Confidence intervals with higher accuracy for short and long-memory linear processes

Masoud M. Nasari1,2 · Mohamedou Ould-Haye1

Received: 20 June 2020 / Revised: 23 July 2021 / Accepted: 18 September 2021 / Published online: 19 November 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

In this paper an easy to implement method of stochastically weighing short and long-memory linear processes is introduced. The method renders asymptotically exact size confidence intervals for the population mean which are significantly more accurate than their classic counterparts for each fixed sample size $n$. It is illustrated both theoretically and numerically that the randomization framework of this paper produces randomized (asymptotic) pivotal quantities, for the mean, which admit central limit theorems with smaller magnitudes of error as compared to those of their leading classic counterparts. An Edgeworth expansion result for randomly weighted linear processes whose innovations do not necessarily satisfy the Cramer condition, is established. Numerical illustrations and applications to real world data are also included.

Keywords Accuracy of the CLT · Confidence intervals · Limit theorems · Edgeworth expansion · Linear processes · Long memory · Time series analysis

1 Introduction

In a nonparametric framework, the central limit theorem (CLT) is a critical tool in drawing inference about the mean of a population. The CLT validates the use of the percentiles of the normal distribution to approximate those of the unknown sampling distribution of a centered partial sum of a given set of data. The next important step
after establishing a CLT is to characterize the departure of the sampling distribution of the underlying partial sum functional from normality and to address the speed at which it vanishes. Expressed as a function of the sample size \( n \), and without restricting the distribution of the data to the symmetrical ones, the error of the CLT is generally known to be of order \( O(1/\sqrt{n}) \). Usually, the vanishing rate \( 1/\sqrt{n} \) is referred to as the first-order accuracy, correctness or efficiency of the CLT. The error of the CLT has been extensively studied by several authors, mainly using the Berry-Esseen inequality and the Edgeworth expansion. Both these methods show that in the i.i.d case, the CLT is usually first-order efficient [cf., e.g., Bhattacharya and Rao (1976), Senatov (1998) and Shorack (2000)]. The same error rate of \( 1/\sqrt{n} \) has also been shown to hold for some classes of weakly dependent data [cf., for example, Götze and Hipp (1983), Lahiri (1993)].

There are a number of methods in the literature that deal with the problem of how to increase the accuracy of the CLT. The tilted approximation [cf. Nielsen (1989)] is a method of reducing the error admitted by the CLT. In this method, which is of theoretical interest, the densities of the partial sums of an i.i.d. data set are approximated by some tilting expressions. The error of this approximation is of order \( O(1/n) \). The drawback of the tilted approximation is that computing the tilting expressions requires the full knowledge of the cumulant generating function of the data which is usually unknown. Also, in the context of i.i.d. data, some other error reduction techniques are built around adjusting the cut-off points, the usual percentiles of standard normal distribution, by some additive correcting factors [cf., for example, Hall (1983)]. This method is called the Edgeworth correction and it requires estimating values of \( r \)-th moments of the data, where \( r \leq k \), for some \( k \geq 3 \). The drawback of this method is that it tends to over-correct the coverage probability of the resulting confidence intervals, when the sample size is relatively small. This issue is the result of the significant deviation between the actual values of the moments of the population from their estimated values, when the sample is of small size. Another Edgeworth correction method was introduced in Eriksson (2008). The approach is essentially a randomization via adding a simulated set of values, from a specific class of distributions, to the original data. To specify the class of distributions from which the additive values are to be generated, it is assumed that the values of the second and third moments of the original data are known. Furthermore, Eriksson (2008) provides only one example of the distributions which can be used to generate the additive random values from. The approach is interesting in general, however, it is unsuitable for use in a nonparametric framework. The bootstrap [cf., for example, Efron (1979)] is another approach to reducing the error in approximating the sampling distribution that requires repeated re-sampling from a given set of data and, hence it can be intrinsically viewed as randomized jackknife. To preserve the dependent nature of time series data, bootstrap technique required some adjustments. Techniques such as block bootstrap, Sieve and augmented Sieve are examples of such adjustments [cf. for example, Lahiri (2003), Kim and Nordman (2011), Götze and Künsch (1996), Poskitt et al. (2015)]. In case of short-memory processes, these adjusted bootstrap techniques seem to outperform the classic CLT in terms of producing confidence intervals with better coverage probability closer to the target. However, when dealing with long-memory data, these various adapted for dependant data versions of the bootstrap suffer from a significant underestimation
of the variance which in turn translates in producing confidence intervals with poor coverage probability even for samples with moderate sizes (cf. numerical studies in Sect. 5).

In this paper we introduce a method to stochastically weighing a given set of data in a multiplicative way that results in more accurate CLTs for linear processes. This approach creates a framework that requires neither re-sampling nor extra knowledge on the distribution of the data except for the same commonly assumed moment conditions for the classic CLT.

The idea of boosting the accuracy of the CLT via creating randomized versions of the (asymptotic) pivots of interest, that continue to possess the pivotal property for the parameter of interest, was first explored in Csörgő and Nasari (2015) for i.i.d. data and in Csörgő et al. (2017) for linear processes. The viewpoint in the latter papers is to create pivots that admit CLTs with smaller magnitude of error thorough replacing a data set by a randomly weighted version of it, where the multiplicative weights are of the form of functionals of symmetric multinomial random variables. A broader randomization framework for i.i.d. data was introduced in Nasari (2016), which is flexible in the sense that it allows choosing the randomizing weights from a virtually unlimited class of random weights, including multinomial weights, with a window parameter to regulate the trade-off between the accuracy of the CLT and the volume of the resulting randomized confidence regions for the mean of the original data. The present paper constitutes a generalization of the randomization approach introduced in Nasari (2016) for i.i.d. data to the case when the data form a short or long-memory linear process.

As will be seen in this paper, introducing a controlled extra source of randomness in conjunction with the random mechanism that produces the original data can enhance the accuracy of the CLT. The refinement in the accuracy results from multiplying the original data by appropriately chosen random factors which do not change the nature of the given data set. More precisely, the multiplicative stochastic weights used in this paper are so that, if the original data set is of short-memory then so is their randomized version. The same continues to hold true for randomized long-memory linear processes. Moreover, in the case of long-memory linear processes, the randomized pivots introduced in this paper tend to yield significantly better probabilities of coverage as compared to a number of the existing bootstrap methods including the Sieve and the augmented Sieve (cf. Sect. 5).

We conclude this section by an outline of the topics addressed in the rest of the paper. In Sect. 2 we introduce a class of randomized pivotal quantities for the value of the common mean of the population from which a linear process structured data set of size \( n \geq 1 \) is drawn. In Sect. 3 we derive a general Edgeworth expansion result for a class of randomly weighted linear processes which, in the univariate case, is an extension of the Edgeworth expansion obtained in Götzte and Hipp (1983) for sums of weakly dependent random vectors. This result, that is also of independent interest, is then used to illustrate the error reduction effect of the randomization framework in Sect. 2 on the CLT of the randomized pivots introduced in it. Section 4 is devoted to establishing the asymptotic validity of the randomized pivots introduced in Sect. 2 via deriving a CLT for them under more relaxed conditions than those required for the Edgeworth expansion result in Sect. 3. Section 4 also contains some discussions on
the resulting randomized confidence intervals and some numerical demonstrations of their improved performance over that of their classic non-randomized counterparts. A numerical comparison of the performances of some methods of bootstrap and the randomization scheme introduced in this paper is presented in Sect. 5 for long-memory data. In Sect. 6, based on historical data, randomized confidence intervals are constructed for the true mean of the producer price index and consumer confidence index of Japan and Sweden. Further results on Studentization of the randomized pivotal quantities are given in Sect. 7. Some concluding remarks are given in Sect. 8. The proofs are given in Appendix.

2 Skewness reduction

In this section we discuss the shortcoming of the leading measures of skewness which are defined primarily for i.i.d. data, when dealing with manipulations (functionals) of dependent data. Then, we consider a generalization of the criterion of skewness so that it is suitable for both stationary and i.i.d. data. This generalized definition of skewness is then used to introduce our approach to increase the accuracy of the CLT for linear processes.

When the sampling distribution of the normalized partial sums of a set of $n, n \geq 1$, univariate i.i.d. random variables $Y_1, \ldots, Y_n$, admits an Edgeworth expansion, the Fisher–Pearson’s measure of skewness $E(Y_1 - E(Y_1))^3/(\text{Var}(Y_1))^{3/2}$ becomes the coefficient of an expression of order $O(1/\sqrt{n})$, as $n \to \infty$. In other words, the measure of skewness is the coefficient of the slowest vanishing term of the error admitted by the CLT. This property motivates viewing the skewness as the most important characteristic of the distribution of the data in measuring the departure from normality. The closer the value of the skewness is to zero, the more symmetrical the sampling distribution of the underlying partial sum of a given data set will be. The commonly used measures of skewness, such as the aforementioned Fisher–Pearson, are defined for the marginal distribution of an i.i.d. set of data with a finite third moment. This definition is suitable only for the i.i.d. case, as it does not account for the dependence when a stationary and dependent data of size $n$ form the summands of a partial sum which admits the CLT. The latter observation calls for a broader definition for skewness that is to take the partial sum of a set of random variables into consideration. This means that skewness should naturally be defined for the sampling distribution of an underlying partial sum functional rather than for the marginal distribution of its summands. The need for such a definition for skewness in the context of this paper stemmed from our need for a proper measure of skewness when studying the problem of boosting the accuracy of pivotal quantities of the form of partial sums for the mean of linear processes. Prior to defining an appropriate measure of skewness for stationary linear processes, we first set up the definition of linear processes of consideration. To formally state our results in this section, and also for the use throughout this paper, we let $X_1, \ldots, X_n$ be the first $n \geq 1$ terms of the linear process $\{X_t : t \geq 1\}$ defined as
where \( \mu, \mu \in \mathbb{R} \), is the mean of the process, \( \{a_k; k \geq 0\} \) 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. white noise innovations with \( \mathbb{E}\zeta_1 = 0, 0 < \sigma^2 := \text{Var}(\zeta_1) < \infty \) and \( \eta := \mathbb{E}\zeta_1^3 \). For throughout use in this paper, we denote the covariance function of the process \( X_t \) as

\[
\gamma_h := \text{Cov}(X_s, X_{s+h}) = \mathbb{E}[(X_s - \mu)(X_{s+h} - \mu)], \quad h \geq 0, \ s \geq 1.
\] (2.2)

In this paper we consider two types of the linear process (2.1). (i) when \( \sum_{k=0}^{\infty} |a_k| = \infty \). In this case we refer to \( X_t \) as a long-memory linear process. In particular, we consider the case when, as \( k \to \infty \), for some \( c > 0 \) we have \( a_k \sim ck^{d-1} \), where \( 0 < d < 1/2 \). We refer to \( d \) as the memory parameter. The other type of linear processes considered in this paper is (ii) when \( \sum_{k=0}^{\infty} |a_k| < \infty \). In this case we refer to \( X_t \) as a short-memory linear process. To unify our notation we define the memory parameter \( d \) for short-memory linear processes as \( d = 0 \).

