Original Article

Journal of
Time Series Analysis

First version received September 2010

Published online in Wiley Online Library: 15 July 2012

(wileyonlinelibrary.com) DOI: 10.1111/j.1467-9892.2012.00808.x

Subsampling inference for the autocovariances and autocorrelations of long-memory heavy- tailed linear time series

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

We provide a self-normalization for the sample autocovariances and autocorrelations of a linear, long-memory time series with innovations that have either finite fourth moment or are heavy-tailed with tail index $2 < \alpha < 4$. In the asymptotic distribution of the sample autocovariance there are three rates of convergence that depend on the interplay between the memory parameter d and α, and which consequently lead to three different limit distributions; for the sample autocorrelation the limit distribution only depends on d. We introduce a self-normalized sample autocovariance statistic, which is computable without knowledge of α or d (or their relationship), and which converges to a non-degenerate distribution. We also treat self-normalization of the autocorrelations. The sampling distributions can then be approximated non-parametrically by subsampling, as the corresponding asymptotic distribution is still parameter-dependent. The subsampling-based confidence intervals for the process autocovariances and autocorrelations are shown to have satisfactory empirical coverage rates in a simulation study. The impact of subsampling block size on the coverage is assessed. The methodology is further applied to the log-squared returns of Merck stock.

Keywords: Linear time series; parameter-dependent convergence rates; self-normalization; subsampling confidence intervals.

1. INTRODUCTION

Long memory models and processes have enjoyed an increasing impact in the time series literature over the past few decades, and there is substantial interest in studying heavy-tailed long-range dependent phenomena. The reason for this is that such processes seem to be useful in describing data from many of the applied sciences, including hydrology (Montanari *et al.*, 1997; Serroukh and Walden, 2000), finance (Deo *et al.*, 2006; Kulik and Soulier, 2011) and network traffic (Cappe *et al.*, 2002). Although some of the statistical literature is concerned with data that are either heavy-tailed or long memory, there are some studies (see the above references) that consider time series exhibiting both effects. Both nonlinear and linear models (e.g. Montanari *et al.*, 1997) are utilized in practice; this article focuses on the latter. Understanding the statistical properties of sample autocovariances (acvs) and autocorrelations (acs) is important in order to obtain proper modelling of data.

Horváth and Kokoszka (2008) (henceforth HK) derived the asymptotic distributions for the sample acvs, demonstrating explicitly that the convergence rate depends upon α, d and their relationship to one another. Moreover, the limit asymptotic distributions depended on α and d – one can obtain a Gaussian, a Stable, or a Rosenblatt process as the limit – and which limit random variable pertains depends upon the interplay of α and d. In fact, three cases arise according to a partition of the α-d parameter space into three regions.

Analogous to this trichotomous situation, a dichotomy was found in McElroy and Politis (2007), which studied the sample mean problem. This situation was addressed through self-normalization, such that the appropriately studentized mean had nondegenerate limit distribution irrespective of α and d, that is, without knowledge of α, d, or their relationship. This was effective in constructing a test statistic computable without prior knowledge of parameters, and quantiles of the limit distribution could be obtained via subsampling (see Jach *et al.*, 2012).

Here we seek to extend this self-normalization strategy to the sample acv and ac problems. In particular, we propose a combination of sample fourth moment and tapered sample acvs to jointly account for both heavy tails and long memory, analogously to how self-normalization was achieved for the sample mean in McElroy and Politis (2007). The fourth moment correctly assesses the rate of growth in the acv due to heavy tails, whereas the tapered sample acvs assess growth due to long memory. In order to validate this technique, it is necessary to extend the theoretical results of HK to two joint weak convergences. For the sample ac statistic, the contribution of the heavy tails actually drops out asymptotically, so that d alone determines the asymptotic behaviour; hence, a different studentization is appropriate.

[a]*US Census Bureau*
[b]*Universidad Carlos III de Madrid*
**Correspondence to: Tucker McElroy, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233.*
[†]*E-mail: tucker.s.mcelroy@census.gov*
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 U.S. Census Bureau.

Of course, these studentizations only resolve half the problem – namely, eliminating the need to know the convergence rate to compute the studentized statistic – since the limit distributions will be unconventional and non-pivotal. As in Jach *et al.* (2012), we propose to use subsampling to empirically estimate the quantiles of the sampling distribution. Subsampling, elucidated in Politis *et al*. (1999), operates by computing the same statistics on a small subsample – typically a contiguous stretch – drawn from the original time series data, with the unknown parameter being replaced by its best large-sample estimate. Consistency of the resulting empirical distribution for the sampling (or asymptotic) distribution is typically enforced through a strong-mixing assumption, together with strict stationarity, but recent literature indicates that the more flexible θ-weak dependence coefficients of Doukhan and Louhichi (1999) are also sufficient to establish subsampling in the context of long range dependence.

So we begin with basic concepts and notation in Section 2, along with a summary of HK, which motivates the self-normalizations that we employ. With our statistics defined, we provide full asymptotic theory in Section 3 for the principal cases of interest; these results are extensions of the mathematics of HK, and contain theorems of independent interest. Section 4 focuses on the subsampling discussion, and we discuss the method's validity under θ-weak dependence. Our empirical studies are in Section 5, where extensive simulations explore the interplay of kurtosis and memory. We also explore the adaptive block-size selection algorithm of Götze and Račkauskas (2001), and we study the log-squared returns of Merck stock. Section 6concludes, and proofs are in the Appendix.

2. SELF-NORMALIZATION FOR AUTOCOVARIANCES AND AUTOCORRELATIONS

As in HK, we consider a linear, long-memory time series $\{X_t\}$ with mean-zero innovations $\{Z_t\}$ that have either finite fourth moment or are heavy-tailed with index $2 < \alpha < 4$. That is, we define

$$X_t = \sum_{j=0}^{\infty} \psi(j) Z_{t-j},$$

where the Z_ts are i.i.d. and either $\mathbb{E}[Z^4] < \infty$ or for some $2 < \alpha < 4$,

$$\mathbb{P}[|Z_t| > x] = x^{-\alpha} L(x), \quad \frac{\mathbb{P}[Z_t > x]}{\mathbb{P}[|Z_t| > x]} \to p \in [0,1],$$

as $x \to \infty$, where $L \in \mathcal{L}$, $\mathcal{L}$ being the set of slowly varying functions (Embrechts *et al.*, 1997). Set $q = 1 - p$ and let Z be a common version of $\{Z_t\}$, which is absolutely continuous by assumption. The cdf of Z, $F(x) = \mathbb{P}[Z \le x]$, will be denoted by F and the right-tail cdf, $\overline{F}(x) = \mathbb{p}[Z > x]$, by $\overline{F}$. Then the heavy-tail conditions, denoted by HT(α), are equivalent to the existence of a sequence a_N such that

$$N\overline{F}(a_N x) \to p x^{-\alpha} \quad \text{and} \quad NF(-a_N x) \to q x^{-\alpha}$$

as $N \to \infty$, for any $x > 0$. This implies that $N\mathbb{P}[|Z| > a_N x] \to x^{-\alpha}$. It follows from results in Embrechts *et al.* (1997) that $a_N = N^{1/\alpha} L(N)$; moreover,

$$a_N^{-2} \sum_{t=1}^{N} (Z_t^2 - b_N) \overset{\mathcal{L}}{\Rightarrow} S,$$

where S is an (α/2)-stable random variable and $b_N = \mathbb{E}[Z^2 \mathbf{1}\{|Z| \le a_N\}]$.

The assumption on long memory is expressed through the behaviour of (square-summable) coefficients $\psi(j)$ satisfying $\psi(j) \sim C_d j^{d-1}$, that is, it is assumed that for $x \ge 0$

$$\psi(x) = x^{d-1} K(x), \quad 0 < d < 0.5,$$

where K is a real-valued function such that $K(x) \to C_d$ as $x \to \infty$, and $\psi_j = \psi(j)$ for integers j. For $x < 0$, set $\psi(x) = 0$.

We are principally interested in the sample acvs based on the available sample $X_1, X_2, \ldots, X_{N+H}$, $N > 1$, $H \ge 0$,

$$\widehat{\gamma}_{h,N} = \frac{1}{N} \sum_{t=1}^{N} X_t X_{t+h}, \quad h = 0, 1, \ldots, H.$$

We use a sample of size $N + H$ rather than N to avoid notational annoyances; this makes no difference to the asymptotic results. If $h < 0$, let $\widehat{\gamma}_{h,N} = \widehat{\gamma}_{-h,N}$. (One can also center by the sample mean, and this makes no difference to the asymptotic results of the study.) Letting $\sigma^2 = \text{var}[Z]$, the population acvs are

$$\gamma_h = \mathbb{E}[X_0 X_h] = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h} = \sigma^2 \Psi_h.$$

HK establish three rates of convergence (Figure 1) for the sample acvs. These rates depend on the interplay between α and d, and are described below.

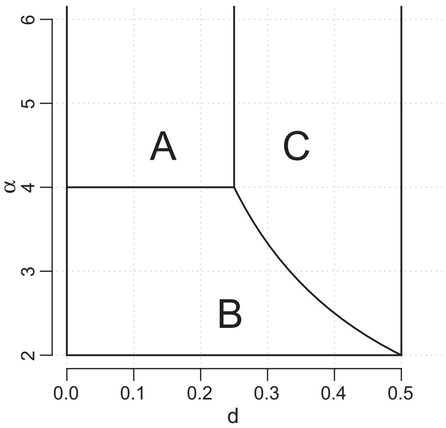

Figure 1. Regions with three different rates of convergence of $\widehat{\gamma}_{h,N} - \gamma_h$. Here $A = \{(d,\alpha)|0 \leq d < 1/4, \alpha > 4\}$ with standard rate $N^{1/2}$; $B = \{(d,\alpha)|2 < \alpha < 4, d < 1/\alpha\}$ with tail-dominated rate $\sim N^{1-2/\alpha}$; $C = \{(d,\alpha)|1/4 < d < 1/2, \alpha > 1/d\}$ with memory-dominated rate N^{1-2d}

Let c_N denote the normalization for $N(\widehat{\gamma}_{N,h} - \gamma_h)$, such that $Nc_N^{-1}(\widehat{\gamma}_{N,h} - \gamma_h)$ converges in distribution to a non - degenerate random variable. From HK we know that under suitable assumptions, we have convergence of the following vector of sample acvs at lags zero through H:

$$Nc_N^{-1}\{\widehat{\gamma}_{h,N} - \gamma_h\}_{h=0}^{H} \xrightarrow{\mathcal{L}} \begin{cases} \{G_h\}_{h=0}^{H} & \text{if } (d,\alpha) \in A \text{ (Theorems 3.2 and 3.3a of HK)} \\ (S - \frac{\alpha}{\alpha-2})\{\Psi_h\}_{h=0}^{H} & \text{if } (d,\alpha) \in B \text{ (Theorem 3.1a of HK)} \\ \sigma^2 C_d^2 \{U_d(1)\}_{h=0}^{H} & \text{if } (d,\alpha) \in C \text{ (Theorems 3.1b and 3.3b of HK)}. \end{cases}$$

In the first case, $\{G_h\}_{h=0}^{H}$ is a mean-zero Gaussian vector with covariances $\mathbb{E}[G_h G_k] = (\eta - 3)\gamma_h\gamma_k + \sum_{\ell=-\infty}^{\infty}(\gamma_\ell\gamma_{\ell-h+k} + \gamma_{\ell+k}\gamma_{\ell-h})$, where η is given by $\mathbb{E}[Z^4] = \eta\sigma^4$. In the third case, $U_d(t)$ is the Rosenblatt process (which does not depend on h)

$$U_d(t) = \int\int\left[\int_0^t (v-x)_+^{d-1}(v-y)_+^{d-1}\mathrm{d}v\right]W(\mathrm{d}x)W(\mathrm{d}y),$$

where W is the standard Brownian Motion (BM) defined on the real line. The rate c_N in the above cases is

$$c_N = \begin{cases} N^{1/2} & \text{if}(d,\alpha) \in A \\ a_N^2 = N^{2/\alpha}L^2(N) & \text{if}(d,\alpha) \in B \\ N^{2d} & \text{if}(d,\alpha) \in C, \end{cases}$$

and it follows that $c_N = \max(N^{1/2}, a_N^2, N^{2d})$. So c_N depends upon α and d and their relationship to one another, and thus this rate cannot be computed without some prior knowledge about the data process.

