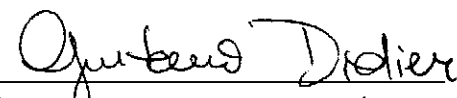


ASYMPTOTIC THEORY FOR THE STATISTICAL ANALYSIS OF
ANOMALOUS DIFFUSION IN SINGLE PARTICLE TRACKING
EXPERIMENTS


AN ABSTRACT

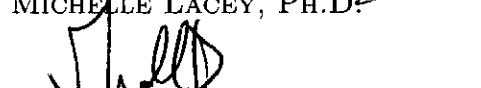
SUBMITTED ON THE FOURTH DAY OF APRIL, 2017
TO THE DEPARTMENT OF MATHEMATICS
OF THE SCHOOL OF SCIENCE AND ENGINEERING OF
TULANE UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
BY

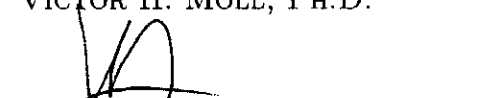

KUI ZHANG

APPROVED: 
GUSTAVO DIDIER, PH.D. (ADVI-
SOR)


MAC HYMAN, PH.D.


MICHELLE LACEY, PH.D.


VICTOR H. MOLL, PH.D.


VINCENT MARTINEZ, PH.D.

Abstract

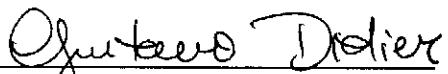
Recent advances in light microscopy have spawned new research frontiers in nanobiophysics by working around the diffraction barrier and allowing for the observation of nanometric biological structures. Microrheology is the study of the properties of complex fluids, such as those found in biology, through the dynamics of small embedded particles, typically latex beads. Statistics based on the recorded sample paths are used by biophysicists to infer rheological properties of the fluid.

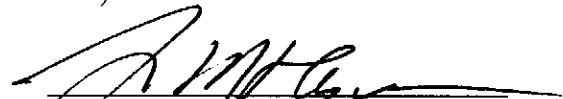
In this dissertation, we provide new probabilistic results and statistical methods for the analysis of single particle tracking experiments in nanobiophysics. In Chapter 2, assuming that the stochastic behavior of tracer particles falls within a broad class of fractional, stationary increment Gaussian processes, we establish the asymptotic distribution of the mean squared displacement ($\widehat{\text{MSD}}$) of particles. This had remained an open problem for years in the nanobiophysical literature. In Chapter 3, by drawing upon the framework laid out in Chapter 2, we construct a general protocol for the statistical testing of fluid heterogeneity and apply it to real data on biofilm eradication. In Chapter 4, assuming that the observed stochastic process is a fractional Gaussian noise, we establish the asymptotic distribution of the statistic proposed in [50] for verifying the mixing property of an underlying dynamical system starting from a single particle path.

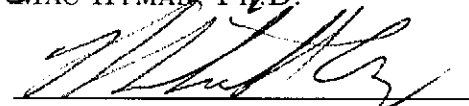
ASYMPTOTIC THEORY FOR THE STATISTICAL ANALYSIS OF
ANOMALOUS DIFFUSION IN SINGLE PARTICLE TRACKING
EXPERIMENTS


A DISSERTATION
SUBMITTED ON THE FOURTH DAY OF APRIL, 2017
TO THE DEPARTMENT OF MATHEMATICS
OF THE SCHOOL OF SCIENCE AND ENGINEERING OF
TULANE UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
BY

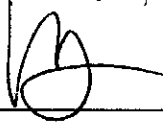

KUI ZHANG

APPROVED: 
GUSTAVO DIDIER, PH.D. (ADVI-
SOR)


MAC HYMAN, PH.D.


MICHELLE LACEY, PH.D.


VICTOR H. MOLL, PH.D.


VINCENT MARTINEZ, PH.D.

© Copyright by Kui Zhang 2017.

All rights reserved.

Acknowledgement

My deepest gratitude goes first and foremost to Professor Gustavo Didier, my supervisor, for his constant encouragement and guidance. He has walked me through all the stages of the writing of this thesis. Without his consistent and illuminating instruction, this thesis could not have reached its present form.

I am also greatly indebted to the professors at the Department of Mathematics: Dr. Lisa J. Fauci, Dr. Robert Herbert, Dr. Mac Hyman, Dr. Alexander Kurganov, Dr. Michelle Lacey, and Dr. Victor H. Moll, who have instructed and helped me a lot in the past five years. In addition, I would like to thank Dr. Vincent Martinez for serving on my dissertation committee.

Last my thanks would go to my beloved family for their loving considerations and great confidence in me all through these years. I also owe my sincere gratitude to my friends and my fellow classmates who gave me their help and time in listening to me and helping me work out my problems.

List of Tables

2.1	Mean and s.d. of \hat{H} for fBm: consecutive lags $(2, \dots, 2^8)$ vs two lags $(h_1 = 2, h_2 = 2^7)$	22
2.2	Mean and s.d. of \hat{H} for ifOU: small lags vs large lags, path length= 2^{12}	22
3.1	Hypotheses	32
3.2	Test statistics	32
3.3	Tests	33
3.4	Intrafluid biofilm heterogeneity testing after treatment with tobramycin at concentration levels 25, 50, 100, 200, 400 $\mu\text{g ml}^{-1}$. 100 independent paths of length 1800 are randomly selected for each concentration level.	45
3.5	Intrafluid biofilm heterogeneity testing after treatment with COS2-NO at concentration levels 1, 2, 4, 8, 16 mg ml^{-1} . 100 independent paths of length 1800 are randomly selected for each concentration level. . .	45
3.6	Interfluid biofilm heterogeneity testing after treatment with COS2-NO at concentration level 16 mg ml^{-1} . 100 independent paths of length 1800 are randomly selected for each group.	45

List of Figures

2.1	$\widehat{\text{MSD}}$ plots, each with 50 paths of size 2^{12} from an ifOU process with parameters $\lambda = 1$ and H (see Definition 2.2.1). Left: $H = 1/4$ (subdiffusive: $\alpha = 1/2$). Right: $H = 1/2$ (diffusive: $\alpha = 1$).	6
2.2	Monte Carlo distribution of $\bar{\mu}_2(1)$ over 5,000 paths of size 2^{10} . Top plots, fBm (left: $\alpha = 0.5$; right: $\alpha = 1.8$. Bottom plots, ifOU (left: $\alpha = 0.5$; right: $\alpha = 1.8$. In both cases, $\lambda = 1$).	20
3.1	$\log_2 - \log_2$ plots of MSD curves as a function of the lags. (a) 20 ifOU paths with length 2^{11} , the diffusion exponent is $\alpha = 0.6$; (b) 20 particle paths with length 1800 were randomly selected from the <i>P. aeruginosa</i> biofilm after treatment with COS2-NO at concentration level 8 mg ml ⁻¹ . The first red line in each plot indicates the fitted slope over small lag values $h = 1, 2$. The estimated diffusion exponent $\hat{\alpha}$ is 1.42 in (a) and 1.10 in (b). The second red line in each plot indicates the fitted slope over large lag values $h = 8, 32$. The estimated diffusion exponent $\hat{\alpha}$ is 0.74 in (a) and 0.69 in (b).	27
3.2	x -axis: α . Bias, standard deviation, and MSE of the estimators $\hat{\xi}$ and $\hat{\xi}_{\text{OLS}}$ as a function of diffusion exponents. For comparison, solid and dashed lines represent $\hat{\xi}$ and $\hat{\xi}_{\text{OLS}}$, respectively. The estimator $\hat{\xi}$ has smaller bias, standard deviation, and MSE than $\hat{\xi}_{\text{OLS}}$	30

3.3	x -axis: α . The effect of the bias correction procedure in the performance of the estimator $\widehat{\xi}$ as a function of the diffusion exponent α . For comparison, the latter is also considered without bias correction. The version of the estimator with bias correction has a significantly smaller bias than that without bias correction.	31
3.4	Monte Carlo standard error of the estimator $\widehat{\alpha}$ (dashed line) and the estimator $\Sigma_{\widehat{\xi}}(\widehat{\alpha}_{\widehat{\text{EM}}})$ (solid line) as a function of the diffusion exponent α . The latter closely matches the former, especially in the subdiffusive range $\alpha < 1$	36
3.5	x -axis: $H = \alpha/2$. Rejection rate of (a) chi-square heterogeneity test of multiple particle paths from a fluid (b) z -test of multiple particle paths from two fluid samples, as a function of the diffusion exponent. For every value of α , each of 500 Monte Carlo runs consisted of generating 50 independent paths and conducting a test at 95% confidence level. The Monte Carlo rejection rate is very close to the theoretical value of 0.05.	37
3.6	Interfluid test power v.s. $\delta_{\alpha} = \alpha_{\text{I}} - \alpha_{\text{II}} $. Top: $\alpha = 0.2$. Middle: $\alpha = 1.0$. Bottom: $\alpha = 1.8$	39
3.7	Bias, standard deviation, and MSE of two sampling strategies: more paths v.s. longer paths. x -axis: total number of data points. Top: $\alpha = 0.6$. Middle: $\alpha = 1.0$. Bottom: $\alpha = 1.8$	43
4.1	Histogram of $\widehat{E}(n)$ with fitted normal curve. Left: real part. Right: imaginary part. Top: $\alpha = 0.6$. Middle: $\alpha = 1.0$. Bottom: $\alpha = 1.8$. . .	63
4.2	x -axis: $\log(N + 1)$. y -axis: log of the sample standard deviation. Left: real part of $\widehat{E}(n)$. Right: imaginary part of $\widehat{E}(n)$	64

Contents

Acknowledgement	ii
List of Tables	iii
List of Figures	iv
1 Introduction	1
2 The asymptotic distribution of the pathwise mean squared displacement in single particle tracking experiments	3
2.1 Introduction	3
2.2 Preliminaries and assumptions	9
2.3 The asymptotic distribution of MSD-based anomalous diffusion parameter estimators	15
2.4 Conclusion	21
3 MSD-based fluid heterogeneity testing	23
3.1 Introduction	23
3.2 Improved pathwise MSD-based estimation	24
3.3 Testing heterogeneity	32
3.4 Ensemble estimation: more or longer paths?	40
3.5 Application in experimental data: heterogeneity of treated <i>P. aeruginosa</i> biofilms	42

3.6	Conclusion	45
4	The asymptotic distribution of a mixing estimator	47
4.1	Introduction	47
4.2	Non-central and central limit theorems	49
4.3	The asymptotic distribution of $\hat{E}(n)$	57
4.4	Conclusion	64
A	Proofs for Chapter 2	65
A.1	Proofs for Section 2.2	65
A.2	Proofs for Section 2.3	68
A.3	Auxiliary results	72
A.4	Additional proofs	74
B	Proofs for Chapter 3	85
B.1	Some lemmas	86
B.2	Proofs of Theorem 3.2.1 and Theorem 3.2.2	96
B.3	A numerical approximation of κ_1 and κ_2	97
C	Proofs for Chapter 4	100
	References	106

Chapter 1

Introduction

Recent advances in light microscopy have spawned new research frontiers in nanobiophysics by working around the diffraction barrier and allowing for the observation of nanometric biological structures. Microrheology is the study of the properties of complex, non-Newtonian fluids, such as those found in biology, by means of the anomalous dynamics of small embedded particles, typically latex beads. Statistics based on the recorded sample paths are used by biophysicists to infer rheological (material) properties of the viscoelastic fluid under scrutiny. The goal of this dissertation is to provide new probabilistic results and statistical methods for the analysis of single particle tracking experiments in nanobiophysics. It comprises three chapters, each pertaining to a different problem.

