BlockSampling-2013-02-14.pdf

Block Sampling under Strong Dependence [1]

Ting Zhang[a], Hwai-Chung Ho[b], Martin Wendler[c], Wei Biao Wu[d]

[a] *Department of Statistics and Actuarial Science, The University of Iowa, USA*

[b] *Institute of Statistical Science, Academia Sinica, Taipei*

[c] *Fakultät für Mathematatik, Ruhr-Universität Bochum, Germany*

[d] *Department of Statistics, The University of Chicago, USA*

February 14, 2013

Abstract

The paper considers the block sampling method for long-range dependent processes. Our theory generalizes earlier ones by Hall et al. (1998) on functionals of Gaussian processes and Nordman and Lahiri (2005) on linear processes. In particular, we allow nonlinear transforms of linear processes. Under suitable conditions on physical dependence measures, we prove the validity of the block sampling method. Its finite-sample performance is illustrated by a simulation study.

1 Introduction

Long memory (strongly dependent, or long-range dependent) processes have received considerable attention in areas including econometrics, finance, geology and telecommunication among others. Let X_i, $i \in \mathbb{Z}$, be a stationary linear process of the form

$$X_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j}, \qquad (1)$$

where ε_i, $i \in \mathbb{Z}$, are independent and identically distributed (iid) random variables with zero mean, finite variance and $(a_j)_{j=0}^{\infty}$ are square summable real coefficients. If $a_i \to 0$ very slowly, say $a_i \sim i^{-\beta}$, $1/2 < \beta < 1$, then there exists a constant $c_\beta > 0$ such that the covariances $\gamma_i = \mathbb{E}(X_0 X_i) = \mathbb{E}(\varepsilon_0^2) \sum_{j=0}^{\infty} a_j a_{i+j} \sim c_\beta \mathbb{E}(\varepsilon_0^2) i^{1-2\beta}$ are not summable,

[1] *Keywords:* Asymptotic normality; Covariance; Hermite processes; Linear processes; Long-range dependence; Rosenblatt distribution

thus suggesting strong dependence. An important example is the fractionally integrated autoregressive moving average (FARIMA) processes (Granger and Joyeux, 1980 and Hosking, 1981). Let K be a measurable function such that $\mathbb{E}[K^2(X_i)] < \infty$, and $\mu = \mathbb{E}K(X_i)$. This paper considers the asymptotic sampling distribution of

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} K(X_i) = \frac{S_n}{n} + \mu, \text{ where } S_n = \sum_{i=1}^{n} [K(X_i) - \mu].$$

In the inference of the mean μ, such as the construction of confidence intervals and hypothesis testing, it is necessary to develop a large sample theory for the partial sum process S_n. The latter problem has a substantial history. Here we shall only give a very brief account. Davydov (1970) considered the special case $K(x) = x$ and Taqqu (1975) and Dobrushin and Major (1979) dealt with another special case in which K can be a nonlinear transform while (X_i) is a Gaussian process. Other contributions can be found in Surgailis (1982), Avram and Taqqu (1987) and Dittmann and Granger (2002), and Wu (2006) contains further references. For general linear processes with nonlinear transforms, under some regularity conditions on K, if X_i is a short memory (or short-range dependent) process with $\sum_{j=0}^{\infty} |a_j| < \infty$, then $S_n/\sqrt{n}$ satisfies a central limit theorem with a Gaussian limiting distribution; if X_i is long-memory (or long-range dependent), then with proper normalization, S_n may have either a non-Gaussian or Gaussian limiting distribution and the normalizing constant may no longer be $\sqrt{n}$ (Ho and Hsing, 1997 and Wu, 2006). In many situations, the non-Gaussian limiting distribution can be expressed as a multiple Wiener-Itô integral (MWI); see equation (2).

The distribution function of a non-Gaussian WMI does not have a close form. This brings considerable inconveniences in the related statistical inference. As a useful alternative, we can resort to re-sampling techniques to estimate the sampling distribution of S_n. Künsch (1989) proved the validity of the moving block bootstrap method for weakly dependent stationary processes. However, Lahiri (1993) showed that, for Gaussian subordinated long-memory processes, the block bootstrapped sample means are always asymptotically Gaussian; thus it fails to recover the non-Gaussian limiting distribution of the multiple

Wiener-Itô integrals. On the other hand, Hall and Jing (1996) proposed a sampling windows method. Hall et al. (1998) showed that, for the special class of processes of nonlinear transforms of Gaussian processes, the latter method is valid in the sense that the empirical distribution functions of the consecutive block sums converge to the limiting distribution of S_n with a proper normalization. Nordman and Lahiri (2005) proved that the same method works for linear processes, an entirely different special class of stationary processes. However, for linear processes, the limiting distribution is always Gaussian. It has been an open problem whether a limit theory can be established for a more general class of long-memory processes.

Here we shall provide an affirmative answer to the above question by allowing functionals of linear processes, a more general class of stationary processes which include linear processes and nonlinear transforms of Gaussian processes as special cases. Specifically, given a realization $Y_i = K(X_i)$, $1 \leq i \leq n$, with both K and X_i being possibly unknown or unobserved, we consider consistent estimation of the sampling distribution of S_n/n. To this end, we shall implement the concept of physical dependence measures (Wu, 2005) which quantify the dependence of a random process by measuring how outputs depend on inputs. The rest of the paper is organized as follows. Section 2 presents the main results and it deals with the asymptotic consistency of the empirical distribution functions of the normalized consecutive block sums. It is interesting to observe that the same sampling windows method works for both Gaussian and non-Gaussian limiting distributions. A simulation study is provided in Section 3, and some proofs are deferred to the Appendix.

2 Main Results

In Section 2.1, we briefly review the asymptotic theory of S_n in Ho and Hsing (1997) and Wu (2006). The block sampling method of Hall and Jing (1996) is described in Section 2.2. With physical dependence measures, Section 2.3 presents a consistency result for empirical sampling distributions. In Section 2.4, we obtain a convergence rate for a variance estimate of $s_l^2 = \|S_l\|^2$, and a subsampling procedure is proposed in Section 2.5 for making statistical inference.

For two positive sequences (a_n) and (b_n), write $a_n \sim b_n$ if $a_n/b_n \to 1$ and $a_n \asymp b_n$ if there exists a constant $C > 0$ such that $a_n/C \leq b_n \leq C a_n$ holds for all large n. Let $\mathcal{C}_A$ (resp. $\mathcal{C}_A^p$) denote the collection of continuous functions (resp. functions having p-th order continuous derivatives) on $A \subseteq \mathbb{R}$. Denote by "$\Rightarrow$" the weak convergence; see Billingsley (1968) for a detailed account for the weak convergence theory on $\mathcal{C}_{[0,1]}$. For a random variable Z, we write $Z \in \mathcal{L}^\nu$, $\nu > 0$, if $\|Z\|_\nu = (\mathbb{E}|Z|^\nu)^{1/\nu} < \infty$, and write $\|Z\| = \|Z\|_2$. For integers $i \leq j$ define $\mathcal{F}_i^j = (\varepsilon_i, \varepsilon_{i+1}, \ldots, \varepsilon_j)$. Write $\mathcal{F}_i^\infty = (\varepsilon_i, \varepsilon_{i+1}, \ldots)$ and $\mathcal{F}_{-\infty}^j = (\ldots, \varepsilon_{j-1}, \varepsilon_j)$. Define the projection operator $\mathcal{P}_j$, $j \in \mathbb{Z}$, by

$$\mathcal{P}_j \cdot = \mathbb{E}(\cdot|\mathcal{F}_{-\infty}^j) - \mathbb{E}(\cdot|\mathcal{F}_{-\infty}^{j-1}).$$

Then $\mathcal{P}_j\cdot$, $j \in \mathbb{Z}$, yield martingale differences.