We are first interested in self-normalization of $N(\widehat{\gamma}_{h,N} - \gamma_h)$ by a statistic that involves the unknown rate c_N. For any fixed h, consider the following normalization, that actually grows at rate c_N (shown in the next section):

$$\widehat{\sigma}_N = \sqrt{S_N^{(4)}} + V_{\Lambda,M},$$

where

$$S_N^{(4)} = \sum_{t=1}^{N} X_t^4$$

is the sample fourth moment, and

$$V_{\Lambda,M} = \sum_{|h|\leq M}\Lambda(h/M)\widehat{\gamma}_{N,h} = \sum_{|h|\leq M}\Lambda(h/M)\frac{1}{N}\sum_{t=1}^{N}X_t X_{t+|h|}.$$

The integer M is the bandwidth, and typically is shrinking relative to the growth of the sample size N. This type of estimator – a tapered sum of sample acvs – is called a HAC (Heteroskedasticity-Autocorrelation) estimator in the econometrics literature (Kiefer and Vogelsang, 2005). In this study we employ the fixed bandwidth-ratio perspective (see McElroy and Politis, 2011b), so we assume that $M = \lfloor \rho N \rfloor$ for some user-defined bandwidth ratio $\rho \in (0,1]$. Although Jach *et al.* (2012) uses $M = N^\rho$ for some $\rho < 1$, and raises $V_{\Lambda,M}$ to the $1/\rho$ power to obtain the correct convergence rate, this was only utilized to ensure a non-random limit of the statistic's denominator. However, for the subsampling applications we have in mind, a random limit presents no difficulties. $\Lambda(\cdot)$ is a taper, which we assume is continuous and even with support on $[-1,1]$, such that $\Lambda(x) = 1$ for $|x| \leq c$ for some $c \in [0,1)$. For many tapers

$c = 0$, but for a flat-top taper (Politis, 2001) we have $c > 0$. Additionally, we assume that $\Lambda(\cdot)$ is continuous and twice differentiable on $(-1, -c)\cup(1,c)$. The derivatives from the left and right of a point x will be denoted by $\dot\Lambda_-(x)$ and $\dot\Lambda_+(x)$ respectively.

Note that this definition of the self-normalization does not depend on h – we can utilize the same studentization for every sample acv. The self-normalized difference is then defined to be

$$T_{h,N} = N\frac{\widehat{\gamma}_{h,N} - \gamma_h}{\sqrt{S_N^{(4)} + V_{\Lambda,M}}} = N\frac{\widehat{\gamma}_{h,N} - \gamma_h}{\widehat{\sigma}_N}.$$

It is proved in Theorem 3 that $T_{h,N}$ converges to a non-degenerate distribution, no matter which of the cases A, B, or C pertains. This non-degeneracy is a prerequisite for the subsampling technique described in Section 4.

But for some applications, more interest focuses on the autocorrelations, although the lag behaviour of acvs and acs is qualitatively the same. The population acs are $\varrho_h = \gamma_h/\gamma_0$, and are typically estimated via the sample acs $\widehat{\varrho}_{h,N} = \widehat{\gamma}_{h,N}/\widehat{\gamma}_{0,N}$. In the following sections, we also present results (from both theory and simulation) for sample acs. Interestingly, a direct application of the sample acv theory to the sample ac case reveals that tail thickness plays no role asymptotically, and so the parameter space instead becomes partitioned along the line $d = 1/4$. The studentization used for the acv case is no longer appropriate, and instead we suggest $\sqrt{N} + V_{\Lambda,M}$. Then

$$Q_{h,N} = N\frac{\widehat{\varrho}_{h,N} - \varrho_h}{\sqrt{N} + V_{\Lambda,M}}$$

is the self-normalized centered sample ac.

3. THEORY

In this section we establish weak convergence of $T_{h,N}$ for the various cases A, B, or C, and weak convergence of $Q_{h,N}$ for the cases $d < 1/4$ and $d > 1/4$. The two main steps for the acvs are the joint convergence results relating the sample autocovariance to each one of the normalizing quantities, $S_N^{(4)}$ and $V_{\Lambda,M}$. As shown in HK, the sample acv statistics can be divided into two portions, the so-called diagonal and off-diagonal contributions to the double summation entailed by a quadratic statistic of a linear process. First, we present a convergence result for the diagonal portion of the sample acf jointly with the sample fourth moment, when $2 < \alpha < 4$ (but memory not constrained). Second, we consider a joint convergence result for the off-diagonal portion of the sample acf and $V_{\Lambda,M}$, for $1/4 < d < 1/2$ (but any tail index $\alpha > 2$). Then we demonstrate how to combine these two results, together with an auxiliary proposition, such that no matter which zone A, B, or C pertains, the asymptotic distribution of the normalized statistic is non-degenerate.

But for the sample acs the situation is different, because the diagonal portion for the autocorrelations is always negligible relative to the off-diagonal portions, even in region B. This leads to a different normalization with asymptotics stated in a separate theorem.

As in HK, expansion of the sample acv yields

$$N^{-1}\sum_{t=1}^{N} X_t X_{t+h} - \gamma_h = N^{-1}\sum_{t=1}^{N}\sum_{j\geq 0}\psi_j\psi_{j+h}(Z_{t-j}^2 - \sigma^2) + N^{-1}\sum_{t=1}^{N}\sum_{k\neq j+h}\psi_j\psi_k Z_{t-j}Z_{t+h-k},$$

and these random variables on the RHS will be denoted by $d_{N,h}$ and $r_{N,h}$, respectively. The diagonal portion $d_{N,h}$ has growth rate matched by that of the square root of the sample fourth moment, whereas the off-diagonal portion remainder $r_{N,h}$ grows at a rate matched by $V_{\Lambda,M}$.

Theorem 1. *Assume that $2 < \alpha < 4$ and $0 \leq d < 1/2$. Then $(Na_N^{-2}d_{N,h}, a_N^{-4}S_N^{(4)})$ converges weakly to $(\Psi_h\cdot U,V)$ jointly in $0 \leq h \leq H$, with U and V both positive stable random variables. In particular, the joint Fourier–Laplace Transform (FLT) of the diagonal portion of the first H sample autocovariances and the fourth sample moment has a limit given by*

$$\begin{aligned}&\mathbb{E}\exp\left\{i\sum_{h=0}^{H}\nu_h N a_N^{-2} d_{N,h} - \frac{\omega^2}{2}a_N^{-4}S_N^{(4)}\right\} \to \exp\left\{g\left(\sum_{h=0}^{H}\nu_h\Psi_h, \omega\sqrt{\sum_j \psi_j^4}\right)\right\}\\ &g(x,y) = \int_0^\infty \left[\exp\{ixz^2 - y^2z^4/2\}(2ixz - 2y^2z^3) - 2ixz\right]z^{-\alpha}dz.\end{aligned} \tag{1}$$

Remark 1. By results in Fitzsimmons and McElroy (2010), the FLT characterizes weak convergence. The marginal distributions are found by setting x or y equal to zero in $g(x,y)$. That is,

$$g(x,0) = -|x|^{\alpha/2}C_{\alpha/2}^{-1}(1 - i\ \mathrm{sign}(x)\tan\pi\alpha/4),$$

for C_α defined by $C_\alpha^{-1} = \Gamma(1-\alpha)\cos(\pi\alpha/2)$. Then $\exp\{g(x,0)\}$ is the characteristic function of the $\alpha/2$ stable distribution (called U in Theorem 1) with scale $C_{\alpha/2}^{-\alpha/2}$, skewness one, and location zero (see Samorodnitsky and Taqqu, 1994). In contrast,

$$g(0,y) = -(y^2/2)^{\alpha/4}\Gamma(1-\alpha/4), \tag{2}$$

so that the exponential is the Laplace Transform (Proposition 1.2.12 of Samorodnitsky and Taqqu, 1994) of an $\alpha/4$ stable random variable with scale $(\Gamma(1-\alpha/4)\cos(\pi\alpha/8))^{4/\alpha}$, skewness one and location zero (called V in Theorem 1).

Theorem 1 is primarily of interest in zone B; incidentally, it also shows that the sample fourth moment is $O_P(a_N^4)$ even in part of zone C, where $2 < \alpha < 4$ but $d > 1/\alpha$. The next result handles the remainder terms.

Theorem 2. *Assume that $\alpha > 2$ and $1/4 < d < 1/2$. Let Λ be a taper of form described in Section 2, with bandwidth $M = \rho N$ and $\rho \in (0,1]$. Then the off-diagonal portion of the sample acvs converge jointly with $V_{\Lambda,M}$ to quadratic functionals of Fractional Brownian Motion (FBM):*

$$\left(\{N^{1-2d} r_{N,h}\}_{h=0}^{H}, N^{-2d}V_{\Lambda,M}\right) \overset{\mathcal{L}}{\Rightarrow} \sigma^2\left(\left\{\int\int c(x,y)W(\mathrm{d}x)W(\mathrm{d}y)\right\}_{h=0}^{H}, \int\int d_\rho(x,y)W(\mathrm{d}x)W(\mathrm{d}y)\right)$$

as $N \to \infty$. The integrand functions are given by

$$\begin{aligned}
c(x,y) &= C_d^2\int_0^1 (z-x)_+^{d-1}(z-y)_+^{d-1}\mathrm{d}z \\
d_\rho(x,y) &= -\frac{C_d^2}{\rho^2 d^2}\int_0^1\int_0^1 \ddot{\Lambda}\left(\frac{r-s}{\rho}\right) f_r(x) f_s(y)\mathrm{d}r\ \mathrm{d}s \\
&\quad - \frac{2C_d^2}{\rho d^2}\dot{\Lambda}_+(c)\int_0^{1-c\rho} f_r(x) f_{r+c\rho}(y)\mathrm{d}r + \frac{2C_d^2}{\rho d^2}\dot{\Lambda}_-(1)\int_0^{1-\rho} f_r(x) f_{r+\rho}(y)\mathrm{d}r \\
f_r(x) &= (r-x)_+^d - (-x)_+^d.
\end{aligned} \tag{3}$$

Note that the limit of the remainders is the same exact variable for all h, indicating full inter-dependence in the limit. Also the second limit variable is absolutely continuous, and positive with probability one. This result is primarily of interest in zone C, but in zone B we see that $V_{\Lambda,M} = O_P(N^{2d})$ as well, which is useful to know; in fact this is true for all $\alpha > 2$. We state this, and some other required auxiliary results, below.

Proposition 1. *Let Λ be a taper of form described in Section 2, with bandwidth $M = \rho N$ and $\rho \in (0,1]$. For all $0 \le d < 1/2$ and all $\alpha > 2$,*

$$N^{-2d}V_{\Lambda,M} \overset{\mathcal{L}}{\Rightarrow} \sigma^2 \int\int \mathrm{d}_\rho(x,y)W(\mathrm{d}x)W(\mathrm{d}y). \tag{4}$$

When $\alpha > 4$, for any $0 \le d < 1/2$,

$$N^{-1}S_N^{(4)} \xrightarrow{P} \mathbb{E}[X^4]. \tag{5}$$

When $\alpha \in (2,4)$ and $d \in [0,1/4)$,

$$Na_N^{-2}r_{N,h} \xrightarrow{P} 0. \tag{6}$$

When $\alpha > 4$ and $d \in (1/4,1/2)$,

$$N^{1-2d}d_{N,h} \xrightarrow{P} 0. \tag{7}$$

The above results can now be combined to address the asymptotic schedules under cases A, B and C of the root $T_{h,N}$.

Theorem 3. *Let $\alpha > 2$ and $0 \le d < 1/2$ such that $d \ne 1/\alpha$, $\alpha \ne 4$, and $d \ne 1/4$. Then we have the following weak convergence, where the limit random variables have been defined in Theorems 1 and 2 (and G_h is the Gaussian defined in Section 2):*

$$\frac{N\{\widehat{\gamma}_{h,N} - \gamma_h\}_{h=0}^{H}}{\sqrt{S_N^{(4)} + V_{\Lambda,M}}} \overset{\mathcal{L}}{\Rightarrow} \frac{G_h}{\sqrt{\mathbb{E}[X^4]}}\mathbf{1}_A(d,\alpha) + \frac{\Psi_h\cdot U}{\sqrt{V}}\mathbf{1}_B(d,\alpha) + \frac{\int\int c(x,y)W(\mathrm{d}x)W(\mathrm{d}y)}{\int\int d_\rho(x,y)W(\mathrm{d}x)W(\mathrm{d}y)}\mathbf{1}_C(d,\alpha)$$

as $N \to \infty$.

Remark 2. Results have not been proved for the boundary cases, where the process belongs to the intersection of two or more zones. The theory is more challenging here, and it is likely that the limit variables become super-imposed on the boundaries of the respective regions, as in McElroy and Politis (2007).

