How the instability of ranks under long memory affects large-sample inference

Shuyang Bai
University of Georgia

Murad S. Taqqu
Boston University

October 10, 2017

Abstract

Under long memory, the limit theorems for normalized sums of random variables typically involve a positive integer called "Hermite rank". There is a different limit for each Hermite rank. From a statistical point of view, however, we argue that a rank other than one is unstable, whereas, a rank equal to one is stable. We provide empirical evidence supporting this argument. This has important consequences. Assuming a higher-order rank when it is not really there usually result in underestimating the order of the fluctuations of the statistic of interest. We illustrate this through various examples involving the sample variance, the empirical processes and the Whittle estimator.

1 Introduction

Suppose that $\mathcal{D}$ is a data set, and one has a statistical model for $\mathcal{D}$ which involves a random stationary sequence $\{X(n)\}$, referred to as noise. Let $T = T(\mathcal{D})$ be a sample statistic of interest. Deriving the asymptotic distribution for the statistic T as the sample size tends to infinity is a standard practice in large sample inference. The asymptotic distribution is useful for reporting confidence intervals, conducting hypothesis tests, etc.

When the construction of the statistic T involves summing the data and if the stationary noise $\{X(n)\}$ is weakly dependent, then the asymptotic distribution of T is typically Gaussian in view of the Central Limit Theorem. This asymptotic distribution can also be a functional of a Gaussian process.

The situation, however, is much more intricate when the strength of dependence in the noise increases significantly. This strong-dependence regime, often called *long memory* or *long-range dependence*, is typically characterized by the following behavior of the variance of partial sums:

$$\text{Var}\Big[\sum_{n=1}^{N} X(n)\Big] \approx N^{2H}, \text{ as } N \to \infty. \tag{1}$$

where $\approx$ means asymptotic equivalence up to some positive constant, the parameter $H \in (1/2, 1)$ is called the *Hurst index*[1]. Normally when the dependence is weak, one expects $H = 1/2$ in (1), that is, the growth

---

[1]The term "Hurst index" is also frequently used for the self-similarity parameter of self-similar processes arising from the normalized limit of the sum of $X(n)$ (see Pipiras and Taqqu [64]). It is also common to introduce the so-called memory parameter d in the context of (2), re-expressed as $\text{Cov}[X(n), X(0)] \approx n^{2d-1}$ $(d = H - 1/2)$. We will use H throughout in order not to switch between parameters and thus to avoid confusion.

of the variance is linear. The superlinear growth in (1) is typically due to the slow decay of the covariance of $\{X(n)\}$:

$$\text{Cov}[X(n), X(0)] \approx n^{2H-2}, \text{ as } n \to \infty, \tag{2}$$

where $-1 < 2H - 2 < 0$. In fact, (2) is also a common characterization of long memory. We refer the reader to the recent monographs Beran et al. [11], Giraitis et al. [39], Samorodnitsky [69] and Pipiras and Taqqu [64] for comprehensive introductions to the notion long memory.

In view of (1), when deriving the asymptotic distribution of the sum, one needs to associate the stronger normalization N^{-H} to $\sum_{n=1}^{N} X(n)$ rather than the standard $N^{-1/2}$ normalization. These limit theorems have been applied in many statistical studies. See Section 3 below.

This paper makes the following basic argument: while these limit theorems are definitely of probabilistic interest, their immediate application to statistical inference can lead to problems. This is because these limit theorems can be unstable, that is, they often cease to hold when $\{X(n)\}$ is slightly perturbed. In particular, the limit theorems under long memory often depend critically on an integer quantity called *rank*, e.g, the *Hermite rank* in the Gaussian context. We will show that the rank is unstable when it takes value greater than one, and it easily collapses to rank one when there is a slight perturbation.

The notion of rank, however, is not relevant when the data is weakly dependent. We indicate that under weak dependence, limit theorems are robust against, for example, a transformation of the data. We illustrate this by considering various types of weak dependence, such as strong mixing, Gaussian subordination and Bernoulli shifts.

The paper is organized as follows. The rank instability issue is discussed in Section 2. In Section 3, we provide some examples on how the instability of rank can affect statistical results. In Section 4, we carry out an empirical study which supports the instability argument. In contrast, we show in Section 5 that the central limit theorems under weak dependence are not subject to such instability issues. Section 6 contains conclusions and suggestions. Some technical extensions are found in the Appendix.

2 The instability of ranks under long memory

In this section, we introduce the notion Hermite rank and point out its instability. We focus on the simple scenario of instantaneous transformation of a Gaussian stationary process. (The non-instantaneous case is somewhat technical and is deferred to Appendix A.) We then address the case where the model involves a non-Gaussian linear (moving-average) process, where the corresponding notion of Hermite rank is called the *Appell rank* or the *power rank*.

Throughout the paper, the notation $a(n) \approx b(n)$ means $\lim_{n \to \infty} a(n)/b(n) = c$ for some generic constant $0 < c < \infty$ that can change from expression to expression. We note that in many places, one can include a slowly varying function in the asymptotic relation, for example, a logarithmic function (see, e.g., Bingham et al. [14]), but for simplicity we do not do that.

2.1 Transformation of Gaussian processes

We want to consider possibly nonlinear finite-variance transformations of Gaussian random variables. To do so, let Z be a standard normal random variable, $\gamma(dx)$ be the standard Gaussian measure $(2\pi)^{-1/2}e^{-x^2/2}dx$ on $\mathbb{R}$, and let

$$L^2(\gamma) = \{G(\cdot) : \ \mathbb{E}G(Z)^2 < \infty\}.$$

It is well-known (see, e.g., see Pipiras and Taqqu [64], Proposition 5.1.3) that $\{\frac{1}{\sqrt{m!}} H_m(\cdot), m \geq 0\}$ forms an orthonormal basis of $L^2(\gamma)$, where $\{H_m(\cdot), \ m \geq 0\}$ are Hermite polynomials defined as $H_0(x) = 1$ and

$$H_m(x) = (-1)^m e^{x^2/2} \frac{d}{dx^m} e^{-x^2/2} \quad \text{for} \quad m \geq 1. \tag{3}$$

Thus $H_1(x) = x$, $H_2(x) = x^2 - 1$ and $H_3(x) = x^3 - 3x$, etc. We can now define the Hermite rank of a function $G \in L^2(\gamma)$.

2

Definition 2.1. Suppose that $G(\cdot) \in L^2(\gamma)$. Let Z be a standard Gaussian random variable. The *Hermite rank* k of $G(\cdot)$ is defined as

$$k = \inf\left\{ m \geq 1 : \; \mathbb{E}G(Z)H_m(Z) = \int_{\mathbb{R}} G(x)H_m(x)\gamma(dx) \neq 0 \right\}. \tag{4}$$

where $H_m(\cdot)$ is the m-th order Hermite polynomial.

Remark 2.2. An alternative way of defining the Hermite rank k is through the starting index of the Hermite expansion of $G(\cdot) - \mathbb{E}G(Z)$, namely,

$$G(\cdot) - \mathbb{E}G(Z) = \sum_{m=k}^{\infty} c_m H_m(\cdot), \quad k \geq 1, \tag{5}$$

for some sequence c_m satisfying $c_k \neq 0$, where the series converges in the $L^2(\gamma)$-sense. By the orthonormality of $\{\frac{1}{\sqrt{m!}}H_m(\cdot)\}$, we have

$$c_m = \frac{\mathbb{E}[G(Z)H_m(Z)]}{m!} \quad \text{for} \quad m \geq 0. \tag{6}$$

Note that $c_0 = \mathbb{E}G(Z)$ since $H_0(Z) = 1$. Furthermore, since the Hermite polynomials $\{H_m(\cdot), 0 \leq m \leq k\}$ form a basis for polynomials of degree less than equal to k, the definition (4) can be re-expressed as

$$k = \inf\left\{ m \geq 1 : \; \mathbb{E}\left[(G(Z) - \mathbb{E}G(Z))Z^m \right] \neq 0 \right\}. \tag{7}$$

Remark 2.3. The Hermite rank of $G(x)$ is the same as that of $G(x) + a$, for any $a \in \mathbb{R}$, since relation (5) involves centering.

Now suppose that $\{X(n)\}$ is a long-memory stationary Gaussian process satisfying (2). We may assume without loss of generality that it is standardized, that is, $\mathbb{E}X(n) = 0$ and $\mathrm{Var}[X(n)] = 1$ The following lemma explains the role that the Hermite rank plays in determining the asymptotic behavior of the covariance of the transformed sequence $\{G(X(n))\}$ (see p.223 of Beran et al. [11]).

Lemma 2.4. *If $G(\cdot)$ has Hermite rank k, then*

$$\mathrm{Cov}[G(X(n)), G(X(0))] \approx \mathrm{Cov}[X(n), X(0)]^k \approx n^{(2H-2)k}. \tag{8}$$

Remark 2.5. Comparing (8) and (2), we note that for functions $G(\cdot)$ that have Hermite rank $k = 1$, the Hurst index of $\{G(X(n))\}$ is the same as the Hurst index of $\{X(n)\}$. In general since $2H - 2 < 0$, the higher the Hermite rank k is, the faster the covariance decays as $n \to \infty$. Note that in view of (2) and (8), for $\{G(X(n))\}$ to have long memory, one needs

$$(2H - 2)k > -1 \iff H > 1 - \frac{1}{2k}.$$

This is natural because when $k > 1$, the covariance of $\{G(X(n))\}$ decays faster than that of $\{X(n)\}$ and thus H must be greater than $1 - 1/(2k)$ in order to ensure that $\{G(X_n)\}$ has long memory.

2.2 Asymptotic behavior

Now returning to the theme of the introduction: suppose that in order to derive the asymptotic distribution of the statistics T of interest, one first needs to obtain the distributional limit as $N \to \infty$ of

$$\frac{1}{A(N)} \sum_{n=1}^{[Nt]} \Big(G(X(n)) - \mathbb{E}G(X(n)) \Big), \quad t \in [0,1], \tag{9}$$

where $G(\cdot) \in L^2(\gamma)$, $A(N)$ is a suitable normalization, and $[\cdot]$ stands for the integer part.

Theorem 2.6. *(Dobrushin and Major [31], Taqqu [76], Breuer and Major [15], Major [57].) Suppose that G has Hermite rank k. Then the following conclusions hold.*
- *Central limit case: suppose that*

$$H < 1 - \frac{1}{2k}.$$

Then $\{G(X(n))\}$ has short memory in the sense that

$$\sigma^2 := \sum_{n=-\infty}^{\infty} \mathrm{Cov}[G(X(n)), G(X(0))]$$

converges absolutely and

$$\frac{1}{N^{1/2}} \sum_{n=1}^{[Nt]} \Big(G(X(n)) - \mathbb{E}G(X(n))\Big) \xrightarrow{f.d.d.} \sigma B(t), \quad t \geq 0, \tag{10}$$

where $\xrightarrow{f.d.d.}$ denotes convergence of the finite-dimensional distributions and $B(t)$ is the standard Brownian motion.
- *Non-central limit case: suppose that*

$$H > 1 - \frac{1}{2k}.$$

Then $\{G(X(n))\}$ has long memory with Hurst index:

$$H_G = (H - 1)k + 1 \in \left(\frac{1}{2}, 1\right). \tag{11}$$

Furthermore, as $N \to \infty$, we have[2]

$$\frac{1}{N^{H_G}} \sum_{n=1}^{[Nt]} \Big(G(X(n)) - \mathbb{E}G(X(n))\Big) \xrightarrow{f.d.d.} c Z_{H_G,k}(t), \tag{12}$$

for some $c \neq 0$, and

$$Z_{H_G,k}(t) = \int_{\mathbb{R}^k}' \left[\int_0^t \prod_{j=1}^k (s - x_j)_+^{\gamma} ds \right] B(dx_1) \ldots B(dx_k), \quad \gamma = H - \frac{3}{2} = \frac{H_G - 1}{k} - \frac{1}{2}, \tag{13}$$

is the so-called k-th order Hermite process, where $\int_{\mathbb{R}^k}' [\; \cdot \;] B(dx_1) \ldots B(dx_k)$ denotes the k-tuple Wiener-Itô integral with respect to the standard Brownian motion $B(\cdot)$. The prime $'$ indicates that one does not integrate on the diagonals $x_i = x_j$.

