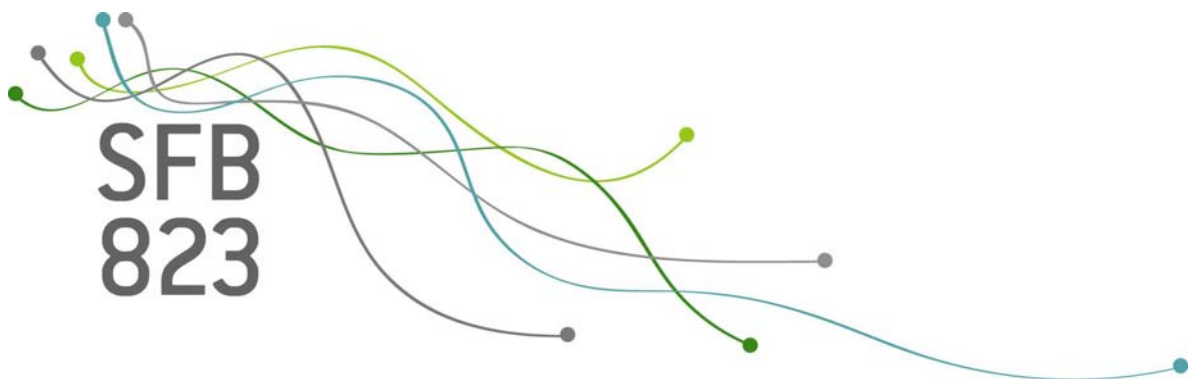# Conditional heavy-tail behavior with applications to precipitation and river flow extremes

Paul Kinsvater, Roland Fried

Nr. 35/2016

SFB
823

Conditional heavy-tail behavior with applications to precipitation and river flow extremes

Paul Kinsvater & Roland Fried
TU Dortmund

Abstract

This article 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 $F_Y(y|\ \mathbf{x}) = P(Y \leqslant y|\ \mathbf{X} = \mathbf{x})$, 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 possible $\mathbf{x}$.

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, $\gamma(\mathbf{x}) = \eta_0 + x \cdot \eta_1$. 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.

Keywords: Heavy tails; Extreme Value Index; Regression model; Relative excesses; Flood frequency; Precipitation

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 by 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 article 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 behavior $\gamma(\mathbf{x}) \neq \gamma_0$ 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 article is organized as follows: Section 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 3 and compared by simulation in Section 4. Applications to French

weekly precipitation and to river flow data from the Mulde basin are presented in Section 5.

2   Heavy tails and relative excesses

In the literature there are different characterizations of the term *heavy-tailed*. One of them often used in the literature on extremes identifies the right tail of a distribution function F as heavy, if, for $y \to \infty$, its tail decay $\bar{F}(y) = 1 - F(y) \to 0$ is of polynomial order. More precisely, F is called heavy-tailed to the right, if

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

for some parameter $\gamma > 0$ called extreme value index and function $L : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\lim_{t \to \infty} L(ty)/L(t) = 1$ for all $y > 0$. γ controls the heaviness of the right tail, with finite (resp. infinite) r-th moment $\int_0^\infty y^r \ dF(y)$ for $\gamma < 1/r$ (resp. $\gamma > 1/r$). The canonical example, the Pareto distribution function with support on $[1, \infty)$, is obtained if we let $L \equiv 1$ and this is why the family of all distributions satisfying (1) is also called *Pareto-type*.

It turns out that this family is very rich containing many common distributions. 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 $\xi = \gamma$. In fact, the family of Pareto-type distributions coincides with the Fréchet domain of attraction [de Haan and Ferreira, 2006, Theorem 1.2.1]: F satisfies (1) if and only if there exist $a_n > 0$, $b_n \in \mathbb{R}$ such that $\lim_{n \to \infty} F^n(a_n y + b_n) = G_\gamma(y)$ for all continuity points y of the extreme value distribution G_γ and $\gamma > 0$.

From a theoretical point of view, the characterization in (1) says absolutely nothing about F restricted to the interval $(-\infty, u]$, for arbitrary constants $u \in \mathbb{R}$. Thus, provided there are no further assumptions on F, 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, ∞).

