Econometric Theory, **21**, 2005, 1087–1111. Printed in the United States of America.
DOI: 10.1017/S0266466605050541

VALIDITY OF THE SAMPLING WINDOW METHOD FOR LONG-RANGE DEPENDENT LINEAR PROCESSES

DANIEL J. NORDMAN AND SOUMENDRA N. LAHIRI
Iowa State University

The sampling window method of Hall, Jing, and Lahiri (1998, *Statistica Sinica* 8, 1189–1204) is known to consistently estimate the distribution of the sample mean for a class of long-range dependent processes, generated by transformations of Gaussian time series. This paper shows that the same nonparametric subsampling method is also valid for an entirely different category of long-range dependent series that are linear with possibly non-Gaussian innovations. For these strongly dependent time processes, subsampling confidence intervals allow inference on the process mean without knowledge of the underlying innovation distribution or the long-memory parameter. The finite-sample coverage accuracy of the subsampling method is examined through a numerical study.

1. INTRODUCTION

This paper considers nonparametric distribution estimation for a class of random processes that exhibit strong or long-range dependence. Here we classify a real-valued stationary time process $\{Y_t\}, t \in \mathbb{Z}$ as long-range dependent (LRD) if its autocovariance function $r(k) = \mathrm{Cov}(Y_t, Y_{t+k})$ can be represented as

$$r(k) = k^{-\alpha} L_1(k), \qquad k \to \infty, \tag{1}$$

for some $0 < \alpha < 1$ and function $L_1 : (0,\infty) \to (0,\infty)$ that is slowly varying at infinity, that is, $\lim_{x\to\infty} L_1(\lambda x)/L_1(x) = 1$ for all $\lambda > 0$. Time series satisfying (1) often find application in astronomy, hydrology, and economics (Beran, 1994; Montanari, 2003; Henry and Zaffaroni, 2003).

For comparison, we note that weakly dependent processes are usually characterized by rapidly decaying, summable covariances (Doukhan, 1994). However, (1) implies that the sum of covariances $\sum_{k=1}^{\infty} r(k)$ diverges under long-range dependence. This feature of strongly dependent data often compli-

The authors thank two referees for comments and suggestions that greatly improved an earlier draft of the paper. This research was partially supported by U.S. National Science Foundation grants DMS 00-72571 and DMS 03-06574 and by the Deutsche Forschungsgemeinschaft (SFB 475). Address correspondence to Dan Nordman, Department of Statistics, Iowa State University, Ames, IA 50011, USA; e-mail: dnordman@iastate.edu.

cates standard statistical inference based on the sample mean $\overline{Y}_n$ of a stretch of observations $Y_1,\dots,Y_n$. For one reason, the variance of a size n sample mean $\overline{Y}_n$ decays to zero at a rate that is both slower than $O(n^{-1})$ and unknown in practice (Beran, 1994). The usual scaling factor $\sqrt{n}$ used with independent or weakly dependent data then fails to produce a limit distribution for $\overline{Y}_n$ under long-range dependence. Even if properly standardized, the sample mean can have normal in addition to nonnormal limit laws across various types of strongly dependent processes (Davydov, 1970; Taqqu, 1975). As a consequence, statistical approximations of the unknown sampling distribution of $\overline{Y}_n$ are necessary under long-range dependence, without making stringent assumptions on the underlying process or the strength of the dependence α, L_1 in (1).

For weakly dependent data, the moving block bootstrap of Künsch (1989) and Liu and Singh (1992) provides accurate nonparametric estimates of the sample mean's distribution. However, the block bootstrap has been shown to break down for a class of LRD processes where the asymptotic distribution of $\overline{Y}_n$ can be nonnormal (cf. Lahiri, 1993). These processes are obtained through transformations of certain Gaussian series (Taqqu, 1975, 1979; Dobrushin and Major, 1979). Although the bootstrap rendition of $\overline{Y}_n$ fails for transformed-Gaussian LRD processes, Hall, Jing, and Lahiri (1998) have shown that their so-called sampling window procedure can consistently approximate the distribution of the normalized sample mean for these same time series. This procedure is a subsampling method that modifies data-blocking techniques developed for inference with weakly dependent (mixing) data (Politis and Romano, 1994; Hall and Jing, 1996). With the aid of subsampling variance estimators, Hall et al. (1998) also developed a Studentized version of the sample mean along with a consistent, subsample-based estimator of its distribution.

In this paper, we establish the validity of the sampling window method of Hall et al. (1998) for a different category of LRD processes: *linear* LRD processes with an unknown innovation distribution. The subsampling method is shown to correctly estimate the distribution of normalized and Studentized versions of the sample mean under this form of long-range dependence, without knowledge of the exact dependence strength α or innovation structure. The results illustrate that subsampling can be applied to calibrate nonparametric confidence intervals for the mean $E(Y_t) = \mu$ of either a transformed-Gaussian LRD process (Hall et al., 1998) *or* a linear LRD series. That is, the same subsampling procedure allows nonparametric interval estimation when applied to two major examples of strongly dependent processes considered in the literature (Beran, 1994, Ch. 3).

The rest of the paper is organized as follows. In Section 2, we frame the process assumptions and some distributional properties of $\overline{Y}_n$. Main results are given in Section 3, where we establish the consistency of subsampling distribution estimation for the sample mean under linear long-range dependence. In Section 4, we report a simulation study on the coverage accuracy of a subsampling confidence interval procedure for the LRD process mean μ. A second numerical study also considers subsampling estimators for the distribution of the Studentized sample mean. In Section 5, we discuss the validity of the subsampling method for weakly dependent linear processes. Proofs of the main results are provided in Section 6.

2. PRELIMINARIES

2.1. Process Assumptions

We suppose that the observed data $\mathcal{Y}_n = \{Y_1, \dots, Y_n\}$ represent a realization from a stationary, real-valued LRD process $\{Y_t\}, t \in \mathbb{Z}$ that satisfies the following assumption.

Assumption L. For independent identically distributed (i.i.d.) innovations $\{\varepsilon_t\}, t \in \mathbb{Z}$ with mean $E(\varepsilon_t) = 0$ and $0 < E(\varepsilon_t^2) < \infty$, it holds that

$$Y_t = \mu + \sum_{j \in \mathbb{Z}} c_{t-j}\,\varepsilon_j, \qquad t \in \mathbb{Z},$$

where $E(Y_t) = \mu \in \mathbb{R}$ and the real sequence $\{c_j\}, j \in \mathbb{Z}$ is square summable $\sum_{j\in\mathbb{Z}} c_j^2 < \infty$ such that the autocovariance function $r(k) = \mathrm{Cov}(Y_t, Y_{t+k})$ admits a representation as in (1).

Assumption L encompasses two popular models for strong dependence: the fractional Gaussian processes of Mandelbrot and van Ness (1968) and the fractional autoregressive integrated moving average (FARIMA) models of Adenstedt (1974), Granger and Joyeux (1980), and Hosking (1981). For FARIMA processes in particular, we permit greater distributional flexibility through possibly non-Gaussian innovations. Note that a LRD FARIMA$(0, d, 0)$ series, $d \in (0, \frac{1}{2})$, admits a casual moving average representation

$$Y_t = \mu + \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}\,\varepsilon_{t-j}, \qquad t \in \mathbb{Z}, \tag{2}$$

involving the gamma function $\Gamma(\cdot)$. More general FARIMA series, for which (1) holds with $\alpha = 1 - 2d$ and constant $L_1(\cdot) = C_1 > 0$, follow from applying an autoregressive moving average (ARMA) filter to a process from (2) (cf. Beran, 1994).

We remark that the results of this paper also hold by stipulating the long-range dependence through regularity conditions, as in Theorem 2.2 of Hall et al. (1998), on the spectral density f of the process $\{Y_t\}$ for which $\lim_{x\to 0} f(x)/\{|x|^{\alpha-1}L_1(1/|x|)\} > 0$ exists finitely. Under certain conditions, this behavior of f at the origin is equivalent to (1) and serves as an alternative description of long-range dependence (Bingham, Goldie, and Teugels, 1987). However, our assumptions here on the LRD linear process are fairly mild, requiring i.i.d. innovations to only have a bounded second moment.

2.2. Distributional Properties of the Sample Mean

In the following discussion, for any two nonzero real sequences $\{s_n\}$ and $\{t_n\}$, we write $s_n \sim t_n$ if $\lim_{n\to\infty} s_n/t_n = 1$. With the proper scaling d_n, the asymptotic distribution of the normalized sample mean is known to be normal for Assumption L processes (cf. Davydov, 1970): as $n \to \infty$,

$$n(\overline{Y}_n - \mu)/d_n \xrightarrow{d} Z, \tag{3}$$

where Z represents a standard normal variable and $\xrightarrow{d}$ denotes convergence in distribution. However, setting confidence intervals for μ, based on $\overline{Y}_n$ and its large-sample normal distribution, becomes complicated for linear LRD processes. The covariance decay rate in (1) implies that the variance of $\overline{Y}_n$ converges to 0 as follows:

$$\mathrm{Var}(\overline{Y}_n) \sim n^{-\alpha} L(n), \tag{4}$$

for $L(\cdot) = 2\{(2-\alpha)(1-\alpha)\}^{-1} L_1(\cdot)$, which is slower than the usual $O(n^{-1})$ rate associated with weakly dependent data. Consequently, the correct scaling $d_n = \{n^{2-\alpha}L(n)\}^{1/2} \sim \{\mathrm{Var}(n\overline{Y}_n)\}^{1/2}$ for $n(\overline{Y}_n - \mu)/d_n$ to have a normal limit depends on the *unknown* quantities $\alpha, L(\cdot)$ from (4).

With additional assumptions on the linear process, unknown quantities in d_n could in principle be estimated directly for interval estimation of μ based on a normal approximation (3). For example, by assuming a constant function $L_1(\cdot) = C_1$ in (1) (along with additional regularity conditions on f), estimates $\hat{\alpha}, \hat{C}_1$ of α, C_1 could be obtained through various periodogram-based techniques (Bardet, Lang, Oppenheim, Philippe, Stoev, and Taqqu, 2003). However, after substituting such estimates directly into d_n from (3), the resulting Studentized mean $G_n = n(\overline{Y}_n - \mu)/\hat{d}_{n,\hat{\alpha},\hat{C}_1}$ may fail to adequately follow a normal distribution. To illustrate, we conducted a small numerical study of the coverage accuracy of confidence bounds for the mean μ of several LRD FARIMA processes, set with a normal approximation for a Studentized mean G_n. For these processes, (1) holds with a function $L_1(\cdot) = C_1$. We obtained a version G_n by estimating α and C_1 in d_n through popular log-periodogram regression (Geweke and Porter-Hudak, 1983) using the first $n^{4/5}$ Fourier frequencies (Hurvich, Deo, and Brodsky, 1998). The coverage probabilities in Table 1 suggest that the normal distribution may not always appropriately describe the behavior of a Studentized sample mean obtained through such plug-in estimates in (3). (The LRD processes in Table 1 involve filtered FARIMA$(0,d,0)$ series, but other simulation results indicate that a plug-in version G_n may produce better confidence intervals with unfiltered FARIMA$(0,d,0)$ series.) We remark that, with Gaussian LRD processes, Beran (1989) developed a modified normal distribution for $\overline{Y}_n$ after Studentization with a periodogram-based estimate of α. However, the approach given was globally parametric in requiring the form of the spectral density f to be known on the entire interval $[0, \pi]$, which is a strong assumption.