Consider the classic pivotal quantity for \( \mu, T_{n,X} : \mathbb{R}^{n+1} \to \mathbb{R} \), that is a normalized partial sum functional, defined as

\[
T_{n,X} = \sum_{i=1}^{n} (X_i - \mu)/(\text{Var}(\sum_{i=1}^{n} X_i))^{1/2}.
\] (2.3)

When \( \mathbb{E}|X_1|^3 < \infty \), we define the skewness measure of the pivotal quantity \( T_{n,X} \) as follows:

\[
\beta_{T_{n,X}} := \mathbb{E}(T_{n,X})^3 = \mathbb{E}\left(\sum_{i=1}^{n} (X_i - \mu)\right)^3/(\text{Var}(\sum_{i=1}^{n} X_i))^{3/2}.
\] (2.4)

In view of the stationarity of \( X_t \), and without loss of generality, we assume that \( \mu = 0 \) and expand the skewness measure \( \beta_{T_{n,X}} \) as follows:

\[
\beta_{T_{n,X}} = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right)^3/(\text{Var}(\sum_{i=1}^{n} X_i))^{3/2}
= \mathbb{E}(X_1^3)/(\sqrt{n}(\gamma_0 + 2 \sum_{h=1}^{n} (1 - h/n)\gamma_h)^{3/2})
+ 3\left(\sum_{h=1}^{n} (1 - h/n)(\mathbb{E}(X_1^2X_{1+h}) + \mathbb{E}(X_1X_{1+h}^2))\right)/(\sqrt{n}(\gamma_0)
+ 2 \sum_{h=1}^{n} (1 - h/n)\gamma_h)^{3/2})
\]
\[ + 6 \left( \sum_{h=1}^{n-1} \sum_{h'=1}^{n-h} \frac{(1 - h + h')}{n} \mathbb{E}(X_1 X_{1+h} X_{1+h+h'}) \right) / \left( \sqrt{n} \gamma_0 \right) + 2 \left( \sum_{h=1}^{n} \left( 1 - \frac{h}{n} \right) \right) \gamma h \right)^{3/2}. \]  

(2.5)

It is shown in Nasari (2021) that, for \( 0 \leq d < 1/2 \), as \( n \to \infty \),

\[ \sqrt{n} \beta_{T_n,X} \to k(d) := \frac{n}{\sigma^3} \left( \frac{1}{1+2d} + \int_0^\infty \left( (1+x)^d - x^d \right)^3 \, dx \right)^{3/2}. \]  

(2.6)

From the preceding result one can see that the skewness of sums of short and long-memory linear processes asymptotically vanishes at the same convergence rate \( \sqrt{n} \) as that of sums of i.i.d. data. The effect of the short or long range dependence appears only in terms of \( d \) in the limiting constant \( k(d) \) in (2.6) which, surprisingly enough, (in absolute value) is decreasing in \( d \in [0, 1/2) \), with \( k(0) = \eta/\sigma^3 \) and \( k(1/2) = 0 \). For more detailed discussion we refer to the aforementioned paper Nasari (2021).

We now introduce a randomization approach to construct more symmetrical, and hence more accurate (cf. Theorem 1), versions of \( T_n,X \), based on the data set \( X_1, \ldots, X_n \), as in (2.1), as follows:

\[ T_{n,X,w}(\theta_n) = \left( \sum_{i=1}^{n} (w_i - \theta_n)(X_i - \mu) \right) \sqrt{n} \mathcal{D}_{n,X,w}, \]  

(2.7)

where,

\[ \mathcal{D}_{n,X,w} := \text{Var} \left( \sum_{i=1}^{n} (w_i - \theta_n) X_i \right) / n \]

\[ = \mathbb{E}(w_1 - \theta_n)^2 \gamma_0 + 2 \mathbb{E}((w_1 - \theta_n)(w_2 - \theta_n)) \sum_{h=1}^{n} (1 - \frac{h}{n}) \gamma h. \]  

(2.8)

The real valued constant \( \theta_n \), to which we shall refer as the window constant, and the random weights \( w_i \) used in the definition of the randomized pivot \( T_{n,X,w}(\theta_n) \), as in (2.7), are to be determined in view of the following scenario, to which we shall refer as the scheme (RS).

### 2.1 The randomization scheme (RS)

Let \( w_1, \ldots, w_n \) be either i.i.d. random variables with \( \mathbb{E}|w_1|^3 < \infty \) or have a symmetric multinomial distribution, i.e., \( \text{Multinomial}(n; 1/n, \ldots, 1/n) \). Furthermore, we assume \( w_i \) are independent from the data \( X_i, 1 \leq i \leq n \). Choose \( \theta_n \) such that \( \theta_n \neq \mathbb{E}(w_1) \) and \( |\beta_{T_{n,X,w}(\theta_n)}| < |\beta_{T_{n,X}}| \), where \( \beta_{T_{n,X,w}(\theta_n)} \) is the skewness of the
partial sum of the randomized data, i.e.,

$$\beta_{T_n,X,w}(\theta_n) := \mathbb{E}\left(\sum_{i=1}^{n}(w_i - \theta_n)X_i\right)^{3/2} / \left(\text{Var}_{X,w}\left(\sum_{i=1}^{n}(w_i - \theta_n)X_i\right)\right)^{3/2}$$  \hspace{1cm} (2.9)

Clearly choosing the window $\theta_n$ according to the randomization scheme (RS) makes the randomized partial sums more symmetrical than their non-randomized counterparts. The effect of this symmetrization on the accuracy of the CLT for $T_n, X, w(\theta_n)$, as in (2.7), is discussed in Sect. 3.

It is important to note that in the context of the scheme (RS) the skewness of the partial sum of a linear process can be reduced while preserving the original dependence structure of the data. This is ensured by excluding the mean of the random weights from the choices of the window constant. Such exclusion also yields shrinking confidence intervals, cf. Remark 2.

### 3 The effect of scheme (RS) on the error of the CLT for randomized linear processes

We use the Edgeworth expansion for a class of linear processes to illustrate the refinement that results from our randomization approach (RS), as in Sect. 2.1, which is primarily designed to reduce $\beta_{T_n,X}$, the skewness of the classic pivot $T_n, X$, as in (2.5). The choice of the Edgeworth expansion is due to the fact that it provides an explicit and direct link between the skewness of a partial sum pivot and the error admitted by its CLT.

The following Theorem 1 gives an Edgeworth expansion for the sampling distribution of partial sums of the random variables $(w_i - \theta_n)(X_i - \mu)$, $1 \leq i \leq n$, where $X_1, \ldots, X_n$ are data from some short-memory linear processes as in (2.1).

**Theorem 1** Assume that the data $X_1, \ldots, X_n$ are as in (2.1). Let the weights $w_1, w_2, \ldots$, be i.i.d. random variables, that are independent from the data. Also, let $\{\theta_n\}$ be a converging sequence of constants such that, $\lim_{n \to \infty} \theta_n \neq \mathbb{E}(w_1)$. Furthermore, assume that

(i) either $\sum_{k=0}^{\infty} a_k(w_{k+1} - \theta_n) \neq 0$, a.s. for all $n \geq 1$, and $\lim_{u \to \infty} \sup_{\theta_n} \left|\mathbb{E}(e^{iu\xi_1})\right| < 1$, or

$X_1 \neq 0$ a.s., and $\lim_{u \to \infty} \sup_{\theta_n} \left|\mathbb{E}(e^{iuw_1})\right| < 1$,

(ii) There exists $\ell > 0$ such that $|a_k| \leq (1/\ell)e^{-\ell k}$, for all $k = 0, 1, \ldots$,

(iii) $\mathbb{E}(|\xi_1|^4 < \infty$, $\mathbb{E}|w_1|^4 < \infty$.

Then, we have for all $x \in \mathbb{R}$,

$$P\left(T_{n,X,w}(\theta_n) \leq x\right) - \Phi(x) = \beta_{T_n,X,w}(\theta_n)H(x) + o(n^{-1/2}),$$  \hspace{1cm} (3.1)

where $T_{n,X,w}(\theta)$ and $\beta_{T_n,X,w}(\theta)$ are, respectively, as in (2.7) and (2.9), $\Phi$ is the distribution function of the standard normal distribution and $H$ is a polynomial of degree 3.
We note that the conditions \( \limsup_{u \to \infty} |\mathbb{E}(e^{iuw})| < 1 \) and \((iv)\) of Theorem 1, pose no practical restriction on the choice of the random weights, as the class of random weights that satisfy these conditions is virtually unlimited.

In view of the relation (3.1), one can see that, while maintaining the same magnitude of error as that of the CLT of \( T_{n,X} \), for each fixed \( n \), we can improve on the accuracy of the CLT by reducing the skewness coefficient using the randomization scheme (RS).

We also note that when the linear process \( X_t \), as in (2.1), is so that its innovations satisfy the Cramer condition, i.e., \( \limsup_{u \to \infty} |\mathbb{E}(e^{iu\xi_1})| < 1 \), and also it satisfies conditions (ii)–(iii) of Theorem 1 as well as the condition

\[
(i') \sum_{k=0}^{\infty} a_k \neq 0,
\]

then, Theorem (2.8) of Götze and Hipp (1983) and its Corollary (2.9), imply that, for all \( x \in \mathbb{R} \), the sampling distribution of \( T_{n,X} \), as in (2.3), admits the following Edgeworth expansion:

\[
P\left(T_{n,X} \leq x\right) - \Phi(x) = \beta_{T_{n,X}} H(x) + o(n^{-1/2}),
\]

(3.2)

where \( \beta_{T_{n,X}} \), as in (2.5), is the skewness of \( T_{n,X} \).

To compare the Edgeworth expansion (3.1) to (3.2), we consider linear processes whose innovations \( \xi_j \) satisfy the Cramer condition and \( \sum_{k=0}^{\infty} a_k \neq 0 \). We also consider weights that are non-degenerate i.i.d. and continuous with a finite fourth moment. In this case the first part of condition (i) in Theorem 1 holds true. Now under the conditions (ii), (iii), and (iv), one can see that (3.1) yields a smaller error when \( \beta_{T_{n,X},w}(\theta_n) \) is small.

**Remark 1** It is worth noting that when constructed using some continuous random weights, so that they satisfy the conventional conditions (iii) and (iv) of Theorem 1, the randomized pivot \( T_{n,X,w}(\theta) \) admits the Edgeworth expansion (3.1) for the linear processes (2.1) even when \( \sum_{k=0}^{\infty} a_k = 0 \), provided that (ii) holds true. This is true, since, in the case that the random weights are continuous, the first part of condition (i) of Theorem 1 holds true even when \( \sum_{k=0}^{\infty} a_k = 0 \). In contrast, the Edgewroth expansion (3.2) for the classic \( T_{n,X} \), as in (2.3), is valid only for short-memory linear processes (2.1) for which we have \( \sum_{k=0}^{\infty} a_k \neq 0 \), provided that the innovation satisfy the Cramer condition and, conditions (ii) and (iii) of Theorem 1 also hold true.

### 4 CLT for randomized linear processes

In this section we establish a CLT for the randomized pivot \( T_{n,X,w}(\theta_n) \), as in (2.7), under less stringent conditions than those assumed in Theorem 1. The CLT in this section is valid for the random weights and window constants as characterized in the randomization scheme (RS) in Sect. 2.1.

