Semi- and non-parametric flood frequency analysis

Long-standing assumptions under investigation

PAUL KINSVATER

technische universität dortmund

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
an der Fakultät für Statistik
der Technischen Universität Dortmund

Juli 2016

Contents

Introduction

Flood frequency analysis is a discipline from hydrology dealing with the estimation of river flow distributions in order to analyze the risk of floods. A flow is the amount of water in m^3 per second passing through a measurement station. The ultimate objective is to determine design characteristics of future flood protection systems, for instance, the hight of a dam for some predefined non-exceedance probability of $p \in (0, 1)$.
Official guidelines from the German Hydrological Society define a non-failure probability of a dam in terms of the distribution function $F(y) = \mathbb{P}(Y \leq y)$ of an annual maximal flow Y at the location of interest. Essentially they claim that the design of a dam should be determined by some high quantile $F^{-1}(p)$, where usually practitioners have to deal with probabilities $p \geq 0.99$, depending on the safety-relevance of the local environment.

In practice the unknown distribution function F needs to be estimated. Let $Y_1, \ldots, Y_n$ denote a sample of annual maximal flows from the past n years. Practitioners typically have to deal with sample lengths n that are rather small compared to p. Often we have that $p > 1 - 1/n$, which means that $F^{-1}(p)$ is supposed to lie beyond the range of observations. Non-parametric sample quantiles $F_n^{-1}(p) = Y_{\lfloor np \rfloor + 1:n}$ computed from order statistics $Y_{1:n} \leq \ldots \leq Y_{n:n}$ are unsatisfactory in such situations. It is therefore of interest to introduce additional assumptions on F. This is where extreme value theory comes into play, with its statistical methods designed for inference on the distribution F at the boundary of, or even beyond the range of observed data.
The majority of practitioners from flood frequency analysis focuses on the block maxima method popularized in the monograph by Gumbel [1958]: Let $Z_1, Z_2, \ldots$ denote a sequence of independent and identically distributed (i.i.d.) random variables and suppose that some mild conditions on $z \mapsto \mathbb{P}(Z_i \leq z)$ are met. Since Fisher and Tippett [1928] it is known that, with increasing block length ℓ, the distribution of block maxima $M_\ell = \max\{Z_1, \ldots, Z_\ell\}$ is approximated by a parametric extreme value distribution G_θ. Assuming that annual maximal flows Y are exactly extreme value distributed, $F = G_\theta$ for some unknown parameter θ, allows us to apply parametric methods, for instance, maximum likelihood estimation $\hat{\theta}_{ML}$ of θ. Plugging in $\hat{\theta}_{ML}$ into $F^{-1} = G_\theta^{-1}$ yields to efficient estimates of high quantiles, provided the parametric assumption is reasonably met.
However, considering annual flows, say, $Y = \max\{Z_{Jan}, Z_{Feb}, \ldots, Z_{Dec}\}$ as maxima over twelve i.i.d. monthly maximal flows Z_{month} within a year does not reflect a realistic scenario. A river flow, just like e.g. temperature and rainfall, is subject to seasonal variability. In the cold season rivers are fed by large masses of melting snow, while warm

periods usually are accompanied by short but heavy rainfalls. Therefore, similar to Lekina et al. [2014], we feel that a parametric model G_θ is not necessarily the right choice to describe the behavior of annual maximal flows. This motivated us to take a closer look at a semi-parametric framework from extreme value theory, referred to as the Pareto-type distributions. These assume only a certain right-tail behavior while the reminder of the distribution is left unspecified. A large part of this dissertation is devoted to the adaption and development of familiar techniques from flood frequency analysis under this semi-parametric framework.

The large estimation uncertainty due to the availability of relatively small sample lengths n is a serious issue in flood frequency analysis, irrespective of whether we apply the classical parametric or the semi-parametric framework. To remedy this problem in the parametric framework, several authors have proposed so-called regional methods, for instance, the Index Flood approach from the seminal work by Dalrymple [1960]. Essentially, they try to increase the local estimation efficiency by gathering observations from many measurement stations in the neighborhood, referred to as *regional* estimation. These methods are based on the idea that stations with similar characteristics (e.g. catchment area and mean elevation) provide somehow similar river flow distributions called the *regional homogeneity* assumption. In the first main contribution of this work we will show how to adapt this concept to the Pareto-type framework, which will then be called regional *heavy-tail homogeneity*.

In some applications practitioners are interested in the joint behavior of flows at two or more river stations. Suppose that the river at some site of interest is fed by two main tributaries and that the subject is to control the confluence of the tributaries, for instance, by construction of a water reservoir [Schulte and Schumann, 2015]. Since catastrophic floods typically occur when flows from both tributaries are simultaneously exceeding extraordinary high levels, not only the margins but also their inter-site dependence is of interest. For the estimation of the joint distribution it is common practice to assume that the corresponding bivariate observations from the past n time units are identically distributed, even though the local environment has changed due to human interventions in nature, for instance, the construction of a dam upstream of one of the tributaries. Several authors have proposed statistical tests for the detection of such change-points in the distribution. Inside the block maxima framework considered here, we are going to present a novel method that is sensitive to changes in extreme value copulas, which will be the second main contribution of this dissertation. Even more, we will present an extension of the procedure that allows to ignore known changes in the marginal distributions.

The third main contribution of this work is devoted to the analysis of conditional heavy-tail behavior and (temporal) trends in the extremes of univariate Pareto-type distributions. In flood frequency analysis a common technique for the detection of temporal trends is the Mann-Kendall test [Yue et al., 2002]. However, this test is rather insensitive against trends in the tail of a distribution, which is of main interest in the applications. We will present procedures that are sensitive to trends in the relevant right tail. It is the aim of our work to demonstrate the opportunities and the limits of modeling non-constant

conditional tail behavior.

The dissertation is organized as follows: In the first chapter we will recapitulate some theoretical foundations from the extreme value literature. We will introduce the families of extreme value and of Pareto-type distributions and we will illustrate them on real data from hydrology. We will also provide some own theoretical results.
Chapter 2 introduces the term heavy-tail homogeneity and its applications in regional flood frequency analysis. We particularly focus on a novel semi-parametric test of regional homogeneity, which, for the detection of heterogeneity in the right tails, turns out to be superior to competing methods from the literature.
In Chapter 3 we will deal with a non-parametric test for change-point detection, which is particularly sensitive to changes in extreme value dependence. The extension of the test is able to ignore known change-points in the marginal series, which allows us to examine the question whether the dependence between river stations has changed or not, irrespective of a possible change in the margins due to the construction of a dam.
Chapter 4 discusses conditional heavy-tail behavior. We present a new estimator for conditional tails based on a model with tail behavior linear in covariates. We particularly focus on detection of non-constant tail behavior, which is a common assumption in flood frequency analysis even in non-stationary approaches [see e.g. Cunderlik and Burn, 2003; El Adlouni et al., 2007; Villarini et al., 2009, and many more].
Further theoretical background from the literature and some own results are summarized in an appendix.

⋆ ⋆ ⋆

First of all, I would like to express my deep gratitude to my advisor Professor Roland Fried for supporting me with countless hours of discussion during these past three years.

I would also like to thank my dissertation reviewers, Professor Walter Krämer and Professor Peter Ruckdeschel, for spending their valuable time during the whole review process.

Special thanks goes to Dr. Axel Bücher and Professor Ivan Kojadinovic. I have learned a lot from both of you during our always pleasant collaboration. I am grateful to Professor Andreas Schumann for providing us with real data from hydrology and for many suggestions from a practitioners point of view. I would also like to thank Professor Stanislav Volgushev, my colleagues at the Department of Mathematics in Bochum and the Department of Statistics in Dortmund for many fruitful discussions and for having a great time.

The financial support of the Deutsche Forschungsgemeinschaft (SFB 823, “Statistical modeling of nonlinear dynamical processes”) is gratefully acknowledged.

Last but not least, I would like to thank my parents, brother, sister and my wife for supporting me throughout the whole life.

Chapter 1

Mathematical preliminaries and a first data application

The first chapter introduces the reader to the basics of our modeling framework. Both sections are devoted to certain families of distributions, which play a central role in extreme value statistics. We begin with the family of multivariate extreme value distributions and the related concept of extreme value dependence. In the second section we will introduce a semi-parametric family of univariate heavy-tailed distributions.

A convenient way to describe any multivariate distribution function $F(\mathbf{y}) = \mathbb{P}(\mathbf{Y} \leq \mathbf{y})$ of a random vector $\mathbf{Y} = (Y_1, \ldots, Y_d)'$ is via a characterization of its margins $F_j(y) = P(Y_j \leq y)$, $j = 1, \ldots, d$, and its inter-site dependence. One common way of describing dependence is based on the following result dating back to the seminal work by Sklar [1959]:

Theorem 1.1 (Sklar's theorem)
Let $C : [0,1]^d \to [0,1]$ be a d-dimensional distribution function with uniform margins called copula and let $F_1, \ldots, F_d : \mathbb{R} \to [0,1]$ be univariate distribution functions. Then the function $F : \mathbb{R}^d \to \mathbb{R}$ defined by

$$F(\mathbf{y}) = C\left(F_1(y_1), \ldots, F_d(y_d)\right), \ \mathbf{y} = (y_1, \ldots, y_d)' \in \mathbb{R}^d, \tag{1.1}$$

abbreviated $F = C(F_1, \ldots, F_d)$, is a distribution function on $\mathbb{R}^d$.
Conversely, for every distribution function F on $\mathbb{R}^d$ there exists a copula C and margins $F_1, \ldots, F_d$ such that (1.1) *holds. If all margins are continuous, then the expression in* (1.1) *is unique and the copula is the distribution function*

$$C(\mathbf{u}) = \mathbb{P}(\mathbf{U} \leq \mathbf{u}), \ \mathbf{u} \in [0,1]^d, \tag{1.2}$$

of the probability transform $\mathbf{U} = (F_1(Y_1), \ldots, F_d(Y_d))'$.

This result allows us to introduce the family of multivariate extreme value distributions via a characterization of all possible margins and all possible copulas, which will be carried out in the next section.

1.1 Multivariate extreme value distributions

Multivariate extreme value distributions play a central role in extreme value statistics. Their importance in the analysis of extremes is comparable with that of the multivariate normal in connection with the analysis of mean values.

Definition 1.2 (MEV distributions)
Let $F : \mathbb{R}^d \to [0,1]$ be a d-dimensional distribution function and suppose that there exist vector sequences $\mathbf{a}_n = (a_{n1}, \ldots, a_{nd})' \in \mathbb{R}^d_+$, $\mathbf{b}_n = (b_{n1}, \ldots, b_{nd})' \in \mathbb{R}^d$, $n \in \mathbb{N}$, and a distribution function $G : \mathbb{R}^d \to [0,1]$ with non-degenerate margins such that

$$\lim_{n\to\infty} F^n\left(a_{n1}z_1 + b_{n1}, \ldots, a_{nd}z_d + b_{nd}\right) = G(\mathbf{z}) \tag{1.3}$$

holds for all continuity points $\mathbf{z} = (z_1, \ldots, z_d)'$ of G. Then we say that F lies in the maximum domain of attraction of the maximum attractor G. G is also called (multivariate) extreme value distribution (for $d > 2$). Accordingly, for $d > 2$, margins and the copula of maximum attractors G are called univariate extreme value distributions and extreme value copula, respectively.

The term *extreme value* stems from the fact that the left-hand side of (1.3) describes the distribution of a componentwise maximum

$$\mathbf{M}_n = \left(\frac{\max_{1\le i\le n}(Z_{11}, \ldots, Z_{n1}) - b_{n1}}{a_{n1}}, \ldots, \frac{\max_{1\le i\le n}(Z_{1d}, \ldots, Z_{nd}) - b_{nd}}{a_{nd}}\right)' \tag{1.4}$$

over a standardized block of n i.i.d. random vectors $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{id})'$, $i = 1, \ldots, n$, with distribution function $F(\mathbf{z}) = \mathbb{P}(\mathbf{Z}_i \le \mathbf{z})$. Therefore, multivariate extreme value distributions are defined as the only possible non-degenerate limits of such componentwise maxima. Surprisingly, every univariate extreme value distribution can be uniquely identified by a three-dimensional parameter vector $\theta = (\mu, \sigma, \xi)'$:

Theorem 1.3 (Fisher and Tippett [1928])
Let G be a univariate extreme value distribution function, that is, the right-hand side of (1.3) *with $d = 1$. Then there exist $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$ and $\xi \in \mathbb{R}$ such that*

$$G(z) = G_{\mu,\sigma,\xi}(z) = \exp\left\{-\left[1 + \xi\frac{z-\mu}{\sigma}\right]^{-1/\xi}\right\},\ z \in \mathbb{I}_{\mu,\sigma,\gamma}, \tag{1.5}$$

where the support of G is given by $\mathbb{I}_{\mu,\sigma,\gamma} = \{z \in \mathbb{R} :\ 1 + \xi(z-\mu)/\sigma > 0\}$. For $\xi = 0$, the distribution function is interpreted as the limit

$$G_{\mu,\sigma,0}(z) = \lim_{\xi\to 0} G_{\mu,\sigma,\xi}(z) = \exp\left\{-\exp\left(\frac{z-\mu}{\sigma}\right)\right\},\ z \in \mathbb{R}.$$

The set $\left\{G_{\mu,\sigma,\xi} :\ (\mu,\sigma,\xi) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}\right\}$ is called the generalized extreme value (GEV) family with parameters μ, σ and ξ called location, scale and shape, respectively. The

family is called general, because it unifies the Fréchet ($\xi > 0$), the Gumbel ($\xi = 0$) and the Reversed-Weibull ($\xi < 0$) sub-families of extreme value distributions.

From Theorem 1.3 we have learned that every margin of a multivariate extreme value distribution is part of the three-parametric GEV family. It turns out that the family of extreme value copulas cannot be parameterized by a finite-dimensional set of parameters. Still, there is some reduction of dimensional complexity possible, which is precisely formulated in a characterization due to Pickands [1981].

Theorem 1.4 (Max-stability and Pickands' characterization)
(i) A d-dimensional copula C is an extreme value copula if and only if it is max-stable, that is, if

$$\left\{C(u_1^{1/n}, \ldots, u_d^{1/n})\right\}^n = C(\mathbf{u}) \text{ for all } \mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d,\ n \geq 1. \tag{1.6}$$

(ii) For every d-dimensional extreme value copula C there exists a convex function $A : \mathcal{S}_{d-1} \to [1/d, 1]$ defined on the $(d-1)$-simplex $\mathcal{S}_{d-1} = \{(t_2, \ldots, t_d) \in [0,1]^{d-1} : \ t_2 + \ldots + t_d \leq 1\}$ and satisfying $\max\{1 - \sum_{j=2}^d t_j, t_2, \ldots, t_d\} \leq A(t_2, \ldots, t_d) \leq 1$ such that

$$C(\mathbf{u}) = \exp\left\{\left(\sum_{j=1}^d \log u_j\right) \cdot A\left(\frac{\log u_2}{\sum_{j=1}^d \log u_j}, \ldots, \frac{\log u_d}{\sum_{j=1}^d \log u_j}\right)\right\} \tag{1.7}$$

holds for all $\mathbf{u} = (u_1, \ldots, u_d)' \in [0,1]^d$. In case of $d = 2$, the converse is also true: Every convex function $A : [0,1] \to [1/2, 1]$ satisfying $\max\{1-t, t\} \leq A(t) \leq 1$ defines a 2-dimensional extreme value copula via equation

$$C(u, v) = \exp\left\{\log(uv) \cdot A\left(\frac{\log v}{\log(uv)}\right)\right\},\ u, v \in [0,1]. \tag{1.8}$$

Theorem 1.4 (ii) states that, provided our variables are extreme value dependent, it suffices to consider the $(d-1)$-dimensional surface $\{(\mathbf{t}, A(\mathbf{t})) : \ \mathbf{t} \in \mathcal{S}_{d-1}\}$ embedded in $[0,1]^d$ instead of the usual copula characterization, which is a d-dimensional surface $\{(\mathbf{u}, C(\mathbf{u})) : \ \mathbf{u} \in [0,1]^d\}$ embedded in $[0,1]^{d+1}$. This reduction of dimensional complexity and a few other helpful properties of extreme value copulas pay off in finite-data applications, for instance, in certain change-point problems presented in Chapter 3.
For an illustration in case of $d = 2$, an exemplary extreme value copula is depicted in Figure 1.1 via its (left) distribution function C and its (right) Pickands dependence function A. The dashed line visualizes the restriction $\max\{1-t, t\} \leq A(t) \leq 1$. Independence (resp. complete dependence) corresponds to the upper (resp. lower) boundary with $A(t) \equiv 1$ (resp. $A(t) = \max\{1-t, t\}$). Loosely speaking, the more the graph of A sags the stronger the amount of extreme value dependence.

Remark 1.5 (Parametric modeling of extreme value dependence)
The Gumbel-Hougaard family $\{C_\vartheta : \ \vartheta \in [1, \infty]\}$ of d-dimensional extreme value copulas

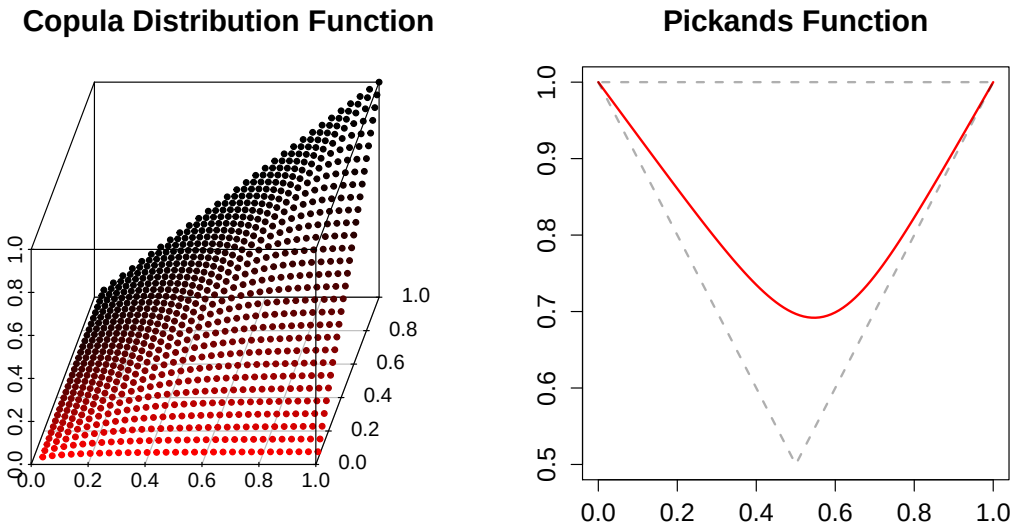

Figure 1.1: (Left) The copula distribution function $(u, v) \mapsto C(u, v)$, $C : [0,1]^2 \to [0,1]$, and (right) its corresponding Pickands dependence function $t \mapsto A(t)$, $A : [0,1] \to [1/2, 1]$ of an exemplary bivariate extreme value copula.

is defined through its Pickands dependence functions by

$$A_\vartheta(\mathbf{t}) = A_\vartheta(t_2, \ldots, t_d) = \left\{ \sum_{j=1}^{d} t_j^\vartheta \right\}^{1/\vartheta}, \; t_1 = 1 - \sum_{j=2}^{d} t_j, \tag{1.9}$$

and where $A_\infty(\mathbf{t}) = \lim_{\vartheta \to \infty} A_\vartheta(\mathbf{t}) = \max\{1 - \sum_{j=2}^{d} t_j, t_2, \ldots, t_d\}$. Note that A_1 and A_∞ correspond to the upper and lower boundaries representing independence and complete positive dependence, respectively. A drawback of this family is that it only covers symmetric (or exchangeable) models in the sense that $C_\vartheta(u_1, \ldots, u_d) = C_\vartheta(u_{\pi(1)}, \ldots, u_{\pi(d)})$ for every permutation π on $\{1, \ldots, d\}$. For instance, this implies that all pairwise dependencies follow the same bivariate copula $C_\vartheta(u, v, 1, \ldots, 1)$, which is a severe restriction for many applications.
Khoudraji [1995] proposed a simple device that is used for an asymmetric extension of symmetric copula models: For $\mathbf{a} = (a_1, \ldots, a_d)' \in [0,1]^d$ we define

$$C_{\mathbf{a},\vartheta}(\mathbf{u}) = \mathbf{u}^{\mathbf{a}} \cdot C_\vartheta(\mathbf{u}^{1-\mathbf{a}}), \; \mathbf{u} \in [0,1]^d, \tag{1.10}$$

where, for notational simplicity, we set $1 - \mathbf{a} = (1 - a_1, \ldots, 1 - a_d)'$ and $\mathbf{u}^{\mathbf{a}} = (u_1^{a_1}, \ldots, u_d^{a_d})'$. It is easy to verify that (1.10) indeed defines a d-dimensional copula (a distribution function with uniform margins). More generally, for arbitrary d-dimensional copulas C_1, C_2 and $\mathbf{a} \in [0,1]^d$ we have that their mixture $C(\mathbf{u}) = C_1(\mathbf{u}^{\mathbf{a}}) \cdot C_2(\mathbf{u}^{1-\mathbf{a}})$ defines a valid cop-

ula. Even more, if C_1 and C_2 are extreme value copulas, so is their mixture C. This is easily verified by the max-stability property of extreme value copulas stated in (1.6).

1.1.1 Estimation of bivariate extreme value distributions

From the previous section we have learned that every multivariate extreme value distribution is determined by the parameter vectors of its GEV margins and by its Pickands dependence function. For simplicity and in view of Chapters 2 and 3, we will restrict ourselves to the case of $d = 2$. Estimation of Pickands dependence functions in arbitrary dimensions is similar and can be found in Gudendorf and Segers [2011].

Let us first start with the estimation of a univariate extreme value distribution. Suppose that $Z_1, \dots, Z_n$ is a sample of independent and identically $GEV(\theta_0)$-distributed random variables with parameter vector $\theta_0 = (\mu_0, \sigma_0, \xi_0)' \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$. A maximum likelihood estimator of θ_0 over $\Theta \subset \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$ is defined by

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} \prod_{i=1}^{n} f_\theta(Z_i), \tag{1.11}$$

where f_θ is the density of the GEV distribution and θ_0 lies in the interior of Θ. The first rigorous proof of the asymptotic normality of $\hat{\theta}_{ML}$ has been published in a recent article:

Theorem 1.6 (Asymptotic normality of the ML estimator; Bücher and Segers [2016])
Let $(Z_i)_{i \geq 1}$ denote a sequence of independent and identically $GEV(\theta_0)$ distributed random variables and let $\Theta \subset \mathbb{R} \times \mathbb{R}_+ \times (-1, \infty)$ be a compact set with θ_0 in its interior. Then, for $n \to \infty$, we have that $\hat{\theta}_{ML} \xrightarrow{a.s.} \theta_0$.
Furthermore, if $\Theta \subset \mathbb{R} \times \mathbb{R}_+ \times (-1/2, \infty)$, then, for $n \to \infty$, we also have that

$$\sqrt{n}\left(\hat{\theta}_{ML} - \theta_0\right) \xrightarrow{D} \mathcal{N}\left(0,\ \mathbf{I}_{\theta_0}^{-1}\right), \tag{1.12}$$

where $\mathbf{I}_{\theta_0} = -\int \frac{d^2}{d\theta d\theta'} \log f_\theta(z)|_{\theta=\theta_0} \cdot f_{\theta_0}(z)\, dz \in \mathbb{R}^{3\times 3}$ is the Fisher information matrix of the GEV distribution [see Beirlant et al., 2006, p. 169 for an explicit expression of $\mathbf{I}_\theta$].

We now turn to the estimation of the Pickands dependence function $A : [0,1] \to [1/2, 1]$. Let $(X, Y)'$ denote a random vector with $\mathbb{P}(X \leq x, Y \leq y) = C\left(F(x), G(y)\right)$ for all $(x, y)' \in \mathbb{R}^2$, whose extreme value copula C is defined through A via equation (1.8). Let $U = F(X)$, $V = G(Y)$ and recall that $C(u, v) = \mathbb{P}(U \leq u, V \leq v)$, $(u, v) \in [0,1]^2$. Then, for arbitrary $t, u \in [0,1]$, we have that

$$\mathbb{P}\left(\max\left\{F(X)^{1/(1-t)}, G(Y)^{1/t}\right\} \leq u\right) = \mathbb{P}\left(U \leq u^{1-t}, V \leq u^t\right) = C(u^{1-t}, u^t) = u^{A(t)}$$

with the convention $u^{1/0} = \lim_{t \searrow 0} u^{1/t} = 0$ for $u \in (0,1)$. This implies that

$$S(t) = \mathbb{E}\left[\max\left\{F(X)^{1/(1-t)}, G(Y)^{1/t}\right\}\right] = \int_0^1 A(t) \cdot u^{A(t)}\, du = \frac{A(t)}{1 + A(t)},$$

which, in turn, gives us

$$A(t) = \frac{S(t)}{1 - S(t)}. \tag{1.13}$$

Let $(X_i, Y_i)'$, $i = 1, \ldots, n$, denote independent copies of $(X, Y)'$. Since $S(t)$ is simply an expected value, equation (1.13) suggests to estimate A via $A_n(t) = S_n(t)/(1 - S_n(t))$ with $S_n(t) = \frac{1}{n}\sum_{i=1}^{n} \max\left\{U_i^{1/(1-t)}, V_i^{1/t}\right\}$, $U_i = F(X_i)$ and $V_i = G(Y_i)$. Because in practice the margins F and G are unknown, we define the so-called pseudo observations $(\hat{U}_i, \hat{V}_i)'$, $i = 1, \ldots, n$, with

$$\hat{U}_i = \frac{1}{n+1}\sum_{\ell=1}^{n} \mathbb{1}\left(X_\ell \leq X_i\right) \text{ and } \hat{V}_i = \frac{1}{n+1}\sum_{\ell=1}^{n} \mathbb{1}\left(Y_\ell \leq Y_i\right), \tag{1.14}$$

and finally set

$$\hat{A}_n(t) = \frac{\hat{S}_n(t)}{1 - \hat{S}_n(t)} \text{ with } \hat{S}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \max\left\{\hat{U}_i^{1/(1-t)}, \hat{V}_i^{1/t}\right\}, \; t \in [0, 1]. \tag{1.15}$$

$\hat{A}_n$ as defined in the last equation is an alternative representation of the madogram estimator from Naveau et al. [2009]. We will thus call $\hat{A}_n$ the madogram estimator of A.
An alternative, so-called endpoint-corrected CFG-estimator studied in Genest and Segers [2009] is probably the most recommended estimator of A in the literature, which has proven to be a good choice in many comparative studies [Genest and Segers, 2009; Bücher et al., 2011]. From our own simulations reported in Appendix A.2 we feel that the madogram estimator is less efficient but still pretty competitive. However, what is helpful for the method presented in Chapter 3 is that the madogram estimator does not need to be end-point corrected and its asymptotic analysis is simpler than that of the CFG-estimator.

Remark 1.7
Note that pseudo observations in (1.14) are defined in accordance with $U_i = F(X_i)$ and $V_i = G(Y_i)$ with margins F and G replaced by their empirical counterparts. The scale $1/(n+1)$ instead of $1/n$ is used to keep pseudo observations away from the boundaries of $[0, 1]^2$. Since we are dealing with multivariate extreme value distributions, one might be also willing to replace the margins by parametric estimates of GEV distributions in the definition of pseudo observations. However, it turns out that estimation of A via empirical marginal distributions can be more efficient, even if the margins F and G are completely known [Genest and Segers, 2009, Sec. 4.2]. If, in addition, one considers the fact that the estimation of GEV parameters is associated with large uncertainty, it is rather advisable to drop the marginal information, that is, to use (1.14), for the estimation of A.

Proposition 1.8 (Asymptotic normality of the madogram estimator)
Suppose that $(X_i, Y_i)'$, $i \geq 1$, is a sequence of i.i.d. extreme value dependent random vectors whose distribution function is continuous and whose Pickands dependence function A is continuously

differentiable on $(0,1)$. *Then, in the space* $\ell^\infty([0,1])$ *equipped with the uniform metric* d_∞ *and for* $n \to \infty$, *we have that*

$$\sqrt{n}\left(\hat{A}_n - A\right) \xrightarrow{D} \left(\{1 + A(t)\}^2 \cdot \int_0^1 \mathbb{C}_C^0(u^{1-t}, u^t)\, du\right)_{t\in[0,1]}, \tag{1.16}$$

where $\mathbb{C}_C^0 = \mathbb{C}_C(0,1,\cdot)$ *is defined in Theorem A.8 of Appendix A.4 with* $d = 2$.

For further details on weak convergence in function spaces $\ell^\infty(I)$ and on empirical copula processes we refer to Appendix A.1 and A.4, respectively. Proposition 1.8, first proven in Bücher et al. [2015], is a direct consequence of Theorem 3.4 from Section 3.5.1. For illustrative purpose, I also give an alternative proof in Appendix A.2, which relies on the functional delta method and on an auxiliary result from Kosorok [2008] concerning Hadamard-differentiability.

1.1.2 Data application part one

We conclude Section 1.1 with the analysis of a data set from the river stations Lichtenwalde1 and Wechselburg1 located at the Mulde river basin in Saxony. Figure 1.2 displays a two-dimensional series $(X_i, Y_i)'$, $i = 1, \ldots, n$, of annual maximal flows observed for $n = 103$ consecutive years (1910–2012). On the left hand side we see the typical heavy-tailed behavior of the marginal series. The inter-site dependence is illustrated on the right hand side via a scatter plot of pseudo observations $(\hat{U}_i, \hat{V}_i)$, $i = 1, \ldots, n$, according to (1.14).

Recall from the introduction of this thesis that the observations can be viewed as componentwise maxima over blocks of, say, 12 bivariate vectors of monthly maximal flows.

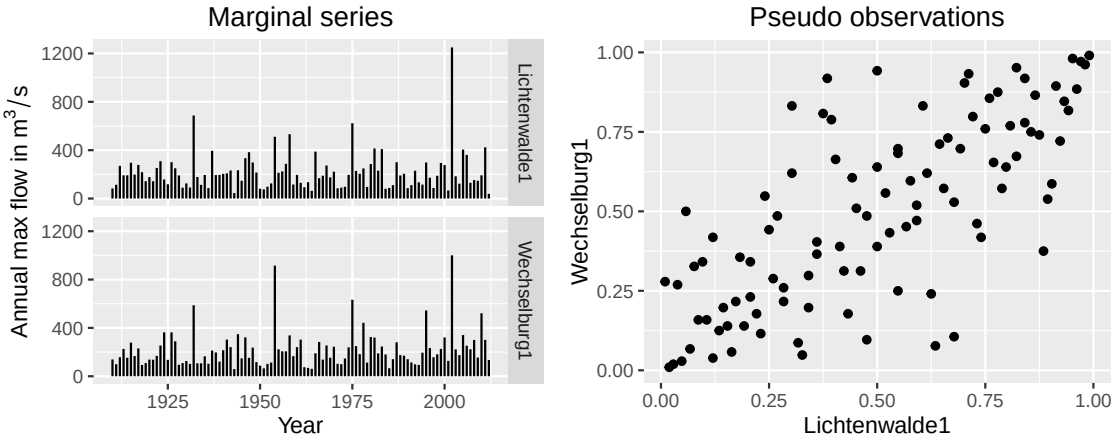

Figure 1.2: (Left) Annual maximal flow series for the period 1910–2012 measured at the neighboring stations Lichtenwalde1 and Wechselburg1. (Right) The corresponding scatter plot of pseudo observations.

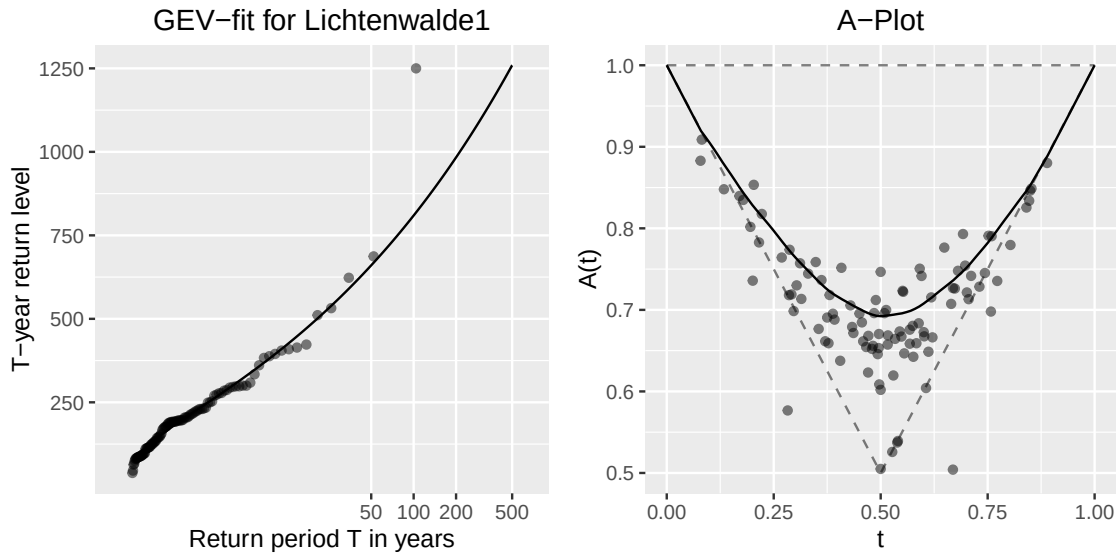

Figure 1.3: (Left) Return level plot for the Lichtenwalde1 series of annual maximal flows and (right) A-plot for the bivariate series of annual maximal flows in Lichtenwalde1 and Wechselburg1. The observation period covers the years 1910–2012.

Thanks to Theorems 1.3 and 1.4, we may conclude that our sample (approximately) follows a bivariate extreme value distribution $H = C(F, G)$ with margins $F \sim GEV(\mu_1, \sigma_1, \xi_1)$ (Lichtenwalde1) and $G \sim GEV(\mu_2, \sigma_2, \xi_2)$ (Wechselburg1), and extreme value copula C defined through its Pickands dependence function A by equation (1.8).
We fitted both marginal series to GEV distributions. The *ML*-estimates are

$$(\hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1) = (150.05,\ 81.30,\ 0.23) \ \text{ and } \ (\hat{\mu}_2, \hat{\sigma}_2, \hat{\xi}_2) = (148.68,\ 73.48,\ 0.29) \tag{1.17}$$

with estimated standard errors $(9.16,\ 7.49,\ 0.09)$ and $(8.34,\ 6.97,\ 0.09)$ obtained from the asymptotic result in (1.12). Our focus lies on the shape parameters ξ_1 and ξ_2. Larger shape values can be interpreted as a greater risk of extraordinary high floods like in the year of 2002. Positive shapes $\xi > 0$ mean that we are in the so-called Fréchet domain with a heavy tail to the right which is further specified in Section 1.2.
Figure 1.3 depicts estimates of (left) the first margin and (right) the Pickands dependence function. The visualization of the second margin is omitted. The points and solid line in both of the plots represent empirical values and estimates of the joint distribution, respectively. More specifically, the left-hand side of Figure 1.3 is called a return level plot with points and solid line defined by

$$\left(\frac{n+1}{n-i+1},\ X_{i:n} \right)_{i=1,\ldots,n} \ \text{ and } \ \left(T,\ \hat{F}^{-1}(1-1/T) \right)_{T \in (1,\ 500]}, \tag{1.18}$$

respectively, where $X_{1:n} \leq \ldots \leq X_{n:n}$ is the ordered sample. In hydrology it is convenient to visualize discharge distributions F via their return level curves $U(T) = F^{-1}(1 - 1/T)$.

This approach points attention to the relevant upper part of the distribution and it brings with a natural interpretation: A flood with a flow level of at least $U(T)$ is expected to occur ones per T time units, for instance, a flood is called *Jahrhundertflut* if its level exceeds $U(T)$ with $T = 100$ years. The order statistics $X_{i:n} = \hat{U}_n(T_{i:n})$ can be considered as empirical return levels for return periods $T_{i:n} = (n+1)/(n-i+1)$, $i = 1, \ldots, n$.
For the left-hand side of Figure 1.3 we set $\hat{F} \sim GEV(\hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1)$ with solid line corresponding to the curve $T \mapsto \hat{F}^{-1}(1 - 1/T)$ and we note that the return periods T on the x-axis are spaced on log-scale. A solid line that is close to the points representing the empirical return levels is interpreted as a good fit of the hypothetical model to the observations. In the case of the Lichtenwalde1 data, the only point far away from the line corresponds to the largest observation. Interestingly enough, the value $X_{n:n} = 1250$ with empirical return period of $T_{n:n} = n + 1$ is not within the 95%-confidence interval $CI = (514.4,\ 1133.8)$ obtained from the estimator

$$\hat{F}^{-1}(p) = \hat{\mu}_1 + \frac{\hat{\sigma}_1}{\hat{\xi}_1} \left\{ (-\log(p))^{-\hat{\xi}_1} - 1 \right\}$$

for the p-quantile for fixed $p = 1 - 1/(n+1) \approx 0.99$. The confidence interval is obtained from the asymptotic normality of the ML estimator (Theorem 1.6) and the delta method.
The right-hand side of Figure 1.3 is a diagnostic tool called A-plot [Cormier et al., 2014], which can be used to visually examine the extreme value dependence assumption. There the solid line corresponds to the graph $\mathbb{G}_{\hat{A}} = \{(t, \hat{A}(t)) : \ t \in [0,1]\}$ of the madogram estimator $\hat{A}$ from (1.15). For the points in the A-plot consider the set

$$\mathbb{M}_C = \left\{ \left(\frac{\log v}{\log(uv)},\ \frac{\log C(u,v)}{\log(uv)} \right) : \ u, v \in [0,1] \right\}$$

defined for arbitrary copulas C. If C is an extreme value copula with Pickands dependence function A, that is, if we have that (1.8) holds, then the set $\mathbb{M}_C$ coincides with the graph $\mathbb{G}_A$ of A. On the other hand, if C is not an extreme value copula, then the set M_C will not coincide with the graph of any Pickands dependence function. The points in the A-plot are defined similar to those in $\mathbb{M}_C$ but with (u, v) and C replaced by their empirical counterparts $(\hat{U}_i, \hat{V}_i)$ and C_n, $i = 1, \ldots, n$. If the extreme value dependence is satisfied, we should expect that most of the points are close to $\mathbb{G}_{\hat{A}}$. Besides this visual examination, we also carried out so-called tests of extreme value dependence [Bücher and Kojadinovic, 2014]. None of the three tests implemented in the `copula` package [Hofert et al., 2015] rejected the null at a nominal level of 5%.

Let us again turn to the estimation of marginal distributions. We initially argued that annual maximal flows are (approximately) GEV distributed, because they can be viewed as maxima over blocks of 12 monthly maximal flows and because of Theorem 1.3. The catch to this matter is that Theorem 1.3 requires that, say, the 12 monthly maxima are independent and identically distributed. We found no evidence against the independence assumption, but, as illustrated in Figure 1.4 with the Lichtenwalde1 data set, there are plausible reasons to drop the idea of identical monthly distributions.
Each of the box plots in Figure 1.4 is computed from $n = 103$ monthly maximal flows of

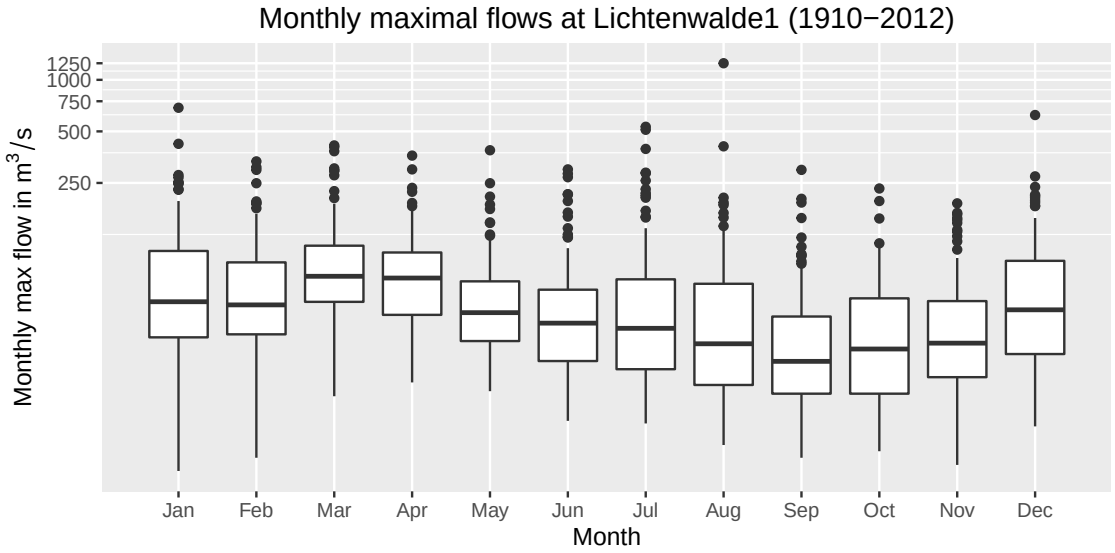

Figure 1.4: Box plots of monthly maximal flows observed at Lichtenwalde1 for each of the 12 months. The y-axis is given on log-scale.

a specific month, with y-axis on log-scale. It seems that the location (median) and dispersion (interquartile range) of the boxes follow a smooth seasonal pattern, for instance, with large location but small dispersion in the spring and vice versa in the summer. Besides the evidence from data exploration, there are also physical reasons for the seasonal variability. In the first months of the year the rivers are fed by large masses of melting snow, which increases the mean flow behavior and sometimes causes winter floods. The physical mechanism behind summer floods is completely different. These typically occur after short but heavy rainfalls, e.g., the severe flood in the summer of 2002.
Summarizing everything up, annual maximal flows from the Mulde river basin follow heavy-tailed distributions, with a rather strong dependence between stations. In hydrology the most popular approach for estimating, say, high quantiles of annual maximal flow distributions is based on the GEV assumption stemming from Theorem 1.3. The latter requires identically distributed observations within blocks, which is reasonably not justified for annual maximal flows. This motivated us to consider estimation of flow distributions under less restrictive, semi-parametric model assumptions.

1.2 A semi-parametric concept of heavy-tailed distributions

The statistical literature provides different characterizations of the term *heavy-tailed*. A definition popularized in extreme value theory calls the right tail of a distribution function F to be heavy, if, for $y \to \infty$, its right tail decay $\bar{F}(y) = 1 - F(y)$ is of polynomial order. More formally, we say that a distribution function F has a heavy tail to the right, if it belongs to the family of *Pareto-type* distribution.

Definition 1.9 (Slowly varying functions and Pareto-type distributions)
(i) A measurable function $L : \mathbb{R}_+ \to \mathbb{R}_+$ is called slowly varying (at infinity) if

$$\lim_{t\to\infty} \frac{L(ty)}{L(t)} = 1 \text{ for all } y > 0. \tag{1.19}$$

(ii) A univariate distribution function F with unlimited support to the right is called a Pareto-type distribution with extreme value index $\gamma > 0$, if there exists a slowly varying function L such that

$$\bar{F}(y) = 1 - F(y) = y^{-1/\gamma} \cdot L(y),\ y > 0. \tag{1.20}$$

The characterization in (1.20) is called semi-parametric, since the parameter γ is of core interest, while the distribution cannot be characterized completely by a finite-dimensional parameter due to the presence of the function L. The extreme value index γ controls the heaviness of the right tail (the polynomial degree of tail decay), with finite (resp. infinite) r-th moment $\int_0^\infty y^r\, dF(y)$ for $\gamma < 1/r$ (resp. $\gamma > 1/r$).

The family of Pareto-type distributions is a very rich class containing many common distribution families. E.g., Student's t_ν with $\gamma = 1/\nu$, Fisher's $F_{m,\ell}$ with $\gamma = 2/\ell$, the $Burr(c, r)$ with $\gamma = 1/(cr)$, the generalized extreme value $GEV(\mu, \sigma, \xi)$ and generalized Pareto distributions $GP(\sigma, \xi)$ with positive shape $\xi = \gamma$. Furthermore, the family of Pareto-type distributions coincides with another important class from extreme value theory [de Haan and Ferreira, 2006, Theorem 1.2.1]:

Theorem 1.10 (Fréchet maximum domain of attraction)
Let F denote a univariate distribution function. The following two statements are equivalent:

(i) F lies in the maximum domain of attraction of an extreme value distribution G with positive shape $\xi > 0$ (Definition 1.2; Fréchet domain)

(ii) F belongs to the family of Pareto-type distributions.

Furthermore, the extreme value index γ of F coincides with the shape ξ of G, that is, $\xi = \gamma$.

The exponential and the normal distribution are typical examples for normal-tailed distributions. These have a right-tail decay of exponential order. The log-normal distribution, which is not of Pareto-type, is also called heavy-tailed by many statisticians, although all its moments are finite. Its right tail decay can be regarded as somewhere between exponential and polynomial. We will continue to call distributions F heavy-tailed to the right, only if they are of Pareto-type.

Typical textbook examples of slowly varying functions are $L(y) = \log(1 + y)$ or simply a constant function $L \equiv c > 0$. They are slower than polynomials in the sense that $y^\alpha L(y) \to \infty$ and $y^{-\alpha} L(y) \to 0$ for $y \to \infty$ and any $\alpha > 0$ [Resnick, 1987, Prop. 0.8]. It is easily seen that if L_1 and L_2 are slowly varying, so is their product $L_1 \cdot L_2$ and their quotient L_1/L_2. With a bit more effort we also show that $L_1 + L_2$ is slowly varying:

Let $y > 0$ be fixed. From the triangular inequality, because of $L_1, L_2 > 0$ and because L_1 and L_2 are slowly varying, we obtain

$$\begin{aligned}\left|\frac{L_1(ty)+L_2(ty)}{L_1(t)+L_2(t)}-1\right| &\leq \left|\frac{L_1(ty)-L_1(t)}{L_1(t)+L_2(t)}\right| + \left|\frac{L_2(ty)-L_2(t)}{L_1(t)+L_2(t)}\right| \\ &< \left|\frac{L_1(ty)}{L_1(t)}-1\right| + \left|\frac{L_2(ty)}{L_2(t)}-1\right| \longrightarrow 0\end{aligned}$$

for $t \to \infty$. This implies (1.19) for $L = L_1 + L_2$.
The previous properties of slowly varying functions are used by Tucker [1968] to prove that sums $X_1 + \ldots + X_d$ over independent and Pareto-type distributed random variables X_j with extreme value index $\gamma_j > 0$ are again Pareto-type distributed, where the extreme value index γ of the sum is obtained from maximization $\gamma = \max\{\gamma_1, \ldots, \gamma_d\}$.
We prove the same statement for maxima $Y = \max\{X_1, \ldots, X_d\}$:

Corollary 1.11
Suppose that the independent random variables X_j are Pareto-type distributed with extreme value index $\gamma_j > 0$, $j = 1, \ldots, d$. Then their maximum $Y = \max\{X_1, \ldots, X_d\}$ is also Pareto-type distributed with extreme value index $\gamma = \max\{\gamma_1, \ldots, \gamma_d\}$.

Proof. Since $\max\{X_1, \ldots, X_d\} = \max\{X_1, \max\{X_2, \ldots, \max\{X_{d-1}, X_d\} \ldots\}\}$, the general result immediately follows from the case of $d = 2$. Let $W = X_1$, $S = X_2$ and

$$F_W(y) = \mathbb{P}(W \leq y) = 1 - y^{-1/\gamma_1} L_1(y) \text{ and } F_S(y) = \mathbb{P}(S \leq y) = 1 - y^{-1/\gamma_2} L_2(y)$$

denote the Pareto-type distribution functions of W and S, respectively. Then, from the independence of W and S, we obtain

$$F_Y(y) = \mathbb{P}(\max\{W, S\} \leq y) = F_W(y) \cdot F_S(y) = 1 - y^{-1/\gamma} \cdot L(y),$$

where we set $\gamma = \max\{\gamma_1, \gamma_2\}$ and

$$L(y) = \begin{cases} L_1(y) + y^{1/\gamma - 1/\gamma_2} L_2(y) - y^{-1/\gamma_2} L_1(y) L_2(y) & \text{, if } \gamma_1 > \gamma_2 \\ L_1(y) + L_2(y) - y^{-1/\gamma} L_1(y) L_2(y) & \text{, if } \gamma_1 = \gamma_2 \\ L_2(y) + y^{1/\gamma - 1/\gamma_1} L_1(y) - y^{-1/\gamma_1} L_1(y) L_2(y) & \text{, if } \gamma_1 < \gamma_2 \end{cases}.$$

We will show that L is a slowly varying function: In case of $\gamma = \gamma_1 > \gamma_2$ (and similar for the third case) we have that

$$\frac{L(ty)}{L_1(ty)} = 1 + (ty)^{1/\gamma - 1/\gamma_2} - (ty)^{-1/\gamma_2} L_2(ty) \longrightarrow 1$$

for $t \to \infty$ and all $y > 0$, since $\gamma > \gamma_2$ and $s^{-1/\gamma_2} L(s) \to 0$ for $s \to \infty$. This implies that

$$\frac{L(ty)}{L(t)} = \frac{L_1(ty)}{L_1(t)} \cdot \frac{L(ty)}{L_1(ty)} \cdot \frac{L_1(t)}{L(t)} \longrightarrow 1$$

for $t \to \infty$ and all $y > 0$. In case of $\gamma = \gamma_1 = \gamma_2$ we have that

$$\frac{L(ty)}{L_1(ty) + L_2(ty)} = 1 - (ty)^{-1/\gamma} \frac{L_1(ty)L_2(ty)}{L_1(ty) + L_2(ty)} \longrightarrow 1$$

for $t \to \infty$ and all $y > 0$, since $\tilde{L} = L_1 \cdot L_2/(L_1 + L_2)$ is again a slowly varying function and because $s^{-1/\gamma}\tilde{L}(s) \to 0$ for $s \to \infty$. This implies that

$$\frac{L(ty)}{L(t)} = \frac{L_1(ty) + L_2(ty)}{L_1(t) + L_2(t)} \cdot \frac{L(ty)}{L_1(ty) + L_2(ty)} \cdot \frac{L_1(t) + L_2(t)}{L(t)} \longrightarrow 1$$

for $t \to \infty$ and all $y > 0$. □

Annual flows Y considered in our applications can be regarded as maxima over winter and summer maximal flows W and S, respectively. From Corollary 1.11 we have learned that the tail behavior of annual flows is determined by that of the season with the heaviest tail, which, for the Mulde river basin in Saxony, is supposed to be the summer season.

