



HAL
open science

Extreme quantile estimation for β -mixing time series and applications

Valérie Chavez-Demoulin, Armelle Guillou

► **To cite this version:**

Valérie Chavez-Demoulin, Armelle Guillou. Extreme quantile estimation for β -mixing time series and applications. Insurance: Mathematics and Economics, 2018, 83, pp.59-74. 10.1016/j.insmatheco.2018.09.004 . hal-02392748

HAL Id: hal-02392748

<https://hal.science/hal-02392748>

Submitted on 4 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Extreme quantile estimation for β -mixing time series and applications

Valérie CHAVEZ-DEMOULIN, Armelle GUILLOU *

Abstract

In this paper, we discuss the application of extreme value theory in the context of stationary β -mixing sequences that belong to the Fréchet domain of attraction. In particular, we propose a methodology to construct bias-corrected tail estimators. Our approach is based on the combination of two estimators for the extreme value index to cancel the bias. The resulting estimator is used to estimate an extreme quantile. In a simulation study, we outline the performance of our proposals that we compare to alternative estimators recently introduced in the literature. Also, we compute the asymptotic variance in specific examples when possible. Our methodology is applied to two datasets on finance and environment.

JELcode: C130

Keywords: Asymptotic normality; β -mixing; Extreme value index; GARCH models; High quantile; Market index; Return level; Value-at-Risk; Wind speed data.

*Valérie Chavez-Demoulin is Professor, Faculty of Business and Economics (HEC), Université de Lausanne, Switzerland (valerie.chavez@unil.ch). Armelle Guillou is Professor, Université de Strasbourg & CNRS, IRMA UMR 7501, France (armelle.guillou@math.unistra.fr). The research was partially funded by a research grant (VKR023480) from VILLUM FONDEN. We thank Laurens de Haan, Cécile Mercadier and Chen Zhou for several discussions about the proof of Theorem 4.2 in de Haan et al. (2016) which has allowed to detect an error in their paper. We thank them for providing us the authorization to mention the corrected version of their estimator that we proposed in (11). Note that only this paper contains this corrected version of their estimator since till this date no erratum of their paper has been published.

1 INTRODUCTION

Quantitative Risk Management (QRM) has become an inevitable field aimed at building models to understand the risks of financial portfolios and environmental hazards. For the most complete treatment of the theoretical concepts and modeling tools of QRM, we refer to the book by McNeil et al. (2015). Building such models is now a crucial task across the banking and insurance industries under the regulatory obligations of the Basel Committee on Banking Supervision and Solvency 2. In financial risk management, research related to regulatory risk measures such as the Value-at-Risk (VaR) has received lot of attention over the last decades. In environmental risk management, and due to the increasing frequency of extreme events (SwissRe 2014, Embrechts et al. 2016) and their disastrous societal impact, estimating risk measures such as the return level is of vital importance. Both these risk measures (VaR and return level) rely on high quantile estimation in the tail region of the observations distribution. In this context of tail modeling, extreme value theory (EVT) offers strong and adequate statistical tools. Classical EVT models are based on the independent and identically distributed (i.i.d.) assumption. This assumption is however very often violated in practice. Financial time series, for instance, show volatility clustering and environmental data typically exhibit serial dependence. Many EVT-related papers for time series address the modeling of such features (see, for instance, McNeil & Frey 2000, Chavez-Demoulin et al. 2014). In this paper, we introduce a new asymptotically unbiased high quantile estimator for stationary time series such as the very commonly used autoregressive (AR), the moving average (MA) and the generalized autoregressive conditional heteroskedasticity (GARCH) models. More precisely, we propose a new estimator of high quantiles for β -mixing stationary time series with heavy-tailed distribution. The estimator is based on an asymptotically unbiased estimator of the extreme value index. The advantage of our estimator is twofold: first, it directly handles the serial feature of such β -mixing time series contrary to other methods that need a pre-filtering of the heteroskedasticity before applying the standard estimator. As an example, we refer to the two step-method of McNeil & Frey (2000). Second, it improves the alternative bias correction procedure proposed by de Haan et al. (2016) for β -mixing series.

Throughout the paper, we assume that (X_1, X_2, \dots) is a β -mixing time series, that is, a series such that

$$\beta(m) := \sup_{p \geq 1} \mathbb{E} \left\{ \sup_{C \in \mathcal{B}_{p+m+1}^\infty} |\mathbb{P}(C | \mathcal{B}_1^p) - \mathbb{P}(C)| \right\} \longrightarrow 0,$$

as $m \rightarrow \infty$, where \mathcal{B}_i^j denotes the σ -algebra generated by X_i, \dots, X_j . Loosely speaking, $\beta(m)$ measures the total variation distance between the unconditional distribution of the future of the

time series and the conditional distribution of the future given the past of the series when both are separated by m time points. Let F be the common marginal distribution function of $X_i, i \in \mathbb{N}$, which is assumed to belong to the Fréchet domain of attraction, that is, the tail quantile function $U := (1/(1 - F))^\leftarrow$ where $^\leftarrow$ denotes the left continuous inverse function, satisfies

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma, \quad \forall x > 0. \quad (1)$$

The estimation of the extreme value index γ has been extensively studied in the case of i.i.d. random variables, but only few papers consider this topic in case of time series with serial dependence features. We can mention, among others, Hsing (1991), Drees (2000) and Drees (2003) and very recently de Haan et al. (2016). As in the i.i.d. context, the simplest estimator for $\gamma > 0$ is the Hill estimator (Hill 1975) defined as

$$\hat{\gamma}_k^H := \frac{1}{k} \sum_{i=1}^k \log X_{n-i+1,n} - \log X_{n-k,n},$$

where $X_{1,n} \leq \dots \leq X_{n,n}$ denote the order statistics and k is an intermediate sequence, that is, a sequence such that $k \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$.

To prove the asymptotic normality of a tail parameter such as the Hill estimator, we need a second order condition which specifies the rate of convergence for the left-hand side in (1) to its limit. This condition can be formulated in different ways, below we state it in terms of the logarithm since it is this formulation that we will use later.

Second order condition (C_{SO}). *Suppose that there exists a positive or negative function A with $\lim_{t \rightarrow \infty} A(t) = 0$ and a real number $\rho < 0$ such that*

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

The rate of convergence for the function A to 0 is crucial if we want to exhibit the bias term of the estimator of a tail parameter. Under the assumption that the intermediate k -sequence is such that $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$, and assuming the following regularity conditions on the β -mixing coefficients :

Regularity conditions (C_R). *There exist $\varepsilon > 0$, a function r and a sequence ℓ_n such that, as $n \rightarrow \infty$,*

$$(a) \quad \frac{\beta(\ell_n)}{\ell_n} n + \ell_n \frac{\log^2 k}{\sqrt{k}} \longrightarrow 0;$$

$$(b) \quad \frac{n}{\ell_n k} \text{Cov} \left(\sum_{i=1}^{\ell_n} \mathbb{1}_{\{X_i > F^\leftarrow(1-kx/n)\}}, \sum_{i=1}^{\ell_n} \mathbb{1}_{\{X_i > F^\leftarrow(1-ky/n)\}} \right) \longrightarrow r(x, y), \quad \forall 0 \leq x, y \leq 1 + \varepsilon;$$

(c) For some constant C :

$$\frac{n}{\ell_n k} \mathbb{E} \left[\left(\sum_{i=1}^{\ell_n} \mathbb{1}_{\{F^{\leftarrow}(1-ky/n) < X_i \leq F^{\leftarrow}(1-kx/n)\}} \right)^4 \right] \leq C(y-x), \quad \forall 0 \leq x < y \leq 1 + \varepsilon \text{ and } n \in \mathbb{N},$$

Drees (2000) has established the asymptotic normality of $\hat{\gamma}_k^H$

$$\sqrt{k}(\hat{\gamma}_k^H - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}, \sigma^2\right), \quad (2)$$

where σ^2 depends on the covariance structure r , but has a simple expression in the i.i.d. context, where it is equal to γ^2 . In practice, the bias term of $\hat{\gamma}_k^H$ can be important depending on whether ρ is close to zero or not, since under the second order condition (C_{SO}), the function $|A|$ is regularly varying at infinity with index ρ . This explains all the literature spread on bias correction in the i.i.d. context, see, e.g., Feuerverger & Hall (1999), Beirlant et al. (1999) and Gomes et al. (2008), among others. On the contrary, in case of stationary β -mixing time series only the very recent paper by de Haan et al. (2016) deals with this problem and proposes a bias-corrected estimator for γ . Their method consists first in estimating the bias term of $\hat{\gamma}_k^H$ and second in subtracting it from $\hat{\gamma}_k^H$. A similar approach is also used in their paper to estimate a high quantile $x_p = U(1/p)$ with $p \rightarrow 0$. The procedure we propose in this paper is an alternative approach to construct bias-corrected tail estimators. First, we introduce a class of estimators for γ which can be viewed as statistical tail functionals, $T(Q_n)$, where Q_n is the tail quantile function defined as $Q_n(t) := X_{n-[kt],n}$, $0 < t < n/k$, and T is a suitable functional. Then, we combine two estimators for γ of this class to cancel the asymptotic bias term. The resulting unbiased estimator for γ can then be used to construct an asymptotically unbiased estimator of a high quantile.

The paper is organized as follows: our approach is described in details in Section 2. Section 3 presents, using some examples, the finite sample performance of our extreme value index and high quantile estimators based on simulation studies. Two real data applications illustrate the use of our estimator in Section 4: one in the financial context of market risk data and the other in the environmental situation of hourly wind speed data. We conclude in Section 5. All the related theoretical proofs are detailed in the appendix.

2 DESCRIPTION OF OUR METHODOLOGY

Goegebeur & Guillou (2013) introduced a class of weighted function estimators for the tail dependence coefficient η in the bivariate extreme value framework. Combining two of their estimators,

they are able to construct an asymptotically unbiased estimator for η . In this paper, we adapt this methodology in the case of β -mixing sequences to estimate a tail parameter such as the extreme value index or an extreme quantile.

2.1 Estimation of the extreme value index

For any measurable function $z : [0, 1] \rightarrow \mathbb{R}$, we consider the functional

$$T_K(z) = \begin{cases} \int_0^1 \log \frac{z(t)}{z(1)} d(tK(t)) & \text{if the right-hand side is defined and finite,} \\ 0 & \text{otherwise,} \end{cases}$$

where K is a function with support on $(0, 1)$. This leads to the following class of estimators for γ :

$$\hat{\gamma}_k(K) = T_K(Q_n) = \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} d(tK(t)).$$

Some assumptions on K are required if we want to derive the asymptotic normality of our class of estimators. They can be formulated as:

Assumption (C_K). *Let K be a function such that $\int_0^1 K(t)dt = 1$. Suppose that K is continuously differentiable on $(0, 1)$ and that there exist $M > 0$ and $\tau \in [0, 1/2)$ such that $|K(t)| \leq M t^{-\tau}$.*

These conditions are not restrictive but are satisfied by the usual weight functions used in the literature, including the power kernel $K(u) = (1 + \nu)u^\nu$, $\nu \geq 0$, and the log-weight function $K(u) = (-\log u)^\nu / \Gamma(1 + \nu)$, $\nu \geq 0$. In particular, we note that the classical Hill estimator $\hat{\gamma}_k^H$ can be viewed as a particular case of our power kernel-type estimator corresponding to $\nu = 0$: $\hat{\gamma}_k^H = \hat{\gamma}_k(K)$ with $K(u) = 1$.

The aim of Theorem 1 is to provide the asymptotic normality of our class of estimators with the explicit bias term.

Theorem 1 *Let (X_1, X_2, \dots) be a stationary β -mixing time series with a continuous common marginal distribution function F and assume (C_{S0}) , (C_R) and (C_K) . Suppose that k is an intermediate sequence such that $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$. We have*

$$\sqrt{k} \{ \hat{\gamma}_k(K) - \gamma \} \xrightarrow{d} \lambda \int_0^1 t^{-\rho} K(t) dt + \gamma \int_0^1 [t^{-1}W(t) - W(1)] d(tK(t)), \quad (3)$$

where $(W(t))_{t \in [0,1]}$ is a centered Gaussian process with covariance function r defined in (C_R) .

In particular

$$\sqrt{k} \{ \hat{\gamma}_k(K) - \gamma \} \xrightarrow{d} \mathcal{N}(\lambda \mathcal{A}\mathcal{B}(K), \mathcal{A}\mathcal{V}(K)),$$

where

$$\begin{aligned}\mathcal{AB}(K) &:= \int_0^1 t^{-\rho} K(t) dt \\ \text{and } \mathcal{AV}(K) &:= \gamma^2 \int_0^1 \int_0^1 \left[\frac{r(t,s)}{ts} - \frac{r(t,1)}{t} - \frac{r(1,s)}{s} + r(1,1) \right] d(tK(t))d(sK(s)).\end{aligned}$$

This result is similar to Theorem 1 in Goegebeur & Guillou (2013) in the i.i.d. context, although in the latter the centered Gaussian process $(W(t))_{t \in [0,1]}$ is a standard Brownian motion.

Some remarks.

