Kernel regression with Weibull-type tails

Tertius de Wet [*]
Yuri Goegebeur [†]
Armelle Guillou [‡]
Michael Osmann [§]

June 12, 2015

Abstract

We consider the estimation of the tail coefficient of a Weibull-type distribution in the presence of random covariates. The approach followed is non-parametric and consists of locally weighted estimation in narrow neighborhoods in the covariate space. We introduce two families of estimators and study their asymptotic behavior under some conditions on the conditional response distribution, the kernel function, the density function of the independent variables, and for appropriately chosen bandwidth and threshold parameters. We also introduce a Weissman-type estimator for estimating upper extreme conditional quantiles. The finite sample behavior of the proposed estimators is examined with a simulation experiment. The practical applicability of the methodology is illustrated on a data set of sea storm measurements.

Keywords: Extreme value statistics, Weibull-type distribution, regression, second order condition.

1 Introduction

The estimation of the tail index of a distribution plays a central role in the area of extreme value statistics. It is typically the first step in a practical data analysis, since such an estimate is needed in e.g. models used to estimate upper extreme quantiles. In this paper, we will consider some estimation problems within the Gumbel class, which is a rich subclass of the max-domain

[*]Department of Statistics and Actuarial Science, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa (email: tdewet@sun.ac.za).

[†]Corresponding author, Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: yuri.goegebeur@imada.sdu.dk).

[‡]Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France (email: armelle.guillou@math.unistra.fr).

[§]Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: mosma@imada.sdu.dk).

The online version of this article contains supplementary material

of attraction. Although different types of tail behavior are possible, all these distributions have in common an extreme value index γ equal to zero and thus differentiating them on the basis of this parameter alone is impossible. To solve this issue, we restrict our study to Weibull-type distributions, which have distribution functions F defined for some $\theta > 0$ as

$$1 - F(y) = \exp(-y^{1/\theta}\ell^*(y)), \qquad y > 0, \tag{1}$$

where ℓ^* is a slowly varying function at infinity, i.e.

$$\frac{\ell^*(\lambda y)}{\ell^*(y)} \to 1 \text{ as } y \to \infty \text{ for all } \lambda > 0.$$

The parameter θ, called the Weibull-tail coefficient, clearly governs the tail behavior, where larger values correspond to a slower decay of $1 - F$ towards zero. Different values of it allow the Weibull-type distributions to cover a large part of the Gumbel class, and hence to constitute a flexible subgroup. The model finds important applications in areas such as hydrology, meteorology, environmental and actuarial science, to name but a few. For a given data set one can evaluate the adequacy of the Gumbel class by performing a test for $\gamma = 0$ versus $\gamma \neq 0$, see e.g. Segers and Teugels (2000). The more specific Weibull-type class can be visually evaluated by an inspection of the Weibull quantile plot, which should become linear in the largest observations. Goegebeur and Guillou (2010) introduced a formal goodness-of-fit test for the semi-parametric Weibull-type model based on a quantification of the linearity of the upper part of the Weibull quantile plot. In the analysis of univariate Weibull-type tails, the estimation of θ and the subsequent estimation of upper extreme quantiles assume a central position. We refer to Broniatowski (1993), Beirlant *et al.* (1995), Girard (2004), Gardes and Girard (2005, 2008a), Diebolt *et al.* (2008), Dierckx *et al.* (2009), Goegebeur *et al.* (2010), Goegebeur and Guillou (2011), and the references therein.

In this paper we will consider the estimation of the Weibull-tail coefficient and of upper extreme quantiles when a *random* covariate X is recorded simultaneously with the variable of interest Y. Thus, we do not consider the classical experimental situation where the researcher can control X and then subsequently observe Y. Instead, X and Y are observed together, and therefore X is considered to be random. Having methodology that allows to take covariate information into account in an extreme value analysis is important as it enables one to differentiate the tail behavior in terms of one or more risk factors. Our approach to this regression on extreme values is non-parametric and based on locally weighted estimation.

The regression analysis of extreme values with *fixed*, i.e. nonrandom, covariates has been extensively considered in the extreme value literature in a parametric, semi-parametric or non-parametric way, but mostly in the case of Pareto-type distributions. On the contrary, the development of extreme value regression methodology with random covariates is still in its infancy. Some methodology has been introduced but again focusing on the case of Pareto-type distributions. One has to wait for Daouia *et al.* (2013) for a non-parametric estimation method that is valid in the general max-domain of attraction. In the present paper, we will develop a methodology for estimating the tail coefficient and upper extreme conditional quantiles for the

class of the Weibull-type distributions, which, as mentioned above, forms a rich subset of the Gumbel class. To the best of our knowledge, this estimation problem has not been considered in the extreme value literature.

The remainder of this paper is organized as follows. In the next section we introduce two classes of estimators for the conditional Weibull-tail coefficient and study their asymptotic properties. The estimation of upper extreme conditional quantiles is considered in Section 3. In Section 4 we examine the small sample behavior of the proposed estimators by a simulation experiment, and in Section 5 we show the results for the real data example. Section 6 concludes the paper. The proofs of the main results are deferred to the appendix, while for other results we refer to the supporting information which is online available.

2 Construction and asymptotic properties

Let (X_i, Y_i), $i = 1, \ldots, n$, be independent copies of the random vector $(X, Y) \in \mathbb{R}^q \times \mathbb{R}_{+,0}$, where X has a distribution with joint density function g, and the conditional survival function of Y given $X = x$, denoted $\bar{F}(y; x) := 1 - F(y; x)$, is of Weibull-type, that is, for some $\theta(x) > 0$,

$$\bar{F}(y; x) = \exp(-y^{1/\theta(x)} \ell^*(y; x)), \qquad y > 0, \tag{2}$$

where ℓ^* is a slowly varying function at infinity. The model can also be defined in terms of the generalized inverse of F, denoted Q, i.e. $Q(p; x) := \inf\{y : F(y; x) \geq p\}$, $0 < p < 1$:

$$Q(p; x) = (-\ln(1 - p))^{\theta(x)} \ell(-\ln(1 - p); x), \tag{3}$$

where ℓ is again a slowly varying function at infinity (Bingham *et al.*, 1987, p 6).

Our approach to the estimation of the Weibull-tail coefficient and of extreme conditional quantiles is based on elemental kernel statistics of the form

$$T_n^{(t)}(x, K) := \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(x - X_i)(\ln Y_i - \ln \omega_n)_+^t \mathbf{1}\{Y_i > \omega_n\}, \tag{4}$$

where $t \geq 0$, $K_{h_n}(x) := K(x/h_n)/h_n^q$, where K is a joint density function on $\mathbb{R}^q$, h_n is a non-random sequence of bandwidths with $h_n \to 0$ if $n \to \infty$, $\mathbf{1}\{A\}$ is the indicator function on the event A and ω_n is a local non-random threshold sequence satisfying $\omega_n \to \infty$ if $n \to \infty$. The parameter t is a tuning parameter, introduced to allow for more flexibility. Since $\theta(x)$ is a tail parameter for the conditional distribution of Y given $X = x$, it is natural to consider statistics like (4) that involve exceedances over a high threshold, and that also occur in some neighborhood of the point of interest $x \in \mathbb{R}^q$.

To obtain the limiting behaviour of (4) we need to introduce some conditions on the conditional quantile function $Q(p; x)$, the density function g of the independent variables and the kernel function K. Concerning $Q(p; x)$, we introduce the so-called second order condition on its

slowly varying function ℓ.

Assumption $(\mathcal{R})$ There exists a constant $\rho(x) < 0$ and a rate function $b(.;x)$ satisfying $b(y;x) \to 0$ as $y \to \infty$, such that for all $\lambda \geq 1$, we have

$$\ln\left(\frac{\ell(\lambda y; x)}{\ell(y; x)}\right) = b(y; x) D_{\rho(x)}(\lambda)(1 + o(1))$$

with $D_{\rho(x)}(\lambda) := \int_1^\lambda t^{\rho(x)-1} dt$, and where $o(1)$ is uniform in $\lambda \geq 1$, as $y \to \infty$.

Since $\ln x \sim x - 1$ for $x \to 1$, we have that Assumption $(\mathcal{R})$ makes the convergence of $\ell(\lambda y; x)/\ell(y; x)$ to its limit, being 1, explicit. In fact, as shown in Geluk and de Haan (1987), $(\mathcal{R})$ implies that $|b(y;x)|$ is regularly varying with index $\rho(x)$, i.e. $|b(\lambda y;x)|/|b(y;x)| \to \lambda^{\rho(x)}$ as $y \to \infty$ for all $\lambda > 0$, so $\rho(x)$ governs the rate of this convergence. If $|\rho(x)|$ is small then the convergence is slow and the estimation of tail quantities is generally difficult. Assumption $(\mathcal{R})$ is well accepted in the literature. It was formulated in a slightly different form in e.g. Diebolt *et al.* (2008) and Goegebeur and Guillou (2011) in the Weibull-type framework, and Gardes and Girard (2008b) invoked it for tail analysis in the Fréchet max-domain of attraction.

As a first step in the development of estimators for $\theta(x)$, we study in Lemma 1, given in the appendix, the local behavior of the conditional expectation of a power transformed excess

$$m(\omega_n, t; x) := \mathbb{E}[(\ln Y - \ln \omega_n)_+^t \mathbf{1}\{Y > \omega_n\}|X = x].$$

Then, to deal with the randomness of X, we have to consider the unconditional expectation

$$\widetilde{m}_n(K, t; x) := \mathbb{E}[K_{h_n}(x - X)(\ln Y - \ln \omega_n)_+^t \mathbf{1}\{Y > \omega_n\}]$$

which is exactly the expectation of our elemental kernel $T_n^{(t)}$. In Lemma 2, given in the appendix, we state its main asymptotic expansion under the following additional assumptions. The density function of the covariate X is assumed to satisfy a Lipschitz condition. For all $x, z \in \mathbb{R}^q$, the Euclidean distance between x and z is denoted by $d(x, z)$.

Assumption $(\mathcal{G})$ There exists $c_g > 0$ such that $|g(x) - g(z)| \leq c_g d(x, z)$ for all $x, z \in \mathbb{R}^q$.

For the weight or kernel function K we assume the following.

Assumption $(\mathcal{K})$ K is a bounded density function on $\mathbb{R}^q$, with support Ω included in the unit hypersphere in $\mathbb{R}^q$.

This assumption is standard in the framework of kernel estimation.

Finally, since the estimation takes place in a narrow neighborhood in the covariate space, we have to introduce a condition on the oscillation of the conditional response distribution, when considered as a function of x. This condition is formulated in terms of the conditional

expectation $m(\omega_n, t; x)$.

Assumption ($\mathcal{F}$) The conditional excess function $m(\omega_n, t; x)$ satisfies that, for $\omega_n \to \infty$, $h_n \to 0$, and some $T > 0$,

$$\Phi(\omega_n, h_n; x) := \sup_{t \in [0,T]} \sup_{z \in \Omega} \left| \frac{m(\omega_n, t; x - h_n z)}{m(\omega_n, t; x)} - 1 \right| \to 0 \text{ if } n \to \infty.$$

This assumption is a smoothness condition on the underlying conditional distribution function (when $t = 0$) and on some conditional expectations of power transformed excesses (when $t > 0$). As illustrated in the supporting information, which is online available, this assumption is satisfied in case $t = 0$ by imposing some more structure on $\bar{F}$ (in particular that $\ell^*(.; x)$ is normalized) and suitable conditions on h_n and ω_n.