Table 1. Coverage probabilities for one-sided 90% lower and upper confidence bounds

		Standard Normal Innovations			Chi-Square Innovations		
n	Filter	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$
100	1	(76.1, 76.1)	(95.0, 95.4)	(98.2, 98.1)	(84.3, 80.8)	(97.4, 94.8)	(98.6, 97.5)
	2	(69.1, 72.7)	(81.0, 82.9)	(91.2, 93.9)	(80.9, 73.3)	(82.3, 78.2)	(95.7, 90.1)
400	1	(83.0, 82.8)	(98.6, 98.6)	(99.8, 99.8)	(92.3, 90.2)	(99.3, 98.9)	(99.9, 100)
	2	(78.1, 79.0)	(84.5, 85.6)	(89.4, 90.0)	(90.5, 88.9)	(87.6, 81.9)	(91.5, 88.7)
900	1	(85.8, 85.6)	(100, 99.6)	(99.9, 100)	(96.0, 95.9)	(99.9, 99.2)	(100, 100)
	2	(81.6, 81.7)	(86.7, 86.9)	(89.4, 88.4)	(93.8, 93.6)	(90.1, 84.5)	(89.0, 87.4)

Note: Confidence bounds are denoted in parentheses $(\cdot,\cdot)$, using the large-sample normal distribution of $G_n = n(\bar{Y}_n - \mu)/\hat{d}_{n,\hat{\alpha},\hat{C}_1}$. Computed probabilities are based on 1,000 simulations from a FARIMA process detailed in Section 4.1.

In Section 3, we show that the sampling window method produces consistent, nonparametric estimates of the finite-sample distribution of the sample mean from strongly dependent linear processes. Subsampling distribution estimators for $\bar{Y}_n$ can then be used to calibrate nonparametric confidence intervals for the process mean μ. Under linear long-range dependence, an advantage of this approach over traditional large-sample theory is that the subsampling confidence intervals may be constructed without making restrictive assumptions on the behavior of f near zero and without estimating the covariance parameter α. Another benefit of the subsampling method is its applicability to other formulations of long-range dependence involving nonlinear processes. That is, the subsampling technique has established validity with transformed-Gaussian LRD series as treated in Hall et al. (1998). For these series, $n(\bar{Y}_n - \mu)/d_n$ may have a nonnormal limit distribution, and a normal approximation for the sample mean might break down.

3. MAIN RESULTS: SUBSAMPLING DISTRIBUTION ESTIMATION

3.1. Result for the Normalized Sample Mean

We briefly present the subsampling estimator of the sampling distribution of the normalized sample mean $T_n = n(\bar{Y}_n - \mu)/d_n$, as prescribed in Hall et al. (1998). Denote the distribution function of T_n as $F_n(x) = P(T_n \le x)$, $x \in \mathbb{R}$. To capture the underlying dependence structure, the subsampling method creates several small-scale replicates of $Y_1, \dots, Y_n$ through data blocks or subsamples. Let $1 \le \ell \le n$ be the block length and denote $\mathcal{B}_i = (Y_i, \dots, Y_{i+\ell-1})$ as the ith overlapping data block, $1 \le i \le N = n - \ell + 1$. Treating each block as a scaled-down copy of the original time series, define the analog of T_n on each block $\mathcal{B}_i$ as $T_{\ell i} = (S_{\ell i} - \ell \bar{Y}_n)/d_\ell$, $1 \le i \le N$, where $S_{\ell i} = \sum_{j=i}^{i+\ell-1} Y_j$ represents a block sum.

The sampling window estimator $\hat{F}_n(x)$ of the distribution $F_n(x)$ is given by $\hat{F}_n(x) = N^{-1} \sum_{i=1}^{N} I\{T_{\ell i} \le x\}$, where $I\{\cdot\}$ denotes the indicator function. The subsampling estimator $\hat{F}_n$ is simply the empirical distribution of the subsample versions $T_{\ell i}$ of T_n. Hall et al. (1998) establish the consistency of $\hat{F}_n$ in estimating F_n with transformed-Gaussian LRD series. The following result extends the consistency of the subsampling estimator $\hat{F}_n$ to include a large class of LRD linear processes. Let $\xrightarrow{p}$ denote convergence in probability.

THEOREM 1. *If Assumption L holds and* $\ell^{-1} + n^{-(1-\delta)}\ell = o(1)$ *for some* $\delta \in (0,1)$*, then*

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F_n(x)| \xrightarrow{p} 0, \quad \text{as } n \to \infty.$$

Note that the correct scaling d_n, d_ℓ for centered sums $n(\bar{Y}_n - \mu)$, $(S_{\ell i} - \ell\mu)$ to have a normal limit in (3) depends on unknown quantities $\alpha, L(\cdot)$. In practice, these scaling factors need to be consistently estimated, and $\hat{F}_n$ must be appropriately modified, to set confidence intervals for μ. We next give a modified subsampling approach for accomplishing this.

3.2. Result for the Studentized Sample Mean

Following the setup in Hall et al. (1998), we first replace d_n in T_n with a data-based construct involving two subsampling variance estimates. To describe the estimate of d_n, let m_{1n}, $m_{2n} \in [1,n]$ denote integers such that for some $\theta \in (0,1)$ we have

$$m_{1n} \sim n^{(1+\theta)/2}, \qquad m_{2n} \sim n^{\theta} \tag{5}$$

as $n \to \infty$, implying further that $m_{1n}^2/m_{2n} \sim n$ holds. For $m \in [1,n]$, define $\tilde{d}_m^2 = (n-m+1)^{-1}\sum_{i=1}^{n-m+1}(S_{mi} - m\bar{Y}_n)^2$ where $S_{mi} = \sum_{j=i}^{i+m-1} Y_j$, $i \geq 1$. Here $\tilde{d}_m^2$ represents a subsampling variance estimator of $\text{Var}(S_{m1})$. Next define $\hat{d}_n^2 = \tilde{d}_{m_{1n}}^4/\tilde{d}_{m_{2n}}^2$ with the smoothing parameters m_{1n}, m_{2n}. We use $\hat{d}_n^2$ as an estimator of d_n^2 to obtain a Studentized version of the sample mean as $T_{1n} = n(\bar{Y}_n - \mu)/\hat{d}_n$.

To calibrate confidence intervals for μ based on T_{1n}, a subsampling estimator $\hat{F}_{1n}$ of the distribution function F_{1n} of T_{1n} can be constructed as follows. For each length ℓ block $\mathcal{B}_i$, $1 \leq i \leq N$, let $\hat{d}_{\ell i}^2$ denote the subsample version of $\hat{d}_n^2$ found by replacing $(Y_1,\ldots,Y_n)$ and n with $\mathcal{B}_i = (Y_i,\ldots,Y_{i+\ell-1})$ and ℓ in the definition of $\hat{d}_n^2$. Analogous to values m_{1n}, m_{2n} used in $\hat{d}_n^2$, each version $\hat{d}_{\ell i}^2$ requires subsample smoothing parameters $m_{1\ell}, m_{2\ell}$ that satisfy (5) with ℓ rather than n. Let $T_{1\ell,i} = (S_{\ell i} - \ell\bar{Y}_n)/\hat{d}_{\ell i}$, $1 \leq i \leq N$ denote the subsample replicates of T_{1n}. The subsampling estimator of F_{1n} is then given by $\hat{F}_{1n}(x) = N^{-1}\sum_{i=1}^{N} I\{T_{1\ell,i} \leq x\}$, $x \in \mathbb{R}$.

We show that the preceding subsampling estimator successfully approximates the distribution F_{1n} of the Studentized sample mean for long-memory linear processes. Hall et al. (1998) give an analogous result for transformed-Gaussian LRD series.

THEOREM 2. *In addition to the conditions of Theorem 1, assume m_{1n}, m_{2n} satisfy (5) and*

$$L^2(x^{1+c})/\{L(x^2)L(x^{2c})\} \to 1 \quad \text{as } x \to \infty \quad \text{for any } c > 0. \tag{6}$$

Then, as $n \to \infty$: (a) $\hat{d}_n^2/\text{Var}(n\bar{Y}_n) \xrightarrow{p} 1$; (b) $T_{1n} \xrightarrow{d} Z$, a standard normal variable; (c) $\sup_{x\in\mathbb{R}}|\hat{F}_{1n}(x) - F_{1n}(x)| \xrightarrow{p} 0$.

Condition (6) represents a weakened version of a similar assumption used by Hall et al. (1998, Thm. 2.5) and implies that the combination of subsampling variance estimators in $\hat{d}_n^2$ can consistently estimate d_n^2 under long-range dependence. For LRD fractional Gaussian and FARIMA processes, the function $L(\cdot)$

is constant in (4) and so easily satisfies (6) (Beran, 1994). Examples of other slowly varying functions that fulfill (6) include $L(x) = \tilde{L}(\log(x)) : (1,\infty) \to (0,\infty)$ based on an arbitrary slowly varying $\tilde{L}$, such as $L(x) = [\log\log(x)]^{c_1}$, $c_1 \in \mathbb{R}$ or $L(x) = \exp\{[\log\log(x)]^{c_2}\}$ for $0 \le c_2 < 1$. However, condition (6) is still restrictive in not permitting general slowly varying functions such as $L(x) = \log(x)$.

In the next section, we outline a procedure for constructing confidence intervals for the mean μ based on the subsampling result in Theorem 2.

3.3. Subsampling Confidence Interval Procedure

Let $\lfloor \cdot \rfloor$ denote the integer part function. For $\beta \in (0,1)$, let $\hat{t}_{\beta,n}$ denote the $\lfloor N\beta \rfloor$th order statistic of the N possible subsample versions $T_{1\ell,i}$, $1 \le i \le N$, of T_{1n}. Here $\hat{t}_{\beta,n}$ represents the β-percentile of the subsampling estimator $\hat{F}_{1n}$ taken as an estimate of the same percentile of F_{1n}. Using T_{1n} and $\hat{F}_{1n}$, we set approximate one-sided lower and upper $100(1-\beta)\%$ confidence bounds for μ as $L_{1-\beta,n} = \bar{Y}_n - n^{-1}\hat{d}_n \hat{t}_{1-\beta,n}$ and $U_{1-\beta,n} = \bar{Y}_n - n^{-1}\hat{d}_n \hat{t}_{\beta,n}$, respectively. These subsampling bounds have asymptotically correct coverage under Theorem 2, namely, $\lim_{n\to\infty} P(\mu > L_{1-\beta,n}) = \lim_{n\to\infty} P(\mu < U_{1-\beta,n}) = 1-\beta$. An approximate two-sided $100(1-\beta)\%$ subsampling confidence interval for μ is then $(L_{1-\beta/2,n}, U_{1-\beta/2,n})$.

