Estimation of tail dependence with application to twin data

Master thesis by
Michael Osmann

May 21, 2012

Supervisors: Yuri Goegebeur and Jacob Hjelmborg

Contents

Abstract

This master thesis consists of a theoretical discussion on univariate and bivariate extreme value statistics along with an application to twin data. We first discuss the fundamental convergence results from extreme value theory, which we use to construct the traditional maximum likelihood estimators of the extreme value index. In order to put our work into the proper framework, attention is paid to the three classes of extreme value distributions situated within the max domain of attraction of the generalized extreme value distribution. Special attention is given to the class of Pareto type distributions, since the methodology of how to construct estimators in the multivariate setting resembles the methodology used to construct estimators within the class of Pareto-type distributions. For the class of Pareto-type distributions we propose an estimator of the extreme value index and an estimator for the second order parameter. For both of these estimators we establish the asymptotic normality.
In the multivariate setting we start by discussing the transformation of the margins to standard Fréchet distributions and the fundamental convergence results. We discuss the domain of attraction to the bivariate extreme value distribution and asymptotic dependence and asymptotic independence. We discuss furthermore the exponent measure, the spectral measure, Pickands dependence function, the dependence measures χ and $\bar{\chi}$, and finally the coefficient of tail dependence. The interpretations of these measures are discussed and we show how they are all connected. For the coefficient of tail dependence we introduce a functional estimator, for which we show how it can be bias corrected. This bias correction requires estimation of the second order parameter τ, so we propose two estimators that can be used to estimate this second order parameter. The consistency of the estimators for the second order parameter are established. We examine the finite sample size behaviour of our estimator for the coefficient of tail dependence, the estimators of the second order condition and estimators of χ and $\bar{\chi}$ using simulations.
The twin data we consider is from the older cohort of the Finnish Twin Cohort Study. For this data we make a full univariate data analysis and estimate the coefficient of tail dependence, the second order parameter τ, and the measures χ and $\bar{\chi}$ for age and sex defined subsets of the data.
Throughout the thesis, results that are from the literature are stated with a reference, while results that are our own are not stated with a reference.

Acknowledgements

I would like to thank my two supervisors Yuri Goegebeur and Jacob v. B. Hjelmborg for helping me write this master thesis during the last 8 months. I would not have been able to write this thesis without their help, and they have both spend a lot of time and effort on this. I am gratefull that they decided to join forces and help me write a thesis with such an interesting topic.

Chapter 1

Preliminaries

This chapter serves to give a short introduction to some of the basic concepts in univariate extreme value statistics. First we will introduce a convergence result which is the foundation of univariate extreme value statistics. It states what form the limiting distribution of a normalized maximum will follow, if it exists. We will then describe shortly two of the classes of extreme value distributions, known as the Gumbel and extremal Weibull families, respectively. Finally, we discuss some simple ways in which the extreme value index can be estimated in practice.

1.1 Classical convergence result

In the following we will consider a sample $\{X_i, 1 \leq i \leq n\}$ of independent and identically distributed (i.i.d.) random variables having a distribution function F_X. In extreme value statistics we consider either the maximum or the minimum of the random sample, where the maximum is given by

$$X_{n,n} := \max\{X_1, X_2, \ldots, X_n\}.$$

We will try to describe the statistical behaviour of this maximum, but it is easy to transform any result we obtain for the maximum to the minimum because of the relation

$$X_{1,n} := \min\{X_1, X_2, \ldots, X_n\} = -\max\{-X_1, -X_2, \ldots, -X_n\}. \tag{1.1}$$

Because of the i.i.d. nature of $X_1, \ldots, X_n$, the distribution of $X_{n,n}$ can be derived exactly for all possible values of n as follows

$$\begin{aligned} F_{X_{n,n}}(x) &= P(X_{n,n} \leq x) \\ &= P(X_1 \leq x, X_2 \leq x, \ldots X_n \leq x) \\ &= P(X_1 \leq x)P(X_2 \leq x) \cdots P(X_n \leq x) \\ &= (F_X(x))^n . \end{aligned}$$

For practical purposes this relation does not help much though, since the distribution of F_X is usually unknown. One could try to estimate the distribution of F_X and use this to estimate $F_{X_{n,n}}$, but small deviations in the estimation of F_X can lead to large deviations in the estimation of $F_{X_{n,n}}$. Instead we will look for approximate families of $F_{X_{n,n}}$ which for large

n can be estimated by use of the extreme data only.
We look at the behaviour of $F_{X_{n,n}}$ as n approaches infinity. If we denote the right endpoint of F_X as x_*, which means that $x_* := \inf\{x : F_X(x) = 1\}$, then for any $x < x_*$ we have that $F_X^n(x) \to 0$ as $n \to \infty$. So the distribution of $X_{n,n}$ is degenerate in the limit. This degeneracy can possibly be avoided if we look at an appropriate normalization, for instance

$$\frac{X_{n,n} - b_n}{a_n}$$

where $(b_n)_{n=1}^{\infty}$ is a sequence of constants and $(a_n)_{n=1}^{\infty}$ is a sequence of positive constants. Appropriate choices of $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ can stabilize the location and scale of $\frac{X_{n,n}-b_n}{a_n}$. It can be shown that the entire range of limit distributions of $\frac{X_{n,n}-b_n}{a_n}$, if they exist, is given by Theorem 1.1.1.

Theorem 1.1.1. (Fisher and Tippet, 1928; Gnedenko, 1943) *Let $X_1, \ldots, X_n$ be i.i.d. random variables with distribution function F_X. If there exists sequences of constants $(b_n)_{n=1}^{\infty}$ and positive constants $(a_n)_{n=1}^{\infty}$ such that*

$$\lim_{n\to\infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F_X^n\left(a_n x + b_n\right) = G(x) \tag{1.2}$$

at all continuity points of G, where G is a non degenerate distribution function, then G should be of the following type

$$G_\gamma(x) = \exp\left(-(1+\gamma x)^{-\frac{1}{\gamma}}\right), \quad 1 + \gamma x > 0, \tag{1.3}$$

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

This family of distribution functions is known as the generalized extreme value (GEV) family, for which the parameter γ is the shape parameter. This parameter is also called the extreme value index and it describes the tail behaviour of F_X, with larger values indicating heavier tails. The family consists of three classes known as the Gumbel, Fréchet and extremal Weibull families which correspond to $\gamma = 0$, $\gamma > 0$ and $\gamma < 0$ respectively. The Fréchet class is also known as the class of Pareto-type models. If the distribution F_X satisfies (1.2)-(1.3) then we say that it belongs to the max domain of attraction of G_γ, denoted $F_X \in \mathcal{D}(G_\gamma)$.
The result in Theorem 1.1.1 has some equivalent formulations. Some of these formulations are based on the tail quantile function $U(y) := Q\left(1 - \frac{1}{y}\right), \quad y > 1$, where Q is the quantile function, defined as $Q(p) := \inf\{x : F_X(x) \geq p\}, \quad p \in (0,1)$. These equivalent formulations are stated in Theorem 1.1.2.

Theorem 1.1.2. (Gnedenko, 1943; de Haan and Ferreira, 2006) *Let $X_1, \ldots, X_n$ be i.i.d. random variables with distribution function F_X. For $\gamma \in \mathbb{R}$ the following statements are equivalent:*

(i) *There exists sequences of real constants $(b_n)_{n=1}^{\infty}$ and positive real constants $(a_n)_{n=1}^{\infty}$ such that*

$$\lim_{n\to\infty} F_X^n\left(a_n x + b_n\right) = G_\gamma(x) = \exp\left(-(1+\gamma x)^{-\frac{1}{\gamma}}\right), \tag{1.4}$$

for all x with $1 + \gamma x > 0$.

(ii) *There is a positive function a such that for all $x > 0$,*

$$\lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \tag{1.5}$$

where for $\gamma = 0$ the right-hand side is interpreted as $\log x$.

(iii) *There is a positive function a such that*

$$\lim_{t \to \infty} t(1 - F_X(a(t)x + U(t))) = (1 + \gamma x)^{-\frac{1}{\gamma}}, \tag{1.6}$$

for all x with $1 + \gamma x > 0$.

(iv) *There exists a positive function f such that*

$$\lim_{t \uparrow x_*} \frac{1 - F_X(t + xf(t))}{1 - F_X(t)} = (1 + \gamma x)^{-\frac{1}{\gamma}} \tag{1.7}$$

for all x for which $1 + \gamma x > 0$.

Moreover, (1.4) *holds with $b_n := U(n)$ and $a_n := a(n)$. Also* (1.7) *holds with $f(t) = a\left(\frac{1}{1-F_X(t)}\right)$.*

As seen in Theorem 1.1.2 the choice of the normalizing constant b_n does not depend on the sign of γ and can be shown to always work, if we choose $b_n = U(n)$. The choice of a_n depends on whether we are dealing with γ positive, negative or equal to zero, so we will address this in the sections dedicated to the corresponding classes.
In order to discuss the extremal Weibull and Fréchet classes, we need the concept of a slowly varying function. Slowly varying functions are special cases of regularly varying functions, so we will give the definition of what it means to be of regular variation. The regularly varying functions will also be needed later in this thesis.

Definition 1.1.3. (Beirlant et al., 2004, Definition 2.1) Let f be an ultimately positive and measurable function on $\mathbb{R}_+$. We say that f is regularly varying at infinity if there exists a real constant ρ for which

$$\lim_{x \to \infty} \frac{f(\lambda x)}{f(x)} = \lambda^\rho \quad \text{for all } \lambda > 0.$$

We write $f \in \mathcal{R}_\rho$ and we call ρ the the index of regular variation. In the case $\rho = 0$, the function will be called slowly varying or of slow variation. We will reserve the symbol l for such functions. The class of all regularly varying functions is denoted by $\mathcal{R}$.

The next two sections will be dedicated to the Gumbel and the extremal Weibull class, while the Fréchet class which is of more importance for this thesis, will be discussed in the next chapter.

1.2 The Gumbel class

The Gumbel class corresponds with the max domain of attraction of G_γ with $\gamma = 0$. The following proposition provides a characterization of the distributions that belong to this class.

Proposition 1.2.1. *(Gnedenko, 1943) Let X be a random variable with distribution function F_X. Then we have for x_* finite or infinite and a, f suitable positive functions, that*

$$F_X \in \mathcal{D}(G_0) \quad \Leftrightarrow \quad \lim_{t \uparrow x_*} \frac{1 - F_X(t + xf(t))}{1 - F_X(t)} = \exp(-x), \quad x \in \mathbb{R} \tag{1.8}$$

$$\Leftrightarrow \quad \lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} = \log(x), \quad x > 0. \tag{1.9}$$

For the Fréchet and extremal Weibull classes it is easy to show that the distributions belonging to those classes satisfy (1.5), but this is not the case for the Gumbel class. This also means that determining the scaling parameter a_n for the distributions in the Gumbel class is more difficult. It can however be determined by the formula

$$a_n = n \int_{U(n)}^{x_*} (1 - F_X(y))\, dy.$$

We will not derive this formula, but simply take it as a fact. For details we refer to de Haan and Ferreira (2006), Corollary 1.2.4.

Example 1.2.2. If we want to determine the parameters a_n and b_n for the exp(1) distribution with distribution function $F_X(x) = 1 - \exp(-x),\ x > 0$, then we must first find the tail quantile distribution of the exponential distribution. The distribution function has quantile function $Q(p) = -\ln(1-p),\ 0 < p < 1$. So

$$U(x) = Q\left(1 - \frac{1}{x}\right) = \log(x), \quad x > 1.$$

This means b_n can be chosen as

$$b_n = U(n) = \log(n)$$

and a_n can be chosen as

$$a_n = n \int_{\log(n)}^{\infty} \exp(-x) dx = n \exp\left(-\log(n)\right) = 1.$$

Since we know the constants a_n and b_n we can also show that the exponential distribution belongs to the max domain of attraction of the Gumbel class. Indeed

$$\begin{aligned} P\left(\frac{X_{n,n} - b_n}{a_n} \le x\right) &= F_X^n(a_n x + b_n) \\ &= F_X^n\left(x + \log(n)\right) \\ &= \left(1 - \exp\left(-x - \log(n)\right)\right)^n \\ &= \left(1 - \frac{\exp(-x)}{n}\right)^n \\ &\to \exp\left(-\exp\left(-x\right)\right) \quad \text{for } n \to \infty. \end{aligned}$$

The convergence of $F_X^n(a_n x + b_n)$ to $G(x)$ is shown in Figure 1.1. The solid line is $G(x)$, the dashed line is for $n = 2$, the dotted line is for $n = 5$ and the dashed dotted line is for $n = 10$. It is clearly seen that when n grows then $F_X^n(a_n x + b_n)$ converges pointwise to $G(x)$.

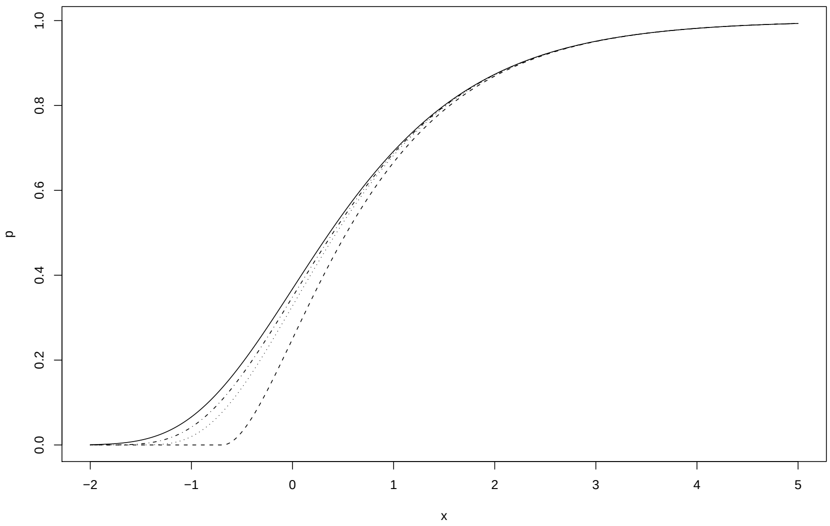

Figure 1.1: The convergence of $F_X^n(a_n x + b_n)$ to $G(x)$ for the standard exponential distribution.

■

1.3 The extremal Weibull class

The extremal Weibull class corresponds with the max domain of attraction of G_γ with $\gamma < 0$. As was the case for the Gumbel class, we have a proposition which provides a characterization of the distributions that belong to this class.

Proposition 1.3.1. *(Gnedenko, 1943) Let X be a random variable with distribution function F_X. Then we have for x_* finite that*

$$F_X \in \mathcal{D}(G_\gamma), \quad \gamma < 0 \quad \Leftrightarrow \quad 1 - F_X\left(x_* - \frac{1}{x}\right) = x^{\frac{1}{\gamma}} l_{F_X}(x), \quad x > 0 \tag{1.10}$$

$$\Leftrightarrow \quad U(x) = x_* - x^{\gamma} l_U(x), \quad x > 1, \tag{1.11}$$

where $l_U(x)$ and $l_{F_X}(x)$ are slowly varying at infinity.

From (1.11) it is easily seen that (1.5) is satisfied when t tends to infinity. Indeed

$$\begin{aligned}\frac{U(tx)-U(t)}{a(t)} &= \frac{x_*-(tx)^\gamma l_U(tx)-(x_*-t^\gamma l_U(t))}{a(t)} \\ &= \frac{t^\gamma l_U(t)}{a(t)}\left(1-x^\gamma \frac{l_U(tx)}{l_U(t)}\right) \\ &\sim -\gamma \frac{t^\gamma l_U(t)}{a(t)} \frac{x^\gamma-1}{\gamma} \\ &\sim \frac{x^\gamma-1}{\gamma}\end{aligned}$$

if we choose $a(t)$ such that $\frac{a(t)}{x_*-U(t)} \to -\gamma$. This indicates that a good choice of a_n would be

$$a_n = a(n) = -\gamma(x_* - U(n)) = -\gamma n^\gamma l_U(n).$$

Example 1.3.2. The reversed Burr distribution has distribution function given by

$$F_X(x) = 1-\left(\frac{\zeta}{\zeta+(1-x)^{-\delta}}\right)^\lambda, \quad x<1; \lambda,\zeta,\delta>0$$

and so the quantile function is

$$Q(p) = 1-\zeta^{-\frac{1}{\delta}}\left((1-p)^{-\frac{1}{\lambda}}-1\right)^{-\frac{1}{\delta}}, \quad 0<p<1.$$

So we find the tail quantile function U to be

$$U(x) = Q\left(1-\frac{1}{x}\right) = 1-\zeta^{-\frac{1}{\delta}}\left(x^{\frac{1}{\lambda}}-1\right)^{-\frac{1}{\delta}}, \quad x>1.$$

The distribution belongs to the max domain of attraction of G_γ with $\gamma = -\frac{1}{\lambda\delta}$. If we consider the reversed Burr distribution with parameters $\lambda = \zeta = \delta = 1$, then we can choose the normalizing constant b_n as

$$b_n = U(n) = 1-(n-1)^{-1}.$$

Since $x_* = 1$ and $\gamma = -1$ we can choose the normalizing constant a_n as

$$a_n = 1-U(n) = (n-1)^{-1}.$$

With these normalizing constants we can show that the reversed Burr distribution with parameters $\lambda = \zeta = \delta = 1$ belongs to the max domain of attraction of the Weibull class. Indeed

$$\begin{aligned}P\left(\frac{X_{n,n}-b_n}{a_n} \le x\right) &= F_X^n(a_n x + b_n) \\ &= F_X^n\left((x-1)(n-1)^{-1}+1\right) \\ &= \left(1-\frac{1}{1+\left(\frac{n-1}{1-x}\right)}\right)^n \\ &= \left(1-\frac{1-x}{n-x}\right)^n \\ &\to \exp(-(1-x)) \quad \text{for } n\to\infty.\end{aligned}$$

The convergence of the reversed Burr distribution to its limit is illustrated in Figure 1.2. The solid line is $G(x)$, the dashed line is for $n = 2$, the dotted line is for $n = 5$, while the dashed dotted line is for $n = 10$. It is clearly seen that when n grows, then $F_X^n(a_n x + b_n)$ converges pointwise to $G(x)$.

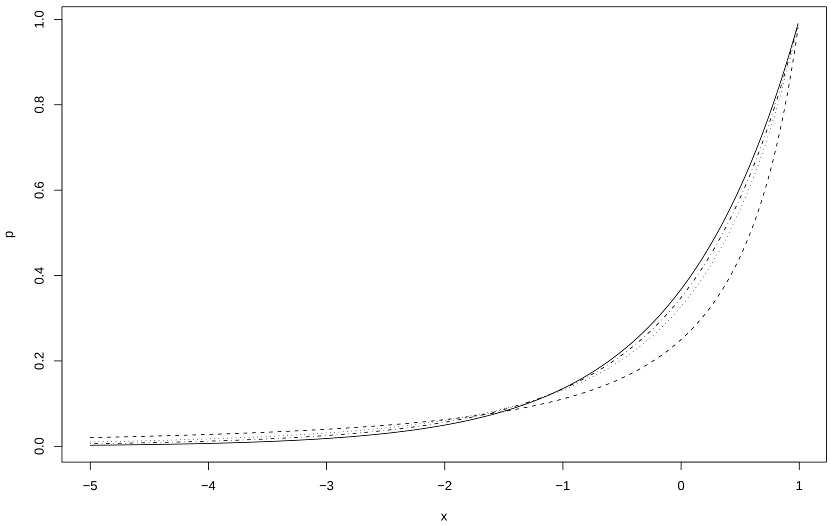

Figure 1.2: The convergence of $F_X^n(a_n x + b_n)$ to $G(x)$ for the reversed Burr distribution with $\lambda = \zeta = \delta = 1$.

■

1.4 Estimation of the extreme value index in practice

In practise we do not know the constants a_n and b_n, so Theorem 1.1.1 is not very usefull if we want to estimate γ. However, if we for some finite $n \in \mathbb{N}$ have that

$$P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) \approx \exp\left(-(1+\gamma x)^{-\frac{1}{\gamma}}\right), \quad 1 + \gamma x > 0,$$

then

$$P\left(X_{n,n} \leq z\right) \approx \exp\left(-\left(1 + \gamma \frac{z - b_n}{a_n}\right)^{-\frac{1}{\gamma}}\right), \quad 1 + \gamma \frac{z - b_n}{a_n} > 0,$$

where $z = b_n + a_n x$. If we let $\mu = b_n$ and $\sigma = a_n$, then we are left with the model

$$P\left(X_{n,n} \leq z\right) \approx \exp\left(-\left(1 + \gamma \frac{z - \mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right), \quad 1 + \gamma \frac{z - \mu}{\sigma} > 0. \tag{1.12}$$

With this model we can easily obtain maximum likelihood estimates of μ, σ and γ. To do this, we divide the data into m blocks and define $z_1, \ldots, z_m$ to be the block maxima of the

m blocks. Under the assumption that $Z_1, \ldots, Z_m$ are independent variables having the GEV distribution we get from (1.12) that the log likelihood is given by

$$\log L(\mu, \sigma, \gamma) = -m \log \sigma - \left(1 + \frac{1}{\gamma}\right) \sum_{i=1}^{m} \log \left(1 + \gamma \frac{z_i - \mu}{\sigma}\right) - \sum_{i=1}^{m} \left(1 + \gamma \frac{z_i - \mu}{\sigma}\right)^{-\frac{1}{\gamma}}. \tag{1.13}$$

The maximum likelihood estimates are then obtained by maximizing (1.13) with respect to μ, σ and γ.

Another popular model is the peaks over threshold model (POT). This model can be derived using Theorem 1.1.2. If we assume that (1.4) is satisfied, then there exists a positive function f such that

$$\begin{aligned} \lim_{t \uparrow x_*} P\left(\frac{X - t}{f(t)} > x \middle| X > t \right) &= \lim_{t \uparrow x_*} \frac{1 - F_X(t + f(t)x)}{1 - F_X(t)}, \quad x > 0 \\ &= (1 + \gamma x)^{-\frac{1}{\gamma}}, \quad 1 + \gamma x > 0. \end{aligned}$$

For t large, we thus have

$$P\left(\frac{X - t}{f(t)} > x \middle| X > t \right) \approx (1 + \gamma x)^{-\frac{1}{\gamma}}, \quad x > 0 \text{ and } 1 + \gamma x > 0,$$

which reduces to

$$P(X - t > z | X > t) \approx \left(1 + \gamma \frac{z}{\sigma}\right)^{-\frac{1}{\gamma}}, \quad z > 0 \text{ and } 1 + \gamma \frac{z}{\sigma} > 0, \tag{1.14}$$

if we set $z = f(t)x$ and $f(t) = \sigma$. From this we are able to get maximum likelihood estimates of γ and σ when we choose a threshold t. If we let $z_1, \ldots, z_k$ denote the k observations which are greater than the threshold t, then we obtain the log likelihood function from (1.14). The log likelihood is given by

$$\log L(\sigma, \gamma) = -k \log \sigma - \left(1 + \frac{1}{\gamma}\right) \sum_{i=1}^{k} \log \left(1 + \gamma \frac{z_i}{\sigma}\right). \tag{1.15}$$

The maximum likelihood estimates are obtained by maximizing (1.15) with respect to γ and σ.

Using maximum likelihood with block maxima or peaks over threshold is an easy way to estimate γ. There are many other ways to estimate γ but we will not go into detail about them. Among the methods of estimating γ for the generalized extreme value distribution are the Pickands estimator (Pickands, 1975), the moment estimator (Dekkers et al., 1989), and the probability-weighted moment estimator (Hosking et al., 1985).

When considering the POT model we have to choose the threshold ourselves. There are several ways to do this, but we will only discuss how to choose the threshold using a mean residual life plot. An introduction to mean residual life plots requires a small lemma about a property of the generalized Pareto distribution.

Lemma 1.4.1. *If $X \sim GPD(\sigma, \gamma)$, then $X - u | X > u \sim GPD(\sigma + \gamma u, \gamma)$.*

Proof. If $X \sim GPD(\sigma, \gamma)$, then $F_X(x) = 1 - \left(1 + \gamma \frac{x}{\sigma}\right)^{-\frac{1}{\gamma}}$. From this we get that

$$\begin{aligned} P(X - u > x | X > u) &= \frac{P(X > u + x, X > u)}{P(X > u)}, \quad x > 0 \\ &= \frac{1 - F_X(u + x)}{1 - F_X(u)} \\ &= \left(\frac{1 + \gamma \frac{x+u}{\sigma}}{1 + \gamma \frac{u}{\sigma}}\right)^{-\frac{1}{\gamma}} \\ &= \left(1 + \gamma \frac{x}{\sigma + \gamma u}\right)^{-\frac{1}{\gamma}}, \end{aligned}$$

which implies that $X - u | X > u \sim GPD(\sigma + \gamma u, \gamma)$. $\square$

If $X \sim GPD(\sigma, \gamma)$ with $\gamma < 1$, then

$$E(X) = \frac{\sigma}{1 - \gamma},$$

while $E(X) = \infty$ for $\gamma \geq 1$. So assuming $\gamma < 1$, it follows from Lemma 1.4.1 that

$$E(X - u | X > u) = \frac{\sigma + \gamma u}{1 - \gamma}, \quad u > 0,$$

and hence the mean excess function is linear in u. The mean residual life plot consists of the points

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} \left(x_{(i)} - u\right) \right) : u < x_{\max} \right\},$$

where $x_{(1)}, \ldots, x_{(n_u)}$ consists of the n_u observations that exceeds u, and $x_{\max}$ is the largest observation. If the GPD approximation is good at threshold u, then it should also be good at a higher threshold, so the mean excess function should be approximately linear in u beyond a good threshold.

Chapter 2

Pareto-type distributions

In this chapter we give an introduction to the Fréchet class. We start by considering the domain of attraction of this class, similar to the discussion of the Gumbel and extremal Weibull classes. Next we turn our attention to the estimation of the extreme value index γ for Pareto-type distributions which satisfy a second order condition. We prove asymptotic normality for a statistic proposed in Goegebeur et al. (2010) and use this to construct a class of estimators for γ. From this class of estimators we construct specific estimators using kernel functions. We end this chapter with a presentation of an estimator of the second order parameter. The asymptotic normality of the latter is established under a third order condition.

2.1 Domain of attraction

The class of Pareto-type models corresponds with the max domain of attraction of G_γ with $\gamma > 0$. The following proposition provides a characterization of the distributions that belong to this class.

Proposition 2.1.1. *(Gnedenko, 1943) Let X be a random variable with distribution function F_X. Then we have for x_* infinite that*

$$F_X \in \mathcal{D}(G_\gamma), \quad \gamma > 0 \quad \Leftrightarrow \quad 1 - F_X(x) = x^{-\frac{1}{\gamma}} l_{F_X}(x), \quad x > 0 \tag{2.1}$$

$$\Leftrightarrow \quad U(x) = x^\gamma l_U(x), \quad x > 1, \tag{2.2}$$

where $l_U(x)$ and $l_{F_X}(x)$ are slowly varying at infinity.

Tail quantile functions of the form (2.2) can be shown to satisfy (1.5) if x tends to infinity, in the following way

$$\begin{aligned}
\frac{U(tx) - U(t)}{a(t)} &= \frac{(tx)^\gamma l_U(tx) - t^\gamma l_U(t)}{a(t)} \\
&= \frac{l_U(t) t^\gamma}{a(t)} \left(\frac{l_U(tx)}{l_U(t)} x^\gamma - 1 \right) \\
&\sim \frac{x^\gamma - 1}{\gamma}
\end{aligned}$$

when choosing $a(t) = \gamma t^{\gamma} l_U(t) = \gamma U(t)$. More generally $a(t)$ can also be chosen as a function satisfying

$$\lim_{t\to\infty} \frac{a(t)}{U(t)} = \gamma.$$

This brings us to how a_n can be chosen as a normalizing constant. If we choose $a_n = a(n) = \gamma U(n)$ then we can use this constant as one of the normalizing constants for the Fréchet class. There exists full equivalence between the Pareto-type models and the extremal Weibull class. If we let X be a random variable with F_X belonging to the max domain of attraction of the extremal Weibull class with x_* as the right endpoint, and put $Y := (x_* - X)^{-1}$, then the Weibull class and the Pareto-type models are linked through the identification

$$F_X \in \mathcal{D}(G_\gamma), \quad \gamma < 0 \quad \Leftrightarrow \quad F_Y \in \mathcal{D}(G_\gamma), \quad \gamma > 0.$$

The equivalence follows easily because

$$1 - F_X\left(x_* - \frac{1}{x}\right) = P\left(X > x_* - \frac{1}{x}\right) = P\left((x_* - X)^{-1} > x\right) = 1 - F_Y(x).$$

Example 2.1.2. The Fréchet distribution has distribution function given by

$$F_X(x) = \exp\left(-x^{-\alpha}\right), \quad x > 0, \alpha > 0.$$

This means it has quantile function

$$Q(p) = (-\log p)^{-\frac{1}{\alpha}}, \quad 0 < p < 1,$$

and hence the tail quantile function is

$$U(x) = \left(-\log\left(1 - \frac{1}{x}\right)\right)^{-\frac{1}{\alpha}}, x > 1.$$

The Fréchet distribution has $\gamma = \frac{1}{\alpha}$ and the normalizing constant a_n can hence be chosen as

$$a_n = \gamma U(n) = \frac{1}{\alpha}\left(-\log\left(1 - \frac{1}{n}\right)\right)^{-\frac{1}{\alpha}}.$$

The normalizing constant b_n can be chosen as

$$b_n = U(n) = \left(-\log\left(1 - \frac{1}{n}\right)\right)^{-\frac{1}{\alpha}}.$$

Concerning the Fréchet distribution with $\alpha = 1$ we see that

$$\begin{aligned} P\left(\frac{X_{n,n} - b_n}{a_n} \le x\right) &= F_X^n\left(a_n x + b_n\right) \\ &= F_X^n\left(\left(-\log\left(1 - \frac{1}{n}\right)\right)^{-1} x + \left(-\log\left(1 - \frac{1}{n}\right)\right)^{-1}\right) \\ &= \left[\left(1 - \frac{1}{n}\right)^n\right]^{\frac{1}{1+x}} \\ &\to \exp\left(-(1+x)^{-1}\right) \quad \text{for } n \to \infty. \end{aligned}$$

The convergence of the Fréchet distribution to its limit is illustrated in Figure 2.1. The solid line is $G(x)$, the dashed line is for $n = 2$, the dotted line is for $n = 5$, while the dashed dotted line is for $n = 10$. It is clearly seen that when n grows, then $F_X^n(a_n x + b_n)$ converges pointwise to $G(x)$.

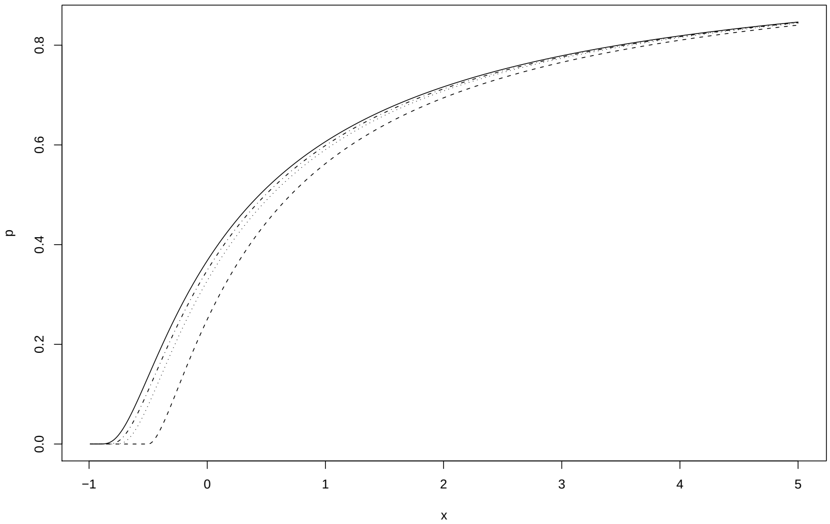

Figure 2.1: The convergence of $F_X^n(a_n x + b_n)$ to $G(x)$ for the Fréchet distribution with $\alpha = 1$.

■

Next we give two examples of distributions that are of Pareto-type.

Example 2.1.3. The Burr distribution has a distribution function given by

$$F_X(x) = 1 - \left(\frac{\zeta}{\zeta + x^\delta}\right)^\lambda, \quad x > 0, \quad \lambda, \zeta, \delta > 0.$$

In order to verify that the Burr distribution is of Pareto-type we start with

$$\begin{aligned} 1 - F_X(x) &= \left(\frac{\zeta}{\zeta + x^\delta}\right)^\lambda \\ &= x^{-\delta\lambda}\left(\frac{\zeta}{\zeta x^{-\delta} + 1}\right)^\lambda. \end{aligned}$$

It is easily seen that $g(x) := \left(\frac{\zeta}{\zeta x^{-\delta}+1}\right)^\lambda$ is slowly varying at infinity since it converges to a constant when $x \to \infty$. So the Burr distribution is of Pareto-type with $\gamma = \frac{1}{\lambda\delta}$.

■

Example 2.1.4. The absolute T distribution has distribution function given by

$$F_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \int_{-x}^{x} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} dt, \quad x > 0, \quad n \in \mathbb{N}.$$

In order to verify that the absolute T distribution is of Pareto-type we start with

$$\begin{aligned}
1 - F_X(x) &= 2\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \int_x^\infty \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} dt \\
&= 2\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \int_x^\infty \left(\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \left(\frac{n}{t^2} + 1\right)^{-\frac{n+1}{2}} dt \\
&= K \int_x^\infty t^{-n-1} \left(nt^{-2} + 1\right)^{-\frac{n+1}{2}} dt,
\end{aligned}$$

where $K := 2\frac{n^{\frac{n}{2}}\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)}$. We are concerned with large values of x, so we make a Taylor series expansion of $(1+x)^{-\frac{n+1}{2}}$ around 0, which yields

$$\begin{aligned}
\left(nt^{-2} + 1\right)^{-\frac{n+1}{2}} =& 1 - \frac{n+1}{2} nt^{-2} + \frac{1}{2}\frac{n+1}{2}\left(\frac{n+1}{2} + 1\right) n^2 t^{-4} \\
& - \frac{1}{6}\frac{n+1}{2}\left(\frac{n+1}{2} + 1\right)\left(\frac{n+1}{2} + 2\right)\left(1 + \tilde{t}\right)^{-\frac{n+1}{2}-3} n^3 t^{-6},
\end{aligned}$$

where $\tilde{t}$ is between 0 and $\frac{n}{t^2}$. From this it follows that

$$\begin{aligned}
1 - F_X(x) =& K\Bigg(\int_x^\infty t^{-n-1} dt - \frac{n(n+1)}{2}\int_x^\infty t^{-n-3} dt \\
& + \frac{n^2(n+1)(n+3)}{8}\int_x^\infty t^{-n-5} dt \\
& - \frac{n^3(n+1)(n+3)(n+5)}{48}\int_x^\infty t^{-n-1}\left(1 + \tilde{t}\right)^{-\frac{n+1}{2}-3} t^{-6} dt\Bigg).
\end{aligned}$$

Since $\left(1+\tilde{t}\right)^{-\frac{n+1}{2}-3} \leq 1$ it follows that $\int_x^\infty t^{-n-1}\left(1+\tilde{t}\right)^{-\frac{n+1}{2}-3} t^{-6} dt \leq \int_x^\infty t^{-n-7} dt$, and hence

$$\begin{aligned}
1 - F_X(x) =& K\left(\frac{x^{-n}}{n} - \frac{n(n+1)}{2(n+2)} x^{-n-2} + \frac{n^2(n+1)(n+3)}{8(n+4)} x^{-n-4} + O\left(x^{-n-6}\right)\right) \\
=& x^{-n} C_0 \left(1 - \frac{n^2(n+1)}{2(n+2)} x^{-2} + \frac{n^3(n+1)(n+3)}{8(n+4)} x^{-4} + O\left(x^{-6}\right)\right), \qquad (2.3)
\end{aligned}$$

where $C_0 := \frac{K}{n}$. Since the function $g(x) := C_0\left(1 - \frac{n^2(n+1)}{2(n+2)}x^{-2} + \frac{n^3(n+1)(n+3)}{8(n+4)}x^{-4} + O\left(x^{-6}\right)\right)$ converges to a constant, when $x \to \infty$, the function is slowly varying at infinity and hence the absolute T distribution is of Pareto-type with $\gamma = \frac{1}{n}$.

■

2.2 Estimation of the extreme value index

In the analysis of Pareto-type models, estimation of γ plays a central role. The asymptotic distribution of the estimator of γ is usually established under the following second order condition on the tail behaviour.

Assumption 2.2.1 (Second order condition)**.** *There exists a positive real parameter γ, a negative real parameter ρ and a function b with $b(t) \to 0$ for $t \to \infty$, of constant sign for large values of t, such that*

$$\lim_{t\to\infty} \frac{\log U(tx) - \log U(t) - \gamma \log x}{b(t)} = \frac{x^\rho - 1}{\rho}, \quad \forall x > 0.$$

The second order condition implies that $|b|$ is regularly varying with index ρ (Geluk and Haan, 1987), so the parameter ρ determines the rate of convergence for $\log U(tx) - \log U(t)$ to its limit $\gamma \log x$, when t tends to infinity. If ρ is close to zero then the convergence is slow and the estimation of tail parameters is practically difficult.
We will now verify that the Burr distribution and the absolute T distribution satisfy the second order condition. That they are of Pareto-type was verified in Example 2.1.3 and Example 2.1.4 respectively.

Example 2.2.2. In order to verify that the Burr distribution satisfies the second order condition we need to find its tail quantile function. The quantile function of the Burr distribution is easily found by inverting the distribution function and it is given by

$$Q(p) = \zeta^{\frac{1}{\delta}} \left((1-p)^{-\frac{1}{\lambda}} - 1 \right)^{\frac{1}{\delta}}, \quad 0 < p < 1.$$

From this we obtain the tail quantile function

$$\begin{aligned} U(x) &= Q\left(1 - \frac{1}{x}\right) \\ &= x^\gamma \zeta^{\frac{1}{\delta}} \left(1 - x^{-\frac{1}{\lambda}}\right)^{\frac{1}{\delta}}, \quad x > 1. \end{aligned}$$

We start with the expression

$$\log U(tx) - \log U(t) - \gamma \log x = \frac{1}{\delta} \log\left(1 - (xt)^{-\frac{1}{\lambda}}\right) - \frac{1}{\delta} \log\left(1 - t^{-\frac{1}{\lambda}}\right).$$

If we make a Taylor series expansion of $\log(1-x)$ around 0, we obtain

$$\begin{aligned} \log U(tx) - \log U(t) - \gamma \log x &= \frac{1}{\delta}\left(-(tx)^{-\frac{1}{\lambda}} - \frac{1}{2}(tx)^{-\frac{2}{\lambda}}\right) - \frac{1}{\delta}\left(-t^{-\frac{1}{\lambda}} - \frac{1}{2}t^{-\frac{2}{\lambda}}\right) + O\left(t^{-\frac{3}{\lambda}}\right) \\ &= \frac{\frac{1}{\lambda\delta} t^{-\frac{1}{\lambda}} \left(x^{-\frac{1}{\lambda}} - 1\right)}{-\frac{1}{\lambda}} + \frac{\frac{1}{\lambda\delta} t^{-\frac{2}{\lambda}} \left(x^{-\frac{2}{\lambda}} - 1\right)}{-\frac{2}{\lambda}} + O\left(t^{-\frac{3}{\lambda}}\right) \qquad (2.4) \\ &= \frac{\gamma t^{-\frac{1}{\lambda}} \left(x^{-\frac{1}{\lambda}} - 1\right)}{-\frac{1}{\lambda}} + O\left(t^{-\frac{2}{\lambda}}\right). \qquad (2.5) \end{aligned}$$

From (2.5) we see that if we choose $\rho = -\frac{1}{\lambda}$ and $b(t) = \gamma t^\rho$, then the Burr distribution satisfies the second order condition. More generally $b(t)$ can be chosen such that $b(t) = \gamma t^\rho (1 + o(1))$. ■

Example 2.2.3. From (2.3) we get for the absolue T distribution that

$$1 - F_X(x) = x^{-\frac{1}{\gamma}} C_0 \left(1 - C_1 x^{-2} + C_2 x^{-4} + O\left(x^{-6}\right)\right),$$

where $C_1 := \frac{n^2(n+1)}{2(n+2)}$ and $C_2 := \frac{n^3(n+1)(n+3)}{8(n+4)}$. In order to find the tail quantile function we have to invert

$$\frac{1}{y} = x^{-\frac{1}{\gamma}} C_0 \left(1 - C_1 x^{-2} + C_2 x^{-4} + O\left(x^{-6}\right)\right).$$

From this we find

$$x = C_0^{\gamma} y^{\gamma} \left(1 - C_1 x^{-2} + C_2 x^{-4} + O\left(x^{-6}\right)\right)^{\gamma}.$$

If we make a Taylor series expansion of $(1-x)^{\gamma}$ around $x = 0$, we obtain

$$\begin{aligned} x =& C_0^{\gamma} y^{\gamma} \Bigg(1 - \gamma\left(C_1 x^{-2} - C_2 x^{-4} + O\left(x^{-6}\right)\right) \\ &+ \frac{1}{2}\gamma(\gamma-1)\left(C_1 x^{-2} - C_2 x^{-4} + O\left(x^{-6}\right)\right)^2 + O\left(x^{-6}\right)\Bigg) \\ =& C_0^{\gamma} y^{\gamma} \Bigg(1 - \gamma C_1 C_0^{-2\gamma} y^{-2\gamma} \left(1 - \gamma C_1 x^{-2} + \left(\gamma C_2 + \frac{\gamma(\gamma-1)}{2} C_1^2\right) x^{-4} + O\left(x^{-6}\right)\right)^{-2} \\ &+ \left(\gamma C_2 + \frac{\gamma(\gamma-1)}{2} C_1^2\right) x^{-4} + O\left(x^{-6}\right)\Bigg). \end{aligned}$$

Now we make a Taylor series expansion of $(1-x)^{-2}$ in which case we obtain

$$\begin{aligned} x =& C_0^{\gamma} y^{\gamma} \Bigg(1 - \gamma C_1 C_0^{-2\gamma} y^{-2\gamma} \left(1 + 2\gamma C_1 x^{-2} + O\left(x^{-4}\right)\right) \\ &+ \left(\gamma C_2 + \frac{\gamma(\gamma-1)}{2} C_1^2\right) x^{-4} + O\left(x^{-6}\right)\Bigg). \end{aligned}$$

If we substitute the right hand side into the place of x, then it follows that

$$\begin{aligned} x =& C_0^{\gamma} y^{\gamma} \Bigg(1 - \gamma C_1 C_0^{-2\gamma} y^{-2\gamma} \\ &+ \left(\gamma C_2 - \frac{\gamma(3\gamma+1)}{2} C_1^2\right) C_0^{-4\gamma} y^{-4\gamma} + O\left(y^{-6\gamma}\right)\Bigg). \end{aligned}$$

So the tail quantile function can be written as

$$U(x) = C_0^{\gamma} x^{\gamma} \left(1 - D_1 x^{-2\gamma} + D_2 x^{-4\gamma} + O(x^{-6\gamma})\right),$$

where $D_1 := \gamma C_1 C_0^{-2\gamma}$, and $D_2 := \left(\gamma C_2 - \frac{\gamma(3\gamma+1)}{2} C_1^2\right) C_0^{-4\gamma}$. We are now ready to verify that the absolute T distribution satisfies the second order condition. We start with the expression

$$\begin{aligned} \log U(xt) - \log U(t) - \gamma \log x =& \log\left(1 - D_1 (xt)^{-2\gamma} + D_2 (xt)^{-4\gamma} + O(t^{-6\gamma})\right) \\ &- \log\left(1 - D_1 t^{-2\gamma} + D_2 t^{-4\gamma} + O(t^{-6\gamma})\right). \end{aligned}$$

By making a Taylor series expansion of $\log(1-x)$ around $x = 0$ we obtain

$$\begin{aligned} \log U(xt) - \log U(t) - \gamma \log x =& - D_1 (xt)^{-2\gamma} + D_2 (xt)^{-4\gamma} - \frac{1}{2}\left(D_1 (xt)^{-2\gamma} - D_2 (xt)^{-4\gamma}\right)^2 \\ &+ D_1 t^{-2\gamma} - D_2 t^{-4\gamma} + \frac{1}{2}\left(D_1 t^{-2\gamma} - D_2 t^{-4\gamma}\right)^2 + O(t^{-6\gamma}) \\ =& - D_1 t^{-2\gamma}\left(x^{-2\gamma} - 1\right) + \left(D_2 - \frac{1}{2} D_1^2\right) t^{-4\gamma}\left(x^{-4\gamma} - 1\right) \\ &+ O(t^{-6\gamma}) \qquad (2.6) \\ =& - D_1 t^{-2\gamma}\left(x^{-2\gamma} - 1\right) + O(t^{-4\gamma}). \qquad (2.7) \end{aligned}$$

From (2.7) we see that if we choose $\rho = -2\gamma$ and $b(t)$ of the form $b(t) = -\rho D_1 t^{\rho}(1 + o(1))$, then the absolute T distribution satisfies the second order condition.

