

## ON OPTIMAL BLOCK RESAMPLING FOR GAUSSIAN-SUBORDINATED LONG-RANGE DEPENDENT PROCESSES

BY QIHAO ZHANG<sup>1,a</sup>, SOUMENDRA N. LAHIRI<sup>2,c</sup> AND DANIEL J. NORDMAN<sup>1,b</sup>

<sup>1</sup>*Department of Statistics, Iowa State University, <sup>a</sup>qihaoz@iastate.edu, <sup>b</sup>dnordman@iastate.edu*

<sup>2</sup>*Department of Mathematics and Statistics, Washington University in St. Louis, <sup>c</sup>s.lahiri@wustl.edu*

Block-based resampling estimators have been intensively investigated for weakly dependent time processes, which has helped to inform implementation (e.g., best block sizes). However, little is known about resampling performance and block sizes under strong or long-range dependence. To establish guideposts in block selection, we consider a broad class of strongly dependent time processes, formed by a transformation of a stationary long-memory Gaussian series, and examine block-based resampling estimators for the variance of the prototypical sample mean; extensions to general statistical functionals are also considered. Unlike weak dependence, the properties of resampling estimators under strong dependence are shown to depend intricately on the nature of nonlinearity in the time series (beyond Hermite ranks) in addition to the long-memory coefficient and block size. Additionally, the intuition has often been that optimal block sizes should be larger under strong dependence (say  $O(n^{1/2})$  for a sample size  $n$ ) than the optimal order  $O(n^{1/3})$  known under weak dependence. This intuition turns out to be largely incorrect, though a block order  $O(n^{1/2})$  may be reasonable (and even optimal) in many cases, owing to nonlinearity in a long-memory time series. While optimal block sizes are more complex under long-range dependence compared to short-range, we provide a consistent data-driven rule for block selection. Numerical studies illustrate that the guides for block selection perform well in other block-based problems with long-memory time series, such as distribution estimation and strategies for testing Hermite rank.

**1. Introduction.** Block-based resampling methods provide useful nonparametric approximations with statistics from dependent data, where data blocks help to capture time dependence (cf. [28]). Considering a stretch from a stationary series  $X_1, \dots, X_n$ , a prototypical problem involves estimating the standard error of the sample mean  $\bar{X}_n = \sum_{t=1}^n X_t/n$ . Subsampling [13, 22, 41] and block-bootstrap [29, 34] use sample averages  $\bar{X}_{i,\ell}$  computed over length  $\ell < n$  data blocks  $\{(X_i, \dots, X_{i+\ell-1})\}_{i=1}^{n-\ell+1}$  within the data  $X_1, \dots, X_n$ ; in both resampling approaches, the empirical variance of block averages, say  $\hat{\sigma}_\ell^2$ , approximates the block variance  $\sigma_\ell^2 \equiv \text{Var}(\bar{X}_\ell)$ . If the series  $\{X_t\}$  exhibits short-range dependence (SRD) with quickly decaying covariances  $r(k) \equiv \text{Cov}(X_0, X_k) \rightarrow 0$  as  $k \rightarrow \infty$  (i.e.,  $\sum_{k=1}^\infty |r(k)| < \infty$ ), then the target variance converges  $n\sigma_n^2 = n \text{Var}(\bar{X}_n) \rightarrow C > 0$  as  $n \rightarrow \infty$  and  $\ell\hat{\sigma}_\ell^2$  is consistent for  $n \text{Var}(\bar{X}_n)$  under mild conditions ( $\ell^{-1} + \ell/n \rightarrow \infty$ ) [42]. Block-based variance estimators have further history in time series analysis (cf. overview in [40]), including batch means estimation in Markov chain Monte Carlo. Particularly for SRD, much research has focused on explaining properties of block-based estimators  $\hat{\sigma}_\ell^2$  (cf. [17, 29, 31, 32, 42, 48]). In turn, these resampling studies have advanced understanding of best block sizes (e.g.,  $O(n^{1/3})$ ) and implementation under SRD [12, 21, 33, 38, 43].

However, in contrast to SRD, relatively little is known about properties of block-based resampling estimators and block sizes under strong or long-range time dependence (LRD).

---

Received June 2022; revised September 2022.

*MSC2020 subject classifications.* Primary 62G09; secondary 62G20, 62M10.

*Key words and phrases.* Block length, block bootstrap, long-memory, subsampling, variance estimation.

For example, recent tests of Hermite rank [9] as well as other approximations with block bootstrap and subsampling under LRD rely on data blocks [4, 6, 10, 26], creating a need for guides in block selection. To develop an understanding of data blocking under LRD, we consider the analog problem from SRD of estimating the variance  $\text{Var}(\bar{X}_n)$  of a sample mean  $\bar{X}_n$  through block resampling (cf. Sections 2–4); block selections in this context extend to broader statistics (cf. Section 5) and provide guidance for distributional approximations with resampling (cf. Section 6). Because long-memory or long-range dependent (LRD) time series are characterized by slowly decaying covariances (i.e.,  $\sum_{k=1}^{\infty} |r(k)| = +\infty$  diverges), optimal block sizes in this problem have intuitively been conjectured as  $O(n^{1/2})$ , which is longer than the best block size  $O(n^{1/3})$  associated with weak dependence [10, 23]. However, this intuition about block selections is misleading. Under general forms of LRD, the best block selections turn out to depend critically on *both* dependence strength (i.e., rate of covariance decay) *and* the nature of nonlinearity in a time series. To illustrate, consider a stationary Gaussian LRD time series  $\{Z_t\}$ , which we may associate with common models for long-memory [20, 35], and suppose  $\{Z_t\}$  has a covariance decay controlled by a long-memory exponent  $\alpha \in (0, 1)$  (described more next). Then the LRD process  $Z_t$  for  $\alpha < 1/2$  can have an optimal block length  $O(n^\alpha)$  while a cousin LRD process  $Z_t + 0.5Z_t^2$  has a best block size  $O(n^{1/2})$  *regardless* of the memory level  $\alpha \in (0, 1/2)$ . That is, classes of LRD processes exist where nonlinearity induces a best block order  $O(n^{1/2})$ . Also, as the optimal block size  $O(n^\alpha)$  for  $Z_t$  ( $\alpha \in (0, 1/2)$ ) illustrates, when the covariance decays slowly as  $\alpha \downarrow 0$ , the best block sizes for a resampling variance estimator under LRD do *not* generally increase with increasing dependence strength. While theory justifies a block length  $O(n^{1/2})$  as optimal in some cases, the forms of theoretically best block sizes can generally be complex under LRD and we also establish a provably consistent data-based estimator of this block size. Numerical studies show that the empirical block selection performs well in variance estimation and provides a guide with good performance in other resampling problems under LRD (e.g., distribution estimation for statistical functionals).

Section 2 describes the LRD framework and variance estimation problem. We consider stationary LRD processes  $X_t = G(Z_t)$  defined by a transformation  $G(\cdot)$  of a LRD Gaussian process  $\{Z_t\}$  with a long-memory exponent  $\alpha \in (0, 1/m)$  (cf. [49, 50]); here, integer  $m \geq 1$  is the so-called Hermite rank of  $G(\cdot)$ , which has a well-known impact on the distributional limit of the sample mean  $\bar{X}_n$  for such LRD processes (e.g., normal if  $m = 1$  [15, 51]). Section 3 provides the large-sample bias and variance of block-based resampling estimators in the sample mean case, which are used to determine MSE-optimal block sizes and a consistent approach to block estimation in Section 4. As complications, best block lengths can depend on the memory exponent  $\alpha$  and a higher order rank beyond Hermite rank  $m$  (e.g., second Hermite rank). Two versions of data blocking are also compared, involving fully overlapping (OL) or nonoverlapping (NOL) blocks; while OL blocks are always MSE-better for variance estimation under SRD [31, 32], this is not true under LRD. Section 5 extends the block resampling to broader statistical functionals under LRD (beyond sample means) and includes the block jackknife technique for comparison. Numerical studies are provided in Section 6 to illustrate block size selections and resampling across problems of variance estimation, distributional approximations, and Hermite rank testing under LRD. Section 7 has concluding remarks, and Supplementary Material [55] provides proofs and supporting results.

We end here with some related literature. Particularly, for Gaussian series  $X_t = Z_t$  (or  $G(x) = x$  with  $m = 1$ ), the computation of (log) block-based variance estimators over a series of large block sizes  $\ell$  can be a graphical device for estimating the long-memory parameter  $\alpha$  (using that  $\text{Var}(\bar{X}_\ell) \approx C_0 \ell^{-\alpha}$  for subsample averages, cf. Section 2) [36, 53]. Relatedly, [18] considered block-average regression-type estimators of  $\alpha$  in the Gaussian case. For distribution estimation with LRD linear processes, [37, 56] studied subsampling, while [27] examined block bootstrap. As perhaps the most closely related works [1, 18, 27] studied optimal

blocks/bandwidths for estimating a sample mean's variance with LRD linear processes (under various assumptions) using data-block averages or related Bartlett-kernel heteroskedasticity and autocorrelation consistent (HAC) estimators. Those results share connections to optimal block sizes here for purely Gaussian series  $X_t = Z_t$  (cf. Section 3.1), but no empirical estimation rules were considered. As novelty, we account for LRD data  $X_t = G(Z_t)$  from general transformations  $G(\cdot)$  (i.e., the pure Gaussian/linear case is comparatively simpler), establish consistent block estimation, provide results for more general statistical functions, and consider the applications of block selections in wider contexts under LRD. In terms of resampling from LRD transformed Gaussian processes, [30] showed the block bootstrap is valid for approximating the full distribution of the sample mean  $\bar{X}_n$  when the Hermite rank is  $m = 1$ , while [23] established subsampling as valid for any  $m \geq 1$ . (While block bootstrap and subsampling differ in their distributional approximations [42], these induce a common block-based variance estimator for the sample mean, as described in Section 2.) Recently, much research interest has also focused on further distributional approximations with subsampling for LRD transformed Gaussian processes; see [4, 6, 10, 26].

## 2. Preliminaries: LRD processes and block-based resampling estimators.

2.1. *Class of LRD processes.* Let  $\{Z_t\}$  be a mean zero, unit variance, stationary Gaussian sequence with covariances satisfying

$$(1) \quad \gamma_Z(k) \equiv \mathbb{E}Z_0Z_k \sim C_0k^{-\alpha}$$

as  $k \rightarrow \infty$  for some given  $0 < \alpha < 1$  and constant  $C_0 > 0$ ; above  $\sim$  denotes that sequences have a ratio of one in the limit. Examples include fractional Gaussian noise with Hurst parameter  $1/2 < H < 1$  having covariances  $\gamma_Z(k) = (|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H})/2$ , which satisfy (1) with  $C_0 = H(2H-1)$  and  $\alpha = 2-2H \in (0, 1)$ , as well as FARIMA processes with difference parameter  $0 < d < 1/2$ , which satisfy (1) with  $\alpha = 1-2d \in (0, 1)$ ; see [20, 35].

Let  $G : \mathbb{R} \rightarrow \mathbb{R}$  be a real-valued function such that  $\mathbb{E}[G(Z_0)]^2 < \infty$  holds for a standard normal variable  $Z_0$ . In which case, the function  $G(Z_0)$  may be expanded as

$$(2) \quad G(Z_0) = \sum_{k=0}^{\infty} \frac{J_k}{k!} H_k(Z_0)$$

in terms of Hermite polynomials,

$$H_k(z) \equiv (-1)^k e^{z^2/2} \frac{d^k}{dz^k} e^{-z^2/2}, \quad k = 0, 1, 2, \dots,$$

and corresponding coefficients  $J_k \equiv \mathbb{E}[G(Z_0)H_k(Z_0)]$ ,  $k \geq 0$ . The first few Hermite polynomials are given by  $H_0(z) = 1$ ,  $H_1(z) = z$ ,  $H_2(z) = z^2 - 1$ ,  $H_3(z) = z^3 - 3z$ , for example, and  $\mathbb{E}H_k(Z_0) = 0$  holds for  $k \geq 1$ . Let  $\mu \equiv \mathbb{E}G(Z_0) = J_0$  denote the mean of  $G(Z_0)$  and define the *Hermite rank* of  $G(\cdot)$  (cf. [49]) as

$$m \equiv \min\{j \geq 1 : J_j \neq 0\}.$$

To avoid degeneracy, we assume  $\text{Var}[G(Z_0)] > 0$  whereby  $m \in [1, \infty)$  is a finite integer.

The target processes of interest are defined as  $X_t \equiv G(Z_t)$  with respect to a stationary Gaussian series  $Z_t$  with covariances as in (1) with  $0 < \alpha < 1/m$ . Such processes  $X_t$  exhibit strong or long-range dependence (LRD) as seen by partial sums  $\sum_{k=1}^n |r(k)|$  of covariances  $r(k) \equiv \text{Cov}(X_0, X_k)$  having a slow decay proportional to

$$(3) \quad \sum_{k=1}^n |r(k)| \sim n^{1-\alpha m} [C_0^m / (1-\alpha m)] \quad \text{as } n \rightarrow \infty, \text{ where } \alpha m \in (0, 1),$$

depending on the Hermite rank  $m$  of the transformation  $G(\cdot)$  and memory exponent  $\alpha \in (0, 1/m)$  under (1). This represents a common formulation of LRD, with partial covariance sums diverging as  $n \rightarrow \infty$  [49]; see [44, 52] for further characterizations.

Suppose  $X_1, \dots, X_n$  is an observed time stretch from the transformed Gaussian series  $X_t \equiv G(Z_t)$ , having sample mean  $\bar{X}_n = n^{-1} \sum_{t=1}^n X_t$ . Setting  $v_{n,\alpha m} \equiv n^{\alpha m} \text{Var}(\bar{X}_n)$ , the process structure (1)–(3) entails a so-called long-run variance as

$$(4) \quad \lim_{n \rightarrow \infty} v_{n,\alpha m} = v_{\infty,\alpha m} \equiv \frac{J_m^2}{m!} \frac{2C_0^m}{(1 - \alpha m)(2 - \alpha m)} > 0$$

(cf. [1, 49, 50]). Under LRD, the variance  $\text{Var}(\bar{X}_n)$  of the sample mean decays at a slower rate  $O(n^{-\alpha m})$  as  $n \rightarrow \infty$  (i.e.,  $\alpha m \in (0, 1)$ ) than the typical  $O(n^{-1})$  rate under SRD. The limit distribution of  $n^{\alpha m/2}(\bar{X}_n - \mu)$  also depends on the Hermite rank  $m \geq 1$  [15, 50].

The development first considers the variance  $v_{n,\alpha m} \equiv n^{\alpha m} \text{Var}(\bar{X}_n)$  of the sample mean (or, equivalently here, its limit (4)) as target of inference under LRD. Resampling results are then extended to broader classes of statistics in Section 5.

**2.2. Block-based resampling variance estimators under LRD.** A block bootstrap “re-creates” the original series  $X_1, \dots, X_n$  by independently resampling  $b \equiv \lfloor n/\ell \rfloor$  blocks from a collection of length  $\ell < n$  data blocks. Resampling from the overlapping (OL) blocks  $\{(X_i, \dots, X_{i+\ell-1}) : i = 1, \dots, n - \ell + 1\}$  of length  $\ell$  within  $X_1, \dots, X_n$  yields the moving block bootstrap [29, 31, 34], while resampling from nonoverlapping (NOL) blocks  $\{(X_{1+\ell(i-1)}, \dots, X_{\ell i}) : i = 1, \dots, b \equiv \lfloor n/\ell \rfloor\}$  gives the NOL block bootstrap [13, 32]. Resampled blocks are concatenated to produce a bootstrap series, say  $X_1^*, \dots, X_{\ell b}^*$ , and the distribution of a statistic from the bootstrap series (e.g.,  $\bar{X}_{\ell b}^* \equiv (\ell b)^{-1} \sum_{i=1}^{\ell b} X_i^*$ ) approximates the sampling distribution of an original-data statistic (e.g.,  $\bar{X}_n$ ). Subsampling [41, 42] is a different approach to approximation that computes statistics from one resampled data block. Both subsampling and bootstrap, though, estimate a sample mean’s variance  $v_{n,\alpha m} \equiv n^{\alpha m} \text{Var}(\bar{X}_n)$  with a common block-based estimator; this is the induced variance of an average under resampling (e.g.,  $\text{Var}_*(\bar{X}_{\ell b}^*)$ ), which has a closed form (cf. [30] under LRD), resembling a batch means estimator [17]. Based on  $X_1, \dots, X_n$ , the OL block-based variance estimator of  $v_{n,\alpha m} \equiv n^{\alpha m} \text{Var}(\bar{X}_n)$  is given by

$$(5) \quad \hat{V}_{\ell,\alpha m,\text{OL}} = \frac{1}{n - \ell + 1} \sum_{i=1}^{n-\ell+1} \ell^{\alpha m} (\bar{X}_{i,\ell} - \hat{\mu}_{n,\text{OL}})^2, \quad \hat{\mu}_{n,\text{OL}} = \frac{1}{n - \ell + 1} \sum_{i=1}^{n-\ell+1} \bar{X}_{i,\ell},$$

where above  $\bar{X}_{i,\ell} = \sum_{j=i}^{i+\ell-1} X_j / \ell$  is the sample average of the  $i$ th data block  $(X_i, \dots, X_{i+\ell-1})$ ,  $i = 1, \dots, n - \ell + 1$ . Essentially, block versions  $\{\ell^{\alpha m/2}(\bar{X}_{i,\ell} - \bar{X}_n)\}_{i=1}^{n-\ell+1}$  of the quantity  $n^{\alpha m/2}(\bar{X}_n - \mu)$  give a sample variance  $\hat{V}_{\ell,\alpha m,\text{OL}}$  that estimates  $v_{\ell,\alpha m} \equiv \ell^{\alpha m} \text{Var}(\bar{X}_\ell) \approx v_{n,\alpha m} \equiv n^{\alpha m} \text{Var}(\bar{X}_n)$  for sufficiently large  $\ell, n$  by (4). The NOL block-based variance estimator is defined as

$$\hat{V}_{\ell,\alpha m,\text{NOL}} = \frac{1}{b} \sum_{i=1}^b \ell^{\alpha m} (\bar{X}_{1+(i-1)\ell,\ell} - \hat{\mu}_{n,\text{NOL}})^2, \quad \hat{\mu}_{n,\text{NOL}} = \sum_{i=1}^b \bar{X}_{1+(i-1)\ell,\ell} / b,$$

using NOL averages  $\bar{X}_{1+(i-1)\ell,\ell}$ ,  $i = 1, \dots, b \equiv \lfloor n/\ell \rfloor$ , where  $\hat{\mu}_{n,\text{NOL}} = \bar{X}_n$  when  $n = b\ell$ .