The subsampling confidence intervals for μ require the selection of subsample lengths ℓ, m_{kn} and $m_{k\ell}$, $k = 1,2$. These are important for the finite-sample performance of the subsampling method. Although best block sizes are unknown, we can modify some proposals made in Hall et al. (1998). In subsampling from transformed-Gaussian type LRD series, Hall et al. (1998) proposed block lengths $\ell = Cn^{1/2}$, $C = 1,3,6,9$. This size $n^{1/2}$ block choice is based on the intuition that subsamples from LRD series should generally be longer compared to blocks for weakly dependent data, for which $\ell \sim Cn^d$, $d \le \frac{1}{3}$ is usually optimal (Künsch, 1989; Hall, Horowitz, and Jing, 1995; Hall and Jing, 1996). That is, a jump in the order of appropriate blocks seems reasonable under long-range dependence, analogous to the sharp increase from length $\ell = 1$ blocks (no blocking) for i.i.d. data to length $\ell = O(n^{1/3})$ blocks for weakly dependent data. Plausible smoothing parameters satisfying (5) are $m_{1n} = \lfloor n^{(1+\theta)/2} \rfloor$, $m_{2n} = \lfloor n^{\theta} \rfloor$ for $\theta \in (0,1)$, and subsample versions $m_{k\ell}$, $k = 1,2$, can be analogously defined with ℓ. Hall et al. (1998) recommend a value of θ near 1 to achieve a smaller bias for the two subsample variance estimators $\tilde{d}^2_{m_{1n}}$, $\tilde{d}^2_{m_{2n}}$ combined in $\hat{d}^2_n$.

We performed a simulation study of the subsampling confidence intervals under linear long-range dependence, investigating various block lengths ℓ. We describe the simulation setup and results in Section 4.

4. NUMERICAL STUDIES OF SUBSAMPLING METHOD

Sections 4.1 and 4.2, respectively, describe the design and results of a simulation study to examine the performance of subsampling confidence intervals with

LRD linear processes. In Section 4.3, we present two examples of subsampling distribution estimation, in addition to confidence intervals, for a linear and a nonlinear LRD time series.

4.1. Data Simulation Design

Let $\{\tilde{Y}_t\}$, $t \in \mathbb{Z}$ represent a FARIMA$(0,d,0)$ series from (2) based on $d = (1-\alpha)/2 \in (0,\frac{1}{2})$ and i.i.d. innovations $\{\varepsilon_t\}$, $t \in \mathbb{Z}$.

To study the coverage accuracy of the subsampling method, we considered FARIMA processes $Y_t = \varphi Y_{t-1} + \tilde{Y}_t + \vartheta \tilde{Y}_{t-1}$, $t \in \mathbb{Z}$, constructed by combining one of the following ARMA filters (specified by φ, ϑ coefficients), α values, and innovation distributions:

- $\varphi = 0.7$, $\vartheta = -0.3$ (Filter 1); $\varphi = -0.7$, $\vartheta = 0.3$ (Filter 2); $\varphi = \vartheta = 0$ (Filter 3);
- $\alpha = 0.1, 0.5, 0.9$;
- ε_t is distributed as standard normal; $\chi_1^2 - 1$; or t_3,

where χ_1^2 and t_3 represent chi-square and t distributions with 1 and 3 degrees of freedom. The preceding framework allows for LRD linear processes $\{Y_t\}$ exhibiting various decay rates α in (1) with Gaussian or non-Gaussian innovations. The non-Gaussian innovations may exhibit skewness (e.g., $\chi_1^2 - 1$) or heavier tails (e.g., t_3). From each LRD FARIMA model, we generated size n time stretches $\mathcal{Y}_n = \{Y_1, \ldots, Y_n\}$ as follows.

A sample $\tilde{\mathcal{Y}}_n = \{\tilde{Y}_1, \ldots, \tilde{Y}_n\}$ from a non-Gaussian FARIMA$(0,d,0)$ process was generated by truncating the moving average expression in (2) after the first $M = 1{,}000$ terms and then using $n + M$ innovations ε_t to build an approximate truncated series (for details, see Bardet, Lang, Oppenheim, Philippe, and Taqqu, 2003, p. 590). Samples $\tilde{\mathcal{Y}}_n$ from a Gaussian series were simulated by the circulant embedding method of Wood and Chan (1994) with FARIMA$(0,d,0)$ covariances (Beran, 1994). Under Filter 3, the desired FARIMA realization is given by $\mathcal{Y}_n = \tilde{\mathcal{Y}}_n$. For FARIMA series involving Filters 1 and 2, generating $\tilde{Y}_t$ innovations as before in the appropriate ARMA model yielded $\mathcal{Y}_n$. We considered sample sizes $n = 100, 400, 900$.

4.2. Coverage Accuracy of Subsampling Intervals

We report here the coverage accuracy of subsampling confidence intervals for the LRD process mean $\mu = 0$ based on a data set $\mathcal{Y}_n$ generated as in Section 4.1. In the subsampling procedure of Section 3.3, we used block sizes $\ell = Cn^{1/2}$, $C \in \{0.5, 1, 2\}$ and $\theta = 0.8$. These ℓ lengths are smaller overall than those considered in Hall et al. (1998), where subsampling intervals performed poorly in numerical studies with overly large C values (e.g., 6,9). Tables 2 and 3 provide coverage probabilities of lower and upper approximate 90% one-sided confidence intervals appearing, respectively, in parenthetical

https://doi.org/10.1017/S0266466605050541 Published online by Cambridge University Press

Table 2. Subsampling coverage probabilities for block choices ℓ with $\theta = 0.8$

		Standard Normal Innovations			Chi-Square Innovations		
n	Filter	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$
$\ell = 0.5n^{1/2}$							
100	1	(93.3, 92.1)	(97.2, 96.8)	(98.4, 97.8)	(94.6, 96.8)	(95.3, 98.6)	(97.0, 99.2)
	2	(81.7, 83.4)	(90.3, 90.7)	(92.0, 91.6)	(83.6, 92.3)	(85.0, 94.6)	(89.3, 94.7)
	3	(87.4, 86.2)	(92.4, 92.5)	(95.9, 96.6)	(89.5, 95.4)	(89.8, 97.4)	(93.8, 99.1)
400	1	(88.2, 91.9)	(94.9, 95.4)	(97.3, 96.4)	(95.0, 96.7)	(94.9, 97.9)	(96.1, 98.3)
	2	(81.2, 80.7)	(87.2, 87.7)	(88.8, 90.5)	(88.1, 92.0)	(88.1, 93.4)	(86.8, 94.3)
	3	(83.7, 82.4)	(90.9, 90.5)	(92.1, 93.5)	(87.7, 92.2)	(90.5, 94.6)	(89.4, 95.6)
900	1	(90.8, 91.9)	(96.0, 97.9)	(98.5, 97.5)	(96.9, 98.0)	(96.0, 97.9)	(96.3, 99.1)
	2	(86.2, 85.7)	(92.3, 92.3)	(93.0, 92.6)	(93.1, 96.8)	(91.7, 95.3)	(89.4, 96.4)
	3	(89.1, 86.8)	(93.2, 94.0)	(95.9, 94.5)	(94.4, 96.9)	(92.1, 96.9)	(93.4, 97.7)
$\ell = n^{1/2}$							
100	1	(84.8, 83.2)	(91.0, 90.1)	(91.7, 91.7)	(88.3, 89.7)	(90.3, 91.7)	(91.6, 94.9)
	2	(73.5, 74.4)	(81.6, 84.9)	(84.9, 86.7)	(80.2, 85.6)	(80.7, 87.7)	(82.0, 89.0)
	3	(77.0, 75.7)	(85.2, 85.1)	(88.5, 89.1)	(80.0, 87.3)	(85.5, 90.2)	(86.5, 93.1)

400	1	(87.1, 84.9)	(95.3, 93.5)	(95.2, 96.0)	(91.7, 95.0)	(92.7, 95.5)	(94.8, 97.4)
	2	(83.1, 80.2)	(89.9, 90.2)	(89.5, 91.8)	(90.2, 91.8)	(88.8, 92.1)	(88.9, 93.0)
	3	(83.3, 82.7)	(89.5, 90.6)	(92.0, 92.4)	(90.8, 91.6)	(91.7, 94.5)	(91.6, 94.6)
900	1	(87.1, 88.1)	(94.3, 94.0)	(95.3, 96.6)	(95.0, 95.4)	(93.0, 96.4)	(95.6, 97.6)
	2	(86.1, 85.9)	(92.3, 91.2)	(91.2, 91.4)	(92.2, 95.6)	(90.8, 95.5)	(89.7, 94.6)
	3	(85.8, 88.1)	(90.9, 92.9)	(93.1, 95.1)	(93.8, 96.1)	(92.6, 95.1)	(92.6, 95.8)
$\ell = 2n^{1/2}$							
100	1	(75.9, 79.0)	(87.6, 88.6)	(89.9, 89.4)	(81.8, 84.4)	(87.9, 87.2)	(89.7, 91.6)
	2	(72.3, 74.8)	(82.1, 83.6)	(85.5, 85.6)	(77.1, 80.7)	(82.1, 85.6)	(86.9, 85.6)
	3	(74.6, 76.6)	(84.8, 84.1)	(86.8, 87.4)	(79.9, 80.2)	(84.8, 85.7)	(84.8, 90.0)
400	1	(80.9, 82.1)	(91.2, 89.5)	(92.3, 93.3)	(89.4, 90.4)	(90.4, 90.9)	(93.0, 94.2)
	2	(80.7, 79.8)	(89.2, 88.1)	(89.1, 88.1)	(88.1, 87.9)	(87.3, 88.4)	(88.0, 91.6)
	3	(81.3, 79.3)	(89.6, 86.3)	(91.5, 89.7)	(86.5, 89.1)	(89.1, 92.8)	(90.1, 91.1)
900	1	(83.4, 82.0)	(91.3, 92.2)	(94.2, 92.6)	(90.1, 93.3)	(91.3, 92.8)	(92.7, 95.6)
	2	(82.2, 80.3)	(88.2, 90.3)	(91.7, 90.9)	(91.1, 93.4)	(90.8, 91.2)	(88.7, 92.1)
	3	(83.4, 82.8)	(89.4, 89.7)	(91.5, 91.5)	(90.3, 92.2)	(88.7, 93.7)	(90.8, 92.1)

