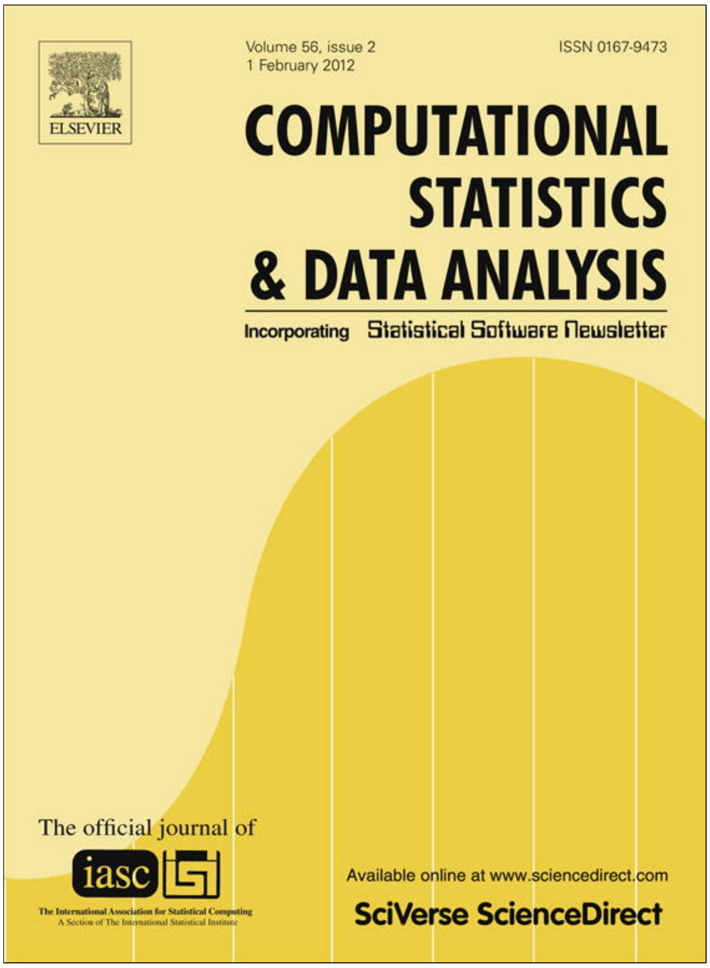
ELSEVIER
Volume 56, issue 2
1 February 2012
ISSN 0167-9473
COMPUTATIONAL STATISTICS & DATA ANALYSIS
Incorporating Statistical Software Newsletter
The official journal of
iasc
The International Association for Statistical Computing
A Section of The International Statistical Institute
Available online at www.sciencedirect.com
SciVerse ScienceDirect

Computational Statistics and Data Analysis 56 (2012) 266–282

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Tail index estimation in the presence of long-memory dynamics

Tucker McElroy [a,*], Agnieszka Jach [b]

[a] Center for Statistical Research and Methodology, US Census Bureau, 4600 Silver Hill Road, Washington, DC 20233-9100, United States
[b] Universidad Carlos III de Madrid, Spain

ARTICLE INFO

Article history:
Received 18 March 2011
Received in revised form 28 July 2011
Accepted 29 July 2011
Available online 7 August 2011

Keywords:
Extreme value theory
Heavy tails
Long-range dependence
Stable distributions

ABSTRACT

Most tail index estimators are formulated under assumptions of weak serial dependence, but nevertheless are applied in practice to long-range dependent time series data. This issue arises because for many time series found in teletraffic and financial econometric applications, both heavy tails and long memory are prevalent features. For a certain class of Heavy-Tail Long-Memory (HTLM) processes, McElroy and Politis (2007a) and Jach et al. (2011) found that the probabilistic behavior of the sample mean depends delicately on the interplay of the tail index and the long memory parameter. In contrast, results in Kulik and Soulier (2011) indicate that the sample quantiles for a related HTLM process are unaffected by long-range dependence. Motivated by these results, we undertake an extensive numerical study to compare the finite-sample performance of several tail index estimators – both those based on sample quantiles, such as the Hill and DEdH (Hill (1975) and Dekkers et al. (1989)) as well as those based on moments, e.g. Meerschaert and Scheffler (1998) – in the HTLM context. Our results largely confirm and expand those of Kulik and Soulier (2011), in that the Hill and DEdH estimators perform well despite the presence of long memory.

Published by Elsevier B.V.

1. Introduction

In this paper, we consider the problem of tail index estimation for time series that have both Heavy Tails (HT) and Long Memory (LM). Both phenomena have been observed throughout the sciences (see Samorodnitsky and Taqqu (1994) pps. 586–590 for an overview). The last several decades have seen a multiplication of papers that address the presence of heavy tails in economic time series—see Embrechts et al. (1997) for a discussion. The older literature on long memory has focused on weakly dependent time series. For example, there is the work of Geweke and Porter-Hudak (1983), Robinson (1995a,b), and many others. More recently there has been some literature addressing the joint presence of heavy tails and long memory: Heyde and Yang (1997), Hall (1997), Rachev and Samorodnitsky (2001), Mansfield et al. (2001), McElroy and Politis (2007a), Kulik and Soulier (2011) and Jach et al. (2011). The paper at hand considers the class of HTLM processes defined in both Kulik and Soulier (2011) and Jach et al. (2011) – which are appropriate for applications in financial econometrics and teletraffic – and provides an extensive simulation study of the performance of leading estimators when long memory is present. We also provide some accompanying theory on asymptotic results for tail index estimators for HTLM processes, as a framework for our numerical results.

By way of background motivation to our study, we note that most tail index estimators have been derived and formulated from an assumption of serial independence. However, if weak serial dependence is present, classical results (see summary treatment in Embrechts et al. (1997) for Hill, Pickands (Pickands, 1975), and DEdH estimators) provide central limit theorems

* Corresponding author. Tel.: +1 301 763 3227; fax: +1 301 763 8399.
E-mail address: tucker.s.mcelroy@census.gov (T. McElroy).

0167-9473/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.csda.2011.07.018

for tail index estimators based on sample quantiles. Intriguingly, these results have been extended to HTLM processes by Kulik and Soulier (2011), with the fascinating outcome that memory has *no* impact on the asymptotic theory (these authors focus on the Hill estimator). This is in marked contrast to results on first and second sample moments for HTLM processes, as explored in McElroy and Politis (2007a) and Jach et al. (2011), where it is found that non-central limit theorems may arise (with non-standard rates of convergence) depending on the interplay between the tail index and the memory parameter. Therefore, it is unclear whether moment-based tail index estimators – such as that of Meerschaert and Scheffler (1998) and McElroy and Politis (2007b) – will still perform well in the presence of long-range dependence. Due to the importance of the issue from a pragmatic data-driven perspective, we have undertaken an extensive numerical investigation to compare several main estimators.

The problem of tail index estimation is most appropriately viewed as semi-parametric. Typically, tail index estimators (such as the Hill and DEdH estimators) have convergence rate depending on a tuning parameter, and omitting proper tuning of the estimator results in poor performance. Fully nonparametric estimators without tuning parameters, such as the moment-based estimator of Meerschaert and Scheffler (1998), will therefore tend to be uncompetitive with the more nuanced semi-parametric methods. Our interest here is not to provide advocacy for nonparametric methods, but to quantify how greatly actual performance differs between rival techniques when long memory is present.

Our main conclusions are that the Hill estimator performs quite well when long memory is present, and the DEdH estimator also does fairly well. The moment-based estimator of Meerschaert and Scheffler (1998), denoted by MS, and the estimator of Stoev et al. (2011), denoted by SMT, fare less well, and yet are not universally outperformed. (The MS and SMT estimators, being nonparametric, cannot be expected to be competitive with the semi-parametric Hill and DEdH estimators.) In summary, both the Hill and DEdH estimators have decent performance despite the presence of long-range dependence, although exact results depend on the precise distributional form of the process and the number of order statistics used. Note that even though we use the term "long memory parameter", both tail index and long memory are defined in a general way compatible with a non-model-based approach to time series analysis.

The rest of the paper is organized as follows. Section 2 introduces the HTLM processes that we study, which are a slight generalization of the HTLM process of Kulik and Soulier (2011), that at the same time encompass the HTLM processes of Jach et al. (2011). We provide some discussion of how these processes may be appropriate for financial and teletraffic applications. Section 3 defines the tail index estimators and summarizes the results of our simulation study. Section 4 provides theoretical results for the estimators; in the case of the quantile-based estimators (i.e., Hill and DEdH), the results are obtained by tedious extensions of the theoretical machinery of Kulik and Soulier (2011), whereas the moment-based MS estimator is derived using the methodology of Jach et al. (2011). Section 5 concludes, and proofs are contained in the Appendix.

2. Heavy-tailed long-memory processes

We now give a definition of HTLM processes; also see McElroy and Politis (2007a), Kulik and Soulier (2011) and Jach et al. (2011). The study of heavy tails and long memory in hydrological time series data was highlighted in Mandelbrot and Wallis (1968), under the terminology of the Noah and Joseph effects, respectively. The first study of long-range dependence for linear processes with stable innovations is Kokoszka and Taqqu (1999), which generalizes the short memory work of Davis and Resnick (1985, 1986). Also see Samorodnitsky and Taqqu (1994) for the stable integral moving average process, and for linear heavy-tailed, long-memory processes. The process considered in this paper has its origin in the Long Memory Stochastic Volatility literature, and is related to the model introduced by Breidt et al. (1998). Let $\{Y_t\}$ be given as the product of a light-tailed serially-dependent process and a heavy-tailed serially-independent process, where the two component processes are independent:

$$Y_t = \sigma(X_t) \cdot Z_t. \tag{2.1}$$

The variables $\{X_t\}$ are a long-memory Gaussian time series, and $\sigma(\cdot)$ is a real-valued function with suitable integrability properties defined below. The series $\{Z_t\}$ is *i.i.d.* with common cumulative distribution function (cdf) F_Z, which has heavy left and right tails of index α

$$\overline{F}_Z(x) = c_+ x^{-\alpha} H(x), \qquad F_Z(-x) = c_- x^{-\alpha} H(x), \tag{2.2}$$

where $\overline{F}_Z(x) = 1 - F_Z(x)$. The variable x is positive, and H is a slowly-varying function—see Embrechts et al. (1997). The parameter $\alpha > 0$ is the tail index. The non-negative constants c_+, c_- give the respective weights of the right and left tails. As for the long-memory portion, we will write $S_t = \sigma(X_t)$. Serial dependence enters into the HTLM process (2.1) through the definition of $\{X_t\}$, which is assumed to be a purely non-deterministic process. In particular, let γ_Y and γ_S denote the autocovariance functions (acfs) of $\{Y_t\}$ and $\{S_t\}$. If $\mathrm{Var}(Z) < \infty$ (i.e., $\alpha > 2$), then

$$\gamma_Y(h) = \mathbb{E}[Z]^2 \gamma_S(h) + \mathbb{E}[S^2]\mathrm{Var}(Z) 1_{\{h=0\}}. \tag{2.3}$$

If $\mathbb{E}[Z] = 0$, then $\{Y_t\}$ is not serially correlated; otherwise $\{S_t\}$ determines the memory structure. The relation of γ_S to γ_X is determined via $\sigma(\cdot)$, according to the Hermite theory expounded in Taqqu (1975) and Taniguchi and Kakizawa (2000). The long-memory property can be defined through the behavior of cumulated autocovariances, noting that the spectral density