**Theorem 2** Let \( X_1, \ldots, X_n \) be the first \( n \) terms of the linear process (2.1) and consider the randomized pivot \( T_{n,X,w}(\theta_n) \), as in (2.7), with the weights \( w_1, \ldots, w_n \) and the
window constant \( \theta_n \) be as specified in the scheme (RS). Then, as \( n \to \infty \), we have for all \( x \in \mathbb{R} \):

\[
P(T_n, X, w(\theta_n) \leq x \mid w_1, \ldots, w_n) \longrightarrow \Phi(x) \quad \text{in probability} \quad (4.1)
\]

and, consequently,

\[
P(T_n, X, w(\theta_n) \leq x) \longrightarrow \Phi(x). \quad (4.2)
\]

### 4.1 Studentization

We introduce the following \( G_{n,X,w}(\theta_n, d) \) as a Studentized version of \( T_{n,X,w}(\theta_n) \) that is valid for short and long-memory processes as specified in Theorem 3.

\[
G_{n,X,w}(\theta_n, d) = n^{-1/2-d} \left( \sum_{i=1}^{n} (w_i - \theta_n)(X_i - \mu) \right) / \sqrt{q^{-2d} S_{n,q,w}^2}, \quad (4.3)
\]

where \( q = O(n^{1/2}) \),

\[
S_{n,q,w}^2 = \mathbb{E}(w_1 - \theta_n)^2 \tilde{\gamma}_0 + \mathbb{E} \left( (w_1 - \theta_n)(w_2 - \theta_n) \right) \sum_{h=1}^{q} \tilde{\gamma}_h \left( 1 - \frac{h}{q} \right), \quad (4.4)
\]

and \( \tilde{\gamma}_s := \sum_{j=1}^{n-s} (X_j - \bar{X}_n)(X_{j+s} - \bar{X}_n)/n, \ 0 \leq s \leq n - 1 \), are the sample autocovariances.

The Studentized pivot \( G_{n,X,w}(\theta_n, d) \) can also be viewed as a randomized version of the Studentized pivotal quantity

\[
T_n^{stu}(d) := n^{1/2-d} (\bar{X}_n - \mu) \sqrt{(q^{-2d}(\tilde{\gamma}_0 + 2 \sum_{h=1}^{q} \tilde{\gamma}_h(1 - h/q)))}^{1/2}. \quad (4.5)
\]

Our Studentizing sequence \( q^{-2d} S_{n,q,w}^2 \) for \( G_{n,X,w}(\theta_n, d) \) is a generalization of \( q^{-2d}(\tilde{\gamma}_0 + 2 \sum_{h=1}^{q} \tilde{\gamma}_h(1 - h/q)) \), the Studentizing sequence for the non-randomized \( T_n^{stu}(d) \). The long run variance estimator \( q^{-2d}(\tilde{\gamma}_0 + 2 \sum_{h=1}^{q} \tilde{\gamma}_h(1 - h/q)) \) (cf. Abadir et al. (2009)) is an adaptation of Bartlett-kernel heteroscedasticity and autocorrelation consistent (HAC) estimator (see, for example, Andrews (1991)) that allows for long-memory. Although, in this paper attention is restricted to our HAC-type Studentizing sequences of the form \( q^{-2d} S_{n,q,w}^2 \), we remark that it is also desirable to use other existing robust sample based estimators such as Robinson’s periodogram based memory and autocorrelation consistent (MAC) estimator [cf. Robinson (2005)], to construct Studentizing sequences for the randomized sum \( n^{-1/2-d} \left( \sum_{i=1}^{n} (w_i - \theta_n)(X_i - \mu) \right) \).

In what follows the notation \( G_{n,X,w}(\theta_n, \hat{d}) \) in the case of a long-memory process will stand for \( G_{n,X,w}(\theta_n, d) \), where the memory parameter \( d \) is replaced by its sample based estimate \( \hat{d} \) as specified in the following Theorem 3 which establishes the CLT for \( G_{n,X,w}(\theta_n, d) \) and it reads as follows.
Theorem 3  Consider $X_1, \ldots, X_n$, the first $n$ terms of the linear process (2.1), and let the weights $w_1, \ldots, w_n$ and the window constants $\theta_n$ be as in Theorem 2.

(A) If the linear process (2.1) is of short-memory, i.e., $\sum_{k=0}^{\infty} |a_k| < \infty$, and $\mathbb{E} \zeta_1^4 < \infty$, then, as $n, q \to \infty$ such that $q = O(n^{1/2})$, we have, for all $x \in \mathbb{R}$,

$$P \left( G_{n,X,w}(\theta_n, 0) \leq x \mid w_1, \ldots, w_n \right) \to \Phi(x) \text{ in probability,}$$

and, consequently,

$$P \left( G_{n,X,w}(\theta_n, 0) \leq x \right) \to \Phi(x), \ t \in \mathbb{R}.$$ 

(B) Let the linear process (2.1) be of long-memory such that $\mathbb{E} \zeta_1^4 < \infty$ and, as $k \to \infty$, $a_k \sim c k^{d-1}$, 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 $x \in \mathbb{R}$,

$$P \left( G_{n,X,w}(\theta_n, d) \leq x \mid w_1, \ldots, w_n \right) \to \Phi(x) \text{ in probability,}$$

and, consequently,

$$P \left( G_{n,X,w}(\theta_n, d) \leq x \right) \to \Phi(x), \ x \in \mathbb{R},$$

$$P \left( G_{n,X,w}(\theta_n, \hat{d}) \leq x \right) \to \Phi(x), \ x \in \mathbb{R},$$

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

When the linear process in Theorem 3 is of long-memory with memory parameter $d$, there are a number of estimators $\hat{d}$ in the literature that can be used to estimate $d$. The MLE of $d$, with the Hasslet and Raftery (1989) method used to approximate the likelihood and, Whittle estimator [cf. Künsch (1987) and Robinson (1995)] are two examples of the estimators of the memory parameter $d$. For more on estimators for the memory parameter and their asymptotic behavior, we refer to Bhansali and Kokoszka (2001), Moulines and Soulier (2003) and references therein.

4.2 Randomized confidence intervals

By virtue of Theorem 3, for the parameter of interest $\mu$, the mean of the linear process (2.1), the Studentized randomized pivotal quantity $G_{n,X,w}(\theta_n, d)$, as in (4.3), yields asymptotically exact size randomized confidence intervals with nominal size $100(1 - \alpha)\%$, $0 < \alpha < 1$, which are of the form:

$$I_{n,X,w}(\theta_n) := \left[ \min \{ A_{n,X,w,\alpha}(\theta_n), B_{n,X,w,\alpha}(\theta_n) \}, \max \{ A_{n,X,w,\alpha}(\theta_n), B_{n,X,w,\alpha}(\theta_n) \} \right],$$

(4.6)
where

\[ A_{n,X,w,\alpha}(\theta_n) = \left( \sum_{i=1}^{n} (w_i - \theta_n) X_i - z_{\alpha/2} n^{1/2+d} \sqrt{q^{-2d} S_{n,q,w}^2} \right) / \left( \sum_{i=1}^{n} (w_i - \theta_n) \right) \]

\[ B_{n,X,w,\alpha}(\theta_n) = \left( \sum_{i=1}^{n} (w_i - \theta_n) X_i + z_{\alpha/2} n^{1/2+d} \sqrt{q^{-2d} S_{n,q,w}^2} \right) / \left( \sum_{i=1}^{n} (w_i - \theta_n) \right), \]

and \( z_{\alpha/2} \) is the 100(1 - \( \alpha/2 \))-th percentile of the standard normal distribution. The length of the randomized confidence interval \( I_{n,X,w}(\theta_n) \) is

\[ \text{length}(I_{n,X,w}(\theta_n)) = 2z_{\alpha/2} n^{1/2+d} \sqrt{q^{-2d} S_{n,q,w}^2} / \left| \sum_{i=1}^{n} (w_i - \theta_n) \right|. \] (4.7)

**Remark 2** Along the lines of the proof of Theorem 3 in Appendix it is shown that for

\[ 0 \leq d < 1/2, q^{-2d} S_{n,q,w}^2 \text{ asymptotically coincides with } \text{Var}(n^{-1/2-d} \sum_{i=1}^{n} (w_i - \theta_n) X_i). \]

In view of the latter asymptotic equivalence together with (4.7), one can see that when \( \theta_n = \mathbb{E}(w_1) \), the length of the randomized confidence interval \( I_{n,X,w}(\theta_n) \) fails to vanish as \( n, q \to \infty \) such that \( q = O(n^{1/2}) \). The window constant \( \theta_n \) cannot be taken to be equal to \( \mathbb{E}(w_1) \) in the framework of the scheme (RS), as in Sect. 2.1. Therefore, (RS) results in randomized confidence intervals whose lengths shrink to zero as \( n, q \to \infty \) such that \( q = O(n^{1/2}) \).

Tables 1, 2 and 3 below provide an empirical comparison of the performance of the randomized confidence interval \( I_{n,X,w}(\theta_n) \), as in (4.6), and that of

\[ I_{n,X} = \tilde{X}_n \pm z_{\alpha/2} n^{-1/2+d} \sqrt{q^{-2d} \sum_{h=1}^{q} \tilde{r}_h (1 - h/q)}, \] (4.8)

which is constructed based on the classic \( T_n^{stu}(d) \), as in (4.5). The lag-length or the bandwidth \( q \) was chosen based on relation (2.14) of Abadir et al. (2009) in each case. More precisely, in Tables 1–3 we let \( q \) be ceiling \((n^{1/3})\) for the examined short-memory linear process.

It is easy to see that the length of \( I_{n,X} \) has the form:

\[ \text{length}(I_{n,X}) = 2z_{\alpha/2} n^{-1/2+d} \sqrt{q^{-2d} \sum_{h=1}^{q} \tilde{r}_h (1 - h/q)}. \] (4.9)

The results in Tables 1–3 are at the nominal level of 95% with \( z_{\alpha/2} = 1.96 \), and based on 2000 replications of the therein specified short-memory ARMA model with (heavily skewed) lognormal \((0, 1)\) innovations \( \xi \). That is, in Table 1, AR(1): \( X_t = \phi X_{t-1} + \xi_t \), in Table 2, AR(2): \( X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \xi_t \), and in Table 3 ARMA \((1, 1)\): \( X_t = \phi X_{t-1} + \delta \xi_{t-1} + \xi_t \). For the randomized confidence intervals in

\( \mathbb{C} \) Springer
Table 1: AR(1): $X_t = 0.8X_{t-1} + \zeta_t$

| $n$ | Skewness of pivot | Length | Coverage | Weights | $\theta_n$ | $I_{n,X,w}(\theta_n)$ | $I_{n,X}$ |
|-----|--------------------|--------|----------|---------|------------|------------------------|-----------|
| 1000 | -3.43              | 1.0    | 0.933    | Multinomial(n;1/n,...,1/n) | 1+0.23 | NA | 4.22 |
| 2000 | -3.17              | 0.87   | 0.947    | Multinomial(n;1/n,...,1/n) | 1+0.18 | NA | 4.74 |
Tables 1–3, 2000 sets of therein specified random weights also generated simultaneously with the data.