From Lemmas 1 and 2 in the appendix, we obtain the following asymptotic expansion in case $t > 0$

$$
\begin{aligned}
\mathbb{E}(T_n^{(t)}(x, K)) &= \frac{\bar{F}(\omega_n; x)}{(-\ln \bar{F}(\omega_n; x))^t}\, g(x)\theta^t(x)\Gamma(t+1) \\
&\quad \times \left\{ 1 + b(-\ln \bar{F}(\omega_n; x); x)\frac{t}{\theta(x)}(1 + o(1)) + O\left(\frac{1}{(-\ln \bar{F}(\omega_n; x))^{1-\varepsilon}}\right) \right. \\
&\quad \left. + O(h_n) + O(\Phi(\omega_n, h_n; x)) \right\},
\end{aligned}
$$
(5)

while for $t = 0$

$$\mathbb{E}(T_n^{(0)}(x, K)) = g(x)\bar{F}(\omega_n; x)(1 + O(h_n) + O(\Phi(\omega_n, h_n; x))).$$
(6)

The leading terms of (5) and (6) motivate our estimators for $\theta(x)$:

$$
\begin{aligned}
\hat{\theta}_n^{(1)}(x; t, K_1, K_2) &= \left(\frac{(-\ln \hat{\bar{F}}(\omega_n; x))^t T_n^{(t)}(x, K_1)}{\Gamma(t+1) T_n^{(0)}(x, K_2)} \right)^{1/t}, \qquad t > 0, \\
\hat{\theta}_n^{(2)}(x; t, K_1, K_2) &= \frac{(-\ln \hat{\bar{F}}(\omega_n; x)) T_n^{(t+1)}(x, K_1)}{(t+1) T_n^{(t)}(x, K_2)}, \qquad t \geq 0,
\end{aligned}
$$

where K_1 and K_2 are kernel functions satisfying ($\mathcal{K}$), $\hat{\bar{F}}(\omega_n; x)$ denotes a non-parametric estimator for $\bar{F}(\omega_n; x)$:

$$\hat{\bar{F}}(\omega_n; x) := \frac{T_n^{(0)}(x, K)}{\hat{g}_n(x)} \quad \text{with} \quad \hat{g}_n(x) := \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x - X_i),$$
(7)

being a classical kernel density estimator for g. The estimators can also be motivated as generalisations of estimators that were initially presented in the univariate independent and identically distributed (i.i.d.) setting, e.g. those by Goegebeur *et al.* (2010), to the regression case.

We can now establish the limiting distributions of $\hat{\theta}_n^{(1)}(x; t, K_1, K_2)$ and $\hat{\theta}_n^{(2)}(x; t, K_1, K_2)$, when appropriately normalized. These are given in Theorems 1 and 2, respectively.

Theorem 1 *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sample of independent copies of the random vector (X, Y) where the conditional quantile function of Y given $X = x$ satisfies (3) with X a random vector having density function g, and assume $(\mathcal{R})$, $(\mathcal{G})$, $(\mathcal{F})$ hold and kernel functions K_1 and K_2 satisfying $(\mathcal{K})$. For all $x \in \mathbb{R}^q$ where $g(x) > 0$, we have that if $h_n \to 0$, $\omega_n \to \infty$ for $n \to \infty$, with $nh_n^q \bar{F}(\omega_n; x) \to \infty$, $\sqrt{nh_n^q \bar{F}(\omega_n; x)}b(-\ln \bar{F}(\omega_n; x); x) \to \lambda$, $nh_n^{q+2}\bar{F}(\omega_n; x) \to 0$, $\sqrt{nh_n^q \bar{F}(\omega_n; x)}/(-\ln(\bar{F}(\omega_n; x)))^{1-\varepsilon} \to 0$ for some small $\varepsilon > 0$, and $nh_n^q \bar{F}(\omega_n; x)\Phi^2(\omega_n, h_n; x) \to 0$, then*

$$\sqrt{nh_n^q \bar{F}(\omega_n; x)g(x)}\, [\hat{\theta}_n^{(1)}(x; t, K_1, K_2) - \theta(x)] \xrightarrow{\mathcal{D}} N\left(\lambda\sqrt{g(x)},\right.$$

$$\left. \frac{\theta^2(x)}{t^2\Gamma^2(t+1)}\left[\Gamma(2t+1)\|K_1\|_2^2 + \Gamma^2(t+1)\|K_2\|_2^2 - 2\Gamma^2(t+1)\|K_1 K_2\|_1\right]\right).$$

Theorem 2 *Under the same assumptions as in Theorem 1, we have*

$$\sqrt{nh_n^q \bar{F}(\omega_n; x)g(x)}\, [\hat{\theta}_n^{(2)}(x; t, K_1, K_2) - \theta(x)] \xrightarrow{\mathcal{D}} N\left(\lambda\sqrt{g(x)},\right.$$

$$\left. \frac{\theta^2(x)\Gamma(2t+1)}{(t+1)\Gamma^2(t+1)}\left[2(2t+1)\|K_1\|_2^2 + (t+1)\|K_2\|_2^2 - 2(2t+1)\|K_1 K_2\|_1\right]\right).$$

Note that the mean of the limiting distribution is not necessarily centered at zero, meaning that the estimators may show some asymptotic bias, which is common for estimators of tail parameters. Similar to the estimation of univariate Weibull-type tails (see e.g. Goegebeur *et al.*, 2010, Gardes and Girard, 2008a) we have that, apart from the factor $\sqrt{g(x)}$, the mean of the asymptotic distribution only depends on λ, and not on other distributional parameters, nor on t. As expected, the asymptotic variance is increasing in $\theta(x)$ (i.e. when the tail of the conditional response distribution becomes more heavy), and also depends on the kernel functions K_1 and K_2, as well as on the tuning parameter t. In the common situation where the kernel functions K_1 and K_2 are assumed to be equal, the asymptotic variance expressions in Theorems 1 and 2 simplify and are given by

$$\begin{aligned}
\mathbb{A}\mathrm{Var}_1 &= \frac{\theta^2(x)\|K\|_2^2}{t^2}\left(\frac{\Gamma(2t+1)}{\Gamma^2(t+1)} - 1\right), \\
\mathbb{A}\mathrm{Var}_2 &= \theta^2(x)\|K\|_2^2\frac{\Gamma(2t+1)}{\Gamma^2(t+1)},
\end{aligned}$$

respectively. If we compare the two variances (without the factor $\theta^2(x)\|K\|_2^2$ which is in common), we can see that no estimator performs uniformly best, although $\hat{\theta}_n^{(1)}(x; t, K, K)$ outperforms $\hat{\theta}_n^{(2)}(x; t, K, K)$ over a wide range of values of the parameter t.

3 Upper extreme quantile estimation

In this section we study the estimation of upper extreme conditional quantiles. Let $\hat{Q}(p; x)$ denote the kernel estimator of the conditional quantile function $Q(p; x)$, i.e.

$$\hat{Q}(p; x) := \inf\{y : \hat{\bar{F}}(y; x) \leq 1 - p\}, \ p \in (0, 1),$$

where $\hat{\bar{F}}$ is defined in (7).

We consider the behavior of $\hat{Q}(1 - \alpha_n; x)$ for $\alpha_n \to 0$ as $n \to \infty$. In the first instance we work under an intermediate order condition $\sqrt{nh_n^q\alpha_n}(-\ln\alpha_n) \to \infty$. Let $V(y; x) := y^{1/\theta(x)}\ell^*(y; x)$, so that $\bar{F}(y; x) = \exp(-V(y; x))$ and assume that $\ell^*(y; x)$ is a normalized slowly varying function (see Bingham *et al.*, 1987, p 15), i.e.

$$\ell^*(y; x) = c(x)e^{\int_1^y \frac{\varepsilon(u; x)}{u} du},$$

for $y \geq 1$, where $c(x) > 0$ and $\varepsilon(t; x) \to 0$ as $t \to \infty$. In terms of $V(y; x)$ we have that

$$V(y; x) = c(x)e^{\int_1^y \frac{\tilde{\varepsilon}(u; x)}{u} du},$$

where $\tilde{\varepsilon}(t; x) \to 1/\theta(x)$ as $t \to \infty$. In the sequel we will say for short that $V(y; x)$ is normalized regularly varying. Note that in this case

$$\frac{yV'(y; x)}{V(y; x)} = \tilde{\varepsilon}(y; x), \quad a.e.$$

Theorem 3 *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sample of independent copies of the random vector (X, Y) where the conditional survival function of Y given $X = x$ satisfies (2) with a normalized regularly varying function $V(y; x)$, X a random vector having density function g, and assume $(\mathcal{G})$, $(\mathcal{F})$ and $(\mathcal{K})$ hold. For all $x \in \mathbb{R}^q$ where $g(x) > 0$, we have that if $h_n \to 0$, $\alpha_n \to 0$ for $n \to \infty$, with $nh_n^q\alpha_n \to \infty$, $nh_n^{q+2}\alpha_n \to 0$, and $nh_n^q\alpha_n\Phi^2(Q(1 - \alpha_n; x)(1 + o(1)), h_n; x) \to 0$ then*

$$(-\ln\alpha_n)\sqrt{nh_n^q\alpha_n g(x)}\left[\frac{\hat{Q}(1 - \alpha_n; x)}{Q(1 - \alpha_n; x)} - 1\right] \xrightarrow{\mathcal{D}} N(0, \theta^2(x)\|K\|_2^2).$$

Similarly to the Weibull-tail coefficient estimators (see Theorems 1 and 2), the asymptotic variance of the extreme quantiles is increasing in $\theta(x)$ and depends on the kernel function K. The condition $nh_n^q\alpha_n \to \infty$ in Theorem 3 implies that α_n should be of a larger order than $1/(nh_n^q)$, and hence, ultimately, $\alpha_n > 1/n$. As a consequence, the estimator $\hat{Q}(1 - \alpha_n; x)$ cannot be used to extrapolate beyond the data range.

For the purpose of estimating quantiles that are further in the upper tail than the $(1 - \alpha_n)-$ quantile, we propose an estimator of $Q(1 - \beta_n; x)$ with $\tau_n := (-\ln\beta_n)/(-\ln\alpha_n) \to \tau \in (1, \infty)$ if $\alpha_n \to 0$, that is in the spirit of the Weissman estimator (Weissman, 1978):

$$\hat{Q}_W(1 - \beta_n; x) := \hat{Q}(1 - \alpha_n; x)\tau_n^{\hat{\theta}(x)},$$

where $\hat{\theta}(x)$ is either $\hat{\theta}_n^{(1)}(x; t, K_1, K_2)$ or $\hat{\theta}_n^{(2)}(x; t, K_1, K_2)$.