Remark 2.7. When the Hermite rank $k = 1$, one has

$$H_G = H$$

and the limit $Z_{H_G,k}(t)$ in Theorem 2.6 is the *fractional Brownian motion* $B_H(t)$, namely, the centered Gaussian process determined by the following covariance structure:

$$\mathrm{Cov}\,[B_H(s), B_H(t)] = \frac{1}{2}\left(|s|^{2H} + |t|^{2H} - |s - t|^{2H}\right).$$

The preceding covariance is shared by all the other Hermite processes. When the Hermite rank $k = 2$, $Z_{H_G,2}(t)$ is called the *Rosenblatt process* (see Rosenblatt [68] and Taqqu [75]). The Hermite process $Z_{H_G,k}(t)$ in (13) admits different representations. See Pipiras and Taqqu [63].

---

[2]In fact, we have weak convergence in the space $D[0,1]$ with uniform metric.

Remark 2.8. The boundary case $H = 1 - \frac{1}{2k}$ typically falls in the central limit theorem regime (convergence to Brownian motion) after modifying the normalization $N^{-1/2}$ to include some slowly varying functions (Theorem 1' of Breuer and Major [15]). In general, the convergence of finite-dimensional distributions $\overset{f.d.d.}{\longrightarrow}$ in the short-memory case cannot be strengthened to weak convergence $\Rightarrow$ in $D[0,1]$ unless some additional assumption is imposed on G, e.g., G being a polynomial (Chambers and Slud [17]).

The long-memory Gaussian $\{X(n)\}$ may be directly used as a model for the long-memory stationary noise. For statistical theory, however, it is often desirable to allow departure from Gaussianity, e.g., to accommodate the situations where the noise distribution is skewed or heavy-tailed. Within the same framework, a way to achieve such flexibility is as follows. Suppose that there is an underlying long-memory *Gaussian* stationary process $\{Y(n)\}$. Assume without loss of generality that $\{Y(n)\}$ is standardized. Now suppose that the noise sequence $\{X(n)\}$ in the model is given by

$$X(n) = F(Y(n)). \tag{14}$$

Remark 2.9. There are different perspectives to interpret (14). First, note that when $F(\cdot)$ is nonlinear, $X(n)$ is non-Gaussian. So $F(\cdot)$ can represent the departure from the ideal Gaussian assumption. Hence when the noise $X(n)$ is modeled by (14) with an unknown $F(\cdot)$, this provides great model flexibility. Note that a proper choice of $F(\cdot)$ can match any marginal distribution for $X(n)$. Second, from the perspective of analysis of robustness, one may view $X(n)$ as a perturbed version of $Y(n)$, where $X(n)$ is close to $Y(n)$, that is, $F(\cdot)$ is close to the identity function.

Following the same statistical inference procedure that leads to (9), we then focus on the distributional limit of

$$\frac{1}{A(N)} \sum_{n=1}^{[Nt]} \Big(G \circ F(Y(n)) - \mathbb{E} G \circ F(Y(n)) \Big)$$

as $N \to \infty$.

Remark 2.10. We emphasize the different roles played by $F(\cdot)$ and $G(\cdot)$. **The function $F(\cdot)$ accounts for an unknown and uncontrollable departure from the Gaussian $Y(n)$. On the other hand, the function $G(\cdot)$ depends on the statistical procedure of interest and is therefore typically precisely known**. For example, G is typically the identity transformation for inference of the mean $\mathbb{E} X(n) = \mathbb{E} F(Y(n))$.

2.3 Basic claim

We are now ready to make the following claim which will be justified below. The case of non-instantaneous (multivariate) F (and also G) will be addressed in Appendix A (the issues remain essentially the same).

Claim 2.11. *It is typically the case that*

(a) the function $G \circ F$ has Hermite rank 1;

(b) the process $\{X(n)\}$ has long memory with the same Hurst index H as $\{Y(n)\}$ in (14).

Justification of the claim: Let Z be standard Gaussian. Since $H_1(Z) = Z$, requiring the function $G \circ F$ to have Hermite rank $k \geq 2$ is equivalent to

$$\mathbb{E}[(G \circ F)(Z)Z] = 0. \tag{15}$$

This requirement is very restrictive, and is, moreover, unrelated to the usual size or smoothness conditions typically imposed on the perturbation F. ***Unlike the precisely known $G(\cdot)$ which is related to the method of inference considered, one has no control nor accurate knowledge of the function $F(\cdot)$***. There is thus no a priori reason that $F(\cdot)$ be such that (15) holds. But if (15) does not hold, then the Hermite rank of $G \circ F$ is 1, which justifies part (a) of the Claim 2.11.

Applying the same reasoning, the perturbation function $F(\cdot)$ is also very likely to be such that

$$\mathbb{E}[F(Z)Z] \neq 0$$

and hence to have Hermite rank 1. Then in view of Lemma 2.4 and Remark 2.5, this justifies part (b) of the Claim 2.11 (b). □

Remark 2.12. The Claim 2.11 indicates not only the instability of a Hermite rank higher than 1, but also the stability of the Hermite rank 1 and hence the Hurst index of the noise model. Then, as suggested by the Claim 2.11, if $G \circ F$ has Hermite rank 1, by Theorem 2.6 and Remark 2.7, one has

$$\frac{1}{N^H} \sum_{n=1}^{[Nt]} \Big(G \circ F(Y(n)) - \mathbb{E}G \circ F(Y(n)) \Big) \Rightarrow cZ_{H,1}(t) = cB_H(t). \tag{16}$$

for some $c > 0$, where $B_H(t)$ is the fractional Brownian motion with Hurst index H. The theorem thus also implies the stability of fractional Brownian motion as the limit.

Remark 2.13. In statistics, one sometimes needs limit theorems for functionals other than the sum. A typical example is the quadratic form $\sum_{n,m=1}^{N} a(n-m)X(n)X(m)$. Limit theorems in this case depends on the not only the "memory" of $\{X(n)\}$ but also the "memory" of the coefficient $a(n)$ (see, e.g., Avram [2] and Terrin and Taqqu [79]). Instead of discussing in general the instability of such quadratic forms, we shall focus in Section 3.3 below on an important statistical application, namely, Whittle estimation.

2.4 The level shift case

One may consider making Claim 2.11 a genuine mathematical statement by, for example, considering $F(\cdot)$ as a *random* element in a suitable function space with a "prior probability model", as long as that model assigns a small probability to the set of $F(\cdot)$ on which (15) happens.

In the following theorem, we consider the simple case where the perturbation is given by a level shift of size z, namely, if $F(y) = z + y$ so that $G \circ F(\cdot) = G(\cdot + z)$. To understand the assumptions, note that we want to exclude the case where $G(\cdot)$ is constant, since then $G(\cdot + z)$ remains equal to $G(\cdot)$. We also want $G(\cdot + z)$ to be in $L^2(\gamma)$.

Theorem 2.14. *Suppose that the function $G(\cdot) \in L^2(\gamma)$ has an arbitrary Hermite rank, $G(\cdot)$ is not constant a.e. , and assume that there exists $\delta > 0$, so that $G(\cdot + z) \in L^2(\gamma)$ for all $|z| < \delta$. Then there exists $\epsilon \in (0, \delta)$, such that the Hermite rank of $G(\cdot + z)$ is 1 for all $z \in (-\epsilon, 0) \cup (0, +\epsilon)$.*

The proof can be found in Bai and Taqqu [6]. In that paper, we also study what happens when the shift tends to zero as the sample size tends to infinity, which is analogous to the near integration analysis of unit roots (see Phillips [62]). In Bai and Taqqu [6], we also consider transformations other than the shift, e.g., the scaling $F(z) = zy$ so that $G \circ F(y) = G(zy)$.

2.5 Transformation of linear processes

Another popular class of models for a stationary, not necessarily Gaussian, noise $\{Y(n)\}$ is the so-called (causal) linear process:

$$Y(n) = \sum_{i=0}^{\infty} a_{n-i}\epsilon_i, \tag{17}$$

where ϵ_i's are assumed to be i.i.d. random variables (not necessarily Gaussian) with mean 0 and variance 1 and $\sum_n a_n^2 < \infty$. When

$$a_n \approx n^{H-3/2}, \quad 1/2 < H < 1 \quad \text{as} \quad n \to \infty, \tag{18}$$

one has $\mathrm{Cov}[Y(n), Y(0)] \approx n^{2H-2}$, and thus $Y(n)$ has long memory with Hurst index H. The well-known fractionally-integrated noise model (see, e.g., Granger and Joyeux [40]) satisfies (18). We shall assume (18) throughout this section.

Theorem 2.6 can be extended to linear processes. In this case, the larger class of polynomials called Appell polynomials (Avram and Taqqu [3]) plays an analogous role to that of the Hermite polynomial in Section 2.1. One can define the so-called *Appell* rank of a function $G(\cdot)$ as in (5), with Hermite polynomials replaced by the Appell polynomials, given that the expansion is valid (for example, when $G(\cdot)$ is a finite-order polynomial). However, in this framework (Surgailis [73]), the class of functions $G(\cdot)$'s that can be treated is rather restrictive. Ho and Hsing [46] greatly extended the allowable $G(\cdot)$'s through a martingale difference approach and introduced a more convenient notion of rank, which we shall call the *power rank*. See also Lévy-Leduc and Taqqu [54].

Given a function $G(\cdot)$ and a random variable Y satisfying $\mathbb{E}G(Y)^2 < \infty$, let

$$G_\infty(y) = \mathbb{E}G(Y + y) \tag{19}$$

given that the expectation exists and suppose that $G_\infty(\cdot)$ has derivatives of order sufficiently high. The power rank of $G(\cdot)$ with respect to Y is defined as

$$\inf\{m \geq 1 : \ G_\infty^{(m)}(0) \neq 0\}, \tag{20}$$

where $G_\infty^{(m)}(y)$ denotes the m-th derivative of $G_\infty(y)$. In fact, the power rank in (20) coincides with the Hermite rank (4) if Y is Gaussian. This was stated in Ho and Hsing [46], and see Bai and Taqqu [6] for a proof. The case where G is a polynomial was treated in Lévy-Leduc and Taqqu [54].

Now we can state the following limit theorems (for simplicity we omit the inclusion of some technical conditions, see Ho and Hsing [46] and Pipiras and Taqqu [64]):

Theorem 2.15 (Ho and Hsing [46]). *Suppose that $\mathbb{E}G(Y(0))^2 < \infty$ and $G(\cdot)$ has power rank $k \geq 1$ with respect to $Y(0)$ in the sense of (20). Under some additional technical conditions, statements exactly analogous to Theorem 2.6 hold with the role of the Hermite rank replaced by the power rank.*

Remark 2.16. Using similar arguments as Section 2.1, one sees that a power rank higher than 1 is also unstable to perturbation: to get a power rank higher than 1, one needs the restrictive condition

$$G_\infty^{(1)}(0) = \frac{d}{dy}\mathbb{E}G(Y + y)|_{y=0} = 0.$$

which can be easily perturbed by compositing G with a transformation before. So an analog of Claim 2.11 may be stated in this context.

Below we provide some further remarks on the instability phenomenon.

Remark 2.17. We mention that Surgailis [74] established some results which can be interpreted as the "robustness" of Theorem 2.15 against additive noise. Roughly speaking, Surgailis [74] showed that if the long memory linear process $Y(n)$ is replaced by $Y'(n) = Y(n) + U(n)$ with $U(n)$ specified as some short-memory models, then the non-central convergence in Theorem 2.15 still holds, where the power (or Appell) rank is now with respect to the distribution of $Y'(n)$. Nevertheless, the instability discussed earlier still applies. Firstly, the rank can still be unstable under a transformation. Secondly, even without considering a transformation perturbation, one has typically no accurate knowledge of the marginal distribution of $Y'(n)$. A change in the distribution will affect the rank defined through (20).

Remark 2.18. It is important to consider not only the limit distribution that one obtains, but also the normalization since the latter corresponds to the magnitude of the fluctuations of the partial sum. When the true rank (Hermite or power) is indeed 1, but one assumes a higher-order Hermite rank from some statistical consideration, this will lead to underestimation of the magnitude of the fluctuations of the partial sum since $H_G < H$ in Theorem 2.6.

