

Multivariate generalized linear-statistics of short range dependent data

Svenja Fischer

*Department of Civil and Environmental Engineering
Ruhr-University Bochum
44801 Bochum, Germany
e-mail: svenja.fischer@rub.de*

Roland Fried

*Faculty of Statistics
TU Dortmund University
44221 Dortmund, Germany
e-mail: fried@statistik.tu-dortmund.de*

and

Martin Wendler

*Department of Mathematics and Computer Science
Ernst Moritz Arndt Universität Greifswald
17847 Greifswald, Germany
e-mail: martin.wendler@uni-greifswald.de*

Abstract: Generalized linear (GL -) statistics are defined as functionals of an U -quantile process and unify different classes of statistics such as U -statistics and L -statistics. We derive a central limit theorem for GL -statistics of strongly mixing sequences and arbitrary dimension of the underlying kernel. For this purpose we establish a limit theorem for U -statistics and an invariance principle for U -processes together with a convergence rate for the remaining term of the Bahadur representation.

An application is given by the generalized median estimator for the tail-parameter of the Pareto distribution, which is commonly used to model exceedances of high thresholds. We use subsampling to calculate confidence intervals and investigate its behaviour under independence and under strong mixing in simulations.

MSC 2010 subject classifications: Primary 62G30, 60G10; secondary 60F17.

Keywords and phrases: GL -statistics, U -statistics, strong mixing, generalized median estimator, tail parameter.

Received December 2014.

Contents

1	Introduction	647
2	Main results	649
2.1	U -statistics and U -processes	651

3	Application: The generalized median estimator	654
4	Preliminary results	662
5	Proofs	669
	Appendix	677
	Acknowledgements	680
	References	680

1. Introduction

Generalized linear statistics (GL -statistics) form a broad class of statistics, which unifies not only the widely used U -statistics but also other classes like L -statistics and even statistics which cannot be assigned to a certain class. GL -statistics were first developed by Serfling (1984), who shows a central limit theorem under independence. In this paper we develop results for GL -statistics of random variables which are short range dependent. An important tool to gain a Central Limit Theorem for GL -statistics are U -statistics with multivariate kernels. Up to now we can find a lot of results for bivariate U -statistics of short range dependent data (cf. Borovkova et al. (2001), Dehling and Wendler (2010) and Wendler (2011b)) but in the multivariate case there occur some additional difficulties caused by the dependencies in the kernel structure.

Now let us introduce some basic assumptions and definitions which we will use throughout the paper.

Let X_1, \dots, X_n be a sequence of random variables with distribution function F . We will assume the random variables to be short range dependent, a detailed definition is given later on. Moreover, let F_n be the empirical distribution function of X_1, \dots, X_n with

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{[X_i \leq x]}, \quad -\infty < x < \infty,$$

and $h(x_1, \dots, x_m)$, for given $m \geq 2$, a kernel, that is a measurable, symmetric function. We define the empirical distribution function H_n of $h(X_{i_1}, \dots, X_{i_m})$ as

$$H_n(x) = \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} 1_{[h(X_{i_1}, \dots, X_{i_m}) \leq x]}, \quad -\infty < x < \infty,$$

and $H_n^{-1}(p) = \inf\{x | H_n(x) \geq p\}$ as the related generalized inverse. Furthermore, let H_F with $H_F(y) = \mathbb{P}_F(h(Y_1, \dots, Y_m) \leq y)$ be the distribution function of the kernel h for independent copies Y_1, \dots, Y_m of X_1 and $0 < h_F < \infty$ the related density (this implies that H_F is continuous).

We define $h_{F; X_{i_2}, \dots, X_{i_k}}$ as the density of $h(Y_{i_1}, X_{i_2}, \dots, X_{i_k}, Y_{i_{k+1}}, \dots, Y_{i_m})$ for $2 \leq k \leq m$ and $i_1 < i_2 < \dots < i_m$.

Definition 1.1.

A generalized L -statistic with kernel h is given by

$$\begin{aligned} T(H_n) &= \int_0^1 H_n^{-1}(t)J(t)dt + \sum_{i=1}^d a_i H_n^{-1}(p_i) \\ &= \sum_{i=1}^{n_{(m)}} \left[\int_{\frac{(i-1)}{n_{(m)}}}^{\frac{i}{n_{(m)}}} J(t)dt \right] H_n^{-1} \left(\frac{i}{n_{(m)}} \right) + \sum_{i=1}^d a_i H_n^{-1}(p_i), \end{aligned}$$

where $n_{(m)} = n(n-1) \cdots (n-m+1)$.

The GL -statistic $T(H_n)$ is a natural estimator of $T(H_F)$, which is defined analogously.

Example 1.1.

Let $h : \mathbb{R}^m \rightarrow \mathbb{R}$ be a measurable function. A U -statistic with kernel h is defined as

$$U_n = \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} h(X_{i_1}, \dots, X_{i_m}).$$

If the random variables are independent and identically distributed, U_n is an unbiased estimator of $\theta = \mathbb{E}(h(X_1, \dots, X_m))$. A U -statistic can be written as a GL -statistic by setting $d = 0$ and $J = 1$.

Example 1.2.

A widely known L -statistic is the α -trimmed mean

$$\bar{X}_{(\alpha)} = \frac{1}{n - 2[n\alpha]} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{(i)},$$

where $X_{(i)}$ is the i th value of the order statistic $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. To rewrite it as a GL -statistic we choose $J(t) = \frac{1}{1-2\alpha}$ for $\alpha < t < 1 - \alpha$ and $J(t) = 0$ everywhere else. As kernel we set $h(x) = x$ and let the sum vanish by the choice $d = 0$.

Example 1.3.

The generalized *Hodges-Lehmann estimator*

$$\text{median} \left(\frac{1}{n} (X_{i_1} + \dots + X_{i_m}), 1 \leq i_1, \dots, i_m \leq n \right)$$

is neither a U -statistic nor an L -statistic, but it is possible to formulate it as a GL -statistic choosing the kernel $h(x_{i_1}, \dots, x_{i_m}) = \frac{1}{n}(x_{i_1} + \dots + x_{i_m})$ and setting $J = 0$, $d = 1$, and $a_1 = 1$. We get the median of the kernel by using the representation via the quantile function $H_n^{-1}(\frac{1}{2})$. Consequently $p_1 = \frac{1}{2}$. The generalized Hodges-Lehmann estimator is the GL -statistic

$$T(H_n) = H_n^{-1} \left(\frac{1}{2} \right).$$

In the following we will consider a special form of short range dependence: strong mixing.

Definition 1.2. Let $(X_n)_{n \in \mathbb{N}}$ be a stationary process. The strong mixing coefficients of (X_n) are

$$\alpha(k) = \sup_{n \in \mathbb{N}} \sup \{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_1^n, B \in \mathcal{F}_{n+k}^\infty \},$$

where \mathcal{F}_a^b is the σ -field generated by X_a, \dots, X_b .

$(X_n)_{n \in \mathbb{N}}$ is called strongly mixing (or α -mixing), if $\alpha(k) \rightarrow 0$ as $k \rightarrow \infty$.

Strong mixing is the weakest among the different forms of mixing since the α -mixing coefficients are always smaller than for example the β -mixing coefficients (cf. Bradley (2007)).

After stating the main results, among others the Central Limit Theorem for GL-statistics, we also provide some results concerning U -statistics and U -processes. In a second step we give an application, the generalized median estimator (GM -estimator) for the tail parameter of the Pareto distribution (cf. Brazauskas and Serfling (2000b) and Brazauskas and Serfling (2000a) under independence). The Pareto distribution is commonly used for modelling heavy tails and exceedances of a threshold (peak over threshold, POT). Especially in hydrology it has wide application when only extreme floods above a certain threshold should be considered in the analysis. There also occurs the need of a robust estimator, needing a downweighting of the influence of extreme floods in short time series. Simulations verify that the generalized median estimator is almost as efficient as the maximum likelihood estimator under independence and for autocorrelated data, but more robust. Short range dependence is up to now seldom modelled in the estimation of parameters under POT, but when considering for example monthly discharges it is very probable to find such dependencies. Our investigation of the generalized median estimator aims at closing this gap and can be extended to other situations, where a robust estimator for dependent data is needed.

Results needed for the proofs of the main results are given in Section 4, the proofs of the results given in Section 2 can be found in Section 5.

2. Main results

An important and well known result concerning quantiles is the representation proposed by Bahadur, which uses the representation of the quantile by the empirical distribution function. A key role plays the remaining term, for which Ghosh (1971) showed the convergence for ordinary quantiles and under independence. In the dependent case results were shown by Sen (1972) for ϕ -mixing and Babu and Singh (1978) and Yoshihara (1995) (with a weaker decay of the mixing coefficients) for strong mixing random variables. Also Sun and Lahiri (2006) as well as Lahiri (2009) with a specified rate show the normal approximation of the term $\sqrt{n}(F_n^{-1} - F^{-1})$ for strong mixing random variables. The

extension to generalized quantiles can be found in Wendler (2011a) with bivariate kernels. In our case we need the convergence of generalized quantiles with a multivariate kernel and strong mixing. The result is stated in the following theorem.

Theorem 2.1.

Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of strong mixing random variables with distribution function F , $\mathbb{E}|X_1|^\rho < \infty$ for a $\rho \geq 1$ and mixing coefficients $\alpha(l) = O(l^{-\delta})$ for a $\delta > \frac{2\rho+1}{\rho}$. Moreover let $h(x_1, \dots, x_m)$ be a Lipschitz-continuous kernel with distribution function H_F and related density $0 < h_F < \infty$ and for all $2 \leq k \leq m$ let $h_{F; X_2, \dots, X_k}$ be bounded. Then we have for the Bahadur representation with $\hat{\xi}_p = H_n^{-1}(p)$

$$\hat{\xi}_p = \xi_p + \frac{H_F(\xi_p) - H_n(\xi_p)}{h_F(\xi_p)} + o_p\left(\frac{1}{\sqrt{n}}\right).$$

Remember that a kernel $h : \mathbb{R}^m \rightarrow \mathbb{R}$ is called Lipschitz-continuous if there exists a constant $L \geq 0$ such that

$$|h(x_1, \dots, x_m) - h(x'_1, \dots, x'_m)| \leq L \|(x_1, \dots, x_m) - (x'_1, \dots, x'_m)\|$$

for all $x_1, \dots, x_m, x'_1, \dots, x'_m \in \mathbb{R}$.

Now we will state the main theorem of our paper, the asymptotic normality of GL -statistics under strong mixing. Under independence this result was proved by Serfling (1984).

Theorem 2.2.

Let $h(x_1, \dots, x_m)$ be a Lipschitz-continuous kernel with distribution function H_F and related density $0 < h_F < \infty$ and for all $2 \leq k \leq m$ and all $i_1 < i_2 < \dots < i_m$ let $h_{F; X_{i_2}, \dots, X_{i_k}}$ be bounded. Moreover let J be a function with $J(t) = 0$ for $t \notin [\alpha, \beta]$, $0 < \alpha < \beta < 1$, and in $[\alpha, \beta]$ let J be bounded and a.e. continuous concerning the Lebesgue-measure and a.e. continuous concerning H_F^{-1} . Additionally, let X_1, \dots, X_n be a sequence of strong mixing random variables with $\mathbb{E}|X_1|^\rho < \infty$ for a $\rho \geq 1$ and mixing coefficients $\alpha(n)$ with $\alpha(n) = O(n^{-\delta})$ for a $\delta \geq 8$. Then the following statement holds for GL -Statistics $T(H_n)$

$$\sqrt{n} (T(H_n) - T(H_F)) \xrightarrow{\mathcal{D}} N(0, \sigma^2),$$

where

$$\begin{aligned} \sigma^2 = & m^2 (\text{Var} (\mathbb{E} (A(Y_1, \dots, Y_m) | Y_1 = X_1)) \\ & + 2 \sum_{j=1}^{\infty} \text{Cov} (\mathbb{E} (A(Y_1, \dots, Y_m) | Y_1 = X_1), \mathbb{E} (A(Y_1, \dots, Y_m) | Y_1 = X_{j+1}))) \end{aligned}$$

with independent copies Y_1, \dots, Y_m of X_1 and

$$A(x_1, \dots, x_m) = - \int_{-\infty}^{\infty} (1_{[h(x_1, \dots, x_m) \leq y]} - H_F(y)) J(H_F(y)) dy$$

$$+ \sum_{i=1}^d a_i \frac{p_i - 1[h(x_1, \dots, x_m) \leq H_F^{-1}(p_i)]}{h_F(H_F^{-1}(p_i))}.$$

For the proof of this theorem, which is given in Section 5, a key tool will be the representation of the kernel A as a U -statistic, see Example 1.1. Additionally also the functional H_n belongs to the class of U -statistics and therefore we make use of several results of the theory of U -statistics. In the following section we will extend some known results for bivariate U -statistics under strong mixing to the multivariate case. We will see that this extension causes some problems concerning the dependencies in the kernels and the solution of these problems is not straightforward.

In the Theorem above the asymptotic variance is influenced by unknown quantities. The conditionals expectation as well as the density of the distribution function of the kernel and also the covariances are not known and cannot be estimated directly. For the simple case of a bivariate kernel Vogel and Wendler (2015) develop an estimator of the long-term variance and show its consistency (Theorem 2.4). The multivariate case is more complicated and a consistency result is beyond the scope of this paper.

Remark 2.1. In the case of bivariate kernels, similar results as Theorems 2.3 and 2.4 can be found in Borovkova et al. (2001) (Theorem 7, Asymptotic Normality of U -Statistics), Dehling and Philipp (2002) (Theorem 5.14, Invariance Principle for U -processes) and Wendler (2011b) (Theorem 3.3.1) for near epoch dependent (NED) sequences of absolutely regular processes. We conjecture that an extension to the multivariate case is possible also under this other type of weak dependence, but detailed proofs are beyond the scope of this paper.