■

We now return to the estimation of γ. The estimator of γ we will consider is based on a kernel statistic with kernel function K. This statistic is given by

$$T_{n,k}(K) := \frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) Z_j, \tag{2.8}$$

where $Z_j := j\left(\log X_{n-j+1,n} - \log X_{n-j,n}\right)$. This statistic will also serve as the basic building block for the ρ estimator we propose in section 2.3. We need some conditions on the kernel function, but first we introduce the following notation

$$\begin{aligned}
\mu(K) &:= \int_0^1 K(u)du, \\
I_1(K,\rho) &:= \int_0^1 K(u)u^{-\rho}du, \\
\sigma^2(K) &:= \int_0^1 K^2(u)du.
\end{aligned}$$

With this notation the kernel function must satisfy

Assumption 2.2.4. *Let K be a function defined on $(0,1)$ such that*

(i) $K(t) = \frac{1}{t}\int_0^t u(v)dv$ *for some function u satisfying* $\left|(k+1)\int_{\frac{j-1}{k+1}}^{\frac{j}{k+1}} u(t)dt\right| \le f\left(\frac{j}{k+1}\right)$ *for some positive continuous and integrable function f defined on* $(0,1)$,

(ii) $\sigma^2(K) < \infty$,

(iii) $\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) = \mu(K) + o\left(\frac{1}{\sqrt{k}}\right)$ *for* $k \to \infty$,

(iv) $\max_{i\in 1,\dots,k}\left|K\left(\frac{i}{k+1}\right)\right| = o\left(\sqrt{k}\right)$ *for* $k \to \infty$,

(v) $\int_0^1 |K(u)|u^{|\rho|-1-\epsilon}du < \infty$ *for some* $\epsilon > 0$.

Next we give an example of a kernel function which satisfies Assumption 2.2.4.

Example 2.2.5. An important special subset of kernel functions which satisfy Assumption 2.2.4 is the kernel $K(t) := t^{\tau}(-\log t)^{\delta}$, where $\tau, \delta \ge 0$ are tuning parameters. This class of kernel functions has as important special cases the Hill kernel $\mathbb{H} := 1$ corresponding to $\tau = \delta = 0$, the power kernels $\mathbb{K}_{\tau}(t) := t^{\tau}, \quad \tau > 0$ and the log kernels $\mathbb{L}_{\delta}(t) := (-\log t)^{\delta}, \quad \delta > 0$.

Lemma 2.2.6. *The function $K(t) := t^{\tau}(-\log t)^{\delta}$ satisfies Assumption 2.2.4.*

The proof of Lemma 2.2.6 can be found in Appendix 2.4.

■

With Assumption 2.2.1 and Assumption 2.2.4 we are able to establish the following result.

Theorem 2.2.7. *Let $X_1, \ldots, X_n$ be i.i.d. random variables according to a distribution satisfying Assumption 2.2.1. If further Assumption 2.2.4 holds, then for $k, n \to \infty$ such that $\frac{k}{n} \to 0$ we have*

$$T_{n,k}(K) \stackrel{\mathcal{D}}{=} \gamma\mu(K) + \gamma\sigma(K)\frac{N_k(K)}{\sqrt{k}} + b\left(\frac{n}{k}\right) I_1(K,\rho)\left(1 + o_{\mathbb{P}}(1)\right), \tag{2.9}$$

where $N_k(K)$ is asymptotically a standard normal random variable.

A proof of this theorem is given in Goegebeur et al. (2010), we will however give an alternative proof of the result.

Proof of Theorem 2.2.7. Let $U_{1,n} \leq \ldots \leq U_{n,n}$ be order statistics from a random sample of size n from the $U(0,1)$ distribution. By using the inverse probability integral transform we find that

$$\begin{aligned} X_{i,n} &\stackrel{\mathcal{D}}{=} Q\left(U_{i,n}\right) \\ &\stackrel{\mathcal{D}}{=} Q\left(1 - U_{n-i+1,n}\right) \\ &= U\left(\frac{1}{U_{n-i+1,n}}\right). \end{aligned}$$

Since the X_i are of Pareto-type it follows that

$$X_{i,n} \stackrel{\mathcal{D}}{=} \left(\frac{1}{U_{n-i+1,n}}\right)^{\gamma} l_U\left(\frac{1}{U_{n-i+1,n}}\right).$$

From this we get

$$\log X_{i,n} \stackrel{\mathcal{D}}{=} -\gamma \log U_{n-i+1,n} + \log l_U\left(\frac{1}{U_{n-i+1,n}}\right).$$

Hence

$$\log X_{n-j+1,n} - \log X_{n-k,n} \stackrel{\mathcal{D}}{=} -\gamma \log \frac{U_{j,n}}{U_{k+1,n}} + \log \frac{l_U\left(\frac{U_{k+1,n}}{U_{j,n}}\frac{1}{U_{k+1,n}}\right)}{l_U\left(\frac{1}{U_{k+1,n}}\right)}.$$

Since $\frac{U_{j,n}}{U_{k+1,n}} \stackrel{\mathcal{D}}{=} V_{j,k}$, where $V_{j,k}$ is the j'th order statistic in a random sample of size k from the $U(0,1)$ distribution, it follows that

$$\begin{aligned} \log X_{n-j+1,n} - \log X_{n-k,n} &\stackrel{\mathcal{D}}{=} -\gamma \log V_{j,k} + \log \frac{l_U\left(\frac{1}{V_{j,k}}\frac{1}{U_{k+1,n}}\right)}{l_U\left(\frac{1}{U_{k+1,n}}\right)} \\ &\stackrel{\mathcal{D}}{=} -\gamma \log\left(1 - V_{k-j+1,k}\right) + \log \frac{l_U\left(\frac{1}{V_{j,k}}\frac{1}{U_{k+1,n}}\right)}{l_U\left(\frac{1}{U_{k+1,n}}\right)}. \end{aligned}$$

Using that the quantile function of the standard exponential distribution is $Q(p) = -\log(1 - p), \quad 0 < p < 1$, and denoting by $E_{1,n} \leq \ldots \leq E_{n,n}$ the order statistics of a random sample of size n from the standard exponential distribution, we get using Assumption 2.2.1 and inspired by Lemma 2.4.3, that

$$\log X_{n-j+1,n} - \log X_{n-k,n} \stackrel{\mathcal{D}}{=} \gamma E_{k-j+1,k} + b_0\left(\frac{1}{U_{k+1,n}}\right)\frac{\left(\frac{1}{V_{j,k}}\right)^{\rho}-1}{\rho} + b_0\left(\frac{1}{U_{k+1,n}}\right)\tilde{R}_{n,k}(j),$$

where $\tilde{R}_{n,k}(j) := \frac{\log U\left(\frac{1}{U_{k+1,n}}\frac{1}{V_{j,k}}\right) - \log U\left(\frac{1}{U_{k+1,n}}\right) - \gamma\log\frac{1}{V_{j,k}}}{b_0\left(\frac{1}{U_{k+1,n}}\right)} - \frac{\left(\frac{1}{V_{j,k}}\right)^{\rho}-1}{\rho}$. Thus

$$\begin{aligned} Z_j &= j\left(\log X_{n-j+1,n} - \log X_{n-j,n}\right) \\ &\stackrel{\mathcal{D}}{=} j\left(\gamma E_{k-j+1,k} - \gamma E_{k-j,k} + b_0\left(\frac{1}{U_{k+1,n}}\right)\frac{\left(\frac{1}{V_{j,k}}\right)^{\rho} - \left(\frac{1}{V_{j+1,k}}\right)^{\rho}}{\rho} + b_0\left(\frac{1}{U_{k+1,n}}\right)R_{n,k}(j)\right), \end{aligned} \tag{2.10}$$

where $R_{n,k}(j) := \tilde{R}_{n,k}(j) - \tilde{R}_{n,k}(j+1)$, with the convention $\tilde{R}_{n,k}(k+1) := 0$ and with b_0 a function satisfying $b_0(t) \sim b(t)$ for $t \to \infty$. Using the Rényi representation (Rényi, 1953) we can express each $E_{j,k}$ as

$$\left\{E_{j,k}\right\}_{j=1,\dots,k} \stackrel{\mathcal{D}}{=} \left\{\sum_{i=1}^{j}\frac{E_{k-i+1}}{k-i+1}\right\}_{j=1,\dots,k},$$

where the $E_1, \dots, E_k$ are independent random variables from a standard exponential distribution. Hence

$$\begin{aligned} E_{k-j+1,k} - E_{k-j,k} &\stackrel{\mathcal{D}}{=} \sum_{i=1}^{k-j+1}\frac{E_{k-i+1}}{k-i+1} - \sum_{i=1}^{k-j}\frac{E_{k-i+1}}{k-i+1} \\ &= \frac{E_j}{j}. \end{aligned} \tag{2.11}$$

Combining (2.10) and (2.11) we find that

$$Z_j \stackrel{\mathcal{D}}{=} \gamma E_j + b_0\left(\frac{1}{U_{k+1,n}}\right) j\frac{\left(\frac{1}{V_{j,k}}\right)^{\rho} - \left(\frac{1}{V_{j+1,k}}\right)^{\rho}}{\rho} + b_0\left(\frac{1}{U_{k+1,n}}\right) jR_{n,k}(j).$$

Let $Y_{1,k} \le \ldots \le Y_{k,k}$ be order statistics of a random sample of size k from the standard strict Pareto distribution. Then we have

$$\begin{aligned} \frac{1}{V_{j,k}} &\stackrel{\mathcal{D}}{=} \frac{1}{1 - V_{k-j+1,k}} \\ &\stackrel{\mathcal{D}}{=} Y_{k-j+1,k}. \end{aligned}$$

Using this we get that

$$Z_j \stackrel{\mathcal{D}}{=} \gamma E_j + b_0\left(Y_{n-k,n}\right) j\frac{Y_{k-j+1,k}^{\rho} - Y_{k-j,k}^{\rho}}{\rho} + b_0\left(Y_{n-k,n}\right) jR_{n,k}(j).$$

Hence

$$\begin{aligned}
T_{n,k}(K) \overset{\mathcal{D}}{=} &\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right)\left(\gamma E_j + b_0\left(Y_{n-k,n}\right) j \frac{Y_{k-j+1,k}^{\rho} - Y_{k-j,k}^{\rho}}{\rho} + b_0\left(Y_{n-k,n}\right) j R_{n,k}(j)\right)\\
=&\gamma \frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) E_j + b_0\left(Y_{n-k,n}\right)\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) j \frac{Y_{k-j+1,k}^{\rho} - Y_{k-j,k}^{\rho}}{\rho}\\
&+ b_0\left(Y_{n-k,n}\right)\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) j R_{n,k}(j)\\
=: &\, T_{n,k}^{(1)} + T_{n,k}^{(2)} + T_{n,k}^{(3)}.
\end{aligned}$$

Using Assumption 2.2.4 (iii) we get for the first term that

$$\begin{aligned}
T_{n,k}^{(1)} &= \gamma \frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) + \gamma \frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right)\left(E_j - 1\right)\\
&= \gamma\mu(K) + o\left(\frac{1}{\sqrt{k}}\right) + \gamma\sigma(K)\frac{\tilde{N}_k(K)}{\sqrt{k}},
\end{aligned} \tag{2.12}$$

where $\tilde{N}_k(K) := \sqrt{k}\frac{\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right)\left(E_j-1\right)}{\sigma(K)}$. The term $\tilde{N}_k(K)$ is according to Lemma 2.4.1 in Appendix 2.4 an asymptotic standard normal random variable. In (2.12) we can combine the $o\left(\frac{1}{\sqrt{k}}\right)$ with $\tilde{N}_k(K)$ to get

$$T_{n,k}^{(1)} = \gamma\mu(K) + \gamma\sigma(K)\frac{N_k(K)}{\sqrt{k}},$$

where $N_k(K)$ is again an asymptotic standard normal random variable.
Since $Y_{i,k} \overset{\mathcal{D}}{=} \frac{1}{1-U_{i,k}}$ and the standard exponential distribution has quantile function $Q(p) = -\log(1-p)$ it follows that $T_{n,k}^{(2)}$ can be written as

$$T_{n,k}^{(2)} \overset{\mathcal{D}}{=} b_0\left(Y_{n-k,n}\right)\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) j \frac{\exp\left(\rho E_{k-j+1,n}\right) - \exp\left(\rho E_{k-j,n}\right)}{\rho}.$$

Using the mean value theorem we find that

$$T_{n,k}^{(2)} \overset{\mathcal{D}}{=} b_0\left(Y_{n-k,n}\right)\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) j \left(E_{k-j+1,n} - E_{k-j,n}\right)\exp\left(\rho Q_{j,k}\right),$$

where $Q_{j,k}$ is a random value between $E_{k-j,k}$ and $E_{k-j+1,k}$, and hence

$$\begin{aligned}
T_{n,k}^{(2)} \stackrel{\mathcal{D}}{=} & b_0\left(Y_{n-k,n}\right) \frac{1}{k} \sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) E_j \exp\left(\rho Q_{j,k}\right) \\
= & b_0\left(Y_{n-k,n}\right) \frac{1}{k} \sum_{j=1}^{k} K\left(\frac{j}{k+1}\right)\left(\frac{j}{k+1}\right)^{-\rho} E_j \\
& + b_0\left(Y_{n-k,n}\right) \frac{1}{k} \sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) E_j\left(\exp\left(\rho Q_{j,k}\right)-\left(\frac{j}{k+1}\right)^{-\rho}\right) \\
= & : T_{n,k}^{(2,1)}+T_{n,k}^{(2,2)}.
\end{aligned}$$

Concerning the term $T_{n,k}^{(2,1)}$ we get

$$\begin{aligned}
T_{n,k}^{(2,1)} = & b_0\left(Y_{n-k,n}\right) \frac{1}{k} \sum_{j=1}^{k} K\left(\frac{j}{k+1}\right)\left(\frac{j}{k+1}\right)^{-\rho} \\
& + b_0\left(Y_{n-k,n}\right) \frac{1}{k} \sum_{j=1}^{k} K\left(\frac{j}{k+1}\right)\left(\frac{j}{k+1}\right)^{-\rho}\left(E_j-1\right),
\end{aligned}$$

so by the law of large numbers it follows that

$$T_{n,k}^{(2,1)}=b_0\left(Y_{n-k,n}\right) I_1(K, \rho)\left(1+o_{\mathbb{P}}(1)\right).$$

We now turn to $T_{n,k}^{(2,2)}$. Note that for $j=1, \ldots, k$ we have that

$$\begin{aligned}
\exp\left(E_{k-j+1,k}\right) & \stackrel{\mathcal{D}}{=} \exp\left(-\log\left(1-U_{k-j+1,k}\right)\right) \\
& \stackrel{\mathcal{D}}{=} \exp\left(-\log\left(U_{j,k}\right)\right) \\
& =\frac{1}{U_{j,k}},
\end{aligned}$$

and hence

$$\begin{aligned}
\left|\exp\left(\rho Q_{j,k}\right)-\left(\frac{j}{k+1}\right)^{-\rho}\right| & \leq \max\left\{\left|\exp\left(\rho E_{k-j,k}\right)-\left(\frac{j}{k+1}\right)^{-\rho}\right|,\left|\exp\left(\rho E_{k-j+1,k}\right)-\left(\frac{j}{k+1}\right)^{-\rho}\right|\right\} \\
& \stackrel{\mathcal{D}}{=} \max\left\{\left|U_{j+1,k}^{-\rho}-\left(\frac{j}{k+1}\right)^{-\rho}\right|,\left|U_{j,k}^{-\rho}-\left(\frac{j}{k+1}\right)^{-\rho}\right|\right\} \\
& \leq \max\left\{\left|U_{j+1,k}^{-\rho}-\left(\frac{j+1}{k+1}\right)^{-\rho}\right|+c_{j,k},\left|U_{j,k}^{-\rho}-\left(\frac{j}{k+1}\right)^{-\rho}\right|\right\},
\end{aligned}$$

where $c_{j,k} = \left(\frac{j+1}{k+1}\right)^{-\rho} - \left(\frac{j}{k+1}\right)^{-\rho}$. From this it follows that

$$\begin{aligned}
&\left|\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) E_j \left(\exp\left(\rho Q_{j,k}\right) - \left(\frac{j}{k+1}\right)^{-\rho}\right)\right| \\
&\quad \leq \frac{1}{k}\sum_{j=1}^{k} \left|K\left(\frac{j}{k+1}\right)\right| E_j \left|U_{j+1,k}^{-\rho} - \left(\frac{j+1}{k+1}\right)^{-\rho}\right| + \frac{1}{k}\sum_{j=1}^{k} \left|K\left(\frac{j}{k+1}\right)\right| c_{j,k} E_j \\
&\quad\quad + \frac{1}{k}\sum_{j=1}^{k} \left|K\left(\frac{j}{k+1}\right)\right| E_j \left|U_{j,k}^{-\rho} - \left(\frac{j}{k+1}\right)^{-\rho}\right| \\
&\quad =: T_{n,k}^{(2,2,1)} + T_{n,k}^{(2,2,2)} + T_{n,k}^{(2,2,3)}.
\end{aligned}$$

According to Lemma 2.4.2 the terms $T_{n,k}^{(2,2,1)}$ and $T_{n,k}^{(2,2,3)}$ are $O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)$. Using the mean value theorem we see that we can write the term $T_{n,k}^{(2,2,2)}$ as

$$T_{n,k}^{(2,2,2)} = \frac{|\rho|}{k+1}\frac{1}{k}\sum_{j=1}^{k} \left|K\left(\frac{j}{k+1}\right)\right| z_{j,k}^{|\rho|-1} E_j,$$

where $z_{j,k}$ is a value between $\frac{j}{k+1}$ and $\frac{j+1}{k+1}$. When $|\rho| \geq 1$ it follows that

$$T_{n,k}^{(2,2,2)} \leq \frac{|\rho|}{k+1}\frac{1}{k}\sum_{j=1}^{k} \left|K\left(\frac{j}{k+1}\right)\right| E_j,$$

and hence by the law of large numbers it follows that $T_{n,k}^{(2,2,2)} = O_{\mathbb{P}}\left(\frac{1}{k}\right)$. When $|\rho| < 1$ we have

$$T_{n,k}^{(2,2,2)} \leq \frac{|\rho|}{k+1}\frac{1}{k}\sum_{j=1}^{k} \left|K\left(\frac{j}{k+1}\right)\right| \left(\frac{j}{k+1}\right)^{|\rho|-1} E_j,$$

which by Assumption 2.2.4 (v) and the law of large numbers implies that $T_{n,k}^{(2,2,2)} = O_{\mathbb{P}}\left(\frac{1}{k}\right)$. So

$$T_{n,k}^{(2)} = b_0\left(Y_{n-k,n}\right) I_1(K,\rho)\left(1 + o_{\mathbb{P}}(1)\right).$$

Concerning the term $T_{n,k}^{(3)}$ we find using Assumption 2.2.4 (i) that

$$\begin{aligned}
\left|T_{n,k}^{(3)}\right| &= \left|b_0\left(Y_{n-k,n}\right)\frac{k+1}{k}\sum_{j=1}^{k} R_{n,k}(j)\int_0^{\frac{j}{k+1}} u(v)dv\right| \\
&= \left|b_0\left(Y_{n-k,n}\right)\frac{k+1}{k}\sum_{j=1}^{k} R_{n,k}(j)\sum_{i=1}^{j}\int_{\frac{i-1}{k+1}}^{\frac{i}{k+1}} u(v)dv\right| \\
&= \left|b_0\left(Y_{n-k,n}\right)\frac{k+1}{k}\sum_{i=1}^{k}\int_{\frac{i-1}{k+1}}^{\frac{i}{k+1}} u(v)dv\sum_{j=i}^{k} R_{n,k}(j)\right| \\
&\leq \left|b_0\left(Y_{n-k,n}\right)\right|\frac{1}{k}\sum_{i=1}^{k} f\left(\frac{i}{k+1}\right)\left|\sum_{j=i}^{k} R_{n,k}(j)\right|.
\end{aligned}$$

For the term $\sum_{j=i}^{k} R_{n,k}(j)$ it follows that

$$\begin{aligned}\sum_{j=i}^{k} R_{n,k}(j) &= \sum_{j=i}^{k} \left(\tilde{R}_{n,k}(j) - \tilde{R}_{n,k}(j+1)\right) \\ &= \tilde{R}_{n,k}(i).\end{aligned}$$

For $\delta, \epsilon > 0$ there exists n_0 such that for any $n \geq n_0$, with arbitrary large probability, for $i = 1, \ldots, k$,

$$\begin{aligned}\left|\sum_{j=i}^{k} R_{n,k}(j)\right| &\leq \epsilon \left(\frac{1}{V_{i,k}}\right)^{\rho} \max\left(\left(\frac{1}{V_{i,k}}\right)^{\delta}, \left(\frac{1}{V_{i,k}}\right)^{-\delta}\right) \\ &= \epsilon V_{i,k}^{-\rho-\delta},\end{aligned}$$

using Lemma 2.4.3. Hence

$$\sup_{i\in\{1,\ldots,k\}} \left|\frac{\sum_{j=i}^{k} R_{n,k}(j)}{V_{i,k}^{-\rho-\delta}}\right| = o_{\mathbb{P}}(1)$$

leading to

$$\left|T_{n,k}^{(3)}\right| \leq b_0\left(Y_{n-k,n}\right) o_{\mathbb{P}}(1) \frac{1}{k}\sum_{i=1}^{k} f\left(\frac{i}{k+1}\right)\left(V_{i,k}^{-\rho-\delta}\right),$$

which by Assumption 2.2.4 (i) and assuming $\delta < |\rho|$ is $o_{\mathbb{P}}\left(b_0\left(Y_{n-k,n}\right)\right)$. Combining the results on $T_{n,k}^{(1)}$, $T_{n,k}^{(2)}$ and $T_{n,k}^{(3)}$ establishes the result. □

Using Theorem 2.2.7 we can create a class of estimators $\hat{\gamma}_k(K) := \frac{T_{n,k}(K)}{\mu(K)}$ for γ in the following way

Proposition 2.2.8. *Let $X_1, \ldots, X_n$ be i.i.d. random variables according to a distribution satisfying Assumption 2.2.1. If further Assumption 2.2.4 holds with $\mu(K) \neq 0$, then for $k, n \to \infty$ such that $\frac{k}{n} \to 0$ and $\sqrt{k} b\left(\frac{n}{k}\right) \to \lambda$ for some constant λ we have*

$$\sqrt{k}\left(\hat{\gamma}_k(K) - \gamma\right) \to N\left(\lambda\frac{I_1(K,\rho)}{\mu(K)}, \gamma^2\frac{\sigma^2(K)}{\mu^2(K)}\right). \tag{2.13}$$

Proof. We have

$$\begin{aligned}\sqrt{k}\left(\hat{\gamma}_k(K) - \gamma\right) &\stackrel{\mathcal{D}}{=} \gamma\frac{\sigma(K)}{\mu(K)} N_k(K) + \sqrt{k} b\left(\frac{n}{k}\right)\frac{I_1(K,\rho)}{\mu(K)}\left(1 + o_{\mathbb{P}}(1)\right) \\ &\to N\left(\lambda\frac{I_1(K,\rho)}{\mu(K)}, \gamma^2\frac{\sigma^2(K)}{\mu^2(K)}\right),\end{aligned}$$

under the conditions of the Proposition. □

We verified in Lemma 2.2.6 that the kernel function $K(t) = t^{\tau}(-\log t)^{\delta}$ satisfies Assumption 2.2.4. This allows us to construct consistent estimators which are asymptotically normal using this kernel. We do so in Corollary 2.2.9.

Corollary 2.2.9. *Let $X_1, \ldots, X_n$ be i.i.d. random variables according to a distribution satisfying Assumption 2.2.1. For $k, n \to \infty$ such that $\frac{k}{n} \to 0$ and $\sqrt{k} b\left(\frac{n}{k}\right) \to \lambda$ for some constant λ we have for the kernel function $K(t) = t^\tau(-\log t)^\delta, \quad \tau, \delta \geq 0$ that*

$$\sqrt{k}\left(\hat{\gamma}_k(K) - \gamma\right) \to N\left(\lambda \frac{(\tau+1)^{\delta+1}}{(\tau-\rho+1)^{\delta+1}}, \gamma^2 \frac{\Gamma(2\delta+1)(\tau+1)^{2\delta+2}}{(2\tau+1)^{2\delta+1}(\Gamma(\delta+1))^2}\right).$$

In particular, we obtain

(i) *For the Hill Kernel*

$$\sqrt{k}\left(\hat{\gamma}_k(\mathbb{H}) - \gamma\right) \to N\left(\lambda \frac{1}{1-\rho}, \gamma^2\right).$$

(ii) *For the Power kernel*

$$\sqrt{k}\left(\hat{\gamma}_k(\mathbb{K}_\tau) - \gamma\right) \to N\left(\lambda \frac{\tau+1}{\tau-\rho+1}, \gamma^2 \frac{(\tau+1)^2}{2\tau+1}\right).$$

(iii) *For the Log kernel*

$$\sqrt{k}\left(\hat{\gamma}_k(\mathbb{L}_\delta) - \gamma\right) \to N\left(\lambda \frac{1}{(1-\rho)^{\delta+1}}, \gamma^2 \frac{\Gamma(2\delta+1)}{(\Gamma(\delta+1))^2}\right).$$

A discussion on when to choose which kernel function is a topic of its own, so we will not spend much time on it since it is not of great importance for this thesis. However, the Hill kernel always has the smallest asymptotic variance. In general, the kernel function for which the asymptotic mean squared error of the resulting γ estimator is minimal depends on the distributional parameters γ and ρ. Concerning the log and power kernel with $\delta = \tau$, we see that the log kernel tends to have a bigger variance than the power kernel, although it suffers from less bias. For a detailed discussion of the performance of γ estimators with kernel functions in the family $K(t) = t^\tau(-\log t)^\delta$ we refer to Gomes et al. (2007).

2.3 Estimation of the second order parameter

The estimation of the second order parameter in the univariate case is not of grave importance to this thesis. We will however in Chapter 4 construct estimators for the second order parameter in the bivariate extreme value framework, which are based on the same ideas as is used to construct the estimator for the second order parameter ρ. In order to construct an estimator for ρ we start with the basic building block $T_{n,k}(K)$ defined in (2.8). By making a Taylor series expansion it follows by Theorem 2.2.7 that

$$T_{n,k}^\alpha(K) \stackrel{\mathcal{D}}{=} \gamma^\alpha \mu^\alpha(K) + \alpha\gamma\mu^{\alpha-1}(K)\sigma(K)\frac{N_k(K)}{\sqrt{k}} + b\left(\frac{n}{k}\right)\alpha\gamma^{\alpha-1}\mu^{\alpha-1}(K)I_1(K,\rho)\left(1 + o_{\mathbb{P}}(1)\right),$$

where $\alpha > 0$ and $K > 0$. The basic idea is to construct a statistic which converges in probability to a function of ρ, which does not depend on the unknown parameter γ. To this end, let $K_1, \dots, K_8$ be kernel functions and define

$$\begin{aligned}
\mathbf{K}^{(1)} &:= (K_1, K_2, K_3, K_4)\,, \\
\mathbf{K}^{(2)} &:= (K_5, K_6, K_7, K_8)\,, \\
\mathbf{K}^{(1,2)} &:= \left(\mathbf{K}^{(1)}, \mathbf{K}^{(2)}\right), \\
\bar{I}_1(K_i, \rho) &:= \frac{I_1(K_i, \rho)}{\mu(K_i)}, \quad i \in \{1, \dots, 8\}\,, \\
\bar{I}_1^{(a)}(K_i, K_j, \rho) &:= \bar{I}_1^a(K_i, \rho) - \bar{I}_1^a(K_j, \rho)\,, \quad a = 1, 2, \quad i, j \in \{1, \dots, 8\}\,.
\end{aligned}$$

Using this notation, we consider the ratio of differences given by

$$\Psi_{n,k}\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_2\right) := \frac{\left(\frac{T_{n,k}(K_1)}{\mu(K_1)}\right)^{\alpha_1} - \left(\frac{T_{n,k}(K_2)}{\mu(K_2)}\right)^{\alpha_1}}{\left(\frac{T_{n,k}(K_3)}{\mu(K_3)}\right)^{\alpha_2} - \left(\frac{T_{n,k}(K_4)}{\mu(K_4)}\right)^{\alpha_2}} \tag{2.14}$$

and the function

$$\psi\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_2, \rho\right) := \gamma^{\alpha_1 - \alpha_2} \frac{\alpha_1 \bar{I}_1^{(1)}(K_1, K_2, \rho)}{\alpha_2 \bar{I}_1^{(1)}(K_3, K_4, \rho)},$$

with $\alpha_1, \alpha_2 > 0$.
If $k, n \to \infty$ such that $\frac{k}{n} \to 0$ and $\sqrt{k} b\left(\frac{n}{k}\right) \to \infty$, then

$$\frac{\left(\frac{T_{n,k}(K_1)}{\mu(K_1)}\right)^{\alpha_1} - \left(\frac{T_{n,k}(K_2)}{\mu(K_2)}\right)^{\alpha_1}}{b\left(\frac{n}{k}\right)} \xrightarrow{P} \alpha_1 \gamma^{\alpha_1 - 1} \bar{I}_1^{(1)}(K_1, K_2, \rho)$$

and

$$\frac{\left(\frac{T_{n,k}(K_3)}{\mu(K_3)}\right)^{\alpha_1} - \left(\frac{T_{n,k}(K_4)}{\mu(K_4)}\right)^{\alpha_2}}{b\left(\frac{n}{k}\right)} \xrightarrow{P} \alpha_2 \gamma^{\alpha_2 - 1} \bar{I}_1^{(1)}(K_3, K_4, \rho)\,.$$

Hence

$$\Psi_{n,k}\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_2\right) \xrightarrow{\mathbb{P}} \psi\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_2, \rho\right).$$

This statistic still depends on γ, but we can get rid of this if we consider a ratio of statistics on the form of (2.14) with appropriately chosen α parameters. So define

$$\Lambda_{n,k}\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l\right) := \frac{\Psi_{n,k}\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_1 + l\right)}{\Psi_{n,k}\left(\mathbf{K}^{(2)}, \alpha_2, \alpha_2 + l\right)}$$

and

$$\Lambda\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \rho\right) := \frac{\psi\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_1 + l, \rho\right)}{\psi\left(\mathbf{K}^{(2)}, \alpha_2, \alpha_2 + l, \rho\right)}$$

where $l > 0$. If we again assume that If $k, n \to \infty$ such that $\frac{k}{n} \to 0$ and $\sqrt{k} b\left(\frac{n}{k}\right) \to \infty$, then clearly

$$\Lambda_{n,k}\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l\right) \xrightarrow{\mathbb{P}} \Lambda\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \rho\right),$$

which does not depend on γ. If the function $\rho \mapsto \Lambda\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \rho\right)$ is bijective, then we obtain the estimator

$$\hat{\rho}\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l\right) := \Lambda^{-1}\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \Lambda_{n,k}\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l\right)\right) \tag{2.15}$$

for the second order parameter. The consistency of this estimator is estblished in Proposition 2.3.1 using a straightforward application of the continuous mapping theorem.

Proposition 2.3.1. *(Goegebeur et al., 2010) Let $X_1, \ldots, X_n$ be i.i.d. random variables according to a distribution satisfying Assumption 2.2.1. Let $K_1, \ldots, K_8$ satisfy Assumption 2.2.4, and suppose $\bar{I}_1^{(1)}(K_1, K_2)$, $\bar{I}_1^{(1)}(K_3, K_4)$, $\bar{I}_1^{(1)}(K_5, K_6)$ and $\bar{I}_1^{(1)}(K_7, K_8)$ are welldefined and nonzero. Then if $k, n \to \infty$ such that $\frac{k}{n} \to 0$ and $\sqrt{k}b\left(\frac{n}{k}\right) \to \infty$ we have $\Lambda_{n,k}\left(\boldsymbol{K}^{(1,2)}, \alpha_1, \alpha_2, l\right) \xrightarrow{\mathbb{P}} \Lambda\left(\boldsymbol{K}^{(1,2)}, \alpha_1, \alpha_2, l, \rho\right)$. Further, if Λ is bijective and Λ^{-1} is continuous then $\hat{\rho}\left(\boldsymbol{K}^{(1,2)}, \alpha_1, \alpha_2, l\right)$ is a consistent estimator for ρ.*

In order to establish asymptotic normality of the estimator of ρ, we need the following third order condition.

Assumption 2.3.2 (Third order condition)**.** *There exists a positive real parameter γ, negative real parameters ρ and β, functions b and $\tilde{b}$ with $b(t) \to 0$ and $\tilde{b}(t) \to 0$ for $t \to \infty$, both of constant sign for large values of t, such that*

$$\lim_{t\to\infty} \frac{\frac{\log U(tx) - \log U(t) - \gamma \log x}{b(t)} - \frac{x^\rho - 1}{\rho}}{\tilde{b}(t)} = \frac{1}{\beta}\left(\frac{x^{\rho+\beta} - 1}{\rho + \beta} - \frac{x^\rho - 1}{\rho}\right), \quad \forall x > 0.$$

The third order condition implies that $|\tilde{b}|$ is regularly varying of index β (de Haan and Ferreira, 2006). The third order contion is not to restrictive. Among distributions of Pareto-type that satisfy the second and third order condition are the Fréchet, the Burr, the GP distributions and the absolute T distribution. This is not a complete list of Pareto-type distributions which satisfy the second and third order condition. As examples, we show that the Burr and the absolute T distribution satisfies the third order condition.

Example 2.3.3. In order to verify that the Burr distribution satisfies the third order condition, it is a good idea to choose $b(t) = \gamma \frac{t^\rho}{1-t^\rho}$. From (2.4) and the choice of $b(t)$ it follows that

$$\frac{\log U(tx) - \log U(t) - \gamma \log x}{b(t)} - \frac{x^\rho - 1}{\rho} = \frac{\frac{\gamma t^\rho (x^\rho - 1)}{\rho} - \frac{1}{2\delta} t^{2\rho}\left(x^{2\rho} - 1\right) + O\left(t^{3\rho}\right)}{\gamma \frac{t^\rho}{1-t^\rho}} - \frac{x^\rho - 1}{\rho} \tag{2.16}$$

$$= -\frac{t^\rho\left(x^\rho - 1\right)}{\rho} + \frac{1}{2\rho} t^\rho\left(x^{2\rho} - 1\right) + O\left(t^{2\rho}\right) \tag{2.17}$$

$$= \rho t^\rho \frac{1}{\rho}\left(\frac{x^{2\rho} - 1}{2\rho} - \frac{x^\rho - 1}{\rho}\right) + O(t^{2\rho}). \tag{2.18}$$

From (2.18) we see that if we choose $\beta = \rho$ and $\tilde{b}(t) = \rho t^\rho (1 + o(1))$ then the Burr distribution satisfies the third order condition.

Example 2.3.4. To verify that the absolute T distribution satisfies the third order condition, it is a good idea to choose $b(t) = -\frac{\rho D_1 t^\rho}{1+2\left(\frac{D_2}{D_1}-\frac{1}{2}D_1\right)t^\rho}$. With this choice of $b(t)$ and (2.6) it follows that

$$\frac{\log U(xt) - \log U(t) - \gamma \log x}{b(t)} - \frac{x^\rho - 1}{\rho} = 2\left(\frac{D_2}{D_1} - \frac{1}{2}D_1\right)\frac{t^\rho\left(x^\rho - 1\right)}{\rho} \tag{2.19}$$

$$- \left(\frac{D_2}{D_1} - \frac{1}{2}D_1\right)\frac{t^\rho\left(x^{2\rho} - 1\right)}{\rho} + O(t^{2\rho}) \tag{2.20}$$

$$= -2\rho\left(\frac{D_2}{D_1} - \frac{1}{2}D_1\right)t^\rho \frac{1}{\rho}\left(\frac{x^{2\rho}-1}{2\rho} - \frac{\left(x^\rho - 1\right)}{\rho}\right) + O(t^{2\rho}). \tag{2.21}$$

From this we see that if we choose $\beta = \rho$ and $\tilde{b}(t)$ on the form $\tilde{b}(t) = -2\rho\left(\frac{D_2}{D_1} - \frac{1}{2}D_1\right)t^\rho(1 + o(1))$, then the absolute T distribution satisfies the third order condition.

We also have to add an extra condition on the kernel function.

Assumption 2.3.5. *Let K be a fuction defined on $(0,1)$ such that Assumption 2.2.4 is satisfied, and the following extra condition.*

(vi) $\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right)\left(\frac{j}{k+1}\right)^{-\rho} = I_1(K,\rho) + o\left(\frac{1}{\sqrt{k}}\right), \quad k \to \infty.$

Lemma 2.3.6. *The kernel function considered in Example 2.2.5 given by $K(t) := t^\tau(-\log t)^\delta$ also satisfies Assumption 2.3.5*

This result can easily be obtained from the proof of Assumption 2.2.4 (iii), and is hence omitted.
Similar to the procedure in Theorem 2.2.7 we can make an asymptotic expansion of the statistic in (2.8) using the third order condition.

Theorem 2.3.7. *(Goegebeur et al., 2010) Let $X_1, \ldots, X_n$ be i.i.d. random variables according to a distribution satisfying Assumption 2.3.2. If Assumption 2.3.5 holds, then for $k, n \to \infty$ such that $\frac{k}{n} \to 0$ we have*

$$\begin{aligned} T_{n,k}(K) \stackrel{\mathcal{D}}{=} &\gamma\mu(K) + \gamma\sigma(K)\frac{N_k(K)}{\sqrt{k}} + b\left(Y_{n-k,n}\right)I_1(K,\rho) + b\left(Y_{n-k,n}\right)\tilde{\sigma}(K,\rho)\frac{P_k(K,\rho)}{\sqrt{k}} \\ &+ b\left(Y_{n-k,n}\right)\tilde{b}\left(Y_{n-k,n}\right)I_2(K,\rho,\beta)\left(1 + o_{\mathbb{P}}(1)\right) + b\left(Y_{n-k,n}\right)O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right), \end{aligned}$$

where $N_k(K)$ and $P_k(K,\rho)$ are asymptotic standard normally distributed random variables.

We will not give a proof of this result, but the line of proof follows the same as the proof of Theorem 2.2.7. The result in Theorem 2.3.7 can be used to obtain the asymptotic expansion

$$\begin{aligned} T^\alpha_{n,k}(K) \stackrel{\mathcal{D}}{=} &\gamma^\alpha\mu^\alpha(K) + \alpha\gamma^\alpha\mu^{\alpha-1}(K)\sigma(K)\frac{N_k(K)}{\sqrt{k}} + b\left(Y_{n-k,n}\right)\alpha\gamma^{\alpha-1}\mu^{\alpha-1}(K)I_1(K,\rho) \\ &+ b\left(Y_{n-k,n}\right)\tilde{b}\left(Y_{n-k,n}\right)\alpha\gamma^{\alpha-1}\mu^{\alpha-1}(K)I_2(K,\rho,\beta)\left(1 + o_{\mathbb{P}}(1)\right) \\ &+ b^2\left(Y_{n-k,n}\right)\frac{\alpha(\alpha-1)}{2}\gamma^{\alpha-2}\mu^{\alpha-2}(K)I_1^2(K,\rho)\left(1 + o_{\mathbb{P}}(1)\right) + b\left(Y_{n-k,n}\right)O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right) \end{aligned}$$

Before we can present the limiting distribution of the ρ estimator presented in (2.15) we need to introduce the following notation, with $i, j \in \{1, \dots, 8\}$.

$$\bar{I}_2(K, \rho, \beta) := \frac{I_2(K, \rho, \beta)}{\mu(K)},$$
$$\bar{I}_2(K_i, K_j, \rho, \beta) := \frac{I_2(K_i, \rho, \beta)}{\mu(K)} - \frac{I_2(K_j, \rho, \beta)}{\mu(K)},$$
$$\bar{\sigma}(K) := \frac{\sigma(K)}{\mu(K)},$$
$$N_k(K_i, K_j) := \bar{\sigma}(K_i) N_k(K_i) - \bar{\sigma}(K_j) N_k(K_j),$$
$$N_k\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_2, \gamma, \rho\right) := \frac{\alpha_1 \gamma_1^{\alpha} N_k(K_1, K_2) - \psi\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_2, \rho\right) \alpha_2 \gamma^{\alpha_2} N_k(K_3, K_4)}{\alpha_2 \gamma^{\alpha_2 - 1} \bar{I}_1^{(1)}(K_3, K_4, \rho)},$$
$$c_1\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_2, \gamma, \rho, \beta\right) := \frac{\alpha_1 \gamma^{\alpha_1 - 1} \bar{I}_2(K_1, K_2, \rho, \beta) - \psi\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_2, \rho\right) \alpha_2 \gamma^{\alpha_2 - 1} \bar{I}_2(K_3, K_4, \rho, \beta)}{\alpha_2 \gamma^{\alpha_2 - 1} \bar{I}_1^{(1)}(K_3, K_4, \rho)},$$
$$c_2\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_2, \gamma, \rho\right)$$
$$:= \frac{\alpha_1(\alpha_1 - 1)\gamma^{\alpha_1 - 2} \bar{I}_1^{(2)}(K_1, K_2, \rho) - \psi\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_2, \rho\right) \alpha_2(\alpha_2 - 1)\gamma^{\alpha_2 - 2} \bar{I}_1^{(2)}(K_3, K_4, \rho)}{\alpha_2 \gamma^{\alpha_2 - 1} \bar{I}_1^{(1)}(K_3, K_4, \rho)},$$
$$N_k\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \gamma, \rho\right)$$
$$:= \frac{N_k\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_1 + l, \gamma, \rho\right) - \Lambda\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \gamma, \rho\right) N_k\left(\mathbf{K}^{(2)}, \alpha_2, \alpha_2 + l, \gamma, \rho\right)}{\psi\left(\mathbf{K}^{(2)}, \alpha_2, \alpha_2 + l, \rho\right)},$$
$$c_1\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \gamma, \rho, \beta\right)$$
$$:= \frac{c_1\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_1 + l, \gamma, \rho, \beta\right) - \Lambda\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \gamma, \rho\right) c_1\left(\mathbf{K}^{(2)}, \alpha_2, \alpha_2 + l, \gamma, \rho, \beta\right)}{\psi\left(\mathbf{K}^{(2)}, \alpha_2, \alpha_2 + l, \rho\right)},$$
$$c_2\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \gamma, \rho\right)$$
$$:= \frac{c_2\left(\mathbf{K}^{(1)}, \alpha_1, \alpha_1 + l, \gamma, \rho\right) - \Lambda\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \gamma, \rho\right) c_2\left(\mathbf{K}^{(2)}, \alpha_2, \alpha_2 + l, \gamma, \rho\right)}{\psi\left(\mathbf{K}^{(2)}, \alpha_2, \alpha_2 + l, \rho\right)},$$
$$v^2\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \gamma, \rho\right) := \operatorname{Var}\left(N_k\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \gamma, \rho\right)\right).$$

With this notation we can obtain a result giving the asymptotic normality of our ρ estimator.