Theorem 4 *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sample of independent copies of the random vector (X, Y) where the conditional survival function of Y given $X = x$ satisfies (2) with a normalized regularly varying function $V(y; x)$, X a random vector having density function g, and assume $(\mathcal{R})$, $(\mathcal{G})$, $(\mathcal{F})$ hold and kernel functions K_1 and K_2 satisfying $(\mathcal{K})$. For all $x \in \mathbb{R}^q$ where $g(x) > 0$, we have that if $h_n \to 0$, $\alpha_n \to 0$, $\tau_n \to \tau \in (1, \infty)$ for $n \to \infty$, with $nh_n^q \alpha_n \to \infty$ and, for some small $\varepsilon > 0$,*

$$\sqrt{nh_n^q \alpha_n} \max \left\{ b(-\ln \alpha_n; x), h_n, \frac{1}{(-\ln \alpha_n)^{1-\varepsilon}}, \Phi(Q(1 - \alpha_n; x)(1 + o(1)), h_n; x) \right\} \to 0,$$

then

$$\frac{\sqrt{nh_n^q \alpha_n g(x)}}{\ln \tau_n} \left[\frac{\hat{Q}_W(1 - \beta_n; x)}{Q(1 - \beta_n; x)} - 1 \right] \xrightarrow{\mathcal{D}} N(0, V^2(x)),$$

where $V^2(x)$ denotes the asymptotic variance of either $\hat{\theta}_n^{(1)}(x; t, K_1, K_2)$ or $\hat{\theta}_n^{(2)}(x; t, K_1, K_2)$.

Remark that the extreme quantile estimator inherits the asymptotic behavior of the estimator for $\theta(x)$ that was used. For what concerns the improvement of $\hat{Q}_W(1 - \beta_n; x)$ over $\hat{Q}(1 - \alpha_n; x)$ in terms of extrapolation: note that one can take e.g. $\beta_n = \alpha_n^{\tau^*}$, for some $\tau^* > 1$, leading to $\beta_n > 1/n^{\tau^*}$, which relaxes the earlier constraint that $\alpha_n > 1/n$. By taking τ^* large one thus improves considerably the order of the quantile that can be estimated. Note also that the only condition that we impose on β_n, namely $(-\ln \beta_n)/(-\ln \alpha_n) \to \tau \in (1, \infty)$, simplifies the interpretation in terms of extrapolation compared to Daouia *et al.* (2013) where they impose several conditions that mix the parameters α_n and β_n together with several functions involving the conditional extreme value index, the conditional quantile and an auxiliary function. In that case, it is very difficult to obtain information about the possible rate of β_n.

4   Simulations

In this section we illustrate our methodology with a simulation experiment. For the practical implementation of our estimators we have to determine both the bandwidth parameter h_n and the threshold ω_n. For the latter, we take, as usual in extreme value statistics, the $(k + 1)$-th largest response observation in the ball $B(x, h_n)$. There is, in fact, some discrepancy between the theory, established in case of a fixed, i.e. non-random, threshold and practical use with a data-driven one. This is often observed in extreme value statistics, see for instance, Smith (1987) and Davison and Smith (1990), in the framework of GPD modelling of excesses, or Wang and Tsai (2009) and Goegebeur *et al.* (2014b), for Pareto-type models. Instead of a nonrandom threshold one could have worked with a random threshold, as was done e.g. by Stupfler (2013) in the general max-domain of attraction. However, this approach with a random threshold would make the theoretical derivations much more difficult. Also, in the general max-domain of attraction, Goegebeur *et al.* (2014a) observed that working with a nonrandom threshold can lead to the same asymptotic distributions as the one obtained under a random threshold (as in Stupfler, 2013), though the practical performance of the approach based on a random threshold is not better. One can expect that similar findings will also apply to the present paper. The

selection of (h_n, k) is carried out using a data driven method which does not require any prior knowledge about the function $\theta(x)$. This approach is based on a two step procedure, where we first select the bandwidth parameter h_n using a cross-validation criterion

$$h_{cv} := \operatorname*{argmin}_{h_n \in \mathcal{H}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\mathbf{1}\{Y_i \leq Y_j\} - \hat{F}_{n,-i}(Y_j; X_i) \right)^2, \qquad (8)$$

where $\mathcal{H}$ is a grid of values for h_n and

$$\hat{F}_{n,-i}(y; x) := \frac{\sum_{j=1, j \neq i}^{n} K_{h_n}(x - X_j) \mathbf{1}\{Y_j \leq y\}}{\sum_{j=1, j \neq i}^{n} K_{h_n}(x - X_j)}.$$

This criterion was introduced in Yao (1999), implemented by Gannoun *et al.* (2002) and considered in an extreme value context by Daouia *et al.* (2011, 2013). Using this bandwidth, we compute in the second step, the estimates for $\theta(x)$ with $k = 5, \ldots, k_{\max}$, where $k_{\max}$ is an appropriately chosen value. The median of these estimates is then the estimate we use for $\theta(x)$. From the simulation results reported in Gardes and Girard (2008a) and Goegebeur *et al.* (2010) for the estimation of θ in the univariate framework, we deduce that the parameter $k_{\max}$ should be chosen relatively small compared to the typical number of observations in $B(x, h_{cv})$, because for Weibull-type distributions the bias of the tail estimators is in some cases important. Concerning the estimation of upper extreme conditional quantiles, we use the estimate for $\theta(x)$ we just obtained and compute the extreme conditional quantiles for $k = 5, \ldots, k_{\max}$. We use again the median of these values as our final conditional quantile estimate.

We report the results for two conditional distributions of Y given $X = x$, which are of Weibull-type:

- The strict Weibull distribution $\mathcal{W}(\alpha(x), \lambda)$,

$$1 - F(y; x) = e^{-\lambda y^{\alpha(x)}}, \quad y > 0; \alpha(x), \lambda > 0,$$

for which $\theta(x) = 1/\alpha(x)$ and $\rho(x) = -\infty$. We consider the case $\lambda = 1$.

- The extended Weibull distribution $\mathcal{EW}(\alpha(x), \beta)$ (Klüppelberg and Villaseñor, 1993),

$$1 - F(y; x) = r(y) e^{-y^{\alpha(x)}},$$

where $\alpha(x) > 0$ and $r(y)$ is a regularly varying function at infinity with index β. Here $\theta(x) = 1/\alpha(x)$, $b(y; x) = -\beta \ln y / [\alpha^2(x) y]$ and $\rho(x) = -1$. We choose $r(y) = 1/y$.

In the simulation experiment we also considered some other conditional Weibull-type distributions, like the Gaussian, Gamma and perturbed Weibull distributions, but these lead to similar results and therefore, for brevity, we omit them from the paper.

For the Weibull-tail coefficient $\theta(x)$ and the covariate distribution we consider two settings:

- Setting 1: $X \sim U(0,1)$ and

$$\theta(x) = \frac{1}{2}\left(\frac{1}{10} + \sin(\pi x)\right)\left(\frac{11}{10} - \frac{1}{2}\exp\left(-64\left(x - \frac{1}{2}\right)^2\right)\right).$$

This function was also considered in Daouia *et al.* (2011), though in the framework of Pareto-type distributions. The function is differentiable with several stationary points.

- Setting 2: $X \sim \text{Beta}(1.25, 1.5)$ and

$$\begin{aligned}\theta(x) &= \frac{1}{4}\left\{1 + \exp(-60(x - 1/4)^2)\mathbf{1}\{3x \in (0,1]\} + \exp(-5/12)\mathbf{1}\{3x \in (1,2]\}\right.\\&\quad \left.+(5 - 6x)(\exp(-5/12)\mathbf{1}\{3x \in (2,5/2]\} - \mathbf{1}\{3x \in (5/2,3)\})\right\},\end{aligned}$$

which was proposed by Gardes and Stupfler (2013), also in the framework of Pareto-type modelling. This function is considered because it has a smooth but non-constant part, a constant part and also some points where it is continuous but not differentiable.

For all the distributions we simulate $N = 500$ samples of size $n = 1000$. As measures of efficiency we use the empirical root mean squared error and the bias, computed over points $x_1, \ldots, x_M$, regularly spread over the covariate space, as follows

$$\text{RMSE}(\widehat{\xi}(.)) := \sqrt{\frac{1}{MN}\sum_{m=1}^{M}\sum_{j=1}^{N}\left(\widehat{\xi}_j(x_m) - \xi(x_m)\right)^2},$$

$$\text{Bias}(\widehat{\xi}(.)) := \frac{1}{M}\sum_{m=1}^{M}\left|\frac{1}{N}\sum_{j=1}^{N}\widehat{\xi}_j(x_m) - \xi(x_m)\right|,$$

where $\widehat{\xi}_j(x_m)$ is the estimator for the tail parameter $\xi(x_m)$, either the Weibull-tail coefficient or an upper extreme conditional quantile, from simulation run j at covariate value x_m. Note that in the bias calculation we use the absolute value to avoid that positive and negative biases at different values of x cancel each other out.

In the estimation of $\theta(x)$ we consider the estimators $\hat{\theta}_n^{(1)}(x; t, K_1, K_2)$ and $\hat{\theta}_n^{(2)}(x; t, K_1, K_2)$ for different values of t, and where both kernel functions are taken as the bi-quadratic kernel function

$$K(x) = \frac{15}{16}\left(1 - x^2\right)^2\mathbf{1}\{x \in [-1, 1]\}.$$

The choice of the kernel function is not very crucial for the practical performance of our estimators. Note that we rerun our simulations for other choices of kernel functions, like e.g. the triweight kernel function $K(x) = 35(1 - x^2)^3/32, x \in [-1, 1]$, and the results do not change a lot. To keep the length of the paper under control, we only report the results for this bi-quadratic function. The bandwidth h_n is selected using the cross validation criterion in (8) on a grid of $h_n \in \{0.05, 0.075, \ldots, 0.15\}$, whereafter the estimate of $\theta(x)$ is computed as described above using $k_{\max} = 25$. Concerning the estimation of upper extreme conditional quantiles, we examine

the behaviour of $\hat{Q}_W(1 - \beta_n; x)$ obtained with $\hat{\theta}_n^{(1)}(x; t, K, K)$ and $\hat{\theta}_n^{(2)}(x; t, K, K)$, denoted as $\hat{Q}_W^{(1)}(1 - \beta_n; x)$ and $\hat{Q}_W^{(2)}(1 - \beta_n; x)$, respectively, and quantile levels $\beta_n = 1/1200$, $1/2000$ and $1/4000$. The considered quantile orders imply a rather serious extrapolation since the local sample sizes are typically much smaller than n. These quantile estimators are compared with the extreme quantile estimator from Daouia $et~al.$ (2013), denoted $\tilde{q}_n^{RP}(\beta_n|x)$, which is developed for the full max-domain of attraction, and therefore it also applies in our context of conditional Weibull-type tails. Concerning the latter we considered constant and linear weights, leading to the estimators $\tilde{q}_n^{RP,1}(\beta_n|x)$ and $\tilde{q}_n^{RP,2}(\beta_n|x)$, respectively, and both were implemented with $J = 3$ and $J = 4$, where J can be seen as the number of elemental Pickands statistics the estimator is based on. For more details we refer to the Daouia $et~al.$ (2013) paper. The data driven method is the one where h and k are selected separately, which is consistent with the data driven method considered for our estimators. Note that for the estimation of the Weibull-tail coefficient $\theta(x)$ there is to date no alternative estimation procedure available as basis for comparison.

