^2 \to \mathbb{R}^2$ defined by $g_\Delta(x, y) = \mathbf{1}_\Delta(x, y)(x, y)^\top$ is obviously integrable, and the function φ in (2.3.2) is given by

$$\varphi(a, b) = \iint_\Delta (x, y)^\top l(x, y; a, b) \,\mathrm{d}x \,\mathrm{d}y = (J(a, b), K(a, b))^\top$$

where $K(a, b) = J(b, a)$ and

$$J(a, b) = \frac{1}{24}\{(2ab - a - b)(b - a + 1) + a(b - 1) + 3\}.$$

Nonparametric estimators of J and K are given by

$$(\hat{J}_n, \hat{K}_n) = \iint_\Delta (x, y)^\top \hat{l}_n(x, y) \,\mathrm{d}x \,\mathrm{d}y,$$

and the method of moment estimators $(\hat{a}_n, \hat{b}_n)$ are defined as the solutions to the equations

$$(\hat{J}_n, \hat{K}_n) = (J(\hat{a}_n, \hat{b}_n), K(\hat{a}_n, \hat{b}_n)).$$

Due to the explicit nature of the functions J and K, these equations can be simplified: if we denote $c_{J,n} := 3(8\hat{J}_n - 1)$ and $c_{K,n} := 3(8\hat{K}_n - 1)$, the estimator $\hat{b}_n$ of b will be a solution of the quadratic equation

$$3(2c_{J,n} + 2c_{K,n} + 3)b^2 + 3(-5c_{J,n} + c_{K,n} - 3)b + 3c_{J,n} - 6c_{K,n} - (c_{J,n} + c_{K,n})^2 = 0$$

that falls into the interval $(0, 1/2)$, and the estimator of a is

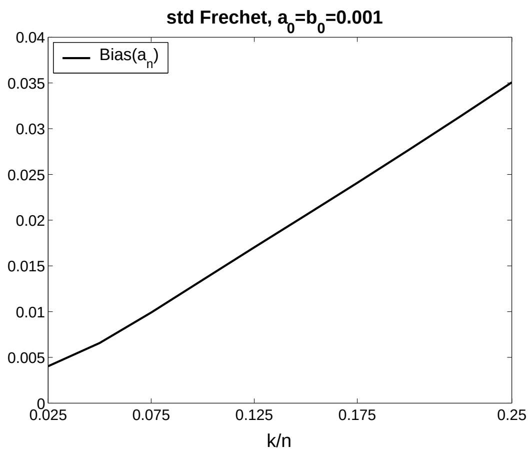
$$\hat{a}_n = \frac{3\hat{b}_n + c_{J,n} + c_{K,n}}{6\hat{b}_n - 3}.$$

In the simulations we used the following models:

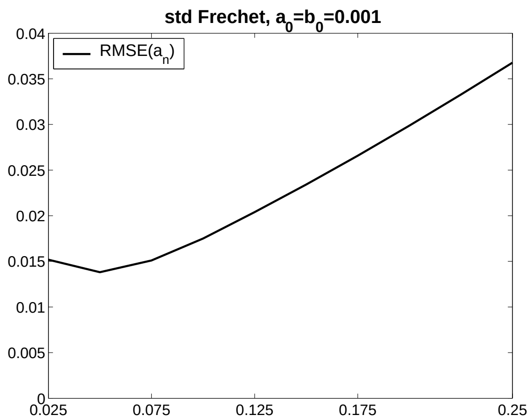
(i) $Z_1, Z_2 \sim$ Fréchet(1), so $\nu = 1$, and $\varepsilon_1, \varepsilon_2 \sim N(0, 1)$ (Figures 2.1, 2.2, 2.3);

(ii) $Z_1, Z_2 \sim t_2$, so $\nu = 1/2$, and $\varepsilon_1, \varepsilon_2 \sim N(0, 0.5^2)$ (Figures 2.4, 2.5, 2.6).

The figures show the bias and the root mean squared error (RMSE) of $\hat{a}_n$ and $\hat{b}_n$ for 1000 samples of size $n = 1000$. The method of moments estimator performs well in general. We see a very good behavior when $a_0 = b_0 \approx 0$. Of course, the heavier the tail of Z_1, the better the performance of the estimator.

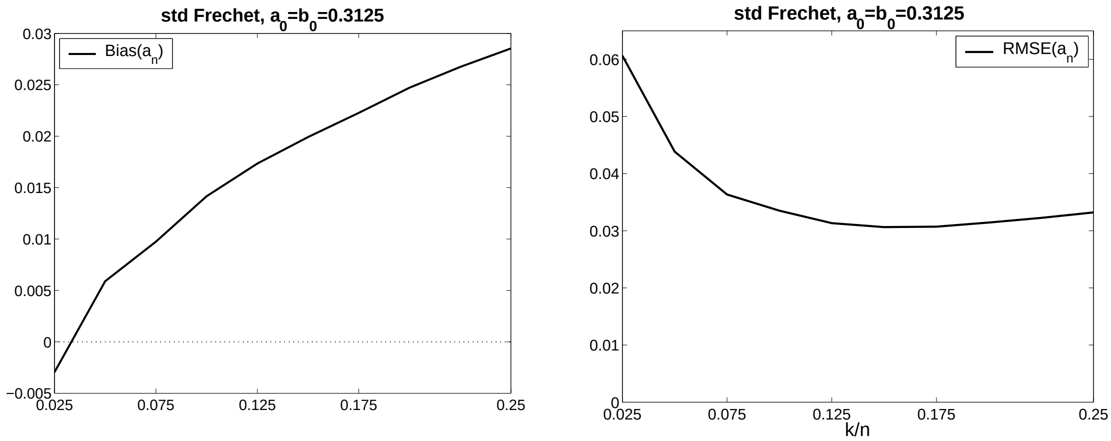


(a) Bias of estimator of a



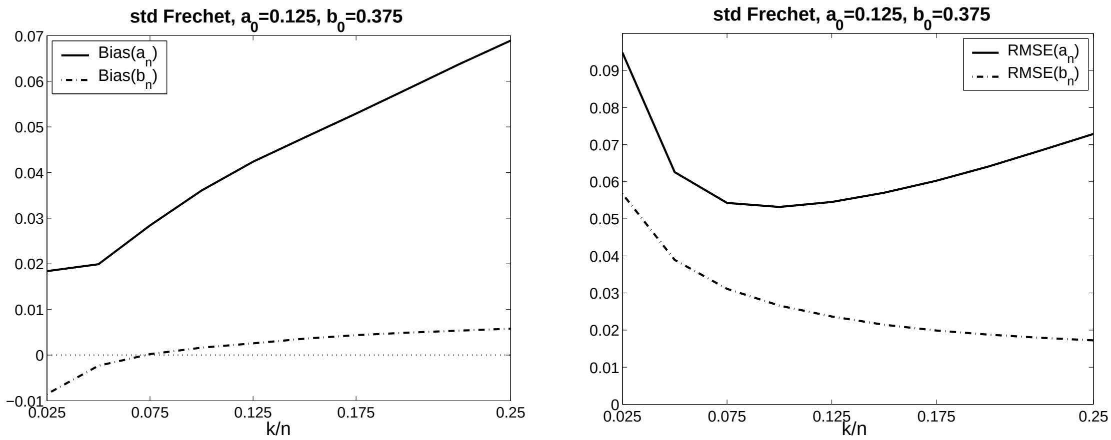
(b) RMSE of estimator of a

Figure 2.1: Model (2.5.1) with $Z_1, Z_2 \sim$ Fréchet(1), $\varepsilon_1, \varepsilon_2 \sim N(0, 1)$, $a_0 = b_0 = 0.001$



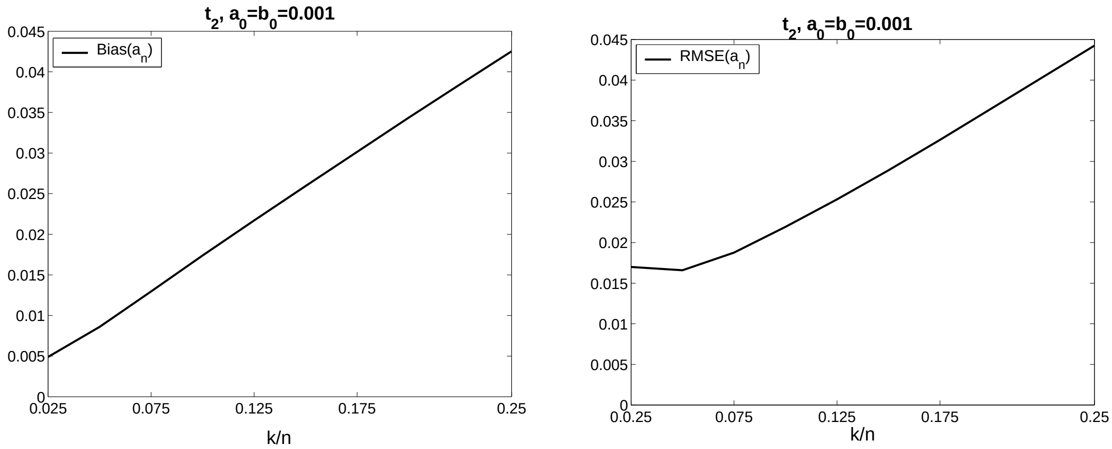
(a) Bias of estimator of a **(b)** RMSE of estimator of a

Figure 2.2: Model (2.5.1) with $Z_1, Z_2 \sim$ Fréchet(1), $\varepsilon_1, \varepsilon_2 \sim N(0,1)$, $a_0 = b_0 = 0.3125$



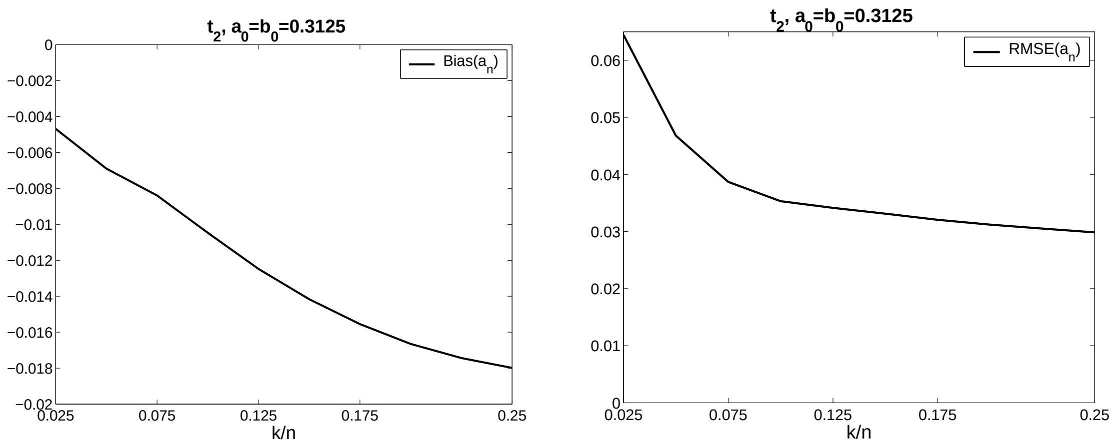
(a) Bias of estimators of a and b **(b)** RMSE of estimators of a and b

Figure 2.3: Model (2.5.1) with $Z_1, Z_2 \sim$ Fréchet(1), $\varepsilon_1, \varepsilon_2 \sim N(0,1)$, $a_0 = 0.125$, $b_0 = 0.375$



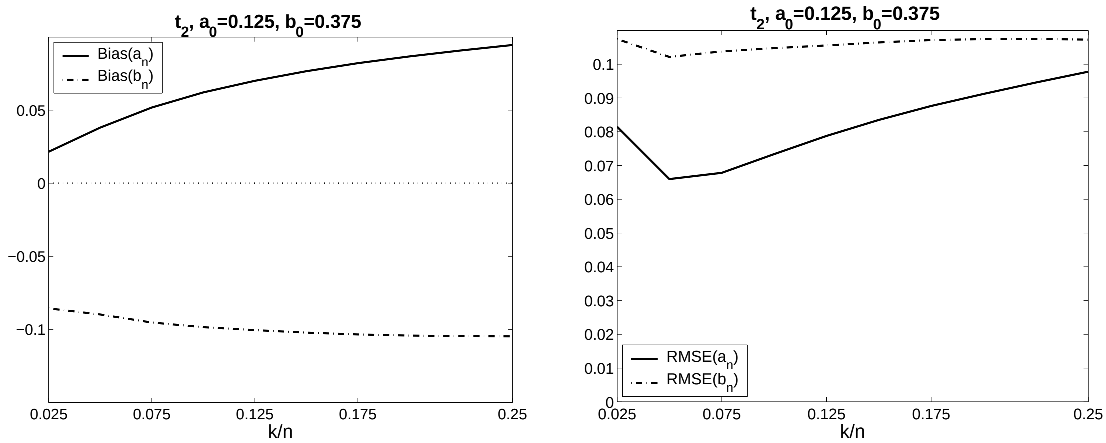
(a) Bias of estimator of a **(b)** RMSE of estimator of a

Figure 2.4: Model (2.5.1) with $Z_1, Z_2 \sim t_2$, $\varepsilon_1, \varepsilon_2 \sim N(0, 0.5^2)$, $a_0 = b_0 = 0.001$



(a) Bias of estimator of a **(b)** RMSE of estimator of a

Figure 2.5: Model (2.5.1) with $Z_1, Z_2 \sim t_2$, $\varepsilon_1, \varepsilon_2 \sim N(0, 0.5^2)$, $a_0 = b_0 = 0.3125$



(a) Bias of estimators of a and b **(b)** RMSE of estimators of a and b

Figure 2.6: Model (2.5.1) with $Z_1, Z_2 \sim t_2$, $\varepsilon_1, \varepsilon_2 \sim N(0, 0.5^2)$, $a_0 = 0.125$, $b_0 = 0.375$

2.6 Example 2: Parallel meta-elliptical model

A random vector (X, Y) is said to be *elliptically distributed* if it satisfies the distributional equality

$$(X, Y)^\top \stackrel{d}{=} \boldsymbol{\mu} + Z\boldsymbol{A}\boldsymbol{U}, \tag{2.6.1}$$

where $\boldsymbol{\mu}$ is a 2×1 column vector, Z is a positive random variable called generating random variable, $\boldsymbol{A}$ is a 2×2 matrix such that $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^\top$ is of full rank, and $\boldsymbol{U}$ is a two-dimensional random vector independent of Z and uniformly distributed on the unit circle $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$. Under the above assumptions, the matrix $\boldsymbol{\Sigma}$ can be written as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \rho\sigma v \\ \rho\sigma v & v^2 \end{pmatrix} \tag{2.6.2}$$

where $\sigma > 0$, $v > 0$, and $-1 < \rho < 1$. The special case $\rho = 0$ yields the subclass of *parallel elliptical distributions*.

By Hult and Lindskog (2002), the distribution of Z satisfies $\mathbb{P}(Z > z) = z^{-\nu} L(z)$ with $\nu > 0$ and L slowly varying if and only if the distribution of (X, Y) is (multivariate) regularly varying with the same index. Under this assumption, the function R of the distribution of (X, Y) was derived in Klüppelberg, Kuhn, and Peng (2007). In case $\rho = 0$, the formula specializes to

$$R(x, y; \nu) = \frac{x \int_{f(x,y;\nu)}^{\pi/2} (\cos\phi)^\nu \mathrm{d}\phi + y \int_0^{f(x,y;\nu)} (\sin\phi)^\nu \mathrm{d}\phi}{\int_{-\pi/2}^{\pi/2} (\cos\phi)^\nu \mathrm{d}\phi}, \tag{2.6.3}$$

with $f(x, y; \nu) = \arctan\{(x/y)^{1/\nu}\}$. Hence, the class of stable tail dependence functions belonging to parallel elliptical vectors with regularly varying generating random variables forms a one-dimensional parametric family indexed by the index of regular variation $\nu \in (0, \infty) = \Theta$ of Z. We will call the corresponding stable tail dependence functions l *parallel elliptical*.

In Fang, Fang, and Kotz (2002), *meta-elliptical distributions* are defined as the distributions of random vectors of the form $(s(X), t(Y))$, where the distribution of (X, Y) is elliptical and s and t are increasing functions. In other words, a distribution is meta-elliptical if and only if its copula is that of an elliptical distribution. Such copulas are called meta-elliptical in Genest, Favre, Béliveau, and Jacques (2007); note that a copula, as a distribution function on the unit square, cannot be elliptical in the sense of (2.6.1). Since a stable tail dependence function l of a bivariate distribution F is only determined by F through its copula C, see (2.2.2), the results in the preceding paragraph continue to hold for meta-elliptical distributions. In case $\rho = 0$, we will speak of parallel meta-elliptical distributions. In case the generating random variable Z is regularly varying with index ν, the function R is given by (2.6.3).

For parallel meta-elliptical distributions, the second-order condition (C2) in Theorem 2.4.2 can be checked via second-order regular variation of Z.

Lemma 2.6.1. *Let F be a parallel meta-elliptical distribution with generating random variable Z. If there exist $\nu > 0$, $\beta < 0$ and a function $A(t) \to 0$ of constant sign near infinity such that*

$$\lim_{t\to\infty} \frac{\mathbb{P}(Z > tx)/\mathbb{P}(Z > t) - x^{-\nu}}{A(t)} = x^{-\nu} \frac{x^\beta - 1}{\beta}, \tag{2.6.4}$$

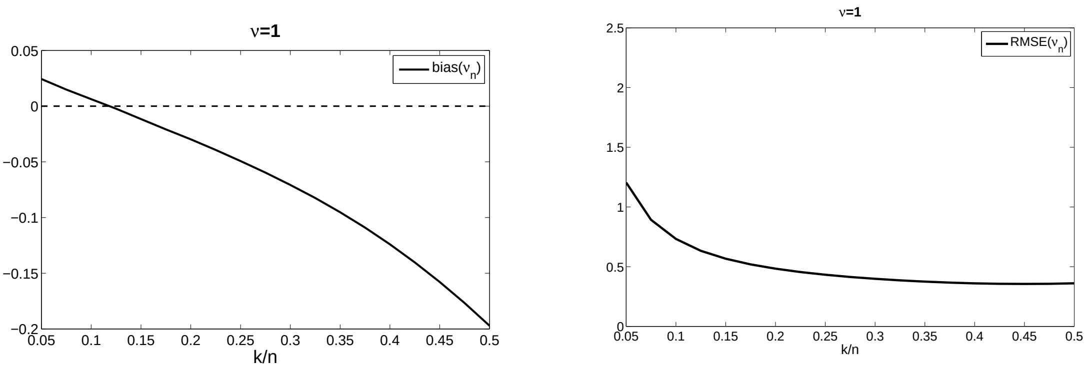
then condition (C2) in Theorem 2.4.2 holds for every $\alpha \in (0, -\beta/\nu)$.

Note that although the generating random variable is only defined up to a multiplicative constant, condition (2.6.4) does makes sense: that is, if (2.6.4) holds for a random variable Z, then it also holds for cZ with $c > 0$, for the same constants ν and β and for the rate function $A^*(t) := A(t/c)$. Note that $|A|$ is necessarily regularly varying with index β, see equation (3.0.3) in Bingham, Goldie, and Teugels (1987).

Now assume that $(X_1, Y_1), \ldots, (X_n, Y_n)$ is a random sample from a bivariate distribution F with parallel elliptical stable tail dependence function l, that is $l \in \{l(\,\cdot\,,\,\cdot\,;\nu) : \nu \in (0,\infty)\}$, where $l(x, y; \nu) = x + y - R(x, y; \nu)$ and $R(x, y; \nu)$ is as in (2.6.3). We will apply the method of moments to estimate the parameter ν. Since l is defined by a limit relation, our assumption on F is weaker than the assumption that F is parallel meta-elliptical with regularly varying Z, which, as explained above, is in turn weaker than the assumption that F itself is parallel elliptical with regularly varying Z. The problem of estimating the R for elliptical distributions was addressed in Klüppelberg, Kuhn, and Peng (2007) and for meta-elliptical distributions in Klüppelberg, Kuhn, and Peng (2008).

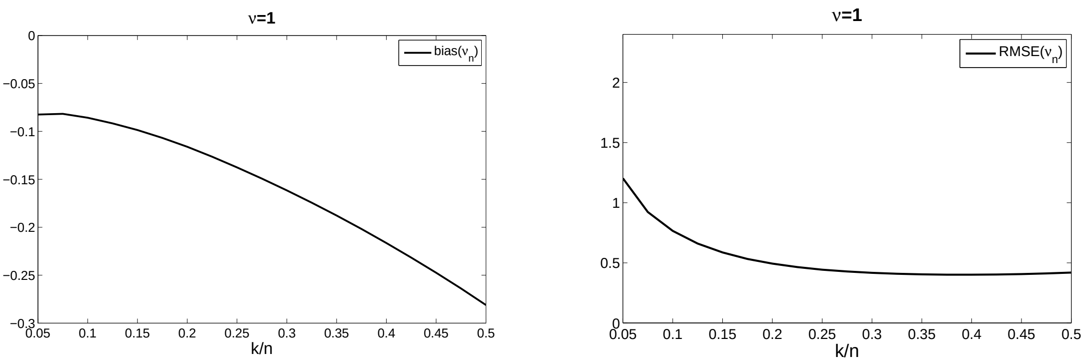
We simulated 1000 random samples of size $n = 1000$ from models for which the assumptions of Theorem 2.4.2 hold, and which have the function $R(\,\cdot\,,\,\cdot\,;\nu)$ as in (2.6.3), with $\nu \in \{1, 5\}$. The three models we used are of the type $(X_1, Y_1)^\top = Z\boldsymbol{U}$. In the first model the generating random variable Z is such that $\mathbb{P}(Z > z) = (1+z^2)^{-1/2}$ for $z \geq 0$, that is the first model is the bivariate Cauchy ($\nu = 1$). In the other two models Z is Fréchet(ν) with $\nu \in \{1, 5\}$.

Figures 2.7, 2.8, 2.9 show the bias and the RMSE of the moment estimator of ν. The auxiliary function $g : [0,1]^2 \to \mathbb{R}$ is $g(x, y) = \mathbf{1}\{x + y \leq 1\}$. For comparison, Figure 2.10 and Figure 2.11 show the plots of the means and RMSE of the parametric and nonparametric estimates $R(1, 1; \hat{\nu}_n)$ and $\hat{R}_n(1, 1) = 2 - \hat{l}_n(1, 1)$ of the upper tail dependence coefficient $R(1, 1)$. We can see that the method of moments estimator of the upper tail dependence coefficient $R(1, 1; \nu)$ performs well. In particular, it is much less sensitive to the choice of k than the nonparametric estimator.



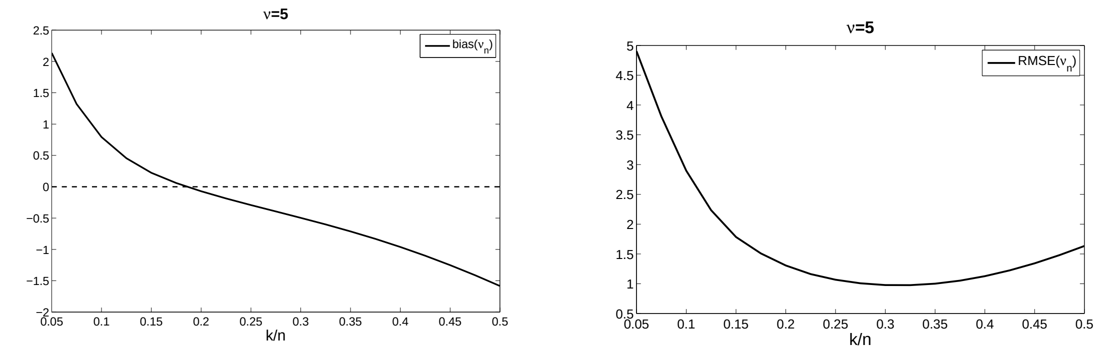
(a) Bias of estimator of ν **(b)** RMSE of estimator of ν

Figure 2.7: Estimation of $\nu = 1$ in the bivariate Cauchy model.