Let Y be a random variable with distribution function $F(y) = \mathbb{P}(Y \leqslant y)$ as in (1) and let $u > 0$ be a real number. The random variable $Z_u = Y/u$ satisfying $Y > u$ is called relative excess over the threshold u. From relation (1) we obtain

$$\mathbb{P}\left(Z_u \leqslant z \mid Y > u\right) = 1 - \frac{\bar{F}(uz)}{\bar{F}(u)} \longrightarrow 1 - z^{-1/\gamma} = P_\gamma(z) \ \text{ for } z > 1, \ u \to \infty. \tag{2}$$

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 given by a Pareto distribution function commonly parameterized $Pareto(\alpha)$ with parameter $\alpha = 1/\gamma > 0$, the so called tail index.

The interest of the present article is in the conditional behavior of extremes. Let $(Y, \mathbf{X})$ be a random element, where Y is now called response and $\mathbf{X} = (1, X_1, \ldots, X_d)'$ a vector of regressors with range on a compact set $\mathcal{X} \subset \mathbb{R}^{d+1}$. We assume that the conditional distribution of Y given $\mathbf{X} = \mathbf{x}$ is of Pareto-type, that is,

$$F_Y(y \mid \mathbf{x}) = P(Y \leqslant y \mid \mathbf{X} = \mathbf{x}) = 1 - y^{-1/\gamma(\mathbf{x})} \cdot L(y \mid \mathbf{x}), \tag{3}$$

3

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

$$\mathcal{H}_{0,tail} : \ \gamma(\mathbf{x}) = \gamma_0 \text{ for some } \gamma_0 > 0 \text{ and all } \mathbf{x} \in \mathcal{X} \tag{4}$$

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 $\gamma(\mathbf{x})$.

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 (3)? A practical solution of this problem is discussed in the following two subsections.

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

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

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

4

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

$$\hat{\boldsymbol{\beta}}_p = \underset{\mathbf{b} \in \mathbb{R}^{d+1}}{\arg\min} \sum_{i=1}^{n} \rho_p \left(g(Y_i) - \mathbf{X}_i' \cdot \mathbf{b} \right), \tag{5}$$

where $\rho_p(y) = y \cdot (p - \mathbb{1}_{\{y \leqslant 0\}})$ 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

$$\hat{u}_p(\mathbf{x}) = \hat{F}_Y^{-1}(p|\,\mathbf{x}) = g^{-1}\left(\mathbf{x}'\hat{\boldsymbol{\beta}}_p\right). \tag{6}$$

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 (3) holds.
Mu and He [2007] proposed the consistent estimator

$$\hat{\lambda}_p = \underset{\lambda \in \mathbb{R}}{\arg\min} \sum_{i=1}^{n} \left[R_n(\mathbf{x}_i, \lambda, p) \right]^2 \tag{7}$$

of λ_p, where

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

and $\hat{\boldsymbol{\beta}}_{p,\lambda}$ is computed by (5) 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 $k < n$:

1. Set $p = p_{k,n} = \frac{n-k}{n+1}$ and compute $\hat{\lambda}$ by (7).

2. Solve (5) with $g = g_{\hat{\lambda}}$ and let $u(\mathbf{x}) = u_{k,n}(\mathbf{x}) = g_{\hat{\lambda}}^{-1}\left(\mathbf{x}'\hat{\boldsymbol{\beta}}_p\right)$.

3. Identify all $1 \leqslant i_1 < \ldots < i_k \leqslant n$ 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 $(Z_{k,j}, (1, X_{k,j})')$.

2.2 How to select the tail sample size 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

$$k^* = \underset{1 \leqslant k < n}{\arg\min}\, D_n(k) = \underset{1 \leqslant k < n}{\arg\min}\, \frac{1}{k} \sum_{j=1}^{k} \left(\hat{U}_{k,j:k}(\hat{\gamma}_{k,n}) - \frac{j}{k+1} \right)^2 \tag{8}$$

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 (8) 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.

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 $\mathbb{P}(Z_{k,j} \leqslant z | \mathbf{X}_{k,j} = \mathbf{x}) = 1 - z^{-1/\gamma(\mathbf{x})}$. 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].

3.1 L-estimation under linearity $\gamma(\mathbf{x}) = \mathbf{x}'\eta$ and related tests

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

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

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

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

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

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

with $\rho_p(u) = u \cdot \left(p - \mathbf{1}_{\{u \leqslant 0\}} \right)$. 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 2 in Appendix A and the model assumptions stated above, we easily obtain the following result:

