Subsampling inference for the mean of heavy-tailed long-memory time series

Agnieszka Jach[a], Tucker McElroy[b] and Dimitris N. Politis[c,*,†]

In this article, we revisit a time series model introduced by MCElroy and Politis (2007a) and generalize it in several ways to encompass a wider class of stationary, nonlinear, heavy-tailed time series with long memory. The joint asymptotic distribution for the sample mean and sample variance under the extended model is derived; the associated convergence rates are found to depend crucially on the tail thickness and long memory parameter. A self-normalized sample mean that concurrently captures the tail and memory behaviour, is defined. Its asymptotic distribution is approximated by subsampling without the knowledge of tail or/and memory parameters; a result of independent interest regarding subsampling consistency for certain long-range dependent processes is provided. The subsampling-based confidence intervals for the process mean are shown to have good empirical coverage rates in a simulation study. The influence of block size on the coverage and the performance of a data-driven rule for block size selection are assessed. The methodology is further applied to the series of packet-counts from ethernet traffic traces.

Keywords: Infinite variance; self-normalization; subsampling; weak dependence; adaptive block size.
JEL classifications: C14; C22.

1. INTRODUCTION

This work builds on a theoretical study by McElroy and Politis (2007a), in which the authors presented a new heavy-tailed long-memory (HTLM) time series process and its properties; here, by long memory we mean that the autocovariances are not absolutely summable. The HTLM series $\{\tilde{X}_t\}$ is defined as

$$\tilde{X}_t = \tilde{\eta} + \tilde{\sigma}_t \tilde{G}_t, \qquad (1)$$

where the series $\{\tilde{G}_t\}$ and $\{\tilde{\sigma}_t\}$ are independent and $\tilde{\sigma}_t^2 = \tilde{\epsilon}_t$ where $\{\tilde{\epsilon}_t\}$ are i.i.d. $\alpha/2$-stable, $\alpha \in (1, 2)$, totally positively-skewed, location-zero random variables with scale $(\cos(\pi\alpha/4))^{2/\alpha}$. The non-deterministic Gaussian series $\{\tilde{G}_t\}$ is stationary with long memory $\beta \in (0, 1)$. The attractiveness of this model is its ability to capture both infinite variance and long memory, two properties eagerly sought by modellers, particularly those working with data from the modern computer networks (Willinger et al., 1998; Cappe et al., 2002; Lee and Fapojuwo, 2005). Likewise, the process has finite autocovariances beyond lag zero, which allows long memory to be well defined (in terms of autocovariances). This feature is not shared by two popular infinite-variance long-memory models, namely the α-stable FARIMA and the linear fractional stable noise (Samorodnitsky and Taqqu, 1994).

Nevertheless, the model of McElroy and Politis (2007a) has two somewhat restrictive characteristics: the marginal distribution is assumed to be both symmetric and α-stable. However, these properties can be mitigated by a modification/extension of the basic model, and this is the first contribution of the current article. The asymmetry of paths that can, for example, mimic the behaviour of the series of packet-counts (such as those shown in Figure 1) can be incorporated by requiring that the series $\{\tilde{G}_t\}$ is a function of a long-memory Gaussian series rather than being one itself. On the other hand, the assumption that the marginals are (exactly) α-stable can be relaxed to instead just requiring them to be in the domain of attraction of the stable law.

Our second contribution is concerned with statistical inference for the mean of the extended model. We first investigate the joint asymptotic behaviour of the sample mean and the sample variance. As expected, the associated rates of convergence depend in an intricate way on the tail and long memory parameters. As in McElroy and Politis (2007a), we then focus on a self-normalized sample mean whose rate of convergence is free of the parameters α and β; the normalization is constructed so as to automatically adapt to these two unknown parameters that are tricky to estimate.

The inferential part of our study is completed by constructing confidence intervals for the mean based on subsampling (Politis et al., 1999). The choice of the subsampling methodology is motivated by the fact that, although α and β are eliminated from the rate

[a]Universidad Carlos III de Madrid,
[b]US Census Bureau,
[c]University of California
*Correspondance to: Dimitris N. Politis, Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, Mail Code 0112, La Jolla, CA 92093-0112, USA.
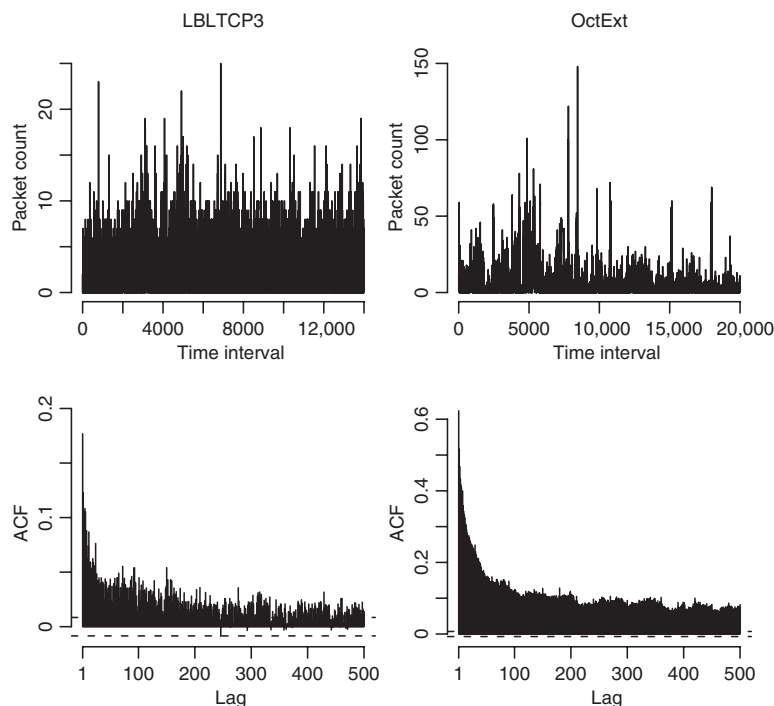†E-mail: dpolitis@ucsd.edu

Figure 1. Two packets counts time series and their autocovariance functions (ACFs). The data are discussed in detail in Section 3

of convergence of the self-normalized sample mean, they are still present (together with other unknown parameters) in the asymptotic distribution. There is no explicit formula for the asymptotic cumulative distribution function (cdf), which precludes us from using it for inference. Our work on the consistency of the subsampling estimator involves the notion of weak dependence (Doukhan and Louhichi, 1999), since the paradigm of strong mixing coefficients is not workable with long memory Gaussian time series. We provide a result of independent interest regarding the consistency of subsampling for self-normalized statistics computed from time series exhibiting the Doukhan and Louhichi (1999) type of dependence, which is weaker than the strong mixing assumption common to the literature.

We conclude our investigation with an in-depth simulation study and an application. For a representative choice of model parameters and the finite-sample settings, we compute the empirical coverage probabilities (CP) for subsampling confidence intervals for the mean. We also devote part of this section to assessing a data-driven rule for selecting the subsampling block size b (Götze and Račkauskas, 2001; Bickel and Sakov, 2008), which is particularly important in the subsequent application to the two series from Figure 1.

The lay-out of the article is as follows. In Section 2.1, we define the extended HTLM model and formulate the limit theorems linked to the self-normalization of the sample mean. The subsampling approach and its justification are described in Section 3, and the empirical study in Section 3. Section 4 contains the summary while the technical proofs are gathered in Appendix A. Appendix B contains a general treatment of subsampling for self-normalized statistics of weakly-dependent time series.

2. THEORETICAL RESULTS

2.1. The general HTLM process

Consider a time series process $\{X_t\}$ of the form

$$X_t = \sigma_t G_t + \eta. \tag{2}$$

We list several assumptions on this process, which generalize the HTLM process of McElroy and Politis (2007a).

ASSUMPTION:
1. The series $\{\sigma_t\}$ and $\{G_t\}$ are independent.
2. The $\{\sigma_t\}$ are i.i.d. random variables with cdf in $D(\alpha)$, i.e. the domain of attraction of an α-stable distribution.
3. The tail index $\alpha \in (1, 2)$ and the mean of σ_t is non-zero.
4. $\{G_t\}$ is a long memory process with parameter $\beta \in (0, 1)$.
5. $\{G_t\}$ is purely non-deterministic with finite variance, and can be written as $G_t = g(V_t)$ for a long memory Gaussian process $\{V_t\}$ and g of unit Hermite rank.

If for convenience we let $Y_t = X_t - \eta$, then Y_t has mean zero. When $\sigma_t^2 = \epsilon_t$ is an $\alpha/2$-stable, totally right-skewed random variable, the marginal distribution of Y_t is symmetric α stable (see McElroy and Politis, 2007a). If $1/\sigma_t^2$ is chi-square distributed with α degrees of freedom, then the marginal distribution of Y_t is student t with α degrees of freedom. More generally, the so-called volatility series σ_t may have any heavy-tailed asymmetric distribution, including the Pareto and Burr distributions, so long as its mean is non-zero (see discussion below).

The second assumption (Ass 2) indicates that there exists a rate $a_n = n^{1/\alpha}K(n)$ with $K \in \mathcal{L}$, where $\mathcal{L}$ denotes the space of slowly-varying functions (Embrechts *et al.* 1997), such that

$$a_n^{-1} \sum_{t=1}^n (\sigma_t - \mu) \stackrel{\mathcal{L}}{\Rightarrow} W,$$

where $\mu = \mathbb{E}\sigma_t$ and W has a (possibly skewed, but centred) α-stable distribution. This rate a_n is unique up to a constant factor, and can be defined by solving $n^{-1} = \mathbb{P}[|\sigma| > a_n]$ for all n (see Propn 2.2.13 of Embrechts *et al.*, 1997); without loss of generality we suppose that $a_n > 0$. Note that since the Hermite rank of g is one by Assumption 5, the mean of G_t is zero, and hence the mean of $\sigma_t G_t$ is zero as well. We assume that $\mu \neq 0$ in Assumption 3; although limiting results can be derived when $\mu = 0$, the asymptotic behaviour is sufficiently different such that self-normalization has to be altered slightly (there is some discussion of this case at the end of the proof of Theorem 1). Our perspective for many applications is that σ_t is positive, or at least right skewed, so that $\mu > 0$ is a reasonable assumption.

The fourth assumption on long memory (Ass 4) is defined more precisely as follows. We first note that the process $\{X_t\}$ has infinite variance (since $\alpha < 2$), but the autocovariances at non-zero lags are perfectly well-defined. Call this sequence $R(h)$ (and informally speaking $R(0) = \infty$) so that $R(h) = \mu^2\gamma(h)$ with $\gamma(h)$ the autocovariance sequence of $\{G_t\}$. Then we assume

$$\sum_{0<|h|<n} R(h) \sim L(n)n^\beta \tag{3}$$

for $L \in \mathcal{L}$ and $\beta \in [0, 1)$. Here the tilde notation is defined by $a_n \sim b_n$ iff $a_n/b_n \to 1$. It is also convenient to assume a growth bound on the absolute sum

$$\sum_{0<|h|<n} |R(h)| = O(L(n)n^\beta). \tag{4}$$

We refer to (3) and (4) together as LM(β) for short, with it being understood that LM(0) denotes the case of absolutely summable autocovariances (at non-zero lags). The rate of partial sum growth in (3) will be denoted by b_n.