3 Examples in Statistics

In this section, we review statistical problems in the literature related to limit theorems involving different ranks. We shall elaborate on some examples: sample variance, empirical processes, Whittle likelihood, and nonparametric estimation, to demonstrate how the asymptotic statistical theories are affected by the instability discussed in Section 2.

3.1 Sample variance

In the context of long memory, with the complexity introduced by the limit theorems, scale estimation becomes a problem. We only discuss the case where the data $\{X(n)\}$ is a Gaussian process, but everything can be extended to a linear process $\{X(n)\}$ (see Section 2.5).

Assume then that $\{X(n)\}$ is a long memory stationary Gaussian process with Hurst index $H \in (1/2, 1)$, unknown mean μ and unknown variance σ^2. Consider the estimation of σ^2 using the sample variance:

$$\widehat{\sigma}_N^2 := \frac{1}{N} \sum_{n=1}^{N} \left(X(n) - \bar{X}_N \right)^2, \tag{21}$$

where $\bar{X}_N = (X(1) + \ldots + X(N))/N$ is the sample mean. In the short memory situation, say if $X(n)$ were i.i.d., it is well known that $\widehat{\sigma}_N^2$ is asymptotically normal. The situation is, however, delicate. Indeed, express (21) as

$$\widehat{\sigma}_N^2 = \frac{1}{N} \sum_{n=1}^{N} (X(n) - \mu)^2 + (\bar{X}_N - \mu)^2 =: U_N + V_N. \tag{22}$$

We can write

$$U_N - \sigma^2 = \frac{1}{N} \sum_{n=1}^{N} (X(n) - \mu)^2 - \sigma^2 = N^{-1/2} \left[\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \left((X(n) - \mu)^2 - \sigma^2 \right) \right], \tag{23}$$

and

$$V_N = (\bar{X}_N - \mu)^2 = N^{2H-2} \left[\left(\frac{1}{N^H} \sum_{n=1}^{N} (X(n) - \mu) \right)^2 \right]. \tag{24}$$

Note that the term $X(n) - \mu$ has Hermite rank $k = 1$, and the term $(X(n) - \mu)^2$ has expectation σ^2 and Hermite rank $k = 2$, since $\mathbb{E}(X(n) - \mu)^2 X(n) = 0$.

Thus when $H < 3/4$, in view of Theorem 2.6, the term in the brackets in the right-hand sides (23) and (24) converge as $N \to \infty$. Since $H < 3/4$ implies $N^{2H-2} \ll N^{-1/2}$, the term V_N is asymptotically negligible, compared to $U_N - \sigma^2$, and hence $N^{1/2}(\widehat{\sigma}_N^2 - \sigma^2)$ has the same limit as $N^{1/2} U_N$ as $N \to \infty$. Thus

$$N^{1/2}(\widehat{\sigma}_N^2 - \sigma^2) \overset{d}{\to} N(0, s_1^2)$$

for some $s_1 > 0$.

When $H > 3/4$, in view of Theorem 2.6 with $k = 2$, we write

$$U_N - \sigma^2 = N^{2H-2} \left[\frac{1}{N^{2H-1}} \sum_{n=1}^{N} \left((X(n) - \mu)^2 - \sigma^2 \right) \right],$$

and V_N is as in (24). Now both U_N and V_N contribute to the limit, where we have by a multivariate version of Theorem 2.6 (see, e.g, Bai and Taqqu [4]):

$$N^{2-2H}(\widehat{\sigma}_N^2 - \sigma^2) \overset{d}{\to} a_H Z_{2H-1,2}(1) + b_H Z_{H,1}(1)^2 \tag{25}$$

where $Z_{2H-1,k}(t), k = 1, 2$ are the Hermite processes in (13) defined by the same Brownian integrator $B(\cdot)$. for some constants $a_H, b_H > 0$. See also Dehling and Taqqu [27].

We now suppose that $\{X(n)\}$ is perturbed by a transformation in the spirit of Claim 2.11, which leads to consider the case where both $X(n)$ and $[X(n) - \mu]^2$ have Hermite rank 1. Then writing

$$U_N - \sigma^2 = N^{H-1} \left[\frac{1}{N^H} \sum_{n=1}^{N} \left((X(n) - \mu)^2 - \sigma^2 \right) \right],$$

and

$$V_N = N^{2H-2} \left[\frac{1}{N^H} \sum_{n=1}^{N} (X(n) - \mu) \right]^2,$$

we can apply Theorem 2.6 with $k = 1$. Since $H < 1$, we have $N^{2H-2} \ll N^{H-1}$, and thus only the term U_N contributes to the limit. Then $N^{1-H}(\widehat{\sigma}_N^2 - \sigma^2)$ has the same limit as

$$N^{1-H}(U_N - \sigma^2) = \frac{1}{N^H} \sum_{n=1}^{N} [(X(n) - \mu)^2 - \sigma^2],$$

namely, $c_H Z_{H,1}(1)$ for some $c_H > 0$, where $Z_{H,1}$ is the fractional Brownian motion. Hence

$$N^{1-H} \left(\widehat{\sigma}_N^2 - \sigma^2 \right) \xrightarrow{d} c_H Z_{H,1}(1), \tag{26}$$

which is different from (25).

Remark 3.1. Under the above perturbation consideration, there is no dichotomy between $H < 3/4$ and $H > 3/4$ in (26), and the normalization is always N^{1-H}, which is of smaller order than both $N^{1/2}$ and N^{2H-2}. In the case $H < 3/4$, however, we get a Gaussian limit with or without perturbation. Hence without the perturbation consideration, there is the danger of underestimating the fluctuation magnitude of the sample variance, namely, taking the fluctuation to be of the order $N^{-1/2}$ when $H < 3/4$ and N^{2H-2} when $H > 3/4$, whereas they are of the order N^{H-1}. We also mention that similar considerations also apply to the study of the asymptotic behavior of sample autocovariance/correlation (see, e.g., Hosking [48], Wu et al. [84] and Lévy-Leduc et al. [56]).

3.2   Empirical processes

Empirical processes play important roles in many statistical problems. We refer the reader to Dehling et al. [28] for an introduction to empirical processes of dependent data. Let $\{X(n)\}$ be a stationary process with marginal cdf $F(x)$. The corresponding centered empirical process is defined as

$$F_N(x) = \frac{1}{N} \sum_{n=1}^{N} \left[I\{X(n) \le x\} - F(x) \right]. \tag{27}$$

When $\{X(n)\}$ is i.i.d., it is well-known that $N^{1/2} F_N(x)$ converges weakly in $D(-\infty, \infty)$ to $B_0(F(x)) = B(F(x)) - F(x)B(1)$, where $B_0(t) = B(t) - tB(1)$ is a Brownian bridge and $B(t)$ is a Brownian motion. Under some weak dependence conditions on $\{X(n)\}$, the process $N^{1/2} F_N(x)$ converges weakly in $D(-\infty, \infty)$ to a centered Gaussian process $G(x)$ with covariance structure given by

$$\mathbb{E} G(x)G(y) = \sum_{n=-\infty}^{\infty} \text{Cov}[I(X(0) \le x), I(X(n) \le y)]. \tag{28}$$

See, e.g., Theorem 4.1 of Dehling and Philipp [25].

When $X(n)$ has long memory, the corresponding weak convergence results become rather different in nature. Indeed, assume that $X(n) = G(Y(n))$ where $\{Y(n)\}$ is a standardized stationary Gaussian process with Hurst index $1/2 < H < 1$. We define the deterministic function

$$J_m(x) = \frac{1}{m!} \mathbb{E} I\{G(Y(0)) \le x\} H_m(Y(0)),$$

where $H_m(\cdot)$ is the m-th order Hermite polynomial. Note that for any fixed $x \in \mathbb{R}$, $J_m(x)$'s are the coefficients of the Hermite expansion (5) of the function $\Delta_x(y) = I\{G(y) \le x\} - F(x)$. We have the following result:

Theorem 3.2 (Theorem 1.1 of Dehling and Taqqu [26]). *Let*

$$k = \inf\{m \geq 1 : \ J_m(x) \neq 0 \text{ for at least one } x \in \mathbb{R}\}, \tag{29}$$

and assume that $H > 1 - \frac{1}{2k}$. Then we have the following weak convergence in $D(-\infty, +\infty)$

$$N^{1-H_k} F_N(\cdot) \Rightarrow c J_m(\cdot) Z_{H_k,k}(1) \tag{30}$$

where $Z_{H_k,k}(\cdot)$ is the Hermite process as in (13), and $H_k = (H-1)k + 1$ as in (11).

Remark 3.3. It is interesting to note that in the long memory case, the limit process

$$\{J_m(x) Z_{H_k,k}(1), \ x \in \mathbb{R}\}$$

is quite degenerate, namely, it has correlation 1 between any different points $x_1, x_2 \in \mathbb{R}$, in contrast to the weak dependence case where the limit Gaussian process $G(x)$ admits a rich correlation structure (see (28)).

By the perturbation argument, one may assume that the rank $k = 1$, regardless of the choice of $G(\cdot)$ in a statistical application of Theorem 3.2. In fact, the definition of rank (29) makes the assumption $k = 1$ even more appealing in this context, because $J_1(x) \neq 0$ for just one point x would make $k = 1$. From this point of view, the only practically relevant convergence in (30) is

$$N^{1-H} F_N(\cdot) \Rightarrow c Z J_1(\cdot), \tag{31}$$

where Z is a standard Gaussian variable, and thus the fluctuation of the empirical process is practically always of the order N^{H-1}. The convergence (30) can be applied to study the asymptotic behavior of U-statistics and V-statistics (see Corollary 1 of Dehling and Taqqu [26]). It is also applied to develop the asymptotic theories of estimation of the probability density function $f = F'$ (see Csörgö and Mielniczuk [23] and Section 3.5).

3.3 Whittle likelihood

In the parametric estimation for time series, the so-called Whittle pseudo-likelihood is a computationally efficient approximation to the Gaussian likelihood, which bypasses the inversion of a covariance matrix in the latter. The resulting Whittle estimator and its semiparametric extensions are found particularly useful in the long memory context for the estimation of the Hurst parameter H. For more details on the background and motivation, we refer to Section 5.5 of Beran et al. [11] or Chapter 10 of Pipiras and Taqqu [64]. We shall focus on the rank instability issue in the asymptotic theory developed in Giraitis and Taqqu [37]. The asymptotic theory in this context depends on the limit theorem for quadratic forms, which is more delicate than the limit theorems for sums. The instability issue in this context exhibits some distinct features compared with the previous cases.

Suppose a stationary time series $\{X(n)\}$ has spectral density (see, e.g., Chapter 1 of Pipiras and Taqqu [64]) $f(\lambda; \theta, \sigma) = \sigma^2 g_\theta(\lambda) > 0$, $\lambda \in (-\pi, \pi)$ so that

$$\text{Cov}[X(n), X(0)] = \int_{-\pi}^{\pi} e^{inx} f_\theta(\lambda; \theta, \sigma) d\lambda,$$

where σ and $\theta = (\theta_1, \ldots, \theta_p)$ are unknown parameters. Assume that the normalization condition (scale identifiability) holds:

$$\int_{-\pi}^{\pi} \log g_\theta(\lambda) d\lambda = 0,$$

under which σ^2 becomes the mean squared error of the one-step prediction by the Kolmogorov's formula (see, e.g., Section 5.8 of Brockwell and Davis [16]). Suppose that we want to estimate the unknown parameter θ. Under long memory, the choice of θ typically includes H. Define

$$a_\theta(n) = \int_{-\pi}^{\pi} e^{in\lambda} \frac{1}{g_\theta(\lambda)} d\lambda \tag{32}$$

and
$$w_\theta(m,n) = \int_{-\pi}^{\pi} g_\theta(\lambda) \frac{\partial^2}{\partial\theta_m \partial\theta_n}[g_\theta(\lambda)]^{-1} d\lambda.$$

The so-called Whittle estimator $\widehat{\theta}_N$ of θ is given by

$$\widehat{\theta}_N = \underset{\theta}{\text{argmin}} \sum_{m,n=1}^{N} a_\theta(m-n)X(m)X(n).$$

If $\{X(n)\}$ is a Gaussian or a linear long-memory process, it was established under some regularity conditions that (see, e.g., Fox and Taqqu [35] and Giraitis and Surgailis [36])

$$N^{1/2}(\widehat{\theta}_N - \theta) \overset{d}{\to} N(0, 4\pi W_\theta^{-1}), \tag{33}$$

where the matrix $W_\theta = (w_\theta(m,n))_{1 \le m,n \le p}$. Note that the standard $N^{1/2}$-convergence rate appears even though $\{X(n)\}$ has long memory. This is due to the dependence cancellation effect from the quadratic coefficient $a_\theta(n)$.