Proposition 1. *Let* $\mathbf{p} = \{p_1, \ldots, p_\ell\} \subset (0, 1)$ *denote a finite set of distinct probabilities and suppose that* $J = E(\mathbf{XX}')$ *and* $H = E\left(\mathbf{XX}' / \mathbf{X}' \boldsymbol{\eta} \right) \in \mathbb{R}^{(d+1) \times (d+1)}$ *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 $\Omega_{\mathbf{p}} = A_{\mathbf{p}} \otimes \left(H^{-1} J H^{-1} \right)$ *and* $A_{\mathbf{p}} \in \mathbb{R}^{\ell \times \ell}$ *is defined through its entries*

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

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 2 from the appendix. 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 2, we have that $H_p = (1 - p) \cdot H$ with H_p defined in (14), which, for $k \to \infty$, gives us

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

$\square$

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

$$\sqrt{k} \left(\hat{\boldsymbol{\eta}}(\mathbf{p}, \mathbf{w}) - \boldsymbol{\eta} \right) \xrightarrow{D} \mathcal{N} \left(0, B_{\mathbf{w}} \Omega_{\mathbf{p}} B_{\mathbf{w}}' \right) = \mathcal{N} \left(0, \Sigma_{\mathbf{p}, \mathbf{w}} \right), \tag{9}$$

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

7

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, p and 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

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

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 (3) 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 1 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 (3). 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 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 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 (9) 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 $H_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \mathbf{X}_{k,j} \mathbf{X}_{k,j}' / \hat{\eta}_{k,n}^{L}(\mathbf{X}_{k,j})$ for J and H, respectively.

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

$$\tau(X, Y) = 4 \cdot \int_{[0,1]^2} C(u, v) dC(u, v) - 1 \tag{10}$$

defines Kendall's correlation coefficient. $\tau(X, Y)$ is a margin-free dependence coefficient with $-1 \leqslant \tau(X, Y) \leqslant 1$ and $\tau(X, Y) = 0$ for independent variables X and Y. The empirical counterpart of (10) is $t_n = \frac{2}{n(n-1)} \cdot S_n$ with

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

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

$$S_{k,n} = \sum_{1 \leqslant i < j \leqslant k} sgn(Z_{k,j} - Z_{k,i}) \cdot sgn(X_{k,j} - X_{k,i}), \tag{11}$$

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 2.1). Critical values are computed based on the presumed approximation $S_{k,n} \overset{D}{\approx} \mathcal{N}(0, \sigma_k^2)$.

9

4 Simulation study

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). 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

$$Y_i = \mathbf{X}_i'\boldsymbol{\mu} + \mathbf{X}_i'\boldsymbol{\sigma} \cdot \varepsilon_i, \ 1 \leqslant i \leqslant n, \tag{12}$$

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

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

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
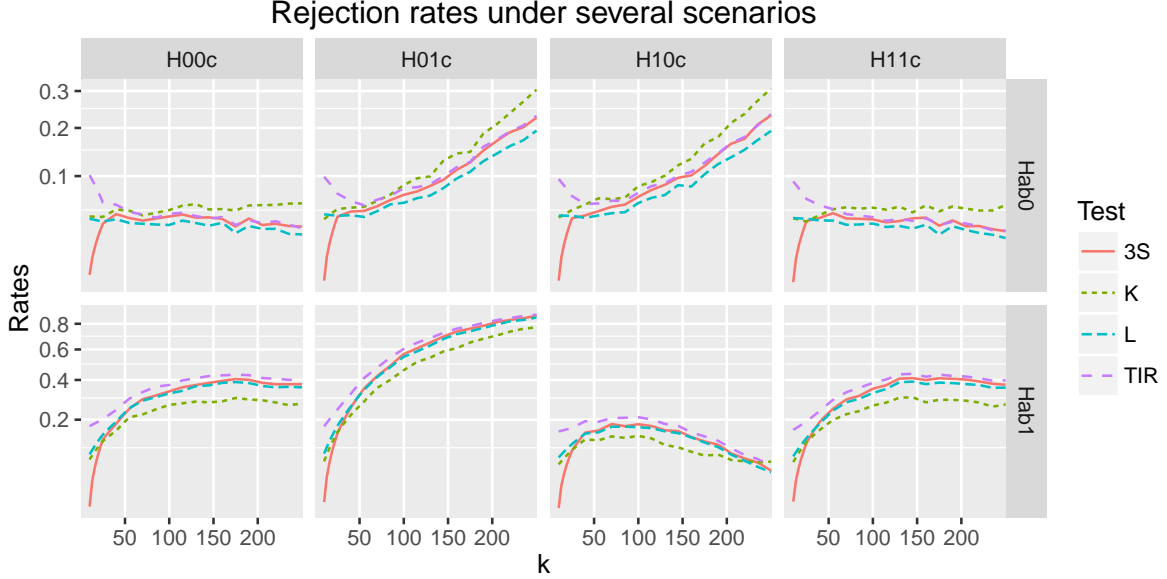
<div align="center">10</div>

Figure 1: Rejection rates of four different tests computed from 4000 samples generated by process (12) 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 (12) 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.