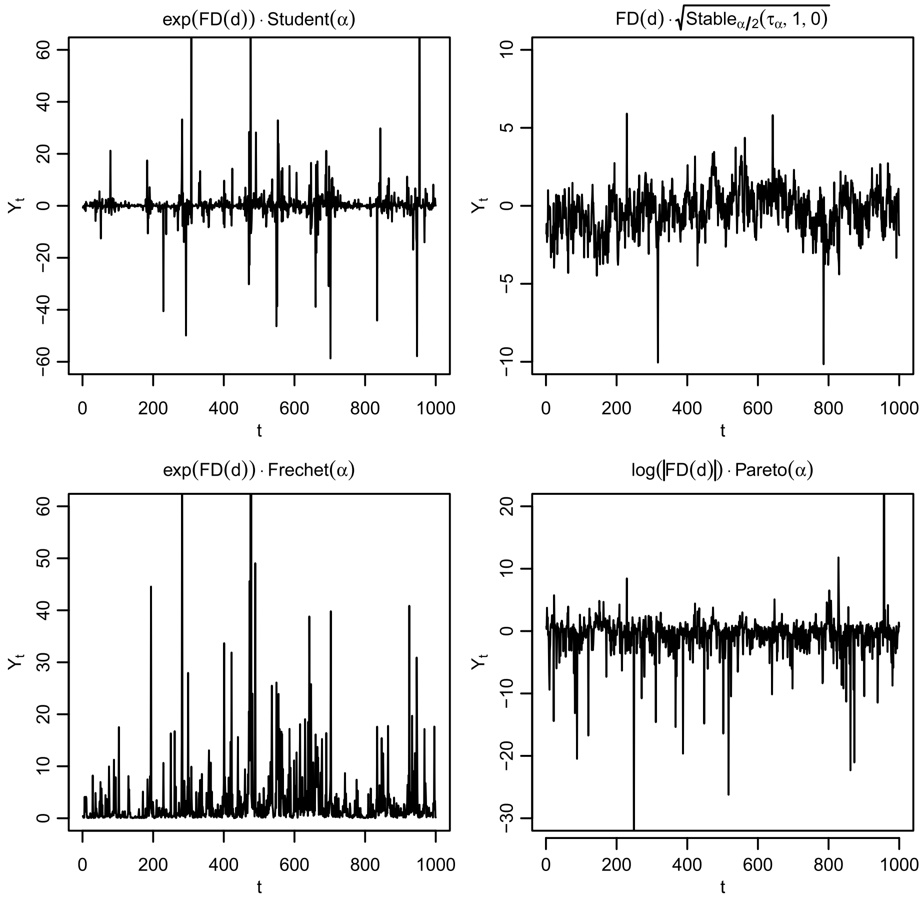

Fig. 2.1. Example of realizations of an HTLM process $Y_t = \sigma(X_t) \cdot Z_t$, where $\{X_t\}$ is a Gaussian fractionally-differenced process with $d = 0.4$ and Z_t has a heavy tail distribution with index $\alpha = 1.9$ (clockwise from top-left they correspond to Processes 4, 1, 6, 7 of Section 3.2). The same realizations of $\{X_t\}$ were used.

is only well-defined when $\mathrm{Var}(Z) < \infty$. We say the Gaussian process $\{X_t\}$ has long memory of parameter β – denoted by LM(β) – if

$$\sum_{0 \le |h| < n} \gamma_X(h) \sim n^{\beta} P(n) \quad \text{and} \quad \sum_{0 \le |h| < n} |\gamma_X(h)| = O(n^{\beta})$$

for some $\beta \in (0, 1)$, and $P(\cdot)$ a slowly-varying function. If the autocovariances are summable, we say the process has short memory, and denote this by LM(0). In terms of how the LM behavior of $\{X_t\}$ is related to that of $\{S_t\}$, the key quantity is the Hermite rank of $\sigma(\cdot)$. Letting this quantity be denoted by q – the definition is given in Taniguchi and Kakizawa (2000) – we know that $\gamma_S(h)$ grows at order $\gamma_X^q(h)$ for large h. The asymptotic results of Kulik and Soulier (2011) and Jach et al. (2011) make assumptions about this Hermite rank, since it affects the distribution theory.

This construction –(2.1), (2.2), and LM(β) – defines the HTLM process of Kulik and Soulier (2011), so long as $\sigma(\cdot) \ge 0$. The positivity assumption of Kulik and Soulier (2011) is motivated by financial applications, but has no real impact on asymptotics (as shown in Section 4). We will allow negative values of $\sigma(\cdot)$ for greater generality. For example, if we restrict Z to have a positive distribution (or more generally, a skewed distribution with nonzero mean), then we obtain the class of processes considered in Jach et al. (2011); these have the form $Y_t = g(X_t) \cdot V_t$, for $\{V_t\}$ *i.i.d.* and heavy-tailed just like $\{Z_t\}$, and with $g(X_t)$ mean zero. Actually, when g is odd, the HTLM process of Jach et al. (2011) is a Kulik–Soulier process (i.e., with $\sigma(\cdot) \ge 0$), since we can define $\sigma(x) = |g(x)|$ and $Z_t = V_t \mathrm{sign}\{g(X_t)\}$ — using the fact that $\mathrm{sign}\{g(X_t)\}$ is independent of $|g(X_t)|$ when $g(\cdot)$ is odd.[1] More generally, when $g(X_t)$ is asymmetric, the two types of HTLM processes are distinct. Nevertheless, both are special cases of the HTLM process set forth here, where $\sigma(\cdot)$ is allowed to take on negative values (Fig. 2.1).

For example, letting $\sigma(\cdot)$ be real-valued and $\{Z_t\}$ positive can allow for asymmetry in the marginal distribution, as well as nonzero serial correlation (2.3). Such a process was used to describe certain kinds of teletraffic data in Jach et al. (2011). On the other hand, if $\sigma(\cdot)$ is non-negative, $\mathbb{E}[Z] = 0$ and $\{Z_t\}$ is symmetric, we can get a symmetric process with no serial

[1] If g is odd and X is symmetric, $\mathbb{P}[g(X) < -y] = \mathbb{P}[g(-X) > y] = \mathbb{P}[g(X) > y]$ implies $g(X)$ is symmetric. Then it can be shown that $|g(X)|$ and $\mathrm{sign}\{g(X)\}$ are independent.

correlation—although the long memory will be manifested in the squares; this can be used to describe the log returns of a financial asset (Breidt et al. (1998) and Deo et al. (2007)).

It can be shown via Breiman's Lemma (Breiman, 1965) that Y shares the variational properties of Z given in (2.2). Namely, when $\mathbb{E}[|S|^{\alpha+\delta}] < \infty$ for some $\delta > 0$, then

$$\mathbb{P}[|Y| > x] \sim \mathbb{E}[|S|^{\alpha}] \cdot \mathbb{P}[|Z| > x].$$

3. Methods and results

3.1. Definition of statistical estimators

We suppose that the observed data $Y_1, Y_2, \ldots, Y_n$ is sampled from the HTLM process of Section 2. We consider the MS, Hill, DEdH, Pickands, and SMT estimators, defined as follows. The MS statistic is given by

$$\widehat{\alpha}_{\mathrm{MS}} = 2\frac{\log(n)}{\log\left(\sum_{t=1}^{n} Y_t^2\right)}, \tag{3.1}$$

which is slightly different from the form presented in Meerschaert and Scheffler (1998). If the process has a nonzero mean or location parameter, we assume that it has already been removed (see Jach et al. (2011) for discussion). To define the Hill and DEdH estimators (Embrechts et al., 1997), introduce the log quantile moments

$$H_{(1)}^{k} = \frac{1}{k}\sum_{j=1}^{k}(\log|Y_{j,n}| - \log|Y_{k+1,n}|), \tag{3.2}$$

$$H_{(2)}^{k} = \frac{1}{k}\sum_{j=1}^{k}(\log|Y_{j,n}| - \log|Y_{k+1,n}|)^2, \tag{3.3}$$

where $|Y_{j,n}|$ is the jth largest of $|Y_1|, \ldots, |Y_n|$, and $k = k(n)$ is a bandwidth parameter satisfying $k(n) \to \infty$ and $n/k(n) \to \infty$ as $n \to \infty$. Then the estimators are given by

$$\widehat{\alpha}_{\mathrm{Hill}}^{k} = (H_{(1)}^{k})^{-1} \tag{3.4}$$

$$\widehat{\alpha}_{\mathrm{DEdH}}^{k} = \left(1 + H_{(1)}^{k} + \frac{1}{2}\left(\frac{\left(H_{(1)}^{k}\right)^2}{H_{(2)}^{k}} - 1\right)^{-1}\right)^{-1}. \tag{3.5}$$

We also define the Pickands estimator (Embrechts et al., 1997), which likewise depends on a bandwidth parameter k

$$\widehat{\alpha}_{\mathrm{Pick}}^{k} = \frac{\log(2)}{\log\left(\frac{|Y_{k,n}| - |Y_{2k,n}|}{|Y_{2k,n}| - |Y_{4k,n}|}\right)}. \tag{3.6}$$

Finally, we consider the SMT estimator, which was introduced by Stoev et al. (2011) and is based on the scaling property of the sample maxima. This estimator falls under the general rate-estimation framework of McElroy and Politis (2007b) (the authors actually suggest the use of a maximum as a diverging statistic). Combining the ideas of Stoev et al. (2011) and McElroy and Politis (2007b) leads to the following estimator

$$\widehat{\alpha}_{\mathrm{SMT}} = \frac{\log(n)}{\log\left(\max_{1\le t\le n}|Y_t|\right)}. \tag{3.7}$$

Although our interest is focused on the tail index α, some of these estimators have been historically formulated in terms of the extreme value index ξ (Embrechts et al., 1997), defined as the reciprocal of the tail index, $\xi = 1/\alpha$. All of the above estimators can then be transformed into respective estimates of ξ via $\widehat{\xi} = 1/\widehat{\alpha}$.

3.2. Definition of simulation processes

In the first part of the simulation study, we compare the relative performance of the above estimators via a mean squared error (MSE) criterion. For a long-memory Gaussian process $\{X_t\}$ with parameter $d = \beta/2$, we choose a fractionally differenced process FD(d) (Samorodnitsky and Taqqu, 1994) with $d \in D = \{0, 0.2, 0.4, 0.45\}$. We will often refer to d as a memory parameter, interchangeably with β. Note that d has no impact on the marginal distribution of these processes; when $d = 0$ we have an i.i.d. process.

For the heavy-tailed portion, we consider the four distributions, Stable, Pareto, Fréchet and Student with tail index $\alpha \in \{1.1, 1.2, \dots, 1.9\}$. For smaller values of α, all estimators perform very well, so we focus on the case that $1 < \alpha < 2$. We then generate $R = 1000$ replications of an HTLM(α, d) process $\{X_t\}$ of length $n = 250, 500, 1000$, though keeping the same samples from FD(d) to compare the performance with respect to the heavy-tailed distributions and value of α. The R code to perform these computations and to reproduce the figures is available by request from the authors. We consider the following processes $Y_t = \sigma(X_t) \cdot Z_t$:

1. $\sigma(t) = t, Z_t = \sqrt{\epsilon_t}$, where ϵ_t follows a Stable distribution with heavy tail index $\alpha/2$, scale $\tau_\alpha = (\cos(\pi\alpha/4))^{2/\alpha}$ (see the comment below), skewness 1 and location 0, written $S_{\alpha/2}(\tau_\alpha, 1, 0)$;
2. $\sigma(t) = \exp(t), Z_t \sim S_\alpha(1, 0, 0)$;
3. $\sigma(t) = t, Z_t = \sqrt{\alpha/\epsilon_t}$, where ϵ_t follows a χ^2 distribution with α degrees of freedom, written χ^2_α;
4. $\sigma(t) = \exp(t)$, Z_t follows a Student distribution with α degrees of freedom, written t_α;
5. $\sigma(t) = \exp(t)$, Z_t follows a Pareto distribution with location parameter 1 and shape parameter α, written Pareto(α);
6. $\sigma(t) = \exp(t)$, Z_t follows a Fréchet distribution with shape parameter α, written Fréchet(α);
7. $\sigma(t) = \log(|t|), Z_t \sim$ Pareto(α);
8. $\sigma(t) = \log(|t|), Z_t \sim t_\alpha$.