Table 3. Subsampling coverage probabilities for block choices ℓ with $\theta = 0.8$ and FARIMA processes generated with t_3-distributed innovations

t-innovations		$\ell = 0.5n^{1/2}$			$\ell = n^{1/2}$			$\ell = 2n^{1/2}$		
n	Filter	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$
100	1	(94.9, 94.6)	(97.4, 97.7)	(98.0, 96.8)	(87.4, 86.9)	(91.6, 90.0)	(92.6, 90.5)	(84.6, 82.4)	(86.3, 87.8)	(87.1, 87.8)
	2	(88.0, 86.8)	(92.1, 89.3)	(91.7, 90.5)	(79.1, 81.9)	(84.0, 84.9)	(84.2, 86.0)	(80.7, 79.4)	(81.8, 81.9)	(85.3, 86.4)
	3	(91.0, 92.5)	(93.6, 94.0)	(96.3, 95.8)	(82.8, 82.3)	(86.4, 85.5)	(89.8, 86.5)	(77.3, 79.8)	(83.7, 85.3)	(86.4, 87.1)
400	1	(95.1, 94.4)	(96.4, 95.5)	(95.8, 95.9)	(94.1, 93.7)	(93.5, 94.1)	(94.6, 96.2)	(88.5, 88.5)	(92.2, 90.9)	(92.5, 93.2)
	2	(91.0, 89.2)	(86.4, 88.6)	(87.8, 88.2)	(89.1, 90.0)	(91.0, 90.5)	(90.7, 90.1)	(87.2, 87.8)	(89.6, 87.2)	(89.5, 89.1)
	3	(91.8, 89.2)	(92.0, 92.6)	(91.3, 92.2)	(89.8, 91.1)	(91.7, 90.6)	(92.1, 92.5)	(88.5, 89.1)	(88.4, 89.9)	(90.5, 90.4)
900	1	(98.1, 97.0)	(97.5, 97.5)	(97.6, 98.1)	(95.1, 94.8)	(95.3, 94.3)	(95.3, 96.1)	(91.3, 92.9)	(91.6, 92.0)	(93.8, 93.7)
	2	(94.8, 95.2)	(93.3, 93.5)	(93.3, 92.3)	(94.4, 92.4)	(93.9, 91.7)	(90.5, 92.5)	(91.8, 91.4)	(90.1, 90.1)	(90.1, 90.8)
	3	(95.8, 95.8)	(94.8, 93.9)	(95.8, 94.4)	(94.6, 93.9)	(92.7, 93.2)	(93.4, 93.0)	(89.2, 92.0)	(91.3, 93.0)	(91.2, 92.0)

pairs $(\cdot,\cdot)$. Table 2 corresponds to FARIMA series with normal and chi-square innovations. Table 3 provides results for t-innovations that have unbounded (third and higher) moments. All coverage probabilities were approximated by an average over 1,000 simulation runs for each considered LRD process.

To summarize our numerical findings:

(1) Subsampling coverage accuracy generally improves with increasing sample size and weaker dependence (increasing α).
(2) Overall, the subsampling method seemed to perform similarly across the innovation processes considered.
(3) Coverage inaccuracy is most apparent under the strongest dependence $\alpha = 0.1$, in the form of undercoverage. Processes under Filter 1 (large, positive autoregressive parameter) also produced instances of overcoverage, most apparent with the smallest block $\ell = 0.5n^{1/2}$. To a larger extent, this latter behavior in coverage probabilities also appeared in Table 1 with the plug-in approach involving direct estimation of α.
(4) The subsampling method performed reasonably well across the block sizes ℓ considered. However, optimal block lengths may depend on the strength of the underlying long-range dependence; $C = 1,2$ values appeared best when $\alpha = 0.5,0.9$ whereas $C = 0.5,1$ seemed better for $\alpha = 0.1$. These findings appear consistent with the simulation results in Hall et al. (1998) with subsampling other LRD processes.
(5) Other simulation studies showed that intervals using a normal approximation for T_{1n}, based on Theorem 2(b), exhibit extreme undercoverage and perform worse than intervals based on the subsampling distribution estimator for T_{1n}. This is because the finite-sample distribution of T_{1n} can exhibit heavy tails and may converge slowly to its asymptotic normal distribution; see also Figure 1.

With subsampling techniques, theoretical investigations of block choices have received much attention for weakly dependent data, and clearly more research is needed to determine theoretically optimal block lengths ℓ under long-range dependence. The block sizes in the simulation study appear to be effective for the considered LRD processes, and similar lengths appear to be appropriate for other types of LRD processes considered in Hall et al. (1998). Results from other simulation studies indicate that smaller order block sizes (e.g., $n^{1/3}$) generally result in overcoverage under long-range dependence, whereas blocks that are excessively long (e.g., $\ell = 9n^{1/2}$) produce undercoverage. Compared to ℓ, the choice of θ appears to be less critical, and repeating the study with $\theta = 0.9$ as in Hall et al. (1998) or $\theta = 0.5,0.7$ led to only slight changes overall.

4.3. Distribution Estimates of Studentized Sample Mean

In theory, the nonparametric subsampling estimators can be applied for inference on the sample mean of different LRD processes, including the linear series considered here and transformed Gaussian processes in Hall et al. (1998). Subsampling confidence intervals for μ require a good subsample-based approximation of the distribution of the Studentized sample mean from Section 3.2. However, the variety of long-range dependence can influence greatly the distribution of the sample mean, leading to both normal (e.g., linear series) and nonnormal limit laws. Because the type of LRD time series could be unknown in practice, we conducted a further numerical study of subsampling distribution estimators of the Studentized sample mean T_{1n} in situations where T_{1n} has a normal and a nonnormal limit.

We applied the subsampling method to two LRD series: a mean zero (linear) Gaussian process Z_t with $\mathrm{Var}(Z_t) = 1$ and spectral density $f(x)$, $0 < |x| \leq \pi$, given by

$$f(x) = \lambda\{1 + \log\log(e + x^{-1})\}|1 - e^{x\iota}|^{-0.9}, \qquad \lambda > 0, \qquad \iota = \sqrt{-1},$$

and a nonlinear, transformed Gaussian $Y_t = G(Z_t)$ series, using the third Hermite polynomial $G(x) = x^3 - 3x$. The covariances $\mathrm{Cov}(Z_t, Z_{t+k})$, $k \in \mathbb{Z}$ satisfy (1) with $\alpha = 0.1$ and nonconstant (up to a scalar multiple) $L(x) = \log\log(x)$; these covariances can be written as a sum of FARIMA$(0, d = 0.45, 0)$ covariances (i.e., $\tilde{C}\int_0^\pi \cos(kx)|1 - e^{x\iota}|^{-0.9}\,\mathrm{dx}$) plus an additional regularly varying component. The process Y_t also exhibits slowly decaying covariances because $G(\cdot)$ has Hermite rank 3 and here $0 < 3\alpha < 1$ (Taqqu, 1975; Beran, 1994). Because of the limit law of the sample mean, the asymptotic distribution of the Studentized sample mean T_{1n} is normal under the Z_t process (e.g., Theorem 2) and nonnormal for the nonlinear series Y_t (Taqqu, 1975, 1979; Hall et al., 1998).

For the preceding two series, we can compare the exact distribution $F_{1n}(x)$ of the subsample-Studentized sample mean T_{1n} and its subsampling estimator $\hat{F}_{1n}(x)$. For each series type, Figure 1 provides the exact distribution F_{1n} of T_{1n} at sample sizes $n = 100, 400, 900$ and $\theta = 0.8$. In each case, the distribution F_{1n} was calculated through simulation (using 15,000 runs) and appears as a thick line in Figure 1. Using a block length $\ell = n^{1/2}$, five subsampling estimates $\hat{F}_{1n}$ of each distribution F_{1n} were computed from five independent size n samples from $\{Z_t\}$ or $\{Y_t\}$; these estimates appear as dotted lines in Figure 1.

In each instance in Figure 1, the finite-sample distribution of T_{1n} exhibits heavy tails. This indicates that confidence intervals for the process mean $E(Z_t) = 0 = E(Y_t)$ set with T_{1n} and a normal approximation to its distribution would be inappropriate. (As stated previously, a normal approximation of T_{1n} is expected to break down for LRD series Y_t.) However, the subsampling estimates appear to adequately approximate the exact distribution of the Studentized sample mean T_{1n}, particularly for larger n. The coverage probabilities listed in Figure 1 additionally suggest that the subsampling method, based on T_{1n} and $\hat{F}_{1n}$, leads to reasonable confidence intervals of the means of both the linear and nonlinear LRD processes.

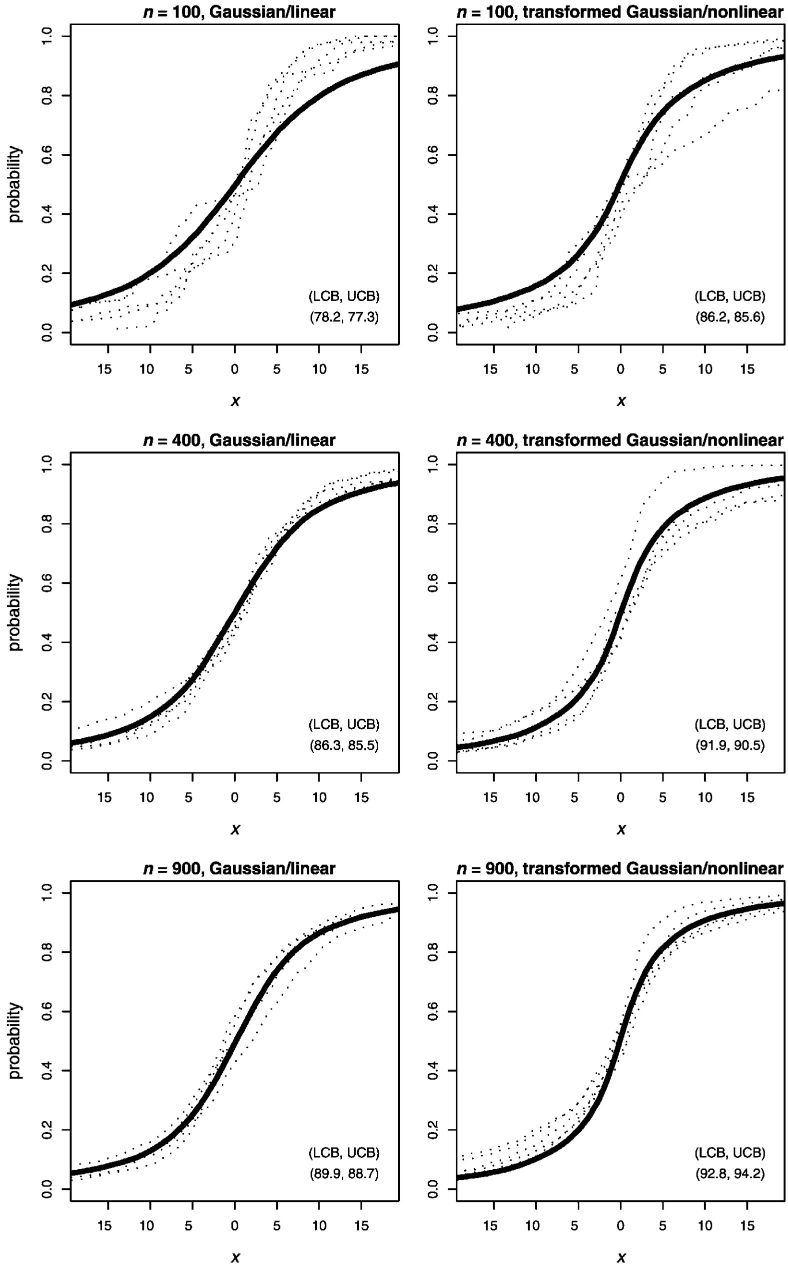

FIGURE 1. Subsampling estimates $\hat{F}_{1n}$ (dotted lines) of the exact distribution F_{1n} (thick line) of Studentized sample mean T_{1n}. Coverage probabilities of 90% one-sided lower and upper confidence bounds (LCB, UCB) for the process mean also appear, based on the subsampling method.

https://doi.org/10.1017/S0266466605050541 Published online by Cambridge University Press

5. THE SUBSAMPLING METHOD UNDER SHORT-RANGE DEPENDENCE

We comment briefly on the subsampling method applied to linear time processes under weak or short-range dependence. A stationary time series $\{Y_t\}$, $t \in \mathbb{Z}$ can be generally called short-range dependent (SRD) if the process autocovariances decay fast enough to be absolutely summable, $\sum_{k=1}^{\infty}|r(k)| < \infty$. Such covariance summability does not hold for LRD processes satisfying (1).