Thus, summarised, in the simulation experiment we compared the small sample behaviour of the Weibull-tail coefficient estimators $\hat{\theta}_n^{(1)}(x; t, K, K)$ and $\hat{\theta}_n^{(2)}(x; t, K, K)$, and that of the extreme conditional quantile estimators $\hat{Q}_W^{(1)}(1 - \beta_n; x)$, $\hat{Q}_W^{(2)}(1 - \beta_n; x)$, $\tilde{q}_n^{RP,1}(\beta_n|x)$ and $\tilde{q}_n^{RP,2}(\beta_n|x)$. The estimator $\hat{\theta}_n^{(1)}(x; t, K, K)$ outperforms $\hat{\theta}_n^{(2)}(x; t, K, K)$ in terms of bias and RMSE, for all values of t and all distributions, and similarly, the extreme quantile estimator $\hat{Q}_W^{(1)}(1 - \beta_n; x)$ has a performance superior to that of $\hat{Q}_W^{(2)}(1 - \beta_n; x)$. For what concerns the Daouia $et~al.$ (2013) estimators, $\tilde{q}_n^{RP,1}(\beta_n|x)$ with $J = 4$ has the best performance. For brevity, in the paper we only comment on the performance of the best estimators, and draw the following conclusions

- In Tables 1 and 2 we show the performance of $\hat{\theta}_n^{(1)}(x; t, K, K)$ and $\hat{Q}_W^{(1)}(1 - \beta_n; x)$, obtained with the data driven strategy on Settings 1 and 2, respectively. The results indicate that in most cases the best estimation results are obtained for small values of the tuning parameter t. As expected, the behaviour of the extreme quantile estimator deteriorates as more extreme quantile orders are considered, though the increase in bias and RMSE is modest.

- Comparison of the results in Tables 1 and 2 with Table 3 indicates that the estimator $\hat{Q}_W^{(1)}(1 - \beta_n; x)$ is better than $\tilde{q}_n^{RP,1}(\beta_n|x)$ in terms of RMSE for all values of β_n and t. In terms of bias it is better for almost all β_n and t.

From these simulations, one can say that the proposed Weissman-type extreme conditional quantile estimator is competitive compared to the Daouia $et~al.$ (2013) estimator, which is to date the only alternative. However, this result is not completely unexpected since the Daouia $et~al.$ (2013) estimator is more general in that it was proposed for the general max-domain of attraction, whereas ours is specific for the class of Weibull-type distributions.

5   Case study: sea storm data

To illustrate the usefulness and practical relevance of the model considered and the methodology developed in this paper we will use the sea level data presented in Chapter 6 of de Haan and

Ferreira (2006); see also de Haan and de Ronde (1998) for a more elaborate discussion. This data set consists of measurements of wave height (HmO) and still water level (SWL), both expressed in meters, recorded during 828 storm events that are relevant for the seawall called Pettemer Zeewering in the province of North Holland (The Netherlands). The data have been extensively analysed in the extreme value literature. For instance, in Draisma *et al.* (1999) the focus was on estimating the tail of the wave height distribution, ignoring the information in the variable still water level. Their analysis indicated that the HmO distribution belongs to the Gumbel class. On the other hand, de Haan and de Ronde (1998) and Draisma *et al.* (2004) took a multivariate approach and estimated the tail dependence between the two variables, i.e. when both variables are simultaneously large. Our approach is complementary to these earlier analyses in that we want to estimate tail parameters of the HmO distribution given a value for the variable SWL, where the latter is not necessarily extreme.

The scatterplot of the data is shown in Figure 1 (a), and clearly indicates that higher waves tend to occur together with higher still water levels. For engineers it can be of interest to study the distribution of wave heights at a given value of still water level, e.g. to decide on the height of a sea wall or, related to this, to assess the danger of overflowing an existing sea wall. We illustrate the adequacy of the Weibull-type model (2) for the variable HmO given that $SWL = 0.6$, by the Weibull quantile plot of the HmO measurements with corresponding SWL in a neighborhood of $SWL = 0.6$, say in the interval [0.5,0.7]; see Figure 1 (b). The Weibull quantile plot clearly becomes linear in the largest observations, which supports the hypothesis of an underlying Weibull-type model for the conditional response distribution. Similar quantile plots were obtained at other positions in the covariate space. Obviously, estimators for parameters related to the conditional distribution of HmO given SWL should take into account that the available SWL measurements are in fact realisations of a random covariate. Based on these considerations we can conclude that the methodology developed in this paper is appropriate for analysing the sea storm data.

As a first step we focus on the estimation of the conditional Weibull-tail coefficient of the variable HmO, considered as a function of SWL. In Figure 1 (c) we show the tail coefficient estimates $\hat{\theta}_n^{(1)}(SWL; 0.2, K, K)$, where K is the bi-quadratic kernel function, as a function of SWL. We focus here only on the estimator $\hat{\theta}_n^{(1)}(SWL; t, K, K)$, since in the simulations it was found to be superior to $\hat{\theta}_n^{(2)}(SWL; t, K, K)$ in terms of bias and RMSE (for all values of t). The value of t used, $t = 0.2$, can also be motivated from the simulation results. The tuning parameters h and k were selected according to the data driven method with $h_n \in \{0.05, 0.075, \ldots, 0.15\}$ and $k_{\max} = 25$, as described in Section 4. The figure indicates that the estimates for the tail coefficient generally follow the pattern in the data in that the estimates tend to be larger at SWL positions where the extreme HmO measurements show larger spacings. To illustrate the extra flexibility of our approach, we also performed a univariate analysis of the tail of the HmO distribution, thus ignoring the information in SWL, using the mean-excess based estimator for θ proposed in Goegebeur and Guillou (2011). Figure 1 (d) shows these univariate estimates for the Weibull-tail coefficient of the variable HmO as a function of k. This plot suggests a stable estimate of θ for k values between 50 and 100, with median 0.42. This median value is also

shown as the solid horizontal reference line in Figure 1 (c), and we also show the univariate 95% bootstrap interval (based on 1000 bootstrap samples). The univariate estimate thus provides some average tail coefficient estimate, and from the bootstrap interval we can conclude that it is clearly not able to provide an adequate description of the conditional tails.

Next, we consider the estimation of upper extreme quantiles of the HmO distribution conditional on a given value of SWL. In Figure 1 (e) we show the Weissman-type quantile estimate $\hat{Q}_W^{(1)}(1 - 1/1200; SWL)$ (solid line) and $\hat{Q}_W^{(1)}(1 - 1/4000; SWL)$ (dotted line) as a function of SWL. This estimate represents the wave height that will be exceeded on average once in 1200 and 4000 storms, respectively, for a given value of SWL. Again the estimate follows the pattern in the data, and it clearly extrapolates beyond the data range, as it should. As a simple goodness-of-fit check we also estimated a less extreme conditional quantile, namely the conditional quantile 1-1/50, and counted the number of observations exceeding that quantile. The proportion exceeding was 0.024, which is in good agreement with the theoretical conditional exceedance probability 0.02. In panel (f) of Figure 1 we show the estimator $\tilde{q}_n^{RP,1}(1/1200|SWL)$ (solid line) and $\tilde{q}_n^{RP,1}(1/4000|SWL)$ (dotted line) with $J = 4$ of Daouia $et\ al.$ (2013). This estimate also follows the pattern in the data reasonably well, though it is generally located slightly below our extreme conditional quantile estimate and also exhibits a larger variability compared to our method. Note also that due to data sparsity the Daouia $et\ al.$ (2013) estimate could not be computed for the very small and large values of SWL. For some values of SWL the estimate $\tilde{q}_n^{RP,1}(1/4000|SWL)$ is slightly below $\tilde{q}_n^{RP,1}(1/1200|SWL)$, which is due to the tuning parameter selection method.

6   Conclusion

We considered the estimation of the tail coefficient and upper extreme conditional quantiles for Weibull-type distributions when there are random covariates. Two families of estimators were introduced and their asymptotic properties were derived. In extreme value statistics the bias of estimators for tail quantities is often an issue. In future research we will focus on the development of bias-corrected estimators, and on related robustness issues.

Appendix

In this appendix, we give the proofs of our main results.

First, we study the local behaviour of the conditional expectation of a power transformed excess.

Lemma 1 *Case (i), $t = 0$:*

$$m(\omega_n, 0; x) = \bar{F}(\omega_n; x).$$

Case (ii), $t > 0$: assume (3) and ($\mathcal{R}$), then for some small $\varepsilon > 0$ and $\omega_n \to \infty$ we have that

$$m(\omega_n, t; x) = \frac{\bar{F}(\omega_n; x)}{(-\ln \bar{F}(\omega_n; x))^t} \theta^t(x)\Gamma(t+1)\left\{1 + b(-\ln \bar{F}(\omega_n; x); x)\frac{t}{\theta(x)}(1 + o(1))\right.$$
$$\left. + O\left(\frac{1}{(-\ln \bar{F}(\omega_n; x))^{1-\varepsilon}}\right)\right\}.$$

Proof of Lemma 1

The case $t = 0$ is trivial so we only focus on the case where $t > 0$. Let $p_n := F(\omega_n; x)$ and $\tilde{p}_n = 1 - (1 - p_n)/\ln(e/(1 - p_n))$. Note that $p_n \leq \tilde{p}_n \leq 1$. From the inverse probability integral transform, and denoting $V \sim U(0,1)$, we have that

$$m(\omega_n, t; x) = \mathbb{E}[(\ln Q(V; x) - \ln Q(p_n; x))_+^t]$$
$$= \int_{p_n}^{\tilde{p}_n} (\ln Q(v; x) - \ln Q(p_n; x))^t dv + \int_{\tilde{p}_n}^{1} (\ln Q(v; x) - \ln Q(p_n; x))^t dv$$
$$=: I_1 + I_2.$$

Concerning I_1, from (3) and ($\mathcal{R}$), we obtain

$$I_1 = (1 - p_n) \int_{\frac{1 - \tilde{p}_n}{1 - p_n}}^{1} \left[\theta(x)\ln\left(1 + \frac{-\ln z}{-\ln(1 - p_n)}\right)\right.$$
$$\left. + b(-\ln(1 - p_n); x)D_{\rho(x)}\left(1 + \frac{-\ln z}{-\ln(1 - p_n)}\right)(1 + o(1))\right]^t dz.$$

Inspired by the inequality

$$\frac{\xi - 1}{2}x^2 \leq D_\xi(1 + x) - x \leq 0, \tag{9}$$

for $\xi \leq 0$ and $x \geq 0$, we write

$$I_1 = \frac{1 - p_n}{(-\ln(1 - p_n))^t} \int_{\frac{1 - \tilde{p}_n}{1 - p_n}}^{1} (-\ln z)^t \{\theta(x)$$
$$+ \theta(x)\frac{-\ln(1 - p_n)}{-\ln z}\left[\ln\left(1 + \frac{-\ln z}{-\ln(1 - p_n)}\right) - \frac{-\ln z}{-\ln(1 - p_n)}\right] + b(-\ln(1 - p_n); x)(1 + o(1))$$
$$+ b(-\ln(1 - p_n); x)\frac{-\ln(1 - p_n)}{-\ln z}\left[D_{\rho(x)}\left(1 + \frac{-\ln z}{-\ln(1 - p_n)}\right) - \frac{-\ln z}{-\ln(1 - p_n)}\right](1 + o(1))\right\}^t dz.$$

From (9) we deduce

$$\sup_{z \in [\frac{1 - \tilde{p}_n}{1 - p_n}, 1]} \left|\frac{-\ln(1 - p_n)}{-\ln z}\left[\ln\left(1 + \frac{-\ln z}{-\ln(1 - p_n)}\right) - \frac{-\ln z}{-\ln(1 - p_n)}\right]\right| \leq \frac{\ln(1 - \ln(1 - p_n))}{2(-\ln(1 - p_n))}$$

and hence

$$\sup_{z\in[\frac{1-\tilde{p}_n}{1-p_n},1]}\left|\frac{-\ln(1-p_n)}{-\ln z}\left[\ln\left(1+\frac{-\ln z}{-\ln(1-p_n)}\right)-\frac{-\ln z}{-\ln(1-p_n)}\right]\right|=o(1),$$

for $p_n\uparrow 1$. Similarly

$$\sup_{z\in[\frac{1-\tilde{p}_n}{1-p_n},1]}\left|\frac{-\ln(1-p_n)}{-\ln z}\left[D_{\rho(x)}\left(1+\frac{-\ln z}{-\ln(1-p_n)}\right)-\frac{-\ln z}{-\ln(1-p_n)}\right]\right|=o(1),$$

for $p_n\uparrow 1$.