Chapter 2 is dedicated to the probabilistic behavior of the so-named mean squared displacement ($\widehat{\text{MSD}}$) of tracer particles. The latter is the main statistic used in a very large proportion of biophysical papers on the analysis of anomalous diffusion. Establishing one key property of the $\widehat{\text{MSD}}$, namely, its asymptotic distribution, remained an open problem for years in the biophysical literature. In this thesis, we solve this problem for a broad class of fractional, stationary increment Gaussian processes that includes fractional Brownian motion and the integrated fractional Ornstein-Uhlenbeck process. The content of the chapter was accepted for the publication in *Journal of*

Time Series Analysis [23].

In Chapter 3, we tackle the problem of fluid heterogeneity. Due to the presence of organic compounds, samples of complex fluids such as those studied in microrheology may be physically heterogeneous both over one fluid sample and across multiple samples. This implies that accurately estimating anomalous diffusion parameters from single particle trajectories is of special interest. In Chapter 3, we draw upon the main result in Chapter 2 to propose a general protocol for the statistical testing of fluid heterogeneity in both senses, and apply it to real data on biofilm eradication. This is part of a collaborative project with David B. Hill (UNC-Chapel Hill).

Whether or not a given dynamical system has the properties of ergodicity and mixing is a central concern in the field of thermodynamics. In [50], a statistic was proposed for verifying the mixing property starting from a single observed particle path. Nevertheless, the asymptotic distribution of this statistic has been an open problem since the paper's publication. In Chapter 4, we solve this problem assuming that the underlying process is a fractional Gaussian noise. In particular, we show that the limiting distribution is Gaussian or non-Gaussian depending on the value of diffusion exponent.

Chapter 2

The asymptotic distribution of the pathwise mean squared displacement in single particle tracking experiments

2.1 Introduction

Abbe's diffraction limit stood for more than a hundred years as a barrier for light microscopy. The resolution limit of roughly 250nm ($1\text{nm} = 10^{-9}\text{m}$) is large compared to organelles in biological cells and most nanostructures. However, in the last twenty years advances in light microscopy technology have spawned new research frontiers by allowing for the observation of nanobiological phenomena *in vitro* and *in vivo* up to resolutions of 10–20nm (e.g., Hell [29, 30], Betzing et al [7], Rust et al [76], Hess et al [31], Westphal et al [88], Berning et al [6], Jones et al [38], Huang et al [33]). Microrheology is a rapidly expanding subfield of nanobiophysics. It consists of the study of the properties of complex fluids, such as those found in biology, through the dynamics of small embedded particles, typically latex beads, tracked and

recorded by means of new light microscopy technology. Microrheology is currently the dominant technique in the study of the physical properties of complex biofluids, of the rheological properties of membranes or the cytoplasm of cells, or of the entire cell (Mason and Weitz [53], Wirtz [89]; see Didier et al. [22] for a broad description of the statistical challenges involved).

The characterization of the diffusive behavior of nanometric particles embedded in viscous, Newtonian fluids is now well-understood both physically and probabilistically. However, in complex fluids particles are expected to display non-classical, or *anomalous*, behavior due to the viscoelasticity of the fluid. As in the early analysis of diffusion, biophysicists dedicate a great deal of attention to the average distance traveled by a particle, namely, the mean-square displacement $\mu_2(t) := EX^2(t)$ (MSD), where $X(t)$ denotes the position of the particle at instant t . For a given time window I , we can express the “local” MSD in the form

$$EX^2(t) \approx \theta t^\alpha, \quad \alpha, \theta > 0, \quad t \in I, \quad (2.1)$$

where the parameters θ and α are called the diffusivity coefficient and diffusion exponent, respectively. The microparticle is said to be sub-, super- or simply diffusive if the α is less than, greater than, or equal to 1, respectively. When $\alpha \neq 1$, the diffusion is commonly named anomalous (see O’Malley and Cushman [67] for a different perspective). The interval I in (2.1) can be of finite length or open-ended, according to the demands of physical analysis. In the former case, (2.1) expresses transient MSD behavior, as observed in polymer physics (Rubinstein and Colby [75], Kremer and Grest [43]). Alternatively, (2.1) describes the asymptotic MSD behavior (see the relation (2.19) for the accurate mathematical depiction of (2.1) in the context of this chapter).

Statistical evidence of anomalous diffusion has turned up in several contexts, including biodiffusion (Valentine et al. [83]), blinking quantum dots (Brokmann et al.

[10], Margolin and Barkai [52]) and fluorescence studies in single-protein molecules (Kou and Xie [42], Kou [41]). The dominant statistical technique in the biophysical literature for estimating the diffusion exponent α is what we will call the sample mean-square displacement ($\widehat{\text{MSD}}$). Suppose that a microrheological experiment generates a tracer bead sample path with observations $X(\Delta j)$, $j = 0, 1, \dots, n$, where $\Delta \in \mathbb{N}$ stands for the sampling rate. The pathwise statistic

$$\bar{\mu}_2(\Delta h) := \frac{1}{n-h} \sum_{j=1}^{n-h} (X(\Delta(j+h)) - X(\Delta j))^2 \quad (2.2)$$

is the $\widehat{\text{MSD}}$ at h , i.e., the statistical counterpart of $\mu_2(\cdot)$ (for notational simplicity, we do not display the dependency of $\bar{\mu}_2$ on n). Under (2.1), and assuming stationary increments, for m lag values $h_1 < \dots < h_m$ and $m \ll n$ one hopes for ergodicity, namely,

$$\bar{\mu}_2(\Delta h_k) \approx EX^2(\Delta h_k), \quad k = 1, \dots, m. \quad (2.3)$$

One then generates $(\widehat{\log \theta}, \widehat{\alpha})$ by means of the linear regression

$$\log \bar{\mu}_2(\Delta h_k) = \log \theta + \alpha \log(\Delta h_k) + \varepsilon_k, \quad k = 1, \dots, m, \quad (2.4)$$

possibly over several independent particle paths, where $\{\varepsilon_k\}_{k=1, \dots, m}$ is a random vector with an unspecified distribution. Plots of $\widehat{\text{MSD}}$ curves as a function of the lag h , sometimes on a log-log scale (see Figure 2.1), are widely reported as part of diffusion analysis (e.g., Valentine et al. [83], Suh et al. [79], Matsui et al [54], Lai et al [44], Lieleg et al. [46]). The choice of lags h_1, \dots, h_m reflects the analyst's visual perception of the range where the slope of the MSD curves stabilize and thus indicate the true diffusive regime and power law (2.1).

The stochastic properties of the $\widehat{\text{MSD}}$ depend on the underlying class of stochastic processes. In the review paper Meroz and Sokolov [57], the authors classify physi-

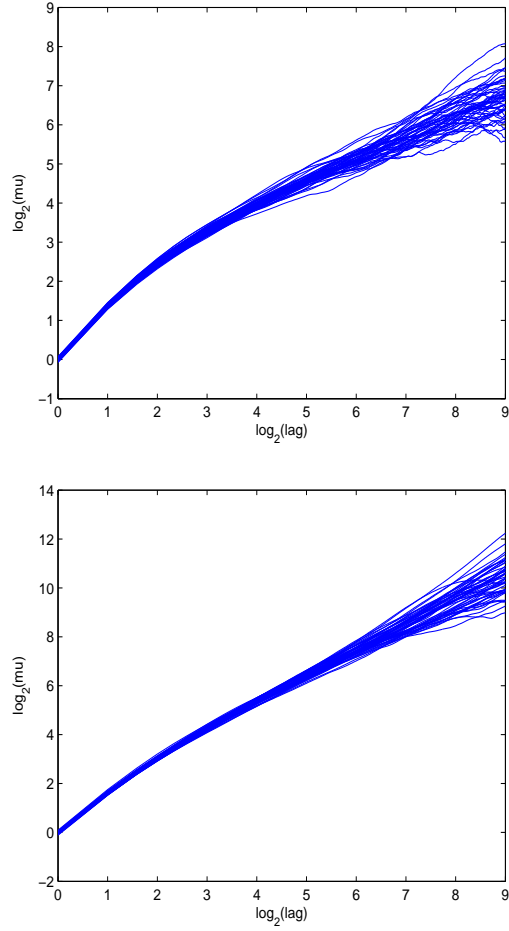


Figure 2.1: $\widehat{\text{MSD}}$ plots, each with 50 paths of size 2^{12} from an ifOU process with parameters $\lambda = 1$ and H (see Definition 2.2.1). Left: $H = 1/4$ (subdiffusive: $\alpha = 1/2$). Right: $H = 1/2$ (diffusive: $\alpha = 1$).

cal models for subdiffusive behavior according to whether one assumes the presence of binding-unbinding events, of geometrical constraints on the particle's movement, or the medium is viscoelastic. This leads to three popular families of stochastic processes, respectively, those of continuous time random walks (Metzler and Klafter [58], Meerschaert and Scheffler [55]), of random walks on fractals (Havlin and Ben-Avraham [28]), and of the celebrated fractional Brownian motion (fBm; see Example 2.2.1). In this chapter, we focus on the latter family, more precisely, that of fractional, stationary increment processes (Barkai et al. [4], Lysy et al. [48]).