On the other hand, Giraitis and Taqqu [37], considered $X(n) = G(Y(n))$, where $\{Y(n)\}$ is long-memory Gaussian and the transformation $G(\cdot)$ is restricted to be a polynomial by Giraitis and Taqqu [37] to avoid some technical difficulties. Define

$$\rho_k = \frac{1}{k!} \sum_{n=-\infty}^{\infty} \mathbb{E}\left[\frac{d^k}{dx^k}G(x+Y(0))G(x+Y(n))\right]\bigg|_{x=0} \nabla a_\theta(n)$$

$$= \sum_{m,n \ge 0, m+n=k} \frac{1}{m!n!} \sum_{r=-\infty}^{\infty} \mathbb{E}\left[G^{(m)}(Y(r))G^{(n)}(Y(0))\right] \nabla a_\theta(r),$$

where ∇ denotes the gradient with respect to θ. In particular,

$$\rho_1 = 2 \sum_{n=-\infty}^{\infty} \mathbb{E}[G'(Y(n))G(Y(n))]\nabla a_\theta(n). \tag{34}$$

Note that in the case $G(x) = x$, namely, the Gaussian case, $\rho_k = 0$ for all $k = 1, 2, \ldots$.

It was established in Corollary 2.1 of Giraitis and Taqqu [37] that under some regularity conditions, if $\rho_1 \ne 0$, then as $N \to \infty$, we have

$$N^{1-H}(\widehat{\theta}_N - \theta) \overset{d}{\to} Z \tag{35}$$

for some centered normal random vector Z. Note that in (35) the convergence rate is the same as that of the sample mean in view of (1). Giraitis and Taqqu [37] also showed that if $\rho_1 = 0$ but some $\rho_k \ne 0$, then (35) needs to be modified resulting in limit theorem with a central and non-central dichotomy similar to Theorem 2.6. See Theorem 2.3 and 3.1 of Giraitis and Taqqu [37].

Now we consider the instability issue. Here the role of Hermite (or power) rank is instead played by

$$k = \inf\{m \in \mathbb{Z}_+ : \rho_m \ne 0\}. \tag{36}$$

There is instability even in the Gaussian case where $G(\cdot)$ is the identity, namely, $G(x) = x$, and where then all $\rho_k = 0$. In that case, we would have (33), but by perturbing $G(\cdot)$ slightly, we would get $\rho_1 \ne 0$ in (34), and thus we would have (35) instead of (33).

The preceding observation raises a question on the applicability of (33) in statistical inference. It turns out that the achievement of the parametric rate $N^{1/2}$, or say the cancellation effect of the quadratic coefficient in (32), critically depends on the Gaussian or linear data-generating assumption, while a disturbance of such an assumption yields instead the rate N^{1-H}, which is the usual slower rate of convergence under long memory. It is unclear whether similar instability issues occur in the semiparametric extensions of the Whittle estimator, e.g., the local Whittle estimator (Künsch [53], Robinson [66]).

3.4 Nonparametric estimation

In this section, we review briefly some nonparametric statistical studies under long memory involving the Hermite or power rank. Assume throughout that $\{X(n)\}$ is a stationary *long-memory* process, typically specified by a Gaussian process, or a linear process, or a transformation of either (we call the model a Gaussian or linear subordination).

In the kernel smoother type nonparametric estimation procedures, a nonlinear transformation of the data is naturally involved. For example, the kernel density estimator of the probability density function $f(x)$ is defined as

$$\widehat{f}(x) = \frac{1}{Nh} \sum_{n=1}^{N} K\left(\frac{x - X(n)}{h}\right), \qquad x \in \mathbb{R}. \tag{37}$$

where N is the sample size, $h > 0$ is the bandwidth parameter, and $K(\cdot)$ is a kernel satisfying $\int_{\mathbb{R}} K(x)dx = 1$. A number of studies have considered the asymptotic behavior of the estimator $\widehat{f}(x)$ as $N \to \infty$ and $h \to 0$. See, for example, Cheng and Robinson [18], Csörgö and Mielniczuk [23] and Ho [45],Wu and Mielniczuk [83]. Another typical class of statistical procedures involving kernel smoothers are the nonparametric local regressions (e.g., Nadaraya-Watson estimator and local polynomial estimators). Some relevant work involving the ranks are Hidalgo [44] , Csörgö and Mielniczuk [24], Masry and Mielniczuk [59], Guo and Koul [41].

We mention that under long memory, the asymptotic behavior of kernel smoothers can be quite different from the short-memory case. In particular, an interesting dichotomy phenomenon appears in the asymptotics depending on how fast the bandwidth b_n tends to 0 with respect to the Hurst index H of $X(n)$. If b_n tends to 0 relatively slowly, one can have a very degenerate behavior such as the kernel density estimate $\widehat{f}(x)$ at different points of x becomes asymptotically perfectly correlated. See, e.g, Csörgö [22] as well as Chapter 5.14 of Beran et al. [11]).

Asymptotic results involving applying limit theorems with different ranks when studying these kernel smoother procedures are due more often to the assumption that $X(n)$ is a transformation of a Gaussian or linear process, than due to the nonlinear transformation produced by $K(\cdot)$ in (37). To obtain a higher-order rank for $\widehat{f}(x)$ when $X(n)$ is Gaussian or linear, one has to be in very special situations, e.g., when focusing on the asymptotic distribution of $\widehat{f}(x)$ in (37) at a fixed point $x = x_0$ while assuming that the true density satisfies $f'(x_0) = 0$ (see, e.g., Theorem 3 of Wu and Mielniczuk [83]). Similar considerations extend to other nonparametric procedures, e.g, the spline regression under long memory noise (Beran and Weiershäuser [10]).

3.5 Miscellaneous

Wavelets are useful tools for analyzing long-memory data due to their natural adaptivity to scaling. Limit theorems involving ranks were applied, e.g., in Clausel et al. [20] and Clausel et al. [21], who studied the asymptotic behaviors of the wavelet coefficients and the wavelet estimation of Hurst index of the Gaussian subordination data.

Some other statistical studies involving higher-order ranks limit theorems are: bivariate U-processes (Lévy-Leduc et al. [55]), change-point test (Zhao et al. [86], Dehling et al. [29]), goodness-of-fit test in regression Koul and Stute [52], normality test (Beran and Ghosh [9]), sign test (Psaradakis [65]), unit root test (Wu [82]).

4 Empirical evidence

In this section, we provide empirical evidence to support the preceding discussion of instability of ranks in the limit theorems under long memory.

Consider the rank of the quadratic transformation

$$G(x) = x^2.$$

It is always 2, in both the Gaussian and linear subordination context. This means that if $\{X(n)\}$ is exactly a centered Gaussian or linear process with Hurst index $H > 1/2$, then the Hurst index of the transformed

series $\{X(n)^2\}$ should be

$$H_G = \max\left(\frac{1}{2}, 2H - 1\right) < H, \tag{38}$$

in view of Theorem 2.6 and 2.15. Note that when $H < 0.75$, the resulting Hurst index is always $H_G = 0.5$ unless in the special case where the sum of covariances of all orders is zero (ainti-persistency).

Here is the question: if $\{X(n)\}$ is a real-life centered stationary data in which displays long memory, does one typically observe the decrease from H to H_G as in (38) when $\{X(n)\}$ is replaced by $\{X(n)^2\}$? If our arguments in the previous sections make practical sense, then the time series $\{X(n)^2\}$ should most likely still possess rank 1, which means that (38) should barely happen. To test this hypothesis, we design the following empirical study which involves $\{X(n)\}$ and $\{X(n)^2\}$. The design is explained in Remark 4.1 below.

Design of the study:

Suppose that we have a collection of M real-life stationary long-memory time series data

$$\{X_m(n), \ n = 1, \ldots, N_m \ m = 1, \ldots, M.\},$$

where n is the time index, and m is the data set index. For each m, we perform the following analysis.

Step 1 For each $m = 1, \ldots, M$, center the data:

$$X_m(n) \leftarrow X_m(n) - \frac{1}{N_m}\sum_{n=1}^{N_m} X_m(n);$$

Step 2 For each $m = 1, \ldots, M$, obtain

the estimated Hurst index $\widehat{H}_m^{(1)}$ of $\{X_m(n) \ n = 1, \ldots, N_m\}$

and

the estimated Hurst index $\widehat{H}_m^{(2)}$ of $\{X_m(n)^2, \ n = 1, \ldots, N_m\}.$

Step 3 For each $m = 1, \ldots, M$, simulate R independent sequences of fractional Gaussian noise (increments of fractional Brownian motion):

$$\{G_{mr}(n), \ n = 1, \ldots, N_m, \ r = 1, \ldots, R\},$$

all with Hurst index $\widehat{H}_m^{(1)}$. Obtain

the estimated Hurst index $\widehat{h}_{mr}^{(1)}$ of $\{G_{mr}(n), \ n = 1, \ldots, N_m\}$

and

the estimated Hurst index $\widehat{h}_{mr}^{(2)}$ of $\{G_{mr}(n)^2, \ n = 1, \ldots, N_m\},$

for each $r = 1, \ldots, R$.

Step 4 For each $m = 1, \ldots, M$, compute

$$\delta_m = \widehat{H}_m^{(2)} - \max(\frac{1}{2}, 2\widehat{H}_m^{(1)} - 1)$$

from the data, and compute

$$\left\{\delta_{mr} := \widehat{h}_{mr}^{(2)} - \max\left(\frac{1}{2}, 2\widehat{h}_{mr}^{(1)} - 1\right), \ r = 1, \ldots, R\right\}$$

from the simulated series. Then compute the relative number of times (percentile) that δ_{mr} is less than or equal to δ_m for $r = 1, \ldots, R$, namely,

$$P_m = \frac{1}{R}\#\{\delta_{mr} : \ \delta_{mr} \le \delta_m, \ r = 1, \ldots, R\} = \widehat{F}_{m,R}(\delta_m),$$

where $\widehat{F}_{m,R}$ is the empirical CDF of $\{\delta_{mr} : \ r = 1, \ldots, R\}$;

13

<u>*Step 5*</u> *Construct the following contrast group: for each $m = 1, \ldots, M$, simulate a fractional Gaussian noise sequence*

$$\{X_m^*(n), \ n = 1, \ldots, N_m^*\}$$

with Hurst index randomly sampled from $\{\widehat{H}_m^{(1)}, \ m = 1, \ldots, M\}$, and length N_m^* randomly sampled from $\{N_m, \ m = 1, \ldots, M\}$.

Then perform the preceding steps 1-4 replacing $\{X_m(n)\}$ by $\{X_m^*(n)\}$, from which one gets $\widehat{H}_m^{(1)*}$, $\widehat{H}_m^{(2)*}$, δ_m^* and P_m^* that correspond to $\widehat{H}_m^{(1)}$, $\widehat{H}_m^{(2)}$, δ_m and P_m respectively.

In our study we set $R = 200$.

Remark 4.1. We explain here the preceding study design. Recall that $\widehat{H}_m^{(1)}$ is the Hurst index estimate of the time series $\{X_m(n)\}$ and $\widehat{H}_m^{(2)}$ is the Hurst index estimate of the squared time series $\{X_m(n)^2\}$. As mentioned before, the goal is to examine whether $\widehat{H}_m^{(1)}$ and $\widehat{H}_m^{(2)}$ behave according to (38). If they behave perfectly according to (38), then

$$\delta_m = \widehat{H}_m^{(2)} - \max\left(\frac{1}{2}, 2\widehat{H}_m^{(1)} - 1\right)$$

should be zero. Both $\widehat{H}_m^{(1)}$ and $\widehat{H}_m^{(2)}$ are random and thus fluctuate as m varies. To get a reference point, we introduce a statistical contrast in Step 3, whereby we simulate R fractional Gaussian noises series and measure their Hurst indices $\widehat{h}_{mr}^{(1)}$ and $\widehat{h}_{mr}^{(2)}$: the first index is for fractional Gaussian noise and the second is for its square. Since these Hurst indices are obtained from fractional Gaussian noises, they indeed obey (38). We want to see how δ_m, which is measured from data, compares to the $\delta_{mr} = \widehat{h}_{mr}^{(2)} - \max\left(\frac{1}{2}, 2\widehat{h}_{mr}^{(1)} - 1\right)$ corresponding to fractional Gaussian noise. This leads us to focus on P_m instead of δ_m. One may view P_m as a "standardized" version of δ_m with respect to the contrast distribution constructed from fractional Gaussian noise, which makes comparison across different data items (different m) more consistent. More technical explanations are given below.

