Bernoulli **23**(4A), 2017, 2558–2586
DOI: 10.3150/16-BEJ819

Randomized pivots for means of short and long memory linear processes

MIKLÓS CSÖRGŐ*, MASOUD M. NASARI** and
MOHAMEDOU OULD-HAYE†

*School of Mathematics and Statistics, Carleton University, Ottawa K1S 5B6, ON, Canada.
E-mail: *mcsorgo@math.carleton.ca; **mmnasari@math.carleton.ca; †ouldhaye@math.carleton.ca*

In this paper, we introduce randomized pivots for the means of short and long memory linear processes. We show that, under the same conditions, these pivots converge in distribution to the same limit as that of their classical non-randomized counterparts. We also present numerical results that indicate that these randomized pivots significantly outperform their classical counterparts and as a result they lead to a more accurate inference about the population mean.

Keywords: central limit theorem; randomized pivots; short and long memory time-series

1. Introduction and background

Recently, Csörgő and Nasari [6] investigated the problem of establishing central limit theorems (CLTs) with improved rates for randomized versions of the Student t-statistic based on i.i.d. observations $X_1, X_2 \dots$. The resulting improvements yield confidence intervals for the population mean μ with a smaller magnitude of error than that of the classical CLT for the Student t-pivot based on an i.i.d. sample of size n, $n \geq 1$. The improvements in hand result from incorporating functionals of multinomially distributed random variables as coefficients for, and independent of, the data. More precisely, Csörgő and Nasari [6] introduced and, via conditioning on the random weights, studied the asymptotic distribution of the randomized pivot for the population mean $\mu := EX_1$, that is defined as follows

$$\frac{\sum_{i=1}^{n} |w_i^{(n)}/m_n - 1/n|(X_i - \mu)}{S_n\sqrt{(w_i^{(n)}/m_n - 1/n)^2}}, \tag{1.1}$$

and can be computed via generating, independently from the data, a realization of the multinomial random weights $(w_1^{(n)}, \dots, w_n^{(n)})$ with, $w_i^{(n)} \geq 0$, $\sum_{i=1}^{n} w_i^{(n)} = m_n$ and associated probability vector $(1/n, \dots, 1/n)$; here S_n^2 is the sample variance.

In Csörgő and Nasari [6], it is shown that, on assuming $E|X_1|^3 < \infty$, the magnitude of the error generated by approximating the sampling distributions of these normalized/Studentized randomized partial sums of i.i.d. observables, as in (1.1), by the standard normal distribution function $\Phi(\cdot)$ can be of order $O(1/n)$. The latter rate is achieved when one takes $m_n = n$, and is to be compared to that of the $O(1/\sqrt{n})$ error rate of the Student t-statistic under the same moment condition.

1350-7265

The present work is an extension of the results in the aforementioned paper Csörgő and Nasari [6] to short and long memory linear processes via creating randomized direct pivots for the mean $\mu = EX_1$ of short and long memory linear processes *à la* (1.1). Adaptation of the randomized version of the Student t-statistic as in Csörgő and Nasari [6]

$$\frac{\sum_{i=1}^{n}(w_i^{(n)}/n - 1/n)X_i}{S_n\sqrt{\sum_{i=1}^{n}(w_i^{(n)}/n - 1/n)^2}} \tag{1.2}$$

to the same context will also be explored (cf. Section 5).

Just like in Csörgő and Nasari [6], in this paper the method of conditioning on the random weights $w_i^{(n)}$s is used for constructing randomized pivots. Viewing the randomized sums of linear processes as weighted sums of the original data, here we derive the asymptotic normality of properly normalized randomized sums of short and long memory linear processes. As will be seen, our conditional CLTs also imply unconditional CLTs in terms of the joint distribution of the observables and the random weights.

The material in this paper is organized as follows. In Section 2 the randomized pivots are introduced and conditional and unconditional CLTs are presented for them. In Section 3, asymptotic confidence intervals of size $1-\alpha$, $0 < \alpha < 1$, are constructed for the population mean $\mu = EX_1$. Also in Section 3, confidence bounds are constructed for some functionals of linear processes. The results in this section are directly applicable to constructing confidence bounds for some functionals of long memory linear processes whose limiting distribution may not necessarily be normal. Section 4 is devoted to presenting our simulations and numerical studies. In Section 5, we study the problem of bootstrapping linear processes and provide a comparison between our results and those obtained by using the bootstrap. The proofs are given in Section 6.

2. CLT for randomized pivots of the population mean

Throughout this section, we let $\{X_i; i \geq 1\}$ be a linear process that, for each $i \geq 1$, is defined by

$$X_i = \mu + \sum_{k=0}^{\infty} a_k \zeta_{i-k} = \mu + \sum_{k=-\infty}^{i} a_{i-k}\zeta_k, \tag{2.1}$$

where μ is a real number, $\{a_k; k \in \mathbb{Z}\}$ is a sequence of real numbers such that $\sum_{k=0}^{\infty} a_k^2 < \infty$ and $\{\zeta_k; k \in \mathbb{Z}\}$ are i.i.d. innovations with $E\zeta_k = 0$ and $0 < \sigma_\zeta^2 := \mathrm{Var}(\zeta_k) < \infty$. Consequently, we have $EX_i = \mu$. Moreover, we assume throughout that the X_is, are non-degenerate and have a finite variance $\gamma_0 := \mathrm{Var}(X_i)$, $i \geq 1$. We note in passing that for some of the results in this paper the existence and finiteness of some higher moments of the data will also be assumed (cf. Theorems 2.2 and 5.2).

For throughout use, we let

$$\gamma_h := \mathrm{Cov}(X_s, X_{s+h}) = E(X_s - \mu)(X_{s+h} - \mu), \qquad h \geq 0, s \geq 1, \tag{2.2}$$

be the autocovariance function of the stationary linear process $\{X_i, i \geq 1\}$ as in (2.1). Moreover, based on the stationary sample $X_1, \ldots, X_n$, $n \geq 1$, on the linear process $\{X_i, i \geq 1\}$, for throughout use we define

$$\begin{aligned} \bar{X}_n &:= \sum_{i=1}^{n} X_i/n, \\ \bar{\gamma}_i &:= \sum_{j=1}^{n-i} (X_j - \bar{X}_n)(X_{j+i} - \bar{X}_n)/n, \qquad 0 \leq i \leq n-1, \end{aligned} \tag{2.3}$$

respectively the sample mean and sample autocovariance.

For throughout use in this paper, for the linear process X_i, the parameter d, $0 \leq d < 1/2$, such that $\lim_{n\to\infty} n^{-2d} \operatorname{Var}(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i)$ is finite, is referred to as the memory parameter, where $d = 0$ corresponds to the case when the linear process is of short memory, that is, $\sum_{k=1}^{\infty} |a_k| < \infty$. While $0 < d < 1/2$ refers to the case when the linear process is of long memory, that is, $\sum_{k=1}^{\infty} |a_k| = \infty$.

We now define the following two randomly weighted versions of the partial sum $\sum_{i=1}^{n} (X_i - \mu)$, where $\{X_i, i \geq 1\}$ is a sequence of linear processes defined in (2.1),

$$\sum_{i=1}^{n} \left(\frac{w_i^{(n)}}{n} - \frac{1}{n} \right) (X_i - \mu), \tag{2.4}$$

$$\sum_{i=1}^{n} \left| \frac{w_i^{(n)}}{n} - \frac{1}{n} \right| (X_i - \mu), \tag{2.5}$$

where the random weights in the triangular array $\{w_1^{(n)}, \ldots, w_n^{(n)}\}_{n=1}^{\infty}$ have a multinomial distribution of size $n = \sum_{i=1}^{n} w_i^{(n)}$ with respective probabilities $1/n$, that is,

$$\left(w_1^{(n)}, \ldots, w_n^{(n)}\right) \stackrel{d}{=} \text{multinomial}\left(n; \frac{1}{n}, \ldots, \frac{1}{n}\right),$$

are independent from the stationary sample $\{X_1, \ldots, X_n\}$, $n \geq 1$, on the process $\{X_i, i \geq 1\}$ as in (2.1).

The just introduced randomized sums in (2.4) and (2.5), which are randomized versions of $\sum_{i=1}^{n} (X_i - \mu)$, can be computed via generating a realization of the multinomial random weights $(w_1^{(n)}, \ldots, w_n^{(n)})$ with $\sum_{i=1}^{n} w_i^{(n)} = n$ with associated probability vector $(1/n, \ldots, 1/n)$. In this context, one way of generating the random weights $w_i^{(n)}$s is to resample from the set of indices $\{1, \ldots, n\}$ of the stationary sample $X_1, \ldots, X_n$ in hand, $n \geq 1$, with replacement n times with respective probabilities $1/n$ so that, for each $1 \leq i \leq n$, $w_i^{(n)}$ is the number of times the index i of X_i is chosen in this resampling process.

Remark 2.1. In view of $\sum_{i=1}^n w_i^{(n)} = n$, one can readily see that for the randomized sum defined in (2.4), we have

$$\sum_{i=1}^{n}\left(\frac{w_i^{(n)}}{n}-\frac{1}{n}\right)(X_i-\mu) = \sum_{i=1}^{n}\left(\frac{w_i^{(n)}}{n}-\frac{1}{n}\right)X_i$$
$$=: \bar{X}_n^* - \bar{X}_n,$$

that is, $\bar{X}_n^* - \bar{X}_n$ forgets about what the value of the population mean $\mu = EX_1$ might be. On the other hand, the randomization used in the sum (2.5) preserves $\mu = EX_1$.

In addition to preserving μ, the randomized sum (2.5) tends to preserve the covariance structure of the data as well, a property that the sum (2.4) fails to maintain (cf. Remark 5.1).

Properly normalized, (2.5) provides a natural direct pivot for the population mean $\mu = EX_1$ (cf. the definition (2.6)).

For throughout use, we introduce the following notations.

Notations. Let $(\Omega_X, \mathfrak{F}_X, P_X)$ denote the probability space of the random variables $X, X_1, \ldots,$ and $(\Omega_w, \mathfrak{F}_w, P_w)$ be the probability space on which $(w_1^{(1)}, (w_1^{(2)}, w_2^{(2)}), \ldots, (w_1^{(n)}, \ldots, w_n^{(n)}), \ldots)$ are defined. In view of the independence of these two sets of random variables, jointly they live on the direct product probability space $(\Omega_X \times \Omega_w, \mathfrak{F}_X \otimes \mathfrak{F}_w, P_{X,w} = P_X.P_w)$. For each $n \geq 1$, we also let $P_{\cdot|w}(\cdot)$ stand for the conditional probability given $\mathfrak{F}_w^{(n)} := \sigma(w_1^{(n)}, \ldots, w_n^{(n)})$ with corresponding conditional expected value and variance, $E_{\cdot|w}(\cdot)$ and $\mathrm{Var}_{\cdot|w}(\cdot)$, respectively.

In a similar fashion to the randomized pivots in the i.i.d. case as in (1.1), in this context we define the randomized t-type statistics, based on the sample $X_1, \ldots, X_n$, $n \geq 1$, of linear processes defined by (2.1), as follows

$$G_n := \frac{\sum_{i=1}^n |w_i^{(n)}/n - 1/n|(X_i-\mu)}{\sqrt{D_n}}, \tag{2.6}$$

where γ_h, $0 \leq h \leq n-1$, are defined in (2.2) and

$$D_n := \gamma_0 \sum_{j=1}^{n}\left(\frac{w_i^{(n)}}{n}-\frac{1}{n}\right)^2 + 2\sum_{h=1}^{n-1}\gamma_h \sum_{j=1}^{n-h}\left|\frac{w_j^{(n)}}{n}-\frac{1}{n}\right|\left|\frac{w_{j+h}^{(n)}}{n}-\frac{1}{n}\right|. \tag{2.7}$$

We note that

$$D_n = \mathrm{Var}_{X|w}\left(\sum_{i=1}^{n}\left|\frac{w_i^{(n)}}{n}-\frac{1}{n}\right|(X_i-\mu)\right).$$

Noting that the normalizing sequence in G_n depends on the parameters γ_h, $0 \leq h \leq n-1$, we now define Studentized versions of it.