Before stating the main result for the autocorrelations, we make an observation that follows immediately from Theorem 3. Noting that

$$\widehat{\varrho}_{h,N} - \varrho_h = \widehat{\gamma}_{0,N}^{-1}\left[(\widehat{\gamma}_{h,N} - \gamma_h) - \varrho_h(\widehat{\gamma}_{0,N} - \gamma_0)\right], \tag{8}$$

it follows that on zone B

$$Na_N^{-2}(\widehat{\varrho}_{h,N} - \varrho_h) \overset{\mathcal{L}}{\Rightarrow} \gamma_0^{-1}(\Psi_h - \rho_h\Psi_0)U$$

as $N \to \infty$. Since $\Psi_h = \rho_h\Psi_0$, this limit random variable is actually zero almost surely. In other words, a_N^2 is not the appropriate rate for the sample acs in zone B, being too large. Actually, a separate analysis – along the lines of that provided in Proposition 4.1 of Davis and Resnick (1986) – is required. It can then be shown that the normalizing rate c_N for the sample acs is actually $\max(N^{1/2}, N^{2d})$, irrespective of $\alpha > 2$. A normalization that also has this growth rate is given by $\sqrt{N} + V_{\Lambda,M}$ – note that the sample fourth moment in $\widehat{\sigma}_N$ has been replaced by N here. The following theorem summarize the asymptotics of the root $Q_{h,N}$.

Theorem 4. *Let $\alpha > 2$ and $0 \le d < 1/2$ such that $d \neq 1/4$. Suppose that $\varrho_h \neq 1$ for $1 \le h \le H$. Then we have the following weak convergence, where the limit random variables have been defined in Theorems 1 and 2 (and G_h is the Gaussian defined in Section 2):*

$$\frac{N\{\widehat{\varrho}_{h,N} - \varrho_h\}_{h=1}^{H}}{\sqrt{N} + V_{\Lambda,M}} \overset{\mathcal{L}}{\Rightarrow} \frac{G_h - \varrho_h G_0}{\gamma_0}\mathbf{1}_{\{d<1/4\}}(d) + \frac{(1-\varrho_h)\int\int c(x,y)W(dx)W(dy)}{\gamma_0\int\int d_\rho(x,y)W(dx)W(dy)}\mathbf{1}_{\{d>1/4\}}(d)$$

as $N \to \infty$.

4. SUBSAMPLING APPROXIMATION OF $\mathbb{P}[T_{h,N} \le x]$ AND $\mathbb{P}[Q_{h,N} \le x]$

To construct confidence intervals for γ_h we need to approximate the sampling distribution of $T_{h,N}$, that is, $L_{h,N}(x) = \mathbb{P}[T_{h,N} \le x]$. Let L_h denote the cdf of the corresponding limiting random variable – which by Theorem 3 depends on α and d. The use of this asymptotic distribution is impractical, as L_h depends on the unknown parameters and the relationship between them. Besides, in two of the three cases there is no known analytic formula for the cumulative distribution function of the limit random variable. Hence, we propose to approximate $L_{h,N}$ non-parametrically via subsampling (Politis *et al.,* 1999). The procedure dictates that we split the sample into overlapping blocks of size b ($b \to \infty$, $b/N \to 0$), written $X_t, X_{t+1}, \ldots, X_{t+b-1}$ for $t = 1,2,\ldots,N-b+1$, and calculate the self-normalized statistic upon each block, treating each block as if it were a full sample. Moreover, the parameter γ_h is replaced by its large-sample estimate $\widehat{\gamma}_{h,N}$. This leads to $N - b + 1$ subsampling statistics

$$T_{h,N,b,t} = b\frac{\widehat{\gamma}_{h,N,b,t} - \widehat{\gamma}_{h,N}}{\widehat{\sigma}_{N,b,t}},$$

where $\widehat{\gamma}_{h,N,b,t} = \sum_{\ell=t}^{t+b-1-h} X_\ell X_{\ell+h}/(b-h)$ and

$$\widehat{\sigma}_{N,b,t} = \sqrt{S_{N,b,t}^{(4)} + V_{\Lambda,\lfloor\rho b\rfloor,t}}.$$

In the last expression

$$S_{N,b,t}^{(4)} = \sum_{\ell=t}^{t+b-1} X_\ell^4$$

and

$$V_{\Lambda,\lfloor\rho b\rfloor,t} = \sum_{|h|\le\lfloor\rho b\rfloor} \Lambda(h/\lfloor\rho b\rfloor)\widehat{\gamma}_{h,N,b,t}.$$

Note that the HAC estimator V_Λ is computed on a bandwidth appropriate for the subsample, namely being equal to the same proportion ρ of the data length, which is now b; hence the bandwidth is $\lfloor\rho b\rfloor$. The sampling distribution $L_{h,N}(x)$ is then approximated by

$$\widehat{L}_{h,N,b}(x) = \frac{1}{N-b+1}\sum_{t=1}^{N-b+1}\mathbf{1}_{\{T_{h,N,b,t}\le x\}},$$

based on which we can construct an $(1-p)$ equal-tailed subsampling confidence intervals for γ_h:

$$CI_{et;1-p}(\gamma_h) = \left[\widehat{\gamma}_{h,N} - \frac{\widehat{\sigma}_N}{N} c_{h,N,b}(1-p/2), \widehat{\gamma}_{h,N} - \frac{\widehat{\sigma}_N}{N} c_{h,N,b}(p/2)\right],$$

where $c_{h,N,b}(1-p) = \inf\{x : \widehat{L}_{h,N,b}(x) \geq 1-p\}$ is the $1-p$ (lower) quantile of the subsampling distribution $\widehat{L}_{h,N,b}$. Alternatively, an $(1-p)$ symmetric subsampling confidence interval for γ_h is

$$CI_{s;1-p}(\gamma_h) = \left[\widehat{\gamma}_{h,N} - \frac{\widehat{\sigma}_N}{N} c_{h,N,b,|\cdot|}(1-p), \widehat{\gamma}_{h,N} + \frac{\widehat{\sigma}_N}{N} c_{h,N,b,|\cdot|}(1-p)\right],$$

$c_{h,N,b,|\cdot|}(1-p) = \inf\{x : \widehat{L}_{h,N,b,|\cdot|}(x) \geq 1-p\}$ and $\widehat{L}_{h,N,b,|\cdot|}(x) = \sum_{t=1}^{N-b+1} \mathbf{1}_{\{|T_{h,N,b,t}| \leq x\}}/(N-b+1)$.

To establish the consistency of the subsampling estimator, that is, $|\widehat{L}_{h,N,b}(x) - L_h(x)| \xrightarrow{p} 0$ as $N \to \infty$ for all x, we will employ a recent result of Jach *et al.* (2012) (Theorem 4 of Appendix B), based on the notion of the θ-weak dependence (Doukhan and Louhichi, 1999; Bardet *et al.*, 2008).

We begin by noting that according to the explicit examples found in Bardet *et al.* (2008), our linear process $\{X_t\}$ is λ-weakly dependent with mixing sequence $\lambda_r = L(r)r^{2d-1}$; and because it is a causal process, it is also θ-weakly dependent. In view of Theorems 1, 2 and 3, the application of Theorem 4 of Jach *et al.* (2012) to the problem here is immediate. However, that result requires that $\lambda_r = O(r^{-a})$ for some $a \geq 1/2$, so $d \leq 1/4$ is a sufficient condition for the consistency result. Examining the proof in Jach *et al.* (2012) shows that the key bound on the subsampling estimator's variance is of the order $\sqrt{b}$ times $q^{-1}\sum_{h=1}^{q-b}\lambda_h$, where $q = N-b+1$, or overall is roughly $O(\sqrt{b}N^{-a})$ for $b = o(N)$. If $a \geq 1/2$ then the variance tends to zero, no matter the particular choice of b. However, if $a < 1/2$ we can choose the blocks to ensure consistency. Suppose that b grows – up to a slowly varying function – of order N^{ζ} for some $\zeta < 1$. Choosing $\zeta < 2a$ when $a < 1/2$ then ensures the subsampling estimator's variance tends to zero.

To summarize, noting that $d = (1-a)/2$: when $d \leq 1/4$ the subsampling procedure is consistent for any choice of blocks b such that $b/N \to 0$; but when $d > 1/4$ we require that $b = O(N^{\zeta})$ (up to slowly varying functions) for some $0 < \zeta < 2-4d$ in order to obtain consistency. The disadvantage of this latter scenario is that d is unknown to us. But in practice block size will be selected by a data-driven technique (described below), that automatically determines on smaller blocks when serial dependence is greater.

Let us proceed under these conditions on b. Denoting the limit of $T_{h,N}$ as Z/W – where Z is either G_h, $\Psi_h{\cdot}U$, or $\int\int c(x,y)W(dx)W(dy)$ and W is either $\sqrt{\mathbb{E}[X^4]}$, $\sqrt{V}$, or $\int\int d_\rho(x,y)W(dx)W(dy)$ according to cases A, B, or C – and writing $\alpha_N = N/c_N$ and $\delta_N = 1/c_N$, we obtain $\alpha_b/\alpha_N \to 0$ and $\tau_b/\tau_N \to 0$ (because $b/N \to 0$), as required by Theorem 4 of Jach *et al.* (2012). Now, since the process $\{X_t\}$ satisfies the weak-dependence condition, we are assured of the consistency of the subsampling procedure.

Turning now to the case of autocorrelations and the sampling distribution of the self-normalized sample autocorrelation $\mathbb{P}[Q_{h,N} \leq x]$, we see that the cdf of the corresponding limiting variable also depends on the unknown parameter d by Theorem 4. Hence, like in the case of the acv, we can resort to subsampling to approximate $\mathbb{P}[Q_{h,N} \leq x]$. In this case the $N-b+1$ subsampling statistics are

$$Q_{h,N,b,t} = b\frac{\widehat{\varrho}_{h,N,b,t} - \widehat{\varrho}_{h,N}}{\sqrt{b} + V_{\Lambda,\lfloor\rho b\rfloor,t}},$$

based on which formulas for the equal-tailed and symmetric subsampling confidence intervals for ϱ_h, analogous to $CI_{et;1-p}(\gamma_h)$ and $CI_{s;1-p}(\gamma_h)$, can be derived. The consistency proof of the subsampling estimator of $\mathbb{P}[Q_{h,N} \leq x]$ for $0 \leq d < 1/4$ follows closely that of the subsampling estimator of $\mathbb{P}[T_{h,N} \leq x]$.

5. SIMULATION STUDY AND APPLICATION

In this section we work with the realizations of the so-called fractionally integrated noise $\{X_t\}$ (Brockwell and Davis, 1991), which is a linear process defined via the difference operator

$$\Delta^d = (1-B)^d = \sum_{j=0}^{\infty} \pi_j B^j,$$

where $BX_t = X_{t-1}$ and

$$\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \prod_{0<k\leq j} \frac{k-1-d}{k}, \quad j = 0, 1, \ldots.$$

$\Gamma(\cdot)$ is the gamma function. We will only consider cases where $0 < d < 0.5$, and we define $\{X_t\}$ as

$$\Delta^d X_t = Z_t, \quad \text{where } Z_t \sim WN(0, \sigma^2).$$

We draw Z_ts from a Student distribution with α degrees of freedom (without the loss of generality, Z_ts will be standardized), $\alpha \in \{2.5, 3.5, 4.5, 5.5\}$ and the values of the memory parameter $d \in \{0.1, 0.2, 0.3, 0.4\}$. In Table 1 we list several values of γ_h and ϱ_h calculated from the formulas of Brockwell and Davis (1991),

Table 1. Selected values for γ_h and ϱ_h in the fractionally integrated noise ($\sigma^2 = 1$)

	γ_h				ϱ_h			
$h\backslash d$	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
1	0.1133	0.2747	0.5642	1.3801	0.1111	0.2500	0.4286	0.6667
5	0.0316	0.1060	0.2999	1.0070	0.0310	0.0965	0.2278	0.4864
10	0.0181	0.0700	0.2274	0.8768	0.0178	0.0637	0.1727	0.4236
15	0.0131	0.0549	0.1933	0.8086	0.0129	0.0499	0.1469	0.3906
20	0.0104	0.0462	0.1723	0.7634	0.0102	0.0420	0.1309	0.3688

$$\gamma_h = \sigma^2 \frac{\Gamma(h+d)\Gamma(1-2d)}{\Gamma(h-d+1)\Gamma(d)\Gamma(1-d)}, \quad \varrho_h = \frac{\Gamma(h+d)\Gamma(1-d)}{\Gamma(h-d+1)\Gamma(d)}.$$

We also take into account several bandwidth ratios of Section 2, $\rho \in \{0.2,0.4,0.6,0.8,1\}$. To compare the effect of ρ we use the same replications of $\{X_t\}$. The replications are generated by an R routine `fracdiff.sim` and there are $R = 500$ of them. Two types of tapers are considered: the Bartlett taper, $\Lambda^B(x) = (1 - |x|)\mathbf{1}_{\{|x|\leq 1\}}(x)$, and a flat-top Trapezoidal taper with $c = 0.5$ (Politis and Romano, 1995)

$$\Lambda^{T,c}(x) = \begin{cases} 1 & \text{if } |x| \leq c \\ (|x|-1)/(c-1) & \text{if } c < |x| \leq 1 \\ 0 & \text{otherwise,} \end{cases}$$

where $c \in [0, 1)$. In the subsampling procedure we look at the block sizes b derived from an adaptive algorithm for the block-size selection of Götze and Račkauskas (2001) and Bickel Sakow (2008) (GRBS) that can be summarized in the following steps:

- Consider values of b given by $b_j = \lfloor q^j N \rfloor$ for $j = 1,2,\ldots$ and $0 < q < 1$. For each b_j, compute $\widehat{L}_{h,N,b_j}$.
- Let δ be the Kolmogorov–Smirnov distance and set $\widehat{b} = \arg\min_{b_j} \delta\left(\widehat{L}_{h,N,b_j}, \widehat{L}_{h,N,b_{j+1}}\right)$.
- The optimal estimator of L_n is $\widehat{L}_{h,N,\widehat{b}}$.