1.2.1 Hill's estimator and Weissman's extrapolation formula

Equation (1.20) is a pure characterization of the right tail behavior of F in the sense that it says absolutely nothing about the behavior of F restricted to the interval $(-\infty, u]$, for arbitrary constants $u \in \mathbb{R}$. It thus might be not a big surprise that statistical inference on the extreme value index γ of a Pareto-type distribution F should be based only on the largest observations representing F on the tail region (u, ∞). Roughly speaking, provided there are no further assumptions, the bulk of the observations representing F on $(-\infty, u]$ contains no information on γ at all.

Let Y be a random variable with distribution function $F(y) = \mathbb{P}(Y \leq y)$ defined in (1.20) and let $u > 0$ be a real number. The random variable Y/u satisfying $Y > u$ is called relative excess over the threshold u. From relation (1.20) it immediately follows that

$$\mathbb{P}\left(\frac{Y}{u} \leq z | \, Y > u\right) = 1 - \frac{\bar{F}(uz)}{\bar{F}(u)} \longrightarrow 1 - z^{-1/\gamma} = P_\gamma(z) \text{ for } u \to \infty. \tag{1.21}$$

In words this means that relative excesses over large thresholds u approximately follow a simple parametric law, which depends only on the extreme value index γ of F. The limit P_γ is called Pareto distribution function commonly parameterized $Pareto(\alpha)$ with parameter $\alpha = 1/\gamma > 0$, the so called tail index. Note that $\gamma = \int_{\mathbb{R}} \log(z) \, dP_\gamma(z)$, which suggests to estimate γ from arithmetic means of log-transformed excesses:

Let $Y_1, \ldots, Y_n$ be independent copies of Y. Hill's popular estimator of γ [Hill, 1975] is defined by

$$H_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \log\left(\frac{Y_{n-i+1:n}}{u_{k,n}}\right), \; u_{k,n} = Y_{n-k:n}, \tag{1.22}$$

for integers $k \leq n$ and where we denote order statistics by $Y_{1:n} \leq \ldots \leq Y_{n:n}$. $H_{k,n}$ is interpreted as a *peaks-over-threshold* (POT) approach with peaks $Y_{n-k+1:n}, \ldots, Y_{n:n}$ over the random threshold $u_{k,n} = Y_{n-k:n}$ and with estimation based on maximization of an approximate likelihood function. This approximation considers the relative excesses $Y_{n-k+1:n}/u_{k,n}, \ldots, Y_{n:n}/u_{k,n}$ as independent and exactly distributed from a limit with distribution function P_γ. From this inaccuracy, a bias arises in Hill's estimator, which heavily depends on the tuning parameter k and which may vanish asymptotically (under appropriate conditions) but still is a big issue in finite-sample applications. The following result is treated in full detail in Resnick [2007, Sec. 9.1.2].

Theorem 1.12 (Consistency of Hill's estimator)
Suppose that $(Y_i)_{i\geq 1}$ is a sequence of i.i.d. variables with distribution function $F(y) = \mathbb{P}(Y_i \leq y)$ defined in (1.20) *and let $k = k_n$ denote a sequence of integers satisfying $k \to \infty$ and $k/n \to 0$ for $n \to \infty$. Then we have $H_{k,n} \xrightarrow{\mathbb{P}} \gamma$. Furthermore, assume that*

$$\lim_{n\to\infty} \sqrt{k} \left(\frac{n}{k} \bar{F} \left(F^{-1}(1 - k/n) \cdot y \right) - y^{-1/\gamma} \right) = 0 \tag{1.23}$$

locally uniformly in $(0, \infty]$ and

$$\lim_{n\to\infty} \sqrt{k} \int_1^\infty \left(\frac{n}{k} \bar{F} \left(F^{-1}(1 - k/n) \cdot s \right) - s^{-1/\gamma} \right) \frac{ds}{s} = 0. \tag{1.24}$$

Then we also have that $\sqrt{k}\,(H_{k,n} - \gamma) \xrightarrow{D} \mathcal{N}(0, \gamma^2)$.

Remark 1.13
Assumptions 1.23 and 1.24 can be substituted by a so-called second order regular variation condition such that the asymptotic normality of $\sqrt{k}(H_{k,n} - \gamma)$ holds with a not necessarily centered limiting distribution [de Haan and Ferreira, 2006, Theorem 3.2.5]. No matter what condition is used, they all rely on detailed information on the tail of the distribution which is however usually not available in practice.

The estimation of γ is only a means to an end. In applications the quantity that is actually of interest is a quantile $F^{-1}(p)$ or, equivalently, a return level $U(T) = F^{-1}(1 - 1/T)$ of F for rather high probabilities $p \in (0, 1)$ or long return periods $T > 1$. In terms of the return level function U, equation (1.20) is equivalent to

$$\lim_{t\to\infty} \frac{U(ts)}{U(t)} = s^\gamma \text{ for all } s > 0. \tag{1.25}$$

Suppose that we have collected n observations from F and that we are interested in the return level $U(T)$ of a period of $T = 2 \cdot n$ time units, which is the double of the available observation length n. Classical non-parametric quantile estimates $\hat{F}^{-1}(p) = Y_{\lfloor np \rfloor + 1:n}$ are unsatisfactory in situations, where it is reasonable to assume that the true quantile $F^{-1}(p)$ lies beyond the range of observations, i.e., $F^{-1}(p) > Y_{n:n}$. They can be used to estimate

a return level $U(t)$ of, say, a period of $t = 0.2 \cdot n$ within the range of the data. If we now look again at relation (1.25), we may hope that $U(2n) \approx U(0.2n) \cdot s^\gamma$ for $s = T/t = 10$. This is exactly the idea of the high quantile estimator by Weissman [1978]: For a sample $Y_1, \ldots, Y_n$ from F, an integer $k < n$ and probabilities $p > 1 - k/n$ we set

$$\hat{F}^{-1}(p) = \hat{F}^{-1}(p;\, k, n, \hat{\gamma}) = u_{k,n} \cdot \left(\frac{k}{n(1-p)} \right)^{\hat{\gamma}}, \tag{1.26}$$

where $\hat{\gamma}$ is any consistent estimator of γ, for instance, $\hat{\gamma} = H_{k,n}$. Note that $u_{k,n} = Y_{n-k:n}$ is a classical non-parametric estimator of $U(t)$ with $t = n/k$. Equation (1.26) is also interpreted as extrapolation beyond the range of observations, with $\hat{\gamma}$ controlling the extrapolation width.

The following result from de Haan and Ferreira [2006, Sec. 4.3] shows that the asymptotic behavior of Weissman's estimator stands in relation to that of estimator $\hat{\gamma}$ used for the extrapolation.

Theorem 1.14 (Consistency of Weissman's estimator)
Suppose there exists a real $\rho < 0$ and a function R satisfying $\lim_{t\to\infty} R(t) = 0$ such that

$$\lim_{t\to\infty} \frac{\frac{U(ts)}{U(t)} - s^\gamma}{R(t)} = s^\gamma \frac{s^\rho - 1}{\rho} \quad \text{for all } s > 0. \tag{1.27}$$

In addition, assume that $k = k_n \to \infty$, $k/n \to 0$, $\sqrt{k}R(n/k) \to \lambda \in \mathbb{R}$ and that

$$\sqrt{k} \left(\hat{\gamma} - \gamma,\ \frac{u_{k,n}}{U(n/k)} - 1 \right)' \xrightarrow{D} (\Gamma,\ Z)'$$

for $n \to \infty$, where $(\Gamma,\ Z)'$ is jointly normal. Then, for any sequence of probabilities $p = p_n$ satisfying $n(1-p)/k \to 0$ and $\log(n(1-p))/\sqrt{k} \to 0$ for $n \to \infty$, we have that

$$\frac{\sqrt{k}}{\log\{k/(n(1-p))\}} \left(\frac{\hat{F}^{-1}(p)}{F^{-1}(p)} - 1 \right) \xrightarrow{D} \Gamma, \tag{1.28}$$

where $\hat{F}^{-1}(p)$ is defined in (1.26).

1.2.2 Data application part two

Let us again look at the Lichtenwalde1 series of annual maximal flows X_i, $i = 1, \ldots, n$, displayed in the top left-hand corner of Figure 1.2. In order to apply the estimation technique from the previous subsection we need to select first an integer $k < n$ representing the tail sample size (number of relative excesses). For that purpose we use a visual method called Hill plot, which displays the graph $k \mapsto H_{k,n}$, $k \in \{1, \ldots, n-1\}$. Recommended choices of k are those who lie in an approximately constant (stable) part of the

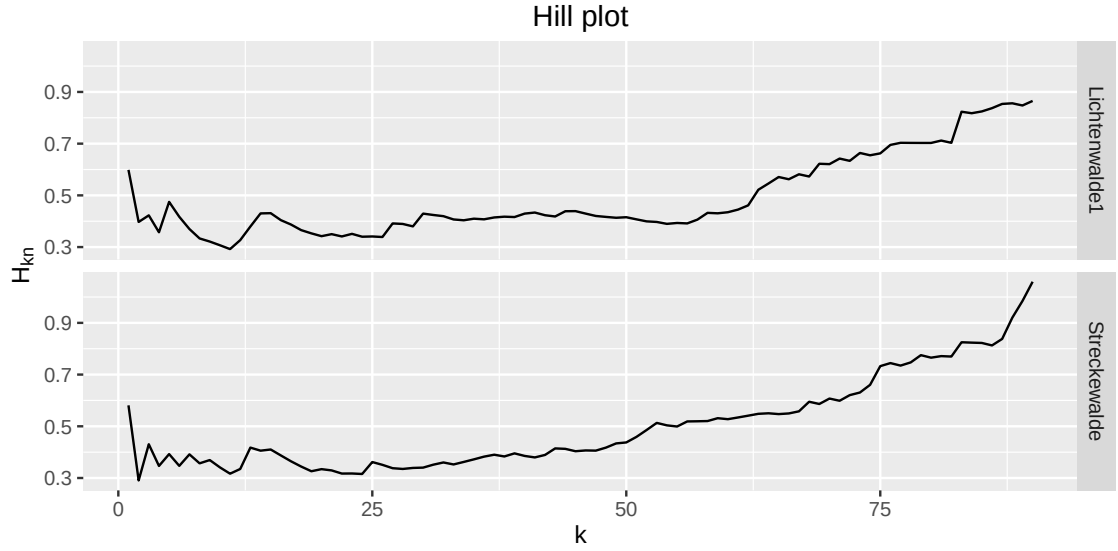

Figure 1.5: Hill plots for the series of annual maximal flows from the stations (top) Lichtenwalde1 and (bottom) Streckewalde.

Hill plot.
Figure 1.5 shows Hill plots for two series of annual maximal flows. The top one corresponds to station Lichtenwalde1. There we can identify a pretty wide stable part in the graph around $k = 40$, with corresponding estimate of γ of around 0.4. For illustrative purpose, we also present the Hill plot of station Streckewalde, where it is hard to identify any reasonably stable part in the graph of its Hill plot. In such cases, we might use alternatives to the usual Hill plot [Drees et al., 2000], or we might use the following simple rule based on the GEV distribution. Suppose that annual maxima are (approximately) GEV distributed. Then, from Remark 3.1 in Gomes and Pestana [2007], we can take

$$k = \lfloor 2n^{2/3} \rfloor \tag{1.29}$$

in $H_{k,n}$, which is asymptotically MSE-optimal for GEV distributions with shape $\xi \neq 1$. Clearly, this GEV-based rule is inconsistent with our intention to present a method that is free of parametric model assumptions. However, from examinations of Hill plots we found that rule (1.29) works out well for many of our annual maximal flow series, e.g., for Lichtenwalde1 with $k = \lfloor 2 \cdot 103^{2/3} \rfloor = 43$.

Finally, we estimated the return level curve $U(T)$, $T > n/k$, for the Lichtenwalde1 series with $k = 43$ and Weissman's extrapolation formula. The result is displayed in Figure 1.6 together with the empirical return levels and the ML estimate based on the GEV assumption (see Section 1.1.2). The curves are reasonably close to each other for return periods of, say, $T < 50$. They deviate much outside the range of observations, which is due to the very different extreme value index estimates $\hat{\xi} = 0.23$ and $\hat{\gamma} = H_{k,n} = 0.42$ for the GEV and semi-parametric approach, respectively (recall that $\xi = \gamma$).

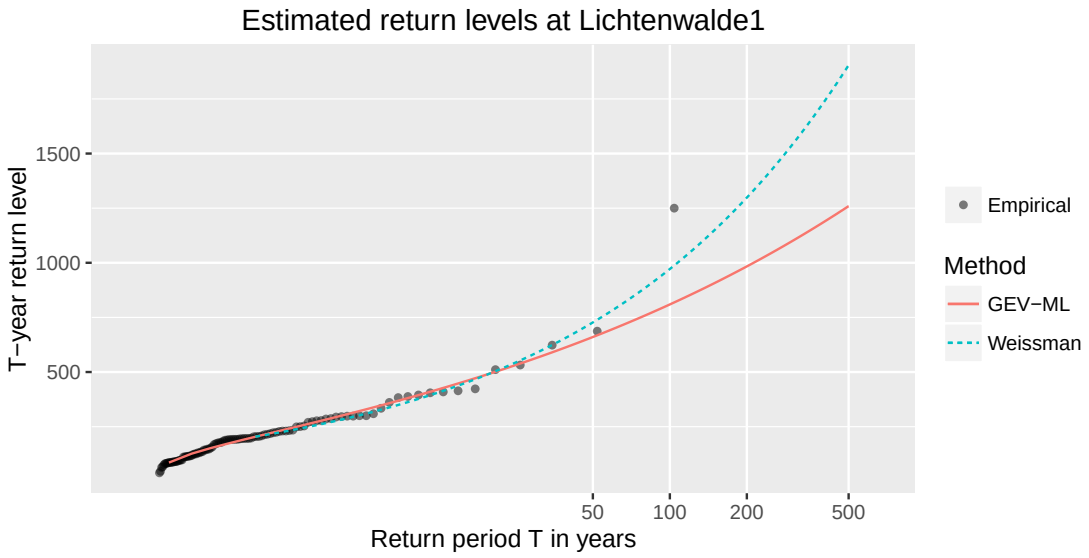

Figure 1.6: Return level plot for the Lichtenwalde1 series of annual maximal flows based on (red) maximum likelihood estimation of GEV parameters and (green) Weissman's extrapolation formula with $k = \lfloor 2n^{2/3} \rfloor$.

Chapter 2

Heavy-tail homogeneity

We assume distribution functions F of Pareto-type, where the right tail behavior of F is characterized by a strictly positive parameter γ, the so-called extreme value index (EVI). In some applications observations from closely related variables are available, with possibly identical EVI γ. If these variables are observed for the same time period, a procedure called BEAR estimator has recently been proposed. We modify this approach allowing for different observation periods and pairwise extreme value dependence of the variables. In addition, we present a new test for equality of the extreme value index. As an application, we discuss regional flood frequency analysis, where we want to combine rather short sequences of univariate observations with very different lengths measured at many stations for joint inference. We illustrate our findings on peak discharges from 18 river stations located at the Mulde basin in Germany, which is known for its severe summer floods, and identify relevant heterogeneous tail behavior, which is not detected by other popular methods.

This chapter is based on the article by Kinsvater et al. [2016].

2.1 Introduction

In environmental sciences we are interested in extreme realizations of a variable Y following some distribution function $F(y) = \mathbb{P}(Y \leq y)$ in order to analyze the frequency of hazardous events such as floods [Dixon et al., 1998; Hosking and Wallis, 2005] or extreme precipitations [Cooley et al., 2007]. Measurements are collected at different locations, with observation lengths for each location being usually rather limited. The analysis is further complicated by the typical heavy-tailed behavior of these quantities. As an application, we will evaluate the flood risk for the Mulde river basin in eastern Germany, which is known for its severe floods, e.g. in 2002 and 2013.
We concentrate on the challenging and practically relevant case of (right) heavy-tailed distributions F. As opposed to parametric or Bayesian procedures mentioned above, we follow a simple and straightforward *peaks-over-threshold* (POT) approach in a semi-parametric framework: We start with independent variables $Y_1, \ldots, Y_n$ from F and de-

note the corresponding ordered values by $Y_{1:n} \leq \ldots \leq Y_{n:n}$. Avoiding stringent model assumptions, we focus on the upper k observations $Y_{n-k+1:n}, \ldots, Y_{n:n}$ representing the right tail of F and assume that these k *excesses* over the random threshold $u_{k,n} = Y_{n-k:n}$ divided by $u_{k,n}$ approximately follow a parametric model (we refer to Section 1.2) with unknown parameter $\gamma > 0$ called extreme value index (EVI).
Useful estimation of γ is very challenging in this semi-parametric framework, especially from small samples n. For estimation, the selection of the tail sample size k is associated with a bias-variance trade-off problem: Smaller numbers k improve the parametric tail approximation in the POT step and thus decrease the bias, while larger numbers increase the effective sample size and thus decrease the variability of estimation. Typically, mean squared error optimal numbers k_{opt} are rather small relative to n, making statistical inference on γ valuable only in sufficiently large samples.
In environmental applications, where we observe the same variables at many sites j with site specific distributions F_j, regional frequency analysis provides methods for pooled estimation to overcome the problem of having only short sequences for each site available. So called *Index Flood* procedures [Hosking and Wallis, 2005, Chapter 1.3] are very popular in hydrology. They are built on the assumption that the quantile functions are identical up to a site-specific scaling factor,

$$\mathcal{H}_{0,IF} : \; F_j^{-1}(p) = \mu_j \cdot G_\theta^{-1}(p) \;\text{ for a given set of sites }\; j \in \{1, \ldots, d\}, \tag{2.1}$$

where $\{G_\theta : \theta \in \Theta\}$ is a predetermined parametric family of distributions and θ, $\mu_j = \mu(F_j)$, $j = 1, \ldots, d$, are unknown parameters for some factor μ (e.g. mean or median) adjusting for the size of each site. Here theory is developed under weaker assumptions than stated in (2.1). Essentially, we suppose that a group of similar distributions shares the same EVI γ, i.e.

$$\mathcal{H}_{0,evi} : \; \gamma_1 = \ldots = \gamma_d = \gamma \;\text{ for some } \gamma > 0, \tag{2.2}$$

where γ_j is the EVI of F_j. We call this assumption *heavy-tail homogeneity*. If $\mathcal{H}_{0,evi}$ holds, γ characterizes the tail decay for the whole region and we call γ the regional extreme value index.
Our approach generalizes the BEAR procedure introduced in Dematteo and Clémençon [2016] to the practically relevant situation where the marginal data sequences are of very different lengths. As an application we consider observations from 18 river stations from the Mulde basin in Saxony, which is a state in Eastern Germany. There the lengths vary between 42 and 100 years of observations per station. The BEAR procedure is based on an asymptotically optimal weighting scheme that allows to decrease the variability in joint estimation. As opposed to these authors we also take the dimension d into account for the bias-variance trade-off problem and we propose a test for hypothesis $\mathcal{H}_{0,evi}$.
An advantage of POT and related methodology is its semi-parametric character, which avoids stringent model assumptions. A drawback of the POT approach is the need for rather long data sequences. The methodology developed in this work remedies this inconvenience by allowing to combine data from dependent variables with very different,

possibly short observation lengths, for joint estimation of the EVI γ. Our main results can be summarized as follows:

- We extend the theoretical results from Dematteo and Clémençon [2016] to the practically relevant case of very different lengths of the marginal samples. This allows us to formulate a joint estimator of γ with an arbitrary weighting of the individual estimators and an asymptotic test for $\mathcal{H}_{0,evi}$.

- For reasonable settings from hydrology with large dimension d, small to moderate marginal sample sizes n_j, $j = 1, \ldots, d$, and under extreme value dependence, the estimation procedure proposed here significantly reduces the estimation error. It turns out to be particularly important to take the dimension d for threshold selection into account to reduce a typically dominant bias.

- Under assumption $\mathcal{H}_{0,IF}$ stated in (2.1), the bias problem is less critical when the proposed test is applied. The nominal level is preserved well in reasonable settings from hydrology. Moreover, when variables are spatially dependent, the new test turns out to be much more powerful for the detection of heterogeneous heavy-tail behavior than competing methods known from the literature.

- In contrast to other popular methods, our test detects deviations from assumption $\mathcal{H}_{0,evi}$ of $d = 18$ river stations located at the Mulde river basin in Germany. The method allows to reduce the detected tail heterogeneity, which is essential for joint tail estimation.

The rest of this chapter is organized as follows. Section 2.2 presents the main theoretical result and Section 2.3 discusses difficulties in finite-sample applications, in particular the bias-variance trade-off problem. A novel semi-parametric test for regional homogeneity is presented in Section 2.4. Section 2.5 reports a simulation study and in Section 2.6 we analyze seasonal maxima from a number of river stations located at the Mulde basin in Germany. We conclude and provide a brief outlook to regional estimation of high quantiles in Section 2.7. Proofs are deferred to Section 2.8.

2.2 Joint estimation of heavy tails

Our goal is to combine information from d sites for joint tail analysis. Let $\mathbf{Y} = (Y_1, \ldots, Y_d)'$ be a random vector with continuous margins $F_j(y) = \mathbb{P}(Y_j \leq y)$, $j = 1, \ldots, d$. Below, the variable Y_j will represent the measurements at station j. Unlike in most of the literature on regional flood frequency analysis, we will not ignore the dependence between the stations. By Theorem 1.1, the joint distribution function F of $\mathbf{Y}$ is uniquely determined by equation (1.1), where $C : [0,1]^d \to [0,1]$ is the corresponding copula, that is, the distribution function of the probability transform $\mathbf{U} = (F_1(Y_1), \ldots, F_d(Y_d))'$. The main assumption of this chapter is as follows:

Assumption 2.1 (Heavy-tail homogeneity)
Each margin F_j, $j = 1, \ldots, d$, is a Pareto-type distribution function with extreme value index $\gamma_j > 0$. Even more, we have that $\mathcal{H}_{0,evi} : \ \gamma_1 = \ldots = \gamma_d = \gamma$ holds.

This allows us to estimate γ from observations of all d sites:
Let $\mathbf{Y}_i = (Y_{i,1}, \ldots, Y_{i,d})'$, $i = 1, \ldots, n$, be independent copies of $\mathbf{Y}$ with i indicating time. Estimation of γ based on polar coordinates and on averaging of local estimates has been proposed in Einmahl et al. [1993] and Dematteo and Clémençon [2016], respectively. However, the usual assumption that all d components $Y_{i,j}$ are observed for the same time units $i \in \{1, \ldots, n\}$ is very restrictive in regional frequency analysis. In our hydrological applications, where data is collected from many sites, this is rarely the case. A more realistic situation is that at least the beginning of recordings are different at the stations. Thus, we assume that we have collected a scheme of observations

$$Y_{a_j+1,j}, Y_{a_j+2,j}, \ldots, Y_{n,j}, \ j = 1, \ldots, d, \tag{2.3}$$

where the integers $1 \leq a_j \leq n$ denote the observation start and $n_j = n - a_j$ the total number of observations at site j. In order to account for possibly very different numbers n_j in the asymptotics, we introduce auxiliary numbers $0 < \tau_j < 1$ and set $a_j = \lfloor n(1-\tau_j) \rfloor$. τ_j is interpreted as the relative sample length available at site j. For $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_d)'$ and $\mathbf{k} = (k_1, \ldots, k_d)' \in \mathbb{N}^d$ with $k_j < n_j$ we set $\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} = \left(H^{(1)}_{k_1,\tau_1,n}, \ldots, H^{(d)}_{k_d,\tau_d,n}\right)'$, where the j-th component $H^{(j)}_{k_j,\tau_j,n}$ is Hill's estimator from (1.22) computed from the k_j largest observations from the j-th marginal sample $Y_{\lfloor n(1-\tau_j) \rfloor + 1,j}, \ldots, Y_{n,j}$.

Assumption 2.2 (Technical assumptions)
The following assumptions, which are also used in Dematteo and Clémençon [2016], are needed in the proof of the main result.

(i) For $j = 1, \ldots, d$, $k_j = k_j(n)$ is an intermediate sequence of integers, i.e. $k_j \to \infty$ and $k_j/n \to 0$ for $n \to \infty$. In addition, $\lim_{n\to\infty} \frac{k_1}{k_j} = c_j$ for some $c_j \in (0, \infty)$.

(ii) We assume that von Mises' condition holds for all $j = 1, \ldots, d$: The derivatives $f_j = F_j'$ exist and satisfy

$$\lim_{x\to\infty} \frac{y f_j(y)}{1 - F_j(y)} = \frac{1}{\gamma_j}, \ j = 1, \ldots, d. \tag{2.4}$$

(iii) For $j = 1, \ldots, d$, $U_j(t) = F_j^{-1}(1 - 1/t)$ and $n \to \infty$ we have

$$\sqrt{k_j} \int_1^\infty \left\{ \frac{n}{k_j} \bar{F}_j \left(U_j(n/k_j) \cdot s \right) - s^{-1/\gamma_j} \right\} \frac{ds}{s} \to 0. \tag{2.5}$$

(iv) For $1 \leq \ell \neq m \leq d$ and $n \to \infty$ we have

$$\sup_{x,y>1} \left| \frac{n}{k_1} \bar{F}_{\ell,m} \left(U_\ell \left(\frac{n}{k_1 x} \right), U_m \left(\frac{n}{k_1 y} \right) \right) - \Lambda_{\ell,m}(x,y) \right| = o\left(\frac{1}{\log k_1} \right), \tag{2.6}$$

where $\bar{F}_{\ell,m}(x,y) = \mathbb{P}\left(Y_\ell > x, Y_m > y\right)$ and existing limits

$$\Lambda_{\ell,m}(x,y) = \lim_{t\to\infty} t \cdot \mathbb{P}\left(Y_\ell > U_\ell(t/x), Y_m > U_m(t/y)\right),\ 1 \leq \ell, m \leq d.$$

Remark 2.3
The assumptions can be viewed as a multivariate extension of those from Theorem 1.12.

(i) Assumption $k_j = o(n)$ is necessary for consistency. For a non-degenerate limiting joint distribution we also need that $k_1/k_j \to c_j \in \mathbb{R}_+$.

(ii) Von Mises' condition is a strengthening of (1.20) [de Haan and Ferreira, 2006, Th. 1.1.11], which is used to establish asymptotic normality of the Hill estimator [Resnick, 2007, Prop. 9.2]. Assumption (2.5), in turn, guarantees that the weak limit of $\sqrt{k_j}(H^{(j)}_{k_j,\tau_j,n} - \gamma_j)$ is centered [Resnick, 2007, Prop. 9.3].

(iii) $\Lambda_{l,m}$ is called upper tail dependence copula [Schmidt and Stadtmüller, 2006] of the joint distribution of $(Y_\ell, Y_m)'$. This concept is used to describe the extremal dependence of a bivariate distribution. A related copula family consists of so called extreme value copulas (EVC), which arise as the only possible limits of copulas of componentwise block maxima (we refer to Section 1.1). In this sense, EVC's themselves characterize extremal dependence. In fact, there is a one-to-one relation between an EVC $C_{\ell,m}$ and the corresponding upper tail dependence copula $\Lambda_{\ell,m}$ of a bivariate distribution given by

$$\Lambda_{\ell,m}(x,y) = (x+y) \cdot \left[1 - A_{\ell,m}\left(\frac{y}{x+y} \right) \right] \quad \text{and} \quad A_{\ell,m}(t) = 1 - \Lambda_{\ell,m}(1-t,t),$$

where $A_{\ell,m}$ is the corresponding Pickands dependence function of $C_{\ell,m}$ (we refer to Theorem 1.4).
The convergence speed in (2.6) is rather slow, because we only need convergence in probability of empirical tail copulas. The same assumption with a faster convergence rate can be used to prove asymptotic normality of empirical tail copulas [Schmidt and Stadtmüller, 2006].

The following theorem extends the main result in Dematteo and Clémençon [2016].

Theorem 2.4 (Joint weak convergence)
We set $\mathbf{1} = (1, \ldots, 1)' \in \mathbb{R}^d$. *Under assumptions 2.1 and 2.2 we have, for* $n \to \infty$, *that*

$$\sqrt{k_1}\left(\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \gamma\mathbf{1}\right) \xrightarrow{D} \mathcal{N}\left(0, \gamma^2 \cdot \Sigma\right) \tag{2.7}$$

holds, where $\Sigma \in \mathbb{R}^{d\times d}$ *is defined componentwise by*

$$\Sigma_{\ell,m} = c_\ell \cdot c_m \cdot (\tau_\ell \wedge \tau_m) \cdot \Lambda_{\ell,m}\left((\tau_\ell c_\ell)^{-1}, (\tau_m c_m)^{-1}\right),\ 1 \le \ell, m \le d, \tag{2.8}$$

and where $x \wedge y = \min(x, y)$. *On the diagonal,* $\ell = m$, *the expression is simplified to* $\Sigma_{\ell,\ell} = c_\ell$.

Let $\mathbb{W} = \{\boldsymbol{w} \in \mathbb{R}^d : \sum_{i=1}^d w_i = 1\}$ denote a set of d-dimensional weightings. Then, from continuity, and for arbitrary $\boldsymbol{w} \in \mathbb{W}$ we have that

$$\sqrt{k_1}\left(\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w}) - \gamma\right) \xrightarrow{D} \mathcal{N}(0, \gamma^2 \boldsymbol{w}'\Sigma\boldsymbol{w}) \text{ for } \hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w}) = \boldsymbol{w}' \cdot \mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n}. \tag{2.9}$$

This suggests to use $\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w}_{opt})$ as a joint estimator of γ, where we set

$$\boldsymbol{w}_{opt} = \operatorname*{arg\,min}_{\boldsymbol{w}\in\mathbb{W}} \lim_{n\to\infty} \operatorname{Var}(\boldsymbol{w}' \cdot \mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n}) = \operatorname*{arg\,min}_{\boldsymbol{w}\in\mathbb{W}} \boldsymbol{w}'\Sigma\boldsymbol{w}.$$

In Dematteo and Clémençon [2016] only non-negative weights $\boldsymbol{w}$ were considered for the minimization problem. Here, however, we do not apply this restriction, which allows us to solve the minimization problem by the Lagrange multiplier technique with solution

$$\boldsymbol{w}_{opt} = \left(\mathbf{1}'\Sigma^{-1}\mathbf{1}\right)^{-1}\Sigma^{-1}\mathbf{1},\quad \mathbf{1} = (1,\ldots,1)' \in \mathbb{R}^d, \tag{2.10}$$

provided Σ is nonsingular.

2.3 Estimation of Σ and selection of k in finite samples

The consistent estimation of the covariance matrix Σ and especially the selection of the integer vector $\mathbf{k}$ is crucial in finite-sample applications. We are particularly interested in applications under the additional extreme value dependence assumption with only short data sequences available.

2.3.1 Estimation of Σ under extreme-value dependence

Recall that the covariance matrix Σ is defined through its components by

$$\Sigma_{\ell,m} = c_\ell \cdot c_m \cdot (\tau_\ell \wedge \tau_m) \cdot \Lambda_{\ell,m}\left((\tau_\ell c_\ell)^{-1}, (\tau_m c_m)^{-1}\right),\ 1 \le \ell, m \le d.$$

It thus suffices to replace c_j, τ_j and the upper tail dependence copulas $\Lambda_{\ell,m}$ by consistent estimates k_1/k_j, n_j/n and $\hat{\Lambda}_{\ell,m}$, respectively.

Let $N = N(\ell, m)$, denote the number of independent copies of $(X_\ell, X_m)'$ that are available for estimation of $\Lambda_{\ell,m}$. In the situation of (2.3) we have $N(\ell, m) = n_\ell \wedge n_m$ the minimum of n_ℓ and n_m. Let u_ℓ and u_m denote two threshold values. The empirical estimator of $\Lambda_{\ell,m}$ studied in Schmidt and Stadtmüller [2006] basically counts the number of pairs (X_ℓ, X_m)

satisfying $X_\ell > u_\ell$ and $X_m > u_m$ divided by the number K of pairs with $X_\ell > u_\ell$ or $X_m > u_m$. This means that, similar to the univariate POT approach (we refer to Section 1.2.1), only a representative number $K = K(\ell, m)$ of observations from the joint tail region $(u_\ell, \infty) \times (u_m, \infty)$ is taken effectively into account for estimation of $\Lambda_{\ell,m}$.

However, we consider componentwise maxima in our application. It is well known that extreme value copulas are the only possible limits of copulas of componentwise maxima of i.i.d. vectors. Thanks to Theorem 1.4, we may assume that our observations are extreme value dependent, which, in turn, allows us to set

$$\Lambda_{\ell,m}(x,y) = (x+y) \cdot \left[1 - A_{\ell,m}\left(\frac{y}{x+y}\right)\right], \tag{2.11}$$

where $A_{\ell,m}$ is the corresponding Pickands dependence function. Several estimators of the Pickands function $A_{\ell,m}$ are known from the literature. In particular, the corrected CFG-estimator $A^{CFG}_{N,c}$ from Genest and Segers [2009] offers high efficiency.
As opposed to the empirical estimator, an estimator of $\Lambda_{\ell,m}$ based on the extreme value dependence assumption, e.g., by plugging in the CFG-estimator, is able to take all N available observations into account. This advantage over the empirical estimator turns out to be crucial for an acceptable type-1 error of the test from Section 2.4 when only small samples N are available. In what follows, we denote the empirical and CFG-based estimators of Σ by $\hat{\Sigma}_{emp}$ and $\hat{\Sigma}_{ev}$, respectively.

Proposition 2.5
Suppose that (2.7) *holds and let* $\hat{\Sigma}$ *satisfy* $\hat{\Sigma} \xrightarrow{\mathbb{P}} \Sigma$ *for* $n \to \infty$. *Further, let* $\hat{\gamma}_{opt} = \hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\hat{\boldsymbol{w}}_{opt})$ *with* $\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})$ *defined in* (2.9) *and* $\hat{\boldsymbol{w}}_{opt} = \left(\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}\right)^{-1}\hat{\Sigma}^{-1}\mathbf{1}$. *Then, for* $n \to \infty$, *we have that*

$$\sqrt{k_1}\left(\hat{\gamma}_{opt} - \gamma\right) \xrightarrow{D} \mathcal{N}\left(0,\ \gamma^2 \cdot (\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1}\right). \tag{2.12}$$

In order to study the gain in efficiency of the optimal weighting scheme, we also consider the joint estimator with weights $\boldsymbol{w}_{ind} = \mathbf{k}/(\mathbf{1}'\mathbf{k})$ in the simulations. Note that this corresponds to optimal weighting under the assumption of upper tail independence, that is, $\Lambda_{\ell,m} \equiv 0$ for $\ell \neq m$.

2.3.2 Selection of k for joint estimation

The choice of the integers $\mathbf{k}$ representing the effective tail sample size is associated with a bias-variance trade-off problem. Several methods were proposed to solve this problem in the univariate setting [Drees et al., 2000]. It soon becomes apparent that the optimal numbers k_j for marginal estimation do not coincide with those $k_j^{(d)}$ for joint estimation from $d > 1$ sites, even under independence:
Suppose that the observations follow a multivariate extreme value distribution with identical marginal distributions $F_1 = \ldots = F_d \sim GEV(\mu, \sigma, \gamma)$ and independent components, that is, with $C(\mathbf{u}) = u_1 \cdots u_d$ in (1.1). Let n_j denote the number of observations of component $j = 1, \ldots, d$. The optimal k_j value that minimizes the asymptotic mean squared

error (MSE) of the marginal Hill estimator from n_j independent observations is given by $k_j^{(1)} = \lfloor 2n_j^{2/3} \rfloor$ [Gomes and Pestana, 2007, Remark 3.1]. In fact, in this simple case, all observations are $N = \sum_{j=1}^{d} n_j$ realizations of the same GEV distribution, which implies that a total number $K = \lfloor 2N^{2/3} \rfloor$ out of N observations should be used and which, in turn, suggests to choose $k_j^{(d)} = \lfloor 2n_j/N^{1/3} \rfloor < k_j^{(1)}$ for the joint estimation. If $n_1 = \ldots = n_d$, we obtain $k_j^{(d)} = \lfloor 2n_j^{2/3}/d^{1/3} \rfloor$, which means that optimal numbers $k_j^{(d)}$ should decrease with increasing dimension d. Indeed, from our simulation results presented in Section 2.5.1 we find that the performance of the joint Hill estimator with $k_j^{(d)} = \lfloor 2n_j^{2/3}/d^{1/3} \rfloor$ is overall superior to that with marginally optimal values $k_j^{(1)} = \lfloor 2n_j^{2/3} \rfloor$.

Let us return to the general case with not necessarily identical margins and dependent components. For a more rigorous thought, suppose that each margin F_j is a member of the Hall-Welsh class [Hall and Welsh, 1985; Gomes and Pestana, 2007] such that

$$F_j^{-1}\left(1 - \frac{1}{t}\right) = M_j \cdot t^{\gamma}\left(1 + \frac{\gamma \beta_j t^{\rho_j}}{\rho_j} + o(t^{\rho_j})\right) \tag{2.13}$$

holds for $t \to \infty$, extreme value index $\gamma > 0$, constants $M_j > 0$ and so-called second order parameters $\rho_j < 0$, $\beta_j \neq 0$, $j = 1, \ldots, d$. The Hall-Welsh class is a rich subset of the Pareto-type distributions. This class allows to study the asymptotic normality of the Hill estimator with not necessarily vanishing bias term. Although not proven here, it is likely that the asymptotic normality of the joint Hill estimator $\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})$ stated in (2.9) is also valid for margins within the Hall-Welsh class, with identical limiting variance but with an additional deterministic bias term. Thus, following Gomes and Pestana [2007, Sec. 3.1] for the bias calculation, the mean squared error of $\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})$ is approximated by

$$MSE(\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})) \approx \gamma^2 \boldsymbol{w}^T \Gamma(\mathbf{k}) \boldsymbol{w} + \gamma^2 \left(\sum_{j=1}^{d} w_j \beta_j \frac{(n_j/k_j)^{\rho_j}}{1 - \rho_j} \right)^2 \tag{2.14}$$

for large n with matrix $\Gamma(\mathbf{k}) \approx \frac{1}{k_1}\Sigma$ defined componentwise by

$$(\Gamma(\mathbf{k}))_{\ell,m} = \frac{\tau_\ell \wedge \tau_m}{k_\ell k_m} \cdot \Lambda_{\ell,m}\left(\frac{k_\ell}{\tau_\ell}, \frac{k_m}{\tau_m}\right), \; 1 \leq \ell, m \leq d.$$

From a theoretical point of view the optimal combination of weights $\boldsymbol{w}$ with $\sum_j w_j = 1$ and integers $\mathbf{k}$ with $1 \leq k_j < n_j$ is achieved by minimizing (2.14) with respect to both $\boldsymbol{w}$ and $\mathbf{k}$. However, this high dimensional and nonlinear minimization problem is computationally expensive and associated with the estimation of the second order parameters β_j and ρ_j. Having our applications from hydrology in mind, we did not further pursue this problem.

2.4 Heavy-tail ANOVA

The statistic from the following proposition can be used to investigate the validity of Assumption 2.1.

Proposition 2.6 (Heavy-tail ANOVA)
Suppose that (2.7) *holds for some positive definite covariance matrix* Σ. *Then, for* $\boldsymbol{w} \in W$, *a weakly consistent estimator* $\hat{\Sigma}$ *of* Σ *and* $n \to \infty$, *we have that*

$$\begin{aligned}\tilde{W}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w}) &= \frac{k_1}{\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})^2} \left(\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})\mathbf{1}\right)' \hat{\Sigma}^{-1} \left(\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})\mathbf{1}\right) \\ &\xrightarrow{D} \lambda_{\boldsymbol{w}} \cdot Z_1^2 + \sum_{j=2}^{d-1} Z_j^2,\end{aligned}$$

where $Z_1, \ldots, Z_{d-1}$ *are i.i.d. standard normal and* $\lambda_{\boldsymbol{w}} = \mathbf{1}^T\Sigma^{-1}\mathbf{1} \cdot \boldsymbol{w}^T\Sigma\boldsymbol{w} \geq 1$. *In addition, let* $W_{\mathbf{k},\boldsymbol{\tau},n} = \tilde{W}_{\mathbf{k},\boldsymbol{\tau},n}(\hat{\boldsymbol{w}}_{opt})$ *with* $\hat{\boldsymbol{w}}_{opt} = (\mathbf{1}^T\hat{\Sigma}^{-1}\mathbf{1})^{-1} \cdot \hat{\Sigma}^{-1}\mathbf{1}$. *Then, for* $n \to \infty$, *we have that* $W_{\mathbf{k},\boldsymbol{\tau},n} \xrightarrow{D} \chi^2_{d-1}$. *Conversely, if* $\gamma_\ell \neq \gamma_m$ *for some* $1 \leq \ell \neq m \leq d$, *we have* $W_{\mathbf{k},\boldsymbol{\tau},n} \xrightarrow{\mathbb{P}} \infty$.

According to these results, $W_{\mathbf{k},\boldsymbol{\tau},n}$ provides an asymptotic significance test of $\mathcal{H}_{0,evi}$ under assumptions 1.-4., which is consistent against arbitrary fixed alternatives.

Heavy-tail ANOVA test:
Reject $\mathcal{H}_{0,evi}$ at a significance level $\alpha \in (0,1)$, if $W_{\mathbf{k},\boldsymbol{\tau},n}$ exceeds the $(1-\alpha)$-quantile of the χ^2 distribution with $d-1$ degrees of freedom.

The performance of this test for finite samples can be very poor, even for large n. A reason for this is the bias of the Hill estimator arising from tail approximation in the POT step. Very different marginal bias terms can lead to a rejection of $\mathcal{H}_{0,evi}$, even if the null hypothesis is true. This bias issue almost vanishes under the classical Index Flood assumption $\mathcal{H}_{0,IF}$ stated in (2.1), which implies that all marginal variables are equal in distribution up to scale and because Hill's estimator is scale invariant. For instance, suppose that each margin F_j is a generalized extreme value distribution $GEV(\mu_j, \sigma_j, \gamma_j)$ with location, scale and shape μ_j, σ_j and γ_j, respectively. This setting is of practical relevance whenever observations $X_{i,j}$ can be considered as block maxima, e.g. in many flood studies and also in the application presented in Section 2.6. Let $\delta_j = \mu_j/\sigma_j$ denote the location-scale ratio of F_j and assume that $\gamma_j > 0$, $j = 1, \ldots, d$. In this GEV setting, hypothesis $\mathcal{H}_{0,IF}$ from (2.1) can be reformulated to

$$\mathcal{H}_{0,IF} = \mathcal{H}_{0,evi} \cap \mathcal{H}_{0,delta}, \tag{2.15}$$

which means that $\mathcal{H}_{0,delta}: \ \delta_1 = \ldots = \delta_d$ holds in addition to $\mathcal{H}_{0,evi}$ from (2.2).

A fundamental question concerning the test is which alternatives can be detected with satisfactory power. The general idea of the test is to quantify the distance between the vector of individual estimates and a vector with all components equal to a weighted average. If the distance is large, we should reject the null hypothesis of equal components.

Our test considers a Mahalanobis distance, which can be seen as an Euclidean distance in new coordinates and which is able to account for different uncertainties of deviations in different directions. In the case of independence, i.e., if we set $\hat{\Sigma} = \mathrm{diag}(\hat{c}_1, \ldots, \hat{c}_d)$ the diagonal matrix with entries $\hat{c}_j = k_1/k_j$, our test can be viewed as a classical analysis of variance (ANOVA) statistic $W = \hat{\gamma}^{-2} \sum_{j=1}^{d} k_j (H^{(j)} - \mathbf{w}'\mathbf{H})^2$, with γ_j interpreted as the approximate population mean of log-transformed peaks divided by the threshold. Interestingly enough, no uniformly most powerful test exists for the problem of testing the equality of $d > 2$ population means, even in the simplest ANOVA setting with independent normal observations, all having the same known variance [Lehmann and Romano, 2005, Sec. 7.3]. For instance, the alternative $\gamma_1 > \gamma_2 = \ldots = \gamma_d$ is detected best by an one-sided two-sample t-test considering the difference between the first and an average of the last $d-1$ components. However, it is known that the classical ANOVA approach is optimal in some maximin sense, which basically means that the test is good in detecting departures $\mathcal{H}_{1,evi}$ from $\mathcal{H}_{0,evi}$ that result from a combination of many small deviations of γ_j values from the average. Alternatively, one could consider a max-type test statistic $M_{\mathbf{k},\boldsymbol{\tau},n}$ (we refer to Appendix A.3) based on the maximum instead of the Euclidean distance. Such max-type tests usually offer better power against a few large deviations in some of the components. In practice the latter situation is less likely to occur, since experienced practitioners should be able to sort out largely deviating variables manually when the group of d variables is selected in a preliminary step. E.g., hydrologists select groups of river stations that share similar site characteristics, which naturally results in rather homogeneous groups of distributions. From this point of view it is more meaningful to apply the Mahalanobis-type test proposed here.
One needs to keep in mind that we de-correlate the data in both these test statistics in order to get a pivotal limiting distribution. Thus, in general, these two tests are constructed to detect many small and a few large deviations, respectively, in terms of linear transformations of the original components, so that the previous remarks on $W_{\mathbf{k},\boldsymbol{\tau},n}$ and $M_{\mathbf{k},\boldsymbol{\tau},n}$ can only be taken as a rough guideline. For a related discussion about maximin optimality of ANOVA and max-type tests we refer to Arias-Castro et al. [2011].

2.5 Simulation study

Motivated by our illustration presented in Section 2.6, we focus on simulations with multivariate extreme value distributed sequences. More precisely, we draw independent vector valued realizations from d dimensional distributions $F = C(F_1, \ldots, F_d)$ with (univariate) extreme value distributed margins $F_j \sim GEV(\mu_j, \sigma_j, \gamma_j)$, positive extreme value index $\gamma_j > 0$ and extreme value copula C from the family

$$C_{\boldsymbol{\theta},\mathbf{a}}(\mathbf{u}) = C_{\theta_1}(\mathbf{u}^{\mathbf{a}}) \cdot C_{\theta_2}(\mathbf{u}^{1-\mathbf{a}}), \tag{2.16}$$

where $\mathbf{u}^{\mathbf{a}} = (u_1^{a_1}, \ldots, u_d^{a_d})$, $1 - \mathbf{a} = (1 - a_1, \ldots, 1 - a_d)$, $\boldsymbol{\theta} = (\theta_1, \theta_2) \in [1, \infty)^2$, $\mathbf{a} = (a_1, \ldots, a_d) \in [0, 1]^d$ and C_θ is the d-dimensional Gumbel(θ) copula. The construction principle (2.16) is known as Khoudraji's device [Khoudraji, 1995; Durante and Salvadori,

2010]. It is used to account for possible asymmetry in the dependence, which is also present in our illustration but not covered by common one-parameter copula families.
Since all considered methods are scale invariant, we pay particular attention to the performance depending on the choice of $\delta_j = \mu_j/\sigma_j$. Recall that under the classical homogeneity assumption stated in (2.1) we have $\gamma_1 = \ldots = \gamma_d = \gamma$ and $\delta_1 = \ldots = \delta_d = \delta$.
Most simulations are carried out for dimension $d = 5$ and the following parameter values:

- $n \in \{50, 100\}$ (maximal sample length)
- $\boldsymbol{\tau} \in \{(1,1,1,1,1)', (1, 0.9, 0.8, 0.7, 0.6)'\}$ (relative marginal sample lengths)
- $\gamma \in \{0.25, 0.5, 0.75\}$ (extreme value index)
- $\delta \in [1,\ 3]$ (location-scale ratio)
- $\boldsymbol{\theta} \in \{(1,1)', (1.5, 2.5)'\}$ (strength of dependence)
- $\mathbf{a} = (0.9, 0.7, 0.5, 0.3, 0.1)'$ (asymmetry of dependence)

These scenarios are supposed to cover many settings from regional flood frequency analysis. We also studied the performance for $d = 10$ and $d = 15$, but many results were qualitatively similar to those for $d = 5$ and are thus not reported in full detail. For $d = m \cdot 5$, $m \in \mathbb{N}$, and $\boldsymbol{\tau}, \mathbf{a} \in \mathbb{R}^5$ from above, the relative sample lengths and asymmetry coefficients were set to $\boldsymbol{\tau}_m = (\boldsymbol{\tau}', \boldsymbol{\tau}', \ldots, \boldsymbol{\tau}')' \in \mathbb{R}^{md}$ and $\mathbf{a}_m = (\mathbf{a}', \mathbf{a}', \ldots, \mathbf{a}')' \in \mathbb{R}^{md}$, respectively.
However, we found that the new test based on statistic $W_{\mathbf{k},\boldsymbol{\tau},n}$ tends to get liberal with increasing dimension d (at constant n). Based on our simulation results, we decided heuristically to multiply the statistic with an asymptotically negligible factor of $1 - d/(5N)$ with $N = min_{1 \le j \le d} n_j$ at the cost of a loss of power.
Simulations were carried out in R [R Core Team, 2015]. In particular, we used code provided by the packages `copula` [Hofert et al., 2015], `fExtremes` [Würtz, 2013], `fgof` [Kojadinovic and Yan, 2012a] and `homtest` [Viglione, 2012] available on CRAN.