2.1 Asymptotic distributions

To study the asymptotic distribution of S_n under strong dependence, we shall introduce the concept of power rank (Ho and Hsing, 1997). Based on K and X_n, let $X_{n,i} = \sum_{j=n-i}^\infty a_j \varepsilon_{n-j} = \mathbb{E}(X_n|\mathcal{F}_{-\infty}^i)$ be the tail process and define functions

$$K_\infty(x) = \mathbb{E}K(x + X_n) \text{ and } K_n(x) = \mathbb{E}K(x + X_n - X_{n,0}).$$

Note that $X_n - X_{n,0} = \sum_{j=0}^{n-1} a_j \varepsilon_{n-j}$ is independent of $X_{n,0}$. Denote by $\kappa_r = K_\infty^{(r)}(0)$, the r-th derivative, if it exists. If $p \in \mathbb{N}$ is such that $\kappa_p \neq 0$ and $\kappa_r = 0$ for all $r = 1, \ldots, p-1$, then we say that K has power rank p with respect to the distribution of X_i. The limiting distribution of S_n can be Gaussian or non-Gaussian. The non-Gaussian limiting distribution here is expressed as MWIs. To define the latter, let the simplex $\mathcal{S}_t = \{(u_1, \ldots, u_r) \in \mathbb{R}^r : -\infty < u_1 < \ldots < u_r < t\}$ and $\{I\!B(u), u \in \mathbb{R}\}$ be a standard two-sided Brownian motion. For $1/2 < \beta < 1/2 + 1/(2r)$, define the *Hermite process*

(Surgailis, 1982 and Avram and Taqqu, 1987) as the MWI

$$Z_{r,\beta}(t) = \int_{\mathcal{S}_t} \int_0^t \prod_{i=1}^r g_\beta(v - u_i) dv\, d\mathbb{B}(u_1) \dots d\mathbb{B}(u_r), \tag{2}$$

where $g_\beta(x) = x^{-\beta}$ if $x > 0$ and $g_\beta(x) = 0$ if $x \le 0$. It is non-Gaussian if $r \ge 2$. Note that $Z_{1,\beta}(t)$ is the fractional Brownian motion with Hurst index $H = 3/2 - \beta$.

Let $\ell(n)$ be a slowly varying function, namely $\lim_{n\to\infty} \ell(un)/\ell(n) = 1$ for all $u > 0$ (Bingham et al., 1987). Assume $a_0 \neq 0$ and a_i has the form

$$a_i = i^{-\beta}\ell(i), \quad i \ge 1, \text{ where } 1/2 < \beta < 1. \tag{3}$$

Under (3), we say that (a_i) is regularly varying with index β. Let $a_i = 0$ if $i < 0$, we need the following regularity condition on K and the process (X_i).

Condition 1. *For a function f and $\lambda > 0$, write $f(x;\lambda) = \sup_{|u|\le\lambda} |f(x+u)|$. Assume $\varepsilon_1 \in \mathcal{L}^{2\nu}$ with $\nu \ge 2$, $K_n \in \mathcal{C}_{\mathbb{R}}^{p+1}$ for all large n, and for some $\lambda > 0$,*

$$\sum_{\alpha=0}^{p+1} \|K_{n-1}^{(\alpha)}(X_{n,0};\lambda)\|_\nu + \sum_{\alpha=0}^{p-1} \|\varepsilon_1^2 K_{n-1}^{(\alpha)}(X_{n,1})\|_\nu + \|\varepsilon_1 K_{n-1}^{(p)}(X_{n,1})\|_\nu = O(1). \tag{4}$$

We remark that in Condition 1 the function K itself does not have to be continuous. For example, if $K(x) = \mathbf{1}_{x\le 0}$; let $a_0 = 1$ and F_ε (resp. f_ε) be the distribution (resp. density) function of ε_i. Then $K_1(x) = F_\varepsilon(-x)$ which is in $\mathcal{C}_{\mathbb{R}}^{p+1}$ if F_ε is so. If $\sup_x |K_{n-1}^{(1+p)}(x)| < \infty$, then for all $0 \le \alpha \le p$, there exists a constant $C > 0$ such that $|K_{n-1}^{(\alpha)}(x)| \le C(1+|x|)^{1+p-\alpha}$, and (4) holds if $\varepsilon_i \in \mathcal{L}^{2\nu(1+p)}$.

Theorem 1. *(Wu, 2006) Assume that K has power rank $p \ge 1$ with respect to X_i and Condition 1 holds with $\nu = 2$. (i) If $p(2\beta - 1) < 1$, let*

$$\sigma_{n,p} = n^H \ell^p(n)\kappa_p \|Z_{p,\beta}(1)\|, \text{ where } H = 1 - p(\beta - 1/2), \tag{5}$$

then in the space $\mathcal{C}_{[0,1]}$ we have the weak convergence

$$\{S_{nt}/\sigma_{n,p},\ 0 \le t \le 1\} \Rightarrow \{Z_{p,\beta}(t)/\|Z_{p,\beta}(1)\|,\ 0 \le t \le 1\}.$$

(ii) If $p(2\beta - 1) > 1$*, then* $D_0 := \sum_{j=0}^{\infty} \mathcal{P}_0 Y_j \in \mathcal{L}^2$*. Assume* $\|D_0\| > 0$*. Then we have*

$$\{S_{nt}/\sigma_n,\ 0 \le t \le 1\} \Rightarrow \{\mathbb{B}(t),\ 0 \le t \le 1\},\ \textit{where } \sigma_n = \|D_0\|\sqrt{n}. \tag{6}$$

The above result can not be directly applied for making statistical inference for the mean $\mu = \mathbb{E}K(X_i)$ since $\sigma_{n,p}$ and σ_n are typically unknown. Additionally, the dichotomy in Theorem 1 causes considerable inconveniences in hypothesis testings or constructing confidence intervals for μ. The primary goal of the paper is to establish the validity of some re-sampling techniques so that the distribution of S_n can be estimated.

2.2 Block sampling

At the outset we assume that $\mu = \mathbb{E}K(X_i) = 0$. The block sampling method by Hall and Jing (1996) can be described as follows. Let l be the block size satisfying $l = l_n \to \infty$ and $l/n \to 0$. Define

$$s_l = \|S_l\|,$$

and the empirical distribution function

$$F_l(x) = \frac{1}{n-l+1} \sum_{i=l}^{n} \mathbf{1}_{B_{i,l} \le x s_l},\ \text{where } B_{i,l} = Y_i + Y_{i-1} + \cdots + Y_{i-l+1}. \tag{7}$$

Recall (5) and (6) for the definitions of $\sigma_{n,p}$ and σ_n, respectively. Lemma 1 asserts that they are asymptotically equivalent to s_n.

Lemma 1. *Under conditions in Theorem 1(i), we have*

$$s_l \sim \sigma_{l,p} = l^H \ell^p(l) \kappa_p \|Z_{p,\beta}(1)\|, \tag{8}$$

as $l \to \infty$*. Under conditions in Theorem 1(ii), we have*

$$s_l \sim \sigma_l = \|D_0\|\sqrt{l}. \tag{9}$$

Under either case, $l\|S_n/n\| = o(s_l)$ *if* $l \asymp n^{r_0}$, $0 < r_0 < 1$.

If s_l is known, we say that the block sampling method is valid if

$$\sup_{x\in\mathbb{R}} |F_l(x) - \mathbb{P}(S_n/s_n \le x)| \to 0 \text{ in probability.} \tag{10}$$

In the long-memory case, the above convergence relation has a deeper layer of meaning since, by Theorem 1, S_n/s_n can have either a Gaussian or non-Gaussian limiting distribution. In comparison, for short-memory processes, typically S_n/s_n has a Gaussian limit. Ideally, we hope that (10) holds for both cases in Theorem 1. Then we do not need to worry about the dichotomy of which limiting distribution to use. As a primary goal of the paper, we in Section 2.3 show that this is indeed the case.

2.3 Consistency of empirical sampling distributions

Theorem 2 confirms that $F_l(\cdot)$ consistently estimates the distribution of S_n/s_n, regardless of whether the limiting distribution of the latter is Gaussian or not. In other words, $F_l(\cdot)$ automatically adapts the limiting distribution of S_n/s_n. Bertail et al. (1999) obtained a result of similar nature for strong mixing processes where the limiting distribution can possibly be non-Gaussian; see also Politis et al. (1999).

Theorem 2. *Assume $\mu = \mathbb{E}Y_i = 0$, $p \ge 1$, $l \asymp n^{r_0}$ for some $0 < r_0 < 1$, and Condition 1 holds with $\nu = 2$. (i) If $p(2\beta - 1) < 1$, then*

$$\sup_{x\in\mathbb{R}} |F_l(x) - \mathbb{P}(Z_{p,\beta}(1) \le x)| \to 0 \textit{ in probability.} \tag{11}$$

(ii) Let $Z \sim N(0,1)$ be standard Gaussian. If $p(2\beta - 1) > 1$, we have

$$\sup_{x\in\mathbb{R}} |F_l(x) - \mathbb{P}(Z \le x)| \to 0 \textit{ in probability.}$$

Hence under either (i) or (ii), we have (10).

Before providing the proof of Theorem 2, we shall introduce some notation. Let $(\varepsilon'_j)_{j\in\mathbb{Z}}$ be an iid copy of $(\varepsilon_j)_{j\in\mathbb{Z}}$, hence $\varepsilon'_i, \varepsilon_l$, $i, l \in \mathbb{Z}$, are iid, and

$$X_i^* = X_i + \sum_{j=-\infty}^{0} a_{i-j}(\varepsilon'_j - \varepsilon_j). \tag{12}$$

We can view X_i^* as a coupled process of X_i with ε_j, $j \le 0$, in the latter replaced by their iid copies ε_j', $j \le 0$. Note that, if $i \le 0$, the two random variables X_i and $X_i^* = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j}'$ are independent of each other. Following Wu (2005), we define the physical dependence measure

$$\tau_{i,\nu} = \|K(X_i) - K(X_i^*)\|_\nu, \tag{13}$$

which quantifies how the process $Y_i = K(X_i)$ forgets the past ε_j, $j \le 0$.