3S: a test based on the three-stage test statistic in (18),

K: the test based on the Kendall's tau statistic (11),

L: a two-sided t-test based on the weak limit of estimator $\hat{\eta}_1$ in (9) and

TIR: a two-sided t-test based on the weak limit of estimator $\hat{\theta}_1$ in (17) with $h_1 = 0$.

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 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 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
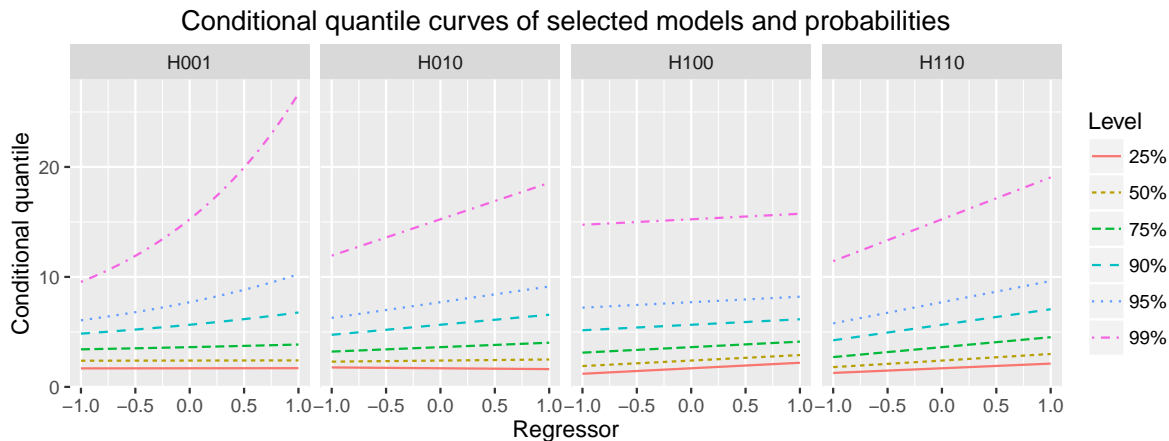
Figure 2: Conditional quantile curves $F^{-1}(p| \ x)$ for probability levels $p \in \{0.25, \ 0.5, \ 0.75, \ 0.9, \ 0.95, \ 0.99\}$ and for a few selected scenarios defined in Section 4.1.

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 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.

4.2 Selection of k

This part of the simulations is devoted to the adaptive selection of k, for instance, rule (8). The simulation results depicted in Figure 3 are computed under $\mathcal{H}_{0,0,1}$ with observations generated by (12), 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 (8), (L^*2) $k = \lfloor 0.75k^* \rfloor$ and (L^*3) $k = \lfloor 1.25k^* \rfloor$. $k = \lfloor 2n^{2/3} \rfloor$ is

Table 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 1. The nominal level is 5%.

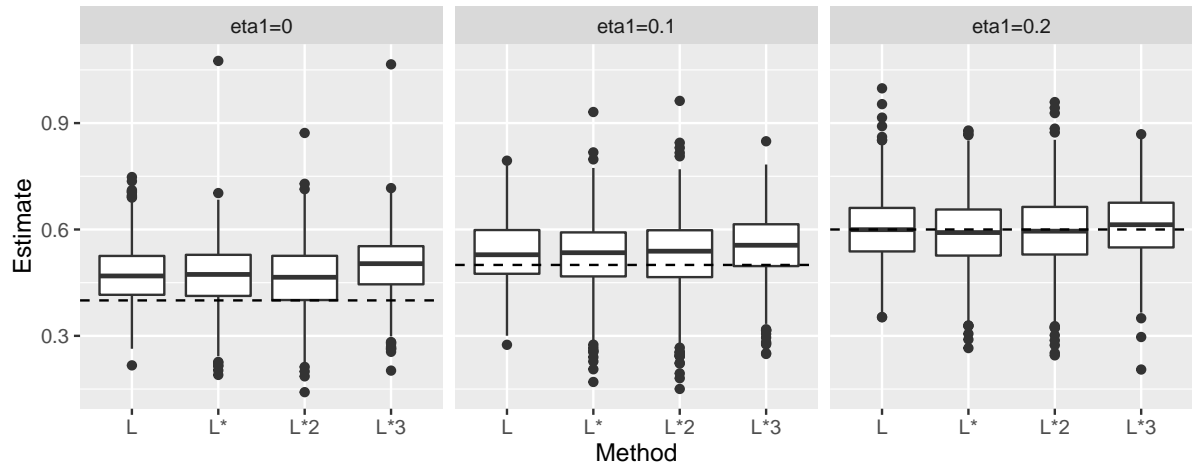|	L	L^*	TIR	TIR^*	K	K^*
$\mathcal{H}_{0,0,0}$	3.9	3.1	4.3	5.3	4.9	4.9
$\mathcal{H}_{0,1,0}$	7.5	10.6	9.5	15.9	10.2	13.8
$\mathcal{H}_{1,0,0}$	7.5	12.0	9.3	15.4	11.0	14.6
$\mathcal{H}_{1,1,0}$	3.3	3.6	4.1	5.3	5.2	5.1