2.5.1 Joint estimation of γ

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent copies with the distribution F given above and $\gamma_1 = \ldots = \gamma_d = \gamma$. In contrast to many other comparable studies from hydrology, where typically $\gamma < 0.3$ is used, we are also interested in more heavy-tailed scenarios with, say, $\gamma = 0.5$ (see also our illustration in Section 2.6). In this case, the L-moment estimator of the shape γ of the GEV distribution is not advisable [Hosking et al., 1985]. This is also confirmed by our simulation results and therefore, we decided to use only a maximum likelihood (ML) based approach $\hat{\gamma}_{ML}$ as a benchmark for the performance of several versions of estimator $\hat{\gamma}_H = \hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})$ from (2.9).
Let $n_j = \lfloor n\tau_j \rfloor$ denote the number of observations available for component j. We consider the following joint estimators of γ:

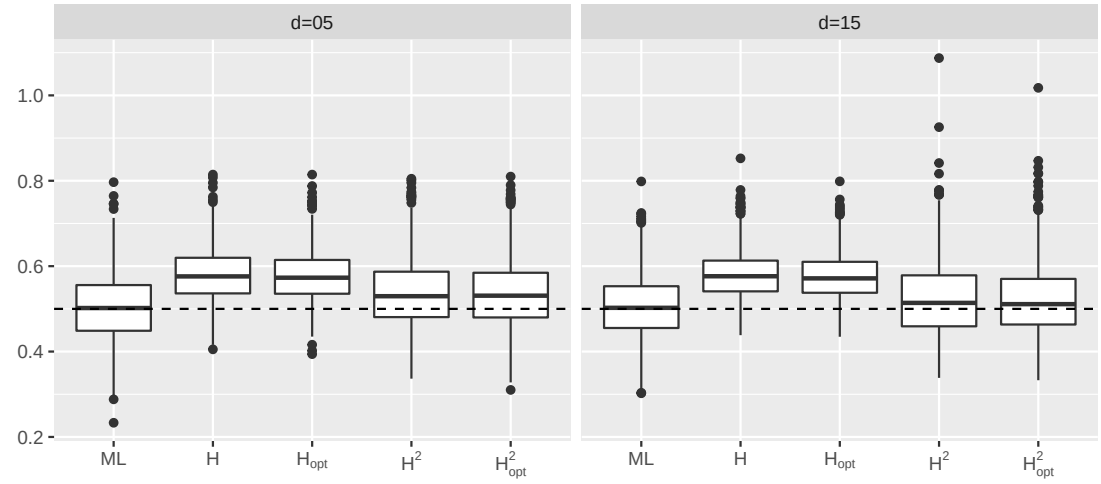

Figure 2.1: Each box plot is derived from 1000 independent realizations of a joint Hill estimator applied on multivariate data with distribution given in the beginning of Section 2.5 and $n = 100$, $\gamma = 0.5$, $\delta = 2$, $\boldsymbol{\theta} = (1.5, 2.5)$, $\mathbf{a} = (0.9, 0.7, 0.5, 0.3, 0.1)$, $\boldsymbol{\tau} = (1, 0.9, 0.8, 0.7, 0.6)$, (left panel) $d = 5$ and (right panel) $d = 15$.

- $\hat{\gamma}_{ML} = \sum_{j=1}^{d} w_j \hat{\gamma}_{ML}^{(j)}$ with $w_j = n_j / \sum_{\ell=1}^{d} n_\ell$ (ML)
- $\hat{\gamma}_H(\boldsymbol{w}_{ind})$ with $k_j = \lfloor 2n_j^{2/3} \rfloor$ (H)
- $\hat{\gamma}_H(\hat{\boldsymbol{w}}_{opt})$ with $\hat{\Sigma} = \hat{\Sigma}_{ev}$ and $k_j = \lfloor 2n_j^{2/3} \rfloor$ (H_{opt})
- $\hat{\gamma}_H(\boldsymbol{w}_{ind})$ with $k_j^{(d)} = \lfloor 2n_j^{2/3}/d^{1/3} \rfloor$ $(H^{(2)})$
- $\hat{\gamma}_H(\hat{\boldsymbol{w}}_{opt})$ with $\hat{\Sigma} = \hat{\Sigma}_{ev}$ and $k_j^{(d)} = \lfloor 2n_j^{2/3}/d^{1/3} \rfloor$ $(H_{opt}^{(2)})$

$\hat{\gamma}_{ML}^{(j)}$ denotes the ML estimator of the GEV distribution applied to the j-th marginal series, $j = 1, \ldots, d$. A simple weighting scheme is applied, which is common practice in hydrology [Hosking and Wallis, 2005]. Extensions that also take spatial dependence into account are computationally difficult because of the complicated likelihood equations. To the best of our knowledge, this problem has not been solved satisfactorily yet.
The performance of four versions of the joint Hill estimator is compared, using simple or asymptotically optimal weights and $k_j = k_j^{(1)} = \lfloor 2n_j^{2/3} \rfloor$ or $k_j = k_j^{(d)} = \lfloor 2n_j^{2/3}/d^{1/3} \rfloor$. We also studied estimators (H_{opt}) and $(H_{opt}^{(2)})$ with $\hat{\Sigma} = \hat{\Sigma}_{emp}$ (not reported here). These, however, are not advisable when the sample lengths n_j are small and dimension d is large because of numerical problems.
We begin with a discussion of the main findings deduced from Figure 2.1. Each of the five boxplots on the left $(d = 5)$ and on the right $(d = 15)$ represents the estimation error

of the above estimators, derived from 1000 repetitions with $n = 100$, $\gamma = 0.5$, $\delta = 2$, $\boldsymbol{\theta} = (1.5, 2.5)'$ and $\boldsymbol{\tau} = (1, 0.9, 0.8, 0.7, 0.6)'$. We want to emphasize the following conclusions that were also present for many other settings: First, the bias of the Hill estimator can be very dominant in the overall estimation error. Second, optimal weighting leads to a small reduction in variability while the bias remains the same as expected. Third, taking the dimension d into account in the choice of $\mathbf{k}$ is important to decrease a possibly dominant bias.
Table 2.1 reports root mean squared errors $\left(E\left[\hat{\gamma} - \gamma\right]^2\right)^{1/2}$ of all five estimators estimated from 1000 independent repetitions for each of many different settings. Generally, the optimal weighting provides only little improvement. As opposed to this, the choice of $k_j^{(d)}$ instead of $k_j^{(1)}$ has a huge impact on the estimation error. In only a few cases, where the bias of Hill's estimator is luckily small (e.g. $\gamma = 0.5$ and $\delta = 3$), the error increases when using $k_j^{(d)}$ instead of $k_j^{(1)}$ because of an increase in variability. In "typical cases", where the bias is dominant, the incorporation of the dimension d into the choice of upper order statistics notably improves the performance of the joint Hill estimator.
The observation that optimal weighting provides only a small decrease in estimation error is a little disappointing. Loosely speaking, joint estimation of γ benefits only a little from the asymptotic theory derived in Section 2.2 in case of small to moderately large samples. This, however, is not true for the test statistic from Proposition (2.6). In fact, the next subsection demonstrates that the established theory is of key importance in order to achieve an acceptable type 1 error rate.

2.5.2 Heavy-tail ANOVA as a test against alternatives of $\mathcal{H}_{0,IF}$

We studied the finite sample performance of the statistic $W_{\mathbf{k},\boldsymbol{\tau},n}$ as a test for the null hypothesis $\mathcal{H}_{0,IF}$ stated in (2.1). Other established tests for $\mathcal{H}_{0,IF}$, which were already compared by simulations in Viglione et al. [2007], are also included in our experiments. The simulation setting used here differs from Viglione et al. [2007] mainly in the following aspects: First, we also take into account possible spatial extreme value dependence. In Viglione et al. [2007] only spatially independent samples are considered, and, second, the marginal distributions there are characterized in terms of L-moments. We will continue to use the (γ, δ) characterization of marginal distributions.
The other competing procedures applied for comparison can be briefly described as follows:

- The statistic of test HW_1 is similar to that of $W_{\mathbf{k},\boldsymbol{\tau},n}$. For HW_1, each marginal sample ratio of L-scale divided by L-location is compared with a regional version computed from the whole data set. $\mathcal{H}_{0,IF}$ is rejected, if the difference between these L-moment ratios is too large.

- HW_2 is similar to HW_1, with an additional term incorporating the distance of L-skewness divided by L-scale. Both, HW_1 and HW_2, are presented in Hosking and Wallis [2005, Chapter 4.3].

Table 2.1: RMSE's estimated from 1000 independent realizations of five joint Hill estimators applied on extreme valued distributed data with distribution given at the beginning of the simulations section in the main file and with $n = 100$, $\boldsymbol{\tau} = (1, 0.9, 0.8, 0.7, 0.6)$ and $\boldsymbol{\theta} = (1.5, 2.5)$.

		$d = 5$					$d = 15$				
γ	Est.	$\delta = \mu/\sigma$									
		1	1.5	2	2.5	3	1	1.5	2	2.5	3
0.25	(ML)	.070	.069	.068	.066	.069	.062	.061	.063	.063	.065
	(H)	.400	.284	.206	.152	.108	.398	.283	.207	.149	.107
	(H_{opt})	.395	.281	.203	.150	.106	.388	.274	.200	.145	.103
	$(H^{(2)})$	.260	.194	.145	.110	.080	.209	.160	.127	.094	.075
	$(H^{(2)}_{opt})$	.258	.192	.143	.108	.079	.204	.155	.122	.091	.070
0.5	(ML)	.080	.079	.079	.078	.078	.077	.074	.077	.075	.072
	(H)	.303	.183	.099	.059	.065	.301	.178	.103	.057	.060
	(H_{opt})	.297	.179	.097	.058	.065	.289	.170	.095	.053	.060
	$(H^{(2)})$	.183	.128	.087	.074	.083	.155	.110	.099	.085	.091
	$(H^{(2)}_{opt})$	.182	.124	.085	.073	.083	.149	.103	.091	.083	.088
0.75	(ML)	.094	.094	.091	.091	.091	.091	.085	.090	.088	.085
	(H)	.236	.122	.085	.121	.164	.237	.116	.086	.120	.168
	(H_{opt})	.231	.118	.084	.121	.164	.225	.107	.084	.121	.170
	$(H^{(2)})$	.161	.123	.111	.133	.154	.154	.130	.130	.140	.162
	$(H^{(2)}_{opt})$	.157	.121	.109	.132	.154	.143	.123	.126	.137	.161

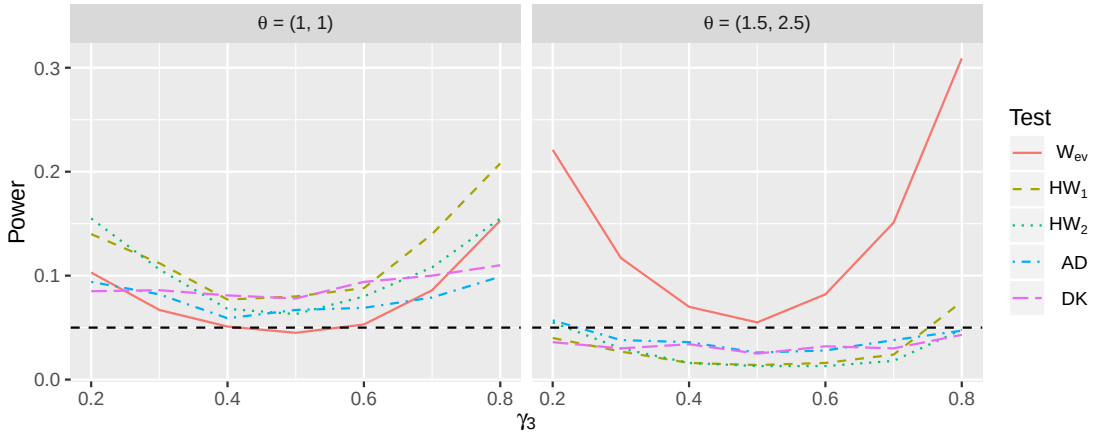

Figure 2.2: Rejection rates of $\mathcal{H}_{1,evi}$ for tests W_{ev}, HW_1, HW_2, AD and DK computed from 4000 samples such that margins $j = 1, 2, 4, 5$ follow a $GEV(\mu = 2, \sigma = 1, \gamma = 0.5)$ and margin $j = 3$ follows a $GEV(\mu = 2, \sigma = 1, \gamma = \gamma_3)$. All 5 margins have sample length $n = 50$ and the spatial dependence corresponds to (2.16) with $\boldsymbol{\theta} = (1, 1)'$ (left) and $\boldsymbol{\theta} = (1.5, 2.5)'$ (right).

- The AD test is based on an Anderson-Darling type distance between marginal empirical distributions and a regional version computed from all available observations. In order to account for possibly different scales under $\mathcal{H}_{0,IF}$, all observations are first divided by their marginal sample median.

- DK is based on a goodness-of-fit statistic proposed by Durbin and Knott [1972]. Just like for AD, all observations are first divided by their marginal sample median. The test is based on the fact that if F is the true distribution function of a continuous random variable X, then $F(X)$ has a uniform distribution on $[0, 1]$.

We studied two versions of test $W_{\mathbf{k},\boldsymbol{\tau},n}$ denoted by W_{emp} and W_{ev}, with empirical estimator $\hat{\Sigma}_{emp}$ and CFG-based estimator $\hat{\Sigma}_{ev}$, respectively, plugged in into the test statistic. Recall that the bias part in the bias-variance trade-off problem is less pronounced, provided $\mathcal{H}_{0,IF}$ holds and due to the scale invariance of the Hill estimator. It is thus more important to account for the variance part. We decided to choose $k_j = \lfloor 2n_j^{2/3} \rfloor$ for arbitrary dimensions d, which was superior to $k_j = k_j^{(d)}$ in our simulation results (not reported here). However, we needed to account for the dimension d by multiplying the statistic with an asymptotically negligible factor $1 - d/(5 \min_j n_j)$. This helped us to reduce the type 1 error at the cost of a slight loss of power. We address the following questions:

(i) How well do the tests keep their nominal level under $\mathcal{H}_{0,IF}$?

(ii) Which test has the largest power against certain alternatives of $\mathcal{H}_{0,IF}$? Specifically, against alternatives (a) $\mathcal{H}_{1,evi} \cap \mathcal{H}_{0,delta}$ or (b) $\mathcal{H}_{0,evi} \cap \mathcal{H}_{1,delta}$ such that $\mathcal{H}_{0,IF}$ holds

Table 2.2: Rejection rates of $\mathcal{H}_0^{IF}$ in % computed from 4000 samples under $\mathcal{H}_0^{IF}$. The nominal level is 5%.

		$\boldsymbol{\tau} = (1, 0.9, 0.8, 0.7, 0.6)$									
		$\boldsymbol{\theta} = (1,1)$					$\boldsymbol{\theta} = (1.5, 2.5)$				
γ	Test	$\delta = \mu/\sigma$									
		1	1.5	2	2.5	3	1	1.5	2	2.5	3
		$n = 50$									
0.25	W_{ev}	8.3	4.0	3.8	4.1	4.0	17.3	10.2	7.8	7.8	6.5
	W_{emp}	14.3	9.8	9.2	9.7	9.2	28.5	21.5	20.8	19.6	18.0
	HW_1	3.6	4.6	5.0	4.9	5.1	0.7	1.0	1.1	1.4	1.0
	HW_2	4.2	4.6	5.0	4.6	4.6	1.4	1.5	1.7	1.8	1.6
	AD	4.7	4.3	4.4	4.6	6.2	2.3	2.7	2.5	2.7	3.5
	DK	6.6	3.8	4.1	5.0	6.2	4.0	2.1	2.3	2.5	2.1
0.5	W_{ev}	5.1	3.8	4.8	5.7	6.4	12.2	6.6	6.2	5.2	6.7
	W_{emp}	11.3	10.0	10.4	11.7	12.3	24.9	19.1	17.9	17.9	18.4
	HW_1	7.6	9.0	9.2	9.5	8.0	1.8	2.0	1.9	1.6	1.6
	HW_2	6.3	7.6	6.9	7.8	7.4	2.2	1.8	2.0	1.8	2.1
	AD	4.3	5.0	6.6	7.3	8.4	2.3	2.2	3.1	3.7	4.0
	DK	4.5	4.6	7.6	10.3	13.4	2.5	2.2	3.2	4.5	4.9
0.75	W_{ev}	4.4	5.1	7.0	8.6	11.5	7.0	6.6	5.9	6.5	7.3
	W_{emp}	10.1	11.2	12.9	14.6	17.8	19.7	18.1	17.1	17.8	18.4
	HW_1	16.6	18.6	17.7	15.9	16.4	4.0	5.4	4.3	4.9	4.0
	HW_2	10.2	12.5	11.5	10.8	12.5	3.4	3.8	3.0	4.0	4.2
	AD	4.9	7.3	10.0	12.2	13.4	2.7	4.0	5.5	5.0	7.0
	DK	4.4	9.2	16.4	22.3	28.0	1.9	4.1	7.4	10.5	12.8
		$n = 100$									
0.25	W_{ev}	3.6	2.9	2.6	3.2	3.5	9.8	7.3	6.0	4.8	4.7
	W_{emp}	6.8	5.8	6.2	6.3	6.5	17.5	14.1	12.3	10.9	11.6
	HW_1	3.3	3.8	4.0	3.9	4.0	0.5	0.9	0.5	0.6	0.7
	HW_2	3.5	3.6	4.3	3.7	4.2	1.2	1.1	1.0	1.0	1.1
	AD	5.1	4.5	4.6	5.1	5.2	2.2	2.7	3.0	3.5	3.1
	DK	7.0	5.2	4.9	5.4	7.1	4.4	2.9	2.2	2.2	2.9
0.5	W_{ev}	3.3	3.5	4.5	5.6	6.3	6.9	5.2	5.6	4.9	5.2
	W_{emp}	6.6	6.5	8.2	9.4	10.0	12.8	11.2	11.5	10.5	10.3
	HW_1	6.4	6.0	6.0	6.5	5.4	0.8	1.2	1.1	0.9	0.8
	HW_2	5.8	5.0	5.4	5.7	4.7	1.1	0.9	1.4	0.9	1.5
	AD	4.0	5.3	6.0	7.1	7.6	1.9	2.6	3.4	3.5	3.3
	DK	5.2	5.2	8.8	11.7	13.9	2.5	2.6	3.2	5.1	7.2
0.75	W_{ev}	3.3	3.9	7.2	8.0	10.3	5.5	5.0	5.1	6.6	6.3
	W_{emp}	7.0	7.9	11.1	12.0	15.0	11.2	11.2	10.0	12.3	11.2
	HW_1	9.6	10.7	11.6	10.6	8.8	2.5	2.3	2.3	1.2	1.4
	HW_2	5.9	6.7	8.0	8.6	7.7	2.3	2.5	2.8	1.7	1.4
	AD	4.8	6.0	9.1	9.9	11.8	2.2	3.0	4.7	4.9	4.9
	DK	4.5	11.0	19.3	25.9	32.2	2.0	4.5	9.2	12.0	15.7

for a group of four margins $j = 1, 2, 4, 5$ and where margin $j = 3$ differs by either $\gamma_3 \neq \gamma$ or $\delta_3 \neq \delta$.

All tests were carried out at a nominal level of $\alpha = 5\%$ and with data drawn from multivariate extreme value distributions presented at the beginning of Section 2.5.

Empirical levels under spatial independence: The left part of Table 2.2 reports rejection rates in percent of all considered tests estimated from 4000 samples under $\mathcal{H}_{0,IF}$ and $\boldsymbol{\theta} = (1, 1)$. The level of W_{emp} is overall not acceptable, whereas test W_{ev} keeps its level reasonably well except for some cases with $\gamma = 0.75$. With increasing heaviness γ of the tails, all other tests fail to get close to the nominal level.

Empirical levels under spatial dependence: The right hand side of Table 2.2 reports rejection rates as before, but with $\boldsymbol{\theta}$ set to $(1.5, 2.5)$. In case of $\mathbf{a} = (0.9, 0.7, 0.5, 0.3, 0.1)$, this leads to an average Spearman's rho for the pairs of about $\rho = 0.5$. Such a strength of dependence is not uncommon in hydrological applications. Test W_{ev} keeps its level reasonably well, except for some settings with $\gamma = 0.25$. In contrast, all other methods are overall far from attaining the nominal level of 5%. We also studied the performance for $\boldsymbol{\tau} = (1, 2)$, which led to an average Spearman's rho of about $\rho = 0.25$. The results were very similar and are therefore not reported here.

Empirical power under $\mathcal{H}_{1,evi} \cap \mathcal{H}_{0,delta}$: Figure 2.2 presents rejection rates of tests W_{ev}, HW_1, HW_2, AD and DK under $\mathcal{H}_{1,evi} \cap \mathcal{H}_{0,delta}$ versus γ_3 estimated from 4000 samples of length $n = 50$ with $\tau = (1, 1, 1, 1, 1)$ such that all but the third component follow a GEV with $\gamma = 0.5$ and $\delta = 2$ and the third component follows a GEV with $\gamma_3 \in \{0.2, 0.3, \ldots, 0.8\}$ and $\delta = 2$. It is remarkable that all tests except W_{ev} have almost no power under positive dependence (right plot of Fig. 2.2), while the power of test W_{ev} is even higher than under independence (left plot of Fig. 2.2). The left plot in Figure 2.3, where we set $n = 100$, confirms these findings.

Empirical power under $\mathcal{H}_{0,evi} \cap \mathcal{H}_{1,delta}$: The right part of Figure 2.3 presents rejection rates of tests W_{ev}, HW_1, HW_2, AD and DK under $\mathcal{H}_{0,evi} \cap \mathcal{H}_{1,delta}$ versus δ_3 estimated from 4000 samples of length $n = 100$ with $\tau = (1, 1, 1, 1, 1)$ such that all but the third component follow a GEV with $\gamma = 0.5$ and $\delta = 2$, while the third component follows a GEV with $\gamma = 0.5$ and $\delta_3 \in \{1.25, 1.5, \ldots, 2.75\}$. Although test W_{ev} is designed to detect deviations from $\mathcal{H}_{0,evi}$, these results indicate that W_{ev} is rather a test for $\mathcal{H}_{0,IF}$. Note that tests AD and DK are way more powerful than W_{ev} against alternatives from $\mathcal{H}_{0,evi} \cap \mathcal{H}_{1,delta}$.

Summing up, the proposed test W_{ev} keeps its level well in reasonable settings from hydrology. Additionally, the new test is the only one that detects deviations from $\mathcal{H}_{0,evi}$ under spatial dependence. On the other hand, test W_{ev} has little power against $\mathcal{H}_{1,delta}$ compared to AD and DK. When hypothesis $\mathcal{H}_{0,IF}$ is rejected by W_{ev}, tests AD and DK can serve as auxiliary tools to indicate whether the deviation from hypothesis $\mathcal{H}_{0,IF}$ is due to $\mathcal{H}_{1,delta}$ or not.

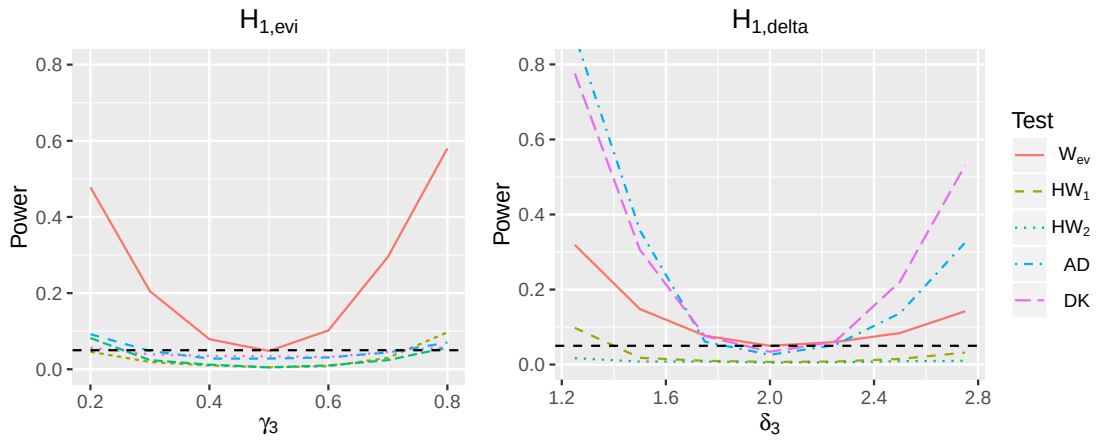

Figure 2.3: Rejection rates of $\mathcal{H}_{1,evi}$ (left) and $\mathcal{H}_{1,delta}$ (right) for tests W_{ev}, HW_1, HW_2, AD and DK computed from 4000 samples such that margins $j = 1, 2, 4, 5$ follow a $GEV(\mu = 2, \sigma = 1, \gamma = 0.5)$. Margin $j = 3$ follows a $GEV(\mu = 2, \sigma = 1, \gamma = \gamma_3)$ (left) and $GEV(\mu = \delta_3, \sigma = 1, \gamma = 0.5)$ (right). The spatial dependence corresponds to (2.16) with $\mathbf{a} = (1.5, 2.5)$. All 5 margins have sample length $n = 100$.

2.6 Application: Summer floods at the Mulde river basin

Many studies in regional flood frequency analysis focus on maximal flows Y (in m^3/sec) observed at several stations of some region of interest. In order to avoid non-stationarity due to seasonal effects, the block maxima method with block length covering one season is applied on each marginal series. Thanks to Theorems 1.3 and 1.4, these marginal series can be modeled by the parametric class of generalized extreme value distributions (GEV) and the spatial dependence by the nonparametric class of extreme value copulas.
Our region of interest is the Mulde river basin in Saxony, Germany. We have monthly data from 116 stations located in Saxony, with between 6 and 100 years of observations per station and an average of about 52 years. Here we focus on the analysis of hydrological summer maxima, namely the maximal river flow Y measured between May and October for each station and year available. There are two reasons for restricting to summer maxima. First, most winter floods are produced from melting snow, whereas summer floods are due to short but heavy rainfalls. These very different meteorological causalities lead to different river flow distributions. Second, extraordinary high flows, which are of particular interest for the flood protection, have been observed only during the summer in that region.
For our data set of 116 stations, the difference between winter and summer peaks is illustrated in Figure 2.4. Each point represents a ML fit $(\hat{\gamma}_{ML}, \hat{\delta}_{ML})$ to the generalized extreme value distribution with $\delta = \mu/\sigma$, where a fit is based on either the series of summer (◯) or winter maxima (△) of the stations. The size of each point is taken proportional to the corresponding sample length available for estimation. Note that winter and summer

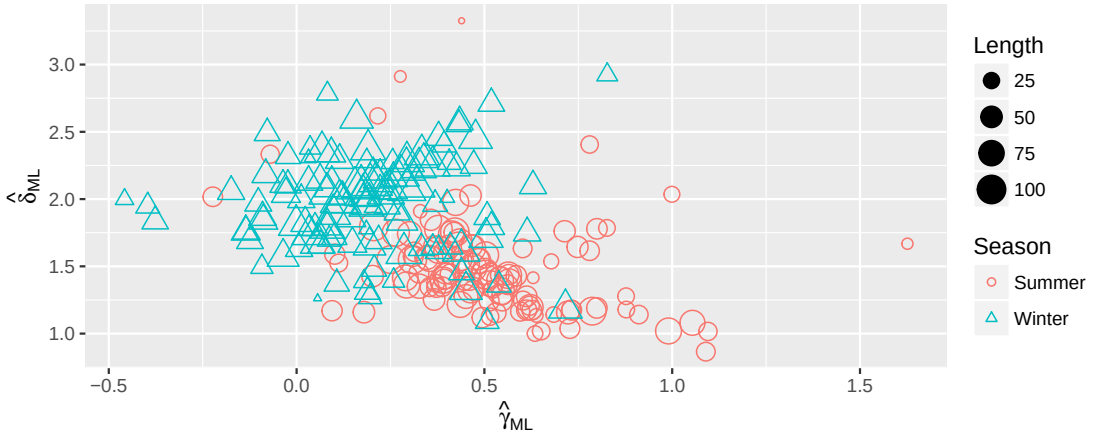

Figure 2.4: Each point represents a maximum likelihood fit $(\hat{\gamma}_{ML}, \hat{\delta}_{ML})$ of the $GEV(\mu, \sigma, \gamma)$ distribution, where $\delta = \mu/\sigma$. We fitted winter $(\triangle)$ and summer maxima $(\bigcirc)$ series of 116 stations that are located in Saxony, Germany. The size of each point was taken proportional to the available sample length at the corresponding station.

maxima are systematically different in distribution and that the range of the summer estimates is covered well by our simulation settings from Section 2.5.
We select a group of 18 stations as possibly homogeneous, which are summarized in Table 2.3 together with some statistics of interest. The selection of this group is based on a canonical correlation analysis (CCA), which is a popular method in regional flood frequency analysis [Ouarda et al., 2001]. The goal of CCA is to identify homogeneous groups, that are stations $j \in \{1, \ldots, d\}$ with marginal distributions F_j satisfying (2.1). This is done by the comparison of a stations basic characteristics (e.g. the hight and size of the catchment area, mean annual precipitation, slope of main channel, ...) and its discharge distribution. A disadvantage of CCA (and other classification techniques) is that the outcome strongly depends on the choice of variables and many tuning parameters. Different hydrologists will usually derive different groups of homogeneous stations. Therefore, it is advisable to check first the homogeneity assumption of a selected group. If not rejected, we can continue with the joint tail estimation.
The last column of Table 2.3 consists of p-values of a goodness-of-fit procedure [Kojadinovic and Yan, 2012b], which evaluates the assumption that a marginal distribution is of GEV type and which is of interest in order to apply a ML based approach for comparative reasons. p-values are computed from 1000 parametric bootstrap replicates. It should be noted that, however, such a goodness-of-fit test has only little power when the number of observations is small ($n \leq 100$). Recall also that the GEV is an asymptotic model for block maxima distributions with blocks over identically distributed observations and block size tending to infinity. The assumption that summer maxima are built from, say, 6 i.i.d. monthly observations is little realistic. Altogether it is not clear whether a GEV model assumption is the right choice for the problem considered here.

Table 2.3: A group of 18 stations was selected based on a canonical correlation analysis (not reported here). The statistics were computed from the corresponding summer maxima series. The last column summarizes p-values to evaluate hypothesis $\mathcal{H}_{0,j} : F_j$ is GEV-distributed.

#	station	obs. years	τ_j	k_j	H_{k_j,n_j}	$\hat{\gamma}_{j,ML}$	$\hat{\delta}_{j,ML}$	GoF p-value
1)	560051	1961-2009	.49	26	.75	.71	1.76	.526
2)	562115	1910-2009	1	43	.54	.42	1.75	.946
3)	563790	1928-2009	.82	37	.63	.48	1.54	.732
4)	564410	1910-2009	1	43	.47	.21	1.80	.185
5)	566010	1936-2009	.74	35	.71	.57	1.39	.261
6)	566040	1926-2009	.84	38	.71	.56	1.43	.435
7)	566100	1961-2009	.49	26	.77	.75	1.64	.469
8)	567400	1960-2009	.50	27	.47	.46	2.03	.168
9)	567451	1910-2009	1	43	.65	.50	1.44	.518
10)	567470	1933-2009	.77	36	.63	.40	1.43	.624
11)	567700	1961-2009	.49	26	.42	.36	1.85	**.005**
12)	567850	1921-2009	.89	39	.50	.38	1.62	.275
13)	568140	1921-2009	.89	39	.58	.47	1.63	.608
14)	568160	1929-2009	.81	37	.72	.45	1.52	.089
15)	568350	1929-2009	.81	37	.59	.36	1.56	.883
16)	576410	1961-2009	.49	26	.65	.48	1.55	.064
17)	576421	1966-2009	.44	24	.66	.37	1.65	.338
18)	577100	1968-2009	.42	12	.42	.33	1.45	.385
mean					.62	.46	1.61	

Recall from the simulation results in Section 2.5.2 that $k_j = k_j^{(1)} = \lfloor 2n_j^{2/3} \rfloor$ is appropriate for test W_{ev}, although this choice is not optimal for the joint estimation. We applied test W_{ev} on the selected group. The resulting p-value of $p = 0.02$ indicates that there is strong evidence against the homogeneity assumption (2.1). In order to reduce the detected heterogeneity, we examined a scatter plot of the 18 pairs $(\hat{\gamma}_{ML}, \hat{\delta}_{ML})$ from Table 2.3. The points corresponding to the station numbers 1, 4, 7 and 8 are quite isolated from the others. Moreover, taking into account the multiple testing, there is some weak evidence that the GEV assumption for station # 11 is violated. Overall we excluded stations 1, 4, 7, 8 and 11 and applied test W_{ev} again. The resulting p-value is $p = 0.22$, making the assumption of homogeneity more plausible than for the larger group considered before. Interestingly enough, none of the competing tests HW_1, HW_2, AD and DK (see Section 2.5.2) rejects the homogeneity hypothesis for the larger group. A reason for this is the large spatial dependence, with an average pairwise Spearman's rho value of about 0.66. Recall that in such a case the competing methods are not able to detect deviations from $\mathcal{H}_{0,evi}$ (right plot in Figure 2.2). The fact that tests AD and DK remain quite powerful against $\mathcal{H}_{1,delta}$ even for dependent data (right plot of Figure 2.3) suggests that the heterogeneity detected by W_{ev} is indeed due to a violation of assumption $\mathcal{H}_{0,evi}$.

The last part of this section deals with the estimation of γ under the assumption that $\gamma_j = \gamma$ holds for all 13 stations j within the smaller group. A recommended rule for the selection of marginally optimal k_j values is based on the examination of plotting methods

[Drees et al., 2000], for instance, the popular Hill plot $(k, H_{k,n})_{1\leq k<n}$. An integer $1 \leq k < n$ is chosen such that the plot is approximately constant (stable) in a neighborhood of k. On the other hand, under the assumption that each margin is GEV distributed, we are able to calculate the asymptotically optimal rate of $k_j = \lfloor 2n_j^{2/3} \rfloor$. Interestingly, for our application, both methods yield very similar results, except for station # 18. For that we found that $k_{18} = 12$ is within a stable region in contrast to $\lfloor 2n_{18}^{2/3} \rfloor = 24$.
Recall from the discussion in Section 2.3.2 and the simulation results in Section 2.5.1 that the marginally optimal k_j values are not optimal for joint estimation. For the joint estimation we apply the simple rule $k_j^{(d)} = \lfloor 2n_j^{2/3}/d^{1/3} \rfloor$, $j \neq 18$, and $k_{18}^{(13)} = \lfloor k_{18}/d^{1/3} \rfloor = 5$ with $d = 13$. Together with the asymptotically optimal weights $\hat{\boldsymbol{w}}_{opt}$ estimated under the extreme value dependence assumption we obtain an estimate of $\hat{\gamma} = 0.43$ with estimated 95% confidence interval $[0.27,\ 0.59]$ derived from the asymptotic normality in (2.9).
Interestingly enough, the same procedure with marginally optimal integers $k_j = k_j^{(1)} = \lfloor 2n_j^{2/3} \rfloor$ (under the GEV assumption) leads to a much heavier index of $\hat{\gamma} = 0.59$ with confidence interval $[0.45,\ 0.73]$. The ML based joint estimator $\hat{\gamma}_{ML} = \sum_{j\in G} w_j \hat{\gamma}_{ML,j}$ with weights $w_j = n_j / \sum_{k\in G} n_k$ proportional to the marginal sample lengths gives us $\hat{\gamma}_{ML} = 0.45$, which supports the first estimate rather than the second one, provided the GEV assumption is reasonably met for this data set.

2.7 Conclusion and outlook on regional estimation of quantiles

The problem of estimating the risk of extreme realizations of heavy-tailed distributions is closely related to the extreme value index estimation problem. Lekina et al. [2014] argue that parametric models are not always appropriate for the estimation of high quantiles in flood frequency analysis. On the other hand, the estimation of nonparametric models is associated with increased uncertainty. Typically, such models are useful only in applications with many data points available. In regional flood frequency analysis, where we observe the same variable at many stations, pooling methods are used to overcome the problem of having only short marginal sequences available. We propose a simple and straightforward peaks-over-threshold approach that allows us to analyze the joint behavior of heavy tails without restricting on parametric models. Even more, our test is able to detect tail heterogeneity, which is hardly detected by competing procedures that are common in the hydrological literature.
Typically, practitioners are not interested in γ, but in the estimation of high quantiles $F_j^{-1}(p)$. In the semi-parametric framework considered here, this is done by plugging in the joint estimator $\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})$ into the extrapolation formula of Weissman [1978]. Weissman's formula is interpreted as extrapolation from moderate to high quantiles, with moderate quantiles estimated by simple order statistics and γ controlling the width of extrapolation. Asymptotic properties of this high quantile estimator are closely related to those of $\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})$. Specifically, asymptotic normality of the high quantile estimator can be deduced from Theorem 1.14.

Proposition 2.7 (Regional estimation of high quantiles)
Suppose that assumptions 2.1 and 2.2 hold and let $j \in \{1, \dots, d\}$ be fixed. Furthermore, suppose that there exists a real $\rho < 0$ and a function R satisfying $\lim_{t\to\infty} R(t) = 0$ such that (1.27) holds for $U = U_j$ with $U_j(t) = F_j^{-1}(1 - 1/t)$ and that $\sqrt{k_1}R(n/k_1) \to \lambda \in \mathbb{R}$. Then, for any sequence of probabilities $p = p_n$ satisfying $n(1-p)/k_1 \to 0$ and $\log(n(1-p))/\sqrt{k_1} \to 0$ for $n \to \infty$, we have that

$$\frac{\sqrt{k_1}}{\log\{k_1/(n(1-p))\}} \left(\frac{\hat{F}_j^{-1}(p)}{F_j^{-1}(p)} - 1 \right) \xrightarrow{D} \mathcal{N}\left(0, \frac{\gamma^2}{c_j} \boldsymbol{w}'\Sigma\boldsymbol{w}\right), \tag{2.17}$$

where we set

$$\hat{F}_j^{-1}(p) = u_j \cdot \left(\frac{k_j}{n_j(1-p)} \right)^{\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})} \tag{2.18}$$

and where $u_j = u_j(k_j, n_j)$ denotes the $(n_j - k_j)$-th largest observation from $Y_{a_j+1,j}, \dots, Y_{n,j}$.

Suppose that we have collected annual maximal flows from d dependent sites and that we are interested in the estimation of, say, the 99% quantile of the distribution F_j of the j-th site. Suppose further that the heavy-tail homogeneity assumption $H_{0,evi} : \gamma_1 = \ldots = \gamma_d = \gamma$ is met for some unknown $\gamma > 0$. Since we are dealing with componentwise maxima, it might be reasonable to assume that the observations stem from a d-variate extreme value distribution with margins $F_j \sim GEV(\mu_j, \sigma_j\xi_j)$. These satisfy $\xi_1 = \ldots = \xi_d = \gamma$ because of $H_{0,evi}$ and because the shape of a GEV distribution coincides with its extreme value index. We thus may estimate the shape ξ_1 of F_1 by averaging over all d maximum likelihood estimates $\hat{\gamma}_j$ of ξ_j. Location μ_1 and scale σ_1 are estimated as usual only from the local observations. Quantile estimation based on this parametric approach will be referred to as GEV-ML procedure.
Recall from the last part of Section 1.1.2 that there is evidence for the GEV being not an appropriate model for annual maximal flows. In such a case, we may prefer the regional Weissman estimator (2.18) built under less restrictive assumptions. As a first illustration, simulation results for the estimation of $F_1^{-1}(0.99)$ in two different scenarios are presented in Figure 2.5. The dashed lines correspond to the true quantiles. On the left-hand side, all the marginal observations were drawn from a GEV distribution, while for the right-hand side a product of two GEV distribution functions was used: Think of annual maximal flows $Y = \max\{W, S\}$ being the maximum over winter and summer maximal flows W and S, respectively. Suppose that W, S are independent and that they follow different GEV distribution functions F_W and F_S. Then the distribution of Y is $F_Y = F_W \cdot F_S$, which is not necessarily of GEV-type. Each of the box plots is computed from $N = 10000$ samples of length $n = 100$, dimension $d \in \{1, 2, 5, 10\}$ on the x-axis and with inter-site dependence generated by a symmetric Gumbel-Hougaard copula with parameter $\vartheta = 2$. For the Weissman estimator we have used $k_j = \lfloor 2n^{2/3}/d^{1/3} \rfloor$ in order to be able to decrease the bias with increasing dimension d (see Section 2.5.1). We conclude:

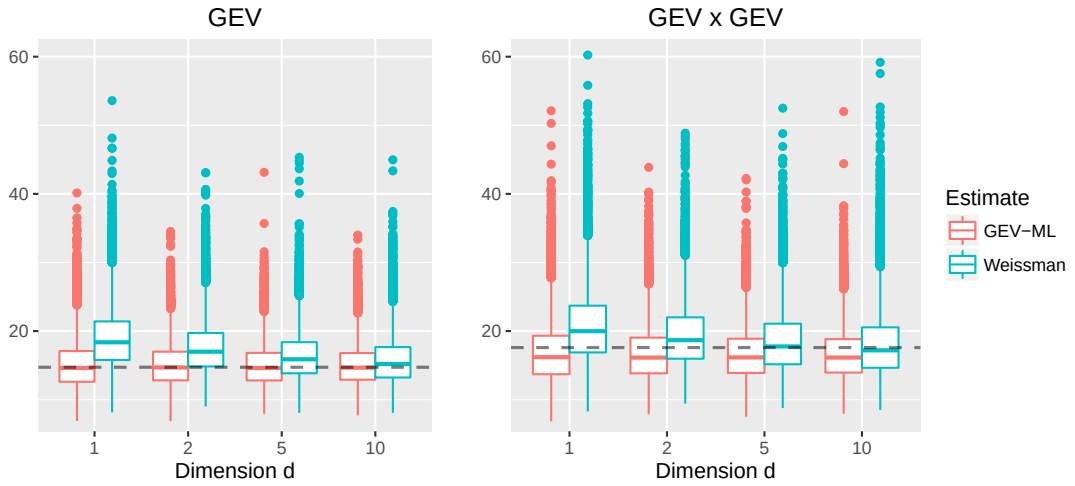

Figure 2.5: Each box plot is computed from 10000 samples of length $n = 100$ and dimension $d \in \{1, 2, 5, 10\}$ drawn from a d-variate distribution with Gumbel-Hougaard copula with parameter $\vartheta = 2$. The marginal observations follow (left) a GEV distribution function and (right) a product of two different GEV distributions. The GEV parameters are selected on the basis of real data from flood frequency analysis.

First note that the GEV-ML approach benefits only very little from the increasing dimension d. A reason for this might be the relatively high inter-site dependence. However, it seems that for the Weissman estimator there is a lot more space for improvement with increasing dimension d making this approach particularly useful in regional estimation. An advantage of the Weissman estimator over parametric GEV approaches is shown on the right-hand side of Figure 2.5. There the GEV assumption is misspecified, which results in a systematic error for the ML approach. This, of course, is not an issue for the Weissman estimator.

2.8 Proofs

2.8.1 Proof of Theorem 2.4

For sake of readability, the proof of Theorem 2.4 is given for dimension $d = 2$. The proof in case of same sample lengths, i.e. $\tau_1 = \tau_2 = 1$, can be found in Dematteo and Clémençon [2016].

Notation: Let $0 < \tau_1, \tau_2 < 1$ be fixed. For ease of presentation, we assume the same beginning and different end points, that is, we observe $X_1, \dots, X_{\lfloor n\tau_1 \rfloor} \sim F_X$ and $Y_1, \dots, Y_{\lfloor n\tau_2 \rfloor} \sim F_Y$ from a sequence $(X_i, Y_i)_{i \geq 1}$ of identically $C(F, G)$ distributed random vectors. In what follows we will write $n\tau$ instead of $\lfloor n\tau \rfloor$, e.g., $X_{n\tau - i:n\tau} = X_{\lfloor n\tau \rfloor - i:\lfloor n\tau \rfloor}$ and $\sum_{i=1}^{n\tau} = \sum_{i=1}^{\lfloor n\tau \rfloor}$. The relation $f_n \sim g_n$, $g_n \neq 0$, means that $f_n / g_n \to 1$ for $n \to \infty$.

Let $a(t) = F^{-1}(1-1/t)$ and $b(t) = G^{-1}(1-1/t)$, $t > 1$, denote the marginal return levels. Recall that the empirical counterparts of $a\left(\frac{n\tau_1}{k_1}\right)$ and $b\left(\frac{n\tau_2}{k_2}\right)$ are given by

$$\hat{a}\left(\frac{n\tau_1}{k_1}\right) = X_{n\tau_1-k_1:n\tau_1} \text{ and } \hat{b}\left(\frac{n\tau_2}{k_2}\right) = Y_{n\tau_2-k_2:n\tau_2}.$$

The Hill estimator of the first component can be rewritten in the form

$$H^{(1)}_{k_1,\tau_1,n} = \frac{1}{k_1}\sum_{i=1}^{k_1} \log \frac{X_{n\tau_1-i+1:n\tau_1}}{X_{n\tau_1-k_1:n\tau_1}} = \int_1^\infty \frac{1}{k_1}\sum_{i=1}^{n\tau_1} \mathbb{1}\left(X_i > \hat{a}\left(\frac{n\tau_1}{k_1}\right)\cdot x\right)\frac{dx}{x} \tag{2.19}$$

and similarly, $H^{(2)}_{k_2,\tau_2,n}$ for the second component.
The first part of the proof considers the arithmetic mean inside the integral in (2.19). Forget first that $\hat{a}, \hat{b}$ instead of a, b are involved in the Hill estimators. We set

$$Z_i^X(x) = \mathbb{1}\left(X_i > a\left(\frac{n\tau_1}{k_1}\right)\cdot x\right) - \bar{F}\left(a\left(\frac{n\tau_1}{k_1}\right)\cdot x\right)$$

and similarly, we define $Z_i^Y(y)$ for the second component. The centered version of the building block involved in (2.19) with $\hat{a}, \hat{b}$ replaced by a, b is given by

$$\left(\frac{1}{\sqrt{k_1}}\sum_{i=1}^{n\tau_1} Z_i^X(x), \frac{1}{\sqrt{k_2}}\sum_{j=1}^{n\tau_2} Z_j^Y(y)\right)' = \left(S_n^X(x,\tau_1),\ S_n^Y(y,\tau_2)\right)' = S_n(x,y).$$

The first goal of our proof is to show weak convergence of S_n towards a Gaussian process in the function space $D^2 = D(\mathbb{R}_+) \times D(\mathbb{R}_+)$. The weak convergence of S_n together with the continuous mapping

$$(\varphi,\ \psi) \mapsto \left(\int_1^\infty \varphi(x)\frac{dx}{x},\ \int_1^\infty \psi(y)\frac{dy}{y}\right),\ D^2 \to \mathbb{R}^2, \tag{2.20}$$

helps us to study the joint convergence of the Hill estimators.
The weak convergence of S_n follows from that of the usual tail empirical process, e.g., in the proof of Dematteo and Clémençon [2016] and a Cramer-Wold device for D^2 [Davidson, 1994, Th. 29.16]: Let $\lambda = (\lambda_1, \lambda_2)' \in \mathbb{R}^2$ and, without loss of generality, let $\tau_1 \le \tau_2$. Then we have that

$$\begin{aligned}\lambda' \cdot S_n(x,y) =& \frac{1}{\sqrt{k_1}}\sum_{i=1}^{n\tau_1}\left[\lambda_1 Z_i^X(x) + \lambda_2\sqrt{\frac{k_2}{k_1}}Z_i^Y(y)\right] \\ &+ \frac{1}{\sqrt{k_2}}\sum_{j=n\tau_1+1}^{n\tau_2}\lambda_2 Z_j^Y(y).\end{aligned} \tag{2.21}$$

Weak convergence of each of the summands on the right-hand side of (2.21) follows from the proof in Dematteo and Clémençon [2016]. Because these two summands are independent, we also have weak convergence of $\lambda' S_n$ for each $\lambda \in \mathbb{R}^2$ and thus, by applying the Cramer-Wold device for D^2, we obtain weak convergence of S_n in D^2.