Under SRD, variance estimators are standardly defined by letting  $\alpha m = 1$  above (e.g., in  $\hat{V}_{\ell,\alpha m,\text{OL}}$  from (5)). Likewise, under LRD, both the target variance  $v_{n,\alpha m} \equiv n^{\alpha m} \text{Var}(\bar{X}_n)$  and block-based estimators  $\hat{V}_{\ell,\alpha m,\text{OL}}$  or  $\hat{V}_{\ell,\alpha m,\text{NOL}}$  are scaled to be comparable, which involves the long-memory exponent  $\alpha m \in (0, 1)$ . In practice,  $\alpha m \in (0, 1)$  is usually unknown.

To develop block-based estimators under LRD, we first consider  $\alpha m \in (0, 1)$  as given. Ultimately, an estimate  $\widehat{\alpha m}_n$  of  $\alpha m$  can be substituted into  $\widehat{V}_{\ell, \alpha m, \text{OL}}$  or  $\widehat{V}_{\ell, \alpha m, \text{NOL}}$  which, under mild conditions, does not change conclusions about best block selection or consistency (cf. Section 4.2).

**3. Properties for block-based resampling estimators under LRD.** Large-sample results for the block-based variance estimators require some extended notions of the Hermite rank of  $G(\cdot)$  in defining  $X_t \equiv G(Z_t) = \sum_{k=0}^{\infty} J_k/k! \cdot H_k(Z_t)$ , for  $J_k = \text{EG}(Z_0)H_k(Z_0)$ ,  $k \geq 1$ . Recalling  $m \equiv \min\{k > 0 : J_k \neq 0\}$  as the usual Hermite rank of  $G(\cdot)$ , define the *second Hermite rank* of  $G(\cdot)$  by the index

$$m_2 \equiv \min\{k > m : J_k \neq 0\}$$

of the next highest nonzero coefficient in the Hermite expansion (2) of  $G(\cdot)$ . In other words,  $m_2$  is the Hermite rank of  $X_t - \mu - J_m H_m(Z_t)/m!$  upon removing the mean and first Hermite rank term from  $X_t = G(Z_t)$ . If the set  $\{k > m : J_k \neq 0\}$  is empty, we define  $m_2 = \infty$ . We also define the *Hermite pair-rank* of a function  $G(\cdot)$  by

$$m_p \equiv \inf\{k \geq m : J_k J_{k+1} \neq 0\};$$

when the above set is empty, we define  $m_p = \infty$ . The Hermite pair-rank  $m_p$  identifies the index of the first consecutive pair of nonzero terms  $(J_k, J_{k+1})$  in the expansion  $X_t = G(Z_t) = \mu + \sum_{k=1}^{\infty} J_k/k! H_k(Z_t)$ . For nondegenerate series  $X_t = G(Z_t)$ , the Hermite rank  $m$  is always finite, but the second rank  $m_2$  and pair-rank  $m_p$  may not be (and  $m_2 = \infty$  implies  $m_p = \infty$ ). For example, both series  $G(X_t) = H_1(Z_t)$  and  $G(X_t) = H_1(Z_t) + H_3(Z_t)$  have Hermite rank  $m = 1$ , pair-rank  $m_p = \infty$ , and second ranks  $m_2 = \infty$  and 3, respectively; the series  $G(X_t) = H_1(Z_t) + H_3(Z_t) + H_4(Z_t)$  and  $G(X_t) = H_3(Z_t) + H_4(Z_t)$  have pair-rank  $m_p = 3$  with Hermite ranks  $m = 1$  and 3, and second ranks  $m_2 = 3$  and 4, respectively.

In what follows, due to combined effects of dependence and nonlinearity in a LRD time series  $X_t = G(Z_t)$ , the Hermite pair-rank  $m_p \in [m, \infty]$  of  $G$  plays a role in the asymptotic variance of resampling estimators (Section 3.2), while the second Hermite rank  $m_2 \in [m + 1, \infty]$  impacts the bias of resampling estimators (Section 3.1).

**3.1. Large-sample bias properties.** Bias expansions for the block resampling estimators require a more detailed form of the LRD covariances than (1) and we suppose that

$$(6) \quad \gamma_Z(k) \equiv \text{Cov}(Z_0, Z_k) = C_0 k^{-\alpha} (1 + k^{-\tau} L(k)), \quad k > 0,$$

holds for some  $\alpha \in (0, 1/m)$  and  $C_0 > 0$  (again  $\gamma_Z(0) = 1$ ) with some  $\tau \in (1 - \alpha m, \infty)$  and slowly varying function  $L : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  that satisfies  $\lim_{x \rightarrow \infty} L(ax)/L(x) = 1$  for any  $a > 0$ . For Gaussian FARIMA (i.e.,  $\alpha = 1 - 2d \in (0, 1)$ ) and fractional Gaussian noise (i.e.,  $\alpha = 2 - 2H \in (0, 1)$ ) processes  $\{Z_t\}$ , one may verify that (6) holds with  $\tau = 1$  for any  $\alpha \in (0, 1/m)$  and  $m \geq 1$ . The statement of bias in Theorem 3.1 additionally requires process constants  $\mathcal{B}_0(m)$ ,  $\mathcal{B}_1(m_2)$  that depend on the first  $m$  and second  $m_2$  Hermite ranks and covariances in (6). These are given as  $\mathcal{B}_1(m_2) \equiv 2 \sum_{j=m_2}^{\infty} (J_j^2/j!) \sum_{k=1}^{\infty} [\gamma_Z(k)]^j$  when  $\alpha m_2 > 1$ ;  $\mathcal{B}_1(m_2) \equiv 2C_0^{m_2} J_{m_2}^2/m_2!$  when  $\alpha m_2 = 1$ ;  $\mathcal{B}_1(m_2) \equiv 2J_{m_2}^2 C_0^{m_2}/[m_2!(1 - m_2\alpha)(2 - m_2\alpha)] = v_{\infty, \alpha m_2}$  from (4) when  $\alpha m_2 < 1$ ; and

$$(7) \quad \mathcal{B}_0(m) \equiv 2C_0^m \frac{J_m^2}{m!} \left\{ \mathcal{I}_{\alpha m} + \sum_{k=1}^{\infty} k^{-\alpha m} \sum_{j=1}^m \binom{m}{j} [L(k)k^{-\tau}]^j \right\} + \sum_{k=m}^{\infty} \frac{J_k^2}{k!},$$

with Euler's generalized constant  $\mathcal{I}_{\alpha m} \equiv \lim_{k \rightarrow \infty} (\sum_{j=1}^k j^{-\alpha m} - \int_0^k x^{-\alpha m} dx) \in (-\infty, 0)$ .



**THEOREM 3.1.** Suppose  $X_t \equiv G(Z_t)$  where the stationary Gaussian process  $\{Z_t\}$  satisfies (6) with memory exponent  $\alpha \in (0, 1/m)$  and where  $G(\cdot)$  has Hermite rank  $m$  and second Hermite rank  $m_2$  (noting  $m_2 > m$  and possibly  $m_2 = \infty$ ). Let  $\widehat{V}_{\ell, \alpha m}$  denote either  $\widehat{V}_{\ell, \alpha m, \text{OL}}$  or  $\widehat{V}_{\ell, \alpha m, \text{NOL}}$  as block resampling estimators of  $v_{n, \alpha m} = n^{\alpha m} \text{Var}(\bar{X}_n)$  based on  $X_1, \dots, X_n$ . If  $\ell^{-1} + \ell/n \rightarrow 0$  as  $n \rightarrow \infty$ , then the bias of  $\widehat{V}_{\ell, \alpha m}$  is given by

$$\begin{aligned} E\widehat{V}_{\ell, \alpha m} - v_{n, \alpha m} &= B_0(m) \left( \frac{\ell^{\alpha m}}{\ell} \right) (1 + o(1)) - v_{\infty, \alpha m} \left( \frac{\ell}{n} \right)^{\alpha m} (1 + o(1)) \\ &\quad + I(m_2 < \infty) B_1(m_2) \left( \frac{\ell^{\alpha m}}{\ell^{\min\{1, \alpha m_2\}}} \right) [\log \ell]^{I(\alpha m_2 = 1)} (1 + o(1)), \end{aligned}$$

where  $I(\cdot)$  denotes the indicator function, the constant  $v_{\infty, \alpha m} \equiv 2J_m^2 C_0^m / [m!(1 - \alpha m)(2 - \alpha m)]$  is from (4), and constants  $B_0(m)$ ,  $B_1(m_2)$  are from (7).

**REMARK 1.** If we switch the target of variance estimation from  $v_{n, \alpha m} = n^{\alpha m} \text{Var}(\bar{X}_n)$  to the limit variance  $v_{\infty, \alpha m} \equiv \lim_{n \rightarrow \infty} v_{n, \alpha m}$  from (4), this does not change the bias expansion in Theorem 3.1 or results in Section 4 on best block sizes for minimizing MSE.

To better understand the bias of a block-based estimator under LRD, we may consider the case of a purely Gaussian LRD series  $X_t = Z_t$  (i.e., no transformation), corresponding to  $G(x) = x$ ,  $m = 1$  and  $m_2 = \infty$ . The bias then simplifies under Theorem 3.1 as

$$(8) \quad E\widehat{V}_{\ell, \alpha m} - v_{n, \alpha m} = B_0(1) \left( \frac{\ell^\alpha}{\ell} \right) (1 + o(1)) - v_{\infty, \alpha m} \left( \frac{\ell}{n} \right)^\alpha (1 + o(1)),$$

depending only on the memory exponent  $\alpha$  of the process  $Z_t$ . This bias form can also hold when  $X_t$  is LRD and linear [18, 27]. However, for a transformed LRD series  $X_t = G(Z_t)$ , the function  $G$  and the underlying exponent  $\alpha$  impact the bias of the block-based estimator in intricate ways. The order of a main bias term in Theorem 3.1 is generally summarized as

$$(9) \quad O\left( \frac{[\log \ell]^{I(\alpha m_2 = 1)}}{\ell^{\min\{1, \alpha m_2\} - \alpha m}} \right),$$

which depends on how the second Hermite rank  $m_2 > m$  of the transformed series  $X_t \equiv G(Z_t)$ , as a type of nonlinearity measure, relates to the long-memory exponent  $\alpha \in (0, 1/m)$ . Small values of  $m_2$  satisfying  $1/\alpha > m_2$  induce the worst bias rates  $O(\ell^{-(m_2 - m)\alpha})$  compared to the best possible bias  $O(\ell^{-(1 - \alpha m)})$  occurring, for example, when  $m_2 = \infty$  (or no terms in the Hermite expansion of  $G(\cdot)$  beyond the first rank  $m$ ). In fact, the largest bias rates arise whenever second Hermite rank terms  $J_{m_2} H_{m_2}(Z_t)/m_2!$  exist in the expansion of  $X_t \equiv G(Z_t)$  (i.e.,  $J_{m_2} \neq 0$ ) and exhibit long-memory under  $\alpha m_2 < 1$  (cf. Section A.1 in [55]). For comparison, block-based estimators in the SRD case [31] exhibit a smaller bias  $O(1/\ell)$  than the best possible bias in (9) under LRD.

**3.2. Large-sample variance properties.** To establish the variance of the block resampling estimators under LRD, we require an additional moment condition regarding the transformed series  $X_t = G(Z_t)$ . For second moments, a simple characterization exists that  $EX_t^2 = E[G(Z_t)]^2 < \infty$  is finite if and only if  $\sum_{k=0}^{\infty} J_k^2/k! < \infty$ . For higher-order moments, however, more elaborate conditions are required to guarantee  $EX_t^4 = E[G(Z_t)]^4 < \infty$  and perform expansions of  $EX_{t_1} X_{t_2} X_{t_3} X_{t_4}$ . We shall use a condition “ $G \in \overline{\mathcal{G}}_4(1)$ ” from [50]. (More generally, Definition 3.2 of [50] prescribes a condition  $G \in \overline{\mathcal{G}}_4(\epsilon)$ , with  $\epsilon \in (0, 1]$ , for moment expansions, which could be applied to derive Theorem 3.2 next. We use  $\epsilon = 1$  for simplicity, where a sufficient condition for  $G \in \overline{\mathcal{G}}_4(1)$  is  $\sum_{k=0}^{\infty} 3^{k/2} |J_k|/\sqrt{k!} < \infty$ , holding

for any polynomial  $G$ , cf. [50]). See the Supplementary Material [55] for more technical details.

To state the large-sample variance properties of block-based estimators  $\widehat{V}_{\ell,\alpha m,\text{OL}}$  or  $\widehat{V}_{\ell,\alpha m,\text{NOL}}$  in Theorem 3.2, we need to introduce more constants. As a function of the Hermite rank, when  $m \geq 2$  and  $\alpha m < 1$ , define a positive scalar

$$\phi_{\alpha,m} \equiv \frac{2}{(1-2\alpha)(1-\alpha)} \left( \frac{2J_m^2 C_0^m}{(m-1)! [1-(m-1)\alpha][2-(m-1)\alpha]} \right)^2.$$

In the case of a Hermite rank  $m = 1$ , define another positive proportionality constant, as a function of  $\alpha \in (0, 1)$  and the type of resampling blocks (OL/NOL)), as

$$a_\alpha \equiv \frac{8J_1^4 C_0^2}{(1-\alpha)^2(2-\alpha)^2} \times \begin{cases} 1 + \frac{(2-\alpha)^2(2\alpha^2+3\alpha-1)}{4(1-2\alpha)(3-2\alpha)} - \frac{\Gamma^2(3-\alpha)}{\Gamma(4-2\alpha)} & \text{if } 0 < \alpha < 1/2, \text{ OL or NOL,} \\ 9/32 & \text{if } \alpha = 1/2, \text{ OL or NOL,} \\ \sum_{x=-\infty}^{\infty} g_\alpha^2(x) & \text{if } 1/2 < \alpha < 1, \text{ NOL,} \\ \int_{-\infty}^{\infty} g_\alpha^2(x) dx & \text{if } 1/2 < \alpha < 1, \text{ OL,} \end{cases}$$

where  $\Gamma(\cdot)$  denotes the gamma function and  $g_\alpha(x) \equiv (|x+1|^{2-\alpha} - 2|x|^{2-\alpha} + |x-1|^{2-\alpha})/2$ ,  $x \in \mathbb{R}$ . In the definition of  $a_\alpha$ ,  $g_\alpha^2(x)$  is summable/integrable when  $\alpha \in (1/2, 1)$  using  $g_\alpha(x) \sim (2-\alpha)(1-\alpha)x^{-\alpha}/2$  as  $x \rightarrow \infty$ . Finally, as a function of any Hermite pair-rank  $m_p \in [1, \infty]$  and  $\alpha \in (0, 1)$ , we define a constant as

$$\lambda_{\alpha,m_p} \equiv \frac{8C_0}{(1-\alpha)(2-\alpha)} \times \begin{cases} \left( \frac{2C_0^{m_p} J_{m_p} J_{m_p+1}}{m_p!} \right)^2 [(1-\alpha m_p)(2-\alpha m_p)]^{-2I(\alpha m_p < 1)} & \text{if } \alpha m_p \leq 1, \\ \left( \sum_{k=m_p}^{\infty} \sum_{j=-\infty}^{\infty} [\gamma_Z(j)]^k J_k J_{k+1}/k! \right)^2 & \text{if } \alpha m_p > 1, \end{cases}$$

with Gaussian covariances  $\gamma_Z(\cdot)$  and  $C_0 > 0$  from (1) and an indicator  $I(\cdot)$  function.

With constants  $\lambda_{\alpha,m_p}$ ,  $\phi_{\alpha,m}$ ,  $a_\alpha > 0$  as above, we may next state Theorem 3.2.