- If we assume that the X_i s are i.i.d. and that $K(t) = 1$ for $t \in (0, 1)$, then Theorem 1 gives the asymptotic normality of the Hill estimator (2) with $\sigma^2 = \gamma^2$.
- If we assume that the X_i s are i.i.d. but nothing on the function K (except that it satisfies condition (C_K)), then the asymptotic variance given in Theorem 1 can be reduced to

$$\mathcal{AV}(K) = \gamma^2 \left\{ \int_0^1 \int_0^1 \frac{\min(s,t)}{ts} d(tK(t))d(sK(s)) - K^2(1) \right\}.$$

Now, our aim is to propose an asymptotically unbiased estimator for γ . For this aim, we propose to use two functions K_1 and K_2 satisfying (C_K) and to consider a mixture of them in the form $K_\Delta(t) = \Delta K_1(t) + (1 - \Delta)K_2(t)$ for $\Delta \in \mathbb{R}$. Clearly K_Δ also satisfies condition (C_K) and hence by Theorem 1, the asymptotic bias of this new estimator $\hat{\gamma}_k(K_\Delta)$ is given by

$$\frac{\lambda}{\sqrt{k}} \mathcal{AB}(K_\Delta) = \frac{\lambda}{\sqrt{k}} \int_0^1 t^{-\rho} K_\Delta(t) dt = \frac{\lambda}{\sqrt{k}} \{ \Delta \mathcal{AB}(K_1) + (1 - \Delta) \mathcal{AB}(K_2) \}.$$

Equating the right-hand side of the above equation to zero leads to the value of Δ eliminating the asymptotic bias

$$\Delta^* = \frac{\mathcal{AB}(K_2)}{\mathcal{AB}(K_2) - \mathcal{AB}(K_1)} \quad \text{provided } \mathcal{AB}(K_1) \neq \mathcal{AB}(K_2). \quad (4)$$

This result is formalized in the next corollary where $\hat{\gamma}_k(K_{\Delta^*})$ is shown to be asymptotically unbiased in the sense that the mean of its limiting distribution is zero, whatever the value of λ .

Corollary 1. *Under the assumptions of Theorem 1 and assuming that K_1 and K_2 satisfy condition (C_K) with $\mathcal{AB}(K_1) \neq \mathcal{AB}(K_2)$, we have*

$$\sqrt{k} \{ \hat{\gamma}_k(K_{\Delta^*}) - \gamma \} \xrightarrow{d} \mathcal{N}(0, \mathcal{AV}(K_{\Delta^*})).$$

An open problem is to determine whether among this class of unbiased estimators we can find the asymptotically unbiased estimator with minimum variance. This question is solved in the i.i.d. framework under a slightly stronger condition than (C_K) (see Theorem 2 and Corollary 4 in Goegebeur & Guillou 2013) where the “optimal” function is given by

$$K_{\Delta_{opt}^*}(t) = \left(\frac{1-\rho}{\rho}\right)^2 - \frac{(1-\rho)(1-2\rho)}{\rho^2}t^{-\rho}, \quad t \in (0, 1). \quad (5)$$

Note that this function can be viewed as a mixture between $K_1(t) := 1$ and $K_{2,\rho}(t) := (1-\rho)t^{-\rho}$ with $t \in (0, 1)$ and Δ^* as in (4). In that case, the minimal variance is given by

$$\mathcal{AV}(K_{\Delta_{opt}^*}) = \gamma^2 \left(\frac{1-\rho}{\rho}\right)^2. \quad (6)$$

In our framework, the covariance structure r is unknown and thus we cannot show that this interesting property is preserved. However, we recommend to use this “optimal” function also in case of β -mixing sequences since it performs very well in practice as illustrated in Section 3.4. From a practical point of view, this cannot be done directly since ρ is unknown. To solve this issue, two natural options can be proposed, either to replace ρ by a canonical choice, or by an external estimator.

The aim of the next corollary is to give the asymptotic normality of our class of estimators for γ in case ρ is replaced by some fixed value $\tilde{\rho}$.

Corollary 2. *Let (X_1, X_2, \dots) be a stationary β -mixing time series with a continuous common marginal distribution function F and assume (C_{S0}) and (C_R) . Suppose that k is an intermediate sequence such that $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$. We have*

$$\sqrt{k}\{\hat{\gamma}_k(K_{\tilde{\Delta}_{opt}^*}) - \gamma\} \xrightarrow{d} \mathcal{N}\left(\lambda \frac{(1-\tilde{\rho})(\tilde{\rho}-\rho)}{\tilde{\rho}(1-\rho)(1-\tilde{\rho}-\rho)}, \mathcal{AV}(K_{\tilde{\Delta}_{opt}^*})\right),$$

where $K_{\tilde{\Delta}_{opt}^*}$ is defined as $K_{\Delta_{opt}^*}$ in (5) with ρ replaced by $\tilde{\rho}$.

Although one clearly loses the bias correction, the extreme value index estimators are not very sensitive to such a misspecification and thus our estimator $\hat{\gamma}_k(K_{\tilde{\Delta}_{opt}^*})$ can still outperform the estimators that are not corrected for bias. Note also that, as expected, if $\tilde{\rho} = \rho$, we recover Corollary 1. However, to keep the asymptotically unbiased property, we can also replace ρ by an external estimator $\hat{\rho}_{k_\rho}$, consistent in probability, which depends on an intermediate sequence k_ρ . This leads to the following general result.

Theorem 2 *Let (X_1, X_2, \dots) be a stationary β -mixing time series with a continuous common marginal distribution function F and assume (C_{S0}) and (C_R) . Let $\hat{\rho}_{k_\rho}$ be an external estimator for ρ , consistent in probability, which depends on an intermediate sequence k_ρ . If k is an intermediate sequence*

such that $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$, then we have

$$\sqrt{k}\{\widehat{\gamma}_k(K_{\widehat{\Delta}_{opt}^*}) - \gamma\} \xrightarrow{d} \mathcal{N}\left(0, \mathcal{AV}(K_{\Delta_{opt}^*})\right),$$

where $K_{\widehat{\Delta}_{opt}^*}$ is defined as $K_{\Delta_{opt}^*}$ in (5) with ρ replaced by $\widehat{\rho}_{k_\rho}$.

Note that this theorem cannot be viewed as a consequence of one of the two first corollaries because $K_{\widehat{\Delta}_{opt}^*}$ depends on $\widehat{\rho}_{k_\rho}$ and thus on n which means that Theorem 1 cannot be used directly. Moreover, it cannot be written as a mixture of the form $\Delta^*K_1 + (1 - \Delta^*)K_{2, \widehat{\rho}_{k_\rho}}$ with Δ^* defined by (4) since this expression would depend on both $\widehat{\rho}_{k_\rho}$ and ρ , which is unknown. This implies that this theorem cannot be viewed as the counterpart of Proposition 2 in Goegebeur & Guillou (2013) where the limiting distribution of the normalized bias-corrected estimator of the tail dependence coefficient is established, but only in cases where the two kernels are assumed to be independent on ρ . Such a result can also be obtained in our framework, but we omit it since we recommend here to use our ‘‘optimal’’ function $K_{\widehat{\Delta}_{opt}^*}$. Note that Gardes & Girard (2008) mention a result similar to our Theorem 2 in their framework as an open problem.

A possible choice for $\widehat{\rho}_{k_\rho}$ is that proposed by Gomes et al. (2002), and also used in de Haan et al. (2016):

$$\widehat{\rho}_k := \frac{-4 + 6S_k^{(2)} + \sqrt{3S_k^{(2)} - 2}}{4S_k^{(2)} - 3} \quad \text{provided } S_k^{(2)} \in \left(\frac{2}{3}, \frac{3}{4}\right), \quad (7)$$

where

$$S_k^{(2)} := \frac{3}{4} \frac{\left[M_k^{(4)} - 24\left(M_k^{(1)}\right)^4\right] \left[M_k^{(2)} - 2\left(M_k^{(1)}\right)^2\right]}{\left[M_k^{(3)} - 6\left(M_k^{(1)}\right)^3\right]^2}$$

with

$$M_k^{(\alpha)} := \frac{1}{k} \sum_{i=1}^k (\log X_{n-i+1, n} - \log X_{n-k, n})^\alpha, \quad \alpha \in \mathbb{N}.$$

In that case, we have the following corollary.

Corollary 3. *Let (X_1, X_2, \dots) be a stationary β -mixing time series with a continuous common marginal distribution function F and assume (C_{S_0}) and (C_R) . Let $\widehat{\rho}_{k_\rho}$ be the external estimator for ρ defined in (7) where the intermediate sequence k_ρ satisfies $\sqrt{k_\rho}A\left(\frac{n}{k_\rho}\right) \rightarrow \infty$. If k is another intermediate sequence such that $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$, then we have*

$$\sqrt{k}\{\widehat{\gamma}_k(K_{\widehat{\Delta}_{opt}^*}) - \gamma\} \xrightarrow{d} \mathcal{N}\left(0, \mathcal{AV}(K_{\Delta_{opt}^*})\right).$$

Note that Theorem 4.1 in de Haan et al. (2016) states a similar result for their asymptotically unbiased extreme value index defined as

$$\hat{\gamma}_{k,k_\rho} = \hat{\gamma}_k^H - \frac{\left[M_k^{(2)} - 2(\hat{\gamma}_k^H)^2 \right] [1 - \hat{\rho}_{k_\rho}]}{2\hat{\gamma}_k^H \hat{\rho}_{k_\rho}}. \quad (8)$$

Compared to this theorem, the assumptions of our Corollary 3 are less constraining, in particular we do not need a third order condition and only $\sqrt{k_\rho} A\left(\frac{n}{k_\rho}\right) \rightarrow \infty$ is required on the intermediate sequence k_ρ . This is due to the fact that we need only the consistency in probability for the ρ -estimator, not its asymptotic normality. However, our rate of convergence \sqrt{k} is smaller than that obtained by de Haan et al. (2016) since they assume $\sqrt{k} A\left(\frac{n}{k}\right) \rightarrow \infty$ in place of our condition $\sqrt{k} A\left(\frac{n}{k}\right) \rightarrow \lambda$. This implies that, although their rate of convergence also has the form \sqrt{k} , their intermediate sequence is larger than ours, taking into account that the function $|A|$ is regularly varying at infinity with index ρ . Despite our slower rate, our estimator is at least as good as the one proposed by de Haan et al. (2016) in finite samples situations as illustrated in Section 3.4.

2.2 Estimation of an extreme quantile

The estimation of an extreme value index is in general only an intermediate goal. In practice, we are much more interested in the estimation of an extreme quantile

$$x_p = U(1/p), \quad (9)$$

$p \rightarrow 0$. As mentioned in Section 1, the VaR and the return level are extreme quantiles consisting of standard risk measures of finance and environment, respectively.

In this section, we illustrate the applicability of our methodology in the case of the estimation of an extreme quantile x_p . To understand heuristically the construction of our estimator, as introduced in Matthys et al. (2004), we start with our second-order condition (C_{SO}), according to which

$$\frac{U(tx)}{U(t)} \simeq x^\gamma \exp \left\{ A(t) \frac{x^\rho - 1}{\rho} \right\}.$$

By setting $tx = 1/p$ and $t = Y_{n-k,n}$ where Y_i is a random variable from a standard Pareto distribution, since $X_{n-k,n} = U(Y_{n-k,n})$, we obtain the following approximation

$$\begin{aligned} x_p &\simeq X_{n-k,n} \left(\frac{1}{pY_{n-k,n}} \right)^\gamma \exp \left\{ A(Y_{n-k,n}) \frac{\left(\frac{1}{pY_{n-k,n}} \right)^\rho - 1}{\rho} \right\} \\ &\simeq X_{n-k,n} \left(\frac{k}{np} \right)^\gamma \exp \left\{ A\left(\frac{n}{k} \right) \frac{\left(\frac{k}{np} \right)^\rho - 1}{\rho} \right\}, \end{aligned}$$

where the last step follows from replacing $Y_{n-k,n}$ by its expected value n/k . The first part on the right-hand side (except the exponential term) is exactly a Weissman-type estimator when γ is replaced by an estimator (see Weissman 1978). Thus this exponential term can be viewed as a correcting term since $A(n/k)$ tends to 0 (and thus the exponential to 1). To take this correcting factor into account, we need to estimate $A(n/k)$. For this aim, we use our Proposition 1 in the appendix according to which, for any $\varepsilon > 0$,

$$\begin{aligned} \log \frac{Q_n(t)}{Q_n(1)} &= \log \frac{Q_n(t)}{U(\frac{n}{k})} - \log \frac{Q_n(1)}{U(\frac{n}{k})} \\ &= -\gamma \log t + \frac{\gamma}{\sqrt{k}} [t^{-1}W(t) - W(1)] + \tilde{A} \left(\frac{n}{k} \right) \frac{t^{-\rho} - 1}{\rho} + \frac{o\left(t^{-\frac{1}{2}-\varepsilon}\right)}{\sqrt{k}}, \end{aligned}$$

where $\tilde{A} \sim A$. This implies that

$$\sqrt{k} \left\{ \hat{\gamma}_k(K_1) - \hat{\gamma}_k(K_{2,\rho}) + A \left(\frac{n}{k} \right) \frac{\rho^2}{(1-\rho)(1-2\rho)} \right\} \xrightarrow{d} \gamma \int_0^1 [t^{-1}W(t) - W(1)] d(t[1 - K_{2,\rho}(t)])$$

which is asymptotically normal $\mathcal{N}(0, \mathcal{AV}(K_1 - K_{2,\rho}))$. Thus we can approximate

$$A \left(\frac{n}{k} \right) \frac{\rho^2}{(1-\rho)(1-2\rho)} \quad \text{by} \quad -[\hat{\gamma}_k(K_1) - \hat{\gamma}_k(K_{2,\rho})]$$

which means that $A \left(\frac{n}{k} \right)$ can be estimated by

$$-\frac{(1-\xi)(1-2\xi)}{\xi^2} [\hat{\gamma}_k(K_1) - \hat{\gamma}_k(K_{2,\xi})]$$

where ξ can be either a consistent estimator for ρ or a canonical negative value.