The ergodicity of the $\widehat{\text{MSD}}$ moments was established in Deng and Barkai [20] for various families of fractional processes (see also Sokolov [78], Metzler et al. [59], Jeon and Metzler [37], Burov et al. [15], Jeon et al. [36], Sandev et al. [77]). Finite sample approximations to the distribution of the $\widehat{\text{MSD}}$ under Gaussianity are provided in Grebenkov [25] (see also Qian et al. [71], Grebenkov [26], Boyer et al. [8], Andreanov and Grebenkov [1], Nandi et al. [65], Boyer et al. [9]). Nevertheless, so far $\widehat{\text{MSD}}$ -based analysis of tracking data has missed one essential feature of statistical methods, namely, the *limiting distribution* of the random vector

$$\left(\bar{\mu}_2(\Delta h_1), \dots, \bar{\mu}_2(\Delta h_m)\right). \quad (2.5)$$

The purpose of this chapter is to fill this gap. We work under the assumption that the particle undergoes a Gaussian process whose stationary increments display a covariance function γ satisfying a decay condition of the type

$$\gamma_h(z)_{1,2} \sim C z^{\alpha-2} h_1 h_2, \quad z \rightarrow \infty, \quad (2.6)$$

for some real constant C , where h_1, h_2 represent lag sizes (see the expressions (2.16) and (2.20) for precise definitions, notation and statements). As in the particular case of fBm, such a particle is not constrained by boundaries (such as those found in a

cell) or a potential.

We assume the availability of just one sample path. This models the situation in which the biophysical samples are physically *heterogeneous* (Valentine et al. [83], Dawson et al. [19], Monnier et al. [61]). Complex biomaterials such as mucus, or simulants such as agarose and hyaluronic acid, are expected to be heterogeneous due to the unequal distribution of chains of polymers. Since the multiple $\widehat{\text{MSD}}$ averages are formed from the same particle path, then even when $|h_2 - h_1|$ is large the associated coefficients (2.2) still display strong correlation (Monnier et al. [61]). Our main result (Theorem 2.3.1) shows that this yields limiting distributions and convergence rates that depend on the diffusion exponent range according to a familiar trichotomy in the literature on fractional processes. When $0 < \alpha \leq 3/2$, the asymptotic distribution is Gaussian, though the case $\alpha = 3/2$ demands a non-standard convergence rate. When $3/2 < \alpha < 2$ the convergence rate depends on the diffusion exponent and the asymptotic distribution is non-Gaussian; this reflects the classical results by M. Rosenblatt [74] and M. Taqqu [80]. This type of result is well-known for fixed sequences of Gaussian, stationary random variables, or for p -variations of shrinking interval size of Gaussian processes (Guyon and León [27], Peltier and Véhel [68], Hosking [32], Bardet [3], Buchmann and Chan [11]). By contrast, we consider the $\widehat{\text{MSD}}$ statistics in the same format found in the biophysical literature, namely, we take the lag limit $h \rightarrow \infty$. Moreover, whereas the related literature on Hermitian processes and random fields often makes use of Wiener-Itô chaos expansions and Malliavin calculus (e.g., Nourdin et al. [66], Réveillac et al. [73]), in this work we develop our results in the style of Rosenblatt's classical arguments as to make the statements and techniques more readily available to the interested reader with a biophysical background. The asymptotic distributions provided allow for a new statistical perspective on the many numerical-experimental results reported by the biophysical community, and make it possible to mathematically compare $\widehat{\text{MSD}}$ -based analysis with that based on other

candidate statistical techniques, e.g., in the Fourier and wavelet domains.

It should be stressed that we do not assume exact self-similarity (see relation (2.9)). Dispensing with the latter property is important because it is often of interest to start from a Newtonian instance, such as the generalized Langevin equation (GLE; e.g., Lysy et al. [48], p.6), to arrive at an anomalous diffusion model that displays non-fractional short range behavior. In particular, we show that our results encompass the stationary-increment process induced by a fractional Ornstein-Uhlenbeck (fOU) velocity process (see Definition 2.2.1). The latter can be regarded as a spectrally simplified model for fractional instances of the GLE (see (2.27)).

The chapter is divided as follows. Section 2.2 contains most definitions and the assumptions used throughout the chapter. We also shed light on the proposed assumptions by showing that they imply the properties (2.1) and (2.6). In Section 2.3, we state and discuss weak limits for the $\widehat{\text{MSD}}$. Furthermore, Monte Carlo experiments are used to illustrate the Gaussian or non-Gaussian nature of the $\widehat{\text{MSD}}$ distribution, and to study the quality of the asymptotic approximation. All proofs can be found in the Appendix A.

2.2 Preliminaries and assumptions

All through the chapter, C is used in bounds to denote a constant that does not depend on the sample size n , and which may change from one line to another. For two sequences of real numbers $\{a_n\}_{n \in \mathbb{N}}$, $\{b_n\}_{n \in \mathbb{N}}$, the expression $a_n \sim b_n$ means that $\frac{a_n}{b_n} \rightarrow 1$ as $n \rightarrow \infty$.

Recall that a stochastic process X is said to have stationary increments when $\{X(t+h) - X(h)\}_{t \in \mathbb{R}}$ has the same finite-dimensional distributions for any time shift $h \in \mathbb{R}$. The stochastic process in (2.1) is assumed to satisfy the following condition.

ASSUMPTION (A1): $X = \{X(t)\}_{t \in \mathbb{R}}$ is a Gaussian, stationary-increment process with harmonizable representation

$$X(t) = C_\alpha \int_{\mathbb{R}} \frac{e^{itx} - 1}{ix} \frac{s(x)}{|x|^{\alpha/2-1/2}} \tilde{B}(dx), \quad (2.7)$$

where $\alpha \in (0, 2)$, $C_\alpha \neq 0$, $\tilde{B}(dx)$ is a \mathbb{C} -valued Brownian measure such that $\tilde{B}(-dx) = \overline{\tilde{B}(dx)}$, $E\tilde{B}(dx)\overline{\tilde{B}(dx')} = 0$ ($x \neq x'$), and $E|\tilde{B}(dx)|^2 = dx$. The function $s(x)$ is a bounded and complex-valued function with $|s(0)|^2 = 1$, and

$$||s(x)|^2 - 1| \leq C_0 x^{\delta_0}, \quad x \in (-\varepsilon_0, \varepsilon_0), \quad (2.8)$$

for constants $C_0, \delta_0, \varepsilon_0 > 0$.

In particular, the representation (2.7) implies that $EX(t) = 0$, $t \in \mathbb{R}$. However, this is inconsequential for modeling, since one can always assume that a single diffusing particle starts at zero. In turn, the condition (2.8) is mild (c.f. Moulines et al. [62], p.302, relation (4)) and plays a technical role in the proof of Proposition 2.2.1 below.

Example 2.2.1. FBm is the only Gaussian, self-similar, stationary-increment process (Taqqu [81], Proposition 2.3). The self-similarity of the fBm $B_H = \{B_H(t)\}_{t \in \mathbb{R}}$ means that, for a Hurst parameter $0 < H \leq 1$, the scaling relation

$$\{B_H(ct)\}_{t \in \mathbb{R}} \stackrel{\mathcal{L}}{=} \{c^H B_H(t)\}_{t \in \mathbb{R}}, \quad c > 0, \quad (2.9)$$

is satisfied. FBm has mean zero, and by Gaussianity, it is characterized by its closed-form covariance function

$$EB_H(s)B_H(t) = \frac{\sigma^2}{2} \{|t|^{2H} + |s|^{2H} - |t-s|^{2H}\}, \quad s, t \in \mathbb{R}. \quad (2.10)$$

In particular, when $\sigma^2 = 1$, we call B_H a standard fBm. Moreover, by taking $s = t$

in (2.10), the expression (2.1) holds at all t as an equality with

$$\alpha = 2H. \quad (2.11)$$

When H is less than, greater than or equal to $1/2$, fBm is sub-, super- or simply diffusive (Brownian motion), respectively. The harmonizable representation of a standard fBm is given by

$$X(t) = C_H \int_{\mathbb{R}} \frac{e^{itx} - 1}{ix} \frac{1}{|x|^{H-1/2}} \tilde{B}(dx), \quad (2.12)$$

where

$$C_H = \sqrt{\pi^{-1} H \Gamma(2H) \sin(H\pi)} \quad (2.13)$$

(see Taqqu [81], p.31, expression (9.8)). Thus, fBm satisfies (A1) with $C_\alpha = C_H$ and $s(x) = 1$.