2.1. U -statistics and U -processes

While examining U -statistics often a technique called Hoeffding decomposition (Hoeffding (1948)) is used. It decomposes the U -statistic into a sum of different terms, which we can examine separately.

Definition 2.1. (Hoeffding decomposition)

Let U_n be a U -statistic with kernel $h = h(x_1, \dots, x_m)$. Then one can write U_n as

$$U_n = \theta + \sum_{j=1}^m \binom{m}{j} \frac{1}{\binom{n}{j}} S_{jn},$$

where

$$\begin{aligned} \theta &= \mathbb{E}(h(Y_1, \dots, Y_m)) \\ \tilde{h}_j(x_1, \dots, x_j) &= \mathbb{E}(h(x_1, \dots, x_j, Y_{j+1}, \dots, Y_m)) - \theta \\ S_{jn} &= \sum_{1 \leq i_1 < \dots < i_j \leq n} g_j(X_{i_1}, \dots, X_{i_j}) \end{aligned}$$

$$\begin{aligned}
 g_1(x_1) &= \tilde{h}_1(x_1) \\
 g_2(x_1, x_2) &= \tilde{h}_2(x_1, x_2) - g_1(x_1) - g_1(x_2) \\
 g_3(x_1, x_2, x_3) &= \tilde{h}_3(x_1, x_2, x_3) - \sum_{i=1}^3 g_1(x_i) - \sum_{1 \leq i < j \leq 3} g_2(x_i, x_j) \\
 &\dots \\
 g_m(x_1, \dots, x_m) &= \tilde{h}_m(x_1, \dots, x_m) - \sum_{i=1}^m g_1(x_i) - \sum_{1 \leq i_1 < i_2 \leq m} g_2(x_{i_1}, x_{i_2}) \\
 &\quad - \dots - \sum_{1 \leq i_1 < \dots < i_{m-1} \leq m} g_{m-1}(x_{i_1}, \dots, x_{i_{m-1}}).
 \end{aligned}$$

for independent copies Y_1, \dots, Y_m of X_1 .

The term $\frac{m}{n} \sum_{i=1}^n g_1(X_i)$ is called the linear part, the remaining parts are called degenerated.

For most of the results in this section we need a regularity condition for the kernel h , which was first developed by Denker and Keller (1986) and is extended for our purpose.

Definition 2.2. A kernel h satisfies the variation condition, if there exists a constant L and an $\epsilon_0 > 0$, such that for all $\epsilon \in (0, \epsilon_0)$

$$\mathbb{E} \left(\sup_{\|(x_1, \dots, x_m) - (X'_1, \dots, X'_m)\| \leq \epsilon} |h(x_1, \dots, x_m) - h(X'_1, \dots, X'_m)| \right) \leq L\epsilon,$$

where the X'_i are independent with the same distribution as X_i and $\|\cdot\|$ is the Euklidean norm.

A kernel h satisfies the extended variation condition, if there additionally exist constants $L' > 0$ and $\delta_0 > 0$, such that for all $\delta \in (0, \delta_0)$ and all $2 \leq k \leq m$

$$\begin{aligned}
 &\mathbb{E} \left(\sup_{|x_{i_1} - Y_{i_1}| \leq \delta} |h(x_{i_1}, X_{i_2}, \dots, X_{i_k}, Y_{i_{k+1}}, \dots, Y_{i_m}) \right. \\
 &\quad \left. - h(Y_{i_1}, X_{i_2}, \dots, X_{i_k}, Y_{i_{k+1}}, \dots, Y_{i_m}) \right) \\
 &\leq L'\delta
 \end{aligned}$$

for independent copies $(Y_n)_{n \in \mathbb{N}}$ of $(X_n)_{n \in \mathbb{N}}$ and all $i_1 < i_2 < \dots < i_m$. If the kernel has dimension $m = 1$, we note that it satisfies the extended variation condition, if it satisfies the variation condition.

Remark 2.2. Every Lipschitz-continuous kernel satisfies the variation condition.

Now we state another main result of this paper, the asymptotic normality of U -statistics under strong mixing. For bivariate U -statistics this result is already known (see Wendler (2011b)), but not for arbitrary dimension m of the kernel h .

Theorem 2.3.

Let $h : \mathbb{R}^m \rightarrow \mathbb{R}$ be a bounded kernel satisfying the extended variation condition. Moreover let $(X_n)_{n \in \mathbb{N}}$ be a sequence of strong mixing random variables with $\mathbb{E}|X_1|^\rho < \infty$ for a $\rho > 0$ and mixing coefficients $\alpha(l) = O(l^{-\delta})$ for a $\delta > \frac{2\rho+1}{\rho}$. Then we have

$$\sqrt{n}(U_n - \theta) \xrightarrow{D} N(0, m^2 \sigma^2)$$

with $\sigma^2 = \text{Var}(g_1(X_1)) + 2 \sum_{j=1}^{\infty} \text{Cov}(g_1(X_1), g_1(X_{1+j}))$.

If $\sigma = 0$ then the statement means convergence to 0 in probability.

The key tool for the proof of this theorem is the Hoeffding decomposition, for which the first term converges against the given distribution while all remaining terms converge towards zero.

As an extension to U -statistics we also analyse U -processes and their convergence. In other words our U -statistic has no longer a fixed kernel h but we have a process $(U_n(t))_{t \in \mathbb{R}}$. Up to now we have had $(H_n(t))_{t \in \mathbb{R}}$ as an example of such a process.

Definition 2.3. Let $h : \mathbb{R}^{m+1} \rightarrow \mathbb{R}$ be a measurable and bounded function, symmetric in the first m arguments and non-decreasing in the last. Suppose that for all $x_1, \dots, x_m \in \mathbb{R}$ we have $\lim_{t \rightarrow \infty} h(x_1, \dots, x_m, t) = 1$, $\lim_{t \rightarrow -\infty} h(x_1, \dots, x_m, t) = 0$. We call the process $(U_n(t))_{t \in \mathbb{R}}$ empirical U -distribution function. As U -distribution function we define $U(t) := \mathbb{E}(h(Y_1, \dots, Y_m, t))$ for independent copies Y_1, \dots, Y_m of X_1 . Then the empirical process is defined as

$$(\sqrt{n}(U_n(t) - U(t)))_{t \in \mathbb{R}}.$$

Analogous to simple U -statistics here the Hoeffding decomposition is an important technique in our proofs. For fixed t we have

$$U_n(t) = \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} h(X_{i_1}, \dots, X_{i_m}, t)$$

and therefore we can decompose $U_n(t)$ analogously to Definition 2.1.

Likewise we will need a new form of the extended variation condition.

Definition 2.4. We say h satisfies the extended uniform variation condition, if the extended variation condition holds for $h(x_1, \dots, x_m, t)$ with a constant not depending on t .

A typical result for processes is the Invariance Principle, a result we also need for our U -processes. For near epoch dependent sequences on absolutely regular processes it was already proved by Dehling and Philipp (2002). A result for strong mixing can be found in Wendler (2011b). Under independence one can find a strong invariance principle in Dehling et al (1987). These results only consider the bivariate case, whereas we also admit multivariate kernels. For our purposes we only need the convergence of the first term of the Hoeffding decomposition, so the proof will be somewhat different.

From now on consider the case where H_n is our empirical U -process, that is $U_n(t)$ has the kernel $g(x_1, \dots, x_m, t) = 1_{[h(x_1, \dots, x_m) \leq t]}$. Therefore $U(t) = \mathbb{E}(1_{[h(Y_1, \dots, Y_m) \leq t]}) = \mathbb{P}(h(Y_1, \dots, Y_m) \leq t) = H_F(t)$ and since H_F has density $h_F < \infty$ we have that H_F is Lipschitz-continuous.

Theorem 2.4.

Let h be a kernel with distribution function H_F and related density $h_F < \infty$. Moreover, let g_1 be the first term of the Hoeffding decomposition of H_n . Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of strong mixing random variables with mixing coefficients $\alpha(l) = O(l^{-6-\gamma})$ for a $0 < \gamma < 1$. Then

$$\left(\frac{m}{\sqrt{n}} \sum_{i=1}^n g_1(X_i, t) \right)_{t \in \mathbb{R}} \xrightarrow{D} (W(t))_{t \in \mathbb{R}},$$

where W is a continuous Gaussian process.

This theorem can be proved in the same way as Theorem 4.1 of Dehling and Philipp (2002) and is therefore omitted.

By using results concerning the convergence of all remaining terms of the Hoeffding decomposition, which is given in Lemma 4.4, we can state the following corollary.

Corollary 2.1.

Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of strong mixing random variables with mixing coefficients $\alpha(l) = O(l^{-\delta})$ for $\delta \geq 8$ and $\mathbb{E}|X_1|^\rho < \infty$ for a $\rho > \frac{1}{4}$. Moreover let h be a Lipschitz-continuous kernel with distribution function H_F and related density $h_F < \infty$ and for all $2 \leq k \leq m$ let $h_{F; X_2, \dots, X_k}$ be bounded. Then

$$\sup_{t \in \mathbb{R}} |\sqrt{n} (H_n(t) - H_F(t))| = O_p(1).$$

The proofs of all results in this section are given in Section 5.

3. Application: The generalized median estimator

The generalized median (GM -) estimator was developed by Brazauskas and Serfling under independence as a robust estimator of the parameters of different distributions, for example the Pareto distribution or Log-Normal distribution (Brazauskas and Serfling (2000b), Brazauskas and Serfling (2000a) and Serfling (2002)).

We will concentrate on the Pareto distribution, which is a very heavy tailed distribution often used in hydrology and other fields for modelling the tail of a distribution. Its distribution function is given by

$$F(x) = \begin{cases} 1 - \left(\frac{\sigma}{x}\right)^\alpha, & x \geq \sigma \\ 0, & x < \sigma \end{cases},$$

where $\alpha > 0$ and $\sigma > 0$. We assume σ to be unknown and estimate it through the minimum of the sample.

We want to expand the GM-estimator to sequences of strong mixing random variables with Pareto distributed margins and estimate the tail index α . Therefore we have to choose a kernel which is median unbiased. Like Brazauskas and Serfling (2000b) we choose the modified maximum likelihood estimator as kernel, which was shown to be median unbiased under independence, and use this result to show its asymptotical median unbiasedness under strong mixing, that is

$$h(x_1, \dots, x_m) = \frac{M_{2m-2}}{2m} \frac{1}{\left(\frac{1}{m} \sum_{i=1}^m \log x_i - \log(\min((x_1, \dots, x_m)))\right)},$$

where M_{2m-2} is the median of the χ^2_{2m-2} -distribution.

Lemma 3.1. *For a sequence of strong mixing, Pareto distributed random variables $(X_n)_{n \in \mathbb{N}}$ with $\mathbb{E}|X_1|^\rho < \infty$ for a $\rho \geq 1$ and mixing coefficients $\alpha(l) = O(l^{-\delta})$ for a $\delta \geq 8$ the kernel*

$h(x_1, \dots, x_m) = \frac{M_{2m-2}}{2m} \frac{1}{\left(\frac{1}{m} \sum_{i=1}^m \log x_i - \log(\min(x_1, \dots, x_m))\right)}$ is asymptotically median unbiased.

Proof.

We have $\mathbb{E}(H_n - H_F)^2 \rightarrow 0$ using the same arguments as in Lemma 2.1. With arguments of Glivenko-Cantelli type this implies

$$\sup_t |\mathbb{E}(H_n(t) - H_F(t))| \rightarrow 0.$$

Following Example 1 of Pollard (1984) the proof is completed. □

The GM-estimator of the parameter α is then given by

$$\hat{\alpha}_{GM} = \text{med}(h(X_{i_1}, \dots, X_{i_m})),$$

which can be expressed as a GL-statistic by choosing $J = 0$, $d = 1$, $a_1 = 1$, and $p_1 = \frac{1}{2}$. Applying Theorem 2.2 we have

$$\sqrt{n}(\hat{\alpha}_{GM} - \alpha) \xrightarrow{D} N(0, \sigma_{GM}^2), \tag{1}$$

$$\begin{aligned} \sigma_{GM}^2 = & \frac{m^2}{h_F^2(\alpha)} \left(\text{Var}(\mathbb{P}(h(Y_1, \dots, Y_m) \leq \alpha | Y_1 = X_1)) \right. \\ & + 2 \sum_{j=1}^{\infty} \text{Cov}(\mathbb{P}(h(Y_1, \dots, Y_m) \leq \alpha | Y_1 = X_1), \\ & \left. \mathbb{P}(h(Y_1, \dots, Y_m) \leq \alpha | Y_1 = X_{j+1})) \right). \end{aligned}$$

The results concerning robustness given by Brazauskas and Serfling (2000a) remain valid since the kernel is unchanged. Additionally one can show that the influence function of the GM-estimator is bounded (cf. Serfling (1984)).

Remark 3.1. In the case of heavy tailed data one often not only deals with exact Pareto distributed data but has Pareto-type tails. That is

$$F(x) = 1 - x^{-\alpha}L(x),$$

with L being a slowly varying function (Vandewalle (2007)). Then, GM - as well as ML -estimator might be biased. In our case usage of a threshold going to infinity as the number of observations increases would be necessary in order to decrease the bias. This would complicate our theorems further, since we would need a formula for U-Statistics derived from a triangular array.