For $q = 0.75$ and $N = 500$ we have block sizes $b \in \{16,21,28,38,50,67,89,119,158,211\}$, which are then doubled as we move from 500 to 1000 observations. In both cases we cover the ratios of b/N from around 3–42%.

We then construct both types – equal-tailed and symmetric – of 95% subsampling confidence intervals for γ_h and ϱ_h, $h \in \{1,5,10,15,20\}$ and calculate the empirical coverage probabilities (cps). In Figures 2 and 3 we show cps for γ_1 as a function of the block-size b (0 corresponds to the adaptively chosen block size) grouped by the memory parameter d, tail index α and bandwidth-controlling ρ, all with Bartlett taper and for the sample size $N = 500$. Analogous results with Trapezoidal taper are captured in Figures 4 and 5.

As we can see, the effect of the bandwidth $M = \lfloor N\rho \rfloor$ on the coverage of γ_1 is negligible (e.g., compare cps within each subplot of Figures 2–5). This situation changes slightly when the lag h and memory parameter d are large. In such cases larger values of ρ are

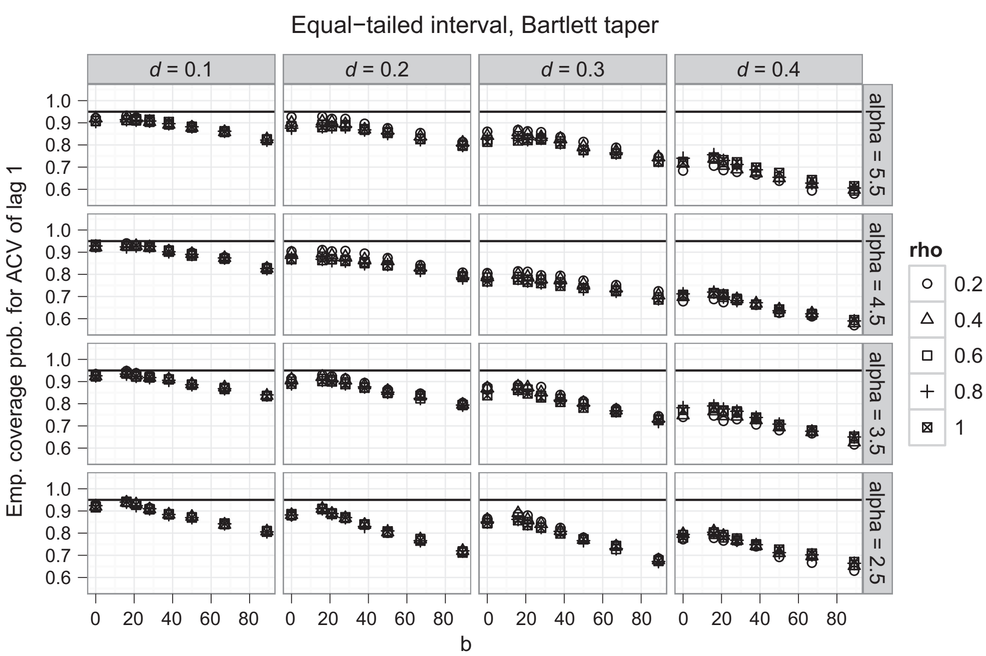

Figure 2. Coverage probabilities for the 95% equal-tailed subsampling confidence intervals for γ_1 versus subsampling block b (value of 0 corresponds to the adaptively chosen b, $N = 500$)

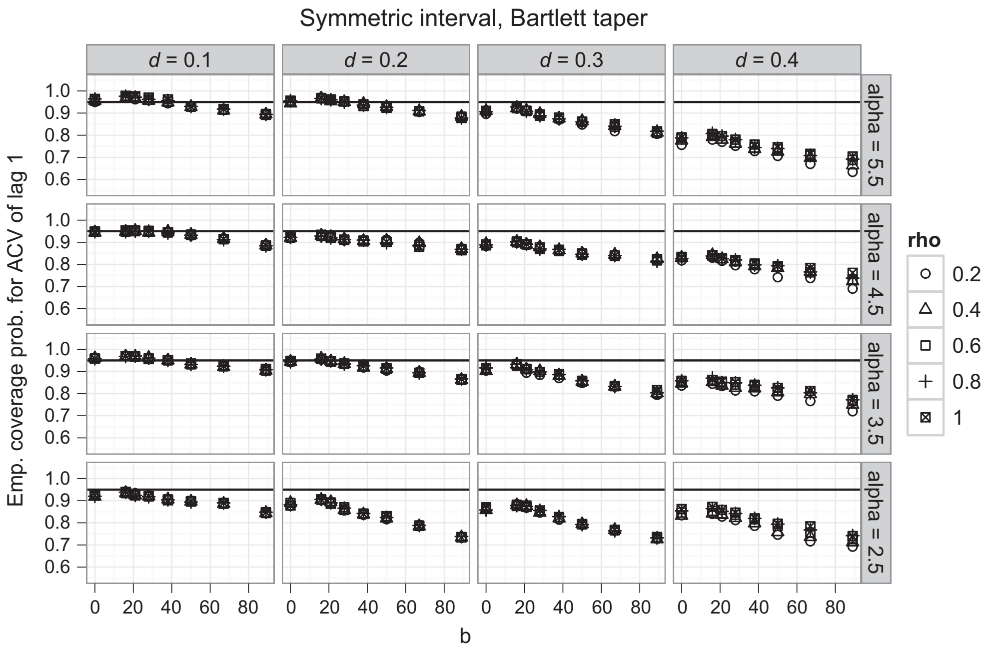

Figure 3. Coverage probabilities for the 95% symmetric subsampling confidence intervals for γ_1 versus subsampling block b (value of 0 corresponds to the adaptively chosen b, $N = 500$)

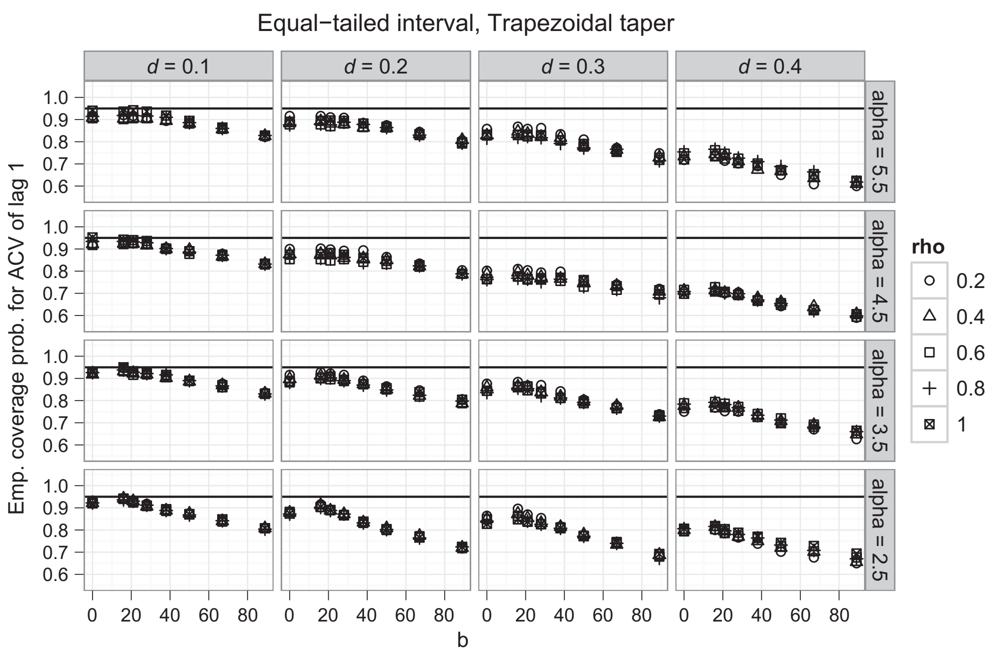

Figure 4. Coverage probabilities for the 95% equal-tailed subsampling confidence intervals for γ_1 versus subsampling block b (value of 0 corresponds to the adaptively chosen b, $N = 500$)

required to achieve satisfactory performance. The choice of the taper has little influence on the results (compare Figures 2 and 3 with 4 and 5 and Table 2 with 3), although cps with Trapezoidal taper are slightly higher. As expected, the equal-tailed confidence intervals produce slightly lower coverage than the symmetric ones (compare Figures 2 and 4 with 3 and 5).

Overall, the empirical cps for acvs are satisfactory and range from around 68% up to around 97%. The poorer results are associated with large d and this is particularly apparent for larger lags of the acv. The quality of the coverage seems to be more influenced by the memory (undercoverage for $d = 0.4$), than by the tail. This might be partially explained by the finite-sample bias in the estimation of γ_h by $\widehat{\gamma}_{h,N}$ and length of the intervals. However, it is apparent that block size has a *greater* impact on coverage when d is high, and in particular that smaller blocks are preferable in these cases. That is, coverage deteriorates (right-hand panels) more

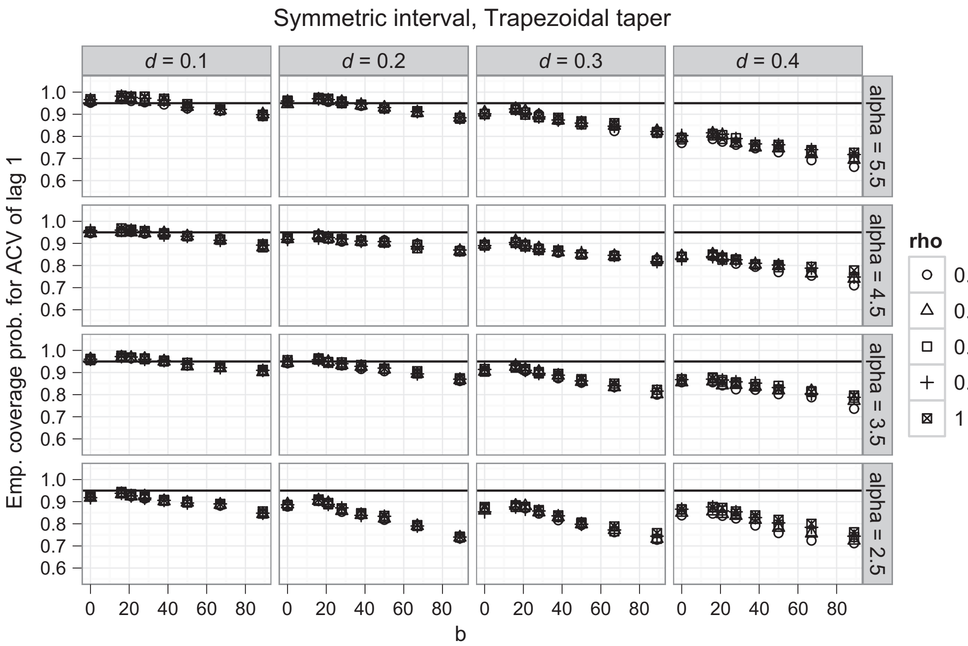

Figure 5. Coverage probabilities for the 95% symmetric subsampling confidence intervals for γ_1 versus subsampling block b (value of 0 corresponds to the adaptively chosen b, $N = 500$)

Table 2. Empirical coverage probabilities for the 95% subsampling confidence intervals for γ_h based on the adaptively chosen block b and Bartlett taper ($\rho = 0.6$, $N = 1000$)

	Equal-tailed				Symmetric			
$\alpha\backslash d$	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
$h = 1$								
5.5	0.936	0.892	0.826	0.744	0.952	0.934	0.894	0.782
4.5	0.940	0.918	0.848	0.694	0.970	0.948	0.914	0.778
3.5	0.958	0.916	0.848	0.756	0.974	0.942	0.902	0.824
2.5	0.942	0.880	0.856	0.770	0.926	0.850	0.828	0.808
$h = 5$								
5.5	0.944	0.894	0.796	0.680	0.964	0.942	0.862	0.740
4.5	0.944	0.914	0.808	0.656	0.968	0.954	0.860	0.728
3.5	0.980	0.932	0.848	0.724	0.988	0.938	0.884	0.778
2.5	0.978	0.938	0.860	0.750	0.986	0.920	0.830	0.768
$h = 10$								
5.5	0.962	0.918	0.800	0.648	0.966	0.936	0.846	0.708
4.5	0.970	0.946	0.828	0.632	0.970	0.954	0.872	0.702
3.5	0.970	0.926	0.860	0.688	0.978	0.946	0.876	0.758
2.5	0.994	0.974	0.854	0.728	0.994	0.970	0.848	0.766
$h = 15$								
5.5	0.954	0.922	0.804	0.640	0.960	0.946	0.856	0.698
4.5	0.970	0.950	0.848	0.614	0.972	0.954	0.888	0.694
3.5	0.978	0.964	0.850	0.676	0.990	0.980	0.874	0.736
2.5	0.988	0.974	0.860	0.708	0.992	0.972	0.856	0.734
$h = 20$								
5.5	0.968	0.954	0.828	0.628	0.974	0.966	0.860	0.684
4.5	0.982	0.964	0.870	0.616	0.978	0.982	0.886	0.676
3.5	0.972	0.948	0.852	0.660	0.980	0.972	0.868	0.722
2.5	0.992	0.984	0.858	0.682	0.998	0.990	0.854	0.726

rapidly as block size increases for cases where d is high, as compared to cases where d is small or moderate (left-hand panels). This empirical finding is consistent with the discussion given in Section 4 – that to control the variance of the subsampling estimator, one should take smaller blocks whenever $d > 1/4$. Also the GRBS rule tends to automatically select the smaller blocks in these cases where there is more long memory.