12

Figure 3: Comparison of L-estimators $\hat{\gamma}^L(1)$ of $\gamma(1) = \eta_0 + \eta_1$ based on 1000 independent samples of DGP (12) 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 (8). 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 $(\eta_1 = 0.2)$ $0.0085, 0.0087, 0.0099, 0.0074$.

the asymptotically optimal choice for 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 (8) 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 (8) was applied. Data were generated under the same scenarios like in the top of Figure 1. Table 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.

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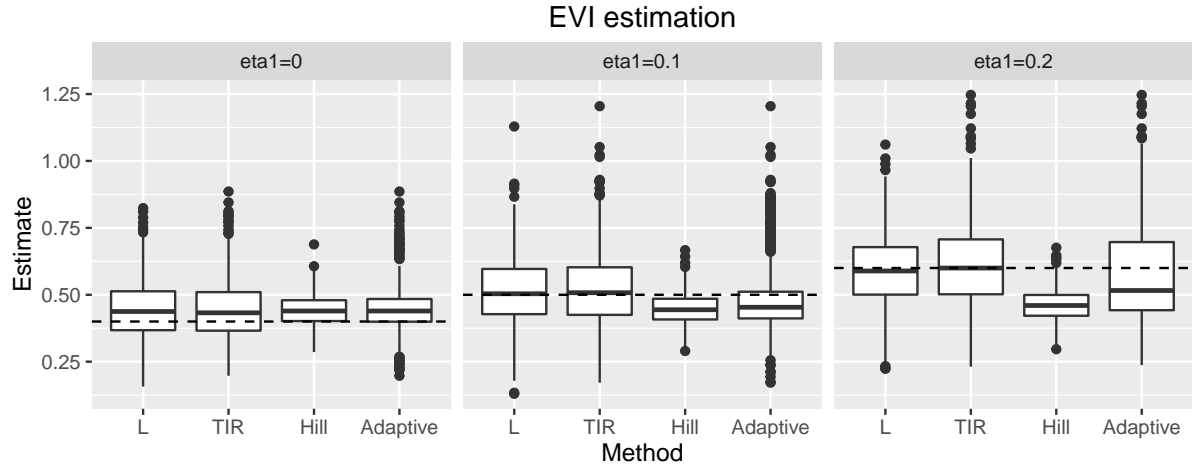
More precisely, we evaluate the following questions:

13

Figure 4: Comparison of estimators of $\gamma(1) = \eta_0 + \eta_1$ based on 1000 independent samples of DGP (12) 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 ($\eta_1 = 0.2$) $0.0166, 0.0234, 0.0238, 0.0300$.

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 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 (12), 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 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 B.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 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), 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.

5   Applications

5.1   Weekly maxima of hourly precipitation in France

The left-hand side of Figure 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 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.