Before comparing the GM -estimator with another estimator we first want to figure out, how well the approximation of the normal distribution works. We therefore estimate the tail index of independent, identically Pareto-distributed with $\alpha = 2$ and $\sigma = 1$ and also of an AR(1)-process with different autocorrelation coefficient $\rho = 0.2, 0.5, 0.8$ and Pareto-distributed margins. We use 500 simulation runs for each setting, except for the case $m = 4$, where we use only 250 runs because of the large computation times. The number of simulated random variables is chosen as $n = 100$ and for “asymptotic” behaviour as $n = 1000$. We then compare $\sqrt{n}(\hat{\alpha}_{GM} - \alpha)$ with the theoretical normal distribution using a QQ-plot. The results can be found in the Appendix (Figures 2-4).

We can see that for a small kernel dimension $m = 2$ and a sample size of $n = 1000$ the GM -estimator is approximately normal distributed for all cases of dependence (independent and $\rho = 0.2, 0.5, 0.8$). If we reduce the sample size to $n = 100$ we see much larger deviations of the tails from the theoretical quantiles. Especially in the case of $\rho = 0.8$ we can definitely say that no normal distribution is present. As the kernel dimension increases, the normality of the estimator gets better. In case of $m = 4$ we can assume normality even for $n = 100$, if the dependence is moderate (independence or $\rho = 0.2$), although we were not able to simulate all 500 runs. Only for the case $\rho = 0.8$ normality has not been reached yet. If we compare the results with the ML -estimator ($m = n$) we can see a better normal behaviour for $\rho = 0.8$ and $n = 1000$. For all other cases the behaviour is similar to the case of $m = 4$.

We can conclude that for large sample sizes of $n = 1000$ the GM -estimator is normal distributed for independent and moderately dependent ($\rho \leq 0.5$) data. For small sample sizes of $n = 100$ normality can be seen for slightly dependent data ($\rho \leq 0.2$).

In the following simulations we compute confidence intervals for the tail index α using subsampling (cf. Politis and Romano (1994)). We show the coverage probability and the length of the confidence interval for different block lengths in subsampling and three different kernel dimensions of the generalized median estimator, that is $m = 2, 3, 4$. The underlying $n=100$ random variables are computed as independent, identically Pareto-distributed with $\alpha = 2$ and $\sigma = 1$ and also from an AR(1)-process with different autocorrelation coefficient $\rho = -0.5, -0.2, 0.2, 0.5$ and Pareto-distributed margins. The simulation is repeated 500 times. The procedure of subsampling is as follows:

Because $\sqrt{n}(\hat{\alpha}_{GM} - \alpha)$ converges towards an unknown distribution, we estimate the quantiles of the distribution the following way: we first choose a blocklength $b = b_n$ with $b_n \rightarrow \infty$ and $\frac{b_n}{n} \rightarrow 0$ for $n \rightarrow \infty$. Then we calculate the GM -estimator of α for each of the $n - b + 1$ subsamples consisting of b consecutive data values, getting a vector of estimates $(\hat{\alpha}_{GM}^1, \dots, \hat{\alpha}_{GM}^{n-b+1})$. Using

$$L_n(t) = \frac{1}{n - b + 1} \sum_{i=1}^{n-b+1} 1_{[\sqrt{b}(\hat{\alpha}_{GM}^i - \hat{\alpha})] \leq t}$$

the quantiles $q_\gamma^* = L_n^{-1}(\gamma)$ are calculated, where $\hat{\alpha}$ is the GM -estimate for α derived from the whole sample.

The confidence interval CI for a confidence level $1 - \gamma$ is then

$$CI = \left[\hat{\alpha} - \frac{q_{1-\frac{\gamma}{2}}^*}{\sqrt{n}}; \hat{\alpha} - \frac{q_{\frac{\gamma}{2}}^*}{\sqrt{n}} \right],$$

resulting from $\mathbb{P} \left(q_{\frac{\gamma}{2}}^* \leq \sqrt{n}(\hat{\alpha} - \alpha) \leq q_{1-\frac{\gamma}{2}}^* \right) \rightarrow 1 - \gamma$.

These results are compared with the case $m = n$ corresponding to the maximum-likelihood (ML) estimator.

We want to emphasize that the Generalized Median estimator is computationally very demanding. Wilde and Grimshaw (2013) developed an algorithm, which approximates the estimator and speeds the calculation up. For single estimates it is therefore easily applicable. Nevertheless, extensive simulations are not easily possible. This was the reason for us to limit our simulations to a single parameter combination and small samples and runs and also using subsampling instead of bootstrap. Exemplary simulations show that the simulation results remain valid also for other parameter choices.

All simulations were done in R 3.0.1 using the packages VGAM and fExtremes and the algorithm of Wilde and Grimshaw (2013) for the generalized median estimator. We need to mention that the results can fluctuate up to 0.02 because of the moderate number of observation runs (500). For the choice $m = 4$ of the kernel dimension the calculations are very time consuming. Due to time constraints, some simulations had to be stopped after 250 runs. When this is the case, we mark the column with (★).

First we investigate the efficiency of the GM -estimator in comparison with the classical maximum-likelihood estimator corresponding to the case $m = n$. For this we have a look at the coverage probability and the length of the confidence interval under data from an ideal model. As expected we see in Tables 1 and 2 that under independence the coverage probability and the length of the confidence interval of the GM -estimator get better for increasing m , being best when $m = n$, the case of the ML -estimator. Nevertheless even for small values of m the efficiency of the GM -estimator is not too bad compared to the ML -estimator.

Under slight positive dependence ($\rho = 0.2$) the GM -estimator with $m = 4$ performs almost as well as the ML -estimator with $m = n$ and the length

TABLE 1

Confidence interval length and coverage probability of the 90% and 95% confidence intervals from 100 independent, identically Pareto(2,1)-distributed random variables using subsampling under different block lengths for 500 repetitions and kernel dimension 2, 3, 4 and n

block length	90% confidence interval		95% confidence interval	
	coverage probability	length	coverage probability	length
m=2				
15	0.776	0.769	0.848	0.894
20	0.738	0.701	0.818	0.795
m=3				
15	0.778	0.736	0.812	0.845
20	0.770	0.674	0.792	0.738
m=4				
15	0.781	0.720	0.843	0.814
20	0.772	0.683	0.805	0.697
m=n				
15	0.834	0.666	0.846	0.734
20	0.792	0.585	0.818	0.658

TABLE 2

Confidence interval length and coverage probability of the 90% and 95% confidence intervals from 100 random variables from an AR(1)-process with $\rho = 0.2$ and Pareto(2,1)-distributed margins using subsampling under different block lengths for 500 repetitions and kernel dimension 2, 3, 4 and n

block length	90% confidence interval		95% confidence interval	
	coverage probability	length	coverage probability	length
m=2				
15	0.756	0.874	0.778	1.005
20	0.756	0.789	0.770	0.878
m=3				
15	0.794	0.850	0.764	0.950
20	0.724	0.779	0.780	0.864
m=4				
15	0.803	0.838	0.811	0.943
20	0.769	0.744	0.776	0.822
m=n				
15	0.790	0.840	0.814	0.994
20	0.770	0.749	0.796	0.853

of the confidence interval is sometimes even smaller. Note that in the case of dependence, the *GM*-estimator for $m = n$ is not the *ML*-estimator, since it was constructed to maximize the likelihood under independence. But because of Theorem 1 we have the same theoretical guarantees as for $m = 2, 3, 4, \dots$. The construction of a maximum-likelihood estimator under dependence is quite demanding (see for example Beran et al. (2013)) and sometimes the pseudo-likelihood method is used. Nevertheless, this estimator for $m = n$ is widely applied also under dependence (for example in hydrological time series, see Li et al. (2005)) since it is the Moment estimator for α and we use it for comparison. In general the coverage probability and also the length of the confidence interval of the *GM*-estimator are not influenced very much by the size of m ; for the

TABLE 3

Confidence interval length and coverage probability of the 90% and 95% confidence intervals from 100 random variables from an AR(1)-process with $\rho = 0.5$ and Pareto(2,1)-distributed margins using subsampling under different block lengths for 500 repetitions and kernel dimension 2, 3, 4 and n

block length	90% confidence interval		95% confidence interval	
	coverage probability	length	coverage probability	length
m=2				
15	0.546	1.362	0.554	1.539
20	0.520	1.127	0.558	1.257
m=3				
15	0.528	1.218	0.542	1.457
20	0.552	1.098	0.554	1.272
m=4 (★)				
15	0.558	1.278	0.536	1.126
20	0.584	1.212	0.539	1.008
m=n				
15	0.600	1.362	0.606	1.560
20	0.596	1.158	0.644	1.354

TABLE 4

Confidence interval length and coverage probability of the 90% and 95% confidence intervals from 100 random variables from an AR(1)-process with $\rho = -0.2$ and Pareto(2,1)-distributed margins using subsampling under different block lengths for 500 repetitions and kernel dimension 2, 3, 4 and n

block length	90% confidence interval		95% confidence interval	
	coverage probability	length	coverage probability	length
m=2				
15	0.778	0.723	0.844	0.803
20	0.748	0.634	0.768	0.715
m=3				
15	0.770	0.663	0.798	0.748
20	0.716	0.602	0.802	0.687
m=4 (★)				
15	0.772	0.713	0.776	0.637
20	0.724	0.622	0.728	0.578
m=n				
15	0.810	0.619	0.850	0.717
20	0.772	0.547	0.812	0.635

smallest choice of m the coverage probability and the length of the confidence interval of the GM-estimator are rather close to that of the case $m = n$. The same can be seen if the dependence is increased ($\rho = 0.5$, Table 3). Although the coverage probability as well as the length of the confidence interval worsen, the same tendential behaviour as for $\rho = 0.2$ is shown. In the presence of dependence often also the GM-estimator with $m = 4$ behaves better, especially for the length of the interval, than the ML-estimator. The deviation in the results for $m = 4$ probably result from the smaller number of runs.

We also investigated the case where $\rho = 0.8$, but the results for a sample size $n = 100$ were very poor for all cases of m with a coverage probability

TABLE 5
Confidence interval length and coverage probability of the 90% and 95% confidence intervals from 100 random variables from an AR(1)-process with $\rho = -0.5$ and Pareto(2,1)-distributed margins using subsampling under different block lengths for 500 repetitions and kernel dimension 2, 3, 4 and n

block length	90% confidence interval		95% confidence interval	
	coverage probability	length	coverage probability	length
m=2				
15	0.508	0.659	0.532	0.744
20	0.452	0.591	0.500	0.655
m=3				
15	0.466	0.608	0.546	0.707
20	0.470	0.569	0.492	0.622
m=4 (*)				
15	0.444	0.627	0.444	0.572
20	0.452	0.593	0.408	0.528
m=n				
15	0.516	0.571	0.542	0.640
20	0.476	0.477	0.512	0.556

always about 0.3 and a length of the confidence interval between 3 and 10, and therefore, having in mind the results of the normality plots, they are omitted here.

If the correlation is negative (Tables 4 and 5), the results are very similar to case of positive dependence. For $\rho = -0.2$ the coverage probability is even better than for $\rho = 0.2$, though the length of the interval is not.

For independence or moderate dependence ($\rho = 0.2$), the coverage probability decreases when the block length b increases. For stronger dependence ($\rho = 0.8$), the longer block length ($b = 20$) gives better results.

In almost every case of dependence and independence we can see that the coverage probability and the length of the confidence interval worsens when changing the dimension from $m = 2$ to $m = 3$. This cannot be explained by the random deviation of the results, which is at most about 0.02 (see above). It seems, that using a smaller dimension and therefore smaller groups leads sometimes to a better estimation.

Additionally we compared the robustness of the ML -estimator ($m = n$) with the GM -estimator for $m = 2$, the most robust case. A basic requirement for a statistical method to be called robust is that it should be little affected by a single observation deviating from the model. Such data scenarios are called one-wild by Tukey (1960) and lead, e.g., to the sensitivity curves of estimators. Our interest is in statistical inference. Therefore, we contaminate a sample by adding a value y_i lying in the interval $(0, 100]$, and calculate the average coverage probability that is

$$CP(i) = \frac{1}{n} \sum_{j=1}^n (1_{[CI1_j, CI2_j]}(\alpha) - 1_{[CI1_j(i), CI2_j(i)]}(\alpha)),$$

where $CI1_j$ and $CI2_j$ are the bounds of the confidence interval calculated

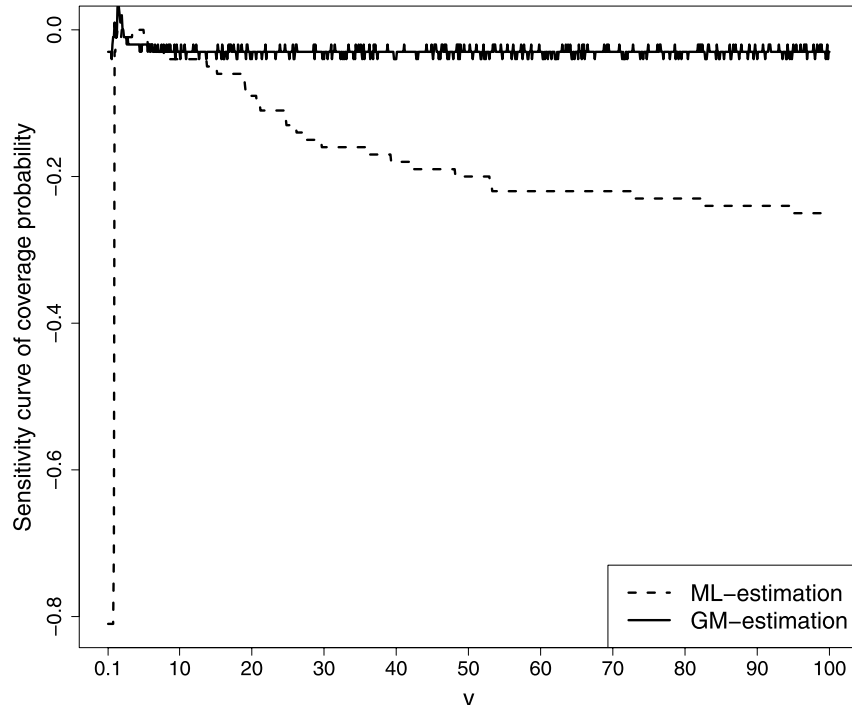


FIG 1. Difference in coverage probability for contamination by one observation y in a sample of size $n = 100$

for the j th sample $(X_j^{(1)}, \dots, X_j^{(n)})$ and $CI1_j(i)$ and $CI2_j(i)$ are the bounds of the confidence interval calculated for the j th sample contaminated by y_i , $(X_j^{(1)}, \dots, X_j^{(n)}, y_i)$, for a confidence level of 0.95 respectively and $j = 1, \dots, 100$. The confidence intervals were again computed by subsampling with a block length of 15. This method is analogous to classical sensitivity curves, but focuses on the coverage probability. The results can be found in Figure 1.