Examples of the 95% symmetric subsampling confidence interval for γ_1 are given in Figure 6. The true value of γ_1 was subtracted from the end-points of $CI_{s;0.95}(\gamma_1)$ so that the intervals are centered at the horizontal axis. When the convergence rate is standard the intervals are short; when it is dominated by the tail, we occasionally have a longer interval, especially when α is close to 2, and the

Table 3. Empirical coverage probabilities for the 95% subsampling confidence intervals for γ_h based on the adaptively chosen block b and Trapezoidal taper ($\rho = 0.6$, $N = 1000$)

	Equal-tailed				Symmetric			
$\alpha \backslash d$	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
$h = 1$								
5.5	0.934	0.882	0.812	0.764	0.960	0.934	0.904	0.802
4.5	0.942	0.904	0.840	0.720	0.970	0.950	0.904	0.798
3.5	0.960	0.906	0.840	0.772	0.976	0.944	0.904	0.838
2.5	0.940	0.876	0.846	0.786	0.922	0.852	0.840	0.824
$h = 5$								
5.5	0.934	0.888	0.784	0.696	0.966	0.944	0.868	0.758
4.5	0.946	0.898	0.792	0.666	0.968	0.948	0.868	0.754
3.5	0.976	0.928	0.826	0.732	0.990	0.944	0.890	0.798
2.5	0.976	0.936	0.864	0.760	0.986	0.922	0.834	0.794
$h = 10$								
5.5	0.970	0.914	0.790	0.686	0.964	0.936	0.858	0.738
4.5	0.970	0.932	0.844	0.664	0.970	0.960	0.888	0.736
3.5	0.970	0.918	0.854	0.716	0.980	0.950	0.886	0.770
2.5	0.992	0.972	0.868	0.762	0.994	0.966	0.852	0.784
$h = 15$								
5.5	0.954	0.920	0.800	0.676	0.964	0.948	0.868	0.724
4.5	0.966	0.932	0.838	0.646	0.970	0.960	0.894	0.724
3.5	0.976	0.952	0.842	0.696	0.988	0.982	0.888	0.764
2.5	0.986	0.974	0.866	0.732	0.992	0.972	0.860	0.758
$h = 20$								
5.5	0.976	0.944	0.832	0.662	0.974	0.970	0.868	0.704
4.5	0.974	0.948	0.864	0.636	0.982	0.982	0.892	0.698
3.5	0.976	0.938	0.846	0.682	0.980	0.972	0.870	0.738
2.5	0.992	0.982	0.866	0.718	0.998	0.992	0.856	0.740

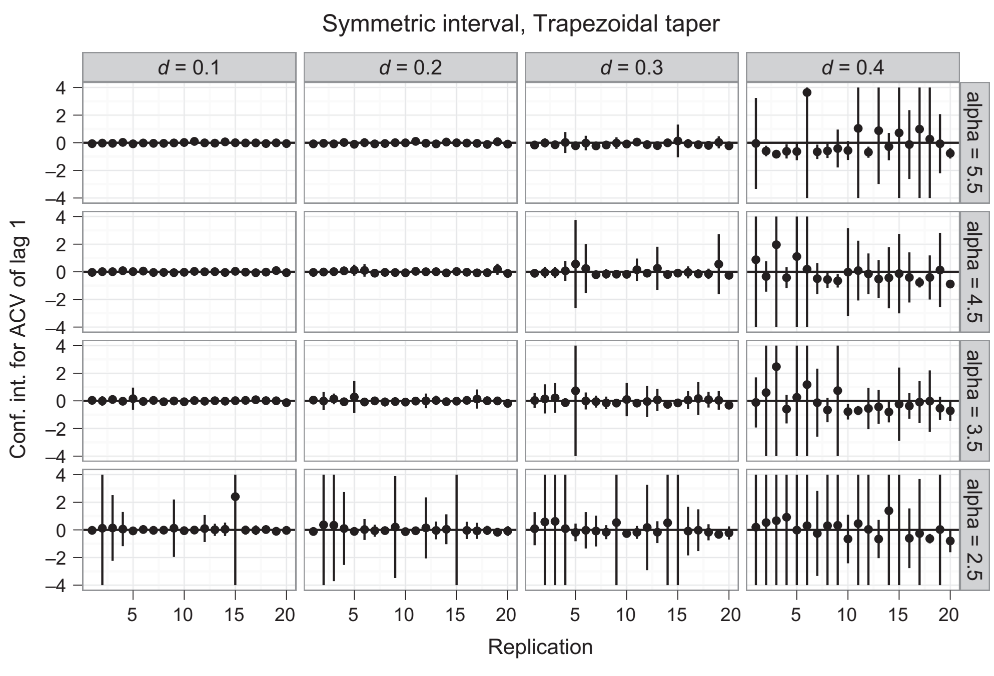

Figure 6. Example of the 95% symmetric subsampling confidence intervals for γ_1, shifted by the true value ($\widehat{\gamma}_{1,N} - \gamma_1$ marked as a black circle, $N = 500$, b adaptively chosen, $\rho = 0.6$)

shorter intervals are centered at the true value; when the memory dynamics dominate, the longer intervals are more abundant and the estimates of γ_1 miss the target more often. The higher variability of $\widehat{\gamma}_1$ in zone C compared to zones A and B, means that when the intervals are short they are not able to trap the true parameter. Consequently, this contributes to a slightly more pronounced undercoverage in cps for $d = 0.4$ and $\alpha = 5.5$.

Table 2 contains the results for the longer series where $N = 1000$, for γ_h with the lag values $h \in \{1,5,10,15,20\}$, with the adaptively selected block size, $\rho = 0.6$ and Bartlett taper. The cps with the same settings but for Trapezoidal taper are displayed in Table 3. The

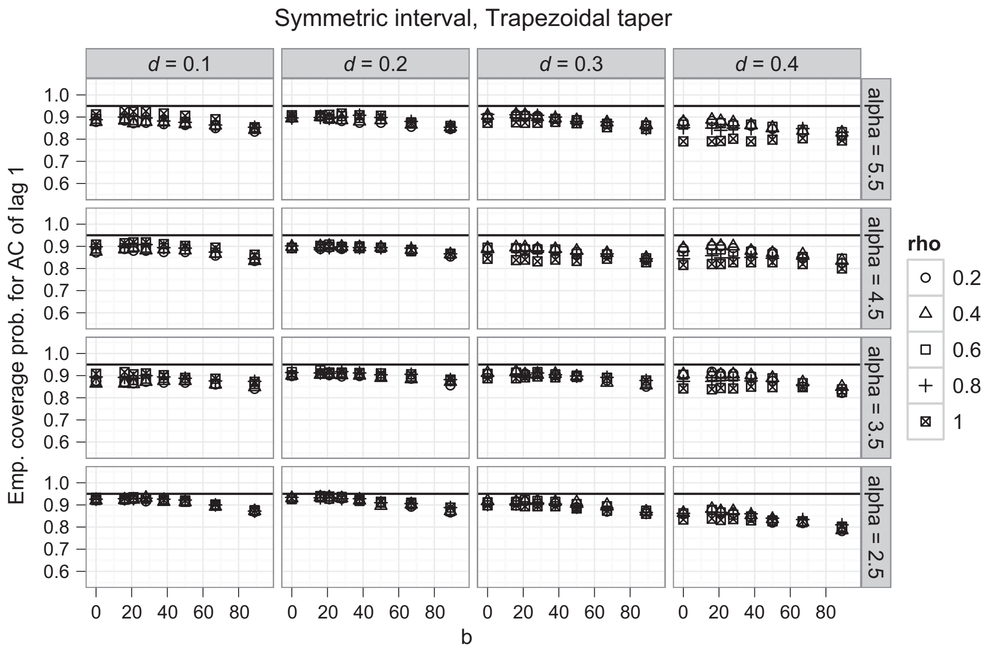

Figure 7. Coverage probabilities for the 95% equal-tailed subsampling confidence intervals for ϱ_1 versus subsampling block b (value of 0 corresponds to the adaptively chosen b, $N = 500$)

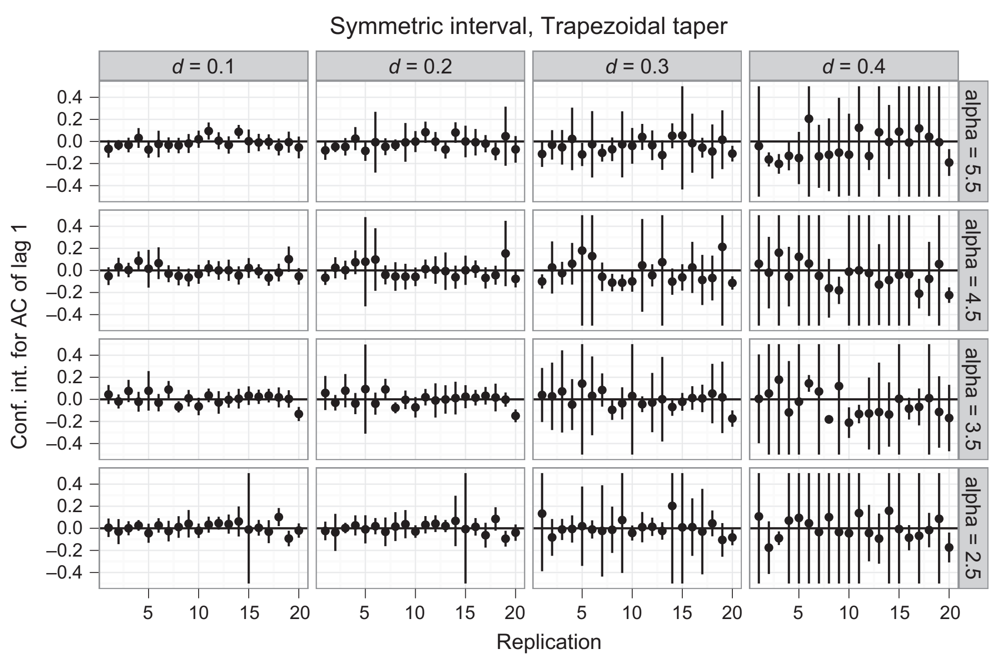

Figure 8. Example of the 95% symmetric subsampling confidence intervals for ϱ_1, shifted by the true value ($\widehat{\varrho}_{1,N} - \varrho_1$ marked as a black circle, $N = 500$, b adaptively chosen, $\rho = 0.6$)

coverage for γ_1 is very similar to that obtained with $N = 500$ observations. Focusing now on the behaviour of cps with respect to h, we notice that when the memory dominates the convergence rate of $\widehat{\gamma}_{h,N}$, the cps decrease with an increasing lag.

Now we discuss the results for the acs. The cps from symmetrical confidence intervals for ϱ_1 with Trapezoidal taper are displayed in Figure 7 and example of those (shifted by the true value) in Figure 8. For $\rho = 0.6$, the cps for the adaptively selected block, range from about 86% to 93% and tend to be lower in the memory-dominated region. The intervals for ϱ_1 in the region with the standard $\sqrt{N}$ convergence rate, are narrower than those in the memory-dominated one. The latter intervals become wider as d increases and have centres that show a finite-sample bias in estimating ϱ_1. We observe little impact of the tail on the cps and the size of the intervals, which is in accordance with Theorem 4.