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

$$\hat{F}^{-1}(p|x) = u_{k,n}(x) \cdot \left(\frac{k}{n(1-p)} \right)^{\hat{\gamma}_0} , \ p > 1 - \frac{k}{n},$$
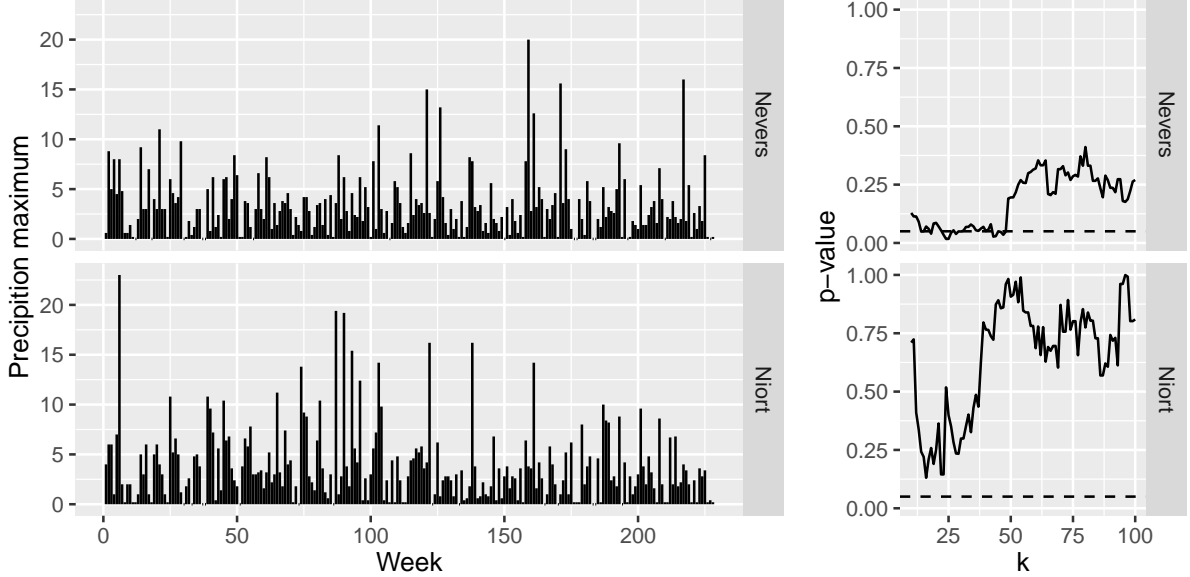
Figure 5: (Left) Weakly maxima of hourly precipitation during the fall season from 1993 to 2011 at the stations Nevers and Niort in France and (right) p-values of the L-test versus k between 10 and 100. The dashed line corresponds to a level of 5%.

where $u_{k,n}(x)$ is defined in Section 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 (8), 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 3.1, which, by Proposition 1, 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

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

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 4.3, we 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 $[0.43, 0.65]$.

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

16

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 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 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 (8) 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.

It is evident 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 6. In Germany the j-th hydrological year starts in the first day of November of the $(j-1)$-th calendar
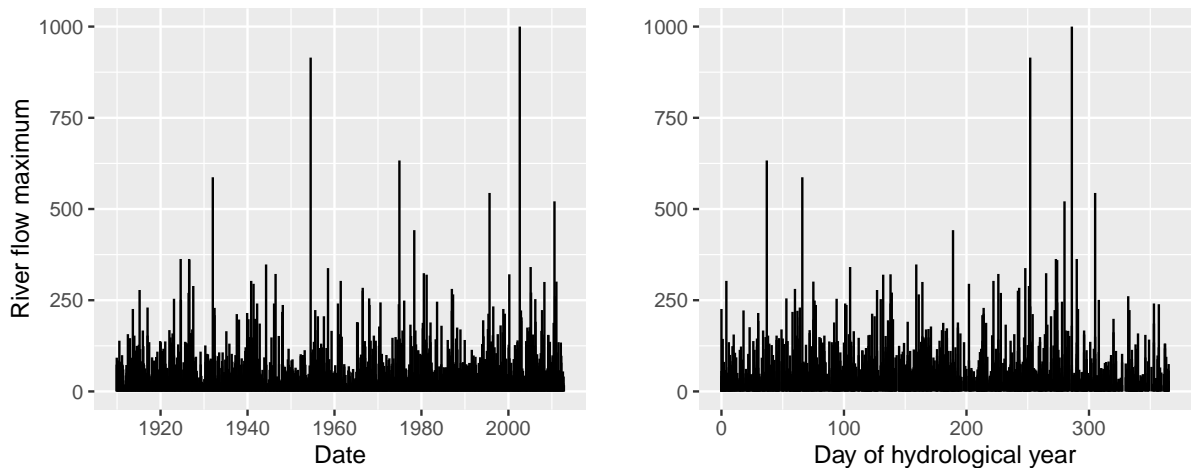

Figure 6: Monthly maximal flows observed at station Wechselburg starting from November 1909 till October 2012 in (left) chronological order and (right) ordered according to their day of appearance in the German hydrological year.

17

year and ends in 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 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 (8) 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 on 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 other 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], RulfovÃ¡ 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 summer (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.

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 [Kinsvater et al., 2016]. 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.