**THEOREM 3.2.** Suppose  $X_t \equiv G(Z_t)$  where the stationary Gaussian process  $\{Z_t\}$  satisfies (1) with  $C_0 > 0$  and memory exponent  $\alpha \in (0, 1/m)$  and where  $G \in \mathcal{G}_4(1)$  has Hermite rank  $m \geq 1$  and Hermite pair-rank  $m_p$  (note  $m_p \geq m$  and possibly  $m_p = \infty$ ). Let  $\widehat{V}_{\ell,\alpha m}$  denote either  $\widehat{V}_{\ell,\alpha m,\text{OL}}$  or  $\widehat{V}_{\ell,\alpha m,\text{NOL}}$  as block resampling estimators of  $v_{n,\alpha m} = n^{\alpha m} \text{Var}(\bar{X}_n)$  based on  $X_1, \dots, X_n$ . If  $\ell^{-1} + \ell/n \rightarrow 0$  as  $n \rightarrow \infty$ , then the variance of  $\widehat{V}_{\ell,\alpha m}$  is given by

$$\text{Var}(\widehat{V}_{\ell,\alpha m}) = \begin{cases} \phi_{\alpha,m} \left( \frac{\ell}{n} \right)^{2\alpha} (1 + o(1)) + r_{n,\alpha,m,m_p} & \text{if } m \geq 2, \\ a_\alpha \left( \frac{\ell}{n} \right)^{\min\{1, 2\alpha\}} [\log n]^{I(\alpha=1/2)} (1 + o(1)) + r_{n,\alpha,m,m_p} & \text{if } m = 1, \end{cases}$$

where  $I(\cdot)$  denotes an indicator function and

$$r_{n,\alpha,m,m_p} \equiv I(m_p < \infty) \frac{\lambda_{\alpha,m_p}}{n^\alpha} \left( \frac{\ell^{\alpha m}}{\ell^{\min\{1, \alpha m_p\}}} [\log \ell]^{I(\alpha m_p=1)} \right)^2 (1 + o(1)).$$

REMARK 2. Above  $r_{n,\alpha,m,m_p}$  represents a second variance contribution, which depends on the Hermite pair-rank  $m_p$  and is nonincreasing in block length  $\ell$  (by  $1/\alpha < m \leq m_p$ ). The value of  $r_{n,\alpha,m,m_p}$  is zero when  $m_p = \infty$  and is largest  $O(n^{-\alpha})$  when the pair-rank assumes its smallest possible value  $m_p = m$ . For example, the series  $X_t = H_m(Z_t) + H_{m+1}(Z_t)$  and  $X_t = H_m(Z_t) + H_{m+2}(Z_t)$  have pair-ranks  $m_p = 1$  and  $\infty$ , respectively, inducing different  $r_{n,\alpha,m,m_p}$  terms. While  $r_{n,\alpha,m,m_p}$  can dominate the variance expression of Theorem 3.2 for some block  $\ell$  sizes, the contribution of  $r_{n,\alpha,m,m_p}$  emerges as asymptotically negligible at an optimally selected block size  $\ell_{\text{opt}}$  (cf. Section 4.1).

By Theorem 3.2, the variance of a resampling estimator  $\widehat{V}_{\ell,\alpha m}$  depends on the block size  $\ell$  through a decay rate  $O((\ell/n)^{2\alpha})$  that, surprisingly, does not involve the exact value of the rank Hermite  $m$ . The reason is that, when  $m \geq 2$ , fourth-order cumulants of the transformed process  $X_t = G(Z_t)$  determine this variance (cf. [55]). Also, any differences in block type ( $\widehat{V}_{\ell,\alpha m,\text{OL}}$  vs.  $\widehat{V}_{\ell,\alpha m,\text{NOL}}$ ) only emerge in a proportionality constant  $a_\alpha$  when  $m = 1$  and  $\alpha \in (1/2, 1)$ ; otherwise,  $a_\alpha$  does not change with block type. Consequently, for processes  $X_t = G(Z_t)$  with strong LRD ( $\alpha < 1/2$ ,  $m \geq 1$ ), there is no large-sample advantage to OL blocks for variance estimation. In contrast, under SRD, OL blocks reduce the variance of a resampling variance estimator by a multiple of 2/3 compared to NOL blocks [29, 31, 32], because the nonoverlap between two OL blocks (e.g.,  $X_1, \dots, X_\ell$  and  $X_{1+i}, \dots, X_{\ell+i}$ ,  $i < \ell$ ) acts roughly uncorrelated. This fails under strong LRD where OL/NOL blocks have the same variance/bias/MSE properties here. Section 6 provides numerical examples. As under SRD, however, OL blocks remain generally preferable (i.e., smaller  $a_\alpha$  for weak LRD,  $\alpha > 1/2$ ).

#### 4. Best block selections and empirical estimation.

4.1. *Optimal block size and MSE.* Based on the large-sample bias and variance expressions in Section 3, an explicit form for the optimal block size  $\ell_{\text{opt}} \equiv \ell_{\text{opt},n}$  can be determined for minimizing the asymptotic mean squared error

$$(10) \quad \text{MSE}_n(\ell) \equiv E(\widehat{V}_{\ell,\alpha m} - v_{n,\alpha m})^2$$

of a block-based resampling estimator  $\widehat{V}_{\ell,\alpha m}$  of  $v_{n,\alpha m} \equiv n^{\alpha m} \text{Var}(\bar{X}_n)$  under LRD.

COROLLARY 4.1. *Under Theorems 3.1–3.2 assumptions, the optimal block size for a resampling estimator  $\widehat{V}_{\ell,\alpha m,\text{OL}}$  or  $\widehat{V}_{\ell,\alpha m,\text{NOL}}$  is given by (as  $n \rightarrow \infty$ )*

$$\ell_{\text{opt},n} = K_{\alpha,m,m_2} \times \begin{cases} n^{\frac{\alpha}{\alpha(1-m)+\min\{1,\alpha m_2\}}} (\log n)^{I(\alpha m_2=1)} & \text{if } 0 < \alpha < 0.5, m \geq 1, \\ n^{0.5} (\log n)^{-0.5+I(\alpha m_2=1)} & \text{if } \alpha = 0.5, m = 1, \\ n^{\frac{1}{3-2\alpha}} & \text{if } 0.5 < \alpha < 1, m = 1, \end{cases}$$

for a constant  $K_{\alpha,m,m_2} > 0$ , changing by block type OL/NOL only when  $m = 1$ ,  $\alpha \in (1/2, 1)$ .

The Appendix provides values for  $K_{\alpha,m,m_2} > 0$ . For LRD processes  $X_t = G(Z_t)$ , the best block lengths  $\ell_{\text{opt}}$  depend intricately on the transformation  $G(\cdot)$  (through ranks  $m$ ,  $m_2$ ) and the memory parameter  $\alpha < 1/m$  of the Gaussian process  $Z_t$ . Optimal blocks *increase* in length whenever the strength of long-memory *decreases* (i.e.,  $\alpha$  increases); as  $\alpha$  moves closer to  $1/m$ , the order of  $\ell_{\text{opt}}$  moves closer to  $O(n)$ . This is a counterintuitive aspect of LRD in resampling. With variance estimation under SRD [31, 32], best block size has a known order  $Cn^{1/3}$  where the process constant  $C > 0$  increases with dependence.

The second Hermite rank  $m_2$  of  $G(\cdot)$  can particularly impact  $\ell_{\text{opt}}$ . Whenever  $\alpha < 1/m_2$ , the optimal block order  $\ell_{\text{opt}} \propto n^{1/(m_2-m+1)}$  does not change, where “ $\propto$ ” denotes proportional



as  $n \rightarrow \infty$ . As a consequence in this case, if an immediate second term  $H_{m+1}(Z_0)$  appears in the Hermite expansion (2) of  $X_0 = G(Z_0)$ , so that the second rank is  $m_2 = m + 1$ , then the optimal block size becomes  $\ell_{\text{opt}} \propto n^{1/2}$ . This suggests that a guess  $\ell_{\text{opt}} = O(n^{1/2})$  often found in the literature for block resampling under LRD can be reasonable, though not by the intuition that slow covariance decay under LRD implies larger blocks compared to those  $O(n^{1/3})$  for SRD. Rather, for transformations  $G(\cdot)$  where  $m_2 = m + 1$  may hold naturally, the choice  $\ell_{\text{opt}} \propto n^{1/2}$  is optimal with sufficiently strong  $\alpha < 1/(m + 1)$  dependence, regardless of the exact Hermite rank  $m$ .

For completeness, we note that  $\text{MSE}_n(\ell) \equiv \mathbb{E}(\widehat{V}_{\ell, \alpha m} - v_{n, \alpha m})^2$  has an optimized order as

$$(11) \quad \text{MSE}_n(\ell_{\text{opt}, n}) \propto \begin{cases} n^{\frac{-2\alpha(\min\{1, \alpha m_2\} - \alpha m)}{\alpha(1-m) + \min\{1, \alpha m_2\}}} (\log n)^{2\alpha I(\alpha m_2=1)} & \text{if } 0 < \alpha < 0.5, m \geq 1, \\ n^{-0.5} (\log n)^{0.5 + I(\alpha m_2=1)} & \text{if } \alpha = 0.5, m = 1, \\ n^{-\frac{2(1-\alpha)}{3-2\alpha}} & \text{if } 0.5 < \alpha < 1, m = 1, \end{cases}$$

at the optimal block  $\ell_{\text{opt}} \equiv \ell_{\text{opt}, n}$ , which also depends on  $m, m_2$  and  $\alpha$  under LRD.

Section 4.2 shows that, under suitable conditions, estimation of the long-memory exponent does not change block considerations or consistency with the resampling variance estimators. Section 4.3 then provides a consistent data-driven method for estimating the block size  $\ell_{\text{opt}}$ .

**4.2. Empirical considerations for long-memory exponent.** We have assumed the memory exponent  $\alpha m \in (0, 1)$  of the LRD process  $X_t = G(Z_t)$  is known in the resampling estimators  $\widehat{V}_{\ell, \alpha m}$  for purposes of characterizing the effect of block size  $\ell$  on the error  $\text{MSE}_n(\ell) \equiv \mathbb{E}(\widehat{V}_{\ell, \alpha m} - v_{n, \alpha m})^2$  from (10). If an appropriate estimator  $\widehat{\alpha m}_n$  of  $\alpha m$  is instead substituted, then the resulting variance estimators can possess similar consistency rates.

Let  $\widehat{V}_\ell \equiv \widehat{V}_{\ell, \widehat{\alpha m}_n}$  denote a block-based estimator of  $v_{n, \alpha m} = n^{\alpha m} \text{Var}(\bar{X}_n)$  found by replacing  $\alpha m$  in  $\widehat{V}_{\ell, \alpha m}$  (e.g.,  $\widehat{V}_{\ell, \alpha m, \text{OL}}$  or  $\widehat{V}_{\ell, \alpha m, \text{NOL}}$  from (5)) with an estimator  $\widehat{\alpha m}_n$  based on  $X_1, \dots, X_n$ . Then a decomposition  $|\widehat{V}_\ell - v_{n, \alpha m}| = O_p(|\widehat{V}_\ell - \widehat{V}_{\ell, \alpha m}|) + O_p(|\widehat{V}_{\ell, \alpha m} - v_{n, \alpha m}|)$  follows. For stating Corollary 4.2 below, define an exponent

$$(12) \quad \kappa \equiv \kappa_{\alpha, m, m_2} \equiv \begin{cases} \frac{\alpha(\min\{1, \alpha m_2\} - \alpha m)}{\alpha(1-m) + \min\{1, \alpha m_2\}} & \text{if } 0 < \alpha < 0.5, m \geq 1, \\ \frac{(1-\alpha)}{3-2\alpha} & \text{if } 0.5 \leq \alpha < 1, m = 1, \end{cases}$$

which relates to the rate of optimized root  $[\text{MSE}_n(\ell_{\text{opt}, n})]^{1/2}$  from (11).

**COROLLARY 4.2.** *Suppose Theorem 3.1–3.2 assumptions. As  $n \rightarrow \infty$ :*

- (i) *if  $|\widehat{\alpha m}_n - \alpha m| \log n \xrightarrow{P} 0$ , then  $|\widehat{V}_\ell - \widehat{V}_{\ell, \alpha m}| = O_p(|\widehat{\alpha m}_n - \alpha m| \log n)$  holds and  $\widehat{V}_\ell$  is consistent for  $v_{n, \alpha m}$ , that is,  $|\widehat{V}_\ell - v_{n, \alpha m}| \xrightarrow{P} 0$ ;*
- (ii) *if  $|\widehat{\alpha m}_n - \alpha m| \log n = o_p(n^{-\kappa})$  for  $\kappa$  in (12), then  $|\widehat{V}_\ell - v_{n, \alpha m}| = O_p([\text{MSE}_n(\ell)]^{1/2})$ , where  $\text{MSE}_n(\ell) \equiv \mathbb{E}(\widehat{V}_{\ell, \alpha m} - v_{n, \alpha m})^2$  is determined by Theorems 3.1–3.2.*

Corollary 4.2(i) helps to separate the effects of block selection  $\ell$ , as our main interest, from those of  $\widehat{\alpha m}_n$  in  $\widehat{V}_\ell$ . Namely, the two resampling versions,  $\widehat{V}_{\ell, \alpha m}$  (i.e., assuming known  $\alpha m$ ) or  $\widehat{V}_\ell$  (i.e., estimating  $\alpha m$ ), can differ at most by an extent determined essentially by the convergence rate of  $\widehat{\alpha m}_n$ . Importantly, the block size  $\ell$  considerations developed for controlling the estimation error of  $\widehat{V}_{\ell, \alpha m}$  then be applied to  $\widehat{V}_\ell$ , as an issue apart from estimating  $\alpha m$ .

If convergence of  $\widehat{\alpha m}_n$  is sufficiently fast in Corollary 4.2(ii), then the same probabilistic bound  $O_p([\text{MSE}_n(\ell)]^{1/2})$  developed for  $|\widehat{V}_{\ell, \alpha m} - v_{n, \alpha m}|$  holds similarly for  $|\widehat{V}_\ell - v_{n, \alpha m}|$ . Consequently, the optimal block for minimizing  $[\text{MSE}_n(\ell)]^{1/2}$  sets a favorable convergence

rate for  $|\widehat{V}_\ell - v_{n,\alpha m}|$  in probability (i.e., given by the root of (11)). On the other hand, if the Corollary 4.2(ii) condition fails for  $\widehat{\alpha m}_n$ , then block  $\ell$  selection remains important, though the best possible convergence rate for  $|\widehat{V}_\ell - v_{n,\alpha m}|$  will be generally limited by an estimation error  $O_p(|\widehat{\alpha m}_n - \alpha m| \log n)$  outside of block selection. However, log-periodogram regression or local Whittle estimation of  $\alpha m$  [44, 45], for example, may be anticipated to satisfy the Corollary 4.2(ii) condition in many cases, given that a cruder estimator of  $\widehat{\alpha m}_n$  (e.g., by combining block-based variance estimators at two different block sizes (cf. [18])) can have a possible rate  $O_p(n^{-\kappa})$  from (11) for Gaussian subordinated processes. Some existing literature is also suggestive. For linear/Gaussian processes, log-periodogram or local Whittle estimators can exhibit a convergence rate  $|\widehat{\alpha m}_n - \alpha m| = O_p(n^{-2/5})$  (cf. [3, 25]), which guarantees  $o_p(n^{-\kappa} / \log n)$  due to  $\kappa \leq 1/4$ . Additionally, the Corollary 4.2(ii) condition also comports with some known rates  $O_p(n^{-\min\{\alpha, 1/2\}})$  from Whittle estimation of  $\widehat{\alpha m}_n$  with subordinated series  $X_t = G(Z_t)$  (cf. Theorem 3, [19]), as the exponent  $\kappa$  in (12) is strictly less than  $\alpha$ .

REMARK 3. Section 4.3 next considers empirical block selection, where we note that milder technical conditions are involved when estimating the memory exponent  $\alpha m$ . Rather than the Corollary 4.2(ii) assumption, we instead require  $|\widehat{\alpha m}_n - \alpha m| \log n = o_p(h^{-2\kappa})$ , where  $h \rightarrow \infty$  as  $n \rightarrow \infty$  represents a smaller time series length (i.e.,  $h = O(n^{1/2})$ ) as a type of user-chosen bandwidth in block selection; see Theorem 4.3. Weaker conditions on estimating  $\alpha m$  are possible because the block selection method is based on subsamples of time series, which are smaller order in size than the original data length  $n$ .

4.3. *Data-driven block estimation.* The block results from Section 4.1 suggest that data-driven choices of block size under LRD have no simple analogs to block resampling in the SRD case. For variance estimation under SRD, several approaches exist for estimating the best block size  $\ell_{\text{opt}}$  through plug-in estimation [12, 33, 43] or empirical MSE-minimization ([21]-method). By exploiting the known block order  $\ell_{\text{opt}} \approx Cn^{1/3}$  under SRD, these methods target the process constant  $C > 0$ . In contrast, optimal blocks under LRD have a form  $\ell_{\text{opt}} \approx Kn^a(\log n)^{\tilde{I}}$  from Corollary 4.1, where  $K \equiv K_{\alpha, m, m_2} > 0$  and  $a \equiv a_{\alpha, m, m_2} \in (0, 1)$  are complicated terms based on  $\alpha, m, m_2$ , while  $\tilde{I} \equiv -0.5I(\alpha = 0.5, m = 1) + I(\alpha m_2 = 1)$  involves indicator functions. Because the order  $n^a$  is unknown in practice, previous strategies to block estimation are not directly applicable in the LRD setting. Plug-in estimation seems particularly intractable under LRD; general plug-in approaches under SRD [33] require known orders for bias/variance in estimation, but these are also unknown under LRD (Theorems 3.1–3.2). Consequently, we consider a modified method for estimating block size  $\ell_{\text{opt}} \equiv \ell_{\text{opt}, n}$  that involves *two* rounds of empirical MSE-minimization ([21]-method). Unlike the SRD case, two rounds become necessary under LRD to estimate both an unknown order exponent  $a$  and scaling term  $K(\log n)^{\tilde{I}}$  in the block size  $\ell_{\text{opt}} \approx Kn^a(\log n)^{\tilde{I}}$ .