By Taylor's theorem

$$
\begin{aligned}
I_1 &= \frac{1-p_n}{(-\ln(1-p_n))^t}\left\{\theta^t(x)\int_{\frac{1-\tilde{p}_n}{1-p_n}}^{1}(-\ln z)^t dz\right.\\
&\quad +t\theta^t(x)\int_{\frac{1-\tilde{p}_n}{1-p_n}}^{1}(-\ln z)^t\frac{-\ln(1-p_n)}{-\ln z}\left[\ln\left(1+\frac{-\ln z}{-\ln(1-p_n)}\right)-\frac{-\ln z}{-\ln(1-p_n)}\right]dz(1+o(1))\\
&\quad +b(-\ln(1-p_n);x)t\theta^{t-1}(x)\int_{\frac{1-\tilde{p}_n}{1-p_n}}^{1}(-\ln z)^t dz(1+o(1))\\
&\quad +b(-\ln(1-p_n);x)t\theta^{t-1}(x)\int_{\frac{1-\tilde{p}_n}{1-p_n}}^{1}(-\ln z)^t\frac{-\ln(1-p_n)}{-\ln z}\\
&\quad \left.\times\left[D_{\rho(x)}\left(1+\frac{-\ln z}{-\ln(1-p_n)}\right)-\frac{-\ln z}{-\ln(1-p_n)}\right]dz(1+o(1))\right\}\\
&=: \frac{1-p_n}{(-\ln(1-p_n))^t}\left\{I_{1,1}+I_{1,2}+I_{1,3}+I_{1,4}\right\}.
\end{aligned}
$$

For $I_{1,1}$, by straightforward calculus, we obtain

$$I_{1,1} = \theta^t(x)\Gamma(t+1)+O\left(\frac{\ln^t(1-\ln(1-p_n))}{-\ln(1-p_n)}\right).$$

Concerning $I_{1,2}$, use (9) to obtain

$$|I_{1,2}| \le \frac{t\theta^t(x)}{2(-\ln(1-p_n))}\int_{\frac{1-\tilde{p}_n}{1-p_n}}^{1}(-\ln z)^{t+1}dz(1+o(1))=O\left(\frac{1}{-\ln(1-p_n)}\right).$$

The terms $I_{1,3}$ and $I_{1,4}$ can be analyzed in a similar way and yield

$$
\begin{aligned}
I_{1,3} &= b(-\ln(1-p_n);x)t\theta^{t-1}(x)\Gamma(t+1)(1+o(1)),\\
|I_{1,4}| &= O\left(\frac{b(-\ln(1-p_n);x)}{-\ln(1-p_n)}\right).
\end{aligned}
$$

15

Collecting the terms gives

$$I_1 = \frac{1 - p_n}{(-\ln(1 - p_n))^t} \left\{ \theta^t(x)\Gamma(t+1) + b(-\ln(1 - p_n); x)t\theta^{t-1}(x)\Gamma(t+1)(1 + o(1)) \right.$$
$$\left. + O\left(\frac{\ln^t(1 - \ln(1 - p_n))}{-\ln(1 - p_n)} \right) \right\}.$$

For what concerns I_2 we introduce the tail quantile function U, defined as $U(y; x) = Q(1 - 1/y; x)$, $y > 1$, and write

$$I_2 = \int_{\tilde{p}_n}^1 \left[\ln U\left(\frac{1 - p_n}{1 - v} \frac{1}{1 - p_n}; x \right) - \ln U\left(\frac{1}{1 - p_n}; x \right) \right]^t dv.$$

The Weibull-type distributions are a subset of the Gumbel max-domain of attraction, and hence for some positive function $a(.; x)$ we have that

$$\lim_{y \to \infty} \frac{U(y; x)}{a(y; x)} \left(\ln U(uy; x) - \ln U(y; x) \right) = \ln u, \ \forall u > 0, \tag{10}$$

see for instance de Haan and Ferreira (2006), p 101. Thus

$$I_2 = \left(\frac{a\left(\frac{1}{1 - p_n}; x \right)}{U\left(\frac{1}{1 - p_n}; x \right)} \right)^t \int_{\tilde{p}_n}^1 \left[\frac{\ln U\left(\frac{1 - p_n}{1 - v} \frac{1}{1 - p_n}; x \right) - \ln U\left(\frac{1}{1 - p_n}; x \right)}{a\left(\frac{1}{1 - p_n}; x \right) / U\left(\frac{1}{1 - p_n}; x \right)} \right]^t dv =: \tilde{a}^t\left(\frac{1}{1 - p_n}; x \right) \tilde{I}_2.$$

Note that $(1 - p_n)/(1 - v) \geq (1 - p_n)/(1 - \tilde{p}_n) \geq 1$, for $v \in [\tilde{p}_n, 1)$.

Now use Corollary B.2.10 in de Haan and Ferreira (2006) p 376, to obtain for any $\varepsilon > 0$ and for some $c > 0$ that for p_n sufficiently large

$$\tilde{I}_2 \leq c \int_{\tilde{p}_n}^1 \left(\frac{1 - p_n}{1 - v} \right)^{t\varepsilon} dv = c \frac{1 - p_n}{1 - t\varepsilon} \frac{1}{(-\ln(1 - p_n))^{1 - t\varepsilon}}(1 + o(1)),$$

provided $t\varepsilon < 1$.

Finally, for $\tilde{a}(.; x)$ we use Lemma S1 (see supporting information) according to which

$$\tilde{a}\left(\frac{1}{1 - p_n}; x \right) \sim \mathbb{E}(\ln Q(V; x) - \ln Q(p_n; x) | V > p_n)$$

if $p_n \uparrow 1$. Now

$$\mathbb{E}(\ln Q(V; x) - \ln Q(p_n; x) | V > p_n) = \frac{\theta(x)}{-\ln(1 - p_n)} \int_0^1 (-\ln z) dz$$
$$+ \theta(x) \int_0^1 \left[\ln\left(1 + \frac{-\ln z}{-\ln(1 - p_n)} \right) - \frac{-\ln z}{-\ln(1 - p_n)} \right] dz$$
$$+ \frac{b(-\ln(1 - p_n); x)}{-\ln(1 - p_n)} \int_0^1 (-\ln z) dz(1 + o(1))$$
$$+ b(-\ln(1 - p_n); x) \int_0^1 \left[D_{\rho(x)}\left(1 + \frac{-\ln z}{-\ln(1 - p_n)} \right) - \frac{-\ln z}{-\ln(1 - p_n)} \right] dz(1 + o(1))$$
$$=: I_{2,1} + I_{2,2} + I_{2,3} + I_{2,4}.$$

16

With arguments similar to those used when considering I_1 we obtain

$$\begin{aligned}
I_{2,1} &= \frac{\theta(x)}{-\ln(1-p_n)}, \\
I_{2,2} &= O\left(\frac{1}{(-\ln(1-p_n))^2}\right), \\
I_{2,3} &= \frac{b(-\ln(1-p_n);x)}{-\ln(1-p_n)}(1+o(1)), \\
I_{2,4} &= O\left(\frac{b(-\ln(1-p_n);x)}{(-\ln(1-p_n))^2}\right).
\end{aligned}$$

Consequently

$$\mathbb{E}(\ln Q(V;x) - \ln Q(p_n;x)|V > p_n) = O\left(\frac{1}{-\ln(1-p_n)}\right).$$

This then gives that

$$I_2 = O\left(\frac{1-p_n}{(-\ln(1-p_n))^{t+1-\delta}}\right),$$

for some small $\delta > 0$.

Combining the results for I_1 and I_2 establishes the result of Lemma 1. $\square$

Now, to deal with the randomness of X we consider the unconditional expectation $\widetilde{m}_n(K,t;x)$ and derive in the next lemma its main asymptotic expansion under further conditions.

Lemma 2 *Assume (3), $(\mathcal{R})$, $(\mathcal{G})$, $(\mathcal{K})$ and $(\mathcal{F})$. For all $x \in \mathbb{R}^q$ where $g(x) > 0$ we have that if $\omega_n \to \infty$ and $h_n \to 0$ then*

$$\widetilde{m}_n(K,t;x) = m(\omega_n,t;x)g(x)(1 + O(h_n) + O(\Phi(\omega_n,h_n;x))).$$

Note: (i) In Lemma 2, the proof of the result for the case $t = 0$ can also be obtained without conditions (3) and $(\mathcal{R})$. (ii) Because (X_i, Y_i), $i = 1, \ldots, n$, are independent and identically distributed random vectors, we have that $\widetilde{m}_n(K,t;x) = \mathbb{E}(T_n^{(t)}(x,K))$.

The proof of Lemma 2 is elementary and therefore it is given in the supporting information.

Since the estimators $\hat{\theta}_n^{(1)}(x;t,K_1,K_2)$ and $\hat{\theta}_n^{(2)}(x;t,K_1,K_2)$ are functions of elemental kernel statistics (4), we need to derive the joint asymptotic behaviour of a vector of such statistics, when appropriately normalized. Let $\widetilde{T}_n^{(t)}(x,K) := (-\ln \bar{F}(\omega_n;x))^t\, T_n^{(t)}(x,K)$, and for some fixed positive integer J,

$$\mathbb{T}_n' := \frac{1}{\bar{F}(\omega_n;x)g(x)}\ [\widetilde{T}_n^{(t_1)}(x,K_1), \ldots, \widetilde{T}_n^{(t_J)}(x,K_J)],$$

and let Σ be a $(J \times J)$ matrix with elements

$$\sigma_{j,k} := \theta^{t_j+t_k}(x)\|K_j K_k\|_1 \Gamma(t_j + t_k + 1).$$

17

Lemma 3 *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sample of independent copies of the random vector (X, Y) where the conditional quantile function of Y given $X = x$ satisfies (3) with X a random vector having density function g, and assume $(\mathcal{R})$, $(\mathcal{G})$, $(\mathcal{F})$ hold and kernel functions $K_1, \ldots, K_J$ satisfying $(\mathcal{K})$. For all $x \in \mathbb{R}^q$ where $g(x) > 0$, we have that if $h_n \to 0$, $\omega_n \to \infty$ for $n \to \infty$, with $n h_n^q \bar{F}(\omega_n; x) \to \infty$, then*

$$\sqrt{n h_n^q \bar{F}(\omega_n; x) g(x)} \ [\mathbb{T}_n - \mathbb{E}(\mathbb{T}_n)] \xrightarrow{\mathcal{D}} N_J(0, \Sigma).$$