Let $F_m(x)$ be the CDF of the random δ_m. Then $F_m(\delta_m)$ follows exactly a *uniform distribution* on $[0, 1]$. If $\{X_m(n)\}$ were indeed generated by fractional Gaussian noise with true Hurst index $\widehat{H}_m^{(1)}$, then the empirical CDF $\widehat{F}_{m,R}$ in Step 4 is a good approximation of F_m. Therefore, if $\{X_m(n)\}$ obeys (38) as the fractional Gaussian noise does, and $\widehat{H}_m^{(1)}$ is a reasonable estimate, then $P_m = \widehat{F}_{m,R}(\delta_m)$ in Step 4 is expected to follow a *uniform distribution* on $[0, 1]$ approximately. On the other hand, if the δ_m computed from the data makes the distribution of $P_m = \widehat{F}_{m,R}(\delta_m)$ skewed towards 1, then this indicates that δ_m tends to be larger than δ_{mr}.

To account for the potential bias due to the estimation of the Hurst index, in Step 5 we replace our original data $\{X_m(n)\}$ by a second contrast group $\{X_m^*(n)\}$ made up of fractional Gaussian noise sequences with similar lengths and Hurst indices. After repeating the same procedure on this contrast group, we can then compare the distribution (histogram) of $\{P_m\}$ obtained from the original data with the distribution of $\{P_m^*\}$ obtained from the contrast group.

These designs may be regarded as simulation-assisted statistical tests where the null hypothesis is the relation (38).

Now we describe the data we use. The *tree ring width* in chronological order has been identified as one of the natural stationary time series data sets which exhibit long memory (see Mandelbrot and Wallis [58] and Pelletier and Turcotte [61]). Since the tree ring width is largely affected by environmental factors, which is explored in dendrochronology (see Schweingruber [70]), it also reflects the long-memory stationary fluctuation of the ecological systems. We shall use the data compiled by The International Tree-Ring Data Bank (ITRDB, `ftp://ftp.ncdc.noaa.gov/pub/data/paleo/treering/chronologies/`) collected from Africa, Asia, Australia, Canada, Europe, Mexico, South America and USA, stored in the Standard Chronology File (*.crn) format. For example, Figure 1 displays the time series extracted from the file ca506.crn in the data bank and its autocorrelation plot. We further select the data according to the following criteria:
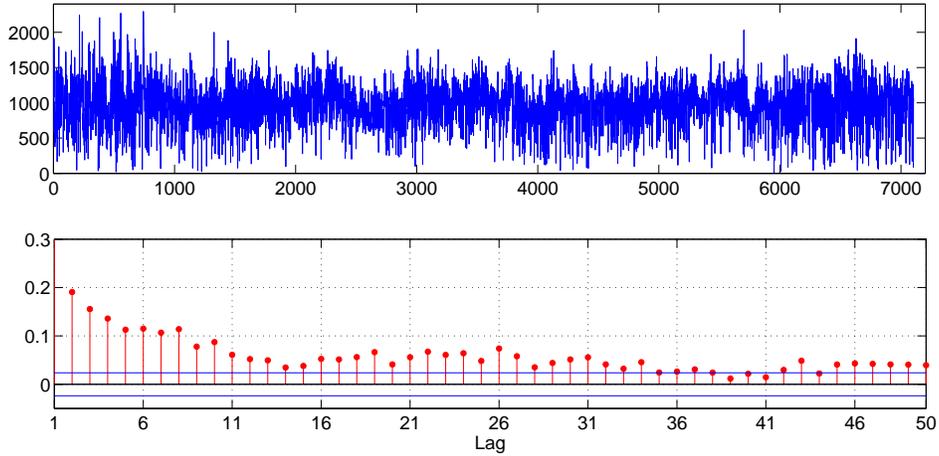
Figure 1: *Up:* Plot of the tree ring time series extracted from ca506.crn in ITRDB. *Down:* autocorrelation plot. The variance aggregation estimate for the Hurst index of the data yields $\widehat{H}^{(1)} = 0.7182$ and the Hurst index for the centered and squared data yields $\widehat{H}^{(2)} = 0.7217$; the local periodogram regression yields $\widehat{H}^{(1)} = 0.7569$ and $\widehat{H}^{(2)} = 0.7801$ respectively; the local Whittle estimate yields $\widehat{H}^{(1)} = 0.7024$ and $\widehat{H}^{(2)} = 0.7061$ respectively.

Criterion 1 The length of the time series is at least 300.

Criterion 2 The time series data is importable by the Tree-Ring Matlab Toolbox[3] (data is usually importable if there is no missing value).

Criterion 3 The estimated Hurst index $\widehat{H}_m^{(1)}$ lies within the interval $[0.6, 0.9]$[4].

To be consistent, we also apply Criterion 1 and Criterion 3 for to the contrast group $\{X_m^*(n),\ n = 1, \ldots, N_m\}$.

We shall use the following three popular estimators of Hurst index:

- Variance aggregation estimator;
- Local periodogram regression estimator (also known as GPH estimator);
- Local Whittle estimator.

For a description and empirical study of these estimators, see Taqqu et al. [78]. There are more sophisticated estimators, for example, the wavelet-type estimators (see, e.g., Faÿ et al. [34]). To minimize finite-sample bias, these methods typically involve complicated choice of some tuning parameters. Since our study design has taken into account the potential bias of the estimator, we shall stick to the three more elementary estimators aforementioned. For the variance aggregation estimator and the local periodogram regression (GPH estimator), we use the implementation by Chu Chen (`http://www.mathworks.com/matlabcentral/fileexchange/19148-hurst-parameter-estimate`, and we use the default parameter settings); For the local Whittle estimate, we use the implementation by Katsumi Shimotsu (`http://shimotsu.web.fc2.com/Site/Matlab_Codes.html`), in which case we choose the frequency cutoff threshold to be $[N^{2/3}]$ with N being the length of the time series).

Observations:

The graphs in the right-hand side of Figure 2, 3 and 4 are as expected, namely, corresponding roughly to a uniform distribution. This indicates that the procedure described in the study is reasonable. In fact, the

---

[3]`http://www.ltrr.arizona.edu/~dmeko/toolbox.html`

[4]Ideally we want the selected data to be stationary and long-range dependent. When the estimate is close to 0.5, the data is likely to have short memory; when the estimate is close to 1, it is likely to be non-stationary.

Estimator	Selected number M	Median δ_m	Median δ_m^*	Median P_m	Median P_m^*
Variance Aggregation	1250	0.0786	0.0104	80.50%	51.00%
Local Periodogram Regression	658	0.0921	-0.0204	86.25%	63.50%
Local Whittle	908	0.0496	-0.0162	80.50%	52.50%

Table 1: Analysis Summary

median of P_m^* is roughly 50% as it should be (see Table 1). As mentioned below, there may be a small bias when using the Local Periodogram Regression method (Figure 3 (right)). See also Taqqu and Teverovsky [77] for an empirical discussion of Whittle-type estimators.

Table 1 summarizes some key statistics of the analysis based on the three different estimators. One can see that for all three estimators, the median of δ_m is consistently smaller than that of the contrast δ_m^*. The median of P_m is significantly smaller than that of the contrast P_m^*. Figure 2, 3 and 4 plot the histograms of $\{P_m\}$ and $\{P_m^*\}$ obtained via the three different estimators. Their results are similar: while $\{P_m^*\}$ are roughly uniformly distributed as expected, the histogram of $\{P_m\}$ is severely skewed towards 1. The contrast in the skewness shows that the δ_m computed from the tree ring data tends to be larger than the $\{\delta_{mr}\}$ computed from the fractional Gaussian noise. In other words, in the case of tree ring data, the Hurst index does *not* tend to decrease as much after squaring as the case of fractional Gaussian noise.

As mentioned in Remark 4.1, if the Hurst index estimate is unbiased, P_m^* is expected to approximately follow a uniform distribution on $[0, 1]$, so that the median is close to $1/2$. However, the estimation bias of Hurst index could distort this uniformity. Indeed, in the Local Periodogram Regression case, the median of P_m^* is 63.5%. But this is still in sharp contrast with the corresponding median of P_m which is 86.25% and hence significantly larger. This indicates that the data is not behaving like fractional Gaussian noise. Thus our design is effective despite the bias inherent in the estimation method.

Remark 4.2. From the analysis above, we conclude that relation (38), or more generally (11), may not make good prediction on real-life data. We note, however, that the estimated Hurst index $\widehat{H}_m^{(2)}$ of $\{X_m(n)^2\}$ tends to be somewhat smaller than the estimated Hurst index $\widehat{H}_m^{(1)}$ of $\{X_m(n)\}$, although for the contrast group $\{X^*(n)\}$ the decrease from $\widehat{H}_m^{(1)*}$ to $\widehat{H}_m^{(2)*}$ is more significant. See Figure 5. A possible explanation is that although $\{X_m(n)^2\}$ actually possesses rank 1 and thus has the same Hurst index as $\{X_m(n)\}$, many of the $\{X_m(n)\}$ may be close to a Gaussian (or linear) process. So they tend to exhibit somewhat the relation (38) when the sample size is moderate. See Bai and Taqqu [6] for an analysis of the interplay between the rank instability effect and the sample size.

Remark 4.3. As a reviewer pointed out, another explanation of the observations found in the study is that the data originally follows a model with a rank higher than 1, in which case squaring does not necessarily lead to a higher-order rank. Although this explanation is allowable in theory, it is less natural than the instability explanation. The reviewer's explanation relies on assuming a special model: the transformation of a Gaussian or linear process with higher-order rank, while ours indicates that a slight perturbation makes the formula (38) unrealistic in practice.

5 Stability of limit theorems under weak dependence

In this section, we demonstrate that the instability phenomenon appearing in the limit theorems under long memory does not typically occur in the short-memory case. This is important because it shows that the transformation considered as "perturbation" in the previous section usually does not make any qualitative difference in short-memory situations and hence may be safely negligible in large sample inference.