The following Studentized statistic, in (2.8), is defined for all short memory linear processes as in (2.1) as well as, for long memory linear processes as defined in (2.1) with $a_k \sim ck^{d-1}$, for some $c>0$, as $k\to\infty$, where $0<d<1/2$. Thus, the Studentized version of G_n as in (2.6) is defined as follows

$$G_n^{\text{stu}}(d) := \frac{\sum_{i=1}^{n} |w_i^{(n)}/n - 1/n|(X_i-\mu)}{\sqrt{D_{n,q,d}}}, \tag{2.8}$$

where $q\to\infty$ in such a way that as $n\to\infty$, $q=O(n^{1/2})$, $\bar{\gamma}_h$ is defined in (2.4), $0\le d<1/2$ and

$$D_{n,q,d} := \left(\frac{q}{n}\right)^{-2d} \bar{\gamma}_0 \sum_{j=1}^{n}\left(\frac{w_i^{(n)}}{n}-\frac{1}{n}\right)^2 + 2\sum_{h=1}^{q}\bar{\gamma}_h \sum_{j=1}^{q-h}\left|\frac{w_j^{(n)}}{n^{1-2d}}-\frac{1}{n^{1-2d}}\right|\left|\frac{w_{j+h}^{(n)}}{q^{1+2d}}-\frac{1}{q^{1+2d}}\right|.$$

We note in passing that in the case of long memory linear processes, when an estimator $\hat{d}$ is used to replace the memory parameter d (cf. Section 4), then the notation $G_n^{\text{stu}}(\hat{d})$ stands for the version of $G_n^{\text{stu}}(d)$, as in (2.8), in which d is replaced by $\hat{d}$. Also, in the case of having a short memory linear process, that is, when $d=0$, the notation $G_n^{\text{stu}}(0)$ stands for the version of $G_n^{\text{stu}}(d)$ in which d is replaced by 0.

It can also be readily seen that, in the case of having a long memory linear process, after estimating d by a proper estimator $\hat{d}$, then $G_n^{\text{stu}}(\hat{d})$, apart from μ that is to be estimated, is computable based on the data $X_1,\dots,X_n$ and the generated multinomial weights $(w_1^{(n)},\dots,w_n^{(n)})$. The same is also true when dealing with short memory linear processes, that is, when $d=0$. In other words, in the case of short memory linear processes, apart from the population mean μ, which is to be estimated, the other elements of the pivot $G_n^{\text{stu}}(0)$ are computable based on the data and the generated multinomial weights.

The following two theorems, namely Theorems 2.1 and 2.2, establish conditional (given the weights) and unconditional CLTs for G_n and $G_n^{\text{stu}}(d)$, respectively. These theorems are valid for classes of both short and long memory data.

We note that throughout this paper $\Phi(\cdot)$ stands for the standard normal distribution function.

Theorem 2.1. *Suppose that $\{X_i, i\ge 1\}$ is a stationary linear process as defined in* (2.1). *As $n\to\infty$, we have for all $t\in\mathbb{R}$,*

$$P_{X|w}(G_n\le t)\longrightarrow \Phi(t) \qquad \textit{in probability } P_w \tag{2.9}$$

and, consequently,

$$P_{X,w}(G_n\le t)\longrightarrow \Phi(t), \qquad t\in\mathbb{R}.$$

We note in passing that, for each $t\in\mathbb{R}$, the convergence (2.9) means that the sequence, in n, of the random variables $P_{X|w}(G_n\le t)$, with respect to P_w, converges in probability to $\Phi(t)$.

Remark 2.2. Theorem 2.1 allows having CLTs, conditionally on the weights, or in terms of the joint distribution of the data and the random weights, for randomized versions of partial sums of linear processes for which there are no CLTs with the standard deviation of the partial sum in hand in its normalizing sequence. Examples of such processes, which are usually the results of *overdifferencing*, are of the form $X_t = Y_t - Y_{t-1}$, where the Y_t are white noise, like, see, for example, the well-known non-invertible moving average MA(1) processes. Randomizing these processes results in randomly weighted partial sums of the original data whose variance, unlike the variance of the original partial sums, goes to infinity as the sample size $n \to \infty$. This phenomenon can be seen to be the result of incorporating the random weights, for then the sum $\sum_{i=1}^{n} |\frac{w_i^{(n)}}{n} - \frac{1}{n}| X_i$ no longer forms a telescoping series as the original non-randomized sum $\sum_{t=1}^{n} X_t = Y_n - Y_0$.

The following result, which is a companion of Theorem 2.1, establishes the asymptotic normality for the Studentized statistics $G_n^{\mathrm{stu}}(d)$.

Theorem 2.2. (A) *Assume that the stationary linear process* $\{X_i, i \geq 1\}$, *as defined in* (2.1), *is of short memory, that is,* $\sum_{k=0}^{\infty} |a_k| < \infty$, *and* $E\zeta_1^4 < \infty$. *Then, as* $n, q \to \infty$ *such that* $q = O(n^{1/2})$, *we have for all* $t \in \mathbb{R}$,

$$P_{X|w}\big(G_n^{\mathrm{stu}}(0) \leq t\big) \longrightarrow \Phi(t) \qquad \textit{in probability } P_w$$

and, consequently,

$$P_{X,w}\big(G_n^{\mathrm{stu}}(0) \leq t\big) \longrightarrow \Phi(t), \qquad t \in \mathbb{R}.$$

(B) *Let the linear process* $\{X_i, i \geq 1\}$, *as defined in* (2.1), *with* $\sum_{k=0}^{\infty} a_k^2 < \infty$, *be of long memory such that* $E\zeta_1^4 < \infty$ *and, as* $k \to \infty$, $a_k \sim ck^{d-1}$, *for some* $c > 0$, *where* $0 < d < 1/2$. *Then, as* $n, q \to \infty$ *such that* $q = O(n^{1/2})$, *we have, for all* $t \in \mathbb{R}$,

$$P_{X|w}\big(G_n^{\mathrm{stu}}(d) \leq t\big) \longrightarrow \Phi(t) \qquad \textit{in probability } P_w,$$

$$P_{X|w}\big(G_n^{\mathrm{stu}}(\hat{d}) \leq t\big) \longrightarrow \Phi(t) \qquad \textit{in probability } P_w,$$

and, consequently,

$$P_{X,w}\big(G_n^{\mathrm{stu}}(d) \leq t\big) \longrightarrow \Phi(t), \qquad t \in \mathbb{R},$$

$$P_{X,w}\big(G_n^{\mathrm{stu}}(\hat{d}) \leq t\big) \longrightarrow \Phi(t), \qquad t \in \mathbb{R},$$

where $\hat{d}$ *is an estimator of the memory parameter* d *such that* $\hat{d} - d = o_{P_X}(1/\log n)$.

3. Randomized confidence intervals for the population mean μ

In this section, we use $G_n^{\mathrm{stu}}(\hat{d})$ as a natural randomized pivot for the population mean μ in a nonparametric way. Based on it, we now spell out asymptotic randomized $1-\alpha$ size confidence intervals for the population mean μ. In what follows, $z_{1-\alpha}$ stands for the solution to $\Phi(z_{1-\alpha}) = 1 - \alpha$.

It is important to note that the randomized confidence intervals (one or two-sided) which we are about to present, henceforth, are valid in terms of the conditional distribution $P_{X|w}$, as well as in terms of the joint distribution $P_{X,w}$.

When the linear process in hand possesses the property of short memory, and if it satisfies the conditions of part (A) of Theorem 2.2, then the asymptotic two-sided $1-\alpha$ size randomized confidence interval for the population mean $\mu = E_X X_1$ has the following form:

$$\begin{aligned} &\frac{\sum_{i=1}^n |w_i^{(n)}/n - 1/n| X_i - z_{1-\alpha/2} D_{n,q,0}^{1/2}}{\sum_{j=1}^n |w_j^{(n)}/n - 1/n|} \\ &\quad \le \mu \le \frac{\sum_{i=1}^n |w_i^{(n)}/n - 1/n| X_i + z_{1-\alpha/2} D_{n,q,0}^{1/2}}{\sum_{j=1}^n |w_j^{(n)}/n - 1/n|}. \end{aligned} \tag{3.1}$$

An asymptotic $1-\alpha$ size randomized two-sided confidence interval for the population mean $\mu = E_X X_1$ of a long range dependent linear process, as defined in (2.1), when it satisfies the conditions in part (B) of Theorem 2.2 is constructed as follows:

$$\begin{aligned} &\frac{\sum_{i=1}^n |w_i^{(n)}/n - 1/n| X_i - z_{1-\alpha/2} D_{n,q,\hat{d}}^{1/2}}{\sum_{j=1}^n |w_j^{(n)}/n - 1/n|} \\ &\quad \le \mu \le \frac{\sum_{i=1}^n |w_i^{(n)}/n - 1/n| X_i + z_{1-\alpha/2} D_{n,q,\hat{d}}^{1/2}}{\sum_{j=1}^n |w_j^{(n)}/n - 1/n|}, \end{aligned} \tag{3.2}$$

where $D_{n,q,0}$ and $D_{n,q,\hat{d}}$ are as defined in (2.8).

3.1. Confidence bounds for the mean of some functionals of long memory linear processes

The following result, namely Corollary 3.1, is a consequence of Theorem 2.2 and Jensen's inequality. Corollary 3.1 gives randomized confidence bounds for $\mu_{\mathcal{G}} := E_X \mathcal{G}(X_i)$, for some measurable functions $\mathcal{G}$, that is, for the mean of certain subordinated functions of the long memory linear process in hand, and it reads as follows.

Corollary 3.1. *Let $\{X_i, i \ge 1\}$ be so that it satisfies the conditions in* (B) *of Theorem* 2.2. *Assume that $E_X|\mathcal{G}(X_i)| < \infty$. As $n, q \to \infty$ in such a way that $q = O(n^{1/2})$, the following holds.*

(A) *If $\mathcal{G}$ is increasing and convex, then, an asymptotic $1-\alpha$ level lower confidence bound for μ_G is*

$$\mu_{\mathcal{G}} \ge \mathcal{G}\left(\frac{\sum_{i=1}^n |w_i^{(n)}/n - 1/n| X_i - z_{1-\alpha} D_{n,q,\hat{d}}^{1/2}}{\sum_{i=1}^n |w_i^{(n)}/n - 1/n|}\right)$$

(B) *If $\mathcal{G}$ is decreasing and convex, then, a randomized asymptotic $1-\alpha$ level lower confidence bound for $\mu_{\mathcal{G}}$ is*

$$\mu_{\mathcal{G}} \geq \mathcal{G}\Bigg(\frac{\sum_{i=1}^{n} |w_i^{(n)}/n - 1/n| X_i + z_{1-\alpha} D_{n,q,\hat{d}}^{1/2}}{\sum_{i=1}^{n} |w_i^{(n)}/n - 1/n|}\Bigg).$$

Remark 3.1. Corollary 3.1 remains valid for functionals of short memory linear processes with $D_{n,q,0}$ replacing $D_{n,q,\hat{d}}$. It is also important to note that the conclusions of (A) and (B) of Corollary 3.1 hold true without making any assumptions about the variance of the subordinated function $\mathcal{G}$. In other words, Corollary 3.1 is valid even when $\mathrm{Var}(\mathcal{G})$ is not finite.