Let $\Delta = 1$ in (2.2). We assume that an experiment produced *one* sequential sample X_1, \dots, X_n , $n \in \mathbb{N}$, of observations from the stochastic process X . For $h \in \mathbb{N}$, let

$$\{h_k := w_k h\}_{k=1, \dots, m} \quad (2.14)$$

be distinct integer-valued increment sizes, where $h_m \leq n - 1$, $w_1 < w_2 < \dots < w_m$ and $m \leq n - 1$. We can define an associated vector of increments

$$\mathbf{Y} = (Y_1(h_1), \dots, Y_{n-h_1}(h_1); Y_1(h_2), \dots, Y_{n-h_2}(h_2); \dots; Y_1(h_m), \dots, Y_{n-h_m}(h_m))^T, \quad (2.15)$$

where

$$Y_i(h_k) := X(i + h_k) - X(i), \quad h_k \in \mathbb{N} \cup \{0\}, \quad k = 1, \dots, m.$$

Since X is a stationary-increment process, then the cross product

$$EY_{j+z}(h_{k_1})Y_j(h_{k_2}) = E(X(j+z+h_{k_1}) - X(j+z))(X(j+h_{k_2}) - X(j))$$

$$= E(X(z + h_{k_1}) - X(z))(X(h_{k_2}) - X(0)) = EY_z(h_{k_1})Y_0(h_{k_2})$$

is not a function of j . Denote the covariance matrix of the increments by $\Sigma_{\mathbf{Y}}(h_1, \dots, h_m)$, where an entry has the form

$$\gamma_h(z)_{k_1, k_2} = EY_{j+z}(h_{k_1})Y_j(h_{k_2}), \quad (2.16)$$

for $k_1, k_2 = 1, \dots, m$. Note that (2.16) satisfies the symmetry relation

$$\gamma_h(-z + h_{k_2} - h_{k_1})_{k_1, k_2} = \gamma_h(z)_{k_1, k_2}, \quad z \in \mathbb{Z}.$$

The self-similarity of fBm (see (2.9)) makes the asymptotic distribution of the $\widehat{\text{MSD}}$ much simpler to establish (see Peltier and V  hel [68], Proposition 4.2). Since we do not assume self-similarity, as in the biophysical literature we need to make the size of the lags themselves go to infinity, though slower than the sample size n . This mathematically expresses what biophysicists do in practice: h has to be large enough for the $\widehat{\text{MSD}}$ regime to become linear, but at the same time cannot be too large because of the increased variance of the $\widehat{\text{MSD}}$. This is illustrated in Figure 2.1 and accurately described in assumption (A2), stated next.

ASSUMPTION (A2): for $h = h(n) \in \mathbb{N} \cup \{0\}$, $n \in \mathbb{N}$,

$$\frac{h(n) \log^2(n)}{n} + \frac{n}{h(n)^{1+\delta/2}} \rightarrow 0, \quad n \rightarrow \infty, \quad (2.17)$$

where

$$\delta = \min(\alpha/2, \delta_0/2) \quad (2.18)$$

(see (2.7) and (2.8) for the definitions of α and δ_0).

As anticipated in the Introduction, the expressions (2.19) and (2.20) in the following proposition give exact mathematical meaning to the heuristic properties (2.1)

and (2.6).

Proposition 2.2.1. *Suppose the assumptions (A1) and (A2) hold.*

(i) *Then, there is a constant $\theta > 0$ such that*

$$\left| \frac{EX^2(h)}{\theta h^\alpha} - 1 \right| \leq Ch^{-\delta}, \quad n \rightarrow \infty, \quad (2.19)$$

where $\delta > 0$ is given by (2.18);

(ii) *moreover, for any $k_1, k_2 = 1, \dots, m$ and $\gamma_h(z)_{k_1, k_2}$ as in (2.16),*

$$\gamma_h(z)_{k_1, k_2} = |z|^{\alpha-2} h^2 w_{k_1} w_{k_2} \{ \tau + g(z, h)_{k_1, k_2} \}, \quad n \rightarrow \infty, \quad (2.20)$$

$$h_m + 1 \leq |z| \leq n, \quad (2.21)$$

where

$$\tau = \tau(\alpha) = \left(\frac{C_\alpha}{C_H} \right)^2 \frac{\alpha(\alpha-1)}{2}, \quad (2.22)$$

C_H is given by (2.13), and the residual function $g(\cdot, \cdot)_{k_1, k_2}$ satisfies

$$|g(z, h)_{k_1, k_2}| \leq C \left(\frac{h}{|z|} \right)^\delta, \quad 0 < h \leq |z|. \quad (2.23)$$

Besides fBm, another model for anomalous diffusion used in this work is what we call the integrated fractional Ornstein-Uhlenbeck (ifOU) process. One major difference between fBm and the ifOU process is that the latter is not exactly self-similar. The ifOU is an example of a fractional process whose $\widehat{\text{MSD}}$ asymptotics are naturally studied under the assumption (A2).

To define the ifOU process, recall that the fractional Ornstein-Uhlenbeck process (fOU) is the a.s. continuous solution to the fBm-driven Langevin equation

$$dV(t) = -\lambda V(t)dt + \sigma dB_H(t), \quad t \geq 0, \quad \lambda > 0, \quad 0 < H < 1 \quad (2.24)$$

(Rao [70], p.78). The a.s. continuous process

$$V(t) = \sigma \int_{-\infty}^t e^{-\lambda(t-u)} dB_H(u), \quad t > 0,$$

solves (2.24) with initial condition $V(0)$ defined by the same integral. When $H = 1/2$, the solution is the classical Ornstein-Uhlenbeck process. We are interested in the stationary-increment counterpart of the fOU process, as put forward in the next definition.

Definition 2.2.1. Given a fOU process $\{V(t)\}_{t \geq 0}$, the associated ifOU process is given by

$$X(t) = \int_0^t V(s) ds, \quad t > 0. \quad (2.25)$$

The integrand in (2.25) is a version of V with continuous paths (see Didier and Fricks [21], p.719, Lemma A.4).

The ifOU is a simple parametric model for anomalous diffusion. This can be seen in the Fourier domain, based on the harmonizable representation

$$X(t) = \sigma \sqrt{\Gamma(2H+1) \sin(\pi H)} \int_{\mathbb{R}} \frac{e^{itx} - 1}{ix} \frac{1}{\lambda + ix} \frac{1}{|x|^{H-1/2}} \tilde{B}(dx). \quad (2.26)$$

The spectral density

$$f_X(x) = \sigma^2 \Gamma(2H+1) \sin(\pi H) \frac{1}{\lambda^2 + x^2} \frac{1}{|x|^{2H-1}}, \quad x \in \mathbb{R} \setminus \{0\}, \quad (2.27)$$

exhibits the short range dependence term $(\lambda^2 + x^2)^{-1}$ besides the fractional term $|x|^{1-2H}$, with a tuning parameter λ .

Let $H \in (0, 1) \setminus \{1/2\}$. By (2.27) and dominated convergence, the covariance function $\gamma(s) = \text{Cov}(V(t), V(t+s))$ of V is continuous. Furthermore, in Cheridito et

al. [16], p.8, it is shown that

$$\gamma(s) = \frac{\sigma^2}{2} \sum_{n=1}^N \lambda^{-2n} \left(\prod_{k=0}^{2n-1} (2H - k) \right) s^{2H-2n} + O(s^{2H-2N-2}), \quad (2.28)$$

for an arbitrary $N \in \mathbb{N}$, where the remainder is taken with respect to $s \rightarrow \infty$.

Therefore,

$$|EX(t_1)X(t_2)| \leq \int_0^{t_1} \int_0^{t_2} |\gamma(s_1 - s_2)| ds_1 ds_2 < \infty, \quad (2.29)$$

whence the integral (2.25) is also well-defined in the mean-square sense (see Cramér and Leadbetter [18], p.86). By (2.26), like fBm the ifOU also satisfies (A1) and the conclusions of Proposition 2.2.1 apply, where the relation between α and H is again given by (2.11). Note that when the ifOU is simply diffusive ($H = 1/2$, or $\alpha = 1$), the expression (2.20) holds with $\tau = 0$.

2.3 The asymptotic distribution of MSD–based anomalous diffusion parameter estimators

The following theorem is the main result of this paper. It gives the asymptotic distribution of a random vector of $\widehat{\text{MSD}}$ entries at different lag values, according to subranges of the diffusion exponent α . For the theorem, recall that in the notation (2.15), the $\widehat{\text{MSD}}$ at a given lag value is given by

$$\bar{\mu}_2(h_k) = \frac{1}{N_k} \sum_{i=1}^{N_k} Y_i^2(h_k), \quad (2.30)$$

where

$$N_k := n - h_k, \quad k = 1, \dots, m. \quad (2.31)$$

Theorem 2.3.1. *Suppose the assumptions (A1) and (A2) hold. Let $m \in \mathbb{N}$ be the*

chosen number of $\widehat{\text{MSD}}$ lag values, and let N_k be as in (2.31), $k = 1, \dots, m$. Then,

$$\left(\frac{N_k}{\eta(N_k)\zeta(h_k)} (\bar{\mu}_2(h_k) - EX^2(h_k)) \right)_{k=1, \dots, m} \xrightarrow{d} \mathbf{Z}, \quad n \rightarrow \infty, \quad (2.32)$$

where

$$\begin{cases} 0 < \alpha < 3/2 : & \eta(n) = \sqrt{n}, \zeta(h) = h^{\alpha+1/2}; \\ \alpha = 3/2 : & \eta(n) = \sqrt{n \log(n)}, \zeta(h) = h^2; \\ 3/2 < \alpha < 2 : & \eta(n) = n^{\alpha-1}, \zeta(h) = h^2. \end{cases} \quad (2.33)$$

In (2.32),

(i) if $0 < \alpha < 3/2$, then $\mathbf{Z} \sim N(0, \Sigma)$, where the entry k_1, k_2 of the matrix $\Sigma = \Sigma(\alpha)$ is given by

$$\Sigma_{k_1, k_2} = 2w_{k_1}^{-\alpha-1/2} w_{k_2}^{-\alpha-1/2} \left(\frac{C_\alpha}{C_H} \right)^4 \|\widehat{G}(y; w_{k_1}, w_{k_2})\|_{L^2(\mathbb{R})}^2, \quad k_1, k_2 = 1, \dots, m, \quad (2.34)$$

$$\widehat{G}(y; w_{k_1}, w_{k_2}) = C_H^2 \frac{(e^{iw_{k_1}y} - 1)(e^{-iw_{k_2}y} - 1)}{|y|^{\alpha+1}}, \quad (2.35)$$

and C_H is given in (2.13);

(ii) if $\alpha = 3/2$, then $\mathbf{Z} \sim N(0, \Sigma)$, where the entry k_1, k_2 of the matrix $\Sigma = \Sigma(\alpha)$ is given by

$$\Sigma_{k_1, k_2} = 4\tau^2, \quad k_1, k_2 = 1, \dots, m. \quad (2.36)$$

(iii) if $3/2 < \alpha < 2$, \mathbf{Z} follows a multivariate Rosenblatt-type distribution whose characteristic function is

$$\phi_{\mathbf{Z}}(\mathbf{t}) = \exp \left\{ \frac{1}{2} \sum_{s=2}^{\infty} \frac{[2i\tau \sum_{k=1}^m t_k]^s}{s} c_s \right\}, \quad (2.37)$$

where, for $s \geq 2$,

$$c_s = c_s(\alpha) = \int_0^1 \int_0^1 \cdots \int_0^1 |x_1 - x_2|^{\alpha-2} |x_2 - x_3|^{\alpha-2} \cdots |x_s - x_1|^{\alpha-2} dx_1 dx_2 \cdots dx_s. \quad (2.38)$$