Proof. (Theorem 2) We first prove (i). Since $Z_{p,\beta}(1)$ has a continuous distribution, by the Glivenko-Cantelli argument (cf. Chow and Teicher, 1997) for the uniform convergence of empirical distribution functions, (11) follows if we can show that, for any fixed x,

$$\mathbb{E}|F_l(x) - \mathbb{P}(Z_{p,\beta}(1) \le x)|^2 = \mathrm{var}(F_l(x)) + |\mathbb{E}F_l(x) - \mathbb{P}(Z_{p,\beta}(1) \le x)|^2 \to 0.$$

Note that $B_{i,l}/s_l \Rightarrow Z_{p,\beta}(1)$ as $n \to \infty$, the second term on the right hand side of the above converges to 0. We now show that the first term

$$\mathrm{var}(F_l(x)) \le \frac{2}{n-l+1} \sum_{i=0}^{n-1} |\mathrm{cov}(\mathbf{1}_{B_{0,l}/s_l \le x},\ \mathbf{1}_{B_{i,l}/s_l \le x})| \to 0. \tag{14}$$

Here we use the fact that $(B_{i,l})_{i \in \mathbb{Z}}$ is a stationary process. To show (14), we shall apply the tool of coupling. Recall (12) for X_i^*. Let $B_{i,l}^* = \sum_{j=i-l+1}^{i} Y_j^*$, where $Y_j^* = K(X_j^*)$. Since $B_{i,l}^*$ and $\mathcal{F}_{-\infty}^0$ are independent, $\mathbb{E}(\mathbf{1}_{B_{i,l}^*/s_l \le x} | \mathcal{F}_{-\infty}^0) = \mathbb{P}(B_{i,l}^*/s_l \le x)$. Hence

$$\begin{aligned} |\mathrm{cov}(\mathbf{1}_{B_{0,l}/s_l \le x}, \mathbf{1}_{B_{i,l}/s_l \le x})| &= |\mathbb{E}[\mathbf{1}_{B_{0,l}/s_l \le x}(\mathbf{1}_{B_{i,l}/s_l \le x} - \mathbf{1}_{B_{i,l}^*/s_l \le x})]| \\ &\le \mathbb{E}|\mathbf{1}_{B_{i,l}/s_l \le x} - \mathbf{1}_{B_{i,l}^*/s_l \le x}|. \end{aligned} \tag{15}$$

For any fixed $\lambda > 0$, by the triangle and the Markov inequalities,

$$\begin{aligned} \mathbb{E}|\mathbf{1}_{B_{i,l}/s_l \le x} - \mathbf{1}_{B_{i,l}^*/s_l \le x}| &\le \mathbb{E}(\mathbf{1}_{|B_{i,l}/s_l - x| \le \lambda}) + \mathbb{E}(\mathbf{1}_{|B_{i,l}/s_l - B_{i,l}^*/s_l| \ge \lambda}) \\ &\le \mathbb{P}(|B_{i,l}/s_l - x| \le \lambda) + \frac{\|B_{i,l} - B_{i,l}^*\|}{\lambda s_l}. \end{aligned} \tag{16}$$

Since $\mathbb{E}(B_{i,l}|\mathcal{F}_1^\infty) = \mathbb{E}(B_{i,l}^*|\mathcal{F}_1^\infty)$ for $i > 2l$, by Lemma 4(ii) and the fact that $B_{i,l}^* - \mathbb{E}(B_{i,l}^*|\mathcal{F}_1^\infty)$ and $B_{i,l} - \mathbb{E}(B_{i,l}|\mathcal{F}_1^\infty)$ are identically distributed, we have

$$\|B_{i,l} - B_{i,l}^*\| \le \|B_{i,l} - \mathbb{E}(B_{i,l}|\mathcal{F}_1^\infty)\| + \|\mathbb{E}(B_{i,l}|\mathcal{F}_1^\infty) - B_{i,l}^*\|$$

$$
\begin{aligned}
&= 2\|B_{i,l} - \mathbb{E}(B_{i,l}|\mathcal{F}_1^\infty)\| \\
&= 2\|S_l - \mathbb{E}(S_l|\mathcal{F}_{l+1-i}^\infty)\| = s_l O[l^{-\varphi_1} + (l/i)^{\varphi_2}]. \qquad (17)
\end{aligned}
$$

Assume without loss of generality that $\varphi_2 < 1$. Otherwise we can replace it by $\varphi_2' = \min(\varphi_2, 1/2)$. By Lemma 4(i) and Lemma 1, we have $\|B_{0,l}\| = O(s_l)$. Recall that $l \asymp n^{r_0}$, $0 < r_0 < 1$, we have

$$
\begin{aligned}
\frac{1}{n}\sum_{i=0}^{n-1} \frac{\|B_{i,l} - B_{i,l}^*\|}{s_l} &= \frac{O(1)}{n}\sum_{i=0}^{2l} O(1) + \frac{O(1)}{n}\sum_{i=2l+1}^{n-1} O[l^{-\varphi_1} + (l/i)^{\varphi_2}] \\
&= O(l/n) + O(l^{-\varphi_1}) + O[(l/n)^{\varphi_2}] = O(n^{-\phi}), \qquad (18)
\end{aligned}
$$

where $\phi = \min(1 - r_0, \varphi_1 r_0, (1 - r_0)\varphi_2)$. Since $\mathbb{P}(|B_{i,l}/s_l - x| \le \lambda) \to \mathbb{P}(|Z_{p,\beta}(1) - x| \le \lambda)$, (14) then follows from (15) and (16) by first letting $n \to \infty$, and then $\lambda \to 0$.

For (ii), by the argument in (i), it suffices to show that

$$
\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{\|B_{i,l} - \mathbb{E}(B_{i,l}|\mathcal{F}_1^\infty)\|}{\sqrt{l}} = 0. \qquad (19)
$$

More specifically, if (19) is valid, then by $\|B_{i,l} - B_{i,l}^*\| \le 2\|B_{i,l} - \mathbb{E}(B_{i,l}|\mathcal{F}_1^\infty)\|$, we have (18) and consequently (14). Let $N > 3l$ and $G_N = B_{N,l} - \mathbb{E}(B_{N,l}|\mathcal{F}_1^\infty)$. Observe that $(\mathcal{P}_k G_N)_{k=-\infty}^N$ is a sequence of martingale differences and $G_N = \sum_{k=-\infty}^N \mathcal{P}_k G_N$, we have

$$
\|G_N\|^2 = \sum_{k=-\infty}^{N} \|\mathcal{P}_k G_N\|^2. \qquad (20)
$$

By (45) and Lemma 2 with $\nu = 2$, we know that the predictive dependence measures $\eta_i = \|\mathcal{P}_0 Y_i\|$ is summable. Recall (13) for $\tau_{n,\nu}$. Let $\tau_n^* = \max_{m\ge n} \tau_{m,2}$. Then τ_n^* is non-increasing and $\lim_{n\to\infty} \tau_n^* = 0$. Since $\|\mathcal{P}_k \mathbb{E}(Y_j|\mathcal{F}_1^\infty)\| \le \|\mathcal{P}_k Y_j\| = \eta_{j-k}$ and $\|Y_j - \mathbb{E}(Y_j|\mathcal{F}_1^\infty)\| \le \tau_{j,2}$, we have

$$
\begin{aligned}
\|\mathcal{P}_k G_N\| &\le \sum_{j=N-l+1}^{N} \|\mathcal{P}_k[Y_j - \mathbb{E}(Y_j|\mathcal{F}_1^\infty)]\| \\
&\le \sum_{j=N-l+1}^{N} \min(2\eta_{j-k},\, \tau_{N-l+1}^*) \le \eta_*, \qquad (21)
\end{aligned}
$$

where $\eta_* = 2\sum_{i=0}^{\infty}\eta_i$. Then, by (20) and the Lebesgue dominated convergence theorem, we have

$$\begin{aligned}
\lim_{n\to\infty}\frac{\|G_N\|^2}{l} &\le \lim_{n\to\infty}\sum_{k=-\infty}^{N}\frac{\eta_*}{l}\|\mathcal{P}_k G_N\| \\
&\le \lim_{n\to\infty}\sum_{k=-\infty}^{N}\frac{\eta_*}{l}\sum_{j=N-l+1}^{N}\min(2\eta_{j-k},\, \tau^*_{N-l+1}) \\
&\le \lim_{n\to\infty}\eta_*\sum_{i=0}^{\infty}\min(2\eta_i,\, \tau^*_{N-l+1}) = 0, \qquad (22)
\end{aligned}$$

since $\tau^*_{N-l+1} \le \tau^*_l \to 0$ as $l\to\infty$ and η_i are summable. Hence $\sum_{N=3l}^{n}\|G_N\|^2 = o(nl)$. Note that $l = o(n)$, (19) follows by the inequality $(\sum_{i=1}^{n}|z_i|/n)^2 \le \sum_{i=1}^{n} z_i^2/n$. ♢