(a) Bias of estimator of ν **(b)** RMSE of estimator of ν

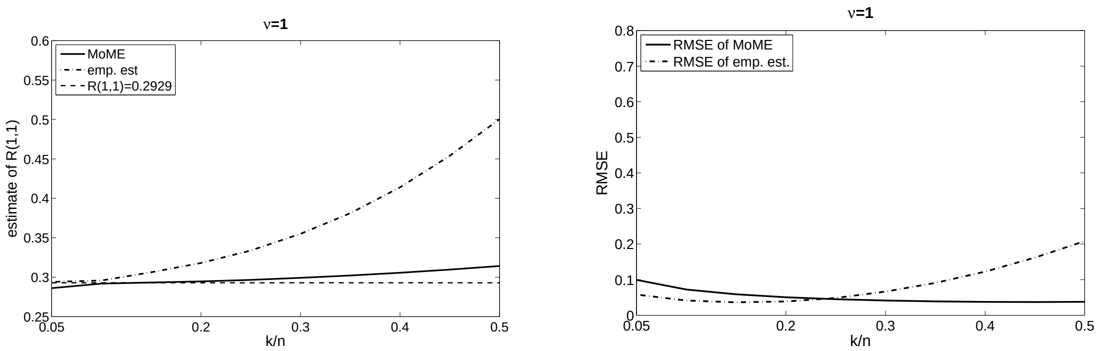
Figure 2.8: Estimation of $\nu = 1$ in the model $(X_1, Y_1)^\top = Z\boldsymbol{U}$, where Z is Fréchet(1).



(a) Bias of estimator of ν

(b) RMSE of estimator of ν

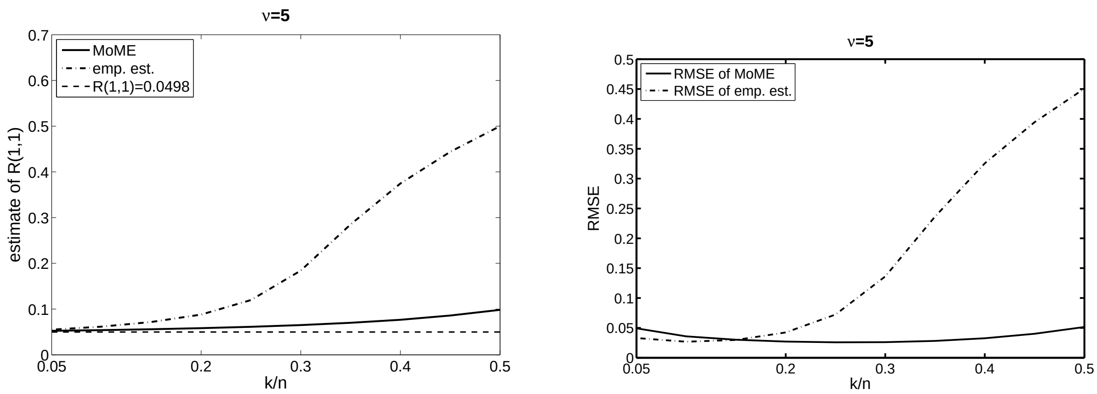
Figure 2.9: Estimation of $\nu = 5$ in the model $(X_1, Y_1)^\top = Z\boldsymbol{U}$, where Z is Fréchet(5).



(a) the mean over 1000 replications of the estimate of $R(1,1;1)$

(b) the RMSE of the estimator of $R(1,1;1)$

Figure 2.10: Estimation of $R(1,1;1)$ in the bivariate Cauchy model.



(a) the mean over 1000 replications of the estimate of $R(1,1;5)$

(b) the RMSE of the estimator of $R(1,1;5)$

Figure 2.11: Estimation of $R(1,1;5)$ in the model $(X_1, Y_1)^\top = Z\boldsymbol{U}$, where Z is Fréchet(5).

2.7 Proofs

Proof of Theorem 2.4.1 First note that

$$\left| \iint_{[0,1]^2} g(x,y)\hat{l}_n(x,y)\,\mathrm{d}x\,\mathrm{d}y - \iint_{[0,1]^2} g(x,y)l(x,y;\theta_0)\,\mathrm{d}x\,\mathrm{d}y \right|$$
$$\leq \sup_{0\leq x,y\leq 1} |\hat{l}_n(x,y) - l(x,y;\theta_0)| \iint_{[0,1]^2} |g(x,y)|\,\mathrm{d}x\,\mathrm{d}y.$$

The second term is finite by assumption, and

$$\sup_{0\leq x,y\leq 1} |\hat{l}_n(x,y) - l(x,y;\theta_0)| \xrightarrow{\mathbb{P}} 0$$

by (2.3.1) and Theorem 1 in Huang (1992), see also Drees and Huang (1998). Therefore, as $n \to \infty$,

$$\iint_{[0,1]^2} g(x,y)\hat{l}_n(x,y)\,\mathrm{d}x\,\mathrm{d}y \xrightarrow{\mathbb{P}} \iint_{[0,1]^2} g(x,y)l(x,y;\theta_0)\,\mathrm{d}x\,\mathrm{d}y = \varphi(\theta_0).$$

Since $\varphi(\theta_0) \in \varphi(\Theta^o)$, which is open, and since φ^{-1} is continuous at $\varphi(\theta_0)$ by assumption, we can apply the function φ^{-1} on both sides of the previous limit relation, so that, by the continuous mapping theorem, indeed $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$. □

For the proof of Theorem 2.4.2 we will need the following lemma, the proof of which follows from Lemma 6.2.1 in Falk, Hüsler, and Reiss (2004).

Lemma 2.7.1. *The function R in* (2.2.3) *is differentiable at $(x,y) \in (0,\infty)^2$ if $H(\{z\}) = 0$ with $z = y/(x+y)$. In that case, the gradient of R is given by $(R_1(x,y), R_2(x,y))^\top$, where*

$$R_1(x,y) = \int_0^z wH(dw), \qquad R_2(x,y) = \int_z^1 (1-w)H(dw). \tag{2.7.1}$$

For $i = 1,\ldots,n$ denote $U_i := 1 - F_1(X_i)$ and $V_i := 1 - F_2(Y_i)$. Let Q_{1n} and Q_{2n} denote the empirical quantile functions of $(U_1,\ldots,U_n)$ and $(V_1,\ldots,V_n)$ respectively, that is

$$Q_{1n}\left(\frac{kx}{n}\right) = U_{\lceil kx\rceil:n}, \qquad Q_{2n}\left(\frac{ky}{n}\right) = V_{\lceil ky\rceil:n},$$

where $U_{1:n} \leq \cdots \leq U_{n:n}$ and $V_{1:n} \leq \cdots \leq V_{n:n}$ are the order statistics and where $\lceil a \rceil$ is the smallest integer not smaller than a. Denote

$$S_{1n}(x) := \frac{n}{k} Q_{1n}\left(\frac{kx}{n}\right), \quad S_{2n}(y) := \frac{n}{k} Q_{2n}\left(\frac{ky}{n}\right)$$

and define

$$\begin{aligned}
\hat{R}_n^1(x,y) &:= \frac{1}{k}\sum_{i=1}^{n} \mathbf{1}\left\{U_i < \frac{k}{n}S_{1n}(x),\ V_i < \frac{k}{n}S_{2n}(y)\right\}, \\
&= \frac{1}{k}\sum_{i=1}^{n} \mathbf{1}\left\{U_i < U_{\lceil kx \rceil:n},\ V_i < V_{\lceil ky \rceil:n}\right\}, \\
&= \frac{1}{k}\sum_{i=1}^{n} \mathbf{1}\left\{R_i^X > n+1-kx,\ R_i^Y > n+1-ky\right\}, \\
R_n(x,y) &:= \frac{n}{k}\mathbb{P}\left(U_1 \leq \frac{kx}{n}, V_1 \leq \frac{ky}{n}\right), \\
T_n(x,y) &:= \frac{1}{k}\sum_{i=1}^{n} \mathbf{1}\left\{U_i < \frac{kx}{n},\ V_i < \frac{ky}{n}\right\}.
\end{aligned}$$

Further, note that

$$\hat{R}_n^1(x,y) = T_n\left(S_{1n}(x), S_{2n}(y)\right).$$

Write $v_n(x,y) = \sqrt{k}\left(T_n(x,y) - R_n(x,y)\right),\ v_{n,1}(x) := v_n(x,\infty)$ and $v_{n,2}(y) := v_n(\infty, y)$. From Proposition 3.1 in Einmahl, de Haan, and Li (2006) we get

$$\begin{aligned}
&(v_n(x,y), x, y \in [0,1]; v_{n,1}(x), x \in [0,1]; v_{n,2}(y), y \in [0,1]) \\
&\xrightarrow{d} (W(x,y), x, y \in [0,1]; W_1(x), x \in [0,1]; W_2(y), y \in [0,1])
\end{aligned}$$

in the topology of uniform convergence, as $n \to \infty$. Invoking the Skorohod construction, see for instance van der Vaart and Wellner (1996), we get a new probability space containing all $\tilde{v}_n, \tilde{v}_{n,1}, \tilde{v}_{n,2}, \tilde{W}, \tilde{W}_1, \tilde{W}_2$ for which it holds

$$(\tilde{v}_n, \tilde{v}_{n,1}, \tilde{v}_{n,2}) \stackrel{d}{=} (v_n, v_{n,1}, v_{n,2}),$$

$$(\tilde{W}, \tilde{W}_1, \tilde{W}_2) \stackrel{d}{=} (W, W_1, W_2)$$

as well as

$$\sup_{0\le x,y\le 1} |\tilde{v}_n(x,y) - \tilde{W}(x,y)| \overset{a.s.}{\to} 0,$$

$$\sup_{0\le x\le 1} |\tilde{v}_{n,j}(x) - \tilde{W}_j(x)| \overset{a.s.}{\to} 0,\ j=1,2.$$

We will work on this space from now on, but keep the old notation (without tildes). The following consequence of the above and Vervaat's lemma (Vervaat (1972)), will be useful

$$\sup_{0\le x\le 1} |\sqrt{k}(S_{jn}(x) - x) + W_j(x)| \overset{a.s.}{\to} 0,\ j=1,2. \tag{2.7.2}$$

Proof of Theorem 2.4.2 In this proof we will write $l(x,y)$ and $R(x,y)$ instead of $l(x,y;\theta_0)$ and $R(x,y;\theta_0)$, respectively.
First we will show that as $n\to\infty$,

$$\left|\sqrt{k}\left(\iint_{[0,1]^2} g(x,y)\hat{L}_n^1(x,y)\mathrm{d}x\mathrm{d}y - \varphi(\theta_0)\right) + \tilde{B}\right| \overset{\mathbb{P}}{\to} 0. \tag{2.7.3}$$

Since for each $x,y\in(0,1]$

$$\left(\hat{L}_n^1 + \hat{R}_n^1\right)(x,y) = \frac{\lceil kx\rceil + \lceil ky\rceil - 2}{k}$$

almost surely, from

$$\left|\frac{\lceil kx\rceil + \lceil ky\rceil - 2}{k} - x - y\right| \le \frac{2}{k}$$

it follows that

$$\begin{aligned}&\left|\sqrt{k}\left(\iint_{[0,1]^2} g(x,y)\hat{L}_n^1(x,y)\mathrm{d}x\mathrm{d}y - \iint_{[0,1]^2} g(x,y)l(x,y)\mathrm{d}x\mathrm{d}y\right)\right.\\
&\quad\left.+\sqrt{k}\left(\iint_{[0,1]^2} g(x,y)\hat{R}_n^1(x,y)\mathrm{d}x\mathrm{d}y - \iint_{[0,1]^2} g(x,y)R(x,y)\mathrm{d}x\mathrm{d}y\right)\right|\\
&= \left|\iint_{[0,1]^2} g(x,y)\sqrt{k}\left(\frac{\lceil kx\rceil + \lceil ky\rceil - 2}{k} - x - y\right)\mathrm{d}x\mathrm{d}y\right| = O\left(\frac{1}{\sqrt{k}}\right)\end{aligned}$$

almost surely. Hence, to show (2.7.3) we will prove

$$\left| \iint_{[0,1]^2} g(x,y)\sqrt{k}\left(\hat{R}_n^1(x,y) - R(x,y)\right) \mathrm{d}x\mathrm{d}y - \tilde{B} \right| \xrightarrow{\mathbb{P}} 0. \tag{2.7.4}$$

First we write

$$\begin{aligned}\sqrt{k}\left(\hat{R}_n^1(x,y) - R(x,y)\right) &= \sqrt{k}\left(\hat{R}_n^1(x,y) - R_n\left(S_{1n}(x), S_{2n}(y)\right)\right) \\ &\quad + \sqrt{k}\big(R_n\left(S_{1n}(x), S_{2n}(y)\right) - R\left(S_{1n}(x), S_{2n}(y)\right)\big) \\ &\quad + \sqrt{k}\big(R\left(S_{1n}(x), S_{2n}(y)\right) - R(x,y)\big).\end{aligned}$$

From the assumption on integrability of g and the proof of Theorem 2.2 in Einmahl, de Haan, and Li (2006), p. 2003, we get

$$\begin{aligned}&\iint_{[0,1]^2} |g(x,y)| \left| \sqrt{k}\left(\hat{R}_n^1(x,y) - R_n(S_{1n}(x), S_{2n}(y))\right) - W(x,y) \right| \mathrm{d}x\mathrm{d}y \\ &\quad \le \sup_{0\le x,y\le 1} \left| \sqrt{k}\left(\hat{R}_n^1(x,y) - R_n(S_{1n}(x), S_{2n}(y))\right) - W(x,y) \right| \\ &\qquad \cdot \iint_{[0,1]^2} |g(x,y)|\, \mathrm{d}x\mathrm{d}y \\ &\quad \xrightarrow{\mathbb{P}} 0\end{aligned} \tag{2.7.5}$$

and, by conditions (C2) and (C3)

$$\begin{aligned}&\iint_{[0,1]^2} |g(x,y)| \left| \sqrt{k}\big(R_n(S_{1n}(x), S_{2n}(y)) - R(S_{1n}(x), S_{2n}(y))\big) \right| \mathrm{d}x\mathrm{d}y \\ &\quad \le \sup_{0\le x,y\le 1} \left| \sqrt{k}\big(R_n(S_{1n}(x), S_{2n}(y)) - R(S_{1n}(x), S_{2n}(y))\big) \right| \\ &\qquad \cdot \iint_{[0,1]^2} |g(x,y)|\, \mathrm{d}x\mathrm{d}y \\ &\quad \xrightarrow{\mathbb{P}} 0.\end{aligned} \tag{2.7.6}$$

Take ω in the Skorohod probability space introduced above such that $\sup_{0\le x\le 1} |W_1(x)|$ and $\sup_{0\le y\le 1} |W_2(y)|$ are finite and (2.7.2) holds. For such ω

we will show by means of dominated convergence that

$$\begin{aligned}\iint_{[0,1]^2} &|g(x,y)|\big|\sqrt{k}\big(R(S_{1n}(x),S_{2n}(y))-R(x,y)\big) \\ &+R_1(x,y)W_1(x)+R_2(x,y)W_2(y)\big|\mathrm{d}x\mathrm{d}y\to 0.\end{aligned} \tag{2.7.7}$$

(i) *Pointwise convergence of the integrand to zero for almost all* $(x,y)\in[0,1]^2$.

Convergence in (x,y) follows from (2.7.2), provided $R(x,y)$ is differentiable. The set of points in which this might fail is by Lemma 2.7.1 equal to

$$D_R := \left\{(x,y)\in[0,1]^2\colon H(\{z\})>0,\; z=\frac{y}{x+y}\right\}.$$

Since H is a finite measure, there can be at most countably many z for which $H(\{z\})>0$. The set D_R is then a union of at most countably many lines through the origin, and hence has Lebesgue measure zero.

(ii) *The domination of the integrand for all* $(x,y)\in[0,1]^2$.

Comparing (2.7.1) and the moment conditions (2.2.4) we see that for all $(x,y)\in[0,1]^2$ it holds that $|R_1(x,y)|\le 1$ and $|R_2(x,y)|\le 1$. Hence for all $(x,y)\in[0,1]^2$,

$$\begin{aligned}&|g(x,y)|\left|\sqrt{k}\big(R\left(S_{1n}(x),S_{2n}(y)\right)-R(x,y)\big)+R_1(x,y)W_1(x)+R_2(x,y)W_2(y)\right| \\ &\quad\le |g(x,y)|\left(\sqrt{k}\left|R\left(S_{1n}(x),S_{2n}(y)\right)-R(x,y)\right|+|W_1(x)|+|W_2(y)|\right).\end{aligned}$$

We will show that the right-hand side in the inequality above is less than or equal to $M|g(x,y)|$ for all $(x,y)\in[0,1]^2$ and some positive constant M (depending on ω). For that purpose we prove

$$\sup_{0\le x,y\le 1}\sqrt{k}|R(S_{1n}(x),S_{2n}(y))-R(x,y)|=O(1).$$

The representation (2.2.1) implies that for all $x,x_1,x_2,y,y_1,y_2\in[0,1]$

$$|R(x_1,y)-R(x_2,y)|\le|x_1-x_2|,$$

$$|R(x,y_1)-R(x,y_2)|\le|y_1-y_2|.$$

By these inequalities and (2.7.2) we now have

$$\begin{aligned}
&\sup_{0\le x,y\le 1} \sqrt{k}\,|R(S_{1n}(x),S_{2n}(y)) - R(x,y)| \\
&\quad\le \sup_{0\le x,y\le 1} \sqrt{k}\,|R(S_{1n}(x),S_{2n}(y)) - R(S_{1n}(x),y)| \\
&\qquad + \sup_{0\le x,y\le 1} \sqrt{k}\,|R(S_{1n}(x),y) - R(x,y)| \\
&\quad\le \sup_{0\le x\le 1} \sqrt{k}\,|S_{1n}(x) - x| + \sup_{0\le y\le 1} \sqrt{k}\,|S_{2n}(y) - y| \\
&\quad= O(1).
\end{aligned}$$

Recalling that $\sup_{0\le x\le 1}|W_1(x)|$ and $\sup_{0\le y\le 1}|W_2(y)|$ are finite completes the proof of the domination, and hence the proof of (2.7.7).
Combining (2.7.5), (2.7.6) and (2.7.7) we get (2.7.4), and therefore (2.7.3) too. Property (2.3.1) provides us with a statement analogous to (2.7.3), but with $\hat{L}_n^1$ replaced by $\hat{l}_n$. That is, we have

$$\left|\sqrt{k}\left(\iint_{[0,1]^2} g(x,y)\hat{l}_n(x,y)\mathrm{d}x\mathrm{d}y - \varphi(\theta_0)\right) + \tilde{B}\right| \xrightarrow{\mathbb{P}} 0. \tag{2.7.8}$$

Using condition (C1) and the inverse mapping theorem we get that φ^{-1} is continuously differentiable in a neighborhood of $\varphi(\theta_0)$ and $D_{\varphi^{-1}}(\varphi(\theta_0))$ is equal to $D_\varphi(\theta_0)^{-1}$. By a routine argument, using the delta method (see for instance Theorem 3.1 in van der Vaart (1998)), (2.7.8) implies

$$\sqrt{k}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathbb{P}} -D_\varphi(\theta_0)^{-1}\tilde{B}$$

and, since $\tilde{B}$ is mean zero normally distributed ($\tilde{B} \overset{d}{=} -\tilde{B}$),

$$\sqrt{k}(\hat{\theta}_n - \theta_0) \xrightarrow{d} D_\varphi(\theta_0)^{-1}\tilde{B}.$$

□

Lemma 2.7.2. *Let H_θ be the spectral measure and $\Sigma(\theta)$ the covariance matrix in* (2.4.2)*. If the mapping $\theta \mapsto H_\theta$ is weakly continuous at θ_0, then $\theta \mapsto \Sigma(\theta)$ is continuous at θ_0.*

Proof of Lemma 2.7.2 Let $\theta_n \to \theta_0$. In view of the expression for $\Sigma(\theta)$ in (2.4.2) and (2.4.3), the assumption that g is integrable and the fact that R, $|R_1|$ and $|R_2|$ are bounded by 1 for all θ and $(x, y) \in [0, 1]^2$, it suffices to show that $R(x, y; \theta_n) \to R(x, y; \theta)$ and $R_i(x, y; \theta) \to R_i(x, y; \theta)$ for $i = 1, 2$ and for almost all $(x, y) \in [0, 1]^2$.

Convergence of R for all $(x, y) \in [0, 1]^2$ follows directly from the representation of R in terms of H in (2.2.3) and the definition of weak convergence. Convergence of R_1 and R_2 in the points $(x, y) \in (0, 1]^2$ for which $H_{\theta_0}(\{y/(x + y)\}) = 0$ follows from Lemma 2.7.1; see for instance in Billingsley (1968), Theorem 5.2(iii) (note that by the moment constraints (2.2.4), $H_\theta/2$ is a probability measure). Since H_{θ_0} can have at most countably many atoms, R_1 and R_2 converge in all $(x, y) \in (0, 1]^2$ except for at most countably many rays through the origin. $\square$

Proof of Corollary 2.4.3 By the continuous mapping theorem, it suffices to show that

$$(\Sigma(\hat{\theta}_n))^{-1/2} D_\varphi(\hat{\theta}_n)\sqrt{k}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I_p),$$

with I_p the $p \times p$ identity matrix. By condition (C1) of Theorem 2.4.2, the map $\theta \mapsto D_\varphi(\theta)$ is continuous at θ_0, so that, by the continuous mapping theorem, $D_\varphi(\hat{\theta}_n) \xrightarrow{\mathbb{P}} D_\varphi(\theta_0)$ as $n \to \infty$. Slutsky's lemma and (2.4.1) yield

$$D_\varphi(\hat{\theta}_n)\sqrt{k}(\hat{\theta}_n - \theta_0) \xrightarrow{d} D_\varphi(\theta_0) D_\varphi(\theta_0)^{-1}\tilde{B} = \tilde{B},$$

as $n \to \infty$. By Lemma 2.7.2 and the assumption that the map $\theta \mapsto H_\theta$ is weakly continuous, $\Sigma(\hat{\theta}_n)^{-1/2} \xrightarrow{\mathbb{P}} \Sigma(\theta_0)^{-1/2}$. Apply Slutsky's lemma once more to conclude the proof. $\square$

Proof of Theorem 2.4.4 We will show that, for the Skorohod construction introduced before the proof of Theorem 2.4.2,

$$\Bigg| \iint_{[0,1]^2} \left(k \left(\hat{l}_n(x, y) - l(x, y; \hat{\theta}_n) \right)^2 \right.$$
$$\left. - \left(B(x, y) - D_{l(x,y;\theta)}(\theta_{\mathcal{H}_0}) D_\varphi(\theta_{\mathcal{H}_0})^{-1}\tilde{B} \right)^2 \right) \mathrm{d}x\mathrm{d}y \Bigg| \xrightarrow{\mathbb{P}} 0$$

as $n \to \infty$. The left-hand side of the previous expression is less than or equal to

$$\sup_{0\le x,y\le 1}\left|\sqrt{k}\left(\hat{l}_n(x,y)-l(x,y;\hat{\theta}_n)\right)-B(x,y)+D_{l(x,y;\theta)}(\theta_{\mathcal{H}_0})D_{\varphi}(\theta_{\mathcal{H}_0})^{-1}\tilde{B}\right|$$

$$\cdot\left(\left|\iint_{[0,1]^2}\left(\sqrt{k}\left(\hat{l}_n(x,y)-l(x,y;\theta_{\mathcal{H}_0})\right)+B(x,y)\right)\mathrm{d}x\mathrm{d}y\right|\right.$$

$$\left.+\iint_{[0,1]^2}\left|\sqrt{k}\left(l(x,y;\theta_{\mathcal{H}_0})-l(x,y;\hat{\theta}_n)\right)-D_{l(x,y;\theta)}(\theta_{\mathcal{H}_0})D_{\varphi}(\theta_{\mathcal{H}_0})^{-1}\tilde{B}\right|\mathrm{d}x\mathrm{d}y\right)$$

$$=: S(I_1+I_2).$$

From (2.7.8) with $g \equiv \mathbf{1}$, $\mathbf{1} \in \mathbb{R}^p$, we read $I_1 \xrightarrow{\mathbb{P}} 0$. We need to prove that $S = O_{\mathbb{P}}(1)$ and $I_2 = o_{\mathbb{P}}(1)$.