Now let $\hat{S}_n$ denote process S_n with a, b replaced by $\hat{a}, \hat{b}$. From de Haan and Ferreira [2006, Th. 2.4.8] and with $n \to \infty$, we also have that

$$\left(\frac{\hat{a}\left(\frac{n\tau_1}{k_1}\right)}{a\left(\frac{n\tau_1}{k_1}\right)}, \frac{\hat{b}\left(\frac{n\tau_2}{k_2}\right)}{b\left(\frac{n\tau_2}{k_2}\right)} \right) \xrightarrow{\mathbb{P}} (1,\ 1). \tag{2.22}$$

Applying the continuous mapping $(S(x,y), p, q) \mapsto (S(px, qy))$, $D^2 \times \mathbb{R}^2 \to D^2$, on $\hat{S}_n$ and the left-hand side of (2.22) gives us S_n. Afterwards, applying the mapping (2.20) on S_n and letting $n \to \infty$ results in

$$\begin{aligned} &\sqrt{k_1} \left(H^{(1)}_{k_1,\tau_1,n} - \int_{\hat{a}\left(\frac{n\tau_1}{k_1}\right)}^{\infty} \frac{n\tau_1}{k_1} \bar{F}(x) \frac{dx}{x}, H^{(2)}_{k_2,n,\tau_2} - \int_{\hat{b}\left(\frac{n\tau_2}{k_2}\right)}^{\infty} \frac{n\tau_2}{k_2} \bar{G}(y) \frac{dy}{y} \right) \\ =&\sqrt{k_1} \left(\int_1^{\infty} \frac{1}{k_1} \sum_{i=1}^{n\tau_1} Z_i^X(x) \frac{dx}{x}, \int_1^{\infty} \frac{1}{k_2} \sum_{i=1}^{n\tau_2} Z_i^Y(y) \frac{dy}{y} \right) \xrightarrow{D} \mathcal{N}(0, \Sigma^*) \end{aligned} \tag{2.23}$$

for some covariance matrix $\Sigma^* \in \mathbb{R}^{2\times 2}$. The limit is normally distributed, because the limit of S_n is a Gaussian process. For the calculation of Σ^*, recall that $c_i = \lim_{n\to\infty} \frac{k_1}{k_i}$ and note that the asymptotic distribution of each of the components on the left-hand side of (2.23) does not depend on τ_1 and τ_2. Thus, the diagonal elements of Σ^* are the same as for $\tau_1 = \tau_2 = 1$ and given by $\Sigma^*_{i,i} = 2c_i\gamma^2$ [Dematteo and Clémençon, 2016].
The calculation of $\Sigma^*_{1,2}$ requires some more effort. For this, recall that return level functions a, b of Pareto-type distributions F are regularly varying with index γ, which, for $n \to \infty$, means that

$$a\left(\frac{n\tau_1}{k_1}\right) \cdot a\left(\frac{n}{k_1}\right)^{-1} \longrightarrow \tau_1^{\gamma} = \tau_1^{\gamma} \text{ and } b\left(\frac{n\tau_2}{k_2}\right) \cdot b\left(\frac{n}{k_1}\right)^{-1} \longrightarrow (c_2\tau_2)^{\gamma}$$

(recall that $c_1 = 1$). For $n \to \infty$ and from Assumption 2.2, we obtain

$$\begin{aligned} &\frac{n}{k_1}\mathbb{P}\left(X_1 > a\left(\frac{n\tau_1}{k_1}\right) x, Y_1 > b\left(\frac{n\tau_2}{k_2}\right) y \right) \\ =&\frac{n}{k_1}\mathbb{P}\left(X_1 > a\left(\frac{n}{k_1}\right) \frac{a\left(\frac{n\tau_1}{k_1}\right)}{a\left(\frac{n}{k_1}\right)} x, Y_1 > b\left(\frac{n}{k_1}\right) \frac{b\left(\frac{n\tau_2}{k_2}\right)}{b\left(\frac{n}{k_1}\right)} y \right) \longrightarrow \Lambda\left(\frac{x^{-1/\gamma}}{\tau_1}, \frac{y^{-1/\gamma}}{c_2\tau_2} \right) \end{aligned}$$

uniformly in x, y, where Λ is the corresponding upper tail dependence copula. Since, in addition,

$$\frac{n}{k_1}\mathbb{P}\left(X_1 > a\left(\frac{n\tau_1}{k_1}\right) x \right) \cdot \mathbb{P}\left(Y_1 > b\left(\frac{n\tau_2}{k_2}\right) y \right) \to 0,$$

we conclude that

$$
\begin{aligned}
& k_1 \cdot \mathbb{E}\left[\int_1^\infty \frac{1}{k_1}\sum_{i=1}^{n\tau_1} Z_i^X(x)\frac{dx}{x}\cdot\int_1^\infty \frac{1}{k_2}\sum_{i=1}^{n\tau_2} Z_i^Y(y)\frac{dy}{y}\right] \\
\sim & \int_1^\infty\int_1^\infty \frac{n(\tau_1\wedge\tau_2)}{k_2}\mathbb{P}\left(X_1>\widehat{a}\left(\frac{n\tau_1}{k_1}\right)x, Y_1>\widehat{b}\left(\frac{n\tau_2}{k_2}\right)y\right)\frac{dxdy}{xy} \\
\to & c_2(\tau_1\wedge\tau_2)\int_1^\infty\int_1^\infty \Lambda\left(\frac{x^{-1/\gamma}}{\tau_1},\frac{y^{-1/\gamma}}{c_2\tau_2}\right)\frac{dxdy}{xy} = \Sigma^*_{1,2}.
\end{aligned}
\tag{2.24}
$$

To sum up: The previous part of the proof shows joint asymptotic normality of the Hill estimators $H^{(1)}_{k_1,\tau_1,n}$ and $H^{(2)}_{k_2,\tau_2,n}$ centered by the random values

$$
\int_{\hat{a}\left(\frac{n\tau_1}{k_1}\right)}^\infty \frac{n\tau_1}{k_1}\bar{F}(x)\frac{dx}{x} \quad\text{and}\quad \int_{\hat{b}\left(\frac{n\tau_2}{k_2}\right)}^\infty \frac{n\tau_2}{k_2}\bar{G}(y)\frac{dy}{y}, \tag{2.25}
$$

respectively, and with limiting covariance matrix Σ^*. In the remaining part of the proof we will replace both terms in (2.25) by γ. Proceeding in the same lines as in the proof of Dematteo and Clémençon [2016], the final step that has to be done is the derivation of the limiting covariance matrix in (2.7).

First, note that $\int_1^\infty s^{-1/\gamma}\frac{ds}{s} = \gamma$, which, together with (2.5), implies that

$$
\gamma = \lim_{n\to\infty}\int_{a\left(\frac{n\tau_1}{k_1}\right)}^\infty \frac{n\tau_1}{k_1}\bar{F}(x)\frac{dx}{x} = \lim_{n\to\infty}\int_{b\left(\frac{n\tau_2}{k_2}\right)}^\infty \frac{n\tau_2}{k_2}\bar{G}(y)\frac{dy}{y}.
$$

We can thus write

$$
\begin{aligned}
& \lim_{n\to\infty}\mathbb{E}\left[k_1\left(H^{(1)}_{k_1,\tau_1,n}-\gamma\right)\left(H^{(2)}_{k_2,\tau_2,n}-\gamma\right)\right] \\
= \; & c_2(\tau_1\wedge\tau_2)\int_1^\infty\int_1^\infty \Lambda\left(\frac{x^{-1/\gamma}}{\tau_1},\frac{y^{-1/\gamma}}{c_2\tau_2}\right)\frac{dxdy}{xy} && (2.26)\\
& - \lim_{n\to\infty}\mathbb{E}\left[k_1\int_{a\left(\frac{n\tau_1}{k_1}\right)}^{\hat{a}\left(\frac{n\tau_1}{k_1}\right)}\frac{n\tau_1}{k_1}\bar{F}(x)\frac{dx}{x}\int_{b\left(\frac{n\tau_2}{k_2}\right)}^{\hat{b}\left(\frac{n\tau_2}{k_2}\right)}\frac{n\tau_2}{k_2}\bar{G}(y)\frac{dy}{y}\right] && (2.27)\\
& - \lim_{n\to\infty}\mathbb{E}\left[k_1\left(H^{(1)}_{k_1,\tau_1,n}-\gamma\right)\int_{b\left(\frac{n\tau_2}{k_2}\right)}^{\hat{b}\left(\frac{n\tau_2}{k_2}\right)}\frac{n\tau_2}{k_2}\bar{G}(y)\frac{dy}{y}\right] && (2.28)\\
& - \lim_{n\to\infty}\mathbb{E}\left[k_1\left(H^{(2)}_{k_2,n,\tau_2}-\gamma\right)\int_{a\left(\frac{n\tau_1}{k_1}\right)}^{\hat{a}\left(\frac{n\tau_1}{k_1}\right)}\frac{n\tau_1}{k_1}\bar{F}(x)\frac{dx}{x}\right] && (2.29)
\end{aligned}
$$

provided all the limits exist and where we used (2.24). It remains to show that

$$
(2.26)-(2.27)-(2.28)-(2.29) = c_2(\tau_1\wedge\tau_2)\gamma^2\cdot\Lambda\left(\tau_1^{-1},(\tau_2c_2)^{-1}\right). \tag{2.30}
$$

First, we will use Lemmas 7.5 and 7.6 from Dematteo and Clémençon [2016] to prove modified versions of Lemmas 7.7, 7.8 and 7.9 from the previous reference. These results

allow us to calculate analytical expressions of (2.27), (2.28) and (2.29), which, in turn, is used to show (2.30) in the final step. It should be noted that the tail measure ν used in Dematteo and Clémençon [2016] satisfies $\nu(x,y) = \Lambda(x^{-1/\gamma}, y^{-1/\gamma})$.
For ease of notation, let $p_{i,n\tau} = \frac{n\tau - i + 1}{n\tau}$, $U_i = F(X_i)$ and $V_i = G(Y_i)$. The densities of F and G will be denoted by f and g, respectively.

Lemma 2.8 (First auxiliary result)
For intermediate sequences k_1, k_2, i.e. $k_j \to \infty$ and $k_j/n \to 0$, we have

$$\mathbb{E}\left[\log X_{n\tau_1 - i + 1:n\tau_1} \int_{b\left(\frac{n\tau_2}{k_2}\right)}^{\hat{b}\left(\frac{n\tau_2}{k_2}\right)} \frac{n\tau_2}{k_2} \bar{F}_Y(y) \frac{dy}{y}\right] = M_{n,\tau_1,\tau_2}(i,k_2) + R_{n,\tau_1,\tau_2,1}(i,k_2) + R_{n,\tau_1,\tau_2,2}(i,k_2)$$

with $M_{n,\tau_1,\tau_2}(i,k_2)$, $R_{n,\tau_1,\tau_2,1}(i,k_2)$ and $R_{n,\tau_1,\tau_2,2}(i,k_2)$ given in the proof and

$$R_{n,\tau_1,\tau_2,1}(k_1,k_2) = O\left(\frac{n^{-3/2}(\log\log n)^{-1/2}(\log n)^{-1}}{a(n\tau_1/k_1)\, b(n\tau_2/k_2)}\right), \tag{2.31}$$

$$R_{n,\tau_1,\tau_2,2}(k_1,k_2) = O\left(\frac{n^{-3/4}(\log\log n)^{-1/4}(\log n)^{-1/2}}{b(n\tau_2/k_2)}\right). \tag{2.32}$$

Proof. With Dematteo and Clémençon [2016, Lem. 7.5] we obtain

$$\begin{aligned}
&\mathbb{E}\left[\log X_{n\tau_1 - i + 1:n\tau_1} \int_{b\left(\frac{n\tau_2}{k_2}\right)}^{\hat{b}\left(\frac{n\tau_2}{k_2}\right)} \frac{n\tau_2}{k_2} \bar{G}(y) \frac{dy}{y}\right] \\
&= \mathbb{E}\left[\left(\log a\left(\frac{n\tau_1}{i}\right) + \frac{p_{i,n\tau_1} - \frac{1}{n\tau_1}\sum_{j=1}^{n\tau_1} 1_{\{U_j \le p_{i,n\tau_1}\}}}{a\left(\frac{n\tau_1}{i}\right) f\left(a\left(\frac{n\tau_1}{i}\right)\right)}\right) \cdot \left(\frac{p_{k,n\tau_2} - \frac{1}{n\tau_2}\sum_{j=1}^{n\tau_2} 1_{\{V_j \le p_{k,n\tau_2}\}}}{b\left(\frac{n\tau_2}{k_2}\right) g\left(b\left(\frac{n\tau_2}{k_2}\right)\right)}\right)\right] \\
&\quad + \mathbb{E}\left[O\left(\frac{T_{n\tau_1}(p_{i,n\tau_1})}{a\left(\frac{n\tau_1}{i}\right)} \int_{b\left(\frac{n\tau_2}{k_2}\right)}^{\hat{b}\left(\frac{n\tau_2}{k_2}\right)} \frac{n\tau_2}{k_2} \bar{G}(y) \frac{dy}{y}\right)\right] + \mathbb{E}\left[\log X_{n\tau_1 - i + 1:n\tau_1} \frac{T_{n\tau_2}(p_{k_2,n\tau_2})}{b\left(\frac{n\tau_2}{k_2}\right)}\right] \\
&= M_{n,\tau_1,\tau_2}(i,k_2) + R_{n,\tau_1,\tau_2,1}(i,k_2) + R_{n,\tau_1,\tau_2,2}(i,k_2),
\end{aligned}$$

where, since $\mathbb{P}(V_j \le p) - p = 0$ for $p \in (0,1)$, the first summand is equal to

$$M_{n,\tau_1,\tau_2}(i,k_2) = \mathbb{E}\left[\frac{\frac{1}{n\tau_1}\sum_{j=1}^{n\tau_1}(1_{\{U_j \le p_{i,n\tau_1}\}} - p_{i,n\tau_1}) \frac{1}{n\tau_2}\sum_{j=1}^{n\tau_2}(1_{\{V_j \le p_{k_2,n\tau_2}\}} - p_{k_2,n\tau_2})}{a(n\tau_1/i) \cdot f(a(n\tau_1/i)) \cdot b(n\tau_2/k_2) \cdot g(b(n\tau_2/k_2))}\right].$$

(2.31) and (2.32) follow directly from Dematteo and Clémençon [2016, Lem. 7.7]. □

Lemma 2.9 (Second auxiliary result)
For $i = 1, \ldots, k$ we have that

$$M_{n,\tau_1,\tau_2}(i,k_2) \sim (\tau_1 \wedge \tau_2)\gamma^2 \frac{n}{ik_2} \mathbb{P}\left(X_1 > a\left(\frac{n\tau_1}{i}\right), Y_1 > b\left(\frac{n\tau_2}{k_2}\right)\right)$$

and in particular

$$M_{n,\tau_1,\tau_2}(k_1,k_2) \sim (\tau_1 \wedge \tau_2)\gamma^2 \frac{1}{k_2} \cdot \Lambda\left(\tau_1^{-1}, (c_2\tau_2)^{-1}\right).$$

Proof. Note that $\bar{F}(a(n\tau_1/i)) = 1 - p_{i,n\tau_1} \sim \frac{i}{n\tau_1}$ and thus, by applying the von Mises condition (2.4), we have

$$a\left(\frac{n\tau_1}{i}\right) \cdot f\left(a\left(\frac{n\tau_1}{i}\right)\right) \sim \gamma^{-1}\frac{i}{n\tau_1} \text{ and } b\left(\frac{n\tau_2}{k_2}\right) \cdot g\left(b\left(\frac{n\tau_2}{k_2}\right)\right) \sim \gamma^{-1}\frac{k_2}{n\tau_2}.$$

This leads to

$$\begin{aligned} M_{n,\tau_1,\tau_2}(i,k_2) =& \mathbb{E}\left[\frac{\frac{1}{n\tau_1}\sum_{j=1}^{n\tau_1}(1_{\{U_j \le p_{i,n\tau_1}\}} - p_{i,n\tau_1})\frac{1}{n\tau_2}\sum_{j=1}^{n\tau_2}(1_{\{V_j \le p_{k_2,n\tau_2}\}} - p_{k_2,n\tau_2})}{a\left(\frac{n\tau_1}{i}\right) f(a\left(\frac{n\tau_1}{i}\right))b\left(\frac{n\tau_2}{k_2}\right) g(b\left(\frac{n\tau_2}{k_2}\right))}\right] \\ =& \frac{\tau_1 \wedge \tau_2}{\tau_1\tau_2}\frac{\mathbb{P}\left(X_1 > a\left(\frac{n\tau_1}{i}\right), Y_1 > b\left(\frac{n\tau_2}{k_2}\right)\right) - (1-p_{i,n\tau_1})(1-p_{k_2,n\tau_2})}{n \cdot a\left(\frac{n\tau_1}{i}\right) f\left(a\left(\frac{n\tau_1}{i}\right)\right) b\left(\frac{n\tau_2}{k_2}\right) g\left(b\left(\frac{n\tau_2}{k_2}\right)\right)}. \\ \sim& (\tau_1 \wedge \tau_2)\gamma^2 \frac{n}{ik_2}\mathbb{P}\left(X_1 > a\left(\frac{n\tau_1}{i}\right), Y_1 > b\left(\frac{n\tau_2}{k_2}\right)\right). \end{aligned}$$

Consequently,

$$\begin{aligned} M_{n,\tau_1,\tau_2}(k_1,k_2) \sim& (\tau_1 \wedge \tau_2)\gamma^2 \frac{1}{k_2}\frac{n}{k_1} P\left(X_1 > a\left(\frac{n\tau_1}{k_1}\right), Y_1 > b\left(\frac{n\tau_2}{k_2}\right)\right) \\ \sim& (\tau_1 \wedge \tau_2)\gamma^2 \frac{1}{k_2} \cdot \Lambda\left(\tau_1^{-1}, (c_2\tau_2)^{-1}\right). \end{aligned}$$

□

Lemma 2.10 (Third auxiliary result)
We have that

$$\lim_{n\to\infty} \sum_{i=1}^{k_1} \frac{n}{ik_2}\mathbb{P}\left(X_1 > a\left(\frac{n\tau_1}{i}\right), Y_1 > b\left(\frac{n\tau_2}{k_2}\right)\right) = c_2\frac{1}{\gamma}\int_1^\infty \Lambda\left(\frac{x^{-1/\gamma}}{\tau_1}, \frac{1}{c_2\tau_2}\right)\frac{dx}{x} \quad (2.33)$$

The proof follows from exactly the same lines as in Dematteo and Clémençon [2016, Lem. 7.9] and is thus omitted.

Now, using the auxiliary Lemmas 2.8, 2.9 and 2.10, we are able to calculate (2.27), (2.28) and (2.29):

Lemma 2.11 (Calculation of (2.27))
We have that

$$(2.27) = c_2(\tau_1 \wedge \tau_2)\gamma^2 \cdot \Lambda\left(\tau_1^{-1}, (c_2\tau_2)^{-1}\right). \quad (2.34)$$

Proof. We use Dematteo and Clémençon [2016, Lem. 7.5] and Lemma 2.9 to get

$$
\begin{aligned}
&\mathbb{E}\left[k_1 \int_{a\left(\frac{n\tau_1}{k_1}\right)}^{\hat{a}\left(\frac{n\tau_1}{k_1}\right)} \frac{n\tau_1}{k_1}\bar{F}(x)\frac{dx}{x} \int_{b\left(\frac{n\tau_2}{k_2}\right)}^{\hat{b}\left(\frac{n\tau_2}{k_2}\right)} \frac{n\tau_2}{k_2}\bar{G}(y)\frac{dy}{y}\right] \\
=&k_1\mathbb{E}\left[\left(\frac{p_{k_1,n\tau_1} - F_{n\tau_1}\left(a\left(\frac{n\tau_1}{k_1}\right)\right)}{a\left(\frac{n\tau_1}{k_1}\right)\cdot f\left(a\left(\frac{n\tau_1}{k_1}\right)\right)} + \frac{T_{n\tau_1}(p_{k_1,n\tau_1})}{a\left(\frac{n\tau_1}{k_1}\right)}\right)\times \right. \\
&\qquad\qquad \left.\left(\frac{p_{k_2,n\tau_2} - G_{n\tau_2}\left(b\left(\frac{n\tau_2}{k_2}\right)\right)}{b\left(\frac{n\tau_2}{k_2}\right)\cdot g\left(b\left(\frac{n\tau_2}{k_2}\right)\right)} + \frac{T_{n\tau_2}(p_{k_2,n\tau_2})}{b\left(\frac{n\tau_2}{k_2}\right)}\right)\right] \\
=&k_1\mathbb{E}\left[\frac{\frac{1}{n\tau_1\tau_2}\sum_{i=1}^{n(\tau_1\wedge\tau_2)}\left[\mathbb{P}\left(X_i > a\left(\frac{n\tau_1}{k_1}\right), Y_i > b\left(\frac{n\tau_2}{k_2}\right)\right) - (1-p_{k_1,n\tau_1})(1-p_{k_2,n\tau_2})\right]}{a\left(\frac{n\tau_1}{k_1}\right)\cdot f\left(a\left(\frac{n\tau_1}{k_1}\right)\right)\cdot b\left(\frac{n\tau_2}{k_2}\right)\cdot g\left(b\left(\frac{n\tau_2}{k_2}\right)\right)}\right] + o(1) \\
=&\frac{k_1}{k_2}k_2\cdot M_{n,\tau_1,\tau_2}(k_1,k_2) + o(1) \overset{Lem\ 2.9}{\longrightarrow} c_2(\tau_1\wedge\tau_2)\gamma^2\cdot\Lambda\left(\tau_1^{-1},(c_2\tau_2)^{-1}\right).
\end{aligned}
$$

□

Lemma 2.12 (Calculation of (2.28) **and** (2.29)**)**
We have that

$$
(2.28) = c_2(\tau_1\wedge\tau_2)\left[\gamma\int_1^\infty \Lambda\left(\frac{x^{-1/\gamma}}{\tau_1},\frac{1}{c_2\tau_2}\right)\frac{dx}{x} - \gamma^2\Lambda\left(\tau_1^{-1},(c_2\tau_2)^{-1}\right)\right]
$$

and

$$
(2.29) = c_2(\tau_1\wedge\tau_2)\left[\gamma\int_1^\infty \Lambda\left(\frac{1}{\tau_1},\frac{y^{-1/\gamma}}{c_2\tau_2}\right)\frac{dy}{y} - \gamma^2\Lambda\left(\tau_1^{-1},(c_2\tau_2)^{-1}\right)\right].
$$

Proof. With the same arguments as in Dematteo and Clémençon [2016, Lem. 7.4] we arrive at

$$
\begin{aligned}
(2.28) &= \lim_{n\to\infty}\sum_{i=1}^{k_1}\left(M_{n,\tau_1,\tau_2}(i,k_2) - M_{n,\tau_1,\tau_2}(k_1,k_2)\right) \\
&\overset{Lem\ 2.9}{=}(\tau_1\wedge\tau_2)\gamma^2\lim_{n\to\infty}\sum_{i=1}^{k_1}\frac{n}{ik_2}\mathbb{P}\left(X_1 > a\left(\frac{n\tau_1}{i}\right), Y_1 > b\left(\frac{n\tau_2}{k_2}\right)\right) \\
&\qquad - \lim_{n\to\infty} k\cdot M_{n,\tau_1,\tau_2}(k_1,k_2) \\
&\overset{Lem\ 2.10}{=}c_2(\tau_1\wedge\tau_2)\left[\gamma\int_1^\infty \Lambda\left(\frac{x^{-1/\gamma}}{\tau_1},\frac{1}{c_2\tau_2}\right)\frac{dx}{x} - \gamma^2\Lambda\left(\tau_1^{-1},(c_2\tau_2)^{-1}\right)\right].
\end{aligned}
$$

□

Finally, we apply Lemmas 2.11 and 2.12 to calculate (2.30):

Lemma 2.13 (Calculation of (2.30)**)**
We have that

$$\int_1^\infty \int_1^\infty \Lambda\left(\frac{x^{-1/\gamma}}{\tau_1}, \frac{y^{-1/\gamma}}{c_2\tau_2}\right)\frac{dxdy}{xy} = \gamma\int_1^\infty \Lambda\left(\frac{x^{-1/\gamma}}{\tau_1}, \frac{1}{c_2\tau_2}\right)\frac{dx}{x} + \gamma\int_1^\infty \Lambda\left(\frac{1}{\tau_1}, \frac{y^{-1/\gamma}}{c_2\tau_2}\right)\frac{dy}{y}$$

and in particular, (2.30) *follows.*

Proof. With $\gamma = \int_1^\infty s^{-1/\gamma}\frac{ds}{s}$ and $\Lambda(tx,ty) = t\Lambda(x,y)$ for all $t,x,y > 0$ [Schmidt and Stadtmüller, 2006] we obtain

$$\begin{aligned}\gamma\int_1^\infty \Lambda\left(\frac{x^{-1/\gamma}}{\tau_1}, \frac{1}{c_2\tau_2}\right)\frac{dx}{x} &= \int_1^\infty\int_1^\infty y^{-1/\gamma}\Lambda\left(\frac{x^{-1/\gamma}}{\tau_1}, \frac{1}{c_2\tau_2}\right)\frac{dxdy}{xy}\\ = \int_1^\infty\int_1^\infty \Lambda\left(\frac{(xy)^{-1/\gamma}}{\tau_1}, \frac{y^{-1/\gamma}}{c_2\tau_2}\right)\frac{dxdy}{xy} &= \int_1^\infty\int_y^\infty \Lambda\left(\frac{x^{-1/\gamma}}{\tau_1}, \frac{y^{-1/\gamma}}{c_2\tau_2}\right)\frac{dxdy}{xy}\end{aligned}$$

and similarly,

$$\gamma\int_1^\infty \Lambda\left(\frac{1}{\tau_1}, \frac{y^{-1/\gamma}}{c_2\tau_2}\right)\frac{dy}{y} = \int_1^\infty\int_x^\infty \Lambda\left(\frac{x^{-1/\gamma}}{\tau_1}, \frac{y^{-1/\gamma}}{c_2\tau_2}\right)\frac{dydx}{yx}.$$

Finally, note that $[1,\infty)^2 = \{(x,y) : y \geq 1, x \geq y\}\dot\cup\{(x,y) : x \geq 1, y \geq x\}$. This completes the proof of Theorem 2.4. □

2.8.2 Proof of Propositions 2.5, 2.6 and 2.7

Proof of Proposition 2.5. Note that

$$\sqrt{k_1}\left(\hat\gamma_{opt} - \gamma\right) = \hat{\boldsymbol{w}}_{opt}'\sqrt{k_1}\left(\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \gamma\mathbf{1}\right).$$

The assertion thus follows from Theorem 2.4, $\hat{\boldsymbol{w}}_{opt} \xrightarrow{\mathbb{P}} \boldsymbol{w}_{opt}$, Slutsky's lemma and from continuity of linear maps. □

Proof of Proposition 2.6. We denote the centered Hill vector by $\mathbf{H}^c_{\mathbf{k},\boldsymbol{\tau},n} = \mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \gamma\mathbf{1}$. Then, for arbitrary weights $\boldsymbol{w} \in \mathbb{W}$, we have that

$$\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \widehat\gamma_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})\cdot\mathbf{1} = A_{\boldsymbol{w}}\cdot\mathbf{H}^c_{\mathbf{k},\boldsymbol{\tau},n},$$

where we set $A_{\boldsymbol{w}} = I_d - \mathbf{1}\cdot\boldsymbol{w}' \in \mathbb{R}^{d\times d}$ with identity matrix $I_d \in \mathbb{R}^{d\times d}$.
Let $\mathbf{Z}$ denote a random vector with distribution $\mathcal{N}(0,\ \mathbf{I}_d)$. By assumption we have that

$$\sqrt{k_1}\left(\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \gamma\mathbf{1}\right) \xrightarrow{D} \gamma\Sigma^{1/2}\mathbf{Z}$$

and, as a byproduct, also that $\widehat\gamma_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w}) \xrightarrow{\mathbb{P}} \gamma$. From Slutsky's lemma and the continuous mapping theorem, we obtain

$$\tilde W_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w}) = \frac{k_1}{\widehat\gamma_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})^2}(A_{\boldsymbol{w}}\mathbf{H}^c_{\mathbf{k},\boldsymbol{\tau},n})'\widehat\Sigma^{-1}A_{\boldsymbol{w}}\mathbf{H}^c_{\mathbf{k},\boldsymbol{\tau},n} \xrightarrow{D} \mathbf{Z}'\left(A_{\boldsymbol{w}}\Sigma^{1/2}\right)'\Sigma^{-1}A_{\boldsymbol{w}}\Sigma^{1/2}\mathbf{Z}.$$

Note that the matrix $B_{\boldsymbol{w}} = \left(A_{\boldsymbol{w}}\Sigma^{1/2}\right)' \Sigma^{-1} A_{\boldsymbol{w}}\Sigma^{1/2}$ is symmetric. The spectral theorem from linear algebra guarantees the existence of a matrix $O_{\boldsymbol{w}}$ with $O'_{\boldsymbol{w}} \cdot O_{\boldsymbol{w}} = O_{\boldsymbol{w}} \cdot O'_{\boldsymbol{w}} = \mathbf{I}_d$ and a diagonal matrix $D_{\boldsymbol{w}}$ containing all eigenvalues of $B_{\boldsymbol{w}}$ on the diagonal, such that $B_{\boldsymbol{w}} = O_{\boldsymbol{w}} D_{\boldsymbol{w}} O'_{\boldsymbol{w}}$ holds. Note also that $O \cdot \mathbf{Z} \stackrel{D}{=} \mathbf{Z}$ for any orthogonal matrix $O \in \mathbb{R}^{d\times d}$. Summing up, we obtain

$$\tilde{W}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w}) \xrightarrow{D} \mathbf{Z}' D_{\boldsymbol{w}} \mathbf{Z}.$$

We continue with the calculation of the d diagonal elements in $D_{\boldsymbol{w}}$, i.e., the eigenvalues of $B_{\boldsymbol{w}}$ and their algebraic multiplicity:

- It immediately follows that $\Sigma^{-1/2}\mathbf{1}$ is an eigenvector of $B_{\boldsymbol{w}}$ with eigenvalue 0.
- Let $V_{\boldsymbol{w}} = span(\Sigma^{-1/2}\mathbf{1}, \Sigma^{1/2}\boldsymbol{w})$ and $V_{\boldsymbol{w}}^{\perp}$ denote the orthogonal complement of $V_{\boldsymbol{w}}$ with respect to the scalar product. Then, for any $\mathbf{x} \in V_{\boldsymbol{w}}^{\perp}$, we have that $B_{\boldsymbol{w}}\mathbf{x} = \mathbf{x}$. $\dim\left(V_{\boldsymbol{w}}^{\perp}\right) \in \{d-2, d-1\}$ implies that 1 is an eigenvalue of $B_{\boldsymbol{w}}$ with algebraic multiplicity of $d-2$ or $d-1$.
- If $\dim\left(V^{\perp}\right) = d-1$, we are done. Otherwise, let $E_{\boldsymbol{w}} = (\Sigma^{-1/2}\mathbf{1}, \Sigma^{1/2}\boldsymbol{w}) \in \mathbb{R}^{d\times 2}$,

$$F_{\boldsymbol{w}} = \begin{pmatrix} -\boldsymbol{w}'\Sigma\boldsymbol{w} & 1 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

and note that $\mathbf{I}_d - B_{\mathbf{w}} = E_{\mathbf{w}} F_{\mathbf{w}} E'_{\mathbf{w}}$. From linear algebra we know that every non-zero eigenvalue of $E_{\mathbf{w}} F_{\mathbf{w}} E'_{\mathbf{w}}$ is necessarily also an eigenvalue of the matrix

$$F_{\mathbf{w}} E'_{\mathbf{w}} E_{\mathbf{w}} = \begin{pmatrix} 1 - \boldsymbol{w}'\Sigma\boldsymbol{w}\mathbf{1}'\Sigma^{-1}\mathbf{1} & 0 \\ \mathbf{1}'\Sigma^{-1}\mathbf{1} & 1 \end{pmatrix}.$$

From the last expression we conclude that $1 - \lambda_{\mathbf{w}} = 1 - \boldsymbol{w}'\Sigma\boldsymbol{w}\mathbf{1}'\Sigma^{-1}\mathbf{1}$ is an eigenvalue of $\mathbf{I}_d - B_{\boldsymbol{w}}$, which implies that $\lambda_{\boldsymbol{w}}$ is an eigenvalue of $B_{\boldsymbol{w}}$.

Next note that $\boldsymbol{w}'_{opt}\Sigma\boldsymbol{w}_{opt}\mathbf{1}'\Sigma^{-1}\mathbf{1} = 1$ for $\boldsymbol{w}_{opt} = (\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1} \cdot \Sigma^{-1}\mathbf{1}$. Finally, from $\hat{\boldsymbol{w}}_{opt} \xrightarrow{\mathbb{P}} \boldsymbol{w}_{opt}$ and the continuous mapping theorem, we have that $W_{\mathbf{k},\boldsymbol{\tau},n} \xrightarrow{D} \chi^2_{d-1}$.
For the remaining part of the proof, let $F_1, \ldots, F_d$ be all of Pareto-type with extreme value indices $\gamma_1, \ldots, \gamma_d$, but with $\gamma_i \neq \gamma_j$ for some $1 \leq i, j \leq d$. From Theorem 1.12 we have that $\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \widehat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})\mathbf{1} \xrightarrow{P} \mathbf{b} \in \mathbb{R}^d$, $\mathbf{b} \neq 0$. From the positive definiteness of Σ and the consistency of $\hat{\Sigma}$ we thus have $W_{\mathbf{k},\boldsymbol{\tau},n}/k_1 \xrightarrow{\mathbb{P}} const. > 0$, which implies that $W_{\mathbf{k},\boldsymbol{\tau},n} \xrightarrow{\mathbb{P}} \infty$. This completes the proof of Proposition 2.6. □

Proof of Proposition 2.7. We only need to verify joint weak convergence of

$$\sqrt{k_1}\left(\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n} - \gamma,\ \frac{u_j(k_j, n_j)}{F_j^{-1}(n_j/k_j)}\right) \tag{2.35}$$

towards a bivariate normal distribution. Then all the assumptions of Theorem 1.14 are valid and the assertion follows.

Recall from the proof of Theorem 2.4 that weak convergence of the first component in (2.35) is based on that of a stochastic process S_n, also called tail empirical process. Similarly, weak convergence of the second component in (2.35) can by shown by a more general result on tail empirical quantile processes [de Haan and Ferreira, 2006, Def. 2.4.3 and Th. 2.4.8]. Thanks to Vervaat's lemma [Vervaat, 1971], these two tail processes are closely related, which means that both components can be viewed as continuous mappings of one process up to some negligible summand and joint weak convergence follows. For details on tail empirical processes we refer to de Haan and Ferreira [2006, Sec. 5.1].

Chapter 3

Detecting change-points in the dependence of extremes

The aim of this chapter is to test whether the dependence between maximal values has changed during the observation period. The procedure is also extended to allow for the detection of changes in the dependence under known abrupt changes in the marginal distributions. We conclude the chapter with illustrations from flood frequency analysis considering bivariate time series of river flows, where the behavior of margins possibly has changed due to the construction of a dam during the observation period and where still it is of interest to test whether the dependence has changed.

This chapter is based on the article by Bücher et al. [2015] (accepted for publication in *Extremes*). The work started with my interest in the madogram estimator (1.15) and related change-point tests. When noticing that this problem is connected with the theory developed in Bücher and Kojadinovic [2016], I started collaborating with the authors of that paper.
My main responsibility lied in the computational and applied part of this work. The theoretical results, many related to those in Bücher and Kojadinovic [2016], have been mostly carried out by Axel Bücher and Ivan Kojadinovic, with their original argumentation summarized in the appendix of Bücher et al. [2015]. Here, however, I have modified some of the details of the proofs and rearranged the statements in a slightly different order.

3.1 Introduction

The study of maximal values is of importance in many environmental applications. Prominent examples are the analysis of floods [Hosking and Wallis, 2005], heavy rainfalls [Cooley et al., 2007] and extreme temperatures [Katz and Brown, 1992]. Many of these problems are intrinsically multivariate; for instance, the severity of a flood depends not only on its peak flow, which is considered in many univariate flood studies, but also on its volume and its duration [Yue et al., 1999]. Catastrophic floods typically occur when more than one of these variables is taking a high value and therefore, the analysis of the joint

behavior is of key importance. In a river system, where flood data are collected from a number of stations, inference at a specific location can be improved by incorporating observations from neighboring stations [Hosking and Wallis, 2005]. Similarly, extreme temperatures are commonly studied at several stations simultaneously.
In such applications, it is common practice to assume that the time series of block maxima is temporarily independent, extreme value distributed and stationary. It is the aim of this paper to develop tests for change-point detection within the distribution of the block maxima. More precisely, assuming that we observe a sample of independent multivariate observations $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, where each $\mathbf{Y}_i$ follows a multivariate extreme value distribution whose distribution function is denoted $H^{(i)}$, we develop a test for the hypothesis

$$\mathcal{H}_0 : \; H^{(1)} = \ldots = H^{(n)} \tag{3.1}$$

against alternatives involving abrupt changes in the extreme value distribution. Since the univariate version of this problem has been treated, for instance, in Jarušková and Rencová [2008] using results from Chapter 1 of Csörgő and Horváth [1997], we will be particularly interested in the multivariate setting throughout this paper.
Outside of the extreme value framework, there is a huge amount of literature on detecting deviations from $\mathcal{H}_0$. We refer to Aue and Horváth [2013] for a recent review. It is useful to note that, by Theorem 1.1, we can rewrite $\mathcal{H}_0$ as $\mathcal{H}_{0,m} \cap \mathcal{H}_{0,c}$, where intersection $\cap$ means that both hypotheses are met and where

$$\mathcal{H}_{0,m} : H^{(1)}, \ldots, H^{(n)} \text{ have same margins,} \tag{3.2}$$

$$\mathcal{H}_{0,c} : H^{(1)}, \ldots, H^{(n)} \text{ have the same copula (i.e., dependence).} \tag{3.3}$$

Roughly speaking, common tests for $\mathcal{H}_0$ can be divided into two groups: tests that are powerful against deviations from $\mathcal{H}_{0,m}$ but which are rather insensitive to deviations from $\mathcal{H}_{0,c}$, and vice versa, see Bücher et al. [2014] for a discussion. In the present setting of observing data from a multivariate extreme value distribution, the tests considered for instance in Jarušková and Rencová [2008] can be used to detect deviations from $\mathcal{H}_{0,m}$, whence it will be our main concern to design a test that is particularly sensitive to deviations from $\mathcal{H}_{0,c}$ when the copulas $C^{(i)}$ are known to be of extreme value type. Denoting the corresponding Pickands dependence function by $A^{(i)}$ (see Theorem 1.4), we can rewrite $\mathcal{H}_{0,c}$ equivalently as

$$\mathcal{H}_{0,A} : \; A^{(1)} = \ldots = A^{(n)}. \tag{3.4}$$

The test statistic in the subsequent sections will be particularly designed for detecting deviations from $\mathcal{H}_{0,A}$.
Note that none of the existing tests for changes in the copula [see, e.g., Quessy et al., 2013; Bücher et al., 2014; Dehling et al., 2014; Kojadinovic et al., 2015, among others] incorporates the information that C is of the extreme-value type, whence an improvement in the power properties seems possible. In fact, our simulation study reported in Section 3.3 suggests that our proposed testing procedure is indeed superior to existing methods.

The test tailored to deal with extreme-value dependence that we propose is however affected by the same limitations as most of the aforementioned more general testing procedures: it can be used to reject $\mathcal{H}_{0,c}$ only if $\mathcal{H}_{0,m}$ holds. In some situations, although there are reasons to consider that $\mathcal{H}_{0,m}$ is not true, it is still of interest to assess whether $\mathcal{H}_{0,c}$ holds or not. For instance, in the hydrological applications to be presented in Section 3.4, the construction of dams during the observation period suggests that there might be potential breaks in the marginal distributions of extreme peak flows or volumes, while it is still of interest to assess whether the dependence between the variables of interest has changed or not. A second contribution of this work is thus to propose an extension of the studied testing procedure similar to that considered in Quessy et al. [2013] that can be used to detect deviations from $\mathcal{H}_{0,c}$ under certain types of simple departures from $\mathcal{H}_{0,m}$. The remainder of this chapter is organized as follows: Section 3.2 introduces the test statistic, its extension and summarizes the main theoretical results. A comparative simulation study is presented in Section 3.3 and an application to hydrological time series is given in Section 3.4. Proofs are deferred to Section 3.5.

3.2 A new change-point test under extreme value dependence

For the remainder of this chapter we will restrict ourselves to the case of two-dimensional time series. All the results below can be extended to arbitrary dimensions $d \geq 2$, which is briefly summarized in Appendix A.5.

Suppose that $(X_i, Y_i)'$, $i = 1, \ldots, n$, is a sequence of independent random vectors with unknown distribution functions

$$\mathbb{P}\left(X_i \leq x,\, Y_i \leq y\right) = H^{(i)}(x,y) = C^{(i)}\left(F^{(i)}(x), G^{(i)}(y)\right),\ (x,y)' \in \mathbb{R}^2,$$

copulas $C^{(i)}$ and continuous margins $F^{(i)},\ G^{(i)}$. We assume that each $C^{(i)}$ is an extreme value copula with Pickands dependence function $A^{(i)} : [0,1] \to [1/2, 1]$ and we aim at developing tests for $\mathcal{H}_0$ in (3.1) that are particularly powerful against alternatives

$$\mathcal{H}_{1,A} :\ A^{(1)} = \ldots = A^{(m)} \neq A^{(m+1)} = \ldots = A^{(n)} \ \text{ for some unkown } \ m. \tag{3.5}$$

3.2.1 Test statistic for $d = 2$ under stationarity of the margins

Our test statistic is based on sequential estimation of the Pickands dependence function: Starting from an adaption of pseudo observations,

$$\hat{U}_{k:\ell,i} = \frac{1}{\ell - k + 2} \sum_{j=k}^{\ell} \mathbb{1}\left(X_j \leq X_i\right) \ \text{ and } \ \hat{V}_{k:\ell,i} = \frac{1}{\ell - k + 2} \sum_{j=k}^{\ell} \mathbb{1}\left(Y_j \leq Y_i\right), \tag{3.6}$$

computed from arbitrary subsamples $(X_i, Y_i)'$, $1 \leq k \leq i \leq \ell \leq n$, we define

$$\hat{S}_{k:\ell}(t) = \frac{1}{\ell - k + 1} \sum_{i=k}^{\ell} \max\left(\hat{U}_{k:l,i}^{1/(1-t)}, \hat{V}_{k:l,i}^{1/t}\right), \quad t \in [0,1], \tag{3.7}$$

where we set $\hat{S}_{k:\ell} \equiv 0$ if $k > \ell$. In analogy to (1.15), we define

$$\hat{A}_{k:\ell}(t) = \frac{\hat{S}_{k:\ell}(t)}{1 - \hat{S}_{k:\ell}(t)}, \quad t \in [0,1]. \tag{3.8}$$

Under $\mathcal{H}_0$, we may expect that some distance between $\hat{A}_{1:k}$ and $\hat{A}_{k+1:n}$ should be small, relative to sampling variability and for any $k = 1, \ldots, n-1$. The opposite should be expected, if alternative $\mathcal{H}_{1,A}$ in (3.5) is met. Since we do not know the potential change-point m in (3.5), it is natural to study the behavior of the stochastic process

$$\mathbb{D}_n(s,t) = \frac{\lfloor ns \rfloor (n - \lfloor ns \rfloor)}{n^{3/2}} \left\{ \hat{A}_{1:\lfloor ns \rfloor}(t) - \hat{A}_{\lfloor ns \rfloor + 1:n}(t) \right\}, \ (s,t) \in [0,1]^2. \tag{3.9}$$

The standardization in (3.9) is not unusual in such CUSUM-type (CUMulative SUMs) change-point tests. It is used to prevent processes like $\mathbb{D}_n$ from exploding for s at the boundaries of $[0,1]$. The following proposition establishes weak convergence of process $\mathbb{D}_n$ under $\mathcal{H}_0$ in (3.1).

Proposition 3.1
Suppose that $\mathcal{H}_0$ holds and that A is continuously differentiable on $(0,1)$. Then, in the normed space $(\ell^\infty([0,1]^2), \|\cdot\|_\infty)$ and for $n \to \infty$, we have that $\mathbb{D}_n \xrightarrow{D} \mathbb{D}_C$, where

$$\mathbb{D}_C(s,t) = \{1 + A(t)\}^2 \int_0^1 s\mathbb{C}_C(s, 1, y^{1-t}, y^t) - (1-s)\mathbb{C}_C(0, s, y^{1-t}, y^t)\, dy. \tag{3.10}$$

$\mathbb{C}_C$ denotes a centered Gaussian process defined in Theorem A.8 and where $d = 2$.

Most test statistics in the change-point literature are based on a uniform (Kolmogorov–Smirnov) or L_2 metric (Cramer-von Mises, Anderson-Darling) of $\mathbb{D}_n$. Throughout this paper, we focus on the hybrid version

$$S_{n,A} = \max_{1 \le k < n} \int_{[0,1]} \{\mathbb{D}_n(k/n, t)\}^2 \, d\mu(t), \tag{3.11}$$

where μ denotes some finite measure on $[0,1]$. In the simulation experiments of Section 3.3 we use $\mu = T^{-1} \sum_{t \in \Gamma} \delta_t$ for some finite grid $\Gamma = \{t_1, \ldots, t_T\} \subset [0,1]$, where δ_t is the Dirac mass at t. A corresponding two-sample version of the test statistic for detecting change-points at some pre-specified point $1 \le k^\star < n$ is defined by

$$S_{n,A}(k^\star) = \int_{[0,1]} \{\mathbb{D}_n(k^\star/n, t)\}^2 \, d\mu(t). \tag{3.12}$$

Note that, for the purpose of testing $\mathcal{H}_0$ in (3.1) and because the map $s \mapsto s/(1-s)$ is one-to-one, we may replace $\hat{A}_{k:\ell}$ by $\hat{S}_{k:\ell}$ in (3.9). However, additional simulations (not reported here) revealed that the version based on $\hat{A}_{k:\ell}$ leads to tests with a better finite-sample behavior.

The limit process $\mathbb{D}_C$ in (3.10) depends in an intractable way on the unknown copula C, and, as a consequence, so will do the limit distributions of the test statistics $S_{n,A}$ in (3.11) and $S_{n,A}(k^\star)$ in (3.12). The following *multiplier bootstrap technique* allows us to construct suitable approximations of p-values:
Let B be some large integer. Theorem A.9 is used to approximate process $\mathbb{C}_C$ appearing in (3.10) by multiplier bootstrap versions $\check{\mathbb{C}}_n^{(b)}$, $b = 1, \ldots, B$. We thus may hope that replacing $\mathbb{C}_C$ and $A(t)$ in (3.10) by $\check{\mathbb{C}}_n^{(b)}$ and a consistent estimator of $A(t)$, respectively, gives us bootstrap approximations of $\mathbb{D}_C$. More precisely, a tedious but straightforward calculation suggests to define

$$\check{\mathbb{D}}_n^{(b)}(s,t) = \{1 + \hat{A}_{1:n}(t)\}^2 \cdot \left[\frac{\lfloor ns \rfloor}{n^{3/2}} \sum_{i=\lfloor ns \rfloor + 1}^{n} \xi_i^{(b)} \hat{w}_{\lfloor ns \rfloor + 1:n,i}(t) - \frac{n - \lfloor ns \rfloor}{n^{3/2}} \sum_{i=1}^{\lfloor ns \rfloor} \xi_i^{(b)} \hat{w}_{1:\lfloor ns \rfloor,i}(t) \right]$$

for $(s,t) \in [0,1]^2$, where, for any $1 \le k \le \ell \le n$,

$$\hat{w}_{k:\ell,i}(t) = \bar{m}_{k:\ell}(t) - \hat{m}_{k:\ell,i}(t) + \{\hat{u}_{k:\ell,i}(t) - \bar{u}_{k:\ell}(t)\} \frac{\hat{a}_{k:\ell}(t)}{\hat{b}_{k:\ell}(t)} + \{\hat{v}_{k:\ell,i}(t) - \bar{v}_{k:\ell}(t)\} \frac{\hat{c}_{k:\ell}(t)}{\hat{d}_{k:\ell}(t)},$$

with $\bar{m}_{k:\ell}$, $\bar{u}_{k:\ell}$ and $\bar{v}_{k:\ell}$ denoting the arithmetic means (over $i = k, \ldots, \ell$) of

$$\hat{m}_{k:\ell,i}(t) = \max(\hat{U}_{k:\ell,i}^{1/(1-t)}, \hat{V}_{k:\ell,i}^{1/t}), \qquad \hat{u}_{k:\ell,i}(t) = \hat{U}_{k:\ell,i}^{\hat{b}_{k:\ell}(t)/(1-t)}, \qquad \hat{v}_{k:\ell,i}(t) = \hat{V}_{k:\ell,i}^{\hat{d}_{k:\ell}(t)/t},$$

respectively, and with

$$\begin{aligned} \hat{a}_{k:\ell}(t) &= \hat{A}_{k:\ell}(t) - t\hat{A}'_{k:\ell,n}(t), & \hat{b}_{k:\ell}(t) &= \hat{A}_{k:\ell}(t) + t, \\ \hat{c}_{k:\ell}(t) &= \hat{A}_{k:\ell}(t) + (1-t)\hat{A}'_{k:\ell,n}(t), & \hat{d}_{k:\ell}(t) &= \hat{A}_{k:\ell}(t) + 1 - t, \end{aligned}$$

where, for some positive sequence $h_n \downarrow 0$ such that $\inf_{n \ge 1} h_n \sqrt{n} > 0$,

$$\hat{A}'_{k:\ell,n}(t) = \min[\max\{A'_{k:\ell,n}(t), -1\}, 1], \quad t \in [0,1], \tag{3.13}$$

with

$$A'_{k:\ell,n}(t) = \frac{1}{2h_n} \left\{ \hat{A}_{k:\ell}(t + h_n) - \hat{A}_{k:\ell}(t - h_n) \right\}, \tag{3.14}$$

for $t \in (h_n, 1 - h_n)$, while $A'_{k:\ell,n}(t) = A'_{k:\ell,n}(h_n)$ for $t \le h_n$ and $A'_{k:\ell,n}(t) = A'_{k:\ell,n}(1 - h_n)$ for $t \ge 1 - h_n$.
The following proposition establishes the asymptotic validity of the above resampling scheme under $\mathcal{H}_0$ in (3.1).