Our final extreme quantile estimator is then

$$\hat{x}_{p,\xi} = X_{n-k,n} \left(\frac{k}{np} \right)^{\hat{\gamma}_k(K_{\Delta_{opt}^*})} \exp \left\{ -\frac{(1-\xi)(1-2\xi)}{\xi^2} [\hat{\gamma}_k(K_1) - \hat{\gamma}_k(K_{2,\xi})] \frac{\left(\frac{k}{np} \right)^\xi - 1}{\xi} \right\}. \quad (10)$$

The aim of the next theorem is to prove that, under suitable assumptions, this estimator is asymptotically unbiased.

Theorem 3 *Let (X_1, X_2, \dots) be a stationary β -mixing time series with a continuous common marginal distribution function F and assume (C_{S0}) and (C_R) . Let $\hat{\rho}_{k_\rho}$ be an external estimator for ρ , consistent in probability, which depends on an intermediate sequence k_ρ . Consider now an intermediate sequence k such that $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$ and assume that $p = p_n$ such that $\frac{k}{np} \rightarrow \infty$, $\frac{\log(np)}{\sqrt{k}} \rightarrow 0$ and $n^{-a} \log p \rightarrow 0$ for all $a > 0$. Then, we have*

$$\frac{\sqrt{k}}{\log \frac{k}{np}} \left(\frac{\hat{x}_{p,\xi}}{x_p} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \mathcal{AV}(K_{\Delta_{opt}^*}) \right),$$

where ξ is either a canonical negative value $\tilde{\rho}$ or an estimator $\hat{\rho}$ consistent in probability such that $|\hat{\rho} - \rho| = O_{\mathbb{P}}(n^{-\varepsilon})$ for some $\varepsilon > 0$.

Note that the assumption $\log(np)/\sqrt{k} \rightarrow 0$ is useful in order to ensure that the rate of convergence for our extreme quantile estimator tends to infinity. The other condition on p , that is, $n^{-a} \log p \rightarrow 0$ for all $a > 0$, is only technical and not binding. This is usual in the context of extreme quantile estimation, (see, for instance, Matthys et al. 2004). In the latter, an extreme quantile estimator in the context of i.i.d. censored observations has been proposed, and its asymptotic normality also requires the condition $|\hat{\rho} - \rho| = O_{\mathbb{P}}(n^{-\varepsilon})$ for some $\varepsilon > 0$.

From the proof of Theorem 3, it becomes clear that the exponential term in (10) does not influence the limiting distribution. However, in finite samples situations, this factor typically leads to improved overall stability of the quantile estimates as a function of k . A similar idea has been pursued by de Haan et al. (2016), with another type of correcting factor than the exponential term, that is

$$\hat{x}_{k,k\rho}(p) = X_{n-k,n} \left(\frac{k}{np} \right)^{\hat{\gamma}_{k,k\rho}} \left(1 - \frac{\left[M_k^{(2)} - 2(\hat{\gamma}_k^H)^2 \right] [1 - \hat{\rho}_{k\rho}]^2}{2\hat{\gamma}_k^H \hat{\rho}_{k\rho}^2} \left[1 - \left(\frac{k}{np} \right)^{\hat{\rho}_{k\rho}} \right] \right). \quad (11)$$

Note that this estimator (11) is in fact slightly different from that included in the latter paper since a personal discussion with the authors allowed to identify an error in the proof of Theorem 4.2 in de Haan et al. (2016), precisely in their Assertion (A.6).

3 EXAMPLES AND SIMULATIONS

Our aim in this section is to compare our estimator with the one proposed by de Haan et al. (2016). Unfortunately, without specifying the covariance structure r in (C_R) , it is impossible to compare our asymptotic variance $\mathcal{AV}(K_{\Delta_{opt}^*})$ with that of the asymptotic unbiased estimator of de Haan et al. (2016) in its full generality. However, in the specific case of i.i.d. observations where we know that $r(s, t) = \min(s, t)$, our asymptotic variance given in (6) is clearly smaller than the asymptotic variance of the extreme value index proposed by de Haan et al. (2016), which is

$$\frac{\gamma^2}{\rho^2} (\rho^2 + (1 - \rho)^2).$$

To complete this comparison, we consider below several models commonly found in practice, with an explicit expression of the covariance structure r for two of the models. This allows us to provide an explicit comparison between the two estimators.

3.1 Autoregressive (AR) model

Let ε_i be i.i.d. variables with a positive Lebesgue density f_ε which is L_1 -Lipschitz continuous, that is,

$$\int |f_\varepsilon(\varepsilon + u) - f_\varepsilon(\varepsilon)| d\varepsilon = O(u) \quad \text{as } u \searrow 0.$$

Assume that

$$1 - F_\varepsilon(\varepsilon) \sim q\varepsilon^{-1/\gamma}\ell(\varepsilon) \quad \text{and} \quad F_\varepsilon(-\varepsilon) \sim (1 - q)\varepsilon^{-1/\gamma}\ell(\varepsilon) \quad \text{as } \varepsilon \rightarrow \infty,$$

for some slowly varying function ℓ and $q \in (0, 1)$. Consider now the stationary solution of the AR(1) equation

$$X_i = \theta X_{i-1} + \varepsilon_i, \quad \theta \in (0, 1). \quad (12)$$

The regularity conditions (C_R) hold with

$$r(x, y) = \min(x, y) + \sum_{m=1}^{\infty} \{c_m(x, y) + c_m(y, x)\}$$

where $c_m(x, y) = \min(x, y\theta^m)$.

Direct but tedious computations lead to the following asymptotic variance for our estimator

$$\mathcal{AV}(K_{\Delta_{opt}^*}) = \gamma^2 \left(\frac{1 - \rho}{\rho} \right)^2 r(1, 1), \quad (13)$$

that is always smaller than that obtained by de Haan et al. (2016) under the same framework, which is

$$\sigma^2(\theta, \gamma, \rho) := \frac{\gamma^2}{\rho^2} \left\{ [(1 - \rho)^2 + \rho^2] r(1, 1) + 2\rho(1 - \rho) \frac{\theta^{\frac{1}{\gamma}} \log \theta^{\frac{1}{\gamma}}}{(1 - \theta^{\frac{1}{\gamma}})^2} \right\}.$$

Note also that, compared with the i.i.d. case, our asymptotic variance $\mathcal{AV}(K_{\Delta_{opt}^*})$ is increased by the factor $r(1, 1) > 1$, see (6). In addition, if $K(t) = 1$ for $t \in (0, 1)$, our estimator $\hat{\gamma}_k(K)$ reduces to the classical Hill estimator $\hat{\gamma}_k^H$ and according to Drees (2000) (see also Stărică 1999), under serial dependence, the asymptotic variance of $\hat{\gamma}_k^H$ is $\gamma^2 r(1, 1)$. The latter value is smaller than $\mathcal{AV}(K_{\Delta_{opt}^*})$, but $\hat{\gamma}_k^H$ is not asymptotically unbiased.

3.2 Moving average (MA) model

Assume that ε_i satisfies the same assumptions as for the AR(1) model and consider this time the stationary solution of the MA(1) equation

$$X_i = \theta\varepsilon_{i-1} + \varepsilon_i. \quad (14)$$

In that case, the regularity conditions (C_R) are also satisfied with

$$r(x, y) = \min(x, y) + \left(1 + \theta^{\frac{1}{\gamma}}\right)^{-1} \left\{ \min\left(x, y\theta^{\frac{1}{\gamma}}\right) + \min\left(y, x\theta^{\frac{1}{\gamma}}\right) \right\}.$$

Again tedious computations show that the same expression for $\mathcal{AV}(K_{\Delta_{opt}^*})$ as that given in (13) is valid for the MA(1) model. Thus the comparison between our asymptotic variance and that obtained in the i.i.d. context and with the classical Hill estimator still remains valid. Concerning the estimator proposed by de Haan et al. (2016), they have obtained the asymptotic variance

$$\sigma^2(\theta, \gamma, \rho) := \frac{\gamma^2}{\rho^2} \left\{ [(1 - \rho)^2 + \rho^2] r(1, 1) + 2\rho(1 - \rho) \frac{\theta^{\frac{1}{\gamma}} \log \theta^{\frac{1}{\gamma}}}{1 + \theta^{\frac{1}{\gamma}}} \right\},$$

which is clearly again larger than our asymptotic variance $\mathcal{AV}(K_{\Delta_{opt}^*})$.

3.3 Generalized autoregressive conditional heteroskedasticity (GARCH) model

We consider the GARCH model defined as

$$X_t = \sigma_t \varepsilon_t \tag{15}$$

where (σ_t) is a function of the history up to time $t-1$ represented by \mathcal{H}_{t-1} . The process of innovations ε_t is a strict white noise with mean zero and variance one and is assumed to be independent of \mathcal{H}_{t-1} . In other words σ_t is \mathcal{H}_{t-1} -measurable, \mathcal{H}_{t-1} being the filtration generated by $(X_s)_{s \leq t-1}$ and therefore $\text{var}(X_t | \mathcal{H}_{t-1}) = \sigma_t^2$. The sequence (X_t) follows a GARCH(p, q) process if, for all t ,

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^p \alpha_j X_{t-j}^2 + \sum_{k=1}^q \beta_k \sigma_{t-k}^2, \quad \alpha_j, \beta_k > 0. \tag{16}$$

This model also satisfies the regularity conditions (C_R) but with a covariance structure r which cannot be explicitly computed. In that case the comparison between the different estimators can be done only by simulation.

In fact, in Section 3.4, we compare, in addition to the GARCH model, all the estimators for the three abovementioned models, through a simulation study. Actually, to be completely honest in the comparison, it is not sufficient to compare the constant in the variance. Rather, it is necessary to take the intermediate sequence into account as explained at the end of Section 2.1. Indeed, the variance of our estimator is $\mathcal{AV}(K_{\Delta_{opt}^*})/k$ whereas that of de Haan et al. (2016) is $\sigma^2(\theta, \gamma, \rho)/\tilde{k}$ with \tilde{k} of a larger order than k due to the different conditions imposed.

3.4 Simulation study

We proceed to a simulation study to assess our extreme value index estimator $\hat{\gamma}_k(K_{\hat{\Delta}_{opt}^*})$ and our high quantile estimator (10) with $p = 0.001$ in five different model cases for which we can simulate the theoretical value of γ and of the true 99.9% quantile. The three first models are the independence, AR(1) and MA(1) models proposed by de Haan et al. (2016) and with ε following the distribution

$$F_\varepsilon(\varepsilon) = \begin{cases} (1-q)(1-\tilde{F}(-\varepsilon)) & \text{if } \varepsilon < 0, \\ 1-q+q\tilde{F}(\varepsilon) & \text{if } \varepsilon > 0, \end{cases}$$

where \tilde{F} is the unit Fréchet distribution function, $q = 0.75$ so that F_ε belongs to the max-domain of attraction with an extreme value index $\gamma = 1$. We generate $N = 5000$ time series of size $n = 1000$ based on i.i.d observations generated from F_ε and we construct series from the three following models for which the theoretical value of γ is 1:

- Model 1: Independence model $X_i = \varepsilon_i$. The theoretical value of $x_{0.001}$ is 749.80.
- Model 2: AR(1) model (12) with $\theta = 0.3$. The theoretical value of $x_{0.001}$ is 1072.26.
- Model 3: MA(1) model (14) with $\theta = 0.3$. The theoretical value of $x_{0.001}$ is 972.85.

Note that the theoretical values are computed by Monte Carlo based on 1000 samples of size 10^6 . We also consider two further models that are GARCH models with realistic parameters provided from our two real cases studied in Section 4. They are

- Model 4: GARCH(1,1) model (15) with standardized Student t innovations with 5.99 degrees of freedom and $\hat{\alpha}_0 = 4.49 \cdot 10^{-06}$, $\hat{\alpha}_1 = 0.195$, $\hat{\beta}_1 = 0.746$. The theoretical value of $x_{0.001}$ is 0.049.
- Model 5: GARCH(1,2) model (15) with standardized Student t innovations with 5.66 degrees of freedom and $\hat{\alpha}_0 = 0.0443$, $\hat{\alpha}_1 = 0.202$, $\hat{\beta}_1 = 0.213$ and $\hat{\beta}_2 = 0.467$. The theoretical value of $x_{0.001}$ is 3.103.

In both GARCH cases, the innovations being Student t , the distribution of X_t belongs to the max-domain of attraction with an extreme value index evaluated at 0.27 for Model 4 and 0.15 for Model 5. We simulate $N = 5000$ time series of size $n = 1000$ for Model 4 (corresponding to the sample size of our real financial data) and $n = 4000$ for Model 5 (corresponding to the size of the wind speed data considered in our application).

To compute our high quantile (10), we need to select the sequence k_ρ . This has been done as follows

$$k_\rho := \sup \left\{ k : k \leq \min \left(m - 1, \frac{2m}{\log \log m} \right) \text{ and } \hat{\rho}_k \text{ exists} \right\},$$

with m being the number of positive observations in the sample.