The fifth assumption (Ass 5) allows us to generalize the simple Gaussian assumption of McElroy and Politis (2007a) to more complicated distributions, e.g. $g(x) = e^x$ and $g(x) = \log|x|$. This is useful, since it allows us to model asymmetric data. We do restrict to unit Hermite rank on g, since otherwise the Hermite rank can play a determining role in the convergence rate of the sample first and second moments [see Taqqu (1975) for the non-heavy-tailed case] and thereby ruins the application of self-normalization without additional unverifiable conditions.

2.2. Self-normalization and limit theorems

Now we focus on the statistical inference (confidence intervals) for the unknown mean η. The correct normalization for $n(\bar{X}_n - \eta)$ is the maximum of a_n and $d_n = \sqrt{n|b_n|}$ and as such requires the knowledge of α and β. Moreover, depending on whether a_n dominates d_n or vice versa, or whether the two rates are equivalent, the normalization produces three different limits. To avoid the cumbersome estimation of α and β we can employ the self-normalization of $n(\bar{X}_n - \eta)$ as suggested by McElroy and Politis (2007a). The objective of self-normalization is to come up with a quantity that grows at a rate involving $c_n = \max\{a_n, d_n\}$, irrespective of the interplay between a_n and d_n, which eventually will facilitate the cancellation of c_n. This is achieved by combining the sample variance of $\{X_t\}$ and a long memory estimator $\widehat{\mathrm{LM}}_n(\rho)$ from McElroy and Politis (2007a). The growth rate of the former is always a_n^2/n and that of the latter is always d_n^2/n. The resulting normalizing statistic is

$$\hat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{t=1}^n (X_t - \bar{X}_n)^2 + \widehat{\mathrm{LM}}_n(\rho)},$$

where

$$\widehat{\mathrm{LM}}_n(\rho) = \left|\sum_{|h|=1}^{\lfloor n^\rho \rfloor} \frac{1}{n-|h|}\sum_{t=1}^{n-|h|}(X_t X_{t+h} - \bar{X}_n^2)\right|^{1/\rho}, \quad \rho \in (0, 1).$$

The self-normalized statistic of McElroy and Politis (2007a) is then the ratio of $(n/c_n)(\bar{X}_n - \eta)$ and $(\sqrt{n}/c_n)\hat{\sigma}_n$, namely

$$T_n = \sqrt{n}\frac{(\bar{X}_n - \eta)}{\hat{\sigma}_n},$$

which is free of the troublesome rate c_n. Next we provide formal justification for why the sample variance of $\{X_t\}$ and $\widehat{LM}_n(\rho)$ are suitable for the self-normalization of the sample mean, and conclude this section with the main limit theorem.

Under some mild regularity conditions on the distribution of ϵ_t, we have the following joint convergence result. It is necessary that either a_n or d_n dominates the other asymptotically in the definition of c_n, or as a third option that a_n/d_n is asymptotic to a constant; but we rule out the case that the ratio oscillates indefinitely. In other words, we assume that $\lim_{n\to\infty} nb_n/a_n^2$ either diverges to infinity or converges to a finite quantity. The limiting quantity – which potentially could be infinite – will be referred to as A, and this assumption will be referred to as the *Balanced Condition*. As discussed in Fitzsimmons and McElroy (2010), weak convergence is characterized by the joint Fourier–Laplace transform for first and second sample moments. Here we also introduce the notation $x^{<\alpha>} = sign(x)|x|^\alpha$ and $C_\alpha^{-1} = \Gamma(1-\alpha)\cos(\pi\alpha/2)$.

THEOREM 1. *Assume the model Assumptions 1 through 6, and $LM(\beta)$ for $\beta \in [0, 1)$. Under the balanced condition, let $A = \lim_{n\to\infty} nb_n/a_n^2$, with $A \in [0, \infty]$. Then we distinguish three cases depending on the value of A:*

$$\left(c_n^{-1}\sum_{t=1}^n Y_t,\ c_n^{-2}\sum_{t=1}^n Y_t^2\right) \overset{\mathcal{L}}{\Rightarrow} \begin{cases} (S, U) & \text{if } A = 0 \\ (V, 0) & \text{if } A = \infty \\ (S+V, U) & \text{if } A \in (0, \infty) \end{cases}$$

as $n \to \infty$. Here S is an α-stable random variable with scale $C_\alpha^{-1/\alpha}(\mathbb{E}|G|^\alpha)^{1/\alpha}\min\{A^{-1/2}, 1\}$, mean-zero and skewness $\mathbb{E}G^{<\alpha>}/\mathbb{E}|G|^\alpha$. Also V is a mean-zero Gaussian random variable with variance $\mu^2\min\{A,1\}/(\beta+1)$, and U is an $\alpha/2$-stable with location parameter zero, skewness one, and scale equal to $C_{\alpha/2}^{-2/\alpha}(\mathbb{E}|G|^\alpha)^{2/\alpha}\min\{A^{-1}, 1\}$. Although V is independent of S and U, these latter variables are dependent. The joint Fourier–Laplace transform of $S + V$, U is (θ real, $\phi > 0$)

$$\mathbb{E}[\exp\{i\theta(S+V) - \phi U\}]$$
$$= \exp\left\{-C_\alpha^{-1}\mathbb{E}|G(\theta + \sqrt{2\phi}N)|^\alpha \min\{A^{-\alpha/2}, 1\}\left(1 - i(r-l)\frac{\mathbb{E}(G(\theta+\sqrt{2\phi}N))^{<\alpha>}}{\mathbb{E}|G(\theta+\sqrt{2\phi}N)|^\alpha}\tan(\pi\alpha/2)\right)1_{\{A<\infty\}}\right\}$$
$$\times \exp\left\{-\frac{\theta^2\mu^2\min\{A,1\}}{2(\beta+1)}1_{\{A>0\}}\right\}.$$

Here N has a standard normal distribution, independent of G, which has the same distribution as the marginal of $\{G_t\}$. Also $r, l > 0$ are constants giving the relative mass of the tails of the distribution of σ.

We can utilize the statistic $\widehat{LM}_n(\rho)$ as the second component in $\hat{\sigma}_n^2$, so long as the slowly varying function L in (3) behaves asymptotically like a constant. We next extend Theorem 3 of McElroy and Politis (2007a) to the case that ϵ_t has cdf in $D(\alpha/2)$, making a correction to the error rate as we do so. We say that a sequence of random variables Z_n are $U_P(e_n)$ iff $Z_n = O_P(e_n K(n))$ for any $K \in \mathcal{L}$. (This implies that $Z_n = o_P(e_n/n^\epsilon)$ for any $\epsilon > 0$.) Then we have the following result.

THEOREM 2. *Assume the same conditions as Theorem 1, and let $\rho \in (0, 1)$. Further more suppose the fourth order cumulant function of $\{G_t\}$ is absolutely summable in its three indices. Then*

$$\frac{\widehat{LM}_n^\rho(\rho)}{b_n^\rho} = \frac{L(n^\rho)}{L^\rho(n)} + U_P(n^{\rho(1-\beta)-2+[2/\alpha\max(\beta+1)]}) + U_P(n^{\rho(1-\beta)-\eta}) + U_P(n^{\rho(1-\beta)+1/\alpha-1}),$$

where $\eta = 1/2$ if $\beta < 1/2$ and $\eta = 1 - \beta$ if $\beta \geq 1/2$.

REMARK 1. There are three error terms in the above result that depend on ρ, all of which must tend to zero to obtain consistency. The first term is negligible for any ρ when $\beta + 1 \geq 2/\alpha$, but otherwise the rate ρ should be taken smaller than unity. For the second error term, when $\beta > 1/2$ it is always true that $\rho(1-\beta) < \eta$, but otherwise we must decrease ρ. For the third term, we must have $\rho < (1 - 1/\alpha)/(1 - \beta)$. In order for ρ to satisfy all three conditions simultaneously, it is sufficient that $\rho < 1$ and $\rho < (1 - 1/\alpha)/(1 - \beta)$. A plot of this surface – the maximal admissible ρ as a function of α and β – is provided in Figure 2.

Assuming that ρ is chosen sufficiently small such that the errors in Theorem 2 tend to zero, as described in Remark 1, we conclude that

$$\frac{\widehat{LM}_n(\rho)}{b_n} = o_P(1) + \frac{L^{1/\rho}(n^\rho)}{L(n)}$$

as $n \to \infty$. In order for this ratio to converge, it is necessary that $L^{1/\rho}(n^\rho)/L(n)$ be asymptotic to a constant; call this quantity D_ρ. So long as $D_\rho \in (0, \infty)$, we can proceed to use $\widehat{LM}_n(\rho)$ as an estimator of the long memory dynamics. The following result summarizes our findings.
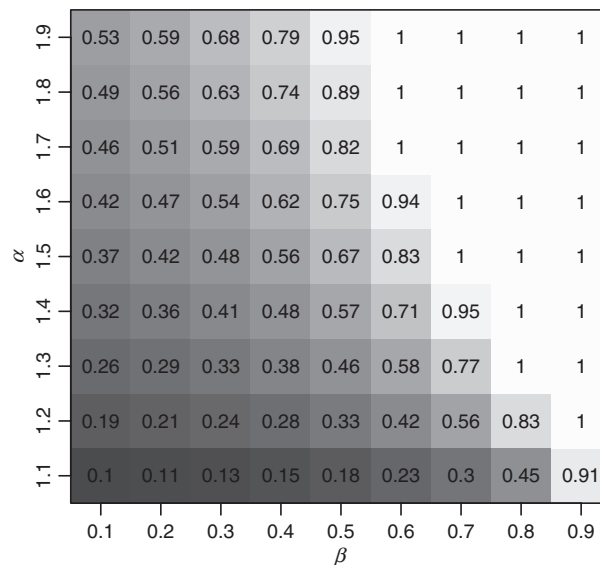
α \ β	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1.9	0.53	0.59	0.68	0.79	0.95	1	1	1	1
1.8	0.49	0.56	0.63	0.74	0.89	1	1	1	1
1.7	0.46	0.51	0.59	0.69	0.82	1	1	1	1
1.6	0.42	0.47	0.54	0.62	0.75	0.94	1	1	1
1.5	0.37	0.42	0.48	0.56	0.67	0.83	1	1	1
1.4	0.32	0.36	0.41	0.48	0.57	0.71	0.95	1	1
1.3	0.26	0.29	0.33	0.38	0.46	0.58	0.77	1	1
1.2	0.19	0.21	0.24	0.28	0.33	0.42	0.56	0.83	1
1.1	0.1	0.11	0.13	0.15	0.18	0.23	0.3	0.45	0.91

Figure 2. Maximal admissible values of $\rho = \min\{(1 - 1/\alpha) / (1 - \beta), 1\}$ (see Rmk 1) as a function of (α, β). Dark shades correspond to small ρs and light to the large ones, with the exact values printed

THEOREM 3. *Assume the same conditions as Theorem 1. Let $A = \lim_{n\to\infty} nb_n/a_n^2$ (assuming the balanced condition) with $A \in [0, \infty]$, and suppose that $D_\rho = \lim_{n\to\infty} L^{1/\rho}(n^\rho)/L(n)$ is a constant in $(0, \infty)$ when $A > 0$. Finally, suppose that ρ is sufficiently small such that the error terms in Theorem 2 vanish. Then we have*

$$\left(c_n^{-1} \sum_{t=1}^{n} (X_t - \eta), c_n^{-2} \sum_{t=1}^{n} (X_t - \bar{X}_n)^2, c_n^{-2} n \widehat{LM}_n(\rho) \right)$$

$$\stackrel{\mathcal{L}}{\Rightarrow} \begin{cases} (S, U, 0) & \text{if } A = 0 \\ (V, 0, D_\rho) & \text{if } A = \infty \\ (S + V, U, \min\{1, A\}D_\rho) & \text{if } A \in (0, \infty) \end{cases}$$

as $n \to \infty$. The definitions of the random variables S, U, V are as in Theorem 1.