In Process 1, the function $\sigma(\cdot)$ is real-valued and the sequence $\{Z_t\}$ is positive α-stable, while in Process 2, $\sigma(\cdot)$ is positive and Z_ts are symmetric α-stable. In Processes 3 and 4, the $\sigma(\cdot)$s are as in 1 and 2, but this time the Z_ts come from the Student distribution. Processes 5 and 6 are defined with the same positive $\sigma(\cdot)$, the Z_ts are also positive, but have different distributions, Pareto and Fréchet. Finally, the last two processes are obtained with the same real-valued $\sigma(\cdot)$, however in one the Z_ts are positive and in the other they are symmetrical around 0. Process 6 was used in a numerical example of Kulik and Soulier (2011). For expositional purposes, we will often abuse the notation in figure/table captions, for example writing $\log(|\mathrm{FD}(d)|) \cdot t_\alpha$ to define Process 8.

We can isolate the impact of long memory by comparing performance across simulations from the same class of process, with the same α, by allowing d to vary. Thus the marginal distribution is fixed, and comparing the $d = 0$ case to $d > 0$ allows us to assess the influence of long range dependence. For example, see the panels of Figs. 3.1 and 3.2.

We note here that the MS estimator based on scaled data satisfies $\log\left(\sum_{t=1}^n (cY_t)^2\right)/\log(n) = \log(c^2)/\log(n) + \log\left(\sum_{t=1}^n Y_t^2\right)/\log(n)$, and thus the statistic based on scaled data is offset from the statistic based on unscaled data by the factor $\log(c^2)/\log(n)$, which tends slowly to zero. This presents a potentially substantial finite-sample bias for the MS estimator (and also likewise for the SMT estimator). This deficiency was part of the motivation for the estimators of McElroy and Politis (2007b). Again, we emphasize that the MS and SMT estimators are cruder than the Hill and DEdH, so comparisons are unfair; results should be viewed as a quantification of the expected discrepancy.

3.3. Simulation results

Before summarizing the results of our simulation study, we describe a method for determining the optimal order statistic cut-off k for the Hill and DEdH estimators. The availability of many replications in the simulation study allows us to select the optimal k by minimizing the MSE. For the rth replication of an HTLM(α, d) process with α fixed, we calculate the MSE of the Hill estimator $\mathrm{MSE}^k(d, r)$ for a range of k values $k = k_{\min}, k_{\min} + 1, \dots, k_{\max}$ with $k_{\min} = 15$ and $k_{\max} = n$, and then average over the replications:

$$\overline{\mathrm{MSE}}^k(d) = \frac{1}{R}\sum_{r=1}^{R} \mathrm{MSE}^k(d, r).$$

Then, since according to the theory there should be no memory effect asymptotically, we average over the d's:

$$\overline{\overline{\mathrm{MSE}}}^k = \frac{\sum_{d \in D} \overline{\mathrm{MSE}}^k(d)}{|D|},$$

where $|D|$ is the number of d's used. The winning $k_{\mathrm{opt}}^{\mathrm{MSE}}$ is the order corresponding to the smallest mean MSE

$$\mathrm{MSE}_{\mathrm{opt}} = \min_{k_{\min} \le k \le k_{\max}} \overline{\overline{\mathrm{MSE}}}^k. \tag{3.8}$$

In Figs. 3.1 and 3.2, the MSE of the Hill estimator (averaged over all replications) as a function of k is shown. In all subplots, there is a range of k values minimizing the MSE irrespective of the memory intensity, which corroborates the theory of Kulik and Soulier (2011) and the current paper. When $\alpha = 1.1$, this range covers more upper orders compared to $\alpha = 1.9$. With the exception of Process 1 with $\alpha = 1.9$, $k_{\mathrm{opt}}^{\mathrm{MSE}}$s for $\alpha = 1.9$ are much smaller compared to those for $\alpha = 1.1$, because as such they guarantee that the effect of the bias will be minimized. Hence, looking at the top panels of Fig. 3.1, we would choose k around 300 and 450, respectively. The minimum-MSE rule selects 287 and 447. For the bottom panels $k_{\mathrm{opt}}^{\mathrm{MSE}}$ is 100 and 56.

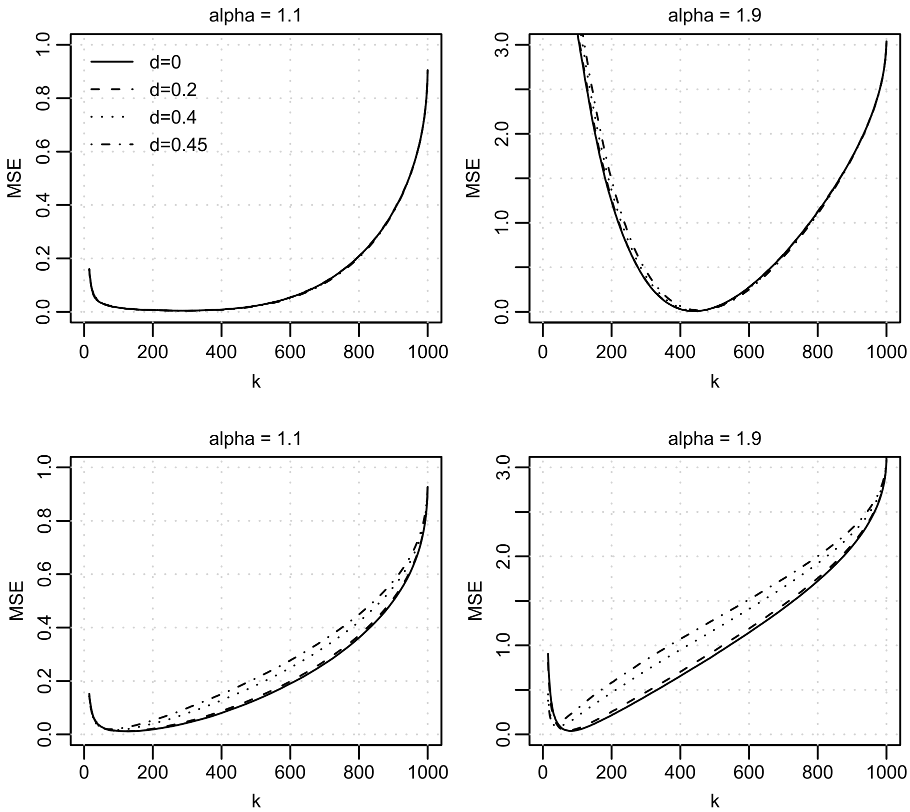

Fig. 3.1. MSE of Hill estimator based on $R = 1000$ replications of length $n = 1000$ from Process 1 ($\mathrm{FD}(d) \cdot \sqrt{S_{\alpha/2}(\tau_\alpha, 1, 0)}$) in the top panel and Process 2 ($\exp(\mathrm{FD}(d)) \cdot S_\alpha(1, 0, 0)$) in the bottom panel, as a function of the upper order k.

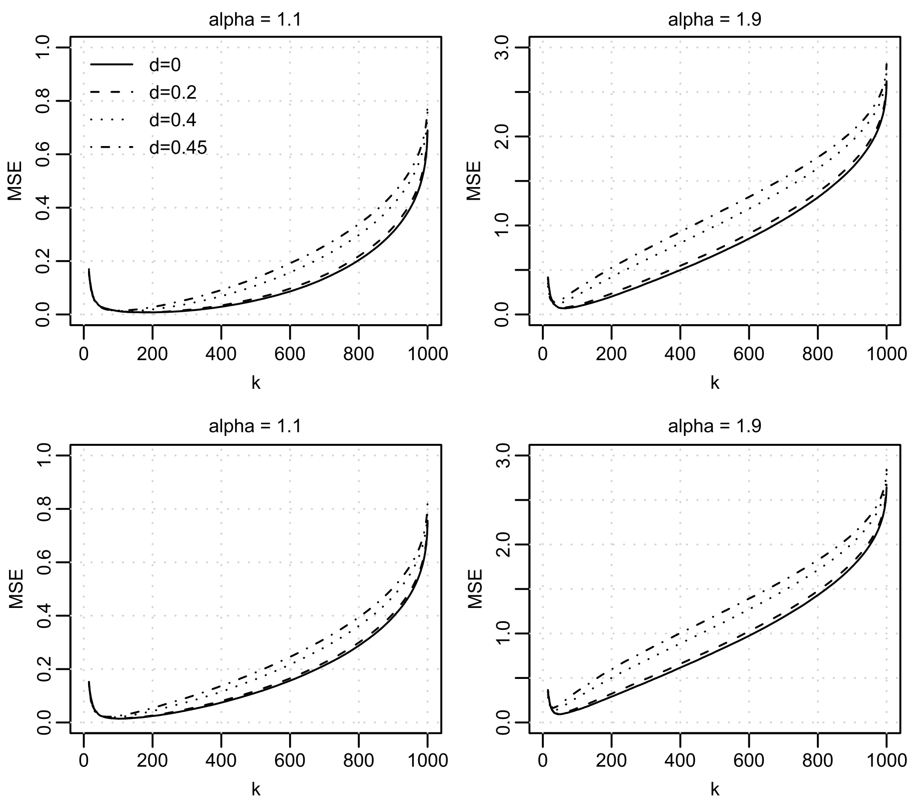

Fig. 3.2. MSE of Hill estimator based on $R = 1000$ replications of length $n = 1000$ from Process 5 ($\exp(\mathrm{FD}(d)) \cdot \mathrm{Pareto}(\alpha)$) in the top panel and Process 6 ($\exp(\mathrm{FD}(d)) \cdot \mathrm{Fréchet}(\alpha)$) in the bottom panel, as a function of the upper order k.

Table 3.1
k_{opt}^{MSE} of Hill and DEdH estimators based on $R = 1000$ replications of length $n = 1000$ corresponding to MSE_{opt}.