We compare our extreme quantile estimator $\hat{x}_{p, \hat{\rho}_{k_\rho}}$ with the one proposed by de Haan et al. (2016) and defined in (11) where $\hat{\rho}_{k_\rho}$ is again the consistent estimator (7) and $\hat{\gamma}_{k, k_\rho}$ is their new estimator of the index defined in (8). To this aim, we compute the absolute value of the mean of the bias (ABias) together with the root mean squared errors (RMSE) based on the N samples, and defined as

$$\text{ABias}(\theta; k) := \left| \frac{1}{N} \sum_{i=1}^N \frac{\hat{\theta}^{(i)}}{\theta} - 1 \right| \text{ and } \text{RMSE}(\theta; k) := \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{\hat{\theta}^{(i)}}{\theta} - 1 \right)^2},$$

where θ is either γ or x_p , and $\hat{\theta}^{(i)}$ is the i -th value ($i = 1, \dots, N$) of the estimator of γ or x_p evaluated at k .

Figure 1, resp. Figure 2, shows the results for the extreme value index, resp. extreme quantile, for each of the five models by row, and by column, the ABias (left) and RMSE (right). The full line corresponds to our estimator, the dotted line to the original Hill (1975) or Weissman (1978) estimator and the dashed line to the de Haan et al. (2016) estimator. Note that in Figure 2, it is not appropriate to compare our curves with that of de Haan et al. (2016), Figures 6, 7 and 8, because our curves are computed from their corrected estimator.

Based on these simulations, we can draw the following conclusions:

- Our extreme value index performs similarly to de Haan et al. (2016) estimator in terms of bias in Models 1, 2 and 3 with a longer stability as a function of k than the Hill estimator, but a minimal value of the RMSE similar. However, for the GARCH models, the Hill estimator performs very poorly, whereas our estimator is at least as good as de Haan et al. (2016) estimator;
- Our high quantile estimator shows a lower bias than those of de Haan et al. (2016) and Weissman especially for the GARCH models. The bias is also less variable than the two others for the lowest values of k . In terms of RMSE it is also very competitive when compared to the two alternative estimators. Both the bias and RMSE of our high quantile show a period of stability for a wider range of k ; an important feature for practical applications.

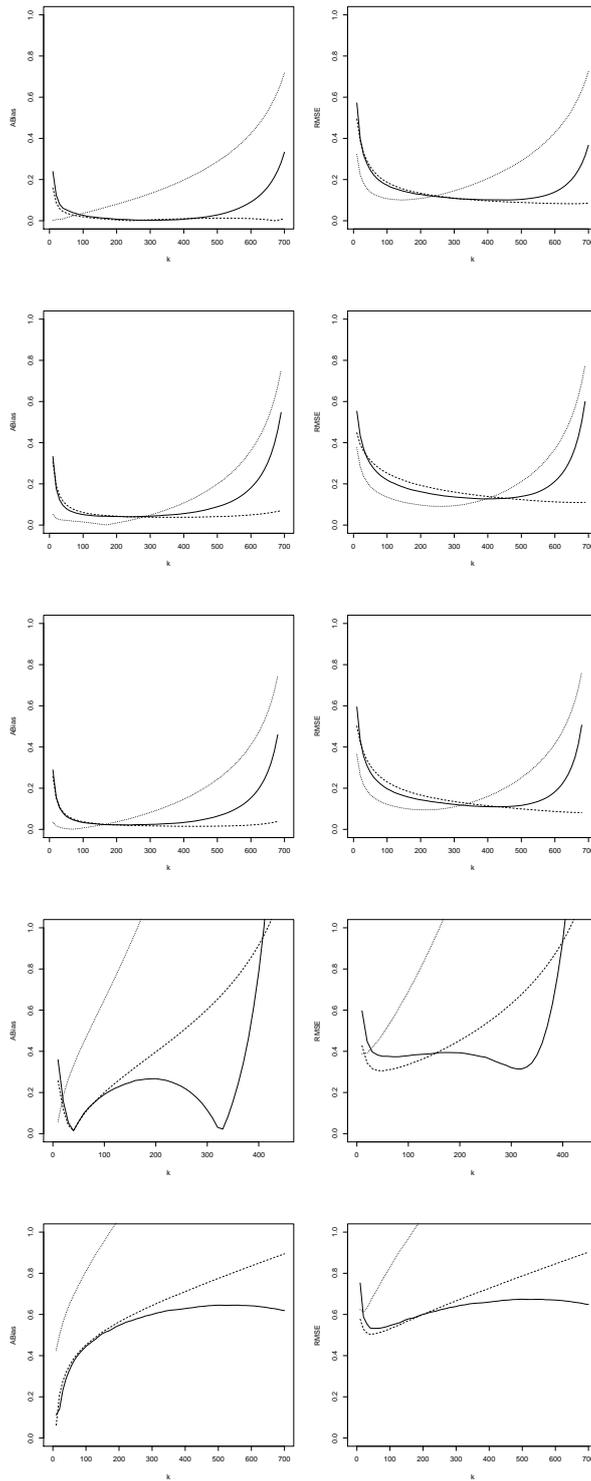


Figure 1: Simulation study: By row, Models 1, 2, 3, 4, 5. By column, ABias (left) and RMSE (right). The full line corresponds to our estimator of the extreme value index, the dashed line to the de Haan et al. (2016) estimator and the dotted line corresponds to the Hill estimator.

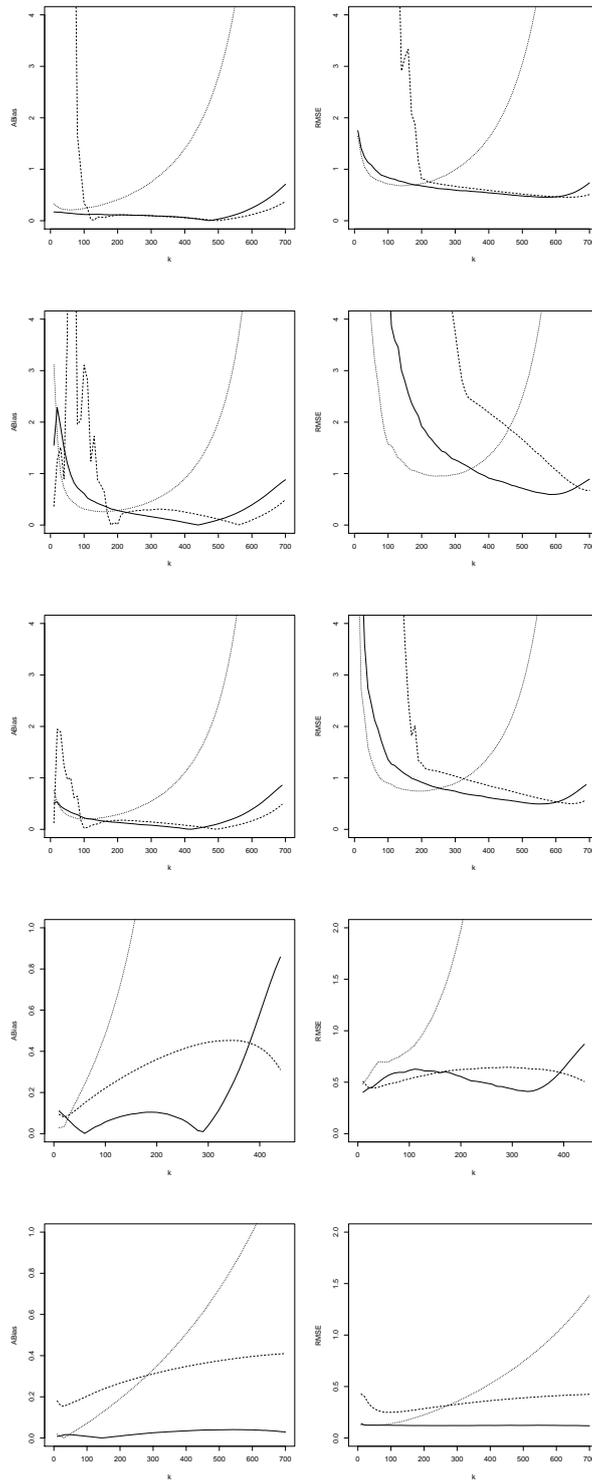


Figure 2: Simulation study: By row, Models 1, 2, 3, 4, 5. By column, ABias (left) and RMSE (right). The full line corresponds to our high quantile estimator, the dashed line to the de Haan et al. (2016) estimator and the dotted line corresponds to the Weissman (1978) estimator.

4 REAL DATA ANALYSIS

In this section, we illustrate the use of our estimator to calculate the daily VaR of a financial index series and the hourly return level of wind speed data. Both the VaR and return level are a high quantile (9) defined for a certain p -level. We use out-of-sample backtesting to assess the efficiency of our high quantile estimation.

4.1 Financial index data

The data in Figure 3 shows the daily negative log-returns X_t for $n = 1000$ values of S&P500 index from 2013-05-09 to 2017-04-27. In a risk management perspective, the Value-at-Risk (VaR) is a common quantity hardwired in the international regulatory framework referred to as the Basel Accords (see Tarullo 2008, for a historical review of the Basel International Settlement). The Basel Accord requires the largest international banks to hold regulatory capital for the trading book based on a 99%-VaR over a 1-day or 10-day holding period. The VaR-based risk capital calculation has received much attention over the last two decades (see McNeil et al. 2015, Chapter 1). The α -VaR for the horizon $h = 1$ day is the quantile x_p (9) with $p = 1 - \alpha$ of the distribution for the index daily log-returns.

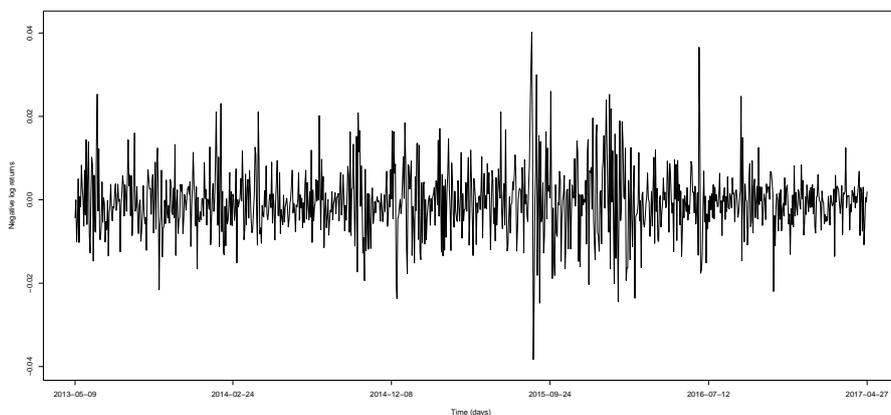


Figure 3: S&P500 index data: daily negative log-returns from 2013-05-09 to 2017-04-27.

Stylized features of financial series such as the S&P500 index returns are heavy tailedness; gaussianity assumption strongly violated; presence of heteroskedasticity or volatility clustering; absence of autocorrelations in returns (see Cont 2006, for more details). To model phenomenon with such characteristics, the GARCH(1,1) model (15) and (16) with $p = q = 1$ is specifically appropriate; see

McNeil et al. (2015), Chapter 4 for a review of GARCH models and Example 5.59, p. 235 for the GARCH fitting of financial time series. We fit a GARCH(1,1) model to our dataset with Student- t innovations. The estimated model is

$$\sigma_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 X_{t-1}^2 + \hat{\beta}_1 \sigma_{t-1}^2,$$

with $\hat{\alpha}_0 = 4.49 \cdot 10^{-06} (1.34 \cdot 10^{-06})$, $\hat{\alpha}_1 = 0.195 (0.0396)$, $\hat{\beta}_1 = 0.746 (0.0440)$ and the parameter of the Student t is $\hat{\nu} = 5.99 (1.143)$ where the value in parentheses is the standard deviation. The S&P500 log-returns being of a stationary β -mixing type, we can use our estimator to calculate the α -VaR for the horizon $h = 1$ day. We start to estimate the extreme value index γ of the loss returns using our estimator $\hat{\gamma}_k(K_{\hat{\Delta}_{opt}^*})$ that we compare to the Hill estimator and the asymptotically unbiased estimator of de Haan et al. (2016) for γ . Figure 4, left panel, shows the different curves of the estimated values against k .