As a consequence, the self-normalized statistic of this article will converge to a non-degenerate distribution for any value of $A \in [0, \infty]$, so long as the assumptions hold true. That is,

$$T_n \stackrel{\mathcal{L}}{\Rightarrow} \frac{S 1_{\{A<\infty\}} + V 1_{\{A>0\}}}{\sqrt{U 1_{\{A<\infty\}} + \min\{1, A\}D_\rho}}. \tag{5}$$

2.3. Subsampling

As seen in Theorem 3, the asymptotic distribution of T_n cannot readily be used to approximate the exact sampling distribution of T_n, $L_n(x) = P(T_n \le x)$ (and in particular its quantiles) as it depends on the unknown parameters. That leaves us with two options, either to estimate α and β explicitly (which we deliberately avoided so far) or to approximate the distribution L_n non-parametrically. The first choice calls for two sophisticated (non-existing as far as we are aware) estimation schemes, one to estimate α in the nonlinear, long-memory process and another to estimate β in the nonlinear, heavy-tailed infinite-variance time series. Moreover, even if such schemes were available, the asymptotic distribution of T_n would still need to be approximated via Monte Carlo because it lacks a computable cdf. Therefore, motivated by good numerical results, we selected the subsampling (Politis *et al.*, 1999) methodology, the most general resampling method.

In subsampling we recompute the statistic T_n over short, overlapping subseries of length b – where b is a block-size dependent on n – described via $X_t, X_{t+1}, \ldots, X_{t+b-1}, t = 1, 2, \ldots, n - b + 1$. This gives rise to $n - b + 1$ subsampling statistics

$$T_{n,b,t} = \sqrt{b} \frac{(\bar{X}_{n,b,t} - \bar{X}_n)}{\hat{\sigma}_{n,b,t}},$$

where $\bar{X}_{n,b,t} = \sum_{l=t}^{t+b-1} X_l / b$ and

$$\hat{\sigma}_{n,b,t}^2 = \frac{1}{b} \sum_{l=t}^{t+b-1} (X_l - \bar{X}_{n,b,t})^2 + \widehat{LM}_{n,b,t}(\rho),$$

with

$$\widehat{LM}_{n,b,t}(\rho) = \left| \sum_{|h|=1}^{\lfloor b^\rho \rfloor} \frac{1}{b - |h|} \sum_{l=t}^{t+b-1-|h|} (X_l X_{l+h} - (\bar{X}_{n,b,t})^2) \right|^{1/\rho}, \quad \rho \in (0,1).$$

These statistics in turn are used to approximate the distribution $L_n(x)$ via

$$\hat{L}_{n,b}(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} 1_{\{T_{n,b,t} \leq x\}}.$$

The approximate two-sided $1 - p$ equal-tailed confidence intervals for η is then defined as

$$I_{et;1-p}(\eta) = \left[\bar{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} c_{n,b}(1 - p/2), \bar{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} c_{n,b}(p/2) \right], \tag{6}$$

where $c_{n,b}(1 - p) = \inf\{x : \hat{L}_{n,b}(x) \geq 1 - p\}$ is the (lower) $1 - p$ quantile of the subsampling distribution. Setting

$$\hat{L}_{n,b,|\cdot|}(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} 1_{\{|T_{n,b,t}| \leq x\}}$$

leads to a two-sided symmetric subsampling interval

$$I_{s;1-p}(\eta) = \left[\bar{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} c_{n,b,|\cdot|}(1 - p), \bar{X}_n + \frac{\hat{\sigma}_n}{\sqrt{n}} c_{n,b,|\cdot|}(1 - p) \right], \tag{7}$$

with $c_{n,b,|\cdot|}(1 - p) = \inf\{x : \hat{L}_{n,b,|\cdot|}(x) \geq 1 - p\}$.

The consistency of the subsampling estimator, in the sense that $|\hat{L}_{n,b}(x) - L_n(x)| \xrightarrow{P} 0$ as $n \to \infty$ for all x, for strongly mixing sequences and statistics with a continuous asymptotic distribution is known (Politis et al., 1999), but otherwise requires a non-standard approach. It is known in the long memory literature (see Guégan and Ladoucette, 2001) that Gaussian long memory processes cannot be strong mixing, since this is equivalent to complete regularity – an impossibility for processes with spectral poles by Theorem 3 of Ibragimov and Rozanov (1978). This is in contrast to Theorem 2.2 of Hall et al. (1998), whose assumptions include complete regularity of the long memory Gaussian time series. Note that Nordman and Lahiri (2005) approaches subsampling through a linearity assumption instead.

Therefore we present an alternative route to establishing subsampling consistency, by considering the θ-weak dependence literature of Doukhan and Louhichi (1999) and Bardet et al. (2008). By combining the proofs of Theorem 11.3.1 of Politis et al. (1999) and Lemma 3.1 of Ango-Nze et al. (2003) it is possible to establish the desired consistency rate under assumptions typical to the literature; for completeness we present the result in Appendix B with proof. The application of Theorem 4 to the problem of our article is immediate, using Theorem 3 to identify Z as $S1_{\{A<\infty\}} + V1_{\{A>0\}}$, W as $\sqrt{U1_{\{A<\infty\}}} + \min\{1, A\}D_\rho$, $\alpha_n = n/c_n$, $\delta_n = \sqrt{n}/c_n$, and L the cdf of the limit variable in (5). Then $\alpha_b/\alpha_n \to 0$ and $\tau_b/\tau_n \to 0$, as required by Theorem 4, follows from $b/n \to 0$. Therefore, so as long as the process satisfies the weak-dependence condition – the other conditions are easily seen to be satisfied – and $b/n + 1/b \to 0$ as $n \to \infty$, we are assured of the consistency of the subsampling procedure. The following result gives supplemental conditions on the data process – when σ_t is positive – to ensure λ- and θ-weak dependence [see Doukhan and Louhichi (1999) and Bardet et al. (2008) for definitions].

PROPOSITION 1. *Under the model assumptions of this section and the assumption that g is invertible and Lipschitz continuous, and that σ_t is a positive sequence, the processes $\{G_t\}$ and $\{X_t\}$ are λ-weakly dependent. If the process $\{V_t\}$ is also causal, then $\{G_t\}$ and $\{X_t\}$ are θ-weakly dependent.*

The validity of subsampling under weak dependence is proved in Appendix B under a condition of polynomial decay of the weak dependence coefficients ϵ_r. In the case of our long memory process $\{X_t\}$, the rate of decay of ϵ_r is inherited from the dependence structure of $\{V_t\}$; as in Bardet et al. (2008), this can be a polynomial rate depending on β. In particular, if g is the identity such that $\{G_t\}$ is actually a long-range dependent Gaussian process, the results in Section 3 of Bardet et al. (2008) indicate that a polynomial rate of decay on ϵ_r can be obtained. In particular, a (causal) Gaussian ARFIMA has $\epsilon_r = O(r^{\beta-1})$ – see Doukhan and Louhichi (1999).

3. SIMULATION STUDY AND APPLICATION TO REAL DATA

3.1. Simulations

We now report results of a simulation study of the model with $\eta = 0$, where the series $\{V_t\}$ follows the Gaussian fractionally differenced (FD) process with parameter $d = \beta/2 \in (0, 0.5)$ and standard normal innovations. Its autocovariance function is

$\gamma_{FD}(n) = \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} n^{2d-1}$ and $\gamma_{FD}(0) = \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}$ – see Samorodnitsky and Taqqu (1994). The sample paths of the FD process were generated in package R via the FFT-based algorithm of Stoev and Taqqu (2004).

For the $\{\epsilon_t\}$ series, we first chose $\alpha/2$-stable totally right-skewed i.i.d. random variables with scale parameter $(\cos(\pi\alpha/4))^{2/\alpha}$ and location 0, and for the function g we set $g(x) = \exp(x) - \exp(0.5)$. As a second selection, we chose Pareto random variables with shape parameter $\alpha/2$ and location 1 and $g(x) = x$.

We generated $R = 500$ realizations of length $n = 500, 1000$ for each of the 16 combinations (α, d), $\alpha \in \{1.2, 1.4, 1.6, 1.8\}$ and $d \in \{0.1, 0.2, 0.3, 0.4\}$. To assess the influence of $\rho \in \{0.6, 0.7, 0.8\}$ we used the same realizations of $\{X_t\}$ for a given pair (α, d).

For $n = 500$, we considered $b \in \{16, 21, 28, 38, 50, 67, 89, 119, 158, 211\}$, that is, we followed the formula $[q^j n]$ (motivated by the discussion given below) with $q = 0.75$ and $j = j_{max}, j_{max} - 1, \ldots, j_{min}$ with $j_{max} = [\log_q(15/n)]$, $j_{min} = 3$. This covered the subsample size from around 3–42% of the sample size. For $n = 1000$ we set $j_{max} = [\log_q(30/n)]$, $j_{min} = 3$ to obtain the same ratios of b/n.

To automatically choose the subsampling block size b, we implemented the adaptive block selection algorithm of Götze and Račkauskas (2001) and Bickel and Sakov (2008) – henceforth GRBS – with Kolmogorov-Smirnov (KS) distance. This scheme has been considered in the m-out-of-n bootstrap for the i.i.d data, but gave satisfactory results in subsampling of our stationary time series. The rule consists of the following steps:

- Consider bs, $b_j = [q^j n]$, $j = 1, 2, \ldots, 0 < q < 1$.
- For each b_j, compute $\hat{L}_{n,b_j}$.
- Let δ be the KS distance and set

$$\hat{b} = \arg\min_{b_j} \delta(\hat{L}_{n,b_j}, \hat{L}_{n,b_{j+1}}).$$

- The estimator of L_n is $\hat{L}_{n,\hat{b}}$.

Heuristics behind the GRBS rule are as follows. If the block size is within the 'right' range of b values, then the corresponding empirical distributions should be 'close' to each other, and therefore values of b associated with small KS distances will be selected by the rule.

Subsampling confidence intervals, equal-tailed (6) and symmetric (7), for η at 95% nominal level were constructed in each case. Figures 3–6 display the CP in the stable case for $n = 500$. The subsample size b is shown on the lower axis ('a' stands for adaptive) and the type of the interval, equal-tailed or symmetric, on the upper. Different plotting characters correspond to different ρs as indicated by the legend, and the solid horizontal line marks the nominal level of 0.95. The remaining CPs, for a particular case of $\rho = 0.7$ and adaptively-selected block size, are given in Table 1.