Proof of $\boldsymbol{S = O_{\mathbb{P}}(1)}$ We have

$$\begin{aligned} S &\le \sup_{0\le x,y\le 1}|B(x,y)|+\sup_{0\le x,y\le 1}\left|\sqrt{k}\left(\hat{l}_n(x,y)-l(x,y;\theta_{\mathcal{H}_0})\right)\right| \\ &\quad+\sup_{0\le x,y\le 1}\left|\sqrt{k}\left(l(x,y;\theta_{\mathcal{H}_0})-l(x,y;\hat{\theta}_n)\right)+D_{l(x,y;\theta)}(\theta_{\mathcal{H}_0})D_{\varphi}(\theta_{\mathcal{H}_0})^{-1}\tilde{B}\right| \\ &=: \sup_{0\le x,y\le 1}|B(x,y)|+S_1+S_2. \end{aligned}$$

From the definition of process B it follows that $|B(x,y)|$ is almost surely bounded. Furthermore, we have

$$\begin{aligned} S_1 &= \sup_{0\le x,y\le 1}\left|\sqrt{k}\left(\hat{R}_n^1(x,y)-R(x,y;\theta_{\mathcal{H}_0})\right)\right|+o(1) \\ &\le \sup_{0\le x,y\le 1}\left|\sqrt{k}\left(\hat{R}_n^1(x,y)-R_n(S_{1n}(x),S_{2n}(y))\right)\right| \\ &\quad+\sup_{0\le x,y\le 1}\left|\sqrt{k}\big(R_n(S_{1n}(x),S_{2n}(y))-R(S_{1n}(x),S_{2n}(y);\theta_{\mathcal{H}_0})\big)\right| \\ &\quad+\sup_{0\le x,y\le 1}\left|\sqrt{k}\big(R(S_{1n}(x),S_{2n}(y);\theta_{\mathcal{H}_0})-R(x,y;\theta_{\mathcal{H}_0})\big)\right|+o(1) \end{aligned}$$

almost surely. In the last part of the proof of Theorem 2.4.2 we have shown that the third term is almost surely bounded, and by the proof of Theorem 2.2 in

Einmahl, de Haan, and Li (2006) we know that the first two terms are bounded in probability. Let M denote a constant (depending on $\theta_{\mathcal{H}_0}$) bounding the gradient of $\theta \to l(x, y; \theta)$ at $\theta_{\mathcal{H}_0}$ in $(x, y) \in [0, 1]^2$. Then by (2.4.1)

$$S_2 \leq M \left\| \sqrt{k}(\hat{\theta}_n - \theta_{\mathcal{H}_0}) \right\| + M \left\| D_\varphi(\theta_{\mathcal{H}_0})^{-1} \tilde{B} \right\| = O_{\mathbb{P}}(1).$$

Proof of $I_2 = o_{\mathbb{P}}(1)$ In Theorem 2.4.2 we have shown

$$T_n := \sqrt{k}(\hat{\theta}_n - \theta_{\mathcal{H}_0}) \xrightarrow{\mathbb{P}} -D_\varphi(\theta_{\mathcal{H}_0})^{-1} \tilde{B} =: N.$$

By Slutsky's lemma, it is also true that $(T_n, N) \xrightarrow{\mathbb{P}} (N, N)$. By Skorohod construction there exists a probability space, call it Ω^*, such that it contains both T_n^* and N_n^*, where $(T_n^*, N_n^*) \stackrel{d}{=} (T_n, N)$ and

$$(T_n^*, N_n^*) \xrightarrow{a.s.} (N^*, N^*). \tag{2.7.9}$$

Set $\hat{\theta}_n^* := T_n^*/\sqrt{k} + \theta_{\mathcal{H}_0} \stackrel{d}{=} T_n/\sqrt{k} + \theta_{\mathcal{H}_0} = \hat{\theta}_n$. Let $\Omega_0^* \subset \Omega^*$ be a set of probability one on which N^* is finite and the convergence in (2.7.9) holds. We will show that on Ω_0^*

$$\begin{aligned} I_2^* &:= \iint_{[0,1]^2} X_n^*(x, y) \mathrm{d}x\mathrm{d}y \\ &:= \iint_{[0,1]^2} \left| \sqrt{k} \left(l(x, y; \hat{\theta}_n^*) - l(x, y; \theta_{\mathcal{H}_0}) \right) - D_{l(x,y;\theta)}(\theta_{\mathcal{H}_0}) N_n^* \right| \mathrm{d}x\mathrm{d}y \end{aligned}$$

converges to zero. Since $I_2^* \stackrel{d}{=} I_2$, the above convergence (namely $I_2^* \xrightarrow{a.s.} 0$) will imply $I_2 \xrightarrow{\mathbb{P}} 0$. To show that I_2^* converges to zero on Ω_0^* we will once more apply the dominated convergence theorem. From now on we work on Ω_0^*.

(i) *Pointwise convergence of $X_n^*(x, y)$ to zero.*

We have that

$$\begin{aligned} X_n^*(x, y) &\leq \left| \sqrt{k} \left(l(x, y; \hat{\theta}_n^*) - l(x, y; \theta_{\mathcal{H}_0}) - D_{l(x,y;\theta)}(\theta_{\mathcal{H}_0})(\hat{\theta}_n^* - \theta_{\mathcal{H}_0}) \right) \right| \\ &\quad + \left| D_{l(x,y;\theta)}(\theta_{\mathcal{H}_0})(T_n^* - N_n^*) \right|. \end{aligned}$$

Because of (2.7.9), differentiability of $\theta \mapsto l(x, y; \theta)$ and continuity of matrix multiplication, the right-hand side of the above inequality converges to zero, for all $(x, y) \in [0, 1]^2$.

(ii) *Domination of* $X_n^*(x, y)$.

Let M be as above. Since the sequences $(T_n^*) = (\sqrt{k}(\hat{\theta}_n^* - \theta_{\mathcal{H}_0}))$ and (N_n^*) are convergent and hence bounded, we have

$$\sup_{0 \le x, y \le 1} X_n^*(x, y) \le M \left\| \sqrt{k}(\hat{\theta}_n^* - \theta_{\mathcal{H}_0}) \right\| + M \|N_n^*\| = O(1).$$

This concludes the proof of the domination, and hence the proof of $I_2 \xrightarrow{\mathbb{P}} 0$. $\square$

Proof of Lemma 2.6.1 Without loss of generality, we can assume that F is itself a parallel elliptical distribution, that is, (X, Y) is given as in (2.6.1) with $\rho = 0$ in (2.6.2). Under the assumptions of the lemma and by Theorem 2.3 in Klüppelberg, Kuhn, and Peng (2007), there exists a function $h : [0, \infty)^2 \to \mathbb{R}$ such that as $t \downarrow 0$ and for all $(x, y) \in [0, \infty)^2$,

$$\frac{t^{-1}\mathbb{P}\left(1 - F_1(X) \le tx, 1 - F_2(Y) \le ty\right) - R(x, y; \nu)}{A(F_2^{\leftarrow}(1 - t))} \to h(x, y). \tag{2.7.10}$$

Moreover, the convergence in (2.7.10) holds uniformly on $\{(x, y) \in [0, \infty)^2 : x^2 + y^2 = 1\}$, and the function h is bounded on that region; see Klüppelberg, Kuhn, and Peng (2007) for an explicit expression of the function h.

Condition (2.6.4) obviously implies that $z \mapsto \mathbb{P}(Z > z)$ is regularly varying at infinity with index $-\nu$. Hence, the same is true for the function $1 - F_2$, see Hult and Lindskog (2002). By Proposition 1.5.7 and Theorem 1.5.12 in Bingham, Goldie, and Teugels (1987), the function $x \mapsto |A(F_2^{\leftarrow}(1 - 1/x))|$ is regularly varying at infinity with index β/ν. Hence, for every $\alpha < -\beta/\nu$ we have $A(F_2^{\leftarrow}(1 - 1/x)) = o(x^{-\alpha})$ as $x \to \infty$, or $A(F_2^{\leftarrow}(1 - t)) = o(t^{\alpha})$ as $t \downarrow 0$. As a consequence, for every $\alpha < -\beta/\nu$ we have as $t \downarrow 0$,

$$t^{-1}\mathbb{P}\left(1 - F_1(X) \le tx, 1 - F_2(Y) \le ty\right) - R(x, y; \nu) = O(t^{\alpha}),$$

uniformly on $\{(x, y) \in [0, \infty)^2 : x^2 + y^2 = 1\}$. Uniformity on $\{(x, y) \in [0, \infty)^2 : x + y = 1\}$ now follows as in the proof of Theorem 2.2 in Einmahl, de Haan, and Li (2006). $\square$

Chapter 3

A Method of Moments Estimator of Tail Dependence in Elliptical Copula Models

[Based on A. Krajina, *A Method of Moments Estimator of Tail Dependence in Elliptical Copula Models*, CentER Discussion Paper Series 2009-42.]

Abstract. An elliptical copula model is a distribution function whose copula is that of an elliptical distribution. The tail dependence function in such a bivariate model has a parametric representation with two parameters: a tail parameter and a correlation parameter. The correlation parameter can be estimated by robust methods based on the whole sample. Using the estimated correlation parameter as plug-in estimator, we then estimate the tail parameter applying a modification of the method of moments approach proposed in the paper by J.H.J. Einmahl, A. Krajina and J. Segers [Bernoulli 14(4), 2008, 1003-1026]. We show that such an estimator is consistent and asymptotically normal. Also, we derive the joint limit distribution of the estimators of the two parameters. By a simulation study, we illustrate the small sample behavior of the estimator of the tail parameter and we compare its performance to that of the estimator proposed in the paper by C. Klüppelberg, G. Kuhn and L. Peng [Scandinavian Journal of Statistics 35(4), 2008, 701-718].

3.1 Introduction

The bivariate elliptical distributions, see for example Fang, Kotz, and Ng (1990); Berman (1992), are frequently used in various areas of statistical application, mainly in different branches of financial mathematics, such as risk management, see Embrechts, McNeil, and Straumann (2002); Landsman and Valdez (2003); Kaynar, Birbil, and Frenk (2007). They are a natural extension of Gaussian and t-distributions, and a family wide enough to capture many traits of real-life problems. A number of recent papers have studied the tail behavior of bivariate elliptical distributions, see Abdous, Fougères, and Ghoudi (2005); Asimit and Jones (2007); Hashorva (2005); Demarta and McNeil (2005). An estimator of the tail dependence function of elliptical distributions was suggested in Klüppelberg, Kuhn, and Peng (2007). To model the tail dependence, a wider class of so-called elliptical copula models can be considered instead of the elliptical distributions, since the (tail) dependence structure does not depend on the marginal distributions. The distribution function from an elliptical copula model is a distribution function which has the copula of an elliptical distribution. The tail dependence of the elliptical copula models was estimated in Klüppelberg, Kuhn, and Peng (2008).

Let (X, Y) be a random vector with continuous distribution function F and marginals F_1, F_2. To study the upper tail dependence structure, the *tail dependence function* of (X, Y) is defined as

$$R(x, y) = \lim_{t \downarrow 0} t^{-1} \mathbb{P} \left(1 - F_1(X) \leq tx, 1 - F_2(Y) \leq ty\right),$$

where $x \geq 0$ and $y \geq 0$, see for example Beirlant, Goegebeur, Segers, and Teugels (2004); de Haan and Ferreira (2006); Falk, Hüsler, and Reiss (2004); Huang (1992). The function R is concave; $1 \leq R(x, y) \leq \min\{x, y\}$, for all $x \geq 0$ and $y \geq 0$; and R is homogeneous of order one: $R(tx, ty) = tR(x, y)$, for all $x \geq 0$, $y \geq 0$ and $t \geq 0$. The upper tail dependence coefficient, $R(1, 1)$, is often used as a simple measure of tail dependence.

For an elliptical copula model the tail dependence function depends only on the distribution function, through its copula. Since the copula of an elliptical distribution, and hence the tail dependence function of an elliptical copula model too, belongs to a two-parameter family, the estimation of the tail dependence function reduces to the estimation of the two copula parameters: the correlation parameter and the tail parameter.

The correlation can be estimated using the whole sample, from the rank correlations, which are independent of the precise model. In Klüppelberg, Kuhn, and Peng (2008), the tail parameter was estimated by matching the empirical tail dependence function and the theoretical one, after plugging in the estimated correlation. Using the estimated correlation coefficient as plug-in estimator, in the present chapter we apply the method of moments procedure from Einmahl, Krajina, and Segers (2008) to estimate the tail parameter. The method provides a computationally straightforward estimator which is obtained as a solution of a single equation. The estimator is consistent and asymptotically normal. An interesting result that does not appear in the similar literature, namely the joint limit distribution of the tail parameter and the correlation parameter, is derived. A simulation study shows that the small sample behavior of the estimator of the tail parameter is comparable to and competitive with the small sample behavior of the estimator derived in Klüppelberg, Kuhn, and Peng (2008).

The chapter is organized as follows. In Section 3.2 we state and describe the model. We formulate the problem and present the estimation method in Section 3.3. The main results are given in Section 3.4. In Section 3.5 the performance of the estimator is illustrated using simulated data. All proofs are deferred to Section 3.6.

3.2 Tail dependence in elliptical copula models

Let (Z_1, Z_2) be an *elliptically distributed* random vector, that is, it satisfies the distributional equality

$$(Z_1, Z_2) \stackrel{d}{=} G\boldsymbol{A}\boldsymbol{U},$$

where $G > 0$ is the generating random variable, $\boldsymbol{A}$ is a 2×2 matrix such that $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^\top$ is of full rank, and $\boldsymbol{U}$ is a two-dimensional random vector independent of G and uniformly distributed on the unit circle $\{(z_1, z_2) \in \mathbb{R}^2 : z_1^2 + z_2^2 = 1\}$.

In this case, the matrix $\mathbf{\Sigma}$ can be written as

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where $\sigma_1 > 0$, $\sigma_2 > 0$, and $-1 < \rho < 1$. The parameter ρ is called the correlation coefficient and coincides with the usual correlation, if second moments exist.

A distribution function F follows an *elliptical copula model* if the copula of F is the same as the copula of some elliptical distribution with generating random variable G and correlation coefficient ρ. This model is also known as the *meta-elliptical model*, as introduced in Fang, Fang, and Kotz (2002). If G is regularly varying with index $\nu > 0$ and if $|\rho| < 1$, the expression (3.2.1) for the tail dependence function R was derived in Klüppelberg, Kuhn, and Peng (2007). (Recall that a random variable G is regularly varying with index $\nu > 0$ if $\mathbb{P}(G > x) = x^{-\nu}L(x)$, and L is a slowly varying function.) Setting $f(x, y; \rho, \nu) = \arctan(((x/y)^{1/\nu} - \rho)/\sqrt{1-\rho^2}) \in [-\arcsin\rho, \pi/2]$ for $x, y > 0$, that expression reads

$$R(x, y; \rho, \nu) = \frac{x \int_{f(x,y;\rho,\nu)}^{\pi/2} (\cos\phi)^\nu \mathrm{d}\phi + y \int_{-\arcsin\rho}^{f(x,y;\rho,\nu)} (\sin(\phi + \arcsin\rho))^\nu \mathrm{d}\phi}{\int_{-\pi/2}^{\pi/2} (\cos\phi)^\nu \mathrm{d}\phi}, \quad (3.2.1)$$

and equivalently,

$$R(x, y; \rho, \nu) = \frac{x \int_{f(x,y;\rho,\nu)}^{\pi/2} (\cos\phi)^\nu \mathrm{d}\phi + y \int_{f(y,x;\rho,\nu)}^{\pi/2} (\cos\phi)^\nu \mathrm{d}\phi}{\int_{-\pi/2}^{\pi/2} (\cos\phi)^\nu \mathrm{d}\phi} \quad (3.2.2)$$

$$= \frac{\int_{-\arcsin\rho}^{\pi/2} \min\{x(\cos\phi)^\nu, y(\sin(\phi + \arcsin\rho))^\nu\} \mathrm{d}\phi}{\int_{-\pi/2}^{\pi/2} (\cos\phi)^\nu \mathrm{d}\phi}. \quad (3.2.3)$$

The expression in (3.2.2) was derived in Klüppelberg, Kuhn, and Peng (2008). The one in (3.2.3) is easily obtained from the above formulas.

An expression for Pickands dependence function $A(x) := 1 - R(1-x, x)$ of the bivariate t-distribution was derived in Demarta and McNeil (2005),

$$A(x) = xF_{t(\nu+1)}\left(\frac{(\frac{x}{1-x})^{\frac{1}{\nu}} - \rho}{\sqrt{1-\rho^2}}\sqrt{\nu+1}\right) + (1-x)F_{t(\nu+1)}\left(\frac{(\frac{1-x}{x})^{\frac{1}{\nu}} - \rho}{\sqrt{1-\rho^2}}\sqrt{\nu+1}\right),$$

where $F_{t(\nu+1)}$ is the distribution function of t-distributed random variable with $\nu+1$ degrees of freedom. It was shown in Asimit and Jones (2007) that Pickands dependence function of an elliptical distribution for which the generating variable G is regularly varying with index $\nu > 0$ is the same. Despite the different appearance, expressions (3.2.1)-(3.2.3) lead to the same Pickands dependence function.

3.3 Estimation

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a continuous distribution function F with marginals F_1 and F_2. Assume that F follows an elliptical copula model with underlying generating variable $G > 0$ and correlation coefficient ρ. To estimate the tail dependence function R, we will estimate the unknown parameters, namely the correlation coefficient ρ and the tail index ν, under the assumptions that $|\rho| < 1$ and that G is regularly varying with index $\nu > 0$.

The above assumption corresponds to asymptotic dependence. If $\rho = 1$ or $\rho = -1$, we get complete dependence, $R(1-x, x) = \min\{1-x, x\}$, for any ν. In case of $-1 < \rho < 1$ and $\nu \downarrow 0$ we have a mixture between complete dependence and independence, $R(1-x, x) = \pi^{-1}(\pi/2 + \arcsin\rho)\min\{1-x, x\}$. If $\nu \uparrow \infty$, then for any ρ we are in the case of asymptotic independence, since then $R(1-x, x) \downarrow 0$.

The estimation consists of two steps. We first estimate the correlation coefficient ρ using Kendall's τ, see Kendall (1938, 1948), and the relation $\tau = (2/\pi)\arcsin\rho$ obtained in Lindskog, McNeil, and Schmock (2003), see also Theorem 4.2 in Hult and Lindskog (2002). Then, using expression (3.2.1) with the consistent estimator $\hat{\rho}$ from the previous step plugged in for the true correlation coefficient ρ, we apply the method of moments estimation procedure introduced in Einmahl, Krajina, and Segers (2008) to estimate ν. A similar approach appears in Klüppelberg, Kuhn, and Peng (2008), where the tail parameter is estimated using the pointwise inverse of $R(x, y; \rho, \nu)$ with respect to ν, after the correlation coefficient ρ in R has been replaced by the same consistent estimator as above.

3.3.1 Estimation of the correlation parameter

Kendall's τ of two random variables X and Y is defined by

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

where (X', Y') is independent of and identically distributed as (X, Y). To estimate τ, we will use the classical estimator

$$\hat{\tau} = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \operatorname{sign}\left((X_i - X_j)(Y_i - Y_j)\right),$$

and define the estimator of ρ by

$$\hat{\rho} := \sin\left(\frac{\pi}{2}\hat{\tau}\right).$$

This is a consistent and asymptotically normal estimator of ρ, with rate of convergence $1/\sqrt{n}$, which follows from the corresponding properties of $\hat{\tau}$, see for instance Lee (1990).

3.3.2 Estimation of the tail parameter

Denote by R_i^X and R_i^Y the rank of X_i among $X_1, \ldots, X_n$ and the rank of Y_i among $Y_1, \ldots, Y_n$, respectively. Then for $1 \le k \le n$,

$$\hat{R}_n(x, y) := \frac{1}{k} \sum_{i=1}^{n} \mathbf{1}\left\{R_i^X > n + \frac{1}{2} - kx,\ R_i^Y > n + \frac{1}{2} - ky\right\}$$

is a nonparametric estimator of R. When studying the asymptotic properties of this estimator, $k = k_n$ is an intermediate sequence, that is, $k \to \infty$ and $k/n \to 0$ as $n \to \infty$.

Denote the parameter space by $\bar{\Theta} := \bar{\Theta}_\rho \times \bar{\Theta}_\nu$, with $\bar{\Theta}_\rho = (-1, 1)$ and $\bar{\Theta}_\nu = (0, \infty)$. Its elements are pairs $\bar{\theta} := (\rho, \nu)$. The tail dependence function of an elliptical copula model belongs to a parametric family $\{R(\cdot, \cdot; \bar{\theta}) \colon \bar{\theta} \in \bar{\Theta}\}$. Given the correlation parameter ρ, it reduces to a single-parameter family $\{R(\cdot, \cdot; \rho, \nu) \colon \nu \in \bar{\Theta}_\nu\}$. We use the approach from Einmahl, Krajina, and Segers (2008) to estimate ν: for a given ρ and an integrable function $g \colon [0, 1]^2 \to \mathbb{R}$, the method of moments estimator of ν is defined as the solution to

$$\iint_{[0,1]^2} g(x, y)\hat{R}_n(x, y)\mathrm{d}x\mathrm{d}y = \iint_{[0,1]^2} g(x, y)R(x, y; \rho, \hat{\nu}_n)\mathrm{d}x\mathrm{d}y. \tag{3.3.1}$$

We can simplify the above equation by an appropriate choice of the function g. Choosing $g(x, y) = \mathbf{1}\{x + y \le 1\}$, $(x, y) \in [0, 1]^2$, reduces the area of integration from the unit square to the triangle $\{(x, y) \in [0, 1]^2 \colon x + y \le 1\}$. Due to homogeneity of R, see for instance Beirlant, Goegebeur, Segers, and Teugels (2004); de Haan and Ferreira (2006), we get that

$$\iint_{[0,1]^2} \mathbf{1}\{x + y \le 1\}R(x,y; \rho, \nu)\mathrm{d}x\mathrm{d}y = \frac{1}{3}\int_{[0,1]} R(1 - x, x; \rho, \nu)\mathrm{d}x.$$

Instead of solving the equation (3.3.1), for a given ρ we define the estimator of ν as the solution to

$$\int_{[0,1]} \hat{R}_n(1 - x, x)\mathrm{d}x = \int_{[0,1]} R(1 - x, x; \rho, \hat{\nu})\mathrm{d}x.$$

That is, for a given $\rho \in \bar{\Theta}_\rho$, we define the estimator of ν as the inverse of the function $\bar{\varphi}_\rho \colon \bar{\Theta}_\nu \to \mathbb{R}$, defined by

$$\bar{\varphi}_\rho(\nu) := \int_{[0,1]} R(1 - x, x; \rho, \nu)\mathrm{d}x,$$

in the point $\int_{[0,1]} \hat{R}_n(1 - x, x)\mathrm{d}x$. However, if $\rho > 0$ this is not possible for all ν, since for positive ρ the function $\bar{\varphi}_\rho$ is not invertible on its whole domain $(0, \infty)$. For each $\rho > 0$ there exists a point $\nu^* = \nu^*(\rho)$, such that the function $\nu \mapsto \bar{\varphi}_\rho(\nu)$ is increasing on $(0, \nu^*(\rho))$ and decreasing on $(\nu^*(\rho), \infty)$, see Figure 3.1.

We will restrict the parameter space so to avoid the fact that ν^* changes with

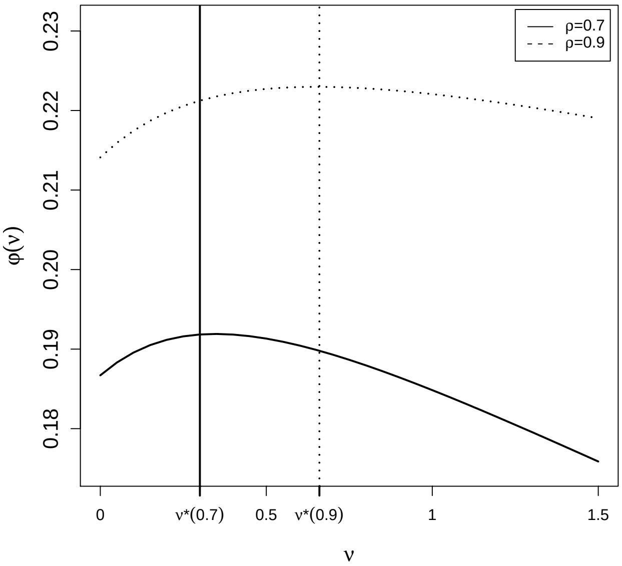


Figure 3.1: Graph of $\nu \mapsto \bar{\varphi}_\rho(\nu)$ for $\rho \in \{0.7, 0.9\}$. The function is decreasing on the interval $(\nu^*(\rho), \infty)$, as indicated by the vertical lines.

ρ, while retaining as much flexibility as possible. We choose some $\rho^* < 1$, numerically approximate the value of $\nu^*(\rho^*)$, and restrict $\bar{\Theta} = (-1,1) \times (0,\infty)$ to $(-1,\rho^*) \times (\nu^*,\infty) =: \Theta_\rho \times \Theta_\nu =: \Theta$. For example, if $\rho^* = 0.9$, we can take $\nu^* = 0.66$, what leads to a parameter space that is appropriate for applications.