Examining the robustness for data which are contaminated by a value y we can see that for the ML -estimator in all three dependence cases (independence, $\rho = 0.2$ and $\rho = 0.8$) the coverage probability flattens for increasing y but does not reach a constant value. This indicates a non-robust behaviour. The opposite can be seen for the GM -estimator, the coverage probability of which becomes constant when y exceeds 5 and only fluctuates between two values. The behaviour of both estimators close to zero is similar. When y decreases towards the lower bound of the distribution, both estimators have large deviations between the contaminated coverage probability and the uncontaminated one. Nevertheless the results concerning the robustness of the GM -estimator with $m = 2$ are confirmed by the simulations. The results for $m = 3, 4, 5$ were very similar, showing also a robust behaviour of the estimator by a constant coverage probability, and are therefore omitted here.

All in all the performance of the GM -estimator in the simulations depends on the chosen kernel dimension m . Concerning asymptotic normality, large samples of about 1000 observations are needed to get good approximations. The ML -estimator performs better than the other methods in case of independent data, but the asymptotic approximation works well also for the other kernel dimensions in this case. For positive and moderately high values of the dependence parameter ρ the choice $m = 4$ works better than the ML -estimator, and for strong positive correlation $m = 2$ seems best.

Looking at the results shown in the Tables 1-5 and Figure 1, and also at the results of Brazauskas and Serfling (2000a) for independent observations, we can give the following recommendations for the choice of m . For small and moderately large positive values of ρ we agree with the recommendation of Brazauskas and Serfling (2000a) to choose $m = 4$, since this leads to a robust and approximately normally distributed estimator with similar confidence intervals compared to the ML -estimator. If ρ is negative, $m = 2$ seems to be the best choice, since it leads to a more robust behaviour and to confidence intervals with a better coverage rate than the other choices of m , at least in the scenarios investigated here. Finally, the sample sizes considered here are apparently not large enough to get reliable estimations in case of very large values of ρ .

4. Preliminary results

In this section we state some results, which will help us to prove our main results.

First of all we want to use the (extended) variation condition not only for the kernel h , but also for the kernels g_k , $1 \leq k \leq m$, of the Hoeffding decomposition. For that the following lemma is helpful.

Lemma 4.1.

If the kernel h satisfies the extended variation condition, then the kernels g_k from Definition 2.1, $1 \leq k \leq m$, satisfy it as well.

Proof.

The proof will be made by mathematical induction. Initially let $k = 1$. We had defined g_1 as $g_1(x_1) = \mathbb{E}(h(x_1, Y_2, \dots, Y_m)) - \theta$. It is

$$\begin{aligned} & \mathbb{E} \left(\sup_{\|(x_1, \dots, x_m) - (X'_1, \dots, X'_m)\| \leq \epsilon} |g_1(x_1) - g_1(X'_1)| \right) \\ & \leq \mathbb{E} \left(\sup_{\|x_1 - X'_1\| \leq \epsilon} \mathbb{E} |(h(y_1, Y_2, \dots, Y_m) - h(y'_1, Y_2, \dots, Y_m))|_{y_1 = x_1, y'_1 = X'_1} \right) \\ & \leq \mathbb{E} \left(\sup_{\|x_1 - X'_1\| \leq \epsilon} |h(x_1, Y_2, \dots, Y_m) - h(X'_1, Y_2, \dots, Y_m)| \right) \\ & \leq L\epsilon, \end{aligned}$$

because h satisfies the variation condition. So g_1 satisfies the extended variation condition.

Now let g_{k-1} satisfy the extended variation condition. We show that g_k also satisfies it:

$$g_k(x_1, \dots, x_k) = \mathbb{E}(h(x_1, \dots, x_k, Y_{k+1}, \dots, Y_m)) - \theta - \sum_{i=1}^k g_1(x_i) - \dots - \sum_{1 \leq i_1 \leq \dots \leq i_{k-1} \leq k} g_{k-1}(x_{i_1}, \dots, x_{i_{k-1}}).$$

The space of the functions satisfying the (extended) variation condition is a vector space (cf. Wendler (2011b)) and since we know that all kernels up to g_{k-1} satisfy the variation condition, it is sufficient to show that

$\mathbb{E}(h(x_1, \dots, x_k, Y_{k+1}, \dots, Y_m)) - \theta$ satisfies the extended variation condition.

$$\begin{aligned} & \mathbb{E} \left(\sup_{|x_1 - Y_1| \leq \delta} \left| \mathbb{E}(h(y_1, X_2, \dots, X_k, Y_{k+1}, \dots, Y_m) | y_1 = x_1) \right. \right. \\ & \qquad \qquad \qquad \left. \left. - \mathbb{E}(h(y'_1, X_2, \dots, X_k, Y_{k+1}, \dots, Y_m) | y'_1 = Y_1) \right| \right) \\ & \leq \mathbb{E} \left(\sup_{|x_1 - Y_1| \leq \delta} \left| h(x_1, X_2, \dots, X_k, Y_{k+1}, \dots, Y_m) \right. \right. \\ & \qquad \qquad \qquad \left. \left. - h(Y_1, X_2, \dots, X_k, Y_{k+1}, \dots, Y_m) \right| \right) \\ & \leq L' \delta, \end{aligned}$$

since h satisfies the extended variation condition. □

Remark 4.1. All results shown before for the extended variation condition without parameter t remain true for the extended uniform variation condition.

To ultimately show the asymptotic normality of U -statistics of strongly mixing random variables, we will first generalize some lemmas proved by Wendler (2011b) respectively Dehling and Wendler (2010) or Wendler (2011a) from the case $m = 2$ to arbitrary m .

First we need a covariance inequality, which we can establish by the coupling technique. A similar result for absolutely regular variables can be found in Yoshihara (1976). Here we will follow Wendler (2011b) and expand the lemma to the case $m \geq 2$, meaning we will treat g_k for $2 \leq k \leq m$. The proof is analogous to Wendler (2011b) using the extended variation condition instead of the ordinary one and is therefore omitted.

Lemma 4.2.

Let $(X_n)_{n \in \mathbb{N}}$ be a strong mixing sequence of random variables with $\mathbb{E}|X_1|^\rho < \infty$ for a $\rho > 0$ and h a bounded kernel, which satisfies the extended variation condition. Moreover set $l = \max\{i_{(2)} - i_{(1)}, i_{(2k)} - i_{(2k-1)}\}$, where $\{i_1, \dots, i_{2k}\} = \{i_{(1)}, \dots, i_{(2k)}\}$ and $i_{(1)} \leq \dots \leq i_{(2k)}$. Then there exists a constant C , such that

for all $2 \leq k \leq m$

$$|\mathbb{E}(g_k(X_{i_1}, \dots, X_{i_k})g_k(X_{i_{k+1}}, \dots, X_{i_{2k}}))| \leq C\alpha^{\frac{\rho}{2\rho+1}}(l).$$

Lemma 4.3.

Let the kernel h be bounded and satisfy the extended variation condition. Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of strong mixing random variables with $\mathbb{E}|X_1|^\rho < \infty$ for a $\rho > 0$ and let $\sum_{l=0}^n l\alpha^{\frac{\rho}{2\rho+1}}(l) = O(n^\gamma)$ for a $\gamma \geq 0$ hold. Then for all $2 \leq k \leq m$

$$\sum_{i_1, \dots, i_{2k}=1}^n |\mathbb{E}(g_k(X_{i_1}, \dots, X_{i_k})g_k(X_{i_{k+1}}, \dots, X_{i_{2k}}))| = O(n^{2k-2+\gamma}).$$

Proof.

Set $\{i_1, \dots, i_{2k}\} = \{i_{(1)}, \dots, i_{(2k)}\}$ with $i_{(1)} \leq \dots \leq i_{(2k)}$. We can rewrite the above sum as

$$\begin{aligned} & \sum_{i_1, \dots, i_{2k}=1}^n |\mathbb{E}(g_k(X_{i_1}, \dots, X_{i_k})g_k(X_{i_{k+1}}, \dots, X_{i_{2k}}))| \\ &= \sum_{l=0}^n \sum_{\substack{i_1, \dots, i_{2k}=1 \\ \max\{i_{(2)}-i_{(1)}, i_{(2k)}-i_{(2k-1)}\}=l}} |\mathbb{E}(g_k(X_{i_1}, \dots, X_{i_k})g_k(X_{i_{k+1}}, \dots, X_{i_{2k}}))| \\ &\leq C \sum_{l=0}^n \sum_{\substack{i_1, \dots, i_{2k} \\ \max\{i_{(2)}-i_{(1)}, i_{(2k)}-i_{(2k-1)}\}=l}} \alpha^{\frac{\rho}{1+2\rho}}(l), \end{aligned}$$

by application of Lemma 4.2.

For a further simplification we calculate via combinatorial arguments the quantity of the terms of the inner sum, that is the quantity of tuples (i_1, \dots, i_{2k}) where $\max\{i_{(2)} - i_{(1)}, i_{(2k)} - i_{(2k-1)}\} = l$. At first there are $(2k)!$ possibilities for a $2k$ -tuple to get the same ordered sequence $i_{(1)}, \dots, i_{(2k)}$. Now we choose $i_{(1)}$ and $i_{(2k)}$ fixed and have n^2 possibilities for doing so. Through the requirement $\max\{i_{(2)} - i_{(1)}, i_{(2k)} - i_{(2k-1)}\} = l$ we can also calculate the remaining possibilities for $i_{(2)}$ and $i_{(2k-1)}$. Suppose $i_{(2)} - i_{(1)} = \max\{i_{(2)} - i_{(1)}, i_{(2k)} - i_{(2k-1)}\} = l$ then $i_{(2)}$ is automatically determined by the established choice of $i_{(1)}$. Because the requirement on the maximum still has to be fulfilled, $i_{(2k-1)}$ can only take l distinct values. In the other case $i_{(2k)} - i_{(2k-1)} = \max\{i_{(2)} - i_{(1)}, i_{(2k)} - i_{(2k-1)}\} = l$ we come to the same result. All remaining values of the k -tuple are arbitrary. Consequently the inner sum altogether is $(2k)! \cdot n^2 l n^{2k-4} = l \cdot (2k)! \cdot n^{2k-2}$ and therefore

$$\begin{aligned} & \sum_{i_1, \dots, i_{2k}=1}^n |\mathbb{E}(g_k(X_{i_1}, \dots, X_{i_k})g_k(X_{i_{k+1}}, \dots, X_{i_{2k}}))| \\ &\leq C' n^{2k-2} \sum_{l=0}^n l\alpha^{\frac{\rho}{1+2\rho}}(l) = O(n^{2k-2+\gamma}). \quad \square \end{aligned}$$

We also need results concerning the remaining terms of the Hoeffding decomposition for U -processes. In this case we of course do not need simple convergence against zero, but since we consider processes need to have convergence of the supremum.

The following lemma was proved by Wendler (2011b) for the case $m = 2$. We will modify the main idea of the proof to obtain a similar result for the degenerated terms of higher dimensional U -processes.

Lemma 4.4.

Let h be a kernel satisfying the extended uniform variation condition, such that the U -distribution function U is Lipschitz-continuous. Moreover let $(X_n)_{n \in \mathbb{N}}$ be a sequence of strong mixing random variables with mixing coefficients $\alpha(l) = O(l^{-\delta})$ for $\delta \geq 8$ and $\mathbb{E}|X_i|^\rho < \infty$ for a $\rho > \frac{1}{4}$. Then for all $2 \leq k \leq m$ and $\gamma = \frac{\delta-2}{\delta}$ we have

$$\sup_{t \in \mathbb{R}} \left| \sum_{1 \leq i_1, \dots, i_k \leq n} g_k(X_{i_1}, \dots, X_{i_k}, t) \right| = o(n^{k-\frac{1}{2}-\frac{\gamma}{8}}) \text{ a.s..}$$

Proof.

We define $Q_n^k(t) := \sum_{1 \leq i_1, \dots, i_k \leq n} g_k(X_{i_1}, \dots, X_{i_k}, t)$.

For $l \in \mathbb{N}$ choose $t_{1,l}, \dots, t_{s-1,l}$ with $s = s_l = O(2^{\frac{5}{8}l})$, such that

$$-\infty = t_{0,l} < t_{1,l} < \dots < t_{s-1,l} < t_{s,l} = \infty$$

and $2^{-\frac{5}{8}l} \leq |U(t_{r,l}) - U(t_{r-1,l})| \leq 2 \cdot 2^{-\frac{5}{8}l}$. Since we required Lipschitz-continuity of U it follows that $2^{-\frac{5}{8}l} \leq C|t_{r,l} - t_{r-1,l}|$. Moreover, because h is non-decreasing in t ,

$\mathbb{E}(h(Y_1, \dots, Y_k, Y_{k+1}, \dots, Y_m, t) | Y_1 = X_{i_1}, \dots, Y_k = X_{i_k})$ is non-decreasing in t for all $2 \leq k \leq m$. We proceed by induction.