Proposition 2.3.8. *(Goegebeur et al., 2010) Let $X_1, \dots, X_n$ be i.i.d. random variables according to a distribution satisfying Assumption 2.3.2. If the kernel functions $K_1, \dots, K_8$ satisfy Assumption 2.3.5 and are such that $\bar{I}_1^{(1)}(K_1, K_2, \rho)$, $\bar{I}_1^{(1)}(K_3, K_4, \rho)$, $\bar{I}_1^{(1)}(K_5, K_6, \rho)$ and $\bar{I}_1^{(1)}(K_7, K_8, \rho)$ are well defined and nonzero, then for $k, n \to \infty$ such that $\frac{k}{n} \to 0$, $\sqrt{k} b\left(\frac{n}{k}\right) \to \infty$, $\sqrt{k} b\left(\frac{n}{k}\right) \tilde{b}\left(\frac{n}{k}\right) \to \lambda_1$ and $\sqrt{k} b^2\left(\frac{n}{k}\right) \to \lambda_2$ we have*

$$\sqrt{k} b\left(\frac{n}{k}\right)\left[\Lambda_{n,k}\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l\right) - \Lambda\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \rho\right)\right]$$
$$\xrightarrow{\mathcal{D}} N\left(\lambda_1 c_1\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \gamma, \rho, \beta\right) + \lambda_2 c_2\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \gamma, \rho\right), v^2\left(\mathbf{K}^{(1,2)}, \alpha_1, \alpha_2, l, \gamma, \rho\right)\right).$$

2.4 Appendix

2.4.1 Proof of Lemma 2.2.6

i)

Since $K(t) = \frac{1}{t}t^{\tau+1}(-\log t)^{\delta}$ it follows that

$$\int_0^t u(v)dv = t^{\tau+1}(-\log t)^{\delta},$$

and hence

$$u(v) = (\tau+1)v^{\tau}(-\log v)^{\delta} - \delta v^{\tau}(-\log v)^{\delta-1}.$$

Now

$$\begin{aligned}\left|(k+1)\int_{\frac{j-1}{k+1}}^{\frac{j}{k+1}} u(t)dt\right| \leq & (k+1)(\tau+1)\int_{\frac{j-1}{k+1}}^{\frac{j}{k+1}} t^{\tau}(-\log t)^{\delta}dt + (k+1)\delta\int_{\frac{j-1}{k+1}}^{\frac{j}{k+1}} t^{\tau}(-\log t)^{\delta-1}dt \\ \leq & \frac{(k+1)}{j}(\tau+1)\int_0^{\frac{j}{k+1}}(-\log t)^{\delta}dt + (k+1)\delta\int_{\frac{j-1}{k+1}}^{\frac{j}{k+1}}(-\log t)^{\delta-1}dt\end{aligned}$$

We distinguish between the two cases $\delta > 1$ and $\delta \leq 1$. We start with the case $\delta > 1$. So

$$\begin{aligned}\left|(k+1)\int_{\frac{j-1}{k+1}}^{\frac{j}{k+1}} u(t)dt\right| \leq & \frac{(k+1)}{j}(\tau+1)\int_0^{\frac{j}{k+1}}(-\log t)^{\delta}dt + \frac{(k+1)}{j}\delta\int_0^{\frac{j}{k+1}}(-\log t)^{\delta-1}dt \\ =: & f\left(\frac{j}{k+1}\right).\end{aligned}$$

Next we show that $\int_0^1 f(x)dx < \infty$ for the case $\delta > 1$.

$$\begin{aligned}\int_0^1 f(x)dx &= (\tau+1)\int_0^1 \frac{1}{x}\int_0^x(-\log t)^{\delta}dtdx + \delta\int_0^1\frac{1}{x}\int_0^x(-\log t)^{\delta-1}dtdx \\ &= (\tau+1)\int_0^1(-\log t)^{\delta}\int_t^1\frac{1}{x}dxdt + \delta\int_0^1(-\log t)^{\delta-1}\int_t^1\frac{1}{x}dxdt \\ &= (\tau+1)\Gamma(\delta+2) + \delta\Gamma(\delta+1) \\ &< \infty.\end{aligned}$$

The case $\delta \leq 1$ follows since

$$\begin{aligned}\left|(k+1)\int_{\frac{j-1}{k+1}}^{\frac{j}{k+1}} u(t)dt\right| \leq & \frac{(k+1)}{j}(\tau+1)\int_0^{\frac{j}{k+1}}(-\log t)^{\delta}dt + \delta\left(-\log\left(\frac{j}{k+1}\right)\right)^{\delta-1} \\ =: & f\left(\frac{j}{k+1}\right).\end{aligned}$$

That $\int_0^1 f(x)dx < \infty$ for the case $\delta \leq 1$ follows by an argument similar to the case $\delta > 1$.

ii)

The second part is easily verified using the following argument.

$$\begin{aligned}\sigma^2(K) &= \int_0^1 K^2(u)du \\ &\leq \Gamma(2\delta+1) \\ &< \infty.\end{aligned}$$

iii)

For the third part we start by considering

$$\begin{aligned}I :&= \left|\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) - \int_0^1 K(u)du\right| \\ &\leq \left|\frac{1}{k+1}\sum_{j=1}^{k}\left(\frac{j}{k+1}\right)^{\tau}\left(-\log\left(\frac{j}{k+1}\right)\right)^{\delta} - \int_{\frac{1}{k+1}}^{1} u^{\tau}(-\log u)^{\delta}du\right| \\ &\quad + \int_0^{\frac{1}{k+1}} u^{\tau}(-\log u)^{\delta}du + O\left(\frac{1}{k}\right) \\ &=: I_1 + I_2 + O\left(\frac{1}{k}\right).\end{aligned}$$

Concerning I_1 we can use the mean value theorem, in which case it follows that

$$I_1 = \left|\frac{1}{k+1}\sum_{j=1}^{k}\left[\left(\frac{j}{k+1}\right)^{\tau}\left(-\log\left(\frac{j}{k+1}\right)\right)^{\delta} - \tilde{u}_j^{\tau}(-\log\tilde{u}_j)^{\delta}\right]\right|,$$

where $\tilde{u}_j$ is a value between $\frac{j}{k+1}$ and $\frac{j+1}{k+1}$. From this we find that

$$\begin{aligned}I_1 &\leq \left|\frac{1}{k+1}\sum_{j=1}^{k}\left(\frac{j}{k+1}\right)^{\tau}\left(\left(-\log\left(\frac{j}{k+1}\right)\right)^{\delta} - (-\log\tilde{u}_j)^{\delta}\right)\right| \\ &\quad + \left|\frac{1}{k+1}\sum_{j=1}^{k}\left(\tilde{u}_j^{\tau} - \left(\frac{j}{k+1}\right)^{\tau}\right)(-\log\tilde{u}_j)^{\delta}\right| \\ &=: I_{11} + I_{12}.\end{aligned}$$

The term I_{11} is easily dealt with if we replace $\tilde{u}_j$ with the value of it that maximizes I_{11} and then telescope out the terms in the sum, in which case we get

$$\begin{aligned}I_{11} &\leq \left|\frac{1}{k+1}\sum_{j=1}^{k}\left(\left(-\log\left(\frac{j}{k+1}\right)\right)^{\delta} - \left(-\log\left(\frac{j+1}{k+1}\right)\right)^{\delta}\right)\right| \\ &= \frac{(\log(k+1))^{\delta}}{k+1}.\end{aligned}$$

A similar argument shows that $I_{12} = O\left(\frac{(\log(k+1))^\delta}{k+1}\right)$.
Concerning the term I_2 we find that

$$
\begin{aligned}
I_2 &\leq \int_{\log(k+1)}^{\infty} z^\delta e^{-z} dz \\
&= \frac{(\log(k+1))^\delta}{k+1} + \delta \int_{\log(k+1)}^{\infty} z^{\delta-1} e^{-z} dz \\
&= \frac{(\log(k+1))^\delta}{k+1} \left(1 + \frac{k+1}{(\log(k+1))^\delta} \delta \int_{\log(k+1)}^{\infty} z^{\delta-1} e^{-z} dz\right).
\end{aligned}
$$

If we can show that $\frac{k+1}{(\log(k+1))^\delta} \delta \int_{\log(k+1)}^{\infty} z^{\delta-1} e^{-z} dz \to 0$ as $k \to \infty$ then $I_2 = O\left(\frac{(\log(k))^\delta}{k}\right)$.
Using l'Hôpital's rule and Leibniz's rule it follows that

$$
\begin{aligned}
\lim_{x\to\infty} \frac{\delta \int_{\log(x)}^{\infty} z^{\delta-1} e^{-z} dz}{\frac{(\log(x))^\delta}{x}} &= \lim_{x\to\infty} \frac{-\delta(\log(x))^{\delta-1} e^{-\log(x)} \frac{1}{x}}{\left(\frac{\delta(\log(x))^{\delta-1} - (\log(x))^\delta}{x^2}\right)} \\
&= \lim_{x\to\infty} \frac{-\delta}{\delta - (\log(x))} \\
&= 0
\end{aligned}
$$

iv)

The fourth condition is trivially satisfied since

$$
\max_{j\in 1,\ldots,k} \left| K\left(\frac{j}{k+1}\right)\right| \leq (\log(k+1))^\delta = o\left(\sqrt{k}\right)
$$

v)

This condition is also trivially satisfied since

$$
\begin{aligned}
\int_0^1 |K(u)| u^{|\rho|-1-\epsilon} du &= \int_0^1 u^{\tau+|\rho|-1-\epsilon} (-\log u)^\delta du \\
&= \frac{\Gamma(\delta+1)}{(\tau+|\rho|-\epsilon)^{\delta+1}} \\
&< \infty,
\end{aligned}
$$

assuming $\epsilon < \tau + |\rho|$.

□

2.4.2 Lemma's needed in the proof of Theorem 2.2.7

Lemma 2.4.1. *(Chernoff et al., 1967; Gomes et al., 2007) Let*

$$
Z_k := \frac{1}{k} \sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) E_j, \tag{2.22}
$$

where E_i are standard exponential random variables and $K(u)$, $0 < u < 1$ is a kernel fuction. Furthermore, let

$$v_k = \sqrt{\frac{1}{k}\sum_{j=1}^{k} K^2\left(\frac{j}{k+1}\right)}. \tag{2.23}$$

Then

$$\frac{\sqrt{k}\left(Z_k - \frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right)\right)}{v_k} \xrightarrow{D} N(0,1) \Leftrightarrow \max_{1\le j\le k}\left|K\left(\frac{j}{k+1}\right)\right| = o\left(\sqrt{k}v_k\right) \tag{2.24}$$

as $k \to \infty$. If further we have

$$\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) = \mu(K) + o\left(\frac{1}{\sqrt{k}}\right), \quad v_k \to \sigma(K) > 0, \tag{2.25}$$

$\mu(K)$ and $\sigma(K)$ finite, and

$$\max_{1\le j\le k}\left|K\left(\frac{j}{k+1}\right)\right| = o\left(\sqrt{k}\right), \quad \text{as } k \to \infty, \tag{2.26}$$

then

$$\frac{\sqrt{k}\left(Z_k - \mu(K)\right)}{\sigma(K)} \xrightarrow{D} N(0,1) \tag{2.27}$$

as $k \to \infty$

Lemma 2.4.2. *(Goegebeur et al., 2010) Denote by $E_1, \dots, E_k$ standard exponential random variables and by $U_{1,k} \le \cdots \le U_{k,k}$ the order statistics of a random sample of size k from $U(0,1)$. Assume that $\int_0^1 |K(u)|\, du < \infty$ in case $|\rho| \ge 1$ and that $\int_0^1 |K(u)|\, u^{|\rho|-1-\epsilon} du < \infty$ for some $\epsilon > 0$ in case $|\rho| < 1$. Then*

$$S_k = \frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right)\left(U_{j,k}^{-\rho} - \left(\frac{j}{k+1}\right)^{-\rho}\right) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right) \tag{2.28}$$

for $k \to \infty$.

Lemma 2.4.3. *(de Haan and Ferreira, 2006) Suppose for a measurable function f and a positive function b we have*

$$\lim_{t\to\infty} \frac{f(tx) - f(t)}{b(t)} = \frac{x^\gamma - 1}{\gamma}$$

for all $x > 0$, where γ is a real parameter. Then for all $\epsilon, \delta > 0$ there is a $t_0 = t_0(\epsilon, \delta)$ such that for t, $tx \ge t_0$,

$$\left|\frac{f(tx) - f(t)}{b_0(t)} - \frac{x^\gamma - 1}{\gamma}\right| \le \epsilon x^\gamma \max\left(x^\delta, x^{-\delta}\right),$$

where

$$b_0(t) := \begin{cases} \gamma f(t), & \gamma > 0, \\ -\gamma(f(\infty) - f(t)), & \gamma < 0, \\ f(t) - t^{-1}\int_0^1 f(s)ds, & \gamma = 0 \end{cases} \tag{2.29}$$

Chapter 3

Multivariate extreme value theory

In this chapter we introduce the basic limit laws in multivariate extreme value theory. After a transformation of the marginal distribution functions to standard Fréchet margins, we discuss the dependence structure between the variables. This discussion starts with the exponent and spectral measure, before we turn our attention to the max domain of attraction in the multivariate framework and asymptotic independence. This is followed by an introduction to several other dependence measures. The measures we consider are the Pickands dependence function and the pair of dependence measures χ and $\bar{\chi}$. We explain the relation between all these dependence measures and discuss ways of getting from one to the other. Finally we introduce the model of Ledford and Tawn (1997) and make the connection between the coefficient of tail dependence η and the other dependence measures discussed previously.

3.1 Limit laws

The results we present in this section will be based on two-dimensional spaces. Generalizations to higher dimensional spaces are obvious, but require heavier notation. Suppose $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. random vectors with distribution function F_{XY}. We define the maximum of a set of vectors of this form as

$$M_n := (\max(X_1, \ldots, X_n), \max(Y_1, \ldots, Y_n)),$$

which is simply the vector of componentwise maxima. We start by deriving an important theorem, which is the foundation of our description of the asymptotic distributions that can occur for an appropriately normalized maximum of the form of M_n. Suppose there exists sequences of constants $(b_n)_{n=1}^{\infty}$, $(d_n)_{n=1}^{\infty}$ and sequences of positive constants $(a_n)_{n=1}^{\infty}$, $(c_n)_{n=1}^{\infty}$ and a distribution function G with nondegenerate marginals such that

$$\lim_{n\to\infty} P\left(\frac{\max(X_1, \ldots, X_n) - b_n}{a_n} \leq x, \frac{\max(Y_1, \ldots, Y_n) - d_n}{c_n} \leq y\right) = G(x, y) \tag{3.1}$$

for all continuity points (x, y) of G. Any limit distribution function G in (3.1) with nondegenerate marginals is called a multivariate extreme value distribution. It follows that

$$\lim_{n\to\infty} P\left(\frac{\max(X_1, \ldots, X_n) - b_n}{a_n} \leq x\right) = G(x, \infty)$$

and

$$\lim_{n\to\infty} P\left(\frac{\max\left(Y_1,\ldots,Y_n\right)-d_n}{b_n}\leq y\right)=G(\infty,y),$$

since (3.1) implies convergence of the marginal distributions. According to Theorem 1.1.2 we can chose the constants a_n, b_n, c_n and d_n such that for some $\gamma_1, \gamma_2 \in \mathbb{R}$, we have

$$G(x,\infty)=\exp\left(-\left(1+\gamma_1 x\right)^{-\frac{1}{\gamma_1}}\right) \tag{3.2}$$

and

$$G(\infty,y)=\exp\left(-\left(1+\gamma_2 y\right)^{-\frac{1}{\gamma_2}}\right). \tag{3.3}$$

It is relevant to note that G is continuous, since the two marginal distributions of G are continuous.
If we let F_X and F_Y be the two marginal distributions of F_{XY} and U_X and U_Y be the two corresponding tail quantile functions, then according to Theorem 1.1.2 there are positive functions $a_X(t)$ and $a_Y(t)$, such that

$$\lim_{t\to\infty}\frac{U_X(tx)-U_X(t)}{a_X(t)}=\frac{x^{\gamma_1}-1}{\gamma_1},\quad \forall x>0$$

and

$$\lim_{t\to\infty}\frac{U_Y(tx)-U_Y(t)}{a_Y(t)}=\frac{x^{\gamma_2}-1}{\gamma_2},\quad \forall x>0.$$

Hence

$$\lim_{n\to\infty}\frac{U_X(nx)-b_n}{a_n}=\frac{x^{\gamma_1}-1}{\gamma_1}$$

and

$$\lim_{n\to\infty}\frac{U_Y(nx)-d_n}{c_n}=\frac{x^{\gamma_2}-1}{\gamma_2},$$

if we choose the constants a_n, b_n, c_n and d_n according to Theorem 1.1.2.
We easily see that (3.1) can be written as

$$G(x,y)=\lim_{n\to\infty}F_{XY}^n\left(a_n x+b_n, c_n y+d_n\right).$$

If $x_n \to u$ and $y_n \to v$ then by the continuity of G and the monotonicity of F_{XY} we have that

$$G(u,v)=\lim_{n\to\infty}F_{XY}^n\left(a_n x_n+b_n, c_n y_n+d_n\right).$$

Applying this result with

$$x_n:=\frac{U_X(nx)-b_n}{a_n},\quad x>0$$

and

$$y_n:=\frac{U_Y(ny)-d_n}{c_n},\quad y>0$$

gives

$$G\left(\frac{x^{\gamma_1}-1}{\gamma_1},\frac{y^{\gamma_2}-1}{\gamma_2}\right)=\lim_{n\to\infty}F_{XY}^n\left(U_1(nx),U_2(ny)\right).$$

These results establish the following theorem.

Theorem 3.1.1. *(de Haan and Ferreira, 2006) Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. random vectors with distribution function F_{XY}. Suppose there exists sequences of real constants $(b_n)_{n=1}^{\infty}$, $(d_n)_{n=1}^{\infty}$ and positive real constants $(a_n)_{n=1}^{\infty}$ and $(c_n)_{n=1}^{\infty}$ such that*

$$\lim_{n\to\infty} F_{XY}^n (a_n x + b_n, c_n y + d_n) = G(x, y)$$

for all (x, y) of G, and the marginals of G are standardized as in (3.2) and (3.3). Then with $F_X(x) := F_{XY}(x, \infty)$, $F_Y(y) := F_{XY}(\infty, y)$ and U_X and U_Y the two corresponding tail quantile functions, we have that

$$\lim_{n\to\infty} F_{XY}^n (U_X(nx), U_Y(ny)) = G_0(x, y) \tag{3.4}$$

for all $x, y > 0$, where

$$G_0(x, y) := G\left(\frac{x^{\gamma_1} - 1}{\gamma_1}, \frac{y^{\gamma_2} - 1}{\gamma_2}\right)$$

and γ_1, γ_2 are the marginal extreme value indices from (3.2) and (3.3).

Remark 3.1.2. The multivariate extreme value distribution function $G\left(\frac{x^{\gamma_1}-1}{\gamma_1}, \frac{y^{\gamma_2}-1}{\gamma_2}\right)$ has marginal distributions which are standard Fréchet, i.e. $F_Z(z) = \exp\left(-\frac{1}{z}\right), \quad z > 0$. This fact simplifies things, because now we only have to discuss the dependence structure between the two variables.

The following Corollary is obtained from Theorem 3.1.1, which we state without proof. For details we refer to de Haan and Ferreira (2006), Corollary 6.1.3 and Corollary 6.1.4

Corollary 3.1.3. *(de Haan and Ferreira, 2006) Under the conditions of Theorem 3.1.1, we have for any (x, y) for which $0 < G_0(x, y) < 1$, that*

$$\lim_{n\to\infty} n \{1 - F : XY (U_X(nx), U_Y(ny))\} = -\log G_0(x, y) \tag{3.5}$$

and

$$\lim_{t\to\infty} t \{1 - F_{XY} (U_X(tx), U_Y(ty))\} = -\log G_0(x, y), \tag{3.6}$$

where t runs through the real numbers.

3.2 The exponent measure and the spectral measure

From Corollary 3.1.3 we can obtain the following usefull theorem.

Theorem 3.2.1. *(de Haan and Ferreira, 2006) Let F_{XY} and G_0 be distribution functions where for $x, y > 0$ with $0 < G_0(x, y) < 1$ we have that*

$$\lim_{n\to\infty} n \{1 - F_{XY} (U_X(nx), U_Y(ny))\} = -\log G_0(x, y),$$

where U_X and U_Y are the tail quantile functions of the marginals of F_{XY}. Then there are set functions $\nu, \nu_1, \nu_2, \ldots$ defined for all Borel sets $A \subset \mathbb{R}_+^2$ with

$$\inf_{x,y\in A} \max(x, y) > 0$$

such that

(i)

$$\nu_n \left\{(s,t) \in \mathbb{R}_+^2 : s > x \ \textit{or}\ t > y\right\} = n\left\{1 - F_{XY}\left(U_X(nx), U_Y(ny)\right)\right\}, \tag{3.7}$$

$$\nu \left\{(s,t) \in \mathbb{R}_+^2 : s > x \ \textit{or}\ t > y\right\} = -\log G_0(x,y). \tag{3.8}$$

(ii) *For all $a > 0$ the set functions $\nu, \nu_1, \nu_2, \ldots$ are finite measures on $\mathbb{R}_+^2 \backslash [0,a]^2$.*

(iii) *For each Borel set $A \subset \mathbb{R}_+^2$ with $\inf_{x,y \in A} \max(x,y) > 0$ and $\nu(\partial A) = 0$,*

$$\lim_{n\to\infty} \nu_n(A) = \nu(A). \tag{3.9}$$

Definition 3.2.2. The measure ν from (3.8) is called the exponent measure of the extreme value distribution G_0, since

$$G_0(x,y) = \exp\left(-\nu\left(A_{x,y}\right)\right)$$

with

$$A_{x,y} := \left\{(s,t) \in \mathbb{R}_+^2 : s > x \text{ or } t > y\right\}.$$

In the following we let $\nu(x,y) := \nu\left(A_{x,y}\right)$

An important property of the exponent measure, which will be needed later in this chapter, is that it is homogeneous of order -1, as given in Theorem 3.2.3.

Theorem 3.2.3. *(de Haan and Ferreira, 2006) For any Borel set $A \subset \mathbb{R}_+^2$ with $\inf_{(x,y)\in A} \max(x,y) > 0$ and $\nu(\partial A) = 0$, and any $a > 0$,*

$$\nu(aA) = a^{-1}\nu(A),$$

where aA is the set obtained by multiplying all elements of A by a.

From the exponent measure we can also obtain the spectral measure. The spectral measure arises when we make a one-to-one transformation $\mathbb{R}_+^2 \backslash \{(0,0)\} \to (0,\infty) \times [0,c]$ for some $c > 0$,

$$\begin{cases} r = r(x,y), \\ d = d(x,y), \end{cases}$$

with the property that for all $a, x, y > 0$, we have

$$\begin{cases} r(ax,ay) = ar(x,y), \\ d(ax,ay) = d(x,y). \end{cases}$$

We can think of r as a radius and d as an angle or a direction. In this thesis we will only consider the transformation

$$\begin{cases} r(x,y) = x + y, \\ d(x,y) = \frac{x}{x+y}, \end{cases}$$

in which case the following theorem can be shown to hold.

Theorem 3.2.4. *(de Haan and Ferreira, 2006) For each limit distribution G from* (3.1), (3.2) *and* (3.3) *there exist a probability distribution (denoted by the distribution function H) concentrated on* $[0,1]$ *with mean* $\frac{1}{2}$ *such that for* $x, y > 0$,

$$\begin{aligned} G\left(\frac{x^{\gamma_1}-1}{\gamma_1}, \frac{y^{\gamma_2}-1}{\gamma_2}\right) &= G_0(x,y) \\ &= \exp\left(-2\int_0^1 \left(\frac{\omega}{x} \vee \frac{1-\omega}{y}\right) dH(\omega)\right), \end{aligned} \tag{3.10}$$

where $\frac{\omega}{x} \vee \frac{1-\omega}{y} := \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right)$.

From (3.10) we see that the limit distributions in (3.1) are characterized solely by the spectral measure H and the marginal extreme value indices. Many more transformations than the one we considered can be chosen in order to construct a spectral measure. In fact there are endless possibilities. The transformation to choose depends on the situation at hand, and in a sense they are all equivalent, since one can be transformed into the other.
From (3.8) and (3.10) we see that the connection between the exponent measure and the spectral measure is given by

$$\nu(x,y) = 2\int_0^1 \left(\frac{\omega}{x} \vee \frac{1-\omega}{y}\right) dH(\omega).$$

However, it is not always obvious how to get from one measure to the other using this relation. In case this is not obvious, and G_0 is absolutely continuous, we can use a method discovered by Coles and Tawn (1991), to compute the spectral density from the exponent measure. In the bivariate case, the point masses of H on 0 and 1 are

$$H(\{0\}) = -\frac{1}{2}\lim_{x\to 0}\frac{\partial \nu}{\partial y}(x,y), \tag{3.11}$$

$$H(\{1\}) = -\frac{1}{2}\lim_{y\to 0}\frac{\partial \nu}{\partial x}(x,y). \tag{3.12}$$

and the density for $0 < \omega < 1$ is given by

$$h(\omega) = -\frac{1}{2}\frac{\partial^2 \nu(x,y)}{\partial x \partial y}\bigg|_{(\omega,1-\omega)}. \tag{3.13}$$

Next we will consider some examples of spectral and exponent measures.

Example 3.2.5. We start by considering two important special cases of H. The first is the distribution function which places a point mass of 1 on $\omega = \frac{1}{2}$. In this case we obtain

$$G_0(x,y) = \exp\left(-\max\left(x^{-1}, y^{-1}\right)\right), \quad x, y > 0,$$

which corresponds to complete dependence between the two variables. Here G_0 is not absolutely continuous, so the method discussed above does not apply. The second case is the distribution function which places point mass of $\frac{1}{2}$ on both $\omega = 0$ and $\omega = 1$. In this case it follows that

$$G_0(x,y) = \exp\left(-\left(x^{-1} + y^{-1}\right)\right), \quad x, y > 0,$$

which corresponds to independence between the two variables. Here G_0 is absolutely continuous, though with a spectral measure putting masses of $\frac{1}{2}$ at 0 and 1.

■

Example 3.2.6. The logistic model (Gumbel, 1960a,b) given by

$$\nu(x,y) = \left(x^{-\frac{1}{\alpha}} + y^{-\frac{1}{\alpha}}\right)^{\alpha}, \quad x,y>0, \quad 0<\alpha<1,$$

is the oldest parametric family of bivariate extreme value dependence structures. It is a versatile model which covers all levels of dependence from independent variables to completely dependent variables. We see that for $\alpha \to 0$ we get

$$\nu(x,y) = \max\left(x^{-1}, y^{-1}\right)$$

and for $\alpha \to 1$ it follows that

$$\nu(x,y) = x^{-1} + y^{-1},$$

which corresponds to complete dependence and independence between the variables, respectively. The logistic model does however not allow for asymmetry in the dependence structure, as the variables are exchangeable.
From the exponent measure we can compute the point mass of H at 0

$$H(\{0\}) = \frac{1}{2}\lim_{x\to 0} y^{-\frac{1}{\alpha}-1}\left(x^{-\frac{1}{\alpha}} + y^{-\frac{1}{\alpha}}\right)^{\alpha-1} = 0,$$

using (3.11). Because of symmetry the point mass of H at 1 is also 0. The spectral density on $(0,1)$ can be found using (3.13). We start by finding

$$\frac{\partial^2 \nu(x,y)}{\partial x \partial y} = -\frac{1-\alpha}{\alpha} x^{-\frac{1}{\alpha}-1} y^{-\frac{1}{\alpha}-1}\left(x^{-\frac{1}{\alpha}} + y^{-\frac{1}{\alpha}}\right)^{\alpha-2}.$$

From this we obtain the spectral density on $(0,1)$

$$h(\omega) = \frac{1}{2}\frac{1-\alpha}{\alpha}\omega^{-\frac{1}{\alpha}-1}(1-\omega)^{-\frac{1}{\alpha}-1}\left(\omega^{-\frac{1}{\alpha}} + (1-\omega)^{-\frac{1}{\alpha}}\right)^{\alpha-2}.$$

■

3.3 Domain of attraction and asymptotic independence

In order to discuss the domain of attraction in the multivariate case we first need to introduce the concept of max stability.

Definition 3.3.1. If there exists sequences of constants $(b_n)_{n=1}^{\infty}$, $(d_n)_{n=1}^{\infty}$ and sequences of positive constants $(a_n)_{n=1}^{\infty}$ and $(c_n)_{n=1}^{\infty}$ such that

$$G^n(a_n x + b_n, c_n y + d_n) = G(x,y), \quad \forall x,y \in \mathbb{R}, \quad \forall n \in \mathbb{N}, \tag{3.14}$$

for some distribution function G. Then G belongs to the class of max stable distributions.

With this definition we are now able to discuss the bivariate max domain of attraction.

Definition 3.3.2. Let $G : \mathbb{R}^2 \to \mathbb{R}_+$ be a max stable distribution function. A distribution function F_{XY} is said to be in the max domain of attraction of G if there exists sequences of constants $(b_n)_{n=1}^{\infty}$, $(d_n)_{n=1}^{\infty}$ and sequences of positive constants $(a_n)_{n=1}^{\infty}$ and $(c_n)_{n=1}^{\infty}$ such that

$$\lim_{n\to\infty} F_{XY}^n\left(a_n x + b_n, c_n y + d_n\right) = G(x,y) \tag{3.15}$$

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

Our next proposition shows that the class of max stable distributions and the class of extreme value distributions coincide.

Proposition 3.3.3. *A distribution function G is max stable if and only if it is an extreme value distribution.*

Proof. Assume G is a max stable distribution. Then by Definition 3.3.1 there exists sequences of constants $(b_n)_{n=1}^{\infty}$, $(d_n)_{n=1}^{\infty}$ and sequences of positive constants $(a_n)_{n=1}^{\infty}$ and $(c_n)_{n=1}^{\infty}$ such that

$$G^n(a_n x + b_n, c_n y + d_n) = G(x, y), \quad \forall x, y \in \mathbb{R}, \quad \forall n \in \mathbb{N}.$$

Since

$$\lim_{n\to\infty} G^n(a_n x + b_n, c_n y + d_n) = G(x, y), \quad \forall x, y \in \mathbb{R},$$

it follows by Theorem 3.1.1 that G is an extreme value distribution.
Now, assume that G is an extreme value distribution. We can without loss of generality assume that G is on the same form as G_0 defined in Theorem 3.1.1. By Definition 3.2.2 and Theorem 3.2.3, it follows that

$$\begin{aligned} G^n(nx, ny) &= \exp(-n\nu(nA_{x,y})), \quad \forall x, y \in \mathbb{R}, \quad \forall n \in \mathbb{N} \\ &= \exp(-\nu(A_{x,y})) \\ &= G(x, y). \end{aligned}$$

So G satisfies Definition 3.3.1 with $a_n = c_n = n$ and $b_n = d_n = 0$, and is hence a max stable distribution. ☐

Next we present a theorem which gives some equivalent formulations of the max domain of attraction condition.

Theorem 3.3.4. *(de Haan and Ferreira, 2006) Let G be a max stable distribution. Let the marginal distribution functions be $\exp\Big(-(1+\gamma_1 x)^{-\frac{1}{\gamma_1}}\Big)$ and $\exp\Big(-(1+\gamma_2 y)^{-\frac{1}{\gamma_2}}\Big)$, and let H be its spectral measure according to the representation of Theorem 3.2.4. Then*

(i) *If the distribution function F_{XY} of the random vector (X, Y) with continuous marginal distribution functions F_X and F_Y is in the max domain of attraction of G, then the following equivalent conditions are fulfilled:*

  (a) *With U_X and U_Y being the tail quantile functions of F_X and F_Y, we have for $x, y > 0$, that*

$$\lim_{t\to\infty} \frac{1 - F_{XY}\left(U_X(tx), U_Y(tx)\right)}{1 - F_{XY}\left(U_X(t), U_Y(t)\right)} = S(x, y) \tag{3.16}$$

  with $S(x,y) := \frac{\log G\left(\frac{x^{\gamma_1}-1}{\gamma_1}, \frac{y^{\gamma_2}-1}{\gamma_2}\right)}{\log G(0,0)}$.

  (b) *For all $r > 1$ and all $s \in [0, 1]$ that are continuity points of H,*

$$\lim_{t\to\infty} P\left(V + W > rt \text{ and } \frac{V}{V+W} \le s \,\middle|\, V + W > t\right) = r^{-1}H(s), \tag{3.17}$$

  where $V := \frac{1}{1-F_X(X)}$ and $W := \frac{1}{1-F_Y(Y)}$

(ii) *Conversely, if the continuous marginal distribution functions* F_X *and* F_Y *are in the domain of attraction of* $\exp\left(-(1+\gamma_1 x)^{-\frac{1}{\gamma_1}}\right)$ *and* $\exp\left(-(1+\gamma_2 y)^{-\frac{1}{\gamma_2}}\right)$*, respectively, and any limit relation* (3.16)-(3.17) *holds for some positive function* S *or some distribution function* H*, then* F_{XY} *is in the max domain of attraction of* G*.*

We saw in Example 3.2.5 that there exists a special case of the spectral measure, where the max stable distribution has independent components. This gives inspiration to the following definition.

Definition 3.3.5. A random vector (X, Y) whose distribution function F_{XY} is in the domain of attraction of a max stable distribution with independent components, is said to have the property of asymptotic independence.

From this definition we are able to obtain the following theorem.

Theorem 3.3.6. *(de Haan and Ferreira, 2006) Let* $F_{XY} : \mathbb{R}^2 \to \mathbb{R}_+$ *be a probability distribution function. Suppose that its marginal distribution functions* $F_X : \mathbb{R} \to \mathbb{R}_+$ *and* $F_Y : \mathbb{R} \to \mathbb{R}_+$ *satisfy*

$$\lim_{n\to\infty} F_X^n(a_n x + b_n) = \exp\left(-(1+\gamma_1 x)^{-\frac{1}{\gamma_1}}\right)$$

and

$$\lim_{n\to\infty} F_Y^n(c_n y + d_n) = \exp\left(-(1+\gamma_2 y)^{-\frac{1}{\gamma_2}}\right)$$

for all x, y *for which* $1+\gamma_1 x > 0$, $1+\gamma_2 y > 0$ *and where* $(b_n)_{n=1}^{\infty}$, $(d_n)_{n=1}^{\infty}$ *are sequences of real constants and* $(a_n)_{n=1}^{\infty}$ *and* $(c_n)_{n=1}^{\infty}$ *are sequences of positive real constants. Let* (X, Y) *be a random vector with distribution function* F_{XY}*. If*

$$\lim_{t\to\infty} \frac{P(X > U_X(t), Y > U_Y(t))}{P(Y > U_Y(t))} = 0, \tag{3.18}$$

then

$$\lim_{n\to\infty} F_{XY}^n(a_n x + b_n, c_n y + d_n) = \exp\left(-(1+\gamma_1 x)^{-\frac{1}{\gamma_1}} - (1+\gamma_2 y)^{-\frac{1}{\gamma_2}}\right)$$

for $1+\gamma_1 x > 0$ *and* $1+\gamma_2 y > 0$*. Hence* X *and* Y *are asymptotically independent. Conversely, asymptotic independence entails* (3.18)*.*

Proof. Assume (3.18) holds. Then also

$$\lim_{t\to\infty} \frac{tP(X > U_X(t), Y > U_Y(t))}{tP(Y > U_Y(t))} = 0.$$

Using Theorem 1.1.2 (i) and (iii) with $x = 0$ we find that

$$\lim_{t\to\infty} tP(Y > U_Y(t)) = 1, \tag{3.19}$$

and hence

$$\lim_{t\to\infty} tP(X > U_X(t), Y > U_Y(t)) = 0.$$

Because of monotonicity, it follows that

$$\lim_{t\to\infty} tP(X > U_X(tx), Y > U_Y(ty)) = 0, \quad \forall x, y > 0,$$

and then also for the set $\tilde{A}_{x,y} := \left\{(s,t) \in \mathbb{R}_+^2 : s > x \text{ and } t > y\right\}$ we have

$$\begin{aligned}\nu\left(\tilde{A}_{x,y}\right) &= \lim_{n\to\infty} \nu_n\left(\tilde{A}_{x,y}\right)\\ &= \lim_{n\to\infty} nP\left(X > U_X(tx), Y > U_Y(ny)\right)\\ &= 0, \quad \forall x,y > 0.\end{aligned}$$

This means that the spectral measure puts its entire mass on the lines $x = 0$ and $y = 0$, i.e.

$$H[\{0\}] = \frac{1}{2} \quad \text{and} \quad H[\{1\}] = \frac{1}{2}.$$

This is equivalent to X and Y being asymptotically independent.
Conversely, assume that X and Y are asymptotically independent. Then

$$G_0(x,y) = \exp\left(-x^{-1} - y^{-1}\right), \quad x,y > 0,$$

and hence for $x = y = 1$ we have

$$G_0(1,1) = \exp(-2).$$

Using Corollary 3.1.3, this implies that

$$\begin{aligned}2 &= \lim_{t\to\infty} t\left(1 - P\left(X \le U_X(t), Y \le U_Y(t)\right)\right) &(3.20)\\ &= \lim_{t\to\infty} t\left(P\left(X > U_X(t)\right) + P\left(Y > U_Y(t)\right) - P\left(X > U_X(t), Y > U_Y(t)\right)\right). &(3.21)\end{aligned}$$

From Theorem 1.1.2 (i) and (iii) it follows that

$$\lim_{t\to\infty} tP\left(X > U_X(t), Y > U_Y(t)\right) = 0,$$

and hence by (3.19), we have that

$$\lim_{t\to\infty} \frac{P\left(X > U_X(t), Y > U_Y(t)\right)}{P\left(Y > U_Y(t)\right)} = 0.$$

□

3.4 Pickands dependence function

Whereas the dependence measures we have discussed previously have straightforward generalizations from the bivariate case to the multidimensional case, this is not true for the following dependence measure. This is strictly a bivariate dependence measure. The dependence measure we are going to discuss is related to the function $L : \mathbb{R}_+^2 \to \mathbb{R}$ given by

$$L(x,y) := -\log G_0\left(\frac{1}{x}, \frac{1}{y}\right). \tag{3.22}$$

This can also be expressed in terms of the exponent measure as

$$L(x,y) = \nu\left\{(s,t) \in \mathbb{R}_+^2 : s > \frac{1}{x} \text{ or } t > \frac{1}{y}\right\}$$

using (3.8), or in terms of the spectral measure as

$$L(x, y) = 2 \int_0^1 (\omega x \vee (1 - \omega)y)\, dH(\omega) \tag{3.23}$$

using (3.10). The function L has the following properties. These are easy to derive from the properties of the exponent and spectral measure and will therefore for brevity not be proven here.

Proposition 3.4.1. *(de Haan and Ferreira, 2006) Let L be as defined in* (3.22)*. Then L has the following properties.*

(i) *Homogeneity of order 1:* $L(ax, ay) = aL(x, y)$, *for all* $a, x, y > 0$.

(ii) $L(x, 0) = L(0, x) = x$, *for all* $x > 0$.

(iii) $x \vee y \leq L(x, y) \leq x + y$, *for all* $x, y > 0$.

(iv) *Let* (X, Y) *be a random vector with distribution function* $G_0(x, y)$. *If* X *and* Y *are independent, then* $L(x, y) = x + y$, *for* $x, y > 0$. *If* X *and* Y *are completely dependent, then* $L(x, y) = x \vee y$ *for* $x, y > 0$.

(v) L *is continuous.*

(vi) $L(x, y)$ *is a convex function:* $L(\lambda (x_1, y_1) + (1 - \lambda)(x_2, y_2)) \leq \lambda L(x_1, y_1) + (1 - \lambda) L(x_2, y_2)$ *for all* $x_1, x_2, y_1, y_2 > 0$ *and* $\lambda \in [0, 1]$.

From the function L we can obtain the Pickands dependence function $A : [0, 1] \to \mathbb{R}$ introduced in Pickands (1981). This function is given by

$$A(t) := -\log G_0\left(\frac{1}{1-t}, \frac{1}{t}\right) = L(1 - t, t). \tag{3.24}$$

If we let $t = \frac{y}{x+y}$ we easily find that

$$L(x, y) = (x + y)A\left(\frac{y}{x+y}\right),$$

and hence Pickands dependence function completely determines the function L.
Pickands dependence function can easily be connected to the spectral measure through the function L. If we combine (3.23) and (3.24) we get

$$\begin{aligned} A(t) &= 2 \int_{[0,1]} (\omega(1 - t) \vee (1 - \omega)t) dH(\omega) \\ &= 2t \int_{[0,t]} (1 - \omega) dH(\omega) + 2(1 - t) \int_{(t,1]} \omega dH(\omega). \end{aligned}$$

Since H has mean $\frac{1}{2}$ we have that $\int_{[0,1]} \omega dH(\omega) = \int_{[0,1]} (1-\omega) dH(\omega) = \frac{1}{2}$. Using this it follows that

$$\int_{(t,1]} \omega dH(\omega) = \frac{1}{2} - H([0, t]) + \int_{[0,t]} (1 - \omega) dH(\omega).$$

Hence

$$A(t) = 2\int_{[0,t]} (1-\omega)dH(\omega) + (1-t)\left(1 - 2H([0,t])\right).$$

The term $\int_{[0,t]}(1-\omega)dH(\omega)$ can also be written as

$$\begin{aligned}
\int_{[0,t]} (1-\omega)dH(\omega) &= \int_{[0,t]}\int_{[\omega,1]} du dH(\omega) \\
&= \int_{[0,1]}\int_{[0,u\wedge t]} dH(\omega)du \\
&= \int_{[0,t]}\int_{[0,u]} dH(\omega)du + \int_{(t,1]}\int_{[0,t]} dH(\omega)du \\
&= \int_{[0,t]} H([0,u])du + (1-t)H([0,t]),
\end{aligned}$$

where $u \wedge t := \min(u,t)$. Hence

$$A(t) = 1 - t + 2\int_{[0,t]} H([0,\omega])d\omega.$$

This means that H can be computed from A through

$$H([0,\omega]) = \begin{cases} \frac{1}{2}\left(1 + A'(\omega)\right) & \text{if } \omega \in [0,1), \\ 1 & \text{if } \omega = 1, \end{cases}$$

where $A'(\omega)$ denotes the right-sided derivative of A.
The point masses of H at 0 and 1 can be computed as

$$H(\{0\}) = \frac{1}{2}\left(1 + A'(0)\right)$$

and

$$H(\{1\}) = \frac{1}{2}\left(1 - A'(1)\right),$$

where $A'(1) = \sup_{0\le t\le 1} A'(t)$.

Example 3.4.2. Using the results from example 3.2.5, we see that Pickands dependence function in the case of completely dependent variables is

$$A(t) = \max(1-t, t),$$

and in the case of independent variables becomes

$$A(t) = 1.$$

This is illustrated in Figure 3.1.