Acknowledgements

A   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 $F(y|\mathbf{x}) = P(Y \leqslant y | \mathbf{X} = \mathbf{x})$ of Y given $\mathbf{X} = \mathbf{x}$ satisfies

$$F^{-1}(p|\mathbf{x}) = \inf\{y : F(y|\mathbf{x}) \geqslant p\} = \mathbf{x}'\boldsymbol{\beta}_p \tag{13}$$

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 (13) 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 $\rho_p(y) = y \cdot (p - \mathbb{1}_{\{y \leqslant 0\}})$ 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 the function space $(\ell^\infty(I))^{d+1}$ [van der Vaart and Wellner, 1996].

Theorem 2 (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*

$$J = \mathbb{E}\left[\mathbf{X}\mathbf{X}'\right] \quad and \quad H_p = \mathbb{E}\left[\mathbf{X}\mathbf{X}' \cdot f(F^{-1}(p|\mathbf{X})|\mathbf{X})\right] \tag{14}$$

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

$$\left(H_p\sqrt{n}\left(\hat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_p\right)\right)_{p \in I} \xrightarrow{D} \mathbb{Z} \tag{15}$$

in $(\ell^\infty(I)))^{d+1}$, *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

$$\lim_{n \to \infty} \text{Cov}\left[\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{p_i} - \boldsymbol{\beta}_{p_i}\right), \sqrt{n}\left(\hat{\boldsymbol{\beta}}_{p_j} - \boldsymbol{\beta}_{p_j}\right)\right] = (p_i \wedge p_j - p_i \cdot p_j) \cdot H_{p_i}^{-1}JH_{p_j}^{-1}.$$

This result is used to prove Proposition 1.

B   Conditional heavy-tail behavior - competing methods

B.1   Tail index regression (TIR) by Wang and Tsai [2009]

Wang and Tsai [2009] study model (3) 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

$$\hat{\theta}_{u_n} = \underset{\theta \in \mathbb{R}^{d+1}}{\arg\min} \sum_{i=1}^{n}\left[\exp(\mathbf{X}_i'\theta) \cdot \log(Y_i/u_n) - \mathbf{X}_i'\theta\right] \cdot \mathbf{1}(Y_i > u_n) \tag{16}$$

with regressor independent threshold $u_n \to \infty$ for $n \to \infty$. (16) 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 (16) 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

$$\sqrt{k} \cdot \hat{\Sigma}_{u_n}^{1/2} \cdot \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{D} \mathcal{N}\left(\mathbf{h}, I_{d+1}\right) \tag{17}$$

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 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 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 $\hat{\Sigma}_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \mathbf{X}_{k,j} \mathbf{X}_{k,j}'$.

B.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:

1. Set $p = p_{k,n} = \frac{n-k}{n+1}$ and compute $\hat{\lambda}$ as in Section 2.1.

2. Let $p_{n-j,n} = \frac{j}{n+1}$ for $j = 1, \ldots, m$ with $m = n - \lfloor n^\eta \rfloor$ and $\eta = 0.1$. For $j = 1, \ldots, m$, estimate $F_Y^{-1}(p_{n-j,n}|\mathbf{x})$ by the right hand side of (6) with $g = g_{\hat{\lambda}}$ and $p = p_{n-j,n}$. Denote these estimates by $\hat{q}_j(\mathbf{x})$, $j = 1, \ldots, m$. If $\hat{q}_j(\mathbf{x})$ is not increasing in j, apply the rearrangement procedure of Chernozhukov et al. [2010].

3. 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 $F_Y(\cdot | \mathbf{x})$. Wang and Li [2013] also propose a test statistic

$$T_n = \frac{1}{n} \sum_{i=1}^{n} (\hat{\gamma}_{k,n}(\mathbf{X}_i) - \hat{\gamma}_p)^2, \quad \hat{\gamma}_p = \frac{1}{n} \sum_{i=1}^{n} \hat{\gamma}(\mathbf{X}_i), \tag{18}$$

as a test for hypothesis $\mathcal{H}_{0,tail}$ in (4). If $\mathcal{H}_{0,tail}$, $E(\mathbf{X}) = (1, 0, \ldots, 0)' \in \mathbb{R}^{d+1}$ and either $\gamma^*(\mathbf{x}) = 0$ or a certain homogeneity assumption are met, Wang and Li [2013] show under additional technical assumptions that $kT_n \xrightarrow{D} \gamma^2 \chi_d^2$ 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].

References

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006). Quantile regression under misspecification, with an application to the u.s. wage structure. *Econometrica 74*(2), 539–563.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.