The case $k = 2$ was treated by Wendler (2011b) and is therefore omitted here.

From now on suppose that the statement of the lemma is valid for $k - 1$.

Together with the above consideration we have for every $t \in [t_{r-1,l}, t_{r,l}]$ and $2^l \leq n < 2^{l+1}$

$$\begin{aligned} & |Q_n^k(t)| \\ = & \left| \sum_{1 \leq i_1 < \dots < i_k \leq n} (\mathbb{E}(h(Y_1, \dots, Y_k, Y_{k+1}, \dots, Y_m, t) | Y_1 = X_{i_1}, \dots, Y_k = X_{i_k}) \right. \\ & \quad - g_1(X_{i_1}, t) - \dots - g_1(X_{i_k}, t) \\ & \quad \left. - g_2(X_{i_1}, X_{i_2}, t) - \dots - g_2(X_{i_{k-1}}, X_{i_k}, t) - \dots - U(t)) \right| \\ \leq & \max \left\{ \left| \sum_{1 \leq i_1 < \dots < i_k \leq n} (\mathbb{E}(h(X_{i_1}, \dots, X_{i_k}, Y_{i_{k+1}}, \dots, Y_m, t_{r,l})) \right. \right. \end{aligned}$$

$$\begin{aligned}
& -g_1(X_{i_1}, t_{r,l}) - \dots - g_1(X_{i_k}, t_{r,l}) \\
& - g_2(X_{i_1},) - \dots - g_2(X_{i_{k-1}}, X_{i_k}, t_{r,l}) - \dots - U(t_{r,l})) \Big|, \\
& \left| \sum_{1 \leq i_1 < \dots < i_k \leq n} (\mathbb{E}(h(X_{i_1}, \dots, X_{i_k}, Y_{i_{k+1}}, \dots, Y_m, t_{r-1,l})) \right. \\
& \quad - g_1(X_{i_1}, t_{r-1,l}) - \dots - g_1(X_{i_k}, t_{r-1,l}) \\
& \quad - g_2(X_{i_1}, X_{i_2}, t_{r-1,l}) - \dots - g_2(X_{i_{k-1}}, X_{i_k}, t_{r-1,l}) \\
& \quad \left. - \dots - U(t_{r-1,l})) \Big| \right\} \\
& + \binom{n-1}{k-1} \max \left\{ \left| \sum_{i=1}^n (g_1(X_i, t_{r,l}) - g_1(X_i, t)) \right|, \right. \\
& \quad \left. \left| \sum_{i=1}^n (g_1(X_i, t) - g_1(X_i, t_{r-1,l})) \right| \right\} \\
& + \binom{n-2}{k-2} \max \left\{ \left| \sum_{i=1}^n (g_2(X_{i_1},) - g_2(X_{i_1}, X_{i_2}, t)) \right|, \right. \\
& \quad \left. \left| \sum_{i=1}^n (g_2(X_{i_1}, X_{i_2}, t) - g_2(X_{i_1}, X_{i_2}, t_{r-1,l})) \right| \right\} \\
& + \dots + \binom{n}{k} |U(t_{r,l}) - U(t_{r-1,l})| \\
& \leq \max\{|Q_n^k(t_{r,l})|, |Q_n^k(t_{r-1,l})|\} \\
& + \binom{n-1}{k-1} \left| \sum_{i=1}^n (g_1(X_i, t_{r,l}) - g_1(X_i, t_{r-1,l})) \right| \\
& + \binom{n-2}{k-2} \left| \sum_{1 \leq i_1 < i_2 \leq n} (g_2(X_{i_1}, X_{i_2}, t_{r,l}) - g_2(X_{i_1}, X_{i_2}, t_{r-1,l})) \right| \\
& + \dots + \binom{n-(k-1)}{k-(k-1)} \left| \sum_{1 \leq i_1 < \dots < i_{k-1} \leq n} (g_{k-1}(X_{i_1}, \dots, X_{i_{k-1}}, t_{r,l}) \right. \\
& \quad \left. - g_{k-1}(X_{i_1}, \dots, X_{i_{k-1}}, t_{r-1,l})) \right| \\
& + \binom{n}{k} |U(t_{r,l}) - U(t_{r-1,l})|.
\end{aligned}$$

Again we will treat the first, second and last term of the sum separately.

For the first term follows

$$\begin{aligned}
 & \mathbb{E} \left(\max_{n=2^l, \dots, 2^{l+1}-1} \max_{r=0, \dots, s} |Q_n^k(t_{r,l})|^2 \right) \\
 & \leq \sum_{r=0}^s \mathbb{E} \left(\left(\sum_{d=0}^l \max_{i=1, \dots, 2^{l-d}} |Q_{2^l+i2^d}^k(t_{r,l}) - Q_{2^l+(i-1)2^d}^k(t_{r,l})| \right)^2 \right) \\
 & \leq \sum_{r=0}^s l \sum_{d=0}^l \sum_{i=1}^{2^{-d}} \mathbb{E} \left(\left(Q_{2^l+i2^d}^k(t_{r,l}) - Q_{2^l+(i-1)2^d}^k(t_{r,l}) \right)^2 \right) \\
 & \leq \sum_{r=0}^s l \sum_{d=0}^l \underbrace{\sum_{i_1, \dots, i_d=1}^{2^{l+1}} |\mathbb{E} (g_k(X_{i_1}, \dots, X_{i_k}, t) g_k(X_{i_{k+1}}, \dots, X_{i_{2k}}, t))|}_{=O((2^{l+1})^{2k-2+\gamma}), \text{ with Lemma 4.3}} \\
 & \leq sl^2 C 2^{(2k-2)(l+1)} \leq C' l^2 2^{(2k-2+\frac{5}{8})l}.
 \end{aligned}$$

For the first inequality we used the so called chaining technique: via the triangular inequality we parted the term Q_n into two differences $Q_{2^l+i2^d} - Q_{2^l+(i-1)2^d}$.

Now we apply the Chebychev inequality getting for every $\epsilon > 0$

$$\begin{aligned}
 & \sum_{l=1}^{\infty} \mathbb{P} \left(\max_{n=2^l, \dots, 2^{l+1}-1} \max_{r=0, \dots, s} |Q_n^k(t_{r,l})| > \epsilon 2^{l(k-\frac{1}{2}-\frac{\gamma}{8})} \right) \\
 & \leq \sum_{l=1}^{\infty} \frac{1}{\epsilon^2 2^{l(2k-1-\frac{\gamma}{4})}} \mathbb{E} \left(\max_{n=2^l, \dots, 2^{l+1}-1} \max_{r=0, \dots, s} |Q_n^k(t_{r,l})|^2 \right) \\
 & \leq \sum_{l=1}^{\infty} \frac{1}{\epsilon^2 2^{l(2k-1-\frac{\gamma}{4})}} C' l^2 2^{(2k-2+\frac{5}{8})l} \leq \sum_{l=1}^{\infty} C' \frac{l^2}{\epsilon^2} 2^{\frac{-3+2\gamma}{8}l} < \infty.
 \end{aligned}$$

Then with the Borel-Cantelli Lemma

$$\mathbb{P} \left(\max_{n=2^l, \dots, 2^{l+1}-1} \max_{r=0, \dots, s} |Q_n^2(t_{r,l})| > \epsilon 2^{l(k-\frac{1}{2}-\frac{\gamma}{8})} \text{ infinitely often} \right) = 0.$$

That is, $\max_{r=0, \dots, s} |Q_n^k(t_{r,l})| = o(n^{k-\frac{1}{2}-\frac{\gamma}{8}})$.

Now we will treat the second term of the sum for which we want to apply Lemma 4.2.1 of Wendler (2011b). For $2^l \leq n < 2^{l+1}$ it follows

$$\begin{aligned}
 & \mathbb{E} \left(\sum_{i=1}^n (g_1(X_i, t_{r,l}) - g_1(X_i, t_{r-1,l})) \right)^4 \\
 & \leq C n^2 (\log n)^2 \max \left\{ \mathbb{E} |g_1(X_i, t_{r,l}) - g_1(X_i, t_{r-1,l})|, C n^{-\frac{3}{4}} \right\}^{1+\gamma} \\
 & \leq C n^2 (\log n)^2 (C n^{-\frac{3}{4}})^{1+\gamma}.
 \end{aligned}$$

By usage of the assumption $|U(t_{r,l}) - U(t_{r-1,l})| \geq 2^{-\frac{5}{8}l} \geq C2^{-\frac{3}{4}l} \geq Cn^{-\frac{3}{4}}$, the last term simplifies to

$$Cn^2(\log n)^2|U(t_{r,l}) - U(t_{r-1,l})|^{1+\gamma}.$$

All in all we get

$$\begin{aligned} & \mathbb{E} \left(\max_{n=2^l, \dots, 2^{l+1}-1} \max_{r=1, \dots, s} \binom{n-1}{k-1} \left| \sum_{i=1}^n (g_1(X_i, t_{r,l}) - g_1(X_i, t_{r-1,l})) \right| \right)^4 \\ & \leq n^{4(k-1)} \sum_{r=1}^s \mathbb{E} \left(\max_{n=2^l, \dots, 2^{l+1}-1} \left| \sum_{i=1}^n (g_1(X_i, t_{r,l}) - g_1(X_i, t_{r-1,l})) \right| \right)^4 \\ & \leq n^{4(k-1)} \sum_{r=1}^s Cn^2(\log n)^2|U(t_{r,l}) - U(t_{r-1,l})|^{1+\gamma} \\ & \leq 2^{4(k-1)(l+1)} Cn^2(\log n)^2 s \left(\max_{r=1, \dots, s} |U(t_{r,l}) - U(t_{r-1,l})| \right)^{1+\gamma} \\ & \leq C'(l+1)^2 2^{(4k-2-\frac{5}{8}\gamma)l}. \end{aligned}$$

Thereby we used Corollary 1 of Moricz (1983) and the assumption $s = O(2^{\frac{5}{8}l})$.

Analogously to the above calculation we again apply the generalized Chebychev Inequality and the Borel-Cantelli Lemma getting

$$\binom{n-1}{k-1} \left| \sum_{i=1}^n (g_1(X_i, t_{r,l}) - g_1(X_i, t_{r-1,l})) \right| = o(n^{k-\frac{1}{2}-\frac{1}{8}\gamma}).$$

For the last term of the sum, using the assumptions and the fact that $\gamma < 1$, we have

$$\begin{aligned} \max_{r=0, \dots, s} \binom{n}{k} |U(t_{r,l}) - U(t_{r-1,l})| & \leq Cn^k 2^{-\frac{5}{8}l} \leq Cn^{k-\frac{5}{8}} < Cn^{k-\frac{4}{8}-\frac{1}{8}\gamma} \\ & = o(n^{k-\frac{4}{8}-\frac{1}{8}\gamma}). \end{aligned}$$

Now the terms including g_2, \dots, g_{k-1} remain. For these we know for $2 \leq j \leq k-1$

$$\sup_{t \in \mathbb{R}} \left| \sum_{1 \leq i_1, \dots, i_j \leq n} g_j(X_{i_1}, \dots, X_{i_j}, t) \right| = o(n^{j-\frac{1}{2}-\frac{\delta-2}{8\delta}})$$

and consequently

$$\begin{aligned} & \binom{n-j}{k-j} \max_{r=1, \dots, s} \left| \sum_{1 \leq i_1 < \dots < i_j \leq n} (g_j(X_{i_1}, \dots, X_{i_j}, t_{r,l}) - g_j(X_{i_1}, \dots, X_{i_j}, t_{r-1,l})) \right| \\ & \leq n^{k-j} \left(\max_{r=1, \dots, s} \left| \sum_{1 \leq i_1 < \dots < i_j \leq n} g_j(X_{i_1}, \dots, X_{i_j}, t_{r,l}) \right| \right) \end{aligned}$$

$$\begin{aligned}
 & + \max_{r=1, \dots, s} \left| \sum_{1 \leq i_1 < \dots < i_j \leq n} g_j(X_{i_1}, \dots, X_{i_j}, t_{r-1, l}) \right| \\
 \leq & n^{k-j} o(n^{j-\frac{1}{2}-\frac{1}{8}\frac{\delta-2}{\delta}}) = o(n^{k-\frac{1}{2}-\frac{1}{8}\frac{\delta-2}{\delta}}).
 \end{aligned}$$

So we could show for arbitrary k and all terms of the sum that they are of order $o(n^{k-\frac{1}{2}-\frac{1}{8}\frac{\delta-2}{\delta}})$. Using mathematical induction the proof is completed. \square

5. Proofs

In this section we give the proofs of the main results stated in Section 2.

Theorem 2.2. For the main proof we have to show that the following three conditions are fulfilled. Serfling (1984) has already proved that these conditions jointly imply asymptotic normality. From there one can see that independence is not required, if these conditions are fulfilled. Some of the lemmas used for proving this theorem can also be found in Choudhury and Serfling (1988).