The effect of the incorporation of the window constants $\theta_n$ on increasing the accuracy of the randomized confidence intervals $I_{n,X,w}(\theta_n)$ is illustrated in Tables 1–3. This is done by including the estimated values of the *scaled* skewness coefficients of the classic pivot $T_{n,X}$, as in (2.3) and its randomized counterpart $T_{n,X,w}(\theta_n)$, as defined in (2.7). The estimators of the skewness are introduced in Nasari (2021). It is shown in the just mentioned paper that the rate at which the skewness of partial sums of short and long-memory processes, with non-symmetric innovations, vanish equals the square root of the sample size. The scaled skewness of the pivot $T_{n,X}$ is estimated as

$$\sqrt{n\hat{\beta}_{T,n,X}} := \frac{S_3(X)}{q^{-2d}\left(\bar{y}_0 + 2\sum_{h=1}^{q}(1 - h/q)\bar{y}_h\right)^{3/2}},$$

with

$$S_3(X) := q^{-3d}\left(\bar{\Delta}(0) + 3\sum_{h=1}^{q}\left(1 - \frac{h}{q}\right)\bar{\Delta}(h) + 6\sum_{h=1}^{q-1}\sum_{h'=1}^{q-h}\left(1 - \frac{h + h'}{q}\right)\bar{\Delta}(h, h')\right)$$

and the scaled skewness of the randomized pivot $T_{n,X,w}(\theta)$ is estimated as

$$\sqrt{n\hat{\beta}_{T,n,X,w}(\theta)} := \frac{S_3(X, w)}{q^{-2d}\left(S^2_{n,q,w}\right)^{3/2}},$$
Table 3  ARMA(1,1): $X_t = 0.8X_{t-1} + 0.8\zeta_{t-1} + \zeta_t$

| n     | Skewness of pivot | Length | Coverage | Weights               | $I_{n,X,w}(\theta_n)$ | $I_{n,X}$ |
|-------|-------------------|--------|----------|-----------------------|------------------------|-----------|
| 1000  |                   | 1.73   | 0.931    | Multinomial(n;1/n,...,1/n) | NA                     | NA        |
|       |                   |        |          | $\theta_n$            | 1+0.23                 | NA        |
| 2000  |                   | 1.52   | 0.943    | Multinomial(n;1/n,...,1/n) | NA                     | NA        |
|       |                   |        |          | $\theta_n$            | 1+0.18                 | NA        |
where $S_{n,q,w}^2$ is as defined in (4.4) and

\[
S_3(X, w) := q^{-3d} \left[ \text{E}(w_1 - \theta)^3 \Delta(0) + 3 \text{E}((w_1 - \theta)^2(w_2 - \theta)) \sum_{h=1}^q \left( 1 - \frac{h}{q} \right) \Delta(h) \right.
\]

\[+ 6 \text{E}((w_1 - \theta)(w_2 - \theta)(w_3 - \theta)) \sum_{h=1}^q \sum_{h'=1}^q \left( 1 - \frac{h+h'}{q} \right) \Delta(h, h') \right],
\]

where, $\Delta(h)$ and $\Delta(h, h')$ are third order sample covariances defined as

\[
\Delta(h) := \frac{1}{n} \sum_{j=1}^{n-h} \left( (X_j - \bar{X})^2(X_{j+h} - \bar{X}) + (X_j - \bar{X})(X_{j+h} - \bar{X})^2 \right),
\]

\[
\Delta(h, h') := \frac{1}{n} \sum_{j=1}^{n-h-h'} (X_j - \bar{X})(X_{j+h} - \bar{X})(X_{j+h+h'} - \bar{X}).
\]

**Remark 3** In the numerical analyses presented in Tables 1–3 the bandwidths $q_n = q$, used to estimate the scaled skewnesses $\sqrt{n} \hat{\beta}_T$, $X$ and $\sqrt{n} \hat{\beta}_T(X, w) (\theta)$, are such that, as $n \to \infty$, $q \to \infty$ and $q = o(\sqrt{n})$ [cf. Nasari (2021)].

### 5 Comparison to bootstrap confidence intervals for long-memory linear processes

In the context of long-memory data, the residual resampling method, known as the Sieve bootstrap [cf., e.g., Poskitt (2008)], is built around approximating a long-memory process by approximating its infinite auto regressive representation by a truncated one. However, this approximation produces error terms which tend to be significant (cf. Table 4). The approximation significantly underestimates the variance of the statistic of interest, which is the sample mean in the context of this paper, based on long-memory linear process structured data. The underestimation of the variance translates in poor coverage probabilities of the confidence intervals constructed using the (raw) Sieve, as can be seen in Table 4. Filtered Sieve introduced by Poskitt et al. (2015) is a modification of the Sieve that tends to improve upon the coverage probability of the Sieve. It essentially consists of applying the raw Sieve to a filtered series (basically to a truncated version of the infinite sum $(1 - B)^{\hat{d}} X_t$, where $B$ is the back-shift operator and $\hat{d}$ is an estimator of the long-memory parameter $d$) and then unfilter the resulting bootstrapped error series $\hat{\xi}_t^*$ (again by considering the truncated sum $(1 - B)^{-\hat{d}} \hat{\xi}_t^*$).

The block bootstrap is another method of drawing a bootstrap sample from a given set of dependent and temporal observations with the goal of preventing the i.i.d. nature of the naive bootstrap from dismissing the dependence and chronological order of the data. The method essentially consists in first partitioning the data into a number of blocks and then resampling from these blocks. This method was first introduced by Künsch (1989) for short-memory linear processes. The validity of this resampling
scheme for both short and long-memory linear processes has been investigated in a number of papers, e.g., Götze and Künsch (1996), Kim and Nordman (2011).

Although, with a relatively large number of replications, the existing bootstrap methods tend to perform well for short-memory data, it is not the case when they are applied to long-memory data. In our numerical study below, we consider three methods of constructing bootstrap confidence intervals for the mean of a population from which long-memory structured linear process data sets are simulated. In Table 4 we compare the performance of the resulting confidence intervals from the three bootstrap methods Sieve, filtered or augmented Sieve and the block bootstrap for long-memory linear processes to those of the randomized confidence interval $I_{n,X,w}(\cdot)$, as in (4.6), and the classic $I_{n,X}$, as in (4.8), in terms of their respective empirical probabilities of coverage.

In Table 4, $I_{\text{AugSiv}}$, $I_{\text{Bloc}}$ and $I_{\text{Siv}}$ stand, respectively, for the augmented Sieve, block-bootstrap and the Sieve confidence intervals at nominal level of 95%. Each of these bootstrap confidence intervals is constructed based on $B = 1000$ bootstrap replications to estimate the cut-off points for each of the 2000 sets of simulated data. In Table 4, $\text{Mult}$ is a short hand notation for the symmetric multinomial distribution, i.e., $\text{Multinomial}(n; 1/n, \cdots, 1/n)$. For all examined confidence intervals, the simulated observations in Table 4, are generated from the fractionally integrated model $X_t = (1 - B)^{-d} \xi_t$, where $\xi$ is the memory parameter and $\xi_t$ have standardized lognormal$(0,1)$ distribution. Similarly to Tables 1–3, to construct the randomized confidence interval $I_{n,X,w}(\theta_n)$, for each simulated set of data, simultaneously, a set of random weights from symmetric multinomial distribution was generated. In Table 4 for the therein considered long-memory linear process with $d = 0.4$, we let $q$ be $\text{ceiling}(n^{1/2-d})$.

**Remark 4** As it is evident from the numerical study presented in Table 4, the randomized confidence interval $I_{n,X,w}(\theta)$, introduced in this paper, as in (4.6), produces significantly more accurate confidence intervals as compared to the classic $I_{n,X}$, as in (4.8), and the three examined bootstrap confidence intervals.

It is noteworthy that the higher accuracy of $I_{n,X,w}(\theta)$ can also be achieved with random weights other than the symmetric multinomial and exponential used in Tables 1–4, provided that they are chosen according to the skewness reduction scheme (RS), as in Sect. 2.1.

**Remark 5** In the numerical illustrations in Tables 1–4 above, for estimating the variance, the bandwidths $q$ were chosen according to formula (2.14) of Abadir et al. (2009) that gives the optimal bandwidths in terms of minimizing the MSE of the long-run variance estimator. More precisely, we took $q = \text{ceiling}(n^{1/3})$ for short-memory in Tables 1–3 and $q = \text{ceiling}(n^{1/2-d})$ in Table 4.

### 6 Application to real world data

**Use case 1** Historical monthly values of the consumer confidence index (CCI) of Swedish consumers form April 2005 to April 2021 were examined. At 5% significance level, the KPSS test [cf. Kwiatkowski et al. (1992)] shows that the time series
### Table 4  Fractionally integrated long-memory with d=0.4

|         | $I_{n,X,w} (\theta_n)$ | $I_{n,X}$ | $I_{B_{loc}}$ | $I_{A_{rgSiv}}$ | $I_{Siv}$ |
|---------|-------------------------|-----------|---------------|-----------------|-----------|
| Weights | $\mathcal{M}ult$       | NA        | NA            | NA              | NA        |
| $\theta_n$ | 1+0.97                 | NA        | NA            | NA              | NA        |
| n=100  | Skewness of pivot      | -2.193    | 2.714         | NA              | NA        |
|        | Length                  | 3.22      | 2.50          | 2.26            | 1.90      |
|        | Coverage                | 0.92      | 0.83          | 0.803           | 0.710     |
| Weight | $\mathcal{M}ult$       | NA        | NA            | NA              | NA        |
| $\theta_n$ | 1+0.97                 | NA        | NA            | NA              | NA        |
| n=200  | Skewness of pivot      | -2.48     | 2.93          | NA              | NA        |
|        | Length                  | 3.20      | 2.45          | 2.25            | 1.83      |
|        | Coverage                | 0.9445    | 0.864         | 0.823           | 0.739     |

Confidence intervals with higher accuracy for short...
Table 5  95% randomized confidence interval for the mean of CCI of Sweden from April 2005 to April 2021

| Weights | Multinomial(193, 1/193,...,1/193) |
|---------|----------------------------------|
| $\theta_n$ | $1+0.97$ |
| Skewness of randomized pivot | $-0.045$ |
| Skewness of classical pivot | $0.053$ |
| 95% randomized confidence interval | $[97.97,102.6]$ |
| Length | $4.67$ |
| Range of data | Min=96.46, Max=103.4 |

Data source: https://data.oecd.org/leadind/consumer-confidence-index-cci.htm

Table 6  95% randomized confidence interval for the mean of the annual PPI of Japan from 1961 to 2020

| weights | Multinomial(60, 1/60,...,1/60) |
|---------|---------------------------------|
| $\theta_n$ | $1+0.96$ |
| Skewness of randomized pivot | $-1.54$ |
| Skewness of classical pivot | $1.75$ |
| 95% randomized confidence interval | $[-3.68,8.56]$ |
| Length | $12.24$ |
| Range of data | Min=-4.83, Max=27 |

Data source: https://data.oecd.org/price/producer-price-indices-ppi.htm#indicator-chart

is stationary, with $p$-value of 0.1125. Using the Arfima package in R, the memory parameter of the time series was estimated as $\hat{d} = 0.41$ which indicates that the data depicts a long memory behaviour. Using multinomial random weights, in view of the randomization scheme (RS) in Sect. 2.1, a 95% confidence interval, of the form of (4.6), is constructed for the true average of monthly CCI of Swedish consumers and it is given in Table 5 below.

Use case 2 Historical annual values of producer price index (PPI) of Japan from 1961 to 2020 was examined. At 5% significance level, the KPSS test shows that the time series is stationary, with $p$-value of 0.09114. Arfima estimated the memory parameter as $\hat{d} = 0.29$ which again indicates that the time series dataset exhibits a long memory pattern. Using multinomial weights, a 95% randomized confidence interval, of the form of (4.6), is constructed for the true mean of the Japanese annual PPI and it is given in the following Table 6.