Bickel, P. J. and E. L. Lehmann (1975, 09). Descriptive statistics for nonparametric models ii. location. *Annals of Statistics 3*(5), 1045–1069.

Bücher, A., P. Kinsvater, and I. Kojadinovic (2015). Detecting breaks in the dependence of multivariate extreme-value distributions. *ArXiv:1505.00954*.

Chavez-Demoulin, V. and A. C. Davison (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 54*(1), 207–222.

Chebana, F., T. B. Ouarda, and T. C. Duong (2013). Testing for multivariate trends in hydrologic frequency analysis. *Journal of Hydrology 486*, 519 – 530.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010). Quantile and probability curves without crossing. *Econometrica 78*(3), 1093–1125.

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.

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

de Haan, L., A. Tank, and C. Neves (2015). On tail trend detection: modeling relative risk. *Extremes 18*(2), 141–178.

Dierckx, G. (2011). Trends and change points in the tail behaviour of a heavy tailed distribution. In *Proc. 58th World Statistical Congress (ISI2011), Dublin*, pp. 290–299.

Dierckx, G. and J. L. Teugels (2010). Change point analysis of extreme values. *Environmetrics 21*(7-8), 661–686.

Dupuis, D. J., Y. Sun, and H. J. Wang (2015). Detecting change-points in extremes. *Statistics and Its Interface 8*(1), 19–31.

Einmahl, J. H. J., L. de Haan, and C. Zhou (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(1), 31–51.

Gardes, L. and S. Girard (2010). Conditional extremes from heavy-tailed distributions: an application to the estimation of extreme rainfall return levels. *Extremes 13*(2), 177–204.

Gomes, M. I. and D. Pestana (2007). A sturdy reduced-bias extreme quantile (var) estimator. *Journal of the American Statistical Association 102*(477), 280–292.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics 3*(5), 1163–1174.

Jarušková, D. and M. Rencová (2008). Analysis of annual maximal and minimal temperatures for some european cities by change point methods. *Environmetrics 19*(3), 221–233.

Kendall, M. G. (1948). *Rank correlation methods.* Charles Griffin London.

Kim, M. and S. Lee (2009). Test for tail index change in stationary time series with pareto-type marginal distribution. *Bernoulli 15*(2), 325–356.

Kinsvater, P., R. Fried, and J. Lilienthal (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 J. Bassett, Gilbert (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Kojadinovic, I. and P. Naveau (2015). Nonparametric tests for change-point detection in the distribution of block maxima based on probability weighted moments. *ArXiv:1507.06121*.

Madsen, H. and D. Rosbjerg (1997). The partial duration series method in regional index-flood modeling. *Water Resources Research 33*(4), 737–746.

Mediero, L., D. Santillán, L. Garrote, and A. Granados (2014). Detection and attribution of trends in magnitude, frequency and timing of floods in spain. *Journal of Hydrology 517*, 1072–1088.

Mu, Y. and X. He (2007). Power transformation toward a linear regression quantile. *Journal of the American Statistical Association 102*(477), 269–279.

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

Roth, M., G. Jongbloed, and T. Buishand (2016). Threshold selection for regional peaks-over-threshold data. *Journal of Applied Statistics 43*(7), 1291–1309.

Rulfová, Z., A. Buishand, M. Roth, and J. Kyselý (2016). A two-component generalized extreme value distribution for precipitation frequency analysis. *Journal of Hydrology 534*, 659 – 668.

Schumann, A. (2005). Hochwasserstatistische bewertung des augusthochwassers 2002 im einzugsgebiet der mulde unter anwendung der saisonalen statistik. *Hydrologie und Wasserbewirtschaftung 49*(4), 200–206.

Strupczewski, W. G., K. Kochanek, E. Bogdanowicz, and I. Markiewicz (2012). On seasonal approach to flood frequency modelling. part i: Two-component distribution revisited. *Hydrological Processes 26*(5), 705–716.

Teugels, J. L. and G. Vanroelen (2004). Box-cox transformations and heavy-tailed distributions. *Journal of Applied Probability 41*, 213–227.

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

Wang, H. and C.-L. Tsai (2009). Tail index regression. *Journal of the American Statistical Association 104*(487), 1233–1240.

Wang, H. J. and D. Li (2013). Estimation of extreme conditional quantiles through power transformation. *Journal of the American Statistical Association 108*(503), 1062–1074.

Wang, H. J., D. Li, and X. He (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.

Yue, S., P. Pilon, and G. Cavadias (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.