Remark 3.2. In reference to studying the mean of functions of stationary long memory Gaussian linear processes, Corollary 3.1 helps avoid dealing with the sampling distributions of processes of functions of stationary long memory Gaussian processes which are known to be relatively complicated, specially when they exhibit non-normal asymptotic distributions (cf. Taqqu [23] and Dobrushin and Major [8], for example). We note that a long memory Gaussian process $\{\eta_i; i \geq 1\}$, that is, $\mathrm{Cov}(\eta_1, \eta_{1+k}) = k^{-\alpha} L(k)$, where $0 < \alpha < 1$ and $L(\cdot)$ is a slowly varying function at infinity, can be viewed as a long memory linear process (cf. Csáki *et al.* [5], for example) that satisfies the conditions of part (B) of Theorem 2.2. Therefore, Corollary 3.1 is directly applicable to constructing randomized confidence bounds for means of subordinated functions of long memory Gaussian processes $\{\eta_i; i \geq 1\}$ without making assumptions concerning their variance, or referring to their Hermit expansions. It is also worth noting that the technique used to derive our results for constructing one sided confidence intervals is not limited to the randomized data as in our context. In fact, this method can be used for any subordinated long or short memory Gaussian processes whose subordinated function has a finite mean.

4. Simulation results

In this section, we examine numerically the performance of $G_n^{\mathrm{stu}}(d)$ and $G_n^{\mathrm{stu}}(\hat{d})$, in view of the CLTs in Theorem 2.2, versus those of their classical counterparts $T_n^{\mathrm{stu}}(d)$ and $T_n^{\mathrm{stu}}(\hat{d})$, defined as

$$T_n^{\mathrm{stu}}(d) := \frac{n^{1/2-d}(\bar{X}_n - \mu)}{\sqrt{q^{-2d}\bar{\gamma}_0 + 2q^{-2d}\sum_{h=1}^{q}\bar{\gamma}_h(1-h/q)}}, \tag{4.1}$$

$$T_n^{\mathrm{stu}}(\hat{d}) := \frac{n^{1/2-\hat{d}}(\bar{X}_n - \mu)}{\sqrt{q^{-2\hat{d}}\bar{\gamma}_0 + 2q^{-2\hat{d}}\sum_{h=1}^{q}\bar{\gamma}_h(1-h/q)}}. \tag{4.2}$$

When the linear process in hand is of short memory, then $T_n^{\mathrm{stu}}(0)$ stands for a version of $T_n^{\mathrm{stu}}(d)$ in which d is replaced by 0. Under the conditions of our Theorem 2.2, from Theorem 3.1 of Giraitis *et al.* [10] and Theorem 2.2 of Abadir *et al.* [1], we conclude that the limiting distribution of $T_n^{\mathrm{stu}}(0)$, $T_n^{\mathrm{stu}}(d)$ and $T_n^{\mathrm{stu}}(\hat{d})$ is standard normal. Thus, $T_n^{\mathrm{stu}}(0)$, $T_n^{\mathrm{stu}}(d)$ and $T_n^{\mathrm{stu}}(\hat{d})$ converge to the same limiting distribution as that of $G_n^{\mathrm{stu}}(0)$, $G_n^{\mathrm{stu}}(d)$ and $G_n^{\mathrm{stu}}(\hat{d})$ under the conditions of Theorem 2.2.

In Tables 1–6, we provide motivating simulation results in preparation for the upcoming in depth numerical studies in Tables 7–12. In the following Tables 1–6 we use packages "arima.sim" and "fracdiff.sim" in R to generate observations from short and long memory non-Gaussian processes, respectively. Tables 1–6 present empirical probabilities of coverage with the normal cut-off points ± 1.96, that is, the nominal probability coverage is 0.95. The results are based on 1000 replications of the data and the multinomial weights $(w_1^{(n)}, \dots, w_n^{(n)})$. The choice of q is made based on relation (2.14) of Abadir *et al.* [1] in each case. More precisely, we let q be $\lceil n^{1/3} \rceil$, where $\lceil \cdot \rceil$ stands for the ceiling function, for the examined short memory linear processes, and for long memory linear processes with $0 < d < 0.25$, we let q be $\lceil n^{1/(3+4d)} \rceil$, and for $0.25 < d < 0.5$, we let q be $\lceil n^{1/2-d} \rceil$ and, when the data are long memory with parameter d, then $\hat{d}$ stands for the MLE approximation of d, with the Hasslet and Raftery [13] method used to approximate the likelihood. This estimator of d is provided in the R package "fracdiff" and it is used in our simulation studies in Tables 4, 6, 10 and 12. We note that there are other commonly used methods of estimating the memory parameter d, such as the Whittle estimator (cf. Künch [16] and Robinson [20]), which is available in the R package "longmemo" using the Beran [3] algorithm. For more on estimators for the memory parameter d and their asymptotic behavior, we refer, to for example, Robinson [21] and Moulines and Soulier [18] and references therein. In Tables 1–6, coverage$G_n^{\mathrm{stu}}(\cdot)$ and length$G_n^{\mathrm{stu}}(\cdot)$ stand, respectively, for the empirical coverage probabilities and the average of the lengths of the randomized confidence intervals for the population mean based on the randomized pivots $G_n^{\mathrm{stu}}(\cdot)$, and coverage$T_n^{\mathrm{stu}}(\cdot)$ and length$T_n^{\mathrm{stu}}(\cdot)$ stand for their non-randomized counterparts constructed based on the classical pivots $T_n^{\mathrm{stu}}(\cdot)$.

The numerical studies in the preceding 6 tables indicate better accuracy of the randomized pivots for both short and long memory linear processes. These tables at the same time address the trade-off between the accuracy of the confidence intervals and their lengths. Improving upon the probabilities of coverage comes at the expense of wider confidence intervals. However, the increase in length of our introduced confidence intervals is a minor drawback when one puts their significantly better probabilities of coverage into perspective. In other words, the relation between the length and the accuracy of our randomized confidence intervals can be described as a balanced one.

We now present a more in depth simulation study for non-Gaussian linear processes. Here, once again, we use the packages "arima.sim" and "fracdiff.sim" in R to generate observations from short and long memory linear processes, respectively. In our numerical studies below, as in Tables 1–12, we use the standardized Lognormal (0, 1) distribution, that is, Lognormal with mean 0 and variance 1, to generate observations from short or long memory linear processes. The choice of Lognormal (0, 1) in our studies is due to the fact it is a heavily skewed distribution.

The following Tables 7–12 are presented to illustrate the significantly better performance of $G_n^{\mathrm{stu}}(d)$, $G_n^{\mathrm{stu}}(\hat{d})$ and $G_n^{\mathrm{stu}}(0)$ over their respective classical counterparts $T_n^{\mathrm{stu}}(d)$, $T_n^{\mathrm{stu}}(\hat{d})$ and $T_n^{\mathrm{stu}}(0)$, in view of our Theorem 2.2. In Tables 7 and 8 we present numerical comparisons between the performance of $G_n^{\mathrm{stu}}(0)$ to that of $T_n^{\mathrm{stu}}(0)$, both as pivots for the population mean, for some moving average and autoregressive processes. The numerical performance of $G_n^{\mathrm{stu}}(d)$ to that of the classical $T_n^{\mathrm{stu}}(d)$ for some long memory linear processes are presented in Tables 9 and 11. Tables 10 and 12 are specified to comparing $G_n^{\mathrm{stu}}(\hat{d})$ to $T_n^{\mathrm{stu}}(\hat{d})$.

Table 1. MA(1): $X_t = W_t - 0.5W_{t-1}$

Distribution	n	coverage$G_n^{\mathrm{stu}}(0)$	length$G_n^{\mathrm{stu}}(0)$	coverage$T_n^{\mathrm{stu}}(0)$	length$T_n^{\mathrm{stu}}(0)$
$W_t \stackrel{d}{=}$ Lognormal(0, 1)	10	0.887	2.366	0.852	1.791
	40	0.950	1.449	0.936	1.151

Table 2. AR(1): $X_t = 0.5X_{t-1} + W_t$

Distribution	n	coverage$G_n^{\mathrm{stu}}(0)$	length$G_n^{\mathrm{stu}}(0)$	coverage$T_n^{\mathrm{stu}}(0)$	length$T_n^{\mathrm{stu}}(0)$
$W_t \stackrel{d}{=}$ Lognormal(0, 1)	12	0.859	2.278	0.817	1.738
	48	0.943	1.461	0.918	1.161

Table 3. Long Memory with $d = 0.2$: $X_t = (1 - B)^{0.2} W_t$

Distribution	n	coverage$G_n^{\mathrm{stu}}(0.2)$	length$G_n^{\mathrm{stu}}(0.2)$	coverage$T_n^{\mathrm{stu}}(0.2)$	length$T_n^{\mathrm{stu}}(0.2)$
$W_t \stackrel{d}{=}$ Lognormal(0, 1)	50	0.916	1.303	0.847	0.211
	200	0.942	0.888	0.906	0.086

Table 4. Long Memory with $d = 0.2$: $X_t = (1 - B)^{0.2} W_t$; estimator $\hat{d}$ used

Distribution	n	coverage$G_n^{\mathrm{stu}}(\hat{d})$	length$G_n^{\mathrm{stu}}(\hat{d})$	coverage$T_n^{\mathrm{stu}}(\hat{d})$	length$T_n^{\mathrm{stu}}(\hat{d})$
$W_t \stackrel{d}{=}$ Lognormal(0, 1)	100	0.863	0.998	0.812	0.190
	400	0.947	0.713	0.913	0.063

Table 5. Long Memory with $d = 0.4$: $X_t = (1 - B)^{0.4} W_t$

Distribution	n	coverage$G_n^{\mathrm{stu}}(0.4)$	length$G_n^{\mathrm{stu}}(0.4)$	coverage$T_n^{\mathrm{stu}}(0.4)$	length$T_n^{\mathrm{stu}}(0.4)$
$W_t \stackrel{d}{=}$ Lognormal(0, 1)	125	0.915	3.042	0.864	0.051
	500	0.944	2.926	0.895	0.016

Table 6. Long Memory with $d = 0.4$: $X_t = (1 - B)^{0.4} W_t$; estimator $\hat{d}$ used

Distribution	n	coverage$G_n^{\mathrm{stu}}(\hat{d})$	length$G_n^{\mathrm{stu}}(\hat{d})$	coverage$T_n^{\mathrm{stu}}(\hat{d})$	length$T_n^{\mathrm{stu}}(\hat{d})$
$W_t \stackrel{d}{=}$ Lognormal(0, 1)	500	0.912	2.689	0.855	0.018
	2000	0.935	2.591	0.907	0.005

Table 7. MA(1): $X_t = W_t - 0.5W_{t-1}$

Distribution	n	$\mathrm{Prop}G_n^{\mathrm{stu}}(0)$	$\mathrm{Prop}T_n^{\mathrm{stu}}(0)$
$W_t \stackrel{d}{=}$ Lognormal(0, 1)	15	0.132	0.000
	30	0.210	0.156

Table 8. AR(1): $X_t = 0.5X_{t-1} + W_t$

Distribution	n	$\mathrm{Prop}G_n^{\mathrm{stu}}(0)$	$\mathrm{Prop}T_n^{\mathrm{stu}}(0)$
$W_t \stackrel{d}{=}$ Lognormal(0, 1)	50	0.118	0.022
	100	0.640	0.512

Table 9. Long Memory with $d = 0.2$: $X_t = (1 - B)^{0.2} W_t$

Distribution	n	$\mathrm{Prop}G_n^{\mathrm{stu}}(0.2)$	$\mathrm{Prop}T_n^{\mathrm{stu}}(0.2)$
$W_t \stackrel{d}{=}$ Lognormal(0, 1)	150	0.590	0.004
	300	0.614	0.068