2.4 Variance estimation

In practice both $\mu = \mathbb{E}K(X_i)$ and s_l are unknown, and Theorem 2 is not directly applicable. We propose to estimate μ by $\bar{Y}_n = \sum_{i=1}^{n} Y_i/n$ and s_l by

$$\tilde{s}_l^2 = \frac{\tilde{Q}_{n,l}}{n-l+1}, \text{ where } \tilde{Q}_{n,l} = \sum_{i=l}^{n}|B_{i,l} - l\bar{Y}_n|^2. \qquad (23)$$

The realized version of $F_l(x)$ in (7) now has the form

$$\tilde{F}_l(x) = \frac{1}{n-l+1}\sum_{i=l}^{n}\mathbf{1}_{B_{i,l}-l\bar{Y}_n \le x\tilde{s}_l},$$

and correspondingly (10) becomes

$$\sup_{x\in\mathbb{R}}|\tilde{F}_l(x) - \mathbb{P}(S_n/\tilde{s}_n \le x)| \to 0 \text{ in probability.} \qquad (24)$$

Note that $\bar{Y}_n - \mu = o_{\mathbb{P}}(s_l/l) = o_{\mathbb{P}}(1)$ from Lemma 1. Hence, by Theorems 1 and 2, (24) holds if the estimates $\tilde{s}_l$ and $\tilde{s}_n$ satisfy $\tilde{s}_l/s_l \to 1$ and $\tilde{s}_n/s_n \to 1$ in probability. With (24), we can construct the two-sided $(1-\alpha)$-th $(0<\alpha<1)$ and the upper one-sided $(1-\alpha)$-th confidence intervals for μ as $[\bar{Y}_n - \tilde{q}_{1-\alpha/2}\tilde{s}_n/n,\, \bar{Y}_n - \tilde{q}_{\alpha/2}\tilde{s}_n/n]$ and $[\bar{Y}_n - \tilde{q}_{1-\alpha}\tilde{s}_n/n,\, \infty)$ respectively, where $\tilde{q}_\alpha$ is the α-th sample quantile of $\tilde{F}_l(\cdot)$. Note that s_n cannot be analogously estimated by (23), we will address the issue of how to construct the confidence intervals for μ without directly having to estimate s_n later in Section 2.5. Theorem 3 provides the asymptotic property of $\tilde{s}_l$, which entails its consistency.

Theorem 3. *Assume that $l \asymp n^{r_0}$, $0 < r_0 < 1$, and Condition 1 holds with $\nu = 4$. (i) If $p(2\beta - 1) < 1$, then there exists a constant $0 < \phi < 1$ such that*

$$\operatorname{var}(\tilde{s}_l^2 / s_l^2) = O(n^{-\phi}). \tag{25}$$

(ii) If $p(2\beta - 1) > 1$, then $\operatorname{var}(\tilde{s}_l^2/s_l^2) \to 0$. (iii) If $p(2\beta - 1) > 1$ and $\tau_{n,4} = O(n^{-\phi_1})$ for some $\phi_1 > 0$, then (25) holds as well.

Proof. For (i), we first consider the case with $\mu = 0$. Then s_l^2 can be estimated by

$$\hat{s}_l^2 = \frac{\hat{Q}_{n,l}}{n-l+1}, \text{ where } \hat{Q}_{n,l} = \sum_{i=l}^{n} B_{i,l}^2,$$

and we show that, for some $\phi > 0$,

$$\operatorname{var}(\hat{s}_l^2 / s_l^2) = O(n^{-\phi}). \tag{26}$$

Recall that $B_{n,l}^* = \sum_{j=n-l+1}^{n} Y_j^*$, where $Y_j^* = K(X_j^*)$. Then $\mathbb{E}(B_{i,l}^2) = \mathbb{E}[(B_{i,l}^*)^2 | \mathcal{F}_{-\infty}^0]$ and $\operatorname{cov}(B_{0,l}^2, B_{i,l}^2) = \mathbb{E}[B_{0,l}^2(B_{i,l}^2 - (B_{i,l}^*)^2)]$ since $B_{0,l}$ and $B_{i,l}^*$ are independent for any $i \geq l$. By the Cauchy-Schwarz inequality,

$$\begin{aligned}
\operatorname{var}(\hat{s}_l^2) &= \frac{1}{n-l+1} \sum_{i=l-n}^{n-l} [1 - |i|/(n-l+1)] \operatorname{cov}(B_{0,l}^2, B_{i,l}^2) \\
&\leq \frac{2}{n-l+1} \sum_{i=0}^{n-1} \|B_{0,l}^2\| \|B_{i,l}^2 - (B_{i,l}^*)^2)\| \\
&\leq \frac{2}{n-l+1} \sum_{i=0}^{n-1} \|B_{0,l}\|_4^2 \|B_{i,l} + B_{i,l}^*\|_4 \|B_{i,l} - B_{i,l}^*\|_4.
\end{aligned} \tag{27}$$

By Lemma 4(ii) and the argument (17) in the proof of Theorem 2(i), for $i > 2l$, we have

$$\|B_{i,l} - B_{i,l}^*\|_4 \leq \|B_{i,l} - \mathbb{E}(B_{i,l}|\mathcal{F}_0^\infty)\|_4 + \|\mathbb{E}(B_{i,l}|\mathcal{F}_0^\infty) - B_{i,l}^*\|_4$$

$$= \quad s_l O[l^{-\varphi_1} + (l/i)^{\varphi_2}], \tag{28}$$

in view of Lemma 1 since $\|B_{i,l}\| \sim s_l$. Again we assume without loss generality that $\varphi_2 < 1$. By Lemma 4(i), $\|B_{0,l}\|_4 = O(\sigma_{l,r})$. So (27) similarly implies (26) via

$$\begin{aligned} \mathrm{var}(\hat{s}_l^2/s_l^2) &= \frac{O(1)}{n}\sum_{i=0}^{2l} O(1) + \frac{O(1)}{n}\sum_{i=2l+1}^{n-1} O[l^{-\varphi_1} + (l/i)^{\varphi_2}] \\ &= O(l/n) + O(l^{-\varphi_1}) + O((l/n)^{\varphi_2}) = O(n^{-\phi}) \end{aligned} \tag{29}$$

with $\phi = \min(1-r_0,\, \varphi_1 r_0,\, (1-r_0)\varphi_2)$ since $l \asymp n^{r_0}$, $0 < r_0 < 1$.

Now we shall show that (26) implies (25). By Lemma 4(i) and the Cauchy-Schwarz inequality,

$$\begin{aligned} \|\hat{Q}_{n,l} - \tilde{Q}_{n,l}\| &= \left\| (n-l+1)(l\bar{Y}_n)^2 - 2l|\bar{Y}_n| \sum_{i=l}^{n} B_{i,l} \right\| \\ &\leq nl^2\|\bar{Y}_n^2\| + \|2l\bar{Y}_n\|_4 l \|Y_1 + \cdots + Y_n\|_4 \\ &= O(l^2 s_n^2/n) \\ &= n s_l^2 O(l^2 s_n^2/(n^2 s_l^2)) \\ &= n s_l^2 O[(l/n)^{2-2H} \ell^{2p}(n)/\ell^{2p}(l)] \\ &= n s_l^2 O(n^{-\theta}), \end{aligned} \tag{30}$$

where $0 < \theta < (2-2H)(1-r_0)$. Hence (25) follows from Lemma 1.