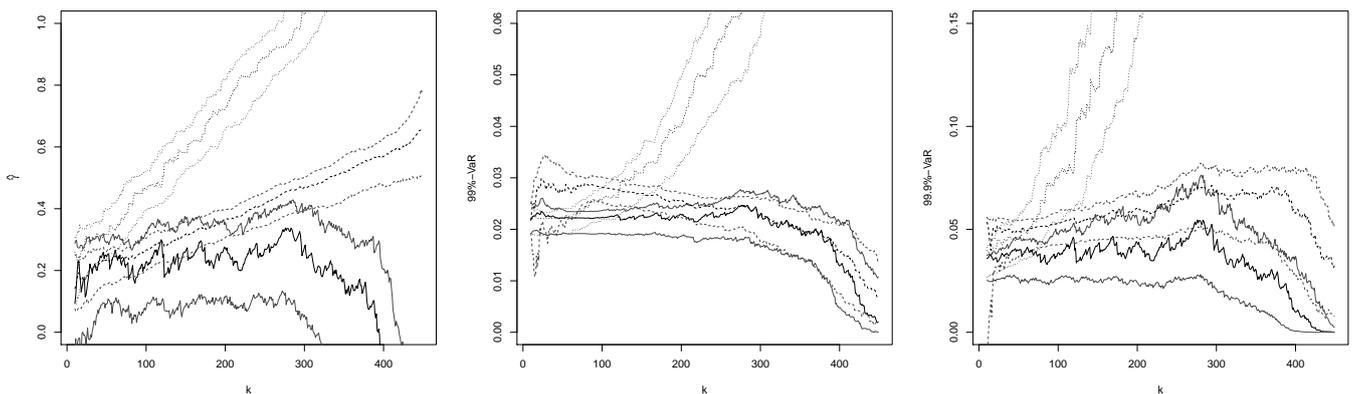


Figure 4: S&P500 index data: estimated values of γ (left panel), 99%-VaR (middle panel) and 99.9%-VaR (right panel) against k using our estimator $\hat{\gamma}_k(K_{\hat{\Delta}_{opt}^*})$ (left panel, full line), the de Haan et al. (2016) estimator (dashed line), the Hill estimator (left panel, dotted line) and the Weissman estimator (middle and right panels, dotted line). The grey lines are the 95%-bootstrap confidence intervals.

Our estimator (full line), even if more variable, looks more stable than the two others. Both the Hill estimator (dotted line) and that of de Haan et al. (2016) (dashed line) increase with k . The grey lines are 95%-confidence intervals calculated using block bootstrap (Bühlmann 2002, Davison & Hinkley 1997). We use blocks of length 200 as suggested by de Haan et al. (2016) and we simulate 99 bootstrap samples. We therefore estimate the 99% (resp. 99.9%)-VaR shown in Figure 4 middle panel (resp. right panel) using our high quantile estimator (full line), de Haan et al. (2016) estimator (dashed line) and Weissman (1978) estimator (dotted line). Whereas the Weissman estimator is

not stable for both levels of α , our estimator seems more stable than the one of de Haan et al. (2016) which slightly decreases over k for the lowest value of α and increases over k for the highest α . The stability of the estimator is essential to decide a value of k which will be used to get the high quantile estimator. The selection of k is equivalent to the choice of the threshold in the EVT peaks-over-threshold method. We arbitrarily choose $k = 80$ that is a value within the window of stable values of k in Figure 4.

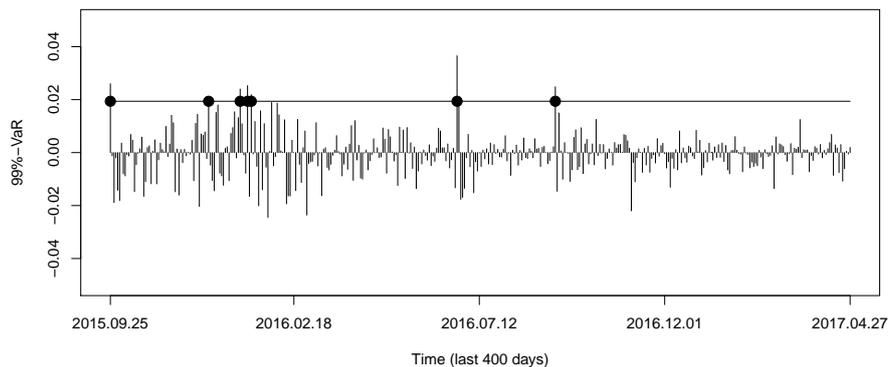


Figure 5: S&P500 index data: the straight line is the 99%-VaR estimate calculated using the historical period 2013-05-09 to 2015-09-23. The points are the times at which the realized log-return is higher than the estimated 99%-VaR.

To assess the efficiency of the VaR estimator, it is standard to use the so-called backtesting procedure (Danielsson 2011) as suggested by the Basel Committee on banking supervision. The backtesting procedure is based on comparing the VaR with the realized value over a certain horizon. A violation is said to occur whenever the estimated VaR is lower than a realized value. As pointed out by Christoffersen (1998) an accurate VaR model satisfies two properties, namely the unconditional coverage property and the independence property. The first property states that a VaR for a $100 \times (1 - \alpha)\%$ coverage rate is valid if the expected frequency of observed violations is also equal to $100 \times (1 - \alpha)\%$. Under the independence property, the violations are supposed to be distributed independently. Since three decades, a variety of tests have been proposed, testing on either the unconditional coverage hypothesis and on the independence hypothesis separately or on both hypotheses jointly. The latter tests are called conditional coverage tests (Christoffersen 1998). In our context assuming that the underlying data generating process is stationary over time, we proceed to the following backtesting procedure of our unconditional VaR estimate: we choose an estimating window of historical data corresponding to the period 2013-05-09 to 2015-09-23 from

which we estimate the 99%-VaR. We compare this 99%-VaR estimate with the 400 realized values of the testing period 2015-09-24 to 2017-04-26 on which we count the number of violations. Figure 5 shows the 99%-VaR estimate (straight line). The expected number of violations of the VaR is 4 and the observed number is 7. The VaR violations are represented by the points in Figure 5. The highest violation during this period happened on 2016-06-23 explained by the Wall Street reaction to Brexit (see, for instance, <http://www.reuters.com/article/us-usa-stocks-idUSKCN0Z918E>). The unconditional coverage Kupiec (1995) test we use is a variation of the Binomial test. It provides a p -value of 0.173. The conditional test of Christoffersen (1998) provides a p -value of 0.355. Both fail to reject the null hypothesis of an accurate VaR model.

4.2 Wind speed data

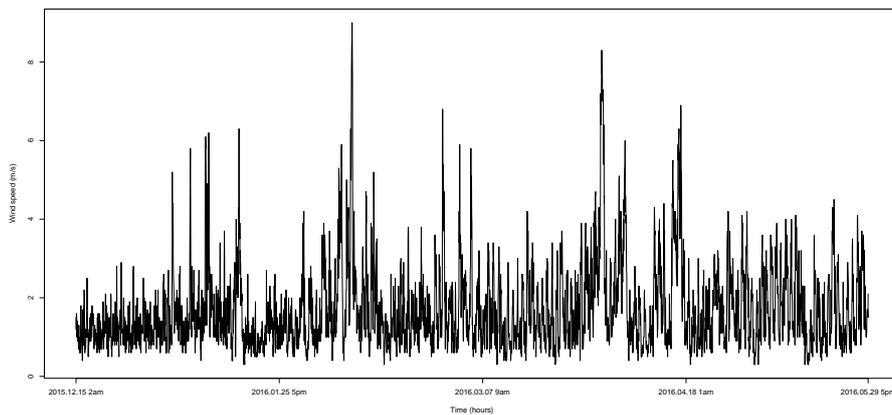


Figure 6: Wind speed data: hourly measurements in Arosa, Switzerland, from 2015.12.15, 2am to 2016.05.29, 5pm.

Figure 6 shows hourly wind speed (m/s) data X_t , (t represents hour) measured in Arosa, Switzerland, from 2015.12.15, 2am to 2016.05.29, 5pm, consisting of $n = 3895$ hourly values. The data were provided by the Federal Office of Meteorology and Climatology, SwissMeteo. The data show evidence of seasonality confirmed by the ACF plot of the series in the top left panel of Figure 7. Several recent papers suggest that wind speed data are of an ARMA-GARCH-type. Lojowska et al. (2010), for instance, claim that artificial wind speeds simulated from ARMA-GARCH models are statistically indistinguishable from the real wind speed time series measurements under observation; Liu et al. (2013) use the ARMA-GARCH model to predict the time series mean and volatility of wind speed. The top left panel of Figure 7 shows that autocorrelations of Arosa wind speeds are significant for a large number of lags but we can notice that the PACF plot (top right panel) has

a significant spike only at lag 1, meaning that all the higher-order autocorrelations are effectively explained by the lag-1 autocorrelation. More formally, we fit different ARIMA models and the model selected according to AIC is an ARIMA(1,1,1) corresponding to a model with one AR term and one MA term and a first difference used to account for a linear trend in the data as follows

$$Y_t = X_t - X_{t-1}$$

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1},$$

where ε_t is a random shock occurring at time t . Fitted to the data, the estimated parameters are $\hat{\phi}_1 = 0.819(0.0125)$, $\hat{\theta}_1 = -0.989(0.0050)$. The ACF in the bottom left panel of Figure 7 shows no serial correlation remaining for the ARIMA residuals

$$e_t = x_t - x_{t-1} - \hat{\phi}_1 x_{t-1} + \hat{\phi}_1 x_{t-2} + \hat{\theta}_1 e_{t-1}, \quad t = 1, \dots, n. \quad (17)$$

However, from the ACF plot of the absolute values for the ARIMA residuals e_1, \dots, e_n , there is evidence of volatility clustering. The absence of autocorrelations for the residuals and the presence of volatility clustering clearly suggests a GARCH-type model. This is confirmed by a Ljung-Box test on the squared ARIMA residuals (17) providing a p -value lower than 10^{-5} , rejecting the null hypothesis that the series is a strict white noise. We fit several GARCH models of different orders

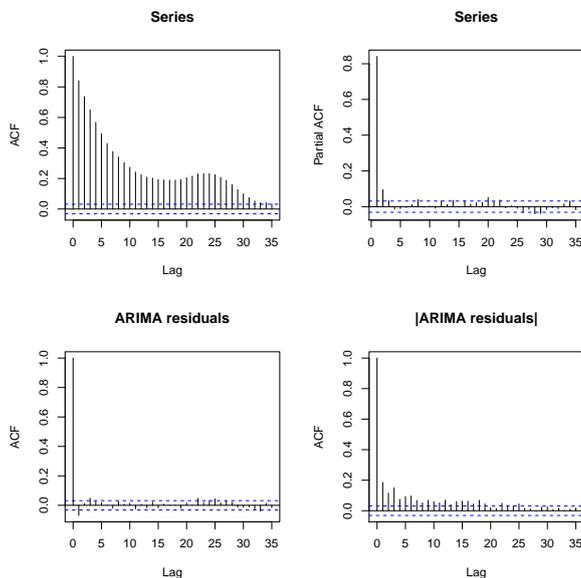


Figure 7: Wind speed data: ACF (top left) and PACF (top right) of the wind speed series and ACF of the ARIMA residuals (bottom left) and of the absolute value of the ARIMA residuals (bottom right).

(p, q) on the ARIMA residuals e_1, \dots, e_n . The ARIMA residuals being heavy-tailed, we use Student

t innovations. The GARCH form is given by (15) where here X_t represents our residuals and σ_t is the ARIMA wind speed residual volatility. We select the orders p and q using AIC. Figure 8 represents the AIC against the total number of parameters estimated for each model, that is $p+q+1$. The AIC is minimized for $p+q+1=4$. Not readable from the graph, the lowest AIC value for $p+q+1=4$ corresponds to the model GARCH(1,2) which is

$$\sigma_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 e_{t-1}^2 + \hat{\beta}_1 \sigma_{t-1}^2 + \hat{\beta}_2 \sigma_{t-2}^2,$$

$\hat{\alpha}_0 = 0.0443(0.00946)$, $\hat{\alpha}_1 = 0.202(0.0279)$, $\hat{\beta}_1 = 0.213(0.0764)$ and $\hat{\beta}_2 = 0.467(0.0787)$ with standardized Student t innovations with degrees of freedom $\hat{\nu} = 5.66$. With such data, a measure of

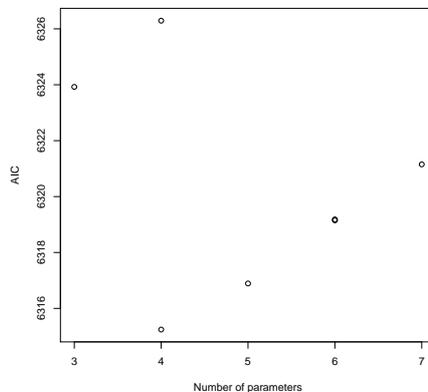


Figure 8: Wind speed data: AIC against the GARCH number of parameters estimated.

extreme events of interest is the $1/p$ -hour return level $r(p)$ with p small. The return level is the value that has a p chance of being exceeded in a given hour. Our proposed estimator for high quantile (9) can be used on the series of the ARIMA residuals e_1, \dots, e_n being GARCH-type removed from seasonality and therefore satisfying the stationary β -mixing conditions. Note that to obtain a return level for the wind speed original data $r_{x_t}(p)$ (in m/s) at time t one can use the ARIMA model and the $1/p$ -hour return level $r_{e_t}(p)$ estimate of the ARIMA residual e_t at time t , that is

$$\hat{r}_{x_t}(p) = \hat{r}_{e_t}(p) - \hat{\theta}_1 e_{t-1} + x_{t-1} + \hat{\phi}_1 x_{t-1} - \hat{\phi}_1 x_{t-2}. \quad (18)$$

To estimate the high quantile of the ARIMA residuals $r_e(p)$, we first estimate γ using our estimator $\hat{\gamma}_k(K_{\hat{\Delta}_{opt}^*})$. The full line of the left panel in Figure 9 shows our estimated values of γ against different values of k . Compared to the Hill (dotted line) and the de Haan et al. (2016) (dashed line) estimators, our estimator looks more stable over k even if it seems more variable. The grey lines are the 95%-confidence intervals calculated using a block bootstrapping method with block length of

size 200 and based on 99 bootstrap samples. For the high quantile estimation with $p = 0.01$ (resp. $p = 0.001$), corresponding to the 100-hour return level (resp. 1000-hour return level), our estimator (full line) stays very stable as shown in the middle panel of Figure 9 (resp. right panel) compared to the Weissman estimator (dotted line) and even to that of de Haan et al. (2016) (dashed line) which shows a slightly decreasing trend (resp. increasing trend).