For every $\rho \in \Theta_\rho$ denote by φ_ρ the restriction of $\bar{\varphi}_\rho$ to Θ_ν, that is, for every $\rho \in \Theta_\rho$,

$$\varphi_\rho(\nu) := \int_{[0,1]} R(1-x, x; \rho, \nu)\mathrm{d}x\,, \quad \nu \in \Theta_\nu.$$

Finally, for $\hat{\rho} \in \Theta_\rho$, we define $\hat{\nu}_n$, the moment estimator of the tail parameter ν, as the solution to

$$\int_{[0,1]} \hat{R}_n(1-x, x)\mathrm{d}x = \int_{[0,1]} R(1-x, x; \hat{\rho}, \hat{\nu})\mathrm{d}x,$$

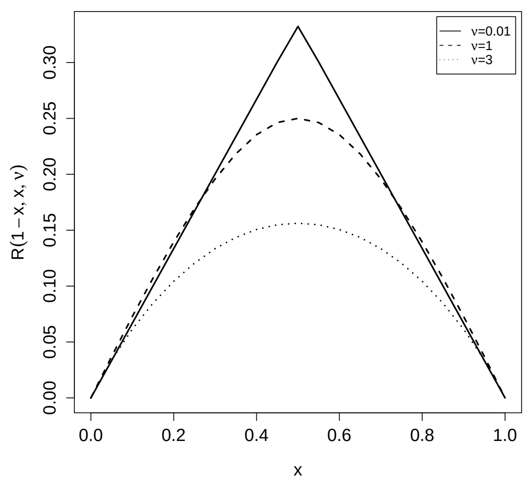
that is,

$$\hat{\nu}_n := \varphi_{\hat{\rho}}^{-1}\left(\int_{[0,1]} \hat{R}_n(1-x, x)\mathrm{d}x\right). \tag{3.3.2}$$

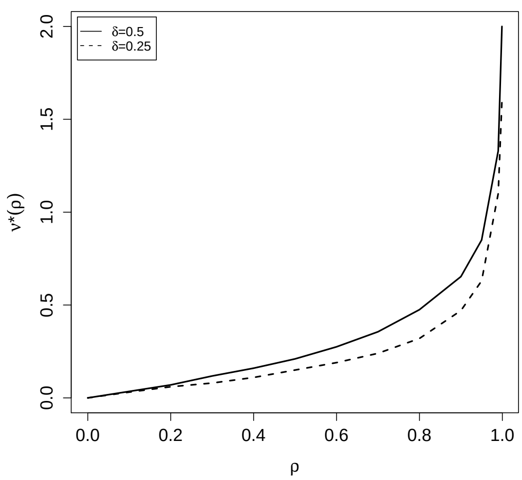
The estimator is well-defined with probability tending to one, as a consequence of the consistency of $\hat{\rho}$ and the uniform consistency of $\hat{R}_n(1-x,x)$. If $\hat{\rho} \notin \Theta_\rho$ or if $\int_{[0,1]} \hat{R}_n(1-x,x)\mathrm{d}x \notin \Theta_\nu$, let $\hat{\nu}_n$ be some fixed value in Θ_ν.

Remark 3.3.1. (i) In the central part of the interval $[0,1]$ the functions $x \mapsto R(1-x,x;\rho,\nu)$, $\rho > 0$, behave in a favorable way, that is, they are decreasing in ν, see Figure 3.2(a). To keep the parameter space as large as possible, we could restrict the area of integration from $[0,1]$ to $[1/2-\delta, 1/2+\delta]$, for some $\delta \in (0,1/2]$, see Figure 3.2(b). However, this may result in a less efficient estimator.

(ii) Note that the set $\Theta = \Theta_\rho \times \Theta_\nu$ is not unique. For any fixed ρ^* we can take $\Theta = (-1,\rho^*) \times (\nu,\infty)$, with $\nu \geq \nu^*(\rho^*)$, see Figure 3.2(b), the solid line. Also, one could fix $\nu^* > 0$ in advance, and appropriately restrict ρ to the interval $(-1,\rho^*(\nu^*))$.



(a) Graphs of $x \mapsto R(1-x,x;\rho,\nu)$ for $\rho = 0.5$ and $\nu \in \{0.01, 1, 3\}$.



(b) Graphs of $\rho \mapsto \nu^*(\rho)$ for $\delta \in \{0.25, 0.5\}$, where $\delta = 0.5$ corresponds to integration over $[0,1]$.

Figure 3.2

3.4 Main results

Let $\hat{\rho}$ and $\hat{\nu}_n$ be as in Section 3.3 and let $\rho_0 \in \Theta_\rho$ and $\nu_0 \in \Theta_\nu$ be the true values of the correlation coefficient and the tail index, respectively. The basic assumption is that

(C0) g is integrable and g and $\Theta = \Theta_\rho \times \Theta_\nu$ are such that φ_ρ is a homeomorphism between Θ_ν and its image, for every $\rho \in \Theta_\rho$.

For some of the results, we will need the following conditions:

(C1) there exists an $\alpha > 0$ such that as $t \to 0$,

$$t^{-1}\mathbb{P}\left(1 - F_1(X_1) \leq tx, 1 - F_2(Y_1) \leq ty\right) - R(x, y) = O(t^\alpha),$$

uniformly on $\{(x, y) \in (0, \infty)^2 : x + y = 1\}$;

(C2) $k = k_n \to \infty$ and $k = o(n^{2\alpha/(1+2\alpha)})$ as $n \to \infty$, with α from (C1).

Proposition 3.4.1. *Assume an elliptical copula model in $\mathbb{R}^2$ with $(\rho_0, \nu_0) \in \Theta$. If (C0) holds, then the function $H\colon \Theta \to \Theta_\rho \times \mathbb{R}$ defined by*

$$H(\rho, \nu) := \left(\rho, \int_{[0,1]} R(1 - x, x; \rho, \nu)\, dx\right), \tag{3.4.1}$$

is continuously differentiable at (ρ_0, ν_0) and its differential in this point is regular.

An application of the inverse mapping theorem yields the following consequence of Proposition 3.4.1. Let $D_f(x)$ denote the differential of f in x.

Corollary 3.4.2. *Assume the situation as in Proposition 3.4.1. Then there exist open neighborhoods $U \subseteq \Theta$ of (ρ_0, ν_0) and $V \subseteq H(\Theta)$ of $H(\rho_0, \nu_0)$ such that the restriction $H|_U\colon U \to V$ is one-to-one. Moreover, its inverse*

$$K := (H|_U)^{-1}\colon V \to U \tag{3.4.2}$$

is continuously differentiable and for the differential of K in $H(\rho_0, \nu_0)$ we have

$$D_K\left(H\left(\rho_0, \nu_0\right)\right) = \left(D_H(\rho_0, \nu_0)\right)^{-1}.$$

Next we present the consistency and asymptotic normality results for $\hat{\nu}_n$ and $(\hat{\nu}_n, \hat{\rho})$, respectively.

Theorem 3.4.3 (Consistency of $\hat{\nu}_n$)**.** *Assume the situation as in Proposition 3.4.1. It holds that*

$$\hat{\nu}_n \xrightarrow{\mathbb{P}} \nu_0, \quad \text{as } n \to \infty,\ k \to \infty,\ k/n \to 0.$$

Denote by W a mean-zero Wiener process on $[0,\infty)^2$ with covariance function

$$\mathbb{E}W(x_1, y_1)W(x_2, y_2) = R(x_1 \wedge x_2, y_1 \wedge y_2; \rho_0, \nu_0), \tag{3.4.3}$$

and for $x, y \in [0, \infty)$ denote

$$W_1(x) := W(x, \infty), \qquad W_2(y) := W(\infty, y). \tag{3.4.4}$$

Further, for $(x,y) \in [0,\infty)^2$ let $\dot{R}_1(x,y)$ and $\dot{R}_2(x,y)$ be the partial derivatives of R in the point (x,y) with respect to the first and second coordinates, respectively.

Finally, define the stochastic process B on $[0,\infty)^2$ by

$$B(x,y) := W(x,y) - \dot{R}_1(x,y)W_1(x) - \dot{R}_2(x,y)W_2(y). \tag{3.4.5}$$

Let $N_\rho \sim \mathbb{N}(0, \sigma_\rho^2)$ be the normal limiting random variable of $\sqrt{n}(\hat{\rho} - \rho_0)$ and denote by $N_\nu \sim \mathbb{N}(0, \sigma_\nu^2)$ the normal random variable $N_\nu := c \int_{[0,1]} B(1-x, x)\mathrm{d}x$, where

$$c := \left(\partial/\partial\nu \int_{[0,1]} R(1-x, x; \rho_0, \nu)\mathrm{d}x \Big|_{\nu=\nu_0} \right)^{-1}. \tag{3.4.6}$$

Theorem 3.4.4 (Asymptotic normality of $(\hat{\nu}_n, \hat{\rho})$)**.** *Let $k/n \to 0$. Assume the situation as in Proposition 3.4.1 and assume that the conditions (C1) and (C2) hold. Then as $n \to \infty$ and $k \to \infty$,*

$$\left(\sqrt{k}(\hat{\nu}_n - \nu_0), \sqrt{n}(\hat{\rho} - \rho_0) \right) \xrightarrow{d} (N_\nu, N_\rho),$$

where N_ν and N_ρ are independent.

Remark 3.4.5. The above results are not tied to the Kendall's tau based estimator of ρ. The consistency and the asymptotic normality of $\hat{\nu}_n$ hold whenever the rate of convergence of the estimator of ρ is faster than $1/\sqrt{k}$.

3.5 Simulation study

We simulated 50 random samples of size $n = 1000$ from two elliptical copula models with correlation coefficient $\rho_0 = 0.3$ and tail parameter $\nu_0 \in \{1, 5\}$.

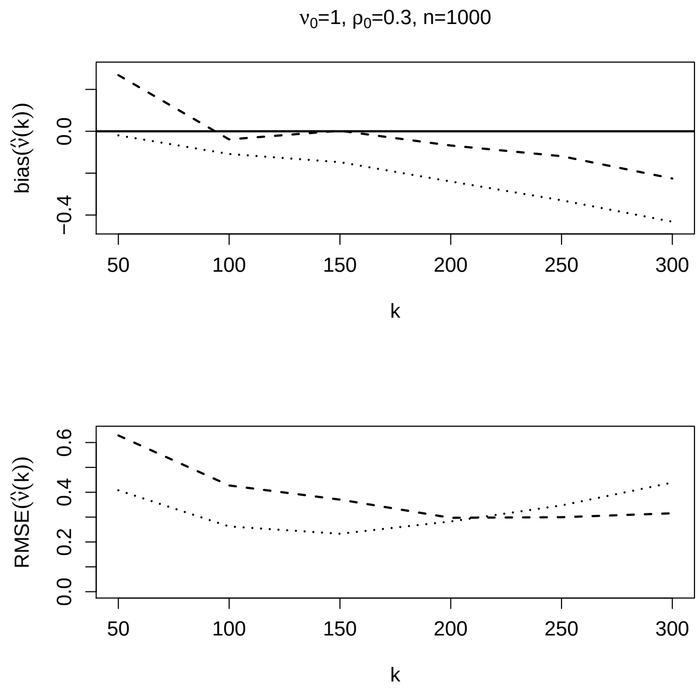
The two estimators that we compare are the MoME, the method of moments estimator $\hat{\nu}_n$ defined in (3.3.2), and the KKP estimator of tail parameter from Klüppelberg, Kuhn, and Peng (2008) with the weight function $m(\psi) = 1 - (4\psi/\pi - 1)^2$, $0 \leq \psi \leq \pi/2$. The KKP estimator is defined by

$$\hat{\nu}_{KKP} := \frac{1}{M(\hat{Q} \cap \hat{Q}^*)} \int_{\hat{Q} \cap \hat{Q}^*} \tilde{\nu}\left(\sqrt{2}\cos\psi, \sqrt{2}\sin\psi\right) M(\mathrm{d}\psi),$$

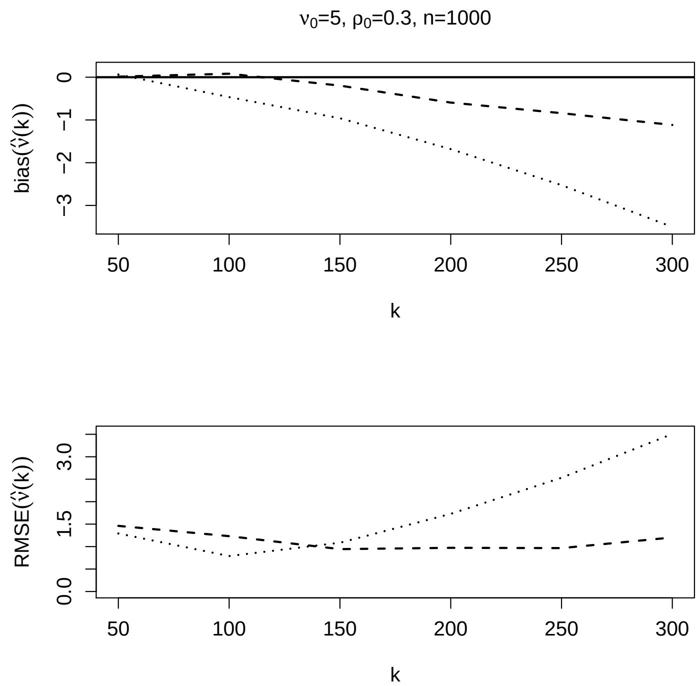
where M is the measure defined by m, $\tilde{\nu}(x, y)$ is the inverse of $R(x, y; \hat{\rho}, \nu)$ with respect to ν in the point $\hat{R}_n(x, y)$, for $x > 0$, $y > 0$, and the sets $\hat{Q}$ and $\hat{Q}^*$ are the subsets of $[0, \pi/2]$ defined in such a way so that $\hat{\nu}_{KKP}$ is well-defined and that it has desired asymptotic properties, see Klüppelberg, Kuhn, and Peng (2008).

In Figure 3.3 we plot for those two estimators the bias and the root mean squared error (RMSE) against the effective sample size k.

The plots show that the MoME has much smaller bias than the KKP estimator. Further, it appears to be more robust with respect to the choice of k, and better than the KKP estimator for k large enough. Also, the value of k after which the MoME performs better gets smaller as the tail parameter that is estimated gets larger.



(a) Elliptical copula model with $\rho_0 = 0.3$ and $\nu_0 = 1$.



(b) Elliptical copula model with $\rho_0 = 0.3$ and $\nu_0 = 5$.

Figure 3.3: The bias and the RMSE of two different estimators of tail coefficient ν; MoME ($- - --$), KKP ($\cdots\cdots$).

3.6 Proofs

Proof of Proposition 3.4.1 To show that the function H is continuously differentiable we will show that its partial derivatives exist and are continuous on Θ. Since $H(\rho,\nu) = (H_1(\rho,\nu), H_2(\rho,\nu))$, where $H_i\colon \Theta \to \mathbb{R}$, $i = 1, 2$, are given by

$$\begin{aligned} H_1(\rho,\nu) &= \rho, \\ H_2(\rho,\nu) &= \int_{[0,1]} R(1-x, x; \rho, \nu)\mathrm{d}x, \end{aligned}$$

we have

$$\begin{aligned} \frac{\partial H_1}{\partial \rho}(\rho,\nu) &= 1, \quad \frac{\partial H_1}{\partial \nu}(\rho,\nu) = 0, \\ \frac{\partial H_2}{\partial \rho}(\rho,\nu) &= c_0^{-1}(1-\rho^2)^{\nu/2} \int_{[0,1]} \frac{x(1-x)}{\left(x^{2/\nu} + (1-x)^{2/\nu} - 2\rho x^{1/\nu}(1-x)^{1/\nu}\right)^{\nu/2}}\mathrm{d}x, \\ \frac{\partial H_2}{\partial \nu}(\rho,\nu) &= c_0^{-2} \int_{[0,1]} (1-x) C\left(\nu, \arctan \frac{\left(\frac{1-x}{x}\right)^{1/\nu} - \rho}{\sqrt{1-\rho^2}}\right)\mathrm{d}x. \end{aligned}$$

The last partial derivative relies on a similar result in Klüppelberg, Kuhn, and Peng (2008); the notation used above also comes from that paper:

$$c_0 = \int_{-\pi/2}^{\pi/2} (\cos\phi)^\nu \mathrm{d}\phi, \quad c_1 = \int_{-\pi/2}^{\pi/2} (\cos\phi)^\nu \ln(\cos\phi)\mathrm{d}\phi,$$

$$\begin{aligned} D(\nu, z) &= c_0 \int_z^{\pi/2} (\cos\phi)^\nu \ln(\cos\phi)\mathrm{d}\phi - c_1 \int_z^{\pi/2} (\cos\phi)^\nu \mathrm{d}\phi, \\ C(\nu, z) &= D(\nu, z) + (\rho + \sqrt{1-\rho^2}\tan z)^{-\nu} D(\nu, \arccos\rho - z). \end{aligned}$$

All four partial derivatives exist and are continuous functions on Θ.

It can be shown that the partial derivative $\partial H_2/\partial\nu$ is negative for all $(\rho,\nu) \in \Theta$, which implies that the differential is regular in every point in Θ. $\square$

Proof of Theorem 3.4.3 Let $p_2\colon \mathbb{R}^2 \to \mathbb{R}$ denote the projection onto the second coordinate and let H and K be the mappings introduced in (3.4.1) and (3.4.2), respectively. Note that ν_0 can be written as

$$\nu_0 = (p_2 \circ K)\left(\rho_0, \int_{[0,1]} R(1-x, x; \rho_0, \nu_0)\mathrm{d}x\right).$$

Moreover, the estimator $\hat{\nu}_n$ has the representation

$$\hat{\nu}_n = (p_2 \circ K)\left(\hat{\rho}, \int_{[0,1]} \hat{R}_n(1-x, x)\mathrm{d}x\right). \tag{3.6.1}$$

The uniform consistency of $\hat{R}_n$, see the proof of Theorem 2.2 in Einmahl, de Haan, and Li (2006), and the equation (3.1) in Einmahl, Krajina, and Segers (2008) imply

$$\int_{[0,1]} \hat{R}_n(1-x, x)\mathrm{d}x \xrightarrow{\mathbb{P}} \int_{[0,1]} R(1-x, x)\mathrm{d}x. \tag{3.6.2}$$

Hence the right-hand side of (3.6.1) is well defined with probability tending to one. Further, from the continuous mapping theorem and Lee (1990) we know that

$$\hat{\rho} \xrightarrow{\mathbb{P}} \rho_0, \tag{3.6.3}$$

as $n \to \infty$. Using (3.6.2), (3.6.3) and continuity of $p_2 \circ K$, we obtain $\hat{\nu}_n \xrightarrow{\mathbb{P}} \nu_0$, as $n \to \infty$, $k \to \infty$ and $k/n \to 0$. $\square$

Some more notation and technical results are needed for the proof of Theorem 3.4.4. For $i = 1, \ldots, n$ denote $U_i := 1 - F_1(X_i)$ and $V_i := 1 - F_2(Y_i)$. Let $U_{1:n} \le \cdots \le U_{n:n}$ and $V_{1:n} \le \cdots \le V_{n:n}$ be the corresponding order statistics and by $\lceil a \rceil$ denote the smallest integer not smaller than a. Define

$$\begin{aligned}
\hat{R}_n^1(x,y) &:= \frac{1}{k}\sum_{i=1}^n \mathbf{1}\left\{R_i^X > n+1-kx,\ R_i^Y > n+1-ky\right\},\\
R_n(x,y) &:= \frac{n}{k}\mathbb{P}\left(U_1 \le \frac{kx}{n}, V_1 \le \frac{ky}{n}\right),\\
T_n(x,y) &:= \frac{1}{k}\sum_{i=1}^n \mathbf{1}\left\{U_i < \frac{kx}{n},\ V_i < \frac{ky}{n}\right\},
\end{aligned}$$

and note that

$$\hat{R}_n^1(x,y) = T_n\left(\frac{n}{k}U_{\lceil kx\rceil:n}, \frac{n}{k}V_{\lceil ky\rceil:n}\right).$$

It is easily seen that

$$\sup_{(x,y)\in[0,n/k]^2} \sqrt{k}\left|\hat{R}_n^1(x,y) - \hat{R}_n(x,y)\right| \le \frac{1}{\sqrt{k}} \to 0,$$

as $n \to \infty$.

Let W, W_1 and W_2 be as in (3.4.3) and (3.4.4). Write

$$v_n(x,y) = \sqrt{k}\left(T_n(x,y) - R_n(x,y)\right), \quad v_{n,1}(x) := v_n(x,\infty) \text{ and } v_{n,2}(y) := v_n(\infty,y).$$

Proposition 3.1 in Einmahl, de Haan, and Li (2006) shows that for any $T > 0$

$$\begin{aligned}&\left(v_n(x,y),(x,y)\in[0,T]^2;\ v_{n1}(x),x\in[0,T];\ v_{n2}(y),y\in[0,T]\right)\\ &\xrightarrow{d} \left(W(x,y),(x,y)\in[0,T]^2;\ W_1(x),x\in[0,T];\ W_2(y),y\in[0,T]\right),\end{aligned}$$

in the topology of uniform convergence, as $n \to \infty$.

Let $F_n(x,y) = (1/n)\sum_{i=1}^n \mathbf{1}\{X_i \le x, Y_i \le y\}$ be the empirical distribution function of F, and let F_{n1} and F_{n2} be the empirical distribution functions of the marginals F_1 and F_2, respectively. Define the empirical process $r_n(x,y) := \sqrt{n}(F_n(x,y) - F(x,y))$, $(x,y) \in [-\infty,\infty]^2 =: \bar{\mathbb{R}}^2$, and denote by W_B a Brownian bridge on $[-\infty,\infty]^2$ with covariance structure

$$\mathbb{E}W_B(x_1,y_1)W_B(x_2,y_2) = F(\min\{x_1,x_2\},\min\{y_1,y_2\}) - F(x_1,y_1)F(x_2,y_2).$$

We know, see e.g. Neuhaus (1971), that $r_n \xrightarrow{d} W_B$ in the topology of uniform convergence, as $n \to \infty$. Hence we obtain for the marginal processes, $r_{nj} \xrightarrow{d} W_{Bj}$, where $r_{nj}(x) := \sqrt{n}(F_{nj}(x) - F_j(x))$, $j = 1,2$, $W_{B1}(x) = W_B(x,\infty)$ and $W_{B2}(x) = W_B(\infty,x)$.

Lemma 3.6.1. *For fixed $(x,y) \in [0,\infty]^2$ and $(t,w) \in [-\infty,\infty]^2$ it holds that as $n \to \infty$, $k \to \infty$, $k/n \to 0$,*

$$\mathbb{E}v_n(x,y)r_n(t,w) \to 0.$$

Proof Fix $(x, y) \in [0, \infty]^2$ and $(t, w) \in \bar{\mathbb{R}}^2$. Then,

$$
\begin{aligned}
\mathbb{E} v_n(x,y) r_n(t,w) &= \mathbb{E}\left[\sqrt{k}\left(T_n(x,y) - R_n(x,y)\right) \cdot \sqrt{n}\left(F_n(x,y) - F(x,y)\right)\right] \\
&= \frac{1}{\sqrt{kn}} \mathbb{E}\left[\sum_{i=1}^{n}\left(\mathbf{1}\left\{U_i < \frac{kx}{n},\ V_i < \frac{ky}{n}\right\} - \mathbb{P}\left(U_1 \leq \frac{kx}{n}, V_1 \leq \frac{ky}{n}\right)\right)\right. \\
&\qquad \left. \cdot \sum_{j=1}^{n}\left(\mathbf{1}\{X_j \leq t, Y_j \leq w\} - F(t,w)\right)\right] \\
&= \frac{1}{\sqrt{kn}} \mathbb{E}\left[\sum\sum_{i,j=1, i\neq j}^{n}\left(\mathbf{1}\left\{U_i < \frac{kx}{n},\ V_i < \frac{ky}{n}\right\} - \mathbb{P}\left(U_1 \leq \frac{kx}{n}, V_1 \leq \frac{ky}{n}\right)\right)\right. \\
&\qquad \left. \cdot \left(\mathbf{1}\{X_j \leq t, Y_j \leq w\} - F(t,w)\right)\right] \\
&\quad + \frac{1}{\sqrt{kn}} \mathbb{E}\left[\sum_{i=1}^{n}\left(\mathbf{1}\left\{U_i < \frac{kx}{n},\ V_i < \frac{ky}{n}\right\} - \mathbb{P}\left(U_1 \leq \frac{kx}{n}, V_1 \leq \frac{ky}{n}\right)\right)\right. \\
&\qquad \left. \cdot \left(\mathbf{1}\{X_i \leq t, Y_i \leq w\} - F(t,w)\right)\right] \\
&=: E_1 + E_2.
\end{aligned}
$$

Using independence of the sample, we get

$$
\begin{aligned}
E_1 &= \frac{1}{\sqrt{kn}} \sum_{i,j=1, i\neq j}^{n} \mathbb{E}\left[\mathbf{1}\left\{U_i < \frac{kx}{n},\ V_i < \frac{ky}{n}\right\} - \mathbb{P}\left(U_1 \leq \frac{kx}{n}, V_1 \leq \frac{ky}{n}\right)\right] \\
&\qquad \cdot \mathbb{E}\left[\mathbf{1}\{X_j \leq t, Y_j \leq w\} - F(t,w)\right] \\
&= 0.
\end{aligned}
$$

Since the factors in the sum in E_2 have the same distribution, we get

$$\begin{aligned}
|E_2| &= \sqrt{\frac{n}{k}}\left|\mathbb{E}\left[\left(\mathbf{1}\left\{U_1 < \frac{kx}{n},\ V_1 < \frac{ky}{n}\right\} - \mathbb{P}\left(U_1 \le \frac{kx}{n}, V_1 \le \frac{ky}{n}\right)\right)\right.\right. \\
&\qquad \left.\left. \cdot\left(\mathbf{1}\{X_1 \le t, Y_1 \le w\} - F(t,w)\right)\right]\right| \\
&\le \sqrt{\frac{n}{k}}\left(\left|\mathbb{E}\left[\mathbf{1}\left\{U_1 < \frac{kx}{n},\ V_1 < \frac{ky}{n}\right\}\mathbf{1}\{X_1 \le t, Y_1 \le w\}\right]\right|\right. \\
&\qquad \left. + \mathbb{P}\left(U_1 \le \frac{kx}{n}, V_1 \le \frac{ky}{n}\right)F(t,w)\right) \\
&\le \sqrt{\frac{n}{k}}\,\mathbb{P}\left(U_1 < \frac{kx}{n}, V_1 < \frac{ky}{n}\right)(1 + F(t,w)) \\
&\le 2\sqrt{\frac{k}{n}}\min\{x,y\} \to 0,
\end{aligned}$$

as $n \to \infty$, $k \to \infty$ and $k/n \to 0$. □

Lemma 3.6.2. *Let $T > 0$. In the topology of uniform convergence, as $n \to \infty$, $k \to \infty$, $k/n \to 0$, the process*

$$(v_n(x,y),(x,y) \in [0,T]^2;\ v_{n1}(x), x \in [0,T];\ v_{n2}(y), y \in [0,T],\ r_n(x,y),(x,y) \in \bar{\mathbb{R}}^2) \tag{3.6.4}$$

converges in distribution to

$$(W(x,y),(x,y) \in [0,T]^2;\ W_1(x), x \in [0,T];\ W_2(y), y \in [0,T],\ W_B(x,y),(x,y) \in \bar{\mathbb{R}}^2), \tag{3.6.5}$$

with

$$(W(x,y),(x,y) \in [0,T]^2;\ W_1(x), x \in [0,T];\ W_2(y), y \in [0,T])$$

and $(W_B(x,y),(x,y) \in \bar{\mathbb{R}}^2)$ independent.