Table 10. Long Memory with $d = 0.2$: $X_t = (1 - B)^{0.2} W_t$; estimator $\hat{d}$ used

Distribution	n	$\mathrm{Prop}G_n^{\mathrm{stu}}(\hat{d})$	$\mathrm{Prop}T_n^{\mathrm{stu}}(\hat{d})$
$W_t \stackrel{d}{=}$ Lognormal(0, 1)	300	0.120	0.000
	600	0.572	0.008

Table 11. Long Memory with $d = 0.4$: $X_t = (1 - B)^{0.4} W_t$

Distribution	n	$\mathrm{Prop}G_n^{\mathrm{stu}}(0.4)$	$\mathrm{Prop}T_n^{\mathrm{stu}}(0.4)$
$W_t \stackrel{d}{=}$ Lognormal(0, 1)	300	0.214	0.000
	600	0.468	0.000

Table 12. Long Memory with $d = 0.4$: $X_t = (1 - B)^{0.4} W_t$; estimator $\hat{d}$ used

Distribution	n	$\mathrm{Prop}G_n^{\mathrm{stu}}(\hat{d})$	$\mathrm{Prop}T_n^{\mathrm{stu}}(\hat{d})$
$W_t \stackrel{d}{=}$ Lognormal(0, 1)	1500	0.064	0.000
	3000	0.262	0.000

In Table 7, for the therein underlined MA(1) process, we generate 500 empirical coverage probabilities of the event that

$$G_n^{(\mathrm{stu})} \in [-1.96, 1.96].$$

Each one of these generated 500 coverage probabilities is based on 500 replications. We then record the proportion of those coverage probabilities that deviate from the nominal 0.95 by no more than 0.01. This proportion is denoted by $\mathrm{Prop}G_n^{(\mathrm{stu})}(0)$. For the same generated data, the same proportion, denoted by $\mathrm{Prop}T_n^{(\mathrm{stu})}(0)$, is also recorded for $T_n^{(\mathrm{stu})}(0)$, that is, the classical counterpart of $G_n^{(\mathrm{stu})}(0)$.

The same idea is used to compute the proportions $\mathrm{Prop}G_n^{(\mathrm{stu})}(0)$, $\mathrm{Prop}G_n^{(\mathrm{stu})}(d)$ and $\mathrm{Prop}G_n^{(\mathrm{stu})}(\hat{d})$ and those of their respective classical counterparts $\mathrm{Prop}T_n^{(\mathrm{stu})}(0)$, $\mathrm{Prop}T_n^{(\mathrm{stu})}(d)$ and $\mathrm{Prop}T_n^{(\mathrm{stu})}(\hat{d})$, in Tables 8–12 for the therein indicated short and long memory processes.

Here again, the choice of q is based on relation (2.14) of Abadir *et al.* [1] in each case. More precisely, we let q be $\lceil n^{1/3} \rceil$ for the examined short memory linear processes, and for long memory linear processes with $0 < d < 0.25$, we let q be $\lceil n^{1/(3+4d)} \rceil$, and for $0.25 < d < 0.5$, we let q be $\lceil n^{1/2-d} \rceil$.

Remark 4.1. It is important to note that, our randomized pivots $G_n^{\mathrm{stu}}(0)$, $G_n^{\mathrm{stu}}(d)$ and $G_n^{\mathrm{stu}}(\hat{d})$, for $\mu = E_X X_1$, significantly outperform their respective classical counterparts $T_n^{\mathrm{stu}}(0)$, $T_n^{\mathrm{stu}}(d)$ and $T_n^{\mathrm{stu}}(\hat{d})$ for short and long memory linear processes. This better performance, most likely, is an indication that the respective sampling distributions of $G_n^{\mathrm{stu}}(0)$, $G_n^{\mathrm{stu}}(d)$ and $G_n^{\mathrm{stu}}(\hat{d})$ approach that of standard normal at a faster speed as compared to that of their respective classical counterparts $T_n^{\mathrm{stu}}(0)$, $T_n^{\mathrm{stu}}(d)$ and $T_n^{\mathrm{stu}}(\hat{d})$. In other words, approximating the sampling distributions of $G_n^{\mathrm{stu}}(0)$, $G_n^{\mathrm{stu}}(d)$ and $G_n^{\mathrm{stu}}(\hat{d})$ by that of standard normal most likely result in smaller magnitudes of error in terms the number of observations n. The difference in the performance is even more evident when comparing $G_n^{\mathrm{stu}}(d)$ and $G_n^{\mathrm{stu}}(\hat{d})$ to $T_n^{\mathrm{stu}}(d)$ and $T_n^{\mathrm{stu}}(\hat{d})$, respectively, for non-Gaussian long memory linear processes (cf. Tables 3–6 and Tables 9–12).

5. On the bootstrap and linear processes

In the classical theory of the bootstrap, constructing an asymptotic bootstrap confidence interval for the population mean, based on i.i.d. data, is done by using the Student t-statistic, and estimating the underlying percentile of the conditional distribution, given the data, by repeatedly and independently resampling from the set of data in hand (cf. Efron and Tibshirani [9], for example).

Under certain conditions, the cutoff points of the conditional distribution, given the data, of the randomized Student t-statistic, as in (1.2) in terms of i.i.d. $X_1, X_2, \ldots,$ are used to estimate those of the sampling distribution of the traditional pivot. For more on bootstrapping i.i.d. data we refer to, for example, Davison and Hinkley [7], Hall [11] and Shao and Tu [22]. Considering that the cutoff points of the randomized t-statistic are unknown, they usually are estimated via drawing $B \geq 2$ independent bootstrap sub-samples. The same approach is also taken when the data form short or long memory processes. For references on various bootstrapping methods to mimic the sampling distribution of statistics based on dependent data, and thus, to capture a characteristic of the population, we refer to Härdle *et al.* [12], Kim and Nordman [14], Kreiss and Paparoditis [15], Nordman and Lahiri [19], Lahiri [17], and references therein.

Extending the i.i.d. based techniques of the bootstrap to fit dependent data by no means can be described as straightforward. Our investigation of the bootstrap, in this section, sheds light on some well known issues that arise when the bootstrap is applied to long memory processes.

In this section, we study the problem of bootstrapping linear processes via the same approach that we used to investigate the asymptotic distribution of G_n, and its Studentized versions $G_n^{\mathrm{stu}}(d)$ and $G_n^{\mathrm{stu}}(\hat{d})$, the direct randomized pivots for the population mean μ.

In the classical method of conditioning on the data, a bootstrap sub-sample becomes an i.i.d. set of observables in terms of the classical empirical distribution even when the original data are dependent. In comparison, conditioning on the weights enables us to trace the effect of randomization by the weights $w_i^{(n)}$ on the stochastic nature of the original sample.

To formally state our results on bootstrapping linear processes, we first consider the bootstrapped sum $\bar{X}_n^* - \bar{X}_n$, where $\bar{X}_n^*$ is the mean of a bootstrap sub-sample $X_1^*, \dots, X_n^*$ drawn with replacement from the sample X_i, $1 \le i \le n$, of linear processes as defined in (2.1). Via (2.4), instead of (2.5) as in (2.6), define

$$T_n^* := \frac{\bar{X}_n^* - \bar{X}_n}{\sqrt{D_{n,w}^*}} = \frac{\sum_{i=1}^n (w_i^{(n)}/n - 1/n) X_i}{\sqrt{D_{n,w}^*}}, \tag{5.1}$$

where

$$\begin{aligned} D_{n,w}^* &:= \mathrm{Var}_{X|w}\bigl(\bar{X}_n^* - \bar{X}_n\bigr) \\ &= \gamma_0 \sum_{j=1}^n \biggl(\frac{w_i^{(n)}}{n} - \frac{1}{n}\biggr)^2 + 2\sum_{h=1}^{n-1} \gamma_h \sum_{j=1}^{n-h} \biggl(\frac{w_j^{(n)}}{n} - \frac{1}{n}\biggr)\biggl(\frac{w_{j+h}^{(n)}}{n} - \frac{1}{n}\biggr) \end{aligned}$$

and, as before in (2.4) and (2.5), $w_i^{(n)}$, $1 \le i \le n$, are the multinomial random weights of size n with respective probabilities $1/n$, and are independent from the observables in hand.

Despite the seeming similarity of the bootstrapped statistic T_n^* to G_n as in (2.6), the two objects are, in fact, very different from each other. Apart from the latter being a direct pivot for the population mean μ, while the former is not, T_n^* can only be used up to short memory processes. In other words, in case of long memory linear processes, it fails to converge in distribution to a nondegenerate limit (cf. Remark 5.1). This is quite disappointing when bootstrap is used to capture the sampling distribution of the classical pivot $T_n(d)$ (for μ) of a long memory linear process by that of its bootstrapped version T_n^*. It should also be kept in mind that, in view of Remark 5.1, for T_n^* the natural normalizing sequence, i.e., $\mathrm{Var}_{X|w}(\bar{X}_n^* - \bar{X}_n) = \mathrm{Var}_{X|w}\{\sum_{i=1}^n (\frac{w_i^{(n)}}{n} - \frac{1}{n}) X_i\}$, fails to provide the same asymptotic distribution as that of the original statistic $T_n^{\mathrm{stu}}(d)$, when the data are of long memory.

Remark 5.1. When dealing with dependent data, T_n^* does not preserve the covariance structure of the original data. This can be explained by observing that the expected values of the coefficients of the covariance γ_h, $h \ge 1$, are $\mathrm{cov}_w(w_1^{(n)}, w_2^{(n)}) = -1/n$. As a result (cf. (6.30) and

(6.31) in the proofs), as $n \to \infty$, one has

$$n^{-1}\gamma_0 \sum_{j=1}^{n}\bigl(w_j^{(n)} - 1\bigr)^2 - \gamma_0 = o_{P_w}(1), \tag{5.2}$$

$$\sum_{h=1}^{n-1}\gamma_h \sum_{j=1}^{n-h}\bigl(w_j^{(n)} - 1\bigr)\bigl(w_{j+h}^{(n)} - 1\bigr) = o_{P_w}(n). \tag{5.3}$$

In view of (5.2) and (5.3), for any linear long memory process, $\{X_i, i \ge 1\}$, as defined in part (B) of Theorem 2.2, with a finite and positive variance, for any $0 < d < 1/2$, as $n \to \infty$, we have

$$\begin{aligned}
&\operatorname{Var}_{X|w}\bigl(n^{1/2-d}\bigl(\bar{X}_n^* - \bar{X}_n\bigr)\bigr)\\
&\quad = \operatorname{Var}_{X|w}\left(n^{1/2-d}\sum_{i=1}^{n}\left(\frac{w_i^{(n)}}{n} - \frac{1}{n}\right)X_i\right) \to 0 \qquad \text{in probability } P_w.
\end{aligned}$$

The latter conclusion implies that T_n^* cannot be used for long memory processes. Hence, T_n^* works *only* for short memory processes.

In view of (5.2) and (5.3), for short memory linear processes, T_n^* can, without asymptotic loss of information in probability-P_w, also be defined as

$$T_n^* := \frac{\sum_{i=1}^{n}(w_i^{(n)}/n - 1/n)X_i}{\sqrt{\gamma_0 \sum_{j=1}^{n}(w_i^{(n)}/n - 1/n)^2}}. \tag{5.4}$$

Thus, the two definitions of T_n^* in (5.1) and (5.4) coincide asymptotically. We note in passing that the asymptotic equivalence of (5.1) and (5.4) does not mean that for a finite number of data, they are equally robust. Obviously, (5.1) is more robust, and it should be used, for studying its behavior for a short memory finite sample of size n.