Process \ α	Hill					DEdH				
	1.1	1.3	1.5	1.7	1.9	1.1	1.3	1.5	1.7	1.9
1	287	369	412	436	447	551	418	204	140	NA
2	100	86	70	60	56	220	225	246	262	287
3	232	188	147	130	110	440	467	558	556	510
4	87	62	49	36	29	185	170	150	131	147
5	136	98	76	60	45	279	258	245	232	221
6	89	69	56	43	37	192	191	186	176	165
7	221	184	148	120	102	446	459	510	505	500
8	124	88	70	58	43	262	239	238	225	220

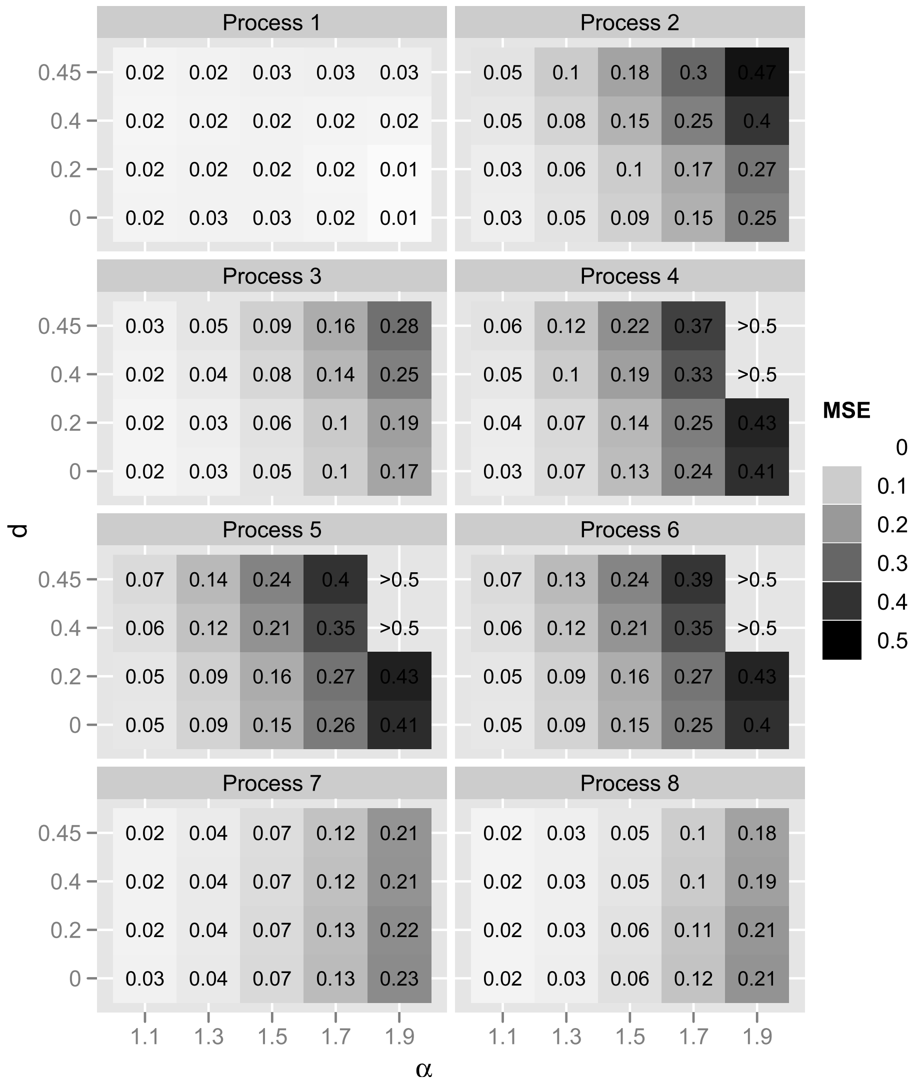

Fig. 3.3. MSE of MS estimator based on $R = 1000$ replications of length $n = 1000$ as a function of α and d for Processes 1–8 (row-wise, from top-left).

The four values for the top of Fig. 3.2 are 136 and 45 and for the bottom, 89 and 37. The complete information on $k_{\text{opt}}^{\text{MSE}}$ in the Hill and DEdH estimators for $n = 1000$ is shown in Table 3.1 ($k_{\text{opt}}^{\text{MSE}}$ for $n = 250, 500$ is available by request from the authors). The number of the upper orders in the Hill estimator decreases with α in all processes except for the first one. The values for the DEdH estimator are generally bigger, however this time $k_{\text{opt}}^{\text{MSE}}$ decreases only in cases 1, 4, 5, 6, 8. Hence compared to the Hill estimator, the tendency in Processes 1, 2, 3, 7 is reversed. In Processes 3, 4 and 7, the tendency is slightly distorted for larger αs.

The MSEs of the four estimators MS, SMT, Hill, DEdH (see Table 3.1 for the optimal k used in Hill and DEdH), as a function of tail index $\alpha \in \{1.1, 1.3, \ldots, 1.9\}$ and memory parameter d for samples of size $n = 1000$, are shown in Figs. 3.3–3.6, one estimator per figure. The values are rounded to two decimals. The Pickands estimator was excluded because of its poor

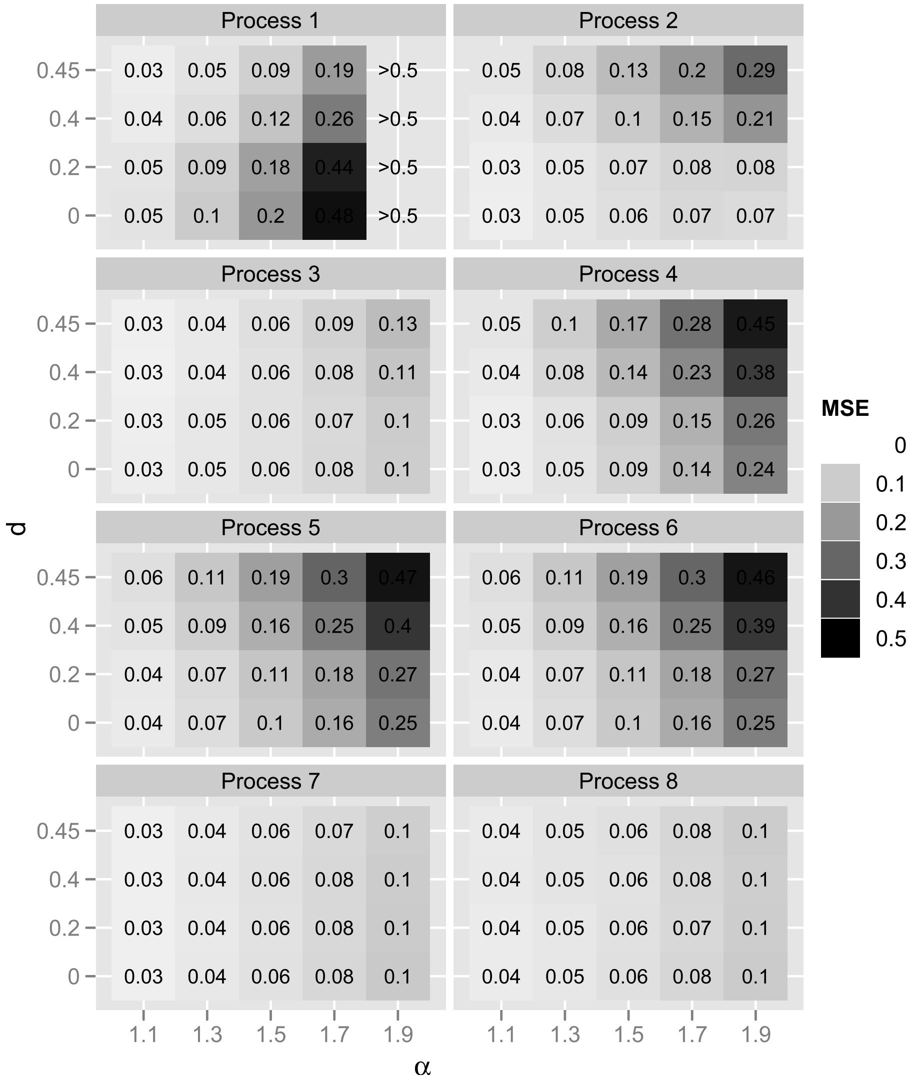

Fig. 3.4. MSE of SMT estimator based on $R = 1000$ replications of length $n = 1000$ as a function of α and d for Processes 1–8 (row-wise, from top-left).

performance. The eight subplots correspond to Processes 1–8 (row-wise from top left). The value of the MSE (rounded to two decimals) is printed on a cross-section of α and d values. For comparison, the color palette in all four figures covers the same range of grays, from 0 (white) to 0.5 (black).

Taking into account all processes and all combinations of α and d, the Hill estimator with $k_{\text{opt}}^{\text{MSE}}$ beats the contenders, and has MSE bounded by approximately 0.2. DEdH (except Process 1) follows closely, and MS and SMT have lower performance. The MS and SMT estimators underperform in Processes 2, 4, 5 and 6 when the values of tail and memory parameters are large. In addition, the SMT estimator behaves poorly in Process 1. DEdH-estimates cannot be computed for Process 1 when α is close to 2 as the estimator breaks down (we are not able to find $k_{\text{opt}}^{\text{MSE}}$). We notice an increasing tendency in the MSEs as the heaviness of the tail decreases; this agrees with the theory (see Section 4). The memory effect on the MSE in finite samples is overall quite small and is more pronounced when α is close to 2. This is manifested by the darker shades in the top-right corners of some of the panels in Figs. 3.3–3.6. As noted by Kulik and Soulier (2011), when the process is extremely heavy-tailed, the effect of long memory is irrelevant. For all estimators, there is almost no long-memory effect in Processes 7 and 8, though it is more pronounced in Processes 5 and 6 (the relative errors between the MSEs for $d = 0.45$ and those for $d = 0$ are around 0.5). In Processes 2 and 4, the memory has the biggest impact of any case. In Process 1, SMT and DEdH estimators are influenced by memory, and in Process 3, the SMT seems to be affected as well. The HTLM processes obtained by specifications 2, 4, 5, and 6 seem to be particularly challenging for all estimators, and thus produce the highest MSEs; this may be due to the exponential form of σ, which generates large variability.

The biases and standard deviations of the estimators MS, SMT, Hill, DEdH as a function of the sample size $n = 250$, 500, 1000, are displayed in Figs. 3.7 and 3.8, respectively. As before, eight subplots correspond to Processes 1–8, and the color palette in both figures covers the same range of grays, from 0 (white) to 0.5 (dark gray). The shading of the negative biases is that of their absolute values (BIASABS in the legend). The values are rounded to two decimals.

Overall, the biases and standard deviations decrease with increasing number of observations. In the MS estimator, except for Process 1, we observe large negative bias (largest in absolute terms compared to other estimators), that is responsible

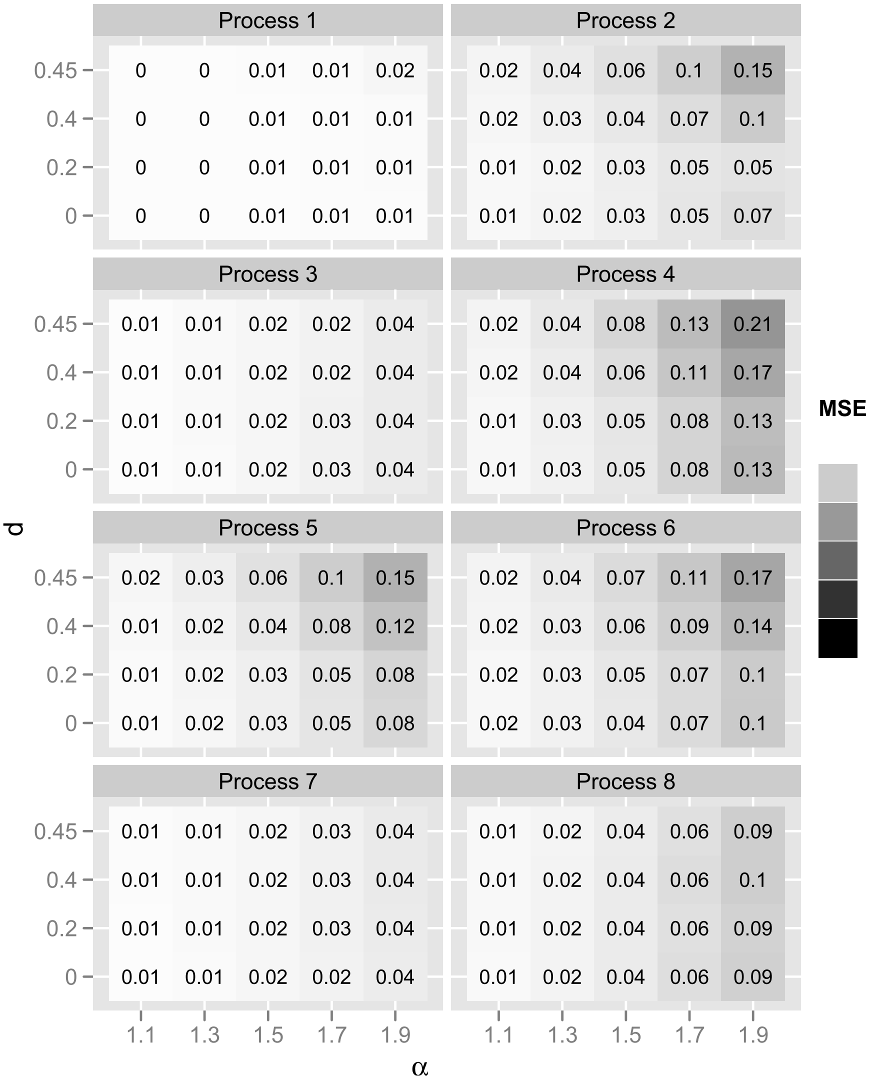

Fig. 3.5. MSE of Hill estimator based on $R = 1000$ replications of length $n = 1000$ as a function of α and d for Processes 1–8 (row-wise, from top-left).