Proof From the weak convergence, and hence tightness, of

$$\big(v_n(x,y),(x,y) \in [0,T]^2;\ v_{n1}(x), x \in [0,T];\ v_{n2}(y), y \in [0,T]\big)$$

and $(r_n(x,y),(x,y)\in\bar{\mathbb{R}}^2)$, we get the tightness of the process in (3.6.4).

By the Cramér-Wold device, see for example Shorack and Wellner (1986), and the univariate Lindeberg-Feller central limit theorem, using Lemma 3.6.1, we get convergence of the finite-dimensional distributions. □

Using the Skorohod construction we get a probability space containing all processes $\tilde{v}_n$, $\tilde{v}_{n1}$, $\tilde{v}_{n2}$, $\tilde{r}_n$, $\tilde{W}$, $\tilde{W}_1$, $\tilde{W}_2$ and $\tilde{W}_B$, where

$$\begin{aligned}(\tilde{v}_n,\tilde{v}_{n1},\tilde{v}_{n2},\tilde{r}_n) &\stackrel{d}{=} (v_n,v_{n1},v_{n2},r_n),\\ (\tilde{W},\tilde{W}_1,\tilde{W}_2,\tilde{W}_B) &\stackrel{d}{=} (W,W_1,W_2,W_B),\end{aligned}$$

and it holds that as $n\to\infty$, $k\to\infty$, $k/n\to 0$,

$$\sup_{(x,y)\in[0,T]^2}|\tilde{v}_n(x,y)-\tilde{W}(x,y)|\to 0 \text{ a.s.,} \tag{3.6.6}$$

$$\sup_{(x,y)\in\bar{\mathbb{R}}^2}|\tilde{r}_n(x,y)-\tilde{W}_B(x,y)|\to 0 \text{ a.s.,} \tag{3.6.7}$$

and the analogous statements hold for marginal processes v_{n1}, v_{n2}, r_{n1} and r_{n2} as well. We work on this space from now on, but keep the old notation (without tilde's).

Lemma 3.6.3. *Assume the situation as in Theorem 3.4.4. On the probability space of the Skorohod construction*

$$\left(\sqrt{k}\left(\int_{[0,1]}\hat{R}_n(1-x,x)\,dx-\int_{[0,1]}R(1-x,x)\,dx\right),\sqrt{n}(\hat{\rho}-\rho_0)\right) \xrightarrow{\mathbb{P}} \left(\int_{[0,1]}B(1-x,x)\,dx,N_\rho\right), \tag{3.6.8}$$

as $n\to\infty$, $k\to\infty$ and $k/n\to 0$, where $\int_{[0,1]}B(1-x,x)\,dx$ and N_ρ are independent, and B is the process defined in (3.4.5).

Proof By Lemma 3.6.2 it is sufficient to show that

$$\sqrt{k}\left(\int_{[0,1]}\hat{R}_n(1-x,x)\mathrm{d}x-\int_{[0,1]}R(1-x,x)\mathrm{d}x\right)\xrightarrow{\mathbb{P}}\int B(1-x,x)\mathrm{d}x, \tag{3.6.9}$$

and

$$\sqrt{n}(\hat{\rho}-\rho_0) \xrightarrow{\mathbb{P}} N_\rho, \tag{3.6.10}$$

since N_ρ is a functional of W_B, by (3.6.12) and the delta method.

For the convergence in (3.6.10) we will first show that $\sqrt{n}(\hat{\tau}-\tau_0) \xrightarrow{\mathbb{P}} N_\tau$, where N_τ is the limiting normal random variable for $\hat{\tau}$, see for example Kendall (1948) or Lee (1990). By the Hoeffding representation of U-statistics and its properties, see for example Lee (1990), we get that

$$\sqrt{n}(\hat{\tau}-\tau_0) = 2\sqrt{n}\left(\iint_{\bar{\mathbb{R}}^2} \Phi(x,y)\mathrm{d}F_n(x,y) - \iint_{\bar{\mathbb{R}}^2} \Phi(x,y)\mathrm{d}F(x,y)\right) + o_{\mathbb{P}}(1),$$

where $\Phi(x,y) = 1 - 2F_1(x) - 2F_2(y) + 4F(x,y)$. Let r_n, r_{n1}, r_{n2}, W_B, W_{B1} and W_{B2} be as defined before Lemma 3.6.1. From integration by parts we get

$$\begin{aligned}\sqrt{n}(\hat{\tau}-\tau_0) &= -8\iint_{\mathbb{R}^2} r_n(x,y)\mathrm{d}F(x,y) + 4\int_{\mathbb{R}} r_{n1}(x)\mathrm{d}F_1(x) \\ &\quad +4\int_{\mathbb{R}} r_{n2}(y)\mathrm{d}F_2(y) + o_{\mathbb{P}}(1).\end{aligned} \tag{3.6.11}$$

Denote

$$N_\tau := -8\iint_{\mathbb{R}^2} W_B(x,y)\mathrm{d}F(x,y) + 4\int_{\mathbb{R}} W_{B1}(x)\mathrm{d}F_1(x) + 4\int_{\mathbb{R}} W_{B2}(y)\mathrm{d}F_2(y). \tag{3.6.12}$$

The result in (3.6.7), its marginal versions and (3.6.11) yield that $\sqrt{n}(\hat{\tau}-\tau_0) \xrightarrow{\mathbb{P}} N_\tau$. Since $\hat{\rho} = \sin((\pi/2)\hat{\tau})$, the delta method yields (3.6.10), where N_ρ is an appropriate function of N_τ. Note that N_ρ is a normally distributed random variable with mean zero and some variance, σ_ρ^2, say. □

Lemma 3.6.4. *Assume the situation as in Proposition 3.4.1. As $n \to \infty$, $k \to \infty$ and $k/n \to 0$,*

$$\frac{\varphi_{\hat{\rho}}^{-1}(\int_{[0,1]} \hat{R}_n(1-x,x)\,dx) - \varphi_{\hat{\rho}}^{-1}(\int_{[0,1]} R(1-x,x;\rho_0,\nu_0)\,dx)}{\int_{[0,1]} \hat{R}_n(1-x,x)\,dx - \int_{[0,1]} R(1-x,x;\rho_0,\nu_0)\,dx} \xrightarrow{\mathbb{P}} c, \tag{3.6.13}$$

where c *is defined in* (3.4.6), *and*

$$\sqrt{k}\left(\varphi_{\hat{\rho}}^{-1}\left(\int_{[0,1]} R(1-x,x;\rho_0,\nu_0)\,dx\right) - \varphi_{\rho_0}^{-1}\left(\int_{[0,1]} R(1-x,x;\rho_0,\nu_0)\,dx\right)\right) \xrightarrow{\mathbb{P}} 0. \tag{3.6.14}$$

Proof Throughout the proof we omit writing the region of integration, $[0,1]$. As before, let H be the function on Θ given by $H(\rho,\nu) = (\rho, \varphi_\rho(\nu))$, let K be its local inverse, and let p_2 be the projection onto the second coordinate. Since $K(\rho,\mu) = (\rho,\nu)$, where ν is such that $\mu = \int_{[0,1]} R(1-x,x;\rho,\nu)\mathrm{d}x$, we see that $(p_2 \circ K)(\rho,\mu) = \varphi_\rho^{-1}(\mu)$. Denote $\mu_0 := \int R(1-x,x;\rho_0,\nu_0)\mathrm{d}x$.

First we prove (3.6.13). Define the function $f\colon [0,1] \to \mathbb{R}$ by

$$f(t) := (p_2 \circ K)\left(\hat{\rho}, \mu_0 + t\left(\int \hat{R}_n(1-x,x)\mathrm{d}x - \mu_0\right)\right).$$

Using the mean value theorem for f on $[0,1]$ we get

$$f(1) - f(0) = (1-0)\cdot f'(t)\big|_{t=t^*}, \quad t^* \in (0,1).$$

Since $f(1) = (p_2 \circ K)(\hat{\rho}, \int \hat{R}_n(1-x,x)\mathrm{d}x) = \varphi_{\hat{\rho}}^{-1}(\int \hat{R}_n(1-x,x)\mathrm{d}x)$ and $f(0) = (p_2 \circ K)(\hat{\rho},\mu_0) = \varphi_{\hat{\rho}}^{-1}(\mu_0)$, we get

$$\varphi_{\hat{\rho}}^{-1}(\textstyle\int \hat{R}_n(1-x,x)\mathrm{d}x) - \varphi_{\hat{\rho}}^{-1}(\mu_0) = \dfrac{\partial}{\partial\mu}(p_2 \circ K)(\hat{\rho},\mu)\big|_{\mu=\mu^*}\left(\textstyle\int \hat{R}_n(1-x,x)\mathrm{d}x - \mu_0\right),$$

with $\mu^* = \mu_0 + t^*(\int \hat{R}_n(1-x,x)\mathrm{d}x - \mu_0)$. Because μ^* is between $\int \hat{R}_n(1-x,x)\mathrm{d}x$ and μ_0, the consistency of $\int \hat{R}_n(1-x,x)\mathrm{d}x$ implies that $\mu^* \xrightarrow{\mathbb{P}} \mu_0$, as $n \to \infty$, $k \to \infty$ and $k/n \to 0$. This, together with the consistency of $\hat{\rho}$ and the continuous differentiability of K, see Corollary 3.4.2, implies that the left-hand side of (3.6.13) converges in probability to $(\partial/\partial\mu)(p_2 \circ K)(\rho_0,\mu_0) = (\partial/\partial\mu)\varphi_{\rho_0}^{-1}(\mu_0)$. By the inverse mapping theorem, this constant equals c.

Next we show that (3.6.14) holds. Similarly, we define the function $f\colon [0,1] \to \mathbb{R}$ by

$$f(t) := (p_2 \circ K)(\rho_0 + t(\hat{\rho} - \rho_0), \mu_0).$$

The mean value theorem applied to f on $[0,1]$ yields

$$f(1) - f(0) = (1-0) \cdot f'(t)\big|_{t=t^*}, \quad t^* \in (0,1).$$

Write $\rho^* := \rho_0 + t^*(\hat{\rho} - \rho_0)$. Since $f(1) = \varphi_{\hat{\rho}}^{-1}(\mu_0)$, and $f(0) = \varphi_{\rho_0}^{-1}(\mu_0)$, the left-hand side of (3.6.14) is equal to

$$\frac{\partial}{\partial \rho} \varphi_\rho^{-1}(\mu_0)\big|_{\rho=\rho^*} \sqrt{k}(\hat{\rho} - \rho_0). \tag{3.6.15}$$

By Corollary 3.4.2, $\rho \mapsto (\partial/\partial\rho)\varphi_\rho^{-1}(\mu_0)$ is continuous, hence it is bounded on a closed neighborhood of ρ_0. The consistency of $\hat{\rho}$ then implies that $(\partial/\partial\rho)\varphi_\rho^{-1}(\mu_0)\big|_{\rho=\rho^*}$ is bounded with probability tending to one. Since the rate of convergence of $\hat{\rho}$ is $1/\sqrt{n}$, the expression in (3.6.15) converges to zero in probability as $n \to \infty$, $k \to \infty$ and $k/n \to 0$. □

Proof of Theorem 3.4.4 Here we again omit writing the region of integration, $[0,1]$, and we write $R(1-x,x)$ instead of $R(1-x,x;\rho_0,\nu_0)$. We have

$$\begin{aligned}
\sqrt{k}\,(\hat{\nu}_n - \nu_0) \quad = \quad & \frac{\varphi_{\hat{\rho}}^{-1}(\int \hat{R}_n(1-x,x)\mathrm{d}x) - \varphi_{\hat{\rho}}^{-1}(\int R(1-x,x)\mathrm{d}x)}{\int \hat{R}_n(1-x,x)\mathrm{d}x - \int R(1-x,x)\mathrm{d}x} \\
& \cdot\sqrt{k}\left(\int \hat{R}_n(1-x,x)\mathrm{d}x - \int_{[0,1]} R(1-x,x)\mathrm{d}x\right) \\
& + \sqrt{k}\left(\varphi_{\hat{\rho}}^{-1}(\int R(1-x,x)\mathrm{d}x) - \varphi_{\rho_0}^{-1}(\int R(1-x,x)\mathrm{d}x)\right).
\end{aligned}$$

By Lemma 3.6.4 it follows that

$$\sqrt{k}\,(\hat{\nu}_n - \nu_0) = c(1 + o_{\mathbb{P}}(1))\sqrt{k}\left(\int \hat{R}_n(1-x,x)\mathrm{d}x - \int R(1-x,x)\mathrm{d}x\right) + o_{\mathbb{P}}(1). \tag{3.6.16}$$

Combining (3.6.8) and (3.6.16) we conclude that

$$\left(\sqrt{k}(\hat{\nu}_n - \nu_0), \sqrt{n}(\hat{\rho} - \rho_0)\right) \xrightarrow{d} (N_\nu, N_\rho),$$

where N_ν and N_ρ are independent, and if σ_R^2 is the variance of $\int B(1-x,x)\mathrm{d}x$, we have that $\sigma_\nu^2 = c^2 \sigma_R^2$. □

Chapter 4

An M-Estimator of Tail Dependence in Arbitrary Dimensions

[Based on joint project with J.H.J. Einmahl and J. Segers, *An M-Estimator of Tail Dependence in Arbitrary Dimensions*, work in progress.]

Abstract. Consider the situation of a random sample from a multivariate distribution in the max-domain of attraction of an extreme-value distribution. Assume that the dependence structure of the extreme-value attractor belongs to a given parametric model. A new estimator for the unknown parameter vector of the model is proposed. The estimator is an extension of the one introduced in J.H.J. Einmahl, A. Krajina and J. Segers [Method of Moments Estimator of Tail Dependence, Bernoulli 14(4), 2008] in two respects: (i) the number of variables is arbitrary; (ii) the number of moment equations can exceed the dimension of the parameter space. More precisely, the estimator is defined as the value of the parameter vector that minimizes the distance between a vector of weighted integrals of the tail dependence function on the one hand and empirical counterparts of these integrals on the other hand. Under minimal conditions, this minimization problem has with probability tending to one a unique, global solution. The estimator is consistent and asymptotically normal. The asymptotic covariance matrix can be estimated consistently as well, allowing for the construction of asymptotic confidence regions. The method, not being likelihood based, applies to discrete

and continuous models alike. We demonstrate the performance and applicability of the estimator on examples.

4.1 Introduction

As the number of variables increases, modeling tail dependence becomes more complex. For instance, in dimension d there are $d(d-1)/2$ bivariate marginals, which in general can be different up to some consistency requirements. In order to simplify the problem, it is customary to impose a parametric model. Here, we will assume that the stable tail dependence function l, defined by

$$l(x) = \lim_{t \downarrow 0} t^{-1} \mathbb{P}\left(1 - F_1(X_{11}) \leq tx_1 \text{ or } \ldots \text{ or } 1 - F_d(X_{1d}) \leq tx_d\right),$$

belongs to a parametric family, $l \in \{l(\cdot;\theta)\colon \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}^p$. The existing estimators of θ are all likelihood based and as such, apply only to d times differentiable functions l; see Joe, Smith, and Weissman (1992); Smith (1994); Ledford and Tawn (1996); de Haan, Neves, and Peng (2008); Coles and Tawn (1991).

In Chapter 2, we introduced the method of moments estimator of the bivariate parametric stable tail dependence function, which did not require differentiability of l. Here we extend that estimator in two directions. First we consider models in arbitrary dimensions. Second, we extend the method of moments estimation to general M-estimation by allowing for more estimating equations than parameters. If $\theta \in \Theta \subseteq \mathbb{R}^p$ is the unknown parameter, $g\colon [0,1]^d \to \mathbb{R}^q$, $q \geq p$ is an auxiliary function and $\hat{l}_n$ is the nonparametric estimator of l, we define $\hat{\theta}_n$, the M-estimator of θ, as the minimizer of the Euclidean distance in $\mathbb{R}^q$ between

$$\int_{[0,1]^d} g(x)\hat{l}_n(x)\mathrm{d}x \quad \text{and} \quad \int_{[0,1]^d} g(x)l(x;\theta)\mathrm{d}x.$$

Under minimal conditions, the unique, global minimizer exists with probability tending to one, and it is a consistent and asymptotically normal estimator of θ.

The absence of smoothness assumptions on l makes it possible to estimate the tail dependence structure of factor models like (1.2.3). The factor model in (1.2.3) is used in our simulation studies. The same tail dependence structure appears in the factor model $X = (X_1, \dots, X_d)$, with

$$X_j = \sum_{i=1}^{r} b_{ij} Z_i + \varepsilon_j, \quad j = 1, \dots, d,$$

where all b_{ij} are nonnegative, $\sum_i b_{ij} = 1$ for all j, Z_i are independent heavy tailed factors (for example, they could be standard Fréchet distributed, $F(x) = \exp\{-1/x\}$), and ε_j, $j = 1, \dots, d$, are independent random variables which are lighter tailed than the main factors and which are independent of them. This kind of factor model is often used in finance, for example in credit risk modeling or in stock returns modeling; see Fama and French (1993); Malevergne and Sornette (2004); Geluk, de Haan, and de Vries (2007).

The organization of this chapter is as follows. The basics of the tail dependence structures in multivariate models are presented in Section 4.2. The M-estimator is defined in Section 4.3. Section 4.4 contains the main theoretical results: consistency and asymptotic normality of the M-estimator, and some consequences of the asymptotic normality result that can be used for construction of confidence regions and testing. In Section 4.5 we apply the M-estimator on the well-known logistic stable tail dependence function. The tail dependence structure of factor models is studied and estimated in Section 4.6. The behavior of the M-estimator is illustrated on two examples. The proofs are deferred to Section 4.7.

4.2 Tail dependence

We will write points in $\mathbb{R}^d$ as $x := (x_1, \dots, x_d)$ and random vectors as $X_i := (X_{i1}, \dots, X_{id})$, for $i = 1, \dots, n$. Let $X_1, \dots, X_n$ be independent random vectors in $\mathbb{R}^d$ with common continuous distribution function F and marginal distribution functions $F_1, \dots, F_d$. We assume that F has a *stable tail dependence function* l, that is, we assume that for all $x = (x_1, \dots, x_d) \in [0, \infty)^d$ the following limit exists:

$$\lim_{t \downarrow 0} t^{-1} \mathbb{P}\left(1 - F_1(X_{11}) \le t x_1 \text{ or } \dots \text{ or } 1 - F_d(X_{1d}) \le t x_d\right) = l(x). \tag{4.2.1}$$

The function l has the following properties:

- $l(x_1, 0, \ldots, 0) = \cdots = l(0, \ldots, 0, x_1) = x_1$ for any $x_1 > 0$;
- $\max\{x_1, \ldots, x_d\} \leq l(x) \leq x_1 + \cdots + x_d$;
- l is convex: $l(\lambda x + (1-\lambda)y) \leq \lambda l(x) + (1-\lambda)l(y)$, for $\lambda \in [0,1]$ and $x, y \in [0,\infty)^d$; and
- l is homogeneous of order one: $l(tx_1, \ldots, tx_d) = t\, l(x_1, \ldots, x_d)$, for all $(x_1, \ldots, x_d) \in [0,\infty)^d$ and $t > 0$.

Let $\Delta_{d-1} := \{w \in \mathbb{R}^d : w_j \geq 0,\ w_1 + \cdots + w_d = 1\}$ be the unit simplex in $\mathbb{R}^d$. A finite Borel measure H on Δ_{d-1} satisfying the d moment conditions

$$\int_{\Delta_{d-1}} w_j H(\mathrm{d}w) = 1, \quad j = 1, \ldots, d, \tag{4.2.2}$$

is called a *spectral* or *angular measure*. There is a one-to-one correspondence between the stable tail dependence function and the spectral measure: it holds that there exists a unique spectral measure H such that

$$l(x) = \int_{\Delta_{d-1}} \max_{j=1,\ldots,d} \{w_j x_j\} H(\mathrm{d}w). \tag{4.2.3}$$

It can be shown that there exists a measure Λ on $[0,\infty]^d \setminus \{(\infty, \ldots, \infty)\}$ such that

(1) $l(x) = \Lambda\left(\{u \in [0,\infty]^d : u_1 \leq x_1 \text{ or } \ldots \text{ or } u_d \leq x_d\}\right)$,

(2) $\Lambda(tA) = t\Lambda(A)$, for any $t > 0$ and any Borel set $A \subset [0,\infty]^d \setminus \{(\infty, \ldots, \infty)\}$, with $tA := \{tx : x \in A\}$,

see for example Resnick (1987); Beirlant, Goegebeur, Segers, and Teugels (2004); de Haan and Ferreira (2006). The measure Λ is called the *exponent measure* and it is yet another way of defining the tail dependence structure. Property (1) above connects the exponent measure and the function l. If μ is the measure Λ after the transformation $(x_1, \ldots, x_d) \mapsto (1/x_1, \ldots, 1/x_d)$, the relationship between the spectral measure H and the exponent measure Λ (and μ) is given by

$$H(B) = \mu\left(\left\{x \in [0,\infty)^d \colon \sum_{j=1}^d x_j \geq 1, x/\sum_{j=1}^d x_j \in B\right\}\right),$$

for any Borel set B on Δ_{d-1}. By (2) we get that for any $t > 0$ and any Borel set B on Δ_{d-1},

$$\frac{1}{t}\,H(B) = \mu\left(\left\{x \in [0,\infty)^d \colon \sum_{j=1}^d x_j \geq t, x/\sum_{j=1}^d x_j \in B\right\}\right),$$

which is a version of the spectral decomposition of the exponent measure, see de Haan and Resnick (1977) or Resnick (1987).