Proof of Lemma 3

We make use of the Cramér-Wold device (see e.g. van der Vaart, 1998, p 16), according to which it is sufficient to prove that for all $\xi \in \mathbb{R}^J$ we have that

$$\Psi_n := \sqrt{n h_n^q \bar{F}(\omega_n; x) g(x)} \ \xi'[\mathbb{T}_n - \mathbb{E}(\mathbb{T}_n)] \xrightarrow{\mathcal{D}} N(0, \xi' \Sigma \xi).$$

From straightforward calculations we obtain

$$
\begin{aligned}
\Psi_n &= \sum_{i=1}^{n} \sqrt{\frac{h_n^q}{n \bar{F}(\omega_n; x) g(x)}} \left[\sum_{j=1}^{J} \xi_j (-\ln \bar{F}(\omega_n; x))^{t_j} K_{j,h_n}(x - X_i)(\ln Y_i - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_i > \omega_n\} \right. \\
&\quad \left. - \mathbb{E}\left(\sum_{j=1}^{J} \xi_j (-\ln \bar{F}(\omega_n; x))^{t_j} K_{j,h_n}(x - X_i)(\ln Y_i - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_i > \omega_n\} \right) \right] \\
&=: \sum_{i=1}^{n} W_i.
\end{aligned}
$$

Note that $W_1, \ldots, W_n$ are i.i.d. random variables, and hence $\mathbb{V}\text{ar}(\Psi_n) = n \mathbb{V}\text{ar}(W_1)$. Now

$$\mathbb{V}\text{ar}(W_1) = \frac{h_n^q}{n \bar{F}(\omega_n; x) g(x)} \sum_{j=1}^{J} \sum_{k=1}^{J} \xi_j \xi_k (-\ln \bar{F}(\omega_n; x))^{t_j + t_k} \mathbb{C}_{j,k},$$

where

$$\mathbb{C}_{j,k} := \mathbb{C}\text{ov}\left(K_{j,h_n}(x - X_1)(\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\}, K_{k,h_n}(x - X_1)(\ln Y_1 - \ln \omega_n)_+^{t_k} \mathbf{1}\{Y_1 > \omega_n\} \right).$$

We then have

$$
\begin{aligned}
\mathbb{C}_{j,k} &= \frac{\|K_j K_k\|_1}{h_n^q} \ \mathbb{E}\left[\frac{1}{h_n^q \|K_j K_k\|_1} K_j\left(\frac{x - X_1}{h_n}\right) K_k\left(\frac{x - X_1}{h_n}\right)(\ln Y_1 - \ln \omega_n)_+^{t_j + t_k} \mathbf{1}\{Y_1 > \omega_n\} \right] \\
&\quad - \mathbb{E}\left[K_{j,h_n}(x - X_1)(\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\} \right] \mathbb{E}\left[K_{k,h_n}(x - X_1)(\ln Y_1 - \ln \omega_n)_+^{t_k} \mathbf{1}\{Y_1 > \omega_n\} \right].
\end{aligned}
$$

Using Lemmas 1 and 2 we obtain that

$$\mathbb{C}_{j,k} = \frac{\|K_j K_k\|_1}{h_n^q} \frac{\bar{F}(\omega_n; x)}{(-\ln \bar{F}(\omega_n; x))^{t_j + t_k}} \ g(x) \theta^{t_j + t_k}(x) \Gamma(t_j + t_k + 1)(1 + o(1)),$$

18

and consequently $\mathbb{V}\mathrm{ar}(\Psi_n) = \xi'\Sigma\xi(1 + o(1))$.

To establish the asymptotic normality of Ψ_n we verify Lyapounov's criterion for triangular arrays of random variables, see e.g. Billingsley (1995), p 362. In the present context this simplifies to proving that $\sum_{i=1}^n \mathbb{E}|W_i|^3 = n\mathbb{E}|W_1|^3 \to 0$. We have

$$
\mathbb{E}|W_1|^3 \leq \left(\frac{h_n^q}{n\bar{F}(\omega_n;x)g(x)}\right)^{3/2}
$$

$$
\times \left\{ \mathbb{E}\left[\left(\sum_{j=1}^J |\xi_j|(-\ln\bar{F}(\omega_n;x))^{t_j}K_{j,h_n}(x-X_1)(\ln Y_1 - \ln\omega_n)_+^{t_j}\mathbf{1}\{Y_1 > \omega_n\}\right)^3\right]\right.
$$

$$
+3\mathbb{E}\left[\left(\sum_{j=1}^J |\xi_j|(-\ln\bar{F}(\omega_n;x))^{t_j}K_{j,h_n}(x-X_1)(\ln Y_1 - \ln\omega_n)_+^{t_j}\mathbf{1}\{Y_1 > \omega_n\}\right)^2\right]
$$

$$
\times\mathbb{E}\left[\sum_{j=1}^J |\xi_j|(-\ln\bar{F}(\omega_n;x))^{t_j}K_{j,h_n}(x-X_1)(\ln Y_1 - \ln\omega_n)_+^{t_j}\mathbf{1}\{Y_1 > \omega_n\}\right]
$$

$$
\left. +4\left[\mathbb{E}\left(\sum_{j=1}^J |\xi_j|(-\ln\bar{F}(\omega_n;x))^{t_j}K_{j,h_n}(x-X_1)(\ln Y_1 - \ln\omega_n)_+^{t_j}\mathbf{1}\{Y_1 > \omega_n\}\right)\right]^3\right\}.
$$

Again by using Lemmas 1 and 2 we obtain that

$$
\mathbb{E}|W_1|^3 = O\left(\left(n\sqrt{nh_n^q\bar{F}(\omega_n;x)}\right)^{-1}\right), \tag{11}
$$

and hence, under the conditions of the lemma $n\mathrm{E}|W_1|^3 \to 0$. $\square$

In order to obtain the limiting distribution of $\hat{\theta}_n^{(i)}(x;t,K_1,K_2)$, $i = 1, 2$, we also need a result concerning $\hat{\bar{F}}(\omega_n;x)$ as defined in (7). Daouia et $al.$ (2011) studied the asymptotic behaviour of $\hat{\bar{F}}(\omega_n;x)$ in the framework of conditional Pareto-type tails, whereas in Daouia et $al.$ (2013) its behaviour was studied for the general max-domain of attraction, though assuming that $\bar{F}(y;x)$ is twice differentiable. Lemma 4 below is essentially a simplified version of Proposition 1 in Daouia et $al.$ (2013) since we only consider a single extreme level ω_n, although we do not assume differentiability of $\bar{F}(y;x)$ nor a max-domain of attraction condition.

Lemma 4 Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a $sample$ of $i.i.d.$ $random$ $vectors,$ and $assume$ $(\mathcal{G})$, $(\mathcal{F})$ and $(\mathcal{K})$ $hold.$ For all $x \in \mathbb{R}^q$ $where$ $g(x) > 0$, we $have$ $that$ if $h_n \to 0$, $\omega_n \to \infty$ for $n \to \infty$, $with$ $nh_n^q\bar{F}(\omega_n;x) \to \infty$, $nh_n^{q+2}\bar{F}(\omega_n;x) \to 0$, and $nh_n^q\bar{F}(\omega_n;x)\Phi^2(\omega_n, h_n;x) \to 0$ $then$

$$
\sqrt{nh_n^q\bar{F}(\omega_n;x)g(x)} \left[\frac{\hat{\bar{F}}(\omega_n;x)}{\bar{F}(\omega_n;x)} - 1\right] \xrightarrow{\mathcal{D}} N(0, \|K\|_2^2).
$$

For the proof of Lemma 4 we refer to the supporting information.

Proof of Theorem 1

First we consider the estimator $\hat{\theta}_n^{(1)}(x; t, K_1, K_2)$ assuming that $\bar{F}(\omega_n; x)$ is known, and introduce

$$\tilde{\theta}_n^{(1)}(x; t, K_1, K_2) := \left(\frac{(-\ln \bar{F}(\omega_n; x))^t T_n^{(t)}(x, K_1)}{\Gamma(t+1) T_n^{(0)}(x, K_2)} \right)^{1/t}.$$

Write

$$r_n \left(\frac{(-\ln \bar{F}(\omega_n; x))^t T_n^{(t)}(x, K_1)}{\Gamma(t+1) T_n^{(0)}(x, K_2)} - \theta^t(x) \right)$$

$$= \frac{\bar{F}(\omega_n; x) g(x)}{T_n^{(0)}(x, K_2)} \left\{ r_n \left[\frac{(-\ln \bar{F}(\omega_n; x))^t T_n^{(t)}(x, K_1)}{\Gamma(t+1) \bar{F}(\omega_n; x) g(x)} - \mathbb{E} \left(\frac{(-\ln \bar{F}(\omega_n; x))^t T_n^{(t)}(x, K_1)}{\Gamma(t+1) \bar{F}(\omega_n; x) g(x)} \right) \right] \right.$$

$$- \theta^t(x) r_n \left[\frac{T_n^{(0)}(x, K_2)}{\bar{F}(\omega_n; x) g(x)} - \mathbb{E} \left(\frac{T_n^{(0)}(x, K_2)}{\bar{F}(\omega_n; x) g(x)} \right) \right]$$

$$\left. + r_n \left[\frac{\mathbb{E}((-\ln \bar{F}(\omega_n; x))^t T_n^{(t)}(x, K_1)) - \theta^t(x) \Gamma(t+1) \mathbb{E}(T_n^{(0)}(x, K_2))}{\Gamma(t+1) \bar{F}(\omega_n; x) g(x)} \right] \right\}$$

$$=: \frac{\bar{F}(\omega_n; x) g(x)}{T_n^{(0)}(x, K_2)} \{ I_3 + I_4 + I_5 \}.$$

From Lemma 3, we have

$$I_3 + I_4 \overset{\mathcal{D}}{\to} N \left(0, \frac{\theta^{2t}(x)}{\Gamma^2(t+1)} \left[\Gamma(2t+1) \|K_1\|_2^2 + \Gamma^2(t+1) \|K_2\|_2^2 - 2\Gamma^2(t+1) \|K_1 K_2\|_1 \right] \right).$$

Concerning I_5, by using Lemmas 1 and 2

$$I_5 = r_n \left\{ b(-\ln \bar{F}(\omega_n; x); x) t\theta^{t-1}(x)(1 + o(1)) + O \left(\frac{1}{(-\ln \bar{F}(\omega_n; x))^{1-\varepsilon}} \right) \right.$$

$$\left. + O(h_n) + O(\Phi(\omega_n, h_n; x)) \right\},$$

and hence under our assumptions, $I_5 \to \lambda \sqrt{g(x)}\, t\theta^{t-1}(x)$.

Further, from Lemma 3 we have that $T_n^{(0)}(x, K_2)/(\bar{F}(\omega_n; x) g(x)) = 1 + o_{\mathbb{P}}(1)$.