To adapt the [21]-method for LRD, we take a collection of subsamples  $(X_i, \dots, X_{i+h-1})$  of length  $h < n, i = 1, \dots, n - h + 1$ . Based on  $X_1, \dots, X_n$ , let  $\widehat{\alpha m}_n$  denote an estimator of  $\alpha m$  for use in resampling estimators to follow and, as in Section 4.2, write  $\widehat{V}_{\tilde{\ell}}$  as a resampling variance estimator based on a pilot block size  $\tilde{\ell}$  (e.g.,  $\tilde{\ell} \propto n^{1/2}$ ). Similarly, let  $\widehat{V}_{\tilde{\ell}}^{(i, h)}$  denote a resampling variance estimator computed on the subsample  $(X_i, \dots, X_{i+h-1})$  using a block length  $\ell < h, i = 1, \dots, n - h + 1$ . For clarity,  $\widehat{V}_{\tilde{\ell}}^{(i, h)}$  and  $\widehat{V}_{\tilde{\ell}}$  correspond to the sample variances of  $\{\ell^{\widehat{\alpha m}_n/2} \bar{X}_{j, \ell}\}_{j=i}^{h-\ell+i}$  and  $\{\tilde{\ell}^{\widehat{\alpha m}_n/2} \bar{X}_{j, \tilde{\ell}}\}_{j=1}^{n-\tilde{\ell}+1}$ , respectively, when based on OL block averages, writing  $\bar{X}_{j, \ell} \equiv \sum_{t=j}^{j+\ell-1} X_t / \ell$ , for any  $j, \ell \geq 1$ . We then define an initial

block-length estimator  $\widehat{\ell}_{\text{opt},h}$  as the minimizer of the empirical MSE,

$$\widehat{\text{MSE}}_{n,h}(\ell) \equiv \frac{1}{n-h+1} \sum_{i=1}^{n-h+1} (\widehat{V}_\ell^{(i,h)} - \widehat{V}_\ell)^2, \quad 1 \leq \ell < h.$$

Here,  $\widehat{\text{MSE}}_{n,h}(\ell)$  estimates  $\text{MSE}_h(\ell)$  in (10), or the MSE of a resampling estimator based on a sample of size  $h$  and block size  $\ell$ , while  $\widehat{\ell}_{\text{opt},h}$  then estimates the minimizer of  $\text{MSE}_h(\ell)$  or the optimal block  $\ell_{\text{opt},h}$  from Corollary 4.1 using “ $h$ ” in place of “ $n$ ” there. Above the pilot estimator  $\widehat{V}_\ell$  plays the role of a target variance to mimic the MSE formulation (10).

Theorem 4.3 establishes important conditions on the subsample size  $h$  and pilot block  $\tilde{\ell}$  for consistent estimation under LRD. For the transformed series  $X_t = G(Z_t)$ , the result involves a general eighth-order moment condition (i.e.,  $G \in \overline{\mathcal{G}}_8(1)$  under Definition 3.2 of [50]) analogous to the fourth-order moment condition described in Section 3.2.

**THEOREM 4.3.** *Assume that  $h \rightarrow \infty$  and  $\tilde{\ell} \rightarrow \infty$  with  $h/\tilde{\ell} + \tilde{\ell}h/n = O(1)$  as  $n \rightarrow \infty$ . Along with Theorem 3.1–3.2 assumptions, suppose also that  $G \in \overline{\mathcal{G}}_8(1)$  and that  $|\widehat{\alpha m}_n - \alpha m| \log n = o_p(h^{-2\kappa})$  for  $\kappa$  in (12). Then, as  $n \rightarrow \infty$ , the empirical MSE,  $\widehat{\text{MSE}}_{n,h}(\ell)$ , has a sequence of minimizers  $\widehat{\ell}_{\text{opt},h}$  such that*

$$\frac{\widehat{\ell}_{\text{opt},h}}{\ell_{\text{opt},h}} \xrightarrow{p} 1 \quad \text{and} \quad \frac{\widehat{\text{MSE}}_{n,h}(\widehat{\ell}_{\text{opt},h})}{\text{MSE}_h(\ell_{\text{opt},h})} \xrightarrow{p} 1,$$

where  $\ell_{\text{opt},h}$  has Corollary 4.1 form with  $\text{MSE}_h(\ell_{\text{opt},h})$  as in (11) (with “ $h$ ” replacing “ $n$ ”).

Theorem 4.3 does not address estimation of the best block size  $\ell_{\text{opt},n}$  for a length  $n$  time series, but rather the optimal block  $\ell_{\text{opt},h}$  for a smaller length  $h < n$  series. Nevertheless, the result establishes a nontrivial first step that, under LRD, some block sizes can be validly estimated through empirical MSE ([21]-method) provided that the subsample size  $h$  and pilot block  $\tilde{\ell}$  are appropriately chosen. In particular, the condition  $h/\tilde{\ell} + \tilde{\ell}h/n = O(1)$  cannot be reduced (related to pilot estimation  $\widehat{V}_\ell$ ) and entails that the largest subsample length possible is  $h = O(n^{1/2})$  within the empirical MSE approach under LRD.

Based on this and to resolve two unknown terms  $n^a$  and  $K(\log n)^{\tilde{I}}$  in  $\ell_{\text{opt},n} \approx Kn^a(\log n)^{\tilde{I}}$ , we use empirical MSE device twice, based on two subsample lengths  $h \equiv h_1 = \lfloor C_1 n^{1/r} \rfloor$  and  $h_2 = \lfloor C_2 n^{\theta/r} \rfloor$ . Here,  $r \geq 2$  and  $0 < \theta < 1$  are constants to control the subsample sizes (i.e.,  $h$  having larger order than  $h_2$ ) with  $C_1, C_2 > 0$ . A common pilot estimate  $\widehat{V}_\ell$  is used for both  $\widehat{\text{MSE}}_{n,h}(\ell)$  and  $\widehat{\text{MSE}}_{n,h_2}(\ell)$ . We denote corresponding block estimates as  $\widehat{\ell}_{\text{opt},h}$  and  $\widehat{\ell}_{\text{opt},h_2}$ , and define an estimator of the target optimal block size  $\ell_{\text{opt},n}$  as

$$\widehat{\ell}_{\text{opt},n} \equiv \left( \frac{\widehat{\ell}_{\text{opt},h}}{C_1^{\widehat{a}_n}} \right)^r \left( \frac{h^{\widehat{a}_n}}{\widehat{\ell}_{\text{opt},h}} \right)^{r-1} \widehat{c}_n, \quad \widehat{c}_n \equiv r \widehat{I}_n \left( \frac{\log h_2}{\log h} \right)^{(r-1)\widehat{I}_n(\log h)/\log(h/h_2)},$$

where

$$\widehat{a}_n \equiv \frac{\log(\widehat{\ell}_{\text{opt},h}/\widehat{\ell}_{\text{opt},h_2})}{\log(h/h_2)}, \quad \widehat{I}_n \equiv \frac{1}{2} \left( \frac{\log(\widehat{\ell}_{\text{opt},h}) - \widehat{a}_n \log h}{\log \log h} + \frac{\log(\widehat{\ell}_{\text{opt},h_2}) - \widehat{a}_n \log h_2}{\log \log h_2} \right)$$

estimate the exponent  $a \equiv a_{\alpha,m,m_2} \in (0, 1)$  and indicator quantity  $\tilde{I} \equiv -0.5I$  ( $\alpha = 0.5, m = 1$ ) +  $I(\alpha m_2 = 1)$  appearing in the Corollary 4.1 expression for  $\ell_{\text{opt},n} \approx Kn^a(\log n)^{\tilde{I}}$ . The estimator  $\widehat{\ell}_{\text{opt},n}$  has three components, where  $\widehat{\ell}_{\text{opt},h}/C_1^{\widehat{a}_n}$  estimates  $Kh^a(\log h)^{\tilde{I}}$  while  $\widehat{\ell}_{\text{opt},h}/h^{\widehat{a}_n}$  captures  $K(\log h)^{\tilde{I}}$  up to a constant, and  $\widehat{c}_n$  is scaling adjustment from  $\log n \approx r \log h$ . The data-driven block estimator  $\widehat{\ell}_{\text{opt},n}$  is provably valid over differing forms for  $\ell_{\text{opt},n}$  under LRD.

COROLLARY 4.4. Let  $h \equiv \lfloor C_1 n^{1/r} \rfloor$  and  $h_2 = \lfloor C_2 n^{\theta/r} \rfloor$  (for  $C_1, C_2 > 0$ ,  $r \geq 2$ ,  $\theta \in (0, 1)$ ), and suppose Corollary 4.3 assumptions hold. Then, as  $n \rightarrow \infty$ , the estimator  $\widehat{\ell}_{\text{opt},n}$  is consistent for  $\ell_{\text{opt},n}$  in that  $\widehat{\ell}_{\text{opt},n}/\ell_{\text{opt},n} \xrightarrow{p} 1$  and, additionally,

$$\widehat{a}_n \xrightarrow{p} a, \quad \widehat{I}_n \xrightarrow{p} \tilde{I}, \quad \widehat{c}_n \xrightarrow{p} [r\theta^{(r-1)/(1-\theta)}] \tilde{I}, \quad \frac{\widehat{\ell}_{\text{opt},h}}{h^{\widehat{a}_n}} \cdot \frac{1}{K(\log h)^{\tilde{I}}} \xrightarrow{p} \theta^{\tilde{I}/(1-\theta)},$$

regarding constants  $\tilde{I} \equiv -0.5I(\alpha = 0.5, m = 1) + I(\alpha m_2 = 1)$ ,  $K \equiv K_{\alpha, m, m_2} > 0$ , and  $a \equiv a_{\alpha, m, m_2}$  for prescribing  $\ell_{\text{opt},n} \approx K n^a (\log n)^{\tilde{I}}$  under Corollary 4.1.

We suggest a first subsample size  $h = C_1 n^{1/2}$  of maximal possible order ( $r = 2$ ). We then take the pilot block to be  $\tilde{\ell} = n^{1/2}$ , representing a reasonable choice under LRD and also satisfying Theorem 4.3, Corollary 4.4 (i.e.,  $h/\tilde{\ell} + h\tilde{\ell}/n = O(1)$  then holds). For a general rule in numerical studies to follow, we chose  $C_1 = 9$ ,  $C_2 = 12$ ,  $\theta = 0.95$  to keep subsamples adequately long under LRD. Based on empirical findings with Gaussian subordinated series (e.g., [5]), we use local Whittle estimation of  $\alpha m$ , for simplicity, with bandwidth  $\lfloor n^{0.7} \rfloor$  (cf. [3]). We also consider a modified block estimation rule

$$(13) \quad \widehat{\ell}_n = \min\{\lfloor n/20 \rfloor, \lfloor \widehat{\ell}_{\text{opt},n} \rfloor\},$$

to avoid overly large block selection in finite sample cases. This variation retains consistency due to  $\ell_{\text{opt},n} = o(n)$  and performs well over a variety of applications in Section 6.

**5. Extending the scope of statistics.** Here, we discuss extending block selection and re-sampling variance estimation to a larger class of statistics defined by functionals of empirical distributions. Using a small notational change to develop this section, let us denote data from an observed time stretch as  $Y_1 = \tilde{G}(Z_1), \dots, Y_n = \tilde{G}(Z_n)$  (rather than  $X_t = G(Z_t)$ ) and let  $F_n \equiv n^{-1} \sum_{t=1}^n \delta_{Y_t}$  denote the corresponding the empirical distribution, where  $\delta_y$  denotes a unit point mass at  $y \in \mathbb{R}$ . Consider a statistic  $T_n \equiv T(F_n)$ , given by a real-valued functional  $T(\cdot)$  of  $F_n$ , which estimates a target parameter  $T(F)$  defined by the process marginal distribution  $F$ . A broad class of statistics and parameters can be expressed through such functionals, with some examples given below.

EXAMPLE 1. Smooth functions  $T_n$  of averages given by

$$T_n \equiv H\left(n^{-1} \sum_{t=1}^n \phi_1(Y_t), \dots, n^{-1} \sum_{t=1}^n \phi_l(Y_t)\right),$$

involving a function  $H: \mathbb{R}^l \rightarrow \mathbb{R}$  of  $l \geq 1$  real-valued functions  $\phi_j: \mathbb{R} \rightarrow \mathbb{R}$  for  $j = 1, \dots, l$ . These statistics include ratios/differences of means as well as sample moments ([32], Chapter 5).

EXAMPLE 2. M-estimators  $T_n$  defined as the solution to

$$\frac{1}{n} \sum_{t=1}^n \psi(Y_t, T_n) = 0$$

for an estimating function with mean zero  $E[\psi(Y_t, T(F))] = 0$ . This includes several types of location/scale or regression estimators investigated in the LRD literature (cf. [7, 8]).

EXAMPLE 3. L-estimators  $T_n$  defined through integrals as

$$T_n = \int x J(F_n(x)) dF_n(x),$$

involving a bounded function  $J : [0, 1] \rightarrow \mathbb{R}$ . These include trimmed means  $J(x) = I(\delta_1 < x < \delta_2)/(\delta_2 - \delta_1)$  (based on the indicator function and trimming proportions  $\delta_1, \delta_2 \in (0, 1)$ ) along with Windsorized averages and Gini indices (cf. [47]).

For a fixed integer  $k \geq 1$ , functionals defined by linear combinations or products of components in “ $k$ -dimensional” marginal distributions might also be considered (i.e., empirical distributions of  $(Y_t, Y_{t+1}, \dots, Y_{t+k})$ ). For simplicity, we use  $k = 1$ . Under regularity conditions [16, 47], statistical functionals  $T_n = T(F_n)$  as above are approximately linear and admit an expansion

$$(14) \quad T_n = T(F) + \frac{1}{n} \sum_{t=1}^n \text{IF}(Y_t, F) + R_n,$$

in terms of the influence function  $\text{IF}(y, F)$ , defined as

$$\text{IF}(y, F) \equiv \lim_{\epsilon \downarrow 0} \frac{T((1 - \epsilon)F + \epsilon \delta_y) - T(F)}{\epsilon}, \quad y \in \mathbb{R},$$

and an appropriately small remainder  $R_n$ ; note that  $E[\text{IF}(Y_t, F)] = 0$  holds. See [14] and [24] for such expansions with LRD Gaussian subordinated processes.

To link to our previous block resampling developments (Section 2), a statistic  $T_n$  as in (14) corresponds approximately to an average  $\bar{X}_n = \sum_{t=1}^n X_t/n$  of transformed LRD Gaussian observations  $X_1, \dots, X_n$ , where  $X_t \equiv \text{IF}(Y_t, F) = \text{IF}(\tilde{G}(Z_t), F) = G(Z_t)$  has Hermite rank denoted by  $m$  with  $\alpha m < 1$  here. That is, under appropriate conditions, the normalized statistic  $n^{\alpha m/2}[T_n - T(F)] = n^{\alpha m/2}\bar{X}_n + o_p(1)$  has a distributional limit determined by  $\bar{X}_n$  (e.g., [49, 50]) with a limiting variance  $\lim_{n \rightarrow \infty} n^{\alpha m} \text{Var}(\bar{X}_n) = v_{\infty, \alpha m}$  given by (4) as before. Results in [5] also suggest that compositions  $X_t \equiv G(Z_t) = \text{IF}(\tilde{G}(Z_t), F)$  may tend to produce Hermite ranks of  $m = 1$ , in which case  $n^{\alpha/2}[T_n - T(F)]$  will be asymptotically normal with asymptotic variance  $v_{\infty, \alpha m}$ .

To estimate  $v_{\infty, \alpha m}$  through block resampling, we would ideally use  $X_1, \dots, X_n$  to obtain a variance estimator as in Section 2, which we denote as  $\hat{V}_{\ell, \alpha m} \equiv \hat{V}_{\ell, \alpha m}(X)$ . Then all estimation and block properties from Sections 3–4 would apply. Unfortunately,  $F$  is generally unknown in practice so that  $\{X_t \equiv \text{IF}(Y_t, F)\}_{t=1}^n$  are unobservable from the data  $Y_1, \dots, Y_n$ . Consequently,  $\hat{V}_{\ell, \alpha m}(X)$  represents an oracle estimator. In Sections 5.1–5.2, we detail two block-based strategies for estimating  $v_{\infty, \alpha m}$  based on either a substitution method or block jackknife. In both cases, these approaches can be as good as the oracle estimator  $\hat{V}_{\ell, \alpha m}(X)$  under some conditions. These resampling results under LRD have counterparts to the SRD case [29, 39], though we nontrivially include L-estimation in addition to M-estimation.

**5.1. Substitution method.** Classical substitution (i.e., plug-in) estimates  $F$  in the influence function  $\text{IF}(y, F)$  with its empirical version  $F_n$  (cf. [39, 47]) and develops observations as  $\hat{X}_1, \dots, \hat{X}_n$  with  $\hat{X}_t \equiv \text{IF}(Y_t, F_n)$ . For example, in a smooth function  $T_n = H(\bar{Y}_n)$  of the average  $\bar{Y}_n$ , we have  $\text{IF}(y, F_n) = H'(\bar{Y}_n)(y - \bar{Y}_n)$ , where  $H'$  denotes the derivative of  $H$ . We denote a resampling variance estimator computed from such observations as  $\hat{V}_{\ell, \alpha m}(\hat{X})$ .

To compare  $\hat{V}_{\ell, \alpha m}(\hat{X})$  to the oracle estimator  $\hat{V}_{\ell, \alpha m}(X)$ , we require bounds between estimated  $\text{IF}(y, F_n)$  and true influence functions  $\text{IF}(y, F)$ . For weakly dependent processes and M-estimators, [29] considered pointwise expansions of  $\text{IF}(y, F) - \text{IF}(y, F_n)$  as linear combinations of other functions in  $y$ . We need to generalize the concept of such expansions to accommodate LRD and more general functionals (e.g., L-estimators) as follows.



CONDITION I. There exist random variables  $U_{1,n}$ ,  $U_{2,n}$ ,  $W_n$  and real constants  $c, d \in \mathbb{R}$ ,  $C > 0$  such that, for any generic real values  $y_1, \dots, y_k$  and  $k \geq 1$ , it holds that

$$\left| \frac{1}{k} \sum_{j=1}^k \text{IF}(y_j, F) - \frac{1}{k} \sum_{j=1}^k \text{IF}(y_j, F_n) + U_{1,n} \right| \leq |W_n| + |U_{2,n}| \left( \int_c^d \left| \frac{1}{k} \sum_{j=1}^k h_\lambda(y_j) \right|^2 d\lambda \right)^{1/2},$$

where  $|W_n| = O_p(n^{-\alpha m})$ ;  $|U_{1,n}|, |U_{2,n}| = O_p(n^{-\alpha m/2})$ ; and, as indexed by  $\lambda \in [c, d]$ ,  $h_\lambda(\cdot)$  denotes a real-valued function such that  $h_\lambda(Y_t) = h_\lambda(\tilde{G}(Z_t))$  has mean zero, variance  $E[h_\lambda(Y_t)]^2 \leq C$ , and Hermite rank of at least  $m$  (the rank of  $G(Z_t) = \text{IF}(Y_t, F)$ ).