In general, the empirical CP are fairly good and range from around 70% up to around 95%. The undercoverage seems to be coming from the subsampling and not from the adaptive GRBS rule, which tends to pick up the optimal subsample size. We see that
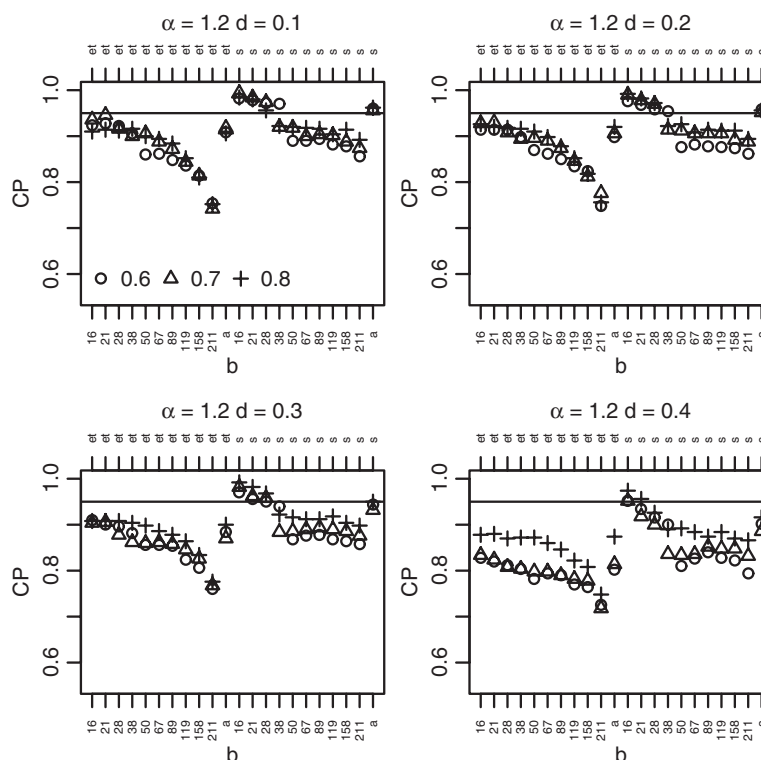

Figure 3. Empirical coverage probabilities (CP) for the 95% subsampling confidence intervals for η based on $R = 500$ replications of length $n = 500$ when $\epsilon_t \sim \alpha/2$-stable, $\alpha = 1.2$, value of ρ indicated by the legend
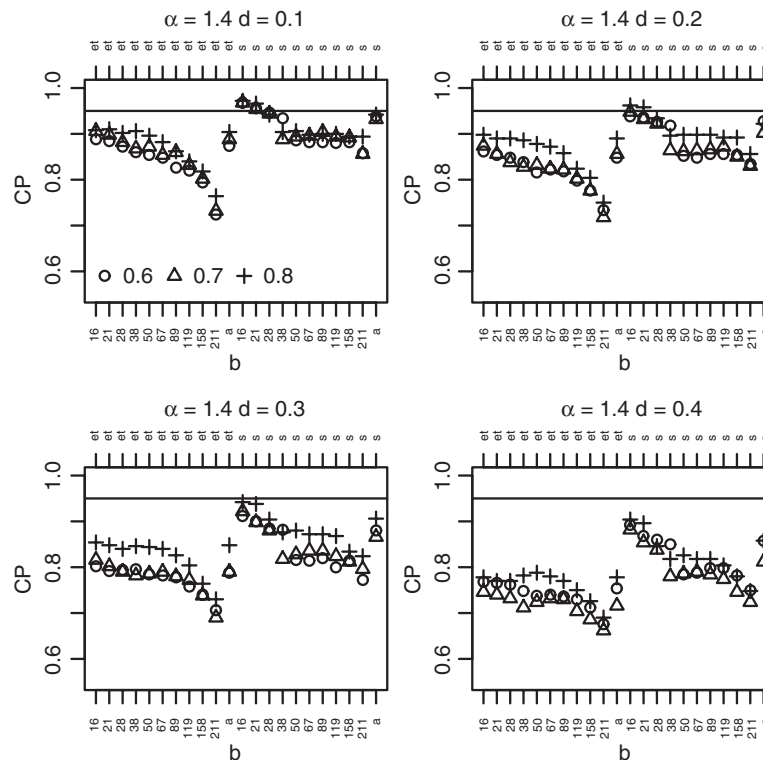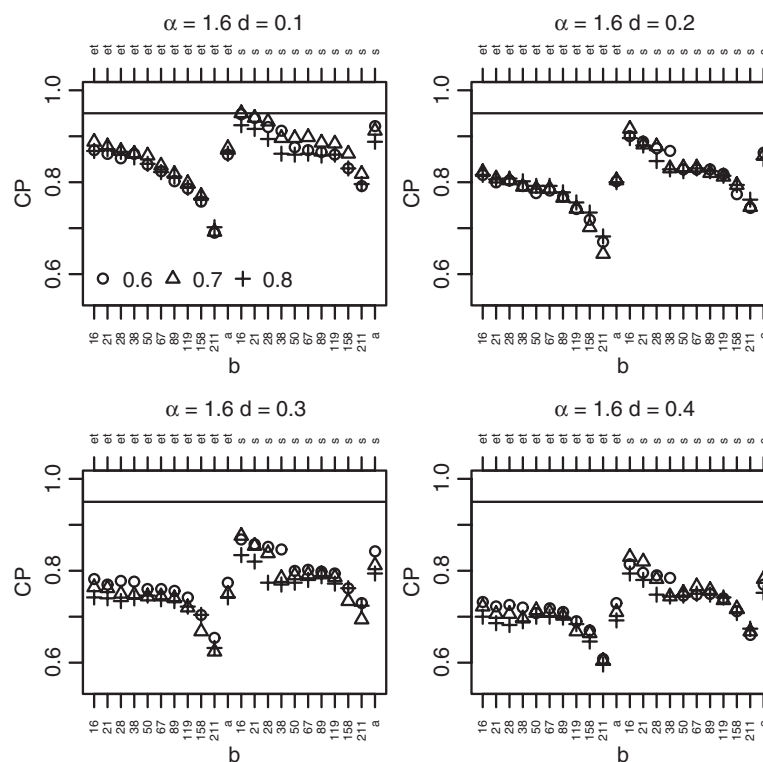
Figure 4. Empirical coverage probabilitie (CP) for the 95% subsampling confidence intervals for η based on $R = 500$ replications of length $n = 500$ when $\epsilon_t \sim \alpha/2$-stable, $\alpha = 1.4$, value of ρ indicated by the legend

Figure 5. Empirical coverage probabilities (CP) for the 95% subsampling confidence intervals for η based on $R = 500$ replications of length $n = 500$ when $\epsilon_t \sim \alpha/2$-stable, $\alpha = 1.6$, value of ρ indicated by the legend
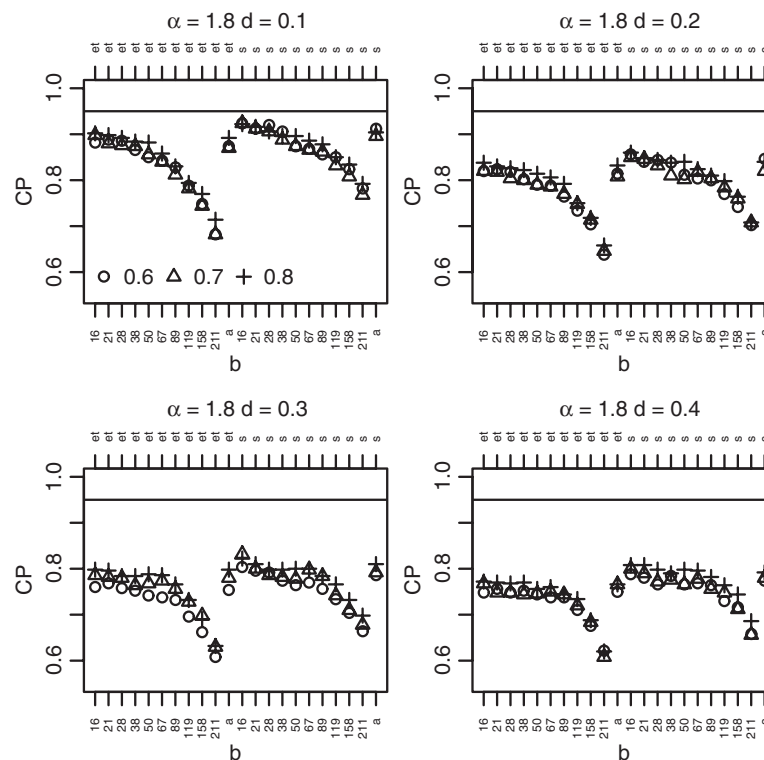
Figure 6. Empirical coverage probabilities (CP) for the 95% subsampling confidence intervals for η based on $R = 500$ replications of length $n = 500$ when $\epsilon_t \sim \alpha/2$-stable, $\alpha = 1.8$, value of ρ indicated by the legend

Table 1. Empirical coverage probabilities for the 95% subsampling confidence intervals for η based on $R = 500$ replications with adaptively selected block size and $\rho = 0.7$

	Equal-tailed				Symmetric			
$\alpha\backslash d$	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
	$\epsilon_t \sim$ Stable, $n = 500$							
1.2	0.918	0.906	0.870	0.814	0.956	0.952	0.932	0.886
1.4	0.888	0.856	0.790	0.716	0.932	0.902	0.866	0.812
1.6	0.874	0.804	0.750	0.710	0.912	0.858	0.812	0.782
1.8	0.870	0.808	0.780	0.766	0.896	0.820	0.792	0.780
	$\epsilon_t \sim$ Pareto, $n = 500$							
1.2	0.928	0.926	0.904	0.908	0.976	0.972	0.960	0.958
1.4	0.924	0.894	0.860	0.840	0.962	0.932	0.902	0.896
1.6	0.842	0.778	0.742	0.716	0.902	0.836	0.806	0.766
1.8	0.856	0.772	0.738	0.728	0.904	0.834	0.786	0.790
	$\epsilon_t \sim$ Stable, $n = 1000$							
1.2	0.934	0.940	0.950	0.948	0.974	0.974	0.984	0.982
1.4	0.886	0.884	0.876	0.750	0.922	0.936	0.926	0.784
1.6	0.880	0.888	0.874	0.666	0.916	0.908	0.876	0.680
1.8	0.922	0.920	0.914	0.760	0.944	0.956	0.920	0.704
	$\epsilon_t \sim$ Pareto, $n = 1000$							
1.2	0.938	0.944	0.936	0.914	0.982	0.984	0.978	0.962
1.4	0.854	0.846	0.818	0.800	0.924	0.898	0.874	0.862
1.6	0.826	0.788	0.752	0.768	0.890	0.858	0.792	0.824
1.8	0.848	0.802	0.776	0.824	0.906	0.834	0.820	0.872

the coverage decreases with increasing block size. As expected, the empirical coverage for symmetric intervals is closer to the nominal level compared to the equal-tailed intervals, because the former intervals tend to be wider, and thus less informative. Likewise, the empirical CP increase as the tail index decreases and decrease with increasing d.