There are many ways to mathematically characterize weak dependence. For an introduction to various notions of weak dependence of stationary processes and corresponding limit theorems, we refer to Doukhan [32]. In this section, we shall mainly look at the following three as examples:

(1) Fast-decaying mixing coefficients under strong mixing conditions;

(2) Fast decaying covariance function in Gaussian subordination model (Theorem 5.2);
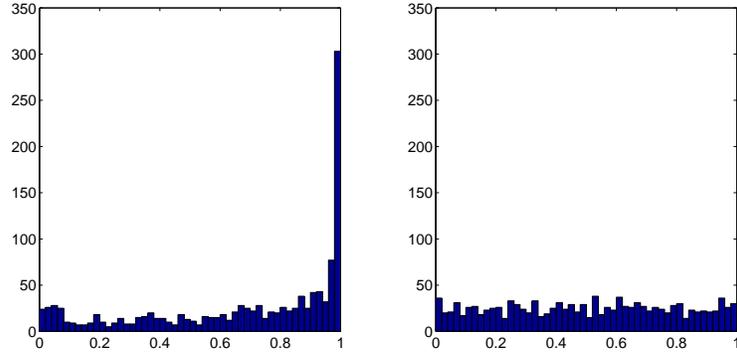
Figure 2: Histogram of $\{P_m\}$ (left) v.s. $\{P_m^*\}$ (right) from the Variance Aggregation Estimator
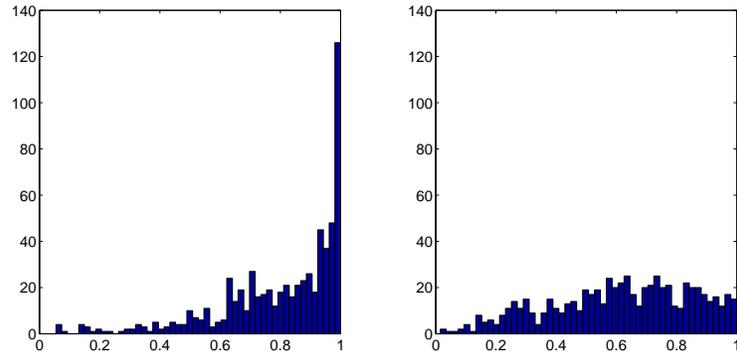

Figure 3: Histogram of $\{P_m\}$ (left) v.s. $\{P_m^*\}$ (right) from the Local Periodogram Regression
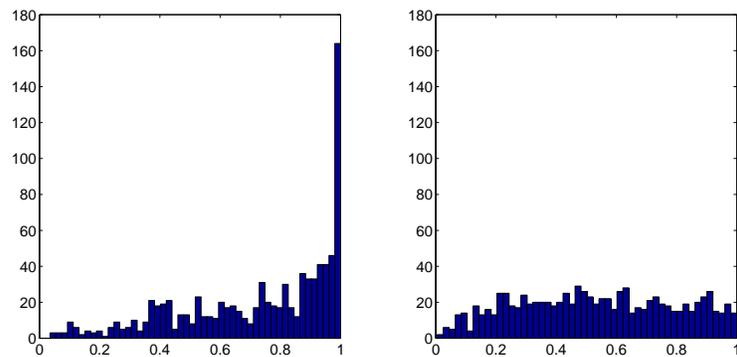

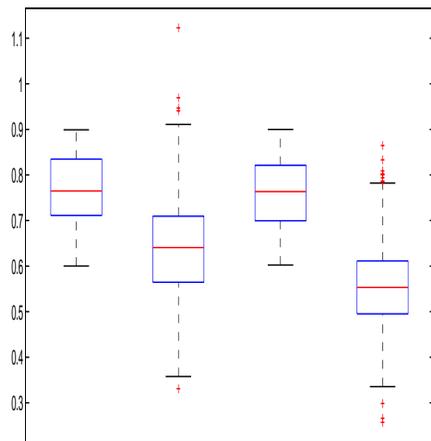Figure 4: Histogram of $\{P_m\}$ (left) v.s. $\{P_m^*\}$ (right) from the Local Whittle Estimator
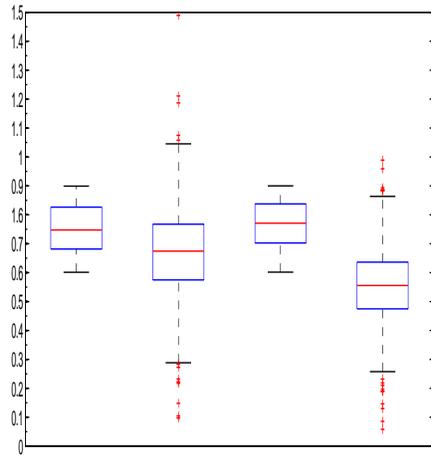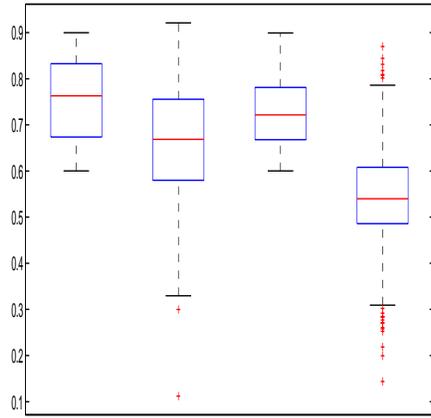
Figure 5: Top to bottom: variance aggregation estimator, local periodogram regression and local Whittle estimator. In each boxplot, from left to right: $\widehat{H}_m^{(1)}$, $\widehat{H}_m^{(2)}$, $\widehat{H}_m^{(1)*}$ and $\widehat{H}_m^{(2)*}$.

(3) Fast decaying physical dependence measure of Wu [81] in Bernoulli shift models.

The first is by far the most widely-used notion for weak dependence which applies to very general stationary processes. The second is mentioned due to its close connection to the considerations in Section 2.1. The third is a convenient criterion under the Bernoulli shift framework which covers a wide range of concrete statistical models.

5.1 Strong mixing conditions

Suppose that $\{Y(n)\}$ is a stationary process with $\mathbb{E}[Y(n)] = 0$ and $\mathrm{Var}[Y(n)] = 1$. Define the σ-field $\mathcal{F}_a^b = \sigma\{Y(n) : a \le n \le b\}$, where $-\infty \le a \le b \le +\infty$. Given two σ-fields $\mathcal{A}, \mathcal{B}$, one can define the following measure of dependence

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup\{|P(A \cap B) - P(A)P(B)| : \ A \in \mathcal{A}, B \in \mathcal{B}\}. \tag{39}$$

Then the α-mixing coefficient of $\{X(n)\}$, first introduced in Rosenblatt [67], is defined as

$$\alpha_Y(n) = \alpha\left(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty\right).$$

When $\alpha_Y(n) \to 0$ as $n \to \infty$, we say that $\{Y(n)\}$ is *strong mixing*. If one assumes that $\alpha_Y(n)$ decays to zero fast enough together with some other regularity conditions, then a central limit theorem for $X(n)$ can be established. We state, as an example, the following central limit theorem due to Ibragimov [49] and Herrndorf [43].

Theorem 5.1. *If $\mathbb{E}|Y(n)|^{2+\delta} < \infty$ for some $\delta > 0$ and*

$$\sum_{n=1}^\infty \alpha_Y(n)^{\delta/(2+\delta)} < \infty, \tag{40}$$

then

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{[Nt]} \left(Y(n) - \mathbb{E}Y(n)\right) \Rightarrow \sigma B(t).$$

where $B(t)$ is a standard Brownian motion, $\Rightarrow$ stands for weak convergence in $D[0, 1]$, and

$$\sigma^2 = \sum_{n=-\infty}^\infty \mathrm{Cov}[Y(n), Y(0)].$$

Now consider the transformation

$$X(n) = F(Y(n), \ldots, Y(n - l)).$$

Let us compare α_X and α_Y. Since $X(n) \in \mathcal{F}_{n-l}^n$, it is easily deduced that for $n > l$, the α-mixing coefficient of $\{X(n)\}$ satisfies

$$\alpha_X(n) \le \alpha_Y(n - l). \tag{41}$$

The relation (41) means that the dependence measured by the α-mixing coefficient after the perturbing tranform $F(\cdot)$ cannot exceed that of the original process $Y(n)$ (up to a fixed lag l). In particular, relation (40) holds for $\alpha_X(n)$. One then only needs $\mathbb{E}|X(n)|^{2+\delta} < \infty$ (which is the case if $F(\cdot)$ has at most linear growth) for Theorem 5.1 to hold.

There are different mixing coefficients than $\alpha(n)$, obtained by modifying the measure of dependence between the σ-fields in (39), for example, the ϕ-mixing coefficient defined through

$$\phi(\mathcal{A}, \mathcal{B}) = \sup\{|P(A|B) - P(A)| : \ A \in \mathcal{A}, B \in \mathcal{B}, P(B) > 0\},$$

the ρ-mixing coefficient defined through

$$\rho(\mathcal{A}, \mathcal{B}) = \sup\{\mathrm{Corr}(X, Y) : \ X \in L^2(\mathcal{A}), Y \in L^2(\mathcal{B})\},$$

and so on. In general, as long as a dependence measure $m(\cdot, \cdot)$ is non-increasing with respect to set inclusion and the mixing coefficient is defined as $m(n) = m(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty)$, then a relation as (41) always holds.

Hence, the central limit theorems under strong mixing conditions is robust against a transformation perturbation.

5.2 Gaussian subordination

Let $\{Y(n)\}$ be a stationary *Gaussian* process, and let

$$X(n) = F(Y(n), \ldots, Y(n-l)).$$

When the covariance function of $Y(n)$ decays fast enough, a central limit theorem always holds for $X(n)$. In particular, we have the following result which is a consequence of Ho and Sun [47].

Theorem 5.2. *Suppose that* $\mathbb{E}X(n)^2 < \infty$ *and*

$$\sum_{n=-\infty}^{\infty} |\mathrm{Cov}\,[Y(n), Y(0)]| < \infty. \tag{42}$$

Then one has

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{[Nt]} \Big(X(n) - \mathbb{E}X(n) \Big) \xrightarrow{f.d.d.} \sigma B(t),$$

where $B(t)$ *is a standard Brownian motion and*

$$\sigma^2 = \sum_{n=-\infty}^{\infty} \mathrm{Cov}[X(n), X(0)].$$

Theorem 5.2 directly expresses the robustness of the central limit theorem against transformation perturbation when the short memory condition (42) is imposed on $Y(n)$.

5.3 Bernoulli shift

Let $\{\epsilon_i\}$ be an i.i.d. sequence of random variables with mean 0 and variance 1. Consider the Bernoulli shift model

$$Y(n) = G_Y(\epsilon_n, \epsilon_{n-1}, \ldots), \tag{43}$$

where G_Y is a non-random measurable function. This specification covers not only the causal linear process (17), but also many nonlinear time series models obtained as solutions of difference equations involving ϵ_i.

Wu [81] introduced the following so-called *physical dependence measure* for a process $\{Y(n)\}$ specified by (43). Let ϵ_0^* be a random variable independent of $\{\epsilon_i\}$ and having the same distribution as ϵ_0. Define

$$\delta_2^X(n) = \|G_Y(\epsilon_n, \ldots, \epsilon_1, \epsilon_0, \epsilon_{-1}, \ldots) - G_Y(\epsilon_n, \ldots, \epsilon_1, \epsilon_0^*, \epsilon_{-1}, \ldots)\|_{L^2(\Omega)}. \tag{44}$$

If (43) is interpreted as a nonlinear system with input $\{\epsilon_n\}$ and output $\{Y(n)\}$, then $\delta_2^Y(n)$ in (44) measures the influence of the lag-n input ϵ_0 on the current output $Y(n)$.

With $\delta_2^Y(n)$, one can state the following central limit theorem, which is a consequence of Theorem 1 and 3 of Wu [81].

Theorem 5.3. *Suppose that*

$$\sum_{n=1}^{\infty} \delta_2^X(n) < \infty. \tag{45}$$

Then one has

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{[Nt]} \Big(X(n) - \mathbb{E}X(n) \Big) \xrightarrow{f.d.d.} \sigma B(t), \quad t \geq 0,$$

where $B(t)$ *is a standard Brownian motion, and*

$$\sigma^2 = \sum_{n=-\infty}^{\infty} \mathrm{Cov}[X(n), X(0)].$$

Remark 5.4. The criterion (45) is typically easier to check for a specific Bernoulli shift model than the criteria based on strong mixing conditions (see Theorem 5.1), while still providing numerous statistical applications.

Now we consider the transformation perturbation. Let

$$X(n) = F(Y(n), \ldots, Y(n - l + 1)) =: G_X(\epsilon_n, \epsilon_{n-1}, \ldots).$$

We need to assume some smoothness condition (compare with the arguments of Claim 2.11) on the perturbation function $F(x_1, \ldots, x_l)$. In particular, suppose that $F(\cdot)$ is Lipschitz, that is,

$$|F(x_1, \ldots, x_l) - F(y_1, \ldots, y_l)| \leq C_F \sum_{i=1}^{l} |x_i - y_i| \tag{46}$$

for some constant $C_F \geq 0$.