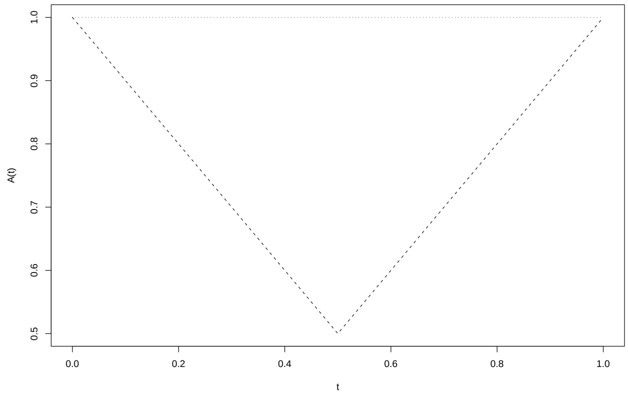

Figure 3.1: Illustration of Pickands dependence function in case of independent variables and completely dependent variables. The dotted line is Pickands dependence function when the variables are independent and the dashed line is when the variables are completely dependent.

In fact, these two cases are also the lower and upper bound of Pickands dependence function, respectively, as can be seen from Proposition 3.4.1(iii)

■

Example 3.4.3. For the logistic model, Pickands dependence function can easily be computed through the exponent measure.

$$\begin{aligned} A(t) &= \nu\left(\frac{1}{1-t}, \frac{1}{t}\right) \\ &= \left((1-t)^{\frac{1}{\alpha}} + t^{\frac{1}{\alpha}}\right)^{\alpha}. \end{aligned}$$

This function is illustrated in Figure 3.2 for different values of α.

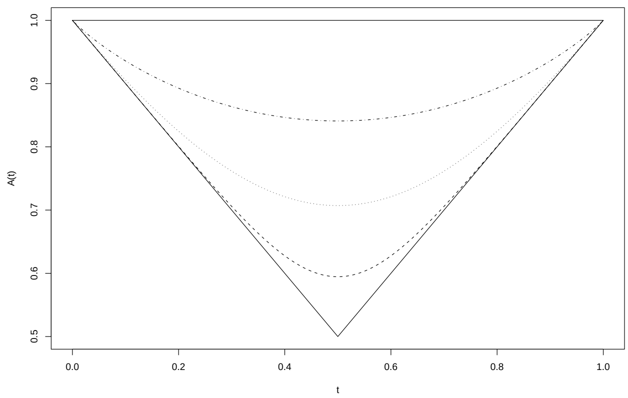

Figure 3.2: Illustration of Pickands dependence function for different values of α in the logistic model. The solid lines are for $\alpha \to 0$ and $\alpha \to 1$, described in Example 3.4.2. The dashed line is for $\alpha = 0.25$, the dotted line is for $\alpha = 0.5$ and the dashed-dotted line is for $\alpha = 0.75$.

■

3.5 The dependence measures χ and $\bar{\chi}$

So far we have focussed on functions that allow us to recover G_0, and hence describe the dependence structure completely. However, similar to classical statistics one can try and summarize the dependence structure in a number of well chosen coefficients, that give a rough, but representative picture of the full dependency structure.
Let (X, Y) be a bivariate random vector with distribution function F_{XY}, and marginal distribution functions F_X and F_Y. For simplicity, we assume that F_X and F_Y are continuous. If we assume that F_X and F_Y are identically, we can define the measure χ by

$$\chi := \lim_{z \uparrow z_*} P\left(Y > z | X > z\right),$$

assuming that the limit exists. A generalization of this is easily obtained if F_X and F_Y are not identical. Since $U_1 := F_X(X)$ and $U_2 := F_Y(Y)$ are uniformly distributed on $(0, 1)$, we can define

$$\chi := \lim_{u \uparrow 1} P\left(U_2 > u | U_1 > u\right), \tag{3.25}$$

assuming that the limit exists. In the following we will use the latter definition of χ. From (3.18) we get that $\chi = 0$ corresponds to asymptotic independence. When $\chi > 0$ we are in the class of asymptotically dependent variables, and here, larger values of χ indicate stronger dependence between the variables. So $\chi > 0$ is a measure of extremal dependence in the class of asymptotic dependent variables. Because the definition of χ is based on the joint distribution of $F_X(X)$ and $F_Y(Y)$ it seems natural natural to make the link between χ and the other dependence measures we have discussed previously, through the copula function. Therefore we will give a short introduction to copulas, and mention some of their basic properties.

Definition 3.5.1. A copula $C : [0, 1]^2 \to [0, 1]$ is a function with the following properties.

(i) For every $x, y \in [0, 1]$

$$C(x, 0) = 0 = C(0, y)$$

and

$$C(x, 1) = x \quad \text{and} \quad C(1, y) = y$$

(ii) For every $x_1, x_2, y_1, y_2 \in [0, 1]$ such that $x_1 \leq x_2$ and $y_1 \leq y_2$,

$$C\left(x_2, y_2\right) - C\left(x_2, y_1\right) - C\left(x_1, y_2\right) + C\left(x_1, y_1\right) \geq 0$$

From the definition of copulas, we easily obtain the following bounds, which are known as the Fréchet-Hoeffding bounds.

Theorem 3.5.2. *Let C be a copula. Then for every $x, y \in [0, 1]$,*

$$\max(x + y - 1, 0) \leq C(x, y) \leq \min(x, y). \tag{3.26}$$

Proof. Let $x_1, x_2, y_1, y_2 \in [0,1]$ be such that $x_1 \leq x_2$ and $y_1 \leq y_2$. Then $t \mapsto C(t, y_2) - C(t, y_1)$ is nondecreasing since

$$C(x_2, y_2) - C(x_2, y_1) \geq C(x_1, y_2) - C(x_1, y_1).$$

Similarly $t \mapsto C(x_2, t) - C(x_1, t)$ is nondecreasing.
Now, let $x, y \in [0,1]$ be arbitrary. Since

$$C(x,y) \leq C(x,1) = x \quad \text{and} \quad C(x,y) \leq C(1,y) = y,$$

we have that $C(x,y) \leq \min(x,y)$. Furthermore

$$1 - x - y + C(x,y) \geq 0 \quad \text{and} \quad C(x,y) \geq 0,$$

so $C(x,y) \geq \max(x+y-1, 0)$. □

The next theorem about copulas is known as Sklar's theorem. It makes the connection between a joint distribution function and its univariate margins.

Theorem 3.5.3. *(Sklar, 1959) Let F_{XY} be a joint distribution function with margins F_X and F_Y. Then there exists a copula C such that for all $x, y \in [-\infty, \infty]$,*

$$F_{XY}(x,y) = C\left(F_X(x), F_Y(y)\right). \tag{3.27}$$

If F_X and F_Y are continuous, then C is unique. Otherwise, C is uniquely determined on $RanF_X \times RanF_Y$, where Ran denotes the range.
Conversely, if C is a copula and F_X and F_Y are distribution functions, then the function $F_{XY}(x,y) := C\left(F_X(x), F_Y(y)\right)$ is a joint distribution function with margins F_X and F_Y.

Now we return to the connection between χ and the other dependence measures. This is found through the function $\chi(u)$, which is defined as

$$\chi(u) := 2 - \frac{\log C(u,u)}{\log u}, \quad 0 < u < 1. \tag{3.28}$$

Concerning this function we have the following proposition, which makes the connection between χ and the exponent measure clear.

Proposition 3.5.4. *For the function $\chi(u)$ defined in* (3.28) *and the measure χ defined in* (3.25) *we have*

(i) $\lim_{u \to 1} \chi(u) = \chi$.

(ii) *In case G is a bivariate extreme value distribution,* $\chi(u) = 2 - \nu(1,1)$.

(iii) $2 - \frac{\log(\max(2u-1,0))}{\log u} \leq \chi(u) \leq 1$.

Proof. *(i)* By making a Taylor series expansion of $\log x$ around 1 we find that

$$\chi(u) = 2 - \frac{1 - C(u,u) + O\left((1 - C(u,u))^2\right)}{1 - u + O\left((1-u)^2\right)}, \quad u \to 1.$$

Using the bounds for copulas, given in Theorem 3.5.2, it follows that

$$\begin{aligned}\chi(u) &= 2 - \frac{1 - C(u,u) + O\left((1-u)^2\right)}{1 - u + O\left((1-u)^2\right)} \\ &= 2 - \frac{1 - C(u,u)}{1-u} + o(1), \quad u \to 1.\end{aligned}$$

From this we obtain

$$\begin{aligned}\chi(u) &= \frac{1 - 2u + P\left(U_1 \leq u, U_2 \leq u\right)}{1-u} + o(1) \\ &= P\left(U_1 > u | U_2 > u\right) + o(1), \quad u \to 1.\end{aligned}$$

Hence $\lim_{u \to 1} \chi(u) = \chi$.

(ii) Concerning the connection between χ and ν, we can without loss of generality assume that G is an extreme value distribution with standard Fréchet margins, so

$$\begin{aligned}\exp(-\nu(x,y)) &= P(X \leq x, Y \leq y) \\ &= P\left(F_X(X) \leq F_X(x), F_Y(Y) \leq F_Y(y)\right) \\ &= C\left(F_X(x), F_Y(y)\right).\end{aligned}$$

If we make the substitutions $u_1 := \exp\left(-\frac{1}{x}\right)$ and $u_2 = \exp\left(-\frac{1}{y}\right)$ it follows that

$$C\left(u_1, u_2\right) = \exp\left(-\nu\left(-\frac{1}{\log u_1}, -\frac{1}{\log u_2}\right)\right).$$

Combining this result with the definition of $\chi(u)$ and using the homogeneity property of ν we find that

$$\begin{aligned}\chi(u) &= 2 - \frac{-\nu\left(-\frac{1}{\log u}, -\frac{1}{\log u}\right)}{\log u} \\ &= 2 - \nu(1,1).\end{aligned}$$

(iii) The bounds for $\chi(u)$ are easily obtained using the bounds for copulas, given in Theorem 3.5.2.

□

Note that Proposition 3.5.4 (*ii*) implies that $\chi(u)$ is constant for bivariate extreme value distributions. So if $\chi(u)$ is not constant, then there is evidence for lack of model fit.
If (U,V) is distributed according to some copula C, and we have some independent observations $(U_1,V_1),\ldots,(U_n,V_n)$ from the copula C, then a natural way to estimate $\chi(u)$ is by

$$\hat{\chi}(u) := 2 - \frac{\log \hat{C}(u,u)}{\log u}, \tag{3.29}$$

where $\hat{C}(u,u)$ is defined by

$$\hat{C}(u,u) := \frac{1}{n}\sum_{i=1}^{n} 1_{\{U_i \leq u, V_i \leq u\}}.$$

The asymptotic distribution of the estimator $\hat{\chi}(u)$ is given in Proposition 3.5.5.

Proposition 3.5.5. *Assume that* (U, V) *are distributed according to the copula* C. *Then we have for* $\hat{\chi}(u)$ *defined in* (3.29) *that*

$$\sqrt{n}\left(\hat{\chi}(u) - \chi(u)\right) \xrightarrow{D} N\left(0, \frac{C(u,u)(1-C(u,u))}{\log^2(u)C^2(u,u)}\right). \tag{3.30}$$

Proof. It follows that

$$E(\hat{C}(u,u)) = C(u,u),$$

and since $\hat{C}(u,u)$ is the estimate of the succes probability for a Bernoulli random variable, we have that

$$\operatorname{var}\left(\hat{C}(u,u)\right) = \frac{C(u,u)(1-C(u,u))}{n}.$$

So by the central limit theorem, it follows that

$$\sqrt{n}\left(\hat{C}(u,u) - C(u,u)\right) \xrightarrow{D} N(0, C(u,u)(1-C(u,u))).$$

From the Delta method, it then follows that

$$\sqrt{n}\left(\hat{\chi}(u) - \chi(u)\right) \xrightarrow{D} N\left(0, \frac{C(u,u)(1-C(u,u))}{\log^2(u)C^2(u,u)}\right),$$

which is the desired result. $\square$

As mentioned previously, when $\chi = 0$ we are in the class of asymptotically independent variables. However at finite levels, the variables are not necessarily independent. This gives rise to the dependence measure $\bar{\chi}$. If we define the survivor copula function $\bar{C}(x,y)$ by

$$\bar{C}(x,y) := P\left(U_1 > x, U_2 > y\right) = 1 - x - y + C(x,y), \quad 0 \leq x, y \leq 1,$$

and let

$$\bar{\chi}(u) := \frac{2\log(1-u)}{\log \bar{C}(u,u)} - 1, \quad 0 < u < 1, \tag{3.31}$$

then we define the measure $\bar{\chi}$ to be

$$\bar{\chi} := \lim_{u \to 1} \bar{\chi}(u), \tag{3.32}$$

assuming that the limit exists. Writing the survival copula in terms of the copula and using the bounds in Theorem 3.5.2 we find that

$$\frac{2\log(1-u)}{\log(\max(0, 1-2u))} - 1 \leq \bar{\chi}(u) \leq 1, \quad 0 < u < 1.$$

Hence, the bounds for $\bar{\chi}$ are given by $-1 \leq \chi \leq 1$.

If (U, V) is distributed according to some copula C, and we have some independent observations $(U_1, V_1), \ldots, (U_n, V_n)$ from the copula C, then the function $\bar{\chi}(u)$ can be estimated by

$$\hat{\bar{\chi}}(u) := \frac{2\log(1-u)}{\log \hat{\bar{C}}(u,u)} - 1, \tag{3.33}$$

where

$$\hat{\bar{C}}(u,u) := \frac{1}{n}\sum_{i=1}^{n} 1_{\{U_i>u, V_i>u\}}. \tag{3.34}$$

The asymptotic distribution of the estimator $\hat{\bar{\chi}}(u)$ is described in Proposition 3.5.6.

Proposition 3.5.6. *Assume that (U,V) are distributed according to the copula C. Then we have for $\hat{\bar{\chi}}(u)$ defined in (3.34) that*

$$\sqrt{n}\left(\hat{\bar{\chi}}(u) - \bar{\chi}(u)\right) \xrightarrow{D} N\left(0, \left(\frac{2\log(1-u)}{\bar{C}(u,u)\log^2(\bar{C}(u,u))}\right)^2\right). \tag{3.35}$$

The proof of this result is analogous to the proof of Proposition 3.5.5. Note that if $u < 1$ but either all $U_i < u$ or all $V_i < u$ then the estimator $\hat{\bar{\chi}}(u) = -1$. This implies that in practice the estimator will always be -1, when u is chosen suffciently large.
When $\chi = 0$, we have that $\bar{\chi}$ is a measure of dependence within the class of asymptotically independent variables. Alternatively, $\chi > 0$ and $\bar{\chi} = 1$ signifies we are in the class of asymptotically dependent variables. So the complete pair $(\chi, \bar{\chi})$ is needed to summarize the extremal dependence. For the pair $(\chi > 0, \bar{\chi} = 1)$, we are in the class of asymptotic dependent variables, where the value of χ is a measure of dependence between the variables. Alternatively, we are in the class of asymptotically independent variables when we have the pair $(\chi = 0, \bar{\chi} \le 1)$, and here, the value of $\bar{\chi}$ is a measure of dependence within this class. This discussion is summarized in Table 3.1.

	Asymptotic independence	Asymptotic dependence
χ	0	$(0,1]$
$\bar{\chi}$	$[-1,1]$	1

Table 3.1: Summary of the pair of measures $(\chi, \bar{\chi})$.

Example 3.5.7. In the case of an extreme value distribution with completely dependent variables, we have that $\nu(1,1) = 1$, and hence $\chi = 1$. The copula needed to find $\bar{\chi}$ is found to be

$$C(u,u) = u.$$

Hence

$$\bar{\chi}(u) = \frac{2\log(1-u)}{\log(1-u)} - 1 = 1,$$

which also implies that $\bar{\chi} = 1$.
When the variables are independent, $\nu(1,1) = 2$, so $\chi = 0$. This time the copula needed to find $\bar{\chi}$ is

$$C(u,u) = u^2$$

Thus

$$\begin{aligned}\bar{\chi}(u) &= \frac{2\log(1-u)}{\log\left((1-u)^2\right)} - 1 \\ &= 0,\end{aligned}$$

so $\bar{\chi} = 0$.

■

Example 3.5.8. Concerning the logistic dependence model, we have that $\nu(1,1) = 2^\alpha$, which means that $\chi = 2 - 2^\alpha$. The copula needed to find $\bar{\chi}$ is

$$C(u,u) = u^{2^\alpha}.$$

From this it follows that

$$\bar{\chi}(u) = \frac{2\log(1-u)}{\log\left(1 - 2u + u^{2^\alpha}\right)} - 1.$$

For $\alpha = 1$ we have that $\bar{\chi}(u) = 0$. For $0 \le \alpha < 1$, we get by using L'Hôpitals rule that

$$\begin{aligned}
\bar{\chi} &= \lim_{u \to 1} \frac{2\left(1 - 2u + u^{2^\alpha}\right)}{(1-u)\left(2 - 2^\alpha u^{2^\alpha - 1}\right)} - 1 \\
&= \lim_{u \to 1} \frac{2\left(-2 + 2^\alpha u^{2^\alpha - 1}\right)}{-2 - 2^\alpha(2^\alpha - 1)u^{2^\alpha - 2} + 2^{2\alpha} u^{2^\alpha - 1}} - 1 \\
&= \frac{-4 + 2^{\alpha+1}}{-2 - 2^\alpha(2^\alpha - 1) + 2^{2\alpha}} - 1 \\
&= 1
\end{aligned}$$

■

Example 3.5.9. Even though the bivariate normal distribution is not a bivariate extreme value distribution, it deserves some attention. We consider the bivariate normal distribution with mean zero, unit variances and correlation coefficient $|\rho| < 1$. It can be shown that in this case, the marginal random variables are asymptotically independent (Sibuya, 1960; de Haan and Ferreira, 2006), i.e. $\chi = 0$. The copula for the bivariate normal distribution with standard normal margins and correlation coefficient $|\rho| < 1$ is for $0 < u, v < 1$,

$$C(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\left(1-\rho^2\right)^{\frac{1}{2}}} \exp\left(-\frac{1}{2\left(1-\rho^2\right)}\left(x^2 + y^2 - 2\rho xy\right)\right) dxdy,$$

where $\Phi^{-1}(\cdot)$ denotes the quantile function of the univariate standard normal distribution. Using this copula we can obtain

$$\chi(u) = 2 - \frac{\log\left(\int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(u)} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(x^2 + y^2 - 2\rho xy\right)\right) dxdy\right)}{\log u}$$

for $0 < u < 1$. As can be seen in Figure 3.3 the convergence of $\chi(u)$ to $\chi = 0$ becomes very slow when ρ increases.

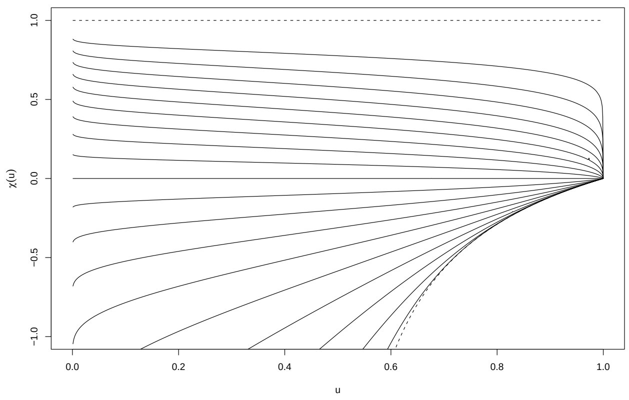

Figure 3.3: Illustration of $\chi(u)$ for the bivariate normal distribution with different values of ρ. The upper and lower bounds for $\chi(u)$ is shown as the dashed lines and the curves in solid correspond to $\rho = -0.9, -0.8, \ldots, 0.9$ (bottom to top).

■

3.6 The model of Ledford and Tawn

Before we introduce the model of Ledford and Tawn, we require some technical preliminaries.

Definition 3.6.1. A function $\mathcal{L} : (0, \infty)^2 \to (0, \infty)$ is called bivariate slowly varying if there exists a function $g : (0, \infty)^2 \to (0, \infty)$ such that

$$\lim_{t\to\infty} \frac{\mathcal{L}\left(tz_1, tz_2\right)}{\mathcal{L}(t,t)} = g\left(z_1, z_2\right), \quad 0 < z_1, z_2 < \infty,$$

where the function g is homogeneous of order zero, i.e.

$$g\left(sz_1, sz_2\right) = g\left(z_1, z_2\right), \quad 0 < s, z_1, z_2 < \infty.$$

Since g is homogeneous of order zero, there exists a function $g_* : (0, 1) \to (0, \infty)$ such that $g\left(z_1, z_2\right) = g_*\left(\frac{z_1}{z_1+z_2}\right)$ for all $0 < z_1, z_2 < \infty$. If

$$\frac{g_*(w)}{g_*(1-w)}, \quad 0 < w < 1$$

is slowly varying at $w \to 0$ and $w \to 1$, then $\mathcal{L}$ is said to be quasi symmetric. Now, let (X, Y) be a random vector with distribution function F_{XY}, and assume that the margins F_X and F_Y are standard Fréchet. Ledford and Tawn (1997) proposed to model the joint survivor function of X and Y as

$$\bar{F}_{XY}(x, y) := P(X > x, Y > y) = \mathcal{L}(x, y)x^{-c_1}y^{-c_2}, \quad c_1, c_2, x, y > 0, \tag{3.36}$$

where $\mathcal{L}$ is a quasi symmetric, bivariately slowly varying function. In the case where $z = x = y$ the model reduces to

$$\bar{F}_{XY}(z,z) = \mathcal{L}(z) z^{-\frac{1}{\eta}}, \quad \eta, z > 0, \tag{3.37}$$

where the parameter $\eta := \frac{1}{c_1+c_2}$ is called the coefficient of tail dependence, and the function $\mathcal{L}(z) := \mathcal{L}(z,z)$ is slowly varying at infinity. This model was originally proposed by Ledford and Tawn (1996). Even though the only restriction on η we have is that $\eta > 0$, we can easily establish that $\eta \leq 1$. Indeed, since $1 - \exp\left(-\frac{1}{z}\right) \sim \frac{1}{z}$, it follows that

$$\begin{aligned} P(X > z, Y > z) &\leq P(X > z) \\ &= 1 - \exp\left(-\frac{1}{z}\right) \\ &\sim \frac{1}{z}. \end{aligned} \tag{3.38}$$

If $\eta > 1$ then (3.37) and (3.38) would make a contradiction.
The connection between η and χ is easily established since

$$P(Y > z | X > z) = \frac{P(Y > z, X > z)}{P(X > z)} \sim \mathcal{L}(z) z^{1-\frac{1}{\eta}}, \quad z \to \infty.$$

If $\lim_{z\to\infty} \mathcal{L}(z) z^{1-\frac{1}{\eta}}$ exists, then

$$\chi = \lim_{z\to\infty} \mathcal{L}(z) z^{1-\frac{1}{\eta}}.$$

The survival function in (3.36) can also be expressed in terms of the survival copula, through

$$\bar{C}\left(\exp\left(-\frac{1}{x}\right), \exp\left(-\frac{1}{y}\right)\right) = P(X > x, Y > y), \quad x, y > 0$$

and hence

$$\bar{C}(u,u) = \mathcal{L}\left(-\frac{1}{\log u}\right)(-\log u)^{\frac{1}{\eta}}, \quad 0 < u < 1.$$

Thus, by (3.31), we get that

$$\begin{aligned} \bar{\chi}(u) &= \frac{2\log(1-u)}{\log\left(\mathcal{L}\left(-\frac{1}{\log u}\right)(-\log u)^{\frac{1}{\eta}}\right)} - 1 \\ &\sim \frac{2\log(1-u)}{\log \mathcal{L}\left(\frac{1}{1-u}\right) + \frac{1}{\eta}\log(1-u)} - 1, \quad u \to 1. \end{aligned}$$

In order to evaluate this limit, we need the following proposition.

Proposition 3.6.2. *(Beirlant et al., 2004) Let $\mathcal{L}$ be slowly varying at infinity. Then*

$$\lim_{x\to\infty} \frac{\log \mathcal{L}(x)}{\log(x)} = 0.$$

By Proposition 3.6.2 the dominant term in the denominator is $\log(1-u)$, hence

$$\bar{\chi}(u) \sim 2\eta - 1, \quad u \to 1.$$

Thus $\bar{\chi} = 2\eta - 1$.
From the connection between η and χ and $\bar{\chi}$, we see that if $\eta = 1$ and $\lim_{z\to\infty} \mathcal{L}(z) = c$, for some constant $0 < c \leq 1$, then $\chi = c$ and $\bar{\chi} = 1$, so the variables are asymptotically dependent of degree $\chi = c$. If $0 < \eta < 1$ or $\eta = 1$ and $\lim_{z\to\infty} \mathcal{L}(z) = 0$, then $\chi = 0$ and the variables are asymptotically independent of degree $\bar{\chi} = 2\eta - 1$. Within the class of asymptotically independent variables, there are three cases. First, $0 < \eta < \frac{1}{2}$ means that the variables are negatively associated, while $\frac{1}{2} < \eta \leq 1$ means the variables are positively associated. For $\eta = \frac{1}{2}$, the variables are near independent if $\mathcal{L}(z) \neq 1$, and independent if $\mathcal{L}(z) = 1$. So the degree of dependence between large values of X and Y are determined by η, where larger values of η indicates stronger association.

Example 3.6.3. If we consider the logistic model, then the joint survivor function $\bar{G}_0(z,z)$ is

$$\bar{G}_0(z,z) = 1 - 2\exp\left(-z^{-1}\right) + \exp\left(-2^{\alpha} z^{-1}\right), \quad z > 0.$$

From Example 3.5.8 we have that $\bar{\chi} = 0$ for $\alpha = 1$. This implies that $\eta = \frac{1}{2}$ and the slowly varying function $\mathcal{L}(z)$ is given by

$$\mathcal{L}(z) = z^2 - 2z^2\exp\left(-z^{-1}\right) + z^2\exp\left(-2z^{-1}\right), \quad z > 0.$$

When $0 \leq \alpha < 1$, then $\bar{\chi} = 1$ and hence $\eta = 1$. In this case the slowly varying function $\mathcal{L}(z)$ is found to be

$$\mathcal{L}(z) = z - 2z\exp\left(-z^{-1}\right) + z\exp\left(-2^{\alpha} z^{-1}\right), \quad z > 0.$$

■

Chapter 4

Estimation of the coefficient of tail dependence and the second order parameter in bivariate extreme value statistics

We start this chapter with an introduction to a class of functional estimators for the coefficient of tail dependence η. For this class of estimators, the asymptotic normality is established under a second order condition on the joint tail behaviour. Then we introduce a class of bias corrected estimators and discuss variance optimality. Next we introduce two ways of estimating the second order parameter τ, which is needed in the bias corrected estimation of η.

4.1 Estimation of the coefficient of tail dependence

Let (X, Y) be a bivariate random vector with distribution function F_{XY}. We can assume, without loss of generality, that the marginal distributions F_X and F_Y of X and Y, respectively, are standard Fréchet. If we assume that F_{XY} satisfies (3.37), and set $Z := \min(X, Y)$, then

$$\begin{aligned} P(Z > z) &= P(X > z, Y > z) \\ &= z^{-\frac{1}{\eta}} \mathcal{L}(z). \end{aligned}$$

This implies that η can be considered as the tail index of a Pareto type model for the random variable Z and hence it can be estimated with classical estimators for the extreme value index-like the Hill (Hill, 1975), moment (Dekkers et al., 1989) or maximum likelihood estimator. In practice, the marginal distributions of a sample of i.i.d. random vectors, $(X_1, Y_1), \ldots (X_n, Y_n)$, are unknown. So we let $\hat{F}_X$ and $\hat{F}_Y$ denote the empirical distribution functions of the X_i and $Y_i, i = 1, \ldots, n$ and define

$$Z_i := \min\left(-\frac{1}{\log\left(\hat{F}_X\left(X_i\right)\right)}, -\frac{1}{\log\left(\hat{F}_Y\left(Y_i\right)\right)} \right), \quad i = 1, \ldots n.$$

Then by the inverse probability integral transform, we have that the Z_i are approximately distributed as the minimum of two standard Fréchet transformed margins. If we order the Z_i, then we let $Z_{1,n} \leq \ldots \leq Z_{n,n}$ be the corresponding ascending order statistics.
For a measureable function $z : [0,1] \to \mathbb{R}$ we define the functional

$$T_K(z) := \int_0^1 \log \frac{z(t)}{z(1)} d(tK(t)),$$

provided the right hand side is defined and finite, and where K is a kernel function. Using this functional we get a class of estimators of η, proposed in Goegebeur and Guillou (2012),

$$\begin{aligned} \hat{\eta}_m(K) :&= T_K(Q_n) \\ &= \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} d(tK(t)), \end{aligned} \tag{4.1}$$

where $Q_n(t) := Z_{n-\lfloor mt \rfloor,n}, 0 < t < \frac{n}{m}$ is the empirical quantile function. This statistic is in fact just a kernel statistic similar to the one we used in Chapter 2, if we assume that $0K(0) = 0$. Indeed

$$\begin{aligned} \hat{\eta}_m(K) &= \sum_{k=1}^m \int_{\frac{k-1}{m}}^{\frac{k}{m}} (\log Q_n(t) - \log Q_n(1))\, d(tK(t)) \\ &= \sum_{k=1}^m (\log Z_{n-k+1,n} - \log Z_{n-m,n}) \int_{\frac{k-1}{m}}^{\frac{k}{m}} d(tK(t)) \\ &= \sum_{k=1}^m \sum_{l=k}^m \left((\log Z_{n-l+1,n} - \log Z_{n-l,n}) \left(\frac{k}{m} K\left(\frac{k}{m}\right) - \frac{k-1}{m} K\left(\frac{k-1}{m}\right)\right)\right) \\ &= \sum_{l=1}^m (\log Z_{n-l+1,n} - \log Z_{n-l,n}) \sum_{k=1}^l \left(\frac{k}{m} K\left(\frac{k}{m}\right) - \frac{k-1}{m} K\left(\frac{k-1}{m}\right)\right) \\ &= \frac{1}{m} \sum_{l=1}^m l \left(\log Z_{n-l+1,n} - \log Z_{n-l,n}\right) K\left(\frac{l}{m}\right). \end{aligned}$$

Just as in Chapter 2, we need some conditions on the kernel function K in order to establish the asymptotic normality of $\hat{\eta}_m(K)$.

Assumption 4.1.1. *Let K be a function defined on $(0,1)$ such that*

(i) *$K(\cdot)$ is continuously differentiable on $(0,1)$,*

(ii) $\int_0^1 (-\log t) d(tK(t)) = 1$,

(iii) *There exists $M > 0$, $0 \leq r < \frac{1}{2}$ and $p < 1$ such that $|K(t)| \leq Mt^{-r}$ and $|K'(t)| \leq Mt^{-p-r}$ for $t \in (0,1)$.*

The conditions on K are not very restrictive. We illustrate this with an example of a kernel function which is much like the one presented in Example 2.2.5.

Example 4.1.2. We consider kernel functions of the form $K(t) := \frac{(\beta+1)^{\alpha+1}}{\Gamma(\alpha+1)}(-\log t)^{\alpha} t^{\beta}$, where $\beta \geq 0$ and $\alpha \geq 1$ or $\alpha = 0$ are tuning parameters. This class has as special cases the Hill estimator $K(t) = 1$, the log weight-type estimator $K(t) = \frac{(-\log t)^{\alpha}}{\Gamma(\alpha+1)}, \alpha \geq 1$ and the weight functions $K(t) = (1+\beta)t^{\beta}, \beta \geq 0$ and $K(t) = (1+\beta)^2 t^{\beta}(-\log t), \beta \geq 0$ proposed in Gomes et al. (2007). We exclude the case $0 < \alpha < 1$ since Assumption 4.1.1 (iii) is violated here.

Lemma 4.1.3. *The kernel function* $K(t) = \frac{(\beta+1)^{\alpha+1}}{\Gamma(\alpha+1)}(-\log t)^{\alpha} t^{\beta}$ *with* $\beta \geq 0$ *and* $\alpha \geq 1$ *or* $\alpha = 0$ *satisfies Assumption 4.1.1.*

A proof of this lemma can be found in Appendix 4.3 ■

In order to prove asymptotic normality of the estimator proposed, we also need an assumption on the second order tail behavior of F_{XY}.

Assumption 4.1.4. *Let* (X, Y) *be a random vector with joint distribution function* F_{XY} *and continuous marginal distributions* F_X *and* F_Y*, respectively. Then*

$$\lim_{t \downarrow 0} \frac{\frac{P(1-F_X(X)<tx, 1-F_Y(Y)<ty)}{q(t)} - c(x,y)}{q_1(t)} =: c_1(x,y) \tag{4.2}$$

exists for all $x \geq 0$, $y \geq 0$ *with* $x + y > 0$, *a positive function* $q \to 0$ *and a function* $q_1 \to 0$ *as* $t \to 0$, *and* c_1 *is a function that is neither a multiple of* c, *nor a constant. We also require that* $c_1(x,x) = \frac{x^{\frac{1}{\eta}}(x^{\tau}-1)}{\tau}$ *and that the convergence is uniform on* $\{(x,y) \in [0,\infty)^2 | x^2 + y^2 = 1\}$.

This assumption is similar to the one used in Draisma et al. (2004). It can be shown that (4.2) implies that q and $|q_1|$ are regularly varying at zero with index $\frac{1}{\eta}$ and $\tau \geq 0$, respectively. The function c is homogeneous of order $\frac{1}{\eta}$, i.e. $c(tx, ty) = t^{\frac{1}{\eta}} c(x,y)$. We can without loss of generality assume that $q(t) = P(1 - F_X(X) < t, 1 - F_Y(Y) < t)$, which of course implies that $c(1,1) = 1$. In what follows, we only consider the case where $\eta < 1$ and $\tau > 0$. There are several examples of copulas which satisfy Assumption 4.1.4. We will consider three of these in section 5.1.

Using Assumption 4.1.1 and 4.1.4 we are able to prove asymptotic normality of the functional estimator $\hat{\eta}_m(K)$ proposed in (4.1). This is stated more formally in Theorem 4.1.5 below.

Theorem 4.1.5. *(Goegebeur and Guillou, 2012) Assume that Assumption 4.1.1 and Assumption 4.1.4 are satisfied with a function* c *that is continuously differentiable and a function* c_1 *that is continuous. If* $m \to \infty$ *such that* $\frac{m}{n} \to 0$ *and* $\sqrt{m} q_1\left(q^{-1}\left(\frac{m}{n}\right)\right) \to \lambda$, *finite, there exists a standard Brownian motion* $\bar{W}$, *such that*

$$\sqrt{m}\left(\hat{\eta}_m(K) - \eta\right) \xrightarrow{d} \eta \int_0^1 \left(t^{-(\eta+1)}\bar{W}(t) + \lambda t^{-\eta}\frac{t^{\eta\tau}-1}{\tau}\right)\nu(dt),$$

where $\nu(dt) := t^{\eta} d(tK(t)) - K(1)\epsilon_1(dt)$, ϵ_1 *denotes the Dirac measure at* 1 *and* $K(1) := \lim_{t\uparrow 1} K(t)$. *In particular* $\sqrt{m}\left(\hat{\eta}_m(K) - \eta\right)$ *is asymptoticallly normal* $N\left(\lambda \mathcal{AB}(K), \mathcal{AV}(K)\right)$ *where*

$$\mathcal{AB}(K) := \frac{\eta}{\tau} \int_0^1 \left(t^{\eta\tau} - 1\right) d(tK(t)), \tag{4.3}$$

$$\mathcal{AV}(K) := \eta^2 \int_0^1 \int_0^1 \frac{\min(s,t)}{st} d(tK(t)) d(sK(s)) - \eta^2 K^2(1). \tag{4.4}$$

A proof of this result requires knowledge about Hadamard differentiability and the functional delta method. These topics would lead us to far, so for a proof of the result we refer to Goegebeur and Guillou (2012). The bias $\mathcal{AB}(K)$ in (4.3) and variance $\mathcal{AV}(K)$ in (4.4) can be simplified, which implies that deriving the mean and variance for a given kernel function becomes much simpler. These simplifications are given in Lemma 4.1.6.

Lemma 4.1.6. *Let K be a kernel function satisfying Assumption 4.1.1. The bias $\mathcal{AB}(K)$ given in* (4.3) *and the variance $\mathcal{AV}(K)$ given in* (4.4) *can also be calculated as*

$$\mathcal{AB}(K) = -\eta^2 \int_0^1 t^{\eta\tau} K(t)dt, \tag{4.5}$$

$$\mathcal{AV}(K) = \eta^2 \int_0^1 K^2(t)dt. \tag{4.6}$$

Proof. From (4.3) we see that

$$\begin{aligned}\mathcal{AB}(K) &= \frac{\eta}{\tau}\int_0^1 t^{\eta\tau+1}K'(t)dt + \frac{\eta}{\tau}\int_0^1 t^{\eta\tau}K(t)dt - \frac{\eta}{\tau}\int_0^1 tK'(t)dt - \frac{\eta}{\tau}\int_0^1 K(t)dt \\ &=: I_1 + I_2 - I_3 - I_4.\end{aligned}$$

By applying integration by parts to I_1 and I_3 we find that

$$I_1 = \frac{\eta}{\tau}K(1) - \eta^2\int_0^1 t^{\eta\tau}K(t)dt - I_2$$

and

$$I_3 = \frac{\eta}{\tau}K(1) - I_4.$$

So $\mathcal{AB}(K) = -\eta^2 \int_0^1 t^{\eta\tau} K(t)dt$.
Concerning the variance in (4.6) we get from (4.4) that

$$\begin{aligned}\mathcal{AV}(K) &= \eta^2\int_0^1 \frac{1}{s}\int_0^s d(tK(t))d(sK(s)) + \eta^2\int_0^1\int_s^1 \frac{1}{t}d(tK(t))d(sK(s)) - \eta^2K^2(1) \\ &=: \eta^2 I_1 + \eta^2 I_2 - \eta^2 K^2(1).\end{aligned}$$

By changing the order of integration in I_2 we get that $I_1 = I_2$, hence $\mathcal{AV}(K) = 2\eta^2 I_1 - \eta^2 K^2(1)$.
Now concerning I_1, we find that

$$I_1 = \int_0^1 sK'(s)K(s)ds + \int_0^1 K^2(s)ds$$

By applying integration by parts we get that

$$\int_0^1 sK'(s)K(s)ds = \frac{1}{2}K^2(1) - \frac{1}{2}\int_0^1 K^2(s)ds$$

hence $\mathcal{AV}(K) = \eta^2 \int_0^1 K^2(t)dt$. □

From direct calculation of (4.5) and (4.6) we get the following Corollary.

Corollary 4.1.7. *Under the assumptions of Theorem 4.1.5, if* $K(t) := \frac{(\beta+1)^{\alpha+1}}{\Gamma(\alpha+1)}(-\log t)^{\alpha} t^{\beta}, \beta \geq 0, \alpha \geq 1$ *or* $\alpha = 0$*, then*

$$\sqrt{m}\left(\hat{\eta}_m(K) - \eta\right) \xrightarrow{d} N\left(-\lambda\eta^2 \frac{(1+\beta)^{\alpha+1}}{(\beta+\eta\tau+1)^{\alpha+1}}, \frac{(\beta+1)^{2\alpha+2}\Gamma(2\alpha+1))}{(2\beta+1)^{2\alpha+1}\Gamma^2(\alpha+1)}\right).$$

As can be seen from Theorem 4.1.5 the estimator of η we have proposed suffers from asymptotic bias, meaning that the center of the normal limit is not zero. This happens even though the estimator is consistent. It is however relatively easy to get rid of this bias. In order to see this, let K_1 and K_2 be two different kernels satisfying Assumption 4.1.1. If $\alpha \in \mathbb{R}$, then

$$K_\alpha(t) := \alpha K_1(t) + (1-\alpha)K_2(t), \quad 0 < t < 1 \tag{4.7}$$

also satisfies Assumption 4.1.1. From Theorem 4.1.5 we see that the asymptotic bias resulting from using this kernel function is

$$\begin{aligned}\mathcal{AB}(K_\alpha) :&= \frac{\eta}{\tau}\int_0^1 (t^{\eta\tau} - 1)\, d(tK_\alpha(t)) \\ &= \alpha\mathcal{AB}(K_1) + (1-\alpha)\mathcal{AB}(K_2).\end{aligned}$$

Under the assumption that $\mathcal{AB}(K_1) \neq \mathcal{AB}(K_2)$ we have that

$$\alpha^* = \frac{\mathcal{AB}(K_2)}{\mathcal{AB}(K_2) - \mathcal{AB}(K_1)} \tag{4.8}$$

leads to a value of α that eliminates bias, that is $\mathcal{AB}(K_{\alpha^*}) = 0$. This result is stated more formally in Proposition 4.1.8.

Proposition 4.1.8. *(Goegebeur and Guillou, 2012) Assume that Assumption 4.1.4 is satisfied with a function c that is continuously differentiable and a function c_1 that is continuous. Furthermore, assume that K_1 and K_2 are kernel functions satisfying Assumption 4.1.1 with $\mathcal{AB}(K_1) \neq \mathcal{AB}(K_2)$. If $m \to \infty$ such that $\frac{m}{n} \to 0$ and $\sqrt{m}q_1\left(q^{-1}\left(\frac{m}{n}\right)\right) \to \lambda$, finite, then*

$$\sqrt{m}\left(T_{K_{\alpha^*}}(Q_n) - \eta\right) \xrightarrow{d} N\left(0, \mathcal{AV}(K_{\alpha^*})\right).$$

Note that the choice of α^* in (4.8) depends on both η and τ, i.e. $\alpha^* = \alpha^*(\tau, \eta)$. Since these are unknown parameters they need to be estimated from the data. Replacing the true parameters η and τ with initial consistent estimators does not change the limiting distribution of the normalized bias corrected estimator, as can be seen from Proposition 4.1.9.

Proposition 4.1.9. *(Goegebeur and Guillou, 2012) Assume (i) that Assumption 4.1.4 is satisfied with a function c that is continuously differentiable and a function c_1 that is continuous, (ii) kernel functions K_1 and K_2 that satisfy Assumption 4.1.1 with $\mathcal{AB}(K_1) \neq \mathcal{AB}(K_2)$ and that are such that α^* is continuously differentiable with respect to η and τ, and (iii) $\tilde{\eta}$ and $\tilde{\tau}$ are initial consistent estimators for η and τ, respectively. Then, if $m \to \infty$ such that $\frac{m}{n} \to 0$ and $\sqrt{m}q_1\left(q^{-1}\left(\frac{m}{n}\right)\right) \to \lambda$, finite, we have that*

$$\sqrt{m}\left(T_{K_{\hat{\alpha}^*}}(Q_n) - \eta\right) \xrightarrow{d} N\left(0, \mathcal{AV}(K_{\alpha^*})\right),$$

where $\hat{\alpha}^* := \alpha^*(\tilde{\eta}, \tilde{\tau})$.

Next we construct an asymptotically unbiased functional estimator for η with minimum variance. This construction is inspired by Drees (1998) and Proposition 3 in Gardes and Girard (2008).