The right-hand partial derivatives of l always exist; indeed, by bounded convergence it follows that for $j = 1, \ldots, d$, as $h \downarrow 0$,

$$\begin{aligned}
&\frac{1}{h}\big(l(x_1, \ldots, x_{j-1}, x_j + h, x_{j+1}, \ldots, x_d) - l(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_d)\big) \\
&= \int_{\Delta_{d-1}} \frac{1}{h}\Big(\max\{w_j x_j + w_j h, \max_{s \neq j}\{w_s x_s\}\} - \max\{w_j x_j, \max_{s \neq j}\{w_s x_s\}\}\Big) H(\mathrm{d}w) \\
&\qquad\qquad \to \int_{\Delta_{d-1}} w_j \mathbf{1}\{w_j x_j \geq \max_{s \neq j}\{w_s x_s\}\} H(\mathrm{d}w). \qquad (4.2.4)
\end{aligned}$$

Similarly, the left-hand partial derivatives exist for all $x \in (0,\infty)^d$. By convexity, the function l is almost everywhere continuously differentiable, with its gradient vector of (right-hand) partial derivatives being given by (4.2.4). Also, if l is differentiable, its partial derivatives are continuous.

4.3 Estimation

Let R_i^j denote the rank of X_{ij} among $X_{1j}, \ldots, X_{nj}$, $i = 1, \ldots, n$, $j = 1, \ldots, d$. For $k \in \{1, \ldots, n\}$, define a nonparametric estimator of l by

$$\hat{l}_n(x) := \frac{1}{k} \sum_{i=1}^{n} \mathbf{1} \left\{ R_i^1 > n + \frac{1}{2} - kx_1 \text{ or } \ldots \text{ or } R_i^d > n + \frac{1}{2} - kx_d \right\}.$$

When we study asymptotic properties of this estimator, $k = k_n$ is an intermediate sequence, that is, $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. In the literature, the stable tail dependence function is often modeled parametrically. We impose that the stable tail dependence function l belongs to some parametric family $\{l(\,\cdot\,;\theta) : \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^p$, $p \geq 1$ and we propose an M-estimator of θ. Let $q \geq p$. Let $g \equiv (g_1, \ldots, g_q)^T : [0,1]^d \to \mathbb{R}^q$ be an integrable function such that $\varphi : \Theta \to \mathbb{R}^q$ defined by

$$\varphi(\theta) := \int_{[0,1]^d} g(x) l(x;\theta) \, \mathrm{d}x \tag{4.3.1}$$

is a homeomorphism between Θ and its image $\varphi(\Theta)$. Let θ_0 denote the true parameter value. The M-estimator $\hat{\theta}_n$ of θ_0 is defined as a minimizer of the criterion function

$$\begin{aligned} Q_{k,n}(\theta) &= \|\varphi(\theta) - \textstyle\int g\hat{l}_n\|^2 \\ &= \sum_{m=1}^{q} \left(\int_{[0,1]^d} g_m(x) \left(\hat{l}_n(x) - l(x;\theta) \right) \mathrm{d}x \right)^2, \end{aligned}$$

where $\|\cdot\|$ is the L_2 norm. In other words, if $\hat{Y}_n = \arg\min_{y \in \varphi(\Theta)} \|y - \int g\hat{l}_n\|$, then $\hat{\theta}_n \in \varphi^{-1}(\hat{Y}_n)$. Later we show that $\hat{\theta}_n$ is, with probability tending to one, unique.

4.4 Main results

4.4.1 Notation

Let W_Λ be a mean-zero Wiener process indexed by Borel sets of $[0,\infty]^d\setminus\{(\infty,\ldots,\infty)\}$ with "time" Λ: its covariance structure is given by

$$\mathbb{E}W_\Lambda(A_1)W_\Lambda(A_2) = \Lambda(A_1 \cap A_2),$$

for any two Borel sets A_1 and A_2 in $[0,\infty]^d \setminus \{(\infty,\ldots,\infty)\}$. Define

$$W_l(x) := W_\Lambda(\{u \in [0,\infty]^d \setminus \{(\infty,\ldots,\infty)\} : u_1 \le x_1 \text{ or } \ldots \text{ or } u_d \le x_d\}).$$

Let $W_j,\ j = 1,\ldots,d$, be the marginal processes

$$W_j(x_j) := W_l(0,\ldots,0,x_j,0,\ldots,0), \ \ x_j \ge 0.$$

Define $l_j(x)$ to be the right-hand partial derivatives of l with respect to x_j, $j = 1,\ldots,d$; if l is differentiable, these are equal to the partial derivatives of l, see (4.2.4). Denote

$$B(x) := W_l(x) - \sum_{j=1}^{d} l_j(x)W_j(x_j),$$

and let

$$\tilde{B} := \int_{[0,1]^d} g(x)B(x)\mathrm{d}x.$$

The distribution of $\tilde{B}$ is zero-mean Gaussian with covariance matrix

$$\begin{aligned}\Sigma &:= \mathbb{E}\Big[\int_{[0,1]^d} g(x)B(x)\mathrm{d}x \cdot \int_{[0,1]^d} g(y)^T B(y)\mathrm{d}y\Big] \\ &= \iint_{([0,1]^d)^2} \mathbb{E}[B(x)B(y)]g(x)g(y)^T \mathrm{d}x\mathrm{d}y \in \mathbb{R}^{q\times q}. \quad (4.4.1)\end{aligned}$$

Note that if l is parametric, Σ depends on the parameter, that is $\Sigma = \Sigma(\theta)$.

Let $\nabla Q_{k,n}(\theta) \in \mathbb{R}^{p\times 1}$ be the gradient vector of $Q_{k,n}$ at θ; for every $x \in [0,1]^d$ let $\nabla l(x;\theta) \in \mathbb{R}^{p\times 1}$ be the gradient vector of $l(x;\cdot)$ in θ; let $\dot{\varphi}(\theta) \in \mathbb{R}^{q\times p}$ be the total derivative of φ at θ; and put

$$V(\theta) := 4\dot{\varphi}(\theta)^T \Sigma(\theta)\dot{\varphi}(\theta) \in \mathbb{R}^{p\times p}.$$

Further let $\mathcal{H}_{k,n}(\theta) \in \mathbb{R}^{p\times p}$ denote the Hessian matrix of $Q_{k,n}$ in θ. Let $\mathcal{H}(\theta)$ be the deterministic and symmetric matrix given by its (i,j)th element, $i = 1,\ldots,p$, $j = 1,\ldots,p$,

$$(\mathcal{H}(\theta))_{ij} = 2\left(\frac{\partial}{\partial\theta_i}\varphi(\theta)\right)^T \left(\frac{\partial}{\partial\theta_j}\varphi(\theta)\right) - 2\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\varphi(\theta)\right)^T (\varphi(\theta_0) - \varphi(\theta)).$$

Observe that

$$\mathcal{H}(\theta_0) = 2\dot{\varphi}(\theta_0)^T\dot{\varphi}(\theta_0),$$

and define

$$M(\theta) := \left(\dot{\varphi}(\theta)^T\dot{\varphi}(\theta)\right)^{-1} \dot{\varphi}(\theta)^T\Sigma(\theta)\dot{\varphi}(\theta) \left(\dot{\varphi}(\theta)^T\dot{\varphi}(\theta)\right)^{-1} \in \mathbb{R}^{p\times p}. \tag{4.4.2}$$

For the proofs of the results we require subsets of the following list of conditions:

(C1) $t^{-1}\mathbb{P}\left(1 - F_1(X_{11}) \leq tx_1 \text{ or } \ldots \text{ or } 1 - F_d(X_{1d}) \leq tx_d\right) - l(x) = O(t^\alpha)$, uniformly in $x \in \Delta_{d-1}$ as $t \downarrow 0$, for some $\alpha > 0$;

(C2) $k = o\left(n^{\frac{2\alpha}{1+2\alpha}}\right)$, for the α of (C1);

(C3) l is differentiable;

(C4) φ is twice continuously differentiable and $\dot{\varphi}(\theta_0)$ is of full rank.

4.4.2 Results

Let $\hat{\Theta}_n$ be the set of minimizers of $Q_{k,n}$,

$$\hat{\Theta}_n := \underset{\theta\in\Theta}{\arg\min}\, \|\varphi(\theta) - \textstyle\int g\hat{l}_n\|^2.$$

Note that $\hat{\Theta}_n$ may be empty or may contain more than one element.

Theorem 4.4.1 (Existence, uniqueness and consistency of $\hat{\theta}_n$)**.** *Let $g : [0,1]^d \to \mathbb{R}^q$ be integrable.*

(i) If φ is a homeomorphism from Θ to $\varphi(\Theta)$ and if there exists $\varepsilon_0 > 0$ such that the set $\{\theta \in \Theta : \|\theta - \theta_0\| \leq \varepsilon_0\}$ is closed, then for every ε such that $\varepsilon_0 \geq \varepsilon > 0$,

$$\mathbb{P}\left(\hat{\Theta}_n \neq \varnothing \text{ and } \hat{\Theta}_n \subseteq \{\theta \in \Theta : \|\theta - \theta_0\| \leq \varepsilon\}\right) \to 1,$$

as $n \to \infty$, $k \to \infty$ and $k/n \to 0$.

(ii) If in addition to the assumptions of (i), condition (C4) holds, then, with probability tending to one, $Q_{k,n}$ has a unique minimizer $\hat{\theta}_n$. Hence $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$, as $n \to \infty$, $k \to \infty$ and $k/n \to 0$.

We prove the asymptotic normality of $\hat{l}_n$. This result is of independent interest and can be found in the literature for $d = 2$ only, see Huang (1992); Drees and Huang (1998). Here it is used as a main tool for asymptotic normality of $\hat{\theta}_n$. The result is stated in an approximation setting, where $\hat{l}_n$ and B are defined on the same probability space obtained by a Skorohod construction. The random quantities involved are only in distribution equal to the original ones, but for convenience this is not expressed in the notation.

Theorem 4.4.2 (Asymptotic normality of $\hat{l}_n$ in arbitrary dimensions)**.** *If the conditions (C1), (C2), (C3) hold, then for every $T > 0$, as $n \to \infty$ and $k \to \infty$,*

$$\sup_{x \in [0,T]^d} \left| \sqrt{k}\left(\hat{l}_n(x) - l(x)\right) - B(x) \right| \xrightarrow{\mathbb{P}} 0. \tag{4.4.3}$$

Theorem 4.4.3 (Asymptotic normality of $\hat{\theta}_n$)**.** *If in addition to the assumptions of Theorem 4.4.1(i), the conditions (C1), (C2), (C4) hold, then as $n \to \infty$ and $k \to \infty$,*

$$\sqrt{k}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, M(\theta_0)). \tag{4.4.4}$$

The following consequence of Theorem 4.4.3 can be used for construction of the confidence regions.

Corollary 4.4.4. *If in addition to conditions of Theorem 4.4.3, the mapping $\theta \mapsto H_\theta$ is weakly continuous at θ_0 and if $M(\theta_0)$ is non-singular, then as $n \to \infty$ and $k \to \infty$,*

$$k(\hat{\theta}_n - \theta_0)^T M(\hat{\theta}_n)^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \chi_p^2. \tag{4.4.5}$$

Let $1 \leq r < p$ and $\theta = (\theta_1, \theta_2) \in \Theta \subset \mathbb{R}^p$, where $\theta_1 \in \mathbb{R}^{p-r}$, $\theta_2 \in \mathbb{R}^r$. We want to test $\theta_2 = \theta_2^*$ against $\theta_2 \neq \theta_2^*$, where θ_2^* corresponds to the submodel. Denote $\hat{\theta}_n = (\hat{\theta}_{1n}, \hat{\theta}_{2n})$, and let $M_2(\theta)$ be the $r \times r$ matrix corresponding to the lower right corner of M, as in (4.4.6) below,

$$M = \left(\begin{array}{c|c} \cdots & \cdots \\ \hline \cdots & M_2 \end{array}\right) \in \mathbb{R}^{p \times p}. \tag{4.4.6}$$

Corollary 4.4.5 (Test)**.** *If the assumptions of Corollary 4.4.4 are satisfied, and $\theta_0 = (\theta_1, \theta_2^*) \in \Theta$ for some θ_1, then as $n \to \infty$ and $k \to \infty$,*

$$k(\hat{\theta}_{2n} - \theta_2^*)^T M_2(\hat{\theta}_{1n}, \theta_2^*)^{-1}(\hat{\theta}_{2n} - \theta_2^*) \xrightarrow{d} \chi_r^2. \tag{4.4.7}$$

The above result can be used for testing for a submodel. For example, we could test for the symmetric logistic model within the asymmetric logistic one, or for the 2-factor model within the 3-factor one.

Remark 4.4.6. The matrices M and M_2 are needed for the computation of the confidence regions and the test statistic. Computing these matrices can be challenging, mostly due to the complicated expression for Σ, see (4.4.1).

4.5 Example 1: Logistic model

The stable tail dependence function corresponding to the multivariate symmetric logistic model is

$$l_\theta(x_1, \ldots, x_d) = \left(x_1^{1/\theta} + \cdots + x_d^{1/\theta}\right)^\theta, \tag{4.5.1}$$

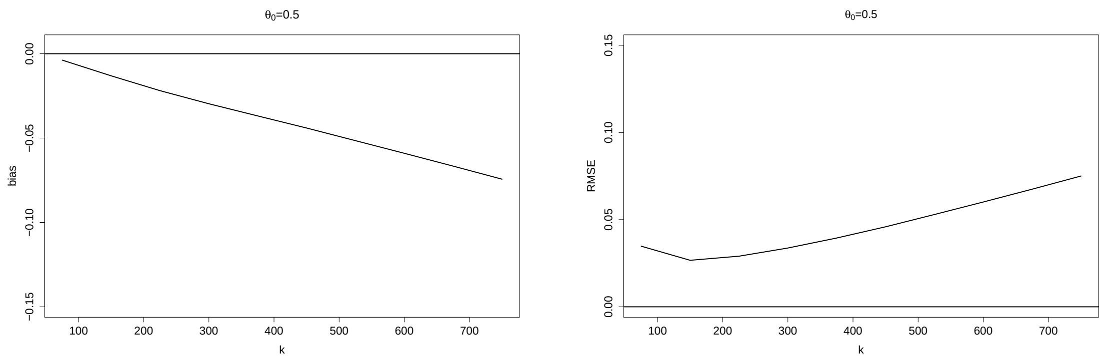
with $\theta \in [0, 1]$, $x_j > 0$, $j = 1, \ldots, d$. The bivariate logistic tail dependence function was first introduced in Gumbel (1960), and it is one of the oldest parametric models of tail dependence. The multivariate logistic distribution function defined by

$$F_\theta(x_1, \ldots, x_d) = \exp\left\{ - \left(\sum_{j=1}^{d} x_j^{-1/\theta} \right)^{\theta} \right\},$$

$x_1 > 0, \ldots, x_d > 0$, with $\theta \in [0, 1]$ has the stable tail dependence function l_θ above.

We simulate 500 samples of size $n = 3000$ from a five-dimensional logistic distribution function with $\theta_0 = 0.5$. We define $\hat{\theta}_n$, an M-estimator of θ_0, by choosing $g_1 \equiv 1$ and $g_2(x) = 2^5 x_1 \cdot \cdots \cdot x_5$. The bias and the Root Mean Squared Error (RMSE) of this estimator are shown in Figure 4.1.

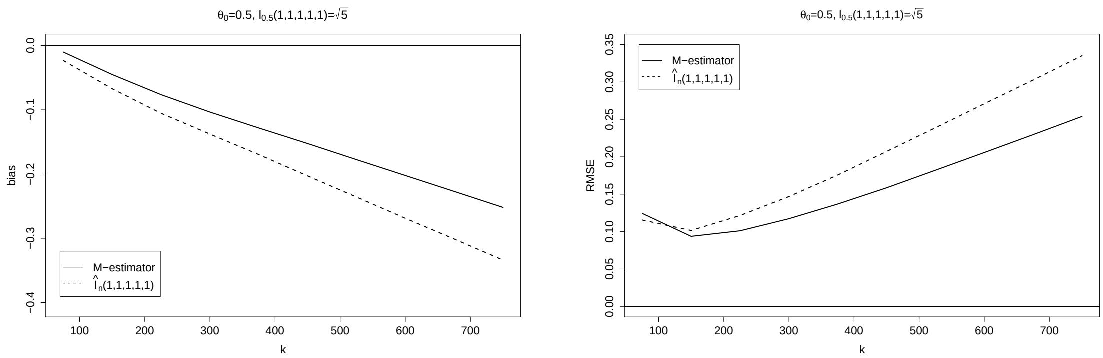
Also, we consider the estimation of $l_\theta(1,1,1,1,1)$, based on $\hat{\theta}_n$ defined above. From (4.5.1) it follows that $l_\theta(1,1,1,1,1) = 5^\theta$. The estimator of this quantity is then defined as $5^{\hat{\theta}_n}$. Since $\theta_0 = 0.5$, the true parameter here is $5^{\theta_0} = \sqrt{5}$. We compare the bias and the RMSEs of this estimator and the nonparametric estimator $\hat{l}_n(1,1,1,1,1)$. Figure 4.2 shows that our estimator performs better than the nonparametric estimator for almost every choice of k.



(a) Bias of estimator of θ

(b) RMSE of estimator of θ

Figure 4.1: Five-dimensional logistic model, $\theta_0 = 0.5$



(a) Bias of estimator of $l_{0.5}(1,1,1,1,1)$

(b) RMSE of estimator of $l_{0.5}(1,1,1,1,1)$

Figure 4.2: Five-dimensional logistic model, $\theta_0 = 0.5$

4.6 Example 2: Discrete spectral measure

Consider the r-factor model, $r \in \mathbb{N}$, in dimension d: $X = (X_1, \ldots, X_d)$ and

$$X_j = \max_{i=1,\ldots,r} \{a_{ij} Z_i\}, \qquad j \in \{1, \ldots, d\}$$

with Z_i independent Fréchet(ν) random variables, $\nu > 0$, and a_{ij} nonnegative constants such that $\sum_j a_{ij} > 0$ for all i. Let $W_i = Z_i^\nu$, $i = 1, \ldots, r$, and note that the W_i are standard Fréchet random variables. Define a d-dimensional random vector $Y = (Y_1, \ldots, Y_d)$ by

$$Y_j := X_j^\nu = \max_{i=1,\ldots,r} \{a_{ij}^\nu W_i\}, \qquad j \in \{1, \ldots, d\}.$$

It is easily seen that as $x \to \infty$,

$$1 - F_{Y_j}(x) = 1 - \exp\left\{ -\frac{\sum_{i=1}^r a_{ij}^\nu}{x} \right\} \sim 1 - \frac{\sum_{i=1}^r a_{ij}^\nu}{x}.$$

Since X_j are monotone transformations of the Y_j, the (tail) dependence structure of X and Y is the same. We will determine the tail dependence function l and the spectral measure H of X.

Lemma 4.6.1. *For $x = (x_1, \ldots, x_d) \in [0, \infty)^d$,*

$$l(x_1, \ldots, x_d) = \sum_{i=1}^{r} \max_{j=1,\ldots,d} \{b_{ij} x_j\}, \quad \textit{with} \quad b_{ij} := \frac{a_{ij}^{\nu}}{\sum_{i=1}^{r} a_{ij}^{\nu}}.$$

Next, we are looking for a measure H on the unit simplex $\Delta_{d-1} = \{w \in [0, \infty)^d : w_1 + \cdots + w_d = 1\}$ such that for all $x \in [0, \infty)^d$,

$$\sum_{i=1}^{r} \max_{j=1,\ldots,d} \{b_{ij} x_j\} = l(x_1, \ldots, x_d) = \int_{\Delta_{d-1}} \max_{j=1,\ldots,d} \{w_j x_j\} \, H(\mathrm{d}w).$$

This H must be a discrete measure with r atoms given by

$$\left(\frac{b_{i1}}{\sum_j b_{ij}}, \ldots, \frac{b_{id}}{\sum_j b_{ij}} \right), \qquad i \in \{1, \ldots, r\},$$

the atom receiving mass $\sum_j b_{ij}$, which is positive by assumption. Note that H is indeed a spectral measure, for

$$\int_{\Delta_d} w_j \, H(\mathrm{d}w) = \sum_{i=1}^{r} b_{ij} = 1, \qquad j \in \{1, \ldots, d\}. \tag{4.6.1}$$

Every discrete spectral measure can arise in this way.

The spectral measure is completely determined by the rd parameters b_{ij}, but by the d moment conditions from (4.6.1), the actual number of parameters is $p = d(r-1)$. The parameter vector $\theta \in \mathbb{R}^p$, which is to be estimated, can be constructed in many ways. For identification purposes, the definition of θ should be unambiguous. We opt for the following approach. Consider the matrix of the coefficients b_{ij},

$$\begin{pmatrix} b_{11} & \cdots & b_{r1} \\ \vdots & \ddots & \vdots \\ b_{1d} & \cdots & b_{rd} \end{pmatrix} \in \mathbb{R}^{d \times r}.$$

In the ith column of this matrix are the coefficients corresponding to the ith factor, $i = 1, \ldots, r$. We define θ by stacking the above columns in decreasing order of their sums, leaving out the column with the lowest sum. (If two columns have the same sum, we order them then in decreasing order lexicographically.)

The definition of the M-estimator of θ involves integrals of the form

$$\int_{[0,1]^d} g_s(x) l(x)\, \mathrm{d}x = \sum_{i=1}^{r} \int_{[0,1]^d} g_s(x) \max_{j=1,\dots,d} \{b_{ij} x_j\}\, \mathrm{d}x,$$

where $g_s : [0,1]^d \to \mathbb{R}$ is integrable and $s = 1, \dots, q$.

A possible choice is $g_s(x) = x_k^m$, where $k \in \{1, \dots, d\}$ and $m \geq 0$.

Lemma 4.6.2. *It holds that*

$$\int_{[0,1]^d} x_k^m l(x)\, dx = \sum_{i=1}^{r} \sum_{j=1}^{d} \frac{b_{ij}}{1 + m(1 - \delta_{jk})} \int_0^1 \left(\frac{b_{ij}}{b_{ik}} x \wedge 1 \right)^m \prod_{l=1}^{d} \left(\frac{b_{ij}}{b_{il}} x \wedge 1 \right) dx,$$

where δ_{jk} *is* 1 *if* $j = k$ *and* 0 *if* $j \neq k$.

The integral on the right-hand side is to be computed numerically.

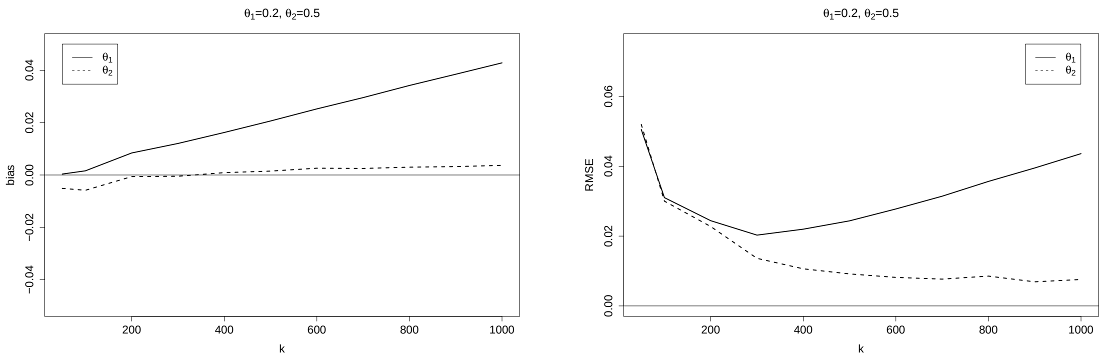
We illustrate the performance of the M-estimator of the unknown parameters on two factor models: a four-dimensional factor model with 2 factors ($p = 4 \times 1 = 4$), and a three-dimensional factor model with 3 factors ($p = 3 \times 2 = 6$).

Four-dimensional model with 2 factors. We simulated 500 samples of size $n = 5000$ from a four-dimensional model

$$\begin{aligned} X_1 &= 0.8 Z_1 \vee 0.2 Z_2 \\ X_2 &= 0.5 Z_1 \vee 0.5 Z_2 \\ X_3 &= 0.3 Z_1 \vee 0.7 Z_2 \\ X_4 &= 0.1 Z_1 \vee 0.9 Z_2, \end{aligned}$$

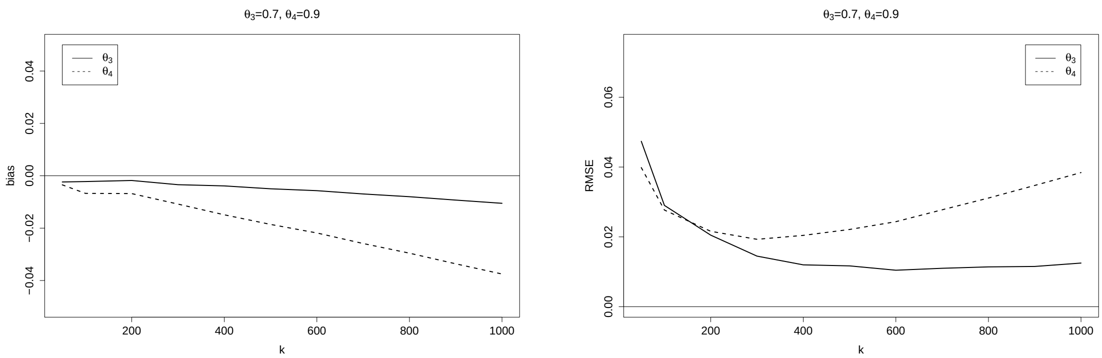
with independent standard Fréchet distributed factors Z_1 and Z_2. The unknown parameter θ consists of $p = d(r - 1) = 4$ elements, and we could choose θ to be the coefficients of the first factor $(0.8, 0.5, 0.3, 0.1)$ or the coefficients of the second factor $(0.2, 0.5, 0.7, 0.9)$, for example. Since the latter quadruple has a higher sum, it in a sense corresponds to a more important factor, and we choose $\theta = (0.2, 0.5, 0.7, 0.9)$.