Remark 2.3.1. It is worthwhile recalling the fact that the constant c_s in (2.38) is, indeed, finite. Indeed, by an application of the Cauchy-Schwarz inequality,

$$c_s \leq \left(\int_0^1 \int_0^1 |x_1 - x_2|^{2\alpha-4} dx_1 dx_2 \right)^{s/2} = \left(\frac{1}{(2\alpha-3)(\alpha-2)} \right)^{s/2}. \quad (2.39)$$

Theorem 2.3.1 allows us to develop the asymptotic distribution of the $\widehat{\text{MSD}}$ -based least square estimator of the diffusivity coefficient and diffusion exponent. Recast the (pathwise) system (2.4) as the regression model

$$Q_n = M_n \beta + \varepsilon_n, \quad (2.40)$$

where

$$Q_n = \begin{pmatrix} \log \bar{\mu}_2(h_1) \\ \vdots \\ \log \bar{\mu}_2(h_m) \end{pmatrix}, \quad \beta = \begin{pmatrix} \log \theta \\ \alpha \end{pmatrix}, \quad M_n = \begin{pmatrix} 1 & \log(h_1) \\ \vdots & \vdots \\ 1 & \log(h_m) \end{pmatrix}, \quad (2.41)$$

and ε has a distribution to be determined. We will denote by

$$\widehat{\beta}_n := (M_n^T M_n)^{-1} M_n^T Q_n = ((\widehat{\log \theta})_n, \widehat{\alpha}_n)^T \quad (2.42)$$

the estimator generated by the ordinary least squares solution to the system (2.40).

The next corollary describes the asymptotic distribution of the least squares estimator (2.42).

Corollary 2.3.1. *Suppose the assumptions (A1) and (A2) hold. Then, as $n \rightarrow \infty$,*

$$\frac{nh^\alpha}{\eta(n)\zeta(h)} \begin{pmatrix} \frac{1}{\log h}((\widehat{\log \theta})_n - \log \theta) \\ \widehat{\alpha}_n - \alpha \end{pmatrix} \xrightarrow{d} \begin{pmatrix} U^T \\ -U^T \end{pmatrix} A\mathbf{Z}, \quad (2.43)$$

where A and \mathbf{Z} , respectively, are as in (A.15) and Theorem 2.3.1, and

$$U^T = \frac{1}{c_w} \left(\sum_{k=1}^m \log(w_k/w_1), \dots, \sum_{k=1}^m \log(w_k/w_m) \right) \quad (2.44)$$

with constant

$$c_w = m \sum_{k=1}^m \log^2(w_k) - \left(\sum_{k=1}^m \log(w_k) \right)^2. \quad (2.45)$$

Corollary 2.3.1 states that the limiting distributions for $\widehat{\beta}_n$ are qualitatively distinct as a function of the underlying diffusion exponent α . In particular, a non-Gaussian limit appears in the superdiffusive range $3/2 < \alpha < 2$. Though probably of little interest in the modeling of viscoelastic diffusion, superdiffusion appears in many other applications (e.g., Brokman et al. [10], Margolin and Barkai [52]; note that in these papers the processes are viewed as following Lévy walk-type dynamics).

Remark 2.3.2. Corollary 2.3.1 can be directly used in the construction of confidence intervals, at least starting from knowledge that α lies in one of the subregions $(0, 3/2)$ or $(3/2, 2)$ of the parameter space. To fix ideas, consider the parameter α ; the ensuing argument can be easily adapted for θ . By Corollary 2.3.1, $\widehat{\alpha}_n - \alpha$ is asymptotically equivalent to

$$-\frac{\eta(n)\zeta(h)}{nh^\alpha} U^T A\mathbf{Z} =: \frac{\eta(n)\zeta(h)}{nh^\alpha} \sigma(\theta, \alpha) Z(\alpha),$$

where $\sigma(\theta, \alpha)$ is a smooth function of (θ, α) defined as to make $Z(\alpha)$ a standardized random variable. When $0 < \alpha < 3/2$, this is clearly possible, since the limiting distribution is Gaussian and $-U^T A$ and Σ are smooth functions of θ and α (see (2.34), (A.14) and (2.44)). When $3/2 < \alpha < 2$, first note that \mathbf{Z} in Theorem 2.3.1 is a

rank 1 random vector. Indeed, recall that the characteristic function of a standardized (mean zero, variance one) Rosenblatt random variable is given by

$$\phi_\alpha(t) = \exp \left\{ \frac{1}{2} \sum_{s=2}^{\infty} (2it\psi(\alpha))^s \frac{c_s}{s} \right\}, \quad \psi(\alpha) := \left(\frac{(2\alpha-3)(\alpha-1)}{2} \right)^{1/2}, \quad (2.46)$$

(see Veillette and Taqqu [84], expression (4)). Let Z_1 be a Rosenblatt random variable with normalizing constant τ (i.e., with the latter in place of $\psi(\alpha)$ in (2.46)), and let $Z_2 = \dots = Z_m := Z_1$. Then, $Ee^{i \sum_{k=1}^m t_k Z_k} = Ee^{i Z_1 \sum_{k=1}^m t_k} = \exp \left\{ \frac{1}{2} \sum_{s=2}^{\infty} [2i\tau \sum_{k=1}^m t_k]^s \frac{c_s}{s} \right\}$. So, denote by \tilde{Z} the limiting Rosenblatt random variable obtained in (2.22). Then, $Z_\alpha := (\psi(\alpha)/\tau)\tilde{Z}$ is standardized and, by (2.22) and (2.46), the coefficient $\psi(\alpha)/\tau$ also depends smoothly on α .

So, the consistency of $\hat{\theta}_n$ and $\hat{\alpha}_n$ for θ and α , respectively, implies that of $\sigma(\hat{\theta}_n, \hat{\alpha}_n)$ for $\sigma(\theta, \alpha)$. When $0 < \alpha < 3/2$, $Z(\alpha) \sim N(0, 1)$, which is independent of α . Then, an approximate $100(1 - \xi)\%$ confidence interval for α is simply

$$\left(\hat{\alpha}_n + \left(\frac{h}{n} \right)^{1/2} \sigma(\hat{\theta}_n, \hat{\alpha}_n) z_{\xi/2}, \hat{\alpha}_n + \left(\frac{h}{n} \right)^{1/2} \sigma(\hat{\theta}_n, \hat{\alpha}_n) z_{1-\xi/2} \right).$$

When $3/2 < \alpha < 2$, by (2.39) and the dominated convergence theorem, the characteristic function $\phi_\alpha(t) =: \phi(t, \alpha)$ of Z_α is continuous with respect to α for t around the origin. Now consider the function $\phi(z, \alpha)$ with domain in z extended to a vicinity of the origin of \mathbb{C} . By applying Theorem 7.1.1 in Lukacs [47] and the uniqueness of analytic continuation, we obtain that $\phi(t, \alpha)$ is continuous with respect to α for all $t \in \mathbb{R}$. Consequently, the cumulative distribution function and the quantile function $z_\varsigma(\alpha)$ are also continuous with respect to α , whence $z_\varsigma(\hat{\alpha}_n) \xrightarrow{P} z_\varsigma(\alpha)$ for $\varsigma \in (0, 1)$. So, an approximate $100(1 - \xi)\%$ confidence interval for α is

$$\left(\hat{\alpha}_n + \left(\frac{h}{n} \right)^{2-\hat{\alpha}_n} \sigma(\hat{\theta}_n, \hat{\alpha}_n) z_{\xi/2}(\hat{\alpha}_n), \hat{\alpha}_n + \left(\frac{h}{n} \right)^{2-\hat{\alpha}_n} \sigma(\hat{\theta}_n, \hat{\alpha}_n) z_{1-\xi/2}(\hat{\alpha}_n) \right) \quad (2.47)$$

(see Veillette and Taqqu [84] for numerical results on the quantiles of the Rosenblatt distribution). In (2.47), we are using the fact $(\frac{h}{n})^{2-\alpha}/(\frac{h}{n})^{2-\hat{\alpha}} \rightarrow 1$, which can be verified by taking logs.

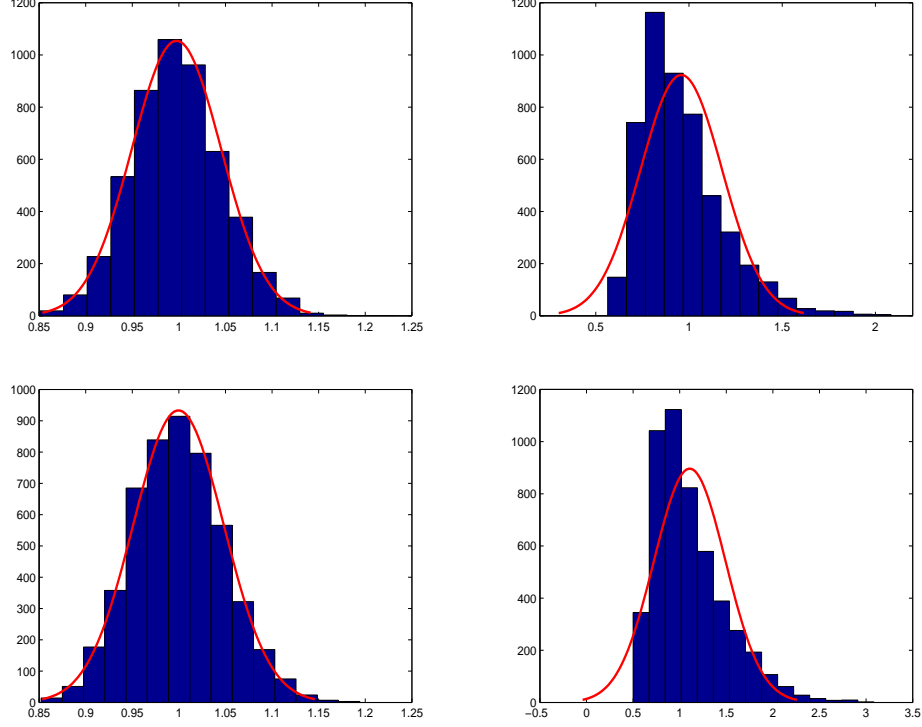


Figure 2.2: Monte Carlo distribution of $\bar{\mu}_2(1)$ over 5,000 paths of size 2^{10} . Top plots, fBm (left: $\alpha = 0.5$; right: $\alpha = 1.8$). Bottom plots, ifOU (left: $\alpha = 0.5$; right: $\alpha = 1.8$). In both cases, $\lambda = 1$.