Setting $\boldsymbol{\epsilon}_n = (\epsilon_n, \ldots, \epsilon_1, \epsilon_0, \epsilon_{-1}, \ldots)$ and $\boldsymbol{\epsilon}_n^* = (\epsilon_n, \ldots, \epsilon_1, \epsilon_0^*, \epsilon_{-1}, \ldots)$, one has by (46) that

$$|G_X(\boldsymbol{\epsilon}_n) - G_X(\boldsymbol{\epsilon}_n^*)| \leq C_F \sum_{i=0}^{l-1} |G_Y(\boldsymbol{\epsilon}_{n-i}) - G_Y(\boldsymbol{\epsilon}_{n-i}^*)|.$$

Therefore, if $\delta_2^X(n)$ and $\delta_2^Y(n)$ are the physical dependence measures of $\{X(n)\}$ and $\{Y(n)\}$ respectively, then

$$\delta_2^X(n) = \|G_X(\boldsymbol{\epsilon}_n) - G_X(\boldsymbol{\epsilon}_n^*)\|_{L^2(\Omega)} \leq C_F \sum_{i=0}^{l-1} \delta_2^Y(n-i).$$

Hence if $\{Y(n)\}$ satisfies the short memory condition

$$\sum_{n=1}^{\infty} \delta_2^Y(n) < \infty,$$

then so does $\{X(n)\}$. This shows the robustness of Theorem 5.3 against a perturbation by any Lipschitz transformation.

Remark 5.5. The proof of Theorem 5.3 is based on a martingale difference approximation method and resorts to the martingale difference central limit theorem. We note, however, that the martingale difference central limit theorem is itself not robust against transformation, since the martingale difference structure in general can be easily disturbed by a transformation. For example, in the stochastic volatility-type models, e.g., the LARCH(∞) model (Giraitis et al. [38]), the return sequence $X(n)$ is a martingale difference, while $|X(n)|$ can exhibit long memory (see Beran et al. [11], Chapter 4.2.8.).

Remark 5.6. Using similar arguments, one can show that the θ-weak dependence criterion (whose definition involves bounded Lipschitz transformation) introduced by Doukhan and Louhichi [33], enjoys a robustness against bounded Lipschitz transformations.

6 Conclusion and suggestions

In this paper, we discussed the instability issue of Hermite rank and other related ranks appearing in limit theorems under long memory. We argued that a rank greater than 1 can be disturbed by a transformation and only a rank equal to 1 is stable. We provided empirical evidence supporting this argument. Such an instability feature has important statistical implications. In particular, assuming a higher-order rank when it is really not there may result in underestimating the order of fluctuation of the statistic of interest.

To address this issue we briefly indicate here some suggestions for performing valid inference. As illustrated, particularly in Section 3, one may adopt the assumption that the rank is always 1, regardless of any nonlinear transformation resulting from the statistical procedure. Here the rank should be understood in

a generalized sense, taking into account situations as (36). Some studies have implicitly done so, although without giving an explanation (see, e.g., Beran [8] and Shao [72]). Recently Beran et al. [12] designed a statistical test based on resampling to distinguish Hermite rank 1 and a higher-order Hermite in the model (14).

Another appealing way out, is to redesign the statistical procedure in a way as to avoid using the fixed-rank limit theorems for inference directly. This may be achieved by combining re-sampling method (see, e.g, Hall et al. [42], Nordman and Lahiri [60], Zhang et al. [85]), Bai and Taqqu [5]), together with suitable self-normalization technique (see, e.g., Shao [71] and Shao [72]). We refer the reader to Jach et al. [50] Betken and Wendler [13] and Bai et al. [7] for approaches of this type.

A   Non-instantaneous transformation of the Gaussian

Let $\{Y(n)\}$ be a standardized stationary long-memory Gaussian process with Hurst index H. We extend here the discussion on instantaneous transformation (14) to the *non-instantaneous* transformation

$$X(n) = F\big(Y(n), Y(n-1), \ldots, Y(n-l)\big), \tag{47}$$

where $X(n) \in L^2(\Omega)$ and l is a finite positive integer. Since the non-instantaneous case is much less treated in the literature, we shall introduce in this section the relevant results in Dobrushin and Major [31], and show that the arguments developed in Section 2.1 continue to be valid.

It is well-known that the Gaussian $Y(n)$ admits the spectral representation (see, e.g., Dobrushin and Major [31])

$$Y(n) = \int_{(-\pi,\pi]} e^{inx} W_Y(dx), \tag{48}$$

where $W_Y(dx)$ is a complex-valued Gaussian measure satisfying

$$\mathbb{E}|W_Y(dx)|^2 = F_Y(dx) \tag{49}$$

and $F_Y(\cdot)$ is the spectral distribution[5] of $Y(n)$. Then $X(0)$ has following Wiener-Itô expansion (see Dobrushin and Major [31], formula (6.1), or Janson [51], Theorem 7.61):

$$X(0) - \mathbb{E}X(0) = \sum_{m=1}^{\infty} \int_{(-\pi,\pi]^m}'' \alpha_m(x_1, \ldots, x_m) W_Y(dx_1) \ldots W_Y(dx_m), \tag{50}$$

where the double prime $''$ indicates the exclusion of the hyper-diagonals $x_p = \pm x_q$ in the multiple stochastic integral. Here $\alpha_m(\cdot)$'s are a.e. unique complex-valued functions in satisfying

$$\alpha_m(x_1, \ldots, x_k) = \overline{\alpha_m(-x_1, \ldots, -x_m)},$$

and

$$\sum_{m=1}^{\infty} m! \|\alpha_m\|_{L^2((-\pi,\pi]^m, F_Y^{\otimes m})}^2 < \infty,$$

where

$$\|\alpha_m\|_{L^2((-\pi,\pi]^m, F_Y^{\otimes m})^2} = \int_{(-\pi,\pi]^m} |\alpha_m(x_1, \ldots, x_m)|^2 F_Y(dx_1) \ldots F_Y(dx_m).$$

The *Hermite rank* of $X(n)$ (or say the Hermite rank of $F(\cdot)$ with respect to $\{Y(n)\}$) is defined as

$$\inf\left\{ m \geq 1 : \|\alpha_m\|_{L^2((-\pi,\pi]^m, F_Y^{\otimes m})} \neq 0 \right\}, \tag{51}$$

The Hermite rank in (51) is also equal to (see Dobrushin and Major [31] Remark 6.3)

$$\inf\left\{ m \geq 1 : \mathbb{E}\Big[(X(0) - \mathbb{E}X(0))Y(n)^m\Big] \neq 0 \text{ for some } n \in \mathbb{Z} \right\}. \tag{52}$$

---

[5]Do not confuse F_Y in (49) with F in (47).

This should be compared to (7).

By Remark 6.1 of Dobrushin and Major [31], the a.e. unique function $\alpha_m(\cdot)$ can further be chosen to be continuous, which we shall assume throughout below. We are now ready to state the following generalization of Theorem 2.6, which follows from Dobrushin and Major [31] Theorem 3, Remark 6.3 and Remark 6.4.

Theorem A.1. *Suppose that $X(n) = F(Y(n), \ldots, Y(n-l))$, and that the Hermite rank in the sense of (51) is k, and that the Hurst index H of $\{Y(n)\}$ satisfies*

$$H > 1 - \frac{1}{2k}.$$

Suppose also that $\alpha_k(\cdot)$ in (50) satisfies

$$\alpha_k(0, \ldots, 0) \neq 0. \tag{53}$$

Then $\{X(n)\}$ has long memory with Hurst index:

$$H_F = (H-1)k + 1 \in \left(\frac{1}{2}, 1\right).$$

Furthermore, as $N \to \infty$, we have

$$\frac{1}{N^{H_F}} \sum_{n=1}^{[Nt]} \Big(X(n) - \mathbb{E}X(n) \Big) \Rightarrow c\alpha_k(0, \ldots, 0) Z_{H_F, k}(t), \tag{54}$$

for some $c \neq 0$, where $Z_{H_F, k}(t)$ is the Hermite process in (13).

Remark A.2. In contrast to Theorem 2.6 where the constant c in (12) is always nonzero, in the non-instantaneous case we need to assume in addition the condition (53). If $\alpha_k(0, \ldots, 0) = 0$, then (54) tells nothing more than that the normalization N^{-H_F} is too strong. In this case, terms with order greater than k may contribute to the asymptotic distribution as well. For example, if in (47) we let

$$X(n) = H_1(Y(n)) - H_1(Y(n-1)) + H_2(Y(n)) = Y(n) - Y(n-1) + Y(n)^2 - 1.$$

Using the spectral representation (48) and Major [57] Theorem 4.3, we have

$$X(0) = \int_{(-\pi, \pi]} (1 - e^{-ix}) W_Y(dx) + \int_{(-\pi, \pi]^2}'' W_Y(dx_1) W_Y(dx_2)$$

so that $\alpha_1(x) = 1 - e^{-ix}$ and $\alpha_1(0) = 0$. On the other hand, the Hermite rank of $X(n)$ is $k = 1$ in view of (50). Now

$$\sum_{n=1}^{N} X(n) = Y(N) - Y(0) + \sum_{n=1}^{N} H_2(Y(n)).$$

Since $Y(n)$ is stationary, $N^{-H}[Y(N) - Y(0)] \xrightarrow{p} 0$, and thus only the term $\sum_{n=1}^{N} H_2(Y(n))$ contributes to in the limit. Hence the limit of suitably normalized $\sum_{n=1}^{N} X(n)$ can be either a Brownian motion if $H \leq 3/4$ or a Hermite process of order 2 if $H > 3/4$ in view of Theorem 2.6.

Remark A.3. Now arguing as in Section 2.1, one notes that a Hermite rank higher than 1 in this non-instantaneous context is also unstable. Recall that the role of $F(\cdot)$ in (47), as in Section 2.1, is to account for an uncontrollable perturbation of the Gaussian model. Suppose that $G(\cdot)$ is a function determined by the statistical procedure of interest. Then one can formulate a statement parallel to Claim 2.11. So the part of Theorem A.1 which is most likely of statistical relevance is just the case $k = 1$, where the limit is fractional Brownian motion and the normalization is N^{-H}. Note that this non-instantaneous consideration includes not only $G(X(n))$ with $X(n)$ defined in (47), but also the case where $G(\cdot)$ is a finite-dimensional multivariate function of the observed time series $\{X(n)\}$, for example $G(X(n), \ldots, X(n-p)) = X(n)X(n-p)$, a term which appears in the sample covariance.

Arguing as in Claim 2.11, condition (53) should be expected to typically hold in practice.

Remark A.4. Using the full generality of Theorem 3 of Dobrushin and Major [31], it is even possible to consider the case $l = \infty$ in (47), namely, including dependence on the infinite past. In this case, however, one encounters major technical difficulties since $F(\cdot)$ with $l = \infty$ may alter the long memory property of $Y(n)$, for example, if $F(\cdot)$ is a linear filter with a slow power-law decay (see, e.g., Section 2.5 below). On the other hand, one may be satisfied with the restriction to $l < \infty$ since $F(\cdot)$ has been introduced only to account for a small perturbation of the Gaussian model, in which case the argument of $F(\cdot)$ is not expected to stretch to the infinite past.

Remark A.5. This discussion can also be extended to the case where $\mathbf{Y}(n)$ is a vector-valued Gaussian stationary noise and $\mathbf{X}(n)$ is also vector-valued. See, e.g., Denaranjo [30] and Arcones [1].

Remark A.6. We mention that the extension of Theorem 2.15 to non-instantaneous transformation of linear processes, that is, an analog of Theorem A.1 when $X(n)$ is linear, is still open. Only central limit theorems involving non-instantaneous filter of linear processes have been considered (see Wu [80] and Cheng and Ho [19]).

References

[1] M.A. Arcones. Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *The Annals of Probability*, pages 2242–2274, 1994.

[2] F. Avram. On bilinear forms in Gaussian random variables and Toeplitz matrices. *Probability Theory and Related Fields*, 79(1):37–45, 1988.

[3] F. Avram and M.S. Taqqu. Noncentral limit theorems and Appell polynomials. *The Annals of Probability*, 15(2):767–775, 1987.

[4] S. Bai and M.S. Taqqu. Multivariate limit theorems in the context of long-range dependence. *Journal of Time Series Analysis*, 34(6):717–743, 2013.

[5] S. Bai and M.S. Taqqu. On the validity of resampling methods under long memory. *To appear in The Annals of Statistics*, 2016.

[6] S. Bai and M.S. Taqqu. Some properties of the Hermite rank. *Preprint, see ArXiv http://arxiv.org/abs/1710.01612*, 2017.

[7] S. Bai, M.S. Taqqu, and T. Zhang. A unified approach to self-normalized block sampling. *Stochastic Processes and their Applications*, 126(8):2465–2493, 2016.

[8] J. Beran. M estimators of location for Gaussian and related processes with slowly decaying serial correlations. *Journal of the American Statistical Association*, 86(415):704–708, 1991.