In the following (5.5), for further study we present the Studentized counterpart of T_n^* as defined in (5.4):

$$T_n^{*^{\text{stu}}} := \frac{\sum_{i=1}^{n}(w_i^{(n)}/n - 1/n)X_i}{\sqrt{\bar{\gamma}_0 \sum_{j=1}^{n}(w_i^{(n)}/n - 1/n)^2}}. \tag{5.5}$$

The following two results are respective counterparts of Theorems 2.1 and 2.2.

Theorem 5.1. *Suppose that $\{X_i, i \ge 1\}$ is a stationary linear process as defined in* (2.1) *with $\sum_{k=0}^{\infty}|a_k| < \infty$. Then, as $n \to \infty$, we have for all $t \in \mathbb{R}$,*

$$P_{X|w}\bigl(T_n^* \le t\bigr) \longrightarrow \Phi(t) \qquad \textit{in probability } P_w \tag{5.6}$$

and, consequently,

$$P_{X,w}\bigl(T_n^* \le t\bigr) \longrightarrow \Phi(t), \qquad t \in \mathbb{R}. \tag{5.7}$$

Theorem 5.2. *Assume that for the stationary linear process* $\{X_i, i \geq 1\}$, *as defined in* (2.1), $\sum_{k=0}^{\infty} |a_k| < \infty$, *and* $E\zeta_1^4 < \infty$. *Then, as* $n \to \infty$, *we have for all* $t \in \mathbb{R}$,

$$P_{X|w}\bigl(T_n^{*^{\mathrm{stu}}} \leq t\bigr) \longrightarrow \Phi(t) \qquad \textit{in probability } P_w \tag{5.8}$$

and, consequently,

$$P_{X,w}\bigl(T_n^{*^{\mathrm{stu}}} \leq t\bigr) \longrightarrow \Phi(t). \tag{5.9}$$

It is also noteworthy that taking the traditional method of conditioning on the data yields the same conclusion on T_n^* as the one in Remark 5.1 for long memory linear processes. In fact, in case of conditioning on the sample, recalling that here without loss of generality $\mu = 0$, one can see that

$$\begin{aligned}
&\mathrm{Var}\bigl(n^{1/2-d}\bigl(\bar{X}_n^* - \bar{X}_n\bigr)|X_1, \dots, X_n\bigr) \\
&\quad = \mathrm{Var}\left(n^{1/2-d} \sum_{i=1}^{n} \left(\frac{w_i^{(n)}}{n} - \frac{1}{n}\right) X_i | X_1, \dots, X_n\right) \\
&\quad = n^{-2d}\left(1 - \frac{1}{n}\right) \frac{\sum_{i=1}^{n} X_i^2}{n} - n^{-1-2d} \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) X_i X_{i+h} \\
&\quad = o_{P_X}(1), \qquad \text{as } n \to \infty; \text{ when } 0 < d < 1/2.
\end{aligned}$$

The preceding convergence to zero takes place when X_is are of long memory.

In the literature, block-bootstrap methods are usually used to modify the randomized t-statistic T_n^* so that it should reflect the dependent structure of the data (cf. Kim and Nordman [14], Kreiss and Paparoditis [15], Lahiri [17] and references therein) that is concealed by the conditional independence of the randomized random variables with common distribution $F_n(x) := n^{-1}\#\{k : 1 \leq k \leq n, X_k \leq x\}$, $x \in \mathbb{R}$, given $X_1, \dots, X_n$. These methods are in contrast to our direct pivot G_n as in (2.6), and its Studentized version $G_n^{\mathrm{stu}}(d)$ as defined in (2.8), that can be used *both* for short and long memory processes without dividing the data into blocks. This is so, since, the random weights $|w_i^{(n)}/n - 1/n|$ in G_n, and in its Studentized version $G_n^{\mathrm{stu}}(d)$ as defined in (2.8), project and preserve the covariance structure of the original sample. To further elaborate on the latter, we note that, as $n \to \infty$, we have $nE_w|(w_1^{(n)}/n - 1/n)(w_2^{(n)}/n - 1/n)| \to 4e^{-2}$ (cf. (6.5) in the proofs). This, in turn, implies that, for $0 \leq d < 1/2$, the term $n^{1-2d} \sum_{h=1}^{n-1} \gamma_h \sum_{j=1}^{n-h} |\frac{w_j^{(n)}}{n} - \frac{1}{n}||\frac{w_{j+h}^{(n)}}{n} - \frac{1}{n}|$ will be neither zero nor infinity in the limit (cf. (6.6) and (6.7) in the proofs). This means that, unlike T_n^*, G_n preserves the covariance structure of the data without more ado. Hence, G_n and its Studentized version $G_n^{\mathrm{stu}}(d)$, as in (2.6) and (2.8) respectively, are natural choices to make inference about the mean of long memory processes, as well as that of short memory ones.

Comparison of our randomization approach to bootstrap

In comparing the use of our direct randomized pivot G_n and its Studentized versions $G_n^{\mathrm{stu}}(0)$ and $G_n^{\mathrm{stu}}(\hat{d})$ to the bootstrap method of constructing confidence intervals, our approach has a number of advantages over the latter. The first advantage is that G_n and its Studentized versions can be used without any adjustment of the data, such as dividing them into blocks, for example. The second advantage, that is also a consequence of the first one, is that in G_n and its Studentized versions, for both long and short memory linear processes, the natural normalization, that is, the standard deviation of the randomized sums as the normalizing sequence, directly yields the CLT. On the other hand, the i.i.d. based bootstrap, represented by T_n^*, as in (5.5), fails to yield a CLT for the bootstrapped linear process with the standard deviation as the normalizing sequence (cf. Remark 5.1), and it is to be replaced by modified versions of it that are computed based on blocks of the original data set as in Kim and Nordman [14]. The third advantage concerns the fact that the Studentized pivots $G_n^{\mathrm{stu}}(0)$ and $G_n^{\mathrm{stu}}(\hat{d})$ are direct pivots for the population mean μ. Recall that T_n^* fails to remember μ (cf. Remark 2.1). We note as well that our approach to making inference about the population mean μ, based on the pivot G_n and its Studentized versions, for both short and long memory linear processes, does not require repeated resampling from the original data. This is in contrast to the block bootstrap methods, where bootstrap samples are to be drawn repeatedly from the data after dividing them into blocks (cf. for example, Lahiri [17]).

Further to the block-bootstrap approach, the sub-sample window approach for long range dependent linear processes as in Nordman and Lahiri [19] is similar in nature to the former. Only, in the latter, replicated bootstrap blocks samples are replaced by overlapping block sub-samples. Treating each block as a scaled-down copy of the original time series, the plug-in analog of a normalized version of Student t-pivot for the mean, that is based on the original n observations, is defined on each block. The empirical distribution of the latter plug-in analogs is shown to be near in probability to the sampling distribution of the original pivot. Consequently, the comments made above on comparing our randomization approach to the bootstrap approach, with the exception of the one concerning repeated resampling, continue to hold true in the sub-sample window method as well.

6. Proofs

In proving Theorems 2.1 and 5.1, we make use of Theorem 2.2 of Abadir *et al.* [2] in which the asymptotic normality of sums of deterministically weighted linear processes are established. In this context, in Theorems 2.1 and 5.1, we view the sums defined in (2.4) and (2.5) as randomly weighted sums of the data on X. Conditioning on the weights $w_i^{(n)}$s, we show that the conditions required for the deterministic weights in the aforementioned Theorem 2.2 of Abadir *et al.* [2] hold true in probability-P_w in this context. The latter, in view of the characterization of convergence in probability in terms of almost sure convergence of subsequences, will enable us to conclude the conditional CLTs, in probability-P_w, in Theorems 2.1 and 5.1. The unconditional CLTs in terms of the joint distribution $P_{X,w}$ will then follow from the dominated convergence theorem.

Employing Slutsky type arguments, we conclude Theorems 2.2 and 5.2 from Theorems 2.1 and 5.1, respectively.

Proof of Theorem 2.1

In view of Theorem 2.2 of Abadir *et al.* [2], the proof of Theorem 2.1 follows if we show that, as $n \to \infty$, the following two statements, namely (6.1) and (6.2), hold true:

$$\frac{\max_{1\le i\le n} |w_j^{(n)}/n - 1/n|}{\sqrt{D_n}} = o_{P_w}(1), \tag{6.1}$$

$$\frac{\sum_{i=1}^n (w_j^{(n)}/n - 1/n)^2}{D_n} = O_{P_w}(1), \tag{6.2}$$

where D_n is as defined in (2.7).

In order to establish (6.1) and (6.2), we first note that it is not difficult to observe that, as $n \to \infty$,

$$n\sum_{j=1}^n \left(\frac{w_j^{(n)}}{n} - \frac{1}{n}\right)^2 \longrightarrow 1 \qquad \text{in probability } P_w. \tag{6.3}$$

Since, for $0 \le d < 1/2$, as $n \to \infty$, we have that

$$0 < \lim_{n\to\infty} \operatorname{Var}\bigl(n^{1/2-d}\bar{X}_n\bigr) = \lim_{n\to\infty} n^{-2d}\left(\gamma_0 + 2\sum_{h=1}^n \gamma_h(1-h/n)\right) < \infty,$$

we will also have

$$n^{1-2d}\sum_{h=1}^n \gamma_h \sum_{j=1}^{n-h} \left|\frac{w_j^{(n)}}{n} - \frac{1}{n}\right|\left|\frac{w_{j+h}^{(n)}}{n} - \frac{1}{n}\right| - 4e^{-2}\lim_{n\to\infty} n^{-2d}\sum_{h=1}^n \gamma_h(1-h/n) = o_{P_w}(1). \tag{6.4}$$

In order to prove (6.4), we first note that

$$\begin{aligned}
&E_w\left(\left|\frac{w_1^{(n)}}{n} - \frac{1}{n}\right|\left|\frac{w_2^{(n)}}{n} - \frac{1}{n}\right|\right) \\
&\quad = -1/n^3 - 2/n^2 E_w \\
&\qquad \times \bigl\{\bigl(w_1^{(n)} - 1\bigr)\bigl(w_2^{(n)} - 1\bigr)\mathbb{1}\bigl(\bigl(w_1^{(n)} - 1\bigr)\bigl(w_2^{(n)} - 1\bigr) < 0\bigr)\bigr\} \\
&\quad = -1/n^3 + 4/n^2(1-1/n)^n\bigl(1 - 1/(n-1)\bigr)^n.
\end{aligned} \tag{6.5}$$

The preceding relation implies that, as $n \to \infty$,

$$\begin{aligned} & n^{-2d} E_w\left(n \sum_{h=1}^{n-1} \gamma_h \left(\sum_{j=1}^{n-h} \left|\frac{w_j^{(n)}}{n} - \frac{1}{n}\right| \left|\frac{w_{j+h}^{(n)}}{n} - \frac{1}{n}\right|\right)\right) \\ &= n^{-2d}\left(\sum_{h=1}^{n-1} \gamma_h \left[\frac{-(n-h)}{n^2} + 4\frac{(n-h)}{n}(1-1/n)^n\big(1-1/(n-1)\big)^n\right]\right) \\ &\longrightarrow 4e^{-2} \lim_{n\to\infty} n^{-2d} \sum_{h=1}^{n} \gamma_h (1-h/n). \end{aligned} \tag{6.6}$$

By virtue of the preceding result, in order to prove (6.4), we need to show that, as $n \to \infty$,

$$n^{1-2d} \sum_{h=1}^{n-1} \gamma_h \left(\sum_{j=1}^{n-h} \left|\frac{w_j^{(n)}}{n} - \frac{1}{n}\right| \left|\frac{w_{j+h}^{(n)}}{n} - \frac{1}{n}\right| - b_{n,h}\right) = o_{P_w}(1), \tag{6.7}$$

where

$$b_{n,h} := \frac{-(n-h)}{n^2} + \frac{4(n-h)}{n}\left(1-\frac{1}{n}\right)^n \left(1 - \frac{1}{n-1}\right)^n. \tag{6.8}$$

In order to show the validity of (6.7), for $\varepsilon > 0$, we write

$$\begin{aligned} & P_w\left(\left|n^{1-2d} \sum_{h=1}^{n-1} \gamma_h \sum_{j=1}^{n-h} \left|\frac{w_j^{(n)}}{n} - \frac{1}{n}\right| \left|\frac{w_{j+h}^{(n)}}{n} - \frac{1}{n}\right| - b_{n,h}\right| > \varepsilon\right) \\ &\le \varepsilon^{-1} n^{-2d} \sum_{h=1}^{n-1} |\gamma_h| E_w^{1/2}\left(n \sum_{j=1}^{n-h} \left|\frac{w_j^{(n)}}{n} - \frac{1}{n}\right| \left|\frac{w_{j+h}^{(n)}}{n} - \frac{1}{n}\right| - b_{n,h}\right)^2. \end{aligned} \tag{6.9}$$

We are now to show that $E_w(n \sum_{j=1}^{n-h} |\frac{w_j^{(n)}}{n} - \frac{1}{n}||\frac{w_{j+h}^{(n)}}{n} - \frac{1}{n}| - b_{n,h})^2$ approaches zero uniformly in $1 \le h \le n-1$, as $n \to \infty$.