for its poor ranking in terms of the MSE. At the same time, its standard deviation is the smallest among all estimators. The SMT estimator has large negative bias in Processes 2, 4, 5 and 6, which dominates its standard deviation. The bias in the estimate for Process 1 is large and positive; in the remaining processes the MSEs are smaller, but attributable more to the standard deviation than to the bias. The Hill estimates slightly undershoot the true tail index, and thus the size of the MSE is mainly due to the spread. The DEdH estimator for sample size $n = 1000$ has a similar behavior, in terms of bias and standard deviation, to the Hill estimator.

3.4. *Selection of k in practice*

Now we consider the issue of k selection in the Hill estimator based on a single realization, to mirror a practical application. We employ the approach of Drees and Kaufmann (1998) and the practical recommendations therein. The optimal k estimator is given by

$$k_{\text{opt}}^{\text{DK}} = (2\hat{\rho}_n + 1)^{-1/\hat{\rho}_n} (2\tilde{\xi}_n^2 \hat{\rho}_n)^{1/(\hat{\rho}_n+2)} \left(\frac{\bar{k}_n(r_n^\eta)}{(\bar{k}_n(r_n))^\eta} \right)^{1/(1-\eta)}, \tag{3.9}$$

where

$$\hat{\rho}_n = \log \left(\frac{\max\limits_{2 \le i \le [\lambda \bar{k}_n(r_n^\eta)]} \sqrt{i} |\hat{\xi}_{n,i} - \hat{\xi}_{n,[\lambda \bar{k}_n(r_n^\eta)]}|}{\max\limits_{2 \le i \le \bar{k}_n(r_n^\eta)} \sqrt{i} |\hat{\xi}_{n,i} - \hat{\xi}_{n,\bar{k}_n(r_n^\eta)}|} \right) \Bigg/ \log(\lambda) - 0.5$$

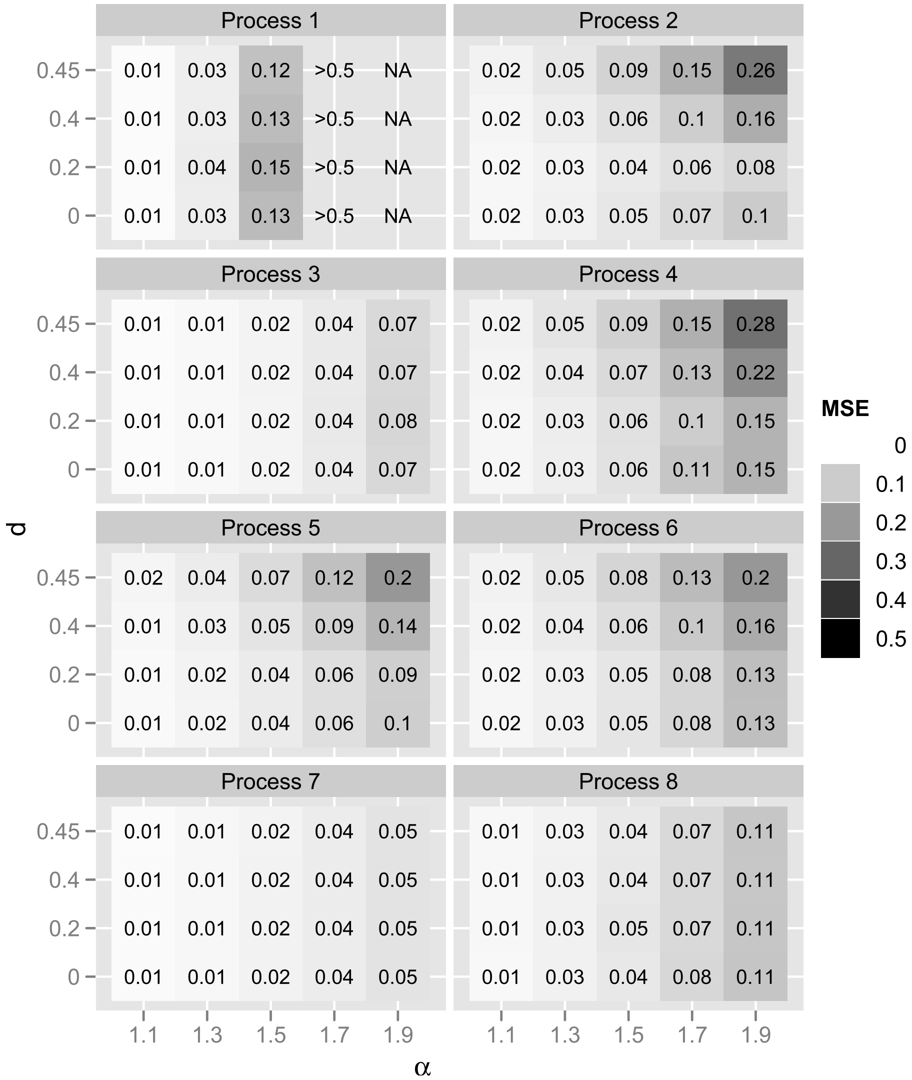

Fig. 3.6. MSE of DEdH estimator based on $R = 1000$ replications of length $n = 1000$ as a function of α and d for Processes 1–8 (row-wise, from top-left).

and the extreme value index Hill estimator $\hat{\xi}_{n,k_n} = H_{(1)}^{k_n}$ (Eq. (3.2)). The so-called stopping time $\bar{k}_n(r_n)$ is defined as

$$\bar{k}_n(r_n) = \min\left\{k \in \{2, \ldots, n\} \mid \max_{2\le i\le k} \sqrt{i}|\hat{\xi}_{n,i} - \hat{\xi}_{n,k}| > r_n\right\}, \tag{3.10}$$

where the random threshold $r_n = 2.5\tilde{\xi}_n n^{0.25}$. The initial estimate $\tilde{\xi}_n$ is equal to $H_{(1)}^{2\sqrt{n^+}}$, where n^+ is the number of positive observations in the sample. The values of the constants η and λ are 0.7 and 0.6, respectively. The interquartile-ranges of $k_{\text{opt}}^{\text{DK}}$ and of the corresponding Hill estimates for Processes 1, 2, 5 and 6 are shown in Figs. 3.9 and 3.10. The rule performs very well with the exception of Process 1 with $\alpha = 1.9$ (this case is also unusual with respect to $k_{\text{opt}}^{\text{MSE}}$, see comments following Table 3.1).

3.5. Illustration on data

Finally, we apply the Hill estimator to two time series that can potentially be described by the HTLM process discussed in this paper. The first series corresponds to the squares of log-returns of the daily adjusted closing prices of the S&P500 index between 25/02/2007 and 10/02/2011. The second series is based on the OctExt Ethernet trace collected in 1994 at the Bellcore Morristown Research and Engineering facility. The time series of packet-counts was obtained by counting the number of packets arriving within consecutive intervals of a fixed length of one second. The Hill plots for both samples of length 1000 are presented in Fig. 3.11. The dotted vertical lines mark $k_{\text{opt}}^{\text{DK}}$, 19 in S&P500 and 18 in OctExt. Corresponding Hill tail estimates are 1.7917 and 2.1209.

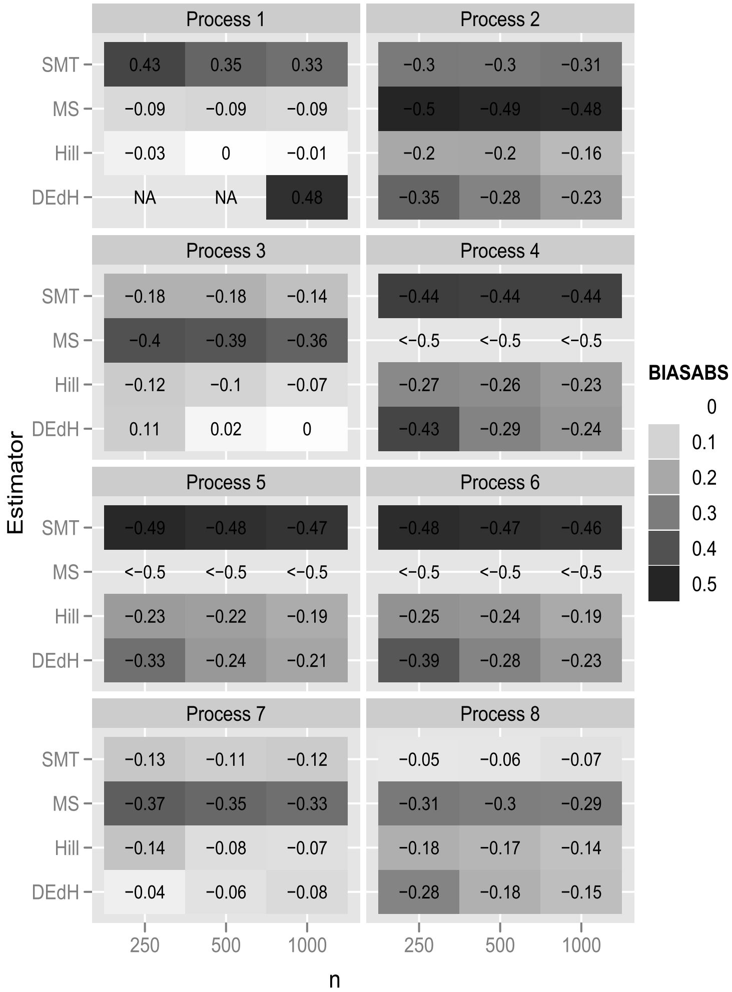

Fig. 3.7. Biases of the four estimators based on $R = 1000$ replications of Processes 1–8 (row-wise, from top-left) with $d = 0.4$ and $\alpha = 1.7$ as a function of the sample size $n = 250, 500, 1000$.

4. Theoretical results

We here provide some asymptotic results for the estimators discussed in Section 3. The poor performance of the Pickands estimator indicates that it is not a real contender, so we focus our efforts on the MS, Hill, DEdH, and SMT estimators of ξ. Recall the definition of $\tau_\alpha = (\cos(\pi\alpha/4))^{2/\alpha}$.