[9] J. Beran and S. Ghosh. Slowly decaying correlations, testing normality, nuisance parameters. *Journal of the American Statistical Association*, 86(415):785–791, 1991.

[10] J. Beran and A. Weiershäuser. On spline regression under Gaussian subordination with long memory. *Journal of Multivariate Analysis*, 102(2):315–335, 2011.

[11] J. Beran, Y. Feng, S. Ghosh, and R. Kulik. *Long-Memory Processes: Probabilistic Properties and Statistical Methods.* Springer, 2013.

[12] J. Beran, S. Möhrle, and S. Ghosh. Testing for Hermite rank in Gaussian subordination processes. *Journal of Computational and Graphical Statistics*, 25(3):917–934, 2016.

[13] A. Betken and M. Wendler. Subsampling for general statistics under long range dependence. *arXiv preprint arXiv:1509.05720*, 2015.

[14] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation*. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 1989.

[15] P. Breuer and P. Major. Central limit theorems for non-linear functionals of Gaussian fields. *Journal of Multivariate Analysis*, 13(3):425–441, 1983.

[16] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer, 1991.

[17] D. Chambers and E. Slud. Central limit theorems for nonlinear functionals of stationary Gaussian processes. *Probability Theory and Related Fields*, 80(3):323–346, 1989.

[18] B. Cheng and P.M. Robinson. Density estimation in strongly dependent non-linear time series. *Statistica Sinica*, 1(2):335–359, 1991.

[19] T. Cheng and H. Ho. On Berry–Esseen bounds for non-instantaneous filters of linear processes. *Bernoulli*, 14(2):301–321, 2008.

[20] M. Clausel, F. Roueff, M.S. Taqqu, and C. Tudor. Large scale behavior of wavelet coefficients of non-linear subordinated processes with long memory. *Applied and Computational Harmonic Analysis*, 32 (2):223–241, 2012.

[21] M. Clausel, F. Roueff, M.S. Taqqu, and C. Tudor. Wavelet estimation of the long memory parameter for hermite polynomial of Gaussian processes. *ESAIM: Probability and Statistics*, 18:42–76, 2014.

[22] S. Csörgö. The smoothing dichotomy in nonparametric regression under long-memory errors. *Statistica neerlandica*, 56(2):132–142, 2002.

[23] S. Csörgö and J. Mielniczuk. Density estimation under long-range dependence. *The Annals of Statistics*, pages 990–999, 1995.

[24] S. Csörgö and J. Mielniczuk. Random-design regression under long-range dependent errors. *Bernoulli*, pages 209–224, 1999.

[25] H. Dehling and W. Philipp. Empirical process techniques for dependent data. In *Empirical process techniques for dependent data*, pages 3–113. Springer, 2002.

[26] H. Dehling and M.S. Taqqu. The empirical process of some long-range dependent sequences with an application to U-statistics. *The Annals of Statistics*, pages 1767–1783, 1989.

[27] H. Dehling and M.S. Taqqu. Bivariate symmetric statistics of long-range dependent observations. *Journal of Statistical Planning and Inference*, 28(2):153–165, 1991.

[28] H. Dehling, T. Mikosch, and M. (editors) Sorensen. *Empirical process techniques for dependent data*. Springer, 2002.

[29] H. Dehling, A. Rooch, and M.S. Taqqu. Non-parametric change-point tests for long-range dependent data. *Scandinavian Journal of Statistics*, 40(1):153–173, 2013.

[30] M.V.S. Denaranjo. Non-central limit theorems for non-linear functionals of k Gaussian fields. *Journal of multivariate analysis*, 44(2):227–255, 1993.

[31] R.L. Dobrushin and P. Major. Non-central limit theorems for non-linear functional of Gaussian fields. *Probability Theory and Related Fields*, 50(1):27–52, 1979.

[32] P. Doukhan. Models, inequalities, and limit theorems for stationary sequences. In *Theory and Applications of Long-Range Dependence*, pages 43–100. Birkhäuser, 2003.

[33] P. Doukhan and S. Louhichi. A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, 84(2):313–342, 1999.

[34] G. Faÿ, E. Moulines, F. Roueff, and M.S. Taqqu. Estimators of long-memory: Fourier versus wavelets. *Journal of econometrics*, 151(2):159–177, 2009.

[35] R. Fox and M.S. Taqqu. Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, pages 517–532, 1986.

[36] L. Giraitis and D. Surgailis. A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probability Theory and Related Fields*, 86(1):87–104, 1990.

[37] L. Giraitis and M.S. Taqqu. Whittle estimator for finite-variance non-Gaussian time series with long memory. *Annals of Statistics*, pages 178–203, 1999.

[38] L. Giraitis, R. Leipus, P.M. Robinson, and D. Surgailis. LARCH, leverage, and long memory. *Journal of Financial Econometrics*, 2(2):177–210, 2004.

[39] L. Giraitis, H.L. Koul, and D. Surgailis. *Large Sample Inference for Long Memory Processes*. World Scientific Publishing Company Incorporated, 2012.

[40] C.W.J. Granger and R. Joyeux. An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis*, 1(1):15–29, 1980.

[41] H. Guo and H.L. Koul. Nonparametric regression with heteroscedastic long memory errors. *Journal of Statistical Planning and Inference*, 137(2):379–404, 2007.

[42] P. Hall, B-Y Jing, and S.N. Lahiri. On the sampling window method for long-range dependent data. *Statistica Sinica*, 8(4):1189–1204, 1998.

[43] N. Herrndorf. A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability*, pages 141–153, 1984.

[44] J. Hidalgo. Non-parametric estimation with strongly dependent multivariate time series. *Journal of Time Series Analysis*, 18(2):95–122, 1997.

[45] H. Ho. On central and non-central limit theorems in density estimation for sequences of long-range dependence. *Stochastic processes and their Applications*, 63(2):153–174, 1996.

[46] H. Ho and T. Hsing. Limit theorems for functionals of moving averages. *The Annals of Probability*, 25 (4):1636–1669, 1997.

[47] H. Ho and T. Sun. A central limit theorem for non-instantaneous filters of a stationary Gaussian process. *Journal of Multivariate Analysis*, 22(1):144–155, 1987.

[48] J.R.M. Hosking. Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *Journal of Econometrics*, 73(1):261–284, 1996.

[49] I.A. Ibragimov. Some limit theorems for stationary processes. *Theory of Probability & Its Applications*, 7(4):349–382, 1962.

[50] A. Jach, T. McElroy, and D.N. Politis. Subsampling inference for the mean of heavy-tailed long-memory time series. *Journal of Time Series Analysis*, 33(1):96–111, 2012.

[51] S. Janson. *Gaussian Hilbert Spaces*. Cambridge Tracts in Mathematics. Cambridge University Press, 1997.

[52] H.L. Koul and W. Stute. Regression model fitting with long memory errors. *Journal of statistical planning and inference*, 71(1):35–56, 1998.

[53] H.R. Künsch. Statistical aspects of self-similar processes. In *Proceedings of the first World Congress of the Bernoulli Society*, volume 1, pages 67–74. VNU Science Press Utrecht, 1987.

[54] C. Lévy-Leduc and M.S. Taqqu. Long-range dependence and the rank of decompositions. *Fractal Geometry and Dynamical Systems in Pure and Applied Mathematics II: Fractals in Applied Mathematics*, 601:289, 2013.

[55] C. Lévy-Leduc, H. Boistard, E. Moulines, M.S. Taqqu, and V.A. Reisen. Asymptotic properties of U-processes under long-range dependence. *The Annals of Statistics*, 39(3):1399–1426, 2011.

[56] C Lévy-Leduc, H. Boistard, E. Moulines, and Reisen V.A. Taqqu, M.S. Robust estimation of the scale and of the autocovariance function of Gaussian short and long-range dependent processes. *Journal of Time Series Analysis*, 32:135–156, 2011.

[57] P. Major. *Multiple Wiener-Itô Integrals: With Applications to Limit Theorems.* Lecture Notes in Mathematics. Springer, 2nd edition, 2014.

[58] B.B. Mandelbrot and J.R. Wallis. Some long-run properties of geophysical records. *Water resources research*, 5(2):321–340, 1969.

[59] E. Masry and J. Mielniczuk. Local linear regression estimation for time series with long-range dependence. *Stochastic Processes and their Applications*, 82(2):173–193, 1999.

[60] D.J. Nordman and S.N. Lahiri. Validity of the sampling window method for long-range dependent linear processes. *Econometric Theory*, 21(06):1087–1111, 2005.

[61] J.D. Pelletier and D.L. Turcotte. Long-range persistence in climatological and hydrological time series: analysis, modeling and application to drought hazard assessment. *Journal of Hydrology*, 203(1):198–208, 1997.

[62] P.C.B. Phillips. Towards a unified asymptotic theory for autoregression. *Biometrika*, pages 535–547, 1987.

[63] V. Pipiras and M.S. Taqqu. Regularization and integral representations of Hermite processes. *Statistics and Probability Letters*, 80(23):2014–2023, 2010.

[64] V. Pipiras and M.S. Taqqu. *Long-Range Dependence and Self-Similarity.* Cambridge University Press, 2017.

[65] Z. Psaradakis. On inference based on the one-sample sign statistic for long-range dependent data. *Computational Statistics*, 25(2):329–340, 2010.

[66] P.M. Robinson. Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, pages 1630–1661, 1995.

[67] M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1):43, 1956.

[68] M. Rosenblatt. Independence and dependence. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab*, volume 2, pages 431–443, 1961.

[69] G. Samorodnitsky. *Stochastic Processes and Long Range Dependence.* Springer, 2016.

[70] F.H. Schweingruber. *Tree rings and environment: dendroecology.* Paul Haupt AG Bern, 1996.

[71] X. Shao. A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):343–366, 2010.

[72] X. Shao. A simple test of changes in mean in the possible presence of long-range dependence. *Journal of Time Series Analysis*, 32(6):598–606, 2011.

[73] D. Surgailis. Zones of attraction of self-similar multiple integrals. *Lithuanian Mathematical Journal*, 22 (3):327–340, 1982.

[74] D. Surgailis. Long-range dependence and Appell rank. *Annals of Probability*, pages 478–497, 2000.

[75] M.S. Taqqu. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Probability Theory and Related Fields*, 31(4):287–302, 1975.

[76] M.S. Taqqu. Convergence of integrated processes of arbitrary Hermite rank. *Probability Theory and Related Fields*, 50(1):53–83, 1979.

[77] M.S. Taqqu and V. Teverovsky. Robustness of Whittle-type estimators for time series with long-range dependence. *Communications in Statistics. Stochastic Models*, 13(4):723–757, 1997.

[78] M.S. Taqqu, V. Teverovsky, and W. Willinger. Estimators for long-range dependence: an empirical study. *Fractals*, 3(04):785–798, 1995.

[79] N. Terrin and M.S. Taqqu. A noncentral limit theorem for quadratic forms of Gaussian stationary sequences. *Journal of Theoretical Probability*, 3(3):449–475, 1990.

[80] W.B. Wu. Central limit theorems for functionals of linear processes and their applications. *Statistica Sinica*, 12(2):635–650, 2002.

[81] W.B. Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(40):14150–14154, 2005.

[82] W.B. Wu. Unit root testing for functionals of linear processes. *Econometric Theory*, 22(01):1–14, 2006.

[83] W.B. Wu and J. Mielniczuk. Kernel density estimation for linear processes. *Annals of Statistics*, pages 1441–1459, 2002.

[84] W.B. Wu, Y. Huang, and W. Zheng. Covariances estimation for long-memory processes. *Advances in Applied Probability*, 42(1):137–157, 2010.

[85] T. Zhang, H-C Ho, M. Wendler, and W.B. Wu. Block sampling under strong dependence. *Stochastic Processes and their Applications*, 123(6):2323–2339, 2013.

[86] W. Zhao, Z. Tian, and Z. Xia. Ratio test for variance change point in linear process with long memory. *Statistical Papers*, 51(2):397–407, 2010.

Shuyang Bai
Department of Statistics
University of Georgia
Athens, GA 30606
bsy9142@uga.edu

Murad S. Taqqu
Department of Mathematics and Statistics
Boston University
Boston, MA 02215
murad@bu.edu