For (iii), by (38) and (45), under $p(2\beta-1) > 1$, for $0 < \varphi_3 < p(2\beta - 1)$, the predictive dependence measure

$$\begin{aligned} \eta_{i,4} &:= \|\mathcal{P}_0 Y_i\|_4 = \|\mathcal{P}_0(L_{n,p} + \kappa_p U_{n,p})\|_4 \leq |\kappa_p| \|\mathcal{P}_0 U_{n,p}\|_4 + \|\mathcal{P}_0 L_{n,p}\|_4 \\ &= O(a_n A_n^{(p-1)/2}) + a_n O(a_n + A_{n+1}^{1/2}(4) + A_{n+1}^{p/2}) \\ &= O(i^{-1-\varphi_3}), \end{aligned}$$

where $L_{n,p}$ is defined in (36). Recall the proof of Theorem 2(ii) for the definition of G_N, $N > 3l$. By (39), $\|G_N\|_4^2 \leq C_4 \sum_{k=-\infty}^{N} \|\mathcal{P}_k G_N\|_4^2$, and the arguments in (21) and (22), there exists a constant $C > 0$ such that

$$\frac{\|G_N\|_4^2}{l} \leq C \eta_{*,4} \sum_{i=0}^{\infty} \min(\eta_{i,4},\, \tau^*_{N-l+1,4}),$$

where $\eta_{*,4} = \sum_{i=0}^{\infty} \eta_{i,4}$ and $\tau^*_{n,4} = \max_{m \geq n} \tau_{m,4}$. As $\tau \to 0$, we have $\sum_{i=0}^{\infty} \min(\eta_{i,4}, \tau) = O(\tau^{\varphi_4})$, where $\varphi_4 = \varphi_3/(1+\varphi_3)$. Similarly as (27), (28) and (29),

$$\begin{aligned}
\mathrm{var}(\hat{s}_l^2/s_l^2) &= \frac{O(1)}{n} \sum_{i=0}^{n-1} \frac{\|G_i\|_4}{\sqrt{l}} \\
&= \frac{O(1)}{n} \sum_{i=1+3l}^{n-1} \frac{\|G_i\|_4}{\sqrt{l}} + O(l/n) \\
&= \frac{O(1)}{n} \sum_{i=1+3l}^{n-1} i^{-\phi_1 \varphi_4/2} + O(l/n) \\
&= O(n^{-\phi_1 \varphi_4/2}) + O(l/n).
\end{aligned}$$

So (26), and hence (25) follows in view of (30).

For (ii), as in the proof Theorem 2(ii), it follows from the Lebesgue dominated convergence theorem since $\tau^*_{m,4} \to 0$ as $m \to \infty$. $\diamondsuit$

2.5 A subsampling approach

Although s_l can be consistently estimated by (23), there is no analogous way to propose a consistent estimate for s_n since one cannot use blocks of size n to estimate it. One way out is to utilize its regularly varying property as in equations (8) and (9). To this end, we choose positive integers n_1 and l_1 such that

$$\frac{l_1}{n_1} \sim \frac{l}{n} \text{ and } \frac{1}{l_1} + \frac{n_1 + l}{n} = O(n^{-\theta}) \text{ for some } \theta > 0. \tag{31}$$

For example, (31) holds if $l = \lfloor n^f \rfloor$, $n_1 = \lfloor n^g \rfloor$ and $l_1 = \lfloor n^{f+g-1} \rfloor$, where $0 < f < 1$, $0 < g < 1$ and $f + g > 1$. Further assume that $\ell(\cdot)$ is *strongly slowly varying* in the sense that $\lim_{k\to\infty} \ell(k)/\ell(k^\alpha) = 1$ for any $\alpha > 0$. It holds for functions like $\ell(k) = (\log\log k)^c$, $c \in \mathbb{R}$, while the slowly varying function $\ell(k) = \log k$ is not strongly slowly varying. Similar conditions were also used in Hall et al. (1998) and Nordman and Lahiri (2005). Note that (31) implies that

$$\lim_{n\to\infty} \frac{s_{l_1} s_n}{s_l s_{n_1}} = 1. \tag{32}$$

Then by Theorem 1 and condition (31), we have

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(S_l/s_l \le x) - \mathbb{P}\{S_n s_{l_1}/(s_{n_1} s_l) \le x\}| \to 0. \tag{33}$$

Hence, the distribution of $S_n s_{l_1}/(s_{n_1} s_l)$ can be approximated by that of S_l/s_l. Note that s_l appears in the denominator of both quantities, in practice we can simply use the sample quantiles of S_l/s_{l_1} to estimate those of S_n/s_{n_1}. In particular, let

$$\tilde{F}_l^\star(x) = \frac{1}{n-l+1}\sum_{i=l}^{n} \mathbf{1}_{B_{i,l}-l\bar{Y}_n \le x\tilde{s}_{l_1,i}},$$

where

$$\tilde{s}_{l_1,i}^2 = \frac{\tilde{Q}_{l,l_1,i}}{l-l_1+1}, \text{ with } \tilde{Q}_{l,l_1,i} = \sum_{j=l_1}^{l} |B_{j+i-l,l_1} - l_1\bar{Y}_n|^2. \tag{34}$$

Since $\lim_{n\to\infty} \tilde{s}_{l_1,i}^2/s_{l_1}^2 = 1$, using the argument in Theorem 2, we have

$$\sup_{x\in\mathbb{R}} |\tilde{F}_l^\star(xs_l/s_{l_1}) - \mathbb{P}(S_n/\tilde{s}_{n_1} \le xs_l/s_{l_1})| \to 0 \text{ in probability}, \tag{35}$$

where $\tilde{s}_{n_1}$ can be obtained by (23). Then confidence intervals for μ can be constructed based on sample quantiles of $\tilde{F}_l^\star(\cdot)$, and we do not need to estimate the nuisance parameter s_l/s_{l_1} in (35). In particular, we can construct the two-sided $(1-\alpha)$-th $(0<\alpha<1)$ and the upper one-sided $(1-\alpha)$-th confidence intervals for μ as $[\bar{Y}_n - \tilde{q}^\star_{1-\alpha/2}\tilde{s}_{n_1}/n,\ \bar{Y}_n - \tilde{q}^\star_{\alpha/2}\tilde{s}_{n_1}/n]$ and $[\bar{Y}_n - \tilde{q}^\star_{1-\alpha}\tilde{s}_{n_1}/n,\ \infty)$ respectively, where $\tilde{q}^\star_\alpha$ is the α-th sample quantile of $\tilde{F}_l^\star(\cdot)$.

3 Simulation Study

Consider a stationary process $Y_i = K(X_i)$, where X_i is a linear process defined in (1) with $a_k = (1+k)^{-\beta}$, $k \ge 0$, and ε_i, $i\in\mathbb{Z}$, are iid innovations. We shall here investigate the finite-sample performance of the block sampling method described in Section 2.5 by considering different choices of the transform $K(\cdot)$, the beta index β, the sample size n and innovation distributions. In particular, we consider the following four processes:

(i) $K(x) = x$, and ϵ_i, $i\in\mathbb{Z}$, are iid $N(0,1)$;

(ii) $K(x) = \mathbf{1}_{\{x \le 1\}}$, and ϵ_i, $i \in \mathbb{Z}$, are iid t_7;

(iii) $K(x) = \mathbf{1}_{\{x \le 0\}}$, and ϵ_i, $i \in \mathbb{Z}$, are iid t_7;

(iv) $K(x) = x^2$, and ϵ_i, $i \in \mathbb{Z}$, are iid Rademacher.

For cases (i) and (ii), the power rank $p = 1$, while for (iii) and (iv), the power rank $p = 2$. If $p = 1$, we let $\beta = 0.75$ and $\beta = 2$, which correspond to long- and short-range dependent processes, respectively. For $p = 2$, we consider three cases: $\beta \in \{0.6, 0.8, 2\}$. The first two are situations of long-range dependence but have different limiting distributions as indicated in Theorems 1 and 2. We use block sizes $l = \lfloor cn^{0.5} \rfloor$, $c \in \{0.5, 1, 2\}$, $n_1 = \lfloor n^{0.9} \rfloor$ and $l_1 = \lfloor ln_1/n \rfloor$. Let $n \in \{100, 500, 1000\}$. The empirical coverage probabilities of lower and upper one-sided 90% confidence intervals are computed based on $5,000$ realizations and they are summarized in Table 1 as pairs in parentheses. We observe the following phenomena. First, the accuracy of the coverage probabilities generally improves as we increase n, or decrease the strength of dependence (increasing the beta index β). Second, the nonlinearity worsens the accuracy, noting that the processes in (ii)–(iv) are nonlinear while the one in (i) is linear.

Acknowledgments

We thank the editor, an associate editor, and three anonymous referees for their helpful comments and suggestions. Hwai-Chung Ho is grateful for the support of the National Science Council Grant 100-2118-M-001-007-MY2. Martin Wendler was supported by the German Academic Foundation (Studienstiftung des deutschen Volkes). Wei Biao Wu was supported by the NSF Grant DMS 0906073 and DMS 1106970.