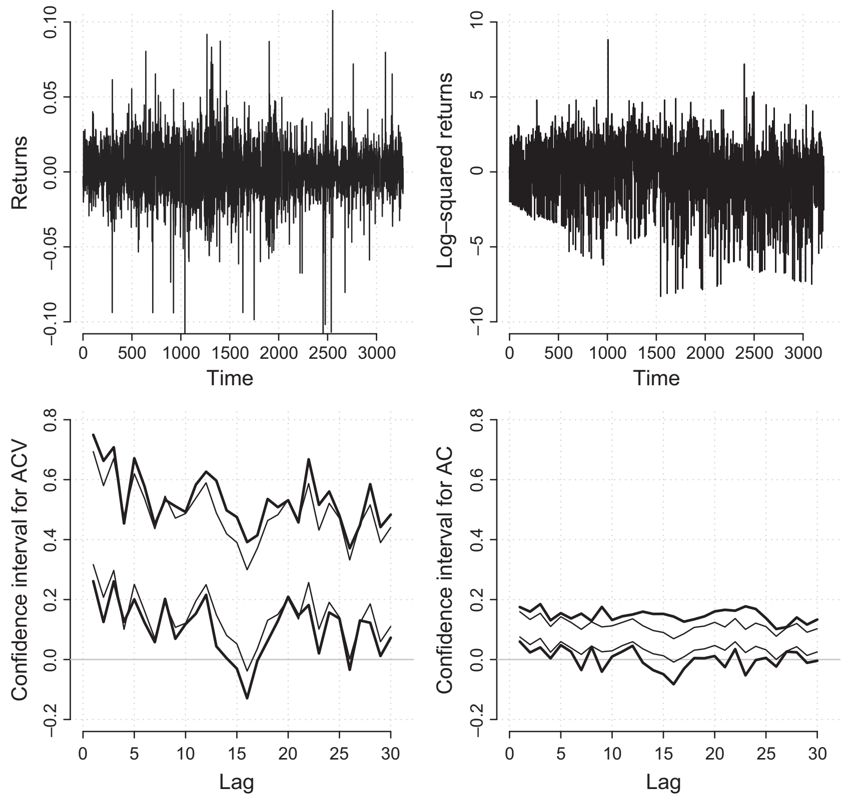

Figure 9. Top: returns and the (centred) log-squared returns. Bottom: 95% confidence intervals for γ_h and ϱ_h, subsampling symmetric (thick line, b adaptively-chosen, $\rho = 0.6$, Trapezoidal taper) and 'standard' (thin line)

Finally, we apply our methodology to the log-squared returns of Merck & Co., Inc. stock, from 3, January 1995 through 31 December, 2007, with $N = 3211$ (61 zero values for returns were removed). This transformed series can be consider as a proxy for the volatility. The returns $\{r_t\}$ and the (centred) volatility series $\{X_t\}$, $X_t = \log(r_t^2)$, are displayed in the upper panels of Figure 9. Linear models for log-squared returns have been proposed in a series of articles (Breidt *et al.*, 1998; Harvey, 1998; Deo *et al.*, 2006) and hence it is not unreasonable to apply the methods of this study to the Merck data. The corresponding 95% subsampling symmetric confidence intervals for the acvs and acs of $\{X_t\}$ (b from GRBS rule, $\rho = 0.6$, Trapezoidal taper) are drawn as thick lines in the bottom panels of the same figure.

We compare the subsampling confidence intervals with the ones obtained by a 'standard' method in which X_ts are considered as a realization from a high-order moving average process [here an MA(20) process]. In such case the 95% confidence interval for γ_h is obtained using the Bartlett formula, as described in Brockwell and Davis (1991): $CI_{0.95}(\gamma_h) = \left[\widehat{\gamma}_{h,N} \pm 1.96\sqrt{v_{hh}}/\sqrt{N}\right]$, where the asymptotic variance v_{ii} is given by

$$v_{ii} = (\eta - 3)\gamma_i^2 + \gamma_0^2 + \gamma_i^2 + 2\sum_{k=1}^{q}\gamma_i^2 + 2\sum_{k=1}^{q-i}\gamma_{k+i}\gamma_{k-i}$$

and for $i > q$, $v_{ii} = v_{qq}$. The 95% confidence for ϱ_h is $CI_{0.95}(\varrho_h) = \left[\widehat{\varrho}_{h,N} \pm 1.96\sqrt{w_{hh}}/\sqrt{N}\right]$ and

$$w_{ii} = \sum_{k=1}^{q-i}\rho_{k+i}^2 + 2\sum_{k=1}^{q-i}\rho_{k+i}\rho_{k-i} + \sum_{k=1}^{q+i}\rho_{k-i}^2 - 4\rho_i\sum_{k=1}^{q-i}\rho_k\rho_{k+i} - 4\rho_i\sum_{k=1}^{q}\rho_k\rho_{k-i} + 4\rho_i^2\sum_{k=1}^{q}\rho_k^2.$$

For $i > q$ $w_{ii} = 1 + 2(\rho_1^2 + \ldots + \rho_q^2)$. In both cases, the unknown quantities were estimated.

The subsampling confidence intervals for acvs and acs are different from the standard ones and tend to be wider. Both types, subsampling and standard intervals for the autocorrelations, suggest some degree of persitance in volatility, however in the subsampling approach fewer ϱ_hs are significant at 95% level compared to the standard approach.

6. SUMMARY

In this study we propose a self-normalization for the sample acvs and acs of a finite-variance linear, long-memory time series, whose innovations may possibly be heavy-tailed with tail index $2 < \alpha < 4$. The aim of this construction is to eliminate the dependence on parameters d and α from the rate of convergence of the sample acvs and on parameter d from the rate of convergence of the sample acs. In the case of acvs, this is achieved by combining the sample fourth moment (tail-adjustment) and the tapered sample acvs (memory-adjustment) – a HAC estimator; in the case of acs, the sample fourth moment is substituted by the sample size.

New theoretical results establish joint weak convergences of the sample acvs with the HAC estimator, as well as the sample fourth moment, and are later utilized in the convergence of the sample acs. The fixed bandwidth-ratio approach is taken in the development of the HAC asymptotics, as in Kiefer and Vogelsang (2005). To our knowledge, this is the first attempt to quantify and estimate variation in the sample acv and ac through a HAC estimator, such that the result is robust with respect to long memory (let alone high kurtosis).

The subsampling approach is employed to approximate the quantiles of the sampling distribution of the self-normalized acv (respectively, ac), as the corresponding nonstandard asymptotic distribution remains dependent upon d and α (respectively, d). The asymptotic properties of the subsampling estimator is here based upon the concept of θ-weak dependence introduced in Doukhan and Louhichi (1999), and consistency is established under some additional conditions. To choose a block size, an adaptive procedure of Götze and Račkauskas (2001) and Bickel and Sakov (2008) is implemented.

In the simulation study, realizations from processes with various combinations of tail and memory parameters, and of two different lengths, 500 and 1000, are investigated; the influence of both tapers and bandwidths is considered as well.

The empirical coverage rates for a 95% nominal level and automatically selected b are adequate, though with some tendency towards under-coverage in the memory-dominated convergence region. Finally, confidence-interval methodology is illustrated on the acvs and acs of log-squared returns of Merck stock, suggesting the presence of persistence in volatility for that time series.

APPENDIX

We begin with two auxiliary lemmas that have independent theoretical interest, and are used in the proofs of the theorems. First define $W_t = Z_t^2 - \sigma^2$, which is mean zero and heavy-tailed.

Lemma 1. *Under the conditions of Theorem* 1,

$$\mathbb{E}\exp\left\{ iva_N^{-2}\sum_{t=1}^N W_t - \frac{\omega^2}{2}a_N^{-4}\sum_{t=1}^N W_t^2 \right\} \to \exp\{g(v,\omega)\}$$

with g defined via eqn(1).

Next, we require another result – a generalization of Lemma 4.2 of HK – which uses the concept of a Donsker interpolation, defined via a collection of random sequences $\{W_N(k)\}$ with the property that for all $s < r$,

$$\sum_{[sN]\le k\le [rN]} W_N(k) \stackrel{\mathcal{L}}{\Rightarrow} W(r) - W(s). \tag{A.1}$$

Lemma 2. *Suppose each* $W_N(k)$ *has all finite moments, with mean zero and variance* N^{-1}*, such that* (A.1) *is satisfied. Let* $c_N(x,y) = \sum_{j\neq k} C_N(j,k) 1_{[j/n, j+1/n)}(x) 1_{[k/n,k+1/n)}(y)$ *be defined for some sequence of coefficients* $C_N(j,k)$*, and* $a_N(x,r) = \sum_k A_N(k,r) 1_{[k/n,k+1/n)}(x)$ *be defined for any* $r \in [0,1]$ *for some sequence of coefficients* $A_N(k,r)$*. Suppose that these functions each converge in* $\mathbb{L}_2$ *to functions* $c(x,y)$ *and* $a(x,r)$ *for each* r *fixed. Then for every* $r \in [0,1]$

$$\left(\sum_k A_N(k,r)W_N(k), \sum_{j\neq k} C_N(j,k)W_N(j)W_N(k)\right) \stackrel{\mathcal{L}}{\Rightarrow} \left(\int_{-\infty}^{\infty} a(x,r)W(dx), \int\int c(x,y)W(\mathrm{d}x)W(\mathrm{d}y)\right)$$

as $N \to \infty$. This result also holds as a joint convergence over any collection $\{r_\ell\}_{\ell=1}^m$.

Next, we present proofs of all the results.

Proof of Lemma 1. *Using integration by parts, we can show that*

$$
\begin{aligned}
&\log \mathbb{E}\exp\left\{ iv a_N^{-2}\sum_{t=1}^{N} W_t - \frac{\omega^2}{2} a_N^{-4}\sum_{t=1}^{N} W_t^2 \right\} \\
&= N\log\left(1+\int_{-\infty}^{\infty}\left[\exp\left\{\frac{iv}{a_N^2}(z^2-\sigma^2)-\frac{\omega^2}{2a_N^4}(z^2-\sigma^2)^2\right\}-1-\frac{iv}{a_N^2}(z^2-\sigma^2)\right]\mathrm{d}F(z)\right) \\
&= N\log\left(1+\int_{0}^{\infty}\left[\exp\left\{\frac{iv}{a_N^2}(z^2-\sigma^2)-\frac{\omega^2}{2a_N^4}(z^2-\sigma^2)^2\right\}\left(\frac{2ivz}{a_N^2}-\frac{2\omega^2 z(z^2-\sigma^2)}{a_N^4}\right)-\frac{2ivz}{a_N^2}\right]\overline{F}(z)\mathrm{d}z\right. \\
&\left.-\int_{-\infty}^{0}\left[\exp\left\{\frac{iv}{a_N^2}(z^2-\sigma^2)-\frac{\omega^2}{2a_N^4}(z^2-\sigma^2)^2\right\}\left(\frac{2ivz}{a_N^2}-\frac{2\omega^2 z(z^2-\sigma^2)}{a_N^4}\right)-\frac{2ivz}{a_N^2}\right]F(z)\mathrm{d}z\right) \\
&= N\log\left(1+\int_{0}^{\infty}\left[\exp\left\{iv(w^2-\frac{\sigma^2}{a_N^2})-\frac{\omega^2}{2}(w^2-\frac{\sigma^2}{a_N^2})^2\right\}\left(2ivw-2\omega^2 w(w^2-\frac{\sigma^2}{a_N^2})\right)-2ivw\right]\overline{F}(a_N w)\mathrm{d}w\right. \\
&\left.+\int_{0}^{\infty}\left[\exp\left\{iv(w^2-\frac{\sigma^2}{a_N^2})-\frac{\omega^2}{2}(w^2-\frac{\sigma^2}{a_N^2})^2\right\}\left(2ivw-2\omega^2 w(w^2-\frac{\sigma^2}{a_N^2})\right)-2ivw\right]F(a_N w)\mathrm{d}w\right).
\end{aligned}
$$

Now use the tail behaviour of the right and left cdfs, along with the Dominated Convergence Theorem and the Taylor series expansion of the logarithm, to obtain convergence to $\exp\{g(v,\omega)\}$, noting that $p + q = 1$.