- (i) For $W_{H_n, H_F}(y) = \left(\frac{\int_0^{H_n(y)} J(t)dt - \int_0^{H_F(y)} J(t)dt}{H_n(y) - H_F(y)} - J(H_F(y)) \right)$ holds $\|W_{H_n, H_F}\|_{L_1} = o_p(1)$ and it is $\|H_n - H_F\|_\infty = O_p(n^{-\frac{1}{2}})$.
- (ii) For the remainder term $R_{p_i, n} = \hat{\xi}_{p_i, n} - \xi_{p_i} + \frac{p_i - H_n(\xi_{p_i})}{h_f(\xi_{p_i})}$ of the Bahadur representation of an empirical quantile holds

$$R_{p_i, n} = o_p(n^{-\frac{1}{2}}).$$

- (iii) For a U -statistic with kernel

$$\begin{aligned}
 A(x_1, \dots, x_m) = & - \int_{-\infty}^{\infty} (1_{[h(x_1, \dots, x_m) \leq y]} - H_F(y)) J(H_F(y)) dy \\
 & + \sum_{i=1}^d a_i \frac{p_i - 1_{[h(x_1, \dots, x_m) \leq H_F^{-1}(p_i)]}}{h_F(H_F^{-1}(p_i))}
 \end{aligned}$$

we have

$$\sqrt{n}(U_n(A) - \theta) \xrightarrow{D} N(0, \sigma^2).$$

Proofs of the conditions

Now we show that conditions (i)-(iii) are satisfied.

For the first part of condition (i) we refer to Lemma 8.2.4.A of Serfling (1980). Although he demands independence of the random variables in his proof this property is not needed. The second part of condition (i) follows from Corollary 2.1.

Condition (ii) is fulfilled by Lemma 2.1.

It remains to show that condition (iii) is satisfied.

For this we apply Theorem 2.3. We merely have to verify, whether A satisfies the assumptions for the kernel, that is (a) A is bounded and (b) satisfies the

extended variation condition. We consider again the kernel A

$$A(x_1, \dots, x_m) = - \int_{-\infty}^{\infty} (1_{[h(x_1, \dots, x_m) \leq y]} - H_F(y)) J(H_F(y)) dy \\ + \sum_{i=1}^d a_i \frac{p_i - 1_{[h(x_1, \dots, x_m) \leq H_F^{-1}(p_i)]}}{h_F(H_F^{-1}(p_i))}.$$

- (a) The boundedness is a result of the continuity of H_F and J and that J vanishes off the interval $[\alpha, \beta]$.
 (b) Now we want to show that A satisfies the extended variation condition. We will treat both terms of the sum separately, at first for arbitrary y_1, \dots, y_m :

$$\mathbb{E} \left(\sup_{\|(x_1, \dots, x_m) - (y_1, \dots, y_m)\| \leq \epsilon} |A(x_1, \dots, x_m) - A(y_1, \dots, y_m)| \right) \\ \leq \mathbb{E} \left(\sup_{\|(x_1, \dots, x_m) - (y_1, \dots, y_m)\| \leq \epsilon} \left| \int_{-\infty}^{\infty} (1_{[h(x_1, \dots, x_m) \leq y]} - H_F(y)) J(H_F(y)) dy \right. \right. \\ \left. \left. - \int_{-\infty}^{\infty} (1_{[h(y_1, \dots, y_m) \leq y]} - H_F(y)) J(H_F(y)) dy \right| \right) \\ + \mathbb{E} \left(\sup_{\|(x_1, \dots, x_m) - (y_1, \dots, y_m)\| \leq \epsilon} \left| \sum_{i=1}^d a_i \frac{p_i - 1_{[h(x_1, \dots, x_m) \leq H_F^{-1}(p_i)]}}{h_F(H_F^{-1}(p_i))} \right. \right. \\ \left. \left. - \sum_{i=1}^d a_i \frac{p_i - 1_{[h(y_1, \dots, y_m) \leq H_F^{-1}(p_i)]}}{h_F(H_F^{-1}(p_i))} \right| \right) \\ \leq \mathbb{E} \left(\underbrace{\sup_{\|(x_1, \dots, x_m) - (y_1, \dots, y_m)\| \leq \epsilon} \left| \int_{-\infty}^{\infty} (1_{[h(x_1, \dots, x_m) \leq y]} \right. \right.} \\ \left. \left. - 1_{[h(y_1, \dots, y_m) \leq y]} J(H_F(y)) dy \right|}_{=: A_1(y_1, \dots, y_m)} \right) \\ + \mathbb{E} \left(\sup_{\|(x_1, \dots, x_m) - (y_1, \dots, y_m)\| \leq \epsilon} \underbrace{\left| \sum_{i=1}^d a_i \frac{p_i - 1_{[h(y_1, \dots, y_m) \leq H_F^{-1}(p_i)]} - 1_{[h(x_1, \dots, x_m) \leq H_F^{-1}(p_i)]}}{h_F(H_F^{-1}(p_i))} \right|}_{=: A_2(y_1, \dots, y_m)} \right).$$

For the verification of the simple variation condition we first treat A_1 getting

$$A_1(X'_1, \dots, X'_m) \\ \leq \left| \int_{-\infty}^{\infty} J(H_F(y)) \right.$$

$$\sup_{\|(x_1, \dots, x_m) - (X'_1, \dots, X'_m)\| \leq \epsilon} \left| \mathbb{1}_{[h(x_1, \dots, x_m) \leq y]} - \mathbb{1}_{[h(X'_1, \dots, X'_m) \leq y]} \right| dy$$

Using the Lipschitz-continuity we have

$$\begin{aligned} & \sup_{\|(x_1, \dots, x_m) - (X'_1, \dots, X'_m)\| \leq \epsilon} \left| \mathbb{1}_{[h(X'_1, \dots, X'_m) \leq y]} - \mathbb{1}_{[h(x_1, \dots, x_m) \leq y]} \right| \\ &= \begin{cases} 1 & , \text{ if } h(X'_1, \dots, X'_m) \in (y - \tilde{L}\epsilon, y + \tilde{L}\epsilon) \\ 0 & , \text{ else.} \end{cases} \end{aligned}$$

One can easily see that $C := \left| \int_{-\infty}^{\infty} J(H_F(y)) dy \right|$ is bounded. Therefore

$$\begin{aligned} & \mathbb{E}(A_1(X'_1, \dots, X'_m)) \\ & \leq \mathbb{E} \left(\sup_{t \in \mathbb{R}} \left| \mathbb{1}_{[h(X'_1, \dots, X'_m) \in (t - \tilde{L}\epsilon, t + \tilde{L}\epsilon)]} \right| \cdot \left| \int_{-\infty}^{\infty} J(H_F(y)) dy \right| \right) \\ & \leq \sup_{t \in \mathbb{R}} \left| \mathbb{E} \left(\mathbb{1}_{[h(X'_1, \dots, X'_m) \in (t - \tilde{L}\epsilon, t + \tilde{L}\epsilon)]} \right) \right| \cdot C \\ & \leq C \cdot \sup_{t \in \mathbb{R}} \left| \mathbb{P} \left(h(X'_1, \dots, X'_m) \in (t - \tilde{L}\epsilon, t + \tilde{L}\epsilon) \right) \right| \\ & \leq C \cdot \sup_{t \in \mathbb{R}} \left| \int_{t - \tilde{L}\epsilon}^{t + \tilde{L}\epsilon} h_F(x) dx \right| \leq C \cdot \left(\sup_{x \in \mathbb{R}} h_F(x) \right) 2\tilde{L}\epsilon \leq L\epsilon, \end{aligned}$$

since h_F is bounded.

The treatment of A_2 is analogous, using the same notation of the supremum as above. Therefore A satisfies the variation condition and using the same arguments for the extended variation condition the proof is finished.

We have shown conditions (i)-(iii) and so the proof of asymptotic normality is completed. \square

Theorem 2.1.

Let be $t \in \mathbb{R}$, $\xi_{nt} = \xi_p + tn^{-\frac{1}{2}}$, $Z_n(t) = \sqrt{n} \frac{H_F(\xi_{nt}) - H_n(\xi_{nt})}{h_F(\xi_p)}$ and $V_n(t) = \sqrt{n} \frac{H_F(\xi_{nt}) - H_n(\hat{\xi}_p)}{h_F(\xi_p)}$.

Using $|p - H_n(\hat{\xi}_p)| \leq \frac{1}{n}$ we obtain

$$V_n(t) = \underbrace{\sqrt{n} \frac{H_F(\xi_p + tn^{-\frac{1}{2}}) - p}{h_F(\xi_p)}}_{=: V'_n(t)} + \underbrace{\sqrt{n} \frac{\overbrace{p - H_n(\hat{\xi}_p)}^{=O(n^{-1})}}{h_F(\xi_p)}}_{=: O(n^{-\frac{1}{2}})} \rightarrow t.$$

Next we will show that $Z_n(t) - Z_n(0) \xrightarrow{P} 0$. One can easily see that

$$\text{Var}(Z_n(t) - Z_n(0))$$

$$= \frac{n}{h_F^2(\xi_p)} \text{Var} \left(\frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} 1_{[h(X_{i_1}, \dots, X_{i_m}) \leq \xi_p + tn^{-\frac{1}{2}}]} - 1_{[h(X_{i_1}, \dots, X_{i_m}) \leq \xi_p]} \right).$$

To find bounds for the right hand side, we define U_n and U'_n as

$$\begin{aligned} U_n &= \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} 1_{[h(X_{i_1}, \dots, X_{i_m}) \leq \xi_p + tn^{-\frac{1}{2}}]} \\ &= \theta + \sum_{j=1}^m \binom{m}{j} \frac{1}{\binom{n}{j}} \sum_{1 \leq i_1 < \dots < i_j \leq n} g_k(X_{i_1}, \dots, X_{i_k}) \\ U'_n &= \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} 1_{[h(X_{i_1}, \dots, X_{i_m}) \leq \xi_p]} \\ &= \theta' + \sum_{j=1}^m \binom{m}{j} \frac{1}{\binom{n}{j}} \sum_{1 \leq i_1 < \dots < i_j \leq n} g'_k(X_{i_1}, \dots, X_{i_k}), \end{aligned}$$

where g_k and g'_k are the related terms of the Hoeffding decomposition as used before.

Therefore we have

$$\begin{aligned} &\sqrt{\text{Var} \left(\frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} 1_{[\xi_p < h(X_{i_1}, \dots, X_{i_m}) \leq \xi_p + tn^{-\frac{1}{2}}]} \right)} \\ &\leq \underbrace{\sqrt{\text{Var}(\theta)}}_{=0} + \underbrace{\sqrt{\text{Var}(\theta')}}_{=0} + \sqrt{\text{Var} \left(\frac{m}{n} \sum_{i=1}^n (g_1(X_i) - g'_1(X_i)) \right)} \\ &+ \sqrt{\text{Var} \left(\frac{\binom{m}{2}}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} g_2(X_i, X_j) \right)} + \sqrt{\text{Var} \left(\frac{\binom{m}{2}}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} g'_2(X_i, X_j) \right)} \\ &+ \dots + \sqrt{\text{Var} \left(\frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} g_m(X_{i_1}, \dots, X_{i_m}) \right)} \\ &+ \sqrt{\text{Var} \left(\frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} g'_m(X_{i_1}, \dots, X_{i_m}) \right)}. \end{aligned}$$

We have shown in the proof of Theorem 2.3 that for all $2 \leq k \leq m$ it is

$$\text{Var} \left(\frac{\binom{m}{k}}{\binom{n}{k}} \sum_{1 \leq i_1 < \dots < i_k \leq n} g_k(X_{i_1}, \dots, X_{i_k}) \right) = O(n^{-2+\gamma})$$

for a $\gamma < 1$, if the kernel is bounded and satisfies the extended variation condition. Analogous to the proof of Corollary 2.1 we know that $g(x_1, \dots, x_m) = 1_{[h(X_{i_1}, \dots, X_{i_m}) \leq \xi_p + tn^{-\frac{1}{2}}]}$ and $g'(x_1, \dots, x_m) = 1_{[h(X_{i_1}, \dots, X_{i_m}) \leq \xi_p]}$ satisfy the extended variation condition.

Applying Proposition 1 of Doukhan et al (2010) on $g_1(X_i) - g'_1(X_i)$ and $p = 2, b = 3$ and using $\|g_1(X_i) - g'_1(X_i)\|_3 < \infty$, since the kernels are bounded, we have

$$\mathbb{E} \left| \sum_{i=1}^n (g_1(X_i) - g'_1(X_i)) \right|^2 \leq Cn,$$

where the constant

$$C = 4 \left(\int_0^1 \left(\min \left\{ \sum_{i \geq 0} 1_{[u < \alpha(i)]}, n \right\} \right)^3 du \right)^{\frac{1}{3}} \|g_1(X_i) - g'_1(X_i)\|_3^2$$

only depends on $\|g_1(X_i) - g'_1(X_i)\|_3$, since Doukhan and Lang (2009) proved

$$\left(\int_0^1 \left(\min \left\{ \sum_{i \geq 0} 1_{[u < \alpha(i)]}, n \right\} \right)^3 du \right)^{\frac{1}{3}} < \infty.$$

So we get

$$\begin{aligned} & \sqrt{\text{Var} \left(\frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} 1_{[\xi_p < h(X_{i_1}, \dots, X_{i_m}) \leq \xi_p + tn^{-\frac{1}{2}}]} \right)} \\ & \leq \sqrt{\frac{m^2}{n^2} Cn + 2(m-1)\sqrt{O(n^{-2+\gamma})}} \leq \frac{Cm^2}{\sqrt{n}} + 2(m-1)O(n^{-1+\gamma/2}), \end{aligned}$$

where the constant C only depends on $\|g_1(X_i) - g'_1(X_i)\|_3$.