Proposition 3.2
Under the conditions of Proposition 3.1 and for $n \to \infty$,

$$\left(\mathbb{D}_n, \check{\mathbb{D}}_n^{(1)}, \ldots, \check{\mathbb{D}}_n^{(B)} \right) \xrightarrow{D} \left(\mathbb{D}_C, \mathbb{D}_C^{(1)}, \ldots, \mathbb{D}_C^{(B)} \right)$$

in $(\ell^\infty([0,1]^2), \|\cdot\|_\infty)^{B+1}$, where $\mathbb{D}_C^{(1)}, \ldots, \mathbb{D}_C^{(B)}$ denote independent copies of $\mathbb{D}_C$.

We define multiplier bootstrap versions $\check{S}_{n,A}^{(b)}$ of the test statistic in (3.11) by

$$\check{S}_{n,A}^{(b)} = \max_{1 \le k < n} \int_{[0,1]} \left\{ \mathbb{D}_n^{(b)}(k/n, t) \right\}^2 d\mu(t).$$

The previous proposition in combination with the continuous mapping theorem allows us to estimate p-values by

$$\hat{p}_{n,B} = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1} \left(S_{n,A}^{(b)} > S_{n,A} \right). \tag{3.15}$$

A test rejecting $\mathcal{H}_0$ if $\hat{p}_{n,B} < \alpha$ for some $\alpha \in (0,1)$ asymptotically keeps its nominal level of α for $n \to \infty$ followed by $B \to \infty$ (see Appendix F in Bücher and Kojadinovic [2016]).

3.2.2 Test statistic for $d = 2$ under known marginal change-points

The test developed in the previous section can be used to reject $\mathcal{H}_{0,c}$ in (3.4) only if we are sure that $\mathcal{H}_{0,m}$ in (3.2) holds. Otherwise, it can be considered only as a test for the more general hypothesis $\mathcal{H}_0$ in (3.1).
In Section 3.4 we will present applications, where it is plausible to believe that $\mathcal{H}_{0,m}$ does not hold and where it is still of interest to test whether $\mathcal{H}_{0,c}$ holds or not. Even more, in our applications it is reasonable to assume that the potential change-point in the margins is known, which allows us to consider the following simple alternatives of $\mathcal{H}_{0,m}$:
For some known $\theta \in (0,1)$, we define

$$\mathcal{H}_{1,m}^{\theta} : \begin{cases} F^{(1)} = \cdots = F^{(\lfloor n\theta \rfloor)} \neq F^{(\lfloor n\theta \rfloor + 1)} = \cdots = F^{(n)} \\ \quad \text{or} \\ G^{(1)} = \cdots = G^{(\lfloor n\theta \rfloor)} \neq G^{(\lfloor n\theta \rfloor + 1)} = \cdots = G^{(n)}, \end{cases} \tag{3.16}$$

which means that one or even both margins change abruptly at the same known point of time $\lfloor n\theta \rfloor$. In the applications of Section 3.4, $\lfloor n\theta \rfloor$ corresponds to the time point right before the construction of a dam.
The aim of this section is to provide a modified version $S_{n,A}^{\theta}$ of the test statistic $S_{n,A}$ from the previous section and to study its distribution under $\mathcal{H}_{1,m}^{\theta} \cap \mathcal{H}_{0,c}$. In the end, the only thing we need to adapt is the calculation of sequential pseudo observations: For $1 \le k \le i \le \ell \le n$ we set

$$\hat{U}_{\theta,k:\ell,i} = \begin{cases} \hat{U}_{k:\ell,i} & \text{if } \lfloor n\theta \rfloor \notin \{k, \dots \ell\}, \\ \hat{U}_{k:\lfloor n\theta \rfloor,i} & \text{if } \lfloor n\theta \rfloor \in \{k, \dots, \ell\} \text{ and } i \le \lfloor n\theta \rfloor, \\ \hat{U}_{\lfloor n\theta \rfloor + 1:\ell,i} & \text{if } \lfloor n\theta \rfloor \in \{k, \dots, \ell\} \text{ and } i > \lfloor n\theta \rfloor, \end{cases} \tag{3.17}$$

and similarly for $\hat{V}_{\theta,k:\ell,i}$. In words, we need to compute pseudo observations from stationary sub-subsamples. For instance, if we use these modified pseudo observations in the definition of the estimator for S, denoting with $\hat{S}_{k:\ell}^{\theta}$, we obtain the representation

$$\hat{S}_{k:\ell}^{\theta}(t) = \begin{cases} \frac{\lfloor n\theta \rfloor - k + 1}{\ell - k + 1} \hat{S}_{k:\lfloor n\theta \rfloor}(t) + \frac{\ell - \lfloor n\theta \rfloor}{\ell - k + 1} \hat{S}_{\lfloor n\theta \rfloor + 1:\ell}(t), & \text{if } k < \lfloor n\theta \rfloor < \ell, \\ \hat{S}_{k:\ell}(t), & \text{otherwise,} \end{cases} \tag{3.18}$$

which seems to be a natural choice under $\mathcal{H}^{\theta}_{1,m} \cap \mathcal{H}_{0,c}$. To complete the story, the madogram estimator $\hat{A}^{\theta}_{k:\ell}$, the statistics $\mathbb{D}^{\theta}_n$, $S^{\theta}_{n,A}$, $S^{\theta}_{n,A}(k^*)$ and their multiplier bootstrap replicates are defined in analogy to those from the previous section, provided the initial pseudo observations $(\hat{U}_{k:\ell,i}, \hat{V}_{k:\ell,i})'$ are replaced by $(\hat{U}_{\theta,k:\ell,i}, \hat{V}_{\theta,k:\ell,i})'$ wherever possible. Thus, from a practical perspective, once the above adapted pseudo observations are computed, the computer code for the simpler setting considered in the previous section can be fully reused.
The following proposition establishes the joint limit distribution of $\mathbb{D}^{\theta}_n$ and its multiplier bootstrap replicates under $\mathcal{H}_{0,c} \cap \mathcal{H}^{\theta}_{1,m}$.

Proposition 3.3
Assume that either $\mathcal{H}_0$ or $\mathcal{H}_{0,c} \cap \mathcal{H}^{\theta}_{1,m}$ holds and that A is continuously differentiable on $(0,1)$. Then, in $(\ell^{\infty}([0,1]^2), \|\cdot\|_{\infty})^{B+1}$ and for $n \to \infty$, we have that

$$\left(\mathbb{D}^{\theta}_n, \check{\mathbb{D}}^{\theta,(1)}_n, \ldots, \check{\mathbb{D}}^{\theta,(B)}_n\right) \xrightarrow{D} \left(\mathbb{D}_C, \mathbb{D}^{(1)}_C, \ldots, \mathbb{D}^{(B)}_C\right),$$

where $\mathbb{D}_C$ is defined in (3.10) *and $\mathbb{D}^{(1)}_C, \ldots, \mathbb{D}^{(B)}_C$ denote independent copies of $\mathbb{D}_C$.*

As previously, the null hypothesis is rejected at the significance level α if $\hat{p}^{\theta}_{n,B} < \alpha$, with $\hat{p}^{\theta}_{n,B}$ defined in analogy to (3.15).

3.3 Simulation study

Simulations were carried out in order to evaluate the finite-sample performance of the tests studied in Sections 3.2.1 and 3.2.2. For the sake of simplicity, we only focused on the test based on statistic $S_{n,A}$ in (3.11) and its adaption $S^{\theta}_{n,A}$ to $\mathcal{H}^{\theta}_{1,m}$ from Section 3.2.2. Results of the corresponding two-sample tests $S_{n,A}(k^*)$ and $S^{\theta}_{n,A}(k^*)$ are strongly related.

The finite-sample performance of the tests based on $S_{n,A}$ and $S^{\theta}_{n,A}$ was compared with that of three other tests for $\mathcal{H}_0$ in (3.1) designed to be particularly sensitive to $\mathcal{H}_{0,c}$ in (3.3):

- $S_{n,C}$: A test based on the empirical copula [Bücher et al., 2014, statistic $\check{S}_n$]
- $S_{n,\rho}$: A test based on Spearman's rho [Kojadinovic et al., 2015, statistic $\tilde{S}_{n,1}$]
- $S_{n,\tau}$: A test based on Kendall's tau [Dehling et al., 2014]

These procedures however do not assume the underlying dependence structures to be of the extreme-value type. The former is sensitive to all kind of changes in the underlying copula, while the latter two are only sensitive to changes in Spearman's rho and Kendall's tau, respectively.

Attention: All the tests mentioned above are procedures for testing $\mathcal{H}_0$ (resp. $\mathcal{H}^{\theta}_{1,m} \cap \mathcal{H}_{0,c}$) designed to be particularly sensitive to departures from $\mathcal{H}_{0,c}$. They should not be used to reject $\mathcal{H}_{0,c}$ unless we are sure that $\mathcal{H}_{0,m}$ (resp. $\mathcal{H}^{\theta}_{1,m}$) holds.

The rejection rates of the tests were estimated from samples drawn from bivariate distributions whose copulas are of the form

$$C_{\mathbf{a},\vartheta}(u,v) = u^{a_1} v^{a_2} C_\vartheta(u^{1-a_1}, v^{1-a_2}), \quad (u,v) \in [0,1]^2, \tag{3.19}$$

where C_ϑ is a symmetric extreme-value copula with parameter $\vartheta \in \mathbb{R}$ and $\mathbf{a} = (a_1, a_2) \in [0,1]^2$ is a parameter controlling the amount of asymmetry of $C_{\mathbf{a},\vartheta}$. The above copula construction principle is frequently referred to as *Khoudraji's device* [Khoudraji, 1995]. As long as C_ϑ is an extreme-value copula, so is its potentially asymmetric version $C_{\mathbf{a},\vartheta}$. Given that there is hardly any practical difference among the existing bivariate symmetric parametric families of extreme-value copulas [see Genest et al., 2011, for more evidence], C_ϑ in (3.19) was taken to be the Gumbel–Hougaard copula with parameter $\vartheta \in [1, \infty)$. As an illustrative example, the copula depicted in Figure 1.1 of the first chapter is exactly of this form.

All the tests considered in our numerical experiments were carried out at the 5% significance level using $B = 1000$ multiplier bootstrap replicates. The values 50, 100, 200 and 400 were considered for the sample size n. The measure μ involved in the definition of $S_{n,A}$ and $S^\theta_{n,A}$ was taken equal to $9^{-1} \sum_{i=1}^{9} \delta_{i/10}$; finer grids did not appear to lead to better behaved tests. The bandwidth h_n in (3.14) was set to $10^{-2}/\sqrt{n}$; additional simulations not reported for the sake of brevity indicated that different choices, in particular smaller values, did not lead to any improvements. With the illustrations of Section 3.4 in mind, the values 0.25 and 0.5 were considered for θ. For the study of empirical power and for fixed $s \in (0,1)$, we define the alternatives

$$\mathcal{H}^s_{1,c} : C_1 = \ldots = C_{\lfloor ns \rfloor} \neq C_{\lfloor ns \rfloor + 1} = \ldots = C_n. \tag{3.20}$$

The computations were carried out using the R statistical system [R Core Team, 2015], and the R packages `copula` [Hofert et al., 2015] and `npcp` [Kojadinovic, 2015].

Empirical levels of the tests based on $S_{n,A}$, $S_{n,C}$, $S_{n,\tau}$ and $S_{n,\rho}$ Columns 5-8 of Table 3.1 report the rejection rates of the four tests estimated from 4000 random samples generated under $\mathcal{H}_0$ from c.d.f. $C_{\mathbf{a},\vartheta}$ in (3.19) for various values of $\mathbf{a}$ and ϑ. Note that, since ranks are invariant with respect to strictly monotone increasing transformations, the choice of the marginal distributions is irrelevant for the results. Standard uniform ones, which means that we are drawing realizations directly from $C_{\mathbf{a},\vartheta}$, are just the most convenient for simulation. The empirical levels of the tests based on $S_{n,A}$, $S_{n,C}$ and $S_{n,\rho}$ are overall reasonably close to the 5% nominal level in all settings for which Kendall's tau of $C_{\mathbf{a},\vartheta}$ given in the fourth column is strictly smaller than 0.6. For tau $\geq$ 0.6, the three tests are overall slightly conservative. On the contrary, the test based on $S_{n,\tau}$ can be way too liberal, especially for $n \leq 100$.

Empirical power of the tests based on $S_{n,A}$, $S_{n,C}$, $S_{n,\rho}$ and $S_{n,\tau}$ under changes in the copula only The right plot of Figure 3.1 displays the rejection rates of the first three tests estimated from 2000 samples of size $n = 100$ generated under $\mathcal{H}_{0,m} \cap \mathcal{H}^s_{1,c}$, where we set $s = 0.5$ such that, for each sample, the first (resp. last) 50 observations were drawn from

n	$\mathbf{a}$	ϑ	τ	$S_{n,A}$	$S_{n,C}$	$S_{n,\rho}$	$S_{n,\tau}$	$S_{n,A}^{0.25}$	$S_{n,A}^{0.5}$
50	$(0,0)$	1	0	4.9	6.3	5.6	6.1	7.6	4.2
		1.25	.2	6.7	6.2	5.8	7.0	7.9	7.0
		1.67	.4	5.8	4.4	3.9	7.8	6.6	6.1
		2.5	.6	4.0	3.7	2.4	11.7	5.6	4.7
		5	.8	3.6	2.4	0.9	20.9	8.2	2.7
	$(0,.3)$	4	.56	4.5	4.7	3.1	7.8	5.2	5.5
100	$(0,0)$	1	0	5.5	5.1	5.8	4.9	7.7	5.4
		1.25	.2	6.3	5.4	6.2	5.2	7.4	6.9
		1.67	.4	6.2	4.3	4.4	5.4	6.2	6.6
		2.5	.6	5.4	3.0	2.9	7.1	6.0	5.5
		5	.8	2.0	2.2	1.0	15.4	4.0	2.6
	$(0,.3)$	4	.56	4.5	4.2	3.8	5.4	4.5	5.0
200	$(0,0)$	1	0	5.0	4.3	4.8	4.3	6.2	5.6
		1.25	.2	6.0	4.8	5.8	5.0	6.4	6.4
		1.67	.4	5.9	4.0	4.9	4.8	6.4	6.2
		2.5	.6	3.6	2.8	3.1	5.3	4.4	4.4
		5	.8	2.6	1.3	2.0	7.4	3.4	3.4
	$(0,.3)$	4	.56	4.8	4.0	4.3	5.1	5.2	5.2
400	$(0,0)$	1	0	5.5	5.1	5.5	5.0	6.0	5.8
		1.25	.2	5.2	4.7	5.1	5.0	5.5	5.4
		1.67	.4	4.9	4.4	5.0	4.6	5.1	5.2
		2.5	.6	5.3	3.9	4.6	5.5	5.5	5.5
		5	.8	3.8	1.9	2.7	5.8	4.3	4.2
	$(0,.3)$	4	.56	4.8	4.0	4.4	4.6	4.8	4.8

Table 3.1: Rejection rates of $\mathcal{H}_0$ in % estimated from 4000 random samples generated under $\mathcal{H}_0$ from c.d.f. $C_{\mathbf{a},\vartheta}$ in (3.19). The column τ gives the value of Kendall's tau of the copula $C_{\mathbf{a},\vartheta}$.

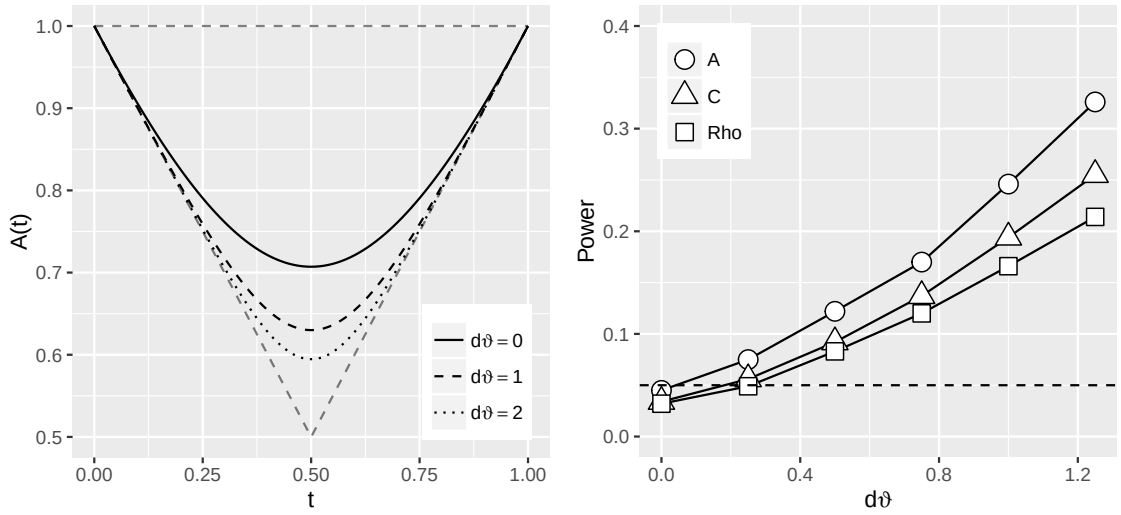

Figure 3.1: (Left) Pickands dependence function of the Gumbel–Hougaard copula with parameter $2 + d\vartheta$. (Right) Rejection rates of the tests based on $S_{n,A}$ (◯), $S_{n,C}$ (△) and $S_{n,\rho}$ (□) versus $d\vartheta$ estimated from 2000 bivariate samples of size $n = 100$ such that, for each sample, the first (resp. last) 50 observations were drawn from a Gumbel–Hougaard copula with parameter 2 (resp. $2 + d\vartheta$).

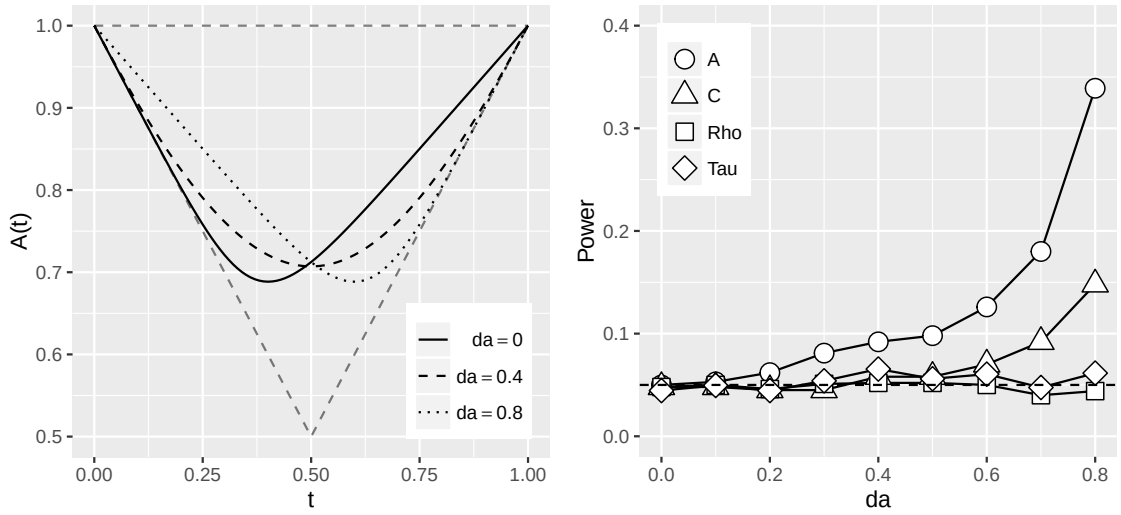

Figure 3.2: (Left) Pickands dependence function associated with the copula $C_{\mathbf{a},\vartheta}$ in (3.19) with $\mathbf{a} = (\max(0.4 - da, 0), \max(da - 0.4, 0))$, for $da \in \{0, 0.4, 0.8\}$, and ϑ set to keep Kendall's tau of $C_{\mathbf{a},\vartheta}$ equal to 0.5. (Right) Rejection rates of the tests based on $S_{n,A}$ (◯), $S_{n,C}$ (△), $S_{n,\rho}$ (□) and $S_{n,\tau}$ (◇) versus da estimated from 2000 bivariate samples of size $n = 200$ such that, for each sample, the first (resp. last) 100 observations were drawn the above mentioned copula with $da = 0$ (resp. $da \in \{0, 0.1, \ldots, 0.8\}$).

(3.19) with $(\mathbf{a}, \vartheta) = (0, 0, 2)$ (resp. $(0, 0, 2 + d\vartheta)$). We did not include the results for the test based on $S_{n,\tau}$ because it does not hold its level for the sample size under consideration. As expected, the test based on $S_{n,A}$ is more powerful than its two competitors in this simple setting.
To investigate the influence of asymmetry on the power of the four tests, as a second experiment, we considered again the copula $C_{\mathbf{a},\vartheta}$ in (3.19), but this time with parameter $\mathbf{a}$ defined as $(\max(0.4 - da, 0), \max(da - 0.4, 0))$, for $da \in \{0, 0.1, \ldots, 0.8\}$, and with parameter $\vartheta = \vartheta(\mathbf{a})$ set to keep Kendall's tau of $C_{\mathbf{a},\vartheta}$ equal to 0.5 for all considered values of $\mathbf{a}$. The corresponding graphs of the Pickands dependence functions for $da \in \{0, 0.4, 0.8\}$ are depicted in the left-hand side of Figure 3.2. The right plot displays the rejection rates of the tests based on $S_{n,A}$, $S_{n,C}$, $S_{n,\rho}$ and $S_{n,\tau}$ versus da estimated from 2000 samples of size $n = 200$ under $\mathcal{H}_{0,m} \cap \mathcal{H}^s_{1,c}$, where we set $s = 0.5$ such that, for each sample, the first (resp. last) 100 observations were drawn from the above mentioned copula with $da = 0$ (resp. $da \in \{0, 0.1, \ldots, 0.8\}$). Although the rejection rates are overall relatively low, the test based on $S_{n,A}$ is by far the best. The fact that the tests based on $S_{n,\tau}$ and $S_{n,\rho}$ have very little (if any) power against such alternatives is due to the fact that Kendall's τ (resp. Spearman's ρ) remains constant (resp. almost constant) in this experiment.

Empirical power of the tests based on $S_{n,A}$, $S_{n,C}$, $S_{n,\rho}$ and $S_{n,\tau}$ under an abrupt change in one margin only: Table 3.2 reports rejection rates of $\mathcal{H}_0$ estimated from 2000 bivariate samples of size n generated under $\mathcal{H}^\theta_{1,m} \cap \mathcal{H}_{0,c}$, where $\mathcal{H}_{0,c}$ and $\mathcal{H}^\theta_{1,m}$ are defined in (3.3) and (3.16), respectively, such that, for each sample, the first $\lfloor n\theta \rfloor$ (resp. last $n - \lfloor n\theta \rfloor$) observations were drawn from a c.d.f. whose copula is the Gumbel–Hougaard, whose first margin is GEV with parameters $\mu = 20$, $\sigma = 10$ and $\gamma = 0.25$ (resp. $\mu = 20 + d\mu$, $\sigma = 10$ and $\gamma = 0.25$), and whose second margin is uniform (the results are unaffected by the choice of the second margin since the test is rank-based).
All tests have little power against such alternatives when the shift $d\mu$ in the location parameter of the first margin is relatively small ($d\mu = 5$). This is a desirable property since the tests were designed to be sensitive to departures from $\mathcal{H}_{0,c}$. Higher rejection rates were obtained for $d\mu = 15$ and when the dependence is moderate or high, in particular if the (scaled) change-point in the first margin is non-central ($\theta = 0.25$). The latter results illustrate the fact that the procedures based on $S_{n,A}$, $S_{n,C}$, $S_{n,\rho}$ and $S_{n,\tau}$ are tests for $\mathcal{H}_0$ and that one should not use them to reject $\mathcal{H}_{0,c}$ unless $\mathcal{H}_{0,m}$ holds. Additional changes in the dispersion or scale parameter of the first margin might even increase the phenomenon.

Empirical levels of the test based on $S^\theta_{n,A}$: A consequence of Proposition 3.3 is that the test based on $S^\theta_{n,A}$ will hold its level asymptotically under one abrupt marginal change only (formally, this is a consequence of the results in Appendix F of Bücher and Kojadinovic, 2016). To evaluate the corresponding finite-sample behavior, we considered again the setting of Table 3.1. Indeed, since ranks are invariant with respect to monotone transformations, samples generated under $\mathcal{H}_0$ can equivalently be regarded as generated from $\mathcal{H}_{0,c} \cap \mathcal{H}^\theta_{1,m}$, provided θ is known. From the last two columns of Table 3.1, we see that the test based on $S^{0.5}_{n,A}$ holds its level equally well as the test based on $S_{n,A}$. The test based on $S^{0.25}_{n,A}$ is however slightly too liberal for $n = 50$, although the agreement of its empirical

			$\theta = 0.5$				$\theta = 0.25$			
$d\mu$	n	τ	$S_{n,A}$	$S_{n,C}$	$S_{n,\rho}$	$S_{n,\tau}$	$S_{n,A}$	$S_{n,C}$	$S_{n,\rho}$	$S_{n,\tau}$
5	50	0	5.5	6.2	6.0	5.7	5.5	6.3	6.2	5.6
		0.25	6.6	6.1	6.2	6.2	5.2	4.1	3.7	5.3
		0.5	4.3	3.3	2.5	8.1	4.4	2.8	2.7	8.3
		0.75	3.9	2.2	1.2	17.8	2.7	2.5	1.1	16.0
	100	0	5.1	5.1	5.4	4.4	5.0	5.9	6.0	3.9
		0.25	5.1	3.6	5.1	4.5	6.7	4.0	4.9	5.0
		0.5	5.5	4.1	4.3	4.9	4.7	2.9	3.3	5.2
		0.75	3.3	1.0	0.6	13.7	4.5	2.8	1.5	9.2
	200	0	5.6	4.8	5.5	4.1	5.6	5.8	5.4	3.8
		0.25	4.8	5.1	5.1	3.6	3.9	3.6	3.4	4.4
		0.5	4.6	2.5	2.9	3.6	5.2	3.4	4.0	3.7
		0.75	4.0	0.8	1.0	20.0	5.0	3.0	2.1	7.4
	400	0	4.6	4.7	4.5	3.2	5.2	4.2	4.5	3.0
		0.25	5.4	4.2	5.5	4.7	5.8	4.1	4.2	3.1
		0.5	5.5	4.3	4.5	5.2	5.5	5.5	5.9	4.6
		0.75	6.6	2.8	1.8	38.2	9.7	6.7	7.4	6.7
15	50	0	4.5	5.6	5.3	9.0	4.7	5.3	5.3	5.7
		0.25	7.6	5.7	5.0	8.1	6.8	6.5	6.8	6.1
		0.5	5.2	2.8	2.1	14.8	8.6	6.7	5.3	7.8
		0.75	18.3	1.5	0.3	63.6	20.0	15.9	4.6	13.0
	100	0	4.3	4.4	4.9	6.5	4.2	4.2	4.6	4.3
		0.25	4.7	3.4	3.9	9.0	5.9	4.3	4.9	4.8
		0.5	7.2	3.8	3.0	23.4	9.0	8.4	6.5	6.9
		0.75	40.6	10.0	1.8	91.9	42.0	36.9	23.0	18.6
	200	0	4.3	3.8	5.4	5.7	4.2	4.3	4.3	3.4
		0.25	6.4	5.5	5.2	10.3	7.7	5.5	6.5	4.8
		0.5	9.3	7.4	6.3	44.1	14.1	15.7	14.7	10.0
		0.75	75.2	56.5	36.8	99.8	79.2	79.3	71.5	43.9
	400	0	5.1	4.9	5.6	5.8	4.9	5.1	5.1	2.9
		0.25	6.4	6.5	6.6	13.9	8.1	8.6	8.5	4.5
		0.5	16.4	16.4	15.8	71.8	22.1	29.6	28.0	14.0
		0.75	98.6	98.0	94.7	100	98.2	99.2	98.3	87.4

Table 3.2: Rejection rates of $\mathcal{H}_0$ in % estimated from 2000 bivariate samples of size n generated under $\mathcal{H}_{1,m} \cap \mathcal{H}_{0,c}$, where $\mathcal{H}_{0,c}$ and $\mathcal{H}_{1,m}$ are defined in (3.3) and (3.16), respectively, such that, for each sample, the first $\lfloor n\theta \rfloor$ (resp. last $n - \lfloor n\theta \rfloor$) observations were drawn from a c.d.f. whose copula is the Gumbel–Hougaard, whose first margin is GEV with parameters $\mu = 20$, $\sigma = 10$ and $\gamma = 0.25$) (resp. $\mu = 20 + d\mu$, $\sigma = 10$ and $\gamma = 0.25$), and whose second margin is standard normal. The value of the parameter of the Gumbel–Hougaard copula is set through its one-to-one relationship with Kendall's tau τ.

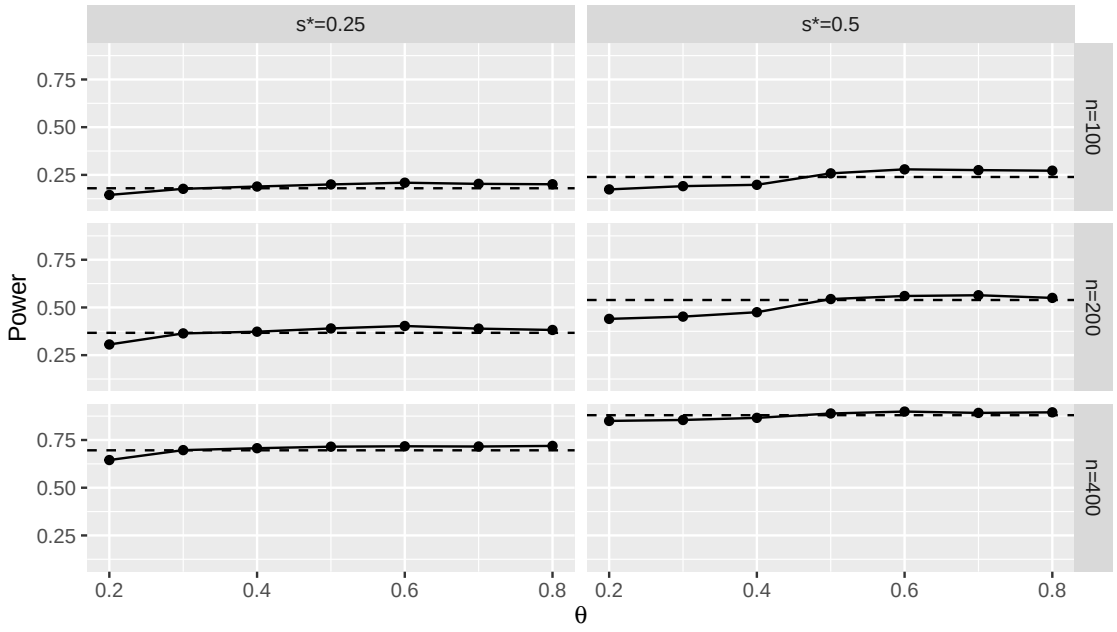

Figure 3.3: Rejection rates of the test based on $S^{\theta}_{n,A}$ ($\circ$) against $\theta \in \{0.2, 0.3, \ldots, 0.8\}$ estimated from 2000 bivariate samples of size $n \in \{100, 200, 400\}$, such that, for each sample, the first $\lfloor ns^\star \rfloor$ (resp. last $n - \lfloor ns^\star \rfloor$) observations are generated from a Gumbel–Hougaard copula with parameter 2 (resp. 3). The dashed line marks the corresponding estimated rejection rate of the test based on $S_{n,A}$.

levels with the 5% nominal level improves as n increases.

Empirical power of the test based on $S^{\theta}_{n,A}$: As a last experiment, we investigated the influence of the value θ on the power of the test based on $S^{\theta}_{n,A}$. Figure 3.3 displays the rejection rates of the test based on $S^{\theta}_{n,A}$ against $\theta \in \{0.2, 0.3, \ldots, 0.8\}$ estimated from 2000 bivariate samples of size $n \in \{100, 200, 400\}$ under $\mathcal{H}_{0,m} \cap \mathcal{H}^{s^*}_{1,c}$ such that, for each sample, the first $\lfloor ns^\star \rfloor$ (resp. last $n - \lfloor ns^\star \rfloor$) observations are generated from a Gumbel–Hougaard copula with parameter 2 (resp. 3). The values 0.25 and 0.5 were considered for $s^\star$. As one can see, the rejections rates are not too much affected by the value of θ. In addition, the power of the test based on $S^{\theta}_{n,A}$ remains overall reasonably close to that of the test based on $S_{n,A}$. From a practical perspective, the latter result suggests that, under $\mathcal{H}_{0,m}$ in (3.2), the somehow "non optimal" use of the test based on $S^{\theta}_{n,A}$ instead that based on $S_{n,A}$ does not incur a large power loss, if any. As a consequence, if one hesitates about which of $\mathcal{H}_{0,m}$ in (3.2) or $\mathcal{H}_{1,m}$ in (3.16) holds, it seems safer to use the test based on $S^{\theta}_{n,A}$ as, should $\mathcal{H}_{1,m}$ be actually true, the latter test is more likely to hold its level by construction, and should $\mathcal{H}_{0,m}$ be true, the power loss, if any, should not be too large.

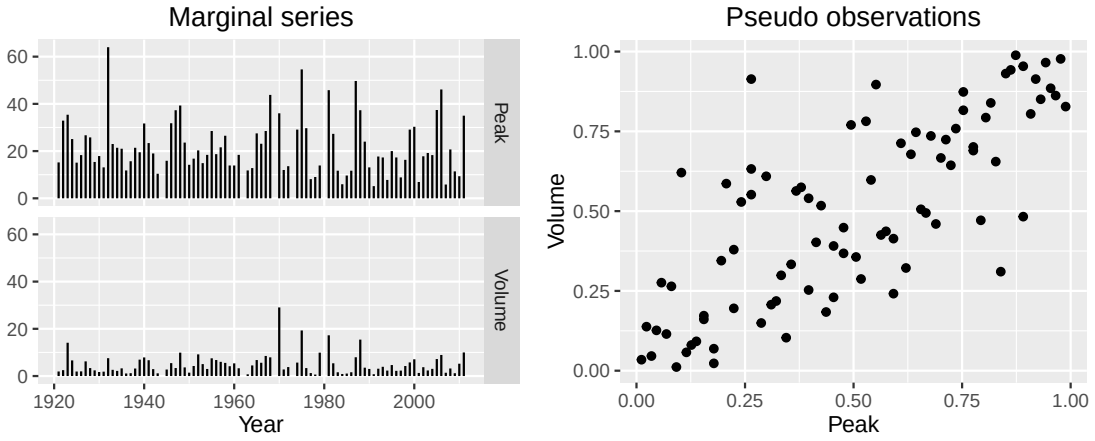

Figure 3.4: Left: Annual maximal peak flows and volumes of discharges measured in Streckewalde, Germany. Right: Corresponding pseudo-observations computed using (1.14).

3.4 Application: Do dams change dependence?

We consider two related applications from flood frequency analysis. The first time series consists of $n = 86$ bivariate annual maxima measured between 1921 and 2011 (five years of data are missing) at a station located on the river Preßnitz in Streckewalde, Germany. The variables of interest are Q, the annual maximal peak flow in m^3/s, and V, the annual maximal volume of discharge in $10^6\, m^3$. The observations are displayed in Figure 3.4. The joint distribution of Q and V is of interest to hydrologists as it can be used to assess the risk of catastrophic flood levels. For a recent case study we refer to Mitková and Halmová [2014].

Because we are dealing with bivariate block maxima, it is natural to assume that the data arise from one or more bivariate extreme-value distributions. The aim of our analysis is to test for possible changes in the dependence between Q and V that might have occurred during the long period of observation. An additional element to be taken into account here is that a dam was built on the river Preßnitz in 1973 (which corresponds to the 48th observation) a few kilometers upstream from the measurement station. We make the hypothesis that, if there are changes in the two components series, then, they are unique and they occurred simultaneously after observation 48 due to the construction of the dam. In other words, we assume that either $\mathcal{H}_{0,m}$ in (3.2) or $\mathcal{H}^{\theta}_{1,m}$ in (3.16) with $\theta = 48/86$ holds. In the former case, it is natural to use the test for change-point detection based on $S_{n,A}$ in (3.11), while in the latter case, the extension from Section 3.2.2 based on $S^{\theta}_{n,A}$ with $\theta = 48/86$ should be preferred. As mentioned in the previous section (see Figure 3.3 and the related discussion), using the test based on $S^{\theta}_{n,A}$ for some value of θ when $\mathcal{H}_{0,m}$ in (3.2) actually holds does not seem to result in a strong power loss, if any. For that reason, we carried out the test based on $S^{\theta}_{n,A}$ with $\theta = 48/86$. The resulting approximate p-value

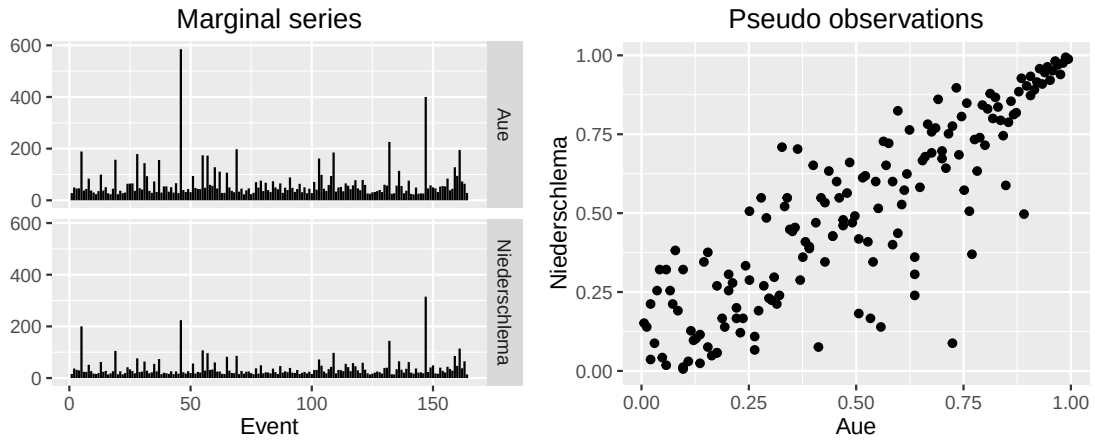

Figure 3.5: Left: Peak flows in m^3/s of 164 summer flood events simultaneously measured at gauges in Aue and Niederschlema, Germany. Right: Corresponding pseudo-observations computed using (1.14).

of 0.068, obtained from $B = 10000$ multiplier bootstrap replicates, indicates that there is some weak evidence of change in the dependence between Q and V. Interestingly enough, the maximum over k within the definition of the test statistic was not obtained for observation 48 but for observation 32 corresponding to year 1953.

The second data set consists of peak flows in m^3/s simultaneously measured at two neighboring stations for $n = 164$ physically independent summer flood events. The two stations are located in Germany, in Aue and Niederschlema, respectively, and the corresponding measurements will thus be denoted by Q_A and Q_N, respectively. The observations, chronologically ordered, span the period 1929-2011 and are displayed in Figure 3.5.

An event was classified as a flood, if each peak flow exceeded the smallest annual maximal peak flow measured between 1929 and 2011 in Aue and Niederschlema, respectively. The period of each flood event was identified by hand and only the largest value (peak flow) was included in the data set. Hence, by construction, the observations are formed subject to a block maximal procedure, with possibly slightly differing block sizes for each of the flood events. It therefore seems sensible to assume that the data-generating distribution(s) are extreme-value distributions.

There were two reasons why only summer events were included in the analysis. First, typical winter floods are produced from melting snow, whereas summer floods are due to short but heavy rainfalls. These very different physical mechanisms lead to different peak flow distributions. Second, very high peak flows, which are of particular interest, almost exclusively occur during the summer time in that region. We refer to Figures 2.4 for more evidence.

The aim of our analysis is to assess whether the dependence between Q_A and Q_N changed during the long observation period. As for the previous illustration, it might be impor-

tant to take into account the fact that dams where constructed on the river Mulde and one of its tributary upstream of the two gauges Aue and Niederschlema. A first dam, called Schönheiderhammer, was put in service in 1980 (which corresponds to observation 108) and a second dam, named Eibenstock, was put into service in 1982. As previously, we make the hypothesis that, if there are changes in the two components series, then, they are unique and they occurred simultaneously after observation 108 due to the construction of the dams (this is in fact a slight simplification, 1980 $\approx$ 1982, which does not have a big influence on the results). Following the same reasoning as for the first illustration, we apply the test based on $S_{n,A}^{\theta}$ with $\theta = 108/164$ and obtain an approximate p-value of 0.195 based on $B = 10000$ multiplier bootstrap replicates. Hence, there is no evidence for a change in the dependence between Q_A and Q_N.

Summing up, we have not found evidence for a change-point in the dependence due to dams in the present applications and in many more that are not reported here. Intuitively, it seems to be obvious that dams have no influence on the copula, since, from a simplified view, it is reasonable to expect that a dam changes margins only in a monotonic way, e.g., a location shift $x \mapsto x - \mu$. It can be easily verified that copulas are invariant to such strictly increasing transformations of marginal variables.

3.5 Proofs

3.5.1 Proofs of Propositions 3.1 and 3.2

For notational simplicity, we set

$$\lambda_n(r,s) = \frac{\lfloor ns \rfloor - \lfloor nr \rfloor}{n} \text{ for } (r,s) \in \Delta = \{(r,s) \in [0,1]^2 : \ r \leq s\}$$

and we define the process $\mathbb{A}_n$ on $\Delta \times [0,1]$ by

$$\mathbb{A}_n(r,s,t) = \sqrt{n}\lambda_n(r,s)\Big(\hat{A}_{\lfloor nr \rfloor + 1:\lfloor ns \rfloor} - A(t)\Big),$$

where, for instance, $\mathbb{A}_n(0,1,t) = \sqrt{n}\big(\hat{A}_n(t) - A(t)\big)$ for $t \in [0,1]$ and

$$\mathbb{D}_n(s,t) = \lambda_n(s,1)\mathbb{A}_n(0,s,t) - \lambda_n(0,s)\mathbb{A}_n(s,1,t), \ \ (s,t) \in [0,1]^2. \tag{3.21}$$

Theorem 3.4
Under the conditions of Proposition 3.1 and for $n \to \infty$, we have that $\mathbb{A}_n \xrightarrow{D} \mathbb{A}_C$ in $\ell^\infty(\Delta \times [0,1])$ equipped with the uniform metric, where

$$\mathbb{A}_C(r,s,t) = \big\{1 + A(t)\big\}^2 \cdot \int_0^1 \mathbb{C}_C\left(r, s, u^{1-t}, u^t\right) \, du$$

and where $\mathbb{C}_C$ is defined in Theorem A.8 with $d = 2$.

Proof. Just like in the first lines of the proof of Proposition 1.8 given in Appendix A.2, we use that

$$\sqrt{n}\lambda_n(r,s)\Big(\hat{S}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t)-S(t)\Big) = -\int_0^1 \mathbb{C}_n(r,s,u^{1-t},u^t)\,du \xrightarrow{D} \int_0^1 \mathbb{C}_C(r,s,u^{1-t},u^t)\,du,$$

where weak convergence holds in $\ell^\infty(\Delta\times[0,1])$ and where $\mathbb{C}_n$ and $\mathbb{C}_C$ are defined in Theorem A.8 with $d=2$. $C(u^{1-t},u^t)=u^{A(t)}$ and some simple arithmetic gives us

$$\mathbb{A}_n(r,s,t) = \frac{1+A(t)}{\int_0^1 C_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(u^{1-t},u^t)\,du}\sqrt{n}\lambda_n(r,s)\Big(\hat{S}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t)-S(t)\Big). \tag{3.22}$$

Using the fact that $\mathbb{C}_n$ is asymptotically uniformly equicontinuous in probability by Theorem A.8, it remains to show that, for any $\delta\in(0,1)$,

$$\begin{aligned}&\sup_{(r,s)\in\Delta,\ s-r\geq\delta,\ t\in[0,1]}\left|\int_0^1 C_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(u^{1-t},u^t)\,du-\frac{1}{1+A(t)}\right|\\ =&\frac{1}{\sqrt{n}}\sup_{(r,s)\in\Delta,\ s-r\geq\delta,\ t\in[0,1]}\frac{1}{\lambda_n(r,s)}\left|\int_0^1 \mathbb{C}_n(r,s,u^{1-t},u^t)\,du\right|\xrightarrow{\mathbb{P}}0.\end{aligned}$$

The latter is a consequence of $\lambda_n(r,s)\geq\delta/2$, $s-r\geq\delta$, n sufficiently large, and $\mathbb{C}_n = O_{\mathbb{P}}(1)$ due to Theorem A.8. Applying the functional version of Slutsky's lemma [van der Vaart and Wellner, 1996, Example 1.4.7] completes the proof of Theorem 3.4. □

Proof of Proposition 3.1. The assertion follows from representation (3.21), the fact that $\lambda_n\to\lambda$ with $\lambda(r,s)=s-r$ for $(s,r)\in\Delta$ and from the continuous mapping theorem. □

Recall the definition of the multipliers $(\xi_i^{(b)})$, $i=1,\dots,n$ and $b=1,\dots,B$. For the proof of the validity of the resampling scheme, we set

$$\check{\mathbb{A}}_n^{(b)}(r,s,t) = \big\{1+\hat{A}_n(t)\big\}^2\cdot\int_0^1 \check{\mathbb{C}}_n^{(b)}\left(r,s,u^{1-t},u^t\right)\,du,$$

where $\check{\mathbb{C}}_n^{(b)}(r,s,u,v)$ is defined in Theorem A.9 with $d=2$ but with partial derivatives $\dot{C}_j$ estimated by $\dot{C}_{j,\lfloor nr\rfloor+1:\lfloor ns\rfloor}$ according to

$$\begin{aligned}\dot{C}_1(u^{1-t},u^t) &= \big\{A(t)-tA'(t)\big\}\cdot u^{A(t)}\\ \dot{C}_2(u^{1-t},u^t) &= \big\{A(t)+(1-t)A'(t)\big\}\cdot u^{A(t)-t}\end{aligned}$$

and with A and A' estimated by $\hat{A}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}$ and $\hat{A}'_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}$ from (3.14), respectively.

Theorem 3.5
Under the conditions of Proposition 3.2, we have that

$$\left(\mathbb{A}_n,\check{\mathbb{A}}_n^{(1)},\dots,\check{\mathbb{A}}_n^{(B)}\right)\xrightarrow{D}\left(\mathbb{A}_C,\mathbb{A}_C^{(1)},\dots,\mathbb{A}_C^{(B)}\right) \tag{3.23}$$

in $\{\ell^\infty(\Delta\times[0,1])\}^{B+1}$, *where* $\mathbb{A}_C^{(1)},\dots,\mathbb{A}_C^{(B)}$ *denote independent copies of* $\mathbb{A}_C$.