For weakly dependent time series fulfilling a mixing condition, subsampling techniques have been developed for inference on the distribution of a variety of statistics, including the sample mean (Carlstein, 1986; Künsch, 1989; Politis and Romano, 1994; Hall and Jing, 1996). However, the sampling window method of this paper applies to linear time processes that may exhibit *either* short-range dependence *or* long-range dependence. In particular, we require no mixing assumptions on the process $\{Y_t\}$ under weak dependence.

THEOREM 3. *Suppose m_{1n}, m_{2n} satisfy (5), $\ell^{-1} + n^{-1}\ell = o(1)$ and Assumption L holds after replacing condition (1) with a condition of weak dependence: $\sum_{k=1}^{\infty}|r(k)| < \infty$ with $\sum_{k\in\mathbb{Z}} r(k) > 0$. Defining $d_n^2 \equiv n\sum_{k\in\mathbb{Z}} r(k)$, the convergence results of both Theorem 1 and Theorem 2 remain valid.*

With the convention that we define $\alpha = 1$ and $L(\cdot) = \sum_{k\in\mathbb{Z}} r(k) > 0$ under short-range dependence, both (4) and the scaling $d_n^2 = n^{2-\alpha}L(n)$ in (3) are correct for short-range dependence; that is, $\sum_{k\in\mathbb{Z}} r(k) = \lim_{n\to\infty} n\,\mathrm{Var}(\overline{Y}_n)$. The same subsampling method can applied to distribution estimation of the sample mean, in addition to interval estimation, under both SRD and LRD classifications of a linear time series.

6. PROOFS

6.1. Proofs of Main Results

In the following discussion, let $\sigma_n^2 = n^2\,\mathrm{Var}(\overline{Y}_n)$. Denote the supremum norm $\|g\|_\infty = \sup\{|g(x)| : x \in \mathbb{R}\}$ for a function $g : \mathbb{R} \to \mathbb{R}$ and let Φ denote the standard normal distribution function. Unless otherwise specified, limits in order symbols are taken letting $n \to \infty$.

We first state a useful result concerning moments of the sample mean $\overline{Y}_n$ from a LRD linear process. Lemma 1(a) follows from the proof of Theorem 18.6.5 in Ibragimov and Linnik (1971) and bounds sums of consecutive filter coefficients in terms of the standard deviation of $n\overline{Y}_n$; part (b) of Lemma 1 corresponds to Lemma 4 of Davydov (1970).

LEMMA 1. *Suppose Assumption L holds. For all $n \in \mathbb{N}$,*

(a) and for all $k \in \mathbb{Z}$,

$$\sigma_n^{-1}\left|\sum_{j=1}^{n} c_{j-k}\right| \leq \omega_n = \left\{\sigma_n^{-1}(4+2\sigma_n^{-1})\sum_{j\in\mathbb{Z}} c_j^2\right\}^{1/2}.$$

(b) $E\{[n(\overline{Y}_n-\mu)]^{2k}\} \leq A_k(\sigma_n^2)^k$ for some $A_k > 0$, if $E(|\varepsilon_0|^{2k}) < \infty$ for a given $k \in \mathbb{N}$.

For the proof of Theorem 1, we define $\tilde{F}_n(x) = N^{-1}\sum_{i=1}^{N} I\{(S_{\ell i}-\ell\mu)/d_\ell \leq x\}$, $x \in \mathbb{R}$. Note that $\tilde{F}_n(x)$ differs from the sampling window estimator $\hat{F}_n(x)$ from Section 3.1 by centering subsample sums with $\ell\mu$ rather than $\ell\overline{Y}_n$. We also require the following result for LRD linear processes with bounded innovations. For these series, Lemma 2(a) shows that standardized subsample sums based on well-separated blocks are asymptotically uncorrelated, whereas Lemma 2(b) establishes the convergence of $\tilde{F}_n$. We defer the proof of Lemma 2 to Section 6.2.

LEMMA 2. *Suppose the conditions of Theorem 1 hold with bounded innovations, that is, $P(|\epsilon_t| \leq B) = 1$ for some $B > 0$. Then, as $n \to \infty$,*

(a) for any nonnegative integers a, b and $0 < \epsilon < 1$,

$$\max_{n\epsilon\leq i\leq n}|E[(S_{\ell 1}^*)^a(S_{\ell i}^*)^b]-E(Z^a)\cdot E(Z^b)| = o(1),$$

where $S_{\ell i}^* = (S_{\ell i}-\ell\mu)/\sigma_\ell$, $i \in \mathbb{N}$, and Z is a standard normal variable. For $a > 0$, $E(Z^a) = (a-1)(a-3)\ldots(1)$ for even a; 0 otherwise.

(b) $E\{\|\tilde{F}_n-\Phi\|_\infty\} \to 0$.

Proof of Theorem 1. We note that (4) and the assumption $\ell^{-1}+n^{-1+\delta}\ell = o(1)$ imply

$$(\ell^2 d_n^2)/(n^2 d_\ell^2) \sim (n/\ell)^{-\alpha}\{L(n)/L(\ell)\} = o(1) \tag{7}$$

because L is positive and $x^\gamma L(x) \to \infty$, $x^{-\gamma}L(x) \to 0$ as $x \to \infty$ for any $\gamma > 0$ (Ibragimov and Linnik, 1971, App. 1). We can bound $\|\hat{F}_n - F_n\|_\infty \leq \|\hat{F}_n-\Phi\|_\infty + \|F_n-\Phi\|_\infty$ and

$$\|\hat{F}_n-\Phi\|_\infty \leq \|\tilde{F}_n-\Phi\|_\infty + \sup_{x\in\mathbb{R}}|\Phi(x+\epsilon)-\Phi(x)| + 2I\{\ell|\overline{Y}_n-\mu|/d_\ell > \epsilon\},$$

for each $\epsilon > 0$. From (7), we find $P(\ell|\overline{Y}_n-\mu|/d_\ell > \epsilon) = o(1)$ by Chebychev's inequality using $\sigma_n^2 \sim d_n^2$ from (4). From the continuity of Φ, it follows that $\|F_n-\Phi\|_\infty = o(1)$ by (3) and also that $\sup_{x\in\mathbb{R}}|\Phi(x+\epsilon)-\Phi(x)| \to 0$ as $\epsilon \to 0$. Hence, it suffices to show $\|\tilde{F}_n-\Phi\|_\infty \xrightarrow{p} 0$ or, equivalently as a result

https://doi.org/10.1017/S0266466605050541 Published online by Cambridge University Press

of Φ's continuity, $|\tilde{F}_n(x) - \Phi(x)| \xrightarrow{p} 0$ for each $x \in \mathbb{R}$. We will prove $E(|\tilde{F}_n(x) - \Phi(x)|) = o(1)$ for $x \in \mathbb{R}$.

Let $E(\varepsilon_t^2) = \tau^2 > 0$. For each $b \in \mathbb{N}$, define variables $\varepsilon_{t,b} = \varepsilon_t I\{|\varepsilon_t| \le b\} - E(\varepsilon_t I\{|\varepsilon_t| \le b\})$ and $Y_{t,b} = \mu + \sum_{j\in\mathbb{Z}} c_{t-j}(\tau\varepsilon_{j,b}/\tau_b)$, where $\tau_b^2 = E(\varepsilon_{0,b}^2)$. (We may assume $\tau_b^2 > 0$ w.l.o.g. in the following discussion.) For each $n, b, i \in \mathbb{N}$, write $d_{n,b}^2$, $S_{\ell i,b} = \sum_{j=i}^{i+\ell-1} Y_{j,b}$, $F_{n,b}$, and $\tilde{F}_{n,b}$ to denote the analogs of d_n^2, $S_{\ell i}$, F_n and $\tilde{F}_n$ with respect to $Y_{1,b},\dots,Y_{n,b}$. Note that both series $\{Y_t\}$ and $\{Y_{t,b}\}$ involve the same linear filter with i.i.d. innovations of mean 0 and variance τ^2 and hence have the same covariances; in particular, we may set $d_{n,b}^2 = d_n^2$ for $b, n \in \mathbb{N}$. For any $\epsilon > 0$, $x \in \mathbb{R}$ and $b, n \in \mathbb{N}$,

$$E(|\tilde{F}_n(x) - \Phi(x)|) \le \{E(|\tilde{F}_n(x) - \tilde{F}_{n,b}(x)|^2)\}^{1/2} + E(|\tilde{F}_{n,b}(x) - \Phi(x)|)$$

$$\equiv A_{1n,b}(x) + A_{2n,b}(x).$$

Letting $D_{\ell,b} = (S_{\ell 1,b} - S_{\ell 1})/d_\ell$,

$$\begin{aligned}
A_{1n,b}^2(x) &\le \frac{1}{N}\sum_{i=1}^{N} E\{(I\{(S_{\ell i} - \ell\mu)/d_\ell \le x\} - I\{(S_{\ell i,b} - \ell\mu)/d_{\ell,b} \le x\})^2\} \\
&= F_\ell(x) + F_{\ell,b}(x) - 2F_{\ell,b}(\min\{x, x + D_{\ell,b}\}) \\
&\le |F_\ell(x) - \Phi(x)| + 3\|F_{\ell,b} - \Phi\|_\infty \\
&\quad + 2\sup_{y\in\mathbb{R}}|\Phi(y+\epsilon) - \Phi(y)| + 2P(|D_{\ell,b}| > \epsilon),
\end{aligned}$$