β	n	Models (i) and (iii) $c = 0.5$	$c = 1$	$c = 2$	Models (ii) and (iv) $c = 0.5$	$c = 1$	$c = 2$
			Model (i)			*Model (ii)*	
0.75	100	(90.8, 91.8)	(87.9, 87.2)	(83.5, 84.9)	(100.0, 96.8)	(99.9, 94.9)	(95.8, 90.5)
	500	(93.5, 94.2)	(91.1, 91.6)	(88.0, 89.7)	(100.0, 98.3)	(99.8, 95.7)	(97.0, 92.1)
	1000	(94.5, 94.2)	(92.6, 91.9)	(90.5, 89.3)	(100.0, 97.8)	(99.8, 95.9)	(96.9, 92.7)
2	100	(93.6, 92.5)	(89.4, 89.3)	(85.6, 85.3)	(100.0, 89.9)	(99.1, 86.3)	(89.5, 84.5)
	500	(93.6, 93.0)	(91.6, 92.3)	(89.8, 90.1)	(100.0, 87.3)	(95.1, 88.8)	(91.8, 87.3)
	1000	(93.1, 93.4)	(91.5, 92.6)	(90.6, 90.3)	(98.4, 88.0)	(95.1, 88.2)	(92.0, 87.7)
			Model (iii)			*Model (iv)*	
0.6	100	(99.3, 98.2)	(99.2, 98.0)	(95.6, 94.7)	(99.2, 77.9)	(98.5, 71.4)	(97.9, 65.4)
	500	(100.0, 99.9)	(99.7, 99.7)	(97.3, 97.8)	(99.1, 88.6)	(98.4, 86.2)	(98.3, 77.0)
	1000	(100.0,100.0)	(99.7, 99.7)	(98.3, 97.3)	(99.1, 91.9)	(98.9, 86.9)	(98.1, 82.1)
0.8	100	(100.0,100.0)	(97.9, 97.6)	(91.5, 90.6)	(97.4, 88.1)	(95.6, 85.0)	(93.3, 78.5)
	500	(99.7, 99.7)	(98.0, 97.3)	(93.2, 94.0)	(96.6, 95.2)	(95.6, 93.4)	(93.1, 88.3)
	1000	(99.4, 99.5)	(97.5, 96.9)	(95.1, 93.4)	(95.6, 96.8)	(94.6, 93.9)	(93.2, 91.5)
2	100	(94.8, 94.7)	(91.8, 89.8)	(86.3, 86.6)	(86.6, 90.7)	(86.7, 89.1)	(84.0, 86.0)
	500	(93.8, 93.6)	(92.1, 91.7)	(89.3, 89.5)	(85.7, 94.7)	(86.7, 93.6)	(86.1, 90.9)
	1000	(93.3, 93.9)	(91.5, 91.3)	(90.6, 89.4)	(87.3, 94.8)	(86.5, 93.6)	(86.2, 90.9)

Table 1: Empirical coverage probabilities of lower and upper (paired in parentheses) one-sided 90% confidence intervals for processes (i)–(iv) with different combinations of beta index β, sample size n and block size $l = \lfloor cn^{0.5} \rfloor$.

4 Appendix

Recall that $\mathcal{F}_i^j = (\varepsilon_i, \varepsilon_{i+1}, \ldots, \varepsilon_j)$, $i \leq j$, $\mathcal{F}_i^\infty = (\varepsilon_i, \varepsilon_{i+1}, \ldots)$ and $\mathcal{F}_{-\infty}^j = (\ldots, \varepsilon_{j-1}, \varepsilon_j)$. In dealing with nonlinear functionals of linear processes, we will use the powerful tool of Volterra expansion (Ho and Hsing, 1997 and Wu, 2006). Define

$$L_{n,p} = K(X_n) - \sum_{r=0}^{p} \kappa_r U_{n,r}, \tag{36}$$

where we recall $\kappa_r = K_\infty^{(r)}(0)$, and $U_{n,r}$ is the Volterra process

$$U_{n,r} = \sum_{0 \leq j_1 < \ldots < j_r < \infty} \prod_{s=1}^{r} a_{j_s} \varepsilon_{n-j_s}. \tag{37}$$

We can view $L_{n,p}$ as the remainder of the p-th order Volterra expansion of $K(X_n)$. Note that $\kappa_r = 0$ if $1 \leq r < p$. In the special case of Gaussian processes, $L_{n,p}$ is closely related to the Hermite expansion. In Lemma 2 we compute the predictive dependence measures for the Volterra process $U_{n,r}$ and for $Y_n = K(X_n)$.

Lemma 2. *Let $\nu \geq 2$, $r \geq 1$ and assume $\varepsilon_i \in \mathcal{L}^\nu$. Let $A_n = \sum_{j=n}^{\infty} a_j^2$. Then*

$$\|\mathcal{P}_0 U_{n,r}\|_\nu^2 = O(a_n^2 A_n^{r-1}). \tag{38}$$

Proof. Let $D_i, i \in \mathbb{Z}$, be a sequence of martingale differences with $D_i \in \mathcal{L}^\nu$. By the Burkholder and the Minkowski inequalities, there exists a constant C_ν which only depends on ν such that, for all $m \geq 1$, we have

$$\|D_1 + \ldots + D_m\|_\nu^2 \leq C_\nu(\|D_1\|_\nu^2 + \ldots + \|D_m\|_\nu^2). \tag{39}$$

We now apply the induction argument and show that, for all $r \geq 1$,

$$\|\mathbb{E}(U_{n,r}|\mathcal{F}_0)\|_\nu^2 = O(A_n^r). \tag{40}$$

Clearly (40) holds with $r = 1$. For any $i_{r+1} < \ldots < i_2$, we have $\mathbb{E}(\prod_{k=2}^{r+1} a_{n-i_k}\varepsilon_{i_k} \mid \mathcal{F}_{i_1}) = 0$ if $i_2 > i_1$; and $\mathbb{E}(\prod_{k=2}^{r+1} a_{n-i_k}\varepsilon_{i_k} \mid \mathcal{F}_{i_1}) = \prod_{k=2}^{r+1} a_{n-i_k}\varepsilon_{i_k}$ if $i_2 \leq i_1$. Hence,

$$\begin{aligned}
\mathbb{E}(U_{n,r}|\mathcal{F}_{i_1}) &= \mathbb{E}\left[\sum_{-\infty < i_{r+1} < \ldots < i_2 \leq n} \prod_{k=2}^{r+1} a_{n-i_k}\varepsilon_{i_k} \middle| \mathcal{F}_{i_1}\right] \\
&= \sum_{i_{r+1} < \ldots < i_2 \leq i_1} \prod_{k=2}^{r+1} a_{n-i_k}\varepsilon_{i_k}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\sum_{i_1=-\infty}^{-1} a_{n-i_1}^2 \|\mathbb{E}(U_{n,r}|\mathcal{F}_{i_1})\|_\nu^2 &= \sum_{i_1=-\infty}^{-1} a_{n-i_1}^2 \left\| \sum_{i_{r+1} < \ldots < i_2 < i_1+1} \prod_{k=2}^{r+1} a_{n-i_k}\varepsilon_{i_k} \right\|_\nu^2 \\
&= \sum_{i_1=-\infty}^{0} a_{n-i_1}^2 \left\| \sum_{i_{r+1} < \ldots < i_2 < i_1} \prod_{k=2}^{r+1} a_{n-i_k}\varepsilon_{i_k} \right\|_\nu^2.
\end{aligned}$$

By (39),

$$\|\mathbb{E}(U_{n,r+1}|\mathcal{F}_0)\|_\nu^2 = \left\| \sum_{i_1=-\infty}^{0} a_{n-i_1}\varepsilon_{i_1} \sum_{i_{r+1} < \ldots < i_2 < i_1} \prod_{k=2}^{r+1} a_{n-i_k}\varepsilon_{i_k} \right\|_\nu^2$$

$$\leq C_\nu \sum_{i_1=-\infty}^{0} a_{n-i_1}^2 \|\varepsilon_{i_1}\|_\nu^2 \left\| \sum_{i_{r+1}<\ldots<i_2<i_1} \prod_{k=2}^{r+1} a_{n-i_k}\varepsilon_{i_k} \right\|_\nu^2$$

$$= C_\nu \|\varepsilon_0\|_\nu^2 \sum_{i_1=-\infty}^{-1} a_{n-i_1}^2 \|\mathbb{E}(U_{n,r}|\mathcal{F}_{i_1})\|_\nu^2.$$

By stationarity, $\|\mathbb{E}(U_{n,r}|\mathcal{F}_{i_1})\|_\nu^2 = \|\mathbb{E}(U_{n-i_1,r}|\mathcal{F}_0)\|_\nu^2$. Then, by the induction hypothesis,

$$\|\mathbb{E}(U_{n,r+1}|\mathcal{F}_0)\|_\nu^2 \leq C_\nu \|\varepsilon_0\|_\nu^2 \sum_{i_1=-\infty}^{0} a_{n-i_1}^2 O(A_{n-i_1}^r) = O(A_n^{r+1}).$$

Hence (40) holds for all $r \geq 1$. By independence, $\mathcal{P}_0 U_{n,r} = a_n \varepsilon_0 \mathbb{E}(U_{n,r-1}|\mathcal{F}_{-1})$, which implies (38) by (40). ♢