Combined, the above gives that

$$r_n \left(\frac{(-\ln \bar{F}(\omega_n; x))^t T_n^{(t)}(x, K_1)}{\Gamma(t+1) T_n^{(0)}(x, K_2)} - \theta^t(x) \right) \overset{\mathcal{D}}{\to} N \left(\lambda \sqrt{g(x)}\, t\theta^{t-1}(x), \right.$$

$$\left. \frac{\theta^{2t}(x)}{\Gamma^2(t+1)} \left[\Gamma(2t+1) \|K_1\|_2^2 + \Gamma^2(t+1) \|K_2\|_2^2 - 2\Gamma^2(t+1) \|K_1 K_2\|_1 \right] \right).$$

A straightforward application of the $\delta-$method yields

$$r_n(\tilde{\theta}_n^{(1)}(x;t,K_1,K_2) - \theta(x)) \xrightarrow{\mathcal{D}} N\left(\lambda\sqrt{g(x)},\right.$$
$$\left.\frac{\theta^2(x)}{t^2\Gamma^2(t+1)}\left[\Gamma(2t+1)\|K_1\|_2^2 + \Gamma^2(t+1)\|K_2\|_2^2 - 2\Gamma^2(t+1)\|K_1K_2\|_1\right]\right).$$

Now consider $\hat{\theta}_n^{(1)}(x;t,K_1,K_2)$:

$$r_n(\hat{\theta}_n^{(1)}(x;t,K_1,K_2) - \theta(x)) = r_n(\tilde{\theta}_n^{(1)}(x;t,K_1,K_2) - \theta(x))$$
$$+ r_n\left(-\ln\frac{\hat{\bar{F}}(\omega_n;x)}{\bar{F}(\omega_n;x)}\right)\left(\frac{T_n^{(t)}(x,K_1)}{\Gamma(t+1)T_n^{(0)}(x,K_2)}\right)^{1/t}.$$

Using Lemma 4 we have

$$r_n\left(-\ln\frac{\hat{\bar{F}}(\omega_n;x)}{\bar{F}(\omega_n;x)}\right) = O_{\mathbb{P}}(1),$$

and by Lemmas 1, 2, and 3

$$\left(\frac{T_n^{(t)}(x,K_1)}{\Gamma(t+1)T_n^{(0)}(x,K_2)}\right)^{1/t} = \frac{\theta(x)}{(-\ln\bar{F}(\omega_n;x))}(1 + o_{\mathbb{P}}(1)).$$

Hence

$$r_n(\hat{\theta}_n^{(1)}(x;t,K_1,K_2) - \theta(x)) = r_n(\tilde{\theta}_n^{(1)}(x;t,K_1,K_2) - \theta(x)) + O_{\mathbb{P}}\left(\frac{1}{(-\ln\bar{F}(\omega_n;x))}\right),$$

from which the result follows. $\square$

Proof of Theorem 2

The proof of this theorem follows arguments similar to those used in the proof of Theorem 1, and is therefore given in the supporting information.

Proof of Theorem 3

The proof of the theorem is inspired by the approach taken by Wretman (1978) in the univariate i.i.d. case. Let $\tilde{r}_n := (-\ln\alpha_n)\sqrt{nh_n^q\alpha_n g(x)}/Q(1-\alpha_n;x)$ and $\breve{r}_n := \theta(x)\sqrt{nh_n^q\alpha_n g(x)}$. By straightforward inversion we obtain that

$$\mathbb{P}(\tilde{r}_n(\hat{Q}(1-\alpha_n;x) - Q(1-\alpha_n;x)) \leq z) =$$
$$\mathbb{P}\left(\breve{r}_n\ln\frac{\hat{\bar{F}}(Q(1-\alpha_n;x) + z/\tilde{r}_n;x)}{\bar{F}(Q(1-\alpha_n;x) + z/\tilde{r}_n;x)} \leq \breve{r}_n\ln\frac{\bar{F}(Q(1-\alpha_n;x);x)}{\bar{F}(Q(1-\alpha_n;x) + z/\tilde{r}_n;x)}\right).$$

Now let

$$Z_n := \breve{r}_n \ln \frac{\hat{\bar{F}}(Q(1-\alpha_n;x) + z/\tilde{r}_n;x)}{\bar{F}(Q(1-\alpha_n;x) + z/\tilde{r}_n;x)} \quad \text{and} \quad z_n := \breve{r}_n \ln \frac{\bar{F}(Q(1-\alpha_n;x);x)}{\bar{F}(Q(1-\alpha_n;x) + z/\tilde{r}_n;x)}.$$

Concerning z_n we have

$$z_n = \breve{r}_n V'\left(Q(1-\alpha_n;x) + \kappa_n \frac{z}{\tilde{r}_n};x\right) \frac{z}{\tilde{r}_n}$$

with $\kappa_n \in (0,1)$, where the last step follows from the mean value theorem. Let $\tilde{Q}(1-\alpha_n;x) := Q(1-\alpha_n;x) + \kappa_n z/\tilde{r}_n$. Note that $\tilde{Q}(1-\alpha_n;x) = Q(1-\alpha_n;x)(1+o(1)) \to \infty$ if $\alpha_n \to 0$, and hence, using the assumption that V is a normalized regularly varying function, we have

$$\frac{\tilde{Q}(1-\alpha_n;x)V'(\tilde{Q}(1-\alpha_n;x);x)}{V(\tilde{Q}(1-\alpha_n;x);x)} \to \frac{1}{\theta(x)} \tag{12}$$

for $\alpha_n \to 0$. Thus

$$z_n = \breve{r}_n \frac{1}{\theta(x)} \frac{V(\tilde{Q}(1-\alpha_n;x);x)}{\tilde{Q}(1-\alpha_n;x)} \frac{z}{\tilde{r}_n}(1+o(1)) = z(1+o(1)).$$

Concerning the random term we have that

$$Z_n = \sqrt{\Delta_n}\theta(x)\sqrt{nh_n^q \bar{F}(Q(1-\alpha_n;x) + z/\tilde{r}_n;x)g(x)} \ln \frac{\hat{\bar{F}}(Q(1-\alpha_n;x) + z/\tilde{r}_n;x)}{\bar{F}(Q(1-\alpha_n;x) + z/\tilde{r}_n;x)}$$

where

$$\Delta_n := \frac{\bar{F}(Q(1-\alpha_n;x);x)}{\bar{F}(Q(1-\alpha_n;x) + z/\tilde{r}_n;x)}.$$

Again using (12) and by the definition of $\tilde{r}_n$, one obtains that $\Delta_n \to 1$ as $\alpha_n \to 0$. Hence by Lemma 4 we have that

$$Z_n \xrightarrow{\mathcal{D}} N(0, \theta^2(x)\|K\|_2^2). \tag{13}$$

Denote by G_n the distribution function of Z_n and by G that of the limiting distribution in (13). Because of the continuity of G we have that the convergence of G_n to G is uniform (see e.g. Lemma 2.11 in van der Vaart, 1998, p 12), and hence $\lim_{n\to\infty} G_n(z_n) = G(z)$. This completes the proof of Theorem 3. $\square$

Proof of Theorem 4

By straightforward rearrangements we obtain

$$\frac{\sqrt{nh_n^q \alpha_n g(x)}}{\ln \tau_n} \ln \frac{\hat{Q}_W(1-\beta_n;x)}{Q(1-\beta_n;x)} = \sqrt{nh_n^q \alpha_n g(x)}(\hat{\theta}(x) - \theta(x))$$

$$+ \frac{\sqrt{nh_n^q \alpha_n g(x)}}{\ln \tau_n} \ln \frac{\hat{Q}(1-\alpha_n;x)}{Q(1-\alpha_n;x)}$$

$$+ \frac{\sqrt{nh_n^q \alpha_n g(x)}}{\ln \tau_n} [\ln Q(1-\alpha_n;x) - \ln Q(1-\beta_n;x) + \theta(x)\ln \tau_n]$$

$$=: I_6 + I_7 + I_8.$$

22

Under the assumptions of the theorem

$$\sqrt{nh_n^q \alpha_n g(x)}(\hat{\theta}(x) - \theta(x)) \xrightarrow{\mathcal{D}} N(0, V^2(x)).$$

For I_7 we use Theorem 3 to obtain

$$I_7 = \frac{(-\ln \alpha_n)\sqrt{nh_n^q \alpha_n g(x)}}{(-\ln \alpha_n)\ln \tau_n} \ln \frac{\hat{Q}(1 - \alpha_n; x)}{Q(1 - \alpha_n; x)} = O_{\mathbb{P}}\left(\frac{1}{(-\ln \alpha_n)}\right).$$

Finally, by using condition $(\mathcal{R})$, the term I_8 can be written as

$$I_8 = -\sqrt{nh_n^q \alpha_n g(x)}b(-\ln \alpha_n; x)\frac{\tau_n^{\rho(x)} - 1}{\rho(x)\ln \tau_n}(1 + o(1)) \to 0. \qquad \square$$

Acknowledgement

References

[1] Beirlant, J., Broniatowski, M., Teugels, J.L., Vynckier, P., 1995. The mean residual life function at great age: applications to tail estimation. *Journal of Statistical Planning and Inference*, 45, 21–48.

[2] Billingsley, P., 1995. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York.

[3] Bingham, N.H., Goldie, C.M., Teugels, J.L., 1987. *Regular Variation*. Cambridge University Press, Cambridge.

[4] Broniatowski, M., 1993. On the estimation of the Weibull tail coefficient. *Journal of Statistical Planning and Inference*, 35, 349–366.

[5] Daouia, A., Gardes, L., Girard, S., Lekina, A., 2011. Kernel estimators of extreme level curves. *Test*, 20, 311–333.

[6] Daouia, A., Gardes, L., Girard, S., 2013. On kernel smoothing for extremal quantile regression. *Bernoulli*, 19, 2557–2589.

[7] Davison, A.C., Smith, R.L., 1990. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society Series B*, 52, 393–442.

[8] de Haan, L., de Ronde, J., 1998. Sea and wind: multivariate extremes at work. *Extremes*, 1, 7–45.

[9] de Haan, L., Ferreira, A., 2006. *Extreme Value Theory: An Introduction.* Springer-Verlag, New York.

[10] Diebolt, J., Gardes, L., Girard, S., Guillou, A., 2008. Bias-reduced estimators of the Weibull tail-coefficient. *Test*, 17, 311–331.

[11] Dierckx, G., Beirlant, J., De Waal, D., Guillou, A., 2009. A new estimation method for Weibull-type tails based on the mean excess function. *Journal of Statistical Planning and Inference*, 139, 1905–1920.

[12] Draisma, G., de Haan, L., Peng, L., Pereira, T.T., 1999. A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes*, 2, 367–404.

[13] Draisma, G., Drees, H., Ferreira, A., de Haan, L., 2004. Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, 10, 251–280.

[14] Gannoun, A., Girard, S., Guinot, C., Saracco, J., 2002. Reference ranges based on nonparametric quantile regression. *Statistics in Medicine*, 21, 3119–3135.

[15] Gardes, L., Girard, S., 2005. Estimating extreme quantiles of Weibull tail distributions. *Communications in Statistics – Theory and Methods*, 34, 1065–1080.

[16] Gardes, L., Girard, S., 2008a. Estimation of the Weibull-tail coefficient with linear combination of upper order statistics. *Journal of Statistical Planning and Inference*, 138, 1416–1427.

[17] Gardes, L., Girard, S., 2008b. A moving window approach for nonparametric estimation of the conditional tail index. *Journal of Multivariate Analysis*, 99, 2368–2388.

[18] Gardes, L., Stupfler, G., 2013. Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes*, 17, 45–75.

[19] Geluk, J.L., de Haan, L., 1987. *Regular Variation, Extensions and Tauberian Theorems.* CWI Tract 40, Center for Mathematics and Computer Science, Amsterdam.

[20] Girard, S., 2004. A Hill type estimator of the Weibull tail coefficient. *Communications in Statistics - Theory and Methods*, 33, 205–234.

[21] Goegebeur, Y., Guillou, A., 2010. Goodness-of-fit testing for Weibull-type behavior. *Journal of Statistical Planning and Inference*, 140, 1417–1436.