where $P(|D_{\ell,b}| > \epsilon) \le \mathrm{Var}(S_{\ell 1} - S_{\ell 1,b})/(\epsilon^2 d_\ell^2)$, and we deduce

$$\mathrm{Var}(S_{\ell 1} - S_{\ell 1,b}) = \sigma_\ell^2\,\mathrm{Var}(\tau^{-1}\varepsilon_0 - \tau_b^{-1}\varepsilon_{0,b}) = 2\sigma_\ell^2\{1 - \tau_b^{-1}\tau^{-1}E(\varepsilon_0\varepsilon_{0,b})\},$$

by the i.i.d. property of innovations. Hence, for any $x \in \mathbb{R}$, $b \in \mathbb{N}$, and $\epsilon > 0$,

$$\limsup_{n\to\infty} E(|\tilde{F}_n(x) - \Phi(x)|)$$

$$\le \left\{2\sup_{y\in\mathbb{R}}|\Phi(y+\epsilon) - \Phi(y)| + 4\epsilon^{-2}\{1 - \tau_b^{-1}\tau^{-1}E(\varepsilon_0\varepsilon_{0,b})\}\right\}^{1/2}$$

using (4), $A_{2n,b}(x) = o(1)$ as $n \to \infty$ by Lemma 2(b), and $|F_\ell(x) - \Phi(x)|, \|F_{\ell,b} - \Phi\|_\infty = o(1)$ as $n \to \infty$ by (3). Because $\lim_{b\to\infty}\tau_b^{-1}E(\varepsilon_0\varepsilon_{0,b}) = \tau^{-1}E(\varepsilon_0^2) = \tau$ and $\sup_{x\in\mathbb{R}}|\Phi(x+\epsilon) - \Phi(x)| \to 0$ as $\epsilon \to 0$, the proof of Theorem 1 is finished. ■

Proof of Theorem 2. Let m denote m_{kn}, $k \in \{1,2\}$, and define $\tilde{d}_{m\mu}^2 = N_m^{-1}\sum_{i=1}^{N_m}(S_{mi} - m\mu)^2$ for $N_m = n - m + 1$. Using Hölder's inequality, (4), and (7), we can show that

$$
\begin{aligned}
&E|\tilde{d}_m^2 - \tilde{d}_{m\mu}^2| \\
&\quad \le \frac{1}{N_m}\sum_{i=1}^{N_m} E|\{(S_{mi} - m\mu) - (S_{mi} - m\bar{Y}_n)\}\{(S_{mi} - m\mu) + (S_{mi} - m\bar{Y}_n)\}| \\
&\quad \le 4m[\mathrm{Var}(\bar{Y}_n)]^{1/2}[\mathrm{Var}(S_{m1}) + m^2\,\mathrm{Var}(\bar{Y}_n)]^{1/2} \\
&\quad = O(m^2 n^{-2} d_n^2 + n^{-1} m d_m d_n) = o(d_m^2). \qquad \textbf{(8)}
\end{aligned}
$$

With the truncated variables from the proof of Theorem 1, let $\tilde{d}_{m\mu,b}^2 = N_m^{-1}\sum_{i=1}^{N_m}(S_{mi,b} - m\mu)^2$ with $S_{mi,b} = \sum_{j=i}^{m+i-1} Y_{j,b}$ for $i,b \in \mathbb{N}$; again the processes $\{Y_t\}$ and $\{Y_{t,b}\}$ have the same covariances for all $b \in \mathbb{N}$. In a fashion similar to (8), we find by Hölder's inequality,

$$
\begin{aligned}
&E|\tilde{d}_{m\mu,b}^2 - \tilde{d}_{m\mu}^2| \\
&\quad \le 2\{E[(S_{m1} - m\mu)^2] + E[(S_{m1,b} - m\mu)^2]\}^{1/2}\{E[(S_{m1} - S_{m1,b})^2]\}^{1/2} \\
&\quad = 4\sigma_m^2\{1 - \tau_b^{-1}\tau^{-1}E(\varepsilon_0\varepsilon_{0,b})\}^{1/2} \qquad \textbf{(9)}
\end{aligned}
$$

using $\sigma_m^2 = \mathrm{Var}(S_{m1}) = \mathrm{Var}(S_{m1,b})$ and $\mathrm{Var}(S_{m1} - S_{m1,b}) = 2\sigma_m^2\{1 - \tau_b^{-1}\tau^{-1}E(\varepsilon_0\varepsilon_{0,b})\}$ from before.

Applying Lemma 2(a) and the bound on $E\{(S_{m1,b} - m\mu)^4\}$ from Lemma 1(b) with (4),

$$
\begin{aligned}
\mathrm{Var}(\tilde{d}_{m\mu,b}^2) &= \sigma_m^4 O\left[N_m^{-2}\sum_{1\le i,j\le N_m}|\mathrm{Cov}\{\sigma_m^{-2}(S_{mi,b} - m\mu)^2, \sigma_m^{-2}(S_{mj,b} - m\mu)^2\}|\right] \\
&= o(d_m), \\
E(\tilde{d}_{m\mu,b}^2) &= \mathrm{Var}(S_{m1,b}) = \sigma_m^2 \sim d_m^2, \qquad \textbf{(10)}
\end{aligned}
$$

holds for each $b \in \mathbb{N}$. Then $\tilde{d}_m^2/d_m^2 \xrightarrow{p} 1$ follows from using (8)–(10) to deduce

$$
\begin{aligned}
&\limsup_{m\to\infty} E\left|\frac{\tilde{d}_m^2}{d_m^2} - 1\right| \\
&\quad \le \limsup_{m\to\infty} \frac{1}{d_m^2}\{E|\tilde{d}_m^2 - \tilde{d}_{m\mu}^2| + E|\tilde{d}_{m\mu,b}^2 - \tilde{d}_{m\mu}^2| + [E(\tilde{d}_{m\mu,b}^2 - d_m^2)^2]^{1/2}\} \\
&\quad \le 4\{1 - \tau_b^{-1}\tau^{-1}E(\varepsilon_0\varepsilon_{0,b})\}^{1/2}, \qquad b \in \mathbb{N},
\end{aligned}
$$

and then applying $\lim_{b\to\infty}\tau_b^{-1}E(\varepsilon_0\varepsilon_{0,b}) = \tau$. Because $\hat{d}_n^2 = \tilde{d}_{m_{1n}}^4/\tilde{d}_{m_{2n}}^2$ and (5) and (6) imply

$$
d_{m_{1n}}^4/(d_{m_{2n}}^2 d_n^2) \sim L^2(m_{1n})/\{L(m_{2n})L(n)\} \sim 1,
$$

the convergence $\hat{d}_n^2/d_n^2 \xrightarrow{p} 1$ now follows. From this and Theorem 1, we find the convergence of $\hat{F}_{1n}$ in probability as in Hall et al. (1998). ■

https://doi.org/10.1017/S0266466605050541 Published online by Cambridge University Press

Proof of Theorem 3. From Corollary 6.1.1.2 of Fuller (1996), we have $\lim_{n\to\infty} n \operatorname{Var}(\overline{Y}_n) = \sum_{k\in\mathbb{Z}} r(k) > 0$. Lemmas 1 and 2 and the same proofs of Theorems 1 and 2 (including Lemma 2) apply with the convention that $\alpha = 1$, $L(\cdot) = \sum_{k\in\mathbb{Z}} r(k) > 0$ and $d_n^2 = n\sum_{k\in\mathbb{Z}} r(k)$ under short-range dependence. The one modification is that (7) still holds if $\ell^{-1} + \ell/n = o(1)$. ■

6.2. Proof of Technical Lemma

Proof of Lemma 2(b). The result follows from Theorem 2.4 of Hall et al. (1998) and its proof after verifying that the conditions required are met: the process $\{Y_t\}$ has all moments finite by assumption; $n(\overline{Y}_n - \mu)/\sigma_n$ converges to a (normal) continuous distribution by (3) which is uniquely determined by its moments; $(\ell^2\sigma_n^2)/(n^2\sigma_\ell^2) = o(1)$ by (4) and (7); and Lemma 2(a) holds. ■

Proof of Lemma 2(a). It suffices to consider only positive $a, b \in \mathbb{N}$, because Lemma 1(b) with (3) implies that $E[(S_{\ell 1}^*)^a] \to E(Z^a)$ as $n \to \infty$ for any nonnegative a.

We establish some additional notation. For $i \in \mathbb{N}$, write the standardized subsample sum $S_{\ell i}^* = \sum_{k\in\mathbb{Z}} d_{k(i)}\varepsilon_k$ using $d_{k(i)} = \sigma_\ell^{-1}\sum_{j=1}^{\ell} c_{(i-1)+j-k}$ and set $E(\varepsilon_t^2) = 1$ throughout the proof because of standardization; we suppress here the dependence of $d_{k(i)}$ on ℓ in our notation. Write the nonnegative integers as $\mathbb{Z}_+ = \{0\} \cup \mathbb{N}$. For $a, b \in \mathbb{N}$ and $1 \le m \le \lfloor (a+b)/2 \rfloor$, denote integer vectors $s^{(m)} = (s_1, \ldots, s_m), t^{(m)} = (t_1, \ldots, t_m) \in (\mathbb{Z}_+)^m$; write a set

$$B_{a,b,m} = \left\{ (s^{(m)}, t^{(m)}) \in (\mathbb{Z}_+)^m \times (\mathbb{Z}_+)^m : \sum_{j=1}^m s_j = a,\ \sum_{j=1}^m t_j = b,\ \min_{1\le j\le m}(s_j + t_j) \ge 2 \right\}$$

and define with $i \in \mathbb{N}$,

$$\Psi_{a,b}(s^{(m)}, t^{(m)}) = \prod_{j=1}^m E(\varepsilon_0^{s_j+t_j}), \quad \Delta_{i,a,b}(s^{(m)}, t^{(m)})$$

$$= \sum_{k_1\ne\cdots\ne k_m\in\mathbb{Z}} \prod_{j=1}^m (d_{k_j(1)})^{s_j}(d_{k_j(i)})^{t_j},$$

where indices in the sum $\sum_{k_1\ne\cdots\ne k_m\in\mathbb{Z}}$ extend over integer m-tuples $(k_1, \ldots, k_m) \in \mathbb{Z}^m$ with distinct components $k_j \ne k_{j'}$ for $1 \le j \ne j' \le m$. We will later use that

$$\sup_{i\in\mathbb{N}} |\Delta_{i,a,b}(s^{(m)}, t^{(m)})| < \infty \tag{11}$$

holds for $a, b \in \mathbb{N}$, $1 \le m \le \lfloor (a+b)/2 \rfloor$, and $(s^{(m)}, t^{(m)}) \in B_{a,b,m}$ so that certain sums $\Delta_{i,a,b}(s^{(m)}, t^{(m)})$ are finitely defined. We omit the proof of (11) here (in light of showing (14) to follow, which uses similar arguments).