Overall, for the longer time series we report higher or similar CP compared to those with $n = 500$.

The effect of ρ is small and perhaps visible to some extent when tails are quite heavy, implying the use of larger values. The size of the intervals seems to be influenced by the tail thickness (the heavier the tails, the wider the interval) and to a lesser extent by the memory parameter (Figure 7).
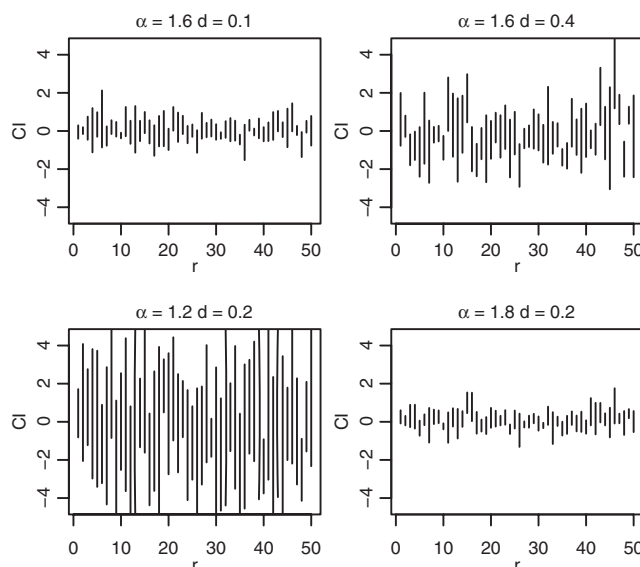
Figure 7. An example of the equal-tailed 95% confidence intervals (CI) for η with adaptively chosen b based on $R = 50$ replications of length $n = 500$ when $\epsilon_t \sim$ stable and $\rho = 0.7$
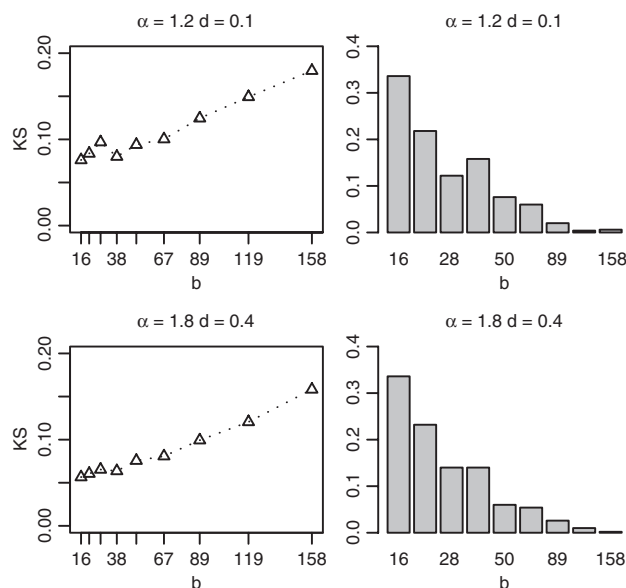

Figure 8. *Left:* Kolmogorov–Smirnov (KS) distances between $\hat{L}_{n,b}(x)$ and $\hat{L}_{n,b+1}(x)$, $b < 211$, averaged over $R = 500$ replications of length $n = 500$ when $\epsilon_t \sim$ stable and $\rho = 0.7$. *Right:* Barcharts of the block sizes selected by the GRBS rule

   In Figure 8 we show how the KS distance between the subsampling distributions for consecutive b's behave as a function of b. Since the distances are small for the relatively small block sizes, the barcharts for the data-driven bs are right skewed.

3.2. A real-data application

Now we apply our subsampling methodology to construct confidence intervals for the mean of time series of packet-counts. Two ethernet traces collected in 1994 that are freely available at http://ita.ee.lbl.gov/html/traces.html were considered: LBL-TCP-3 from the Lawrence Berkeley Laboratory and OctExt from the Bellcore Morristown Research and Engineering facility. More details about these data sets can be found in Paxson and Floyd (1995) and Leland *et al.* (1994). The first data set contains one million rows and two columns, while the second consists of 32,441 rows and eight columns. In both cases, the first two columns are often of main interest as they give packet arrival times (in seconds) and packet sizes (in bytes). Time series of packet-counts used in our study were obtained by counting the number of packets arriving within consecutive intervals of a fixed length (0.5s for LBL-TCP-3 and 1s for OctExt). In LBL-TCP-3 we took the first 14,000 observations (almost the entire series), and in the second one the first 20,000 (just under 20% of the total length). Both series and their ACFs are displayed in Figure 1 suggesting the presence of long memory, the possibility of infinite variance, and asymmetry, and as such are in line with our extended long memory time series model with heavy
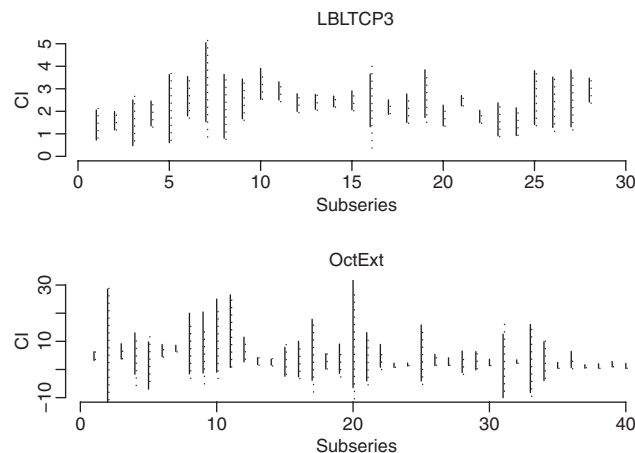
Figure 9. Equal-tailed (solid) and symmetric (dotted) 95% confidence intervals (CI) for η with an adaptively-chosen b for the consecutive 1000-long subseries of byte-counts with $\rho = 0.8$

tails. Similar time series obtained from these two traces (either byte-counts or packet-counts but with different time slots than ours) have been analysed and/or modelled by other writers. For instance, the packet-count per 50 ms series from LBL-TCP-3 was modeled with α-stable FARIMA by Harmantzis and Hatzinakos (2001). The authors estimated tail and memory parameters to be 1.86 and 0.31 respectively. Veitch and Abry (2001) considered the byte-count series (per 1s) derived from OctExt. Assuming – somewhat unrealistically – Gaussian marginals, the authors estimated the quantity $d + 0.5$ (which equals $(\beta + 1)/2$, roughly the rate of growth of partial sums for a long memory process) over non-overlapping blocks, and found that the estimates were between 0.5 and 1. Bates and McLaughlin (2000) found that the same time series had heavy tails by fitting generalized Pareto distribution.

Our time series were split into non-overlapping blocks of length 500, leading to 28 subseries for LBL-TCP-3 and 40 for OctExt, and then a 95% confidence interval for the mean was calculated in each chunk. This was done to explore whether the mean seems to change over time. The results are displayed in Figure 9.

We note that the midpoint of the intervals changes somewhat over time, but even more noteworthy is the fluctuation in interval width – as borne out in our simulation studies, heavier tails and/or longer memory tends to increase the uncertainty in our estimate of the mean. With LBL-TCP-3 the intervals do not include zero, so we can conclude that a mean effect in each block is actually present, whereas for OctExt we cannot always conclude that the mean is significantly different from zero.

4. SUMMARY

In this article we introduced a quite general HTLM model in equation 2 by allowing $\{\sigma_t\}$ to have cdf in $D(\alpha)$ and $\{G_t\}$ to be of the form $g(V_t)$ for a long memory Gaussian process $\{V_t\}$ and function g of unit Hermite rank. This can be applied to teletraffic data or to financial data such as log returns, where σ_t might be taken to be symmetric of mean zero and G_t to be the exponential of a long memory Gaussian process. In the new setting, we obtained the joint asymptotic distribution for the sample first and second moments of $\{X_t - \eta\}$ along with the long memory estimator. As in McElroy and Politis (2007a), we constructed a self-normalized sample mean statistic whose rate of convergence did not depend on either tail or memory parameter.

The subsampling approach was employed to approximate the quantiles of the sampling distribution of the self-normalized sample mean, as the parameter-dependent asymptotic distribution could not serve that purpose. The foundation for the methodology's use was based on the concept of weak dependence introduced in Doukhan and Louhichi (1999). To choose the subsampling block size, the block selection algorithm of GRBS was considered. Its implementation with the time series data was also novel.

The design of the simulation study encompassed models with various tail and memory parameters, falling under regimes where either the tail or memory dynamics dominated the variability of the sample mean. We investigated stable and Pareto distributions for $\{\epsilon_t\}$, the first with the exponential function g and the second with the identity map, as well as three user-defined values of ρ. The empirical coverage rates for a 95% nominal level ($n = 500$ and $n = 1000$) and automatically selected b were adequate, though with some tendency towards under-coverage. The confidence interval methodology was illustrated on two data sets, allowing us to monitor the behaviour of the mean in consecutive blocks of packet-count data.

APPENDIX A: PROOFS

PROOF OF THEOREM 1.   The proof consists of splicing Theorem 2 of McElroy and Politis (2007a) with Proposition 6.1 *of* McElroy and Politis (2007b). It follows from our assumptions on the distribution of σ_t and Theorem 2.2.8 of Embrechts *et al.* (1997) that

$$\bar{F}_\sigma(x) \sim c_+ x^{-\alpha} H(x), \quad F_\sigma(-x) \sim c_- x^{-\alpha} H(x) \tag{A1}$$

as $x \to \infty$ for some constants $c_+, c_- > 0$ and $H \in \mathcal{L}$. Here $\bar{F}(x) = 1 - F(x)$. Then we may define the normalization rate a_n via $n^{-1} = \mathbb{P}[|\sigma| > a_n]$, and thus obtain the relation $(c_+ + c_-)H(a_n) \sim a_n^\alpha n^{-1}$; moreover, $a_n \sim n^{1/\alpha}K(n)$ for some $K \in \mathcal{L}$. The mass in the two tails is balanced as follows: $\bar{F}_\sigma(x)/\bar{F}_{|\sigma|}(x) \sim c_+/(c_+ + c_-)$, which we denote by $r > 0$ (for the right tail), while $F_\sigma(-x)/\bar{F}_{|\sigma|}(x) \sim c_-/(c_+ + c_-)$, which will be denoted by $l > 0$. Of course $r + l = 1$. Another preliminary result we need (which follows from Taqqu, 1975) is that $\sum_{t=1}^n G_t$ obeys a central limit theorem at rate c_n, with limiting variance given by $c_n^{-2}\sum_{t,s=1}^n \gamma(t-s) \sim 1/(\beta+1)$ – also see McElroy and Politis (2007a) for the variance calculation; furthermore, $\sum_{t=1}^n |G_t|^\alpha$, $\sum_{t=1}^n G_t^2$, and $\sum_{t=1}^n (G_t)^{<\alpha>}$ obey laws of large numbers (LLNs). The LLNs results follow from Taqqu (1975) also – without having to assume any higher order moment conditions – by finding that the relevant Hermite rank in each case is at least one. Now using Fitzsimmons and McElroy (2010) we begin with the joint Fourier–Laplace transforms

$$\mathbb{E}\exp\left\{i\theta c_n^{-1}\sum_{t=1}^n Y_t - \phi c_n^{-2}\sum_{t=1}^n Y_t^2\right\} = \mathbb{E}\left[\exp\left\{ic_n^{-1}\sum_{t=1}^n \sigma_t Z_t\right\}\right],$$

with $Z_t = G_t(\theta + \sqrt{2\phi}N_t)$ and $\{N_t\}$ i.i.d standard normal, and independent of $\{\sigma_t\}$ and $\{G_t\}$. The above equality follows by conditioning on these latter variables. Next, we expand by conditioning on $\{G_t\}$ and $\{N_t\}$, using our assumptions about σ_t

$$\mathbb{E}\left[\exp\left\{ic_n^{-1}\sum_{t=1}^n \sigma_t Z_t\right\}\right] = \mathbb{E}\left[\Pi_{t=1}^n \mathbb{E}[\exp\{ic_n^{-1}\sigma_t Z_t\}|\{G_t\},\{N_t\}]\right]$$

$$= \mathbb{E}\exp\left\{\sum_{t=1}^n \log\left(1 + i\mu c_n^{-1}Z_t + \int_{-\infty}^\infty [\exp\{ic_n^{-1}Z_t x\} - 1 - ic_n^{-1}Z_t x]f_\sigma(x)dx\right)\right\}.$$