Lemma 3. *Assume $r \in \mathbb{N}$, $r(2\beta-1) < 1$, and $\varepsilon_i \in \mathcal{L}^\nu$, $\nu \geq 2$. Let $T_{n,r} = \sum_{i=1}^n U_{i,r}$.*

(i) Let $(c_i)_{i\in\mathbb{N}}$ be a real valued sequence. Then

$$\left\| \sum_{i=1}^n c_i U_{i,r} \right\|_\nu^2 = O\left(n^{1-r(2\beta-1)} \ell^{2r}(n) \sum_{i=1}^n c_i^2 \right).$$

(ii) Assume that $n \leq N$ and $\varphi \in (0, \beta - 1/2)$. Then

$$\frac{\|T_{n,r} - \mathbb{E}(T_{n,r}|\mathcal{F}_{-N}^\infty)\|_\nu}{n^{1-r(\beta-1/2)}\ell^r(n)} = O[(n/N)^\varphi].$$

(iii) If additionally the coefficients $a_0 \neq 0$, $a_j = c_j j^{-\beta}$, $j \geq 1$, where $1/2 < \beta < 1$ and $c_j = c + O(j^{-\phi})$ for some $\phi > 0$, then we have for some $\varphi_2 > 0$,

$$\frac{\|T_{n,r}\|^2}{n^{2-r(2\beta-1)} c^{2r} \|Z_{r,\beta}(1)\|^2 \|\varepsilon_1\|^{2r}} = 1 + O(n^{-\varphi_2}). \tag{41}$$

Proof. For (i), we use the following decomposition with the help of the projection operator

$$\sum_{i=1}^n c_i U_{i,r} = \sum_{j=1}^n \sum_{l=0}^\infty \sum_{i=1}^n \mathcal{P}_{-ln-j+i}(c_i U_{i,r}).$$

Note that both $\{\sum_{i=1}^n \mathcal{P}_{-ln-j+i}(c_i U_{i,r})\}_{l\in\mathbb{N}}$ and $\{\mathcal{P}_{-ln-j+i}(U_{1,r})\}_{i=1}^n$ form martingale differences, for any $j \in \{1, \ldots, n\}$, we have

$$\left\| \sum_{l=0}^\infty \sum_{i=1}^n \mathcal{P}_{-ln-j+i}(c_i U_{i,r}) \right\|_\nu^2 \leq C \sum_{l=0}^\infty \left\| \sum_{i=1}^n \mathcal{P}_{-ln-j+i}(c_i U_{i,r}) \right\|_\nu^2$$

$$\leq \quad C \sum_{l=0}^{\infty} \sum_{i=1}^{n} c_i^2 \left\| \mathcal{P}_0(U_{ln+j,r}) \right\|_{\nu}^2 .$$

Hence by Lemma 2, we have

$$\left\| \sum_{l=0}^{\infty} \sum_{i=1}^{n} \mathcal{P}_{-ln-j+i}(c_i U_{i,r}) \right\|_{\nu}^2 \leq C \sum_{i=1}^{n} c_i^2 \left(a_j^2 A_j^{r-1} + \sum_{l=1}^{\infty} a_{j+ln}^2 A_{j+ln}^{r-1} \right) = \sum_{i=1}^{n} c_i^2 O\left(j^{-2\beta-(r-1)(2\beta-1)} \ell^{2r}(j) \right).$$

Then by the triangle inequality we can get

$$\begin{aligned} \left\| \sum_{i=1}^{n} c_i U_{i,r} \right\|_{\nu}^2 &= \left\| \sum_{j=1}^{n} \sum_{l=0}^{\infty} \sum_{i=1}^{n} \mathcal{P}_{-ln-j+i}(c_i U_{i,r}) \right\|_{\nu}^2 \\ &\leq C \sum_{i=1}^{n} c_i^2 \Big(\sum_{j=1}^{n} j^{-\beta-(r-1)(\beta-\frac{1}{2})} \ell^r(j) \Big)^2 \\ &= O\Big(n^{1-r(2\beta-1)} \ell^{2r}(n) \sum_{i=1}^{n} c_i^2 \Big). \end{aligned}$$

For (ii), we define the future projection operator $\mathcal{Q}_j := \mathbb{E}(\cdot|\mathcal{F}_j^{\infty}) - \mathbb{E}(\cdot|\mathcal{F}_{j+1}^{\infty})$ and obtain

$$T_{n,r} - E(T_{n,r}|\mathcal{F}_{-N}^{\infty}) = \sum_{j=N+1}^{\infty} \mathcal{Q}_{-j}(T_{n,r}).$$

Note that $\mathcal{Q}_j(T_{n,r}) = \sum_{i=1}^{n} a_{i-j} \varepsilon_j \mathbb{E}(U_{i,r-1}|\mathcal{F}_{j+1}^{\infty})$ which forms a sequence of martingale differences for $j \in \mathbb{Z}$, we have

$$\begin{aligned} \left\| T_{n,r} - \mathbb{E}(T_{n,r}|\mathcal{F}_{-N}^{\infty}) \right\|_{\nu}^2 &\leq C \sum_{j=N}^{\infty} \left\| a_{i+j} \varepsilon_{-j} \mathbb{E}(U_{i,r-1}|\mathcal{F}_{-j+1}^{\infty}) \right\|_{\nu}^2 \\ &\leq C \sum_{j=N}^{\infty} \left\| a_{i+j} U_{i,r-1} \right\|_{\nu}^2 . \end{aligned}$$

Hence by using part (i) of this lemma, we have

$$\begin{aligned} \left\| T_{n,r} - \mathbb{E}(T_{n,r}|\mathcal{F}_{-N}^{\infty}) \right\|_{\nu}^2 &\leq C \sum_{j=N}^{\infty} \sum_{i=1}^{n} a_{i+j}^2 n^{1-(r-1)(2\beta-1)} \ell^{2(r-1)}(n) \\ &\leq C N^{-(2\beta-1)} \ell^2(N) n^{2-(r-1)(2\beta-1)} \ell^{2(r-1)}(n). \end{aligned}$$

Therefore by the the slow variation of $\ell(\cdot)$, we have for some $\varphi \in (0, \beta - 1/2)$,

$$\frac{\|T_{n,r} - \mathbb{E}(T_{n,r}|\mathcal{F}_{-N}^{\infty})\|_\nu}{n^{1-r(\beta-1/2)}\ell^r(n)} = O\left(\frac{n^{\beta-\frac{1}{2}}\ell(N)}{N^{\beta-\frac{1}{2}}\ell(n)}\right) = O((n/N)^{\varphi}).$$

We now prove (iii). Without loss of generality let the constant c in the stated condition be 1 and assume $\|\varepsilon_1\| = 1$. For $\beta \in (1/2,\, 1/2 + 1/(2r))$, define $a_{i,\beta} = i^{-\beta}$ if $i \geq 1$, $a_{i,\beta} = 1$ if $i = 0$ and $a_{i,\beta} = 0$ if $i < 0$. Let $\beta_k \in (1/2,\, 1/2 + 1/(2r))$, $1 \leq k \leq r$, and define

$$T_{n,\beta_1,\ldots,\beta_r} = \sum_{j_r<\ldots<j_1\leq n} \sum_{i=1}^{n} \prod_{k=1}^{r} a_{i-j_k,\beta_k}\varepsilon_{j_k}.$$

Using the approximations that, for $1/2 < \beta < 1$, $\sum_{i=1}^{n} i^{-\beta} = n^{1-\beta}/(1-\beta) + O(1)$ and $\sum_{i=n_1}^{n_2} i^{-\beta} = (n_2^{1-\beta} - n_1^{1-\beta})/(1-\beta) + O(n_2^{-\beta} + n_1^{-\beta})$ when $n_1, n_2 \geq 1$, by elementary but tedious calculations, we have for some $\varphi_3 > 0$ that

$$\frac{\|T_{n,\beta_1,\ldots,\beta_r}\|^2}{n^{2-\sum_{k=1}^{r}(2\beta_k-1)}\zeta_{\beta_1,\ldots,\beta_r}} = 1 + O(n^{-\varphi_3}), \tag{42}$$

where, recall that $\mathcal{S}_t = \{(u_1, \ldots, u_r) \in \mathbb{R}^r : -\infty < u_1 < \ldots < u_r < t\}$,

$$\zeta_{\beta_1,\ldots,\beta_r} = \int_{\mathcal{S}_1} \left[\int_0^1 \prod_{k=1}^{r} g_{\beta_k}(v - u_k)dv\right]^2 du_1 \ldots du_r.$$