Proof. According to Theorem A.9, if we prove that, for any $\delta \in (0,1)$ (see also the proof of Proposition 4.3 in Bücher et al. [2014]),

$$\sup_{(r,s)\in\Delta, s-r\geq\delta, (t,u)\in[0,1]^2} \left|\dot{C}_{j,\lfloor nr\rfloor+1:\lfloor ns\rfloor}(u^{1-t},u^t) - \dot{C}_j(u^{1-t},u^t)\right| \xrightarrow{\mathbb{P}} 0,\ j=1,2, \tag{3.24}$$

we get, for $\mathbb{I}_n(r,s,t) = \int_0^1 \mathbb{C}_n(r,s,u^{1-t},u^t)du$ and $\check{\mathbb{I}}_n^{(b)}(r,s,t) = \int_0^1 \check{\mathbb{C}}_n^{(b)}(r,s,u^{1-t},u^t)du$ and from the continuous mapping theorem,

$$\left(\mathbb{I}_n, \check{\mathbb{I}}_n^{(1)}, \ldots, \check{\mathbb{I}}_n^{(B)}\right) \xrightarrow{D} \left(\mathbb{I}_C, \mathbb{I}_C^{(1)}, \ldots, \mathbb{I}_C^{(B)}\right) \tag{3.25}$$

in $\{\ell^\infty(\Delta\times[0,1])\}^{B+1}$, where $\mathbb{I}_C^{(1)},\ldots,\mathbb{I}_C^{(B)}$ denote independent copies of $\mathbb{I}_C$ defined by $\mathbb{I}_C(r,s,t) = \int_0^1 \mathbb{C}_C(r,s,u^{1-t},u^t)du$ and with $\mathbb{C}_C$ defined in Theorem A.8 with $d=2$. As a byproduct of Proposition 1.8, we also have that $\hat{A}_n \xrightarrow{\mathbb{P}} A$ in $\ell^\infty([0,1])$. From a functional version of Slutsky's lemma [van der Vaart and Wellner, 1996, Example 1.4.7], we obtain (3.25) jointly with $\hat{A}_n \xrightarrow{D} A$, which gives us (3.23) by another application of the continuous mapping theorem.
It remains to show (3.24) in order to complete the proof. For that, we first show that

$$\begin{aligned}\sup_{(r,s)\in\Delta, s-r\geq\delta, t\in[0,1]} \left|\hat{A}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t) - A(t)\right| &= \sup_{(r,s)\in\Delta, s-r\geq\delta, t\in[0,1]} \left|\frac{\mathbb{A}_n(r,s,t)}{\sqrt{n}\lambda_n(r,s)}\right| \\ \leq \sup_{(r,s)\in\Delta, s-r\geq\delta, t\in[0,1]} \delta^{-1}\left|n^{-1/2}\mathbb{A}_n(r,s,t)\right| &\xrightarrow{\mathbb{P}} 0,\end{aligned}$$

which follows from Theorem 3.4. For the same result with the derivate, let us first consider $t \in [h_n, 1-h_n]$. Then, we have that

$$\begin{aligned}\sup_{(r,s)\in\Delta, s-r\geq\delta, t\in[h_n,1-h_n]} \left|A'_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t) - A'(t)\right| \leq \sup_{t\in[h_n,1-h_n]} &\left|\frac{A(t+h_n)-A(t-h_n)}{2h_n} - A'(t)\right| \\ + \sup_{(r,s)\in\Delta, s-r\geq\delta, t\in[h_n,1-h_n]} &\left|\frac{\mathbb{A}_n(r,s,t+h_n)-\mathbb{A}_n(r,s,t-h_n)}{2h_n\sqrt{n}\lambda_n(r,s)}\right|.\end{aligned} \tag{3.26}$$

By the mean value theorem, we can replace the difference quotient in the first summand by $A'(r_t)$, where $|r_t - t| \leq h_n \to 0$ uniformly in $t \in (0,1)$. Since, in addition, A' is uniformly continuous on $(0,1)$ (A' is monotone increasing, bounded and continuous), we have that the first summand converges to zero.
The second summand converges in probability to zero, because $\inf_n h_n\sqrt{n} > 0$, $\lambda_n(r,s) \geq \delta/2$ for $s-r \geq \delta$ and n sufficiently large and because of the asymptotic equicontinuity in probability of $\mathbb{A}_n$ implied by Theorem 3.4. For $t \in [0,h_n)$ (and similarly for $t \in (1-h_n,1]$), we have

$$\left|A'_{\lfloor nr\rfloor+1:\lfloor ns\rfloor} - A'(t)\right| \leq \left|\frac{A(2h_n)-1}{2h_n} - A'(t)\right| + \left|\frac{\mathbb{A}_n(r,s,2h_n)}{2h_n\sqrt{n}\lambda_n(r,s)}\right|, \tag{3.27}$$

which again converges uniformly in probability to zero. □

Proof of Proposition 3.2. The only part left to show is that, for $\check{\mathbb{D}}_n^{(b)}$ from Section 3.2.1,

$$\check{\mathbb{D}}_n^{(b)}(s,t) = \lambda_n(s,1)\check{\mathbb{A}}_n^{(b)}(0,s,t) - \lambda_n(0,s)\check{\mathbb{A}}_n^{(b)}(s,1,t), \;\; (s,t) \in [0,1]^2.$$

The assertion then follows from (3.23) and the continuous mapping theorem. From the definitions of $\hat{a}_n, \hat{b}_n, \hat{c}_n$ and $\hat{d}_n$ given in Section 3.2.1, we have that

$$\begin{aligned}\int_0^1 \check{\mathbb{C}}_n^{(b)}(r,s,y^{1-t},y^t)\,dy = &\int_0^1 \check{\mathbb{B}}_n^{(b)}(r,s,y^{1-t},y^t)\,dy \\ &- \hat{a}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t)\int_0^1 y^{\hat{b}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t)-1}\check{\mathbb{B}}_n^{(b)}(r,s,y^{1-t},1)\,dy \\ &- \hat{c}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t)\int_0^1 y^{\hat{d}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t)-1}\check{\mathbb{B}}_n^{(b)}(r,s,1,y^t)\,dy \\ =&(I_1 - I_2 - I_3)(r,s,t),\end{aligned}$$

where I_1, I_2 and I_3 are defined in an obvious manner. Note that we have the identity

$$\begin{aligned}\check{\mathbb{B}}_n^{(b)}\left(r,s,u^{1-t},u^t\right) = \frac{1}{\sqrt{n}}\sum_{i=\lfloor nr\rfloor+1}^{\lfloor ns\rfloor}\xi_i^{(b)}\Big[&\mathbb{1}\left\{\hat{m}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor,i}(t)\le u\right\} \\ &- \frac{1}{\lfloor ns\rfloor-\lfloor nr\rfloor}\sum_{j=\lfloor nr\rfloor+1}^{\lfloor ns\rfloor}\mathbb{1}\left\{\hat{m}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor,j}(t)\le u\right\}\Big].\end{aligned}$$

Consequently, from $\int_0^1 \mathbb{1}(m\le u)du = (1-m)$ for $0\le m\le 1$, we get

$$I_1(r,s,t) = \frac{1}{\sqrt{n}}\sum_{i=\lfloor nr\rfloor+1}^{\lfloor ns\rfloor}\xi_i^{(b)}\left\{\bar{m}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t) - \hat{m}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor,i}(t)\right\}$$

and, from $\int_0^1 u^{b-1}\mathbb{1}(m\le u)\,dy = (1-m^b)/b$, we also have that

$$I_2(r,s,t) = \frac{\hat{a}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t)}{\hat{b}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t)}\frac{1}{\sqrt{n}}\sum_{i=\lfloor nr\rfloor+1}^{\lfloor ns\rfloor}\xi_i^{(b)}\left\{\bar{u}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t) - \hat{u}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor,i}(t)\right\}$$

and, similarly,

$$I_3(r,s,t) = \frac{\hat{c}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t)}{\hat{d}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t)}\frac{1}{\sqrt{n}}\sum_{i=\lfloor nr\rfloor+1}^{\lfloor ns\rfloor}\xi_i^{(b)}\left\{\bar{v}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor}(t) - \hat{v}_{\lfloor nr\rfloor+1:\lfloor ns\rfloor,i}(t)\right\}.$$

□

3.5.2 Proof of Proposition 3.3

Let $\theta \in (0,1)$ be fixed. We define the process $\mathbb{A}_n^\theta$ on $\Delta \times [0,1]$ by

$$\mathbb{A}_n^\theta(r,s,t) = \sqrt{n}\lambda_n(r,s)\Big(\hat{A}^\theta_{\lfloor nr\rfloor+1:\lfloor ns\rfloor} - A(t)\Big),$$

where $\hat{A}^\theta_{k:\ell}$ is defined in Section 3.2.2. Note that, for $\mathbb{D}_n^\theta$ defined in Section 3.2.2,

$$\mathbb{D}_n^\theta(s,t) = \lambda_n(s,1)\mathbb{A}_n^\theta(0,s,t) - \lambda_n(0,s)\mathbb{A}_n^\theta(s,1,t), \;\; (s,t) \in [0,1]^2$$

and, by setting

$$\check{\mathbb{A}}_n^{\theta,(b)}(r,s,t) = \big\{1 + \hat{A}_n^\theta(t)\big\}^2 \cdot \int_0^1 \check{\mathbb{C}}_n^{\theta,(b)}\left(r,s,u^{1-t},u^t\right)\, du,$$

we also get that

$$\check{\mathbb{D}}_n^{\theta,(b)}(s,t) = \lambda_n(s,1)\check{\mathbb{A}}_n^{\theta,(b)}(0,s,t) - \lambda_n(0,s)\check{\mathbb{A}}_n^{\theta,(b)}(s,1,t), \;\; (s,t) \in [0,1]^2.$$

Proof of Proposition 3.3. Let us first walk through a sketch of the proof, followed by the technical details at the end of the section.
We will show that

$$\sup_{s\in[0,\theta],t\in[0,1]} \Big|\mathbb{A}_n^\theta(s,1,t) - \{\mathbb{A}_n(s,\theta,t) + \mathbb{A}_n(\theta,1,t)\}\Big| = o_{\mathbb{P}}(1) \tag{3.28}$$

and

$$\sup_{s\in[\theta,1],t\in[0,1]} \Big|\mathbb{A}_n^\theta(0,s,t) - \{\mathbb{A}_n(0,\theta,t) + \mathbb{A}_n(\theta,s,t)\}\Big| = o_{\mathbb{P}}(1). \tag{3.29}$$

This gives us

$$\begin{aligned}\mathbb{D}_n^\theta(s,t) = {} & \lambda_n(s,1)\left\{\mathbb{A}_n(0,s\wedge\theta,t) + \mathbb{A}_n(s\wedge\theta,s,t)\right\} \\ & - \lambda_n(0,s)\left\{\mathbb{A}_n(s,s\vee\theta,t) + \mathbb{A}_n(s\vee\theta,1,t)\right\} + o_{\mathbb{P}}(1),\end{aligned} \tag{3.30}$$

uniformly in $(s,t) \in [0,1]^2$. Similarly, for the bootstrap versions, we can show that

$$\begin{aligned}\mathbb{D}_n^{\theta,(b)}(s,t) = {} & \lambda_n(s,1)\left\{\mathbb{A}_n^{(b)}(0,s\wedge\theta,t) + \mathbb{A}_n^{(b)}(s\wedge\theta,s,t)\right\} \\ & - \lambda_n(0,s)\left\{\mathbb{A}_n^{(b)}(s,s\vee\theta,t) + \mathbb{A}_n^{(b)}(s\vee\theta,1,t)\right\} + o_{\mathbb{P}}(1)\end{aligned} \tag{3.31}$$

uniformly in $(s,t) \in [0,1]^2$. Finally, note that, under $\mathcal{H}^\theta_{1,m} \cap \mathcal{H}_{0,c}$, the right-hand sides of (3.30) and (3.31), $b = 1,\ldots,B$, behaves like under $\mathcal{H}_0$ and thus, converge jointly towards the desired limit by Theorem 3.4.

Let us now turn to the technical details: (3.28), (3.29) and their bootstrap versions. We will show (3.28). The proof of (3.29) is very similar. Similar to the lines starting from (3.22), we obtain

$$\begin{aligned}\mathbb{A}_n^{\theta}(s,1,t) &= -\frac{1+A(t)}{\int_0^1 C_{\lfloor ns\rfloor+1:n}(u^{1-t},u^t)du}\int_0^1 \mathbb{C}_n(s,\theta,u^{1-t},u^t)+\mathbb{C}_n(\theta,1,u^{1-t},u^t)\,du\\ &= -\left\{1+A(t)\right\}^2\int_0^1 \mathbb{C}_n(s,\theta,u^{1-t},u^t)+\mathbb{C}_n(\theta,1,u^{1-t},u^t)\,du+o_{\mathbb{P}}(1)\end{aligned}$$

uniformly in $(s,t)\in[0,1]^2$. Using representation (3.22) for both, $\mathbb{A}_n(s,\theta,t)$ and $\mathbb{A}_n(\theta,1,t)$, proves (3.28).
The bootstrap version of (3.28), that is,

$$\sup_{s\in[0,\theta],t\in[0,1]}\left|\check{\mathbb{A}}_n^{\theta,(b)}(s,1,t)-\left\{\check{\mathbb{A}}_n^{(b)}(s,\theta,t)+\check{\mathbb{A}}_n^{(b)}(\theta,1,t)\right\}\right|=o_{\mathbb{P}}(1), \tag{3.32}$$

follows, if, for $d\check{\mathbb{B}}_n^{\theta,(b)}(r,s,u,v)=\check{\mathbb{B}}_n^{\theta,(b)}(s,u,v)-\check{\mathbb{B}}_n^{\theta,(b)}(r,u,v)$ and similarly $d\check{\mathbb{B}}_n^{(b)}$, we show that

$$\sup_{s\in[0,\theta],u,v\in[0,1]}\left|d\check{\mathbb{B}}_n^{\theta,(b)}(s,1,u,v)-\left\{d\check{\mathbb{B}}_n^{(b)}(s,\theta,u,v)+d\check{\mathbb{B}}_n^{(b)}(\theta,1,u,v)\right\}\right|=o_{\mathbb{P}}(1), \tag{3.33}$$

in combination with

$$\sup_{s\in[0,\theta],t\in[0,1]}\left|\hat{A}^{\theta}_{\lfloor ns\rfloor+1:n}(t)-A(t)\right|=o_{\mathbb{P}}(1) \tag{3.34}$$

and

$$\sup_{s\in[0,\theta],t\in[0,1]}\left|\hat{A}'^{\theta}_{\lfloor ns\rfloor+1:n}(t)-A'(t)\right|=o_{\mathbb{P}}(1). \tag{3.35}$$

Starting with (3.33), the left-hand side is dominated by

$$\begin{aligned}\sup_{(s,u,v)\in[0,\theta]\times[0,1]^2}&\left|\frac{\lfloor n\theta\rfloor-\lfloor ns\rfloor}{n-\lfloor ns\rfloor}C_{\lfloor ns\rfloor+1:\lfloor n\theta\rfloor}(u,v)\right.\\ &\left.+\frac{n-\lfloor n\theta\rfloor}{n-\lfloor ns\rfloor}C_{\lfloor n\theta\rfloor+1:n}(u,v)-C(u,v)\right|\times\sup_{s\in[0,\theta]}\left|\frac{1}{\sqrt{n}}\sum_{i=\lfloor ns\rfloor+1}^{n}\xi_i^{(b)}\right|\\ &\le\sup_{(s,u,v)\in[0,\theta]\times[0,1]^2}\frac{|\mathbb{C}_n(s,\theta,u,v)+\mathbb{C}_n(\theta,1,u,v)|}{\sqrt{n}\lambda_n(s,1)}\times O_{\mathbb{P}}(1)=o_{\mathbb{P}}(1)\end{aligned}$$

and because, under $\mathcal{H}_{1,m}^{\theta}\cap\mathcal{H}_{0,c}$, the processes $(s,u,v)\mapsto\mathbb{C}_n(s,\theta,u,v)$ on $[0,\theta]\times[0,1]^2$, and $(u,v)\mapsto\mathbb{C}_n(\theta,1,u,v)$ on $[0,1]^2$ behave like those under $\mathcal{H}_0$. The fact that

$$\sup_{s\in[0,\theta]}\left|\frac{1}{\sqrt{n}}\sum_{i=\lfloor ns\rfloor+1}^{n}\zeta_i^{(b)}\right|=O_{\mathbb{P}}(1)$$

is due to Donsker's theorem on weak convergence of partial sum processes. Next, in order to show (3.34), observe that

$$\left|\hat{A}^{\theta}_{\lfloor ns\rfloor+1:n}(t) - A(t)\right| = \frac{1}{\sqrt{n}\lambda_n(s,1)}\left|\mathbb{A}^{\theta}_n(s,1,t)\right|,$$

where $\sup_{s\in[0,\theta],t\in[0,1]}\left|\mathbb{A}^{\theta}_n(s,1,t)\right| = O_{\mathbb{P}}(1)$ due to (3.28) and Theorem 3.4. Finally, for (3.35), we can use decompositions in analogy to (3.26) and (3.27), where the first summand is the same and where the second summand vanishes with the same argument as before. □

Chapter 4

Conditional heavy-tail behavior

This chapter deals with the right-tail behavior of a response distribution F_Y conditional on a regressor vector $\mathbf{X} = \mathbf{x}$ restricted to the heavy-tailed case of Pareto-type conditional distributions $F_Y(y|\ \mathbf{x}) = P(Y \leq y|\ \mathbf{X} = \mathbf{x})$, with heaviness of the right tail characterized by the conditional extreme value index $\gamma(\mathbf{x}) > 0$. We particularly focus on testing the hypothesis $\mathcal{H}_{0,tail}:\ \gamma(\mathbf{x}) = \gamma_0$ of constant tail behavior for some $\gamma_0 > 0$ and all $\mathbf{x}$.
When considering $\mathbf{x}$ as a time index, the term *trend analysis* is commonly used. In the recent past several such trend analyses in extreme value data have been published, mostly focusing on time-varying modeling of location and scale parameters of the response distribution. In many such environmental studies a simple test against trend based on Kendall's tau statistic is applied. This test is powerful when the center of the conditional distribution $F_Y(y|\mathbf{x})$ changes monotonically in $\mathbf{x}$, for instance, in a simple location model $\mu(\mathbf{x}) = \mu_0 + x \cdot \mu_1$, $\mathbf{x} = (1, x)'$, but the test is rather insensitive against monotonic tail behavior, say, $\gamma(\mathbf{x}) = \gamma_0 + x \cdot \gamma_1$. This has to be considered, since for many environmental applications the main interest is on the tail rather than the center of a distribution. Our work is motivated by this problem and it is our goal to demonstrate the opportunities and the limits of detecting and estimating non-constant conditional heavy-tail behavior with regard to applications from hydrology.
We present and compare four different procedures by simulations and illustrate our findings on real data from hydrology: Weekly maxima of hourly precipitation from France and monthly maximal river flows from Germany.

4.1 Introduction

In recent years considerable attention has been devoted to the analysis of abrupt change-points and smooth changes in the distribution of environmental variables Y such as amounts of precipitation, sea storm heights and river flows. While change-points are motivated by human intervention, for instance, the relocation of a measurement station or the construction of a river dam, the analysis of smooth changes has gained attention due to the climate change debate. In the latter context the term *trend* is used, which

is usually associated with a smooth monotonic change over time. More generally, the conditional distribution of Y given some regressor variables $\mathbf{X} = \mathbf{x}$, $\mathbf{x} \in \mathcal{X}$, may be of interest. Then the interest might be in the change of the conditional distribution over the regressor space $\mathcal{X}$.

For many environmental applications the main interest is in the frequency of hazardous events, e.g., extreme precipitations and floods. Accordingly, there is a number of articles introducing methodology for change-points [Jarušková and Rencová, 2008; Kim and Lee, 2009; Dierckx and Teugels, 2010; Dupuis et al., 2015; Bücher et al., 2015; Kojadinovic and Naveau, 2015] and regression/trend analysis [Chavez-Demoulin and Davison, 2005; Wang and Tsai, 2009; Gardes and Girard, 2010; Dierckx, 2011; Wang et al., 2012; Wang and Li, 2013; Einmahl et al., 2016; de Haan et al., 2015] of extremes, just to name a few recent contributions. For a case study and an overview of many flood trend analyses we refer to Mediero et al. [2014].

Our work is motivated by hydrological applications, where we aim at detecting smooth monotonic relationships between covariates $\mathbf{X}$ and the upper tail behavior of river discharges or precipitations Y, in particular, temporal trends in the tail behavior. The methods considered here are limited to the case of heavy-tailed response distributions F_Y, which are characterized by a right tail behavior decreasing of polynomial order controlled by the so-called extreme value index $\gamma > 0$.

From a methodological point of view, this chapter is related to Wang and Tsai [2009] and Wang and Li [2013]. These authors propose different tail estimation procedures, the former based on parametric extreme value index regression and the latter based on quantile regression in the tail region. We study a new procedure that can be viewed as L-estimation from regression quantiles. This, in turn, is a regression analogue of ordinary L-statistics, with "L" shorthand for linear combination of order statistics. It is known that estimation from certain L-statistics offers both robustness and high efficiency [Bickel and Lehmann, 1975].

Our main interest is in testing the hypothesis $\mathcal{H}_{0,tail} : \gamma(\mathbf{x}) = \gamma_0$ for some unknown $\gamma_0 > 0$ of a constant heavy-tail behavior over all possible regressor values $\mathbf{x} \in \mathcal{X}$. For that purpose, we also study a modification of Kendall's tau test statistic, where we apply the popular Mann-Kendall test (see Kendall [1948]; Yue et al. [2002]; Chebana et al. [2013]; Mediero et al. [2014] and the references therein) to a properly selected upper fraction of the sample.

We compare the performance of four different procedures that are constructed to detect deviations from $\mathcal{H}_{0,tail}$ and that are supposed to hold their nominal level in an asymptotic sense with sample size tending to infinity. Besides the power of the tests, it is equally important to study their nominal level under $\mathcal{H}_{0,tail}$ in finite-sample experiments. It turns out that, under $\mathcal{H}_{0,tail}$, the avoidance of a false alarm (rejection of $\mathcal{H}_{0,tail}$) is particularly challenging if a location $\mu(\mathbf{x})$ or scale parameter $\sigma(\mathbf{x})$ of the conditional distribution is not constant in $\mathbf{x}$. This is studied in more detail in our simulations section.

The importance of avoiding those false alarms is highlighted in another simulation experiment concerned with the comparison of estimation errors: It is highlighted that the additional source of uncertainty originating from the estimation of non-constant tail be-

havior $\gamma(\mathbf{x}) \neq \gamma_0$ is large. Since sample lengths are very limited in many applications from hydrology, it is often less erroneous (in terms of MSE) to choose a simpler model and work under $\mathcal{H}_{0,tail}$, even in experiments with a pronounced violation of the simplification.
The remainder of this chapter is organized as follows: Section 4.2 introduces the model and describes the idea of selecting samples from the tails. New methods for the analysis of conditional tails are presented in Section 4.3 and compared by simulation in Section 4.4. Applications to French weekly precipitation and to river flow data from the Mulde basin are presented in Section 4.5.

4.2 Heavy tails and relative excesses

All methods considered here are limited to right heavy-tailed response distributions according to Definition 1.9: Let $(Y, \mathbf{X})$ be a random element, where Y is a real response and $\mathbf{X} = (1, X_1, \ldots, X_d)'$ a vector of regressors with range on a compact set $\mathcal{X} \subset \mathbb{R}^{d+1}$. Throughout the article we assume that the conditional distribution of Y given $\mathbf{X} = \mathbf{x}$ is of Pareto-type, that is,

$$F_Y(y|\,\mathbf{x}) = P(Y \leq y|\,\mathbf{X} = \mathbf{x}) = 1 - y^{-1/\gamma(\mathbf{x})} \cdot L(y|\,\mathbf{x}), \tag{4.1}$$

where $\gamma : \mathcal{X} \to \mathbb{R}_+$ is strictly positive and $L(\,\cdot\,|\,\mathbf{x})$ a slowly varying function for each $\mathbf{x} \in \mathcal{X}$. Firstly, our main interest is in the statistical inference on γ, particularly, in testing hypothesis

$$\mathcal{H}_{0,tail}:\ \gamma(\mathbf{x}) = \gamma_0 \text{ for some } \gamma_0 > 0 \text{ and all } \mathbf{x} \in \mathcal{X} \tag{4.2}$$

of heavy-tail behavior constant in $\mathbf{x}$ and, secondly, we are also interested in the estimation of the conditional tail behavior under additional parametric assumptions on $\gamma(\mathbf{x})$.

Recall from Section 1.2.1 that relative excesses from Pareto-type distributions follow an approximately parametric law, which allows to construct estimators of the extreme value index and of high quantiles. Suppose that the sample $(Y_i, \mathbf{X}_i)$, $i = 1, \ldots, n$, consists of independent copies of $(Y, \mathbf{X})$. The first question to be answered for the analysis of conditional heavy tails is: How to select relative excesses under assumption (4.1)? A practical solution to this problem is discussed in the following two subsections.

4.2.1 How to choose the threshold conditional on $\mathbf{X} = \mathbf{x}$

In usual tail analysis a threshold $u \in \mathbb{R}$ is set to split the support of a univariate distribution F into a lower moderate and an upper extreme part (right tail). A natural choice is a quantile $u_p = F^{-1}(p)$ for some high probability $p \in (0, 1)$. Because here we consider conditional distributions, it is meaningful to choose a conditional quantile $u_p(\mathbf{x}) = F_Y^{-1}(p|\,\mathbf{x})$ in analogy to the unconditional case.

In practice the conditional distribution is unknown and thus $F_Y^{-1}(p|\,\mathbf{x})$ needs to be estimated. Here we follow a parametric quantile regression approach: Suppose that the conditional p-quantile of $g(Y)$ given $\mathbf{X} = \mathbf{x}$ follows a linear model

$$F_{g(Y)}^{-1}(p|\,\mathbf{x}) = \inf\left\{z :\ F_{g(Y)}(z|\,\mathbf{x}) \geq p\right\} = \mathbf{x}'\boldsymbol{\beta}_p,$$

where g is a monotone increasing function on the domain of Y and where $\boldsymbol{\beta}_p \in \mathbb{R}^{p+1}$ is an unknown parameter vector called p-th regression quantile.

Example 4.1
(i: Location-scale model) The following data generating process is frequently applied in the quantile regression literature. Let $\mathbf{X}$ be a random vector on $\mathbb{R}^{d+1}$ and $\boldsymbol{\mu}, \boldsymbol{\sigma} \in \mathbb{R}^{d+1}$ such that $\mathbf{X}'\boldsymbol{\sigma} > 0$ almost surely. Let ε be a random variable independent of $\mathbf{X}$ and define $Y = \mathbf{X}'\boldsymbol{\mu} + \mathbf{X}'\boldsymbol{\sigma} \cdot \varepsilon$. Then we have

$$F_Y^{-1}(p|\,\mathbf{x}) = \mathbf{x}'\boldsymbol{\mu} + \mathbf{x}'\boldsymbol{\sigma} \cdot F_\varepsilon^{-1}(p) = \mathbf{x}'\boldsymbol{\beta}_p$$

with $\boldsymbol{\beta}_p = \boldsymbol{\mu} + \boldsymbol{\sigma} F_\varepsilon^{-1}(p)$.
(ii: Conditional Pareto) Suppose now that Y given $\mathbf{X} = \mathbf{x}$ follows a Pareto-type distribution defined in (4.1) with $L(y|\,\mathbf{x}) = \sigma(\mathbf{x})^{1/\gamma(\mathbf{x})}$, $\sigma(\mathbf{x}) = \mathbf{x}'\boldsymbol{\xi} > 0$ and $\gamma(\mathbf{x}) = \mathbf{x}'\boldsymbol{\theta} > 0$ for some deterministic vectors $\boldsymbol{\xi}, \boldsymbol{\theta} \in \mathbb{R}^{d+1}$. This distribution is also called two-parametric Pareto with scale $\sigma(\mathbf{x})$ and shape $\alpha(\mathbf{x}) = 1/\gamma(\mathbf{x})$. Then we have

$$F_{\log(Y)}^{-1}(p|\,\mathbf{x}) = \sigma(\mathbf{x}) - \log(1-p) \cdot \gamma(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}_p$$

with $\boldsymbol{\beta}_p = \boldsymbol{\xi} - \log(1-p) \cdot \boldsymbol{\theta}$.

A consistent M-estimator of $\boldsymbol{\beta}_p$ studied in the seminal article of Koenker and Bassett [1978] is defined by

$$\hat{\boldsymbol{\beta}}_p = \underset{\mathbf{b}\in\mathbb{R}^{d+1}}{\arg\min} \sum_{i=1}^{n} \rho_p\left(g(Y_i) - \mathbf{X}_i' \cdot \mathbf{b}\right), \tag{4.3}$$

where $\rho_p(y) = y \cdot (p - \mathbb{1}_{\{y\leq 0\}})$ is the p-quantile loss function. Since conditional quantiles are invariant up to monotone increasing transformations, i.e. $F_{g(Y)}^{-1}(p|\,\mathbf{x}) = g\left(F_Y^{-1}(p|\,\mathbf{x})\right)$, we set

$$\hat{u}_p(\mathbf{x}) = \hat{F}_Y^{-1}(p|\,\mathbf{x}) = g^{-1}\left(\mathbf{x}'\hat{\boldsymbol{\beta}}_p\right). \tag{4.4}$$

If we let $p = p_{k,n} = \frac{n-k}{n+1}$ with corresponding estimator denoted by $u_{k,n} = \hat{u}_{p_{k,n}}$, we almost get k out of n elements $(Y_i, \mathbf{X}_i)$ with $Y_i > u_{k,n}(\mathbf{X}_i)$ (in simulations mostly between $k-2$ and $k+2$). In what follows, we neglect this small deviation from k. For notational simplicity, we suppose that we get exactly k out of n excesses if we choose $p = p_{k,n}$.

The assumption that the conditional quantile is linear after some known transformation g might be too restrictive. A more flexible approach studied in Mu and He [2007] and also

applied in Wang and Li [2013] is based on the family $\{g_\lambda : \mathbb{R}_+ \to \mathbb{R} \mid \lambda \in \mathbb{R}\}$ of Box-Cox transformations

$$g_\lambda(y) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & , \text{if } \lambda \neq 0 \\ \log(y) & , \text{if } \lambda = 0 \end{cases}.$$

In the previous reference it is assumed that the conditional p-quantile of $g_\lambda(Y)$ given $\mathbf{X} = \mathbf{x}$ follows a linear model, where the parameter $\lambda = \lambda_p$ is unknown. Interestingly enough, Teugels and Vanroelen [2004] showed that the extreme value index $\gamma^*(\mathbf{x})$ of $g_\lambda(Y)$ conditional on $\mathbf{X} = \mathbf{x}$ satisfies $\gamma^*(\mathbf{x}) = \lambda \cdot \gamma(\mathbf{x}) \in \mathbb{R}$, provided (4.1) holds.
Mu and He [2007] proposed the consistent estimator

$$\hat{\lambda}_p = \underset{\lambda \in \mathbb{R}}{\arg\min} \sum_{i=1}^{n} [R_n(\mathbf{x}_i, \lambda, p)]^2 \tag{4.5}$$

of λ_p, where

$$R_n(\mathbf{x}, \lambda, p) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}(\mathbf{x}_j \leq \mathbf{x}) \cdot \left[p - \mathbf{1}\left(g_\lambda(Y_j) \leq \mathbf{x}_i' \hat{\boldsymbol{\beta}}_{p,\lambda}\right)\right]$$

and $\hat{\boldsymbol{\beta}}_{p,\lambda}$ is computed by (4.3) with $g = g_\lambda$.
In summary, the following routine can be applied to select k out of n relative excesses from a sample $(Y_i, \mathbf{X}_i)$, $i = 1, \ldots, n$, and for a fixed number $k < n$:

(i) Set $p = p_{k,n} = \frac{n-k}{n+1}$ and compute $\hat{\lambda}$ by (4.5).

(ii) Solve (4.3) with $g = g_{\hat{\lambda}}$ and let $u(\mathbf{x}) = u_{k,n}(\mathbf{x}) = g_{\hat{\lambda}}^{-1}\left(\mathbf{x}' \hat{\boldsymbol{\beta}}_p\right)$.

(iii) Identify all $1 \leq i_1 < \ldots < i_k \leq n$ with $Y_{i_j} > u(\mathbf{X}_{i_j})$ and let $(Z_{k,j}, \mathbf{X}_{k,j})$, $j = 1, \ldots, k$, denote the sample of relative excesses $Z_{k,j} = Y_{i_j} / u(\mathbf{X}_{i_j})$ with corresponding regressors $\mathbf{X}_{k,j} = \mathbf{X}_{i_j}$.

For single regressors $\mathbf{X} = (1, X)'$ we write $(Z_{k,j}, X_{k,j})$ instead of $\left(Z_{k,j}, (1, X_{k,j})'\right)$.

4.2.2 How to select k

After discussing the shape of the threshold function u for fixed k, we now turn to the selection of k representing the number of relative excesses included in the tail analysis. Wang and Tsai [2009] proposed a data driven selection of k based on the minimization of a discrepancy measure. Similar to them, we let

$$k^* = \underset{1 \leq k < n}{\arg\min}\, D_n(k) = \underset{1 \leq k < n}{\arg\min} \frac{1}{k} \sum_{j=1}^{k} \left(\hat{U}_{k,j:k}(\hat{\gamma}_{k,n}) - \frac{j}{k+1} \right)^2 \tag{4.6}$$

where $\hat{U}_{k,1:k}(\hat{\gamma}_{k,n}) < \ldots < \hat{U}_{k,k:k}(\hat{\gamma}_{k,n})$ are order statistics from a sample computed by $\hat{U}_{k,j}(\hat{\gamma}_{k,n}) = \exp\left(-\log(Z_{k,j}) / \hat{\gamma}_{k,n}(\mathbf{X}_{k,j})\right)$, $j = 1, \ldots, k$, and $\hat{\gamma}_{k,n}(\mathbf{x})$ is an estimator of $\gamma(\mathbf{x})$

computed from $(Z_{k,j}, \mathbf{X}_{k,j})$, $j = 1, \ldots, k$.
The minimization in (4.6) is interpreted as a solution to a trade-off problem: On the one hand, large numbers k worsen the approximation of $Z_{k,j}$ being $Pareto(1/\gamma(\mathbf{X}_{k,j}))$ distributed and thus, of $U_{k,j} = \exp\left(-\log(Z_{k,j})/\gamma_{k,n}(\mathbf{X}_{k,j})\right)$ being uniformly distributed. On the other hand, too small numbers k decrease the efficiency of estimator $\hat{\gamma}$, which, in turn, deteriorates the approximation of $\hat{U}_{k,j}(\hat{\gamma}_{k,n})$ being uniformly distributed.

4.3 New estimator and tests

In this section we suppress the previous approximation and instead simply assume that the sample $(Z_{k,j}, \mathbf{X}_{k,j})$, $j = 1, \ldots, k$, consists of independent and identically distributed variables with $\mathbb{P}(Z_{k,j} \leq z |\, \mathbf{X}_{k,j} = \mathbf{x}) = 1 - z^{-1/\gamma(\mathbf{x})}$. A similar idea and some theoretical background for this simplification is presented in Beirlant et al. [2006, Chap. 7.3]. A more rigorous justification in a related problem is given in Wang and Tsai [2009]. There it is shown that the asymptotic normality of their estimator remains valid also without the previous simplification but with an additional bias $\mathbf{h}$ included in the mean of the limiting distribution. For practical reasons, since the estimation of $\mathbf{h}$ requires detailed information on the tail that is very hard to obtain, the bias usually is set to zero in finite-sample applications [Resnick, 2007; Wang and Tsai, 2009; Wang and Li, 2013].

4.3.1 L-estimation of linear models $\gamma(\mathbf{x}) = \mathbf{x}'\boldsymbol{\eta}$ and related tests

Let $(Z, \mathbf{X})$ be a random element on $\mathbb{R} \times \mathbb{R}^{d+1}$ satisfying

$$F_Z(z |\, \mathbf{x}) = \mathbb{P}(Z \leq z |\, \mathbf{X} = \mathbf{x}) = 1 - z^{-1/\gamma(\mathbf{x})} \;\text{ with }\; \gamma(\mathbf{x}) = \mathbf{x}'\boldsymbol{\eta}$$

for all $\mathbf{x} \in \mathcal{X}$ and some deterministic vector $\boldsymbol{\eta} = (\eta_0, \eta_1, \ldots, \eta_d)' \in \mathbb{R}^{d+1}$. It follows that

$$F^{-1}_{\log(Z)}(p |\, \mathbf{x}) = -\gamma(\mathbf{x}) \cdot \log(1-p) = \mathbf{x}'\boldsymbol{\beta}_p$$

for $\boldsymbol{\beta}_p = -\log(1-p) \cdot \boldsymbol{\eta}$ and all $p \in (0,1)$. In words, conditional quantiles are linear in covariates $\mathbf{x}$, which allows us to estimate $\gamma(\mathbf{x})$ by linear quantile regression [Koenker, 2005]: Let $(Z_j, \mathbf{X}_j)$, $j = 1, \ldots, k$, denote independent copies of $(Z, \mathbf{X})$ and

$$\hat{\boldsymbol{\beta}}_p = \operatorname*{arg\,min}_{\mathbf{b} \in \mathbb{R}^{d+1}} \sum_{j=1}^{k} \rho_p\left(\log(Z_j) - \mathbf{X}_j' \cdot \mathbf{b}\right)$$

with $\rho_p(u) = u \cdot \left(p - \mathbf{1}_{\{u \leq 0\}}\right)$. By setting $\hat{\boldsymbol{\eta}}_p = -\hat{\boldsymbol{\beta}}_p / \log(1-p)$ we obtain an estimator of $\boldsymbol{\eta}$ for each $p \in (0,1)$. Restricting on one probability p obviously is not a satisfactory solution to our estimation problem. Instead we are going to gather information from estimates $\hat{\boldsymbol{\eta}}_p$ for multiple probabilities $p \in \mathbf{p} \subset (0,1)$. From Theorem A.11 in Appendix A.6 and the model assumptions stated above, we easily obtain the following result:

Proposition 4.2
Let $\mathbf{p} = \{p_1, \ldots, p_\ell\} \subset (0,1)$ *denote a finite set of distinct probabilities and suppose that* $J = E(\mathbf{X}\mathbf{X}')$ *and* $H = E\left(\mathbf{X}\mathbf{X}'/\mathbf{X}'\boldsymbol{\eta}\right) \in \mathbb{R}^{(d+1)\times(d+1)}$ *exist with* H *positive definite. Then, under the assumptions from above and for* $k \to \infty$*, we have that*

$$\sqrt{k}\left(\hat{\boldsymbol{\eta}}'_{p_1} - \boldsymbol{\eta}', \ldots, \hat{\boldsymbol{\eta}}'_{p_\ell} - \boldsymbol{\eta}'\right)' \xrightarrow{D} \mathcal{N}(0, \Omega_{\mathbf{p}}),$$

where $\Omega_{\mathbf{p}} = A_{\mathbf{p}} \otimes (H^{-1}JH^{-1})$ *and* $A_{\mathbf{p}} \in \mathbb{R}^{\ell\times\ell}$ *is defined through its entries*

$$a_{ij} = a(p_i, p_j) = \frac{p_i \wedge p_j - p_i \cdot p_j}{(1-p_i)(1-p_j)\log(1-p_i)\log(1-p_j)},\ 1 \le i, j \le \ell,$$

and with $\otimes$ *denoting the Kronecker product.*

Proof. Recall that $\hat{\boldsymbol{\eta}}_p = -\log(1-p)^{-1}\hat{\boldsymbol{\beta}}_p$ and $\boldsymbol{\eta} = -\log(1-p)^{-1}\boldsymbol{\beta}_p$ for all $p \in (0,1)$. Weak convergence towards a multivariate normal distribution follows directly from Theorem A.11. It remains to verify that $\Omega_{\mathbf{p}}$ is the corresponding covariance matrix.
Under the conditions from the present section we have that $f(\cdot|\mathbf{x})$ and $F(\cdot|\mathbf{x})$ are the density and distribution function, respectively, of $\log(Z)$ conditional on $\mathbf{X} = \mathbf{x}$, which is exponential with parameter $1/\gamma(\mathbf{x})$. Following the notation from Theorem A.11, we have that $H_p = (1-p)\cdot H$ with H_p defined in (A.8), which, for $k \to \infty$, gives us

$$\operatorname{Cov}\left[\sqrt{k}\left(\hat{\boldsymbol{\eta}}_{p_i} - \boldsymbol{\eta}\right), \sqrt{k}\left(\hat{\boldsymbol{\eta}}_{p_j} - \boldsymbol{\eta}\right)\right] \longrightarrow a(p_i, p_j)\cdot H^{-1}JH^{-1}.$$

□

As a direct application, we are able to derive the limiting distribution of so called L-estimators $\hat{\boldsymbol{\eta}}(\mathbf{p}, \mathbf{w}) = \sum_{i=1}^{\ell} w_i \cdot \hat{\boldsymbol{\eta}}_{p_i}$ of $\boldsymbol{\eta}$, where $\mathbf{w} = (w_1, \ldots, w_\ell)$ is a vector of weights satisfying $\sum_{i=1}^{\ell} w_i = 1$. We obtain

$$\sqrt{k}\left(\hat{\boldsymbol{\eta}}(\mathbf{p}, \mathbf{w}) - \boldsymbol{\eta}\right) \xrightarrow{D} \mathcal{N}\left(0, B_{\mathbf{w}}\Omega_{\mathbf{p}}B'_{\mathbf{w}}\right) = \mathcal{N}\left(0, \Sigma_{\mathbf{p},\mathbf{w}}\right), \tag{4.7}$$

where $B_{\mathbf{w}} = (w_1 \cdot I_{d+1}, \ldots, w_\ell \cdot I_{d+1}) \in \mathbb{R}^{(d+1)\times(d+1)\ell}$ and $I_{d+1} \in \mathbb{R}^{(d+1)\times(d+1)}$ is the identity on $\mathbb{R}^{(d+1)}$.
As a second application, it is straightforward to construct test statistics for linear hypotheses of the form $\mathcal{H}:\ C\boldsymbol{\eta} = 0$ vs. $\mathcal{A}:\ C\boldsymbol{\eta} \neq 0$, where $C \in \mathbb{R}^{m\times(d+1)}$ is a given matrix. In the simulations section, where we use $d = 1$, the statistic $T^L = \sqrt{k}\hat{\eta}_1/\hat{\sigma}_1$ as a test for $\mathcal{H}_{0,tail}$ vs. $\mathcal{H}_{1,tail}$ is referred to as the L-test.
We close this section with three important remarks:

i): Selection of ℓ, $\mathbf{p}$ and $\mathbf{w}$
Next to be answered is how to choose the number $\ell \in \mathbb{N}$, a set of probabilities $\mathbf{p} \in (0,1)^\ell$ and the corresponding weights $\mathbf{w} = (w_1, \ldots, w_\ell)'$ with $\sum_{j=1}^{\ell} w_j = 1$. Let us first consider the last issue. Suppose that ℓ and $\mathbf{p} \in (0,1)^\ell$ are fixed and let $\hat{\boldsymbol{\eta}}(\mathbf{p}, \mathbf{w}) = (\hat{\eta}_0, \hat{\eta}_1, \ldots, \hat{\eta}_d)'$. Then, for arbitrary weights $\mathbf{w}$ and for each component $\hat{\eta}_j$, we have that

$$\operatorname{Var}\left[\hat{\eta}_j\right] \approx c_j/k \cdot \mathbf{w}' A_{\mathbf{p}}\mathbf{w},\ j = 0, \ldots, d,$$

where $c_j > 0$ does not depend on $\mathbf{p}$ and $\mathbf{w}$. It is therefore sensible if we choose

$$\mathbf{w}_{opt} = \mathbf{w}_{opt}(\mathbf{p}) = \arg\min_{\mathbf{w}} \mathbf{w}' A_{\mathbf{p}} \mathbf{w} = \left(\mathbf{1}' A_{\mathbf{p}}^{-1} \mathbf{1}\right)^{-1} \cdot A_{\mathbf{p}}^{-1} \mathbf{1},$$

where $\mathbf{1} = (1, \ldots, 1)' \in \mathbb{R}^{\ell}$. This is the solution of the previous minimization problem obtained by the Lagrange multipliers technique. Note also that the optimal solution $\mathbf{w}_{opt}$ is the same for all $d+1$ components.
Let us now turn to the selection of the number ℓ and the set $\mathbf{p} \in (0,1)^{\ell}$. From a theoretical point of view, since $w_j = 0$ is possible in the previous minimization, the more probabilities p_j we include the better the estimation. However, from several simulation experiments in the semi-parametric setting (4.1) we found that the choice of a moderate number of, say, $\ell = 20$ probabilities equally spaced in the upper half $[1/2, 1 - 1/40]$ performs well and including additional probabilities did not improve the efficiency.

ii): Deterministic regressors and non-identically distributed observations
So far, we considered samples $(Z_j, \mathbf{X}_j)$, $j = 1, \ldots, k$, as independent and identically distributed, but sometimes this framework does not cover the actual problem: Suppose that $\mathbf{X}_i = \mathbf{x}_i$ is a deterministic sequence of regressors, for instance, regression over the (rescaled) time domain. Rescaling is needed for technical reasons. Then, in many situations, the observations Z_j, $j = 1, \ldots, k$, may still be considered as independent but not identically distributed (i.ni.d.). Thanks to the theory on quantile regression processes based on sequences of such observations [Koenker, 2005, Sec. 4.3], it is still possible to apply the results from the previous section: Let Z_j have a Pareto distribution with extreme value index $\gamma(\mathbf{x}_j) = \mathbf{x}_j' \boldsymbol{\eta} > 0$, $j \in \mathbb{N}$. Then the assertion of Proposition 4.2 holds even in this i.ni.d. case if we define J and H by

$$J = \lim_{k\to\infty} \frac{1}{k} \sum_{j=1}^{k} \mathbf{x}_j \mathbf{x}_j' \text{ and } H = \lim_{k\to\infty} \frac{1}{k} \sum_{j=1}^{k} \mathbf{x}_j \mathbf{x}_j' / \mathbf{x}_j \boldsymbol{\eta},$$

provided these two limits exist. So, from a computational point of view, there is no difference to the former case of i.i.d. observations.

iii): Application to samples from the conditional tail
Initially we started with random elements $(Y, \mathbf{X})$ with conditional distribution defined in (4.1). Assuming that $\gamma(\mathbf{x}) = \mathbf{x}' \boldsymbol{\eta} > 0$ holds for all $\mathbf{x} \in \mathcal{X}$ and some unknown $\boldsymbol{\eta} \in \mathbb{R}^{d+1}$, our main interest is in statistical inference on $\boldsymbol{\eta}$. Following the introductory lines of Section 4.3, it is sensible to apply estimator $\hat{\boldsymbol{\eta}}(\mathbf{p}, \mathbf{w})$ on the sample $(Z_{k,j}, \mathbf{X}_{k,j})$, $j = 1, \ldots, k$, from Subsection 4.2.1, which are approximately distributed like $(Z, \mathbf{X})$. The estimator is denoted by $\hat{\boldsymbol{\eta}}_{k,n}^{L} = \hat{\boldsymbol{\eta}}_{k,n}^{L}(\mathbf{p}, \mathbf{w})$. It is left for future research to prove that a statement similar to (4.7) holds also in this approximate setting, presumably with an additional bias $\mathbf{h}$ in the mean of the limit but with the same limiting covariance matrix. In applications it is common to ignore a possible bias $\mathbf{h}$ and the covariance matrix is estimated by plugging in $J_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \mathbf{X}_{k,j} \mathbf{X}_{k,j}'$ and $H_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \mathbf{X}_{k,j} \mathbf{X}_{k,j}' / \hat{\boldsymbol{\eta}}_{k,n}^{L}(\mathbf{X}_{k,j})$ for J and H, respectively.

4.3.2 Kendall's tau tail-test

In the single regressor setting $\mathbf{X} = (1, X)'$, Kendall's tau test [Kendall, 1948] is a simple rank based test against dependence. More precisely, let C denote the copula of (X, Y). Then we have that

$$\tau(X, Y) = 4 \cdot \int_{[0,1]^2} C(u, v) dC(u, v) - 1 \tag{4.8}$$

defines Kendall's correlation coefficient. $\tau(X, Y)$ is a margin-free dependence coefficient with $-1 \leq \tau(X, Y) \leq 1$ and $\tau(X, Y) = 0$ for independent variables X and Y. The empirical counterpart of (4.8) is $t_n = \frac{2}{n(n-1)} \cdot S_n$ with

$$S_n = \sum_{1 \leq i < j \leq n} sgn(Y_j - Y_i) \cdot sgn(X_j - X_i)$$

and sign function

$$sgn(z) = \begin{cases} 1 \ , \ z > 0 \\ 0 \ , \ z = 0 \\ -1 \ , \ z < 0 \end{cases} .$$

The statistic S_n is used to test $\mathcal{H}_{0,ind}$: X and Y are independent. Under $\mathcal{H}_{0,ind}$ it is known that the distribution of S_n is well approximated by $\mathcal{N}(0, \sigma_n^2)$ with $\sigma_n^2 = n(n+1)(2n+5)/18$, provided $P(X_i = X_j) = P(Y_i = Y_j) = 0$ for $i \neq j$ [Kendall, 1948; Yue et al., 2002]. Because here the interest is in the tail behavior of Y conditional on X, we propose the test statistic

$$S_{k,n} = \sum_{1 \leq i < j \leq k} sgn(Z_{k,j} - Z_{k,i}) \cdot sgn(X_{k,j} - X_{k,i}), \tag{4.9}$$

which is Kendall's test for the sample $(Z_{k,j}, X_{k,j})$, $j = 1, \ldots, k$, of relative excesses and their regressors (see Section 4.2.1). Critical values are computed based on the presumed approximation $S_{k,n} \overset{D}{\approx} \mathcal{N}(0, \sigma_k^2)$.