Remark 6 In Tables 5 and 6, randomized and classical pivots respectively refer to the pivotal quantities (4.3) and (4.5). The randomized confidence intervals constructed in these tables are of the form (4.6), which are built using the randomized pivot (4.3).
7 Complete studentization

In the Studentizing sequence \( q^{-2d} \hat{S}^2_{n,q,w} \) used in the definition of the Studentized randomized pivot \( G_{n,X,w}(\theta_n, d) \), as defined in (4.3), we used the actual values of the moments of the weights, i.e., \( \mathbb{E}(w_1), \mathbb{E}(w_1)^2 \) and \( \mathbb{E}(w_1 w_2) \). The motive of the use of such a partial Studentization is justified considering that in the framework created by the scheme (RS), the distribution of the weights are usually known. Despite the validity of this reasoning, it is also desirable to investigate the asymptotic behavior of the completely Studentized version of \( G_{n,X,w}(\theta_n, d) \). In the following Theorem 4 we establish a CLT result for the completely Studentized randomized pivotal quantity

\[
\hat{G}_{n,X,w}(\theta_n, d) = n^{-1/2-d} \left( \sum_{i=1}^{n} (w_i - \theta_n)(X_i - \mu) \right) / \sqrt{q^{-2d} \hat{S}^2_{n,q,w}}, \tag{7.1}
\]

where \( q = O(n^{1/2}) \),

\[
\hat{S}^2_{n,q,w} = \frac{1}{n} \sum_{j=1}^{n} (w_j - \theta_n)^2 \bar{\gamma}_0 + 2q^{-1} \sum_{h=1}^{q} \bar{\gamma}_h \sum_{j=1}^{q-h} (w_j - \theta_n)(w_{j+h} - \theta_n),
\]

where \( \bar{\gamma}_s, 0 \leq s \leq n - 1 \), are the sample autocovariances, as defined right after (4.3).

**Theorem 4** Consider \( X_1, \ldots, X_n \), the first \( n \) terms of the linear process (2.1), and let the weights \( w_1, \ldots, w_n \) and the window constants \( \theta_n \) be as in Theorem 2.

(A) If the linear process (2.1) is of short-memory, i.e., \( \sum_{k=0}^{\infty} |a_k| < \infty \), and \( \mathbb{E} \xi_1^4 < \infty \), then, as \( n, q \to \infty \) such that \( q = O(n^{1/2}) \), we have, for all \( x \in \mathbb{R} \),

\[
P(\hat{G}_{n,X,w}(\theta_n, 0) \leq x \mid w_1, \ldots, w_n) \longrightarrow \Phi(x) \text{ in probability},
\]

and, consequently,

\[
P(\hat{G}_{n,X,w}(\theta_n, 0) \leq x) \longrightarrow \Phi(x), \quad t \in \mathbb{R}.
\]

(B) Let the linear process (2.1) be of long-memory such that \( \mathbb{E} \xi_1^4 < \infty \) 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 \( x \in \mathbb{R} \),

\[
P(\hat{G}_{n,X,w}(\theta_n, d) \leq x \mid w_1, \ldots, w_n) \longrightarrow \Phi(x) \text{ in probability},
\]

\[
P(\hat{G}_{n,X,w}(\theta_n, \hat{d}) \leq x \mid w_1, \ldots, w_n) \longrightarrow \Phi(x) \text{ in probability},
\]

and, consequently,

\[
P(\hat{G}_{n,X,w}(\theta_n, d) \leq x) \longrightarrow \Phi(x), \quad x \in \mathbb{R},
\]
where \( \hat{d} \) is an estimator of the memory parameter \( d \) such that \( \hat{d} - d = o_P(1/\log n) \).

The CLTs in the preceding Theorem 4 are counterparts of those in Theorem 3.

### 8 Concluding remarks

The randomization approach introduced in the paper produces pivots for the theoretical mean of a given set of stationary dependent data, which are more accurate than their existing counterparts. The randomization procedure does not change the covariance structure of the original data, i.e., if the original sample is of short memory, then so is its randomized version. The same is also true in case of having a set of long memory data. The superior accuracy of the randomized pivots is illustrated mathematically and empirically. The extra random elements can be generated independently form the data and they can be simulated from a virtually unlimited class of distributions. The improved performance of the randomized pivots, to a large extent, is independent from the distribution from which the extra randomness is generated.

### Appendix

**Proof of Theorem 1**

We first define the following \( \sigma \)-fields to be used in the proof of this theorem.

\[
D_j := \sigma(w_{j-1}, \xi_j), \quad j = 0, \pm 1, \pm 2, \ldots,
\]

\[
D_{j,n} := \sigma(\xi_n, w_{j-1}, \xi_j), \quad j = 0, \pm 1, \pm 2, \ldots,
\]

\[
D_{q}^p := \sigma((w_{j-1}, \xi_j), \quad p \leq j \leq q).
\]

In order for the definition of \( D_j \) to hold true, we extended the weights \( w_1, w_2, \ldots \), to

\[
\ldots, w_{-1}, w_0, w_1, \ldots
\]

Theorem 1 results from Theorem (2.8), and its Corollary (2.9), of Götte and Hipp (1983). More precisely, we show that under the conditions of Theorem 1, conditions (2.3) - (2.6) of Götte and Hipp (1983) hold true. In our case, (2.4) of Götte and Hipp (1983) holds true trivially as \( D_{n-\infty} \) is independent from \( D_{n+m}^{\infty} \), for all \( m \geq 1 \). To see why (2.3) holds, observe that

\[
\mathbb{E}|(w_n - \theta_n)(X_n - \sum_{k=0}^{m} a_k \xi_{n-k})| \leq \mathbb{E}|w_1 - \theta_n| \mathbb{E}|X_n - \sum_{k=0}^{m} a_k \xi_{n-k}|
\]

\[
\leq \mathbb{E}|w_1 - \theta_n| \mathbb{E}||\xi_1| (1/\ell) e^{-\ell m} \leq (c_n/\ell) e^{-\ell m},
\]

(9.1)
where $c_n = E|w_1 - \theta_n|E|\xi_1|$. If $c_n \leq 1$, then (2.3) of Götzte and Hipp (1983) clearly holds. If $c_n > 1$, then $c_n$ will be bounded above by some constant $c > 0$, therefore the right hand side of (9.1) is bounded above by $(c/\ell)e^{-\ell m/c}$. This means that (2.3) of Götzte and Hipp (1983) still holds but with $\ell/c$ instead of $\ell$. In conclusion, condition (2.3) of Götzte and Hipp (1983) holds true in our context of Theorem 1.

We now show that (2.5) of Götzte and Hipp (1983) also holds true in the context of Theorem 1, noting first that in what follows $E.(/F)$ stands for conditional expected value given the $\sigma$-field $F$. Condition (2.5) of Götzte and Hipp in our context corresponds to the following statement: For any $b > 0$, there exists $\ell > 0$ such that for $|u| > b$ and all $m, n$ such that $1/b < m < n$, we have

$$E\left| E\left( \exp\left( iu \left[ \sum_{k=0}^{m} a_k (w_{n+k} - \theta_n) \right] + iu w_{n-1} X_{n-1} \right) / D_j, j \neq n \right) \right| \leq e^{-\ell}. \quad (9.2)$$

To show that (9.2) holds true under the assumptions of our Theorem 1, we let $\Psi_\xi$ and $\Psi_w$ respectively denote the characteristic functions of $\xi_n$ and $w_n$. Observe now that the left hand side of (9.2) is bounded above by

$$\mathbb{E} \left| \mathbb{E} \left( \exp \left( iu \left[ \sum_{k=0}^{m} a_k (w_{n+k} - \theta_n) \right] + iu w_{n-1} X_{n-1} \right) / D_j, j \neq n \right) \right|$$

$$= \mathbb{E} \left| \mathbb{E} \left( \exp \left( iu \left[ \sum_{k=0}^{m} a_k (w_{n+k} - \theta_n) \right] ight) + iu w_{n-1} X_{n-1} / D_{j,n}, j \neq n \right) / D_j, j \neq n \right|$$

$$= \mathbb{E} \left| \mathbb{E} \left( \exp \left( iu \left[ \sum_{k=0}^{m} a_k (w_{n+k} - \theta_n) \right] \right) \right) \right| \times \mathbb{E} \left( \exp \left( iu w_{n-1} X_{n-1} \right) / D_{j,n}, j \neq n \right) / D_j, j \neq n \right|$$

$$= \mathbb{E} \left| \Psi_\xi \left( u \sum_{k=0}^{m} a_k (w_{n+k} - \theta_n) \right) \Psi_w (u X_{n-1}) \right|$$

$$\rightarrow \mathbb{E} |\Psi_w (u X_1)| \mathbb{E} |\Psi_\xi (u Z_n)| \text{ as } m \rightarrow \infty, \quad (9.3)$$

where

$$Z_n = \sum_{k=0}^{\infty} a_k (w_{k+1} - \theta_n).$$

The preceding statement is valid in view of the dominated convergence theorem combined with the use of the well known fact that if $X$ and $Y$ are independent, $h(., .)$ is a bounded function and $g(y) = \mathbb{E}(h(X, y))$ then $\mathbb{E}(h(X, Y)|Y) = g(Y)$.
Now, if the first part of assumption (i) of Theorem 1 holds then, since for all \( n \geq 1 \)
\( Z_n \neq 0 \) a.s., for \( |u| > b \) for a given \( b \), we let \( \delta > 0 \) and \( \epsilon > 0 \) be such that
\( P(|uZ_n| > \epsilon) > 1/2 \) and \( |\Psi_\xi(uZ_n)|_{1|uZ_n|>\epsilon} \leq 1 - 2\delta \). Then we have

\[
\mathbb{E}\left|\Psi_\xi(uZ_n)\right| = \mathbb{E}\left(\left|\Psi_\xi(uZ_n)\right|_{1|uZ_n|>\epsilon}\right) + \mathbb{E}\left(\left|\Psi_\xi(uZ_n)\right|_{1|uZ_n|\leq\epsilon}\right) \\
\leq (1 - 2\delta) P(|uZ_n| > \epsilon) + 1 - P(|uZ_n| > \epsilon) \\
= 1 - 2\delta(1 - P(|uZ_n| > \epsilon)) < 1 - \delta.
\]

In summary, for any \( b > 0 \) there exists \( \delta > 0 \) such that for all \( |u| > b \), we have
\( \mathbb{E}(|\Psi_\xi(uZ_n)|) < 1 - \delta \). Therefore for \( m \) sufficiently large, for all \( |u| > b \), we have

\[
\mathbb{E}\left|\Psi_\xi\left(u \sum_{k=0}^{m} a_k w_{k+1}\right)\right| \leq 1 - \delta = e^{-\ell},
\]

for some \( \ell > 0 \), which completes the proof of (9.2). We note that in our proof for
(9.2), without loss of generality, we can take \( b = \ell \) (possibly by having to work with
\( \zeta_j/\ell \) instead of \( \zeta_j \)) to be in the same context as in (2.5) of Götte and Hipp (1983).
This means that (2.5) of Götte and Hipp (1983) holds true under the first part of
of assumption (i) of Theorem 1.