Now the integral can be broken into two portions corresponding to the positive and negative real numbers respectively, and using integration by parts and (A1) we obtain

$$\int_0^\infty [\exp\{ic_n^{-1}Z_t x\} - 1](ic_n^{-1}Z_t)\bar{F}_\sigma(x)dx - \int_{-\infty}^0 [\exp\{ic_n^{-1}Z_t x\} - 1](ic_n^{-1}Z_t)F_\sigma(x)\,dx$$

$$= i\,\text{sign}\,(Z_t)\int_0^\infty [\exp\{i\,\text{sign}\,(Z_t)y\} - 1]\bar{F}_\sigma\left(\frac{yc_n}{|Z_t|}\right)dx$$

$$- i\,\text{sign}\,(Z_t)\int_0^\infty [\exp\{-i\,\text{sign}\,(Z_t)y\} - 1]F_\sigma\left(\frac{-yc_n}{|Z_t|}\right)dx.$$

Now by (A1) we have, for any fixed $y > 0$

$$\bar{F}_\sigma\left(\frac{yc_n}{|Z_t|}\right) \sim c_+ y^{-\alpha}c_n^{-\alpha}|Z_t|^\alpha H(c_n), \quad F_\sigma\left(\frac{-yc_n}{|Z_t|}\right) \sim c_- y^{-\alpha}c_n^{-\alpha}|Z_t|^\alpha H(c_n)$$

as $n \to \infty$ for each t. As argued in McEslroy and Politis (2007a), we can take the limit over n inside the integral expression by the dominating convergence theorem so long as we have an appropriate dominating function. Now near the origin $|\exp\{\pm iy\} - 1| = O(y)$, so that the overall order of the integrand is $y^{1-\alpha}$, an integrable function near zero (since $1 < \alpha < 2$). Furthermore, direct calculation shows that

$$\int_0^\infty [\exp\{\pm iy\} - 1]y^{-\alpha}dy = C_\alpha^{-1}(\tan(\pi\alpha/2) \pm i).$$

Here $C_\alpha^{-1} = \Gamma(1-\alpha)\cos(\pi\alpha/2)$ by definition (see Propy 1.2.15 of Samorodnitsky and Taqqu, 1994). Therefore after some algebra our joint Fourier–Laplace transform is asymptotic to (using Taylor series and the fact that $c_n \to \infty$)

$$\mathbb{E}\exp\left\{i\mu c_n^{-1}\sum_{t=1}^n Z_t - \frac{(c_+ + c_-)H(c_n)}{C_\alpha c_n^\alpha}\sum_{t=1}^n |Z_t|^\alpha(1 - i(r-l)\text{sign}(Z_t)\tan(\pi\alpha/2))\right\}. \tag{A2}$$

At this point we break into the three separate cases, according to the value of A – recall that the balanced condition implies that A exists in the set $[0, \infty]$. First, suppose that $A = 0$, i.e. the tail index dominates. Then $c_n \sim a_n$ and $d_n/a_n \to 0$. In the first term of (A2), we have $\sum_{t=1}^n Z_t = \theta\sum_{t=1}^n G_t + \sqrt{2\phi}\sum_{t=1}^n G_t N_t$. The first term here is $O_P(d_n)$ by (3) since $\mathbb{E}G_t = 0$. The second term – conditional on $\{G_t\}$ – takes the form $\exp\{-\phi\sum_{t=1}^n G_t^2\}$ in the characteristic function. Since G_t^2 has non-zero mean, this sum is $O_P(n)$ so long as $\{G_t^2\}$ obeys the LLN. Thus when $A = 0$ we see that $c_n^{-1}\sum_{t=1}^n Z_t = a_n^{-1}\sum_{t=1}^n Z_t \to^P 0$ (since $n/a_n^2 \to 0$). As for the second term of (A2), note that $a_n^{-\alpha}H(a_n) \sim n^{-1}/(c_+ + c_-)$ for each t, so that we obtain

$$\exp\left\{-C_\alpha^{-1}\left(\mathbb{E}|Z|^\alpha - i(r-l)\mathbb{E}(Z)^{<\alpha>}\tan(\pi\alpha/2)\right)\right\}$$

as the final limit. This uses the variance computation of Taqqu (1975) for partial sums of a function of a Gaussian, where this function has Hermite rank at least 1. The computation is not affected by the presence of the i.i.d $\{N_t\}$ sequence.

Next, suppose that $A = \infty$ so that $c_n \sim d_n$ and $a_n/d_n \to 0$. In this case $c_n^{-2} \sum_{t=1}^{n} G_t^2 \to^P 0$ as well (since $\beta > 0$), and as a result $c_n^{-1} \sum_{t=1}^{n} Z_t \Rightarrow^{\mathcal{L}} \mathcal{N}(0, \theta^2/(\beta + 1))$ since $\mathbb{E}G_t = 0$, as determined earlier in the proof. Moreover $c_n^{\alpha}/n \to \infty$, so the second term of (A2) is $o_P(1)$. Thus the final limit is actually just $\exp\left\{-\frac{\theta^2\mu^2}{2(\beta+1)}\right\}$, by the definition of weak convergence. Note that this argument is valid even when $\mu = 0$, although the limit is degenerate. In the $\mu = 0$ case, a different convergence rate is needed; it can be shown that c_n is required to satisfy $c_n^{\alpha} \sim nH(c_n)$ as $n \to \infty$. In this case the limiting distribution would be the same as the $A = 0$ case where the tail index dominates; also, $c_n = a_n$ clearly has the needed asymptotic behaviour. But our choice of c_n as the maximum of a_n and d_n will not produce the correct rate in this case; essentially, the mean zero structure in σ_t sufficiently destroys the long memory in G_t such that the partial sums (and sample second moments) behave like the $A = 0$ case, even though $A = \infty$.

Finally, suppose that $A \in (0, \infty)$ so that $c_n \sim a_n \max\{1, \sqrt{A}\} \sim d_n \max\{1/\sqrt{A}, 1\}$. In this case we still have $c_n^{-2} \sum_{t=1}^{n} G_t^2 \to^P 0$ so that in the first term of (A2) we obtain $c_n^{-1} \sum_{t=1}^{n} Z_t \Rightarrow^{\mathcal{L}} \mathcal{N}(0, \theta^2 \min\{A, 1\}/(\beta + 1))$. As for the second term, we essentially multiply by $\max\{1, \sqrt{A}\}$ raised to the $-\alpha$ power. The final result is then

$$\exp\left\{-\frac{\theta^2\mu^2 \min\{1, A\}}{2(\beta + 1)}\mathbf{1}_{\{A>0\}} - \frac{\mathbb{E}|Z|^{\alpha}}{C_{\alpha}}\min\{A^{-\alpha/2}, 1\}\left(1 - \mathrm{i}(r - I)\frac{\mathbb{E}(Z)^{<\alpha>}}{\mathbb{E}|Z|^{\alpha}}\tan(\pi\alpha/2)\right)\mathbf{1}_{\{A<\infty\}}\right\},$$

where $Z = G(\theta + \sqrt{2\phi}N)$. In computing $\mathbb{E}|Z|^{\alpha}$ the variables G and $\theta + \sqrt{2\phi}N$ separate, because of independence. We here make a connection to Theorems 1 and 2 of McElroy and Politis (2007a) by specialising to the case that G is Gaussian. Then $\mathbb{E}(Z)^{<\alpha>} = 0$ and $\mathbb{E}|G|^{\alpha}$ is

$$(\gamma(0))^{\alpha/2}2^{\alpha/2}\pi^{-1/2}\Gamma((1+\alpha)/2) = \frac{(\gamma(0))^{\alpha/2}2^{\alpha/2}\pi^{1/2}}{\Gamma((1-\alpha)/2)\cos(\pi\alpha/2)} = \frac{(\gamma(0))^{\alpha/2}2^{-\alpha/2}\Gamma(1-\alpha/2)}{\Gamma(1-\alpha)\cos(\pi\alpha/2)},$$

via Samorodnitsky and Taqqu (1994) (p. 142) followed by 8.334.2 and 8.335.1 of Gradshteyn and Ryzhik (1994). Thus $C_{\alpha}^{-1}\mathbb{E}|G|^{\alpha} = (\gamma(0)/2)^{\alpha/2}\Gamma(1-\alpha/2)$, and this final Gamma function is accounted for in the scale, since it is easy to show that $a_n^{-\alpha} \sim n^{-1}\Gamma(1-\alpha/2)$ from the definition of σ_t in McElroy and Politis (2007a). This concludes the proof.