4.4 Simulation study

4.4.1 Detection of conditional heavy-tail behavior

In this section we focus our attention on the problem of testing hypothesis $\mathcal{H}_{0,tail}$ of constant heavy-tail behavior stated in (4.2). The following questions are the main sources of our motivation:

i) Is it realistically possible to distinguish between trends in $\gamma(\mathbf{x})$ and trends in, say, conditional location $\mu(\mathbf{x})$ or scale $\sigma(\mathbf{x})$ at constant shape $\gamma(\mathbf{x}) = \gamma_0$? In other words: Do the tests keep their nominal level under $\mathcal{H}_{0,tail}$ even in more challenging scenarios than conditional distributions constant in $\mathbf{x}$?

ii) Under what circumstances is it possible to detect deviations from $\mathcal{H}_{0,tail}$ with satisfactory power?

It has to be considered that models with non-constant tail behavior $\gamma(\mathbf{x}) \not\equiv \gamma_0$ suffer from this additional source of uncertainty, especially when the focus is on the right tail, say, the estimation of high quantiles of the conditional distribution. On the one hand, it might be important to take a non-constant $\gamma(\mathbf{x})$ into account. On the other hand, it might be even more important to preserve simplicity in order to keep estimation uncertainty as small as possible.
We restrict our attention to the case $d = 1$. We consider scenarios with non-constant conditional location, scale or shape. Anyway, data are generated independently in i by

$$Y_i = \mathbf{X}_i'\boldsymbol{\mu} + \mathbf{X}_i'\boldsymbol{\sigma} \cdot \varepsilon_i,\ 1 \leq i \leq n, \tag{4.10}$$

where $\mathbf{X}_i = (1, X_i)'$ is a random vector with second component X_i uniformly distributed on $\mathcal{X} = [-1,\ 1]$, and with nonrandom parameter vectors $\boldsymbol{\mu} = (\mu_0, \mu_1)'$ and $\boldsymbol{\sigma} = (\sigma_0, \sigma_1)'$. Furthermore, the variables ε_i satisfy

$$P(\varepsilon_i \leq y |\ \mathbf{X}_i = \mathbf{x}) = \exp\left(-\left[1 + \gamma(\mathbf{x}) \cdot y\right]^{-1/\gamma(\mathbf{x})}\right),\ y > -1/\gamma(\mathbf{x}),$$

for some nonrandom vector $\boldsymbol{\eta} = (\eta_0, \eta_1)'$. All in all, this means that the variables Y_i conditional on $\mathbf{X}_i = \mathbf{x}$ are generalized extreme value (GEV) distributed with location $\mu(\mathbf{x}) = \mathbf{x}'\boldsymbol{\mu}$, scale $\sigma(\mathbf{x}) = \mathbf{x}'\boldsymbol{\sigma} > 0$ and shape $\gamma(\mathbf{x}) = \mathbf{x}'\boldsymbol{\eta} > 0$. Recall that the shape of the GEV is also its extreme value index. Since we are dealing with relative excesses, all the methods are scale but not location invariant. Studying many river flow time series from Saxony in Germany we found that a ratio of about $\mu_0/\sigma_0 = 2$ is common, which, for simplicity, is used throughout the simulation experiments.
For convenience, we denote the hypotheses of constant location and constant scale by

$$\mathcal{H}_{0,loc} :\ \mu_1 = 0 \ \text{ and } \ \mathcal{H}_{0,scale} :\ \sigma_1 = 0,$$

respectively. Corresponding alternatives with non-zero slope $\mu_1 \neq 0$ and $\sigma_1 \neq 0$ are denoted by $\mathcal{H}_{1,loc}$ and $\mathcal{H}_{1,scale}$, respectively. Intersections are abbreviated as

$$\mathcal{H}_{a,b,c} = \mathcal{H}_{a,loc} \cap \mathcal{H}_{b,scale} \cap \mathcal{H}_{c,tail} \ \text{ for } a, b, c \in \{0, 1\},$$

where $\mathcal{H}_{0,tail} :\ \eta_1 = 0$ and $\mathcal{H}_{1,tail} :\ \eta_1 \neq 0$.
We compare the finite-sample performance of

3S: a test based on the three-stage test statistic in (A.12),

K: the test based on the Kendall's tau statistic (4.9),

L: a two-sided t-test based on the weak limit of estimator $\hat{\eta}_1$ in (4.7) and

TIR: a two-sided t-test based on the weak limit of estimator $\hat{\theta}_1$ in (A.11) with $h_1 = 0$.

Figure 4.1: Rejection rates of four different tests computed from 4000 samples generated by process (4.10) under (top) scenarios $\mathcal{H}_{ab0}$ involving the null $\mathcal{H}_{0,tail}$ and (bottom) scenarios $\mathcal{H}_{ab1}$ involving the alternative $\mathcal{H}_{1,tail}$. Samples were generated by (4.10) with sample length $n = 500$ and parameters set to $\mu_0 = 2$, $\sigma_0 = 1$ and $\eta_0 = 0.4$. Under $\mathcal{H}_{1,b,c}$, $\mathcal{H}_{a,1,c}$ and $\mathcal{H}_{a,b,1}$ we set $\mu_1 = \mu_0/4 = 0.5$, $\sigma_1 = \sigma_0/4 = 0.25$ and $\eta_1 = \eta_0/2 = 0.2$, respectively. Note that the y-axis is on square root scale.

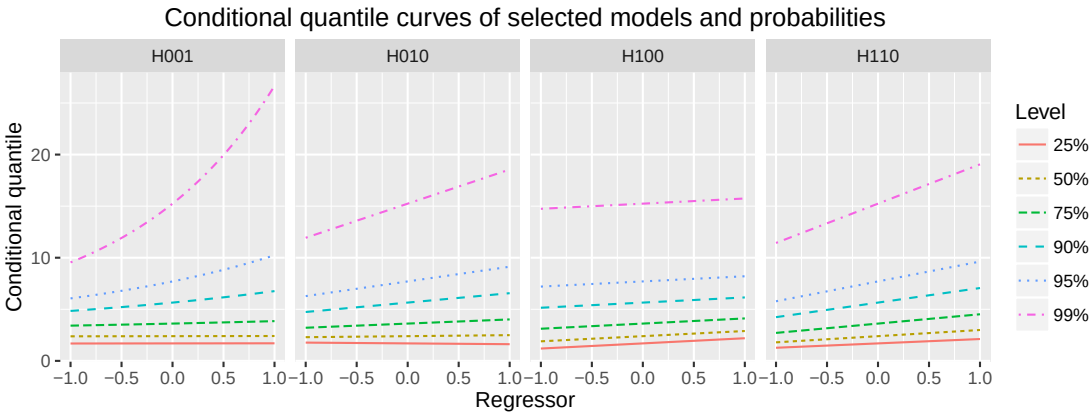

Figure 4.2: Conditional quantile curves $F^{-1}(p|\ x)$ for probability levels $p \in \{0.25,\ 0.5,\ 0.75,\ 0.9,\ 0.95,\ 0.99\}$ and for a few selected scenarios defined in Section 4.4.1.

Table 4.1: Rejection rates of hypothesis $\mathcal{H}_{0,tail}$ computed from 4000 samples under several scenarios under the null identical with those from the top of Figure 4.1. The nominal level is 5%.

	L	L^*	TIR	TIR^*	K	K^*
$\mathcal{H}_{0,0,0}$	3.9	3.1	4.3	5.3	4.9	4.9
$\mathcal{H}_{0,1,0}$	7.5	10.6	9.5	15.9	10.2	13.8
$\mathcal{H}_{1,0,0}$	7.5	12.0	9.3	15.4	11.0	14.6
$\mathcal{H}_{1,1,0}$	3.3	3.6	4.1	5.3	5.2	5.1

These tests are used to check $\mathcal{H}_{0,tail}$ vs. $\mathcal{H}_{1,tail}$ at a nominal level of 5%. The simulation results are presented in Figure 4.1. We computed rejection rates (y-axis) of the null for different values of the tuning parameter k (x-axis) and for six scenarios $\mathcal{H}_{a,b,c}$, involving (top) three $\mathcal{H}_{a,b,0}$ under the null and (bottom) the others $\mathcal{H}_{a,b,1}$ under the alternative. Rejection rates are computed from 4000 independent samples for each scenario with sample length $n = 500$. In all experiments we set $\mu_0 = 2$, $\sigma_0 = 1$ and $\eta_0 = 0.4$. Under scenarios $\mathcal{H}_{1,b,c}$, $\mathcal{H}_{a,1,c}$ and the alternatives $\mathcal{H}_{a,b,1}$ we used $\mu_1 = \mu_0/4 = 0.5$, $\sigma_1 = \sigma_0/4 = 0.25$ and $\eta_1 = \eta_0/2 = 0.2$, respectively. Note also that the rejection rates on the y-axis are given on square root scale.
The performance under the null is presented on the upper half of Figure 4.1. It is particularly interesting to study the impact of the tuning parameter k on the size of the tests. Recall that our tail model assumptions are built in such a way that the justification of the approximation improves with smaller values of k. Indeed, the size of the tests is close to 5% for k being around 50 to 100 in all the considered scenarios under the null. The performance under $\mathcal{H}_{0,1,0}$ and $\mathcal{H}_{1,0,}$ is particularly interesting. There the tests K and TIR are too liberal for all values of the tuning parameter k. The overall best performance under the null is given by test L followed by $3S$, which hold their nominal level of 5% for a reasonable range of k values.
Under the alternatives $\mathcal{H}_{a,b,1}$ presented on the lower half of Figure 4.1, we observe that the power of tests $3S$, L and TIR is very similar, with a slight advantage of test TIR. The fact that test K has the lowest power is not surprising, since this test is based on the weakest model assumptions.
The same experiments with sample lengths $n = 200$ and $n = 1000$ led to qualitatively the same results and are thus not reported here.

4.4.2 Selection of k

This part of the simulations is devoted to the adaptive selection of k, for instance, rule (4.6). The simulation results depicted in Figure 4.3 are computed under $\mathcal{H}_{0,0,1}$ with observations generated by (4.10), where X is uniformly distributed on $[-1, 1]$, $\boldsymbol{\mu} = (2, 0)'$, $\boldsymbol{\sigma} = (1, 0)'$ and $\boldsymbol{\eta} = (0.4, \eta_1)'$. We compared several versions of the L-estimator $\hat{\gamma}^L(1)$ of $\gamma(1) = \eta_0 + \eta_1$: The L-estimator with (L) $k = \lfloor 2n^{2/3} \rfloor$, (L^*) k^* from (4.6), (L^*2) $k = \lfloor 0.75k^* \rfloor$ and (L^*3) $k = \lfloor 1.25k^* \rfloor$. $k = \lfloor 2n^{2/3} \rfloor$ is the asymptotically optimal choice for

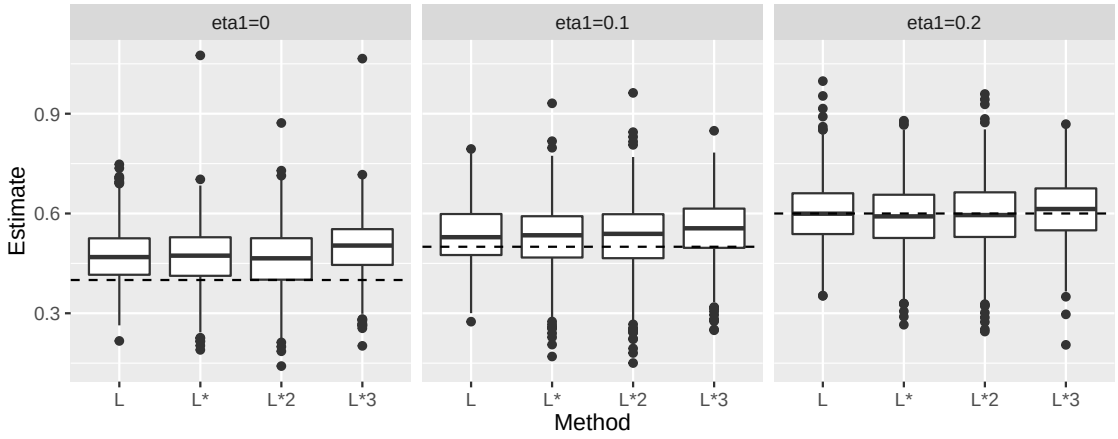

Figure 4.3: Comparison of L-estimators $\hat{\gamma}^L(1)$ of $\gamma(1) = \eta_0 + \eta_1$ based on 1000 independent samples of DGP (4.10) with $\boldsymbol{\mu} = (2,0)'$, $\boldsymbol{\sigma} = (1,0)'$, $\boldsymbol{\eta} = (0.4, \eta_1)$ and sample length $n = 500$ for different selection rules of k: (L) $k = \lfloor 2n^{2/3} \rfloor$, (L^*) $k = k^*$, (L^*2) $k = \lfloor 0.75 \cdot k^* \rfloor$ and (L^*3) $k = \lfloor 1.25 \cdot k^* \rfloor$, where k^* is a data-adaptive rule defined in (4.6). The dashed line corresponds to the true value. The corresponding mean squared errors are given by $(\eta_1 = 0)$ $0.0113, 0.0126, 0.0124, 0.0176$, $(\eta_1 = 0.1)$ $0.0091, 0.0096, 0.0104, 0.0108$ and $(\eta_1 = 0.2)$ $0.0085, 0.0087, 0.0099, 0.0074$.

independent and identically GEV-distributed observations [Gomes and Pestana, 2007] and indeed, this choice led to the best results in our simple scenarios with GEV innovations. In practice, however, we may not always expect that the observations stem from a known parametric family and it may be preferable to choose a data adaptive rule. Overall, we found that the performance of the L-estimator with $k = k^*$ from (4.6) is quite similar to that with the asymptotically optimal choice. The modifications (L^*2) and (L^*3) perform worse.
Finally, we have also compared the size of the L, TIR and K tests with $k = \lfloor 2n^{2/3} \rfloor$ and those with a data-adaptive rule for k. For the TIR test we have used the rule proposed in Wang and Tsai [2009] and for the remaining two tests the rule (4.6) was applied. Data were generated under the same scenarios like in the top of Figure 4.1. Table 4.1 presents the simulation results computed from 4000 independent samples of size $n = 500$. The size of the tests is reasonably close to the nominal level of $\alpha = 5\%$ under scenarios $\mathcal{H}_{0,0,1}$ and $\mathcal{H}_{1,1,0}$. The tests are too liberal under $\mathcal{H}_{0,1,0}$ and $\mathcal{H}_{1,0,0}$, which is even worse with data-adaptive selection of k.

4.4.3 Estimation of conditional heavy-tail behavior

In view of the potential applications with its focus on high quantiles and typically rather limited observation lengths, we may ask: Is it meaningful to consider conditional heavy-tail behavior in hydrological applications, or is it better to rely on less complex models

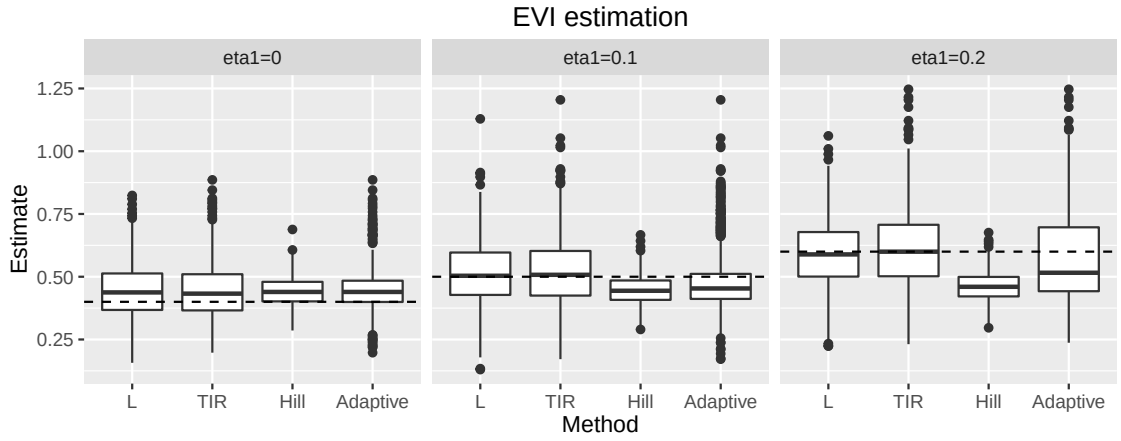

Figure 4.4: Comparison of estimators of $\gamma(1) = \eta_0 + \eta_1$ based on 1000 independent samples of DGP (4.10) with $\boldsymbol{\mu} = (2,0)'$, $\boldsymbol{\sigma} = (1,0)'$, $\boldsymbol{\eta} = (0.4, \eta_1)$ and with sample length $n = 500$ and $k = \lfloor n^{2/3} \rfloor$. The corresponding mean squared errors are given by $(\eta_1 = 0)$ $0.0140, 0.0145, 0.0046, 0.0089$, $(\eta_1 = 0.1)$ $0.0144, 0.0175, 0.0060, 0.0142$ and $(\eta_1 = 0.2)$ $0.0166, 0.0234, 0.0238, 0.0300$.

(work under $\mathcal{H}_{0,tail}$) even if this simplification is not true?
More precisely, we evaluate the following questions:

i) What is the effect of conditional heavy-tail behavior $\gamma(x) = \eta_0 + x\eta_1$ on quantiles of the conditional distribution?

ii) What about estimation efficiency? Under what circumstances (sample length, degree of heavy-tail variability) is it worthwhile to estimate non-constant heavy-tail behavior?

Figure 4.2 illustrates the shape of the conditional distribution as a function of the regressor $x \in [-1, 1]$ on different quantile levels. We have selected a few scenarios, which were already applied in the simulations from the previous section. Particular attention should be paid to scenario $\mathcal{H}_{0,0,1}$ with trend $\gamma(x) = 0.4 + 0.2x$ in the shape but constant location $\mu(x) = 2$ and scale $\sigma(x) = 1$. Note that the lower 75% of the conditional distribution is almost unchanged over the whole regressor space, while, say, the 99%-quantile drastically increases by more than 150%. In contrast to that, a pure trend in location $(\mathcal{H}_{1,0,0})$ or in scale $(\mathcal{H}_{0,1,0})$ has a rather moderate effect on the different quantiles of the conditional distribution.
For the evaluation of question ii), suppose that hypothesis $\mathcal{H}_{0,0,1}$ is met with observations generated by (4.10), where again X is uniformly distributed on $[-1, 1]$, $\boldsymbol{\mu} = (2,0)'$, $\boldsymbol{\sigma} = (1,0)'$ and $\boldsymbol{\eta} = (0.4, \eta_1)'$. Think of $[-1, 1]$ representing the rescaled time period in which we have collected our observations. Suppose that we are interested in todays heavy-tail behavior, that is, in the estimation of $\gamma(1) = \eta_0 + \eta_1$ at time $x = 1$. We compare the following estimators:

L: Estimator $\hat{\gamma}^L(1) = \hat{\eta}_0^L + \hat{\eta}_1^L$ with $\hat{\boldsymbol{\eta}}^L$ defined in Section 4.3.1.

TIR: Estimator $\hat{\gamma}^{TIR}(1) = \exp(\hat{\theta}_0^{TIR} + \hat{\theta}_1^{TIR})$ with $\hat{\boldsymbol{\theta}}^{TIR}$ defined in Section A.7.1.

Hill: The usual Hill estimator from Hill [1975], which assumes that $\eta_1 = 0$ holds.

Ad.: An adaptive procedure, which applies the TIR-estimator if the TIR-based test rejects $\mathcal{H} : \eta_1 = 0$ at a level of 10%. Otherwise, the Hill estimator is used.

Figure 4.4 shows the simulation results for scenario $\mathcal{H}_{0,0,1}$ with (left) $\eta_1 = 0$, (middle) $\eta_1 = 0.1$ and (right) $\eta_1 = 0.2$. Boxplots are computed from 1000 independent repetitions with sample length $n = 500$ and a fixed effective sample size of $k = \lfloor n^{2/3} \rfloor$. As expected, we observe that the estimation of an additional trend η_1 (estimators L, TIR and partly the adaptive method) has to be paid by a large increase of estimation variability. The Hill estimator is the only one that assumes $\eta_1 = 0$, which results in an increasing estimation bias with increasing $\eta_1 > 0$. In terms of mean squared errors (see caption of Figure 4.4), estimators L and TIR are preferable over Hill only in scenario $\eta_1 = 0.2$. The adaptive method is preferable in none of the considered cases. The same experiments with alternative rules for the selection of k and also with $n = 1000$ (not reported here) did not change our conclusions. For $n = 200$ (not reported here) the simple Hill estimator is preferable in all three scenarios. Summing up, the estimation of non-constant conditional heavy-tail behavior is useful only if η_1 is large relative to n, which, for typical applications from hydrology, presumably is not the case.

4.5 Applications

4.5.1 Weekly maxima of hourly precipitation in France

The left-hand side of Figure 4.5 displays two series of $n = 228$ weekly maxima of hourly precipitation measured during the fall season at the stations Nevers and Niort in France. We are going to analyze whether the right-tail behavior changes over time or not. Recently, Kojadinovic and Naveau [2015] found some evidence for change-points in both time series. Even more, if these maxima are regarded as (approximately) GEV distributed, the approach from the previous reference suggests that the change-point in the Nevers series is due to a change in the tail behavior and that of the Niort series is due to a change in location and scale.

The fact that a change in the tail behavior is present only in the first series is also confirmed by our analysis. The right-hand side of Figure 4.5 depicts the p-values of the L-test versus $k \in \{10, 11, \ldots, 100\}$ for both of the series. Small p-values for a wide range of k values between 20 and 50 suggest that there is indeed a change in the tail behavior of the first series. The graph for the second series does not show evidence for a change in the right-tail behavior. In addition, from the application of the usual Mann-Kendall test [Kendall, 1948; Yue et al., 2002] with resulting p-values of $p = 0.20$ for Nevers and $p = 0.03$ for Niort, respectively, we confirm that there is evidence for a monotonic change in the location of the second series.

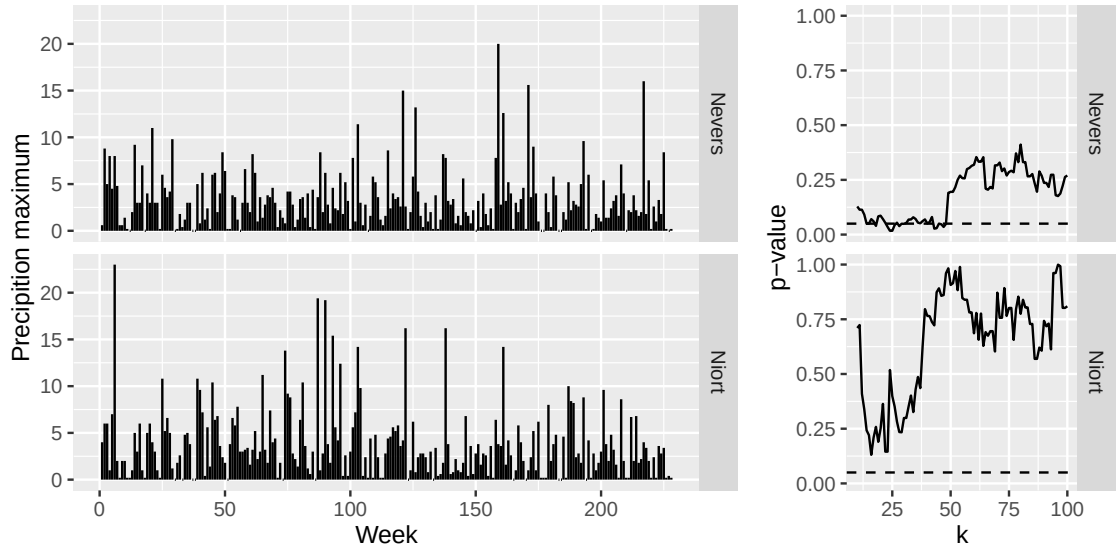

Figure 4.5: (Left) Weakly maxima of hourly precipitation during the fall season from 1993 to 2011 at the stations Nevers and Niort in France and (right) p-values of the L-test versus k between 10 and 100. The dashed line corresponds to a level of 5%.

For practitioners, the previous analysis raises the question of how to include the results into, say, the estimation of high quantiles. Let us start with the second series from Niort. Since there is no evidence against $\gamma(x) = \gamma_0$ for all x but instead some evidence against constant location, we slightly modify the extrapolation formula of Weissman [1978] and set

$$\hat{F}^{-1}(p|x) = u_{k,n}(x) \cdot \left(\frac{k}{n(1-p)}\right)^{\hat{\gamma}_0}, \; p > 1 - \frac{k}{n},$$

where $u_{k,n}(x)$ is defined in Section 4.2.1 and $\hat{\gamma}_0 = \frac{1}{k}\sum_{j=1}^{k} \log(Z_{k,j})$ is computed from relative excesses $Z_{k,j}$ above $u_{k,n}(X_{k,j})$. For instance, if we choose $k = k^* = 84$ from (4.6), we obtain $\hat{\gamma}_0 = 0.52$ with estimated 95%- confidence interval of $[0.41, 0.63]$.
We continue with the time series from Nevers. At first, since we have found some weak evidence against constant tail behavior, we might want to apply the L-estimator from Section 4.3.1, which, by Proposition 4.2, for $k = k^* = 97$ and for time axis rescaled on $x \in [-1, 1]$, gives us $\hat{\gamma}(x) = \hat{\eta}_0 + \hat{\eta}_1 x$ with

$$\hat{\boldsymbol{\eta}} = (0.56, 0.14)' \text{ and estimated variability } \widehat{\mathrm{Var}}\left[\sqrt{k}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})\right] = \begin{pmatrix} 0.34 & 0.16 \\ 0.16 & 0.99 \end{pmatrix}.$$

Note that the 95%-confidence interval for η_1 is rather wide with $[-0.04, 0.24]$. From our experience gained from simulation experiments reported in Section 4.4.3, we would suggest to follow a simpler approach. Because of the large uncertainty of $\hat{\eta}_1$ relative to that of

$\hat{\eta}_0$ and because of the very limited sample length of $n = 228$, it is expected that the overall estimation error decreases if we mistakenly set $\eta_1 = 0$, that is, if we estimate the tail behavior under the assumption $\gamma(x) = \gamma_0$. Applying the same estimator as for the Niort series with $k = k^* = 97$, we obtain $\hat{\gamma}_0 = 0.54$ with estimated 95%-confidence interval of $[0.43, 0.65]$.

4.5.2 Monthly maximal flows at the Mulde river basin in Germany

We analyze river flow series from 16 stations located at the Mulde basin in Germany. A convenient way to eliminate temporal dependence, which is strongly present in the raw data, is by considering only monthly maximal flows. For illustrative purpose, our longest time series of monthly maximal flows is depicted on the left-hand side of Figure 4.6. The series was observed at station Wechselburg1 from November 1909 to October 2012.
Besides the popular annual maxima approach, where only the largest out of twelve monthly maxima in each year is taken into account, there is an increasing interest in the hydrological literature on methods based on all values above some selected threshold [see, e.g., Cunnane, 1973; Madsen and Rosbjerg, 1997; Roth et al., 2016, and the references therein]. Practitioners usually choose a threshold such that, on average, more than one value per year is left for the estimation of the tail. The hope from this is an increase in estimation efficiency, compared to estimation based only on annual maxima. However, in the previous references it is assumed that the observations are identically distributed. Recently, Einmahl et al. [2016] showed consistency of classical tail estimators under slightly weaker assumptions called *heteroscedastic extremes*, but still they need that the tail behavior, i.e., the extreme value index γ is the same for all observations.
In analogy to the previous subsection, we first check whether the extreme value index of monthly maximal flows is constant over the whole observation period. We computed p-values of the test based on the L-estimator for all 16 time series, with (4.6) employed for a data-adaptive selection of k. Ignoring the multiple testing issue, weak evidence against stationary tail behavior is found only for the series from station Streckewalde with a p-value of 0.051. So, for the moment, it seems safe to assume that most of our series are stationary in their tail behavior.

Recall from the discussion in Section 1.1.2 and, in particular, from descriptive evidence in Figure 1.4 that a serious source of non-stationarity is due to seasonal variability within a year. In what follows, this is further investigated in terms of tail behavior:
We rearranged the monthly maximal flows according to their appearance within a hydrological year, which, for the series from Wechselburg1, is depicted on the right-hand side of Figure 4.6. In Germany the j-th hydrological year starts in the first day of November of the $(j-1)$-th calendar year and ends on the last day of October of the j-th calendar year. November first and October 31st correspond to day 1 and 366, respectively, on the x-axis of the right-hand side of Figure 4.6. Let $x_i \in \{1, 2, \ldots, 366\}$ denote the hydrological day of the monthly maximal flow Y_i with corresponding extreme value index $\gamma(x_i) > 0$, which is supposed to depend only on the hydrological day x_i, $i = 1, \ldots, n = 1236$. At first, a linear model $\gamma(x) = \eta_0 + \eta_1 x$ cannot be plausibly assumed because of the natural

period length of one year. One would rather expect a smooth model with the endpoint constraint $\gamma(1) = \gamma(366)$, which is not covered by the methodology considered here. Still, our L-test is able to detect non-constant tail behavior $\gamma(x) \neq \gamma_0$ if $\gamma(x)$ exhibits a monotonic behavior on a broad part of the regressor space. In fact, from descriptive data analysis, we suspect that $\gamma(x)$ increases from early winter to the middle of summer followed by an decrease in October. For empirical evidence, we computed the L-test with k selected by rule (4.6) for all 16 series. Again ignoring the multiple testing, evidence against stationary tail behavior is found for stations Niederstriegis1, Nossen1 and Borstendorf with p-values of about 0.001, 0.04 and 0.03, respectively. Weak evidence was found for station Goeritzhain with a p-value of about 0.09.
In Great Britain, Switzerland, the United States and some other countries from the northern hemisphere the hydrological year starts in the first October and ends in the last day of September. Interestingly enough, if we compute x_i according to this alternative definition and if we apply the same procedure on the modified series, we even obtain six p-values below the 5% and two others below the 10% significance level.

Since practitioners usually are interested in estimation and not in testing, the question is how to proceed with the analysis. Estimation under seasonal variability of river flows and related problems have already been addressed, for instance, in Schumann [2005], Strupczewski et al. [2012], Rulfov et al. [2016] and the references therein: Two or more groups of homogeneous observations, say, winter and summer flows, are identified. Afterwards, distributions are estimated under the i.i.d. assumption individually for each group. The final model is constructed assuming independence between the groups. It thus might be of interest to check whether there is evidence against stationary tail behavior during winter and summer, respectively.
The L-test with data-adaptive selection of k applied to flows from the hydrological sum-

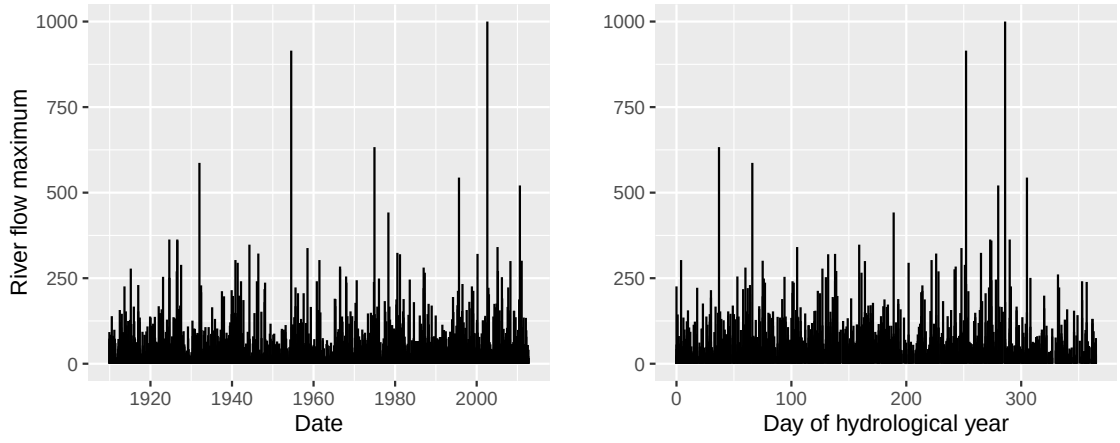

Figure 4.6: Monthly maximal flows observed at station Wechselburg starting from November 1909 till October 2012 in (left) chronological order and (right) ordered according to their day of appearance in the German hydrological year.

mer (May till October) does not provide evidence against stationary tail behavior during the summer. The application to flows from the hydrological winter (November till April) gives us only one p-value below 5% and another one below the 10% significance level. At first, this result sounds logical, since it is consistent with the idea that heterogeneity is mainly caused by the diversity of physiological causalities: Melting snow in the winter and heavy rainfalls in the summer time. We thus may expect that observations within each season are homogeneous in their tail behavior. But note also that the decreased evidence against stationary tail behavior might be also explained by a lack of power, since the sample is cut in half for testing on each season. E.g., for the Wechselbur1 series, instead of $n = 1236$ only $n/2 = 618$ observations are left for estimation of winter and summer distributions, respectively.

4.6 Conclusion and outlook

The analysis of trends in hydrological time series is motivated by a changing climate and by anthropogenic interference with nature, for instance, the dynamic process of urbanization during the past centuries. Little attention has been devoted to the analysis of trends in the tails in the applied literature, even though the primal interest lies on, say, high quantiles of distributions. Our work tries to fill this gap in case of heavy tails of Pareto-type. It turns out that satisfactory inference on non-constant tail-behavior is difficult under typical circumstances in hydrology, because of the rather limited observation lengths. In many of the scenarios considered in our simulations it is advisable to ignore trends in tail-behavior in order to reduce the dominating estimation variability at the cost of a rather small bias. For instance, we believe that estimation of annual maximal flow distributions based on the block maxima method should be carried out under the assumption of stationary tail behavior.
Our work might be extended to regional estimation under the assumption of regional heavy-tail homogeneity. For statistical inference in such a regional setting it is, in contrast to pure local estimation studied here, of practical importance to derive theory under semi-parametric assumptions in order to be able to estimate the dependence between local estimates. This indeed is a challenging problem left for future research.

Appendix A

Additional results and further technical details

A.1 Convergence of random elements in metric spaces

Let $(\mathbb{S}, d)$ denote a metric space and let us first assume that the space is complete and separable. The latter means that there exists a countable dense set in $\mathbb{S}$. Relevant examples are $\mathbb{S} = \mathbb{R}^d$ equipped with the euclidean distance and the space $\mathbb{S} = \mathbb{C}(I)$, $I \subset \mathbb{R}^p$, of continuous functions $f : I \to \mathbb{R}$ equipped with the uniform metric $d(f,g) = \|f - g\|_\infty = \sup_{t \in I} |f(t) - g(t)|$. Countable and dense subsets are given by $\mathbb{Q}^d$ and by the subset of continuous functions that are piecewise affine linear between grid points from $I \cap \mathbb{Q}^p$, respectively.
Let $\mathcal{B}$ be the Borel-σ-Algebra of $(\mathbb{S}, d)$ and let $(\Omega, \mathcal{A}, \mathbb{P})$ denote any probability space. A function $Y : \Omega \to \mathbb{S}$ is called measurable (or random element), if $Y^{-1}(B) \in \mathcal{A}$ for any $B \in \mathcal{B}$. In case of $\mathbb{S} = \mathbb{R}^d$ we also call Y a random vector or random variable if $d = 1$. For $\mathbb{S} = \mathbb{C}(I)$ we call Y a random function or stochastic process.

Definition A.1
Let $Y_n, Y : \Omega \to \mathbb{S}$, $n \in \mathbb{N}$, be random elements and let $\mathbb{P} \circ Y^{-1} : \mathcal{B} \to [0,1]$ denote the image measure with $\mathbb{P} \circ Y^{-1}(B) = \mathbb{P}(\{\omega \in \Omega : \ Y(\omega) \in B\})$. For $n \to \infty$, we define:

(i) (Almost sure convergence)
$Y_n \xrightarrow{a.s.} Y$, if $\mathbb{P}\left(\{\omega \in \Omega : \ Y_n(\omega) \to Y(\omega)\}\right) = 1$.

(ii) (Convergence in probability)
$Y_n \xrightarrow{\mathbb{P}} Y$, if $\mathbb{P}\left(\{\omega \in \Omega : \ d(Y_n(\omega), Y(\omega)) > \varepsilon\}\right) \to 0$ for any $\varepsilon > 0$.

(iii) (Convergence in distribution; weak convergence)
$Y_n \xrightarrow{D} Y$, if

$$\mathbb{E}\left[f(Y_n)\right] = \int_{\mathbb{S}} f \, d\mathbb{P} \circ Y_n^{-1} \longrightarrow \int_{\mathbb{S}} f \, d\mathbb{P} \circ Y^{-1} = \mathbb{E}\left[f(Y)\right] \tag{A.1}$$

for any $f \in C_b(\mathbb{S}) = \{h : \mathbb{S} \to \mathbb{R} : \ h \ \text{continuous and bounded}\}$.

Almost sure convergence always implies convergence in probability, which, in turn, always implies convergence in distribution,

$$Y_n \xrightarrow{a.s.} Y \ \Rightarrow \ Y_n \xrightarrow{\mathbb{P}} Y \ \Rightarrow \ Y_n \xrightarrow{D} Y.$$

Convergence in distribution is thus also called weak convergence.
If $Y_n \xrightarrow{D} Y$ on $\mathbb{S}$, then, for a second separable metric space $\mathbb{T}$, any continuous map $g : \mathbb{S} \to \mathbb{T}$, and from definition (A.1), we also obtain that $g(Y_n) \xrightarrow{D} g(Y)$ on $\mathbb{T}$, since $f \in C_b(\mathbb{T})$ implies that $f \circ g \in C_b(\mathbb{S})$. Even more, the following result called continuous mapping theorem states that we only need almost sure continuity with respect to the limiting image measure $\mathbb{P} \circ Y^{-1}$.

Theorem A.2 (Continuous mapping theorem)
Let $\xrightarrow{*}$ denote either $\xrightarrow{a.s.}$, $\xrightarrow{\mathbb{P}}$ or $\xrightarrow{D}$. Suppose that Y_n, Y, $n \in \mathbb{N}$, are random elements on the metric space $\mathbb{S}$ such that $Y_n \xrightarrow{*} Y$ for $n \to \infty$ and, for a second metric space $\mathbb{T}$, let $g : \mathbb{S} \to \mathbb{T}$ satisfy

$$\mathbb{P} \circ Y^{-1}\Big(\{s \in \mathbb{S} : \ g \ \textit{is continuous in } s\}\Big) = 1.$$

Then, for $n \to \infty$, we also have that $g(Y_n) \xrightarrow{*} g(Y)$ on $\mathbb{T}$.

In many situations we would like to deduce joint convergence from convergence of the components: Let ξ_n, ξ and η_n, η, $n \in \mathbb{N}$, denote two sequences of random elements on the metric spaces $(\mathbb{S}_1, d_1)$ and $(\mathbb{S}_2, d_2)$, respectively. We can always consider the tuples $Y_n = (\xi_n, \eta_n), Y = (\xi, \eta)$ as random elements on the product space

$$\mathbb{S} = \mathbb{S}_1 \times \mathbb{S}_2 = \{(s_1, s_2) : \ s_1 \in \mathbb{S}_1, s_2 \in \mathbb{S}_2\}$$

equipped with the metric $d = d_1 + d_2$. Then we have that marginal convergence of deterministic sequences $s_{1,n} \to s_1$ on $\mathbb{S}_1$ and $s_{2,n} \to s_2$ on $\mathbb{S}_2$ is equivalent to convergence of the tuples $(s_{n,1}, s_{n,2}) \to (s_1, s_2)$ on the product space $\mathbb{S}$. We thus immediately obtain that marginal almost sure convergence $\xi_n \xrightarrow{a.s.} \xi$ and $\eta_n \xrightarrow{a.s.} \eta$ implies joint convergence $Y_n \xrightarrow{a.s.} Y$, since almost sure events are closed under intersections. The same assertion also holds with convergence in probability, but not necessarily with weak convergence: For a counter example, let $\xi_n = \xi$ be standard normal and let $\eta_n = (-1)^n \xi$. Then we have $\eta_n \overset{D}{=} \eta$, which immediately implies $\eta_n \xrightarrow{D} \xi$. But we do not have joint weak convergence, since $\mathbb{E}(\xi_n \eta_n) = (-1)^n \not\to$ contradicts (A.1).
If ξ_n and η_n are independent for each $n \in \mathbb{N}$, then, of course, marginal weak convergence also implies joint weak convergence. Another frequently applied condition is as follows:

Theorem A.3 (Slutsky's lemma)
Suppose, for $n \to \infty$, we have that $\xi_n \xrightarrow{D} \xi$ on $\mathbb{S}_1$ and $\eta_n \xrightarrow{\mathbb{P}} a$ on $\mathbb{S}_2$, where $a \in \mathbb{S}_2$ is a

deterministic element. Then, for $n \to \infty$*, we also have that* $(\xi_n, \eta_n) \xrightarrow{D} (\xi, a)$ *on the product space* $\mathbb{S}_1 \times \mathbb{S}_2$.

Let us now turn to the space $\mathbb{S} = \ell^\infty(I) = \left\{f : I \to \mathbb{R} : \ \sup_{t\in I} |f(t)| < \infty\right\}$ of bounded real-valued functions on some index set I. We equip ℓ^∞ with the uniform metric d_∞. This space plays a central role in empirical process theory. E.g., $\ell^\infty([0,1])$ covers the well-known cádlág space $D([0,1])$ of right-continuous functions with existing left limits and with the empirical distribution F_n as a typical element in it. A difficulty is that $(\ell^\infty(I), d_\infty)$ is not separable [Billingsley, 1968, p. 216], which carries with measurability problems. Although it is natural to study the empirical process $\mathbb{F}_n = \sqrt{n}(F_n - F)$ as a random element in $\ell^\infty([0,1])$, it can be shown that $\mathbb{F}_n$ is not Borel-measurable with respect to the uniform metric [van der Vaart and Wellner, 1996, Chap. 1] and the definition of weak convergence in (A.1) does not make sense anymore for $Y_n = \mathbb{F}_n$.
One classical way to solve this problem consists of replacing d_∞ by Skorohod's d_0-metric, which makes $D[0,1]$ separable and complete [Billingsley, 1968]. An alternative, modern solution is due to Hoffman-Jørgensen [1994]: Suppose that the limiting element Y in (A.1) is measurable and replace the expectations on the left-hand side of (A.1) by outer expectations

$$\mathbb{E}^*[Z] = \inf\{\mathbb{E}[U] : \ U \geq Z, \ U \text{ measurable and } \mathbb{E}[U] < \infty\},$$

which are also defined for non-measurable maps $Z : \Omega \to \mathbb{R}$. It turns out that most common results can be extended to the outer expectation case, including the previous theorems, so, usually, there will be no practical difference when dealing with weak convergence of series of non-measurable random elements, provided the limit is measurable. For further details and extensions of Theorems A.2 and A.3 to $\mathbb{S} = \ell^\infty(I)$ we refer to Chapter 1 in van der Vaart and Wellner [1996].
The following theorem extends the classical delta method to arbitrary metric spaces [van der Vaart and Wellner, 1996, Sec. 3.9].

Theorem A.4 (Functional delta method)
Let $\mathbb{S}$ *and* $\mathbb{T}$ *be metric spaces,* $s \in \mathbb{S}_0 \subset \mathbb{S}$ *and, for* $n \to \infty$*, let* $g : \mathbb{S}_0 \to \mathbb{T}$ *be a map satisfying*

$$\frac{g(s + t_n h_n) - g(s)}{t_n} \longrightarrow g'_s(h) \tag{A.2}$$

for some linear map $g'_s : \mathbb{S} \to \mathbb{T}$ *(the derivative in s) and all sequences* $t_n \in \mathbb{R}$, $h_n \in \mathbb{S}$ *satisfying* $t_n \to 0$, $h_n \to h \in \mathbb{S}$ *and* $s + t_n h_n \in \mathbb{S}_0$*. Suppose further that* $r_n(Y_n - s) \xrightarrow{D} Z$ *on* $\mathbb{S}$*, where* $\mathbb{P}(Z \in \mathbb{S}_0) = 1$ *and* $r_n \to \infty$*. Then, for* $n \to \infty$*, we also have that*

$$r_n\Big(g(Y_n) - g(s)\Big) \xrightarrow{D} g'_s(Z)$$

on $\mathbb{T}$.

Property (A.2) is called Hadamard-differentiability of g in s tangentially to $\mathbb{S}_0$. In case of $\mathbb{S} = \mathbb{R}^k$ and $\mathbb{T} = \mathbb{R}^m$, Hadamard-differentiability is equivalent to usual differentiability

with linear function $g'_s = Dg_s$ given by the Jacobi-matrix of g in s.
Note that every function $g : J \to \mathbb{R}$, $J \subset \mathbb{R}$, can be extended to a map $\tilde{g} : f \mapsto g(f)$ defined on a subset $\mathbb{S}_0 \subset \ell^\infty(I)$, provided the range $f(I) \subset J$ for all $f \in \mathbb{S}_0$. The following result states that if g is differentiable with derivative g' and if J actually covers a little bit more than the range of all functions from $\mathbb{S}_0$, then we also have that $\tilde{g}$ is Hadamard-differentiable with derivative $\tilde{g}'_f(h) = g'(f) \cdot h$.

Theorem A.5 (Lemma 12.2 in Kosorok [2008])
Let $g : B \subset \mathbb{R} \to \mathbb{R}$ be differentiable with continuous derivative g' on all closed subsets of B and let

$$\mathbb{S}_0 = \left\{ f \in \ell^\infty(I) : \ f(I)^\delta \subset B \ \text{for some } \delta > 0 \right\} \subset \ell^\infty(I),$$

where $f(I)^\delta = \{y \in \mathbb{R} : \ |y - x| \leq \delta \text{ for some } x \in f(I)\}$ denotes the δ-enlargement of the range $f(I)$ of f. Then the map $\tilde{g} : \mathbb{S}_0 \to \ell^\infty(I)$, $f \mapsto g(f)$, is Hadamard-differentiable on $\mathbb{S}_0$ with derivative $\tilde{g}'_f(h) = g'(f) \cdot h$ in $f \in \mathbb{S}_0$.

In fact, following the arguments in the proof of Kosorok [2008], it is easily verified that Theorem A.5 generalizes to differentiable maps $g : B \subset \mathbb{R}^m \to \mathbb{R}^k$ for arbitrary $m, k \in \mathbb{N}$. Firstly, the main argument in the proof is that continuous functions on closed and bounded sets obtain their maximum. This is also the case for arbitrary $m \in \mathbb{N}$ and $k = 1$. Secondly, a function $g : B \subset \mathbb{R}^m \to \mathbb{R}^k$ is always of the form $g = (g_1, \ldots, g_k)$ with $g_j : B \to \mathbb{R}$ and convergence in (A.2) is equivalent to componentwise convergence. In other words, $\tilde{g}$ is Hadamard differentiable if and only if $\tilde{g}_1, \ldots, \tilde{g}_k$ are Hadamard differentiable. The Hadamard derivative $D\tilde{g}_f : \mathbb{S}_0 \to \ell^\infty(I)^k$ in $f \in \ell^\infty(I)^m$ and with $\mathbb{S}_0 \subset \ell^\infty(I)^m$ of $\tilde{g} : f \mapsto g(f)$ is given by $D\tilde{g}_f(h) = Dg_f \cdot h$, where Dg_x is the Jacobi matrix of g in $x \in \mathbb{R}^m$.

Theorem A.5 is used to prove Proposition 1.8 in the following section.

A.2 The madogram estimator - theory and praxis

The madogram estimator serves as a basic building block for the change-point test in Chapter 3. With this section, we would like to give a deeper insight into theory and finite-sample properties of this simple estimator. We start with a proof of Proposition 1.8.