Theorem 1. *Suppose that $\{Y_t\}$ is an HTLM process (2.1) such that $\{X_t\}$ is LM(β) with $\beta \in [0, 1)$. Also suppose that Z_t satisfies (2.2), and $\alpha \in (1, 2)$. Then as $n \to \infty$*

$$\log n \left(\widehat{\xi}_{\mathrm{MS}} - \xi - \frac{\log K(n)}{2 \log n} \right) \overset{\mathcal{L}}{\Longrightarrow} \frac{1}{2} \log W,$$

for some slowly varying function $K(n)$. Here W is $\alpha/2$-stable with scale $\Gamma(1-\alpha/2)^{2/\alpha}\,\tau_\alpha$, skewness 1, and location zero. The variance of the limit variable is $(\xi^2 - 1/4)\pi^2/6$.

This establishes the consistency at rate $\log n$ of the MS estimator, when $K(\cdot) \sim 1$. Otherwise, when $K(\cdot)$ does not converge, there is bias and the convergence rate can be arbitrarily slow; however, the estimator is still consistent because $\log K(n)/\log n$ always tends to zero for any slowly-varying $K(\cdot)$. It is interesting that this general purpose estimator retains its asymptotic properties under long memory. However, it is tailored to the infinite variance case $\alpha < 2$. When $\alpha > 2$, the variance of Y is finite and Theorem 1 is no longer valid; use of the MS statistic will then be inconsistent. In order to handle the case of $\alpha \geq 2$ consistently, one must consider higher-order sample moments, as discussed in McElroy and Politis (2007b)—this is a weakness of the MS statistic. Turning to the quantile-based estimators, we require a second-order condition on the slowly-varying function $H(\cdot)$ in (2.2), which appears in Drees (1998).

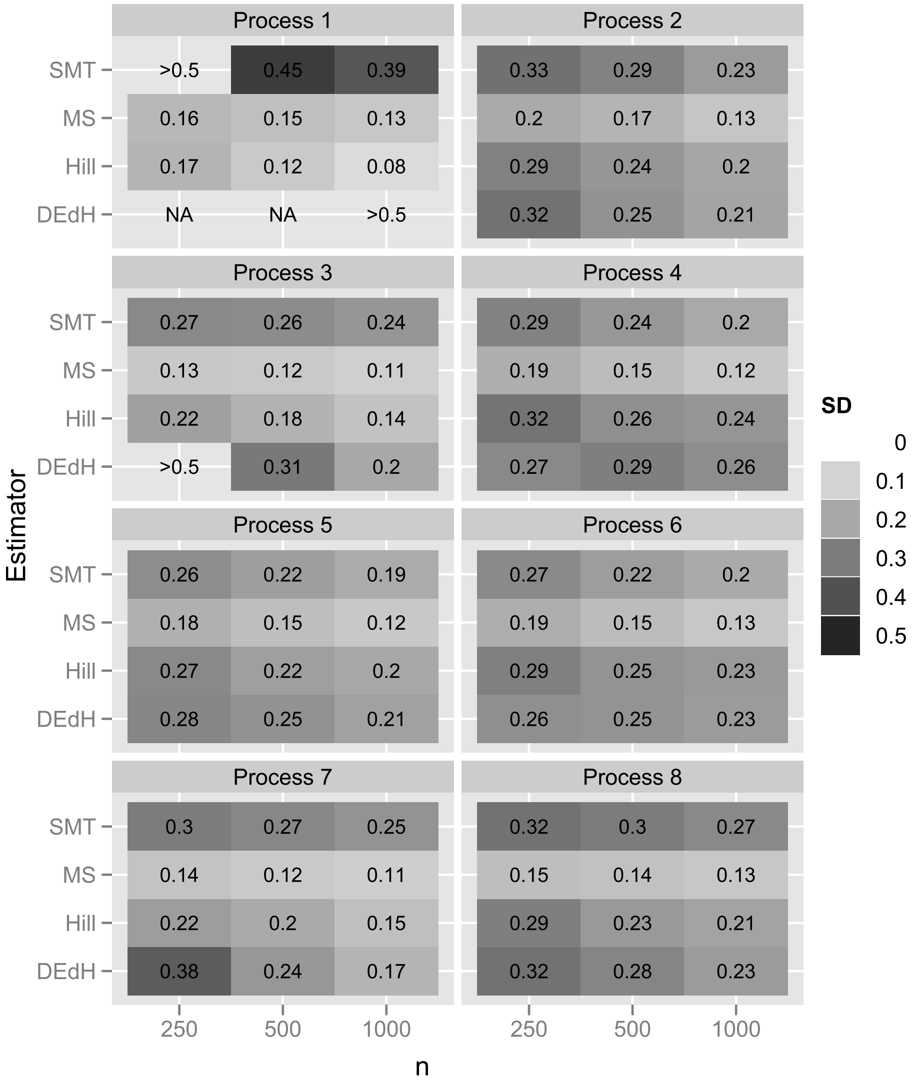

Fig. 3.8. Standard deviations of the four estimators based on $R = 1000$ replications of Processes 1–8 (row-wise, from top-left) with $d = 0.4$ and $\alpha = 1.7$ as a function of the sample size $n = 250, 500, 1000$.

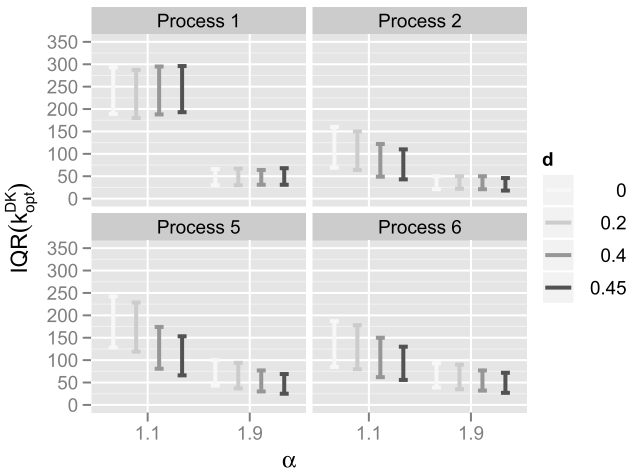

Fig. 3.9. Interquartile range of the upper orders $k_{\text{opt}}^{\text{DK}}$ based on $R = 1000$ replications of length $n = 1000$ from Processes 1, 2, 5 and 6.

Assumption (SO). There exists a bounded non-increasing function $\eta^*(\cdot)$ on $[0, \infty)$, slowly varying at infinity such that $\lim_{t\to\infty} \eta^*(t) = 0$, and there exists a measurable function $\eta(\cdot)$ such that for $z > 0$,

$$H(z) = c \exp \int_1^z \frac{\eta(s)}{s} \mathrm{d}s \tag{4.1}$$

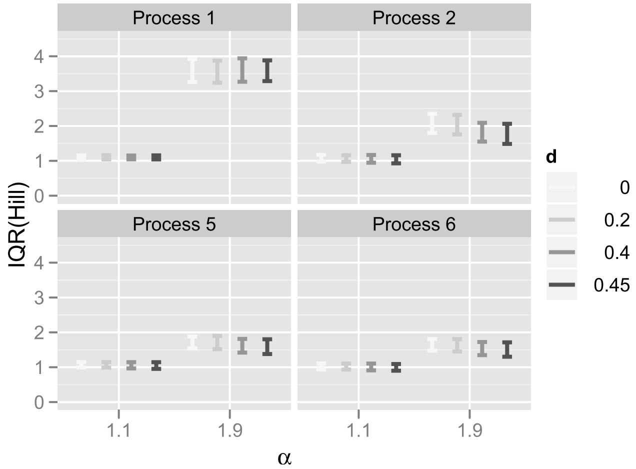

Fig. 3.10. Interquartile range of Hill estimator with $k_{\text{opt}}^{\text{DK}}$ based on $R = 1000$ replications of length $n = 1000$ from Processes 1, 2, 5 and 6.

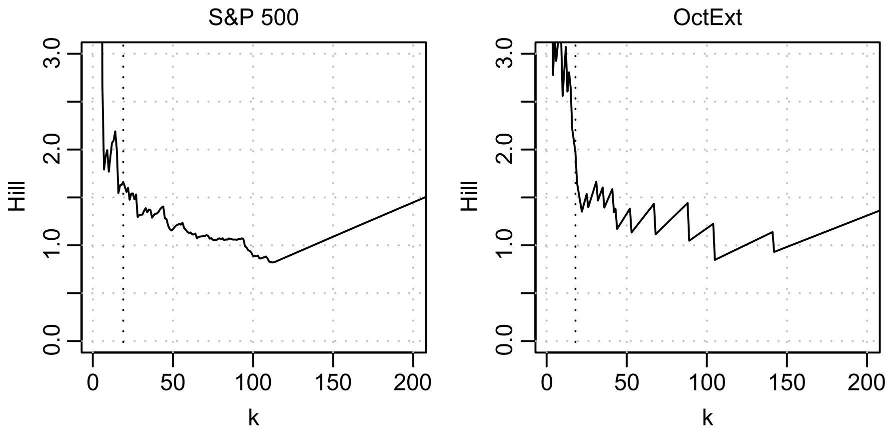

Fig. 3.11. Hill plots of the squared log-returns of S&P500 index and of packet-counts from OctExt teletraffic trace. Vertical lines correspond to orders $k_{\text{opt}}^{\text{DK}}$, 19 in S&P500 and 18 in OctExt.

$$\exists C > 0, \quad \forall s \geq 0, \quad |\eta(s)| \leq C\eta^*(s). \tag{4.2}$$

If (4.1) and (4.2) hold, we will say that $H(\cdot)$ is second order slowly varying with rate function $\eta^*(\cdot)$.

We also require a moment condition on $\sigma(X)$ for our next results, which are similar to those found in Kulik and Soulier (2011). Recall that $S = \sigma(X)$. Suppose that there exists $\epsilon > 0$ such that for a given $\zeta > 0$

$$\mathbb{E}[S^{2\zeta+\epsilon} 1_{\{S>0\}}] < \infty \qquad \mathbb{E}[(-S)^{2\zeta+\epsilon} 1_{\{S<0\}}] < \infty. \tag{4.3}$$

Refer to this as condition $M(\zeta)$, then Eq. (8) of Kulik and Soulier (2011) corresponds to $M(\alpha)$. The variation in Y can be assessed through the following functions:

$$G_n(x, s) = \frac{\mathbb{P}[\sigma(x)Z > (1+s)u_n]}{\mathbb{P}[Y > u_n]}$$

$$G(x) = \frac{c_+\sigma^\alpha(x) 1_{\{\sigma(x)>0\}} + c_-[-\sigma(x)]^\alpha 1_{\{\sigma(x)<0\}}}{c_+^Y}.$$

Then we require that the Hermite rank eventually stabilizes, in the following sense.

Assumption (H). Denote by $J_n(m, s)$, $m \geq 1$, the Hermite coefficients of $G_n(\cdot, s)$ and let $q_n(s)$ be the Hermite rank of $G_n(\cdot, s)$. Define

$$q_n = \inf_{s \geq 0} q_n(s),$$

the Hermite rank of the class of functions $\{G_n(\cdot, s), s \geq 0\}$. Let q be the Hermite rank of G. We assume that $q_n = q$ for n sufficiently large.

Next, define $T(x) = (1+x)^{-\alpha}$ and $\widetilde{B}$ to be a Brownian Bridge process. With $k = k(n)$ the bandwidth for Hill and DEdH, we have the following theorem.