Let us come back to

$$\begin{aligned} & \text{Var}(Z_n(t) - Z_n(0)) \\ & \leq \frac{n}{h_F^2(\xi_p)} \left(\frac{Cm^2}{\sqrt{n}} + 2(m-1)O(n^{-1+\gamma/2}) \right)^2 \\ & \leq \frac{m^2}{h_F^2(\xi_p)} C^2 + \frac{4m^2(m-1)}{h_F^2(\xi_p)} C\sqrt{n}O(n^{-1+\gamma/2}) + 4(m-1)^2O(n^{-2+\gamma}) \\ & \leq \frac{m^2}{h_F^2(\xi_p)} C^2 + \frac{4m^2(m-1)}{h_F^2(\xi_p)} CO(n^{-\frac{1}{2}+\gamma/2}) + 4(m-1)^2O(n^{-2+\gamma}). \end{aligned}$$

Since $|g_1(X_i) - g'_1(X_i)| \leq 1$ for all X_i and

$$|g_1(X_i) - g'_1(X_i)| \xrightarrow{P} 0$$

the constant C converges to zero in probability and therefore

$$\text{Var}(Z_n(t) - Z_n(0)) \xrightarrow{P} 0.$$

Applying the Chebychev inequality we then have $Z_n(t) - Z_n(0) \xrightarrow{P} 0$.

Altogether we have for $t \in \mathbb{R}$ and every $\epsilon > 0$

$$\begin{aligned} \mathbb{P}(\sqrt{n}(\hat{\xi}_p - \xi_p) \leq t, Z_n(0) \geq t + \epsilon) &= \mathbb{P}(Z_n(t) \leq V_n(t), Z_n(0) \geq t + \epsilon) \\ &\leq \mathbb{P}\left(|Z_n(t) - Z_n(0)| \geq \frac{\epsilon}{2}\right) + \mathbb{P}\left(|V_n(t) - t| \geq \frac{\epsilon}{2}\right) \rightarrow 0, \end{aligned}$$

and analogously

$$\mathbb{P}\left(\sqrt{n}(\hat{\xi}_p - \xi_p) \geq t, Z_n(0) \leq t\right) \rightarrow 0.$$

Using Lemma 1 of Ghosh (1971) the proof is completed. \square

Theorem 2.3.

The proof makes use of the Hoeffding decomposition

$$\sqrt{n}(U_n - \theta) = \sqrt{n} \sum_{j=1}^m \binom{m}{j} \frac{1}{\binom{n}{j}} S_{jn}.$$

We show that the linear part $\frac{m}{\sqrt{n}} \sum_{i=1}^n g_1(X_i)$ is asymptotically normal and that the remaining terms converge to 0 in probability.

If $(X_i)_{i \in \mathbb{N}}$ is strong mixing then this also applies to $(g_1(X_i))_{i \in \mathbb{N}}$, because g_1 is measurable (Korolyuk and Borovskikh (1993)), and the mixing coefficients are smaller or equal to the original ones. With these considerations and observing that $(g_1(X_i))_{i \in \mathbb{N}}$ is strong mixing with mixing coefficients $\alpha(l) = O(l^{-\delta})$ for a $\delta > 2$ and moreover $\mathbb{E}(g_1(X_i)) = 0$ and $g_1(X_i)$ is bounded (because h is bounded) we can apply Theorem 1.6 of Ibragimov (1961) getting $\sigma < \infty$ and

$$\frac{m}{\sqrt{n}} \sum_{i=1}^n g_1(X_i) \xrightarrow{D} N(0, m^2 \sigma^2).$$

It remains to show that the remaining terms of the Hoeffding decomposition are of order $o_P(1)$. For this we apply Lemma 4.3 and show $\sum_{l=0}^n l \alpha^{\frac{\rho}{2\rho+1}}(l) = O(n^\gamma)$ for a $\rho \geq 0$.

Using the assumption $\alpha(l) = O(l^{-\delta})$ for a $\delta > \frac{2\rho+1}{\rho}$ we get for a $\gamma < 1$

$$\sum_{l=0}^n l \alpha^{\frac{\rho}{2\rho+1}}(l) \leq \sum_{l=1}^n l^{1-\delta \frac{\rho}{2\rho+1}} = O(n^\gamma).$$

Now it is for all $2 \leq k \leq m$

$$\begin{aligned} & \text{Var} \left(\sqrt{n} \binom{m}{k} \binom{n}{k}^{-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} g_k(X_{i_1}, \dots, X_{i_k}) \right) \\ & \leq \frac{m^{2k} k^{\frac{k}{2}}}{n^{2k-1}} \sum_{1 \leq i_1 < \dots < i_k \leq n} \sum_{1 \leq i_{k+1} < \dots < i_{2k} \leq n} \left| \mathbb{E} (g_k(X_{i_1}, \dots, X_{i_k}) g_k(X_{i_{k+1}}, \dots, X_{i_{2k}})) \right| \\ & \leq \frac{m^{2k} k^{\frac{k}{2}}}{n^{2k-1}} \sum_{i_1, \dots, i_{2k}=1}^n \left| \mathbb{E} (g_k(X_{i_1}, \dots, X_{i_k}) g_k(X_{i_{k+1}}, \dots, X_{i_{2k}})) \right| \\ & = O(n^{2k-2+\gamma-(2k-1)}) = O(n^{-1+\gamma}). \end{aligned}$$

And so

$$\text{Var} \left(\sqrt{n} \binom{m}{k} \binom{n}{k}^{-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} g_k(X_{i_1}, \dots, X_{i_k}) \right) \xrightarrow{n \rightarrow \infty} 0$$

and with the Chebychev inequality we obtain

$$\sqrt{n} \binom{m}{k} \binom{n}{k}^{-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} g_k(X_{i_1}, \dots, X_{i_k}) \xrightarrow{P} 0 \text{ for } n \rightarrow \infty.$$

Using the Theorem of Slutsky we get the result of the theorem. □

Corollary 2.1.

Using the Hoeffding decomposition we obtain

$$\begin{aligned} & \sup_{t \in \mathbb{R}} \left| \sqrt{n} (H_n(t) - H_F(t)) \right| \\ & = \sup_{t \in \mathbb{R}} \left| \sqrt{n} \left(H_F(t) + \sum_{j=1}^m \binom{m}{j} \frac{1}{\binom{n}{j}} S_{jn,t} - H_F(t) \right) \right| \\ & = \sup_{t \in \mathbb{R}} \left| \frac{m}{\sqrt{n}} \sum_{i=1}^n g_1(X_i, t) + \sqrt{n} \frac{\binom{m}{2}}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} g_2(X_i, X_j, t) \right. \\ & \quad \left. + \dots + \sqrt{n} \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} h_m(X_{i_1}, \dots, X_{i_m}, t) \right| \\ & \leq \sup_{t \in \mathbb{R}} \left| \frac{m}{\sqrt{n}} \sum_{i=1}^n g_1(X_i, t) \right| + \sup_{t \in \mathbb{R}} \left| \sqrt{n} \frac{\binom{m}{2}}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} g_2(X_i, X_j, t) \right| \\ & \quad + \dots + \sup_{t \in \mathbb{R}} \left| \sqrt{n} \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} h_m(X_{i_1}, \dots, X_{i_m}, t) \right|. \end{aligned}$$

For the first term of the sum we get, using Theorem 2.4 and the Continuous Mapping theorem,

$$\sup_{t \in \mathbb{R}} \left| \frac{m}{\sqrt{n}} \sum_{i=1}^n g_1(X_i, t) \right| \rightarrow \|W\|_\infty.$$

Since W is a continuous Gaussian process we have $\|W\|_\infty = O_p(1)$.

For the remaining results we want to apply Lemma 4.4. Therefore the kernel of the U -process $g(x_1, \dots, x_m, t) = 1_{[h(x_1, \dots, x_m) \leq t]}$ has to satisfy the extended uniform variation condition. This can be shown using the Lipschitz-continuity of h :

$$\begin{aligned} & \sup_{\|(x_1, \dots, x_m) - (X'_1, \dots, X'_m)\| \leq \epsilon} \left| 1_{[h(X'_1, \dots, X'_m) \leq t]} - 1_{[h(x_1, \dots, x_m) \leq t]} \right| \\ &= \begin{cases} 1 & , \text{ if } h(X'_1, \dots, X'_m) \in (t - L\epsilon, t + L\epsilon) \\ 0 & , \text{ else} \end{cases} \end{aligned}$$

and so

$$\begin{aligned} & \mathbb{E} \left(\sup_{\|(x_1, \dots, x_m) - (X'_1, \dots, X'_m)\| \leq \epsilon} \left| 1_{[h(X'_1, \dots, X'_m) \leq t]} - 1_{[h(x_1, \dots, x_m) \leq t]} \right| \right) \\ & \leq \sup_{t \in \mathbb{R}} \left| \mathbb{E} (1_{[h(X'_1, \dots, X'_m) \in (t - L\epsilon, t + L\epsilon)]}) \right| \\ & \leq \sup_{t \in \mathbb{R}} \left| \int_{t - L\epsilon}^{t + L\epsilon} h_F(x) dx \right| \leq 2L\epsilon (\sup_{x \in \mathbb{R}} h_F(x)) \leq L'\epsilon, \end{aligned}$$

since h_F is bounded.

Using the arguments above we can also show that g satisfies the extended uniform variation condition. For arbitrary $2 \leq k \leq m$ and $i_1 < i_2 < \dots < i_m$

$$\begin{aligned} & \mathbb{E} \left(\sup_{|x_1 - Y_{i_1}| \leq \epsilon} \left| 1_{[h(Y_{i_1}, X_{i_2}, \dots, X_{i_k}, Y_{i_{k+1}}, \dots, Y_{i_m}) \leq t]} \right. \right. \\ & \qquad \qquad \qquad \left. \left. - 1_{[h(x_1, X_{i_2}, \dots, X_{i_k}, Y_{i_{k+1}}, \dots, Y_{i_m}) \leq t]} \right| \right) \\ & \leq \sup_{t \in \mathbb{R}} \left| \int_{t - L\epsilon}^{t + L\epsilon} h_{F; X_{i_1}, \dots, X_{i_k}}(x) dx \right| \leq L\epsilon. \end{aligned}$$

Applying Lemma 4.4 we get for $2 \leq k \leq n$

$$\begin{aligned} & \sup_{t \in \mathbb{R}} \left| \sqrt{n} \frac{\binom{m}{k}}{\binom{n}{k}} \sum_{1 \leq i_1, \dots, i_k \leq n} g_k(X_{i_1}, \dots, X_{i_k}, t) \right| \\ & \leq \sqrt{nn}^{-k} o_p(n^{k - \frac{1}{2} - \frac{\delta - 2}{8\delta}}) = o_p(n^{-\frac{\delta - 2}{8\delta}}). \end{aligned}$$

With Slutsky's Theorem the proof is completed. □

Appendix

In this section the QQ-plots for the simulation of the asymptotic normality of the GM -estimator for the tail index of a Pareto distribution are given. Since the calculations for $n = 1000$ and $m = 4$ are very time intensive we use only 250 runs. These cases are marked with (*).

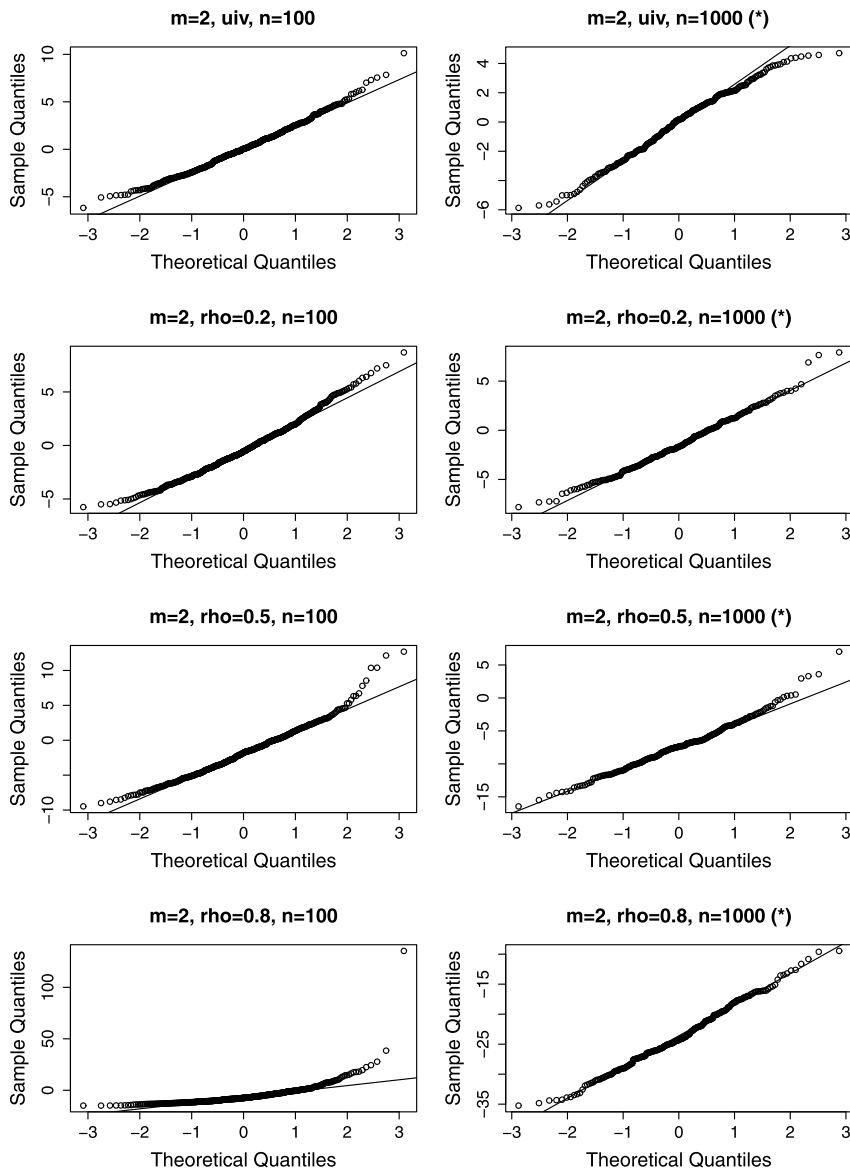


FIG 2. QQ-Plot of the error term $\sqrt{n}(\hat{\alpha}_{GM} - \alpha)$ for a kernel dimension $m = 2$ and different dependencies (i.i.d. and $AR(1)$ with autocorrelation ρ)