Alternatively, when the second part of the assumption (i) of Theorem 1 holds, we
show (2.5) of Götte and Hipp (1983) continues to hold true. Recalling that in this
case the weights satisfy the Cramer condition rather than the innovations, to establish
(2.5) of Götte and Hipp (1983) in this case, we first note that the left hand side of
(9.3), i.e., \( \mathbb{E}\left|\Psi_\xi\left(u \sum_{k=0}^{m} a_k (w_{n+k} - \theta_n)\right) \Psi_w(uX_{n-1})\right| \), is bounded above, uniformly
in \( m \geq 1 \), by \( \mathbb{E}|\Psi_w(uX_1)| \). Since in this case we have \( X_1 \neq 0 \) a.s., for given \( b \) and
for \( |u| > b \), we let \( \delta > 0 \) and \( \epsilon > 0 \) be such that \( P(|uX_1| > \epsilon) > 1/2 \) and
\( |\Psi_w(uX_1)|_{1|uX_1|>\epsilon} \leq 1 - 2\delta \). Then, we have

\[
\mathbb{E}\left|\Psi_w(uX_1)\right| = \mathbb{E}\left(\left|\Psi_w(uX_1)\right|_{1|uX_1|>\epsilon}\right) + \mathbb{E}\left(\left|\Psi_w(uX_1)\right|_{1|uX_1|\leq\epsilon}\right) \\
\leq (1 - 2\delta) P(|uX_1| > \epsilon) + 1 - P(|uX_1| > \epsilon) \\
= 1 - 2\delta(1 - P(|uX_1| > \epsilon)) < 1 - \delta.
\]

The preceding relation implies that, for any \( b > 0 \) there exists \( \delta > 0 \) such that for all
\( |u| > b \), we have \( \mathbb{E}(|\Psi_w(uX_1)|) < 1 - \delta \), which implies that (2.5) of Götte and Hipp
(1983) holds true under the second part of of assumption (i) of Theorem 1.

Finally, in our context, condition (2.6) of Götte and Hipp (1983) is clearly satisfied
with our choice of \( D_j = \sigma(w_{j-1}, \zeta_j) \) in view of the Dynkin’s \( \pi - \lambda \) theorem [cf.,
e.g., Billingsley (2012)].

Now the proof of Theorem 1 is complete.

\(\blacksquare\)

**Proof of Theorem 2**

Our arguments below to prove this theorem are valid for both i.i.d. and symmetric
multinomial weights as specified in Theorem 2. We also note that we only give the
proof for the conditional statement (4.1) as it implies the unconditional statement (4.2), in view of the dominated convergence theorem.

For the use in the proof of this theorem, as well as in the proof of Theorem 3 we define

\[ K := \lim_{n \to \infty} \mathbb{E}(w_1 - \theta_n)^2, \] (9.4)

\[ K' := \lim_{n \to \infty} \mathbb{E}\left((w_1 - \theta_n)(w_2 - \theta_n)\right). \] (9.5)

We note that \( K \) and \( K' \) are positive constants and \( K' < K \).

For the ease of the notation, without loss of generality we assume that \( \mu = 0 \) and also define

\[ \mathfrak{R}_{n,w} := \sum_{i=1}^{n} (w_i - \theta_n)^2 + 2 \sum_{h=1}^{n-1} \sum_{j=1}^{n-h} (w_j - \theta_n)(w_{j+h} - \theta_n) \]

\[ = \text{Var}\left( \sum_{i=1}^{n} (w_i - \theta_n)X_i \mid w_1, \ldots, w_n \right). \]

In view of Theorem 2.2 of Abadir et al. (2014), the proof of Theorem 2 will result if we show that as \( n \to \infty \),

\[ \mathfrak{R}_{n,w} / (n \mathbb{D}_{n,X,w}) - 1 = o_p(1), \] (9.6)

\[ \max_{1 \leq i \leq n} |w_i - \theta_n| / \sqrt{n \mathbb{D}_{n,X,w}} = o_p(1), \] (9.7)

and

\[ \sum_{i=1}^{n} (w_i - \theta_n)^2 / (n \mathbb{D}_{n,X,w}) = O_p(1). \] (9.8)

We remark that our condition (9.8) relates to condition \((ii)\) of Theorem 2.2 of Abadir et al. (2014). The latter condition which is intended for constant weights for linear processes reads as follows: \( \exists C > 0 \) such that

\[ \frac{1}{\sigma^2} \sup_{n \geq 1} \sum_{j=1}^{n} z_{nj}^2 \leq C, \]

where \( z_{nj} \) are non-random weights, \( \sigma^2 = \lim_{n \to \infty} \sum_{i=1}^{n} z_{nj}X_i \) and \( X_i, 1 \leq i \leq n \), are the first \( n \) terms of the linear process (2.1). The preceding condition can, conveniently and equivalently, be replaced by

\[ \frac{1}{\sigma^2} \sum_{j=1}^{n} z_{nj}^2 = O(1), \quad \text{as } n \to \infty. \] (9.9)
Hence, in the stochastic context of our Theorem 2, the deterministic statement (9.8) is replaced by condition (9.9).

We now establish (9.6) by writing

\[
\mathcal{R}_{n,w} / (n \mathcal{D}_{n,w} - 1)
\]

\[
= \frac{\gamma_0 n^{-2d} \sum_{i=1}^{n} (w_i - \theta_n)^2 + 2 n^{-2d} \sum_{h=1}^{n} \gamma_h \sum_{j=1}^{n-h} (w_j - \theta_n)(w_{j+h} - \theta_n)}{\gamma_0 n^{-2d} \mathbb{E}(w_1 - \theta_n)^2 + 2 n^{-2d} \mathbb{E}((w_1 - \theta_n)(w_2 - \theta_n)) \sum_{h=1}^{n} (1 - h/n) \gamma_h} - 1
\]

\[
= \left[ \gamma_0 n^{-2d} \left( \frac{\sum_{i=1}^{n} (w_i - \theta_n)^2}{n} - \mathbb{E}(w_1 - \theta_n)^2 \right) + 2 n^{-2d} \sum_{h=1}^{n} \gamma_h \left( \frac{1}{n} \sum_{j=1}^{n-h} (w_j - \theta_n)(w_{j+h} - \theta_n) - \mathbb{E}((w_1 - \theta_n)(w_2 - \theta_n)) \right) \right]^{-1}
\]

\[
\times \left[ \gamma_0 n^{-2d} \mathbb{E}(w_1 - \theta_n)^2 + 2 n^{-2d} \mathbb{E}((w_1 - \theta_n)(w_2 - \theta_n)) \sum_{h=1}^{n} (1 - h/n) \gamma_h \right]^{-1}.
\]

(9.10)

Considering that, as \( n \to \infty \), for i.i.d. and symmetric multinomial weights we have

\[
\gamma_0 n^{-2d} \mathbb{E}(w_1 - \theta_n)^2 + 2 n^{-2d} \mathbb{E}((w_1 - \theta_n)(w_2 - \theta_n)) \sum_{h=1}^{n} (1 - h/n) \gamma_h
\]

\[
\to \begin{cases} 
\gamma_0 (K - K') + K's_\mathcal{X}^2 > 0, & \text{when } d = 0, \\
K's_\mathcal{X}^2 > 0, & \text{when } 0 < d < 1/2,
\end{cases}
\]

(9.11)

where

\[
s_\mathcal{X}^2 := \lim_{n \to +\infty} \text{Var}(n^{1/2-d} \mathcal{X}_n) = \lim_{n \to +\infty} n^{-2d} \left\{ \gamma_0 + 2 \sum_{h=1}^{n-1} \gamma_h (1 - h/n) \right\}.
\]

From (9.11), the relation (9.6) will follow if we show that, as \( n \to \infty \),

\[
\frac{1}{n} \sum_{i=1}^{n} (w_i - \theta_n)^2 - \mathbb{E}(w_1 - \theta_n)^2 = o_P(1)
\]

(9.12)

and

\[
n^{-2d} \sum_{h=1}^{n} \gamma_h \frac{1}{n} \sum_{j=1}^{n-h} \left( (w_j - \theta_n)(w_{j+h} - \theta_n) - \mathbb{E}((w_1 - \theta_n)(w_2 - \theta_n)) \right) = o_P(1).
\]

(9.13)
When the weights are i.i.d., (9.12) is a consequence of the law of large numbers for row-wise i.i.d. triangular arrays of random variables, in view of the assumption that $\mathbb{E}|w_1|^3 < \infty$, as in scheme (RS) and also in Theorem 2.

We now prove that (9.12) continues to hold for the case when $(w_1, \ldots, w_n) \overset{d}{=} \text{Multinomial}(n; 1/n, \ldots, 1/n)$, by writing

$$P\left( \left| \frac{1}{n} \sum_{i=1}^{n} ((w_i - \theta_n)^2 - \mathbb{E}(w_1 - \theta_n)^2) \right| > \epsilon \right)$$

$$\leq \epsilon^{-2} \frac{n}{n^2} \mathbb{E} \left( (w_1 - \theta_n)^2 - \mathbb{E}(w_1 - \theta_n)^2 \right)^2$$

$$+ \epsilon^{-2} \frac{n(n-1)}{n^2} \mathbb{E} \left[ (w_1 - \theta_n)^2 - \mathbb{E}(w_1 - \theta_n)^2 \right] \left( (w_2 - \theta_n)^2 - \mathbb{E}(w_2 - \theta_n)^2 \right].$$

(9.14)

The first expectation above on the right hand side of (9.14) is $O(1)$, as $n \to \infty$. This conclusion is a consequence of conditions (i) of Theorem 2. To show the asymptotic negligibility of the right hand side of (9.14), it remains to show that the second expectation in it is $o(1)$, as $n \to \infty$. To do so we write,

$$\mathbb{E} \left[ (w_1 - \theta_n)^2 - \mathbb{E}(w_1 - \theta_n)^2 \right] \left( (w_2 - \theta_n)^2 - \mathbb{E}(w_2 - \theta_n)^2 \right]$$

$$= \mathbb{E} \left( (w_1 - \theta_n)^2(w_2 - \theta_n)^2 \right) - \mathbb{E}^2(w_1 - \theta_n)^2$$

$$= \mathbb{E} \left( (w_1 w_2)^2 - 2(w_1)^2w_2 + \theta_n^2(w_1)^2 - 2\theta_n w_1(w_2)^2 + 4\theta_n^2 w_1 w_2 - 2\theta_n^3 w_1 + \theta_n^2(w_2)^2 - 2\theta_n^3 w_2 + \theta_n^4 \right)$

$$- \mathbb{E}^2(w_1)^2 - 4\mathbb{E}(w_1)\theta_n^2 - \theta_n^4 + 4\mathbb{E}^2(w_1)^2\mathbb{E}(w_1)\theta_n$$

$$- 2\mathbb{E}^2(w_1) + 4\theta_n^3\mathbb{E}(w_1)$$

$$\sim \left( 4 - 4\theta^* + 2(\theta^*)^2 - 4\theta^* + 4(\theta^*)^2 - 2(\theta^*)^3 + 2(\theta^*)^2 - 2(\theta^*)^3 + (\theta^*)^4 \right)$$

$$- 4 - 4(\theta^*)^2 - (\theta^*)^4 + 8\theta^* - 4(\theta^*)^2 + 4(\theta^*)^3$$

$$= 0.$$ 

(9.15)

Therefore, (9.12) holds true for symmetric multinomial weights too.