To study the finite-sample properties of $\widehat{\text{MSD}}$ -based estimation and the quality of the asymptotic approximations described in Theorem 2.3.1 and Corollary 2.3.1, Monte Carlo experiments were conducted based on sub- and superdiffusive instances of fBm and ifOU processes. Figure 2.2 displays the Monte Carlo distributions of the $\widehat{\text{MSD}}$, i.e., histograms and best Gaussian fit. The plots reflect the results described in Theorem 2.3.1: for the subdiffusive Hurst parameter value of $H = 0.25$ ($\alpha = 0.5$), the distribution is distinctively Gaussian; by contrast, for the strongly superdiffusive value $H = 0.9$ ($\alpha = 1.8$), the Rosenblatt-like attractor skews the finite-sample distribution. Moreover, Table 2.1 displays results for $\hat{H} = \hat{\alpha}_{\widehat{\text{MSD}}}/2$ under a fBm. The simulations

encompass two distinct situations. In the first one, we follow the common practice in microrheology of taking a large number of consecutive lag values such as $h_{\min}, h_{\min} + 1, \dots, h_{\max} - 1, h_{\max}$, where $h_{\min} = 2$ and $h_{\max} = 128$. In the second one, we pick only two lag values, namely $h_1 = 2$ and $h_2 = 128$. The results show that dropping most of the $\widehat{\text{MSD}}$ s has little impact on the performance of $\widehat{\beta}_n$. Moreover, simulation studies not shown provide evidence that in both subdiffusive and superdiffusive ranges ($H = 0.25, 0.5, 0.75, 0.9$ or $\alpha = 0.5, 1, 1.5, 1.8$, respectively), a pairwise combination of two low lag values (such as $h_1 = 1$ and $h_2 = 2$) leads to the best statistical performance, as measured by the Monte Carlo mean-squared error.

As expected, though, the results for the ifOU are quite distinct. Since the latter is not exactly self-similar, the MSD curves display the asymptotic flattening effect revealed in Figure 2.1 for $H = 0.25$ ($\alpha = 0.50$). This is what drives biophysicists to use larger lags when modeling anomalous diffusion data in the first place. Table 2.2 illustrates this effect in the estimation of $\widehat{\beta}_n$ based on triples of consecutive lag values: the estimation bias decreases as the chosen lags increase. However, since the variance also increases with the lag, due to the smaller number of terms in $\bar{\mu}_2(h)$, the choice of regression lags with the lowest mean-squared error turns out to be $[2^5, 2^6, 2^7]$. A different phenomenon emerges in the superdiffusive range. Table 2.2 shows that for very high values of H or α , it may be optimal to use lower regression lag values. This is so because the bias for large values of H or α is very large. Though not displayed, the same issue appears under different parameter values in the superdiffusive range, and its cause is a matter for future investigation.

2.4 Conclusion

In this chapter, we establish the asymptotic distribution of the $\widehat{\text{MSD}}$ -based estimator widely used in the biophysical literature on anomalous diffusion. We assume that the particle undergoes a Gaussian, stationary-increment process, and take a path-

wise approach, i.e., only one particle path is available. Depending on the diffusion exponent of the underlying process, the $\widehat{\text{MSD}}$ -based estimator has Gaussian or non-Gaussian limiting distribution, as well as different convergence rates. The asymptotic distributions provided allow for a new statistical perspective on the many numerical-experimental results reported by the biophysical community. We illustrate our results analytically and computationally based on fractional Brownian motion and the integrated fractional Ornstein-Uhlenbeck process.

Table 2.1: Mean and s.d. of \widehat{H} for fBm: consecutive lags $(2, \dots, 2^8)$ vs two lags $(h_1 = 2, h_2 = 2^7)$

n	$H = 0.25$ ($\alpha = 0.5$)				$H = 0.90$ ($\alpha = 1.8$)			
	$\widehat{E}_{\text{cons}}(H)$	$\widehat{\text{sd}}_{\text{cons}}(H)$	$\widehat{E}_2(H)$	$\widehat{\text{sd}}_2(H)$	$\widehat{E}_{\text{cons}}(H)$	$\widehat{\text{sd}}_{\text{cons}}(H)$	$\widehat{E}_2(H)$	$\widehat{\text{sd}}_2(H)$
2^9	0.2363	0.0723	0.2376	0.0527	0.8380	0.1067	0.8459	0.0890
2^{10}	0.2437	0.0508	0.2446	0.0359	0.8583	0.0772	0.8644	0.0648
2^{11}	0.2468	0.0361	0.2476	0.0253	0.8715	0.0574	0.8751	0.0507
2^{12}	0.2482	0.0252	0.2486	0.0177	0.8792	0.0462	0.8814	0.0415

Table 2.2: Mean and s.d. of \widehat{H} for ifOU: small lags vs large lags, path length= 2^{12}

lags	$H = 0.25$ ($\alpha = 0.5$)		$H = 0.90$ ($\alpha = 1.8$)	
	$\widehat{E}(H)$	$\widehat{\text{sd}}(H)$	$\widehat{E}(H)$	$\widehat{\text{sd}}(H)$
$[2^3, 2^4, 2^5]$	0.3218	0.0272	0.8817	0.0266
$[2^4, 2^5, 2^6]$	0.2964	0.0354	0.8441	0.0392
$[2^5, 2^6, 2^7]$	0.2756	0.0488	0.8315	0.0506
$[2^6, 2^7, 2^8]$	0.2591	0.0742	0.8239	0.0704

Chapter 3

MSD-based fluid heterogeneity testing

3.1 Introduction

The experimental and statistical difficulties involved in estimating the parameters $\xi = (\log \theta, \alpha)$ based on the regression system (2.4) have been pointed out by many authors. A non-exhaustive list of issues includes limited fluorophore lifetimes, proteins diffusing out of the field of view, finite-resolution imaging and motion blurring due to camera integration times, measurement errors, the presence of drifts and intra-path correlation [5, 13, 39, 56, 60, 71, 86].

Notwithstanding the difficulties involved, the potential heterogeneity of fluid samples in fields such as microrheology implies that estimating ξ from single trajectories is of great interest [12, 14, 49, 86]. This calls for accurate results on the stochastic behavior of the $\widehat{\text{MSD}}$, and accordingly, a wealth of literature on the subject has developed. Starting from an underlying fractional stochastic process, several properties of the $\widehat{\text{MSD}}$ such as ergodicity and approximations to its finite sample distribution were established by many authors [1, 8, 9, 15, 20, 25, 25, 26, 36, 37, 59, 65, 71, 77].

Nevertheless, until recently, the limiting distribution of the $\widehat{\text{MSD}}$ was unknown.

This means that it was impossible, for example, to construct asymptotically valid confidence intervals for the anomalous diffusion parameters starting from a single observed particle path. This difficulty was overcome in Chapter 2 of this thesis under mild assumptions for the case where the particle motion is a fractional Gaussian, stationary increment process [23]. In this chapter, we propose a particle path-based general protocol for the testing of fluid heterogeneity that builds upon the $\widehat{\text{MSD}}$'s asymptotic distribution. The latter already accounts for intra-path correlation and thus can be used to estimate the parameter vector $\boldsymbol{\xi}$ from a single observed path. The protocol tests heterogeneity in two different experimental situations, namely, when there are (i) multiple paths from the same fluid sample; (ii) ensembles of paths from different fluids. To illustrate the use of the protocols, we make inferences on fluid viscoelasticity with data from the David B. Hill Lab (UNC-Chapel Hill) on biofilm eradication, as first reported and described in [72].

The chapter is organized as follows. In Section 3.2, assuming a single observed path of realistic length, we propose an $\widehat{\text{MSD}}$ -based estimation protocol that displays improved performance in terms of bias and variance by comparison to common $\widehat{\text{MSD}}$ -based protocols. In Section 3.3, assuming multiple observed paths, we draw upon the protocol proposed in Section 3.2 to develop intra- and interfluid heterogeneity hypotheses tests. In Section 3.4, we investigate the performance of the ensemble $\widehat{\text{MSD}}$ -based estimator in under two sampling methods: measure more paths, or measure longer paths. In Section 3.5, we test the procedures put forward in Section 3.3 by applying them to experimental data. All proofs can be found in Appendix B.

3.2 Improved pathwise MSD-based estimation

In the analysis of fluid heterogeneity, it is especially important to make efficient use of the information contained in each particle path data set. For this reason, in this

section we assume a single observed path

$$X_1, X_2, \dots, X_n, \quad n \in \mathbb{N}, \quad (3.1)$$

is available, and that the underlying stochastic process $\{X(t)\}_{t \in \mathbb{R}}$ satisfies assumption (A1) and (A2) in Chapter 2.

Fitting (2.4) by means of ordinary least squares (OLS) regression is the most intuitive way of constructing an estimator of the diffusion parameter vector $\boldsymbol{\xi}$ (see (2.1)). This roughly corresponds to the common practice in the biophysical literature, both in experimental and methodological work [14, 46, 49, 83]. Throughout this chapter, $\widehat{\boldsymbol{\xi}}_{\text{OLS}} = (\widehat{\log \theta}_{\text{OLS}}, \widehat{\alpha}_{\text{OLS}})$ denotes this standard estimator. However, $\widehat{\boldsymbol{\xi}}_{\text{OLS}}$ has at least two significant shortcomings: finite sample bias and suboptimal performance in the presence of correlation among the regression disturbance terms.