Note that $\|Z_{r,\beta}(1)\|^2 = \zeta_{\beta,\ldots,\beta}$ if $1/2 < \beta < 1/2 + 1/(2r)$. We now show that (42) implies (41). To this end, for notational clarity, we only consider $r = 2$. The general case similarly follows. Let $\phi > 0$ be such that $\phi < 1/2 + 1/(2r) - \beta$, hence $\phi + \beta < 1/2 + 1/(2r)$. Writing

$$a_{i-j_1}a_{i-j_2} - a_{i-j_1,\beta}a_{i-j_2,\beta} = (a_{i-j_1} - a_{i-j_1,\beta})a_{i-j_2} + a_{i-j_1,\beta}(a_{i-j_2} - a_{i-j_2,\beta}). \tag{43}$$

By the stated condition, for $j_1 \geq 1$, $a_{j_1} - a_{j_1,\beta} = O(j_1^{-\beta-\phi})$. Hence $a_j - a_{j,\beta} = O(a_{j,\beta+\phi})$ for $j \in \mathbb{Z}$. Applying (42) to the case with $\beta_1 = \beta$ and $\beta_2 = \phi + \beta$, we have

$$\begin{aligned}\left\|\sum_{j_2<j_1\leq n} \sum_{i=1}^{n}(a_{i-j_1} - a_{i-j_1,\beta})a_{i-j_2}\varepsilon_{j_1}\varepsilon_{j_2}\right\|^2 &= \sum_{j_2<j_1\leq n}\left[\sum_{i=1}^{n}(a_{i-j_1} - a_{i-j_1,\beta})a_{i-j_2}\right]^2 \\ &= O(1)\left\|\sum_{j_2<j_1\leq n}\sum_{i=1}^{n} a_{i-j_1,\beta_2}a_{i-j_2,\beta_1}\varepsilon_{j_1}\varepsilon_{j_2}\right\|^2\end{aligned}$$

$$= O(n^{2-(2\beta_1-1)-(2\beta_2-1)}).$$

A similar relation can be obtained by replacing $(a_{i-j_1} - a_{i-j_1,\beta})a_{i-j_2}$ in the preceding equation by $a_{i-j_1,\beta}(a_{i-j_2} - a_{i-j_2,\beta})$. Hence, by (43), $\|T_{n,2} - T_{n,\beta,\beta}\|^2 = O(n^{2-(2\beta_1-1)-(2\beta_2-1)})$, which by (42) implies (41) since $\beta_1 = \beta$ and $\beta_2 = \phi + \beta$. ♢

Lemma 4. *Assume Condition 1 holds with $\nu \geq 2$ and K has power rank $p \geq 1$ with respect to the distribution of X_i such that $r(2\beta - 1) < 1$. Then we have: (i) $\|S_n\|_\nu = O(\sigma_{n,p})$; and (ii) there exists $\varphi_1, \varphi_2 > 0$ such that*

$$\frac{\|S_n - \mathbb{E}(S_n|\mathcal{F}_{-N}^{\infty})\|_\nu}{\sigma_{n,p}} = O(n^{-\varphi_1}) + O[(n/N)^{\varphi_2}].$$

Proof. Recall Lemma 3 for $T_{n,p}$. Observe that $Y_n = L_{n,p} + \kappa_p U_{n,p}$ and $S_n = S_n(L^{(p)}) + \kappa_p T_{n,p}$. Since

$$\begin{aligned}\|S_n - \mathbb{E}(S_n|\mathcal{F}_{-N}^{\infty})\|_\nu &\leq \|S_n(L^{(p)}) - \mathbb{E}(S_n(L^{(p)})|\mathcal{F}_{-N}^{\infty})\|_\nu + \|\kappa_p T_{n,p} - \mathbb{E}(\kappa_p T_{n,p}|\mathcal{F}_{-N}^{\infty})\|_\nu \\ &\leq 2\|S_n(L^{(p)})\|_\nu + |\kappa_p| \|T_{n,p} - \mathbb{E}(T_{n,p}|\mathcal{F}_{-N}^{\infty})\|_\nu,\end{aligned}$$

by Lemma 3, it suffices to show that

$$\frac{\|S_n(L^{(p)})\|_\nu}{\sigma_{n,p}} = O(n^{-\varphi_1}). \tag{44}$$

By the argument of Theorem 5 in Wu (2006), Condition 1 with $\nu \geq 2$ implies that

$$\|\mathcal{P}_0 L_{n,p}\|_\nu^2 = a_n^2 O(a_n^2 + A_{n+1}(4) + A_{n+1}^p), \tag{45}$$

where $A_n(4) = \sum_{t=n}^{\infty} a_t^4$ and $A_n = \sum_{t=n}^{\infty} a_t^2$. Let $\theta_i = |a_i|[|a_i| + A_{i+1}^{1/2}(4) + A_{i+1}^{p/2}]$ if $i \geq 0$ and $\theta_i = 0$ if $i < 0$. Theorem 5 and Lemma 2 in Wu (2006) consider only the case $\nu = 2$, but the case $\nu > 2$ can be proved analogously using the Burkholder inequality. Write $\Theta_n = \sum_{k=0}^{n} \theta_k$ and $\Xi_{n,p} = n\Theta_n^2 + \sum_{i=1}^{\infty}(\Theta_{n+i} - \Theta_i)^2$. By (39), since $\mathcal{P}_k S_n(L^{(p)})$, $k = -\infty, \ldots, n-1, n$, for martingale differences, we have

$$\|S_n(L^{(p)})\|_\nu^2 \leq C_\nu \sum_{k=-\infty}^{n} \|\mathcal{P}_k S_n(L^{(2)})\|_\nu^2$$

$$
\begin{aligned}
&\leq C_\nu \sum_{k=-\infty}^{n} \left(\sum_{i=1}^{n} \theta_{i-k} \right)^2 \\
&= O(\Xi_{n,p}).
\end{aligned}
$$

By (i), (ii) and (iii) of Corollary 1 in Wu (2006), we have $\Xi_{n,p}^{1/2}/\sigma_{n,p} = O(n^{1/2-\beta}\ell(n))$ if $(p+1)(2\beta-1) < 1$ and $\Xi_{n,p}^{1/2}/\sigma_{n,p} = O(n^{p(\beta-1/2)-1/2}\ell_0(n))$ if $(p+1)(2\beta-1) \geq 1$. Here ℓ_0 is a slowly varying function. Note that both $1/2-\beta$ and $p(\beta-1/2)-1/2$ are negative, (44) follows. ♢

References

AVRAM, F. AND TAQQU, M. S. (1987). Noncentral limit theorems and Appell polynomials. *Ann. Probab.*, **15**, 767–775.

BERTAIL, P., POLITIS, D. N. AND ROMANO, J. P. (1999). On subsampling estimators with unknown rate of convergence. *J. Amer. Statist. Assoc.*, **94**, 569–579.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BINGHAM, N. H., GOLDIE, C. M. AND TEUGELS, J. M. (1987). *Regular Variation.* Cambridge University Press, Cambridge, UK.

CHOW, Y. S. AND TEICHER, H. (1997). *Probability Theory: Independence, Interchangeability, Martingales.* Third Edition. Springer, New York.

DAVYDOV, Y. A. (1970). The invariance principle for stationary processes. *Theory Probab. Appl.*, **15**, 487–498.

DITTMANN, I. AND GRANGER, C. W. J. (2002). Properties of nonlinear transformations of fractionally integrated processes. *J. Econometrics*, **110**, 113–133.

DOBRUSHIN, R. L. AND MAJOR, P. (1979). Non-central limit theorems for nonlinear functionals of Gaussian fields. *Z. Wahrsch. Verw. Gebiete*, **50**, 27–52.

GRANGER, C. W. J. AND JOYEUX, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Series Anal.*, **1**, 15–29.

HALL, P. AND JING, B.-Y. (1996). On sample reuse methods for dependent data. *J. Roy. Statist. Soc. Ser. B*, **58**, 727–737.

HALL, P., JING, B.-Y. AND LAHIRI, S. N. (1998). On the sampling window method for long-range dependent data. *Statist. Sinica*, **8**, 1189–1204.

HO, H.-C. AND HSING, T. (1997). Limit theorems for functionals of moving averages. *Ann. Probab.*, **25**, 1636–1669.

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

KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, **17**, 1217–1241.

LAHIRI, S. N. (1993). On the moving block bootstrap under long range dependence. *Statist. Probab. Lett.*, **18**, 405–413.

NORDMAN, D. J. AND LAHIRI, S. N. (2005). Validity of the sampling window method for long-range dependent linear processes. *Econometric Theory*, **21**, 1087–1111.

POLITIS, D. N., ROMANO, J. P. AND WOLF, M. (1999). *Subsampling.* Springer, New York.

SURGAILIS, D. (1982). Domains of attraction of self-similar multiple integrals. *Litovsk. Mat. Sb.*, **22**, 185–201.

TAQQU, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete*, **31**, 53–83.

WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci.* **102**, 14150–14154.

WU, W. B. (2006). Unit root testing for functionals of linear processes. *Econometric Theory*, **22**, 1–14