Some basic, yet not quite trivial, calculations that also include the use of the moment generating function of the multinomial distribution show that

$$E_w\left(\left(\frac{w_1^{(n)}}{n} - \frac{1}{n}\right)\left(\frac{w_2^{(n)}}{n} - \frac{1}{n}\right)\right)^2 = O\big(n^{-4}\big) \tag{6.10}$$

and

$$\begin{aligned} & E_w\left(\left|\left(\frac{w_1^{(n)}}{n} - \frac{1}{n}\right)\left(\frac{w_2^{(n)}}{n} - \frac{1}{n}\right)\left(\frac{w_3^{(n)}}{n} - \frac{1}{n}\right)\left(\frac{w_4^{(n)}}{n} - \frac{1}{n}\right)\right|\right) \\ &= 3/n^6 - 6/n^7 + 8/n^4(1-3/n)^n\big(1-1/(n-3)\big)^n \\ &\quad + 8/n^4(1-1/n)^n\big\{\big(n(n-2)\big)/(n-1)^2 - 1 + \big(1-3/(n-1)\big)^n\big\}. \end{aligned} \tag{6.11}$$

By virtue of (6.10) and (6.11), we have that

$$
\begin{aligned}
&E_w\left(n\sum_{j=1}^{n-h}\left|\frac{w_j^{(n)}}{n}-\frac{1}{n}\right|\left|\frac{w_{j+h}^{(n)}}{n}-\frac{1}{n}\right|-b_{n,h}\right)^2 \\
&\quad = E_w\left(n\sum_{j=1}^{n-h}\left|\frac{w_j^{(n)}}{n}-\frac{1}{n}\right|\left|\frac{w_{j+h}^{(n)}}{n}-\frac{1}{n}\right|\right)^2 - b_{n,h}^2 \\
&\quad = n^2(n-h)O\bigl(n^{-4}\bigr) \\
&\qquad + n^2(n-h)(n-h-1)\bigl(3/n^6-6/n^7+8/n^4(1-3/n)^n\bigl(1-1/(n-3)\bigr)^n \\
&\qquad + 8/n^4(1-1/n)^n\bigl\{\bigl(n(n-2)\bigr)/(n-1)^2-1+\bigl(1-3/(n-1)\bigr)^n\bigr\}\bigr)-b_{n,h}^2.
\end{aligned}
$$

Some further algebra shows that the right-hand side of the preceding relation can be bounded above by

$$
n^3O\bigl(n^{-4}\bigr)+3n^{-2}+8/n(1-1/n)^n\bigl(1-1/(n-1)\bigr)^n \to 0 \qquad \text{as } n\to\infty.
$$

It is important to note that the left-hand side of the preceding convergence dose not depend on h.

Incorporating now the preceding relation into (6.9) yields (6.7).

In conclusion, for both short and long memory data, (6.1) results from (6.28) below, (6.3) and (6.4). The relation (6.2) follows from (6.3) and (6.4). Now the proof of Theorem 2.1 is complete.

Prior to establishing the proof of Theorem 2.2, we first define

$$
s_X^2 := \lim_{n\to\infty}\mathrm{Var}_X\bigl(n^{1/2-d}\bar{X}_n\bigr)=\lim_{n\to\infty}n^{-2d}\left\{\gamma_0+2\sum_{h=1}^{n-1}\gamma_h(1-h/n)\right\}, \tag{6.12}
$$

and note that under regular moment conditions, such as those assumed in Theorem 2.2, the conclusion $0<s_X^2<\infty$ is valid for $0\le d<1/2$.

Proof of Theorem 2.2

Considering that in this theorem the data are linear processes that can be of short memory, or posses the property of long range dependence, here, the proofs are given in a general setup that includes both cases.

Prior to stating the details of the proof of Theorem 2.2, we note that, when the X_is form a long memory process, in view of the in probability-P_X asymptotic equivalence of the estimator $\hat{d}$ to d as stated in the assumptions of this theorem, we present our proofs for $G_n^{\mathrm{stu}}(d)$ rather than for $G_n^{\mathrm{stu}}(\hat{d})$.

The proof of both parts of this theorem will follow if, under their respective conditions, one shows that for $0\le d<1/2$, as $n,q\to\infty$ such that $q=O(n^{1/2})$, the following two statements hold true:

$$\frac{n^{1-2d}\gamma_0\sum_{j=1}^{n}(\frac{w_j^{(n)}}{n}-\frac{1}{n})^2+2n^{1-2d}\sum_{h=1}^{n-1}\gamma_h\sum_{j=1}^{n-h}|(\frac{w_j^{(n)}}{n}-\frac{1}{n})(\frac{w_{j+h}^{(n)}}{n}-\frac{1}{n})|}{n^{-2d}\gamma_0(1-\frac{1}{n})+2n^{1-2d}\sum_{h=1}^{n-1}\gamma_h\{\frac{-(n-h)}{n^3}+\frac{4(n-h)}{n^2}(1-\frac{1}{n})^n(1-\frac{1}{n-1})^n\}} \longrightarrow 1 \quad \text{in probability } P_w, \tag{6.13}$$

$$P_{X|w}\left\{\left|\frac{\frac{n\bar{\gamma}_0\sum_{j=1}^{n}(\frac{w_j^{(n)}}{n}-\frac{1}{n})^2}{q^{2d}}+2q^{-1-2d}\sum_{h=1}^{q}\bar{\gamma}_h\sum_{j=1}^{q-h}|(w_j^{(n)}-1)(w_{j+h}^{(n)}-1)|}{\frac{\bar{\gamma}_0(1-1/n)}{q^{2d}}+2q^{-2d}\sum_{h=1}^{q}\bar{\gamma}_h(1-\frac{h}{q})\{\frac{-1}{n}+4(1-\frac{1}{n})^n(1-\frac{1}{n-1})^n\}}-1\right|>\varepsilon\right\} = o_{P_w}(1), \tag{6.14}$$

where, ε is an arbitrary positive number.

Due to the following two conclusions, namely (6.15) and (6.16), (6.13) and (6.14) will, in turn, imply Theorem 2.2. We have, as $n\to\infty$,

$$n^{-2d}\left\{\gamma_0+2\sum_{h=1}^{n-1}\gamma_h(1-h/n)\right\}\to s_X^2, \tag{6.15}$$

and, as $n,q\to\infty$ such that $q=O(n^{1/2})$,

$$q^{-2d}\left\{\bar{\gamma}_0+2\sum_{h=1}^{q}\bar{\gamma}_h(1-h/q)\right\}\to s_X^2 \qquad \text{in probability } P_X, \tag{6.16}$$

where, s_X^2 is as defined in (6.12). In the context of Theorem 2.2, the conclusion (6.16) results from Theorem 3.1 of Giraitis *et al.* [10]. This is so, since, in Theorem 2.2 we assume that the data have a finite forth moment, $n,q\to\infty$ in such a way that $q=O(n^{1/2})$ and, in the case of long memory, in part (B) of Theorem 2.2 we consider long memory linear processes for which we have $a_i\sim ci^{d-1}$, as $i\to\infty$.

In order to prove (6.13), we note that, as $n\to\infty$,

$$n^{-2d}\gamma_0\left(1-\frac{1}{n}\right)+2n^{1-2d}\sum_{h=1}^{n-1}\gamma_h b_{n,h} \to \begin{cases}(1-4e^{-2})\gamma_0+4e^{-2}s_X^2, & \text{when } d=0;\\ 4e^{-2}s_X^2, & \text{when } 0<d<1/2,\end{cases}$$

where $b_{n,h}$ is as in (6.8). Considering that here we have $\lim_{n\to\infty}n^{-2d}\sum_{h=1}^{n}\gamma_h<\infty$, (6.30) and (6.7) imply (6.13), as $n\to\infty$.

In order to establish (6.14), in view of (6.30) and the fact that, under the conditions of Theorem 2.2, as $n \to \infty$, $\bar{\gamma}_0 - \gamma_0 = o_{P_X}(1)$, we conclude that, as $n \to \infty$,

$$P_{X|w}\left(\left|n^{1-2d}\left(\frac{q}{n}\right)^{-2d}\bar{\gamma}_0\sum_{i=1}^{n}\left(\frac{w_i^{(n)}}{n}-\frac{1}{n}\right)^2 - q^{-2d}\bar{\gamma}_0(1-1/n)\right| > \varepsilon\right) \to 0 \qquad \text{in probability } P_w,$$

where $0 \le d < 1/2$ and $\varepsilon > 0$ is arbitrary.

We proceed with the proof of (6.14) by showing that, as $n, q \to \infty$, the following relation holds true: for arbitrary $\varepsilon_1, \varepsilon_2 > 0$, as $n, q \to \infty$, in such a way that $q = O(n^{1/2})$,

$$P_w\left\{P_{X|w}\left(q^{-2d}\left|\sum_{h=1}^{q}\bar{\gamma}_h B_{n,q}\right| > 2\varepsilon_1\right) > 2\varepsilon_2\right\} \to 0, \tag{6.17}$$

where

$$B_{n,q}(h) := q^{-1}\sum_{j=1}^{q-h}\left|\left(w_j^{(n)}-1\right)\left(w_{j+h}^{(n)}-1\right)\right| - b_{n,q,h},$$

$$\begin{aligned} b_{n,q,h} &:= E_w\left(q^{-1}\sum_{j=1}^{q-h}\left|\left(w_j^{(n)}-1\right)\left(w_{j+h}^{(n)}-1\right)\right|\right) \\ &= \frac{-(q-h)}{nq} + \frac{4(q-h)}{q}\left(1-\frac{1}{n}\right)^n\left(1-\frac{1}{n-1}\right)^n. \end{aligned}$$

In order to establish (6.17), without loss of generality, we first assume $\mu = E_X X_1 = 0$, and for each $1 \le h \le q$ define

$$\gamma_h^* := \frac{1}{n}\sum_{i=1}^{n-h} X_i X_{i+h}. \tag{6.18}$$

Observe now that the left-hand side of (6.17) is bounded above by

$$\begin{aligned} &P_w\left\{P_{X|w}\left(q^{-2d}\left|\sum_{h=1}^{q}\left(\bar{\gamma}_h - \gamma_h^*\right)B_{n,q}(h)\right| > \varepsilon_1\right) > \varepsilon_2\right\} \\ &\quad + P_w\left\{P_{X|w}\left(q^{-2d}\left|\sum_{h=1}^{q}\gamma_h^* B_{n,q}(h)\right| > \varepsilon_1\right) > \varepsilon_2\right\}. \end{aligned} \tag{6.19}$$