From Figure 9, we can reasonably choose any value of k between 200 and 600 without leading much to variable results. We chose $k = 450$ for our estimator and as for the financial data, we proceed to an estimation and testing of the 100-hour return level. The straight line in Figure 10 (top panel) shows our 100-hour return level estimate for the ARIMA residuals calculated from the estimating period 2015.12.15 at 2am to 2016.04.05 at 7am. We compare this 99%-quantile estimate with the 1200 observed residuals of the testing period 2016.04.05 at 8am to 2016.05.29 at 4pm. The expected number of 100-hour return level violations is 12 and the observed number is 11. Both the Kupiec unconditional coverage test (p -value = 0.768) and the Christoffersen conditional test (p -value = 0.864) are not rejected. The bottom panel of Figure 10 shows the 100-hour return level for the original wind speed data (in m/s) calculated using (18). Confidence intervals are not shown in the plot for clarity sake. Because of the two-step method used (filtering the seasonality using ARIMA model and then applying our estimator on the residuals), a convenient way to get confidence intervals is by proceeding to a block bootstrap of the original data and applying the two-step method for each resample.

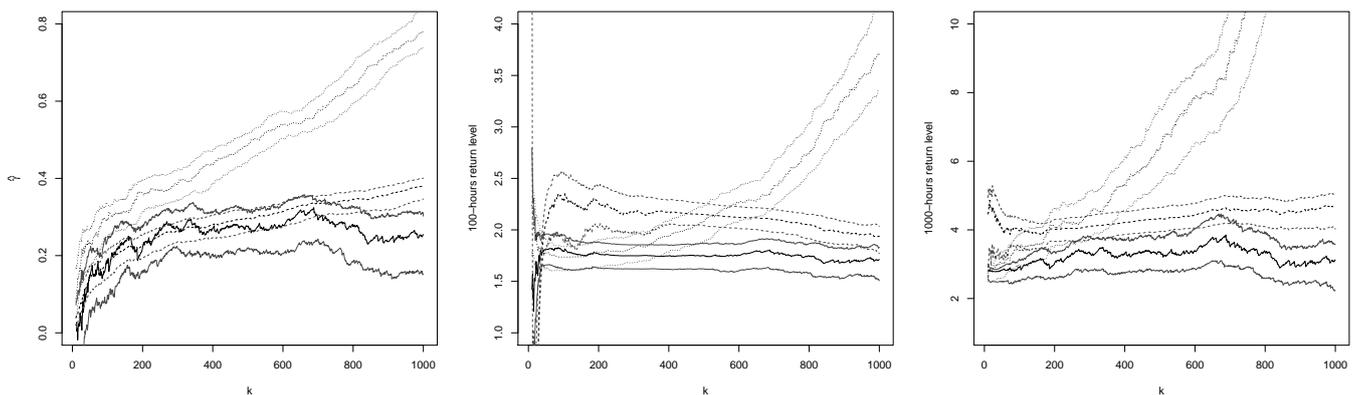


Figure 9: Wind speed data: estimated values of γ (left panel), 100-hour return level (middle panel) and 1000-hour return level (right panel) against k using our estimator $\hat{\gamma}_k(K_{\hat{\Delta}_{opt}^*})$ (left panel, full line), the de Haan et al. (2016) estimator (dashed line), the Hill estimator (left panel, dotted line) and the Weissman estimator (middle and right panels, dotted line). The grey lines are the 95%-bootstrap confidence intervals.

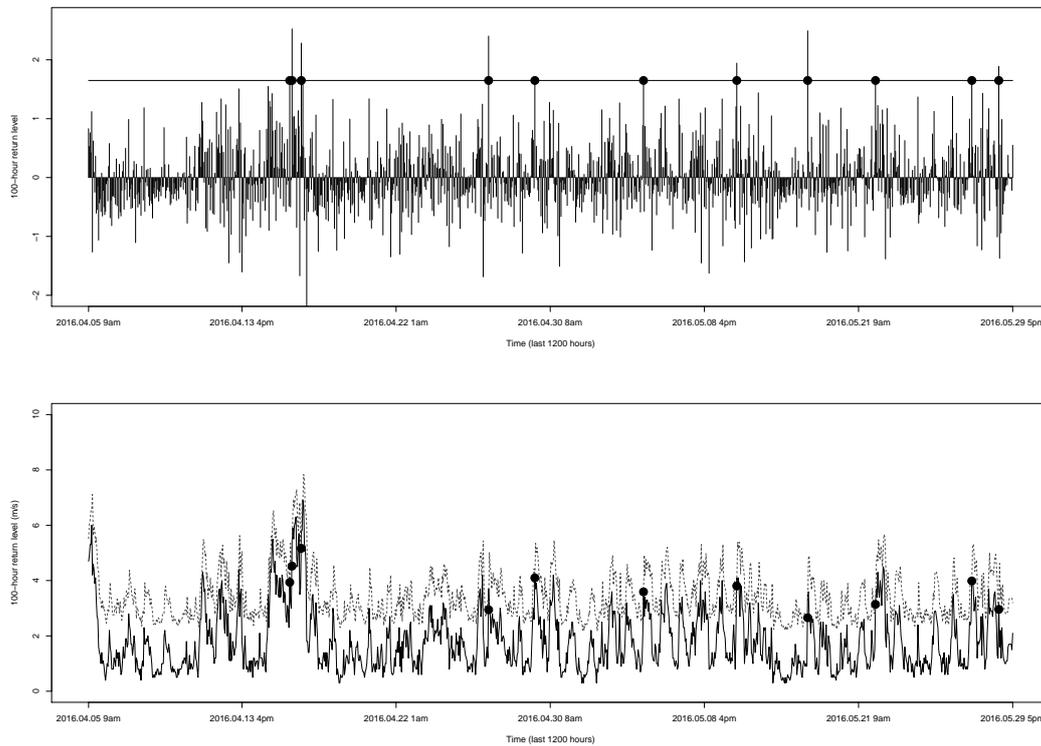


Figure 10: Wind speed data: the straight line of the top panel shows our 100-hour return level estimate calculated from the estimating period 2015.12.15 at 2am to 2016.04.05 at 7am of the ARIMA residuals. The dashed line of the bottom panel shows the high quantile estimate of the wind speed data in m/s. The points are the times at which the ARIMA residual (top) and wind speed value (bottom) is higher than the estimated return level.

5 CONCLUSION

In this paper we have introduced a new asymptotically unbiased estimator of high quantiles for β -mixing stationary time series. Comparing the new procedure to the alternative proposed by de Haan et al. (2016), our high quantile estimator provides, in addition to lower ABias and RMSE in general, more stability over k , an important feature expected in this type of approach to be applicable in practice. In application, the new high quantile estimator can be proposed to any other stationary β -mixing heavy-tailed time series for which high quantiles needed to be calculated. This concerns heavy-tailed autoregressive data encountered in network traffic forecasting for instance and many other applications data in climate change.

APPENDIX: PROOFS OF THE RESULTS

Before establishing our Theorem 1, we need a result similar to Proposition A.1 in de Haan et al. (2016) but under the weak assumptions of our Theorem 1, excluding (C_K) . In particular, a third order condition is not assumed.

Proposition 1. *Let (X_1, X_2, \dots) be a stationary β -mixing time series with a continuous common marginal distribution function F and assume (C_{S0}) and (C_R) . Suppose that k is an intermediate sequence such that $\sqrt{k}A(n/k) = O(1)$. For a given $\varepsilon > 0$, under a Skorohod construction, there exist a function $\tilde{A} \sim A$, and a centered Gaussian process $(W(t))_{t \in [0,1]}$ with covariance function r , such that, as $n \rightarrow \infty$*

$$\sup_{t \in (0,1]} t^{\frac{1}{2} + \varepsilon} \left| \sqrt{k} \left(\log \frac{Q_n(t)}{U\left(\frac{n}{k}\right)} + \gamma \log t \right) - \gamma t^{-1} W(t) - \sqrt{k} \tilde{A} \left(\frac{n}{k} \right) \frac{t^{-\rho} - 1}{\rho} \right| \longrightarrow 0 \quad a.s.$$

Proof of Proposition 1. It is similar to that of Proposition A.1 in de Haan et al. (2016) but assuming that $\sqrt{k}A\left(\frac{n}{k}\right) = O(1)$ and without a third order condition, thus below we only give the main differences. In our context, the key inequality is the following: for all $\varepsilon, \delta > 0$, there exists some positive number $u_0 = u_0(\varepsilon, \delta)$ such that for $ux \geq u_0$:

$$\left| \frac{\log U(ux) - \log U(u) - \gamma \log x}{\tilde{A}(u)} - \frac{x^\rho - 1}{\rho} \right| \leq \varepsilon x^\rho \max(x^\delta, x^{-\delta}), \quad (19)$$

see, for instance, Theorem B.2.18 in de Haan & Ferreira (2006). Now, using the representation $X_i = U(Y_i)$ where Y_i follows a standard Pareto distribution, (Y_1, Y_2, \dots) is a stationary β -mixing series satisfying the regularity conditions (C_R) . Then, according to Drees (2003), since $Q_n(t) = U(Y_{n-[kt],n})$ and under a Skorohod construction, there exists a centered Gaussian process $(W(t))_{t \in [0,1]}$ with a covariance function r such that for $\varepsilon > 0$, as $n \rightarrow \infty$

$$\sup_{t \in (0,1]} t^{\frac{1}{2} + \varepsilon} \left| \sqrt{k} \left(t \frac{Y_{n-[kt],n}}{n/k} - 1 \right) - t^{-1} W(t) \right| \longrightarrow 0 \quad a.s.$$

This convergence combining with inequality (19) entails that

$$\left| \log Q_n(t) - \log U\left(\frac{n}{k}\right) - \gamma \log \left(\frac{k}{n} Y_{n-[kt],n} \right) - \tilde{A} \left(\frac{n}{k} \right) \frac{\left(\frac{k}{n} Y_{n-[kt],n} \right)^\rho - 1}{\rho} \right| \leq \varepsilon \left| \tilde{A} \left(\frac{n}{k} \right) \right| \left(\frac{k}{n} Y_{n-[kt],n} \right)^{\rho + \delta}$$

for sufficiently large $n > n_0(\varepsilon, \delta)$, with probability 1.

Consequently

$$t^{\frac{1}{2} + \varepsilon} \left| \sqrt{k} \left(\log \frac{Q_n(t)}{U\left(\frac{n}{k}\right)} + \gamma \log t \right) - \gamma t^{-1} W(t) - \sqrt{k} \tilde{A} \left(\frac{n}{k} \right) \frac{t^{-\rho} - 1}{\rho} + \sqrt{k} \tilde{A} \left(\frac{n}{k} \right) \frac{1}{\rho} \left\{ t^{-\rho} - \left(\frac{k}{n} Y_{n-[kt],n} \right)^\rho \right\} - \gamma \left\{ \sqrt{k} \left(\log \left(\frac{k}{n} Y_{n-[kt],n} \right) + \log t \right) - t^{-1} W(t) \right\} \right| \leq \varepsilon \sqrt{k} \left| \tilde{A} \left(\frac{n}{k} \right) \right| t^{\frac{1}{2} + \varepsilon} \left(\frac{k}{n} Y_{n-[kt],n} \right)^{\rho + \delta}.$$

Choosing $\delta \in (0, -\rho)$, Proposition 1 then follows similarly as Proposition A.1 in de Haan et al. (2016) since ε can be arbitrarily close to 0. ■

Proof of Theorem 1. From Proposition 1, we can easily infer that

$$\begin{aligned} \sqrt{k} \left\{ \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} d(tK(t)) + \gamma \int_0^1 \log t d(tK(t)) \right\} &= \gamma \int_0^1 [t^{-1}W(t) - W(1)] d(tK(t)) \\ &+ \sqrt{k} \tilde{A} \left(\frac{n}{k} \right) \int_0^1 \frac{t^{-\rho} - 1}{\rho} d(tK(t)) + o(1) \int_0^1 t^{-\frac{1}{2}-\varepsilon} d(tK(t)). \end{aligned}$$

Using integration by parts, we have

$$\begin{aligned} \int_0^1 \log t d(tK(t)) &= -1, \\ \text{and } \int_0^1 \frac{t^{-\rho} - 1}{\rho} d(tK(t)) &= \int_0^1 t^{-\rho} K(t) dt, \end{aligned}$$

from which (3) follows under (C_K) by taking $0 < \varepsilon < \frac{1}{2} - \tau$. The covariance structure r of the centered Gaussian process $(W(t))_{t \in [0,1]}$ now achieves the proof of Theorem 1. ■

Proof of Corollary 1. It is a direct consequence of Theorem 1 since by construction the bias of $\hat{\gamma}_k(K_{\Delta^*})$ is null. ■