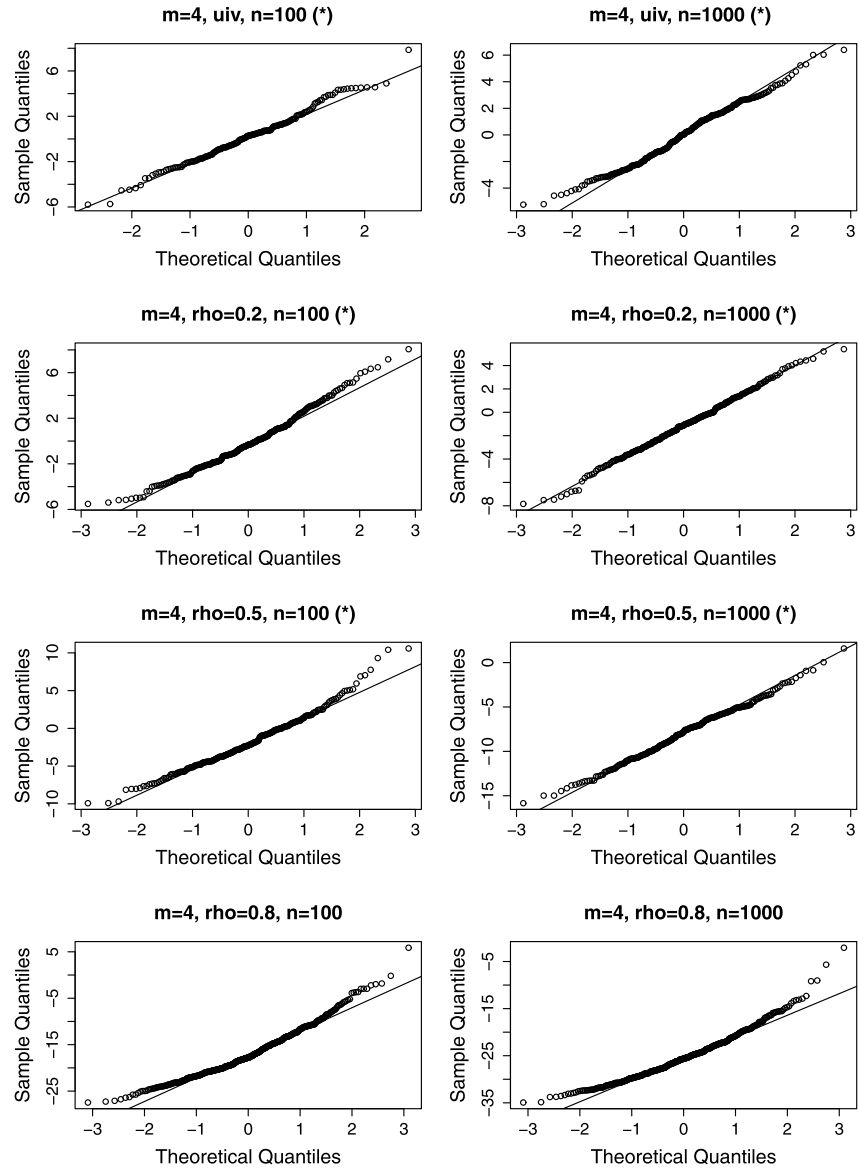


FIG 3. QQ-Plot of the error term $\sqrt{n}(\hat{\alpha}_{GM} - \alpha)$ for a kernel dimension $m = 4$ and different dependencies (i.i.d. and AR(1) with autocorrelation ρ)

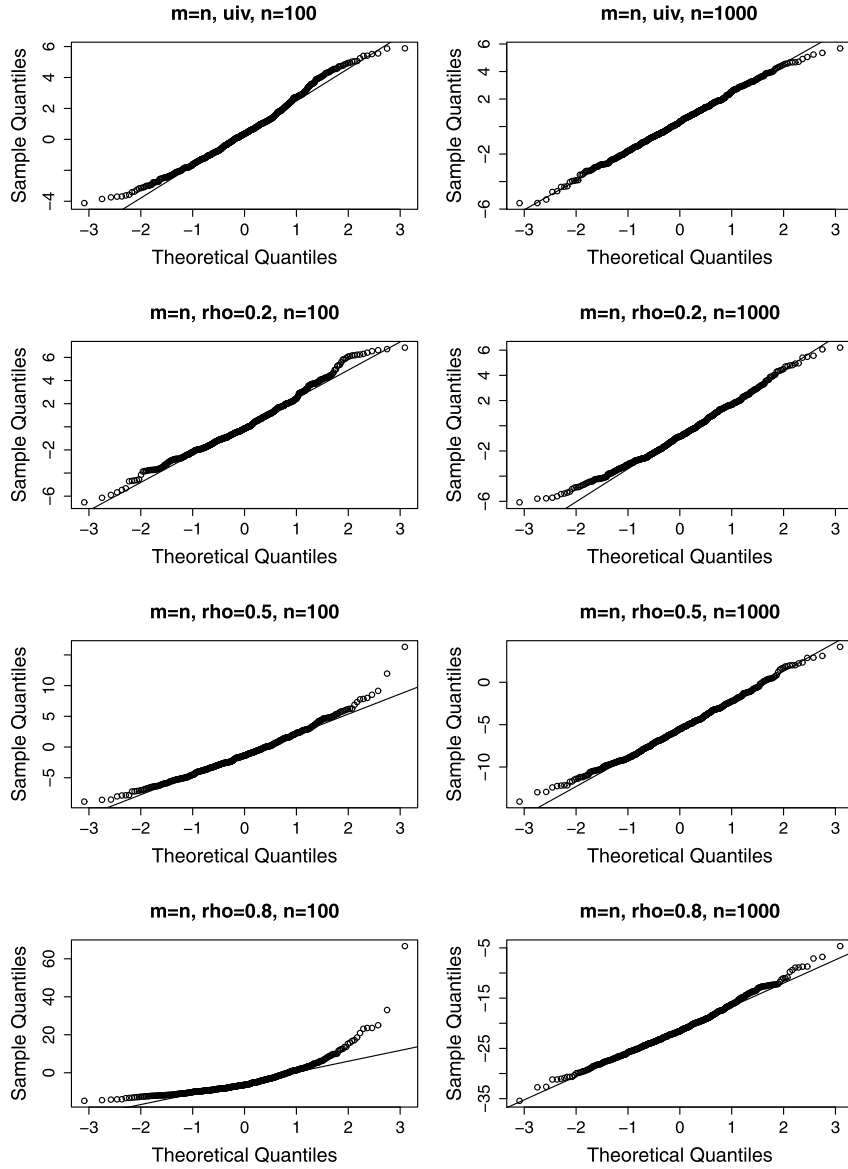


FIG 4. QQ-Plot of the error term $\sqrt{n}(\hat{\alpha}_{GM}-\alpha)$ for a kernel dimension $m = n$ (ML-estimator) and different dependencies (i.i.d. and AR(1) with autocorrelation ρ)

Acknowledgements

The financial support of the Deutsche Forschungsgemeinschaft (SFB 823, ‘Statistical modelling of nonlinear dynamic processes’) is gratefully acknowledged. We want to thank the anonymous referee and the associate editor for the helpful comments and ideas which helped us to improve our paper.

References

- Babu GJ, Singh K (1978) On Deviations Between Empirical and Quantile Processes for Mixing Random Variables. *Journal of Multivariate Analysis* 8:532–549 [MR0520961](#)
- Bahadur RR (1966) A Note on Quantiles in Large Samples. *Annals of Mathematical Statistics* 35
- Beran, J, Feng, Y, Ghosh, S and Kulik, R (2013) *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Springer Science and Business Media [MR3075595](#)
- Borovkova S, Burton R, Dehling H (2001) Limit Theorems for Functionals of Mixing Processes with Applications to U -Statistics and Dimension Estimation. *Transactions of the American Mathematical Society* 353:4261–4318 [MR1851171](#)
- Bradley R (2007) *Introduction to Strong Mixing Conditions*. Kendrick Press 1st edition
- Brazauskas V, Serfling R (2000a) Robust and Efficient Estimation of the Tail Index of a Single-Parameter Pareto Distribution. *North American Actuarial Journal* 4:12–27 [MR1989836](#)
- Brazauskas V, Serfling R (2000b) Robust Estimation of Tail Parameters for Two-Parameter Pareto and Exponential Models via Generalized Quantile Statistics. *Extremes* 3.3:231–249 [MR1856199](#)
- Choudhury J, Serfling RJ (1988) Generalized Order Statistics, Bahadur Representations, and Sequential Nonparametric Fixed-Width Confidence Intervals. *Journal of Statistical Planning and Inference* 19:269–282 [MR0955393](#)
- Dehling H, Philipp W (2002) Empirical Process Techniques for Dependent Data. In: *Empirical Process Techniques for Dependent Data*, pp 3–115 [MR1958777](#)
- Dehling H, Wendler M (2010) Central Limit Theorem and the Bootstrap for U -Statistics of Strongly Mixing Data. *Journal of Multivariate Analysis* 101:126–137 [MR2557623](#)
- Dehling H, Denker M, Philipp W (1987) The Almost Sure Invariance Principle for the Empirical Process of U -Statistics Structure. *Annales de l’Institut Henri Poincaré (B): Probability and Statistics* 23:349–382 [MR0891707](#)
- Denker M, Keller G (1986) Rigorous Statistical Procedures for Data from Dynamical Systems. *Journal of Statistical Physics* 44:67–93 [MR0854400](#)
- Doukhan G, Lang G (2009) Evaluation for Moments of a Ratio with Applications to Regression Estimation. *Bernoulli* 15:1259–1286 [MR2597592](#)
- Doukhan P, Klesov O, Lang G (2010) Rates of Convergence in Some SLLN

- Under Weak Dependence Conditions. *Acta Scientiarum Mathematicarum* 76:683–695 [MR2789692](#)
- Ghosh JK (1971) A new Proof of the Bahadur Representation of Quantiles and an Application. *Annals of Mathematical Statistics* 42:1957–1961 [MR0297071](#)
- Hoeffding W (1948) A Class of Statistics with Asymptotically Normal Distribution. *Annals of Mathematical Statistics* 19:293–325 [MR0026294](#)
- Ibragimov I (1961) Some Limit Theorems for Stationary Processes. *Theory of Probability and its Applications* 7:349–382
- Korolyuk V, Borovskikh Y (1993) *The Theory of U -Statistics*. Springer
- Lahiri S N, Sun S (2009) A Berry-Esseen Theorem for Sample Quantiles under Weak Dependence. *The Annals of Applied Probability* 19.1:108–126 [MR2498673](#)
- Li Y, Cai W, and Campbell E P (2005) Statistical Modeling of Extreme Rainfall in Southwest Western Australia. *Journal of Climate* 18: 852–863.
- Moricz F (1983) A General Moment Inequality for the Maximum of the Rectangular Partial Sums of Multiple Series. *Acta Mathematica Hungaria* 41:337–346 [MR0703745](#)
- Politis DN, Romano JP (1994) Large Sample Confidence Regions based on Subsamples under Minimal Assumptions. *The Annals of Statistics* 22.4:2031–2050 [MR1329181](#)
- Pollard D (1984) *Convergence of Stochastic Processes*. Springer [MR0762984](#)
- Sen P K (1972) On the Bahadur Representation of Sample Quantiles for Sequences of ϕ -mixing Random Variables. *Journal of Multivariate Analysis* 2:77–95 [MR0303581](#)
- Serfling R (1980) *Approximation Theorems of Mathematical Statistics*. John Wiley and Sons [MR0595165](#)
- Serfling R (1984) Generalized L -, M -, and R -Statistics. *The Annals of Statistics* 12.1:76–86 [MR0733500](#)
- Serfling R (2002) Efficient and Robust Fitting of Lognormal Distributions. *North American Actuarial Journal* 6:95–116 [MR1987777](#)
- Sun S (2006) The Bahadur Representation of Sample Quantiles under Weak Dependence. *Statistic and Probability Letters* 76:1238–1244 [MR2269350](#)
- Sun S, Lahiri S N (2006) Bootstrapping the Sample Quantile of a Weakly Dependent Sequence. *Sankhyā: The Indian Journal of Statistics* 68.1: 130–166 [MR2301568](#)
- Tukey, J W (1960), A Survey of Sampling from Contaminated Distributions. *Contributions to Probability And Statistics* 2:448–485 [MR0120720](#)
- Vandewalle, B, Beirlant, J, Christmann, A and Hubert, M (2007) A Robust Estimator for the Tail Index of Pareto-type Distributions. *Computational Statistics and Data Analysis* 51:6252–6268 [MR2408592](#)
- Vogel D, Wendler M (2015) Studentized Sequential U -Quantiles under Dependence with Applications to Change-Point Analysis. [arXiv:1503.04161](#)
- Wendler M (2011a) Bahadur Representation for U -Quantiles of Dependent Data. *Journal of Multivariate Analysis* 102:1064–1079 [MR2793876](#)
- Wendler M (2011b) Empirical U -Quantiles of Dependent Data. PhD thesis, Ruhr-Universität Bochum

- Wilde S, Grimshaw S (2013) Efficient Computation of Generalized Median Estimators. *Computational Statistics* 28:307–317 [MR3044785](#)
- Yoshihara K (1976) Limiting Behavior of U -Statistics for Stationary, Absolutely Regular Processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 35:237–252 [MR0418179](#)
- Yoshihara K (1995) The Bahadur Representation of Sample Quantiles for Sequences of strongly mixing Random Variables. *Statistic and Probability Letters* 24:299–304 [MR1353887](#)