For context, if we set  $\alpha m = 1$  above and skip the notion of Hermite rank, then Condition I would include, as a special case, an assumption used by [29] with weakly dependent processes. However, under LRD, we need to explicitly incorporate Hermite ranks in bounds. If we define  $m_y \geq 1$  as the Hermite rank of an indicator function  $I(Y_t \leq y) = I(\tilde{G}(Z_t) \leq y)$  for  $y \in \mathbb{R}$ , then the smallest rank  $m^* \equiv \{m_y : y \in \mathbb{R}\}$  is known to be useful for describing convergence of the empirical distribution  $[F_n(\cdot) - F(\cdot)]$  (cf. [14]). One general way to ensure any function  $h_\lambda(\cdot)$  appearing in Condition I has Hermite rank of at least  $m$  (the rank of  $\text{IF}(Y_t, F)$ ) is that  $m = m^*$ . The reason is that  $m^*$  sets a lower bound on the Hermite rank of any function of  $Y_t$  (cf. (2.5) of [14]). Such equality  $m = m^*$  appears implicit in work of [24] on statistical functionals under LRD and holds automatically when  $m = 1$ . We show next that the statistics  $T_n$  in Examples 1–3 can satisfy Condition I.

THEOREM 5.1. For  $Y_t = \tilde{G}(Z_t)$ , suppose  $X_t \equiv G(Z_t) = \text{IF}(Y_t, F)$  has Hermite rank  $m \geq 1$  with  $\alpha m < 1$ . Then Condition I holds if the functional  $T_n$  is as in:

- (i) Example 1 (smooth function) where  $\phi_1, \dots, \phi_l$  are bounded functions; first partial derivatives of  $H : \mathbb{R}^l \rightarrow \mathbb{R}$  are Lipschitz in a neighborhood of  $(E[\phi_1(Y_t)], \dots, E[\phi_l(Y_t)])$ ; and either  $m = m^*$  holds or  $m = \min\{\text{Hermite rank of } \phi_j(Y_t) : 1 \leq j \leq l\}$ .
- (ii) Example 2 (M-estimation) where a constant  $C > 0$  and a neighborhood  $N_0$  of  $T(F)$  exist such that  $|\psi(y, \theta)| \leq C$  on  $\mathbb{R} \times N_0$ ;  $\dot{\psi} \equiv \partial\psi/\partial\theta$  exists and  $|\dot{\psi}(y, \theta)| \leq C$  on  $\mathbb{R} \times N_0$ ;  $|\dot{\psi}(y, \theta_1) - \dot{\psi}(y, \theta_2)| \leq C|\theta_1 - \theta_2|$  for  $y \in \mathbb{R}$ ,  $\theta_1, \theta_2 \in N_0$ ;  $E\dot{\psi}(Y_t, T(F)) \neq 0$ ; and either  $m = m^*$  holds or the Hermite rank of  $\psi(Y_t, \theta)$  remains the same for  $\theta \in N_0$ .
- (iii) Example 3 (L-estimation) where  $J$  is bounded and Lipschitz on  $[0, 1]$  with  $J(t) = 0$  when  $t \in [0, \delta_1] \cup [\delta_2, 1]$  for some  $0 < \delta_1 < \delta_2 < 1$ ; and either  $m = m^*$  holds or  $m \leq \min\{m_y : y_1 \leq y \leq y_2\}$  for some real  $y_1 < y_2$  with  $0 < F(y_1) < \delta_1 < \delta_2 < F(y_2) < 1$ .

Theorem 5.1 assumptions for Examples 1–2, dropping Hermite rank conditions, match those of [29]. Smooth function statistics in Example 1 have influence functions  $X_t = \text{IF}(Y_t, F)$  as a linear combination of the baseline functions  $\phi_j(Y_t)$ ,  $1 \leq j \leq l$ , so that the smallest Hermite rank among these typically gives the Hermite rank  $m$  of  $\text{IF}(Y_t, F)$ . In M-estimation, the Hermite rank of  $X_t \equiv G(Z_t) = \text{IF}(Y_t, F)$  matches that of  $\psi(Y_t, T(F))$  and it is sufficient that  $\psi(Y_t, \theta)$  maintains the same rank  $m$  in a  $\theta$ -neighborhood of  $T(F)$ ; the latter condition is mild and implies that the rank of  $\dot{\psi}(Y_t, T(F))$  must be at least  $m$ , which is important as  $\dot{\psi}(\cdot, T(F))$  arises in Condition I under M-estimation. To illustrate with a standard normal  $Y_t = Z_t$ , M-estimation of the process mean uses  $\psi(Z_t, \theta) = Z_t - \theta$  with a constant Hermite rank of 1 as a function of  $\theta$  and a derivative  $\dot{\psi}(Y_t, T(F)) = -1$  of infinite rank; similarly, Huber-estimation uses  $\psi(Z_t, \theta) = \max\{-c, \min\{Z_t - \theta, c\}\}$  (for some  $c > 0$ ), which has constant rank 1 for  $\theta$  in a neighborhood of  $T(F) = 0$  here, while the derivative  $\dot{\psi}(Z_t, T(F)) = I(|Z_t| \leq c)$  has rank 2. For general L-estimation, conditions on the Hermite ranks  $m_y$  of indicator functions  $I(Y_t \leq y)$  (or the empirical distribution  $F_n(y)$ ) are necessary,

particularly when trimming percentages  $\delta_1, \delta_2$  are involved; in this case, we may use the rank  $m_y$  of  $F_n(y)$  over a  $y$ -region  $([y_1, y_2])$  that is not trimmed away.

Theorem 5.2 establishes that the oracle resampling estimator  $\widehat{V}_{\ell, \alpha m}(X)$  (true influence) and the plug-in version  $\widehat{V}_{\ell, \alpha m}(\widehat{X})$  (estimated influence) are often close to the extent that the latter is as good as the former. Blocks can be either OL/NOL below.

**THEOREM 5.2.** *For  $Y_t = \tilde{G}(Z_t)$ , suppose  $X_t \equiv G(Z_t) = \text{IF}(Y_t, F)$  has Hermite rank  $m \geq 1$  with  $\alpha m < 1$ , Condition 1 holds, and  $\ell^{-1} + \ell/n \rightarrow 0$  as  $n \rightarrow \infty$ . Then*

$$\widehat{V}_{\ell, \alpha m}(\widehat{X}) = \widehat{V}_{\ell, \alpha m}(X) + O_p((\ell/n)^{\alpha m/2}) + O_p(n^{-\alpha m/2}).$$

Theorem 5.2 is the LRD analog of a result by [29] for weakly dependent processes (i.e., setting  $\alpha m = 1$  above). As in the SRD case, the difference between estimators is often no larger than the estimation error  $O_p((\ell/n)^{\min\{\alpha, 1/2\}})$  from the standard deviation of the oracle  $\widehat{V}_{\ell, \alpha m}(X)$  (Theorem 3.2). Consequently, optimal block orders and convergence rates for  $\widehat{V}_{\ell, \alpha m}(X)$  (Section 4.1) generally apply to the substitution version  $\widehat{V}_{\ell, \alpha m}(\widehat{X})$ . The block rule of Section 4.3 can also be applied to  $\widehat{X}_1, \dots, \widehat{X}_n$ , which we illustrate in Section 6.

**5.2. Block jackknife (BJK) method.** For estimating the asymptotic variance  $v_{\infty, \alpha m}$  of the functional  $T_n$ , a block jackknife (BJK) estimator is possible under LRD. BJK uses only OL data blocks, as NOL blocks are generally invalid (Remark 4.1, [29]). For  $j = 1, \dots, N \equiv n - \ell + 1$ , we compute the functional  $T_n^{(j)}$  after removing observations in  $j$ th OL block  $(Y_j, \dots, Y_{j+\ell-1})$  from the data  $(Y_1, \dots, Y_n)$ . The BJK estimator of  $v_{\infty, \alpha m}$  is then

$$\widehat{V}_{\ell, \alpha m, \text{OL}}^{\text{BJK}} = \frac{(N-1)^2 \ell^{\alpha m}}{\ell^2} \frac{1}{N} \sum_{j=1}^N (T_n^{(j)} - \bar{T}_n)^2, \quad \bar{T}_n \equiv \frac{1}{N} \sum_{j=1}^N T_n^{(j)}.$$

Unlike the plug-in method (Section 5.1), BJK does not involve influence functions, but uses repeated evaluations of the functional. For the sample mean statistic  $T_n = \sum_{t=1}^n Y_t/n$ , the BJK estimator matches the plug-in estimator  $\widehat{V}_{\ell, \alpha m}(\widehat{X}) \equiv \widehat{V}_{\ell, \alpha m, \text{OL}}(\widehat{X})$  with OL blocks (cf. [29]). More generally, these two estimators may differ, though not substantially, as shown in Theorem 5.3. To state the result, for each OL data block  $j = 1, \dots, N$ , we define a remainder  $S_n^{(j)} \equiv T_n^{(j)} - T_n - M_n^{(j)}$ , due to a type of Taylor expansion of  $T_n^{(j)}$  about  $T_n$ , where

$$M_n^{(j)} \equiv \frac{1}{n-\ell} \sum_{\substack{1 \leq t \leq n, \\ t \notin [j, j+\ell-1]}} \widehat{X}_t - \frac{1}{n} \sum_{t=1}^n \widehat{X}_t$$

involves an average of estimated values  $\{\widehat{X}_i \equiv \text{IF}(Y_i, F_n)\}_{i=1}^n$  after removing the  $j$ th block.

**THEOREM 5.3.** *For  $Y_t = \tilde{G}(Z_t)$ , suppose  $X_t \equiv G(Z_t) = \text{IF}(Y_t, F)$  has Hermite rank  $m \geq 1$  with  $\alpha m < 1$ , and that the OL block plug-in estimator  $\widehat{V}_{\ell, \alpha m, \text{OL}}(\widehat{X})$  is consistent. Then*

$$\widehat{V}_{\ell, \alpha m}^{\text{BJK}} = \widehat{V}_{\ell, \alpha m, \text{OL}}(\widehat{X}) + O_p(\ell/n)$$

*holds as  $n \rightarrow \infty$  if  $\ell^{\alpha m} \sum_{j=1}^N [S_n^{(j)}]^2/N = O_p(\ell^4/[n^2(N-1)^2])$ ; the latter is true under Theorem 5.1 assumptions for Examples 1–3.*

The above difference  $O_p(\ell/n)$  between BJK and plug-in estimators holds similarly under weak dependence (akin to setting  $\alpha m = 1$  above), which improves the bound  $O_p(\ell^{3/2}/n)$  originally given by [29] (Theorem 4.2). Theorems 5.2–5.3 show that BJK can also differ no more from the oracle estimator  $\widehat{V}_{\ell, \alpha m, \text{OL}}(X)$  than the plug-in estimator  $\widehat{V}_{\ell, \alpha m, \text{OL}}(\widehat{X})$ .

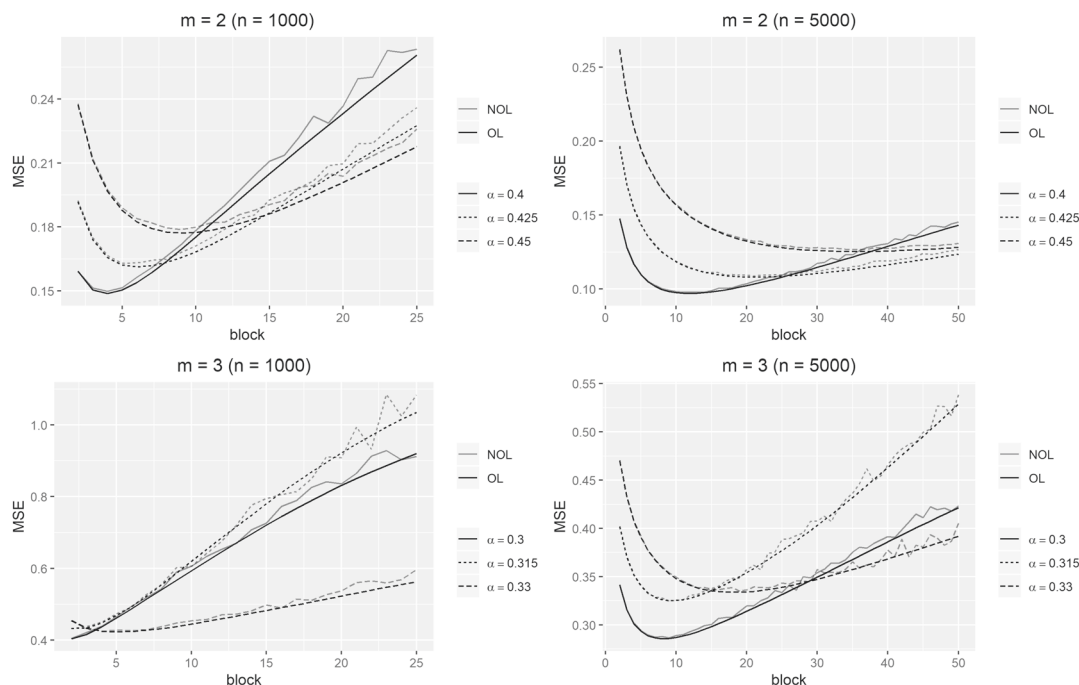


FIG. 1. MSE curves (over block length  $\ell$ ) for resampling estimators with different LRD processes  $(m, \alpha)$ .

## 6. Numerical illustrations and applications.

**6.1. Illustration of MSE over block sizes.** Here, we describe an initial numerical study of the MSE-behavior of resampling variance estimators under LRD. In particular, results of Section 3 suggest that OL/NOL resampling blocks should induce identical large-sample performances under strong dependence (e.g.,  $\alpha < 1/2$ ) and that optimal blocks should generally *decrease* in size as the covariance strength increases (cf. Section 4.1). LRD series were generated as  $X_t = H_2(Z_t)$  or  $X_t = H_3(Z_t)$ , using three values of the memory parameter with  $\alpha < 1/m$  for  $m = 2$  or  $m = 3$ , based on a standardized fractional Gaussian process  $Z_t$  with covariances as in (1) (i.e.,  $H = (2 - \alpha)/2$ ). For each simulated series, OL/NOL block-based estimators  $\hat{V}_{\ell, \alpha m}$  of the variance  $v_{n, \alpha m} = n^{\alpha m} \text{Var}(\bar{X}_n)$  were computed over a sequence of block sizes  $\ell$ . Repeating this procedure over 3000 simulation runs and averaging differences  $(\hat{V}_{\ell, \alpha m} - v_{n, \alpha m})^2 / v_{n, \alpha m}^2$  produced approximations of standardized MSE-curves  $E(\hat{V}_{\ell, \alpha m} - v_{n, \alpha m})^2 / v_{n, \alpha m}^2$ , as shown in Figure 1 with sample sizes  $n = 1000$  or  $5000$ . The MSE curves are quite close between OL/NOL blocks, particularly as sample sizes increase to  $n = 5000$ , in agreement with theory. Also, as suggested by Section 4.1, MSEs should improve at the best block choice as covariance strength increases under LRD ( $\alpha \downarrow$ ), which is visible in Figure 1. Table 1 presents best block lengths from the figure, showing that optimal blocks decrease for these LRD processes with decreasing  $\alpha$ . The Supplementary Material [55] provides additional simulation studies to further illustrate bias/variance behavior of resampling estimators.

**6.2. Resampling variance estimation by empirical block size.** We next examine empirical block choices for resampling variance estimation of the sample mean and provide comparison to other approaches under LRD. Application to another functional is then considered.

We use the data-based rule (13) of Section 4.3 for choosing a block size. We first compare resampling estimators  $\hat{V}_\ell$  of the sample mean's variance  $v_{n, \alpha m}$  between block selections

TABLE 1  
Optimal OL block sizes for LRD series with Hermite ranks  $m = 2$  and 3

$\alpha$	$m = 2$			$m = 3$		
	0.400	0.425	0.450	0.300	0.315	0.330
$n = 1000$	4	6	9	2	3	5
$n = 5000$	12	21	35	8	9	18

$\ell = \widehat{\ell}_n$  and  $\ell = \lfloor n^{1/2} \rfloor$ , where the latter represents a reasonable choice under LRD by theory in Section 4.1. OL blocks are used along with local Whittle estimation  $\widehat{\alpha m}$  of the memory parameter  $\alpha m$  (Section 4.3). Similar to Section 6.1, we simulated samples from LRD processes defined by  $X_t = H_m(Z_t)$  for  $m = 2$  with  $\alpha = 0.20, 0.45$  or  $m = 3$ ,  $\alpha = 0.20, 0.30$  and approximated the MSE  $E(\widehat{V}_\ell - v_{n,\alpha m})^2 / v_{n,\alpha m}^2$  using 500 simulations. Table 2 provides these results. Estimated blocks  $\widehat{\ell}_n$  are generally better than the default  $\lfloor n^{1/2} \rfloor$ , though the latter is also competitive. The default seems preferable with a small sample size and particularly strong dependence (e.g.,  $n = 500$ ,  $m = 3$ ,  $\alpha = 0.2$ ), but empirical block selections show improved MSEs with increased sample sizes  $n = 1000, 2000$  under LRD.