We might also comment here on the asymptotic behaviour when $\mu = 0$. In this case the mean of G_t need not be zero, but we at least need to suppose that the LLNs for the functionals of G_t still hold. Our viewpoint is that c_n is a generic rate to be determined. Then we still arrive at (A2) only with the first term set to zero; clearly we require $H(c_n)c_n^{-\alpha} \sim n^{-1}$ to obtain convergence, i.e. $c_n \sim a_n$. Note this should be our definition, independent of whether tails or memory dominates. In other words, we can self-normalize the partial sums by the square root of the sample second moment, since they grow at the same rate – no matter the amount of long memory present. Essentially, letting $\mu = 0$ destroys enough of the long memory behaviour in the partial sums, such that the true value of β is irrelevant to the asymptotics. The FLT in question therefore converges to that of (S, U), corresponding to the $A = 0$ case of the scenario where $\mu \neq 0$. Of course, this has repercussions for self-normalization – use of a $\widehat{LM}_n(\rho)$ term in the denominator will force the entire statistic to zero in probability when memory dominates tail, and thus needs to be omitted. $\quad\square$

PROOF OF THEOREM 2.   Since all the assumptions of Theorem 1 are in place – including $\mu \neq 0$ – all the conclusions of that result apply. (We do not pursue an analysis of the $\mu = 0$ case, though adaptations of the following argument could be made.) First by simple algebra we obtain

$$\frac{1}{n-|h|}\sum_{t=1}^{n-|h|}\left(X_tX_{t+h} - \bar{X}_n^2\right) = \frac{1}{n-|h|}\sum_{t=1}^{n-|h|}(X_t - \bar{X}_n)(X_{t+|h|} - \bar{X}_n) + \frac{\bar{X}_n}{n-|h|}\left(2|h|\bar{X}_n - \sum_{t=1}^{|h|}(X_t + X_{t+n-|h|})\right).$$

The second term on the right-hand side $O_P(1/n)$ since $\bar{X}_n$ is consistent for the mean η, and $|h| = o(n)$. So we can work instead with the first term on the right-hand side, and clearly we can replace X_t with Y_t. Then we obtain

$$\sum_{|h|=1}^{M}\frac{1}{n-|h|}\sum_{t=1}^{n-|h|}(Y_t - \bar{Y})(Y_{t+|h|} - \bar{Y}) = 2\sum_{h=1}^{M}\left(\frac{1}{n-h}\sum_{t=1}^{n-h}Y_tY_{t+h} - \bar{Y}\frac{1}{n-h}\sum_{t=1}^{n-h}(Y_t + Y_{t+h}) + \bar{Y}^2\right).$$

Here $M = \lfloor n^{\rho}\rfloor$. Now from Theorem 1 we know that $\bar{Y} = O_P(n^{-1}c_n)$, so that

$$\widehat{LM}_n^{\rho}(\rho) = 2\sum_{h=1}^{M}\frac{1}{n-h}\sum_{t=1}^{n-h}Y_tY_{t+h} + O_P(Mn^{-2}c_n^2).$$

Now by the definition of c_n, we have $O_P(Mn^{-2}c_n^2) = U_P(n^{\rho-2+\max[2/\alpha,(\beta+1)]})$. Dividing by b_n^{ρ} yields the first error term of Theorem 2. What remains we split into three terms

$$2\mu^2\sum_{h=1}^{M}\gamma_h + 2\sum_{h=1}^{M}\frac{1}{n-h}\sum_{t=1}^{n-h}(\sigma_t\sigma_{t+h} - \mu^2)G_tG_{t+h} + 2\mu^2\sum_{h=1}^{M}\frac{1}{n-h}\sum_{t=1}^{n-h}(G_tG_{t+h} - \gamma_h).$$

Clearly the first term equals $\sum_{0<|h|\leq M}R_h = b_{M+1} \sim L(n^{\rho})n^{\beta\rho}$. The third term is asymptotic to $2\mu^2n^{-1}\sum_{h=1}^{M}\sum_{t=1}^{n}(G_tG_{t+h} - \gamma_h)$, which has variance

$$\frac{4\mu^4}{n^2}\sum_{h,k=1}^{M}\sum_{t,s=1}^{n}\text{cov}(G_tG_{t+h},G_sG_{s+k})$$

$$=\frac{4\mu^4}{n^2}\sum_{h,k=1}^{M}\sum_{t,s=1}^{n}(\mathbb{E}[G_tG_{t+h}G_sG_{s+k}]-\gamma_h\gamma_k)$$

$$=\frac{4\mu^4}{n^2}\sum_{h,k=1}^{M}\sum_{t,s=1}^{n}(\text{cum}(G_0,G_k,G_{t-s},G_{t-s+h})+\gamma_{t-s}\gamma_{t-s+h-k}+\gamma_{t-s-k}\gamma_{t-s+h})$$

$$=\frac{4\mu^4}{n^2}\sum_{h,k=1}^{M}\sum_{|l|<n}(1-|l|/n)(\text{cum}(G_0,G_k,G_l,G_{l+h})+\gamma_l\gamma_{l+h-k}+\gamma_{l-k}\gamma_{l+h}).$$

The summation involving the fourth order cumulant function is finite by our hypothesis; this function is identically zero in the Gaussian case. (Incidentally, there is a mistake in the proof of Theorem 3 of McElroy and Politis (2007a), where it is incorrectly stated that $\text{Var}[G_0G_h]=2\gamma_h^2$.) Now using the Cauchy–Schwartz inequality twice and the fact that

$$\sum_{|l|<n}\gamma_l^2=\begin{cases}O(1) & \text{if }\beta<1/2\\U(\log n) & \text{if }\beta=1/2\\U(n^{2\beta-1}) & \text{if }\beta>1/2,\end{cases}$$

we see that the variance of the third term of our statistic is of order $U(n^{-1}M^2n^{\xi})$ with $\xi=0$ if $\beta\le 1/2$ and $\xi=2\beta-1$ if $\beta>1/2$. Hence the overall order of this term is $U_P(n^{\rho-\eta})$, which corresponds to the second error term of Theorem 2. Finally, for any fixed $h>0$ consider

$$n^{-1}\sum_{t=1}^{n}(\sigma_t\sigma_{t+h}-\mu^2)G_tG_{t+h}.$$

Let $Z_t=\sigma_t\sigma_{t+h}-\mu^2$, and split the sum into $h+1$ subseries over the indices $I_j=\{j,j+h+1,j+2h+2,\dots\}$ for $j=1,2,\dots,$ $h+1$. Then the $\{Z_t\}$ series restricted to any of these $h+1$ index sets will consist of i.i.d random variables. Then on I_j the Z_t are i.i.d and in $D(\alpha)$ – see Theorem 3.3(iv) of Cline (1983) and Davis and Resnick (1986) – and are also independent of $\mathcal{G}$, the σ-algebra of $\{G_t\}$. We can then apply the method of Theorem 2 (note that the regular variation property (A1) of the product variables follows from their being in $D(\alpha)$, although the slowly-varying function will be different from H) to $n^{-1}\sum_{t\in I_j}Z_tG_tG_{t+h}$, arriving at (A2) with appropriate substitutions (the mean term is zero, since we have already centred). Clearly $n^{-\alpha}\sum_{t\in I_j}|G_tG_{t+h}|^{\alpha}=O_P(n^{1-\alpha})$ by the Markov inequality; since this term raised to the $1/\alpha$ power forms the limiting scale component, the overall order is $O_P(n^{1/\alpha-1})$. This order of approximation holds for each h, which completes the proof by summing over $|h|\le M$. □

PROOF OF THEOREM 3. For the sample variance term, we can replace $\bar{X}_n$ with η asymptotically, using the same arguments as in Theorem 2 of McElroy and Politis (2007a). In the case $A=0$, it is clear that $c_n^{-2}n\widehat{LM}_n(\rho)=O_P(nb_n/a_n)$ since the slowly varying functions are overwhelmed by the polynomial growth of a_n, and this quantity tends to zero as $n\to\infty$. When $A>0$ we must assume that D_ρ is finite and non-zero; then $c_n2=\max\{1/A,1\}nb_n$ so that $c_n^{-2}n\widehat{LM}_n(\rho)\to\min\{A,1\}D_\rho$. Since the distribution of this third term is always a point mass, it converges jointly with the partial sums and sums of squares variables as well. □

PROOF OF PROPOSITION 1. First we note that the Gaussian process $\{V_t\}$ is λ-weakly dependent by Bardet et al. (2008). Then for any bounded Lipschitz function h defined on $\mathbb{R}^u$ for some integer u, the function $h\circ g$ defined via $(h\circ g)(x_1,\dots,x_u)=h(g(x_1),\dots,g(x_u))$ is Lipschitz as well, with Lip $(h\circ g)\le$ Lip $h\cdot\sup_{y\ne x}\sum_{j=1}^{u}\|g(y_j)-g(x_j)\|/\sum_{j=1}^{u}\|y_j-x_j\|$. Then it follows that $\{G_t=g(V_t)\}$ is λ-weakly dependent.

Next, for any Lipschitz function h on $\mathbb{R}^u$, define h^{σ} to be the function $h^{\sigma}(x_{i_1},\dots,x_{i_u})=h(\sigma_{i_1}x_{i_1},\dots,\sigma_{i_u}x_{i_u})$, where $\{\sigma_t\}$ for $t\in\mathbb{Z}$ denotes an infinite string of positive numbers. Then we have Lip $h^{\sigma}\le$ Lip $h\cdot\sup_{y\ne x}\sum_{j=1}^{u}\sigma_j\|y_j-x_j\|/\sum_{j=1}^{u}\|y_j-x_j\|$. So for arbitrary bounded Lipschitz functions h_1,h_2 defined on $\mathbb{R}^u$ and $\mathbb{R}^v$,

$$\|\text{cov}\{h_1(X_{i_1},\dots,X_{i_u}),h_2(X_{j_1},\dots,X_{j_v})\}\|$$

$$\le\int\|\text{cov}\{h_1^{\sigma}(G_{i_1},\dots,G_{i_u}),h_2^{\sigma}(G_{j_1},\dots,G_{j_v})\}\|\Pi_{k,l}p(\sigma_{i_k})p(\sigma_{j_l})d\sigma.$$

This is obtained by conditioning on the volatility series, using its independence from $\{G_t\}$ and its serial independence, along with the triangle inequality. The notation $p(\sigma_t)$ denotes the marginal density of the random variable σ_t. Now the covariance integrand is bounded by $\psi(u,v,\text{Lip }h_1^{\sigma},\text{Lip }h_2^{\sigma})\epsilon_r$ with $r\ge j_1-i_u$, and since $\{G_t\}$ is λ-weakly dependent the function ψ is quadratic in the Lipschitz constants. So substituting our bound for Lip h_1^{σ} and Lip h_2^{σ}, we see that the integration against the volatility densities only acts on the constants $\sup_{y\ne x}\sum_{j=1}^{u}\sigma_j\|y_j-x_j\|/\sum_{j=1}^{u}\|y_j-x_j\|$, which by linearity integrate to μ. Hence the covariance expression for $\{X_t\}$ is bounded by ϵ_r times

$$u \, v \, \text{Lip} \, h_1 \, \text{Lip} \, h_2 \mu^2 + u \, \text{Lip} \, h_1 \mu + v \, \text{Lip} \, h_2 \mu,$$

which shows that the series is λ-weakly dependent. If in addition $\{V_t\}$ is causal, it is θ-weakly dependent with bounding constant $\psi(u, v, \text{Lip} \, g_1, \text{Lip} \, g_2) = v \, \text{Lip} \, g_2$ for functions g_1, g_2. Tracing through the same arguments shows that $\{G_t\}$ is θ-weakly dependent and that the covariance for $\{X_t\}$ is bounded by ϵ_r times $v \, \text{Lip} \, h_2 \mu$, so that $\{X_t\}$ is also θ-weakly dependent. $\square$

APPENDIX B: VALIDITY OF SUBSAMPLING UNDER Θ-WEAK DEPENDENCE

We here develop a theoretical result regarding the validity of the subsampling procedure for self-normalized statistics computed from strictly stationary time series with θ-weak dependence. It is emphasized that this result is developed in a context much more general than that of the rest of the article, i.e. general statistics from a general processes satisfying the weak dependence condition of Doukhan and Louhichi (1999).

Consider a statistic $\hat{\theta}_n$ as an estimator of unknown parameter θ (not to be confused with the nomenclature for the weak dependence), which has a studentization $\hat{\sigma}_n$ such that $\tau_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n$ converges weakly to some limit random variable with cdf denoted by L. Let the sampling cdf be denoted L_n, which is computed from a sample of size n drawn from a stationary θ-weakly dependent process with coefficients ϵ_r.