Theorem 2. *Suppose that* $\{Y_t\}$ *is an HTLM process* (2.1) *satisfying* (2.2) *and condition* $M(\alpha(\beta+1))$ *given in* (4.3)*, and is* LM(β)*. Also suppose Assumptions* (H) *and* (SO) *(so that* (4.1) *and* (4.2) *hold), and that the rate function* η^* *in* (SO) *is regularly varying at infinity with index* $-\alpha\beta$*. Finally, suppose that* $\sqrt{k(n)}\eta^*[Q(1-k(n)/n)] \to 0$ *as* $n \to \infty$ *with* Q *the quantile function of the marginal* Y_t*. Then we have the following convergences:*

$$\sqrt{k(n)}\left(\widehat{\xi}^k_{\mathrm{HILL}} - \xi\right) \stackrel{\mathcal{L}}{\Longrightarrow} \int_0^\infty (1+s)^{-1}\widetilde{B}(T(s))\mathrm{d}s$$

$$\sqrt{k(n)}\left(\widehat{\xi}^k_{\mathrm{DEdH}} - \xi\right) \stackrel{\mathcal{L}}{\Longrightarrow} \int_0^\infty \frac{(\alpha^2\log(1+s) + (1-2\alpha))}{1+s}\widetilde{B}(T(s))\mathrm{d}s.$$

These limiting distributions are normal with variance ξ^2 *and* $1+\xi^2$*, respectively.*

This result shows that LM has no impact on the asymptotic behavior. Finally, we consider the SMT estimator.

Theorem 3. *Suppose that* $\{Y_t\}$ *is an HTLM process* (2.1) *such that* $\{X_t\}$ *is* LM(β) *with* $\beta \in [0,1)$*. Also suppose that* Z_t *satisfies* (2.2)*, and* $\alpha \in (1,\infty)$*. Then as* $n \to \infty$

$$\log n\left(\widehat{\xi}_{\mathrm{SMT}} - \xi - \frac{\log K(n)}{\log n}\right) \stackrel{\mathcal{L}}{\Longrightarrow} \log G,$$

for some slowly varying function $K(n)$*. Here* G *is a Fréchet distribution of parameter* α*, and scale* $\mathbb{E}|\sigma(X)|^\alpha$*. The variance of* $\log G$ *is* $\xi^2\pi^2/6$*.*

5. Conclusion

In this paper, we considered the problem of tail index estimation for a broad class of nonlinear, heavy-tailed processes $\{Y_t\}$, $Y_t = \sigma(X_t)\cdot Z_t$, for which the property of long memory can be incorporated into the levels or into the volatility. We considered four estimators – two moment-based, MS and SMT, and two quantile-based, Hill and DEdH – for which the asymptotic theory, pioneered elsewhere, was completed. Somewhat counterintuitively, that theory demonstrates the lack of impact of the long-memory dynamics on the tail index estimation.

Our simulation study led to the conclusion that for well-chosen orders k, the Hill estimator performs very well and has small MSEs for a wide range of processes with various parameter-combinations. The other quantile estimator performs well too in general, but performance breaks down when the marginal distribution of $\{Y_t\}$ is stable symmetric (Process 1). The maximum-based SMT estimator also behaves poorly in this case, but not MS, which has very small MSEs. The performance of the two rate-estimators seems to be undermined by the finite-sample bias (scale-related), and thus it is recommended – if they are used at all – to use their modified versions proposed by McElroy and Politis (2007b). Practitioners should be cautious about using these nonparametric estimators, especially in light of the superior performance of the quantile-based semi-parametric estimators.

Disclaimer

This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the US Census Bureau.

Acknowledgments

A. Jach was partially supported by grants SEJ2007-64500 and CCG10-UAM/ESP-5494.

Appendix

Proof of Theorem 1. We apply Theorem 1 of Jach et al. (2011) to obtain $a_n^{-2}\sum_{t=1}^n Y_t^2 \stackrel{\mathcal{L}}{\Longrightarrow} W$, where $a_n = n^{1/\alpha}K(n)$ is determined by the inverse cdf of Z at $1-1/n$. In the case that Z_t has mean zero, we can still apply the above theorem, since in the proof the analysis of the sample second moment remains true even if $\mathbb{E}[Z]=0$. Let $W_n = a_n^{-2}\sum_{t=1}^n Y_t^2$. By simple algebra

$$\log n\left(\widehat{\xi}_{\mathrm{MS}} - \xi - \frac{\log|K(n)|}{2\log n}\right) = \frac{1}{2}\log W_n,$$

which has the stated convergence. For the variance, first define the following function of a positive real variable t, which is defined for t sufficiently small (in particular, for any $t < \alpha/2$) as $\phi(t) = \mathbb{E}[W^t]$. Since this is differentiable in a small neighborhood of zero, $\dot{\phi}(0) = \mathbb{E}[\log W]$ and $\ddot{\phi}(0) = \mathbb{E}[\log^2 W]$. On the other hand, a formula for $\phi(t)$ is given in Samorodnitsky and Taqqu (1994), and is due to Hardin (1984)

$$\phi(t) = \frac{2^{t-1}\Gamma(1-2t/\alpha)}{\int_0^\infty tx^{-(t+1)}\sin^2 x\mathrm{d}x}(\sec(\alpha\pi/4))^{2t/\alpha}\cos(t\pi/2).$$

The (natural) logarithm yields

$$(t-1)\log 2+\log\Gamma(1-2t/\alpha)-\log\int_0^\infty tx^{-(t+1)}\sin^2 x\mathrm{d}x+\frac{2t}{\alpha}\log\sec(\alpha\pi/4)+\log\cos(t\pi/2).$$

Some elementary real analysis of the integrand in the third term shows that it is integrable; using integration by parts and a change of variable, it can be expressed as

$$\frac{2^t}{2}\int_0^\infty x^{-t}\sin x\mathrm{d}x=2^{t-1}\frac{\Gamma(2-t)\cos(t\pi/2)}{1-t}.$$

Substituting this, we obtain

$$\log\phi(t)=\log\Gamma(1-2t/\alpha)-\log\Gamma(2-t)+\log(1-t)+\frac{2t}{\alpha}\log\sec(\alpha\pi/4).$$

Now a simple calculation shows that

$$\mathrm{Var}[\log W]=\frac{\mathrm{d}^2}{\mathrm{d}t^2}\log\phi(t)|_{t=0}=(\ddot{\Gamma}(1)-\dot{\Gamma}(1)^2)\left(\frac{4}{\alpha^2}-1\right).$$

Recognizing that $\ddot{\Gamma}(1)-\dot{\Gamma}(1)^2=\pi^2/6$ completes the proof. $\square$

Proof of Theorem 2. The result is an extension of Theorems 2.2 and 2.6 of Kulik and Soulier (2011) to the case of a signed $\sigma(\cdot)$ function. We supply some details, although the reader is referred to the above paper for notation. First we derive the variational properties of Y:

$$\mathbb{P}[Y>x]\sim x^{-\alpha}H(x)\left(c_+\int_0^\infty z^\alpha p_S(z)\mathrm{d}z+c_-\int_{-\infty}^0(-z)^\alpha p_S(z)\mathrm{d}z\right)$$
$$\mathbb{P}[Y<-x]\sim x^{-\alpha}H(x)\left(c_-\int_0^\infty z^\alpha p_S(z)\mathrm{d}z+c_+\int_{-\infty}^0(-z)^\alpha p_S(z)\mathrm{d}z\right)$$

as $x\to\infty$, which depends on (4.3). Now the quantities in parentheses can be viewed as the new constants $c_+^X, c_-^X>0$ that weigh the right and left tails of the cdf of Y. Then we can proceed by analyzing the right tail empirical distribution function (edf) defined via

$$\widetilde{T}_n(s)=\frac{1}{n\overline{F}_Y(u_n)}\sum_{t=1}^n 1_{\{Y_t>(1+s)u_n\}} \tag{A.1}$$

for $s\geq 0$. Its mean is the function $T_n(s)=\overline{F}_Y(u_n(1+s))/\overline{F}_Y(u_n)$. The sequence u_n is any that tends to infinity as $n\to\infty$, such that $T_n(s)\to(1+s)^{-\alpha}$, which will be denoted by $T(s)$. However, by the above calculations it is easy to see that

$$T_n(s)\sim\frac{u_n^{-\alpha}(1+s)^{-\alpha}H\left(u_n(1+s)\right)c_+^Y}{u_n^{-\alpha}H(u_n)c_+^Y}$$

for any sequence u_n tending to infinity. Note that we could also define the left tail edf, but we avoid its analysis since the Hill and DEdH estimators are based on upper order statistics. The right empirical process is defined as $e_n(s)=\widetilde{T}_n(s)-T_n(s)$. We can extend Lemma 2.1 of Kulik and Soulier (2011) as follows.

$$\begin{aligned}G_n(x,s)&=\frac{\mathbb{P}\left[Z>\frac{1+s}{\sigma(x)}u_n\right]1_{\{\sigma(x)>0\}}+\mathbb{P}\left[Z<\frac{1+s}{\sigma(x)}u_n\right]1_{\{\sigma(x)<0\}}}{\mathbb{P}[Y>u_n]}\\&\to T(s)\frac{\sigma^\alpha(x)1_{\{\sigma(x)>0\}}+c_-[-\sigma(x)]^\alpha 1_{\{\sigma(x)<0\}}}{c_+^Y}.\end{aligned}$$

So this limit is $T(s)G(x)$. Hence $\mathbb{E}[G(X)]=1$ and

$$\lim_{n\to\infty}\mathbb{E}\left[\sup_{s\geq 0}|G_n(X,s)-G(X)T(s)|^p\right]=0.$$

From here, we can establish the analogues of Eqs. (39) and (40) of Kulik and Soulier (2011), replacing $\sigma^\alpha(\cdot)$ by $G(\cdot)$ and making other appropriate adjustments—throughout the denominator of $\overline{F}_Z(u_n)$ in Kulik and Soulier (2011) must be replaced by $\overline{F}_Y(u_n)$. Then we conclude that

$$\frac{1}{n\overline{F}_Y(u_n)}\sum_{j=1}^n\mathbb{P}\left[\widetilde{\sigma}(X_j)Z>u_n|\mathcal{X}\right]\xrightarrow{P}1,$$

where $\mathcal{X}$ stands for the sigma-algebra generated by $\{X_t\}$. The rest of the "*i.i.d.*" portion of the proof of Theorem 2.2 of Kulik and Soulier (2011) is the same, noting that Lemmas 4.1 and 4.2, and Eq. (59), of Kulik and Soulier (2011) still hold true in our case.

Turning to the "Long-Memory" portion in the proof, we apply Assumption (H); by (4.3) we have $G(\cdot)$ square integrable with respect to Gaussian measure, and the same conclusions hold with straightforward modifications. In the tightness argument,

$$\begin{aligned} Q_n(s) &= \frac{1}{n\overline{F}_Y(u_n)} \sum_{j=1}^{n} \left\{ F_Z\left(-\frac{1+s}{\sigma(X_j)}u_n\right) 1_{\{\sigma(X_j)>0\}} + \overline{F}_Z\left(\frac{1+s}{\sigma(X_j)}u_n\right) 1_{\{\sigma(X_j)<0\}} \right\} \\ &\sim T(s)\frac{1}{n}\sum_{j=1}^{n} G(X_j) \xrightarrow{P} T(s). \end{aligned}$$

The fact that $T(\cdot)$ is continuous and decreasing can now be used in the same fashion as in Kulik and Soulier (2011). Similarly, the long-memory portion can be handled, by adjusting the denominators appropriately. These are the adjustments needed to the proof of Theorem 2.2.