In Figures 4.3 and 4.4 we show the bias and the RMSE of the M-estimator based on $q = 5$ moment equations, with auxiliary functions $g_i(x) = x_i$, for $i = 1, 2, 3, 4$ and $g_5 \equiv 1$. Estimation in this particular example benefited from the extension of the method of moments estimator to the M-estimator. Adding a fifth moment equation via $g_5 \equiv 1$ to $g_i(x) = x_i$, $i = 1, 2, 3, 4$, reduced the RMSE of the estimator in most cases and for most values of k.



(a) **(b)**

Figure 4.3: Four-dimensional 2-factor model, estimation of $\theta_1 = 0.2$, $\theta_2 = 0.5$



(a) **(b)**

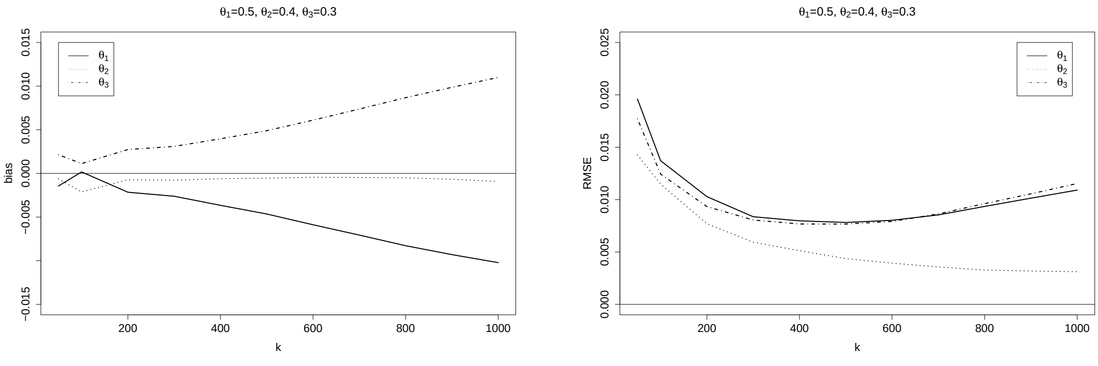
Figure 4.4: Four-dimensional 2-factor model, estimation of $\theta_3 = 0.7$, $\theta_4 = 0.9$

Three-dimensional model with 3 factors. We simulated 500 samples of size $n = 5000$ from a three-dimensional model

$$\begin{aligned} X_1 &= 0.2Z_1 \vee 0.5Z_2 \vee 0.3Z_3 \\ X_2 &= 0.5Z_1 \vee 0.4Z_2 \vee 0.1Z_3 \\ X_3 &= 0.3Z_1 \vee 0.3Z_2 \vee 0.4Z_3, \end{aligned}$$

with independent Z_1, Z_2 and Z_3 following the standard Fréchet distribution. The unknown parameter θ consists of $p = d(r-1) = 6$ elements. According to the above described method for constructing the parameter vector, $\theta = (0.5, 0.4, 0.3, 0.2, 0.5, 0.3)$. The $q = p = 6$ auxiliary functions are $g_i(x) = x_i$, $g_{i+3}(x) = x_i^2$, $i = 1, 2, 3$. The estimator is a method of moments estimator since the number of the functions g we used corresponds to the number of parameters.

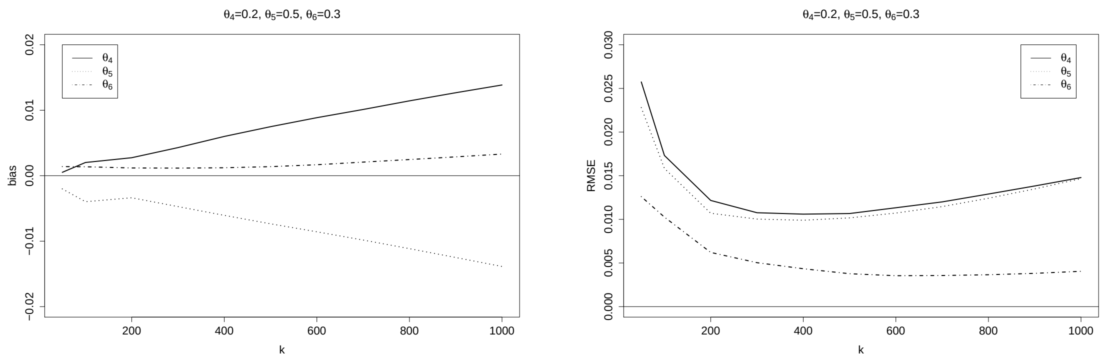
Figure 4.5 and Figure 4.6 show the bias and the RMSE of the estimator of θ. The 6 parameters seem to have been estimated rather well, the root mean squared errors being in the range from 0.01 up to 0.015. However, unlike in the previous example, adding an extra equation corresponding to $g_7 \equiv 1$ did not prove effective here, in the sense that the smaller RMSEs appeared mostly in the case of method of moments estimation.



(a) **(b)**

Figure 4.5: Three-dimensional 3-factor model, estimation of $\theta_1 = 0.5$, $\theta_2 = 0.4$, $\theta_3 = 0.3$

Remark 4.6.3. The three examples we have presented show good performance of the estimator, but they also show how this performance can depend on the choice of function g. The optimal choice of g is a difficult issue, which we do not



(a) **(b)**

Figure 4.6: Three-dimensional 3-factor model, estimation of $\theta_4 = 0.2$, $\theta_5 = 0.5$, $\theta_6 = 0.3$

address. The choice of g in the examples is driven by computational simplicity, see Lemma 4.6.2, for example.

4.7 Proofs

The asymptotic properties of the nonparametric estimator $\hat{l}_n$ are required for the proofs of the asymptotic properties of the M-estimator $\hat{\theta}_n$. Consistency of $\hat{l}_n$ (see (4.7.1)) for dimension $d = 2$ was shown in Huang (1992), cf. Drees and Huang (1998). In particular, it was shown that for every $T > 0$, as $n \to \infty$, $k \to \infty$ and $k/n \to 0$,

$$\sup_{(x_1,x_2)\in[0,T]^2} |\hat{l}_n(x_1,x_2) - l(x_1,x_2)| \xrightarrow{\mathbb{P}} 0.$$

The proof translates straightforwardly to general dimension d, and together with integrability of g yields consistency of $\int g\hat{l}_n$ for $\int gl = \varphi(\theta_0)$. Before the proof of Theorem 4.4.1, a technical result is needed.

Lemma 4.7.1. *If $k/n \to 0$ and if in addition to the assumptions of Theorem 4.4.1 condition (C4) holds, then as $n \to \infty$ and $k \to \infty$, on some closed neighborhood of θ_0:*

(i) $\mathcal{H}_{k,n}(\theta) \xrightarrow{\mathbb{P}} \mathcal{H}(\theta)$ *uniformly in θ, and*

(ii) $\mathbb{P}\left(\mathcal{H}_{k,n}(\theta) \text{ is positive definite}\right) \to 1.$

Proof (i) The Hessian matrix of $Q_{k,n}$ in θ is a $p \times p$ matrix with elements $(\mathcal{H}_{k,n}(\theta))_{ij}$, $i = 1, \ldots, p$, $j = 1, \ldots, p$ given by

$$\begin{aligned}
(\mathcal{H}_{k,n}(\theta))_{ij} &:= \frac{\partial^2 Q_{k,n}(\theta)}{\partial\theta_j \partial\theta_i} \\
&= 2\sum_{m=1}^{q} \int_{[0,1]^d} g_m(x) \frac{\partial}{\partial\theta_j} l(x;\theta) \mathrm{d}x \cdot \int_{[0,1]^d} g_m(x) \frac{\partial}{\partial\theta_i} l(x;\theta) \mathrm{d}x \\
&\quad -2\sum_{m=1}^{q} \int_{[0,1]^d} g_m(x) \frac{\partial^2}{\partial\theta_j \partial\theta_i} l(x;\theta) \mathrm{d}x \cdot \int_{[0,1]^d} g_m(x)(\hat{l}_n(x) - l(x;\theta)) \mathrm{d}x \\
&= 2\left(\frac{\partial}{\partial\theta_i}\varphi(\theta)\right)^T \left(\frac{\partial}{\partial\theta_j}\varphi(\theta)\right) - 2\left(\frac{\partial^2}{\partial\theta_i \partial\theta_j}\varphi(\theta)\right)^T \\
&\quad \cdot \left(\int_{[0,1]^d} g(x)\hat{l}_n(x) - \varphi(\theta)\right).
\end{aligned}$$

The consistency of $\int g\hat{l}_n$ for $\varphi(\theta_0)$ implies

$$\begin{aligned}
(\mathcal{H}_{k,n}(\theta))_{ij} &\xrightarrow{\mathbb{P}} 2\left(\frac{\partial}{\partial\theta_i}\varphi(\theta)\right)^T \left(\frac{\partial}{\partial\theta_j}\varphi(\theta)\right) - 2\left(\frac{\partial^2}{\partial\theta_i \partial\theta_j}\varphi(\theta)\right)^T (\varphi(\theta_0) - \varphi(\theta)) \\
&=: (\mathcal{H}(\theta))_{ij}.
\end{aligned}$$

Since we assumed that there exists $\varepsilon_0 > 0$ such that the set $\{\theta \in \Theta : \|\theta - \theta_0\| \leq \varepsilon_0\} =: B_{\varepsilon_0}(\theta_0)$ is closed, and since φ is assumed to be C^2, the second derivatives of φ are uniformly bounded on $B_{\varepsilon_0}(\theta_0)$, and hence, the convergence above is uniform on $B_{\varepsilon_0}(\theta_0)$.
(ii) For $\theta = \theta_0$ we get

$$(\mathcal{H}(\theta_0))_{ij} = 2\left(\frac{\partial}{\partial\theta_i}\varphi(\theta)\Big|_{\theta=\theta_0}\right)^T \left(\frac{\partial}{\partial\theta_j}\varphi(\theta)\Big|_{\theta=\theta_0}\right),$$

that is,

$$\mathcal{H}(\theta_0) = 2\dot{\varphi}(\theta_0)^T \dot{\varphi}(\theta_0).$$

Since $\dot{\varphi}(\theta_0)$ is assumed to be of full rank, $\mathcal{H}(\theta_0)$ is positive definite. For θ close to θ_0, $\mathcal{H}(\theta)$ is also positive definite. Due to the uniform convergence of $\mathcal{H}_{k,n}(\theta)$ to $\mathcal{H}(\theta)$ on $B_{\varepsilon_0}(\theta_0)$, the matrix $\mathcal{H}_{k,n}(\theta)$ is also positive definite on $B_{\varepsilon_0}(\theta_0)$ with probability tending to one. $\square$

Proof of Theorem 4.4.1 (i) Fix $\varepsilon > 0$ such that $0 < \varepsilon \le \varepsilon_0$. Since φ is a homeomorphism, there exists $\delta > 0$ such that $\theta \in \Theta$ and $\|\varphi(\theta) - \varphi(\theta_0)\| \le \delta$ implies $\|\theta - \theta_0\| \le \varepsilon$. In other words, for every $\theta \in \Theta$ such that $\|\theta - \theta_0\| > \varepsilon$, we have $\|\varphi(\theta) - \varphi(\theta_0)\| > \delta$. Hence, on the event

$$A_n = \{\|\varphi(\theta_0) - \textstyle\int g\hat{l}_n\| \le \delta/2\},$$

for every $\theta \in \Theta$ with $\|\theta - \theta_0\| > \varepsilon$, necessarily

$$\|\varphi(\theta) - \textstyle\int g\hat{l}_n\| \ge \|\varphi(\theta) - \varphi(\theta_0)\| - \|\varphi(\theta_0) - \int g\hat{l}_n\| > \delta - \delta/2 = \delta/2 \ge \|\varphi(\theta_0) - \int g\hat{l}_n\|.$$

As a consequence, on the event A_n, we have

$$\inf_{\theta:\|\theta-\theta_0\|>\varepsilon} \|\varphi(\theta) - \textstyle\int g\hat{l}_n\| > \min_{\theta:\|\theta-\theta_0\|\le\varepsilon} \|\varphi(\theta) - \int g\hat{l}_n\|.$$

Hence, on the event A_n, the "argmin" set $\hat{\Theta}_n$ is non-empty and is contained in the closed ball of radius ε centered at θ_0. Finally, $\mathbb{P}(A_n) \to 1$ by weak consistency of $\int g\hat{l}_n$ for $\int gl = \varphi(\theta_0)$.
(ii) In the proof of (i) we have seen that with probability tending to one the proposed M-estimator exists and it is contained in a closed ball around θ_0. In Lemma 4.7.1 we have shown that the criterion function is with probability tending to one strictly convex on such a closed ball around θ_0, and hence, with probability tending to one, the minimizer of the criterion function is unique. □

For $i = 1, \dots, n$ let

$$U_i := (U_{i1}, \dots, U_{id}) := (1 - F_1(X_{i1}), \dots, 1 - F_d(X_{id})),$$

and denote

$$\begin{aligned} Q_{nj}(u_j) &:= U_{\lceil nu_j \rceil:n,j},\ j = 1, \dots, d, \\ S_{nj}(x_j) &:= \frac{n}{k} Q_{nj}\left(\frac{kx_j}{n}\right),\ j = 1, \dots, d, \\ S_n(x) &:= (S_{n1}(x_1), \dots, S_{nd}(x_d)), \end{aligned}$$

where $U_{1:n,j} \leq \cdots \leq U_{n:n,j}$ are the order statistics of $U_{1j}, \ldots, U_{nj}$, $j = 1, \ldots, d$, and $\lceil a \rceil$ is the smallest integer not smaller than a. Write

$$\begin{aligned}
V_n(x) &:= \frac{n}{k}\mathbb{P}\left(U_{11} \leq \frac{kx_1}{n} \text{ or } \ldots \text{ or } U_{1d} \leq \frac{kx_d}{n}\right), \\
T_n(x) &:= \frac{1}{k}\sum_{i=1}^{n} \mathbf{1}\left\{U_{i1} < \frac{kx_1}{n} \text{ or } \ldots \text{ or } U_{id} < \frac{kx_d}{n}\right\}, \\
\hat{L}_n(x) &:= \frac{1}{k}\sum_{i=1}^{n} \mathbf{1}\left\{U_{i1} < \frac{k}{n}S_{n1}(x_1) \text{ or } \ldots \text{ or } U_{id} < \frac{k}{n}S_{nd}(x_d)\right\}, \\
&= \frac{1}{k}\sum_{i=1}^{n} \mathbf{1}\left\{R_i^1 > n+1-kx_1 \text{ or } \ldots \text{ or } R_i^d > n+1-kx_d\right\},
\end{aligned}$$

and note that

$$\hat{L}_n(x) = T_n(S_n(x)).$$

Since

$$\sup_{x\in[0,1]^d} \sqrt{k}\left|\hat{l}_n(x) - \hat{L}_n(x)\right| \leq \frac{d}{\sqrt{k}} \to 0, \tag{4.7.1}$$

the asymptotic properties of $\hat{l}_n$ and $\hat{L}_n$ are the same. With the notation $v_n(x) = \sqrt{k}(T_n(x) - V_n(x))$, we have the following result.

Proposition 4.7.2. *Let $T > 0$ and denote $A_x := \{u \in [0,\infty]^d : u_1 \leq x_1 \text{ or} \cdots \text{or } u_d \leq x_d\}$. There exists a sequence of processes $\tilde{v}_n$ such that for all n $\tilde{v}_n \stackrel{d}{=} v_n$ and there exist a Wiener process $W_l(x) := W_\Lambda(A_x)$ such that as $n \to \infty$,*

$$\sup_{x\in[0,2T]^d} |\tilde{v}_n(x) - W_l(x)| \xrightarrow{\mathbb{P}} 0. \tag{4.7.2}$$

The result follows from Theorem 3.1 in Einmahl (1997). From the proofs there it follows that a single Wiener process, instead of the sequence in the original statement of the theorem, can be used, and that convergence holds almost surely, instead of in probability, once the Skorohod construction is introduced. From now on, we work on this new (Skorohod) probability space, but keep the old notation, without the tildes. In particular we have convergence of the marginal processes:

$$\sup_{x_j\in[0,2T]} |v_{nj}(x) - W_j(x_j)| \to 0 \text{ a.s.}, \; j = 1, \ldots, d,$$

where $v_{nj}(x_j) := v_n((0,\dots,0,x_j,0,\dots,0))$. Vervaat's lemma, Vervaat (1972), implies

$$\sup_{x_j\in[0,2T]} |\sqrt{k}(S_{nj}(x_j)-x_j)+W_j(x_j)| \to 0 \text{ a.s.}, \; j=1,\dots,d. \tag{4.7.3}$$

Proof of Theorem 4.4.2 Write

$$\begin{aligned}
&\sqrt{k}\left(\hat{L}_n(x)-l(x)\right)\\
&= \sqrt{k}\,(T_n(S_n(x))-V_n(S_n(x)))+\sqrt{k}\,(V_n(S_n(x))-l(S_n(x)))\\
&\quad +\sqrt{k}\,(l(S_n(x))-l(x))\\
&=: D_1(x)+D_2(x)+D_3(x).
\end{aligned}$$

Proof of $\sup_{x\in[0,T]^d}|D_1(x)-W_l(x)| \xrightarrow{\mathbb{P}} 0.$

We have

$$D_1(x)=\sqrt{k}\,(T_n(S_n(x))-V_n(S_n(x)))=v_n(S_n(x)).$$

It holds that

$$\begin{aligned}
&\sup\nolimits_{x\in[0,T]^d}|D_1(x)-W_l(x)|\\
&\qquad\le \sup_{x\in[0,T]^d}|D_1(x)-W_l(S_n(x))|+\sup_{x\in[0,T]^d}|W_l(S_n(x))-W_l(x)|\,.
\end{aligned}$$

Because of (4.7.3), this is with probability tending to one less than or equal to

$$\sup_{y\in[0,2T]^d}|v_n(y)-W_l(y)|+\sup_{x\in[0,T]^d}|W_l(S_n(x))-W_l(x)|\,.$$

Both terms tend to zero in probability, the first one by Proposition 4.7.2, the second one because of the uniform continuity of W_l and (4.7.3).

Proof of $\sup_{x\in[0,T]^d}|D_2(x)| \xrightarrow{\mathbb{P}} 0.$

Because of (4.7.3), with probability tending to one, $\sup_{x\in[0,T]^d}|D_2(x)|$ is less than or equal to $\sup_{y\in[0,2T]^d}\sqrt{k}|V_n(y)-l(y)|$, which in turn, because of conditions (C1) and (C2), is equal to

$$\sqrt{k}\, O\left(\left(\frac{k}{n}\right)^{\alpha}\right) = O\left(\left(\frac{k}{n^{2\alpha/(1+2\alpha)}}\right)^{\frac{1}{2}+\alpha}\right) = o(1).$$

Proof of $\sup_{x\in[0,T]^d} |D_3(x) + \sum_{j=1}^d l_j(x)W_j(x_j)| \xrightarrow{\mathbb{P}} 0.$

Due to the existence of the first derivatives, we can use the mean value theorem to write

$$\frac{1}{\sqrt{k}} D_3(x) \quad = \quad l\left(S_n(x)\right) - l(x) = \sum_{j=1}^{d} (S_{nj}(x_j) - x_j) \cdot l_j(\xi_n),$$

with ξ_n between x and $S_n(x)$. Therefore

$$\sup_{x\in[0,T]^d} |D_3(x) + \sum_{j=1}^{d} l_j(x)W_j(x_j)| \leq \sum_{j=1}^{d} |l_j(\xi_n)\sqrt{k}(S_{nj}(x_j) - x_j) + l_j(x)W_j(x_j)|.$$

Note that all the terms on the right-hand side of the above inequality can be dealt with in the same way. Consider the first term. It is bounded by

$$\begin{aligned} & \sup_{x\in[0,T]^d} |l_1(\xi_n)| \cdot \sup_{x_1\in[0,T]} |\sqrt{k}(S_{n1}(x_1) - x_1) + W_1(x_1)| \\ & \quad + \sup_{x\in[0,T]^d} |l_1(\xi_n) - l_1(x)| \cdot \sup_{x_1\in[0,T]} |W_1(x_1)| \\ =: & \quad D_4 \cdot D_5 + D_6 \cdot D_7. \end{aligned}$$

Since l_1 is continuous on $[0, 2T]^d$, it is uniformly continuous and bounded. We have $D_5 \xrightarrow{\mathbb{P}} 0$ by (4.7.3), so $D_4 \cdot D_5$ also converges to zero in probability. The uniform continuity of l_1 and the fact that almost surely $D_7 < \infty$ yield $D_6 \cdot D_7 \xrightarrow{\mathbb{P}} 0$.

Applying (4.7.1) completes the proof. □

Proposition 4.7.3. *If the conditions (C1), (C2) hold, then as $n \to \infty$ and $k \to \infty$,*

$$\sqrt{k} \int_{[0,1]^d} g(x) \left(\hat{l}_n(x) - l(x)\right) dx \xrightarrow{d} \tilde{B}. \tag{4.7.4}$$

Proof Throughout the proof we write $l(x)$ instead of $l(x;\theta_0)$. Also, since l does not need to be differentiable, we will use notation $l_j(x)$, $j=1,\ldots,d$, to denote the right-hand partial derivatives here. Let $D_1(x)$, $D_2(x)$, $D_3(x)$ be as in the proof of Theorem 4.4.2 and take $T=1$. Then

$$\begin{aligned}
&\Big|\sqrt{k}\;\Big(\int_{[0,1]^d} g(x)\hat{L}_n(x)\mathrm{d}x - \int_{[0,1]^d} g(x)l(x)\mathrm{d}x\Big) - \tilde{B}\Big| \\
&\leq \sup_{x\in[0,1]^d} |D_1(x) - W_l(x)| \int_{[0,1]^d} |g(x)|\mathrm{d}x \\
&\quad + \sup_{x\in[0,1]^d} |D_2(x)| \int_{[0,1]^d} |g(x)|\mathrm{d}x \\
&\quad + \int_{[0,1]^d} |g(x,y)| \cdot \left| D_3(x) + \sum_{j=1}^d l_j(x)W_j(x_j)\right| \mathrm{d}x.
\end{aligned}$$

The first two terms on the right hand side converge to zero in probability due to integrability of g and uniform convergence of $D_1(x)$ and $D_2(x)$, which was shown in the proof of Theorem 4.4.2. The third term needs to be treated separately, as the condition on continuity (and existence) of partial derivatives is no longer assumed to hold.

Let ω be a point in the Skorohod probability space introduced before the proof of Theorem 4.4.2 such that for all $j=1,\ldots,d$,

$$\sup_{x_j\in[0,1]} |W_j(x_j)| < +\infty \text{ and } \sup_{x_j\in[0,1]} |\sqrt{k}(S_{nj}(x_j) - x_j) + W_j(x_j)| \to 0.$$

For such ω we will show by means of dominated convergence that

$$\int_{[0,1]^d} |g(x)| \cdot \left| \sqrt{k}\,(l(S_n(x)) - l(x)) + \sum_{j=1}^d l_j(x)W_j(x_j)\right| \mathrm{d}x \to 0. \tag{4.7.5}$$

Proof of the pointwise convergence. If l is differentiable, convergence of the above integrand to zero follows from the definition of partial derivatives and (4.7.3). Since this might fail only on a set of Lebesgue measure zero, the convergence of the integrand to zero holds almost everywhere on $[0,1]^d$.