For comparison against resampling variance estimators, we also consider the Bartlett-kernel heteroskedasticity and autocorrelation consistent (HAC) estimator [54] and the memory and autocorrelation consistent (MAC) estimator [46], whose large-sample properties have been studied for the sample mean with linear LRD processes (cf. [1, 18]), but not for transformed LRD series  $X_t = G(Z_t)$ . As numerical suggestions from [1], we implemented HAC and MAC estimators of the sample mean's variance using bandwidths  $\lfloor n^{1/5} \rfloor$ ,  $\lfloor n^{4/5} \rfloor$ , respectively; the HAC approach further used local Whittle estimation of the memory parameter  $\alpha m$ , like the resampling estimator. The MSEs of HAC/MAC estimators are given in Table 3 (approximated from 500 simulation runs) for comparison against the resampling estimators in Table 2 with the same processes. For the process  $X_t = H_2(Z_t)$  with  $\alpha = 0.45$ , HAC/MAC estimators emerge as slightly better than the resampling approach with estimated block sizes, though the resampling estimator outperforms HAC/MAC estimators as the dependence increases (smaller  $\alpha$ ) or as the Hermite rank increases ( $m = 3$ ). With small sample sizes  $n$  and strong dependence, the HAC estimator can exhibit large MSEs, indicating that the bandwidth  $\lfloor n^{1/5} \rfloor$  is perhaps too small for the nonlinear LRD series in these settings. In comparison, the empirical block selections with resampling estimators show consistently reasonable MSE-performance among all cases, which is appealing.

We further consider a different statistical functional with resampling estimators and empirical blocks  $\widehat{\ell}_n$ . In the notation of Section 5, we simulated stretches  $Y_1, \dots, Y_n$  of LRD processes defined by  $Y_t = H_2(Z_t)$  or  $Y_t = \sin(Z_t)$  and considered an L-estimator  $T_n$  as a 40%

TABLE 2  
MSE for resampling estimators  $\widehat{V}_\ell$  of sample mean's variance based on blocks  $\ell = \widehat{\ell}_n$  or  $\ell = \lfloor n^{1/2} \rfloor$  with LRD series  $X_t \equiv G(Z_t)$  with  $G = H_2(Z_t)$  ( $m = 2$ ) and  $G(Z_t) = H_3(Z_t)$  ( $m = 3$ )

$G/m$	$\alpha$	$n = 500$		$n = 1000$		$n = 2000$	
		$\widehat{\ell}_n$	$\lfloor n^{1/2} \rfloor$	$\widehat{\ell}_n$	$\lfloor n^{1/2} \rfloor$	$\widehat{\ell}_n$	$\lfloor n^{1/2} \rfloor$
$m = 2$	0.20	0.294	0.316	0.236	0.248	0.180	0.214
	0.45	0.270	0.312	0.269	0.293	0.280	0.292
$m = 3$	0.20	0.917	0.754	0.805	0.858	0.455	0.495
	0.30	0.270	0.294	0.396	0.413	0.393	0.393

TABLE 3  
MSE for HAC and MAC estimators of sample mean's variance with LRD series  $X_t \equiv G(Z_t)$  with  $G = H_2(Z_t)$  ( $m = 2$ ) and  $G(Z_t) = H_3(Z_t)$  ( $m = 3$ )

$G/m$	$\alpha$	$n = 500$		$n = 1000$		$n = 2000$	
		HAC	MAC	HAC	MAC	HAC	MAC
$m = 2$	0.20	8.297	0.336	0.989	0.250	0.286	0.174
	0.45	0.256	0.252	0.265	0.257	0.279	0.267
$m = 3$	0.20	26.43	1.667	1.492	0.797	1.020	0.804
	0.30	0.515	0.442	0.379	0.378	0.369	0.358

trimmed mean based on the empirical distribution  $F_n$  (i.e.,  $\delta_1 = 1 - \delta_2 = 0.2$  in Example 3, Section 5). For either process, the influence function  $X_t \equiv \text{IF}(Y_t, F) = Y_t I(F^{-1}(0.2) < Y_t < F^{-1}(0.8))/0.6$  has Hermite rank  $m = 1$ , where  $F$  and  $F^{-1}$  denote the distribution and quantile functions, respectively, of  $Y_t$ . To estimate the variance, say  $v_{n,\alpha m}$ , of  $n^{\alpha m/2} T_n$ , we apply the substitution method (Section 5.1). That is, using estimated influences  $\hat{X}_t \equiv \text{IF}(Y_n, F_n) = Y_t I(F_n^{-1}(0.2) < Y_t < F_n^{-1}(0.8))/0.6$ , we obtain an estimator  $\hat{\alpha m}$  of the memory-parameter by local Whittle estimation and compute a plug-in resampling variance estimator  $\hat{V}_\ell(\hat{X})$ . Table 4 provides MSEs (i.e.,  $E(\hat{V}_\ell(\hat{X}) - v_{n,\alpha m})^2 / v_{n,\alpha m}^2$ ) approximated from 500 simulation runs) with block choices  $\ell = \hat{\ell}_n$  or  $\lfloor n^{1/2} \rfloor$  over sample sizes  $n = 500, 1000, 2000$ . The empirical block selections perform better than the default  $\lfloor n^{1/2} \rfloor$  with the plug-in variance estimator here, though the choice  $\lfloor n^{1/2} \rfloor$  appears also reasonable.

6.3. *Resampling distribution estimation by empirical block size.* Block selection also plays an important role in other resampling inference, such as approximating full sampling distributions with block bootstrap for purposes of tests and confidence intervals. While optimal block sizes for distribution estimation are difficult and unknown under LRD, we may apply blocking notions developed here for guidance. For distributional approximations of sample means and other statistics as in Section 5, the block bootstrap is valid with transformed LRD series when a normal limit exists (e.g., Hermite rank  $m = 1$ ) [30]. Such normality may occur commonly in practice [5] and can be further assessed as described in Section 6.4.

To study empirical blocks for distribution estimation with the bootstrap, we consider two LRD processes as  $Y_t = \sin(Z_t)$  or  $Y_t = Z_t + 20^{-1} H_2(Z_t)$  defined by Gaussian  $\{Z_t\}$  as before with memory exponent  $\alpha$ . Based on a size  $n$  sample, block bootstrap is applied to approximate the distribution of  $\Delta_n \equiv n^{\alpha m/2} |T_n - \theta|$ , where  $T_n \equiv T(F_n)$  represents either the sample mean or the 40% trimmed mean (Example 3, Section 5), while  $\theta \equiv T(F)$  denotes the corresponding process mean or trimmed mean parameter. In sample mean case,

TABLE 4  
MSE for plug-in resampling estimators  $\hat{V}_\ell(\hat{X})$  of the variance of the 40% trimmed mean statistic  $T_n$  based on blocks  $\ell = \hat{\ell}_n$  or  $\ell = \lfloor n^{1/2} \rfloor$  with LRD series  $Y_t \equiv H_2(Z_t)$  or  $Y_t = \sin(Z_t)$

$Y_t$	$\alpha$	$n = 500$		$n = 1000$		$n = 2000$	
		$\hat{\ell}_n$	$\lfloor n^{1/2} \rfloor$	$\hat{\ell}_n$	$\lfloor n^{1/2} \rfloor$	$\hat{\ell}_n$	$\lfloor n^{1/2} \rfloor$
$H_2(Z_t)$	0.20	0.463	0.494	0.383	0.413	0.366	0.402
	0.45	0.568	0.603	0.565	0.585	0.565	0.575
$\sin(Z_t)$	0.20	0.608	0.641	0.597	0.613	0.590	0.606
	0.30	0.561	0.589	0.565	0.582	0.550	0.565



TABLE 5

Empirical coverage probabilities of 95% block bootstrap confidence intervals for the process mean or 40% trimmed mean, based on block sizes  $\ell = \hat{\ell}_n$  or  $\ell = \lfloor n^{0.5} \rfloor$ , with LRD series as  $\sin(Z_t)$  or  $Z_t + 20^{-1}H_2(Z_t)$

$\ell$		$\alpha = 0.20$		$\alpha = 0.50$		$\alpha = 0.80$	
		$n = 1000$	5000	$n = 1000$	5000	$n = 1000$	5000
Mean							
$\sin(Z_t)$	$\hat{\ell}_n$	0.820	0.866	0.932	0.944	0.964	0.958
	$\lfloor n^{1/2} \rfloor$	0.766	0.814	0.878	0.914	0.958	0.934
$Z_t + \frac{1}{20}H_2(Z_t)$	$\hat{\ell}_n$	0.848	0.864	0.922	0.930	0.942	0.960
	$\lfloor n^{1/2} \rfloor$	0.806	0.854	0.914	0.894	0.938	0.952
Trimmed Mean							
$\sin(Z_t)$	$\hat{\ell}_n$	0.798	0.848	0.934	0.942	0.964	0.950
	$\lfloor n^{1/2} \rfloor$	0.752	0.792	0.892	0.920	0.962	0.946
$Z_t + \frac{1}{20}H_2(Z_t)$	$\hat{\ell}_n$	0.864	0.876	0.930	0.916	0.956	0.980
	$\lfloor n^{1/2} \rfloor$	0.828	0.852	0.918	0.908	0.876	0.754

we compute  $\widehat{\alpha m}$  using local Whittle estimation with data stretch  $X_1 = Y_1, \dots, X_n = Y_n$  and define a bootstrap average  $\bar{X}_{b\ell}^*$  by resampling  $b \equiv \lfloor n/\ell \rfloor$  OL data blocks of length  $\ell$  (see Section 2.2); the bootstrap version of  $\Delta_n$  is then  $\Delta_n^* \equiv b^{1/2} \ell^{\widehat{\alpha m}/2} |\bar{X}_{b\ell}^* - E_* \bar{X}_{b\ell}^*|$  (cf. [30]), where  $E_* \bar{X}_{b\ell}^* = \sum_{i=1}^{n-\ell+1} \bar{X}_{i,\ell} / (n - \ell + 1)$  is a bootstrap expected average. In the trimmed mean case, the estimator  $\widehat{\alpha m}$  and the bootstrap approximation  $\Delta_n^*$  are similarly defined from estimated values  $\hat{X}_t \equiv \text{IF}(Y_n, F_n) = Y_t I(F_n^{-1}(0.2) < Y_t < F_n^{-1}(0.8)) / 0.6$ ,  $t = 1, \dots, n$ . We construct 95% bootstrap confidence intervals for  $\theta$  by approximating the 95th percentile of  $\Delta_n$  with the bootstrap counterpart from  $\Delta_n^*$  (based on 200 bootstrap recreations). Note that, for these processes and statistics, the effective Hermite rank is  $m = 1$  (i.e., the rank of  $X_t = \text{IF}(Y_t, F)$ ) so that the bootstrap should be valid in theory.

We used the empirical rule (13) as a guide for selecting a block length  $\ell$ . Table 5 shows the empirical coverages of 95% bootstrap intervals with samples of size  $n = 1000$  or  $n = 5000$  (based on 500 simulation runs). For strongest LRD  $\alpha = 0.02$ , bootstrap intervals exhibit undercoverage, as perhaps expected, though accuracy improves with increasing sample size  $n$  in this case. The bootstrap performs well in the other cases of long memory. The coverage rates of bootstrap intervals are closer to the nominal level with empirically chosen blocks  $\hat{\ell}_n$  compared to a standard choice  $\lfloor n^{1/2} \rfloor$ , for both the sample mean and trimmed mean. This suggests that the driven-data rule for blocks provides a reasonable guidepost for resampling distribution estimation, as an application beyond variance estimation.

**6.4. A test of Hermite rank/normality.** In a concluding numerical example, we wish to illustrate that data blocking has impacts for inference under LRD beyond the resampling. One basic application of data blocks is for testing the null hypothesis that the Hermite rank is  $m = 1$  for a transformed LRD process  $X_t = G(Z_t)$  against the alternative  $m > 1$ . This type of assessment has practical value in application. For example, analyses in financial econometrics can involve LRD models with assumptions about  $m$  (cf. [11]). More generally, inference from sample averages under LRD may use normal theory only if  $m = 1$  [49, 50]. Even considering resampling approximations under LRD, the block bootstrap (i.e., full data recreation) becomes valid when  $m = 1$  [30], while subsampling (i.e., small scale recreation) should be used instead if  $m > 1$  [4, 10, 21].

**Algorithm 1:** Hermite rank test for  $m = 1$  (normality) from LRD series**Data:** Given a LRD sample  $X_1, \dots, X_n$ .Set initializations: block size  $\ell$ ; number  $M$  of resamples; and significance level $\alpha_{\text{sig}} = 0.05$ .Step 1. Calculate a test statistic  $T_0$  for normality (e.g., Anderson–Darling) from block averages;Step 2. Estimate  $\widehat{\alpha m}$  the memory parameter  $\alpha m$  or Hurst Index  $\widehat{H} = 1 - \widehat{\alpha m}/2$ ;Step 3. **for**  $k = 1, \dots, M$  **do**(i) Simulate a fractional Brownian motion sample  $\{B_{\widehat{H}}^*(\frac{1}{n}), \dots, B_{\widehat{H}}^*(\frac{n}{n})\}$  with Hurst index  $\widehat{H}$ ;(ii) Obtain a bootstrap sample  $X_1^*, \dots, X_n^*$  as  $X_1^* = B_{\widehat{H}}^*(\frac{1}{n})$  and $X_j^* = B_{\widehat{H}}^*(\frac{j}{n}) - B_{\widehat{H}}^*(\frac{j-1}{n}), j = 2, \dots, n$ ;(iii) Calculate the  $k$ th bootstrap test statistic,  $T_k^*$ , for normality from block averages in  $X_1^*, \dots, X_n^*$ ;**end**Step 4. Compute  $\widehat{q}$  as the  $1 - \alpha_{\text{sig}}$  sample percentile of  $\{T_1^*, \dots, T_M^*\}$ ;Step 5. Reject if  $T_0 > \widehat{q}$ .

Based on data  $X_1, \dots, X_n$  from a LRD process  $X_t = G(Z_t)$ , a simple assessment of  $H_0 : m = 1$  can be based on data blocks of  $\ell$  as follows. The idea is to make averages (say)  $W_i \equiv \sum_{j=1}^{\ell} X_{j+\ell(i-1)}/\ell$  of length  $\ell$  blocks,  $i = 1, \dots, b \equiv \lceil n/\ell \rceil$ , as in Section 2.2, and then check their agreement to normality. Letting  $\Phi(\cdot)$  denote a standard normal cdf, we compare the collection of residuals  $R_i \equiv \Phi((W_i - \bar{W})/S_W)$  to a uniform(0, 1) distribution, where  $\bar{W}$ ,  $S_W$  are the average and standard deviation of  $\{W_1, \dots, W_b\}$ . In a usual fashion, we can assess uniformity by applying a Kolmogorov–Smirnov statistic or an Anderson–Darling [2] statistic (e.g.,  $A \equiv -b - b^{-1} \sum_{i=1}^b (2i-1) \log\{R_{(i)}[1 - R_{(b+1-i)}]\}$  for ordered  $R_{(i)}$ ). The distribution of such a test statistic then requires calibration under the null  $H_0 : m = 1$ . A central limit theorem [51] for LRD processes  $X_t = G(Z_t)$  when  $m = 1$  gives

$$(15) \quad \frac{1}{\sqrt{\text{Var}(\bar{X}_n)}} \frac{1}{n} \sum_{i=1}^{\lfloor nt \rfloor} (X_i - \mu) \xrightarrow{d} cB_H(t), \quad 0 \leq t \leq 1,$$

as  $n \rightarrow \infty$ , where  $B_H(t)$  denotes fractional Brownian motion with Hurst index  $H = 1 - \alpha/2 \in (0, 1)$  and  $c > 0$  is a process constant. Note that (15) no longer holds under LRD when  $m > 1$ , which aids in testing. The property (15) suggests a simple bootstrap procedure for recreating the null distribution of residual-based test statistics, given in Algorithm 1, because such statistics are invariant to the location-scale used in a bootstrap sample.

The role of data blocking for tests of Hermite rank  $H_0 : m = 1$  with LRD series  $X_t = G(Z_t)$  may be traced to recent work of [9]. Those authors test for  $m = 1$  (normality) with a cumulant-based two-sample  $t$ -test, using two samples generated from data by different OL block resampling approaches. Our block-based test is different and perhaps more basic. To briefly compare these tests, we use data generation settings from [9] where  $X_t = G(Z_t)$  with  $G(z) = z + 20^{-1}H_2(z) + (20\sqrt{3})^{-1}H_3(z)$  (i.e.,  $m = 1$ ) or  $G(z) = \cos(z)$  (i.e.,  $m = 2$ ), and  $Z_t$  denotes a standardized FAIRMA(0,  $(1 - \alpha)/2$ , 0) Gaussian process for  $\alpha = 0.2, 0.8$ . The test in [9] uses block lengths  $\ell = n^{1/4}$  or  $\ell = n^{1/2}$ , where some best-case results provided there assume the memory parameter to be known. To facilitate comparison against these, we simply use a similar block  $\ell = n^{1/2}$  for our test, as a reasonable choice under LRD, and consider both OL/NOL blocks; we also use local Whittle estimation of the memory parameter along with

TABLE 6

Rejection rates of nominal 5% tests of Hermite rank  $m = 1$  from block-based statistic (Algorithm 1 with OL/NOL blocks  $\ell = \lfloor n^{1/2} \rfloor$ ) with LRD series  $X_t \equiv G(Z_t)$  using  $G_1(z) = z + 20^{-1}H_2(z) + (20\sqrt{3})^{-1}H_3(z)$  or  $G_2(z) = \cos(z)$ . Results from the test of [9] (OL blocks  $\ell = n^{1/2}$  or  $n^{1/4}$ ) are included

$G/m$	Testing Method		$\alpha = 0.20$			$\alpha = 0.80$		
			$n = 400$	1000	10000	$n = 400$	1000	10000
$G_1, m = 1$	Algorithm 1	OL	0.06	0.06	0.07	0.05	0.08	0.07
		NOL	0.10	0.06	0.08	0.06	0.05	0.06
	[9]-test	$\ell = n^{1/2}$	0.04	0.05	0.03	0.01	0.03	0.03
		$\ell = n^{1/4}$	0.03	0.04	0.04	0.07	0.06	0.03
$G_2, m = 2$	Algorithm 1	OL	0.60	0.85	0.99	0.13	0.14	0.13
		NOL	0.45	0.68	0.99	0.06	0.09	0.07
	[9]-test	$\ell = n^{1/2}$	0.11	0.23	0.26	0.13	0.13	0.20
		$\ell = n^{1/4}$	0.19	0.24	0.32	0.24	0.16	0.27

200 resamples in Algorithm 1. Table 6 lists power (based on 500 simulation runs) of our test using a 5% nominal level compared to test findings of [9] (Table 2); we report an Anderson–Darling statistic in Table 6 though a Kolmogorov–Smirnov statistic produced similar results. Both the proposed test and the [9]-test maintain the nominal size for the LRD processes with  $m = 1$ , but our block-based test has much larger power for the LRD process defined by  $m = 2$  and  $\alpha = 0.2$ . The process defined by  $m = 2$  and  $\alpha = 0.8$  in Table 6 is actually SRD; as both our test and the [9] test are block-based assessments of normality, both tests should maintain their sizes in this case and our test performs a bit better. This illustrates that data-blocking has potential for assessments beyond usage in resampling.