As for the extension of Theorem 2.6 of Kulik and Soulier (2011) – making the replacement of $\overline{F}_Z$ by $\overline{F}_Y$ and σ^α by G in the proof – we obtain the same results. The assumption of (SO) allows us to swap T for T_n in the limit theory, and then we can examine the so-called practical process of Kulik and Soulier (2011). So under the other conditions of our theorem, we get $\sqrt{k(n)}(\widehat{T}_n - T) \overset{\mathcal{L}}{\Longrightarrow} \widetilde{B}\circ T$ as a convergence of stochastic processes, where $k(n) = n\overline{F}_Y(u_n)$ implicitly defines u_n. The quantile log moments are

$$H_n^{(1)} = \int_0^\infty (1+s)^{-1}\widehat{T}_n(s)\mathrm{d}s \qquad H_n^{(2)} = \int_0^\infty \frac{2\log(1+s)}{1+s}\widehat{T}_n(s)\mathrm{d}s.$$

Since $\int_0^\infty (1+s)^{-1}T(s)\mathrm{d}s = \xi$ and $\int_0^\infty \log(1+s)(1+s)^{-1}T(s)\mathrm{d}s = 2\xi^2$, we obtain at once the result for the Hill estimator, whereas for the DEdH we have

$$\begin{aligned} \sqrt{k(n)}\left(\widehat{\xi}^k_{\mathrm{DEdH}} - \xi\right) = &\sqrt{k(n)}\left(H_n^{(1)} - \xi\right) + \frac{\sqrt{k(n)}\left(H_n^{(1)} - \xi\right)\left(H_n^{(1)} + \xi\right)}{\left([H_n^{(1)}]^2 - H_n^{(2)}\right)} - \frac{\sqrt{k(n)}\left(H_n^{(2)} - 2\xi^2\right)}{2\left([H_n^{(1)}]^2 - H_n^{(2)}\right)} \\ &\overset{\mathcal{L}}{\Longrightarrow} \int_0^\infty (1+s)^{-1}\widetilde{B}(T(s))\mathrm{d}s + \frac{2\xi\int_0^\infty (1+s)^{-1}\widetilde{B}(T(s))\mathrm{d}s}{-\xi^2} - \frac{\int_0^\infty \frac{\log(1+s)}{1+s}\widetilde{B}(T(s))\mathrm{d}s}{-\xi^2} \end{aligned}$$

by the functional limit theorem. This last convergence simplifies to the expression stated in Theorem 2; it can be written as

$$\xi\int_0^1 \frac{\widetilde{B}(t)}{t}(1 - 2\alpha - \alpha\log t)\,\mathrm{d}t.$$

Letting $r(s) = 1 - 2\alpha - \alpha\log s$, the variance of the limit is calculated to be

$$\begin{aligned} &\xi^2\int_0^1\int_0^t \left(t^{-1} - 1\right) r(s)\mathrm{d}s\,r(t)\mathrm{d}t + \xi^2\int_0^1\int_t^1 \left(s^{-1} - 1\right) r(s)\mathrm{d}s\,r(t)\mathrm{d}t \\ &= \xi^2\int_0^1 (1-t)r(t)[r(t)+\alpha]\mathrm{d}t + \xi^2\int_0^1 r(t)\left(\mathrm{t}r(t) - (1-2\alpha)(1+\log t) - \alpha(1 - t - \log^2 t/2)\right)\mathrm{d}t \\ &= \xi^2\int_0^1 r(t)\left(\alpha\log^2 t/2 - (1-\alpha)\log t\right)\mathrm{d}t \\ &= \xi^2(1+\alpha^2). \end{aligned}$$

This agrees with the variance stated in Theorem 3.1 of Dekkers and de Haan (1989) and completes the proof. □

Proof of Theorem 3. Since we are working with absolute values, note that $|Y_t| = |\sigma(X_t)|\,|Z_t|$, and denote the sample maximum via the notation $\vee_{t=1}^n |Y_t|$. Let a_n denote a sequence that satisfies $n^{-1} = \mathbb{P}[|Z| > a_n] = (c_+ + c_-)a_n^{-\alpha}H(a_n)$; as in the proof of Theorem 1, we know that $a_n = n^\xi K(n)$ for some slowly-varying $K(\cdot)$. Then using conditioning, for any $x > 0$, we obtain

$$\begin{aligned} \mathbb{P}\left[a_n^{-1}\vee_{t=1}^n |Y_t| > x\right] &= \mathbb{E}\left\{\mathbb{P}\left[a_n^{-1}\vee_{t=1}^n |Y_t| > x|\mathcal{X}\right]\right\} \\ &= 1 - \mathbb{E}\exp\left\{\sum_{t=1}^n \log\left(1 - \mathbb{P}\left[a_n^{-1}|Y_t| > x|\mathcal{X}\right]\right)\right\} \end{aligned}$$

$$= 1 - \mathbb{E}\exp\left\{\sum_{t=1}^{n}\log\left(1 - \mathbb{P}\left[|Z| > a_n x/|\sigma(X_t)|\,|\mathcal{X}\right]\right)\right\}$$

$$= 1 - \mathbb{E}\exp\left\{\sum_{t=1}^{n}\log\left(1 - n^{-1}x^{-\alpha}|\sigma(X_t)|^{\alpha}H\left(a_n x/|\sigma(X_t)|\right)/H(a_n)\right)\right\}.$$

It is easy to see that Lemma 4.1 of Kulik and Soulier (2011) holds under (SO), in the following sense:

$$\forall t \geq 1, \forall z > 0, \quad \left|\frac{H(zt)}{H(t)} - 1\right| \leq C\eta^*(t)z^{\rho}(z \vee z^{-1})^{\epsilon},$$

for some $\rho \leq 0$ and any $\epsilon > 0$; C is a positive constant. This, together with the Taylor series expansion of $\log(1-x)$, produces

$$\mathbb{P}\left[a_n^{-1}\vee_{t=1}^{n}|Y_t| > x\right] \sim 1 - \mathbb{E}\exp\left\{-x^{-\alpha}n^{-1}\sum_{t=1}^{n}|\sigma(X_t)|^{\alpha}\right\}$$

as $n \to \infty$. This is because all higher-order terms in the log expansion are bounded in probability of order $\eta^*(a_n)$, which tends to zero; moreover the function $e^{-x^{-\alpha}}$ produces sufficient integrability to allow application of the Dominated Convergence Theorem. Next, we deduce from the proof of Theorem 5.2.1 of Taniguchi and Kakizawa (2000) that $\mathrm{Var}(\sum_{t=1}^{n}|\sigma(X_t)|^{\alpha})$ is $O(n^{q(\beta-1)})$ if $q(1-\beta) < 1$ and is $O(n)$ otherwise. Here q denotes the Hermite rank of $|\sigma(\cdot)|$. In any event, $n^{-1}\sum_{t=1}^{n}|\sigma(X_t)|^{\alpha} \xrightarrow{P} \mathbb{E}|\sigma(X)|^{\alpha}$, and due to the Dominated Convergence Theorem we can write

$$\mathbb{P}\left[a_n^{-1}\vee_{t=1}^{n}|Y_t| > x\right] \to 1 - \exp\left\{-x^{-\alpha}\mathbb{E}|\sigma(X)|^{\alpha}\right\}.$$

Since this is the upper right tail of the Fréchet distribution, the proof is completed along the lines of the proof of Theorem 1. As for the variance, we easily find that $\mathbb{E}[G^t] = \Gamma(1 - t/\alpha)$ for t in a small neighborhood of zero, from which the stated variance formula follows immediately. □

References

Breidt, F.J., Crato, N., de Lima, P., 1998. The detection and estimation of long memory in stochastic volatility. Journal of Econometrics 83, 325–348.
Breiman, L., 1965. On some limit theorems similar to the arc-sin law. Theory of Probability and its Applications 10, 351–360.
Davis, R.A., Resnick, S.I., 1985. Limit theory for moving averages of random variables with regularly varying tail probabilities. The Annals of Probability 13 (1), 179–195.
Davis, R.A., Resnick, S.I., 1986. Limit theory for the sample covariance and correlation functions of moving averages. The Annals of Statistics 14 (2), 533–558.
Dekkers, A., Einmahl, J., de Haan, L., 1989. A moment estimator for the index of an extreme-value distribution. Annals of Statistics 17, 1833–1855.
Dekkers, A., de Haan, L., 1989. On the estimation of the extreme-value index and large quantile estimation. The Annals of Statistics 17, 1795–1832.
Deo, R., Hsieh, M., Hurvich, C., Soulier, P., 2007. Stochastic volatility models with long memory. In: Doukhan, P., Bertail, P., Soulier, P. (Eds.), Dependence in Probability and Statistics. Springer, NY.
Drees, H., 1998. Optimal rates of convergence for estimates of the extreme value index. The Annals of Statistics 26, 434–448.
Drees, H., Kaufmann, E., 1998. Selecting the optimal sample fraction in univariate extreme value estimation. Stochastic Processes and their Applications 75, 149–172.
Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer-Verlag, Berlin.
Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221–238.
Hall, P., 1997. Defining and measuring long-range dependence. Fields Institute Communications 11, 153–160.
Hardin, C., 1984. Skewed stable variables and processes. Technical Report TR79. University of North Carolina, Center for Stochastic Processes.
Heyde, C.C., Yang, Y., 1997. On defining long-range dependence. Journal of Applied Probability 34, 939–944.
Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. Annals of Statistics 3, 1163–1174.
Jach, A., McElroy, T., Politis, D., 2011. Subsampling inference for the mean of heavy-tailed long memory time series. Journal of Time Series Analysis doi:10.1111/j.1467-9892.2011.00742.x.
Kokoszka, P.S., Taqqu, M.S., 1999. Discrete time parametric models with long memory and infinite variance. Mathematical and Computer Modelling 293, 203–215.
Kulik, R., Soulier, P., 2011. The tail empirical process of some long memory sequences. Stochastic Processes and their Applications 121, 109–134.
Mandelbrot, B.B., Wallis, J.R., 1968. Noah, Joseph and operational hydrology. Water Resources Research 4, 909–918.
Mansfield, P., Rachev, S.T., Samorodnitsky, G., 2001. Long strange segments of a stochastic process and long range dependence. The Annals of Applied Probability 11, 878–921.
McElroy, T., Politis, D.N., 2007a. Self-normalization for heavy-tailed time series with long memory. Statistica Sinica 17 (1), 199–220.
McElroy, T., Politis, D.N., 2007b. Computer-intensive rate estimation, diverging statistics, and scanning. Annals of Statistics 35, 1827–1848.
Meerschaert, M., Scheffler, H., 1998. A simple robust estimator for the thicknes of heavy tails. Journal of Statistical Planning and Inference 71, 19–34.
Pickands, J., 1975. Statistical inference using extreme order statistics. Annals of Statistics 3, 119–131.
Rachev, S.T., Samorodnitsky, G., 2001. Long strange segments in a long range-dependent moving average. Stochastic Processes and their Applications 93, 119–148.
Robinson, P.M., 1995a. Log-periodogram regression of time series with long range dependence. The Annals of Statistics 23, 1048–1072.
Robinson, P.M., 1995b. Gaussian semiparametric estimation of long range dependence. The Annals of Statistics 23, 1630–1661.
Samorodnitsky, G., Taqqu, M.S., 1994. Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Chapman & Hall, NY.
Stoev, S., Michailidis, G., Taqqu, M., 2011. Estimating heavy-tail exponents through max self-similarity. IEEE Transactions on Information Theory 57, 1615–1635.
Taniguchi, M., Kakizawa, Y., 2000. Asymptotic Theory of Statistical Inference for Time Series. Springer-Verlag, NY.
Taqqu, M.S., 1975. Weak convergence to fractional Brownian motion and to the Rosenblatt process. Z. Wahr. Verw. Geb. 50, 53–83.