Proof of Proposition 1.8. Recall the definitions of $S(t)$ and $\hat{S}_n(t)$ from (1.13) and (1.15), respectively. We first show weak convergence of the process $\sqrt{n}(\hat{S}_n - S)$ on $(\ell^\infty([0,1]), d_\infty)$: Note that $\int_0^1 \mathbb{1}(m \leq u)\, du = 1 - m$ and $C(u^{1-t}, u^t) = u^{A(t)}$ for $0 \leq u, m \leq 1$, which gives us

$$\hat{S}_n(t) = 1 - \int_0^1 C_n(u^{1-t}, u^t)\, du \quad \text{and} \quad S(t) = 1 - \int_0^1 C(u^{1-t}, u^t)\, du.$$

The continuity of $\dot{A}$ is sufficient for the existence and continuity of $\dot{C}_1$ and $\dot{C}_2$ on $(0,1) \times [0,1]$ and $[0,1] \times (0,1)$, respectively [Segers, 2012, Example 5.3], which, by Theorem A.8, is sufficient for the weak convergence of the empirical copula process $\sqrt{n}(C_n - C) =$

$\mathbb{C}_n^0 \xrightarrow{D} \mathbb{C}_C^0$ on $\left(\ell^\infty([0,1]^2), d_\infty\right)$. Thus, from the continuity of the map $f \mapsto \int_0^1 f(u^{1-t}, u^t)du$, $\left(\ell^\infty([0,1]^2), d_\infty\right) \to (\ell^\infty([0,1]), d_\infty)$, and from the continuous mapping theorem,

$$\sqrt{n}\left(\hat{S}_n - S\right) = -\left(\int_0^1 \mathbb{C}_n^0(u^{1-t}, u^t)\, du\right)_{t\in[0,1]} \xrightarrow{D} \left(\int_0^1 \mathbb{C}_C^0(u^{1-t}, u^t)\, du\right)_{t\in[0,1]} \tag{A.3}$$

on $(\ell^\infty([0,1]), d_\infty)$. The minus sign on the right-hand side of (A.3) can be dropped because the limit is a centered Gaussian process.
For the next step of the proof, recall that $A(t) = g(S(t))$, where the function $g : (0,1) \to (0,1)$, $s \mapsto s/(1-s)$, is differentiable with continuous derivative $g'(s) = 1/(1-s)^2$. Thanks to Theorem A.5, the extension $\tilde{g} : \mathbb{S} \to \ell^\infty([0,1])$, $f \mapsto g(f)$, defined on

$$\mathbb{S}_0 = \{f \in \ell^\infty([0,1]) : \ f(I)^\delta \subset (0,1) \text{ for some } \delta > 0\}$$

is Hadamard-differentiable with derivative $\tilde{g}'_s(h) = g'(s) \cdot h$, $h \in \ell^\infty([0,1])$. Recall that a Pickands dependence function A satisfies $1/2 \le A(t) \le 1$ for all $t \in [0,1]$. Since $S(t) = A(t)/(1+A(t))$, we also have that $1/4 \le S(t) \le 2/3$, which implies that $S \in \mathbb{S}_0$. In addition, from combinatoric arguments it is straightforward to show that

$$\frac{1}{6} < \frac{1}{n}\sum_{i=1}^{n} \frac{i^2}{(n+1)^2} \le \hat{S}_n(t) \le \frac{2}{n}\left(\frac{n(n+1)}{2} - \frac{n/2(n/2+1)}{2}\right) < \frac{3}{4},$$

which means that $\hat{S}_n$ takes its values in $\mathbb{S}_0$. Thanks to Theorem A.5 we conclude that $\tilde{g}$ is Hadamard-differentiable and we are thus in position to apply the functional delta method: We conclude that

$$\sqrt{n}(\hat{A}_n - A) = \sqrt{n}\left(\tilde{g}(\hat{S}_n) - \tilde{g}(S)\right) \xrightarrow{D} \tilde{g}'_S(Z) = g'(S) \cdot Z,$$

where Z is the weak limit on the right-hand side of (A.3). Because of $S = A/(1+A)$, we have that $g'(S) = (1+A)^2$. This completes the proof of Proposition 1.8. □

Remark A.6
It is straightforward to extend the madogram estimator to arbitrary dimensions $d \ge 2$, since we also have that

$$A(\mathbf{t}) = \frac{S(\mathbf{t})}{1 - S(\mathbf{t})} \quad \text{for} \quad S(\mathbf{t}) = \mathbb{E}\left[\max\left\{U_1^{1/t_1}, \ldots, U_d^{1/t_d}\right\}\right],$$

where $\mathbf{U} = (U_1, \ldots, U_d)'$ satisfies $C(\mathbf{u}) = \mathbb{P}(\mathbf{U} \le \mathbf{u})$. At the cost of a more complex notation but without any additional technical difficulties, it is easy to extend Proposition 1.8 to arbitrary dimensions $d \ge 2$.

Let us finish this section with a closer look at the finite-sample performance of the madogram estimator: Figure A.1 depicts mean squared errors $MSE(t) = \mathbb{E}\left[\left(\hat{A}(t) - A(t)\right)^2\right]$ of different estimators $\hat{A}(t)$ of the Pickands dependence function $A(t)$ versus $t \in [0,1]$ estimated from 4000 independent samples of size n and for four different scenarios according to (3.19). Endpoint-corrections are used, if necessary, so that the estimators satisfy

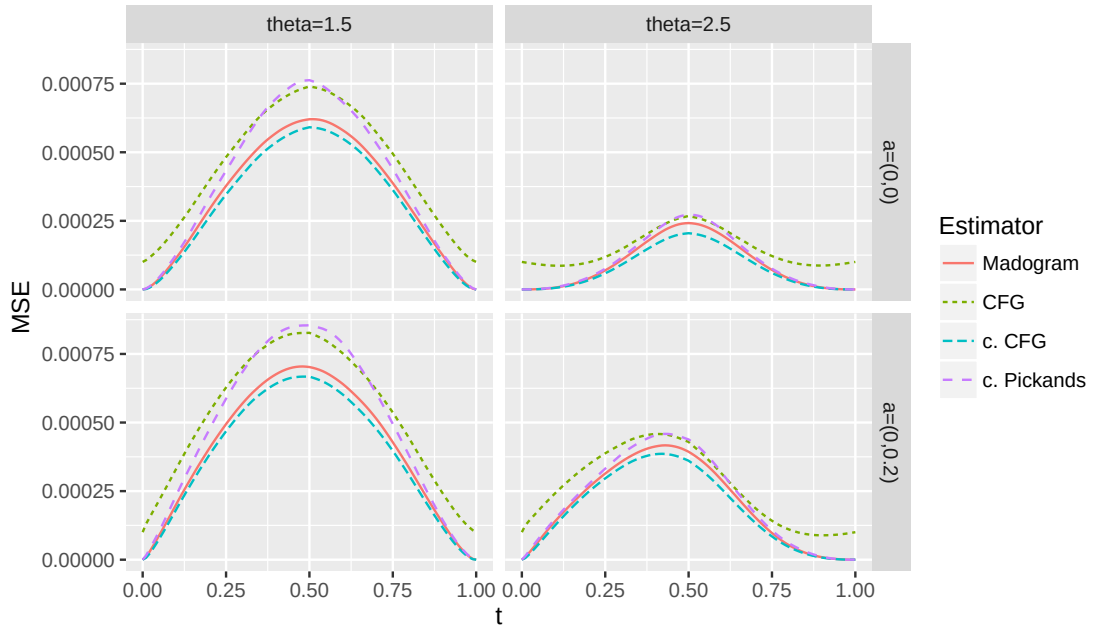

Figure A.1: Mean squared errors (MSE) of the madogram, CFG, endpoint-corrected CFG and endpoint-corrected Pickands estimators of $A(t)$ versus $t \in [0,1]$ computed from 4000 independent samples of size $n = 100$ drawn from a bivariate distribution with copula (3.19). We used four different scenarios with (left) $\vartheta = 1.5$, (right) $\vartheta = 2.5$, (top) $\mathbf{a} = (0,0)$ and (bottom) $\mathbf{a} = (0, 0.2)$.

the endpoint constraints $A(0) = A(1) = 1$ [Genest and Segers, 2009]. For instance, these modifications significantly improve the finite-sample performance of the Pickands and CFG estimators. Endpoint-correction of the madogram estimator is not necessary, since it fulfills the endpoint constraints by construction. From the results in Figure A.1 we find that the madogram estimator is pretty competitive in terms of MSE. It outperforms the uncorrected CFG and the corrected Pickands estimator. The difference in performance of the madogram estimator is small compared to the corrected CFG estimator, which is recommended by many authors as the best available estimator.

A.3 A max-type test of heavy-tail homogeneity

As an alternative to the heavy-tail ANOVA statistic $W_{\mathbf{k},\boldsymbol{\tau},n}$ defined in Proposition 2.6, practitioners might be also interested in a max-type test statistic (see the discussion in Section 2.4). Let

$$M_{\mathbf{k},\boldsymbol{\tau},n} = \left\| \hat{U}' \hat{\Sigma}^{-1/2} \frac{\sqrt{k_1}}{\hat{\gamma}} \left(\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\hat{\boldsymbol{w}}_{opt}) \mathbf{1} \right) \right\|_{\infty}^{2},$$

where $\hat{U} = (\hat{\mathbf{u}}_1 \dots \hat{\mathbf{u}}_{d-1}) \in \mathbb{R}^{d\times(d-1)}$ consists of pairwise orthogonal vectors $\mathbf{u}_1, \dots, \mathbf{u}_{d-1}$ with length 1, all orthogonal to $\hat{\mathbf{u}}_d = \hat{\Sigma}^{-1/2}\mathbf{1}$. These can be obtained from the Gram-Schmidt algorithm.

Theorem A.7
Under the assumptions of Proposition 2.6 and for $n \to \infty$, we have that

$$M_{\mathbf{k},\boldsymbol{\tau},n} \xrightarrow{D} \|\mathbf{Z}_{d-1}\|_\infty^2 = \max\{Z_1^2, \dots, Z_{d-1}^2\}, \tag{A.4}$$

where $\mathbf{Z}_{d-1} = (Z_1, \dots, Z_{d-1})'$ is a vector of independent $\mathcal{N}(0,1)$ distributed random variables.

Note that the limiting distribution function can be written as G^{d-1}, where G is the distribution function of a χ_1^2 distribution with 1 degree of freedom. Thus, the $(1-\alpha)$-quantile of the limiting distribution is given by the $(1-\alpha)^{1/(d-1)}$-quantile of the χ_1^2 distribution.

Proof. Recall first that (see the proof of Proposition 2.6)

$$\frac{\sqrt{k_1}}{\hat{\gamma}} \left(\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\hat{\boldsymbol{w}}_{opt})\mathbf{1}\right) \xrightarrow{D} A\Sigma^{1/2}\mathbf{Z}_d,$$

where $A = A_{\boldsymbol{w}_{opt}} = \mathbf{I}_d - \mathbf{1}\boldsymbol{w}_{opt}'$ and with identity matrix $\mathbf{I}_d \in \mathbb{R}^{d\times d}$. The proof follows from a singular value decomposition of the matrix $E = \Sigma^{-1/2}A\Sigma^{1/2}$:
Note that $E \cdot E' = B$, where $B = B_{\boldsymbol{w}_{opt}}$ is defined in the proof of Proposition 2.6. Thus we can decompose $E = UV'$, where the columns of $U \in \mathbb{R}^{d\times(d-1)}$ and $V \in \mathbb{R}^{d\times(d-1)}$ consist of $d-1$ orthonormal eigenvectors of EE' and $E'E$, respectively. From the proof of Proposition 2.6 and since we use $\boldsymbol{w} = \boldsymbol{w}_{opt}$, we also now that all the eigenvectors correspond to the eigenvalue of 1. Note also that $U'U = V'V = I_{d-1}$. Bringing all these things together completes the proof of (A.4). □

A.4 Empirical copula processes and multiplier bootstrap

Let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{id})'$, $i = 1, \dots, n$, denote a sequence of i.i.d. random vectors with distribution function

$$\mathbb{P}(\mathbf{Y}_i \le \mathbf{y}) = C\left(F_1(y_1), \dots, F_d(y_d)\right), \ \mathbf{y} = (y_1, \dots, y_d)' \in \mathbb{R}^d,$$

whose margins $F_1, \dots, F_d$ are continuous. Recall that $\mathbb{P}(\mathbf{U}_i \le \mathbf{u}) = C(\mathbf{u})$, where we call $\mathbf{U}_i = (F_1(Y_{i1}), \dots, F_d(Y_{id}))'$ the probability transform of $\mathbf{Y}_i$. For $1 \le k \le i \le \ell \le n$, we define sequential pseudo observations $\hat{\mathbf{U}}_{k:\ell,i} = \left(\hat{U}_{k:\ell,i1}, \dots, \hat{U}_{k:\ell,id}\right)'$ by

$$\hat{U}_{k:\ell,ij} = \frac{1}{\ell - k + 1} \sum_{m=k}^{\ell} \mathbb{1}\left(U_{mj} \le U_{ij}\right), \ j = 1, \dots, d$$

and, for $1 \leq k \leq \ell \leq n$, the sequential empirical copula by

$$C_{k:\ell}(\mathbf{u}) = \frac{1}{\ell - k + 1} \sum_{m=k}^{\ell} \mathbb{1}\left(\hat{\mathbf{U}}_{k:\ell,m} \leq \mathbf{u}\right), \ \mathbf{u} \in [0,1]^d.$$

Let $\Delta = \{(r,s) \in [0,1]^2 : \ s \geq r\}$. The process $\mathbb{C}_n$ defined on $\Delta \times [0,1]^d$ by

$$\mathbb{C}_n(r,s,\mathbf{u}) = \frac{\lfloor ns \rfloor - \lfloor nr \rfloor}{n^{1/2}} \left(C_{\lfloor nr \rfloor + 1:\lfloor ns \rfloor}(\mathbf{u}) - C(\mathbf{u})\right), \ (r,s,\mathbf{u}) \in \Delta \times [0,1]^d,$$

is called two-sided sequential empirical copula process. Bücher and Kojadinovic [2016] prove weak convergence of $\mathbb{C}_n$ towards a Gaussian process even under temporal dependence. In case of independence, as considered here, the weak convergence result can be restated as follows:

Theorem A.8 (Bücher and Kojadinovic [2016])
Suppose that all first order partial derivative $\dot{C}_j = dC/du_j$ of C exists and is continuous on $\{\mathbf{u} \in [0,1]^d : \ 0 < u_j < 1\}$, $j = 1, \ldots, d$. For $\mathbf{u} = (u_1, \ldots, u_d)'$, let $\mathbf{u}^{(j)}$ denote the vector of whose components are equal to one, except the jth which is equal to u_j. Then, for $n \to \infty$, we have that $\mathbb{C}_n \xrightarrow{D} \mathbb{C}_C$ on $\ell^\infty\left(\Delta \times [0,1]^d\right)$ equipped with the uniform metric, where $\mathbb{C}_C$ is defined through

$$\mathbb{C}_C(r,s,\mathbf{u}) = \{\mathbb{B}_C(s,\mathbf{u}) - \mathbb{B}_C(r,\mathbf{u})\} - \sum_{j=1}^{d} \dot{C}_j(\mathbf{u})\{\mathbb{B}_C(s,\mathbf{u}^{(j)}) - \mathbb{B}_C(r,\mathbf{u}^{(j)})\} \tag{A.5}$$

and where $\mathbb{B}_C$ is a centered Gaussian process on $\ell^\infty\left([0,1]^{d+1}\right)$ with covariance kernel

$$\mathbb{E}\{\mathbb{B}_C(r,\mathbf{u})\mathbb{B}_C(s,\mathbf{v})\} = (r \wedge s)\{C(u_1 \wedge v_1, \ldots, u_d \wedge v_d) - C(\mathbf{u})C(\mathbf{v})\}.$$

The usual empirical copula process and its weak limit are denoted by $\mathbb{C}_n^0 = \mathbb{C}_n(0,1,\cdot)$ and $\mathbb{C}_C^0 = \mathbb{C}_C(0,1,\cdot)$, respectively.

Note that the distribution of the limit $\mathbb{C}_C$ depends on the unknown copula C in a very complicated way, making it virtually impossible to calculate, say, critical values of test statistics based on $\mathbb{C}_n$. This problem may be solved by the multiplier bootstrap technique, initially proposed by Scaillet [2005] for $\mathbb{C}_C^0$ and extended by Bücher and Kojadinovic [2016] to the sequential case and serial dependence.

Theorem A.9 (Bücher and Kojadinovic [2016])
Let $\xi_i^{(b)}$, $i \geq 1$, $b = 1, \ldots, B$, denote independent standard normal distributed variables, independent of (X_i, Y_i), $i \geq 1$. Let, for $b = 1, \ldots, B$,

$$\check{\mathbb{B}}_n^{(b)}(s,\mathbf{u}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \xi_i^{(b)} \left\{\mathbb{1}\left(\hat{\mathbf{U}}_{1:\lfloor ns \rfloor} \leq \mathbf{u}\right) - C_{1:\lfloor ns \rfloor}(\mathbf{u})\right\}$$

and let $\dot{C}_{j,n}$ denote an estimator of $\dot{C}_j$ satisfying

$$\sup_{\mathbf{u}\in[0,1]^d, u_j\in[\delta,1-\delta]} \left|\dot{C}_{j,n}(\mathbf{u}) - \dot{C}_j(\mathbf{u})\right| \xrightarrow{\mathbb{P}} 0 \tag{A.6}$$

for some $\delta \in (0, 1/2)$, $j = 1, \ldots, d$. Define $\check{\mathbb{C}}_n^{(b)}$ in analogy to (A.5) *with $\mathbb{B}_C$ and $\dot{C}_j$ replaced by $\check{\mathbb{B}}_n^{(b)}$ and $\dot{C}_{j,n}$. Then, under the assumptions of Theorem A.8 and for $n \to \infty$, we have that*

$$\left(\mathbb{C}_n, \check{\mathbb{C}}_n^{(1)}, \ldots, \check{\mathbb{C}}_n^{(B)}\right) \xrightarrow{D} \left(\mathbb{C}_C, \mathbb{C}_C^{(1)}, \ldots, \mathbb{C}_C^{(B)}\right)$$

in $\left(\ell^\infty\left(\Delta \times [0,1]^d\right)\right)^{B+1}$, where $\mathbb{C}_C^{(1)}, \ldots, \mathbb{C}_C^{(B)}$ denote independent copies of $\mathbb{C}_C$.

In analogy to $\mathbb{C}_n^0$, we set $\check{\mathbb{C}}_n^{0,(b)} = \check{\mathbb{C}}_n^{(b)}(0, 1, \cdot)$.

A.5 Change-point test statistic for $d \geq 2$

In Sections 3.2.1 and 3.2.2, we restricted ourselves to the case $d = 2$. Results for arbitrary dimension $d \geq 2$ can be established at the cost of a more complex notation but without significant additional mathematical difficulties. We give the main steps of the generalization hereafter. Let $\mathbf{Y} = (Y_1, \ldots, Y_d)$ be a random vector with distribution function and extreme value copula of the form (1.1) and (1.7), respectively, and suppose that A is continuously differentiable on the interior of $\mathcal{S}_{d-1}$ with partial derivatives $\dot{A}_j(\mathbf{t}) = \partial A(\mathbf{t})/\partial t_j$, $j = 2, \ldots, d$, $\mathbf{t} = (t_2, \ldots, t_d)' \in \mathcal{S}_{d-1}$. With the notation $U_j = F_j(Y_j)$, $j = 1, \ldots, d$, and $t_1 = t_1(\mathbf{t}) = 1 - \sum_{j=2}^d t_j$, $\mathbf{t} \in \mathcal{S}_{d-1}$, we have, just as for $d = 2$,

$$A(\mathbf{t}) = \frac{S(\mathbf{t})}{1 - S(\mathbf{t})} \quad \text{and} \quad S(\mathbf{t}) = \mathbb{E}\left[\max\left\{U_1^{1/t_1}, \ldots, U_d^{1/t_d}\right\}\right],$$

with the convention that $u^{1/0} = 0$ for all $u \in (0, 1)$.
Let $\mathbf{Y}_i$, $i = 1, \ldots, n$, be independent copies of $\mathbf{Y}$ and let $\hat{\mathbf{U}}_{k:\ell,i} = (\hat{U}_{k:\ell,i1}, \ldots, \hat{U}_{k:\ell,id})'$ be d-variate generalizations of the subsample pseudo-observations in (3.6). We define a CUSUM-type process $\mathbb{D}_n$ on $\ell^\infty([0, 1] \times \mathcal{S}_{d-1})$ by

$$\mathbb{D}_n(s, \mathbf{t}) = \frac{\lfloor ns \rfloor (n - \lfloor ns \rfloor)}{n^{3/2}} \left\{\hat{A}_{1:\lfloor ns \rfloor}(\mathbf{t}) - \hat{A}_{\lfloor ns \rfloor + 1:n}(\mathbf{t})\right\},$$

where, for $1 \leq k \leq \ell \leq n$, $\hat{A}_{k:\ell}(\mathbf{t}) = \hat{S}_{k:\ell}(\mathbf{t})/\{1 - \hat{S}_{k:\ell}(\mathbf{t})\}$, and

$$\hat{S}_{k:\ell}(\mathbf{t}) = \frac{1}{\ell - k + 1} \sum_{i=k}^{\ell} \max\left\{\hat{U}_{k:\ell,i1}^{1/t_1}, \ldots, \hat{U}_{k:\ell,id}^{1/t_d}\right\}$$

with the convention that $\hat{S}_{k:\ell} = 0$ if $k > \ell$.
Let us introduce some additional notation. For any $y \in [0, 1]$ and $\mathbf{t} \in \mathcal{S}_{d-1}$, we define $\mathbf{y}^\mathbf{t}$ to be the vector $(y^{t_1}, \ldots, y^{t_d}) \in [0, 1]^d$ with the convention that $0^0 = 1$. Furthermore, for any $\mathbf{u} \in [0, 1]^d$ and any $j = 1, \ldots, d$, $\mathbf{u}^{(j)}$ denotes the vector of $[0, 1]^d$ whose components are all equal to one, except the jth which is equal to u_j.

Proposition A.10

Suppose that all of the above conditions are met. Then, for $n \to \infty$, we have that $\mathbb{D}_n \xrightarrow{D} \mathbb{D}_C$ on $\ell^\infty([0,1] \times \mathcal{S}_{d-1})$, where

$$\mathbb{D}_C(s, \mathbf{t}) = \{1 + A(\mathbf{t})\}^2 \times \int_0^1 s\mathbb{C}_C(s, 1, \mathbf{y}^\mathbf{t}) - (1-s)\mathbb{C}_C(0, s, \mathbf{y}^\mathbf{t})\, dy$$

and where $\mathbb{C}_C$ is defined in Theorem A.8.

The proof is almost identical to that of Proposition 3.1. For a corresponding bootstrap approximation of the limit $\mathbb{D}_C$, let $\xi_i^{(b)}$, $i = 1, \ldots, n$, $b = 1, \ldots, B$, be i.i.d. standard normal multipliers. Furthermore, from (1.7), we have that, for any $y \in (0,1)$ and $\mathbf{t} \in \mathcal{S}_{d-1}$,

$$\dot{C}_j(\mathbf{y}^\mathbf{t}) = \begin{cases} \mathbf{y}^{A(\mathbf{t})-t_1} \left\{ A(\mathbf{t}) - \sum_{k=2}^d t_k \dot{A}_k(\mathbf{t}) \right\}, & j = 1, \\ \mathbf{y}^{A(\mathbf{t})-t_j} \left\{ A(\mathbf{t}) + \dot{A}_j(\mathbf{t}) - \sum_{k=2}^d t_k \dot{A}_k(\mathbf{t}) \right\}, & j = 2, \ldots, d. \end{cases}$$

The above quantities can be estimated consistently by plugging in subsample estimators of A and $\dot{A}_j = dA/dt_j$, $j = 2, \ldots, d$, respectively, namely $\hat{A}_{k:\ell}$ and

$$\dot{A}_{j,k:\ell,n}(\mathbf{t}) = \frac{1}{2h_n} \left\{ \hat{A}_{k:\ell}(\mathbf{t} + h_n \mathbf{e}_j) - \hat{A}_{k:\ell}(\mathbf{t} - h_n \mathbf{e}_j) \right\}, \quad j = 2, \ldots, d,$$

with $\mathbf{t} \pm h_n \mathbf{e_j} = (t_2, \ldots, t_{j-1}, t_j \pm h_n, t_{j+1}, \ldots, t_d)'$ and a sequence $h_n \downarrow 0$ such that $\inf_{n \geq 1} h_n \sqrt{n} > 0$ (boundary effects can be dealt with by generalizing the approach adopted below (3.14)). Then, analogously to the bivariate case, we define

$$\check{\mathbb{D}}_n^{(b)}(s, \mathbf{t}) = \{1 + \hat{A}_{1:n}(\mathbf{t})\}^2 \times \left\{ \frac{\lfloor ns \rfloor}{n^{3/2}} \sum_{i=\lfloor ns \rfloor + 1}^{n} \xi_i^{(b)} \hat{w}_{\lfloor ns \rfloor + 1:n,i}(\mathbf{t}) - \frac{n - \lfloor ns \rfloor}{n^{3/2}} \sum_{i=1}^{\lfloor ns \rfloor} \xi_i^{(b)} \hat{w}_{1:\lfloor ns \rfloor,i}(\mathbf{t}) \right\},$$

where, for $1 \leq k \leq \ell \leq n$,

$$\hat{w}_{k:\ell,i}(\mathbf{t}) = \bar{m}_{k:\ell}(\mathbf{t}) - \hat{m}_{k:\ell,i}(\mathbf{t}) + \sum_{j=1}^d \frac{(\hat{u}_{k:\ell,ij}(\mathbf{t}) - \bar{u}_{k:\ell,j}(\mathbf{t}))\hat{a}_{k:\ell,j}(\mathbf{t})}{\hat{b}_{k:\ell,j}(\mathbf{t})},$$

with $\bar{m}_{k:\ell}$ and $\bar{u}_{k:\ell,j}$ denoting the arithmetic mean over $i = k, \ldots, \ell$ of

$$\hat{m}_{k:\ell,i}(\mathbf{t}) = \max\left(\hat{\mathbf{U}}_{k:\ell,i}^{1/\mathbf{t}}\right) \text{ and } \hat{u}_{k:\ell,ij}(\mathbf{t}) = \hat{U}_{k:\ell,ij}^{\hat{b}_{k:\ell,j}/t_j},$$

and where

$$\hat{a}_{k:\ell,j}(\mathbf{t}) = \begin{cases} \hat{A}_{k:\ell}(\mathbf{t}) - \sum_{j'=2}^d t_{j'} \dot{A}_{j',k:\ell,n}(\mathbf{t}), & j = 1, \\ \hat{A}_{k:\ell}(\mathbf{t}) + \dot{A}_{j,k:\ell,n}(\mathbf{t}) - \sum_{j'=2}^d t_{j'} \dot{A}_{j',k:\ell,n}(\mathbf{t}), & j = 2, \ldots, d, \end{cases}$$
$$\hat{b}_{k:\ell,j}(\mathbf{t}) = \hat{A}_{k:\ell}(\mathbf{t}) + 1 - t_j.$$

Test statistics and corresponding multiplier bootstrap replicates can be defined analogously to Section 3.2.1, as functionals of $\mathbb{D}_n$ and $\check{\mathbb{D}}_n^{(b)}$, $b = 1, \ldots, B$, respectively. In addition, generalizations adapted to known breaks in the margins can be obtained by computing pseudo observations from the subsamples determined by the marginal change-points, as explained in Section 3.2.2. We omit the details for the sake of brevity.

A.6 Quantile regression process

Let Y denote a random variable called response and $\mathbf{X} = (1, X_1, \ldots, X_d)'$ a random vector called regressor with support covered by a compact set $\mathcal{X} \subset \mathbb{R}^{d+1}$. Throughout this section we suppose that the conditional distribution $F(y|\mathbf{x}) = P(Y \leq y|\ \mathbf{X} = \mathbf{x})$ of Y given $\mathbf{X} = \mathbf{x}$ satisfies

$$F^{-1}(p|\ \mathbf{x}) = \inf\{y :\ F(y|\mathbf{x}) \geq p\} = \mathbf{x}'\boldsymbol{\beta}_p \tag{A.7}$$

for all $\mathbf{x} \in \mathcal{X}$, probabilities $p \in I \subset [\varepsilon, 1-\varepsilon]$ and an unknown vector-valued function $p \mapsto \boldsymbol{\beta}_p$, $p \in I$, with $\boldsymbol{\beta}_p \in \mathbb{R}^{d+1}$ called p-th regression quantile [Koenker and Bassett, 1978]. The left-hand side of (A.7) is called generalized inverse or quantile of $F(\cdot|\mathbf{x})$ in $p \in I$. It coincides with the usual inverse of a function, provided the inverse exists. Theoretical aspects and many applications of linear quantile regression are presented in Koenker [2005].
Let $(Y_i, \mathbf{X}_i)$, $i = 1, \ldots, n$, denote independent copies of $(Y, \mathbf{X})$. Estimator

$$\hat{\boldsymbol{\beta}}_p = \underset{\mathbf{b} \in \mathbb{R}^{d+1}}{\arg\min} \sum_{i=1}^{n} \rho_p \left(Y_i - \mathbf{X}_i' \cdot \mathbf{b}\right)$$

with $\rho_p(y) = y \cdot (p - \mathbb{1}_{\{y \leq 0\}})$ is called empirical regression quantile. The following result establishes asymptotic normality of $\sqrt{n}\left(\hat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_p\right)$ uniformly in $p \in I$, i.e., in $\left(\ell^\infty(I)\right)^{d+1}$.

Theorem A.11 (Angrist et al. [2006])
Suppose that, uniformly in $\mathbf{x} \in \mathcal{X}$, the conditional density $f(y|\mathbf{x})$ exists, is bounded and uniformly continuous in y. Suppose further that $\mathbb{E}\|\mathbf{X}\|^{2+\delta} < \infty$ for some $\delta > 0$ and that

$$J = \mathbb{E}\left[\mathbf{X}\mathbf{X}'\right] \quad \textit{and} \quad H_p = \mathbb{E}\left[\mathbf{X}\mathbf{X}' \cdot f(F^{-1}(p|\mathbf{X})|\mathbf{X})\right] \tag{A.8}$$

exist with H_p positive definite for all $p \in I$. Then, for $n \to \infty$, we have that

$$\left(H_p \sqrt{n}\left(\hat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_p\right)\right)_{p \in I} \xrightarrow{D} \mathbb{Z} \tag{A.9}$$

in $\left(\ell^\infty(I)\right))^{d+1}$, where $\mathbb{Z}$ is a centered Gaussian process with $\mathbb{E}[\mathbb{Z}(p)\mathbb{Z}(q)'] = (p \wedge q - p \cdot q) \cdot J$.

The previous result allows us to estimate the joint distribution of several empirical regression quantiles. Let $\mathbf{p} = \{p_1, \ldots, p_\ell\} \subset I$ denote a set of probabilities. Then, for $n \to \infty$, we immediately obtain that

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{p_1} - \boldsymbol{\beta}_{p_1}, \ldots, \hat{\boldsymbol{\beta}}_{p_\ell} - \boldsymbol{\beta}_{p_\ell}\right)' \xrightarrow{D} \mathcal{N}\left(0, \Sigma_{\mathbf{p}}\right),$$

where $\Sigma_{\mathbf{p}}$ is defined piecewise through

$$\lim_{n \to \infty} \mathrm{Cov}\left[\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{p_i} - \boldsymbol{\beta}_{p_i}\right),\ \sqrt{n}\left(\hat{\boldsymbol{\beta}}_{p_j} - \boldsymbol{\beta}_{p_j}\right)\right] = (p_i \wedge p_j - p_i \cdot p_j) \cdot H_{p_i}^{-1} J H_{p_j}^{-1}.$$

This result is used to prove Proposition 4.2.

A.7 Conditional heavy-tail behavior - competing methods

A.7.1 Tail index regression (TIR) by Wang and Tsai [2009]

Wang and Tsai [2009] study model (4.1) with $\alpha(\mathbf{x}) = 1/\gamma(\mathbf{x}) = \exp(\mathbf{x}'\theta)$ for some unknown parameter vector $\theta \in \mathbb{R}^{d+1}$. They propose the estimator

$$\hat{\theta}_{u_n} = \underset{\theta \in \mathbb{R}^{d+1}}{\arg\min} \sum_{i=1}^{n} \left[\exp(\mathbf{X}_i'\theta) \cdot \log(Y_i/u_n) - \mathbf{X}_i'\theta\right] \cdot \mathbf{1}(Y_i > u_n) \tag{A.10}$$

with regressor independent threshold $u_n \to \infty$ for $n \to \infty$. (A.10) can be viewed as an approximate maximum likelihood approach based on the weak approximation of $\log(Y/u_n)$ given $\mathbf{X} = \mathbf{x}$ and $Y > u_n$ to an exponential distribution with mean $1/\alpha(\mathbf{x})$. Let $k = \sum_{i=1}^{n} \mathbf{1}(Y_i > u_n)$ be the effective sample size in (A.10) and $\hat{\Sigma}_{u_n} = \frac{1}{k}\sum_{i=1}^{n} \mathbf{X}_i\mathbf{X}_i'\mathbf{1}(Y_i > u_n)$. Under certain technical assumptions, Wang and Tsai [2009] prove

$$\sqrt{k} \cdot \hat{\Sigma}_{u_n}^{1/2} \cdot \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{D} \mathcal{N}\left(\mathbf{h}, I_{d+1}\right) \tag{A.11}$$

for some vector $\mathbf{h}$ and $(d+1)$-dimensional identity matrix I_{d+1}. The estimation of the bias $\mathbf{h}$ requires detailed information on the tail, which is hardly available and thus set to zero in applications.
However, Wang and Tsai [2009] do not consider regressor dependent thresholds u_n like in Section 4.2.1, which in practice is important to account for regression effects in e.g. the center of the distribution. In order to reduce this problem, we suggest to apply their estimation procedure on the sample $(Z_{k,j}, \mathbf{X}_{k,j})$, $j = 1, \ldots, k$, as given in Section 4.2.1. That is, replace $\hat{\theta}_{u_n}$ by

$$\hat{\boldsymbol{\theta}}_{k,n}^{TIR} = \underset{\boldsymbol{\theta} \in \mathbb{R}^{d+1}}{\arg\min} \sum_{j=1}^{k} \left[\exp(\mathbf{X}_{k,j}'\boldsymbol{\theta}) \cdot \log(Z_{k,j}) - \mathbf{X}_{k,j}'\boldsymbol{\theta}\right]$$

and $\hat{\Sigma}_{u_n}$ by $\hat{\Sigma}_{k,n} = \frac{1}{k}\sum_{j=1}^{k} \mathbf{X}_{k,j}\mathbf{X}_{k,j}'$.

A.7.2 Three-stage procedure by Wang and Li [2013]

An alternative regression approach focusing on high conditional quantiles $F_Y^{-1}(p|\,\mathbf{x})$, $p \in [1-\varepsilon, 1)$, for some small number $\varepsilon > 0$ is proposed in Wang and Li [2013]. Their method is based on the assumption that

$$F_{g_\lambda(Y)}^{-1}(p|\,\mathbf{x}) = \mathbf{x}'\beta_p$$

holds for some $\lambda \in \mathbb{R}$, Box-Cox transformation g_λ, regression quantiles $\beta_p \in \mathbb{R}^{d+1}$ and all $p \in [1-\varepsilon, 1)$. They propose an estimator of $\gamma(\mathbf{x})$ based on a three-stage procedure:

(i) Set $p = p_{k,n} = \frac{n-k}{n+1}$ and compute $\hat{\lambda}$ as in Section 4.2.1.

(ii) Let $p_{n-j,n} = \frac{j}{n+1}$ for $j = 1, \dots, m$ with $m = n - \lfloor n^\eta \rfloor$ and $\eta = 0.1$. For $j = 1, \dots, m$, estimate $F_Y^{-1}(p_{n-j,n} | \mathbf{x})$ by the right hand side of (4.4) with $g = g_{\hat{\lambda}}$ and $p = p_{n-j,n}$. Denote these estimates by $\hat{q}_j(\mathbf{x})$, $j = 1, \dots, m$. If $\hat{q}_j(\mathbf{x})$ is not increasing in j, apply the rearrangement procedure of Chernozhukov et al. [2010].

(iii) For some integer $k < m$, estimate $\gamma(\mathbf{x})$ by

$$\hat{\gamma}_{k,n}(\mathbf{x}) = \frac{1}{k - \lfloor n^\eta \rfloor} \sum_{j=\lfloor n^\eta \rfloor}^{k} \log(\hat{q}_{n-j}) - \log(\hat{q}_{n-k}).$$

Thus $\hat{\gamma}_{k,n}(\mathbf{x})$ is Hill's estimator [Hill, 1975] applied to the sample of $\hat{q}(\mathbf{x})$ values, which can be seen as pseudo observations from $F_Y(\,\cdot \mid \mathbf{x})$. Wang and Li [2013] also propose a test statistic

$$T_n = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{\gamma}_{k,n}(\mathbf{X}_i) - \hat{\gamma}_p \right)^2, \; \hat{\gamma}_p = \frac{1}{n} \sum_{i=1}^{n} \hat{\gamma}(\mathbf{X}_i), \tag{A.12}$$

as a test for hypothesis $\mathcal{H}_{0,tail}$ in (4.2). If $\mathcal{H}_{0,tail}$, $E(\mathbf{X}) = (1, 0, \dots, 0)' \in \mathbb{R}^{d+1}$ and either $\gamma^*(\mathbf{x}) = 0$ or a certain homogeneity assumption are met, Wang and Li [2013] show under additional technical assumptions that $kT_n \xrightarrow{D} \gamma^2 \chi^2_d$ holds. They also derive the limiting distribution under heterogeneity, which in practice involves the estimation of additional parameters. For more details we refer to Wang and Li [2013, Th. 3.3 and Cor. 3.1].

Bibliography

Angrist, J., Chernozhukov, V., and Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the u.s. wage structure. *Econometrica*, 74(2):539–563.

Arias-Castro, E., Cands, E. J., and Plan, Y. (2011). Global testing under sparse alternatives: Anova, multiple comparisons and the higher criticism. *Annals of Statistics*, 39(5):2533–2556.

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.

Bickel, P. J. and Lehmann, E. L. (1975). Descriptive statistics for nonparametric models ii. location. *Annals of Statistics*, 3(5):1045–1069.

Billingsley, P. (1968). *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, first edition.

Bücher, A., Dette, H., and Volgushev, S. (2011). New estimators of the pickands dependence function and a test for extreme-value dependence. *Annals of Statistics*, 39(4):1963–2006.

Bücher, A., Kinsvater, P., and Kojadinovic, I. (2015). Detecting breaks in the dependence of multivariate extreme-value distributions. *ArXiv:1505.00954*.

Bücher, A. and Kojadinovic, I. (2014). An overview of nonparametric tests of extreme-value dependence and of some related statistical procedures. *ArXiv:1410.6784*.

Bücher, A. and Kojadinovic, I. (2016). A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. *Bernoulli*, 22(2):927–968.

Bücher, A., Kojadinovic, I., Rohmer, T., and Segers, J. (2014). Detecting changes in cross-sectional dependence in multivariate time series. *Journal of Multivariate Analysis*, 132(0):111 – 128.

Bücher, A. and Segers, J. (2016). On the maximum likelihood estimator for the Generalized Extreme-Value distribution. *ArXiv e-prints*.

Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):207–222.

Chebana, F., Ouarda, T. B., and Duong, T. C. (2013). Testing for multivariate trends in hydrologic frequency analysis. *Journal of Hydrology*, 486:519 – 530.

Chernozhukov, V., Fernndez-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840.

Cormier, E., Genest, C., and Nešlehová, J. G. (2014). Using b-splines for nonparametric inference on bivariate extreme-value copulas. *Extremes*, 17(4):633–659.

Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, UK.

Cunderlik, J. M. and Burn, D. H. (2003). Non-stationary pooled flood frequency analysis. *Journal of Hydrology*, 276(14):210 – 223.

Cunnane, C. (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology*, 18(3):257 – 271.

Dalrymple, T. (1960). Flood-frequency analyses, manual of hydrology: Part 3. Technical report, USGPO.

Davidson, J. (1994). *Stochastic Limit Theory : An Introduction for Econometricians*. Oxford University Press.

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

de Haan, L., Tank, A., and Neves, C. (2015). On tail trend detection: modeling relative risk. *Extremes*, 18(2):141–178.

Dehling, H., Vogel, D., Wendler, M., and Wied, D. (2014). Testing for changes in the rank correlation of time series. *ArXiv:1203.4871*.

Dematteo, A. and Clémençon, S. (2016). On tail index estimation based on multivariate data. *Journal of Nonparametric Statistics*, 28(1):152–176.

Dierckx, G. (2011). Trends and change points in the tail behaviour of a heavy tailed distribution. In *Proc. 58th World Statistical Congress (ISI2011), Dublin*, pages 290–299.

Dierckx, G. and Teugels, J. L. (2010). Change point analysis of extreme values. *Environmetrics*, 21(7-8):661–686.

Dixon, M. J., Tawn, J. A., and Vassie, J. M. (1998). Spatial modelling of extreme sea-levels. *Environmetrics*, 9(3):283–301.

Drees, H., de Haan, L., and Resnick, S. (2000). How to make a hill plot. *Annals of Statistics*, 28(1):254–274.

Dupuis, D. J., Sun, Y., and Wang, H. J. (2015). Detecting change-points in extremes. *Statistics and Its Interface*, 8(1):19–31.

Durante, F. and Salvadori, G. (2010). On the construction of multivariate extreme value models via copulas. *Environmetrics*, 21(2):143–161.

Durbin, J. and Knott, M. (1972). Components of cramer-von mises statistics. i. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):290–307.

Einmahl, J., Dehaan, L., and Huang, X. (1993). Estimating a multidimensional extreme-value distribution. *Journal of Multivariate Analysis*, 47(1):35 – 47.

Einmahl, J. H. J., de Haan, L., and Zhou, C. (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1):31–51.

El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R., and Bobe, B. (2007). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, 43(3). W03410.

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

Gardes, L. and Girard, S. (2010). Conditional extremes from heavy-tailed distributions: an application to the estimation of extreme rainfall return levels. *Extremes*, 13(2):177–204.

Genest, C., Kojadinovic, I., Nešlehová, J., and Yan, J. (2011). A goodness-of-fit test for bivariate extreme-value copulas. *Bernoulli*, 17(1):253–275.

Genest, C. and Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *Annals of Statistics*, 37(5B):2990–3022.

Gomes, M. I. and Pestana, D. (2007). A sturdy reduced-bias extreme quantile (var) estimator. *Journal of the American Statistical Association*, 102(477):280–292.

Gudendorf, G. and Segers, J. (2011). Nonparametric estimation of an extreme-value copula in arbitrary dimensions. *Journal of Multivariate Analysis*, 102(1):37 – 47.

Gumbel, E. J. (1958). *Statistics of extremes*. Columbia University Press, New York.

Hall, P. and Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. *Annals of Statistics*, 13(1):331–341.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5):1163–1174.

Hofert, M., Kojadinovic, I., Mächler, M., and Yan, J. (2015). *copula: Multivariate Dependence with Copulas*. R package version 0.999-13.

Hoffman-Jørgensen, J. (1994). *Probability with a View Towards Statistics*. Chapman and Hall, New York.

Hosking, J. R. M. and Wallis, J. R. (2005). *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261.

Jarušková, D. and Rencová, M. (2008). Analysis of annual maximal and minimal temperatures for some european cities by change point methods. *Environmetrics*, 19(3):221–233.

Katz, R. and Brown, B. (1992). Extreme events in a changing climate: Variability is more important than averages. *Climatic Change*, 21(3):289–302.

Kendall, M. G. (1948). *Rank correlation methods.* Charles Griffin London.

Khoudraji, A. (1995). *Contributions à l'étude des copules et à la modélisation des valeurs extrêmes bivariées*. PhD thesis, Université Laval, Québec, Canada.

Kim, M. and Lee, S. (2009). Test for tail index change in stationary time series with pareto-type marginal distribution. *Bernoulli*, 15(2):325–356.

Kinsvater, P., Fried, R., and Lilienthal, J. (2016). Regional extreme value index estimation and a test of tail homogeneity. *Environmetrics*, 27(2):103–115.

Koenker, R. (2005). *Quantile regression*. Cambridge university press.

Koenker, R. and Bassett, Gilbert, J. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Kojadinovic, I. (2015). *npcp: Some nonparametric tests for change-point detection in (multivariate) observations*. R package version 0.1-6.

Kojadinovic, I. and Naveau, P. (2015). Nonparametric tests for change-point detection in the distribution of block maxima based on probability weighted moments. *ArXiv:1507.06121*.

Kojadinovic, I., Quessy, J.-F., and Rohmer, T. (2015). Testing the constancy of spearman's rho in multivariate time series. *Annals of the Institute of Statistical Mathematics*, pages 1–26.

Kojadinovic, I. and Yan, J. (2012a). *fgof: Fast Goodness-of-fit Test*. R package version 0.2-1.

Kojadinovic, I. and Yan, J. (2012b). Goodness-of-fit testing based on a weighted bootstrap: A fast large-sample alternative to the parametric bootstrap. *Canadian Journal of Statistics*, 40(3):480–500.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer Science+Business Media, Inc. Springer e-books, New York, NY.

Lekina, A., Chebana, F., and Ouarda, T. (2014). Weighted estimate of extreme quantile: an application to the estimation of high flood return periods. *Stochastic Environmental Research and Risk Assessment*, 28(2):147–165.

Madsen, H. and Rosbjerg, D. (1997). The partial duration series method in regional index-flood modeling. *Water Resources Research*, 33(4):737–746.

Mediero, L., Santillán, D., Garrote, L., and Granados, A. (2014). Detection and attribution of trends in magnitude, frequency and timing of floods in spain. *Journal of Hydrology*, 517:1072–1088.

Mitková, V. B. and Halmová, D. (2014). Joint modeling of flood peak discharges, volume and duration: a case study of the Danube River in Bratislava. *Journal of Hydrology and Hydromechanics*, 62(3):186–196.

Mu, Y. and He, X. (2007). Power transformation toward a linear regression quantile. *Journal of the American Statistical Association*, 102(477):269–279.

Naveau, P., Guillou, A., Cooley, D., and Diebolt, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika*, 96(1):1–17.

Ouarda, T. B., Girard, C., Cavadias, G. S., and Bobée, B. (2001). Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology*, 254(1–4):157–173.

Pickands, III, J. (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, volume 49, pages 859–878, 894–902. With a discussion.

Quessy, J.-F., Saïd, M., and Favre, A.-C. (2013). Multivariate Kendall's tau for change-point detection in copulas. *The Canadian Journal of Statistics*, 41:65–82.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Resnick, S. I. (1987). *Extreme values, regular variation and point processes*. Springer.

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

Roth, M., Jongbloed, G., and Buishand, T. (2016). Threshold selection for regional peaks-over-threshold data. *Journal of Applied Statistics*, 43(7):1291–1309.

Rulfov, Z., Buishand, A., Roth, M., and Kysel, J. (2016). A two-component generalized extreme value distribution for precipitation frequency analysis. *Journal of Hydrology*, 534:659 – 668.

Scaillet, O. (2005). A Kolmogorov-Smirnov type test for positive quadrant dependence. *Canadian Journal of Statistics*, 33(3):415–427.

Schmidt, R. and Stadtmüller, U. (2006). Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics*, 33(2):307–335.

Schulte, M. and Schumann, A. (2015). Downstream-directed performance assessment of reservoirs in multi-tributary catchments by application of multivariate statistics. *Water Resources Management*, 29(2):419–430.

Schumann, A. (2005). Hochwasserstatistische bewertung des augusthochwassers 2002 im einzugsgebiet der mulde unter anwendung der saisonalen statistik. *Hydrologie und Wasserbewirtschaftung*, 49(4):200–206.

Segers, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli*, 18(3):764–782.

Serinaldi, F. and Kilsby, C. G. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources*, 77:17–36.

Sklar, A. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.

Strupczewski, W. G., Kochanek, K., Bogdanowicz, E., and Markiewicz, I. (2012). On seasonal approach to flood frequency modelling. part i: Two-component distribution revisited. *Hydrological Processes*, 26(5):705–716.

Teugels, J. L. and Vanroelen, G. (2004). Box-cox transformations and heavy-tailed distributions. *Journal of Applied Probability*, 41:213–227.

Tucker, H. G. (1968). Convolutions of distributions attracted to stable laws. *Annals of Mathematical Statistics*, 39(5):1381–1390.

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

Vervaat, W. (1971). Functional central limit theorems for processes with positive drift and their inverses. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 23(4):245–253.

Viglione, A. (2012). *homtest: Homogeneity tests for Regional Frequency Analysis*. R package version 1.0-5.

Viglione, A., Laio, F., and Claps, P. (2007). A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research*, 43(3).

Villarini, G., Smith, J. A., Serinaldi, F., Bales, J., Bates, P. D., and Krajewski, W. F. (2009). Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Advances in Water Resources*, 32(8):1255–1266.

Wang, H. and Tsai, C.-L. (2009). Tail index regression. *Journal of the American Statistical Association*, 104(487):1233–1240.

Wang, H. J. and Li, D. (2013). Estimation of extreme conditional quantiles through power transformation. *Journal of the American Statistical Association*, 108(503):1062–1074.

Wang, H. J., Li, D., and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, 107(500):1453–1464.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364):812–815.

Würtz, D. (2013). *fExtremes: Rmetrics - Extreme Financial Market Data*. R package version 3010.81.

Yue, S., Ouarda, T., Bobe, B., Legendre, P., and Bruneau, P. (1999). The Gumbel mixed model for flood frequency analysis. *Journal of Hydrology*, 226(12):88–100.

Yue, S., Pilon, P., and Cavadias, G. (2002). Power of the mann-kendall and spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology*, 259(1-4):254–271.