We now show that the first term in (6.19), that is, the remainder, is asymptotically negligible. To do so, we note that we have

$$\sum_{h=1}^{q}\big(\bar{\gamma}_h-\gamma_h^*\big)B_{n,q}(h)=-\frac{\bar{X}_n}{n}\sum_{h=1}^{q}B_{n,q}(h)\sum_{i=1}^{n-h}X_i-\frac{\bar{X}_n}{n}\sum_{h=1}^{q}B_{n,q}(h)\sum_{i=1}^{n-h}X_{i+h}$$
$$+\bar{X}^2\sum_{h=1}^{q}B_{n,q}(h) \tag{6.20}$$
$$\sim-\bar{X}^2\sum_{h=1}^{q}B_{n,q}(h)\qquad\text{uniformly in } h \text{ in probability } P_{X|w},$$

where, in the preceding conclusion, generically, $Y_n\sim Z_n$ in probability-P means $Y_n=Z_n(1+o_P(1))$. The approximation in (6.20) is true since, for example, for $\varepsilon>0$

$$P_X\Bigg(\bigcup_{1\le h\le q}\bigg|\bar{X}_n-\frac{\sum_{i=1}^{n-h}X_i}{n}\bigg|>\varepsilon\Bigg)\le q\,P_X\Bigg(\bigg|\frac{\sum_{i=n-h+1}^{n}X_i}{n}\bigg|>\varepsilon\Bigg)$$
$$\le\varepsilon^{-4}q\frac{(h-1)^4}{n^4}E\big(X_1^4\big)$$
$$\le\varepsilon^{-4}\frac{q^5}{n^4}E_X\big(X_1^4\big)\to0\qquad\text{as } n\to\infty.$$

The preceding is true since $1\le h\le q$ and $q=O(n^{1/2})$, as $n,q\to\infty$.

We note that for $0\le d<1/2$, as $n\to\infty$, we have that $n^{1/2-d}\bar{X}_n=O_{P_X}(1)$. The latter conclusion, in view of the equivalence in (6.20), implies that, for each $\varepsilon_1,\varepsilon_2>0$, there exists $\varepsilon>0$ such that

$$P_w\Bigg\{P_{X|w}\Bigg(q^{-2d}\bigg|\sum_{h=1}^{q}\big(\bar{\gamma}_h-\gamma_h^*\big)B_{n,q}(h)\bigg|>\varepsilon_1\Bigg)>\varepsilon_2\Bigg\}$$
$$\sim P_w\Bigg\{\frac{q^{-2d}}{n^{1-2d}}\sum_{h=1}^{q}\big|B_{n,q}(h)\big|>\varepsilon\Bigg\} \tag{6.21}$$
$$\le\varepsilon^{-1}\frac{q^{-2d}}{n^{1-2d}}\sum_{h=1}^{q}E_w\big(\big|B_{n,q}(h)\big|\big).$$

Observing now that $\sup_{n\ge2}\sup_{1\le h\le q}E_w(|B_{n,q}(h)|)\le10$, we can bound the preceding relation above by

$$10\varepsilon^{-1}\frac{q^{1-2d}}{n^{1-2d}}\longrightarrow0,$$

as $n, q \to \infty$ in such away that $q = O(n^{1/2})$. This means that the first term in (6.19) is asymptotically negligible and, as a result, (6.17) follows when the second term in the former relation is also asymptotically negligible. To prove this negligibility, we first define

$$\gamma_h^{**} := \frac{1}{n} \sum_{i=1}^{n} X_i X_{i+h}. \tag{6.22}$$

Now, observe that

$$\begin{aligned}
&P_X\left\{ \bigcup_{1\le h\le q} \left|\gamma_h^{**} - \gamma_h^{*}\right| > \varepsilon \right\} \\
&\quad \le q P\left\{ \frac{1}{n}\left| \sum_{i=n-h+1}^{n} X_i X_{i+h} \right| > \varepsilon \right\} \\
&\quad \le \varepsilon^{-2} \frac{q^3}{n^2} E_X\left(X_1^4\right) \to 0,
\end{aligned}$$

as $n, q \to \infty$ such that $q = O(n^{1/2})$, hence, as $n, q \to \infty$ such that $q = O(n^{1/2})$, using a similar argument to arguing (6.19) and (6.21), with γ_h^* replacing $\bar{\gamma}_h$ and γ_h^{**} replacing γ_h^* therein, we arrive at

$$\begin{aligned}
&P_w\left\{ P_{X|w}\left(q^{-2d} \left| \sum_{h=1}^{q} \gamma_h^{*} B_{n,q}(h) \right| > \varepsilon_1 \right) > \varepsilon_2 \right\} \\
&\quad \sim P_w\left\{ P_{X|w}\left(q^{-2d} \left| \sum_{h=1}^{q} \gamma_h^{**} B_{n,q}(h) \right| > \varepsilon_1 \right) > \varepsilon_2 \right\}.
\end{aligned}$$

Therefore, in order to prove (6.17), it suffices to show that, as $n, q \to \infty$ so that $q = O(n^{1/2})$,

$$P_w\left\{ P_{X|w}\left(q^{-2d} \left| \sum_{h=1}^{q} \gamma_h^{**} B_{n,q}(h) \right| > \varepsilon_1 \right) > \varepsilon_2 \right\} \to 0,$$

where γ_h^{**} is defined in (6.22). The latter relation, in turn, follows from the following two conclusions: as $n, q \to \infty$ so that $q = O(n^{1/2})$,

$$\sup_{1\le h,h'\le q} E_w\left(\left|B_{n,q}(h) B_{n,q}\left(h'\right)\right|\right) = o(1) \tag{6.23}$$

and

$$q^{-4d} \sum_{h=1}^{q} \sum_{h'=1}^{q} \left|E_X\left(\gamma_h^{**}\gamma_{h'}^{**}\right)\right| = O(1). \tag{6.24}$$

To prove (6.23), we use the Cauchy inequality to write

$$
\begin{aligned}
&E_w\big(\big|B_{n,q}(h)B_{n,q}(h')\big|\big)\\
&\quad\le E_w\big(B_{n,q}(h)\big)^2\\
&\quad\le \frac{q-h}{q^2}E_w\big(\big(w_1^{(n)}-1\big)\big(w_2^{(n)}-1\big)\big)^2\\
&\qquad+\frac{(q-h)(q-h-1)}{q^2}E_w\big|\big(w_1^{(n)}-1\big)\big(w_2^{(n)}-1\big)\big(w_3^{(n)}-1\big)\big(w_4^{(n)}-1\big)\big|\\
&\qquad-b_{n,q,h}^2\\
&\quad\le (q-1)/q^2O(1)+3/n^2+8/n(1-1/n)^n\big(1-1/(n-1)\big)^n.
\end{aligned}
$$

We note that the right-hand side of the preceding relation does not depend on h and it approaches zero as $n\to\infty$. The latter conclusion implies (6.23).

In order to establish (6.24), we define

$$
H := \lim_{s\to\infty} s^{-2d}\sum_{\ell=-s}^{s}|\gamma_\ell|.
$$

Observe that $H<\infty$. We now carry on with the proof of (6.24), using a generalization of an argument used in the proof of Proposition 7.3.1 of Brockwell and Davis [4] as follows:

$$
\begin{aligned}
&q^{-4d}\sum_{h=1}^{q}\sum_{h'=1}^{q}\big|E_X\big(\gamma_h^{**}\gamma_{h'}^{**}\big)\big|\\
&\quad\le q^{-2d}\sum_{h=1}^{q}|\gamma_h|q^{-2d}\sum_{h'=1}^{q}|\gamma_{h'}|\\
&\qquad+\left(\frac{q}{n}\right)^{1-2d}n^{-2d}\sum_{k=-n}^{n}|\gamma_h|q^{-2d}\sum_{L=-q}^{q}|\gamma_{k+L}| \qquad (6.25)\\
&\qquad+\frac{1}{n}\sum_{k=-n}^{n}q^{-2d}\sum_{h'=1}^{q}|\gamma_{k+h'}|q^{-2d}\sum_{h=1}^{q}|\gamma_{k-h}|\\
&\qquad+\frac{q^{-2d}}{n^{1-2d}}n^{-2d}\sum_{i=1}^{n}\sum_{k=-n}^{n}|a_ia_{i+k}|q^{-d}\sum_{h=1}^{q}|a_{i+h}|q^{-d}\sum_{h'=1}^{q}|a_{i+k-h'}|.
\end{aligned}
$$

It is easy to see that, as $n\to\infty$, and consequently $q\to\infty$, the right-hand side of the inequality (6.25) converges to the finite limit $3H^2$. Now the proof of (6.24) and also that of Theorem 2.2 are complete.

Proof of Corollary 3.1

Due to the similarity of parts (A) and (B), we only give the proof for part (A) of Corollary 3.1.

In order to establish part (A), we first construct an asymptotic $1-\alpha$ size one-sided randomized confidence bound for the parameter $\mu_X = E_X X$ using part (B) of Theorem 2.2, as follows:

$$\mu_X \geq \frac{\sum_{i=1}^n |w_i^{(n)}/n - 1/n| X_i - D_{n,q,\hat{d}}^{1/2} z_{1-\alpha}}{\sum_{j=1}^n |w_j^{(n)}/n - 1/n|}. \tag{6.26}$$

Now, since the function $\mathcal{G}$ is an increasing function, we conclude that (6.26) is equivalent to having

$$\mathcal{G}(\mu_X) \geq \mathcal{G}\left(\frac{\sum_{i=1}^n |w_i^{(n)}/n - 1/n| X_i - D_{n,q,\hat{d}}^{1/2} z_{1-\alpha}}{\sum_{j=1}^n |w_j^{(n)}/n - 1/n|}\right).$$

Employing Jenssen's inequality at this stage yields conclusion (A) of Corollary 3.1. Now the proof of Corollary 3.1 is complete.

Proof of Theorem 5.1

Without loss of generality here, we assume that $\mu = 0$, and note that

$$\begin{aligned} &\mathrm{Var}_{X|w}\left(\sum_{i=1}^n \left(\frac{w_i^{(n)}}{n} - \frac{1}{n}\right) X_i\right) \\ &\quad = \gamma_0 \sum_{j=1}^n \left(\frac{w_j^{(n)}}{n} - \frac{1}{n}\right)^2 + 2\sum_{h=1}^{n-1} \gamma_h \sum_{j=1}^{n-h} \left(\frac{w_j^{(n)}}{n} - \frac{1}{n}\right)\left(\frac{w_{j+h}^{(n)}}{n} - \frac{1}{n}\right). \end{aligned}$$

Now, in view of Theorem 2.2 of Abadir *et al.* [2], it suffices to show that, as $n \to \infty$,

$$\begin{aligned} &\frac{\max_{1\leq i\leq n}(w_i^{(n)}/n - 1/n)^2}{\gamma_0 \sum_{j=1}^n (w_j^{(n)}/n - 1/n)^2 + 2\sum_{h=1}^{n-1} \gamma_h \sum_{j=1}^{n-h} (w_j^{(n)}/n - 1/n)(w_{j+h}^{(n)}/n - 1/n)} \\ &\quad = o_{P_w}(1). \end{aligned} \tag{6.27}$$

Noting that $\gamma_0 > 0$, the proof of the preceding statement results from the following two conclusions: as $n \to \infty$,

$$n \max_{1\leq i\leq n} \left(\frac{w_i^{(n)}}{n} - \frac{1}{n}\right)^2 = o_{P_w}(1) \tag{6.28}$$

and

$$n\gamma_0 \sum_{j=1}^{n}\Big(\frac{w_j^{(n)}}{n}-\frac{1}{n}\Big)^2 + 2n\sum_{h=1}^{n-1}\gamma_h\sum_{j=1}^{n-h}\Big(\frac{w_j^{(n)}}{n}-\frac{1}{n}\Big)\Big(\frac{w_{j+h}^{(n)}}{n}-\frac{1}{n}\Big)-\gamma_0$$
$$= o_{P_w}(1). \tag{6.29}$$

To prove (6.28), for $\varepsilon > 0$, in what follows we employ Bernstien's inequality and write

$$P_w\bigg(\max_{1\le i\le n}\bigg|\frac{w_i^{(n)}}{n}-\frac{1}{n}\bigg| > \frac{\varepsilon}{\sqrt{n}}\bigg) \le n P_w\bigg(\bigg|\frac{w_1^{(n)}}{n}-\frac{1}{n}\bigg| > \frac{\varepsilon}{\sqrt{n}}\bigg)$$
$$\le n\exp\bigg\{-n^{1/2}\frac{\varepsilon^2}{n^{-1/2}+\varepsilon}\bigg\} = o(1),$$

as $n\to\infty$. Now the proof of (6.28) is complete.