We propose an improved estimation protocol that addresses these issues. It involves

- (i) a finite sample bias correction in the regression system of equations. In statistical problems involving scaling, it is well known that estimation bias is present even if the scaling relation in (2.1) holds as an equality. This is a consequence of the elementary fact that the logarithm of the ensemble average and the ensemble average of the logarithm are generally distinct (e.g., [62–64, 85]). In the context of (2.4), this means that

$$\mathbb{E} \log \bar{\mu}_2(h) \neq \log \mathbb{E} \bar{\mu}_2(h) = \alpha \log h + \log \theta, \quad h \in \mathbb{N}.$$

Let $h_k = w_k h$, $w_1 < w_2 < \dots < w_m$ be the regression lag values chosen by the analyst, where $h \in \mathbb{N}$. By reinterpreting $\log \bar{\mu}_2(h)$ itself as an estimator of $\alpha \log h + \log \theta$, we can reexpress its bias as a function of the diffusion exponent α , the path length n and the lag parameter h . This is done in the following

theorem.

Theorem 3.2.1. *For $0 < \alpha < 2$ and a finite n ,*

$$\begin{aligned} \mathbb{E} \log \bar{\mu}_2(h) - \alpha \log h - \log \theta &= -\frac{1}{4n} \sum_{i=-n+1}^{n-1} \left(1 - \frac{|i|}{n}\right) \times \\ &\times \left(\left| \frac{i}{h} + 1 \right|^\alpha - 2 \left| \frac{i}{h} \right|^\alpha + \left| \frac{i}{h} - 1 \right|^\alpha \right)^2 + O(h^{-\delta}) + o(\varrho^2(h, n, \alpha)), \end{aligned} \quad (3.2)$$

for some $\delta > 0$ (see (2.17)) and

$$\varrho(h, n, \alpha) = \begin{cases} (\frac{h}{n})^{1/2}, & 0 < \alpha < 3/2; \\ (\frac{h \log n}{n})^{1/2}, & \alpha = 3/2; \\ (\frac{h}{n})^{2-\alpha}, & 3/2 < \alpha < 1. \end{cases} \quad (3.3)$$

The term of $O(h^{-\delta})$ on the right hand side of (3.2) is determined by the short memory (high-frequency) component of the spectral density of the diffusion process, where the short memory effect can be observed in the \log_2 - \log_2 plots of MSD curves vs lag values. FIG. 3.1(a) is a plot of simulated ifOU ($\alpha = 0.6$) processes and FIG. 3.1(b) is a plot of experimental data (CSO2-NO at 8 mg ml⁻¹). In FIG. 3.1(b), when we choose small lag values to regress the slope of the \log_2 of MSD curve, $\hat{\alpha}$ is 1.42 where the true α is 0.6. If we use larger lag values, the estimate is 0.76, i.e., much closer to 0.6. FIG. 3.1(b) shows a similar trend, that is, the use of large lag values reduces the bias of $\hat{\alpha}$. On the other hand, the lag values cannot be chosen too large by comparison to the path length n . Otherwise, this leads to the inclusion of $\widehat{\text{MSD}}$ terms $\bar{\mu}_2(h)$ with large variance due to the small number of terms in the average.

- (ii) an optimal regression procedure in the presence of correlated errors. In any regression problem with correlated errors, a GLS (generalized least squares) procedure is expected to outperform OLS in terms of mean squared error (MSE)

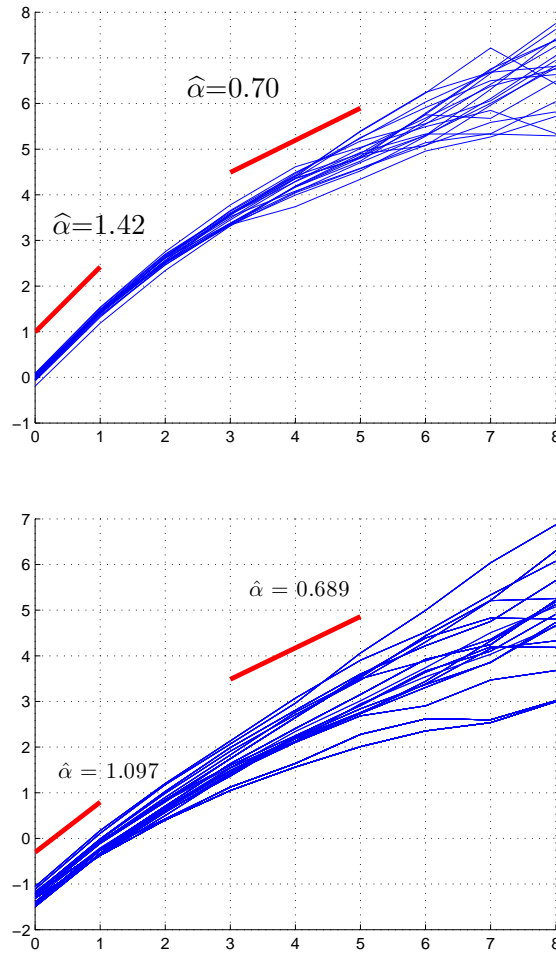


Figure 3.1: $\log_2 - \log_2$ plots of MSD curves as a function of the lags. (a) 20 ifOU paths with length 2^{11} , the diffusion exponent is $\alpha = 0.6$; (b) 20 particle paths with length 1800 were randomly selected from the *P. aeruginosa* biofilm after treatment with COS2-NO at concentration level 8 mg ml^{-1} . The first red line in each plot indicates the fitted slope over small lag values $h = 1, 2$. The estimated diffusion exponent $\hat{\alpha}$ is 1.42 in (a) and 1.10 in (b). The second red line in each plot indicates the fitted slope over large lag values $h = 8, 32$. The estimated diffusion exponent $\hat{\alpha}$ is 0.74 in (a) and 0.69 in (b).

[17]. Therefore, let

$$\widehat{\boldsymbol{\xi}}_{\text{GLS}} = (\widehat{\log \theta}_{\text{GLS}}, \widehat{\alpha}_{\text{GLS}})^T = (X^T \widetilde{\Sigma}^{-1} X)^{-1} X^T \widetilde{\Sigma}^{-1} \mathbf{y}, \quad (3.4)$$

where

$$X = \begin{pmatrix} 1 & \log h_1 \\ 1 & \log h_2 \\ \vdots & \vdots \\ 1 & \log h_m \end{pmatrix}$$

is the so-called design matrix and

$$\widetilde{\Sigma}(\alpha) = (\widetilde{\sigma}_{k_1, k_2})_{k_1, k_2=1, \dots, m} \quad (3.5)$$

is the unknown finite sample covariance matrix of $\mathbf{y} = \left(\log(\bar{\mu}_2(h_k)) \right)_{k=1, \dots, m}^T$.

In addition, $\widehat{\boldsymbol{\xi}}_{\text{GLS}}$ is the best linear unbiased estimator with covariance matrix

$$\Sigma_{\widehat{\boldsymbol{\xi}}_{\text{GLS}}}(\alpha) = (X^T \widetilde{\Sigma}^{-1}(\alpha) X)^{-1}. \quad (3.6)$$

An estimator of each entry $\widetilde{\sigma}_{k_1, k_2}$, $k_1, k_2 = 1, \dots, m$ of the matrix $\widetilde{\Sigma}$ is given by the following theorem.

Theorem 3.2.2. *For $0 < \alpha < 2$ and a finite n ,*

$$\begin{aligned} \widetilde{\sigma}_{k_1, k_2} = & \frac{1}{2n} \sum_{i=-n+1}^{n-1} \left(1 - \frac{|i|}{n} \right) \left\{ \left| \frac{i}{\sqrt{h_{k_1} h_{k_2}}} + \sqrt{\frac{h_{k_1}}{h_{k_2}}} \right|^\alpha - \right. \\ & - \left| \frac{i}{\sqrt{h_{k_1} h_{k_2}}} + \sqrt{\frac{h_{k_1}}{h_{k_2}}} - \sqrt{\frac{h_{k_2}}{h_{k_1}}} \right|^\alpha - \left| \frac{i}{\sqrt{h_{k_1} h_{k_2}}} \right|^\alpha + \\ & + \left. \left| \frac{i}{\sqrt{h_{k_1} h_{k_2}}} - \sqrt{\frac{h_{k_2}}{h_{k_1}}} \right|^\alpha \right\}^2 + O(h^{-\delta}) + o(\varrho^2(h, n, \alpha)). \end{aligned} \quad (3.7)$$

However, the presupposes knowledge of the covariance matrix $\tilde{\Sigma}(\alpha)$ in (3.5), is a condition not met in our framework. An approximation to $\tilde{\Sigma}(\alpha)$, which in turn can be estimated by $\tilde{\Sigma}(\hat{\alpha}_{\text{OLS}})$. Hence, an improved estimator can be constructed by GLS based on $\tilde{\Sigma}(\hat{\alpha}_{\text{OLS}})$ and the bias-corrected regression equation system. Hereinafter, the resulting estimator will be denoted by

$$\hat{\xi} = (\widehat{\log \theta}, \hat{\alpha}). \quad (3.8)$$

Its construction is schematically displayed below in the form of pseudo-code.

Generating the improved pathwise estimator $\hat{\xi}$
<p>Input:</p> <ul style="list-style-type: none"> • one observed particle path (see (3.1)) of length n; • regression lag values $h_k, k = 1, \dots, m$ (typically, $h_k = hw_k, w_1 \leq \dots \leq w_m, h = \sqrt{n}$); • the expression for the asymptotic covariance matrix $\tilde{\Sigma}(\alpha)$ as a function of α; <p>Step 1: obtain the estimator $\hat{\alpha}_{\text{OLS}}$ by means of OLS over the chosen lag values;</p> <p>Step 2: use $\hat{\alpha}_{\text{OLS}}$ to correct the bias of $\log \bar{\mu}_2(\cdot)$ (see (3.2));</p> <p>Step 3: estimate the asymptotic covariance matrix $\tilde{\Sigma}(\alpha)$ by means of $\tilde{\Sigma}(\hat{\alpha}_{\text{OLS}})$;</p> <p>Step 4: obtain the estimator $\hat{\xi}$ by means of $\tilde{\Sigma}(\hat{\alpha}_{\text{OLS}})$-based GLS on the bias-corrected regression equation.</p>

To compare the performance of $\hat{\xi}$ and $\hat{\xi}_{\text{OLS}}$, we generate 1000 independent paths of length 2^{10} and estimate the diffusion exponent based on the two methods. FIG. 3.2 displays the results in terms of Monte Carlo bias, standard deviation and MSE. The improved estimator $\hat{\xi}$ outperforms the usual estimator $\hat{\xi}_{\text{OLS}}$ by any of the three

criteria.