ASSUMPTION A. *As $n \to \infty$ we have $L_n(x) \to L(x)$. In addition, $\alpha_n(\hat{\theta}_n - \theta)$ converges weakly to Z and $\delta_n \hat{\sigma}_n$ converges weakly to W, for positive sequences $\{\alpha_n\}$ and $\{\delta_n\}$ satisfying $\tau_n = \alpha_n/\delta_n$. Here, H and W are two random variables, where W is positive with probability one.*

Let the subsampling distribution estimator based on the self-normalized statistic $\tau_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n$ be denoted by $\hat{L}_{n,b}(x)$, as defined in Section 2.3. Also denote the (lower) $1 - p$ quantile of the subsampling distribution by $c_{n,b}(1 - p)$; the $1 - p$ quantile of $L(x)$ is denoted $c(1 - p)$. Then we have the following result, which is proved next. The theorem as stated results in one-sided confidence bounds; the construction of confidence intervals (either equal-tailed or symmetric) is immediate in view of the discussion in Politis *et al.* (1999).

THEOREM 4. *Assume Assumption A. Also, assume $\alpha_b/\alpha_n \to 0$, $\tau_b/\tau_n \to 0$, $b/n \to 0$, and $b \to \infty$ as $n \to \infty$. Assume the time series is strictly stationary and θ-weakly dependent with rate $\epsilon_r = O(r^{-a})$ for $a \geq 1/2$.*

[i] *If x is a continuity point of L, then $\hat{L}_{n,b}(x) \to^P L(x)$.*
[ii] *If L is continuous, then $\sup_x |\hat{L}_{n,b}(x) - L(x)| \to^P 0$.*
[iii] *If L is continuous at $c(1 - p)$, then as $n \to \infty$,*

$$\mathbb{P}\left[\tau_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n \leq c_{n,b}(1 - p)\right] \to 1 - p.$$

Thus the asymptotic coverage of the interval $[\hat{\theta} - \hat{\sigma}_n \tau_n^{-1} c_{n,b}(1 - p), \infty)$ is the nominal $1 - p$.

PROOF OF THEOREM 4. The proof is a splicing of Theorem 11.3.1 of Politis *et al. (1999)* with Lemma 3.1 of Ango-Nze *et al. (2003)*; but we repeat some of the calculations of the latter work, due to its non-accessibility. We focus on the proof of [i], since [ii] and [iii] will follow as in the proof of Theorem 3.2.1 of Politis *et al. (1999)*, utilizing the assumed continuity of L at value x. Letting $Z_{n,b,t} = \tau_b(\hat{\theta}_{n,b,t} - \theta)/\hat{\sigma}_{n,b,t}$, define

$$I_{b,t} = 1_{\{Z_{n,b,t} \leq x\}}$$

for $t = 1, 2, \ldots, n - b + 1$. So we can define a version of the subsampling estimator with $\hat{\theta}_n$ replaced by the unknown θ, and call this $U_{n,b}(x)$; in other words, let $U_{n,b}(x) = (n - b + 1)^{-1} \sum_{t=1}^{n-b+1} I_{b,t}$. As in the former part of the proof of Theorem 11.3.1 of Politis *et al.* (1999), the difference between $L_{n,b}(x)$ and $U_{n,b}(x)$ tends to zero in probability for each x, so that we can focus on the latter expression. (The content of Assumption A is used in this part of the argument.) Expanding the variance of $U_{n,b}(x)$ with $q = n - b + 1$ yields

$$\text{var} U_{n,b}(x) = q^{-2} \left(\sum_{|h|<q} (q - |h|)\gamma_I(h) \right),$$

where $\gamma_I(h) = \text{cov}(I_{b,t}, I_{b,t+h})$. This expression does not depend on t due to strict stationarity; we suppress dependence on b and n in this notation.

Note that $\text{var} U_{n,b}(x)$ tends to zero as long as $q^{-1} \sum_{h=b}^{q} \gamma_I(h) \to 0$ as $n \to \infty$ (since $b/q \to 0$ and $\gamma_I(h)$ is a bounded sequence). We expand the covariance function $\gamma_I(h)$ as in the proof of Lemma 3.1 of Ango-Nze *et al.* (2003), in particular obtaining their Equation (3.5). Specifically, let $h_x(t) = 1_{\{t \leq x\}}$ and $h_{x,\delta}(t) = 1_{\{t \leq x\}} + (1 - (t - x)/\delta)1_{\{x < t \leq x+\delta\}}$ for some $\delta > 0$; this second function is a piecewise

linear approximation to the former. The central issue with θ-weak dependence is that statistics should ultimately be Lipschitz functions in the data, and the indicator function violates this condition. However, one can approximate it if there is suitable continuity in the limit cdf. Note that $I_{b,t} = h_x(Z_{n,b,t})$. Then we have

$$\gamma_I(h) = \text{cov}(h_x(Z_{n,b,t}) - h_{x,\delta}(Z_{n,b,t}), h_x(Z_{n,b,t+h}))$$
$$+ \text{cov}(h_{x,\delta}(Z_{n,b,t}), h_x(Z_{n,b,t+h}) - h_{x,\delta}(Z_{n,b,t+h}))$$
$$+ \text{cov}(h_{x,\delta}(Z_{n,b,t}), h_{x,\delta}(Z_{n,b,t+h})).$$

For the first term on the right-hand side, use the Cauchy–Schwarz inequality to obtain the bound of var $(1_{\{x < Z_{n,b,t} \le x+\delta\}})$, which in turn is bounded by $\mathbb{P}[x < Z_{n,b,t} \le x + \delta] = L_b(x + \delta) - L_b(x)$ by strict stationarity. The same argument applies to the second term, since $\mathbb{P}[x < Z_{n,b,t+h} \le x + \delta] = L_b(x + \delta) - L_b(x)$ as well. These quantities can be made suitably small by choosing δ small enough due to the assumed continuity at x.

The final term can be handled via the definition of θ-weak dependence, since $h_{x,\delta}$ is a Lipschitz function of the data. We then have that

$$\text{cov}(h_{x,\delta}(Z_{n,b,t}), h_{x,\delta}(Z_{n,b,t+h})) = O(\sqrt{b}\epsilon_{h-b+1}/\delta)$$

where ϵ_r is the weak dependence rate in the assumptions of the theorem. Note that, up to the factor $\sqrt{b}/\delta$, the above is just like the bound entailed by Lemma A.0.2 of Politis *et al.* (1999).

Using the choice $\delta = (n^{-2a}b)^{1/4}$ as in Ango-Nze *et al.* (2003), it then follows that $\gamma_I(h) \to 0$ as $h \to \infty$. A Cesaro-mean argument now implies that var $U_{n,b}(x)$ tends to zero, and the theorem is proven. □

REFERENCES

Ango-Nze, P., Dupoiron, S. and Rios, R. (2003) *Subsampling under weak dependence conditions*. Technical Report DT2003-42, CREST.

Bardet, J. -M., Doukhan, P., Lang, G. and Ragache, N. (2008) *Dependent Lindeberg central limit theorem and some applications*, ArXiv0701872 preprint.

Bates, S. and McLaughlin, S. (2000) The estimation of stable distribution parameters from teletraffic data. *IEEE Transactions on Signal Processing*, **48**, 865–70.

Bickel, P. J. and Sakov, A. (2008) On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica*, **18**, 967–85.

Cappe, O., Moulines, E., Pesquet, J.-C., Petropulu, A. P. and Yang, X. (2002) Long-range dependence and heavy-tail modeling for teletraffic data. *Signal Processing*, **19**, 14–27.

Cline, D. B. H. (1983) *Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data*. Ph.D. thesis, Department of Statistics, Colorado State University.

Davis, R. A. and Resnick, S. I. (1986) Limit theory for the sample covariance and correlation functions of moving averages. *The Annals of Statistics*, **14**, 533–58.

Doukhan, P. and Louhichi, S. (1999) A new weak dependence and applications to moment inequalities. *Stochastic Processes and Their Applications*, **84**, 313–43.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer-Verlag.

Fitzsimmons, P. and McElroy, T. (2010). On joint Fourier-Laplace transforms. *Communications in Statistics - Theory and Methods*, **39**, 1883–5.

Götze, F. and Račkauskas, A. (2001) Adaptive choice of bootstrap sample sizes. In *State of the Art in Probability and Statistics, IMS Lecture Notes Monogr. Ser., 36* (eds. A. van der Vaart, M. de Gunst, C. Klaassen. Cambridge: Cambridge University Press, pp. 286–309.

Gradshteyn, I. S. and Ryzhik, I. M. (1994) *Tables of Integrals, Series, and Products*. San Diego: Academic Press.

Guégan, D. and Ladoucette, S. (2001) Non-mixing properties of long memory sequences. *C.R. Acad. Sci. Paris*, **333**, 373–6.

Hall, P., Jing, B. and Lahiri, S. (1998) On the sampling window method for long-range dependent data. *Statistica Sinica*, **8**, 1189–204.

Harmantzis, F. and Hatzinakos, D. (2001) *Network traffic modeling and simulation using stable FARIMA processes*. Technical Report. Presented at INFORMS Annual Meeting, Miami.

Ibragimov, I. and Rozanov, Y. (1978) *Gaussian Random Processes*. New York: Springer.

Lee, I. W. C. and Fapojuwo, O. (2005) Stochastic processes for computer network traffic modeling. *Computer Communications*, **29**, 1–23.

Leland, W. E., Taqqu, M. S., Willinger, W. and Wilson, D.V. (1994) On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, **2**, 1–15.

McElroy, T. and Politis, D. N. (2007a) Self-Normalization for heavy-tailed time series with long memory. *Statistica Sinica*, **17**, 199–220.

McElroy, T. and Politis, D. N. (2007b) Computer-intensive rate estimation, diverging statistics, and scanning. *Annals of Statistics*, **35**, 1827–48.

Nordman, D. and Lahiri, S. (2005) Validity of the sampling window method for long-range dependent linear processes. *Econometric Theory*, **21**, 1087–111.

Paxson, V. and Floyd, S. (1995) Wide-area traffic: the failure of Poisson modelling. *IEEE/ACM Trans. Networking*, **3**, 226–4.

Politis, D. N., Romano, J. P. and Wolf, M. (1999) *Subsampling*. New York: Springer.

Samorodnitsky, G. and Taqqu, M. S. (1994) *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. NY: Chapman & Hall.

Stoev, S. and Taqqu, M. (2004) Simulation methods for linear fractional stable motion and FARIMA using the fast Fourier transform. *Fractals*, **12**, 95–121.

Taqqu, M. (1975) Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, **31**, 287–302.

Veitch, D. and Abry, P. (2001) A statistical test for the time constancy of scaling exponents. IEEE Trans. on Sig. Proc., **49** (10), 2325–34.

Willinger, W., Paxson, V. and Taqqu, M. S. (1998) Self-similarity and heavy tails: structural modeling of network traffic. In *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (eds. R. Feldman R. Adler and M. S. Taqqu). Birkhäuser: Boston, pp. 27–53.