Theorem 4.1.10. *(Goegebeur and Guillou, 2012) Let* $\alpha_{opt} := \frac{(1+\eta\tau)^2}{\eta^2\tau^2}$, $K_1(t) = 1$ *and* $K_2(t) = (1+\eta\tau)t^{\eta\tau}$. *Then* $K_{\alpha_{opt}}(\cdot)$ *defined as in (4.7) is the asymptotically unbiased weight function with minimum variance among unbiased weight functions satisfying Assumption 4.1.1.*

Corollary 4.1.11. *Under the assumptions of Theorem 4.1.5 and Theorem 4.1.10 we have that*

$$\sqrt{m}\left(\hat{\eta}_m\left(K_{\alpha_{opt}}\right)-\eta\right) \xrightarrow{d} N\left(0, \eta^2\frac{(1+\eta\tau)^2}{\eta^2\tau^2}\right).$$

4.2 Estimation of the second order parameter

Estimation of the second order parameter is important because a consistent estimate of this parameter is required in order to obtain a bias corrected estimate of η. In order to create consistent estimators of τ, we start by defining the function

$$S(x,y) := \sum_{i=1}^{n} 1\left\{X_i \geq x, Y_i \geq y\right\}, \tag{4.9}$$

where $1\{A\}$ denotes the indicator funtion of the event A. Using this function we introduce the estimtor $\hat{\tau}_k\left(x,y,\tilde{\eta}\right)$ for τ, defined as

$$\hat{\tau}_k\left(x,y,\tilde{\eta}\right) := -\frac{1}{\log 2}\log\left|\frac{\left(\frac{S\left(X_{n-\lfloor kx\rfloor+1,n},Y_{n-\lfloor ky\rfloor+1,n}\right)}{m_1}\right)^a - 2^{\frac{a}{\tilde{\eta}}}\left(\frac{S\left(X_{n-\left\lfloor\frac{kx}{2}\right\rfloor+1,n},Y_{n-\left\lfloor\frac{ky}{2}\right\rfloor+1,n}\right)}{m_1}\right)^a}{\left(\frac{S\left(X_{n-\lfloor 2kx\rfloor+1,n},Y_{n-\lfloor 2ky\rfloor+1,n}\right)}{m_2}\right)^a - 2^{\frac{a}{\tilde{\eta}}}\left(\frac{S\left(X_{n-\lfloor kx\rfloor+1,n},Y_{n-\lfloor ky\rfloor+1,n}\right)}{m_2}\right)^a}\right|, \tag{4.10}$$

where $a \neq 0, k := ne^{-(\log n)^\delta}, 0<\delta<1, m_1 := \left\lfloor nq\left(\frac{k}{n}\right)\right\rfloor, m_2 := \left\lfloor nq\left(\frac{2k}{n}\right)\right\rfloor$, and $\tilde{\eta}$ is an initial consistent estimator for η. Our estimator $\hat{\tau}$ is consistent as can be seen in Proposition 4.2.1.

Proposition 4.2.1. *Assume that Assumption 4.1.4 is satisfied with a function c that has continuous first order partial derivatives and a continuous function c_1. If $n \to \infty$, $k = ne^{-(\log n)^\delta}, 0<\delta<1, m_1 = \left\lfloor nq\left(\frac{k}{n}\right)\right\rfloor$ and $m_2 = \left\lfloor nq\left(\frac{2k}{n}\right)\right\rfloor$, then $\hat{\tau}_k\left(x,y,\eta\right) \xrightarrow{P} \tau$. This result continues to hold when η is replaced by a consistent estimator $\tilde{\eta}$ that satsfies*

$$\frac{\tilde{\eta}-\eta}{q_1\left(\frac{k}{n}\right)} \xrightarrow{P} 0.$$

Proof. From Lemma 6.1 in Draisma et al. (2004) and following the method of proof in Drees (1998) we get that

$$\frac{S\left(X_{n-\lfloor kx\rfloor+1,n},Y_{n-\lfloor ky\rfloor+1,n}\right)}{m_1} = c(x,y)+q_1\left(\frac{k}{n}\right)c_1(x,y)\left(1+o_P(1)\right)+\frac{W(x,y)}{\sqrt{m_1}}+o_P\left(\frac{1}{\sqrt{m_1}}\right),$$

where W is a Gaussian process with mean zero and covariance structure given by

$$E\left(W\left(x_1,y_1\right)W\left(x_2,y_2\right)\right)=c\left(x_1\wedge x_2,y_1\wedge y_2\right).$$

By making a Taylor series expansion we find that

$$\begin{aligned}\left(\frac{S\left(X_{n-\lfloor kx\rfloor+1,n},Y_{n-\lfloor ky\rfloor+1,n}\right)}{m_1}\right)^a =&c^a(x,y)+ac^{a-1}(x,y)\Bigg(q_1\left(\frac{k}{n}\right)c_1(x,y)\left(1+o_P(1)\right)\\ &+\frac{W(x,y)}{\sqrt{m_1}}+o_P\left(\frac{1}{\sqrt{m_1}}\right)\Bigg).\end{aligned}\tag{4.11}$$

Hence

$$\begin{aligned}&\left(\frac{S\left(X_{n-\lfloor kx\rfloor+1,n},Y_{n-\lfloor ky\rfloor+1,n}\right)}{m_1}\right)^a-2^{-\frac{a}{\eta}}\left(\frac{S\left(X_{n-\left\lfloor\frac{kx}{2}\right\rfloor+1,n},Y_{n-\left\lfloor\frac{ky}{2}\right\rfloor+1,n}\right)}{m_1}\right)^a\\ &=ac^{a-1}(x,y)\left(q_1\left(\frac{k}{n}\right)c_1(x,y)\left(1+o_P(1)\right)+\frac{W(x,y)}{\sqrt{m_1}}+o_P\left(\frac{1}{\sqrt{m_1}}\right)\right)\\ &-2^{-\frac{a}{\eta}}ac^{a-1}\left(\frac{x}{2},\frac{y}{2}\right)\left(q_1\left(\frac{k}{n}\right)c_1\left(\frac{x}{2},\frac{y}{2}\right)\left(1+o_P(1)\right)+\frac{W\left(\frac{x}{2},\frac{y}{2}\right)}{\sqrt{m_1}}+o_P\left(\frac{1}{\sqrt{m_1}}\right)\right).\end{aligned}$$

For the k sequence we have chosen

$$q\left(\frac{k}{n}\right)=e^{-\frac{1}{\eta}(\log n)^\delta}l_q\left(e^{-(\log n)^\delta}\right)$$

and

$$q_1\left(\frac{k}{n}\right)=e^{-\tau(\log n)^\delta}l_{q_1}\left(e^{-(\log n)^\delta}\right)$$

for some slowly varying functions l_q and l_{q_1}. This implies that

$$\begin{aligned}\sqrt{m_1}q_1\left(\frac{k}{n}\right)&=\sqrt{\left\lfloor ne^{-\frac{1}{\eta}(\log n)^\delta}l_q\left(e^{-(\log n)^\delta}\right)\right\rfloor}e^{-\tau(\log n)^\delta}l_{q_1}\left(e^{-(\log n)^\delta}\right)\\ &\geq\sqrt{n}e^{-\left(\frac{1}{2\eta}+\tau\right)(\log n)^\delta}\sqrt{l_q\left(e^{-(\log n)^\delta}\right)}l_{q_1}\left(e^{-(\log n)^\delta}\right)-e^{-\tau(\log n)^\delta}l_{q_1}\left(e^{-(\log n)^\delta}\right)\\ &\to\infty,\end{aligned}$$

for $n\to\infty$. Thus

$$\begin{aligned}&\left(\frac{S\left(X_{n-\lfloor kx\rfloor+1,n},Y_{n-\lfloor ky\rfloor+1,n}\right)}{m_1}\right)^a-2^{-\frac{a}{\eta}}\left(\frac{S\left(X_{n-\left\lfloor\frac{kx}{2}\right\rfloor+1,n},Y_{n-\left\lfloor\frac{ky}{2}\right\rfloor+1,n}\right)}{m_1}\right)^a\\ &=ac^{a-1}(x,y)q_1\left(\frac{k}{n}\right)c_1(x,y)\left(1+o_P(1)\right)\\ &-2^{-\frac{a}{\eta}}ac^{a-1}\left(\frac{x}{2},\frac{y}{2}\right)q_1\left(\frac{k}{n}\right)c_1\left(\frac{x}{2},\frac{y}{2}\right)\left(1+o_P(1)\right)\\ &=q_1\left(\frac{k}{n}\right)a\left(c^{a-1}(x,y)c_1(x,y)-2^{-\frac{a}{\eta}}c^{a-1}\left(\frac{x}{2},\frac{y}{2}\right)c_1\left(\frac{x}{2},\frac{y}{2}\right)+o_P(1)\right).\end{aligned}\tag{4.12}$$

Similarly

$$\left(\frac{S\left(X_{n-\lfloor 2kx\rfloor+1,n},Y_{n-\lfloor 2ky\rfloor+1,n}\right)}{m_2}\right)^a-2^{-\frac{a}{\eta}}\left(\frac{S\left(X_{n-\lfloor kx\rfloor+1,n},Y_{n-\lfloor ky\rfloor+1,n}\right)}{m_2}\right)^a$$
$$=q_1\left(\frac{2k}{n}\right)a\left(c^{a-1}(x,y)c_1(x,y)-2^{-\frac{a}{\eta}}c^{a-1}\left(\frac{x}{2},\frac{y}{2}\right)c_1\left(\frac{x}{2},\frac{y}{2}\right)+o_P(1)\right). \tag{4.13}$$

By the regular variation of $|q_1|$ we find that

$$\frac{\left(\frac{S\left(X_{n-\lfloor kx\rfloor+1,n},Y_{n-\lfloor ky\rfloor+1,n}\right)}{m_1}\right)^a-2^{-\frac{a}{\eta}}\left(\frac{S\left(X_{n-\left\lfloor \frac{kx}{2}\right\rfloor+1,n},Y_{n-\left\lfloor \frac{ky}{2}\right\rfloor+1,n}\right)}{m_1}\right)^a}{\left(\frac{S\left(X_{n-\lfloor 2kx\rfloor+1,n},Y_{n-\lfloor 2ky\rfloor+1,n}\right)}{m_2}\right)^a-2^{-\frac{a}{\eta}}\left(\frac{S\left(X_{n-\lfloor kx\rfloor+1,n},Y_{n-\lfloor ky\rfloor+1,n}\right)}{m_2}\right)^a}\xrightarrow{P}2^{-\tau}.$$

Concerning the second part of the proof, we have from a first order Taylor series expansion that

$$\hat{\tau}\left(x,y,\tilde{\eta}\right)=\hat{\tau}(x,y,\eta)+\left.\frac{d\hat{\tau}(x,y,\eta)}{d\eta}\right|_{\eta^*}\left(\tilde{\eta}-\eta\right),$$

where η^* is a random value between η and $\tilde{\eta}$. The term $\left.\frac{d\hat{\tau}(x,y,\eta)}{d\eta}\right|_{\eta^*}$ is easily found to be

$$\left.\frac{d\hat{\tau}(x,y,\eta)}{d\eta}\right|_{\eta^*}=\frac{a}{\eta^{*2}}2^{\frac{a}{\eta^*}}\left(\frac{\left(\frac{S\left(X_{n-\left\lfloor \frac{kx}{2}\right\rfloor+1,n},Y_{n-\left\lfloor \frac{ky}{2}\right\rfloor+1,n}\right)}{m_1}\right)^a}{\left(\frac{S\left(X_{n-\lfloor kx\rfloor+1,n},Y_{n-\lfloor ky\rfloor+1,n}\right)}{m_1}\right)^a-2^{-\frac{a}{\eta^*}}\left(\frac{S\left(X_{n-\left\lfloor \frac{kx}{2}\right\rfloor+1,n},Y_{n-\left\lfloor \frac{ky}{2}\right\rfloor+1,n}\right)}{m_1}\right)^a}\right.$$
$$\left.-\frac{\left(\frac{S\left(X_{n-\lfloor kx\rfloor+1,n},Y_{n-\lfloor ky\rfloor+1,n}\right)}{m_1}\right)^a}{\left(\frac{S\left(X_{n-\lfloor 2kx\rfloor+1,n},Y_{n-\lfloor 2ky\rfloor+1,n}\right)}{m_2}\right)^a-2^{-\frac{a}{\eta^*}}\left(\frac{S\left(X_{n-\lfloor kx\rfloor+1,n},Y_{n-\lfloor ky\rfloor+1,n}\right)}{m_2}\right)^a}\right).$$

By inserting the terms (4.11), (4.12) and (4.13) into this expression it follows that

$$\left.\frac{d\hat{\tau}(x,y,\eta)}{d\eta}\right|_{\eta^*}=O_P\left(\frac{1}{q_1\left(\frac{k}{n}\right)}\right),$$

and hence the result follows. □

Even though we in (4.10) have a consistent estimator for τ, there still exists some unresolved problems with this estimator. In practice we do not know the function q, so we can not determine the constants m_1 and m_2. A way to partially solve this, is to note that the estimator in (4.10) also can be written as

$$\hat{\tau}_k\left(x,y,\tilde{\eta}\right):=-\frac{1}{\log 2}\log\left|\left(\frac{m_2}{m_1}\right)^a\frac{S\left(X_{n-\lfloor kx\rfloor+1,n},Y_{n-\lfloor ky\rfloor+1,n}\right)^a-2^{\frac{a}{\tilde{\eta}}}S\left(X_{n-\left\lfloor \frac{kx}{2}\right\rfloor+1,n},Y_{n-\left\lfloor \frac{ky}{2}\right\rfloor+1,n}\right)^a}{S\left(X_{n-\lfloor 2kx\rfloor+1,n},Y_{n-\lfloor 2ky\rfloor+1,n}\right)^a-2^{\frac{a}{\tilde{\eta}}}S\left(X_{n-\lfloor kx\rfloor+1,n},Y_{n-\lfloor ky\rfloor+1,n}\right)^a}\right|.$$

Since q is a function of regular variation with index $\frac{1}{\eta}$, it is a function of the form $q(t) = l_q(t)t^{\frac{1}{\eta}}$, where l_q is a slowly varying function. Thus it seems natural to approximate $m := \frac{m_2}{m_1}$ by

$$\hat{m} := 2^{\frac{1}{\tilde{\eta}}}, \quad k = 1, \dots, n,$$

where $\tilde{\eta}$ is a consistent estimator for η. In the estimator proposed in (4.10) we also have to use an estimate for η which would be better to avoid. There is however a way of eliminating the necessity of knowing the function q and using an estimate for η as can be seen from another estimator of τ we propose. This estimator is also based on the function S defined in (4.9), and is given by

$$\tilde{\tau}_k(z) = \frac{1}{\log 2} \log \left| \frac{b(a+l)}{a(b+l)} \frac{\Xi_1(z,a,l)}{\Xi_2(z,b,l)} \right|, \tag{4.14}$$

where

$$\Xi_1(z,a,l) := \frac{\left(\frac{S(X_{n-\lfloor kz\rfloor+1,n}, Y_{n-\lfloor kz\rfloor+1,n})}{S(X_{n-\lfloor 2kz\rfloor+1,n}, Y_{n-\lfloor 2kz\rfloor+1,n})}\right)^a - \left(\frac{S(X_{n-\lfloor 2kz\rfloor+1,n}, Y_{n-\lfloor 2kz\rfloor+1,n})}{S(X_{n-\lfloor 4kz\rfloor+1,n}, Y_{n-\lfloor 4kz\rfloor+1,n})}\right)^a}{\left(\frac{S(X_{n-\lfloor 2kz\rfloor+1,n}, Y_{n-\lfloor 2kz\rfloor+1,n})}{S(X_{n-\lfloor 4kz\rfloor+1,n}, Y_{n-\lfloor 4kz\rfloor+1,n})}\right)^{a+l} - \left(\frac{S(X_{n-\lfloor 4kz\rfloor+1,n}, Y_{n-\lfloor 4kz\rfloor+1,n})}{S(X_{n-\lfloor 8kz\rfloor+1,n}, Y_{n-\lfloor 8kz\rfloor+1,n})}\right)^{a+l}}, \quad a, a+l \neq 0$$

and

$$\Xi_2(z,b,l) := \frac{\left(\frac{S(X_{n-\lfloor kz\rfloor+1,n}, Y_{n-\lfloor kz\rfloor+1,n})}{S(X_{n-\lfloor 2kz\rfloor+1,n}, Y_{n-\lfloor 2kz\rfloor+1,n})}\right)^b - \left(\frac{S(X_{n-\lfloor 2kz\rfloor+1,n}, Y_{n-\lfloor 2kz\rfloor+1,n})}{S(X_{n-\lfloor 4kz\rfloor+1,n}, Y_{n-\lfloor 4kz\rfloor+1,n})}\right)^b}{\left(\frac{S(X_{n-\lfloor 4kz\rfloor+1,n}, Y_{n-\lfloor 4kz\rfloor+1,n})}{S(X_{n-\lfloor 8kz\rfloor+1,n}, Y_{n-\lfloor 8kz\rfloor+1,n})}\right)^{b+l} - \left(\frac{S(X_{n-\lfloor 8kz\rfloor+1,n}, Y_{n-\lfloor 8kz\rfloor+1,n})}{S(X_{n-\lfloor 16kz\rfloor+1,n}, Y_{n-\lfloor 16kz\rfloor+1,n})}\right)^{b+l}}, \quad b, b+l \neq 0.$$

This estimator is based on the same ideas as was used to construct the ρ estimator presented in Section 2.3.

Proposition 4.2.2. *Assume that Assumption 4.1.4 is satisfied with a function c that has continuous first order partial derivatives and a continuous function c_1. If $k, n \to \infty$ such that $\frac{k}{n} \to 0$ and $\sqrt{m_1} q_1\left(\frac{k}{n}\right) \to \infty$, for m_1 defined as in Proposition 4.2.1, then $\tilde{\tau}_k(z) \xrightarrow{P} \tau$.*

Proof. Let

$$T_1 := \frac{S\left(X_{n-\lfloor kz\rfloor+1,n}, Y_{n-\lfloor kz\rfloor+1,n}\right)}{S\left(X_{n-\lfloor 2kz\rfloor+1,n}, Y_{n-\lfloor 2kz\rfloor+1,n}\right)}.$$

Then

$$T_1^a = \left(\frac{c(z,z) + q_1\left(\frac{k}{n}\right) c_1(z,z)\left(1 + o_P(1)\right) + \frac{W(z,z)}{\sqrt{m1}} + o_P\left(\frac{1}{\sqrt{m_1}}\right)}{c(2z,2z) + q_1\left(\frac{k}{n}\right) c_1(2z,2z)\left(1 + o_P(1)\right) + \frac{W(2z,2z)}{\sqrt{m1}} + o_P\left(\frac{1}{\sqrt{m_1}}\right)}\right)^a$$

and since $\sqrt{m_1} q_1\left(\frac{k}{n}\right) \to \infty$, it follows that

$$T_1^a = \left(\frac{c(z,z) + q_1\left(\frac{k}{n}\right) c_1(z,z)\left(1 + o_P(1)\right)}{c(2z,2z) + q_1\left(\frac{k}{n}\right) c_1(2z,2z)\left(1 + o_P(1)\right)}\right)^a.$$

By using Assumption 4.1.4 and making a Taylor series expansion we get

$$T_1^a = 2^{-\frac{a}{\eta}} + a2^{-\frac{a}{\eta}} q_1\left(\frac{k}{n}\right) \frac{z^\tau - (2z)^\tau}{\tau} \left(1 + o_P(1)\right).$$

Similarly, it follows that

$$\begin{aligned} T_2^a &:= \left(\frac{S\left(X_{n-\lfloor 2kz\rfloor+1,n}, Y_{n-\lfloor 2kz\rfloor+1,n}\right)}{S\left(X_{n-\lfloor 4kz\rfloor+1,n}, Y_{n-\lfloor 4kz\rfloor+1,n}\right)}\right)^a \\ &= 2^{-\frac{a}{\eta}} + a2^{-\frac{a}{\eta}} q_1\left(\frac{k}{n}\right) \frac{(2z)^\tau - (4z)^\tau}{\tau} \left(1 + o_P(1)\right) \end{aligned}$$

and

$$\begin{aligned} T_3^a &:= \left(\frac{S\left(X_{n-\lfloor 4kz\rfloor+1,n}, Y_{n-\lfloor 4kz\rfloor+1,n}\right)}{S\left(X_{n-\lfloor 8kz\rfloor+1,n}, Y_{n-\lfloor 8kz\rfloor+1,n}\right)}\right)^a \\ &= 2^{-\frac{a}{\eta}} + a2^{-\frac{a}{\eta}} q_1\left(\frac{k}{n}\right) \frac{(4z)^\tau - (8z)^\tau}{\tau} \left(1 + o_P(1)\right). \end{aligned}$$

Hence

$$T_1^a - T_2^a = a2^{-\frac{a}{\eta}} q_1\left(\frac{k}{n}\right) z^\tau \frac{(1-2^\tau)^2}{\tau} \left(1 + o_P(1)\right)$$

and

$$T_2^a - T_3^a = a2^{-\frac{a}{\eta}} q_1\left(\frac{k}{n}\right) 2^\tau z^\tau \frac{(1-2^\tau)^2}{\tau} \left(1 + o_P(1)\right).$$

We have that

$$\frac{T_1^a - T_2^a}{q_1\left(\frac{k}{n}\right)} \xrightarrow{P} a2^{-\frac{a}{\eta}} z^\tau \frac{(1-2^\tau)^2}{\tau}$$

and

$$\frac{T_2^a - T_3^a}{q_1\left(\frac{k}{n}\right)} \xrightarrow{P} a2^{-\frac{a}{\eta}} 2^\tau z^\tau \frac{(1-2^\tau)^2}{\tau}$$

so

$$\begin{aligned} \Xi_1(z,a,l) &= \frac{T_1^a - T_2^a}{T_2^{a+l} - T_3^{a+l}} \\ &\xrightarrow{P} \frac{a}{a+l} 2^{\frac{l}{\eta}} 2^{-\tau}. \end{aligned}$$

Similarly we find for

$$\begin{aligned} T_4^a &:= \left(\frac{S\left(X_{n-\lfloor 8kz\rfloor+1,n}, Y_{n-\lfloor 8kz\rfloor+1,n}\right)}{S\left(X_{n-\lfloor 16kz\rfloor+1,n}, Y_{n-\lfloor 16kz\rfloor+1,n}\right)}\right)^a \\ &= 2^{-\frac{a}{\eta}} + a2^{-\frac{a}{\eta}} q_1\left(\frac{k}{n}\right) \frac{(8z)^\tau - (16z)^\tau}{\tau} \left(1 + o_P(1)\right). \end{aligned}$$

that

$$\begin{aligned}\Xi_2(z,b,l) &= \frac{T_1^b - T_2^b}{T_3^{b+l} - T_4^{b+l}} \\ &\xrightarrow{P} \frac{b}{b+l} 2^{\frac{l}{\eta}} 4^{-\tau}.\end{aligned}$$

Hence

$$\frac{\Xi_1(z,a,l)}{\Xi_2(z,b,l)} \xrightarrow{P} \frac{a(b+l)}{b(a+l)} 2^{\tau}.$$

□

4.3 Appendix

4.3.1 Proof of Lemma 4.1.3

(i) We see that

$$\frac{d}{dt}K(t) = \frac{(\beta+1)^{\alpha+1}\beta}{\Gamma(\alpha+1)}(-\log t)^{\alpha} t^{\beta-1} - \frac{(\beta+1)^{\alpha+1}\alpha}{\Gamma(\alpha+1)} t^{\beta-1}(-\log t)^{\alpha-1}$$

is continuous on $(0,1)$, so (i) is satisfied.

(ii) Concerning the second condition, we have that

$$\begin{aligned}\int_0^1 (-\log t) d(tK(t)) &= \frac{(1+\beta)^{\alpha+2}}{\Gamma(\alpha+1)} \int_0^1 (-\log t)^{\alpha+1} t^{\beta} dt - \frac{(1+\beta)^{\alpha+1}\alpha}{\Gamma(\alpha+1)} \int_0^1 (-\log t)^{\alpha} t^{\beta} dt \\ &= \frac{1}{\Gamma(\alpha+1)} \int_0^{\infty} z^{\alpha+1} \exp(-z) dz - \frac{\alpha}{\Gamma(\alpha+1)} \int_0^{\infty} z^{\alpha} \exp(-z) dz \\ &= 1.\end{aligned}$$

(iii) Let $\alpha \geq 1$ and $\beta \geq 0$. The inequality $\left|\frac{(\beta+1)^{\alpha+1}}{\Gamma(\alpha+1)}(-\log t)^{\alpha} t^{\beta}\right| \leq M t^{-r}$ can easily be seen to be satisfied for all t if $r > 0$ and $M \geq \frac{(\beta+1)^{\alpha+1}\alpha}{r\Gamma(\alpha)}$. Similarly, the inequality $\left|\frac{(\beta+1)^{\alpha+1}\beta}{\Gamma(\alpha+1)}(-\log t)^{\alpha} t^{\beta-1} - \frac{(\beta+1)^{\alpha+1}\alpha}{\Gamma(\alpha+1)} t^{\beta-1}(-\log t)^{\alpha-1}\right| \leq M t^{-p-r}$ is satisfied when $p + r > 1$ and $M \geq \max\left(\frac{(\beta+1)^{\alpha+1}\alpha}{\Gamma(\alpha)}, \frac{(\beta+1)^{\alpha+1}\alpha(\beta+\alpha-2}{(p+r)\Gamma(\alpha)}\right)$.
When $\alpha = 0$ and $\beta \geq 0$ we have that the inequality $\left|(\beta+1)t^{\beta}\right| \leq M t^{-r}$ is satisfied for any $0 \leq r < \frac{1}{2}$ and $M \geq \beta + 1$. Also, the inequality $\left|(\beta+1)\beta t^{\beta-1}\right| \leq M t^{-p-r}$ is satisfied for all t if $p + r > 1$ and $M \geq \beta(\beta+1)$.

Chapter 5

Simulation study

Simulation experiments play an important role in testing how an estimator performs. This is because simulating some data allows us to test the finite sample size behaviour of the estimator on data where we know the true value of the parameter. We start this chapter with an introduction to Archimedian copulas and present an algorithm for generating random vectors from Archimedian copulas. Then we present three examples of Archimedian copulas for which we show that they satisfy Assumption 4.1.4 and derive the true values of the first order parameter η and the second order parameter τ. We then generate data from these copulas and examine how well the two estimators $\tilde{\tau}_k$ and $\hat{\tau}_k$ perform in the estimation of τ, and how well the estimator $\hat{\eta}_m$ performs in the estimation of η. In the estimation of η we use bias correction. Finally we examine whether or not the estimators $\hat{\chi}(u)$ and $\hat{\bar{\chi}}(u)$ can be used to estimate the values of χ and $\bar{\chi}$, if we choose a large threshold u.

5.1 Copula examples and simulation of data

We start this section with an introduction to a certain class of copulas known as Archimedian copulas.

Definition 5.1.1. Let ϕ be a continuous, strictly decreasing function from $[0,1] \to [0,\infty]$ such that $\phi(1) = 0$. The pseudo-inverse of ϕ is the function $\phi^{[-1]} : [0,\infty] \to [0,1]$ given by

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t), & 0 \leq t \leq \phi(0), \\ 0, & \phi(0) \leq t \leq \infty. \end{cases} \tag{5.1}$$

From this definition we can construct the class of Archimedian copulas. This is done in Theorem 5.1.2.

Theorem 5.1.2. *(Alsina et al., 2005) Let ϕ be a continuous, strictly decreasing function from $[0,1] \to [0,\infty]$ such that $\phi(1) = 0$, and let $\phi^{[-1]}$ be the pseudo-inverse of ϕ defined by (5.1). Then the function $C : [0,1]^2 \to [0,1]$ given by*

$$C(u,v) = \phi^{[-1]}(\phi(u) + \phi(v)) \tag{5.2}$$

is a copula if and only if ϕ is convex.

The function ϕ is called the generator function, since it is used to generate a copula.

Definition 5.1.3. For a convex, continuous and strictly decreasing function $\phi : [0,1] \to [0,\infty]$, we define the pseudo inverse of ϕ' as the function $\phi'^{[-1]} : [-\infty, 0] \to [0,1]$ given by

$$\phi'^{[-1]}(t) = \begin{cases} \phi'^{(-1)}(t), & \phi'(0) \leq t \leq \phi'(1), \\ 0, & \text{otherwise.} \end{cases} \tag{5.3}$$

It is straightforward to simulate data from an Archimedian copula as can be seen from Algorithm 5.1.4.

Algorithm 5.1.4. (Genest and MacKay, 1986b) Algorithm for generating a pair of uniform $(0,1)$ random variates with a copula that has generator function ϕ:

(i) Generate two independent uniform $(0,1)$ variates u and t.

(ii) Set $w = \phi'^{[-1]}\left(\frac{\phi'(u)}{t}\right)$.

(iii) Set $v = \phi^{[-1]}(\phi(w) - \phi(u))$.

(iv) The desired pair is (u, v).

Theorem 5.1.5. *If ϕ is a function that satisfies the assumptions in Theorem 5.1.2, then Algorithm 5.1.4 produces two random variates from the copula given in* (5.2).

Proof. Let $U, T \sim U(0,1)$ be independent and consider

$$P(U \leq u, V \leq v) = P((U,T) \in A),$$

where

$$A := \left\{(z,t) : z \leq u \text{ and } \phi^{[-1]}\left(\phi\left(\phi'^{[-1]}\left(\frac{\phi'(z)}{t}\right)\right) - \phi(z)\right) \leq v\right\}.$$

We need to show that $P(U \leq u, V \leq v) = C(u,v)$. Using that ϕ is a strictly decreasing function we find that

$$A = \left\{(z,t) : z \leq u \text{ and } \phi\left(\phi'^{[-1]}\left(\frac{\phi'(z)}{t}\right)\right) \geq \phi(v) + \phi(z)\right\}.$$

The constraint $\phi\left(\phi'^{[-1]}\left(\frac{\phi'(z)}{t}\right)\right) \geq \phi(v) + \phi(z)$ implies that $\phi(0) \geq \phi(v) + \phi(z)$. Hence $z \geq \phi^{-1}(\phi(0) - \phi(v))$ and thus

$$A = \left\{(z,t) : \phi^{-1}(\phi(0) - \phi(v)) \leq z \leq u \text{ and } \phi\left(\phi'^{[-1]}\left(\frac{\phi'(z)}{t}\right)\right) \geq \phi(v) + \phi(z)\right\}.$$

If $u < \phi^{-1}(\phi(0) - \phi(v))$ then $A = \emptyset$ which means that $P(U \leq u, V \leq v) = 0$. Otherwise it follows that

$$A = \left\{(z,t) : \phi^{-1}(\phi(0) - \phi(v)) \leq z \leq u \text{ and } t \leq \frac{\phi'(z)}{\phi'\left(\phi^{-1}(\phi(v) + \phi(z))\right)}\right\}$$

and hence

$$\begin{aligned} P((U,T) \in A) &= \int_{\phi^{-1}(\phi(0)-\phi(v))}^{u} \int_{0}^{\frac{\phi'(z)}{\phi'(\phi^{-1}(\phi(v)+\phi(z)))}} 1dtdz \\ &= \int_{\phi^{-1}(\phi(0)-\phi(v))}^{u} \frac{\phi'(z)}{\phi'\left(\phi^{-1}(\phi(v)+\phi(z))\right)} dz. \end{aligned}$$

Since

$$\frac{\partial}{\partial z} C(z,v) = \frac{\phi'(z)}{\phi'\left(\phi^{-1}(\phi(v)+\phi(z))\right)}$$

for $z > \phi^{-1}(\phi(0)-\phi(v))$ we get that

$$P((U,T) \in A) = C(u,v) - C\left(\phi^{-1}(\phi(0)-\phi(v)), v\right).$$

Clearly

$$\begin{aligned} C\left(\phi^{-1}(\phi(0)-\phi(v)), v\right) &= \phi^{[-1]}\left(\phi\left(\phi^{-1}(\phi(0)-\phi(v))\right) + \phi(v)\right) \\ &= 0, \end{aligned}$$

which finishes the proof. □

One of the examples of Archimedian copulas we will consider which satisfies Assumption 4.1.4, has a singular component, i.e. it is not absolutely continuous. In Theorem 5.1.6 we present a nice description of the singular component of an Archimedian copula.

Theorem 5.1.6. *(Genest and MacKay, 1986a) The copula $C(u,v)$ generated by a generator function ϕ has a singular component if and only if $\frac{\phi(0)}{\phi'(0)} \neq 0$. In that case, $\phi(U)+\phi(V) = \phi(0)$ with probability $-\frac{\phi(0)}{\phi'(0)}$.*

Next we present three examples of copulas that are Archimedian and satisfy Assumption 4.1.4. Since we restrict ourselves to bivariate distribution functions with continuous margins in Assumption 4.1.4, it suffices to verify that the copulas satisfy the assumption. This follows because we can transform the margins to any other continuous distribution without altering the dependence structure. In all three examples, we assume that the marginal distributions are continuous.

Example 5.1.7. The Ali-Mikhail-Haq copula has generator function

$$\phi(t) = \log\left(\frac{1-(1-\theta)t}{t}\right), \quad -1 \leq \theta \leq 1, 0 \leq t \leq 1$$

and hence has copula function given by

$$C(u,v) := \frac{uv}{1-\theta(1-u)(1-v)}, \quad -1 \leq \theta \leq 1, 0 \leq u,v \leq 1.$$

From this we find that the survival copula is

$$\bar{C}(u,v) := 1-u-v+\frac{uv}{1-\theta(1-u)(1-v)}, \quad -1 \leq \theta \leq 1, 0 \leq u,v \leq 1.$$

If $\theta = 0$, then the Ali-Mikhail-Haq copula is just the product copula. This copula does not have any second order behaviour, so we have to exclude the case $\theta = 0$.
Since F_X and F_Y are continuous, we have that $F_X(X)$ and $F_Y(Y)$ are uniformly distributed. This implies that

$$\begin{aligned} P\left(1 - F_X(X) < tx, 1 - F_Y(Y) < ty\right) &= \bar{C}(1 - tx, 1 - ty) \\ &= t^2 xy \frac{1 + \theta - \theta t(x+y)}{1 - \theta t^2 xy}, \quad 0 < tx, ty < 1. \end{aligned}$$

By making a Taylor series expansion of $\frac{1}{1-z}$ around 0 we find that

$$\begin{aligned} P\left(1 - F_X(X) < tx, 1 - F_Y(Y) < ty\right) =& t^2 xy\big(1 + \theta - \theta t(x+y) \\ &+ (1+\theta)\theta t^2 xy - \theta^2 t^3 xy(x+y) + O\left(t^4\right)\big). \end{aligned}$$

We have to consider the cases $\theta = -1$ and $-1 < \theta < 1$ seperately. In the case $-1 < \theta \leq 1$ we get

$$\frac{P\left(1 - F_X(X) < tx, 1 - F_Y(Y) < ty\right)}{P\left(1 - F_X(X) < t, 1 - F_Y(Y) < t\right)} = xy \frac{1 + \theta - \theta t(x+y) + O\left(t^2\right)}{1 + \theta - 2\theta t + O\left(t^2\right)},$$

which by a Taylor series expansion is found to be

$$\frac{P\left(1 - F_X(X) < tx, 1 - F_Y(Y) < ty\right)}{P\left(1 - F_X(X) < t, 1 - F_Y(Y) < t\right)} = xy \left(1 - t\frac{\theta}{1+\theta}\left(x + y - 2\right) + O(t^2)\right).$$

From this we easily find that Assumption 4.1.4 is satisfied with $c(x,y) = xy$, $\eta = \frac{1}{2}$, $q_1(t) = -2\frac{\theta}{(\theta+1)}t$, $\tau = 1$ and $c_1(x,y) = \frac{1}{2}xy(x+y-2)$.
In the case where $\theta = -1$ we find that

$$\frac{P\left(1 - F_X(X) < tx, 1 - F_Y(Y) < ty\right)}{P\left(1 - F_X(X) < t, 1 - F_Y(Y) < t\right)} = xy \frac{x + y - t^2 xy(x+y) + O\left(t^3\right)}{2 - 2t^2 + O\left(t^3\right)}$$

By making a Taylor series expansion it follows that

$$\frac{P\left(1 - F_X(X) < tx, 1 - F_Y(Y) < ty\right)}{P\left(1 - F_X(X) < t, 1 - F_Y(Y) < t\right)} = xy \left(\frac{x+y}{2} - t^2 \frac{xy(x+y) - (x+y)}{2} + O(t^3)\right).$$

Hence, Assumption 4.1.4 is satisfied with $c(x,y) = \frac{xy(x+y)}{2}$, $\eta = \frac{1}{3}$, $q_1(t) = -2t^2$, $\tau = 2$ and $c_1(x,y) = \frac{1}{4}xy(xy(x+y) - (x+y))$.

■

Example 5.1.8. The Gumbel-Hougaard copula has generator function

$$\phi(t) = \log\left(1 - \theta \log(t)\right), \quad 0 \leq \theta \leq 1, 0 \leq t \leq 1$$

and hence has a copula function given by

$$C(u,v) = uv \exp(-\theta \log u \log v), \quad 0 < \theta \leq 1, 0 \leq u, v \leq 1,$$

The survival copula is easily found to be

$$\bar{C}(u,v) = 1 - u - v + uv \exp(-\theta \log u \log v).$$

Assuming that F_X and F_Y are continuous it follows that

$$\begin{aligned} P\left(1-F_X(X)<tx, 1-F_Y(Y)<ty\right) =& \bar{C}(1-tx,1-ty) \\ =& -1+tx+ty+(1-tx)(1-ty) \\ & \cdot \exp(-\theta\log(1-tx)\log(1-ty)). \end{aligned}$$

If we expand this through a Taylor series, we get

$$\begin{aligned} P\left(1-F_X(X)<tx, 1-F_Y(Y)<ty\right) = t^2xy\Bigg(& 1-\theta+\frac{1}{2}\theta t(x+y) \\ & +\theta\left(\frac{x^2+y^2}{6}-\frac{1}{4}xy+\frac{1}{2}\theta xy\right)t^2+O\left(t^3\right)\Bigg). \end{aligned}$$

We start by considering the case $0<\theta<1$, in which case

$$\frac{P\left(1-F_X(X)<tx, 1-F_Y(Y)<ty\right)}{P\left(1-F_X(X)<t, 1-F_Y(Y)<t\right)} = xy\left(1+\frac{\theta}{1-\theta}\frac{1}{2}(x+y-2)t+O\left(t^2\right)\right)$$

is obtained through a Taylor series expansion. From this, we find that Assumption 4.1.4 is satisfied with $c(x,y)=xy$, $\eta=\frac{1}{2}$, $q_1(t)=\frac{\theta}{1-\theta}t$, $\tau=1$ and $c_1(x,y)=\frac{1}{2}xy(x+y-2)$.
In the case where $\theta=1$ we get

$$\frac{P\left(1-F_X(X)<tx, 1-F_Y(Y)<ty\right)}{P\left(1-F_X(X)<t, 1-F_Y(Y)<t\right)} = xy\left(\frac{x+y}{2}+\left(\frac{x^2+y^2}{6}+\frac{1}{4}xy-\frac{7x+7y}{24}\right)t+O\left(t^2\right)\right).$$

Hence Assumption 4.1.4 is satisfied with $c(x,y)=xy\frac{x+y}{2}$, $\eta=\frac{1}{3}$, $q_1(t)=\frac{7}{12}t$, $\tau=1$ and $c_1(x,y)=\frac{1}{14}xy(4x^2+4y^2+6xy-7x-7y)$.

■

Example 5.1.9. The next copula we consider has generator function

$$\phi(t)=\frac{1-t}{1+(\theta-1)t}$$

and is given by

$$C(u,v)=\max\left(\frac{\theta^2uv-(1-u)(1-v)}{\theta^2-(\theta-1)^2(1-u)(1-v)},0\right), \quad 1\le\theta<\infty, 0\le u,v\le 1.$$

We will refer to this copula as Nelsen (4.2.8) because our only reference to this copula is from (Nelsen, 2006). From this it follows that the survival copula is

$$\bar{C}(u,v)=1-u-v+\max\left(\frac{\theta^2uv-(1-u)(1-v)}{\theta^2-(\theta-1)^2(1-u)(1-v)},0\right).$$

This is an example of a distribution which is not absoutely continuous, even though it has continuous margins. In fact it follows from Theorem 5.1.6 that a point (U,V) has probability $\frac{1}{\theta}$ of lying on the arc defined by $(\theta-1)^2uv+2(\theta-1)uv+u+v-1=0$.
In the case where $\theta=1$, the copula reduces to $C(u,v)=\max(u+v-1,0)$, which does not

have any second order behaviour, and hence we have to exclude this case. Assuming that F_X and F_Y are continuous it follows that

$$\begin{aligned}P\left(1-F_X(X)<tx,1-F_Y(Y)<ty\right)=&\bar{C}(1-tx,1-ty)\\=&-1+tx+ty+\max\left(\frac{\theta^2(1-tx)(1-ty)-t^2xy}{\theta^2-(\theta-1)^2t^2xy},0\right).\end{aligned}$$

Since we are only interested in what happens when $t\to 0$, we can assume that

$$P\left(1-F_X(X)<tx,1-F_Y(Y)<ty\right)=-1+tx+ty+\frac{\theta^2(1-tx)(1-ty)-t^2xy}{\theta^2-(\theta-1)^2t^2xy}.$$

If we expand this through a Taylor series, we get

$$P\left(1-F_X(X)<tx,1-F_Y(Y)<ty\right)=t^2xy\left(2\frac{\theta-1}{\theta}-\frac{(\theta-1)^2}{\theta^2}(x+y)t+O\left(t^2\right)\right),$$

and hence

$$\frac{P\left(1-F_X(X)<tx,1-F_Y(Y)<ty\right)}{P\left(1-F_X(X)<t,1-F_Y(Y)<t\right)}=xy\left(1-\frac{\theta-1}{\theta}\frac{1}{2}(x+y-2)t+O\left(t^2\right)\right)$$

is obtained through a Taylor series expansion. From this, we find that Assumption 4.1.4 is satisfied with $c(x,y)=xy$, $\eta=\frac{1}{2}$, $q_1(t)=-\frac{\theta-1}{\theta}t$, $\tau=1$ and $c_1(x,y)=\frac{1}{2}xy(x+y-2)$.

■

5.2 Estimation of the second order parameter τ

Using the copulas discussed in Example 5.1.7, Example 5.1.8 and Example 5.1.9 we examine the finite sample size behaviour of the estimators we have proposed in Chapter 4. For each copula we generated 1000 samples of size 5000 and 50000, and computed the estimators $\tilde{\tau}_k$ and $\hat{\tau}_k$ for $k=10,20,\ldots,5000$ and $k=100,200,\ldots,50000$, respectively. For each value of k, the mean and the mean squared error is estimated. The estimation of τ does not depend on the marginal distributions but only on the copula. This means that there is no need to try the estimation of τ using different marginal distributions. Concerning the estimator $\hat{\tau}_k\left(x,y,\tilde{\eta}\right)$ we start by finding an estimate for η. This is done using the bias corrected weight function described in Theorem 4.1.10 where we fix the value $\eta\tau=1$. We then use this value to estimate m_1 and m_2 as discussed in Chapter 4. We also fix the values of x and y at $x=y=\frac{1}{8}$. For the estimator $\tilde{\tau}$ we fix the value of z at $z=\frac{1}{16}$. The choice of the value of the tuning parameter a in $\hat{\tau}_k$ is found to give the best results when a is chosen small. We have therefore chosen to present the results for $a=0.01,0.02,0.03,0.04$. The estimator $\tilde{\tau}_k$ also performs the best when a, b and l are chosen small.
The estimators $\hat{\tau}_k$ and $\tilde{\tau}_k$ are examined using the Ali-Mikhail-Haq copula in Figure 5.1 - Figure 5.4, using the Gumbel-Hougaard copula in Figure 5.5 - Figure 5.8 and using the Nelsen (4.2.8) copula in Figure 5.9 - Figure 5.12. From these Figures we can draw the following conclusions.

(i) The estimator $\tilde{\tau}_k$ has some severe problems with bias and as a result of this, the MSE of this estimator is also very high. The estimator $\hat{\tau}_k$ does not have anywhere near the

same bias issues as $\tilde{\tau}_k$. In most of the cases we have explored this estimator has a stable region where there is a relatively small bias. The estimator does however have no stable region in the case of the Gumbel-Hougaard copula with $\theta = 1$ and the case of the Nelsen (4.2.8) copula with $\theta = \frac{3}{2}$. In the Gumbel-Hougaard case this might be because we are in the limiting case of the copula, and in Nelsen (4.2.8) case it might be because a very large portion of the data lie in the singular component of the copula.