**7. Concluding remarks.** While block-based resampling methods provide useful estimation under data dependence, their performance is intricately linked to a block length parameter, which is important to understand. This problem has been extensively investigated under weak or short-range dependence (SRD) (cf. [32], Chapter 3), though relatively little has been known for long-range dependence (LRD), especially outside the pure Gaussian case (cf. [27]). For general long-range dependent (LRD)  $X_t = G(Z_t)$  processes, based on transforming a LRD Gaussian series  $Z_t$ , results here showed that properties and best block sizes with resampling variance estimators under LRD can intricately depend on covariance strength and the structure of nonlinearity in  $G(\cdot)$ . The long-memory guess  $O(n^{1/2})$  for block size [23] may have optimal order at times, owing more to such nonlinearity rather than intuition about LRD. Additionally, we provided a data-based rule for best block selection, which was shown to be consistent under complex cases for blocks with LRD. While we focused on a variance estimation problem with resampling under LRD, block selection for distribution estimation is also of interest, though seemingly requires further difficult study of distributional expansions for statistics from LRD series  $X_t = G(Z_t)$ . However, we showed that the block selections developed here can provide helpful benchmarks for choosing block size with resampling or other block-based inference problems under LRD.

The current work may also suggest future possibilities toward estimating the Hermite  $m$  rank (or other ranks) under LRD. No estimators of  $m$  currently exist; instead, only estimation of the long-memory exponent  $\alpha m$  of  $X_t = G(Z_t)$  has been possible, which depends on the covariance decay rate  $\alpha < 1/m$  of  $Z_t$ . Results here established that the variance of a block resampling estimator depends only on  $\alpha$ , apart from the Hermite rank  $m$  itself. This suggests

that some higher-order moment estimation may be investigated for separately estimating the memory coefficient  $\alpha$  and Hermite rank  $m$  under LRD.

#### APPENDIX A: COEFFICIENT OF OPTIMAL BLOCK SIZE

The coefficient  $K_{\alpha,m,m_2}$  of Corollary 4.1 is presented in cases with notation:  $A \equiv \mathcal{B}_0^2(m)$ ,  $B \equiv (2C_0^{m_2} J_{m_2}^2 / m_2!)^2$ ,  $C \equiv (\mathcal{B}_0(m) + 2 \sum_{j=m_2}^{\infty} \sum_{k=1}^{\infty} [\gamma_Z(k)]^j J_j^2 / j!)^2$  and  $D \equiv (2C_0^{m_2} / [(1 - m_2\alpha)(2 - m_2\alpha)](J_{m_2}^2 / m_2!))^2$  related to (7);  $E \equiv v_{\infty,\alpha m}^2$  in (4); and  $F \equiv a_\alpha$  from Theorem 3.2.

Case 1:  $m_2 = \infty$

$$K_{\alpha,m,m_2} = \begin{cases} \frac{-(1-2\alpha)\sqrt{AE} + \sqrt{(1-2\alpha)^2 AE + 4\alpha(1-\alpha)A(E+F)}}{2\alpha(E+F)} \\ \quad \text{if } 0 < \alpha < 0.5, m = 1, \\ \left(\frac{A}{F}\right)^{0.5} \\ \quad \text{if } \alpha = 0.5, m = 1, \\ \left(\frac{2A(1-\alpha)}{F}\right)^{\frac{1}{3-2\alpha}} \\ \quad \text{if } 0.5 < \alpha < 1, m = 1 \\ \left(\frac{A(1-\alpha m)}{F\alpha}\right)^{\frac{1}{2(1+\alpha-\alpha m)}} \\ \quad \text{if } 0 < \alpha < \frac{1}{m}, m \geq 2. \end{cases}$$

Case 2:  $m_2 < \infty$  with  $\alpha m_2 > 1$

$$K_{\alpha,m,m_2} = \begin{cases} \frac{-(1-2\alpha)\sqrt{CE} + \sqrt{(1-2\alpha)^2 CE + 4\alpha(1-\alpha)C(E+F)}}{2\alpha(E+F)} \\ \quad \text{if } \frac{1}{m_2} < \alpha < 0.5, m = 1, \\ \left(\frac{C}{F}\right)^{0.5} \\ \quad \text{if } \alpha = 0.5, m = 1, \\ \left(\frac{2C(1-\alpha)}{F}\right)^{\frac{1}{3-2\alpha}} \\ \quad \text{if } \max\left\{\frac{1}{m_2}, 0.5\right\} < \alpha < 1, m = 1, \\ \left(\frac{C(1-\alpha m)}{F\alpha}\right)^{\frac{1}{2(1+\alpha-\alpha m)}} \\ \quad \text{if } \frac{1}{m_2} < \alpha < \frac{1}{m}, m \geq 2. \end{cases}$$

Case 3:  $m_2 < \infty$  with  $\alpha m_2 = 1$

$$K_{\alpha, m, m_2} = \begin{cases} \frac{-(1-2\alpha)\sqrt{BE} + \sqrt{(1-2\alpha)^2 BE + 4\alpha(1-\alpha)B(E+F)}}{2\alpha(E+F)} \\ \text{if } 0 < \alpha = \frac{1}{m_2} < 0.5, m = 1, \\ \left(\frac{B}{F}\right)^{0.5} \\ \text{if } \alpha = \frac{1}{m_2} = 0.5, m = 1, \\ \left(\frac{B(1-\alpha m)}{F\alpha}\right)^{\frac{1}{2(1+\alpha-\alpha m)}} \\ \text{if } 0 < \alpha = \frac{1}{m_2} < 0.5, m \geq 2. \end{cases}$$

Case 4:  $m_2 < \infty$  with  $\alpha m_2 < 1$

$$K_{\alpha, m, m_2} = \begin{cases} \left(\frac{(m_2-2) + \sqrt{(m_2-2)^2 + (m_2-1)(E+F)D}}{E+F}\right)^{\frac{1}{\alpha m_2}} \\ \text{if } 0 < \alpha < \frac{1}{m_2}, m = 1, \\ \left(\frac{D(m_2-m)}{F}\right)^{\frac{1}{2\alpha(1+m_2-m)}} \\ \text{if } 0 < \alpha < \frac{1}{m_2}, m \geq 2. \end{cases}$$

## APPENDIX B: PROOF OF THEOREM 3.2

The [Appendix](#) considers the proof for the large-sample variance (Theorem 3.2) of resampling estimators. Proofs of other results are shown in the Supplementary Material [55]. To derive variance expansions of Theorem 3.2, we first consider the OL block variance estimator  $\widehat{V}_{\ell, \alpha m, \text{OL}} = \sum_{i=1}^N \ell^{\alpha m} (\bar{X}_{i, \ell} - \mu)^2 / N - \ell^{\alpha m} (\widehat{\mu}_{n, \text{OL}} - \mu)^2$  from (5), expressed in terms of the process mean  $\mathbb{E}X_t = \mu$ , the number  $N = n - \ell + 1$  of blocks, the block averages  $\bar{X}_{i, \ell} \equiv \sum_{j=i}^{i+\ell-1} X_j / \ell$  (integer  $i$ ), and  $\widehat{\mu}_{n, \text{OL}} = \sum_{i=1}^N \bar{X}_{i, \ell} / N$ . Due to mean centering, we may assume  $\mu = 0$  without loss of generality. We then write the variance of  $\widehat{V}_{\ell, \alpha m, \text{OL}}$  as

$$\text{Var}(\widehat{V}_{\ell, \alpha m, \text{OL}}) = v_{1, \ell} + v_{2, \ell} - 2c_{\ell},$$

$$c_{\ell} \equiv \text{Cov}\left(\frac{1}{N} \sum_{i=1}^N \ell^{\alpha m} \bar{X}_{i, \ell}^2, \ell^{\alpha m} \widehat{\mu}_{n, \text{OL}}^2\right) = c_{a, \ell} + c_{b, \ell},$$

$$v_{1, \ell} \equiv \text{Var}\left(\frac{1}{N} \sum_{i=1}^N \ell^{\alpha m} \bar{X}_{i, \ell}^2\right) = v_{1a, \ell} + v_{1b, \ell},$$

$$v_{2, \ell} \equiv \text{Var}(\ell^{\alpha m} \widehat{\mu}_{n, \text{OL}}^2) = v_{2a, \ell} + v_{2b, \ell},$$



where each variance/covariance component  $v_{1,\ell}$ ,  $v_{2,\ell}$  and  $c_\ell$  is decomposed into two further subcomponents

$$\begin{aligned}
 v_{1a,\ell} &\equiv 2 \frac{\ell^{2\alpha m}}{N} \sum_{k=-N}^N \left(1 - \frac{|k|}{N}\right) [\text{Cov}(\bar{X}_{0,\ell}, \bar{X}_{k,\ell})]^2, \\
 v_{2a,\ell} &\equiv 2 \ell^{2\alpha m} [\text{Var}(\hat{\mu}_{n,\text{OL}})]^2, \\
 v_{1b,\ell} &\equiv \frac{\ell^{2\alpha m}}{N} \sum_{k=-N}^N \left(1 - \frac{|k|}{N}\right) \text{cum}(\bar{X}_{0,\ell}, \bar{X}_{0,\ell}, \bar{X}_{k,\ell}, \bar{X}_{k,\ell}), \\
 v_{2b,\ell} &\equiv \ell^{2\alpha m} \text{cum}(\hat{\mu}_{n,\text{OL}}, \hat{\mu}_{n,\text{OL}}, \hat{\mu}_{n,\text{OL}}, \hat{\mu}_{n,\text{OL}}), \\
 c_{a,\ell} &\equiv 2(\ell^{2\alpha m}/N) \sum_{i=1}^N [\text{Cov}(\bar{X}_{i,\ell}, \hat{\mu}_{n,\text{OL}})]^2,
 \end{aligned}
 \tag{16}$$

and  $c_{b,\ell} \equiv (\ell^{2\alpha m}/N^3) \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N \text{cum}(\bar{X}_{i,\ell}, \bar{X}_{j,\ell}, \bar{X}_{k,\ell}, \bar{X}_{k,\ell})$ , which denote sums involving fourth-order cumulants ( $v_{1b,\ell}$ ,  $v_{2b,\ell}$ ,  $c_{b,\ell}$ ) or sums involving covariances ( $v_{1a,\ell}$ ,  $v_{2a,\ell}$ ,  $c_{a,\ell}$ ). The second decomposition step follows from the product theorem for cumulants (e.g.,  $\text{Cov}(Y_1 Y_2, Y_3 Y_4) = \text{Cov}(Y_1, Y_3) \text{Cov}(Y_2, Y_4) + \text{Cov}(Y_1, Y_4) \text{Cov}(Y_2, Y_3) + \text{cum}(Y_1, Y_2, Y_3, Y_4)$  for arbitrary random variables with  $\text{E}Y_i = 0$  and  $\text{E}Y_i^4 < \infty$ ). Note that  $\text{E}X_t^4 = \text{E}[G(Z_0)]^4 < \infty$  and  $G \in \bar{\mathcal{G}}_4(1)$  imply these variance components exist finitely for any  $n, \ell$  (see (S.9) or Lemma 3 of the Supplementary Material [55]). Collecting terms, we have

$$\text{Var}(\hat{V}_{\ell,\alpha m,\text{OL}}) = \Gamma_{\ell,\text{OL}} + \Delta_{\ell,\text{OL}},$$

where  $\Gamma_{\ell,\text{OL}} \equiv v_{1a,\ell} + v_{2a,\ell} - 2c_{a,\ell}$  and  $\Delta_{\ell,\text{OL}} \equiv v_{1b,\ell} + v_{2b,\ell} - 2c_{b,\ell}$  denote sums over covariance-terms  $\Gamma_{\ell,\text{OL}}$  or sums of fourth-order cumulant terms  $\Delta_{\ell,\text{OL}}$ . In the NOL estimator case  $\hat{V}_{\ell,\alpha m,\text{NOL}}$ , the variance expansion is similar  $\text{Var}(\hat{V}_{\ell,\alpha m,\text{NOL}}) = \Gamma_{\ell,\text{NOL}} + \Delta_{\ell,\text{NOL}}$  with the convention that  $\Gamma_{\ell,\text{NOL}}$ ,  $\Delta_{\ell,\text{NOL}}$  are defined by replacing the OL block number  $N = n - \ell + 1$ , averages  $\bar{X}_{i,\ell}$  (or  $\bar{X}_{j,\ell}$ ,  $\bar{X}_{k,\ell}$ ) and estimator  $\hat{\mu}_{n,\text{OL}} = \sum_{i=1}^N \bar{X}_{i,\ell}/N$  with the NOL counterparts  $b = \lfloor n/\ell \rfloor$ ,  $\bar{X}_{1+(i-1)\ell,\ell}$  (or  $\bar{X}_{1+(j-1)\ell,\ell}$ ,  $\bar{X}_{1+(k-1)\ell,\ell}$ ) and  $\hat{\mu}_{n,\text{NOL}} = \sum_{i=1}^b \bar{X}_{1+(i-1)\ell,\ell}/b$  in  $v_{1a,\ell}$ ,  $v_{2a,\ell}$ ,  $c_{a,\ell}$ ,  $v_{1b,\ell}$ ,  $v_{2b,\ell}$ ,  $c_{b,\ell}$ .

Let  $\Gamma_\ell$  denote either counterpart  $\Gamma_{\ell,\text{OL}}$  or  $\Gamma_{\ell,\text{NOL}}$ , and  $\Delta_\ell$  denote either  $\Delta_{\ell,\text{OL}}$  or  $\Delta_{\ell,\text{NOL}}$ . Theorem 3.2 then follows by establishing that

$$\Gamma_\ell = \begin{cases} o\left(\left(\frac{\ell}{n}\right)^{\min\{1, 2\alpha m\}} [\log n]^{I(2\alpha m=1)}\right) = o\left(\left(\frac{\ell}{n}\right)^{2\alpha}\right) & \text{if } m \geq 2, \\ a_\alpha \left(\frac{\ell}{n}\right)^{\min\{1, 2\alpha\}} [\log n]^{I(\alpha=1/2)} (1 + o(1)) & \text{if } m = 1, \end{cases}$$

and

$$\Delta_\ell - r_{n,\alpha,m,m_p} = \begin{cases} \phi_{\alpha,m} \left(\frac{\ell}{n}\right)^{2\alpha} (1 + o(1)) & \text{if } m \geq 2, \\ o\left(\left(\frac{\ell}{n}\right)^{\min\{1, 2\alpha\}} [\log n]^{I(\alpha=1/2)}\right) & \text{if } m = 1, \end{cases}$$

where  $I(\cdot)$  denotes an indicator function and  $r_{n,\alpha,m,m_p}$  is defined in Theorem 3.2. For reference, when the Hermite rank  $m \geq 2$ , the contribution of  $\Delta_\ell \propto (\ell/n)^{2\alpha}$  dominates the variance of  $\hat{V}_{\ell,\alpha m,\text{OL}}$  or  $\hat{V}_{\ell,\alpha m,\text{NOL}}$ ; when  $m = 1$ , the contribution of  $\Gamma_\ell \propto (\ell/n)^{\min\{1, 2\alpha\}} [\log n]^{I(\alpha=1/2)}$  instead dominates the variance in Theorem 3.2.

To establish these expansions of  $\Gamma_\ell$ ,  $\Delta_\ell$ , we require a series of technical lemmas (Lemmas 1–4), involving certain graph-theoretic moment expansions. To provide some illustration, Lemma 1 is briefly outlined in Appendix C; the remaining lemmas are described in the Supplementary Material [55]. Define an order constant  $\tau_{\ell,m} \equiv (\ell/n)^{2\alpha}$  if  $m \geq 2$  and  $\tau_{\ell,m} \equiv (\ell/n)^{\min\{1,2\alpha\}} [\log n]^{I(\alpha=1/2)}$  if  $m = 1$ ; we suppress the dependence of  $\tau_{\ell,m}$  on  $n$  and  $\alpha$  for simplicity. Then the above expansion of  $\Gamma_\ell$  follows directly from Lemma 4 (i.e.,  $\Gamma_\ell = a_\alpha \tau_{\ell,1}(1 + o(1))$  if  $m = 1$  and  $\Gamma_\ell = o(\tau_{\ell,m})$  if  $m \geq 2$ ). For handling  $\Delta_\ell$ , Lemma 1 gives that  $v_{1b,\ell} = r_{n,\alpha,m,m_p} + \phi_{\alpha,m} \tau_{\ell,m}(1 + o(1))$  when  $m \geq 2$  and  $v_{1b,\ell} = r_{n,\alpha,m,m_p} + o(\tau_{\ell,m})$  when  $m = 1$ . Combined with this, the expansion of  $\Delta_\ell$  then follows from Lemmas 2 and 3, which respectively show that  $c_{b,\ell} = o(\tau_{\ell,m})$  and  $v_{2b,\ell} = o(\tau_{\ell,m})$  for any  $m \geq 1$ .  $\square$

#### APPENDIX C: LEMMA 1 (DOMINANT FOURTH-ORDER CUMULANT TERMS)

In the proof of Theorem 3.2 (Appendix B), recall  $v_{1b,\ell}$  from (16) represents a sum of fourth-order cumulants from OL block averages, where the version with NOL blocks is  $v_{1b,\ell} \equiv b^{-1} \ell^{2\alpha m} \sum_{k=-b}^b (1 - |k|/b) \text{cum}(\bar{X}_{0,\ell}, \bar{X}_{0,\ell}, \bar{X}_{k\ell,\ell}, \bar{X}_{k\ell,\ell})$ . Lemma 1 provides an expansion of  $v_{1b,\ell}$  under LRD, which is valid in either OL/NOL block case.