Considering that here we have $\sum_{h=1}^{\infty}\gamma_h < \infty$, the proof of (6.29) will follow from the following two statements: as $n\to\infty$,

$$n\sum_{i=1}^{n}\Big(\frac{w_i^{(n)}}{n}-\frac{1}{n}\Big)^2-(1-1/n) = o_{P_w}(1) \tag{6.30}$$

and

$$n\sum_{h=1}^{n-1}\gamma_h\sum_{j=1}^{n-h}\Big(\frac{w_j^{(n)}}{n}-\frac{1}{n}\Big)\Big(\frac{w_{j+h}^{(n)}}{n}-\frac{1}{n}\Big) = o_{P_w}(1). \tag{6.31}$$

To prove (6.30), with $\varepsilon > 0$, we first use Chebyshev's inequality followed by some algebra involving the use of the moment generating function of the multinomial distribution to arrive at

$$P_w\bigg(\bigg|n\sum_{i=1}^{n}\Big(\frac{w_i^{(n)}}{n}-\frac{1}{n}\Big)^2-(1-1/n)\bigg| > \varepsilon\bigg)$$
$$\le \varepsilon^{-2}n^2 E_w\bigg(\sum_{i=1}^{n}\Big(\frac{w_i^{(n)}}{n}-\frac{1}{n}\Big)^2-\frac{(1-1/n)}{n}\bigg)^2$$
$$\le \varepsilon^{-2}n^2\Big(1-\frac{1}{n}\Big)^{-2}\bigg\{\frac{1-1/n}{n^6}+\frac{(1-1/n)^4}{n^3}+\frac{(n-1)(1-1/n)^2}{n^4}+\frac{4(n-1)}{n^4}+\frac{1}{n^2}$$
$$-\frac{1}{n^3}+\frac{n-1}{n^6}+\frac{4(n-1)}{n^5}-\frac{(1-1/n)^2}{n^2}\bigg\}\longrightarrow 0 \qquad \text{as } n\to\infty.$$

The latter completes the proof of (6.30).

In order to establish (6.31), with $\varepsilon > 0$, we write

$$
\begin{aligned}
&P_w\left(n\left|\sum_{h=1}^{n}\gamma_h\sum_{j=1}^{n-h}\left(\frac{w_j^{(n)}}{n}-\frac{1}{n}\right)\left(\frac{w_{j+h}^{(n)}}{n}-\frac{1}{n}\right)\right|>\varepsilon\right)\\
&\quad\le\varepsilon^{-1}\sum_{h=1}^{n-1}|\gamma_h|E_w^{1/2}\left(n\sum_{j=1}^{n-h}\left(\frac{w_j^{(n)}}{n}-\frac{1}{n}\right)\left(\frac{w_{j+h}^{(n)}}{n}-\frac{1}{n}\right)\right)^2.
\end{aligned}
\tag{6.32}
$$

Observe now that

$$
\begin{aligned}
&E_w\left(n\sum_{j=1}^{n-h}\left(\frac{w_1^{(n)}}{n}-\frac{1}{n}\right)\left(\frac{w_2^{(n)}}{n}-\frac{1}{n}\right)\right)^2\\
&=n^2(n-h)E_w\left(\left(\frac{w_1^{(n)}}{n}-\frac{1}{n}\right)\left(\frac{w_2^{(n)}}{n}-\frac{1}{n}\right)\right)^2\\
&\quad+n^2(n-h)(n-h-1)\\
&\quad\times E_w\left(\left(\frac{w_1^{(n)}}{n}-\frac{1}{n}\right)\left(\frac{w_2^{(n)}}{n}-\frac{1}{n}\right)\left(\frac{w_3^{(n)}}{n}-\frac{1}{n}\right)\left(\frac{w_4^{(n)}}{n}-\frac{1}{n}\right)\right)\\
&\le n^3O\left(n^{-4}\right)+n^3O\left(n^{-6}\right)\to 0\qquad\text{as } n\to\infty.
\end{aligned}
$$

The preceding conclusion is true, since $E_w((\frac{w_1^{(n)}}{n}-\frac{1}{n})(\frac{w_2^{(n)}}{n}-\frac{1}{n}))^2=O(n^{-4})$ and $E_w((\frac{w_1^{(n)}}{n}-\frac{1}{n})(\frac{w_2^{(n)}}{n}-\frac{1}{n})(\frac{w_3^{(n)}}{n}-\frac{1}{n})(\frac{w_4^{(n)}}{n}-\frac{1}{n}))=O(n^{-6})$. Incorporating now the latter two results into (6.32), the conclusion (6.31) follows. Now the proof of Theorem 5.1 is complete.

Proof of Theorem 5.2

In order to prove Theorem 5.2, using a Slutsky type argument, it suffices to show that the Studentizing sequence of $T_n^{*^{\text{stu}}}$, asymptotically in n, in a hierarchical way, coincides with the right normalizing sequence, that is, with the one in the denominator of T_n^* defined in (5.4).

Considering that, as $n\to\infty$, we have that $\bar{\gamma}_0-\gamma_0=o_{P_X}(1)$, where $0<\gamma_0<\infty$, the proof of this theorem follows if, for $\varepsilon_1, \varepsilon_2>0$, we show that

$$
\begin{aligned}
&P_w\left\{P_{X|w}\left(\bar{\gamma}_0\left|n\sum_{i=1}^{n}\left(\frac{w_i^{(n)}}{n}-\frac{1}{n}\right)-(1-1/n)\right|>\varepsilon_1\right)>\varepsilon_2\right\}\\
&\quad=o(1),\qquad\text{as } n\to\infty.
\end{aligned}
$$

To establish the preceding relation, we note that its left-hand side is bounded above by

$$P_w\left\{E_X(\bar{\gamma}_0)\left(\left|n\sum_{i=1}^{n}\left(\frac{w_i^{(n)}}{n}-\frac{1}{n}\right)-(1-1/n)\right|\right)>\varepsilon_1\varepsilon_2\right\}$$
$$\leq P_w\left\{\left(\left|n\sum_{i=1}^{n}\left(\frac{w_i^{(n)}}{n}-\frac{1}{n}\right)-(1-1/n)\right|\right)>\frac{\varepsilon_1\varepsilon_2}{\gamma_0}\right\}.$$

The rest of the proof is similar to that of (6.30). Now the proof of this theorem is complete.

Acknowledgments

The authors wish to express their appreciation to the Editor and the Area Editor for their interest in, and careful attention to, this paper. We also wish to express our gratitude to the two referees and the Associate Editor for their careful reading of our original manuscript and for their instructive comments and suggestions that have led to improvements of our paper. In particular, their comments resulted in a significant improvement in our Theorem 2.1 and also in our numerical studies in Tables 1–6.

Work supported in part by an NSERC grant of Miklós Csörgő.

References

[1] Abadir, K.M., Distaso, W. and Giraitis, L. (2009). Two estimators of the long-run variance: Beyond short memory. *J. Econometrics* **150** 56–70. MR2525994

[2] Abadir, K.M., Distaso, W., Giraitis, L. and Koul, H.L. (2014). Asymptotic normality for weighted sums of linear processes. *Econometric Theory* **30** 252–284. MR3177798

[3] Beran, J. (1994). *Statistics for Long-Memory Processes*. *Monographs on Statistics and Applied Probability* **61**. New York: Chapman & Hall. MR1304490

[4] Brockwell, P.J. and Davis, R.A. (1991). *Time Series*: *Theory and Methods*, 2nd ed. New York: Springer.

[5] Csáki, E., Csörgő, M. and Kulik, R. (2013). Strong approximations for long memory sequences based partial sums, counting and their Vervaat processes. Available at arXiv:1302.3740.

[6] Csörgő, M. and Nasari, M.M. (2015). Inference from small and big data sets with error rates. *Electron. J. Stat.* **9** 535–566. MR3326134

[7] Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and Their Application*. *Cambridge Series in Statistical and Probabilistic Mathematics* **1**. Cambridge: Cambridge Univ. Press. MR1478673

[8] Dobrushin, R.L. and Major, P. (1979). Non-central limit theorems for nonlinear functionals of Gaussian fields. *Z. Wahrsch. Verw. Gebiete* **50** 27–52. MR0550122

[9] Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. *Monographs on Statistics and Applied Probability* **57**. New York: Chapman & Hall. MR1270903

[10] Giraitis, L., Kokoszka, P., Leipus, R. and Teyssière, G. (2003). Rescaled variance and related tests for long memory in volatility and levels. *J. Econometrics* **112** 265–294. MR1951145

[11] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. New York: Springer. MR1145237

[12] Härdle, W., Horwitz, J. and Kreiss, J.P. (2003). Bootstrap methods for time series. *Int*. *Stat*. *Rev*. **71** 435–459.
[13] Hasslet, J. and Raftery, A.E. (1989). Space–time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Applied Statistics* **38** 1–50.
[14] Kim, Y.M. and Nordman, D.J. (2011). Properties of a block bootstrap under long-range dependence. *Sankhya A* **73** 79–109. MR2887088
[15] Kreiss, J.-P. and Paparoditis, E. (2011). Bootstrap methods for dependent data: A review. *J*. *Korean Statist*. *Soc*. **40** 357–378. MR2906623
[16] Künsch, H. (1987). Statistical aspects of self-similar processes. In *Proceedings of the* 1*st World Congress of the Bernoulli Society*, *Vol*. 1 (*Tashkent*, 1986) 67–74. Utrecht: VNU Sci. Press. MR1092336
[17] Lahiri, S.N. (2003). *Resampling Methods for Dependent Data*. *Springer Series in Statistics*. New York: Springer. MR2001447
[18] Moulines, E. and Soulier, P. (2003). Semiparametric spectral estimation for fractional processes. In *Theory and Applications of Long-Range Dependence* (P. Dukhan, G. Oppenheim and M. Taqqu, eds.) *Theory an Applications of Long-Range Dependendence*. Birkhäsuer: Boston, MA. MR1956053
[19] 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. MR2200986
[20] Robinson, P.M. (1995). Gaussian semiparametric estimation of long range dependence. *Ann*. *Statist*. **23** 1630–1661. MR1370301
[21] Robinson, P.M. (1997). Large-sample inference for nonparametric regression with dependent errors. *Ann*. *Statist*. **25** 2054–2083. MR1474083
[22] Shao, J. and Tu, D.S. (1995). *The Jackknife and Bootstrap*. *Springer Series in Statistics*. New York: Springer. MR1351010
[23] Taqqu, M.S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z*. *Wahrsch*. *Verw*. *Gebiete* **31** 287–302. MR0400329

Received May 2014 and revised January 2016