(ii) When sample size increases from 5000 to 50000 there is a clear improvement of the estimators performance. In some of the cases $\tilde{\tau}_k$ reduces its bias and has a stable region, but the MSE still remains quite high. For $\hat{\tau}_k$ we see that the problematic case of the Nelsen (4.2.8) copula with $\theta = \frac{3}{2}$ the estimation is improved and that there is a small stable region.

(iii) Concerning the estimator $\hat{\tau}_k$, the stable regions, with no or little bias does not always contain the same values of k. Furthermore, the stable regions does on some occasions have a fairly large MSE, meaning that the variance is very high. This causes some problems, since this means that it is more difficult to choose a region for k in a situation where real data is used.

5.3 Estimation of the first order parameter η

For the copulas considered in Example 5.1.7, Example 5.1.8 and Example 5.1.9 we generated 1000 samples of size 5000 and 50000, and computed $T_{K_{\alpha_{\mathrm{opt}}}}$ for $m = 10, 20, \ldots, 5000$ and $m = 100, 200, \ldots, 50000$, respectively. In Figures 5.13 till 5.18 we show the sample mean (left) and the empirical mean squared error (right) as a function of m with initial estimates used for $\eta\tau$ (solid), $\eta\tau = 1$ (dashed) and the true value of $\eta\tau$ (dotted). In the case where initial estimates are used for the value of $\eta\tau$, the estimate of τ is the median of the estimator $\hat{\tau}_k$ with $k = \lfloor n^{0.95} \rfloor, \lfloor n^{0.95} \rfloor + 1, \ldots, \lfloor n^{0.975} \rfloor$. The estimate of η that is used is the median of $T_{K_{\alpha_{\mathrm{opt}}}}$ with $\eta\tau = 1$ and $m = 10, 11, \ldots, \frac{n}{4}$. If the estimate of $\eta\tau$ happens to be negative, then the absolute value is used. From the simulation results we are able to draw the following conclusions.

(i) The estimator $T_{K_{\alpha_{\mathrm{opt}}}}$ with initial estimates of η and τ has in general a performance which is far worse than using the true value of η and τ or using $\eta\tau = 1$. The cause of this is probably the negative bias of the estimator $\hat{\tau}_k$. When we get an estimate of τ which is close to zero, then the estimator $T_{K_{\alpha_{\mathrm{opt}}}}$ has a large variance, which causes the estimate of η to be very erratic.

(ii) The estimator $T_{K_{\alpha_{\mathrm{opt}}}}$ exhibit similar behaviour when true values of η and τ are used and when $\eta\tau = 1$, both in terms of bias and MSE. This implies that it is reasonable in practice to use $\eta\tau = 1$ instead of estimating τ for the bias corrected estimation of η.

5.4 Estimation of the dependence measures χ and $\bar{\chi}$

Using the copulas in Example 5.1.7, Example 5.1.8 and Example 5.1.9 we generated 1000 samples of size 5000 and 50000. For each copula we estimated the values of $\chi(u)$ and $\bar{\chi}(u)$

for $u = 0, 0.002, 0.004, \ldots, 1$ using the estimators in (3.29) and (3.31), respectively. This was done using the R-package "POT" by Ribatet (2006). In Figures 5.19 till Figure 5.21 we show the sample mean of $\hat{\chi}(u)$ and $\hat{\bar{\chi}}(u)$ for the sample size of 5000 (left) and 50000 (right). The true values of χ and $\bar{\chi}$ are shown with horizontal lines. We use the plots of the sample mean of $\hat{\chi}(u)$ and $\hat{\bar{\chi}}(u)$ to examine how well estimates of the functions $\chi(u)$ and $\bar{\chi}(u)$ can be used to obtain estimates of χ and $\bar{\chi}$, if the value of u is chosen wisely. From the results, we are able to draw the following conclusions.

(i) The estimator $\hat{\chi}(u)$ always gets very close to the true value of χ for large values of u. However, this might be because $\chi = 0$ for all the examples and $\hat{\chi}(u) = 0$ for u large enough. But it seems that u has to be chosen fairly large, if the value of $\hat{\chi}(u)$ should be close to the true value of χ. The estimator $\hat{\bar{\chi}}(u)$ converges in all cases to -1, when u tends to 1. For values of u, that are smaller than 1 it seems that we get an estimate of $\bar{\chi}$, which also suffers from some bias.

(ii) For the estimator $\hat{\chi}(u)$ there is very little difference between a sample size of 5000 and 50000. The estimator $\hat{\bar{\chi}}(u)$ behaves better when the sample size is large, since the degeneracy towards the value -1 happens later. The bias, is however roughly the same for the two sample sizes before the degeneracy occurs.

(iii) The estimation of $\bar{\chi}$ through $\hat{\bar{\chi}}(u)$ seems to be worse than estimating η with the bias corrected estimator for η, and then calculating $\bar{\chi}$ through the relation $\bar{\chi} = 2\eta - 1$.

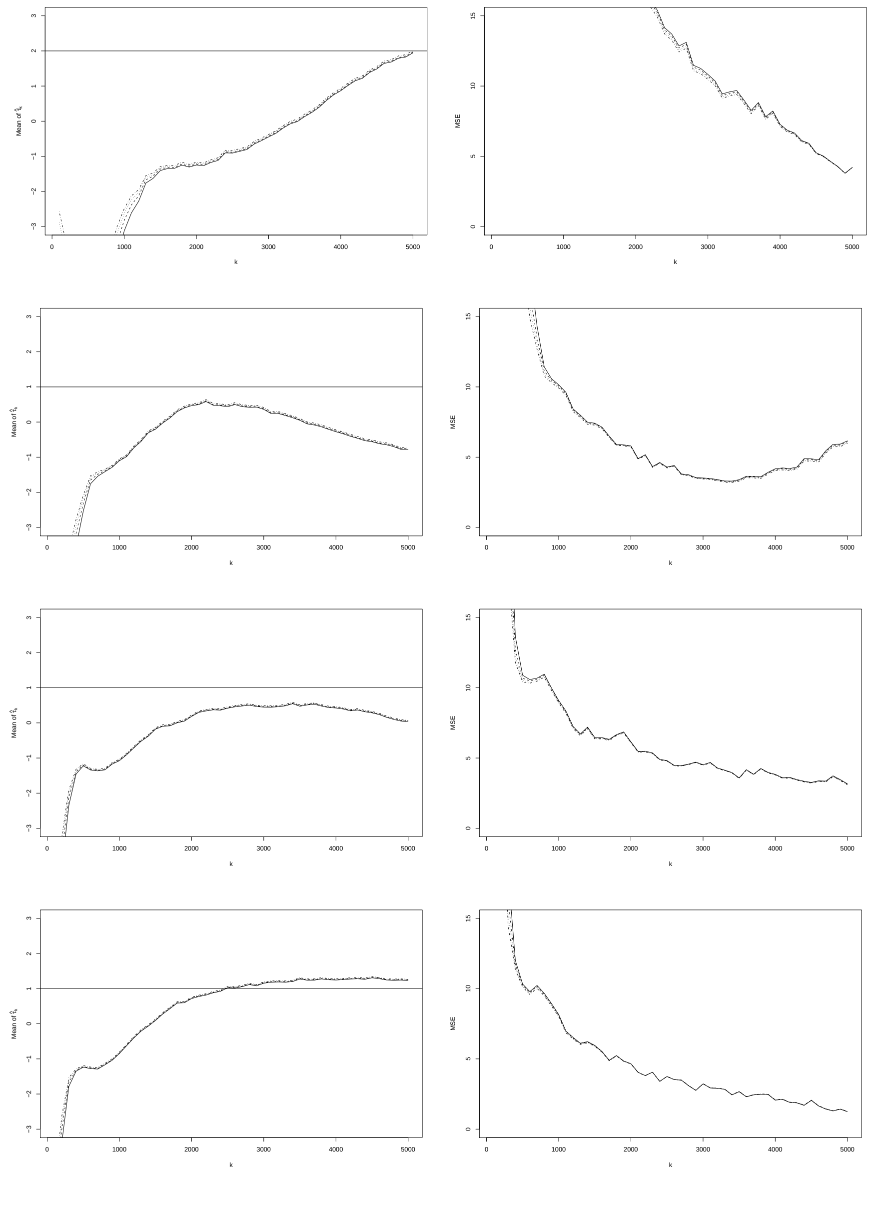

Figure 5.1: Simulation of $\hat{\tau}_k$ with sample size of 5000 for different values of k for the Ali-Mikhail-Haq copula with $\theta = -1$ (row 1), $\theta = -0.5$ (row 2), $\theta = 0.5$ (row 3) and $\theta = 1$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The tuning parameter a in the model is $a = 0.01$ (black, solid), $a = 0.02$ (black, dashed), $a = 0.03$ (black, dashed-dotted), $a = 0.04$ (black, dotted). The horizontal black solid line is the true value of τ.

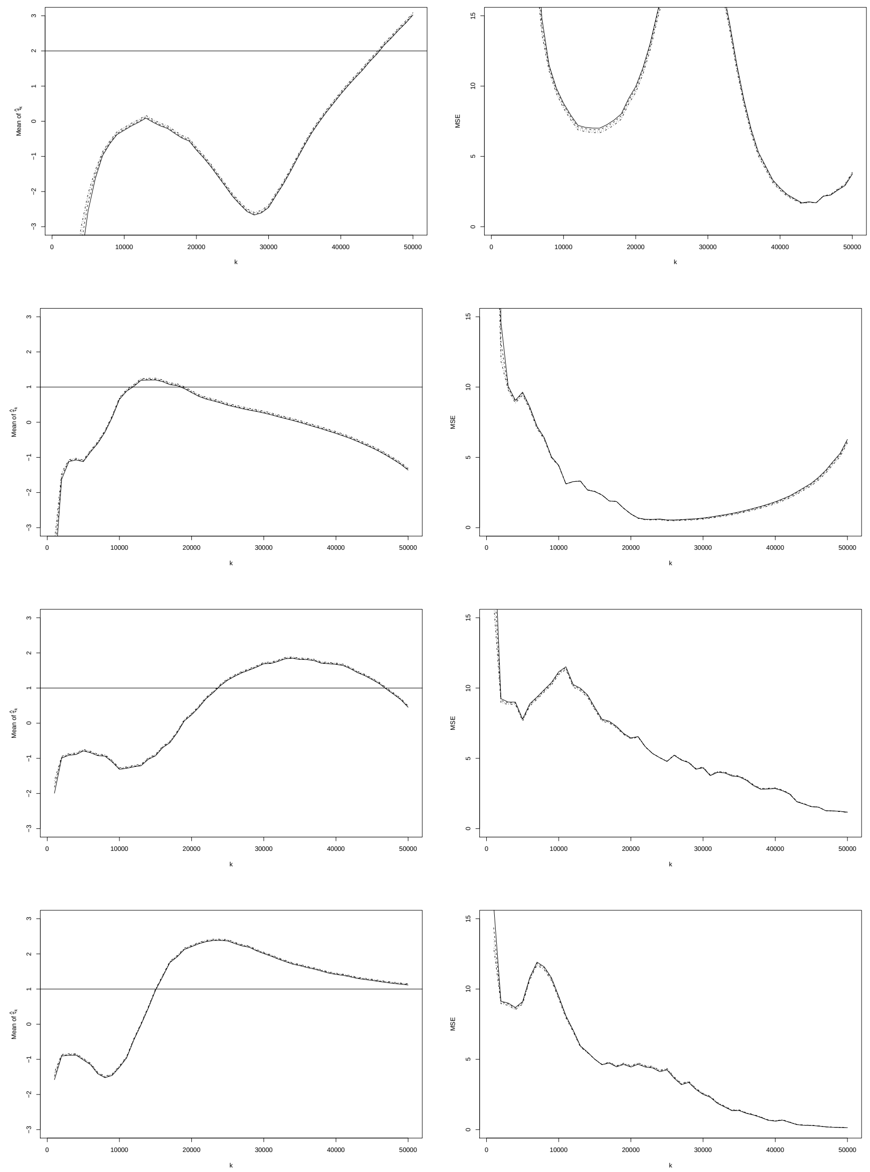

Figure 5.2: Simulation of $\hat{\tau}_k$ with sample size of 50000 for different values of k for the Ali-Mikhail-Haq copula with $\theta = -1$ (row 1), $\theta = -0.5$ (row 2), $\theta = 0.5$ (row 3) and $\theta = 1$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The tuning parameter a in the model is $a = 0.01$ (black, solid), $a = 0.02$ (black, dashed), $a = 0.03$ (black, dashed-dotted), $a = 0.04$ (black, dotted). The horizontal black solid line is the true value of τ.

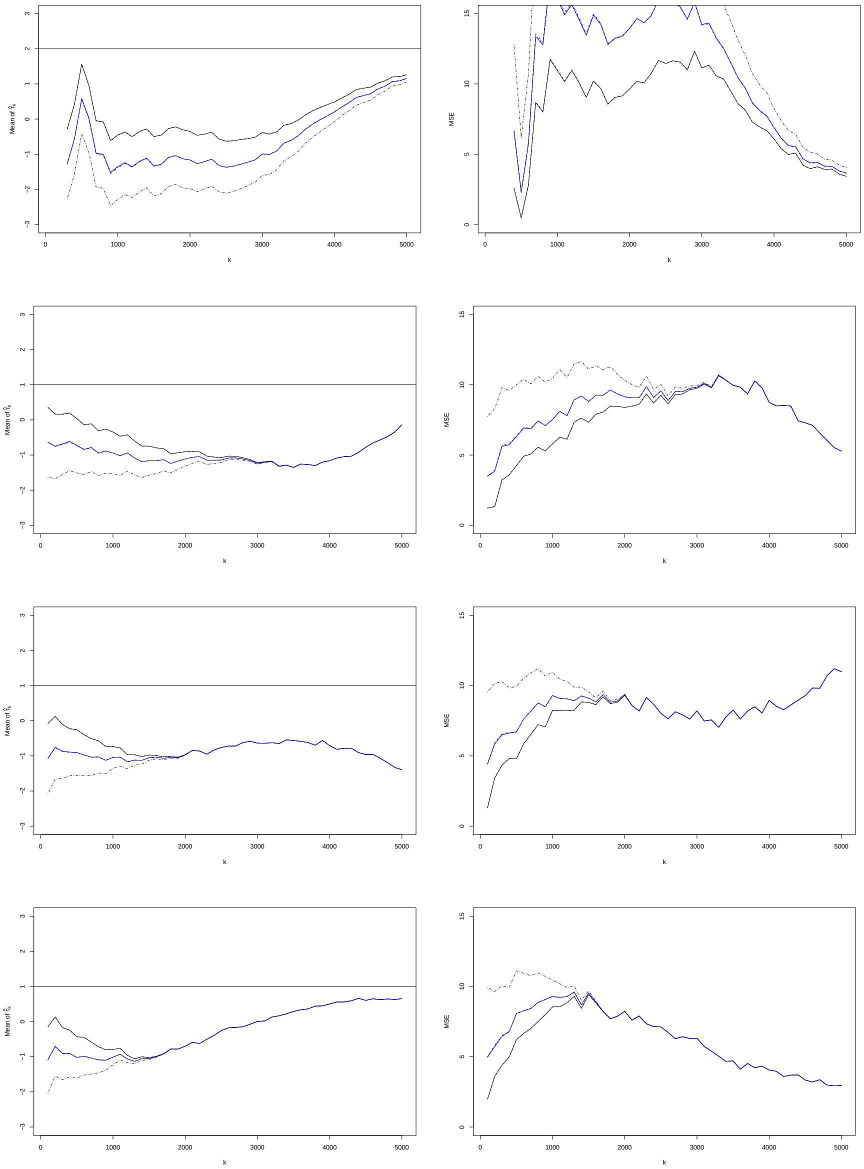

Figure 5.3: Simulation of $\tilde{\tau}_k$ with sample size of 5000 for different values of k for the Ali-Mikhail-Haq copula with $\theta = -1$ (row 1), $\theta = -0.5$ (row 2), $\theta = 0.5$ (row 3) and $\theta = 1$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The tuning parameters a, b and l in the model are $a = 0.01$, $b = 0.02$, $l = 0.01$ (black, solid), $a = 0.01$, $b = 0.02$, $l = 0.02$ (black, dashed), $a = 0.02$, $b = 0.01$, $l = 0.01$ (black, dashed-dotted), $a = 0.02$, $b = 0.01$, $l = 0.02$ (black, dotted), $a = b = 0.01$, $l = 0$ (blue, solid), $a = b = 0.02$, $l = 0$ (blue, dashed), $a = b = 0.03$, $l = 0$ (blue, dashed-dotted), $a = b = 0.04$, $l = 0$ (black, dotted). The horizontal black solid line is the true value of τ.

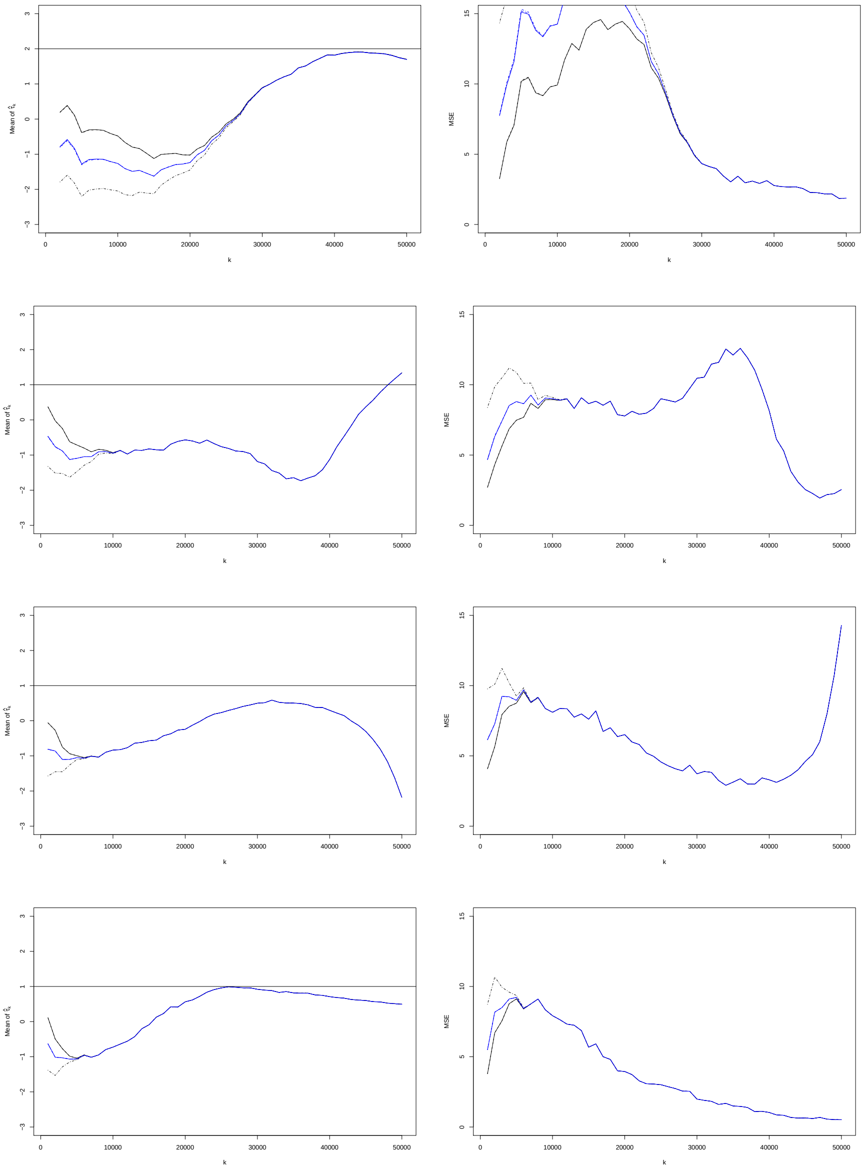

Figure 5.4: Simulation of $\tilde{\tau}_k$ with sample size of 50000 for different values of k for the Ali-Mikhail-Haq copula with $\theta = -1$ (row 1), $\theta = -0.5$ (row 2), $\theta = 0.5$ (row 3) and $\theta = 1$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The tuning parameters a, b and l in the model are $a = 0.01$, $b = 0.02$, $l = 0.01$ (black, solid), $a = 0.01$, $b = 0.02$, $l = 0.02$ (black, dashed), $a = 0.02$, $b = 0.01$, $l = 0.01$ (black, dashed-dotted), $a = 0.02$, $b = 0.01$, $l = 0.02$ (black, dotted), $a = b = 0.01$, $l = 0$ (blue, solid), $a = b = 0.02$, $l = 0$ (blue, dashed), $a = b = 0.03$, $l = 0$ (blue, dashed-dotted), $a = b = 0.04$, $l = 0$ (black, dotted). The horizontal black solid line is the true value of τ.

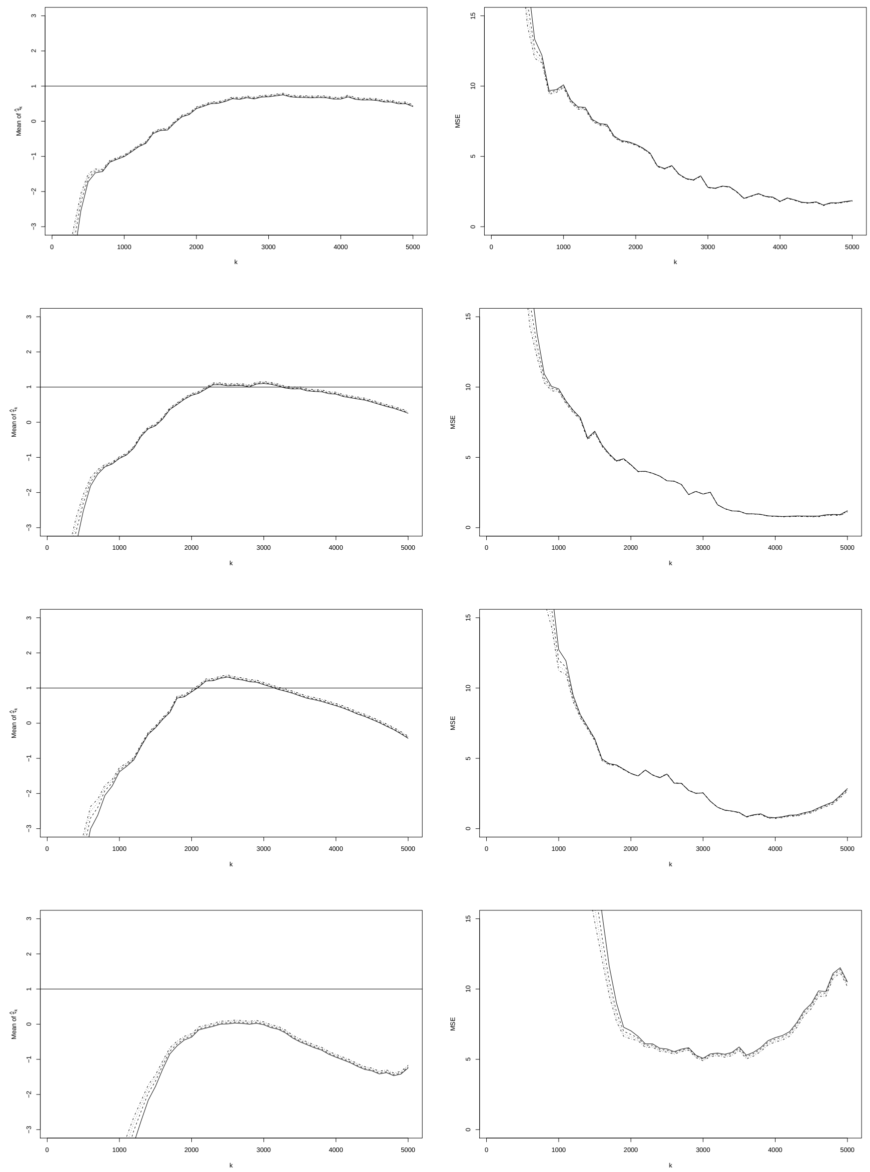

Figure 5.5: Simulation of $\hat{\tau}_k$ with sample size of 5000 for different values of k for the Gumbel-Hougaard copula with $\theta = 0.25$ (row 1), $\theta = 0.5$ (row 2), $\theta = 0.75$ (row 3) and $\theta = 1$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The tuning parameter a in the model is $a = 0.01$ (black, solid), $a = 0.02$ (black, dashed), $a = 0.03$ (black, dashed-dotted), $a = 0.04$ (black, dotted). The horizontal black solid line is the true value of τ.

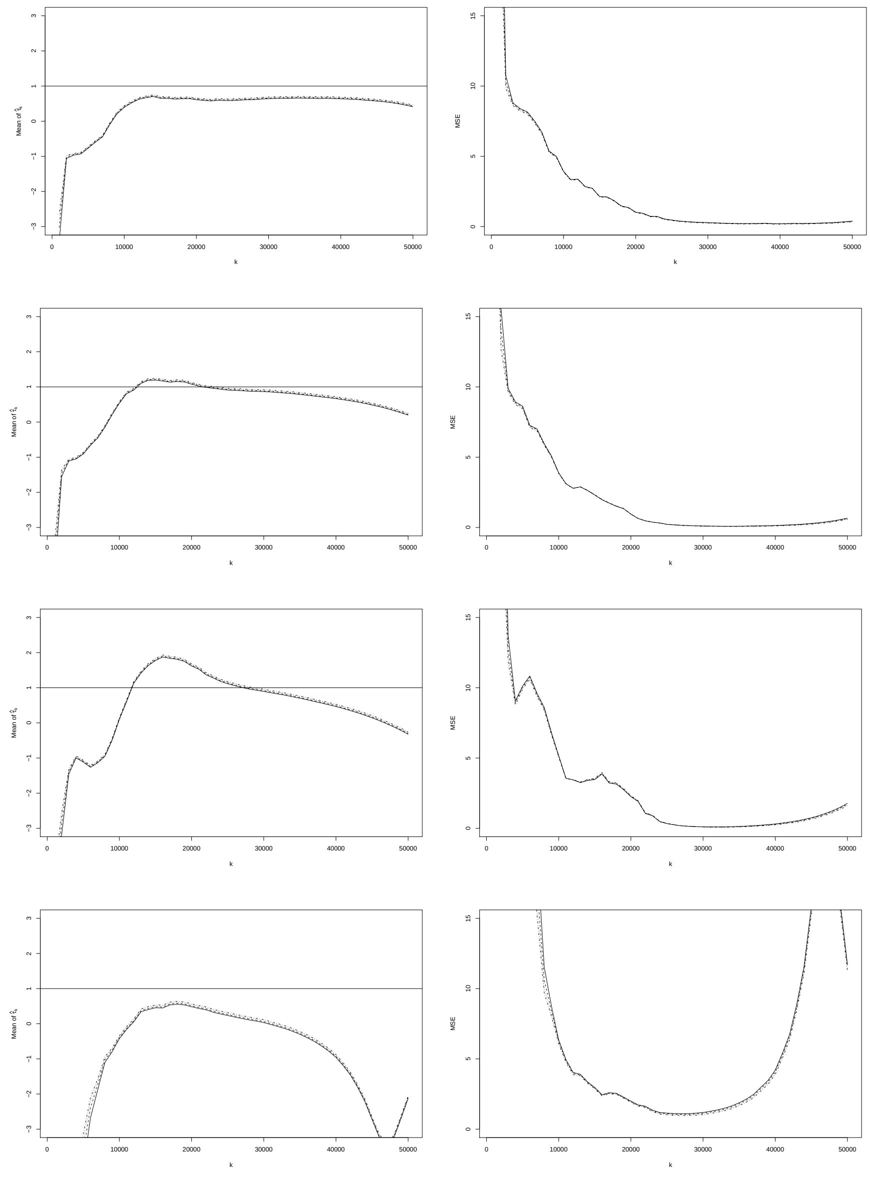

Figure 5.6: Simulation of $\hat{\tau}_k$ with sample size of 50000 for different values of k for the Gumbel-Hougaard copula with $\theta = 0.25$ (row 1), $\theta = 0.5$ (row 2), $\theta = 0.75$ (row 3) and $\theta = 1$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The tuning parameter a in the model is $a = 0.01$ (black, solid), $a = 0.02$ (black, dashed), $a = 0.03$ (black, dashed-dotted), $a = 0.04$ (black, dotted). The horizontal black solid line is the true value of τ.

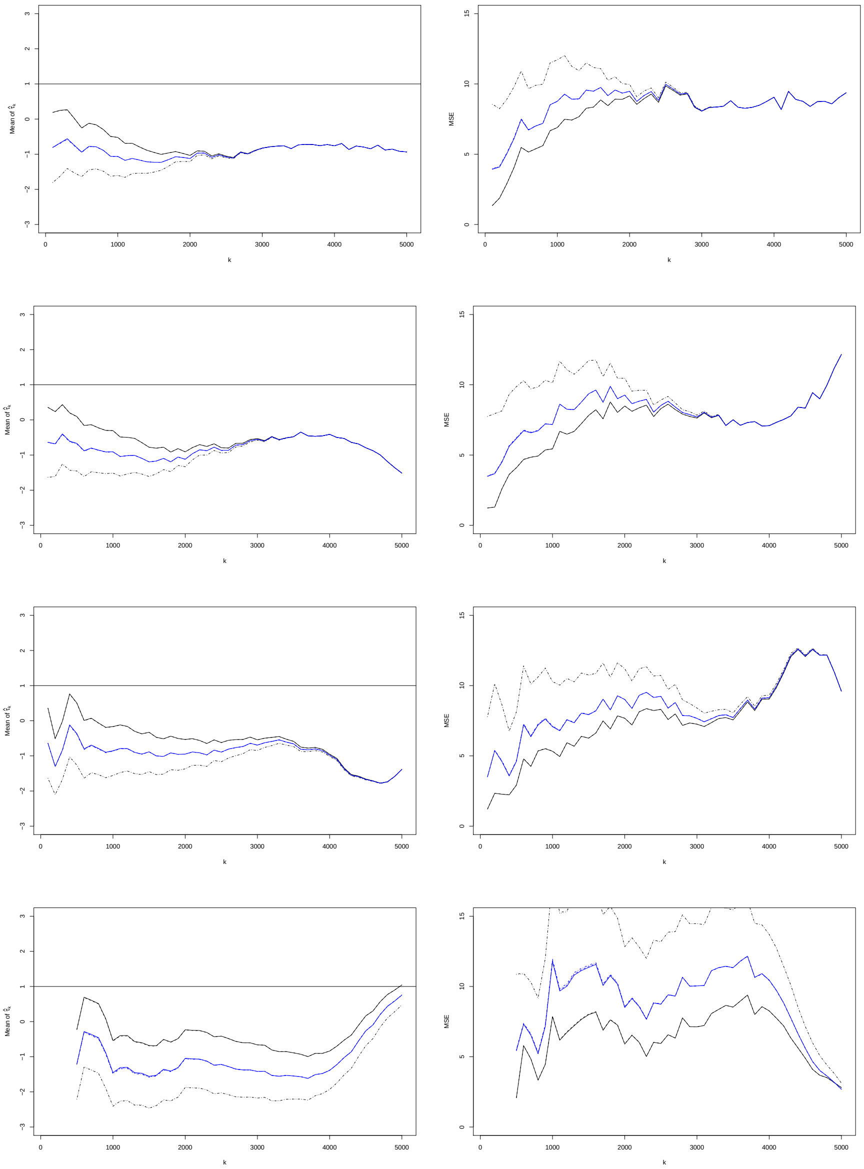

Figure 5.7: Simulation of $\tilde{\tau}_k$ with sample size of 5000 for different values of k for the Gumbel-Hougaard copula with $\theta = 0.25$ (row 1), $\theta = 0.5$ (row 2), $\theta = 0.75$ (row 3) and $\theta = 1$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The tuning parameters a, b and l in the model are $a = 0.01$, $b = 0.02$, $l = 0.01$ (black, solid), $a = 0.01$, $b = 0.02$, $l = 0.02$ (black, dashed), $a = 0.02$, $b = 0.01$, $l = 0.01$ (black, dashed-dotted), $a = 0.02$, $b = 0.01$, $l = 0.02$ (black, dotted), $a = b = 0.01$, $l = 0$ (blue, solid), $a = b = 0.02$, $l = 0$ (blue, dashed), $a = b = 0.03$, $l = 0$ (blue, dashed-dotted), $a = b = 0.04$, $l = 0$ (black, dotted). The horizontal black solid line is the true value of τ.

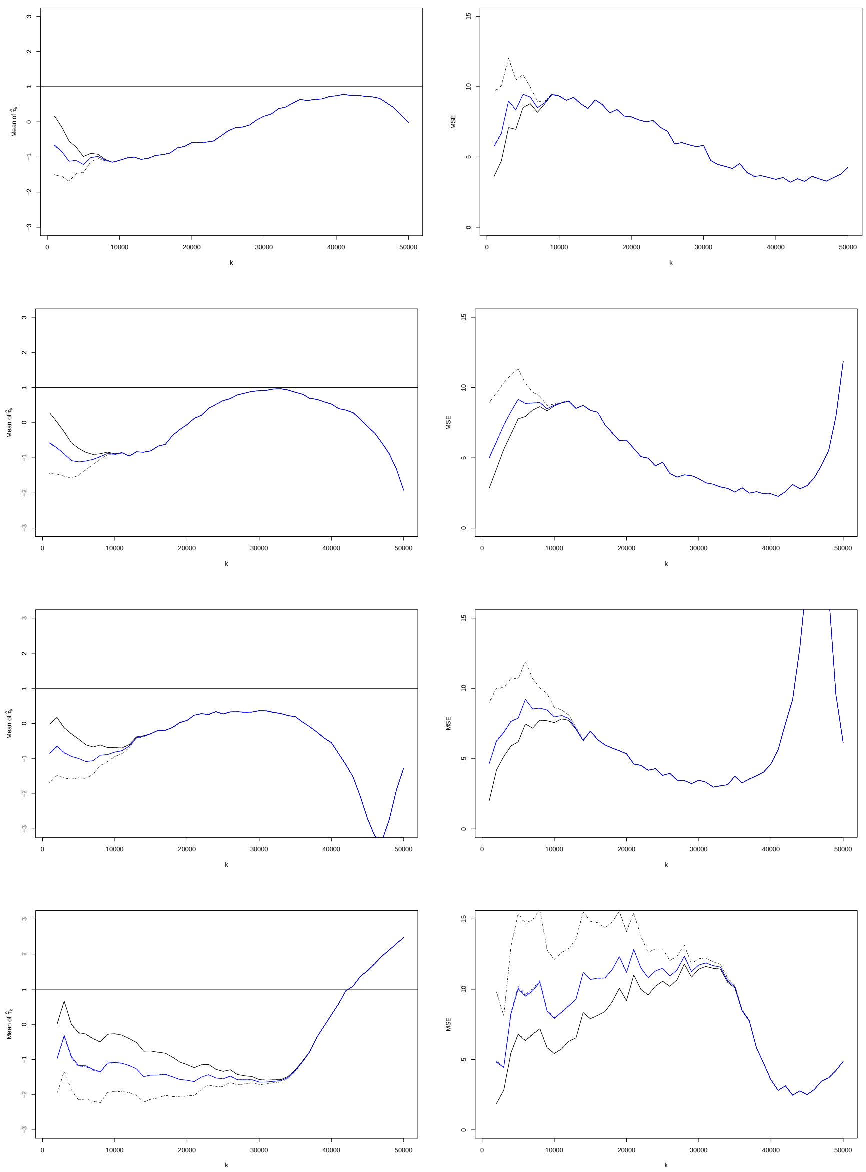

Figure 5.8: Simulation of $\tilde{\tau}_k$ with sample size of 50000 for different values of k for the Gumbel-Hougaard copula with $\theta = 0.25$ (row 1), $\theta = 0.5$ (row 2), $\theta = 0.75$ (row 3) and $\theta = 1$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The tuning parameters a, b and l in the model are $a = 0.01$, $b = 0.02$, $l = 0.01$ (black, solid), $a = 0.01$, $b = 0.02$, $l = 0.02$ (black, dashed), $a = 0.02$, $b = 0.01$, $l = 0.01$ (black, dashed-dotted), $a = 0.02$, $b = 0.01$, $l = 0.02$ (black, dotted), $a = b = 0.01$, $l = 0$ (blue, solid), $a = b = 0.02$, $l = 0$ (blue, dashed), $a = b = 0.03$, $l = 0$ (blue, dashed-dotted), $a = b = 0.04$, $l = 0$ (black, dotted). The horizontal black solid line is the true value of τ.

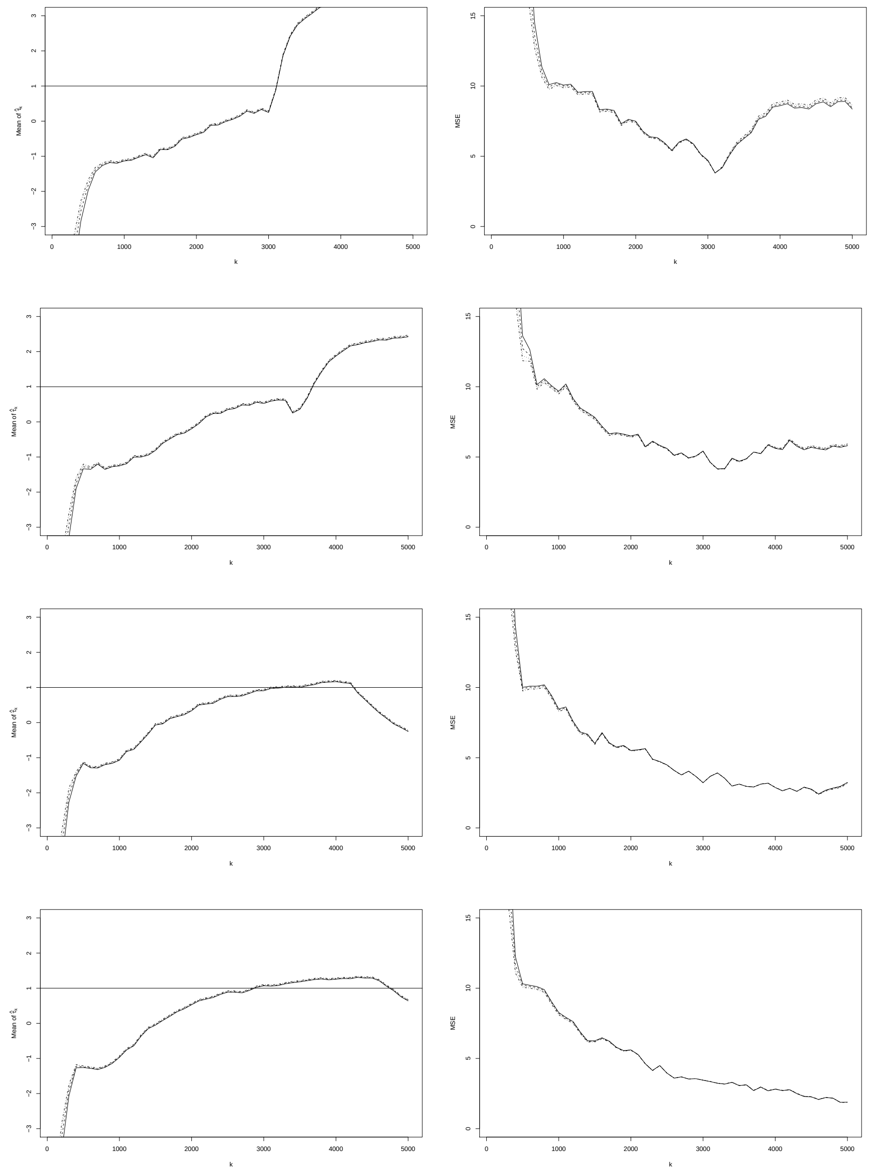

Figure 5.9: Simulation of $\hat{\tau}_k$ with sample size of 5000 for different values of k for the Nelsen (4.2.8) copula with $\theta = 1.5$ (row 1), $\theta = 2$ (row 2), $\theta = 5$ (row 3) and $\theta = 10$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The tuning parameter a in the model is $a = 0.01$ (black, solid), $a = 0.02$ (black, dashed), $a = 0.03$ (black, dashed-dotted), $a = 0.04$ (black, dotted). The horizontal black solid line is the true value of τ.

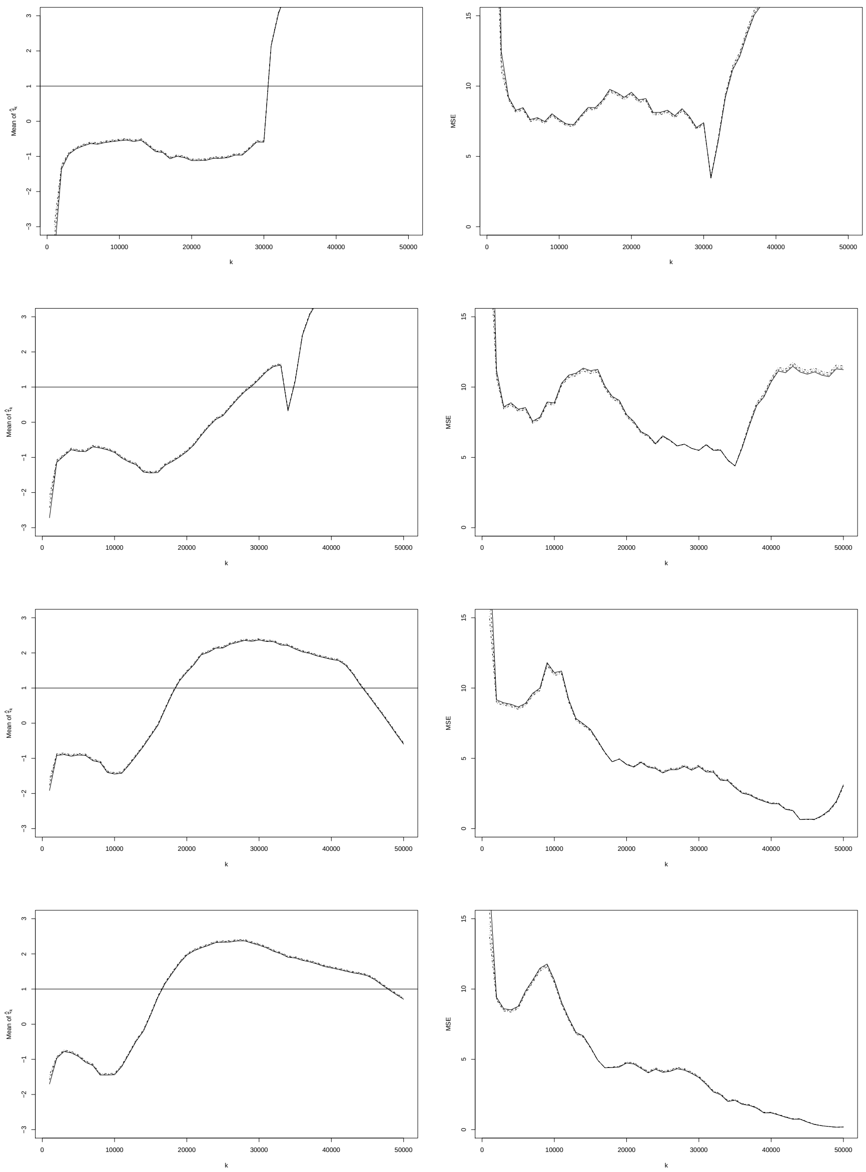

Figure 5.10: Simulation of $\hat{\tau}_k$ with sample size of 50000 for different values of k for the Nelsen (4.2.8) copula with $\theta = 1.5$ (row 1), $\theta = 2$ (row 2), $\theta = 5$ (row 3) and $\theta = 10$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The tuning parameter a in the model is $a = 0.01$ (black, solid), $a = 0.02$ (black, dashed), $a = 0.03$ (black, dashed-dotted), $a = 0.04$ (black, dotted). The horizontal black solid line is the true value of τ.