Proof of the domination. Note that from expressions for (one-sided) partial derivatives (4.2.4), and the moment conditions (4.6.1) it follows that $|l_j(x)| \leq 1$,

for all $x \in [0,1]^d$ and all $j = 1, \dots, d$.
We get

$$\begin{aligned} |g(x)| \quad \cdot \quad & \left| \sqrt{k} \left(l\left(S_n(x)\right) - l(x) \right) + \sum_{j=1}^{d} l_j(x) W_j(x_j) \right| \\ & \leq |g(x)| \cdot \left(\sqrt{k} |l(S_n(x)) - l(x)| + \sum_{j=1}^{d} |W_j(x_j)| \right). \end{aligned}$$

Using the definition of function l and uniformity of $1 - F_j(X_{1j})$, we have for all $j = 1, \dots, d$

$$|l(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_d) - l(x_1, \dots, x_{j-1}, x'_j, x_{j+1}, \dots, x_d)| \leq |x_j - x'_j|.$$

Hence, we can write

$$\begin{aligned} \sup_{x \in [0,1]^d} \quad & \sqrt{k} |l(S_n(x)) - l(x)| \\ \leq \quad & \sup_{x \in [0,1]^d} \sqrt{k} |l(S_n(x)) - l(x_1, S_{n2}(x_2), \dots, S_{nd}(x_d))| \\ & + \sup_{x \in [0,1]^d} \sqrt{k} |l(x_1, S_{n2}(x_2), S_{n3}(x_3), \dots, S_{nd}(x_d)) \\ & \quad - l(x_1, x_2, S_{n3}(x_3), \dots, S_{nd}(x_d))| \\ & + \cdots \\ & + \sup_{x \in [0,1]^d} \sqrt{k} |l(x_1, \dots, x_{d-1}, S_{nd}(x_d)) - l(x)| \\ \leq \quad & \sum_{j=1}^{d} \sup_{x_j \in [0,1]} \sqrt{k} |S_{nj}(x_j) - x_j| = O(1). \end{aligned}$$

Since for all $j = 1, \dots, d$ we have $\sup_{x_j \in [0,1]} |W_j(x_j)| < +\infty$, the proof of (4.7.5) is complete. This together with (4.7.1) finishes the proof of the proposition. □

Lemma 4.7.4. *If in addition to assumptions of Theorem 4.4.1, conditions (C1), (C2), (C4) hold, then as* $n \to \infty$ *and* $k \to \infty$*,*

$$\sqrt{k} \nabla Q_{k,n}(\theta_0) \xrightarrow{d} N(0, V(\theta_0)).$$

Proof The gradient vector of $Q_{k,n}$ with respect to θ in θ_0 is

$$\nabla Q_{k,n}(\theta_0) = \left(\frac{\partial}{\partial \theta_1} Q_{k,n}(\theta) \Big|_{\theta=\theta_0}, \ldots, \frac{\partial}{\partial \theta_p} Q_{k,n}(\theta) \Big|_{\theta=\theta_0} \right)^T,$$

where for $i = 1, \ldots, p$,

$$\frac{\partial}{\partial \theta_i} Q_{k,n}(\theta) \Big|_{\theta=\theta_0} = -2 \sum_{m=1}^{q} \int_{[0,1]^d} g_m(x) \frac{\partial}{\partial \theta_i} l(x;\theta) \Big|_{\theta=\theta_0} \mathrm{d}x \\ \cdot \int_{[0,1]^d} g_m(x)(\hat{l}_n(x) - l(x;\theta_0))\mathrm{d}x.$$

Using vector notation we obtain

$$\nabla Q_{k,n}(\theta_0) = -2\dot{\varphi}(\theta_0)^T \cdot \int_{[0,1]^d} g(x)(\hat{l}_n(x) - l(x;\theta_0))\mathrm{d}x.$$

Equation (4.7.1) and the proof of Proposition 4.7.3 imply that

$$\sqrt{k}\nabla Q_{k,n}(\theta_0) = -2\dot{\varphi}(\theta_0)^T \cdot \int_{[0,1]^d} g(x)\sqrt{k}\left(\hat{l}_n(x) - l(x;\theta_0)\right)\mathrm{d}x \xrightarrow{d} -2\dot{\varphi}(\theta_0)^T \tilde{B}.$$

The limit distribution of $\sqrt{k}\nabla Q_{k,n}(\theta_0)$ is, hence, zero-mean Gaussian with covariance matrix $V(\theta_0) = 4\dot{\varphi}(\theta_0)^T \Sigma(\theta_0)\dot{\varphi}(\theta_0)$. □

Proof of Theorem 4.4.3 Consider the function $f(t) := \nabla Q_{k,n}(\theta_0 + t(\hat{\theta}_n - \theta_0))$, $t \in [0,1]$. The mean value theorem yields

$$\nabla Q_{k,n}(\hat{\theta}_n) = \nabla Q_{k,n}(\theta_0) + \mathcal{H}_{k,n}(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0),$$

for some $\tilde{\theta}_n$ between θ_0 and $\hat{\theta}_n$. First note that with probability tending to one, $0 = \nabla Q_{k,n}(\hat{\theta}_n)$, which follows from the fact that $\hat{\theta}_n$ is a minimizer of $Q_{k,n}$ and that with probability tending to one $\hat{\theta}_n$ is in an open ball around θ_0. By the consistency of $\hat{\theta}_n$ we have that $\tilde{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$, and since the convergence of $\mathcal{H}_{k,n}$ to $\mathcal{H}$ is uniform on a neighborhood of θ_0, we get that $\mathcal{H}_{k,n}(\tilde{\theta}_n) \xrightarrow{\mathbb{P}} \mathcal{H}(\theta_0)$. Hence,

$$\sqrt{k}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, M(\theta_0)).$$

□

Proof of Corollary 4.4.4 As in Lemma 7.2 in Einmahl, Krajina, and Segers (2008), we can see that if $\theta \mapsto H_\theta$ is weakly continuous at θ_0, then $\theta \mapsto \Sigma(\theta)$ is continuous at θ_0. This, together with condition (C4), yields that $\theta \mapsto V(\theta)$ is continuous at θ_0. Assumption (C4) also implies that $\theta \mapsto \mathcal{H}(\theta)$ is continuous at θ_0, which, with the positive definiteness of $\mathcal{H}(\theta)$ in a neighborhood of θ_0, shows that if $\theta \mapsto H_\theta$ is weakly continuous at θ_0, then $\theta \mapsto M(\theta) = \mathcal{H}(\theta)^{-1}V(\theta)\mathcal{H}(\theta)^{-1}$ is continuous at θ_0. Hence, we obtain

$$M(\hat{\theta}_n)^{-1/2}\sqrt{k}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I_p),$$

which yields (4.4.4). □

Proof of Theorem 4.4.5 Theorem 4.4.3 and the arguments used in the proof of Corollary 4.4.4 imply that as $n \to \infty$,

$$M_2^{-1/2}(\hat{\theta}_1, \theta_2^*)\sqrt{k}(\hat{\theta}_2 - \theta_2^*) \xrightarrow{d} N(0, I_r), \tag{4.7.6}$$

and hence (4.4.7). □

Proof of Lemma 4.6.1

$$\begin{aligned}
l(x_1, \ldots, x_d) &= \lim_{t\to\infty} t\mathbb{P}\left(1 - F_1(X_1) \le x_1/t \text{ or } \ldots \text{ or } 1 - F_d(X_d) \le x_d/t\right) \\
&= \lim_{t\to\infty} t\mathbb{P}\left(1 - F_{Y_1}(Y_1) \le x_1/t \text{ or } \ldots \text{ or } 1 - F_{Y_d}(Y_d) \le x_d/t\right) \\
&= \lim_{t\to\infty} t\mathbb{P}\left(Y_1 \ge \frac{t\sum_{i=1}^r a_{i1}^\nu}{x_1} \text{ or } \ldots \text{ or } Y_d \ge \frac{t\sum_{i=1}^r a_{id}^\nu}{x_d}\right) \\
&= \lim_{t\to\infty} t\mathbb{P}\left(\bigcup_{1\le j\le d}\bigcup_{1\le i\le r}\left\{W_i \ge \frac{t\sum_{i=1}^r a_{ij}^\nu}{a_{ij}^\nu x_j}\right\}\right) \\
&= \lim_{t\to\infty} t\mathbb{P}\left(\bigcup_{1\le i\le r}\left\{W_i \ge \min_{1\le j\le d}\frac{t\sum_{i=1}^r a_{ij}^\nu}{a_{ij}^\nu x_j}\right\}\right) \\
&= \lim_{t\to\infty} t\sum_{i=1}^r \mathbb{P}\left(W_i \ge \min_{1\le j\le d}\frac{t\sum_{i=1}^r a_{ij}^\nu}{a_{ij}^\nu x_j}\right) \\
&= \lim_{t\to\infty} \sum_{i=1}^r t\left(1 - \exp\left\{-\frac{1}{t}\max_{1\le j\le d}\frac{a_{ij}^\nu x_j}{\sum_{i=1}^r a_{ij}^\nu}\right\}\right)
\end{aligned}$$

$$= \sum_{i=1}^{r} \max_{1 \leq j \leq d} \left\{ \frac{a_{ij}^{\nu} x_j}{\sum_{i=1}^{r} a_{ij}^{\nu}} \right\} =: \sum_{i=1}^{r} \max_{1 \leq j \leq d} \{b_{ij} x_j\}$$

as required. □

Proof of Lemma 4.6.2 Fix $i \in \{1, \ldots, r\}$. We have

$$\int_{[0,1]^d} x_k^m \max_{1 \leq j \leq d} \{b_{ij} x_j\} \mathrm{d}x = \sum_{j=1}^{d} \int_{[0,1]^d} x_k^m (b_{ij} x_j) \mathbf{1}\left(b_{ij} x_j \geq \max_{l \neq j} \{b_{il} x_l\} \right) \mathrm{d}x.$$

Write the integral as a double integral, the outer integral with respect to $x_j \in [0,1]$ and the inner integral with respect to $x_{-j} = (x_l)_{l \neq j} \in \mathbb{R}^{d-1}$ over the relevant domain. We find

$$\int_{[0,1]^d} x_k^m \max_{1 \leq j \leq d} \{b_{ij} x_j\} \, \mathrm{d}x = \sum_{j=1}^{d} \int_0^1 b_{ij} x_j \int_{0 < x_l < \frac{b_{ij}}{b_{il}} x_j \wedge 1} x_k^m \, \mathrm{d}x_{-j} \, \mathrm{d}x_j.$$

After some long, but elementary computations, this simplifies to the stated expression. □

Bibliography

Abdous, B., A.-L. Fougères, and K. Ghoudi (2005). Extreme behavior for bivariate elliptical distributions. *The Canadian Journal of Statistics 33*, 317–334.

Ané, T. and C. Kharoubi (2003). Dependence structure and risk measure. *Journal of Business 76*, 411–438.

Asimit, A. V. and B. L. Jones (2007). Extreme behavior of bivariate elliptical distributions. *Insurance: Mathematics & Economics 41*, 53–61.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. Chichester: Wiley.

Berman, S. M. (1992). *Sojourns and extremes of stochastic processes*. Belmont: Wadsworth & Brooks/Cole Advanced Books & Software.

Billingsley, P. (1968). *Convergence of Probability Measures*. New York: John Wiley & Sons.

Bingham, N. C., C. M. Goldie, and J. L. Teugels (1987). *Regular Variation*. Cambridge: Cambridge University Press.

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.

Coles, S. G. and J. A. Tawn (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society, Series B 53*, 377–392.

Danielsson, J., L. de Haan, L. Peng, and C. G. de Vries (2002). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis 76*, 226–248.

de Haan, L. (1990). Fighting the archenemy with mathematics. *Statistica Neerlandica 44*, 45–68.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction.* New York: Springer.

de Haan, L., C. Neves, and L. Peng (2008). Parametric tail copula estimation and model testing. *Journal of Multivariate Analysis 99*, 1260–1275.

de Haan, L. and S. Resnick (1977). Limit theory for multidimensional sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 40*, 317–337.

Dekkers, A. L. M. and L. de Haan (1993). Optimal choice of sample fraction in extreme-value estimation. *Journal of Multivariate Analysis 47*, 173–195.

Dekkers, A. L. M., J. H. J. Einmahl, and L. de Haan (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics 17*, 1833–1855.

Demarta, S. and A. J. McNeil (2005). The t copula and related copulas. *International Statistical Review 73*, 111–129.

Drees, H. and X. Huang (1998). Best attainable rates of convergence for estimates of the stable tail dependence function. *Journal of Multivariate Analysis 64*, 25–47.

Drees, H. and E. Kaufmann (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications 75*, 149–172.

Einmahl, J. H. J. (1997). Poisson and Gaussian approximation of weighted local empirical processes. *Stochastic Processes and their Applications 70*, 31–58.

Einmahl, J. H. J., L. de Haan, and D. Li (2006). Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *Annals of Statistics 34*, 1987–2014.

Einmahl, J. H. J., A. Krajina, and J. Segers (2008). A method of moments estimator of tail dependence. *Bernoulli 14*, 1003–1026.

Einmahl, J. H. J., A. Krajina, and J. Segers (2009). An M-estimator of tail dependence in arbitrary dimensions. Work in progress.

Einmahl, J. H. J. and J. R. Magnus (2008). Records in athletics through extreme-value theory. *Journal of the American Statistical Association 103*, 1382–1391.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.

Embrechts, P., A. McNeil, and D. Straumann (2002). Correlation and dependence in risk management: Properties and pitfalls. In *Risk Management: Value at Risk and Beyond*, pp. 176–223. Cambridge University Press.

Falk, M., J. Hüsler, and R.-D. Reiss (2004). *Laws of Small Numbers: Extremes and Rare Events*. Basel: Birkhäuser-Verlag.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*, 3–56.

Fang, H. B., K. T. Fang, and S. Kotz (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis 82*, 1–16.

Fang, K.-T., S. Kotz, and K.-W. Ng (1990). *Symmetric Multivariate and Related Distributions*. London: Chapman and Hall.

Fisher, R. A. and L. H. C. Tippet (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society 24*, 180–190.

Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*. Malaber: Krieger.

Geluk, J. L., L. de Haan, and C. G. de Vries (2007). Weak & strong financial fragility. TI 2007-023/2.

Genest, C., A. C. Favre, J. Béliveau, and C. Jacques (2007). Meta-elliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research 43*. W09401 doi:10.1029/2006WR005275.

Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics 44*, 423–453.

Gomes, M. I. and O. Oliveira (2002). The bootstrap methodology in statistics of extremes - choice of the optimal sample fraction. *Extremes 4*, 331–358.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Society 55*, 698–707.

Hashorva, E. (2005). Extremes of asymptotically spherical and elliptical random vectors. *Insurance: Mathematics & Economics 36*, 285–302.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics 3*, 1163–1174.

Huang, X. (1992). *Statistics of Bivariate Extreme Values*. Ph. D. thesis, Tinbergen Institute Research Series 22, Erasmus University Rotterdam.

Hult, H. and F. Lindskog (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability 34*, 587–608.

Joe, H., R. L. Smith, and I. Weissman (1992). Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society, Series B 54*, 171–183.

Katz, R., M. Parlange, and P. Naveau (2002). Extremes in hydrology. *Advances in Water Resources 25*, 1287–1304.

Kaynar, B., S. I. Birbil, and J. B. G. Frenk (2007). Application of a general risk management model to portfolio optimization problems with elliptical distributed returns for risk neutral and risk averse decision makers. Preprint available on http://hdl.handle.net/1765/10151.

Kendall, M. (1938). A new measure of rank correlation. *Biometrika 30*, 81–89.

Kendall, M. (1948). *Rank Correlation Methods*. London: Charles Griffin & Company Limited.

Klüppelberg, C., G. Kuhn, and L. Peng (2007). Estimating tail dependence of elliptical distributions. *Bernoulli 13*, 229–251.

Klüppelberg, C., G. Kuhn, and L. Peng (2008). Semi-parametric models for the multivariate tail dependence function - the asymptotically dependent case. *Scandinavian Journal of Statistics 35*, 701–718.

Kotz, S. and S. Nadarajah (2000). *Extreme Value Distributions*. London: Imperial College Press.

Krajina, A. (2009). A method of moments estimator of tail dependence in elliptical copula models. *CentER Discussion Paper Series No*. 2009-42.

Landsman, Z. M. and E. A. Valdez (2003). Tail conditional expectations for elliptical distributions. *North American Actuarial Journal 7*, 55–71.

Ledford, A. W. and J. A. Tawn (1996). Statistics for near independence in multivariate extreme values. *Biometrika 83*, 169–187.

Lee, A. (1990). *U-Statistics*. New York: Marcel Dekker.

Lindskog, F., A. McNeil, and U. Schmock (2003). Kendall's tau for elliptical distributions. In G. Bol, G. Nakhaeizadeh, S. Rachev, T. Ridder, and K.-H. Vollmer (Eds.), *Credit Risk: Measurement, Evaluation and Management*, pp. 149–156. Heidelberg: Physica-Verlag.

Malevergne, Y. and D. Sornette (2004). Tail dependence of factor models. *The Journal of Risk 6*, 71–116.

Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Annals of Mathematical Statistics 42*, 1285–1295.

Perrin, O., H. Rootzén, and R. Taesler (2006). A discussion of statistical methods for estimation of extreme wind speeds. *Theoretical and Applied Climatology 85*, 203–215.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics 3*, 119–131.

Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes.* New York: Springer.

Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer-Verlag.

Rootzén, H. and N. Tajvidi (1995). Extreme value statistics and wind storm losses: A case study. *Scandinavian Actuarial Journal 1*, 70–94.

Shorack, G. R. and J. A. Wellner (1986). *Empirical Processes with Applications to Statistics*. New York: Wiley.

Smith, R. L. (1987). Estimating tails of probability distributions. *Annals of Statistics 15*, 1174–1207.

Smith, R. L. (1994). Multivariate threshold methods. In J. Galambos, J. Lechner, and E. Simiu (Eds.), *Extreme Value Theory and Applications*, pp. 225–248. Kluwer.

Tiago de Oliveira, J. (1980). Bivariate extremes: foundations and statistics. In Krishnaiah (Ed.), *Multivariate Analysis V*, Amsterdam, pp. 349–366. Springer-Verlag.

Tiago de Oliveira, J. (1989). Statistical decision for bivariate extremes. In J. Hüsler and R. D. Reiss (Eds.), *Extreme Value Theory: Proceedings, Oberwolfach 1987*, Berlin, pp. 246–261. Lecture Notes in Statistics 51.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge Unversity Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.

Vervaat, W. (1972). Functional central limit theorems for processes with positive drift and their inverses. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 23*, 249–253.

Volkovich, Y. (2009). *Stochastic analysis of web page ranking*. Ph. D. thesis, University of Twente, Enschede.

von Mises, R. (1936). La distribution de la plus grande de n valeurs. *Revue Mathematique de l'Union Interbalcanique 1*, 141–160.

Vrac, M. and P. Naveau (2007). Stochastic downscaling of precipitation: From dry events to heavy rainfalls. *Water Resources Research 43*. W07402, doi:10.1029/2006WR005308.

Nederlandse Samenvatting

Een M-schatter voor multivariate staartafhankelijkheid

Dit proefschrift richt zich op het schatten van de staartafhankelijkheidsstructuur in een semiparametrisch model, verkregen door de aanname dat de stabiele staartafhankelijkheidsfunctie tot een parametrische familie behoort. Onderzoek naar de staartafhankelijkheid behoort tot het deelgebied van de kansrekening en statistiek dat *extreme-waardentheorie* wordt genoemd. Extreme-waardentheorie levert de theoretische fundering voor het modelleren van gebeurtenissen die bijna nooit plaatsvinden. Men is geïnteresseerd in deze gebeurtenissen vanwege hun mogelijk grote gevolgen, zoals in het geval van een grote overstroming of het instorten van de aandelenbeurs. Voorbeelden van extreme gebeurtenissen komt men tegen in gebieden als metereologie (overstromingen, stormen, zware regenval, grote bosbranden), financiering, schadeverzekering en herverzekering, internet page ranking, etc. Dit proefschrift bevat vier hoofdstukken. Een korte beschrijving van de hoofdstukken volgt.

In de inleiding worden enkele kernideeën uit de extreme-waardentheorie gepresenteerd, met nadruk op concepten die we in dit proefschrift gebruiken. Laat $X_1, \ldots, X_n$ een aselecte steekproef zijn uit een continue, d-dimensionale verdelingsfunctie F, $X_i = (X_{i1}, \ldots, X_{id})$, $i = 1, \ldots, n$; $F_1, \ldots, F_d$ zijn de marginale verdelingsfuncties van F. We nemen aan dat de limiet

$$\lim_{t \downarrow 0} t^{-1} \mathbb{P}\left(1 - F_1(X_{11}) \leq tx_1 \text{ or } \ldots \text{ or } 1 - F_d(X_{1d}) \leq tx_d\right) =: l(x), \qquad (4.7.7)$$

bestaat voor iedere $x = (x_1, \ldots, d) \in [0, \infty)^d$. De functie l heet de *stabiele staartafhankelijkheidsfunctie*, en is één van de begrippen die worden gebruikt om de staartafhankelijkheidsstructuur te beschrijven.

Een nieuwe schatter voor de tweedimensionale stabiele staartafhankelijkheidsfunctie wordt geïntroduceerd in Hoofdstuk 2. Aannemende dat de stabiele staartafhankelijkheidsfunctie tot een parametrische familie met onbekende parameter θ uit parameterruimte $\Theta \subseteq \mathbb{R}^p$ behoort, definiëren we een schatter $\hat{\theta}_n$ van θ als de oplossing van

$$\int_{[0,1]^2} g(x)\hat{l}_n \mathrm{d}x = \int_{[0,1]^2} g(x) l(x; \hat{\theta}_n)(x) \mathrm{d}x,$$

waarbij $\hat{l}_n$ de niet-parametrische schatter van l is en $g\colon [0,1]^2 \to \mathbb{R}^p$ een door ons gekozen hulpfunctie. Merk op dat we niet veronderstellen dat F_1, F_2 of de copula van F parametrisch zijn: het model is semiparametrisch. We bewijzen dat de schatter consistent en asymptotisch normaal is onder algemene voorwaarden. We eisen niet dat de partiële afgeleiden van l (naar x) continu zijn, of zelfs bestaan. Dit is een standaard vereiste voor asymptotische normaliteit van alle andere schatters voor l, zowel de niet-parametrische als de meest aannemelijke schatter. De afwezigheid van de aannames met betrekking tot de partiële afgeleiden laat het schatten van de staartafhankelijkheidsstructuur in een grotere klasse van modellen toe. Wij schatten bijvoorbeeld de discrete twee-punts spectraalmaat die bij een bivariaat twee-factormodel hoort.

Elliptische verdelingen vormen een familie van modellen die veel gebruikt worden op het gebied van financiering en verzekering. Bivariate elliptische verdelingen leveren een expliciete uitdrukking van de functie l; ieder model met dezelfde copula leidt tot dezelfde functie l. Die verdelingen vormen de klasse van *elliptische copulamodellen*, ook bekend als de meta-elliptische verdelingen. De functie l van een elliptisch copulamodel hangt af van twee parameters, ρ en ν, die verschillende betekenissen en eigenschappen hebben en daarom op verschillende manieren worden behandeld in de schattingsprocedure. Aangezien de correlatiecoëfficiënt ρ afhangt van de gehele copula, schatten we hem met behulp van de volledige steekproef. Vervolgens substitueren we die schatter in de uitdrukking voor l en schatten we de staartparameter ν met behulp van de momentenmethode gepresenteerd in Hoofdstuk 2. De toepassing is van belang aangezien de elliptische modellen vaak gebruikt worden. Echter, zoewel het implementeren van de schatter als het afleiden van zijn asymptotische eigenschappen is niet triviaal. De schattingsprocedure, de asymptotische resultaten voor de schatter van het paar (ν, ρ) en een simulatiestudie worden gepresenteerd in Hoofdstuk 3.

In Hoofdstuk 4 nemen we weer aan dat l parametrisch is, zodat we een semiparametrisch model hebben. We breiden de schatter van Hoofdstuk 2 in twee richtingen uit. Ten eerste maken we de procedure mogelijk voor een willekeurige dimensie d, $d \geq 2$. Ten tweede gebruiken we M-schatters in plaats van de momentenschatters. Alle bestaande schatters zijn niet-parametrisch, of meest aannemelijke schatters. Het kan erg moeilijk zijn om deze meest aannemelijke schatters uit te rekenen vanwege de ingewikkelde vorm van de aannemelijkheidsfunctie; bovenal zijn de modelveronderstellingen erg restrictief, aangezien ze hogere orde differentieerbaarheid van l omvatten. Wij kiezen een semiparametrische benadering omdat dit ons in staat stelt structuur op te leggen aan de afhankelijkheid en eventueel de schattingsfout te verminderen. Zonder differentieerbaarheid van l op te leggen, definiëren we een M-schatter van θ. We bewijzen dat de schatter consistent is en een normale limietverdeling heeft en we presenteren een toets voor een submodel binnen een gekozen semiparametrisch model. De simulatiestudies voor enkele voorbeelden, waaronder twee verschillende factormodellen, laten zien dat de schatter goed presteert in dimensies hoger dan twee.