**LEMMA 1.** *Suppose the assumptions of Theorem 3.2 ( $G$  has Hermite rank  $m \geq 1$  and Hermite pair-rank  $m \leq m_p \leq \infty$ ) with positive constants  $\phi_{\alpha,m}$ ,  $\lambda_{\alpha,m_p}$  there. Then*

$$v_{1b,\ell} = r_{n,\alpha,m,m_p} + \begin{cases} \phi_{\alpha,m} \left(\frac{\ell}{n}\right)^{2\alpha} (1 + o(1)) & \text{if } m \geq 2, \\ o\left(\left(\frac{\ell}{n}\right)^{\min\{1,2\alpha\}} [\log n]^{I(\alpha=1/2)}\right) & \text{if } m = 1, \end{cases}$$

where  $I(\cdot)$  denotes an indicator function and  $r_{n,\alpha,m,m_p}$  is from Theorem 3.2.

The proof of Lemma 1 involves a standard, but technical, graph-theoretic representation of the fourth-order cumulant among Hermite polynomials  $H_{k_1}(Y_1)$ ,  $H_{k_2}(Y_2)$ ,  $H_{k_3}(Y_3)$ ,  $H_{k_4}(Y_4)$  ( $k_1, k_2, k_3, k_4 \geq 1$ ) at a generic sequence  $(Y_1, Y_2, Y_3, Y_4)$  of marginally standard normal variables, with covariances  $\text{E}Y_i Y_j = r_{ij} = r_{ji}$  for  $1 \leq i < j \leq 4$ . Namely, it holds that

$$\text{cum}[H_{k_1}(Y_1), H_{k_2}(Y_2), H_{k_3}(Y_3), H_{k_4}(Y_4)] = \prod_{i=1}^4 k_i! \sum_{A \in \mathcal{A}_c(k_1, k_2, k_3, k_4)} g(A) r(A),$$

where above  $\mathcal{A}_c(k_1, k_2, k_3, k_4)$  denotes the collection of all path-connected multigraphs from a generic set of four points/vertices  $p_1, p_2, p_3, p_4$ , such that point  $p_i$  has degree  $k_i$  for  $i = 1, 2, 3, 4$ ; in this definition,  $\text{cum}[H_{k_1}(Y_1), H_{k_2}(Y_2), H_{k_3}(Y_3), H_{k_4}(Y_4)] = 0$  holds whenever  $\mathcal{A}_c(k_1, k_2, k_3, k_4)$  is empty. Each multigraph  $A \equiv (v_{12}, v_{13}, v_{14}, v_{23}, v_{24}, v_{34}) \in \mathcal{A}_c(k_1, k_2, k_3, k_4)$  is defined by distinct counts  $v_{ij} = v_{ji} \geq 0$ , interpreted as the number of graph lines connecting points  $p_i$  and  $p_j$ ,  $1 \leq i < j \leq 4$ . Then  $g(A) \equiv 1/[\prod_{1 \leq i < j \leq 4} (v_{ij}!)]$  represents a so-called multiplicity factor, while  $r(A) \equiv \prod_{1 \leq i < j \leq 4} r_{ij}^{v_{ij}}$  represents a weighted product of covariances among variables in  $(Y_1, Y_2, Y_3, Y_4)$  (cf. [50]). Membership  $A \in \mathcal{A}_c(k_1, k_2, k_3, k_4)$  requires degrees  $k_i = \sum_{j:j \neq i} v_{ij}$  for  $i = 1, 2, 3, 4$  (e.g.,  $k_2 = v_{12} + v_{23} + v_{24}$ ) as well as a path connection in  $A$  between any two points  $p_i$  and  $p_j$ ; see, for example, Figure 2.

With this background and using the Hermite expansion (2) for  $G \in \bar{\mathcal{G}}_4(1)$ , the sum  $v_{1b,\ell}$  of fourth-order cumulants from OL block averages (16) can be reduced to a weighted sum of terms

$$\vartheta_{\ell,n}(A) \equiv \frac{\ell^{2\alpha m - 4}}{N} \sum_{k=-N}^N \left(1 - \frac{|k|}{N}\right) \sum_{t_1=1}^{\ell} \sum_{t_2=1}^{\ell} \sum_{t_3=k}^{k+\ell-1} \sum_{t_4=k}^{k+\ell-1} \prod_{1 \leq i < j \leq 4} \frac{1}{v_{ij}!} [\gamma_Z(t_i - t_j)]^{v_{ij}}$$

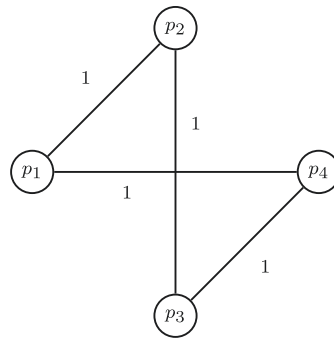


FIG. 2. An example of multigraph  $A \equiv (1, 0, 1, 1, 0, 1) \in \mathcal{A}_c(m, m, m, m)$  for  $m = 2$ .

over all path-connected graphs  $A \equiv (v_{12}, v_{13}, v_{14}, v_{23}, v_{24}, v_{34})$  with degrees  $k_1, \dots, k_4 \geq m$ ; above  $\gamma_Z(\cdot)$  is the Gaussian covariance function from (1). The path-connected property implies  $v_{13} + v_{14} + v_{23} + v_{24} \geq 1$ . Combining  $\vartheta_{\ell,n}(A)$  over all such graphs  $A$  with  $v_{13} + v_{14} + v_{23} + v_{24} = 1$  determines  $r_{n,\alpha,m,m_p}$  in Lemma 1. For  $m \geq 2$ , contributions  $\vartheta_{\ell,n}(A)$  from two graphs  $A \in \mathcal{A}_c(m, m, m, m)$  with  $v_{13} + v_{14} + v_{23} + v_{24} = 2$  yield  $\phi_{\alpha,m}(\ell/n)^{2\alpha}(1 + o(1))$  in Lemma 1;  $\mathcal{A}_c(m, m, m, m)$  is empty when  $m = 1$ . Remaining expressions in Lemma 1 follow from all other graph contributions  $\vartheta_{\ell,n}(A)$ . See the Supplementary Material [55] for more details.

**Acknowledgments.** The authors are grateful to two anonymous referees and an Associate Editor for constructive comments that improved the quality of this paper.

**Funding.** Research was supported by NSF Grants DMS-1811998, DMS 2015390, and DMS-2131233.

## SUPPLEMENTARY MATERIAL

**Proofs and other technical details** (DOI: [10.1214/22-AOS2242SUPPA](https://doi.org/10.1214/22-AOS2242SUPPA); .pdf). A supplement [55] contains proofs and technical details along with further numerical results.

**Code** (DOI: [10.1214/22-AOS2242SUPPB](https://doi.org/10.1214/22-AOS2242SUPPB); .zip). R code.

## 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 <https://doi.org/10.1016/j.jeconom.2009.02.010>
- [2] ANDERSON, T. W. and DARLING, D. A. (1952). Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. *Ann. Math. Stat.* **23** 193–212. MR0050238 <https://doi.org/10.1214/aoms/1177729437>
- [3] ANDREWS, D. W. K. and SUN, Y. (2004). Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica* **72** 569–614. MR2036733 <https://doi.org/10.1111/j.1468-0262.2004.00501.x>
- [4] BAI, S. and TAQQU, M. S. (2017). On the validity of resampling methods under long memory. *Ann. Statist.* **45** 2365–2399. MR3737895 <https://doi.org/10.1214/16-AOS1524>
- [5] BAI, S. and TAQQU, M. S. (2018). How the instability of ranks under long memory affects large-sample inference. *Statist. Sci.* **33** 96–116. MR3757507 <https://doi.org/10.1214/17-STS633>
- [6] BAI, S., TAQQU, M. S. and ZHANG, T. (2016). A unified approach to self-normalized block sampling. *Stochastic Process. Appl.* **126** 2465–2493. MR3505234 <https://doi.org/10.1016/j.spa.2016.02.007>
- [7] BERAN, J. (1991).  $M$  estimators of location for Gaussian and related processes with slowly decaying serial correlations. *J. Amer. Statist. Assoc.* **86** 704–708. MR1147095
- [8] BERAN, J. (2010). Long-range dependence. *Wiley Ser. Comput. Stat.* **2** 26–35.

- [9] BERAN, J., MÖHRLE, S. and GHOSH, S. (2016). Testing for Hermite rank in Gaussian subordination processes. *J. Comput. Graph. Statist.* **25** 917–934. MR3533645 <https://doi.org/10.1080/10618600.2015.1056345>
- [10] BETKEN, A. and WENDLER, M. (2018). Subsampling for general statistics under long range dependence with application to change point analysis. *Statist. Sinica* **28** 1199–1224. MR3821001
- [11] BREIDT, F. J., CRATO, N. and DE LIMA, P. (1998). The detection and estimation of long memory in stochastic volatility. *J. Econometrics* **83** 325–348. MR1613867 [https://doi.org/10.1016/S0304-4076\(97\)00072-9](https://doi.org/10.1016/S0304-4076(97)00072-9)
- [12] BÜHLMAN, P. and KÜNSCH, H. R. (1999). Block length selection in the bootstrap for time series. *Comput. Statist. Data Anal.* **31** 295–310.
- [13] CARLSTEIN, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.* **14** 1171–1179. MR0856813 <https://doi.org/10.1214/aos/1176350057>
- [14] DEHLING, H. and TAQQU, M. S. (1989). The empirical process of some long-range dependent sequences with an application to  $U$ -statistics. *Ann. Statist.* **17** 1767–1783. MR1026312 <https://doi.org/10.1214/aos/1176347394>
- [15] 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 <https://doi.org/10.1007/BF00535673>
- [16] FERNHOLZ, L. T. (2012). *Von Mises Calculus for Statistical Functionals*. Springer Science & Business Media. MR0713611 <https://doi.org/10.1007/978-1-4612-5604-5>
- [17] FLEGAL, J. M. and JONES, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38** 1034–1070. MR2604704 <https://doi.org/10.1214/09-AOS735>
- [18] GIRAITIS, L., ROBINSON, P. M. and SURGAILIS, D. (1999). Variance-type estimation of long memory. *Stochastic Process. Appl.* **80** 1–24. MR1670119 [https://doi.org/10.1016/S0304-4149\(98\)00062-3](https://doi.org/10.1016/S0304-4149(98)00062-3)
- [19] GIRAITIS, L. and TAQQU, M. S. (1999). Whittle estimator for finite-variance non-Gaussian time series with long memory. *Ann. Statist.* **27** 178–203. MR1701107 <https://doi.org/10.1214/aos/1018031107>
- [20] GRANGER, C. W. J. and JOYEUX, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Series Anal.* **1** 15–29. MR0605572 <https://doi.org/10.1111/j.1467-9892.1980.tb00297.x>
- [21] HALL, P., HOROWITZ, J. L. and JING, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82** 561–574. MR1366282 <https://doi.org/10.1093/biomet/82.3.561>
- [22] HALL, P. and JING, B. (1996). On sample reuse methods for dependent data. *J. Roy. Statist. Soc. Ser. B* **58** 727–737. MR1410187
- [23] HALL, P., JING, B.-Y. and LAHIRI, S. N. (1998). On the sampling window method for long-range dependent data. *Statist. Sinica* **8** 1189–1204. MR1666245
- [24] HÖSSJER, O. and MIELNICZUK, J. (1995). Delta method for long-range dependent observations. *J. Non-parametr. Stat.* **5** 75–82. MR1342943 <https://doi.org/10.1080/10485259508832635>
- [25] HURVICH, C. M., DEO, R. and BRODSKY, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *J. Time Series Anal.* **19** 19–46. MR1624096 <https://doi.org/10.1111/1467-9892.00075>
- [26] JACH, A., MCELROY, T. S. and POLITIS, D. N. (2016). Corrigendum to 'Subsampling inference for the mean of heavy-tailed long-memory time series' [MR2877610]. *J. Time Series Anal.* **37** 713–720. MR3536227 <https://doi.org/10.1111/jtsa.12191>
- [27] KIM, Y. M. and NORDMAN, D. J. (2011). Properties of a block bootstrap under long-range dependence. *Sankhya A* **73** 79–109. MR2887088 <https://doi.org/10.1007/s13171-011-0003-3>
- [28] KREISS, J.-P. and LAHIRI, S. N. (2012). Bootstrap methods for time series. In *Handbook of Statist.* **30** 3–26. Elsevier.
- [29] KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241. MR1015147 <https://doi.org/10.1214/aos/1176347265>
- [30] LAHIRI, S. N. (1993). On the moving block bootstrap under long range dependence. *Statist. Probab. Lett.* **18** 405–413. MR1247453 [https://doi.org/10.1016/0167-7152\(93\)90035-H](https://doi.org/10.1016/0167-7152(93)90035-H)
- [31] LAHIRI, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Ann. Statist.* **27** 386–404. MR1701117 <https://doi.org/10.1214/aos/1018031117>
- [32] LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer Series in Statistics. Springer, New York. MR2001447 <https://doi.org/10.1007/978-1-4757-3803-2>
- [33] LAHIRI, S. N., FURUKAWA, K. and LEE, Y.-D. (2007). A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods. *Stat. Methodol.* **4** 292–321. MR2380557 <https://doi.org/10.1016/j.stamet.2006.08.002>
- [34] LIU, R. Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap (East Lansing, MI, 1990)*. Wiley Ser. Probab. Math. Statist. Probab. Math. Statist. 225–248. Wiley, New York. MR1197787

- [35] MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437. MR0242239 <https://doi.org/10.1137/1010093>
- [36] MONTANARI, A., TAQQU, M. S. and TEVEROVSKY, V. (1999). Estimating long-range dependence in the presence of periodicity: An empirical study. *Math. Comput. Modelling* **29** 217–228.
- [37] 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 <https://doi.org/10.1017/S0266466605050541>
- [38] NORDMAN, D. J. and LAHIRI, S. N. (2014). Convergence rates of empirical block length selectors for block bootstrap. *Bernoulli* **20** 958–978. MR3178523 <https://doi.org/10.3150/13-BEJ511>
- [39] PAPARODITIS, E. and POLITIS, D. N. (2002). The tapered block bootstrap for general statistics from stationary sequences. *Econom. J.* **5** 131–148. MR1909304 <https://doi.org/10.1111/1368-423X.t01-1-00077>
- [40] POLITIS, D. N. (2003). The impact of bootstrap methods on time series analysis. *Statist. Sci.* **18** 219–230. MR2026081 <https://doi.org/10.1214/ss/1063994977>
- [41] POLITIS, D. N. and ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050. MR1329181 <https://doi.org/10.1214/aos/1176325770>
- [42] POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling. Springer Series in Statistics*. Springer, New York. MR1707286 <https://doi.org/10.1007/978-1-4612-1554-7>
- [43] POLITIS, D. N. and WHITE, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Rev.* **23** 53–70. MR2041534 <https://doi.org/10.1081/ETC-120028836>
- [44] ROBINSON, P. M. (1995). Log-periodogram regression of time series with long range dependence. *Ann. Statist.* **23** 1048–1072. MR1345214 <https://doi.org/10.1214/aos/1176324636>
- [45] ROBINSON, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630–1661. MR1370301 <https://doi.org/10.1214/aos/1176324317>
- [46] ROBINSON, P. M. (2005). Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction. *Econometric Theory* **21** 171–180. MR2153861 <https://doi.org/10.1017/S0266466605050115>
- [47] SHAO, J. (2003). *Mathematical Statistics*. Springer.
- [48] SONG, W. T. and SCHMEISER, B. W. (1995). Optimal mean-squared-error batch sizes. *Manage. Sci.* **41** 110–123.
- [49] TAQQU, M. S. (1974/75). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* **31** 287–302. MR0400329 <https://doi.org/10.1007/BF00532868>
- [50] TAQQU, M. S. (1977). Law of the iterated logarithm for sums of non-linear functions of Gaussian variables that exhibit a long range dependence. *Z. Wahrsch. Verw. Gebiete* **40** 203–238. MR0471045 <https://doi.org/10.1007/BF00736047>
- [51] TAQQU, M. S. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Z. Wahrsch. Verw. Gebiete* **50** 53–83. MR0550123 <https://doi.org/10.1007/BF00535674>
- [52] TAQQU, M. S. (2002). Fractional Brownian motion and long-range dependence. In *Theory and Applications of Long-Range Dependence* (P. Doukhan, G. Oppenheim and M. S. Taqqu, eds.). Springer Science & Business Media. MR1956042
- [53] TEVEROVSKY, V. and TAQQU, M. (1997). Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator. *J. Time Series Anal.* **18** 279–304. MR1456641 <https://doi.org/10.1111/1467-9892.00050>
- [54] WHITE, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48** 817–838. MR0575027 <https://doi.org/10.2307/1912934>
- [55] ZHANG, Q., LAHIRI, S. N. and NORDMAN, D. J. (2022). Supplement to “On optimal block resampling for Gaussian-subordinated long-range dependent processes.” <https://doi.org/10.1214/22-AOS2242SUPPA>, <https://doi.org/10.1214/22-AOS2242SUPPB>
- [56] ZHANG, T., HO, H.-C., WENDLER, M. and WU, W. B. (2013). Block sampling under strong dependence. *Stochastic Process. Appl.* **123** 2323–2339. MR3038507 <https://doi.org/10.1016/j.spa.2013.02.006>