[22] Goegebeur, Y., Guillou, A., 2011. A weighted mean excess function approach to the estimation of Weibull-type tails. *Test*, 20, 138–162.

[23] Goegebeur, Y., Beirlant, J., de Wet, T., 2010. Generalized kernel estimators for the Weibull tail coefficient. *Communications in Statistics – Theory and Methods*, 39, 3695–3716.

[24] Goegebeur, Y., Guillou, A., Osmann, M., 2014a. A local moment type estimator for the extreme value index in regression with random covariates. *Canadian Journal of Statistics*, 42, 487–507.

[25] Goegebeur, Y., Guillou, A., Schorgen, A., 2014b. Nonparametric regression estimation of conditional tails - the random covariate case. *Statistics*, 48, 4, 732-755.

[26] Klüppelberg, C., Villaseñor, J.A., 1993. Estimation of distribution tails - a semiparametric approach. Deutschen Gesellschaft für Versicherungsmathematik XXI, 213–235.

[27] Segers, J., Teugels, J., 2000. Testing the Gumbel hypothesis by Galton's ratio. *Extremes*, 3, 291-303.

[28] Smith, R. L., 1987. Estimating tails of probability distributions. *Annals of Statistics*, 15, 3, 1174-1207.

[29] Stupfler, G., 2013. A moment estimator for the conditional extreme-value index. *Electronic Journal of Statistics*, 7, 2298–2343.

[30] van der Vaart, A.W., 1998. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.

[31] Wang, H., Tsai, C.L., 2009. Tail index regression. *Journal of the American Statistical Association*, 104, 1233–1240.

[32] Weissman, I., 1978. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73, 812–815.

[33] Wretman, J., 1978. A simple derivation of the asymptotic distribution of a sample quantile. *Scandinavian Journal of Statistics*, 5, 123–124.

[34] Yao, Q., 1999. Conditional predictive regions for stochastic processes. Technical report, University of Kent at Canterbury.

Strict Weibull distribution

t	$\hat{\theta}_n^{(1)}(x; t, K, K)$		$\hat{Q}_W^{(1)}(1 - 1/1200; x)$		$\hat{Q}_W^{(1)}(1 - 1/2000; x)$		$\hat{Q}_W^{(1)}(1 - 1/4000; x)$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.2	**0.041**	**0.102**	**0.110**	**0.237**	**0.120**	**0.259**	**0.132**	**0.287**
0.4	0.046	**0.102**	0.120	0.239	0.130	0.261	0.143	0.289
0.6	0.053	0.105	0.130	0.243	0.141	0.265	0.156	0.294
0.8	0.059	0.108	0.139	0.249	0.152	0.271	0.168	0.301
1.0	0.065	0.111	0.150	0.255	0.163	0.278	0.181	0.309

Extended Weibull distribution

t	$\hat{\theta}_n^{(1)}(x; t, K, K)$		$\hat{Q}_W^{(1)}(1 - 1/1200; x)$		$\hat{Q}_W^{(1)}(1 - 1/2000; x)$		$\hat{Q}_W^{(1)}(1 - 1/4000; x)$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.2	**0.041**	**0.101**	**0.113**	**0.231**	**0.124**	**0.253**	**0.139**	**0.283**
0.4	0.049	0.103	0.127	0.237	0.140	0.260	0.157	0.290
0.6	0.057	0.106	0.140	0.243	0.154	0.267	0.174	0.298
0.8	0.063	0.110	0.153	0.250	0.169	0.274	0.190	0.307
1.0	0.069	0.113	0.164	0.257	0.182	0.282	0.205	0.315

Table 1: Setting 1: Performance of $\hat{\theta}_n^{(1)}(x; t, K, K)$, $\hat{Q}_W^{(1)}(1 - 1/1200; x)$, $\hat{Q}_W^{(1)}(1 - 1/2000; x)$ and $\hat{Q}_W^{(1)}(1 - 1/4000; x)$. The results are averaged over 500 Monte Carlo simulations, with $n = 1000$. The numbers in bold indicate the value of t with smallest bias, respectively RMSE.

Strict Weibull distribution

t	$\hat{\theta}_n^{(1)}(x;t,K,K)$		$\hat{Q}_W^{(1)}(1-1/1200;x)$		$\hat{Q}_W^{(1)}(1-1/2000;x)$		$\hat{Q}_W^{(1)}(1-1/4000;x)$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.2	**0.040**	**0.095**	**0.094**	**0.208**	**0.103**	**0.228**	**0.113**	**0.254**
0.4	0.047	0.097	0.106	**0.208**	0.115	**0.228**	0.128	**0.254**
0.6	0.053	0.099	0.116	0.211	0.127	0.231	0.142	0.257
0.8	0.060	0.103	0.127	0.215	0.139	0.235	0.155	0.262
1.0	0.066	0.107	0.136	0.220	0.150	0.241	0.167	0.269

Extended Weibull distribution

t	$\hat{\theta}_n^{(1)}(x;t,K,K)$		$\hat{Q}_W^{(1)}(1-1/1200;x)$		$\hat{Q}_W^{(1)}(1-1/2000;x)$		$\hat{Q}_W^{(1)}(1-1/4000;x)$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.2	**0.037**	**0.094**	**0.095**	**0.201**	**0.105**	**0.220**	**0.118**	**0.247**
0.4	0.045	0.096	0.108	0.203	0.120	0.224	0.136	0.251
0.6	0.052	0.099	0.120	0.208	0.134	0.228	0.151	0.256
0.8	0.059	0.102	0.132	0.213	0.146	0.235	0.165	0.263
1.0	0.065	0.106	0.142	0.219	0.157	0.241	0.178	0.271

Table 2: Setting 2: Performance of $\hat{\theta}_n^{(1)}(x;t,K,K)$, $\hat{Q}_W^{(1)}(1-1/1200;x)$, $\hat{Q}_W^{(1)}(1-1/2000;x)$ and $\hat{Q}_W^{(1)}(1-1/4000;x)$. The results are averaged over 500 Monte Carlo simulations, with $n=1000$. The numbers in bold indicate the value of t with smallest bias, respectively RMSE.

		$\tilde{q}_n^{RP,1}(1/1200|x)$		$\tilde{q}_n^{RP,1}(1/2000|x)$		$\tilde{q}_n^{RP,1}(1/4000|x)$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
Setting 1	Strict Weibull	0.207	0.327	0.249	0.373	0.309	0.437
	Extended Weibull	0.161	0.299	0.192	0.342	0.237	0.400
Setting 2	Strict Weibull	0.197	0.288	0.240	0.332	0.302	0.393
	Extended Weibull	0.145	0.257	0.175	0.297	0.218	0.352

Table 3: Performance of $\tilde{q}_n^{RP,1}(1/1200|x)$, $\tilde{q}_n^{RP,1}(1/2000|x)$ and $\tilde{q}_n^{RP,1}(1/4000|x)$ with $J=4$ in Setting 1 and Setting 2. The results are averaged over 500 Monte Carlo simulations, with $n=1000$.
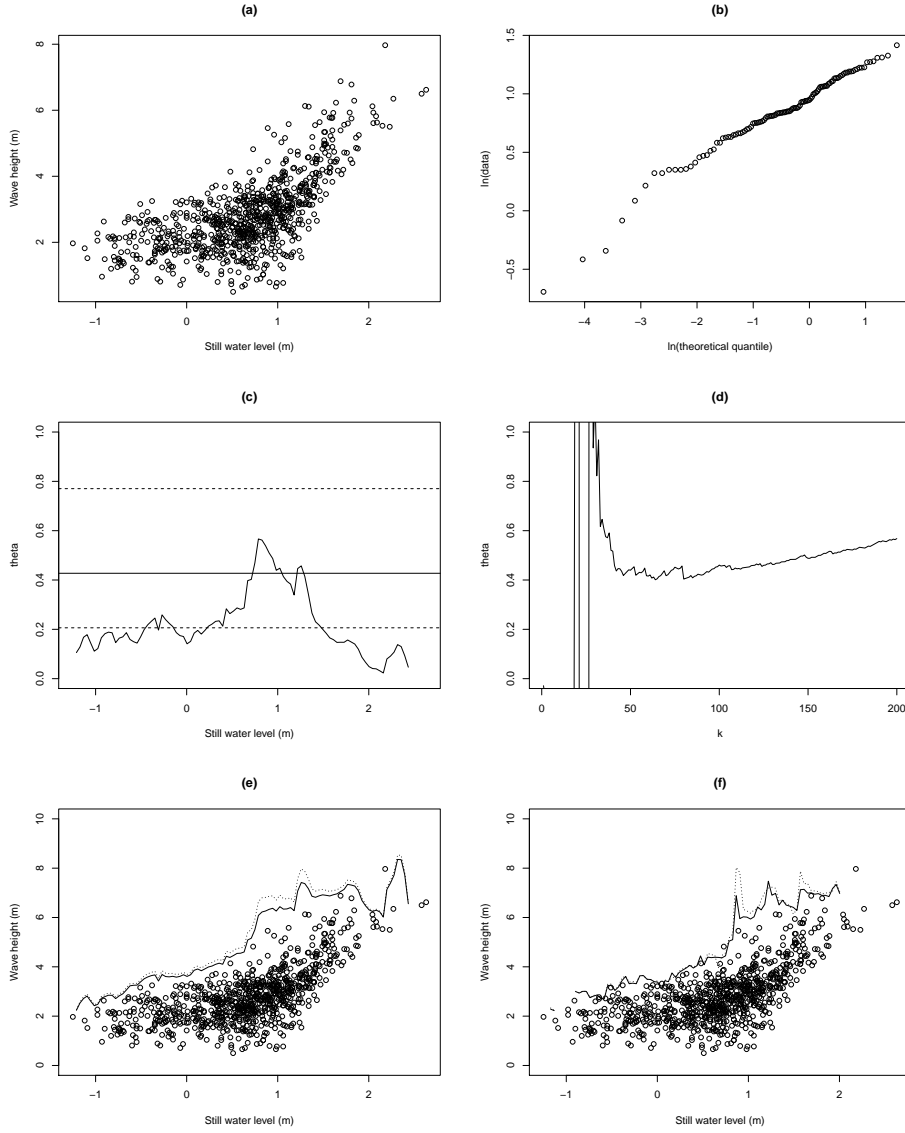
Figure 1: Sea storm data: (a) scatterplot of wave height (HmO) versus still water level (SWL), (b) Weibull quantile plot of the HmO observations with $0.5 \leq SWL \leq 0.7$, (c) $\hat{\theta}_n^{(1)}(SWL; 0.2, K, K)$ as a function of SWL (the solid horizontal reference line represents the estimate obtained in a univariate analysis ignoring SWL, and the dashed horizontal reference lines represent a univariate 95% bootstrap interval), (d) univariate Weibull-tail coefficient estimate for HmO as a function of k, (e) $\hat{Q}_W^{(1)}(1 - 1/1200; SWL)$ (solid line) and $\hat{Q}_W^{(1)}(1 - 1/4000; SWL)$ (dotted line) versus SWL and (f) Daouia $et\ al.$ (2013) estimate $\tilde{q}_n^{RP,1}(1/1200|SWL)$ (solid line) and $\tilde{q}_n^{RP,1}(1/4000|SWL)$ (dotted line) with $J = 4$ versus SWL.