As for (9.13), when weights are i.i.d., for all $\epsilon > 0$, we have

$$P \left( n^{-2d} \sum_{h=1}^{n} \gamma_h \frac{1}{n} \sum_{j=1}^{n-h} ((w_j - \theta_n)(w_{j+h} - \theta_n) - \mathbb{E}((w_1 - \theta_n)(w_2 - \theta_n))) > \epsilon \right)$$

$$\leq \epsilon^{-1} n^{-2d} \sum_{h=1}^{n} |\gamma_h|^{1/2}$$
Now observe that for all $h \geq 1$,

$$
\mathbb{E} \left( \frac{1}{n} \sum_{j=1}^{n-h} (w_j - \theta_n)(w_{j+h} - \theta_n) - \mathbb{E}((w_1 - \theta_n)(w_2 - \theta_n)) \right)^2
$$

$$
= \frac{1}{n^2} \sum_{j=1}^{n-h} \mathbb{E} \left( (w_j - \theta_n)(w_{j+h} - \theta_n) - \mathbb{E}^2(w_1 - \theta_n) \right)^2
$$

$$
= \frac{n - h}{n^2} (\mathbb{E}((w_1 - \theta_n)(w_2 - \theta_n))^2 - \mathbb{E}^4(w_1 - \theta_n))
$$

Substituting the preceding relation into the right hand side of (9.16), in view of conditions (i) and (ii) of this theorem, we conclude that

$$
\epsilon^{-1}n^{-2d} \sum_{h=1}^{n} |\gamma_h| \frac{\sqrt{n-h}}{n} \left( \mathbb{E}^2((w_1 - \theta_n)^2) - \mathbb{E}^4(w_1 - \theta_n) \right)^{1/2}
$$

$$
= o(1), \text{ as } n \to \infty.
$$

The preceding convergence to zero holds uniformly in $h$ since $\sup_{1 \leq h \leq n-1} \sqrt{n-h}/n = \sqrt{n-1}/n \to 0$ and $n^{-2d} \sum_{h=1}^{\infty} |\gamma(h)|$ is bounded. Now the proof of (9.13) in the case of i.i.d. weights is complete.

The proof of (9.13) when the weights are symmetric multinomial also begins with the inequality (9.16). In this case, in view of condition (i), as $n \to \infty$, we have

$$
\mathbb{E} \left( (w_1 - \theta_n)(w_2 - \theta_n) \right)^2 - \mathbb{E}^2 \left( (w_1 - \theta_n)(w_2 - \theta_n) \right) = O(1),
$$

(9.17)

$$
\mathbb{E} \left\{ (w_1 - \theta_n)(w_2 - \theta_n) - \mathbb{E} \left( (w_1 - \theta_n)(w_2 - \theta_n) \right) \right\}
$$

$$
\times \left\{ (w_3 - \theta_n)(w_4 - \theta_n) - \mathbb{E} \left( (w_3 - \theta_n)(w_4 - \theta_n) \right) \right\}
$$

$$
\sim \left\{ 1 - 2\theta^* + (\theta^*)^2 - 2\theta^* + 4(\theta^*)^2 - 2(\theta^*)^3 + (\theta^*)^2 - 2(\theta^*)^3 + (\theta^*)^4 - 1 - 4(\theta^*)^2 - (\theta^*)^4 + 4\theta^* - 2(\theta^*)^2 + 4(\theta^*)^3 \right\}
$$

$$
= 0.
$$

(9.18)
These last two approximations imply that the right hand side of (9.16) is, asymptotically in $n$ and uniformly in $h$, negligible which means that (9.13) holds for symmetric multinomial weights. This completes the proof of (9.6).

We now give a unified argument for both i.i.d. and symmetric multinomial weights to prove (9.7). The proof is also valid for both short and long-memory data and it begins with observing that, as $n \to \infty$, we have

$$
\frac{1}{\sqrt{n}} \max_{1 \leq j \leq n} |w_j - \theta_n| = o_P(1).
$$

(9.19)

Considering that $\theta_n$ is bounded, the weight $w_j$, for each $n$, are identically distributed and, since, in the scheme (RS) the weights are so that $\sup_{n \geq 1} E|w_1|^3 < \infty$, the proof of (9.19) follows from the following argument.

$$
P\left( \max_{1 \leq j \leq n} (w_j)^2 > n\epsilon \right) \leq n P\left( (w_1)^2 > \epsilon \right) \\
\leq \epsilon^{-3/2} n^{-1/2} E|w_1|^3 \leq \epsilon^{-3/2} n^{-1/2} E|w_1|^3 \to 0.
$$

In view of the conditions (i) and (ii) of Theorem 2, for some finite number $A$, we have

$$
E((w_1 - \theta_n)^2) + 2E\left((w_1 - \theta_n)(w_2 - \theta_n)\right) \sum_{h=1}^{n} (1 - \frac{h}{n}) \gamma_h \\
\to K \gamma_0 + 2K' \sum_{h=1}^{\infty} \gamma_h = \begin{cases} A', & \text{when } X_j \text{ of short-memory}, \\ \infty, & \text{when } X_j \text{ of long-memory}. \end{cases}
$$

Recall that $K' < K$. This, in turn, implies that $A > 0$. This together with (9.19) completes the proof of (9.7).

To complete the proof of Theorem 2, we need to prove (9.8). The validity of (9.8), for both i.i.d. and symmetric multinomial weights, by virtue of (9.12) and conditions (ii) of Theorem 2, results from the following weak law of large numbers, as $n \to \infty$.

$$
\frac{1}{n} \sum_{j=1}^{n} (w_j - \theta_n)^2 \\
E((w_1 - \theta_n)^2) + 2E\left((w_1 - \theta_n)(w_2 - \theta_n)\right) \sum_{h=1}^{n} (1 - \frac{h}{n}) \gamma_h \\
\to P \frac{K}{\gamma_0 K + 2K' \sum_{h=1}^{\infty} \gamma_h} = \begin{cases} A' > 0, & \text{when } X_j \text{ of short-memory}, \\ 0, & \text{when } X_j \text{ of long-memory}. \end{cases}
$$

This completes the proof of (9.8) as well as that of Theorem 2.

\[\square\]

**Proof of Theorem 3**

In the proof of this theorem, as well as in the proof of Theorem 4, for the ease of the notation we let $P_{X|w}(\cdot)$ stand for the conditional probability $P(\cdot|w_1, \cdots, w_n)$.
As \( n, q \to \infty \), in such a way that \( q = O(n^{1/2}) \), under the conditions of this theorem, from the approximation (2.10) of Theorem 2.1 of Abadir et al. (2009), that holds true for \( d \) and \( \hat{d} \), we have

\[
q^{-2d} \gamma_0 + 2q^{-2d} \sum_{h=1}^{q} \tilde{\gamma}_h (1 - \frac{h}{q}) \longrightarrow s_X^2,
\]

where \( s_X^2 \) is as defined right after (9.11). This, in turn, implies that, as \( n, q \to \infty \), in such a way that \( q = O(n^{1/2}) \), we also have

\[
q^{-2d} s_{n,q,w}^2 = q^{-2d} \mathbb{E}(w_1 - \theta_n)^2 \gamma_0 + 2q^{-2d} \mathbb{E}\left((w_1 - \theta_n)(w_2 - \theta_n)\right) \sum_{h=1}^{q} \tilde{\gamma}_h (1 - \frac{h}{q})
\]

\[\xrightarrow{P} \begin{cases} \gamma_0(K - K') + K's_X^2 > 0, & \text{when } d = 0, \\
K's_X^2 > 0, & \text{when } 0 < d < 1/2, \end{cases} \tag{9.20} \]

where the constant \( K \) is as in (9.4) and \( K' \) is as in (9.5). The preceding convergence means that the Studentizing sequence \( q^{-2d} s_{n,q,w}^2 \) asymptotically coincides with the limit of \( nD_{n,X,w} \), which is the normalizing sequence for \( T_{n,X,w}(\theta_n) \), as in (2.7). In Theorem 2 we showed \( T_{n,X,w}(\theta_n) \) has standard normal limiting distribution. Considering that

\[
G_{n,X,w}(\theta_n, d) = T_{n,X,w}(\theta_n) \sqrt{\frac{D_{n,X,w}/(q^{-2d} s_{n,q,w}^2)}} \tag{9.21}
\]

we conclude, from Slutsky theorem, that \( G_{n,X,w}(\theta_n, d) \) also convergence to standard normal. The relations (9.20) and (9.21) are also true when the memory parameter \( d \) is replaced by its estimator \( \hat{d} \), provided that \( \hat{d} - d = o_p(1/\log n) \).

Now the proof of Theorem 3 is complete. \( \square \)

**Proof of Theorem 4**

The proof of Theorem 4, for both i.i.d. and symmetric multinomial weights, will follow if we show that

\[
P_{X|w}\left\{ |(q^{-2d} \hat{S}_{n,q,w})/(q^{-2d} s_{n,q,w}^2) - 1| > \epsilon \right\}
\]

\[= P_{X|w}\left\{ \left| \frac{q^{-2d} \gamma_0 \frac{1}{n} \sum_{j=1}^{n} (w_j - \theta_n)^2 + 2q^{-1-2d} \sum_{h=1}^{q} \tilde{\gamma}_h \sum_{j=1}^{q-h} (w_j - \theta_n)(w_{j+h} - \theta_n)}{q^{-2d} \gamma_0 \mathbb{E}(w_1 - \theta_n)^2 + 2q^{-2d} \mathbb{E}\left((w_1 - \theta_n)(w_2 - \theta_n)\right) \sum_{h=1}^{q} \tilde{\gamma}_h (1 - \frac{h}{q})} - 1| > \epsilon \right\}
\]

\[= o_P(1), \quad \text{as } n \to \infty, \tag{9.22} \]

\( \square \) Springer
where $\epsilon$ is an arbitrary positive constant. By virtue of (9.20), the proof of (9.22), as $n, q \to \infty$ such that $q = O(n^{1/2})$, will follow from (9.12) combined with

$$P_{X|w}(\lfloor q^{-2d} \sum_{h=1}^{q} \tilde{y}_h B_{n,q,w}(h) \rfloor > \epsilon) = o_P(1), \quad (9.23)$$

where

$$B_{n,q,w}(h) := q^{-1} \sum_{j=1}^{q-h} \left( (w_j - \theta_n)(w_{j+h} - \theta_n) - \mathbb{E}((w_1 - \theta_n)(w_2 - \theta_n)) \right).$$

The proof of (9.23) is a modification of the proof of (6.14) in Csörgő et al. (2017).

For convenient reference here we give the proof of (9.23) for both i.i.d. and symmetric multinomial weights. To do so, without loss of generality, we assume that $\mu = \mathbb{E}X_1 = 0$ and, for each $1 \leq h \leq q$, we define

$$\gamma_h^* := \frac{1}{n} \sum_{i=1}^{n-h} X_i X_{i+h}. \quad (9.24)$$

Observe that, for $\epsilon_1, \epsilon_2 > 0$, we have

$$P \left\{ P_{X|w}(q^{-2d} \sum_{h=1}^{q} \tilde{y}_h B_{n,q,w}(h) \rfloor > 2\epsilon_1) > \epsilon_2 \right\}$$

$$\leq P \left\{ P_{X|w}(q^{-2d} \sum_{h=1}^{q} (\tilde{y}_h - \gamma_h^*) B_{n,q,w}(h) \rfloor > \epsilon_1) > \epsilon_2 \right\}$$

$$+ P \left\{ P_{X|w}(q^{-2d} \sum_{h=1}^{q} \gamma_h^* B_{n,q,w}(h) \rfloor > \epsilon_1) > \epsilon_2 \right\}. \quad (9.25)$$

We now show that the first term in (9.25) is asymptotically negligible, noting first that

$$\sum_{h=1}^{q} (\tilde{y}_h - \gamma_h^*) B_{n,q,w}(h) = -\tilde{X}_n \sum_{h=1}^{q} B_{n,q,w}(h) \sum_{i=1}^{n-h} X_i - \tilde{X}_n \sum_{h=1}^{q} B_{n,q,w}(h) \sum_{i=1}^{n-h} X_{i+h}$$

$$+ \tilde{X}^2 \sum_{h=1}^{q} B_{n,q,w}(h)$$

$$\sim -\tilde{X}^2 \sum_{h=1}^{q} B_{n,q,w}(h) \text{ uniformly in } h \text{ in probability}$$

$$- P_{X|w}. \quad (9.26)$$
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 (9.26) is true since, for \( \varepsilon > 0 \) we have

\[
P\left( \bigcup_{1 \leq h \leq q} \left| \tilde{X}_n \bigg( \sum_{i=1}^{n-h} X_i \right) \bigg| > \varepsilon \right) \leq q \left( \frac{1}{n} \sum_{i=n-h+1}^{n} X_i \right) > \varepsilon \)
\[
\leq \varepsilon^{-4} \frac{(h-1)^4}{n^4} \mathbb{E}(X_1^4)
\]
\[
\leq \varepsilon^{-4} q^5 \mathbb{E}(X_1^4) \to 0, \text{ as } n \to \infty.
\]