Proof of Corollary 2. According to Theorem 1, we only need to check the bias term. Recall that

$$K_{\tilde{\Delta}_{opt}^*}(t) = \left(\frac{1 - \tilde{\rho}}{\tilde{\rho}} \right)^2 - \frac{(1 - \tilde{\rho})(1 - 2\tilde{\rho})}{\tilde{\rho}^2} t^{-\tilde{\rho}},$$

from which we deduce that

$$\int_0^1 t^{-\rho} K_{\tilde{\Delta}_{opt}^*}(t) dt = \frac{(1 - \tilde{\rho})(\tilde{\rho} - \rho)}{\tilde{\rho}(1 - \rho)(1 - \tilde{\rho} - \rho)}.$$

This achieves the proof of Corollary 2. ■

Proof of Theorem 2. Let $\Delta_{opt}^* := \left(\frac{1-\rho}{\rho} \right)^2$ and $\hat{\Delta}_{opt}^* := \left(\frac{1-\hat{\rho}_{k\rho}}{\hat{\rho}_{k\rho}} \right)^2$. We consider the decomposition

$$\sqrt{k} \left(\hat{\gamma}_k(K_{\hat{\Delta}_{opt}^*}) - \gamma \right) = \sqrt{k} \left(\hat{\gamma}_k(K_{\Delta_{opt}^*}) - \gamma \right) + \sqrt{k} \left(\hat{\gamma}_k(K_{\hat{\Delta}_{opt}^*}) - \hat{\gamma}_k(K_{\Delta_{opt}^*}) \right). \quad (20)$$

According to Corollary 1 we have

$$\sqrt{k} \left(\hat{\gamma}_k(K_{\Delta_{opt}^*}) - \gamma \right) \xrightarrow{d} \mathcal{N} \left(0, \mathcal{AV}(K_{\Delta_{opt}^*}) \right).$$

To prove Theorem 2, it is thus sufficient to show that the second term in (20) is $o_{\mathbb{P}}(1)$. For this aim, note that

$$\begin{aligned}
& \sqrt{k} \left(\widehat{\gamma}_k(K_{\widehat{\Delta}_{opt}^*}) - \widehat{\gamma}_k(K_{\Delta_{opt}^*}) \right) \\
&= \sqrt{k} \left\{ \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} d(tK_{\widehat{\Delta}_{opt}^*}(t)) - \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} d(tK_{\Delta_{opt}^*}(t)) \right\} \\
&= \sqrt{k} \left\{ \widehat{\Delta}_{opt}^* \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} dt - \Delta_{opt}^* \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} dt \right. \\
&\quad \left. + (1 - \widehat{\Delta}_{opt}^*) \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} d(tK_{2, \widehat{\rho}_{k\rho}}(t)) - (1 - \Delta_{opt}^*) \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} d(tK_{2, \rho}(t)) \right\} \\
&= \sqrt{k} \left(\widehat{\Delta}_{opt}^* - \Delta_{opt}^* \right) \left\{ \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} dt - \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} d(tK_{2, \rho}(t)) \right\} \\
&\quad + \sqrt{k} \left(1 - \widehat{\Delta}_{opt}^* \right) \left\{ \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} d(tK_{2, \widehat{\rho}_{k\rho}}(t)) - \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} d(tK_{2, \rho}(t)) \right\} \\
&= \left(\widehat{\Delta}_{opt}^* - \Delta_{opt}^* \right) \left\{ \sqrt{k} (\widehat{\gamma}_k(K_1) - \gamma) - \sqrt{k} (\widehat{\gamma}_k(K_{2, \rho}) - \gamma) \right\} \\
&\quad + \left(1 - \widehat{\Delta}_{opt}^* \right) \sqrt{k} \left\{ \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} d(tK_{2, \widehat{\rho}_{k\rho}}(t)) - \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} d(tK_{2, \rho}(t)) \right\} \\
&=: T_1 + T_2.
\end{aligned}$$

We will study the two terms separately.

Term T_1 . Using the consistency in probability of $\widehat{\rho}_{k\rho}$ and the convergences

$$\begin{aligned}
\sqrt{k} (\widehat{\gamma}_k(K_1) - \gamma) &\xrightarrow{d} \mathcal{N}(\lambda/(1-\rho), \mathcal{AV}(K_1)) \\
\sqrt{k} (\widehat{\gamma}_k(K_{2, \rho}) - \gamma) &\xrightarrow{d} \mathcal{N}(\lambda \mathcal{AB}(K_{2, \rho}), \mathcal{AV}(K_{2, \rho}))
\end{aligned}$$

coming from Theorem 1, we have $T_1 = o_{\mathbb{P}}(1)$.

Term T_2 . For $\varepsilon \in (0, 1/2)$, uniformly for $t \in (0, 1]$:

$$\log \frac{Q_n(t)}{Q_n(1)} = \gamma(-\log t) + \frac{\gamma}{\sqrt{k}} [t^{-1}W(t) - W(1)] + \widetilde{A} \left(\frac{n}{k} \right) \frac{t^{-\rho} - 1}{\rho} + \frac{o(1)}{\sqrt{k}} t^{-\frac{1}{2} - \varepsilon}.$$

This implies that

$$\begin{aligned}
& \sqrt{k} \left\{ \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} d(tK_{2,\hat{\rho}_{k\rho}}(t)) - \int_0^1 \log \frac{Q_n(t)}{Q_n(1)} d(tK_{2,\rho}(t)) \right\} \\
&= \gamma \sqrt{k} \left\{ \int_0^1 (-\log t) d(tK_{2,\hat{\rho}_{k\rho}}(t)) - \int_0^1 (-\log t) d(tK_{2,\rho}(t)) \right\} \\
&+ \gamma \left\{ \int_0^1 [t^{-1}W(t) - W(1)] d(tK_{2,\hat{\rho}_{k\rho}}(t)) - \int_0^1 [t^{-1}W(t) - W(1)] d(tK_{2,\rho}(t)) \right\} \\
&+ \sqrt{k} \tilde{A} \left(\frac{n}{k} \right) \left\{ \int_0^1 \frac{t^{-\rho} - 1}{\rho} d(tK_{2,\hat{\rho}_{k\rho}}(t)) - \int_0^1 \frac{t^{-\rho} - 1}{\rho} d(tK_{2,\rho}(t)) \right\} \\
&+ o(1) \int_0^1 t^{-\frac{1}{2}-\varepsilon} d(tK_{2,\hat{\rho}_{k\rho}}(t)) + o(1) \int_0^1 t^{-\frac{1}{2}-\varepsilon} d(tK_{2,\rho}(t)) \\
&= \gamma \sqrt{k} \left\{ \left[-t \log t \left(K_{2,\hat{\rho}_{k\rho}}(t) - K_{2,\rho}(t) \right) \right]_0^1 + \int_0^1 K_{2,\hat{\rho}_{k\rho}}(t) dt - \int_0^1 K_{2,\rho}(t) dt \right\} \\
&+ \gamma \left\{ \int_0^1 [t^{-1}W(t) - W(1)] \left(K_{2,\hat{\rho}_{k\rho}}(t) - K_{2,\rho}(t) \right) dt + \int_0^1 [t^{-1}W(t) - W(1)] t \left(K'_{2,\hat{\rho}_{k\rho}}(t) - K'_{2,\rho}(t) \right) dt \right\} \\
&+ \sqrt{k} \tilde{A} \left(\frac{n}{k} \right) \left\{ \left[t \frac{t^{-\rho} - 1}{\rho} \left(K_{2,\hat{\rho}_{k\rho}}(t) - K_{2,\rho}(t) \right) \right]_0^1 + \int_0^1 t^{-\rho} \left(K_{2,\hat{\rho}_{k\rho}}(t) - K_{2,\rho}(t) \right) dt \right\} \\
&+ o(1) \left\{ \left[t^{\frac{1}{2}-\varepsilon} \left(K_{2,\hat{\rho}_{k\rho}}(t) - K_{2,\rho}(t) \right) \right]_0^1 + \left(\frac{1}{2} + \varepsilon \right) \int_0^1 t^{-\frac{1}{2}-\varepsilon} \left(K_{2,\hat{\rho}_{k\rho}}(t) - K_{2,\rho}(t) \right) dt \right\} \\
&+ o(1) \int_0^1 t^{-\frac{1}{2}-\varepsilon} d(tK_{2,\rho}(t)) \\
&= \gamma \left\{ \int_0^1 [t^{-1}W(t) - W(1)] \left(K_{2,\hat{\rho}_{k\rho}}(t) - K_{2,\rho}(t) \right) dt + \int_0^1 [t^{-1}W(t) - W(1)] t \left(K'_{2,\hat{\rho}_{k\rho}}(t) - K'_{2,\rho}(t) \right) dt \right\} \\
&+ o_{\mathbb{P}}(1).
\end{aligned}$$

Now, note that for $\varepsilon \in (0, 1/4)$ and $\tilde{\rho}$ a random value between ρ and $\hat{\rho}_{k\rho}$, we have

$$\begin{aligned}
& \bullet \left| \int_0^1 [t^{-1}W(t) - W(1)] \left[K_{2,\hat{\rho}_{k\rho}}(t) - K_{2,\rho}(t) \right] dt \right| \\
&\leq \int_0^1 |t^{-1}W(t) - W(1)| \left| K_{2,\hat{\rho}_{k\rho}}(t) - K_{2,\rho}(t) \right| dt \\
&\leq (1 - \hat{\rho}_{k\rho}) \int_0^1 |t^{-1}W(t) - W(1)| |t^{-\hat{\rho}_{k\rho}} - t^{-\rho}| dt + |\hat{\rho}_{k\rho} - \rho| \int_0^1 |t^{-1}W(t) - W(1)| dt \\
&\leq (1 - \hat{\rho}_{k\rho}) \sup_{t \in (0,1]} \left| t^{\frac{1}{2}+\varepsilon} [t^{-1}W(t) - W(1)] \right| \sup_{t \in (0,1]} t^{\frac{1}{4}} |t^{-\hat{\rho}_{k\rho}} - t^{-\rho}| \int_0^1 t^{-\frac{3}{4}-\varepsilon} dt \\
&\quad + |\hat{\rho}_{k\rho} - \rho| \int_0^1 |t^{-1}W(t) - W(1)| dt \\
&\leq \frac{4}{1 - 4\varepsilon} |\hat{\rho}_{k\rho} - \rho| (1 - \hat{\rho}_{k\rho}) \sup_{t \in (0,1]} \left| t^{\frac{1}{2}+\varepsilon} [t^{-1}W(t) - W(1)] \right| \sup_{t \in (0,1]} (-\log t) t^{\frac{1}{4}-\tilde{\rho}} \\
&\quad + |\hat{\rho}_{k\rho} - \rho| \int_0^1 |t^{-1}W(t) - W(1)| dt \\
&= o_{\mathbb{P}}(1),
\end{aligned}$$

since $\sup_{t \in (0,1]} t^{\frac{1}{2} + \varepsilon} t^{-1} |W(t)| = O(1)$ a.s.. Similarly, we have

$$\begin{aligned}
& \bullet \left| \int_0^1 [t^{-1}W(t) - W(1)] t \left[K'_{2, \hat{\rho}_{k_\rho}}(t) - K'_{2, \rho}(t) \right] dt \right| \\
& \leq \int_0^1 |t^{-1}W(t) - W(1)| t \left| K'_{2, \hat{\rho}_{k_\rho}}(t) - K'_{2, \rho}(t) \right| dt \\
& \leq |\hat{\rho}_{k_\rho}|(1 - \hat{\rho}_{k_\rho}) \int_0^1 |t^{-1}W(t) - W(1)| |t^{-\hat{\rho}_{k_\rho}} - t^{-\rho}| dt + |\hat{\rho}_{k_\rho} - \rho|(1 - \hat{\rho}_{k_\rho} - \rho) \int_0^1 |t^{-1}W(t) - W(1)| dt \\
& = o_{\mathbb{P}}(1).
\end{aligned}$$

This achieves the proof of Theorem 2. ■

Proof of Corollary 3. According to Theorem 2.1 in Gomes et al. (2002), $\hat{\rho}_{k_\rho}$ is consistent in probability as soon as the intermediate sequence k_ρ satisfies $\sqrt{k_\rho} A\left(\frac{n}{k_\rho}\right) \rightarrow \infty$ and the second order condition (C_{SO}) hold. Combining this result with our Theorem 2, Corollary 3 follows. ■