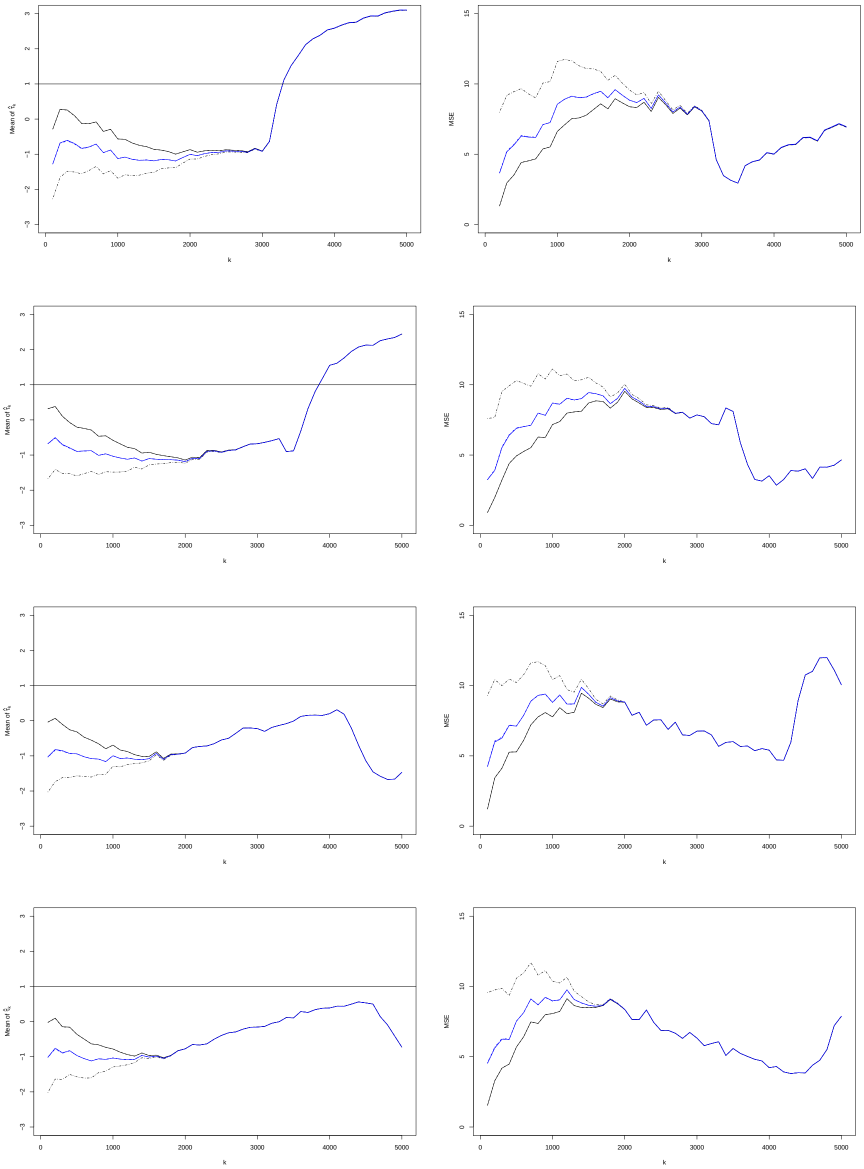

Figure 5.11: Simulation of $\tilde{\tau}_k$ with sample size of 5000 for different values of k for the Nelsen (4.2.8) copula with $\theta = 1.5$ (row 1), $\theta = 2$ (row 2), $\theta = 5$ (row 3) and $\theta = 10$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The tuning parameters a, b and l in the model are $a = 0.01$, $b = 0.02$, $l = 0.01$ (black, solid), $a = 0.01$, $b = 0.02$, $l = 0.02$ (black, dashed), $a = 0.02$, $b = 0.01$, $l = 0.01$ (black, dashed-dotted), $a = 0.02$, $b = 0.01$, $l = 0.02$ (black, dotted), $a = b = 0.01$, $l = 0$ (blue, solid), $a = b = 0.02$, $l = 0$ (blue, dashed), $a = b = 0.03$, $l = 0$ (blue, dashed-dotted), $a = b = 0.04$, $l = 0$ (black, dotted). The horizontal black solid line is the true value of τ.

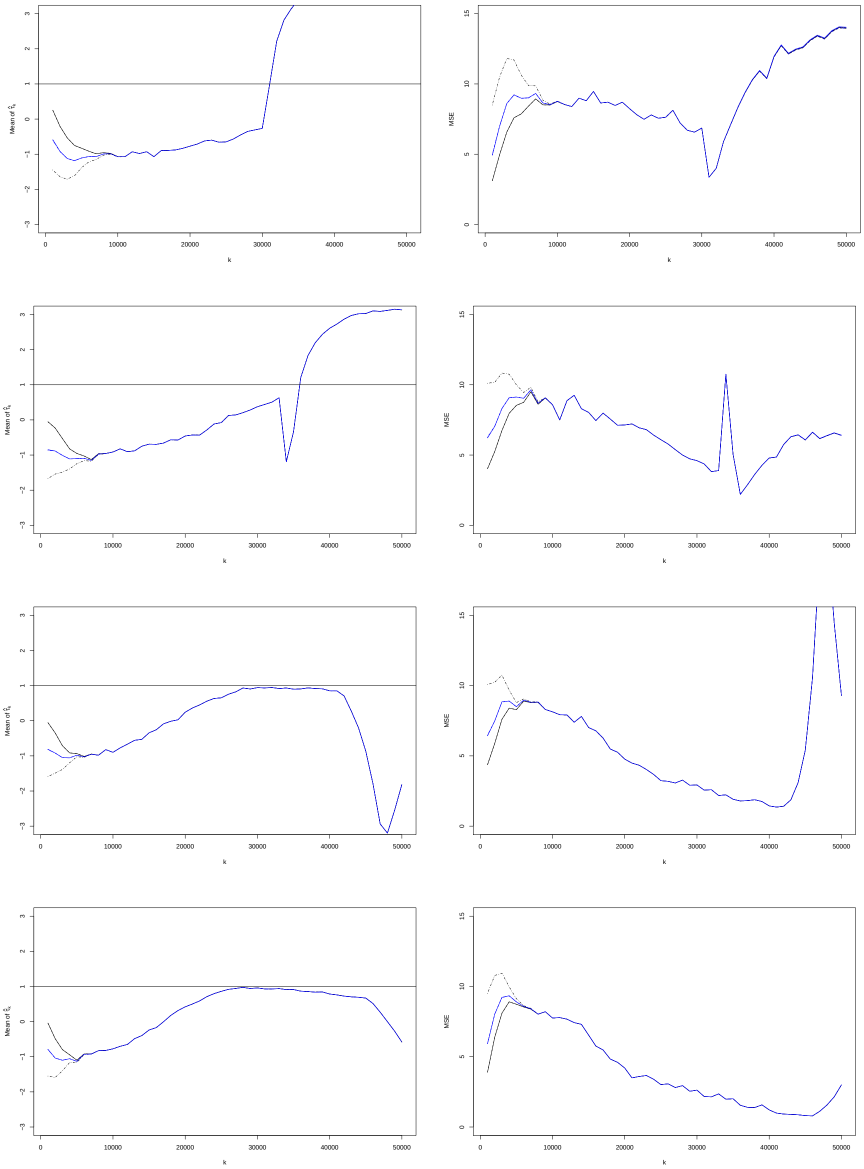

Figure 5.12: Simulation of $\tilde{\tau}_k$ with sample size of 50000 for different values of k for the Nelsen (4.2.8) copula with $\theta = 1.5$ (row 1), $\theta = 2$ (row 2), $\theta = 5$ (row 3) and $\theta = 10$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The tuning parameters a, b and l in the model are $a = 0.01$, $b = 0.02$, $l = 0.01$ (black, solid), $a = 0.01$, $b = 0.02$, $l = 0.02$ (black, dashed), $a = 0.02$, $b = 0.01$, $l = 0.01$ (black, dashed-dotted), $a = 0.02$, $b = 0.01$, $l = 0.02$ (black, dotted), $a = b = 0.01$, $l = 0$ (blue, solid), $a = b = 0.02$, $l = 0$ (blue, dashed), $a = b = 0.03$, $l = 0$ (blue, dashed-dotted), $a = b = 0.04$, $l = 0$ (black, dotted). The horizontal black solid line is the true value of τ.

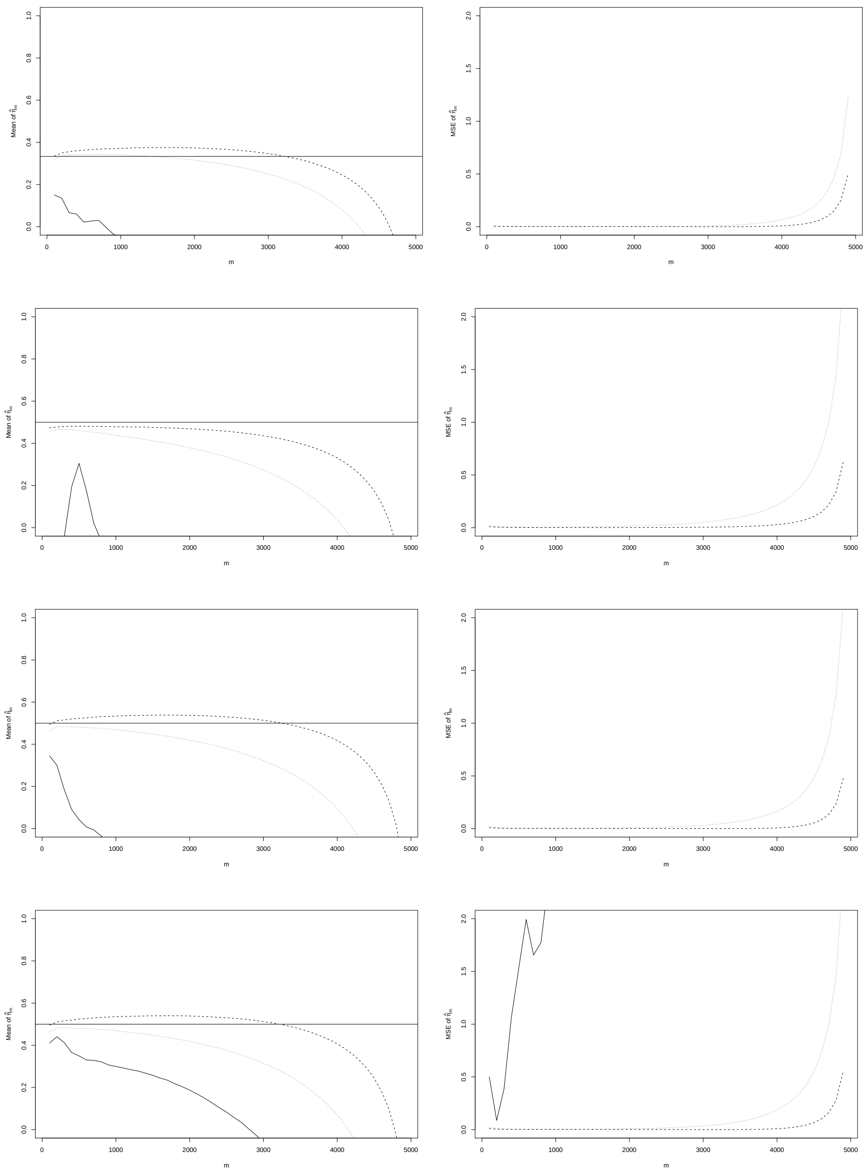

Figure 5.13: Simulation of $\hat{\eta}_m$ with sample size of 5000 for different values of m for the Ali-Mikhail-Haq copula with $\theta = -1$ (row 1), $\theta = -0.5$ (row 2), $\theta = 0.5$ (row 3) and $\theta = 1$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The values of $\eta\tau$ are: initial estimates of η and τ are used (solid), $\eta\tau = 1$ (dashed), true values of η and τ are used (dotted). The horizontal solid line is the true value of η.

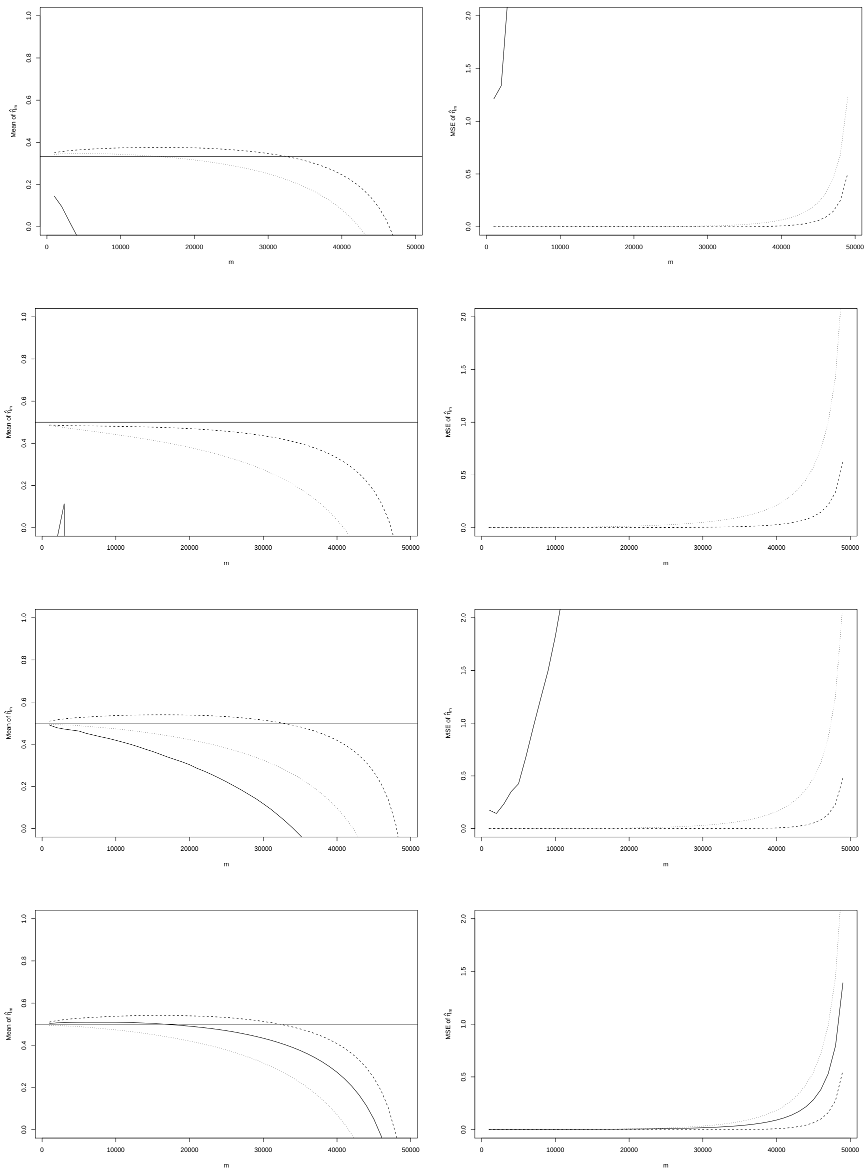

Figure 5.14: Simulation of $\hat{\eta}_m$ with sample size of 50000 for different values of m for the Ali-Mikhail-Haq copula with $\theta = -1$ (row 1), $\theta = -0.5$ (row 2), $\theta = 0.5$ (row 3) and $\theta = 1$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The values of $\eta\tau$ are: initial estimates of η and τ are used (solid), $\eta\tau = 1$ (dashed), true values of η and τ are used (dotted). The horizontal solid line is the true value of η.

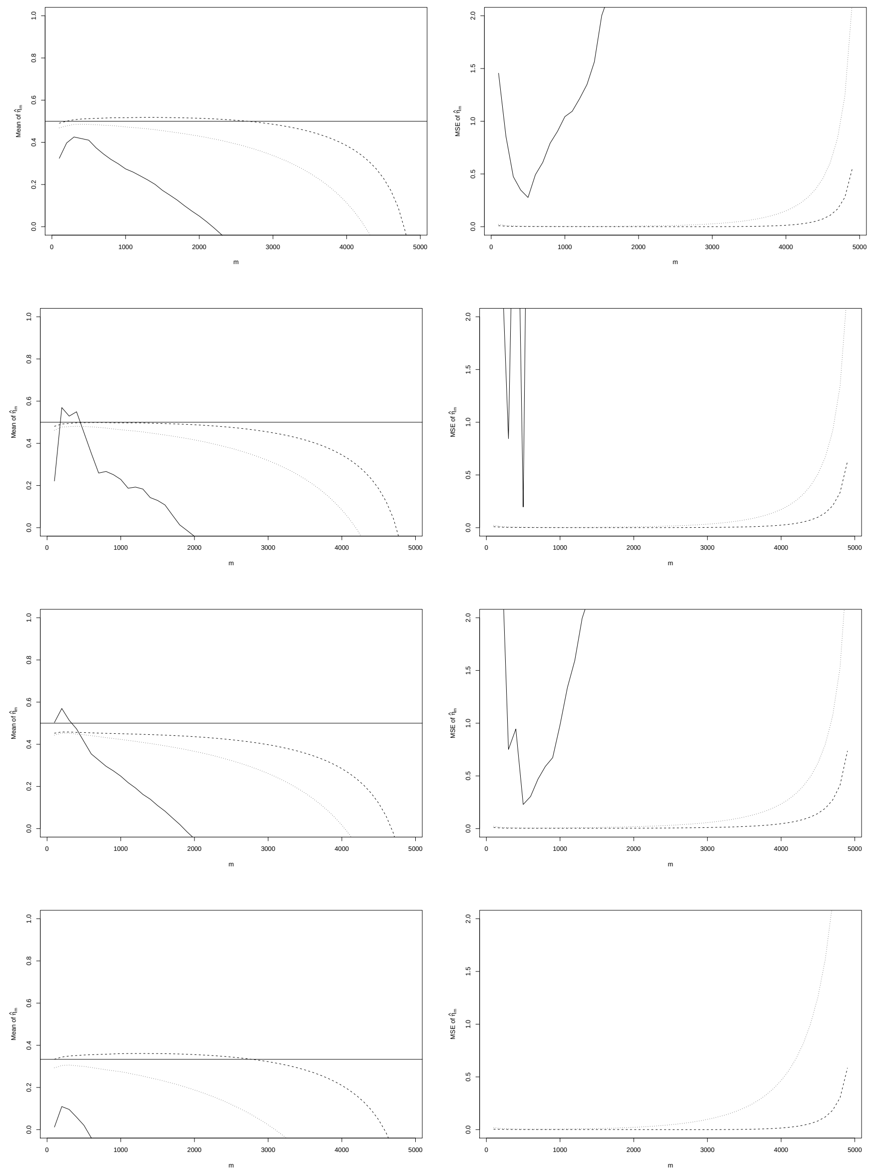

Figure 5.15: Simulation of $\hat{\eta}_m$ with sample size of 5000 for different values of m for the Gumbel-Hougaard copula with $\theta = 0.25$ (row 1), $\theta = 0.5$ (row 2), $\theta = 0.75$ (row 3) and $\theta = 1$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The values of $\eta\tau$ are: initial estimates of η and τ are used (solid), $\eta\tau = 1$ (dashed), true values of η and τ are used (dotted). The horizontal solid line is the true value of η.

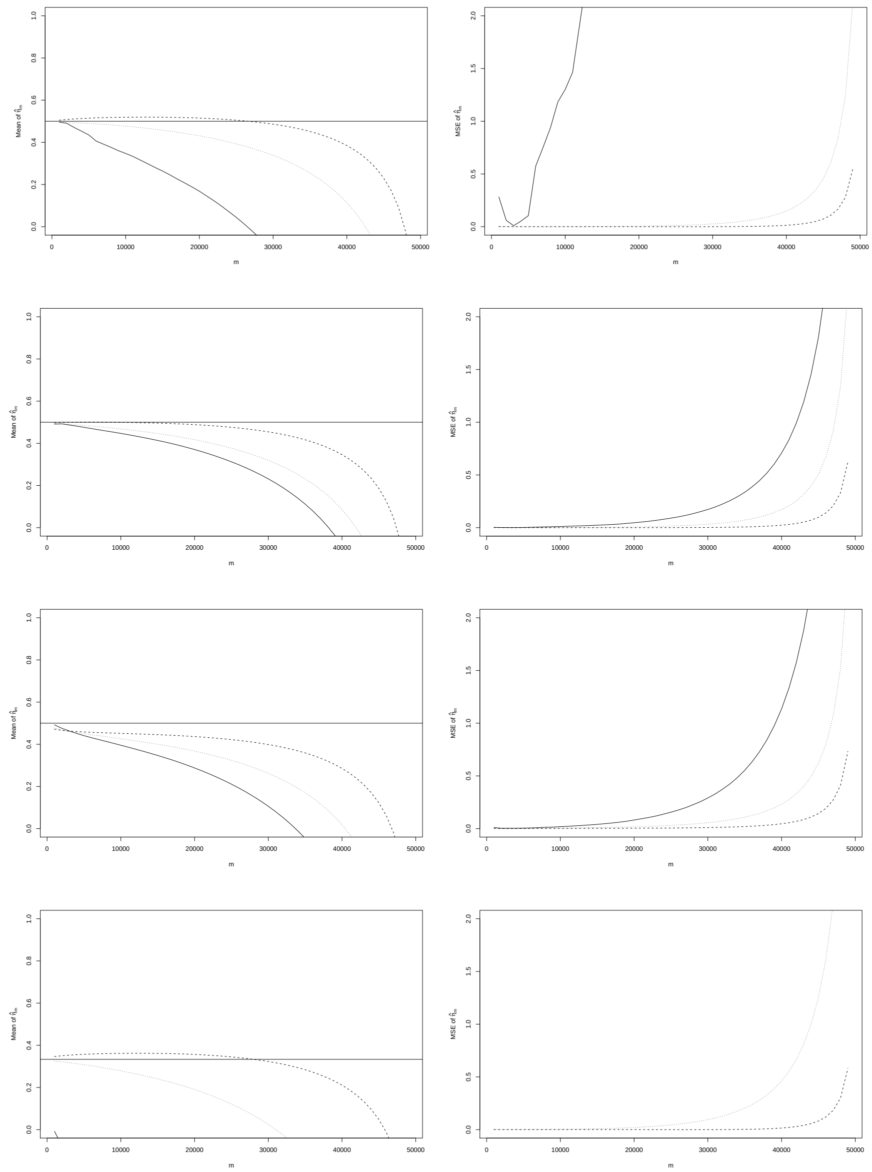

Figure 5.16: Simulation of $\hat{\eta}_m$ with sample size of 50000 for different values of m for the Gumbel-Hougaard copula with $\theta = 0.25$ (row 1), $\theta = 0.5$ (row 2), $\theta = 0.75$ (row 3) and $\theta = 1$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The values of $\eta\tau$ are: initial estimates of η and τ are used (solid), $\eta\tau = 1$ (dashed), true values of η and τ are used (dotted). The horizontal solid line is the true value of η.

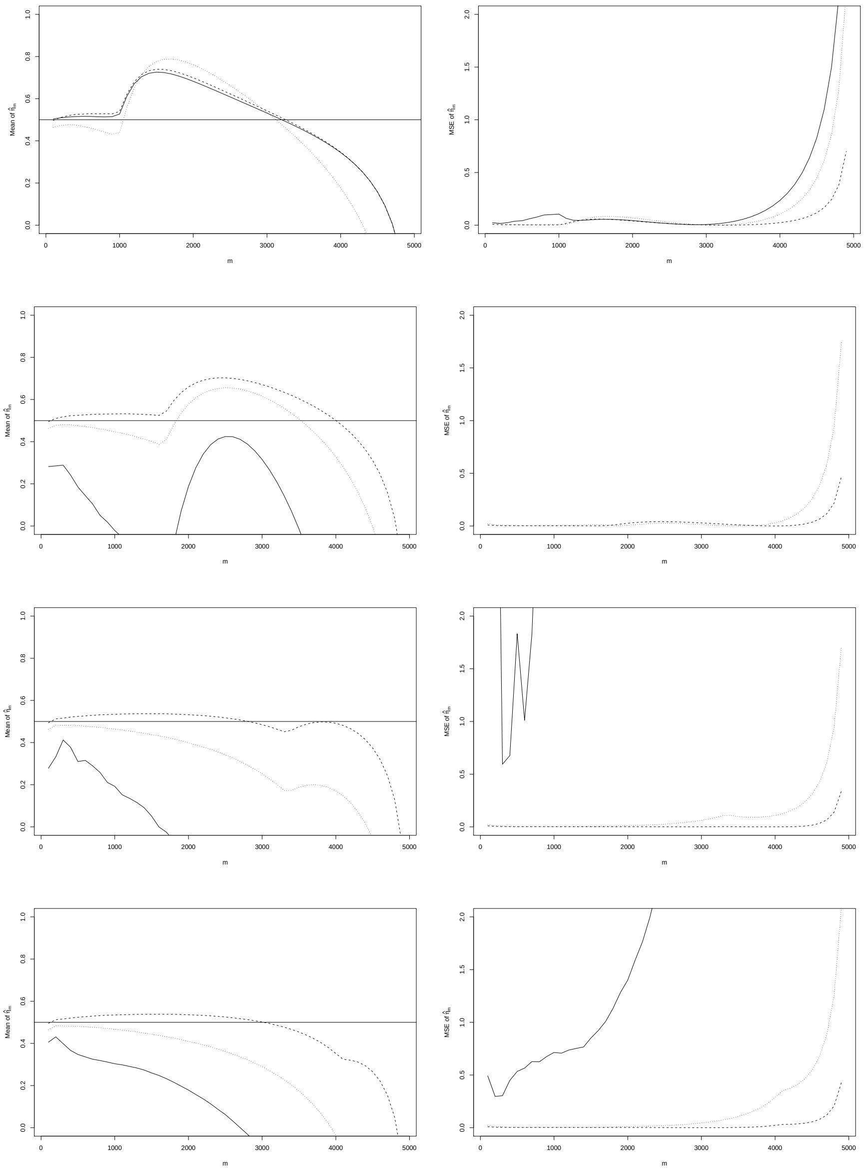

Figure 5.17: Simulation of $\hat{\eta}_m$ with sample size of 5000 for different values of m for the Nelsen (4.2.8) copula with $\theta = 1.5$ (row 1), $\theta = 2$ (row 2), $\theta = 5$ (row 3) and $\theta = 10$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The values of $\eta\tau$ are: initial estimates of η and τ are used (solid), $\eta\tau = 1$ (dashed), true values of η and τ are used (dotted). The horizontal solid line is the true value of η.

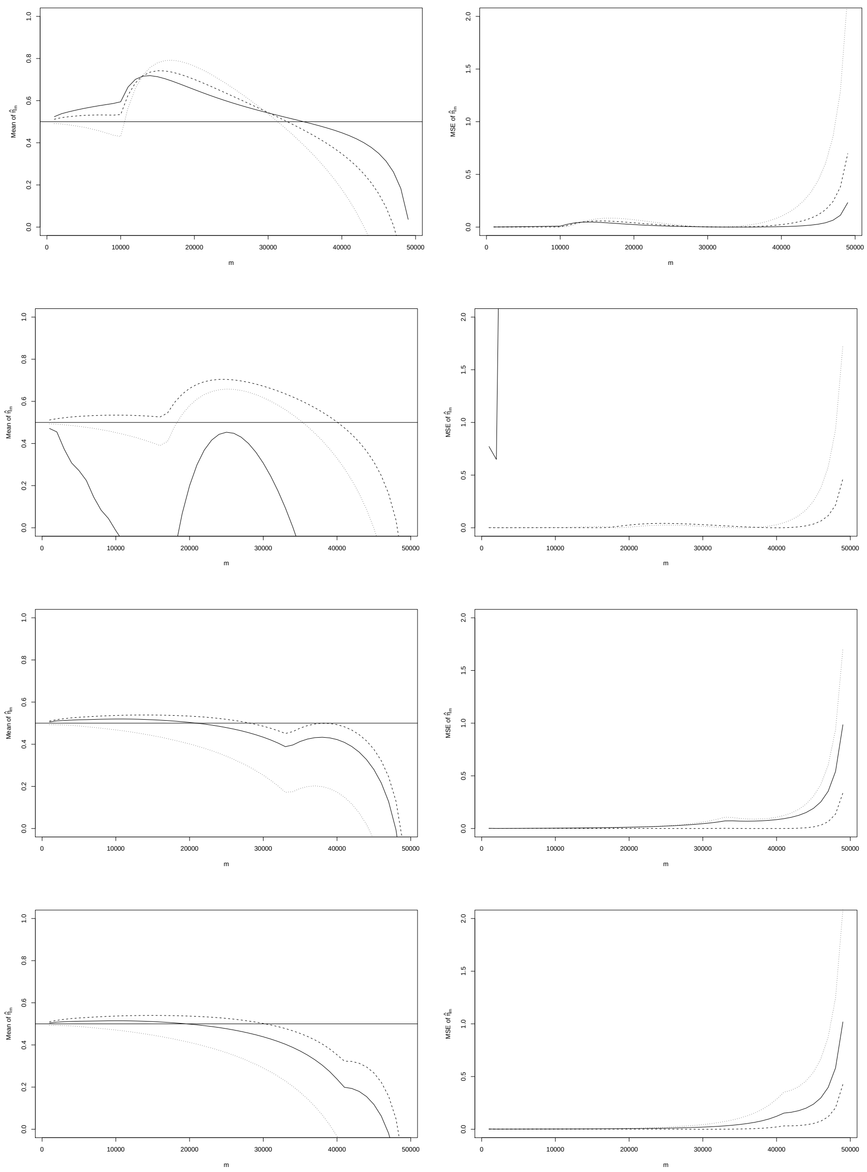

Figure 5.18: Simulation of $\hat{\eta}_m$ with sample size of 50000 for different values of m for the Nelsen (4.2.8) copula with $\theta = 1.5$ (row 1), $\theta = 2$ (row 2), $\theta = 5$ (row 3) and $\theta = 10$ (row 4). On the left is the sample mean of the estimator and on the right is the empirical mean squared error. The values of $\eta\tau$ are: initial estimates of η and τ are used (solid), $\eta\tau = 1$ (dashed), true values of η and τ are used (dotted). The horizontal solid line is the true value of η.

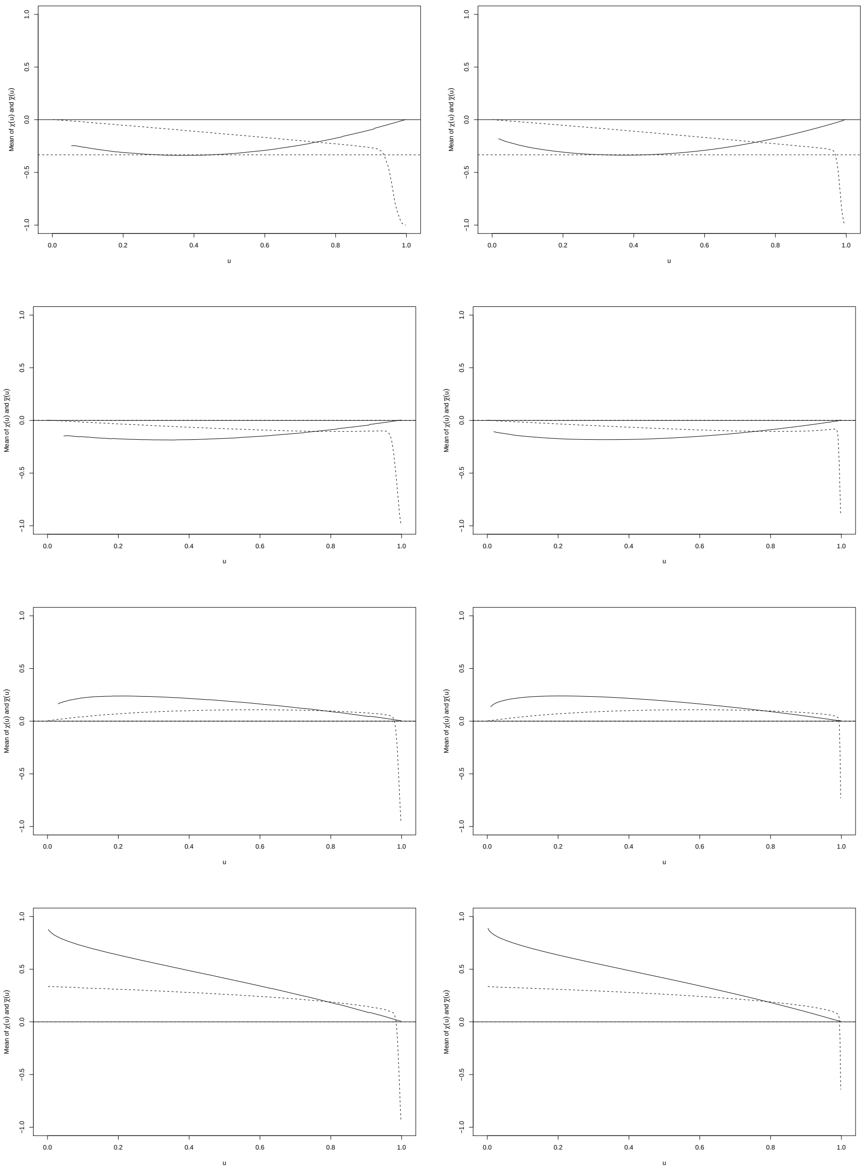

Figure 5.19: Simulation of $\hat{\chi}(u)$ (solid line) and $\hat{\bar{\chi}}(u)$ (dashed line) with a sample size of 5000 (left) and 50000 (right) for the Ali-Mikhail-Haq copula with $\theta = -1$ (row 1), $\theta = -0.5$ (row 2), $\theta = 0.5$ (row 3) and $\theta = 1$ (row 4). The horizontal solid line is the true value of χ, while the horizontal dashed line is the true value of $\bar{\chi}$. In the cases where $\chi = \bar{\chi}$, the true value of this is plotted using a horizontal solid line.

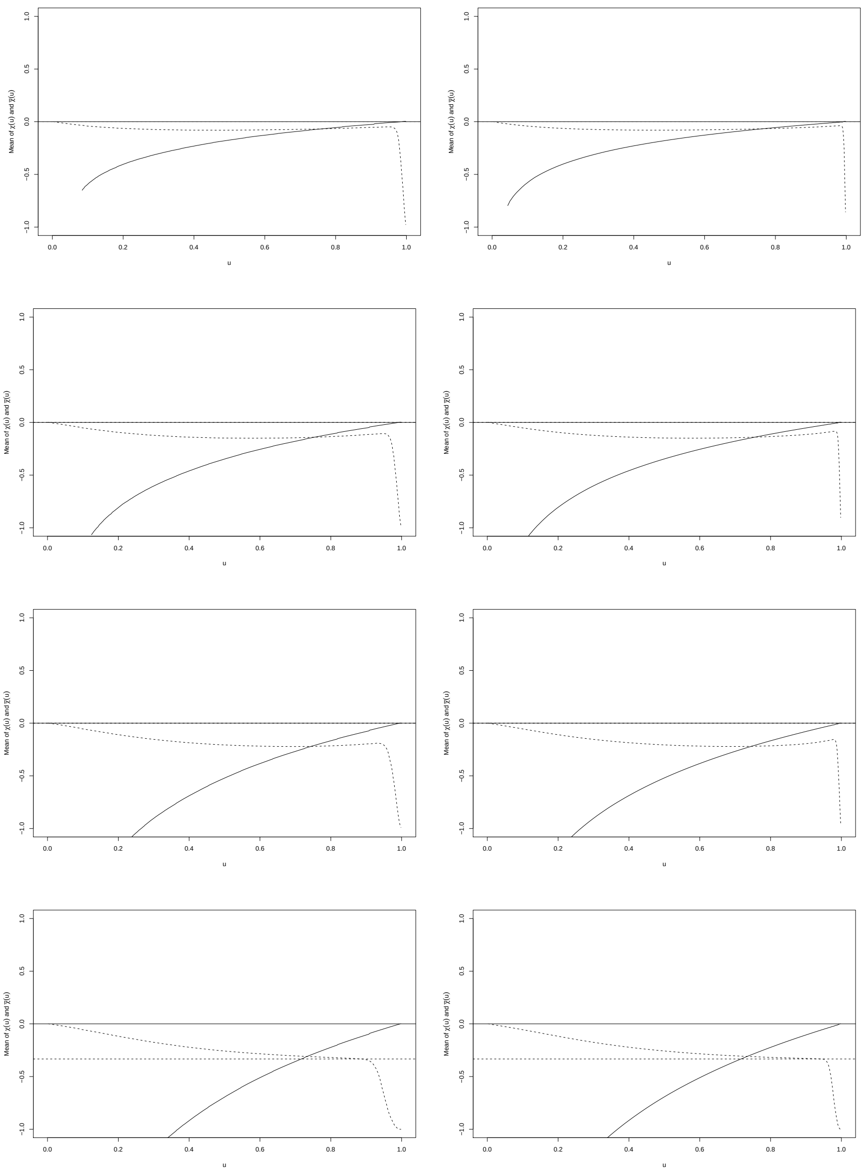

Figure 5.20: Simulation of $\hat{\chi}(u)$ (solid line) and $\hat{\bar{\chi}}(u)$ (dashed line) with a sample size of 5000 (left) and 50000 (right) for the Gumbel-Hougaard copula with $\theta = 0.25$ (row 1), $\theta = 0.5$ (row 2), $\theta = 0.75$ (row 3) and $\theta = 1$ (row 4). The horizontal solid line is the true value of χ, while the horizontal dashed line is the true value of $\bar{\chi}$. In the cases where $\chi = \bar{\chi}$, the true value of this is plotted using a horizontal solid line.

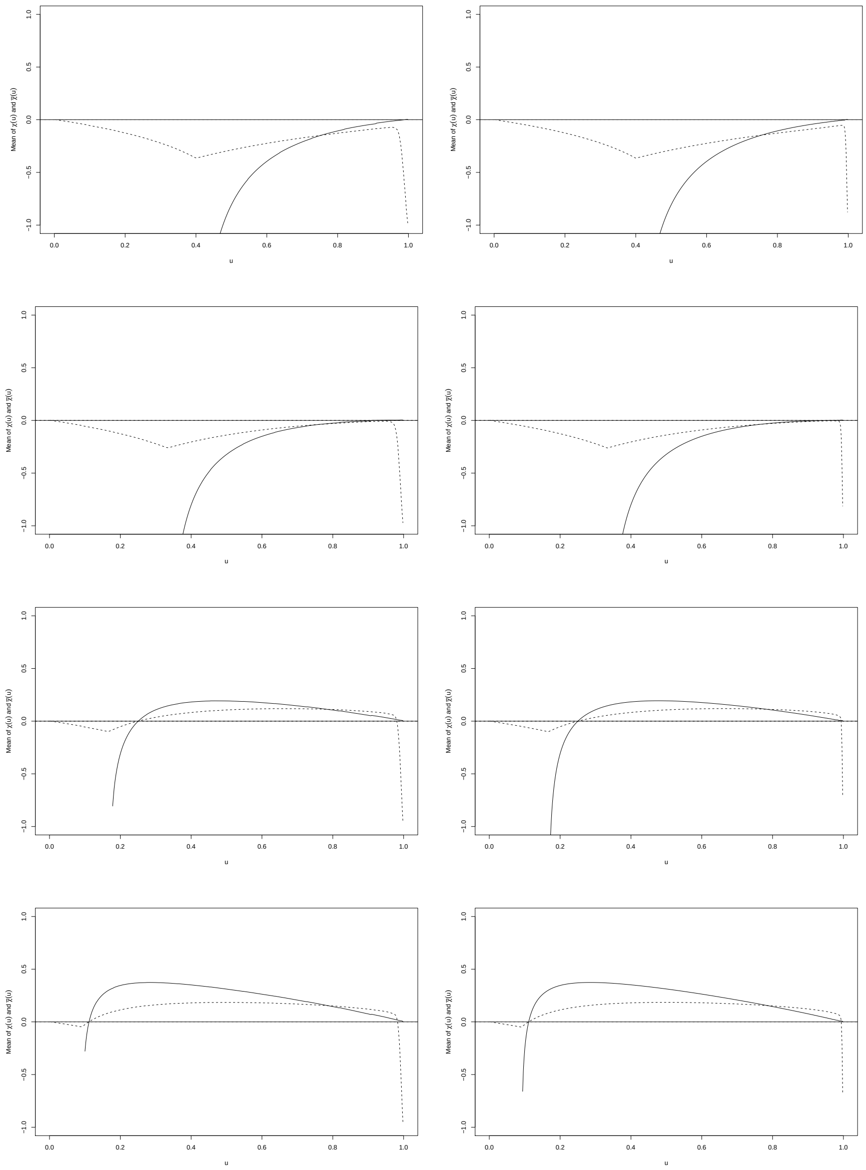

Figure 5.21: Simulation of $\hat{\chi}(u)$ (solid line) and $\hat{\bar{\chi}}(u)$ (dashed line) with a sample size of 5000 (left) and 50000 (right) for the Nelsen (4.2.8) copula with $\theta = 1.5$ (row 1), $\theta = 2$ (row 2), $\theta = 5$ (row 3) and $\theta = 10$ (row 4). The horizontal solid line is the true value of χ, while the horizontal dashed line is the true value of $\bar{\chi}$. In the cases where $\chi = \bar{\chi}$, the true value of this is plotted using a horizontal solid line.

Chapter 6

Estimation of taildependence in BMI twindata

In this chapter we analyze the BMI of Finnish twins from the older cohort of the Finnish Twin Cohort Study. We start by describing the data and then we move on to analyzing the univariate behaviour of the BMI. This analysis is based on the methods described in Chapter 1. We finish with a multivariate analysis of the data and estimate the various dependence measures described in Chapter 3 and Chapter 4. Both the univariate and multivariate analysis is carried out for the lower and the upper tail of the distribution of BMI. The main interest is to investigate the dependency structure in the lower and the upper tail of BMI using monozygotic and dizygotic twin pairs.

6.1 Description of the data

The data we consider are derived from the older cohort of the Finnish Twin Cohort Study (Kaprio and Koskenvuo, 2002). A baseline questionaire was sent to all same-sex Finnish twins in 1975, who were born before 1958 where both twins were still alive. In 1981 a follow-up questionaire was sent to all the twins which had received the questionaire in 1975, regardless of whether or not they had returned the original questionaire. The response rate to the first questionaire was 89%, while it was 84% in the follow-up questionaire. Twin pairs that had participated in at least one of the two first questionaires received a third questionaire in 1990. In this survey twins born before 1930 were excluded and both twins still had to be alive. The response rate in this survey was 77%. In the third questionaire the twins were also asked what their weight was when they were 20 years old, when they were 30 years old, what their weight was 12 months ago and what it was five years ago. These measurements are thus based on the subjects memory rather than an actual measurement. The zygosity of each twin pair was determined in the 1975 and 1981 surveys based on questions regarding the similarity of appearance of the twins at an early school age. The reliability of this method of classifying the zygosity of a twin pair was tested in a small study where 104 twin pairs participated (Sarna et al., 1978). Each pair was classified as monozygotic (MZ) or dizygotic (DZ) using the same method as in the questionaire before classifying them as MZ or DZ using 11 bloodmarkers. The observed agreement between the two methods of classification was 100%. The probability

of misclassification for a twin pair was estimated to be less than 2%.
The height and weight of each individual were self reported, and they were asked in an identical way in each survey. The height was reported in centimeters and weight was reported in kilograms. The BMI is then calculated according to the formula

$$\mathrm{BMI} = \frac{\mathrm{Weight}}{\mathrm{Height}^2},$$

where the weight is given in kilograms and the height in centimaters. Several years after the 1990 questionaire, the validity of self reported height and weight was examined in a small subsample of the people who participated in the 1990 questionaire (Korkeila et al., 1998). A clinical examination showed that correlation of self reported BMI and measured BMI was 89% for men and 90% for women, which indicates a good reliability of self reported BMI.
Our study is based on a total of 4349 twin pairs, where we have up to 7 measurements at different times. Overall we have 57524 measurements on individuals or 28762 measurements on twin pairs, all aged 18-60. In Table 6.1 there is a short summary of the data, which we have split up into the four groups “men MZ”, “men DZ”, “women MZ” and “women DZ”. For each group we have the number of individual measurements, the number of measurements on twin pairs, the age range, the age mean, the age median, the BMI range, the BMI mean, the BMI median, the BMI variance and the covariance between the two twins. In both the measures concerning the age and BMI, there are repeated measurements since some of the indiviuals have reported their age and BMI more than once.