Proof of Lemma 2. *Using the Cramer–Wold device, we must examine*

$$
\begin{aligned}
&\sum_{\ell=1}^{m}\alpha_\ell\sum_{k}A_N(k,r_\ell)W_N(k)+\beta\sum_{j\neq k}C_N(j,k)W_N(j)W_N(k) \\
&=\sum_{\ell=1}^{m}\alpha_\ell\int a_N(x,r_\ell)W_N(\mathrm{d}x)+\beta\int\int c_N(x,y)W_N(\mathrm{d}x)W_N(\mathrm{d}y).
\end{aligned}
$$

If we could substitute $a(x,r_\ell)$ for $a_N(x,r_\ell)$ and $c(x,y)$ for $c_N(x,y)$, the result would follow by the functional central limit theorem assumed for the Donsker interpolation. The discrepancy is

$$
\sum_{\ell=1}^{m}\alpha_\ell\int[a(x,r_\ell)-a_N(x,r_\ell)]W_N(\mathrm{d}x)+\beta\int\int[c(x,y)-c_N(x,y)]W_N(\mathrm{d}x)W_N(\mathrm{d}y),
$$

which we must show tends to zero as $N \to \infty$. To show the first part is asymptotically negligible, we can take a second moment and use standard arguments, along with the assumed $\mathbb{L}_2$ convergence. For the second part, we can mimic the argument of Lemma 4 of Surgailis (1982); also see Lemma 4.2 of HK.

Proof of Theorem 1. *We begin by evaluating the joint FLT described in the theorem, introducing an independent i.i.d. standard normal time series* $\{N_t\}$*:*

$$
\mathbb{E}\exp\left\{i\sum_{h=0}^{H}v_h N a_N^{-2}d_{N,h}-\frac{\omega^2}{2}a_N^{-4}S_N^{(4)}\right\}=\mathbb{E}\exp\left\{i\sum_{h=0}^{H}v_h N a_N^{-2}d_{N,h}+i\omega a_N^{-2}\sum_{t=1}^{N}X_t^2 N_t\right\}.
$$

In the second term we can expand X_t^2, and retain only the diagonal portion given by

$$
i\omega a_N^{-2}\sum_j \psi_j^2\sum_{t=1}^{N}Z_{t-j}^2 N_t = i\omega a_N^{-2}\sum_j\psi_j^2\sum_{t=1}^{N}W_{t-j}N_t + i\omega\sigma^2 a_N^{-2}\Psi_0\sum_{t=1}^{N}N_t.
$$

Here $W_t = Z_t^2 - \sigma^2$ by definition. Since $2 < \alpha < 4$, $a_N^{-2}\sum_{t=1}^{N}N_t = o_P(1)$. For the off-diagonal term in the expansion of the sum of X_t^2, it is shown to be asymptotically negligible:

$$
a_N^{-2}\sum_{j\neq k}\psi_j\psi_k\sum_{t=1}^{N}Z_{t-j}Z_{t-k}N_t \xrightarrow{P} 0
$$

as $N \to \infty$. This is established by noting that the above quantity is the normalized partial sum of $\sum_{j \neq k} \psi_j \psi_k Z_{t-j} Z_{t-k} N_t$, which is mean zero and uncorrelated in time, so that the sum is $O_P(N)$. So the FLT is asymptotic to

$$\mathbb{E}\exp\left\{ i \sum_{h=0}^{H} v_h a_N^{-2} \sum_j \psi_j \psi_{j+h} \sum_{t=1}^{N} W_{t-j} - \frac{\omega^2}{2} a_N^{-4} \sum_{t=1}^{N} \left(\sum_j \psi_j^2 W_{t-j} \right)^2 \right\}.$$

Once again, we can expand the square in the second term, obtaining

$$a_N^{-4} \sum_{t=1}^{N} \sum_j \psi_j^4 W_{t-j}^2 + a_N^{-4} \sum_{t=1}^{N} \sum_{j \neq k} \psi_j^2 \psi_k^2 W_{t-j} W_{t-k}.$$

For any $h \neq 0$ the series $\{W_t W_{t+h}\}$ is heavy-tailed of index $\alpha/2$, with rate of growth for the partial sums given by a_N^2, utilizing Theorem 3.3 of Davis and Resnick (1986); also see Lemma 1 of McElroy and Politis (2007). Now in the process $\{\sum_j \psi_j^2 W_{t-j}\}$ the MA coefficients satisfy $\psi_j^2 \sim C_d^2 j^{2d-2}$, and hence are summable (since $d < 1/2$). Although the coefficients do not satisfy condition (4.2) of Davis and Resnick (1986), this assumption is not really needed in their Proposition 4.2, as the proof reveals. Applying this proposition shows that the off-diagonal terms are $O_P(a_N^{-2})$, which only leaves $a_N^{-4} \sum_{t=1}^{N} \sum_j \psi_j^4 W_{t-j}^2$. Hence the FLT reduces to

$$\mathbb{E}\exp\left\{ i \sum_{h=0}^{H} v_h a_N^{-2} \sum_j \psi_j \psi_{j+h} \sum_{t=1}^{N} W_{t-j} - \frac{\omega^2}{2} a_N^{-4} \sum_{t=1}^{N} \sum_j \psi_j^4 W_{t-j}^2 \right\}.$$

Now using truncation arguments – since the sequences $\{\psi_j \psi_{j+h}\}$ are summable in j for each h, and $\{\psi_j^4\}$ is summable as well – as in the proof of Theorem 4.1 of Davis and Resnick (1985), the result follows from Lemma 1. The derivations for g in Remark 1 are given here for eqn (2):

$$\begin{aligned} g(0,y) &= \int_0^\infty \exp\{-y^2 z^4/2\}(-2y^2) z^{3-\alpha} \mathrm{d}z \\ &= -|y|^{\alpha/2} \int_0^\infty \exp\{-z^2/2\} z^{(2-\alpha)/2} \mathrm{d}z \\ &= -|y|^{\alpha/2} \frac{\sqrt{2\pi}}{2} 2^{(2-\alpha)/4} \pi^{-1/2} \Gamma(1-\alpha/4) \end{aligned}$$

by p.142 of Samorodnitsky and Taqqu (1994).

PROOF OF THEOREM 2. *First, note that a representation for a Fractional Brownian Motion (FBM) – in terms of W – is given by*

$$B(t) = \int_{-\infty}^{\infty} f_t(x) W(\mathrm{d}x) \tag{A.2}$$

by Proposition 7.2.6 of Samorodnitsky and Taqqu (1994), where $f_t(x) = (t-x)_+^d - (-x)_+^d = d \int_0^t (v-x)_+^{d-1} \mathrm{d}v$. We will consider this representation of FBM for $0 \leq t \leq 1$, observing that $B(0) = 0$. Letting $S_n = \sum_{t=1}^{n} X_t$ for any positive integer n, the interpolated process $S_{[Nr]}$ satisfies a functional central limit theorem under a linear hypothesis and second moments condition – see results in Marinucci and Robinson (2000):

$$V_N^{-1/2} S_{[N\cdot]} \overset{\mathcal{L}}{\Rightarrow} B$$

as $N \to \infty$ on the Skorohod space of functions on [0,1]. The normalizing variance V_N is defined to be $V_N = \mathrm{Var}(S_N)$, and is asymptotic to $N^{2d-1} C_d^2 \sigma^2$ by standard calculations. (This result requires greater than 2 moments and $0 \leq d < 1/2$.)

Note that the conditions of the theorem guarantee that Lemma 2 holds, and we will extend this result slightly. Let $\sum_{j \neq k} C_N(j,k) W_N(j) W_N(k)$ be denoted by C_N, and consider the sequence of stochastic processes defined via $\sum_k A_N(k,r) W_N(k)$, called $A_N(r)$ for short. Also let $A(r) = \int a(x,r) W(\mathrm{d}x)$, so we have $A_N(r) \overset{\mathcal{L}}{\Rightarrow} A(r)$ for each $r \in [0,1]$. If $A_N(\cdot)$ is a tight sequence of stochastic processes, then

$$(A_N(\cdot), C_N) \overset{\mathcal{L}}{\Rightarrow} \left(A, \int \int c(x,y) W(\mathrm{d}x) W(\mathrm{d}y) \right)$$

as a convergence of stochastic processes, since the finite-dimensional distributions converge by Lemma 2, and the tightness of C_N (trivial, since it is bounded in probability).

We intend to apply these results with $A_N(k,r) = N^{-d} \sum_{t=1}^{[Nr]} \psi_{t-k}$ and $a(x,r) = C_d f_r(x)/d$, noting that

$$N^{-d-1/2} S_{[Nr]} = N^{-d-1/2} \sum_k Z_k \sum_{t=1}^{[Nr]} \psi_{t-k} = N^{-1/2} \sum_k A_N(k,r) Z_k.$$

Of course, Lemma 2 has a stringent moment assumption on the Donsker sequence. We resolve this in the same manner as in HK: first establish the results using an approximating bounded sequence of input variables – as in their Lemma 4.1 – and then check that the error arising from this substitution is negligible, as in their Lemma 5.5. (The calculations for the partial sum process are similar, and actually easier, then those for $r_{N,h}$.)

So modulo this approximation argument, we identify $N^{-1/2}Z_k/\sigma$ with the Donsker interpolation process $W_N(k)$, so that then $N^{-d-1/2}S_{[Nr]} = A_N(r)$ in the notation above. We know from other results cited above, that the normalized sum obeys a functional central limit theorem, and hence $A_N(\cdot)$ is tight. Moreover,

$$\sum_{t=1}^{[Nr]} \psi_{t-k} = \sum_{t=1}^{[Nr]} (t-k)_+^{d-1} K(t-k) = N^{-1} \sum_{t=1}^{[Nr]} N^d (t/n - k/n)_+^{d-1} K(t-k),$$

and so $a_N(x,r) \sim \int_0^r (v - x)_+^{d-1} \mathrm{d}v C_d = a(x,r)$, with error of order 1/N by properties of the Riemann integral. We also have $\mathbb{L}_2$ convergence of $a_N(\cdot,r)$ to $a(\cdot,r)$ for each $r \in [0,1]$.

Following the development of Lemma 5.4 of HK, we can let $C_N(j,k) = \sum_{h=0}^H \nu_h \sum_{t=1}^N N^{1-2d} \psi_{t-k}\psi_{t+h-j}$ with $c_N(x,y)$ defined accordingly; then $c_N \to c$ in $\mathbb{L}_2$ (we can also apply Lemma 4.3 of HK, since $d > 1/4$ by assumption). Then we may apply Lemma 2, as well as its extension to stochastic processes, learning that

$$\left(N^{-d-1/2} S_{[Nr]}, N^{1-2d} r_{N,h}\right) \overset{\mathcal{L}}{\Rightarrow} \left(\sigma \frac{C_d}{d} \int f_r(x) W(\mathrm{d}x), \sigma^2 \int\int c(x,y) W(\mathrm{d}x) W(\mathrm{d}y)\right)$$

jointly in $0 \le h \le H$, as a functional central limit theorem in r. Of course the left hand limit is just $B(r)$, the FBM process, scaled by $\sigma C_d/d$. Now we use approximation results for $V_{\Lambda,M}$ from McElroy and Politis (2011a), obtaining

$$\begin{aligned} N^{-2d} V_{\Lambda,M} \overset{\mathcal{L}}{\Rightarrow} & -\sigma^2 \frac{C_d^2}{\rho^2 d^2} \int_0^1 \int_0^1 \ddot{\Lambda}\left(\frac{r-s}{\rho}\right) B(r)B(s)\mathrm{d}r\mathrm{d}s \\ & - \frac{2\sigma^2 C_d^2}{\rho d^2} \dot{\Lambda}_+(c) \int_0^{1-c\rho} B(r)B(r+c\rho)\mathrm{d}r + \frac{2\sigma^2 C_d^2}{\rho d^2} \dot{\Lambda}_-(1) \int_0^{1-\rho} B(r)B(r+\rho)\mathrm{d}r. \end{aligned}$$

Utilizing the expression (A.2), we arrive at the results of the theorem and the proof is complete. Note that the HAC limits in McElroy and Politis (2011b) involve functionals of Fractional Brownian Bridge (FBB), rather than FBM, since the acv in that work is centered by the sample mean. The FBB is given by

$$\widetilde{B}(t) = B(t) - tB(1) = \int_{-\infty}^{\infty} g_t(x) W(\mathrm{d}x) \qquad (A.3)$$

where $g_t(x)$ is defined by $g_r(x) = (r-x)_+^d - (-x)_+^d - r(1-x)_+^d + r(-x)_+^d$. □

PROOF OF PROPOSITION 1. *First, eqn* (4) *follows from Marinucci and Robinson* (2000) *and Theorem* 5 *of McElroy and Politis* (2011*a*), *since only* $2+\epsilon$ *moments are required in those results. The sample fourth moment satisfies a Law of Large Numbers whenever* $\mathbb{E}[Z4] < \infty$, *so eqn* (5) *follows. Third, eqn* (6) *follows from Lemma* 5.2 *of HK, whereas eqn* (7) *follows from Lemma* 5.3 *of the same work.*

PROOF OF THEOREM 3. *To organize results, consider four regions of the parameter space, each one a subset of* $\alpha \in (2,\infty)$ *cross* $d \in [0,1/2)$. *Region I* $= (2,4)\times(1/4,1/2)$; *Region II* $= (2,4)\times[0,1/4)$; *Region III* $= (4,\infty)\times[0,1/4)$; *Region IV* $= (4,\infty)\times(1/4,1/2)$. *Then we can summarize the asymptotics of* $d_{N,h}$, $r_{N,h}$, *and* $S_N^{(4)}$ *on each region (the convergence of* $V_{\Lambda,M}$ *is the same in all four regions, though in Regions I and IV it converges jointly with* $r_{N,h}$).