Proof of Theorem 3. It is equivalent to show the asymptotic normality of

$$\begin{aligned}
\frac{\sqrt{k}}{\log \frac{k}{np}} \log \frac{\hat{x}_{p, \xi}}{x_p} &= \frac{\sqrt{k}}{\log \frac{k}{np}} \left\{ \log X_{n-k, n} + \hat{\gamma}_k(K_{\hat{\Delta}_{opt}^*}) \log \frac{k}{np} - \log x_p \right. \\
&\quad \left. - \frac{(1 - \xi)(1 - 2\xi)}{\xi^2} [\hat{\gamma}_k(K_1) - \hat{\gamma}_k(K_{2, \xi})] \frac{\left(\frac{k}{np}\right)^\xi - 1}{\xi} \right\} \\
&= \sqrt{k} \left(\hat{\gamma}_k(K_{\hat{\Delta}_{opt}^*}) - \gamma \right) + \frac{\sqrt{k}}{\log \frac{k}{np}} \log \frac{Q_n(1)}{U\left(\frac{n}{k}\right)} - \frac{\sqrt{k}}{\log \frac{k}{np}} \left\{ \log \frac{U\left(\frac{1}{p}\right)}{U\left(\frac{n}{k}\right)} - \gamma \log \frac{k}{np} \right\} \\
&\quad - \frac{(1 - \xi)(1 - 2\xi)}{\xi^2} \frac{\sqrt{k} [\hat{\gamma}_k(K_1) - \hat{\gamma}_k(K_{2, \xi})] \frac{\left(\frac{k}{np}\right)^\xi - 1}{\xi}}{\log \frac{k}{np}} \\
&= \sqrt{k} \left(\hat{\gamma}_k(K_{\hat{\Delta}_{opt}^*}) - \gamma \right) + \frac{\sqrt{k}}{\log \frac{k}{np}} \log \frac{Q_n(1)}{U\left(\frac{n}{k}\right)} - \frac{\sqrt{k}}{\log \frac{k}{np}} \tilde{A}\left(\frac{n}{k}\right) \frac{\left(\frac{k}{np}\right)^\rho - 1}{\rho} \\
&\quad - \frac{\sqrt{k}}{\log \frac{k}{np}} \tilde{A}\left(\frac{n}{k}\right) \left\{ \frac{\log U\left(\frac{1}{p}\right) - \log U\left(\frac{n}{k}\right) - \gamma \log \frac{k}{np}}{\tilde{A}\left(\frac{n}{k}\right)} - \frac{\left(\frac{k}{np}\right)^\rho - 1}{\rho} \right\} \\
&\quad - \frac{(1 - \xi)(1 - 2\xi)}{\xi^2} \frac{\sqrt{k} [\hat{\gamma}_k(K_1) - \hat{\gamma}_k(K_{2, \xi})] \frac{\left(\frac{k}{np}\right)^\xi - 1}{\xi}}{\log \frac{k}{np}} \\
&=: T_3 + T_4 - T_5 - T_6 - T_7.
\end{aligned}$$

For this aim, we will study the five terms separately. According to Theorem 2, we have

$$T_3 \xrightarrow{d} \mathcal{N}(0, \mathcal{AV}(K_{\hat{\Delta}_{opt}^*})).$$

Now, according to Proposition 1, we have almost surely

$$\begin{aligned} & \left| \sqrt{k} \log \frac{Q_n(1)}{U(\frac{n}{k})} - \gamma W(1) \right| \\ & \leq \sup_{t \in (0,1]} t^{\frac{1}{2} + \varepsilon} \left| \sqrt{k} \left(\log \frac{Q_n(t)}{U(\frac{n}{k})} + \gamma \log t \right) - \gamma t^{-1} W(t) - \sqrt{k} \tilde{A} \left(\frac{n}{k} \right) \frac{t^{-\rho} - 1}{\rho} \right| = o(1), \end{aligned}$$

from which we deduce that

$$T_4 \xrightarrow{\mathbb{P}} 0.$$

Clearly, under our assumptions, we also have

$$T_5 \longrightarrow 0.$$

Now, according to the inequality (19)

$$\begin{aligned} |T_6| & \leq \frac{\sqrt{k} |\tilde{A}(\frac{n}{k})|}{\log \frac{k}{np}} \left| \frac{\log U(\frac{1}{p}) - \log U(\frac{n}{k}) - \gamma \log \frac{k}{np} - \frac{(\frac{k}{np})^\rho - 1}{\rho}}{\tilde{A}(\frac{n}{k})} \right| \\ & \leq \frac{\sqrt{k} |\tilde{A}(\frac{n}{k})|}{\log \frac{k}{np}} \varepsilon \left(\frac{k}{np} \right)^{\rho + \delta} \\ & = o(1) \end{aligned}$$

for any $0 < \delta < -\rho$.

Finally, to treat T_7 two cases have to be considered: either ξ is a canonical negative value $\tilde{\rho}$ or an estimator $\hat{\rho}$ consistent in probability such that $|\hat{\rho} - \rho| = O_{\mathbb{P}}(n^{-\varepsilon})$ for some $\varepsilon > 0$.

If a canonical negative value $\tilde{\rho}$ is used, then according to Theorem 1, two times applied, we have

$$\sqrt{k} [\hat{\gamma}_k(K_1) - \hat{\gamma}_k(K_{2,\tilde{\rho}})] = O_{\mathbb{P}}(1).$$

This immediately implies that $T_7 = o_{\mathbb{P}}(1)$.

If an estimator $\hat{\rho}$ consistent in probability is used, we consider the decomposition

$$\begin{aligned} T_7 & = \frac{(1 - \hat{\rho})(1 - 2\hat{\rho})}{\hat{\rho}^2} \frac{\sqrt{k} [\hat{\gamma}_k(K_1) - \hat{\gamma}_k(K_{2,\hat{\rho}})]}{\log \frac{k}{np}} \frac{\left(\frac{k}{np} \right)^\rho - 1}{\rho} \\ & \quad + \frac{(1 - \hat{\rho})(1 - 2\hat{\rho})}{\hat{\rho}^2} \frac{\sqrt{k} [\hat{\gamma}_k(K_1) - \hat{\gamma}_k(K_{2,\hat{\rho}})]}{\log \frac{k}{np}} \left\{ \frac{\left(\frac{k}{np} \right)^{\hat{\rho}} - 1}{\hat{\rho}} - \frac{\left(\frac{k}{np} \right)^\rho - 1}{\rho} \right\}. \end{aligned}$$

Note that

$$\begin{aligned} \sqrt{k} [\hat{\gamma}_k(K_1) - \hat{\gamma}_k(K_{2,\hat{\rho}})] & = \sqrt{k} [\hat{\gamma}_k(K_1) - \gamma] - \sqrt{k} [\hat{\gamma}_k(K_{2,\rho}) - \gamma] - \sqrt{k} [\hat{\gamma}_k(K_{2,\hat{\rho}}) - \hat{\gamma}_k(K_{2,\rho})] \\ & = O_{\mathbb{P}}(1), \end{aligned}$$

where we use Theorem 1 for the two first-terms of the right-hand side and the proof of Theorem 2 (term T_2) for the last term. This implies that

$$\begin{aligned} T_7 &= o_{\mathbb{P}}(1) + o_{\mathbb{P}}(1) \left\{ \frac{\left(\frac{k}{np}\right)^{\hat{\rho}} - 1}{\hat{\rho}} - \frac{\left(\frac{k}{np}\right)^{\rho} - 1}{\rho} \right\} \\ &= o_{\mathbb{P}}(1) + o_{\mathbb{P}}(1) \int_1^{k/(np)} s^{\rho-1} [s^{\hat{\rho}-\rho} - 1] ds. \end{aligned}$$

Inspired by de Haan & Rootzén (1993), we study this integral by using the inequality

$$\left| \frac{e^x - 1}{x} - 1 \right| \leq e^{|x|} - 1, \quad \forall x \in \mathbb{R},$$

from which we deduce that

$$\begin{aligned} &\left| \int_1^{k/(np)} s^{\rho-1} [s^{\hat{\rho}-\rho} - 1] ds - (\hat{\rho} - \rho) \int_1^{k/(np)} s^{\rho-1} \log s ds \right| \\ &\leq \left| (\hat{\rho} - \rho) \int_1^{k/(np)} s^{\rho-1} \log s \left\{ \frac{\exp\{(\hat{\rho} - \rho) \log s\} - 1}{(\hat{\rho} - \rho) \log s} - 1 \right\} ds \right| \\ &\leq |\hat{\rho} - \rho| \int_1^{k/(np)} s^{\rho-1} \log s \{ \exp(|\hat{\rho} - \rho| \log s) - 1 \} ds \\ &\leq |\hat{\rho} - \rho| \left\{ \exp\left(|\hat{\rho} - \rho| \log \frac{k}{np}\right) - 1 \right\} \int_1^{k/(np)} s^{\rho-1} \log s ds \\ &= O_{\mathbb{P}}\left((\hat{\rho} - \rho)^2 \log \frac{k}{np} \int_1^{k/(np)} s^{\rho-1} \log s ds \right) \\ &= o_{\mathbb{P}}(1). \end{aligned}$$

This implies that

$$\int_1^{k/(np)} s^{\rho-1} [s^{\hat{\rho}-\rho} - 1] ds = (\hat{\rho} - \rho) \int_1^{k/(np)} s^{\rho-1} \log s ds + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

This entails $T_7 = o_{\mathbb{P}}(1)$ and thus achieving the proof of Theorem 3. ■

Acknowledgements. The authors would like to thank the editor, associate editor and two referees for valuable comments that have improved the paper.

References

Beirlant, J., Dierckx, G., Goegebeur, Y. & Matthys, G. (1999), ‘Tail index estimation and an exponential regression model’, *Extremes* **2**, 177–200.

- Bühlmann, P. (2002), ‘Bootstraps for time series’, *Statistical Science* **17**, 52–72.
- Chavez-Demoulin, V., Embrechts, P. & Sardy, S. (2014), ‘Extreme-quantile tracking for financial time series’, *Journal of Econometrics* **181**, 44–52.
- Christoffersen, P. (1998), ‘Evaluating interval forecasts’, *International Economic Review* **39**, 841–862.
- Cont, R. (2006), Volatility clustering in financial markets: Empirical facts and agent-based models, *in* ‘Long memory in economics’, A Kirman & G Teysiere (eds.): Springer.
- Danielsson, J. (2011), *Financial Risk Forecasting*, Wiley.
- Davison, A. & Hinkley, D. (1997), *Bootstrap Methods and Their Application*, Cambridge University Press.
- de Haan, L. & Ferreira, A. (2006), *Extreme Value Theory. An Introduction*, Springer Series in Operations Research and Financial Engineering, New York.
- de Haan, L., Mercadier, C. & Zhou, C. (2016), ‘Adapting extreme value statistics to financial time series: dealing with bias and serial dependence’, *Finance and Stochastics* **20**, 321–354.
- de Haan, L. & Rootzén, H. (1993), ‘On the estimation of high quantiles’, *Journal of Statistical Planning and Inference* **35**, 1–13.
- Drees, H. (2000), ‘Weighted approximations of tail processes for β -mixing random variables’, *Annals of Applied Probability* **10**, 1274–1301.
- Drees, H. (2003), ‘Extreme quantile estimation for dependent data, with applications to finance’, *Bernoulli* **9**, 617–657.
- Embrechts, P., Koch, E. & Robert, C. (2016), ‘Space-time max-stable models with spectral separability’, *Probability, Analysis and Number Theory. In honour of N.H. Bingham. C.M. Goldie and A. Mijatović (Eds.) Advances in Applied Probability* **48A**, 77–97.
- Feuerverger, A. & Hall, P. (1999), ‘Estimating a tail exponent by modelling departure from a Pareto distribution’, *Annals of Statistics* **27**, 760–781.
- Gardes, L. & Girard, S. (2008), ‘A moving window approach for nonparametric estimation of the conditional tail index’, *Journal of Multivariate Analysis* **99**, 2368–2388.

- Goegebeur, Y. & Guillou, A. (2013), ‘Asymptotically unbiased estimation of the coefficient of tail dependence’, *Scandinavian Journal of Statistics* **40**, 174–189.
- Gomes, M., de Haan, L. & Peng, L. (2002), ‘Semi-parametric estimation of the second order parameter in statistics of extremes’, *Extremes* **5**, 387–414.
- Gomes, M., de Haan, L. & Rodrigues, L. (2008), ‘Tail index estimation for heavy-tailed models: accomodation of bias in weighted log-excesses’, *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **70**, 31–52.
- Hill, B. (1975), ‘A simple general approach to inference about the tail of a distribution’, *Annals of Statistics* **3**, 1163–1174.
- Hsing, T. (1991), ‘On tail index estimation using dependent data’, *Annals of Statistics* **19**, 1547–1569.
- Kupiec, P. (1995), ‘Techniques for verifying the accuracy of risk management models’, *The Journal of Derivatives* **3**, 73–84.
- Liu, H., Shi, J. & Erdem, E. (2013), ‘An integrated wind power forecasting methodology: interval estimation of wind speed, operation probability of wind turbine, and conditional expected wind power output of a wind farm’, *International Journal of Green Energy* **10**, 151–176.
- Lojowska, A., Kurowicka, D., Papaefthymiou, G. & van der Sluis, L. (2010), Advantages of ARMA-GARCH wind speed time series modeling, in ‘IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems’, pp. 83–88.
- Matthys, G., Delafosse, E., Guillou, A. & Beirlant, J. (2004), ‘Estimating catastrophic quantile levels for heavy-tailed distributions’, *Insurance: Mathematics and Economics* **34**, 517–537.
- McNeil, A. & Frey, R. (2000), ‘Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach’, *Journal of Empirical Finance* **7**, 271–300.
- McNeil, A., Frey, R. & Embrechts, P. (2015), *Quantitative Risk Management: Concepts, Techniques and Tools*, Revised Edition. Princeton University Press, Princeton, New Jersey.
- Stărică, C. (1999), On the tail empirical process of solutions of stochastic difference equations, Technical report, Chalmers University and University of Pennsylvania.

- SwissRe (2014), ‘Natural catastrophes and man-made disasters in 2013: large losses from floods and hail; Haiyan hits the Philippines’, *Sigma* **1/2014**. Swiss Re, Zurich.
- Tarullo, D. (2008), Banking on Basel. The future of international financial regulation, Technical report, Peterson Institute for International Economics, Washington, DC.
- Weissman, I. (1978), ‘Estimation of parameters and large quantiles based on the k largest observations’, *Journal of the American Statistical Association* **73**, 812–815.