	Men		Women	
	MZ	DZ	MZ	DZ
N (Individuals)	8212	16354	11836	21122
N (Twin pairs)	4106	8177	5918	10561
Age (Range)	(18.01 - 61.00)	(18.04 - 60.85)	(18.04 - 60.89)	(18.02 - 61.00)
Age (Mean)	34.98	35.23	34.18	34.34
Age (Median)	33.81	34.14	32.75	32.81
BMI (Range)	(14.03 - 38.40)	(15.19 - 39.79)	(13.67 - 39.06)	(13.12 - 39.92)
BMI (Mean)	24.00	24.37	22.28	22.65
BMI (Median)	23.67	24.02	21.63	22.05
BMI (Variance)	8.89	9.35	11.09	11.44
BMI (Covariance)	6.09	3.95	7.77	4.59

Table 6.1: Summary of data from the Finnish twins.

6.2 Univariate analysis

A univariate analysis of the twins serves several purposes. Firstly, it is interesting to see if the BMI varies with age, since this affects whether or not it is necessary to do a multivariate analysis on age defined subsets of the data. Secondly, an estimate of γ can give an indication of which class of extreme value distributions the maximum or minimum BMI comes from. This is relevant because larger values of γ means that the tail is heavier, and estimation of

parameters become more difficult if the tails are very light since there are fewer extreme values.
In Figure 6.1 a plot of the BMI versus age of Finnish males and females are shown. In this figure the age of each of the subjects has been rounded down to the nearest integer. For both males and females we see a clear tendency of growth in BMI as the subjects get older. This suggests that the BMI data is not stationary. For the very young men and women (18-19 years old) there are very few subjects with extremely high or low BMI. Also, from these plots it seems that the upper tails are much heavier than the lower tails since more subjects are further from the center of the distribution. This is also not unexpected since the distribution of BMI has a finite left endpoint.

Upper tail An estimate of γ in the upper tail in the GEV framework for each age group can be seen in Figure 6.2, while an estimate of γ in the POT framework can be seen in Figure 6.4. In order to determine the threshold in the POT framework, a Mean residual life plot for both men and women are presented in Figure 6.3. The mean residual life plots are constructed using all the male and female subjects respectively. Based on the mean residual life plots it seems reasonable to choose a threshold for the men at around 28, and 27 for the women, since this is where the mean residual life plots start to become approximately linear. We can not use these thresholds directly since we have not taken the age of the subjects into consideration. Instead we use these thresholds to see how deep we have to get into the data in order to obtain the extreme values. For men a threshold of 28 means that 11.22% of the data should be considered extreme and for the women a threshold of 27 means that 9.80% should be considered extreme. So for both men and women we will define the threshold for each age group to be the 0.9 empirical quantile. The estimates of γ in both the GEV framework and the POT framework are maximum likelihood estimates from the likelihood equations in (1.13) and (1.15), respectively. Concerning the GEV estimates of γ we made 20 blocks and assigned all subjects randomly to one of these blocks. This was done in such an order that all the blocks had the same size whenever possible. In both the GEV famework and the POT framework we used the R package "ismev" from (Coles, 2001). It can be seen from the plots of γ versus age, that the age of each subject does not influence the estimate of γ very much. The estimate of γ is for both men and women fairly close to 0, which suggests that the distribution function of BMI is in the domain of attraction of the Gumbel class. The estimates of γ and the pointwise confidence intervals are based on the assumption that all the individuals are independent. This is not exactly true, since the two individuals in each twinpair can not be considered independent. We do have independence between the pairs, though.

Lower tail An estimate of γ for the lower tail of the distribution can be obtained in the same way as for the upper tail, by using the relation described in (1.1). In Figure 6.5 a plot of the estimate of γ versus age is shown for the Finnish men and the Finnish women in the GEV framework, while a plot of the estimate of γ in the POT framework can be seen in Figure 6.7. Again we use a mean residual life plot to find a percentage based threshold. When we consider the lower tail, the mean residual life plot consists of the mean shortfall below a threshold instead of the mean excess above a threshold, but otherwise it is exactly the same. Based on the mean residual life plots in Figure 6.6 a threshold around 19 for men and

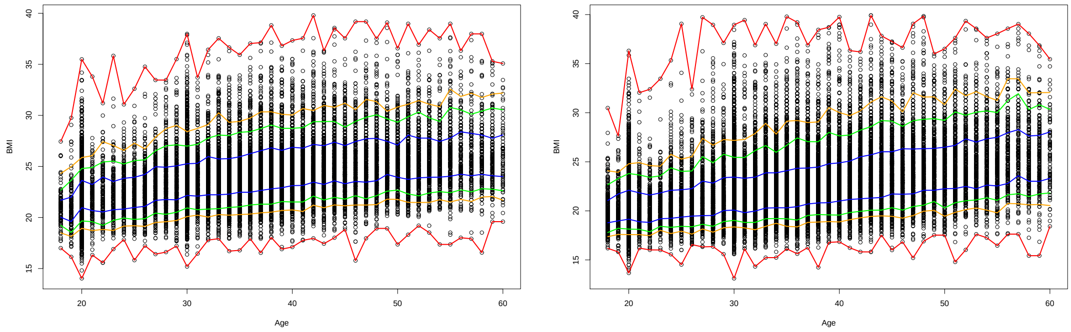

Figure 6.1: Plot of BMI versus age for Finnish males (left) and females (right). Empirical quantiles are given by the lines, Min and Max (red), 0.05 and 0.95 (orange), 0.1 and 0.9 (green), 0.25 and 0.75 (blue).

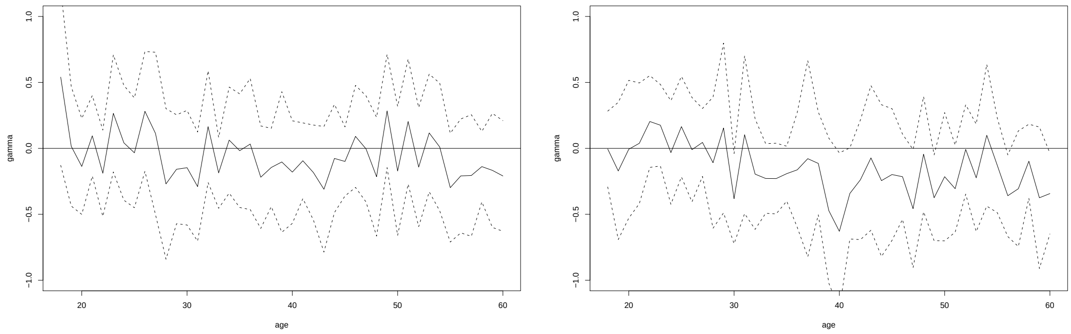

Figure 6.2: Upper tail: Estimates of γ in the GEV framework for Finnish males (left) and females (right) with a pointwise 95% confidence interval.

18 for women seems reasonable, since this is where the mean residual life plots begin to be approximately linear. These thresholds corresponds to selecting approximately 2% of the data and 4% of the data for men and women, respectively. However, if we only include such few observations then the variance of the estimate is very high, so a threshold of 5% for both men and women have been chosen. In the GEV framework we used 20 blocks for both men and women. Based on the estimates of γ in both the GEV framework and the POT framework, it seems likely that the distribution of BMI in the lower tail belongs to the Gumbel Class. Theoretically we should have a finite left endpoint for the distribution of BMI. For simplicity, and because the estimates of γ vary quite alot when we take age into account, we ignore the age of the subjects in our attempt to estimate this left endpoint. The left endpont can be estimated using the constraint in (1.12) for the GEV framework, while it can be done using the constraint in (1.14) for the POT framework. For men we find that the left endpoint of the distribution is 6.84 in the GEV framework while it is 8.33 in the POT framework. For women the left endpoint is estimated to be 11.82 in the GEV framework, whle it is estimated to be 7.96 in the POT framework.

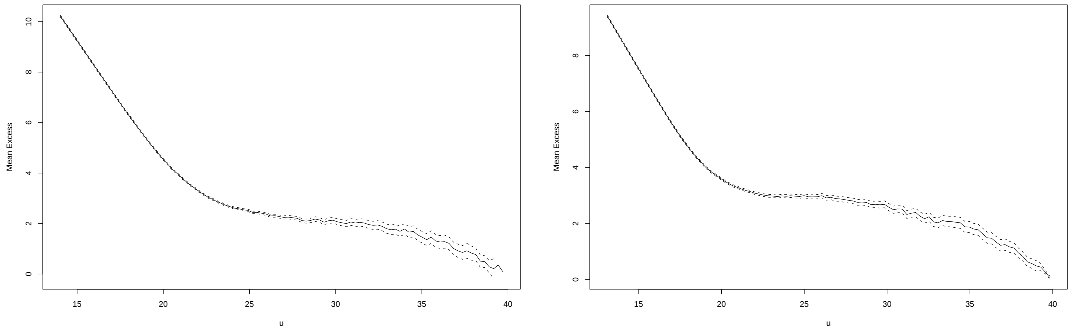

Figure 6.3: Upper tail: Mean residual life plots of Finnish males (left) and females (right) with a pointwise 95% confidence interval.

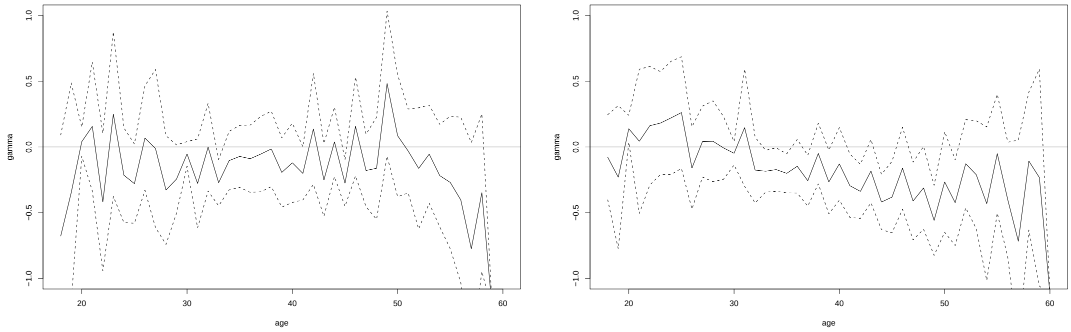

Figure 6.4: Upper tail: Estimation of γ in the POT framework for Finnish males (left) and females (right) with a pointwise 95% confidence interval.

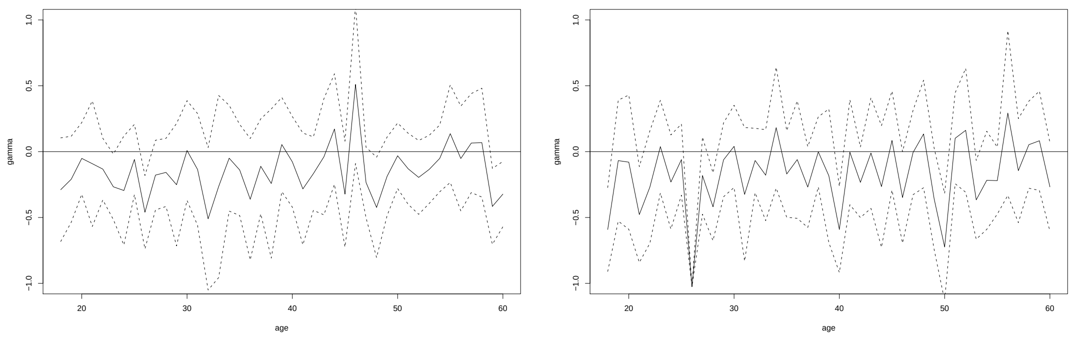

Figure 6.5: Lower tail: Estimates of γ in the GEV framework for Finnish males (left) and females (right) with a pointwise 95% confidence interval.

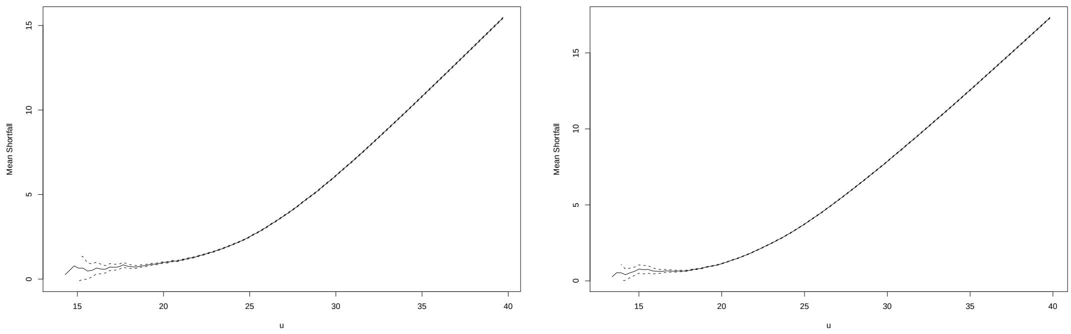

Figure 6.6: Lower tail: Mean residual life plots of Finnish males (left) and females (right) with a pointwise 95% confidence interval.

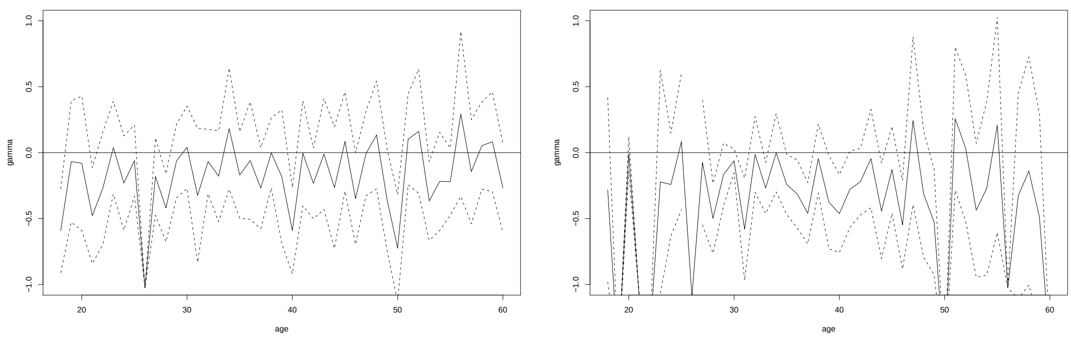

Figure 6.7: Lower tail: Estimates of γ in the POT framework for Finnish males (left) and females (right) with a pointwise 95% confidence interval.

6.3 Multivariate analysis

The main goal of the analysis of the Finnish twin data, is to examine the extremal tail dependence of BMI for monozygotic and dizygotic twins. Such an analysis is very interesting because the effect of the genes in cases of extreme overweight or extreme underweight can be studied. If the tail dependence is much stronger for the monozygotic twins compared to the dizygotic twins, then we can say that genes play a major role. If the tail dependence is approximately equal for the two types of twins, then the data suggests that genes are not the main contributing factor to either extreme overweight or extreme underweight. For both the monozygotic and the dizygotic twins we split them up into several different groups, by age and sex. These are the groups defined in Table 6.2 to Table 6.5. We split the twins into these different groups partly because the BMI depends on age, but also because differences between sex and age groups can be relevant. For each group we estimate the dependence parameters η, χ, $\bar{\chi}$ and τ. It should be noted that for all these estimates, the assumption of independent data is violated. This is because we have several observations for some of the twin pairs, so for the estimates we sometimes have several observations from the same twinpairs.
In Figure 6.8 we see scatterplots, where we have split both men and women up into monozygotic and dizygotic twins. From these plots we get an indication that the monozygotic twins have a higher tail dependence than the dizygotic twins. This is because the extreme datapoints tend to lie closer to the 45 degree line for the monozygotic twins compared to the dizygotic twins. Here, the age of the twins has not been taken into account.

Upper tail In Table 6.2 and Table 6.3 we see the estimates of the tail dependence parameters in the upper tail for the monozygotic twins and the dizygotic twins, respectively. The estimates of η are computed using the estimator $\hat{\eta}_m$ where we use bias correction. For the bias correction we fix the value $\eta\tau = 1$, since this seemed to work best in the simulation studies. The estimate of η that we use is the median of the $\hat{\eta}_m$ estimates we get for $m = 5, \dots, \lfloor \frac{n}{4} \rfloor$, where n denotes the number of twin pairs in the group. The estimates of η also gives us direct estimates of $\bar{\chi}$ using the relation $\bar{\chi} = 2\eta - 1$. The estimates of $\bar{\chi}$ obtained in this way is found in the column $\bar{\chi}_1$. Estimates of χ and $\bar{\chi}$ are found using (3.29) and (3.34), respectively, where we have chosen the threshold $u = 0.9$. These estimates of $\bar{\chi}$ are found in the column $\bar{\chi}_2$. Furthermore, we also have a column $\bar{\chi}_3$, which contains an estimate of $\bar{\chi}$ where the threshold u was chosen to be 0.95. The estimates of τ are obtained using the estimator $\hat{\tau}_k$ with $k = n^{0.95}$, where n is the number of twin pairs in the group. For the estimates of η and $\bar{\chi}_1$ we have used a non parametric bootstrap to obtain standard deviations of the estimates. In these bootstraps 1000 iterations were used. The standard deviations of the estimates of χ and $\bar{\chi}_2$ are found using Proposition 3.5.5 and Proposition 3.5.6, respectively. The estimates of $\chi(u)$ and $\bar{\chi}(u)$ and their standard deviations are computed with the R package "POT" (Ribatet, 2006).
Concerning the estimates of η we see that there is a clear tendency of the estimate to be higher for the monozygotic twins compared to the dizygotic twins. This is true for all the different groupings of the data we have considered. However, the difference in the η estimates gets lower when the subject gets older. This can for instance be seen for the males aged $40 - 60$, where the η estimate is 0.697 for the monozygotic twins while it is 0.650 for the dizygotic twins. This indicates that the effect of the genes on extreme obesity diminish as the subjects gets older. The pair of measures $(\chi, \bar{\chi})$ does not fit perfect into the discussion summarized

in Table 3.1. This is because we require either ($\chi > 0, \bar{\chi} = 1$) or ($\chi = 0, \bar{\chi} \leq 1$), which we do not have. Still it is quite clear that the monozygotic twins have higher values of both χ and $\bar{\chi}$ than the dizygotic twins do, implying a higher tail dependence in the upper tail. The estimates of τ we see in Table 6.2 and Table 6.3 are not very usefull, which was also expected from the simulation study. The values just seem arbitrary, and do not offer much insight to the data. We have however decided to include them anyway to show that we have made an attempt to make a full data analysis.

Lower tail

The estimates of the dependence measures for the lower tail can be found in Table 6.4 and Table 6.5 for monozygotic and dizygotic twins, respectively. Concerning the dependence measure η, we see that the estimates are much higher for the monozygotic twins than for the dizygotic twins in all cases except for men aged $40 - 60$, where the estimate is highest for the dizygotic twins. The fact that it is higher for the dizygotic twins is unexpected, as we would expect the estimates of η to be similar for the two types of twins if the genes did not contribute significantly, but we would never expect the dizygotic twins to have a higher degree of tail dependence. For the dependence measures χ and $\bar{\chi}$, we also see that the estimates tend to be higher for the MZ twins than the DZ twins, meaning that the there is a stronger dependence between extreme underweight for MZ twins. The estimates of τ are again not very usefull.

In Figure 6.9 and Figure 6.10 we have added plots of $\chi(u)$ and $\bar{\chi}(u)$ for the upper and lower tail, respectively. From these plots we see how the dependency structure changes when we move deeper into the tails. We see that both $\chi(u)$ and $\bar{\chi}(u)$ tend to be higher for the monozygotic twins no matter how deep into the tail we go, indicating that the BMI dependency is stronger for monozygotic twins for all levels of BMI.
As mentioned in Section 3.5 we can use a plot of the function $\chi(u)$ as a measure of goodness of fit to a bivariate extreme value distribution. In Figure 6.11 plots of $\chi(u)$ for block maxima of all MZ twins and DZ twins can be seen for both the lower and the upper tail. The plots are made by dividing the data into 400 blocks. In all the plots the function $\chi(u)$ is fairly stable, which is a necessary condition for a good fit to the extreme value distribution. The same plots for the other groups show similar behaviour, and has hence been omitted from the thesis.

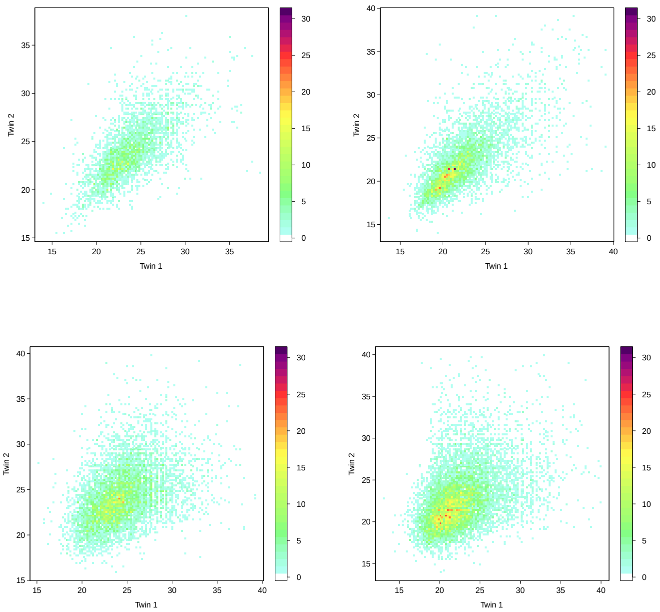

Figure 6.8: Scatterplot of twins for men (left), women (right), MZ (top), DZ (bottom) with the density of the points represented by color.

	N	η	$\bar{\chi}_1$	χ	$\bar{\chi}_2$	$\bar{\chi}_3$	τ
Overall	20048	0.796 (0.0245)	0.592 (0.0490)	0.492 (0.0399)	0.535 (0.0223)	0.556 (0.0274)	2.631
Men	8212	0.781 (0.0405)	0.563 (0.0810)	0.445 (0.0630)	0.509 (0.0346)	0.512 (0.0428)	-0.835
Men (18-29)	2241	0.833 (0.0726)	0.667 (0.1451)	0.441 (0.1209)	0.491 (0.0661)	0.576 (0.0820)	0.665
Men (30-39)	3356	0.871 (0.0739)	0.741 (0.1479)	0.408 (0.0998)	0.468 (0.0536)	0.476 (0.0669)	-0.520
Men (40-60)	2615	0.697 (0.0828)	0.394 (.1656)	0.387 (0.1142)	0.428 (0.0602)	0.416 (0.0762)	-1.570
Women	11836	0.805 (0.0323)	0.610 (0.0645)	0.504 (0.0514)	0.562 (0.0293)	0.597 (0.0357)	1.325
Women (18-29)	3484	0.836 (0.0678)	0.672 (0.1357)	0.452 (0.0978)	0.484 (0.0529)	0.523 (0.0658)	-1.646
Women (30-39)	5071	0.819 (0.0493)	0.638 (0.0985)	0.445 (0.0794)	0.522 (0.0443)	0.542 (0.0544)	0.505
Women (40-60)	3281	0.774 (0.0531)	0.549 (0.1062)	0.472 (0.0989)	0.497 (0.0545)	0.544 (0.0675)	2.163

Table 6.2: Upper tail: Extremal dependency estimates for MZ twins.

	N	η	$\bar{\chi}_1$	χ	$\bar{\chi}_2$	$\bar{\chi}_3$	τ
Overall	37476	0.606 (0.0209)	0.212 (0.0419)	0.261 (0.0317)	0.272 (0.0155)	0.256 (0.0208)	2.880
Men	16354	0.600 (0.0306)	0.200 (0.0612)	0.219 (0.0487)	0.219 (0.0233)	0.198 (0.0322)	7.450
Men (18-29)	4356	0.600 (0.0499)	0.199 (0.0998)	0.217 (0.0922)	0.299 (0.0455)	0.249 (0.0611)	-2.130
Men (30-39)	6657	0.552 (0.0360)	0.104 (0.0721)	0.162 (0.0764)	0.189 (0.0366)	0.172 (0.0511)	5.890
Men (40-60)	5341	0.650 (0.0606)	0.299 (0.1212)	0.135 (0.0860)	0.163 (0.0408)	0.156 (0.0574)	-0.844
Women	21122	0.623 (0.0240)	0.246 (0.0480)	0.253 (0.0423)	0.272 (0.0206)	0.291 (0.0274)	0.498
Women (18-29)	6191	0.568 (0.0497)	0.136 (0.0994)	0.238 (0.0785)	0.234 (0.0380)	0.233 (0.0516)	0.077
Women (30-39)	8835	0.569 (0.0433)	0.139 (0.0867)	0.191 (0.0663)	0.203 (0.0317)	0.184 (0.0441)	-0.385
Women (40-60)	6096	0.668 (0.0413)	0.337 (0.0827)	0.166 (0.0790)	0.256 (0.0383)	0.257 (0.0515)	-0.394

Table 6.3: Upper tail: Extremal dependency estimates for DZ twins.

	N	η	$\bar{\chi}_1$	χ	$\bar{\chi}_2$	$\bar{\chi}_3$	τ
Overall	20048	0.798 (0.0262)	0.597 (0.0525)	0.419 (0.0566)	0.585 (0.0226)	0.574 (0.0274)	-3.539
Men	8212	0.871 (0.0380)	0.742 (0.0759)	0.475 (0.0868)	0.546 (0.0350)	0.600 (0.0428)	0.034
Men (18-29)	2241	0.958 (0.0715)	0.916 (0.1431)	0.591 (0.1595)	0.556 (0.0672)	0.686 (0.0826)	0.430
Men (30-39)	3356	0.787 (0.0541)	0.573 (0.1082)	0.455 (0.1366)	0.567 (0.0550)	0.592 (0.0700)	-0.474
Men (40-60)	2615	0.487 (0.0706)	-0.027 (0.1412)	0.302 (0.1627)	0.495 (0.0612)	0.416 (0.0762)	-0.314
Women	11836	0.784 (0.0321)	0.568 (0.0642)	0.372 (0.0748)	0.543 (0.0291)	0.527 (0.0356)	0.646
Women (18-29)	3484	0.795 (0.0573)	0.589 (0.1147)	0.373 (0.1380)	0.484 (0.0529)	0.535 (0.0658)	0.664
Women (30-39)	5071	0.761 (0.0493)	0.522 (0.0986)	0.297 (0.1170)	0.510 (0.0441)	0.453 (0.0546)	7.192
Women (40-60)	3281	0.888 (0.0573)	0.775 (0.1146)	0.475 (0.1370)	0.490 (0.0544)	0.600 (0.0676)	-0.777

Table 6.4: Lower tail: Extremal dependency estimates for MZ twins.

	N	η	$\bar{\chi}_1$	χ	$\bar{\chi}_2$	$\bar{\chi}_3$	τ
Overall	37476	0.649 (0.0176)	0.298 (0.0352)	0.195 (0.0444)	0.347 (0.0156)	0.352 (0.0203)	1.275
Men	16354	0.664 (0.0307)	0.328 (0.0614)	0.203 (0.0670)	0.335 (0.0236)	0.367 (0.0306)	-1.975
Men (18-29)	4356	0.558 (0.0451)	0.117 (0.0902)	0.107 (0.1332)	0.275 (0.0454)	0.234 (0.0614)	0.134
Men (30-39)	6657	0.629 (0.0514)	0.259 (0.1029)	0.103 (0.1081)	0.276 (0.0368)	0.245 (0.0496)	1.166
Men (40-60)	5341	0.804 (0.0628)	0.607 (0.1256)	0.213 (0.1168)	0.294 (0.0411)	0.364 (0.0537)	0.583
Women	21122	0.629 (0.0242)	0.258 (0.0483)	0.152 (0.0598)	0.309 (0.0207)	0.263 (0.0276)	-0.392
Women (18-29)	6191	0.631 (0.0457)	0.262 (0.0915)	0.151 (0.1106)	0.202 (0.0379)	0.176 (0.0528)	1.096
Women (30-39)	8835	0.632 (0.0331)	0.264 (0.0662)	0.123 (0.0933)	0.301 (0.0320)	0.239 (0.0431)	-1.762
Women (40-60)	6096	0.690 (0.0479)	0.381 (0.0957)	0.140 (0.1116)	0.233 (0.0382)	0.305 (0.0508)	0.614

Table 6.5: Lower tail: Extremal dependency estimates for DZ twins.

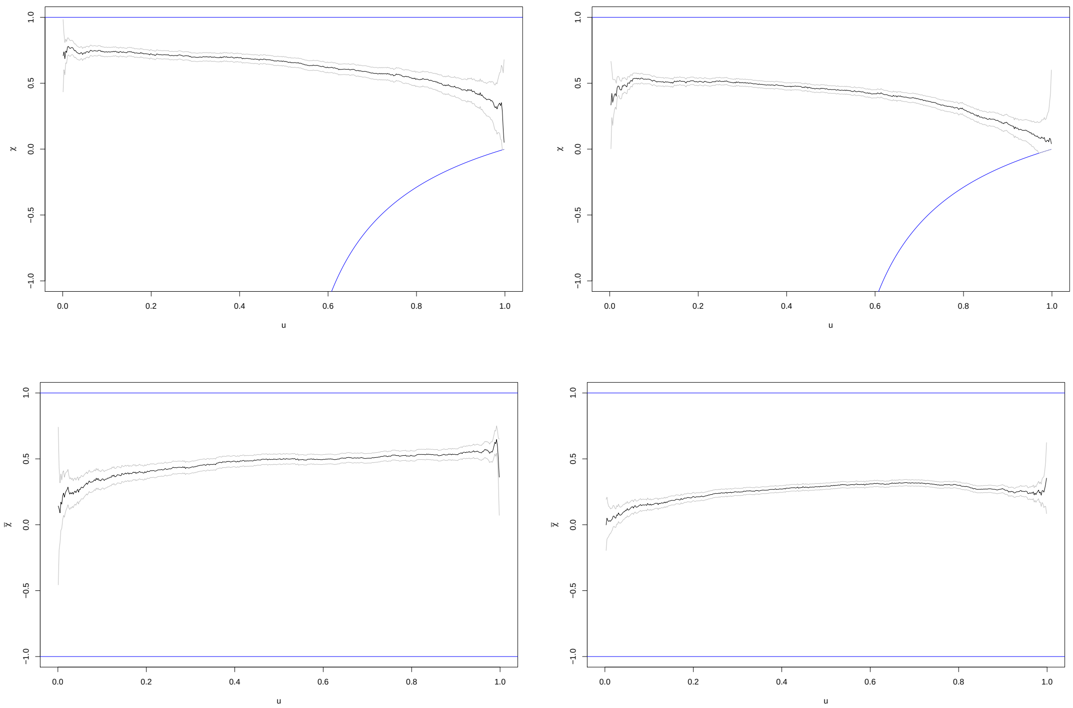

Figure 6.9: Upper tail: Plots of $\chi(u)$ (top) and $\bar{\chi}(u)$ (bottom) for all MZ twins (left) and DZ twins (right) with 95% confidence intervals (Grey) and lower and upper bound of $\chi(u)$ (Blue).

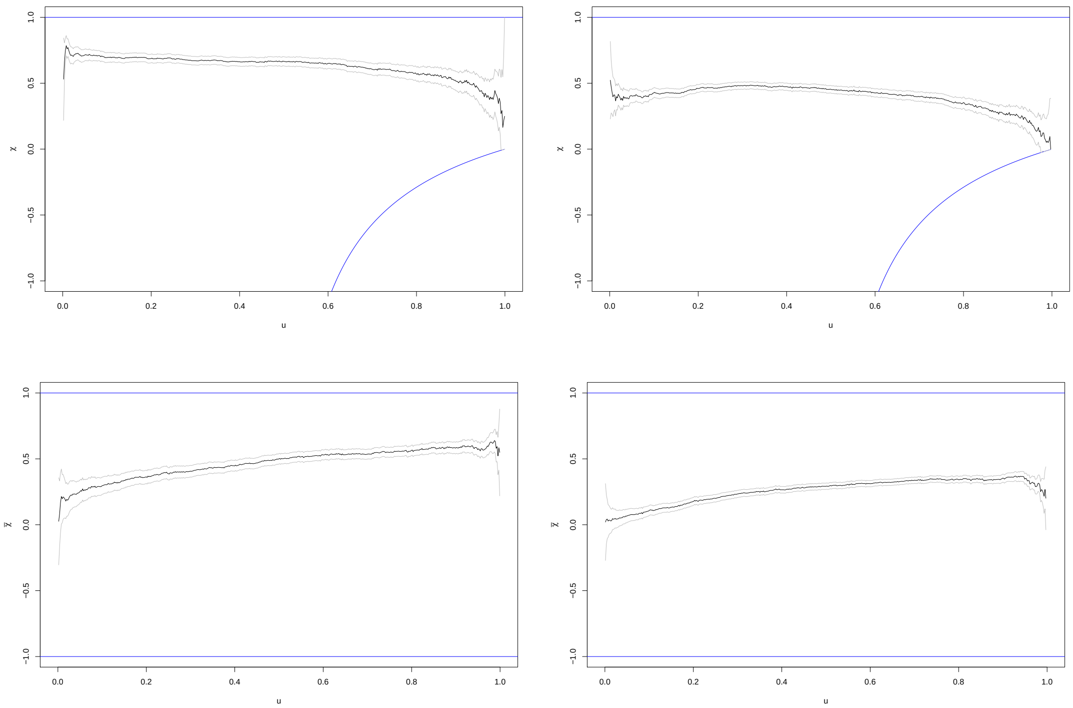

Figure 6.10: Lower tail: Plots of $\chi(u)$ (top) and $\bar{\chi}(u)$ (bottom) for all MZ twins (left) and DZ twins (right) with 95% confidence intervals (Grey) and lower and upper bound of $\chi(u)$ (Blue).

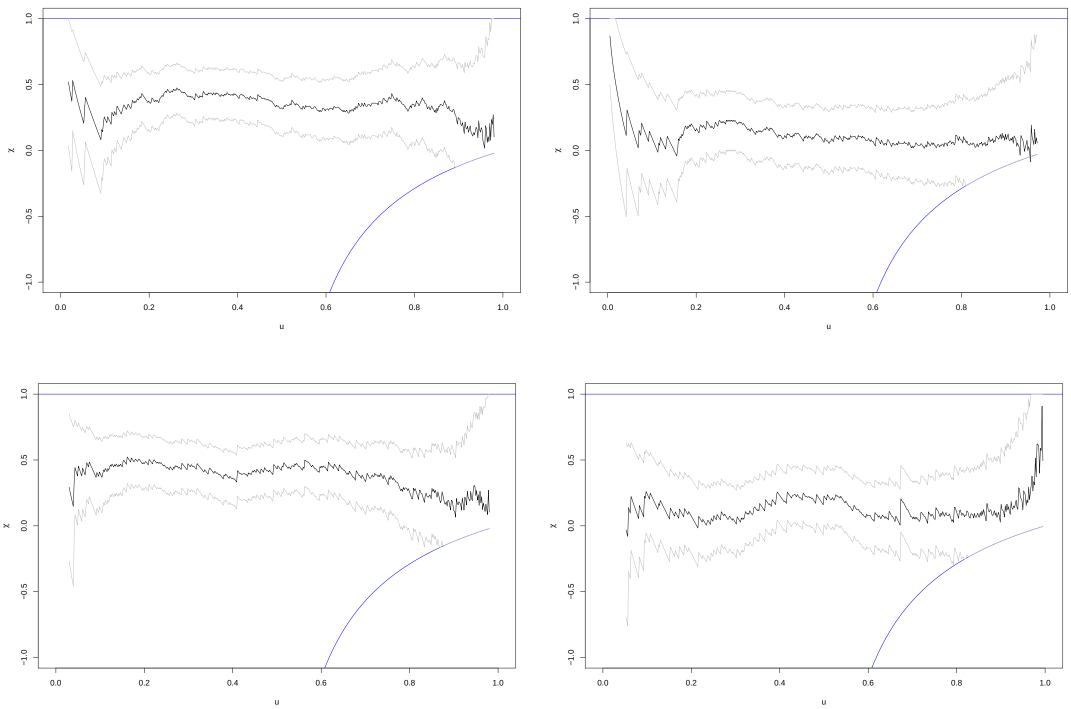

Figure 6.11: Plots of $\chi(u)$ for block maxima of all MZ twins (left), DZ twins (right), upper tail (top), lower tail (bottom) with 95% confidence intervals (Grey) and lower and upper bound of $\chi(u)$ (Blue).

Epilogue

In this thesis we have given a thorough introduction to the fundamental convergence results in extreme value theory. We have provided some simple ways of estimating the extreme value index through maximum likelihood. More sophisticated ways of estimating the extreme value index are available in the literature, but describing these methods in detail, would have led us too far from the main objective of this thesis. We described the max-domain of attraction for the Gumbel class, the extremal Weibull class and the Fréchet class. Extra attention was paid to the Fréchet class, since estimation of parameters in the bivariate case resembles estimation of parameters in the Fréchet class.

After the introduction to the univariate framework in extreme value statistics we discussed multivariate extreme value theory, where we focussed mainly on the bivariate case. Here we discussed transformation of the marginal distributions to standard Fréchet distributions. Transformations to other margins like standard Pareto distributions, exponential distributions, uniform distributions and Gumbel distributions are also possible, and would have led to different simulation results. However, the most common practice is to use transformation to standard Fréchet margins, which is why we opted to do this. Exploring the other ways of transformation would also have been a possibility, but would have required more focus to this area. Concerning the dependency structure in the bivariate case, we have discussed several dependence measures, and shown how they are all connected. For the coefficient of tail dependence and the measures χ and $\bar{\chi}$ we have discussed ways of estimating these. Most of our effort was put into estimating the coefficient of tail dependence, while estimation of χ and $\bar{\chi}$ was considered a minor detail. In the estimation of the coefficient of tail dependence, we also included bias correction which required an estimate of the second order parameter τ. We proposed two ways of estimating τ, where neither proved to be very good. These estimators could possibly be improved, if asymptotic normality was established under a third order condition similar to the second order condition in Assumption 4.1.4. This would allow us to make a bias corrected estimator for τ, and determine values of the tuning paramters, that would theoretically minimize the variance of the estimator. Another possibility of improving the estimator of τ could be to implement an automated threshold selection. This idea is discussed in (Peng, 2010). Since the estimation of τ did not work very well in practice, we decided to use the canonical choice $\tau\eta = 1$ in the estimation of η, throughout the rest of the thesis. This choice was made, since it seemed to perform even better than using the true values of η and τ in the simulations.

For the twin data we made a full univariate analysis, where we found that the lower tail of BMI was lighter than the upper tail. We estimated the extreme value index with maximum likelihood using both the block maxima approach and the peaks over threshold approach. In the bivariate analysis we estimated the coefficient of tail dependence, the dependence measures χ and $\bar{\chi}$ and the second order parameter τ for age and sex defined subsets of the data. The estimation of the second order parameter did not work very well and did not offer any insight to the data. For the other estimates, there were a clear tendency for the dependency measures to be higher for the monozygitc twins, indicating that genes play a major role in extreme obesity and extreme underweight.

In the analysis of the twin data it would have been interesting to have a unified model for the BMI which took the age and sex of the subjects into account. This could have been obtained if an extreme value regression model had been fitted to the data instead of the model we fitted. This is however, an approach we did not include in the theoretical discussion about extreme value theory. An introduction to this area of extreme value theory would be a master thesis in itself, so applying both the approach we took and this approach was not an option. Studying extreme obesity and underweight among twins with a unified model, could be an idea for further study, though.

Bibliography

Alsina, C., Frank, M., and Schweizer, B. (2005). Associative functions on intervals: A primer of triangular norms. *World Scientific, Hackensack, in press.*

Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004). *Statistics of extremes - theory and applications.* John Wiley & Sons Inc, West Sussex.

Chernoff, H., Gastwirth, J., and Johns, M. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *The Annals of Mathematical Statistics*, 38:52–72.

Coles, S. (2001). *An introduction to statistical modeling of extreme values.* Springer Verlag, Berlin.

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

de Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction.* Springer Verlag, Berlin.

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

Draisma, G., Drees, H., Ferreira, A., and De Haan, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, 10:251–280.

Drees, H. (1998). A general class of estimators of the extreme value index. *Journal of Statistical Planning and Inference*, 66:95–112.

Fisher, R. A. and Tippet, L. H. C. (1928). On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical society*, 24:180–190.

Gardes, L. and Girard, S. (2008). A moving window approach for nonparametric estimation of the conditional tail index. *Journal of Multivariate Analysis*, 99:2368–2388.

Geluk, J. and Haan, L. (1987). Regular variation, extensions and tauberian theorems. *CWI Tracts*, 40.

Genest, C. and MacKay, J. (1986a). The joy of copulas: bivariate distributions with uniform marginals. *American Statistician*, pages 280–283.

Genest, C. and MacKay, R. (1986b). Copules archimédiennes et families de lois bidimensionnelles dont les marges sont données. *Canadian Journal of Statistics*, 14:145–159.

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

Goegebeur, Y., Beirlant, J., and de Wet, T. (2010). Kernel estimators for the second order parameter in extreme value statistics. *Journal of Statistical Planning and Inference*, 140:2632–2652.

Goegebeur, Y. and Guillou, A. (2012). Asymptotically unbiased estimation of the coefficient of tail dependence. *Scandinavian Journal of Statistics (To appear)*.

Gomes, M., Miranda, C., and Viseu, C. (2007). Reduced bias tail index estimation and the jackknife methodology. *Statistica Neerlandica*, 61:243–270.

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

Gumbel, E. (1960b). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statist. Univ. Paris*, 9:171–173.

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

Hosking, J., Wallis, J., and Wood, E. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27:251–261.

Kaprio, J. and Koskenvuo, M. (2002). Genetic and environmental factors in complex diseases: the older finnish twin cohort. *Twin Research*, 5:358–365.

Korkeila, M., Kaprio, J., Rissanen, A., Koskenvuo, M., and Sørensen, T. (1998). Predictors of major weight gain in adult finns: stress, life satisfaction and personality traits. *International journal of obesity*, 22:949–957.

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

Ledford, A. and Tawn, J. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 59:475–499.

Nelsen, R. (2006). *An introduction to copulas*. Springer Verlag.

Peng, L. (2010). A practical way for estimating tail dependence functions. *Statistica Sinica*, 20:365.

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

Pickands, J. (1981). Multivariate extreme value distributions. In *Proceedings 43rd Session International Statistical Institute*.

Rényi, A. (1953). On the theory of order statistics. *Acta Mathematica Hungarica*, 4:191–231.

Ribatet, M. A. (2006). A user's guide to the pot package (version 1.0).

Sarna, S., Kaprio, J., Sistonen, P., and Koskenvuo, M. (1978). Diagnosis of twin zygosity by mailed questionnaire. *Human Heredity*, 28:241–254.

Sibuya, M. (1960). Bivariate extreme statistics, i. *Annals of the Institute of Statistical Mathematics*, 11:195–210.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.