Because the innovations are i.i.d. with $E(\varepsilon_t) = 0$, it holds that $E(\prod_{j=1}^{a+b} \varepsilon_{k_j}) = 0$ for $a, b \in \mathbb{N}$ and integers $(k_1, \dots, k_{a+b}) \in \mathbb{Z}^{a+b}$ unless, for each $1 \le j \le a+b$, there exists some j' such that $k_j = k_{j'}$, implying less than $\lfloor (a+b)/2 \rfloor$ distinct integer values among $(k_1, \dots, k_{a+b})$. Using this and

$$E[(S_{\ell 1}^*)^a (S_{\ell i}^*)^b] = \sum_{(k_1,\dots,k_{a+b}) \in \mathbb{Z}^{a+b}} E\left(\prod_{j=1}^{a} d_{k_j(1)} \varepsilon_{k_j} \cdot \prod_{j=a+1}^{a+b} d_{k_j(i)} \varepsilon_{k_j} \right) \tag{12}$$

for any $a, b, i \in \mathbb{N}$, we rewrite (12) as a sum over collections of integer indices $(k_1, \dots, k_{a+b})$ with $1 \le m \le \lfloor (a+b)/2 \rfloor$ distinct values:

$$\sum_{m=1}^{\lfloor (a+b)/2 \rfloor} \sum_{(W_1,\dots,W_m)} \left[\sum_{(k_1,\dots,k_{a+b}) \in \mathbb{Z}^{a+b}} E\left(\prod_{j=1}^{a} d_{k_j(1)} \varepsilon_{k_j} \cdot \prod_{j=a+1}^{a+b} d_{k_j(i)} \varepsilon_{k_j} \right) \right.$$
$$\left. \times I\{k_j = k_{j'} \Leftrightarrow j, j' \in W_h\} \right],$$

where the sum $\sum_{(W_1,\dots,W_m)}$ is taken over all size m partitions $(W_1, \dots, W_m)$ of $\{1, \dots, a+b\} = \cup_{h=1}^{m} W_h$. The indicator function $I\{\cdot\}$ in the bracketed sum in the preceding expression signifies that, for a given partition $(W_1, \dots, W_m)$, we sum terms in (12) over integer indices $(k_1, \dots, k_{a+b}) \in \mathbb{Z}^{a+b}$ satisfying $k_j = k_{j'}$ if and only if $j, j' \in W_h$, $h = 1, \dots, m$. We can more concisely write (12) as

$$E[(S_{\ell 1}^*)^a (S_{\ell i}^*)^b] = \sum_{m=1}^{\lfloor (a+b)/2 \rfloor} \sum_{(W_1,\dots,W_m)} \Psi_{a,b}(s^{(m)}, t^{(m)}) \cdot \Delta_{i,a,b}(s^{(m)}, t^{(m)}), \tag{13}$$

$$s^{(m)}(W_1, \dots, W_m) \equiv s^{(m)} = (s_1, \dots, s_m), \qquad s_h = \sum_{j=1}^{a} I\{j \in W_h\},$$

$$t^{(m)}(W_1, \dots, W_m) \equiv t^{(m)} = (t_1, \dots, t_m),$$

$$t_h = \sum_{j=a+1}^{a+b} I\{j \in W_h\}, \qquad h = 1, \dots, m,$$

in terms of vectors $s^{(m)}, t^{(m)} \in (\mathbb{Z}_+)^m$ as a function of a partition $(W_1, \dots, W_m)$. By the nature of the partitions $(W_1, \dots, W_m)$, the vectors $(s^{(m)}, t^{(m)})$ in (13) are elements of $B_{a,b,m}$ for some $1 \le m \le \lfloor (a+b)/2 \rfloor$ (any set W_h in a partition $(W_1, \dots, W_m)$ has at least two elements by the restriction $m \le \lfloor (a+b)/2 \rfloor$ so that $\min_{1 \le j \le m}(s_j + t_j) \ge 2$ follows).

To help identify the most important terms in the summand (13), we define a count $\mathcal{C}(s^{(m)}, t^{(m)}) = \sum_{j=1}^{m} I\{s_j = t_j = 1\}$ as a function of $(s^{(m)}, t^{(m)}) \in (\mathbb{Z}_+)^m \times (\mathbb{Z}_+)^m$ and also a special indicator function $\mathbb{I}_{a,b}(s^{(m)}, t^{(m)}) = I\{a, b$ even, $m = (a+b)/2$, $\mathcal{C}(s^{(m)}, t^{(m)}) = 0\}$. Now to show Lemma 2(a), it suffices to establish for any $a, b \in \mathbb{N}$, $1 \le m \le \lfloor (a+b)/2 \rfloor$, and $(s^{(m)}, t^{(m)}) \in B_{a,b,m}$ that

$$\max_{n\epsilon \le i \le n} |\Delta_{i,a,b}(s^{(m)}, t^{(m)}) - \mathbb{I}_{a,b}(s^{(m)}, t^{(m)})| = o(1). \tag{14}$$

If either a or b is odd, so that $E(Z^a)E(Z^b) = 0$, Lemma 2(a) follows immediately from (13) and (14). In the case that a, b are even, we find that the dominant component in (13) involves the sum over partitions $(W_1, \ldots, W_m)$ with $m = (a+b)/2$ and corresponding $(s^{(m)}, t^{(m)})$ that satisfy $\mathcal{C}(s^{(m)}, t^{(m)}) = 0$ or equivalently $s_j, t_j \in \{0,2\}$, $s_t + t_j = 2$ for $1 \le j \le m = (a+b)/2$; in this instance, $\Psi_{a,b}(s^{(m)}, t^{(m)}) = 1$ (by $E(\varepsilon_0^2) = 1$) and a size $m = (a+b)/2$ partition $(W_1, \ldots, W_{(a+b)/2})$ of $\{1, \ldots, a+b\}$ is formed by a size $a/2$ a partition of $\{1, \ldots, a\}$ and a size $b/2$ partition of $\{a+1, \ldots, b+a\}$; there are exactly $(a-1)(a-3)\ldots(1) \times (b-1)(b-3)\ldots(1)$ such partitions possible. So for even a, b with $m = (a+b)/2$, it holds that

$$\sum_{m=(a+b)/2, (W_1, \ldots, W_m)} \Psi_{a,b}(s^{(m)}, t^{(m)}) \cdot \mathbb{I}_{a,b}(s^{(m)}, t^{(m)}) = E(Z^a) \cdot E(Z^b),$$

which with (13) and (14) implies that Lemma 2(a) follows for a, b even.

We now focus on proving (14) by treating two cases: $\mathbb{I}_{a,b}(s^{(m)}, t^{(m)}) = 0$ or 1. For $i \in \mathbb{N}$, define subsample sum covariances $r_\ell(i) \equiv \mathrm{Cov}(S^*_{\ell 1}, S^*_{\ell i}) = \sum_{j \in \mathbb{Z}} d_{j(1)} \cdot d_{j(i)}$. Although we cannot assume that $\sum_{k \in \mathbb{Z}} |c_k| < \infty$ or $\sum_{k \in \mathbb{Z}} |d_{k(i)}| < \infty$ under long-range dependence, it holds that

$$\sum_{k \in \mathbb{Z}} d^2_{k(i)} = \mathrm{Var}(S^*_{\ell i}) = 1, \qquad \sum_{k \in \mathbb{Z}} |d_{k(i)}|^s \le \omega_\ell^{s-2} \qquad i \in \mathbb{N}, s \ge 2, \tag{15}$$

by applying Lemma 1(a). Similarly, if $a, b \in \mathbb{N}$, $1 \le m \le \lfloor (a+b)/2 \rfloor$, $(s^{(m)}, t^{(m)}) \in B_{a,b,m}$ with $\min_{1 \le j \le m} \max\{s_j, t_j\} \ge 2$, then

$$\max_{n\epsilon \le i \le n} |\Delta_{i,a,b}(s^{(m)}, t^{(m)})| \le \max_{n\epsilon \le i \le n} \left(\prod_{j=1}^{m} \sum_{k \in \mathbb{Z}} |d_{k(1)}|^{s_j} |d_{k(i)}|^{t_j} \right) \le \omega_\ell^{a+b-2m} \tag{16}$$

follows from (15), where $\omega_\ell = o(1)$.

Case 1. $\mathbb{I}_{a,b}(s^{(m)}, t^{(m)}) = 0$. We show that (14) holds with an induction argument on m. Consider first the possibility $m = 1$, for which $(s^{(m)}, t^{(m)}) = (s_1, t_1) = (a, b)$. If $\mathcal{C}(s^{(m)}, t^{(m)}) > 0$ and $m = 1$, then $a = b = 1$; in this case, $\Delta_{i,a,b}(s^{(m)}, t^{(m)}) = r_\ell(i)$, $i \in \mathbb{N}$. For $0 < \epsilon < 1$ and large n such that $n\epsilon/2 > \ell$, the growth rate in (1) with (4) gives

$$\begin{aligned} \max_{n\epsilon \le i \le n} |r_\ell(i)| &= \sigma_\ell^{-2} \max_{n\epsilon \le i \le n} \left| \sum_{j,j'=1}^{\ell} \mathrm{Cov}(Y_j, Y_{i-1+j'}) \right| \\ &\le \sigma_\ell^{-2} \ell^2 \max_{n\epsilon \le i \le n} \max_{|j| \le \ell} |r(i+j)| \qquad (17) \\ &= O(\ell^\alpha / L(\ell) \cdot |n\epsilon - \ell|^{-\alpha} M_{n,\epsilon}) \\ &= O((n/\ell)^{-\alpha} \{L(n)/L(\ell)\}\{M_{n,\epsilon}/L(n)\}) = o(1), \end{aligned}$$

defining $M_{n,\epsilon} = \sup\{L(tn): \epsilon/2 < t < 2\}$ and using (7) with $M_{n,\epsilon}/L(n) \to 1$ by Taqqu (1977, Lem. A1). If $\mathcal{C}(s^{(m)}, t^{(m)}) = 0$ and $m = 1$, then either $a = s_1 > 1$ or $b = t_1 > 1$ for $a, b \in \mathbb{N}$, and (14) follows from (16) and $O(\omega_\ell^{a+b-2m}) = o(1)$.

Now assume that, for some $T \in \mathbb{N}$, (14) holds whenever $a, b \in \mathbb{N}$ with $\mathbb{I}_{a,b}(s^{(m)}, t^{(m)}) = 0$ and $m \le \min\{T, \lfloor (a+b)/2 \rfloor\}$. We let $a, b \in \mathbb{N}$, $(s^{(m)}, t^{(m)}) \in B_{a,b,m}$, $m = T + 1 \le \lfloor (a+b)/2 \rfloor$ with $\mathbb{I}_{a,b}(s^{(m)}, t^{(m)}) = 0$ in the following discussion and show that (14) must hold by the induction assumption.