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

We note that for \( 0 \leq d < 1/2 \), as \( n \to \infty \), we have that \( n^{1/2-d} \tilde{X}_n = O_P(1) \). The latter conclusion, in view of the equivalence in (9.26), implies that, for each \( \varepsilon_1, \varepsilon_2 > 0 \), there exists \( \varepsilon > 0 \) such that

\[
P \left( \bigcup_{1 \leq h \leq q} \left( |\gamma_h - \gamma_h^*| B_{n,q,w}(h) \right) > \varepsilon_1 \right) > \varepsilon_2
\]
\[
\sim P \left( \frac{q^{-2d}}{n^{1-2d}} \sum_{h=1}^{q} \left( |B_{n,q,w}(h)| \right) \right)
\]
\[
\leq \varepsilon^{-1} \frac{q^{-2d}}{n^{1-2d}} \sum_{h=1}^{q} \mathbb{E}\left(|B_{n,q,w}(h)|\right).
\]

(9.27)

From condition (ii) of Theorem 2, that is also assumed in Theorem 3, there exists a constant \( \mathcal{L} \) whose value does not depend on \( n \) such that \( \sup_{n \geq 2} \sup_{1 \leq h \leq q} \mathbb{E}\left(|B_{n,q,w}(h)|\right) < \mathcal{L} \). Hence, (9.27) can be bounded above by

\[
\mathcal{L} \varepsilon^{-1} \frac{q^{1-2d}}{n^{1-2d}} \to 0,
\]

as \( n, q \to \infty \) in such away that \( q = O(n^{1/2}) \). This means that the first term in (9.25) is asymptotically negligible. To establish (9.22) we show that the second term in (9.25) is also asymptotically negligible. To prove this negligibility, we first define

\[
\gamma_h^{**} := \frac{1}{n} \sum_{i=1}^{n} X_i X_{i+h}.
\]

(9.28)

Now, observe that

\[
P \left( \bigcup_{1 \leq h \leq q} \left| \gamma_h^{**} - \gamma_h^* \right| > \varepsilon \right) \leq q \left( \frac{1}{n} \sum_{i=n-h+1}^{n} X_i X_{i+h} \right) > \varepsilon \]
\[
\leq \varepsilon^{-2} \frac{q^3}{n^2} \mathbb{E}(X_1^4) \to 0,
\]
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 an argument similar to those used for (9.25) and (9.27), with \( \gamma_h^* \) replacing \( \bar{\gamma}_h \) and \( \gamma_{hh}^* \) replacing \( \gamma_h^* \) therein, we arrive at

\[
P \{ P_X | w(q^{-2d} \sum_{h=1}^q \gamma_h^* B_{n,q,w}(h) > \varepsilon_1) > \varepsilon_2 \}
\]

\[
\sim P \{ P_X | w(q^{-2d} | \sum_{h=1}^q \gamma_{hh}^* B_{n,q,w}(h) | > \varepsilon_1) > \varepsilon_2 \}.
\]

Therefore, in order to prove (9.22), it suffices to show that, as \( n, q \to \infty \) so that \( q = O(n^{1/2}) \),

\[
P \{ P_X | w(q^{-2d} | \sum_{h=1}^q \gamma_{hh}^* B_{n,q,w}(h) | > \varepsilon_1) > \varepsilon_2 \} \to 0.
\]

The preceding relation, in turn, follows from the following two conclusions: as \( n, q \to \infty \) so that \( q = O(n^{1/2}) \),

\[
\sup_{1 \leq h, h' \leq q} \mathbb{E}( | B_{n,q,w}(h) B_{n,q,w}(h') |) = o(1)
\]

(9.29)

and

\[
q^{-4d} \sum_{h=1}^q \sum_{h'=1}^q | \mathbb{E}( \gamma_{hh}^* \gamma_{h'h'}^* ) | = O(1).
\]

(9.30)

To prove (9.29), using the Cauchy inequality we write

\[
\mathbb{E}( | B_{n,q,w}(h) B_{n,q,w}(h') |)
\]

\[
\leq \mathbb{E} \left( (w_1 - \theta_n)(w_2 - \theta_n) - \mathbb{E}( (w_1 - \theta_n)(w_2 - \theta_n) ) \right)^2 
\]

\[
+ \frac{(q-h)(q-h-1)}{q^2} \mathbb{E} \left( (w_3 - \theta_n)(w_4 - \theta_n) - \mathbb{E}( (w_3 - \theta_n)(w_4 - \theta_n) ) \right) 
\]

\[
\times \left( (w_3 - \theta_n)(w_4 - \theta_n) - \mathbb{E}( (w_3 - \theta_n)(w_4 - \theta_n) ) \right).
\]

(9.31)

In case of i.i.d. weights, (9.31) is bounded above by \( q^{-1} \mathbb{E}^2(w_1 - \theta_n)^2 \) that vanishes as \( q, n \to \infty \). In case of symmetric multinomial weights, from (9.17) and (9.18) we can see that (9.31) has an upper bound of the form \( 2K_n/q + k_n \), where \( K_n = O(1) \) and \( k_n = o(1) \), that also vanishes as \( n, q \to \infty \). The latter conclusion completes the proof of (9.31) and that of (9.29).
In order to establish (9.30), 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 (9.30), using a generalization of an argument used in the proof of Proposition 7.3.1 of Brockwell and Davis (2009) as follows:

\[ q^{-4d} \sum_{h=1}^{q} \sum_{h'=1}^{q} |E(\gamma_h^{**} \gamma_{h'}^{**})| \leq q^{-2d} \sum_{h=1}^{q} |\gamma_h| q^{-2d} \sum_{h'=1}^{q} |\gamma_{h'}| \]

\[ + \left( \frac{q}{n} \right)^{1-2d} n^{-2d} \sum_{k=-n}^{n} |\gamma_{h'}| q^{-2d} \sum_{L=-q}^{q} |\gamma_{k+L}| \]

\[ + \frac{1}{n} n^{-2d} \sum_{k=-n}^{n} |\gamma_{k} \gamma_{h'}| q^{-2d} \sum_{h=1}^{q} |\gamma_{k-h}| \]

\[ + \frac{q^{-2d}}{1-2d} q^{-2d} \sum_{i=1}^{n} \sum_{k=-n}^{n} |a_i a_{i+k}| q^{-d} \sum_{h=1}^{q} |a_i \gamma_{h}| q^{-d} \sum_{h'=1}^{q} |a_{i+k-h'}|. \quad (9.32) \]

It is easy to see that, as \( n \to \infty \), and consequently \( q \to \infty \), the (whole) right hand side of the inequality (9.32) converges to the finite limit \( 3H^2 \). Now the proof of (9.30) and also that of Theorem 4 are complete. \( \square \)

**Acknowledgements** The authors wish to thank Miklós Csörgő for his constructive comments on the first draft of the paper. The authors also wish to thank the two anonymous referees for their invaluable remarks and suggestions that helped improve the first version of this manuscript.

**References**

Abadir KM, Distaso W, Giraitis L (2009) Two estimators of the long-run variance: beyond short-memory. J Econ 150:56–70

Abadir KM, Distaso W, Giraitis L, Koul HL (2014) Asymptotic normality for weighted sums of linear processes. Econ Theor 30:252–284

Andrews DWK (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59:817–858

Barndorff-Nielsen OE, Cox DR (1989) Asymptotic techniques for use in statistics. Chapman and Hall, London, New York

Bhansali R J, Kokoszka PS (2001) Estimation of the long-memory parameter: a review of recent developments and an extension. In: Basawa IV, Heyde CC, Taylor RL (eds) Proceedings of the Symposium on Inference for Stochastic Processes. IMS Lecture Notes, pp 125–150

Bhattacharya RN, Rao RR (1976) Normal approximation and asymptotic expansions. Wiley, New York, London, Sydney, Toronto

Billingsley P (2012) Probability and measure, 3rd edn. Wiley, New York

Brockwell PJ, Davis RA (2009) Time series: theory and methods, 2nd edn. Springer, New York

Csörgő M, Nasari MM (2015) Inference from small and big data sets with error rates. Electron J Stat 9:535–566
Csörgő M, Nasari MM, Ould-Haye M (2017) Randomized pivots for means of short and long-memory linear processes. Bernoulli 23:2558–2586
Efron B (1979) Another look at the Jackknife. Ann Stat 7:1–26
Eriksson M (2008) A simulation method for skewness correction. U.U.D.M. Project Report 2008:22, Uppsala University
Götze F, Hipp C (1983) Asymptotic expansions for sums of weakly dependent random vectors. Z Wahrsch Verw Gebiete 64:211–239
Götze F, Künsch HR (1996) Second-order correctness of the blockwise bootstrap for stationary observations. Ann Stat 5:1914–1933
Hall P (1983) Inverting an Edgeworth expansion. Ann Stat 11:569–576
Hassle J, Raftery AE (1989) Space-time modelling with longmemory dependence: assessing Ireland’s wind power resource. Appl Stat 38:1–50
Kim YM, Nordman DJ (2011) Properties of a block bootstrap under long-range dependence. Indian J Stat 73–A:79–109
Künsch HR (1987) Statistical aspect of self-similar processes. In: Prohorov YA, Sazonov VV (eds) Proceedings of the 1st World Congress of the Bernoulli Society 1. VNU Science Press, pp 67–74
Künsch HR (1989) The jackknife and the bootstrap for general stationary observations. Ann Stat 3:1217–1241
Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. J Econom 54(1–3):159–178
Lahiri SN (1993) Refinements in asymptotic expansions for sums of weakly dependent random vectors. Ann Probab 21:791–799
Lahiri SN (2003) Resampling methods for dependent data. Springer, New York
Moulines E, Soulier Ph (2003) Semiparametric spectral estimation for fractional processes: In Theory and applications of long-range dependence. In Dukhan P, Oppenheim G, Taqqu M (eds) Theory an applications of long-range dependence. Birkhäuser, Boston
Nasari MM (2016) Pivotal quantities with arbitrary small skewness. arXiv:1605.05985
Nasari MM, Ould-Haye M (2021) A consistent estimator for skewness of partial sums of dependent data. Stat Probab Lett 171:109021
Poskitt DS (2008) Properties of the Sieve bootstrap for fractionally integrated and non-invertible processes. J Time Ser Anal 29:224–250
Poskitt D, Simone G, Martin G (2015) Higher-order improvements of the Sieve bootstrap for fractionally integrated processes. J Econ 188:94–110
Robinson PM (1995) Guassian semiparametric estimation of long range dependence. Ann Stat 23:1630.5-1661.5
Robinson PM (2005) Robust covariance matrix estimation: HAC estimates with long-memory/antipersistence correction. Econom Theor 21:171–180
Senatov VV (1998) Normal approximation: new results, methods and problems. VSP, Utrecht, Tokyo
Shorack GR (2000) Probability for statisticians. Springer, Berlin

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.