Region I: $Na_N^{-2} d_{N,h} \overset{\mathcal{L}}{\Rightarrow} \Psi_h \cdot U$ by Theorem 1, and also $a_N^{-4} S_N^{(4)} \overset{\mathcal{L}}{\Rightarrow} V$. Also $N^{1-2d} r_{N,h} \overset{\mathcal{L}}{\Rightarrow} \sigma^2 \int\int c(x,y)W(\mathrm{d}x)W(\mathrm{d}y)$ by Theorem 2. Note that Region I is split between cases B and C, along the curve $d = 1/\alpha$.

Region II: $Na_N^{-2} d_N \overset{\mathcal{L}}{\Rightarrow} \Psi_h \cdot U$ by Theorem 1, as well as $a_N^{-4} S_N^{(4)} \overset{\mathcal{L}}{\Rightarrow} V$. But $a_N^{-2} r_{N,h} \xrightarrow{P} 0$ by eqn (6). Region II is wholly contained within case B.

Region III: $N^{1/2}(\widehat{\gamma}_{h,N} - \gamma_h) \overset{\mathcal{L}}{\Rightarrow} G_h$ by Theorem 3.3a of HK, so the analysis of $d_{N,h}$ and $r_{N,h}$ becomes irrelevant. Also $N^{-1} S_N^{(4)} \xrightarrow{P} \mathbb{E}[X^4]$ by eqn (5). Region III coincides with case A.

Region IV: $N^{1-2d} d_{N,h} \xrightarrow{P} 0$ by eqn (7), while $N^{1-2d} r_{N,h} \overset{\mathcal{L}}{\Rightarrow} \sigma^2 \int\int c(x,y)W(\mathrm{d}x)W(\mathrm{d}y)$ by Theorem 2. Also $N^{-1} S_N^{(4)} \xrightarrow{P} \mathbb{E}[X^4]$ by eqn (5). Region IV is wholly contained within case C.

These results can now be spliced with some book-keeping. The dominant rate in case B (Region II and part of Region I) is Na_N^{-2}; since $a_N^{-2} N^{2d} \to 0$ (the slowly varying function L grows slower than any polynomial power of N), the remainder term $r_{N,h}$ dies off, as does the normalization $V_{\Lambda,M}$. This produces the case B asymptotics.

For case A, the appropriate rate is $N^{1/2}$, noting that $N^{-1/2} N^{2d} \to 0$ here, guaranteeing that $V_{\Lambda,M}$ has no impact. For case C, the appropriate rate is N^{1-2d}, and this time $N^{-2d} a_N^2 \to 0$, in both Regions I and IV. Thus the normalization by $S_N^{(4)}$ has no impact. This completes the proof.

Proof of Theorem 4. *We prove the result for the marginals, noting that the joint result follows by Cramer–Wold fairly simply. Let $h \geq 1$ be fixed, and let $\phi_j = \psi_{j+h} - \varrho_h \psi_j$ for each j be defined, its dependency on h being suppressed in the notation. Then from eqn* (8) *we obtain the decomposition*

$$\widehat{\varrho}_{h,N} - \varrho_h = \widehat{\gamma}_{0,N}^{-1} N^{-1} \sum_{t=1}^{N} \sum_{j \geq 0} \psi_j \phi_j W_{t-j} + \widehat{\gamma}_{0,N}^{-1} N^{-1} \sum_{t=1}^{N} \sum_{j \neq \ell} \psi_j \phi_\ell Z_{t-j} Z_{t-\ell}.$$

Label the terms on the right as $F_{N,h}$ and $E_{N,h}$ respectively, which correspond to diagonal and off-diagonal terms for the autocorrelations. Although in the final analysis the convergences only depend upon whether d is less than or greater than 1/4, it is convenient to proceed according to zones A, B and C, so that we can utilize prior results. Focusing on the numerator of $Q_{h,N}$, we note that in zone A the convergence $\sqrt{N}(\widehat{\varrho}_{h,N} - \varrho_h) \stackrel{\mathcal{L}}{\Rightarrow} \gamma_0^{-1}[G_h - \varrho_h G_0]$ follows from Theorem 3, which works out to be Gaussian with variance

$$(1 + 2\varrho_h^2) \sum_j \gamma_j^2 - 4\varrho_h \sum_j \gamma_j \gamma_{j+h} + \sum_j \gamma_{j-h} \gamma_{j+h} \tag{A.4}$$

(all terms involving the fourth moment η cancel out) divided by γ_0^2. In zone C we obtain the convergence

$$N^{1-2d}(\widehat{\varrho}_{h,N} - \varrho_h) \stackrel{\mathcal{L}}{\Rightarrow} \frac{(1-\varrho_h)}{\gamma_0} \int \int c(x,y) W(\mathrm{d}x) W(\mathrm{d}y), \tag{A.5}$$

and the limit is nonzero by assumption on ϱ_h; also note that this limit holds jointly with the limit of $V_{\Lambda,M}$. On zone B, we obtain $N a_N^{-2} F_{N,h} \xrightarrow{P} 0$ by Theorem 1. So it remains to consider $E_{N,h}$ on zone B; when $d > 1/4$ we can apply Theorem 2 to obtain the same limit as in (A.5), and again the result is joint with $V_{\Lambda,M}$. Hence the behavior on this slice of zone B is identical to the behavior in zone C. But when $d < 1/4$ some additional analysis is needed, along the lines of Proposition 4.1 of Davis and Resnick (1986). Letting $\xi_t = \sum_j \sum_{k \neq 0} \psi_j \phi_{j-k} Z_{t-j} Z_{t-j+k}$ (which depends on h), we have $E_{N,h} = \widehat{\gamma}_{0,N}^{-1} N^{-1} \sum_{t=1}^{N} \xi_t$. These $\{\xi_t\}$ are a covariance stationary sequence with autocovariance function v_n given by

$$v_n = \mathbb{E}[\xi_n \xi_0] = \sigma^4 \sum_j \sum_{k \neq 0} \psi_j \phi_{j-k} \left[\psi_{j-n} \phi_{j-n-k} + \psi_{j-n-k} \phi_{j-n} \right] \tag{A.6}$$

after simplifications. This sequence is absolutely summable when $d < 1/4$ – HK shows it decays at rate n^{4d-2} – and in fact $\sum_n v_n$ equals (A.4). We then establish a CLT for the partial sum $\sum_{t=1}^{N} \xi_t$ by first truncating the MA representation of the variables as

$$\xi_t^{(m)} = \sum_{j=0}^{m} \sum_{k \neq 0} \psi_j \phi_{j-k} Z_{t-j} Z_{t-j+k};$$

it is immediate from (A.6) that this new sequence $\{\xi_t^{(m)}\}$ is m-dependent. So long as the sum of its autocovariances is non-zero [and we can choose m sufficiently large to guarantee this is always true, given its limit of (A.4)], we can then apply Theorem 6.4.2 of Brockwell and Davis (1991) to conclude the CLT holds. Relaxing the truncation (letting $m \to \infty$) now follows standard arguments (Proposition 6.3.9 of Brockwell and Davis, 1991), noting that the coefficient sequences are absolutely summable. Hence, $\sqrt{N} E_{N,h} \stackrel{\mathcal{L}}{\Rightarrow} \mathcal{N}(0, \sum_n v_n / \gamma_0^2)$. So clearly the sample ac limit in zone B where $d < 1/4$ is identical to that of zone A. Splicing all results together, we have either a CLT for the sample acs when $d < 1/4$, or the functional BM limit when $d > 1/4$, and α is irrelevant. The rate of convergence is $N^{1-\max\{2d,1/2\}}$.

As for the denominator, we recall from Proposition 1 that $\sqrt{N}/V_{\Lambda,M} \xrightarrow{P} 0$ when $d > 1/4$ and $V_{\Lambda,M}/\sqrt{N} \xrightarrow{P} 0$ when $d < 1/4$. Moreover in the former case the convergence of the numerator of $Q_{h,N}$ is joint with the HAC estimator (at rate N^{-2d}), which establishes the $d > 1/4$ case of Theorem 4. But when $d < 1/4$ the studentized statistic is asymptotic to just $\sqrt{N}$ times $\widehat{\varrho}_{h,N} - \varrho_h$, which tends to $\gamma_0^{-1}(G_h - \rho_h G_0)$. This concludes the proof. □

Acknowledgements

We are thankful for the comments from an anonymous referee that improved the manuscript.

REFERENCES

Bardet, J.-M., Doukhan, P., Lang, G. and Ragache, N. (2008). *Dependent Lindeberg Central Limit Theorem and some applications*. ESAIM: Probab. Statist.**12**, 154–72.

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

Breidt, F. J., Crato, N. and de Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. *J. Econometrics* **83**, 325–48.

Brockwell, P. J. and Davis, R. A. (1991). *Time series: theory and methods*. Springer-Verlag, New York.

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.

Davis, R. A. and Resnick, S. I. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.* **13**, 179–95.

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.

Deo, R., Hurvich, C. and Lu, Y. (2006). Forecasting realized volatility using a long memory stochastic volatility model: estimation, prediction and seasonal adjustment. *J. Econometrics* **131**, 29–58.

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

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer-Verlag, Berlin.

Fitzsimmons, P. and McElroy, T. (2010). On joint Fourier-Laplace Transforms. *Commun. Statist. Theor. Meth.* **39**, 1883–5.

Götze, F. and Račkauskas, A. (2001). Adaptive choice of bootstrap sample sizes. In *State of the art in probability and statistics: Festschrift for Willem R. van Zwet, Papers from the symposium held at the University of Leiden, Leiden, March 23-26, 1999. Lecture Notes Monogr. Ser., 36, Cambridge* (eds. M.D. Gunst, C. Klassen and A.V.D. Vaart). Beach Wood: Cambridge University Press, pp. 286–309.

Harvey, A. (1998). Long memory in stochastic volatility. In *Forecasting volatility in financial markets, London* (eds J. Knight and S. Satchell). New York: Butterworth-Heinemann.

Horváth, L. and Kokoszka, P. (2008). Sample autocovariances of long-memory time series. *Bernoull*, **14**, 405–18.

Jach, A., McElroy, T. and Politis, D. (2012). Subsampling inference for the mean of heavy-tailed longmemory time series. *J. Time Ser. Anal.* **33**, 96–111.

Kiefer, N. and Vogelsang, T. (2005). A new asymptotic theory for heteroscedasticity-autocorrelation robust tests. *Econometric Theory* **21**, 1130–64.

Kulik, R. and Soulier, P. (2011). The tail empirical process of some long memory sequences. *Stochastic Process. Appl.* **121**, 109–34.

Marinucci, D. and Robinson, P. (2000). Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications*, **86**, 103–20.

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

McElroy, T. and Politis, D. N. (2011a). *Distribution theory for the studentized mean for long, short, and negative memory time series. Technical Report*. University of California, San Diego.

McElroy, T. and Politis, D. N. (2011b). Fixed-b asymptotics for the studentized mean from time series with short, long or negative memory. *Econometric Theory* **28**, 1–11.

Montanari, A., Rosso, R. and Taqqu, M. (1997). Fractionally differenced ARIMA models applied to hydrologic time series: identification, estimation, and simulation. *Water Resour. Res.*, **33**, 1035–44.

Politis, D. (2001). On nonparametric function estimation with infinite-order flat-top kernels. In *Probability and statistical models with applications, UK* (eds C. Charalambides, M. Koutras and N.Balakrishnan). Boca Raton: Chapman and Hall/CRC, pp. 469–83.

Politis, D. and Romano, J. (1995). Bias-corrected nonparametric spectral estimation. *J. Time Ser. Anal.* **16**, 67–104.

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

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable non-Gaussian random processes: stochastic models with infinite variance*. Chapman & Hall, New York.

Serroukh, A. and Walden, A. T. (2000). Wavelet scale analysis of bivariate time series II: statistical properties for linear processes. *J. Nonparamet. Statist.* **13**, 37–56.

Surgailis, D. (1982). Zones of attraction of self-similar multiple integrals. *Lietuvos Matematikos Rinkinys*, **22**, 185–201.