If $\mathcal{C}(s^{(m)}, t^{(m)}) = 0$, then $\min_{1 \le j \le m} \max\{s_j, t_j\} \ge 2$ holds and (14) follows from (16) and $O(\omega_\ell^{a+b-2m}) = o(1)$ because $a + b > 2m$, which we verify. If a, b are even, then $m < (a+b)/2$ must hold by $\mathbb{I}_{a,b}(s^{(m)}, t^{(m)}) = 0$; if a or b is odd, then $a + b > 2m$ follows because $(s^{(m)}, t^{(m)}) \in B_{a,b,m}$ and $\max_{1 \le j \le m}\{s_j + t_j\} > 2$ must hold (the alternative by $(s^{(m)}, t^{(m)}) \in B_{a,b,m}$ with $\mathcal{C}(s^{(m)}, t^{(m)}) = 0$ is that $s_j, t_j \in \{0,2\}$ for $1 \le j \le m$, so that $\sum_{j=1}^m s_j = a$ and $\sum_{j=1}^m t_j = b$ are even, a contradiction).

Consider now the possibility that $\mathcal{C}(s^{(m)}, t^{(m)}) > 0$. Because $m = T + 1 > 1$ with $a, b > 1$ necessarily, say (w.l.o.g.) that components $s_m = t_m = 1$ in $s^{(m)} = (s_1, \ldots, s_m)$, $t^{(m)} = (t_1, \ldots, t_m)$. Using $r_\ell(i) = \sum_{k_m \in \mathbb{Z}} d_{k_m(1)}^{s_m} \cdot d_{k_m(i)}^{t_m}$, $i \in \mathbb{N}$, we can algebraically rewrite the sum

$$\Delta_{i,a,b}(s^{(m)}, t^{(m)})$$

$$= \sum_{k_1 \ne \cdots \ne k_{m-1} \in \mathbb{Z}} \left(\prod_{j=1}^{m-1} d_{k_j(1)}^{s_j} \cdot d_{k_j(i)}^{t_j} \right) \left(\sum_{k_m \in \mathbb{Z}} d_{k_m(1)}^{s_m} \cdot d_{k_m(i)}^{t_m} - \sum_{h=1}^{m-1} d_{k_h(1)}^{s_m} \cdot d_{k_h(i)}^{t_m} \right)$$

$$= r_\ell(i) \cdot \Delta_{i,a-1,b-1}(s_0^{(m-1)}, t_0^{(m-1)}) - \sum_{j=1}^{m-1} \Delta_{i,a,b}(s_j^{(m-1)}, t_j^{(m-1)}), \qquad \textbf{(18)}$$

where $(s_0^{(m-1)}, t_0^{(m-1)}) \in B_{a-1,b-1,m-1}$ for $s_0^{(m-1)} = (s_1, \ldots, s_{m-1})$, $t_0^{(m-1)} = (t_1, \ldots, t_{m-1})$, and $(s_j^{(m-1)}, t_j^{(m-1)}) \in B_{a,b,m-1}$ for $s_j^{(m-1)} = s_0^{(m-1)} + e_j^{(m-1)}$, $t_j^{(m-1)} = t_0^{(m-1)} + e_j^{(m-1)}$, writing the jth coordinate vector $e_j^{(m-1)} = (I\{1 = j\}, \ldots, I\{m-1 = j\}) \in (\mathbb{Z}_+)^{m-1}$, $1 \le j \le m-1$. Note that $\mathbb{I}_{a,b}(s_j^{(m-1)}, t_j^{(m-1)}) = 0$ for $1 \le j \le m-1$ because $m - 1 < (a+b)/2$. By the induction assumption on terms $\Delta_{i,a,b}(s_j^{(m-1)}, t_j^{(m-1)})$, $j > 0$, in (18) along with (11) and (17), we find that (14) holds whenever $\mathbb{I}_{a,b}(s^{(m)}, t^{(m)}) = 0$ with $m = T + 1$. This completes the induction proof and the treatment of Case 1.

Case 2. $\mathbb{I}_{a,b}(s^{(m)}, t^{(m)}) = 1$. Here a, b are even, and the components of $(s^{(m)}, t^{(m)})$ satisfy $s_j \in \{0,2\}$, $t_j = 2 - s_j$ for $1 \le j \le m = (a+b)/2$. By (15), $\sum_{k \in \mathbb{Z}} d_{k(1)}^{s_j} \cdot d_{k(i)}^{t_j} = 1$ follows for any $1 \le j \le m = (a+b)/2$, $i \in \mathbb{N}$. Using this and algebra similar to (18), we can iteratively write $\Delta_{i,a,b}(s^{(m)}, t^{(m)})$ as a sum by parts equal to

https://doi.org/10.1017/S0266466605050541 Published online by Cambridge University Press

$$\sum_{k_1 \neq \cdots \neq k_{m-1} \in \mathbb{Z}} \left(\prod_{j=1}^{m-1} d_{k_j(1)}^{s_j} \cdot d_{k_j(i)}^{t_j} \right) \left(1 - \sum_{h=1}^{m-1} d_{k_h(1)}^{s_m} \cdot d_{k_h(i)}^{t_m} \right)$$

$$= \sum_{k_1 \neq \cdots \neq k_{m-2} \in \mathbb{Z}} \left(\prod_{j=1}^{m-2} d_{k_j(1)}^{s_j} \cdot d_{k_j(i)}^{t_j} \right) \left(1 - \sum_{h=1}^{m-2} d_{k_h(1)}^{s_{m-1}} \cdot d_{k_h(i)}^{t_{m-1}} \right)$$

$$- R_{m-1,i,a,b}(s^{(m)}, t^{(m)})$$

$$\vdots$$

$$= 1 - \sum_{h=1}^{m-1} R_{h,i,a,b}(s^{(m)}, t^{(m)}),$$

$$R_{h,i,a,b}(s^{(m)}, t^{(m)}) = \sum_{k_1 \neq \cdots \neq k_h \in \mathbb{Z}} \left(\prod_{j=1}^{h} d_{k_j(1)}^{s_j} \cdot d_{k_j(i)}^{t_j} \right) \left(\sum_{j=1}^{h} d_{k_j(1)}^{s_{h+1}} \cdot d_{k_j(i)}^{t_{h+1}} \right),$$

where $1 \leq h \leq m - 1 = (a + b - 2)/2$. With an argument as in (16), we find a uniform bound $|R_{h,i,a,b}(s^{(m)}, t^{(m)})| \leq (a + b)\omega_\ell^2/2 = o(1)$ for each $i \in \mathbb{N}$, $1 \leq h \leq m - 1$. Hence, (14) follows for Case 2. The proof of Lemma 2(a) is now complete. ■

REFERENCES

Adenstedt, R.K. (1974) On large-sample estimation for the mean of a stationary sequence. *Annals of Statistics* 2, 1095–1107.

Bardet, J., G. Lang, G. Oppenheim, S. Philippe, & M.S. Taqqu (2003) Generators of long-range dependent processes: A survey. In P. Doukhan, G. Oppenheim, & M.S. Taqqu (eds.), *Theory and Applications of Long-Range Dependence*, pp. 579–624. Birkhäuser.

Bardet, J., G. Lang, G. Oppenheim, A. Philippe, S. Stoev, & M.S. Taqqu (2003) Semi-parametric estimation of the long-range dependence parameter: A survey. In P. Doukhan, G. Oppenheim, & M.S. Taqqu (eds.), *Theory and Applications of Long-Range Dependence*, pp. 527–556. Birkhäuser.

Beran, J. (1989) A test of location for data with slowly decaying serial correlations. *Biometrika* 76, 261–269.

Beran, J. (1994) *Statistical Methods for Long Memory Processes.* Chapman & Hall.

Bingham, N.H., C.M. Goldie, & J.L. Teugels (1987) *Regular Variation.* Cambridge University Press.

Carlstein, E. (1986) The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Annals of Statistics* 14, 1171–1179.

Davydov, Y.A. (1970) The invariance principle for stationary processes. *Theory of Probability and Its Applications* 15, 487–498.

Dobrushin, R.L. & P. Major (1979) Non-central limit theorems for non-linear functionals of Gaussian fields. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 50, 827–842.

Doukhan, P. (1994) Mixing: properties and examples. *Lecture Notes in Statistics* 85, Springer-Verlag.

Fuller, W. (1996) *Introduction to Statistical Time Series*, 2nd ed. Wiley.

Granger, C.W.J. & R. Joyeux (1980) An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15–29.

Geweke, J. & S. Porter-Hudak (1983) The estimation and application of long-memory time series models. *Journal of Time Series Analysis* 4, 221–238.

Hall, P., J.L. Horowitz, & B.-Y. Jing (1995) On blocking rules for the bootstrap with dependent data. *Biometrika* 82, 561–574.

Hall, P. & B.-Y. Jing (1996) On sample reuse methods for dependent data. *Journal of the Royal Statistical Society, Series B* 58, 727–737.

Hall, P., B.-Y. Jing, & S.N. Lahiri (1998) On the sampling window method for long-range dependent data. *Statistica Sinica* 8, 1189–1204.

Henry, M. & P. Zaffaroni (2003) The long-range dependence paradigm for macroeconomics and finance. In P. Doukhan, G. Oppenheim, & M.S. Taqqu (eds.), *Theory and Applications of Long-Range Dependence*, pp. 417–438. Birkhäuser.

Hosking, J.R.M. (1981) Fractional differencing. *Biometrika* 68, 165–176.

Hurvich, C.M., R. Deo, & J. Brodsky (1998) The mean square error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series. *Journal of Time Series Analysis* 19, 19–46.

Ibragimov, I.A. & Y.V. Linnik (1971) *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff.

Künsch, H.R. (1989) The jackknife and bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1261.

Lahiri, S.N. (1993) On the moving block bootstrap under long range dependence. *Statistics and Probability Letters* 18, 405–413.

Liu, R.Y. & K. Singh (1992) Moving blocks jackknife and bootstrap capture weak dependence. In R. LePage and L. Billard (eds.), *Exploring the Limits of Bootstrap,* pp. 225–248. Wiley.

Mandelbrot, B.B. & J.W. van Ness (1968) Fractional Brownian motions, fractional noises and applications. *SIAM Review* 10, 422–437.

Montanari, A. (2003) Long-range dependence in hydrology. In P. Doukhan, G. Oppenheim, & M.S. Taqqu (eds.), *Theory and Applications of Long-Range Dependence*, pp. 461–472. Birkhäuser.

Politis, D.N. & J.P. Romano (1994) Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics* 22, 2031–2050.

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

Taqqu, M.S. (1977) Law of the iterated logarithm for sums of non-linear functions of Gaussian variables that exhibit a long range dependence. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 40, 203–238.

Taqqu, M.S. (1979) Convergence of integrated processes of arbitrary Hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 50, 53–83.

Wood, A.T.A. & G. Chan (1994) Simulation of stationary Gaussian processes in $[0,1]^d$. *Journal of Computational and Graphical Statistics* 3, 409–432.