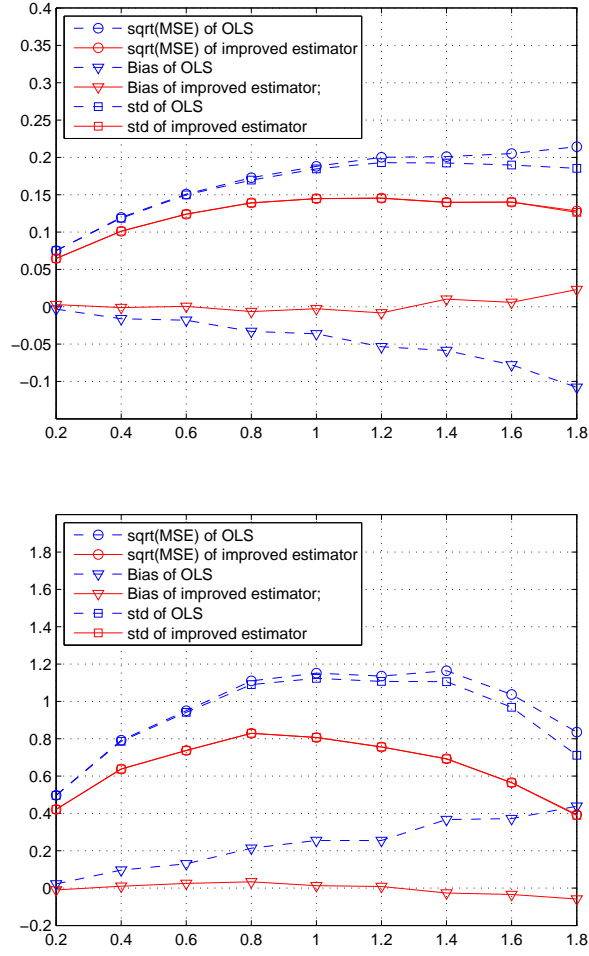


Figure 3.2: x -axis: α . Bias, standard deviation, and MSE of the estimators $\hat{\xi}$ and $\hat{\xi}_{\text{OLS}}$ as a function of diffusion exponents. For comparison, solid and dashed lines represent $\hat{\xi}$ and $\hat{\xi}_{\text{OLS}}$, respectively. The estimator $\hat{\xi}$ has smaller bias, standard deviation, and MSE than $\hat{\xi}_{\text{OLS}}$.

To study the effect of bias correction on the performance of the estimator $\hat{\xi} = (\widehat{\log \theta}, \hat{\alpha})$, we generated 1000 independent paths of length 2^{10} and estimated the diffusion exponent α in the presence or absence of bias correction. The results are shown in FIG. 3.3. As expected, the bias of $\hat{\alpha}$ is smaller after bias correction. In addition, it is near zero for all values of the parameter α , with the exception of the range of extreme superdiffusivity ($\alpha = 1.8$).

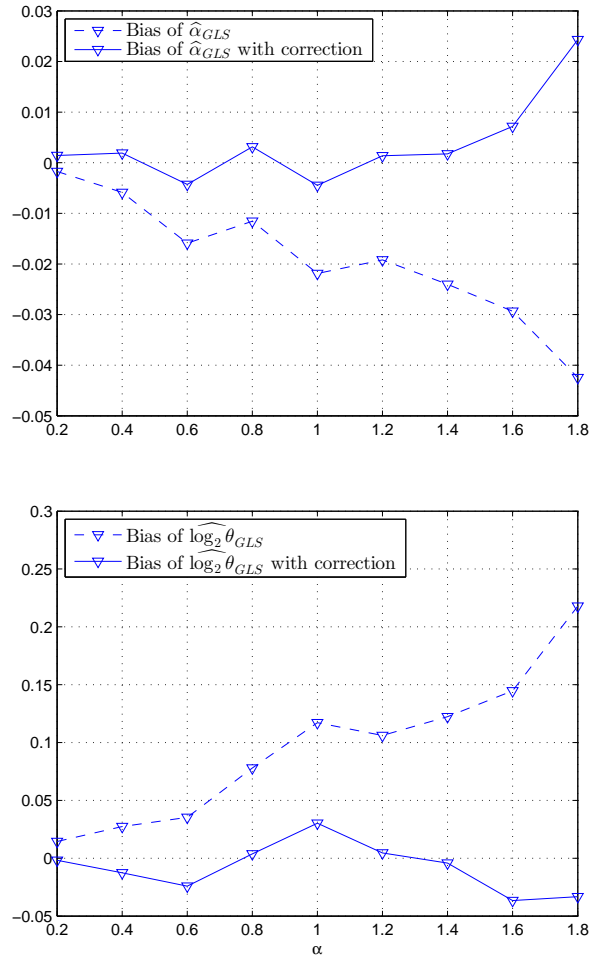


Figure 3.3: x -axis: α . The effect of the bias correction procedure in the performance of the estimator $\hat{\xi}$ as a function of the diffusion exponent α . For comparison, the latter is also considered without bias correction. The version of the estimator with bias correction has a significantly smaller bias than that without bias correction.

3.3 Testing heterogeneity

Single particle tracking experiment often generates multiple particle paths. Given data from the latter, fluid heterogeneity can be tested in at least two senses. First, assuming local physical homogeneity, whether different regions of the fluid are heterogeneous; second, assuming global physical homogeneity of each fluid sample, whether two samples from each fluid are heterogeneous. Hereinafter, we will call these two senses intra- and interfluid heterogeneity, respectively.

Tables 3.1, 3.2 and 3.3 display the proposed framework: namely, for each type of fluid heterogeneity, they show the appropriate hypotheses and the proposed test statistics and test rejection regions.

Table 3.1: Hypotheses

heterogeneity	H_0	H_a
intrafluid	$\xi_1 = \cdots = \xi_\nu$	$\xi_i \neq \xi_j$ for some $1 \leq i, j \leq \nu$
interfluid	$\xi_I = \xi_{II}$	$\xi_I \neq \xi_{II}$

Table 3.2: Test statistics

heterogeneity	test statistic
intrafluid	S_1^2, S_2^2 (see (3.10))
interfluid	T_1, T_2 (see (3.16))

Table 3.3: Tests

heterogeneity	rejection	parameter	sample
	region	range	size
intrafluid	R_{intra} (see (3.11))	$0 < \alpha \leq 3/2$	all ν
intrafluid	R_{intra} (see (3.15))	$3/2 < \alpha < 2$	large ν
interfluid	R_{inter} (see (3.17))	$0 < \alpha < 2$	large ν_I, ν_{II}

Consider first the setting of intrafluid homogeneity testing, i.e., suppose $\nu \in \mathbb{N}$ bead diffusion paths of length n from a single fluid sample are available. If the fluid is physically homogeneous, then it is expected to generate particle paths with (nearly) identical parameter values $\boldsymbol{\xi}$. The alternative is that for some pair i, j , $\boldsymbol{\xi}_i \neq \boldsymbol{\xi}_j$. These two possibilities, labeled H_0 and H_a , respectively, are described in the row “intrafluid” in Table 3.1.

Starting from the ν particle paths, let $\widehat{\boldsymbol{\xi}}_i$, $i = 1, \dots, \nu$, be the vector-valued estimators put forward in Section 3.2. Also, let

$$\widehat{\boldsymbol{\zeta}}_i = \begin{pmatrix} \widehat{\zeta}_{i,1} \\ \widehat{\zeta}_{i,2} \end{pmatrix} = \Sigma_{\widehat{\boldsymbol{\xi}}_{\text{GLS}}}^{-1/2}(\alpha) \widehat{\boldsymbol{\xi}}_i, \quad i = 1, \dots, \nu, \quad (3.9)$$

where $\Sigma_{\widehat{\boldsymbol{\xi}}_{\text{GLS}}}^{-1/2}(\alpha)$ is given by (3.6). Then, for every i , the two components $\widehat{\zeta}_{i,1}$ and $\widehat{\zeta}_{i,2}$ are standardized and uncorrelated. Let

$$S_j^2 = \frac{1}{\nu - 1} \sum_{i=1}^{\nu} (\widehat{\zeta}_{i,j} - \bar{\widehat{\zeta}}_{\cdot,j})^2, \quad j = 1, 2 \quad (3.10)$$

be the sample variances of $\{\widehat{\zeta}_{i,1}\}_{i=1,\dots,\nu}$ and $\{\widehat{\zeta}_{i,2}\}_{i=1,\dots,\nu}$, which converge to ν independent and identically distributed normal random variables at rate $\sqrt{\frac{n}{h}}$ when $0 < \alpha \leq 3/2$, or at rate $\sqrt{\frac{n}{\log(n)h}}$ when $\alpha = 3/2$ (see Corollary 2.3.1). Then, under H_0 , as $n \rightarrow \infty$, the random vector $\left((\nu - 1)S_1^2, (\nu - 1)S_2^2\right)$ approaches, in law, a vec-

tor of independent chi-squared distributions with $\nu - 1$ degrees of freedom. To test heterogeneity at significance level ϵ , we can use Bonferroni-type correction and reject the null hypothesis H_0 if

$$R_{\text{intra}} : (\nu - 1)S_1^2 > \chi_{\nu-1, \epsilon/2}^2 \text{ or } (\nu - 1)S_2^2 > \chi_{\nu-1, \epsilon/2}^2 \quad (3.11)$$

(see Tables 3.2 and 3.3, “intrafluid” rows). Since, in (3.9), $\Sigma_{\hat{\xi}}(\alpha)$ is also a function of the unknown parameter value α , we can replace the latter with its ensemble estimator $\hat{\alpha}_{\widehat{\text{EM}}}$, where we define the ensemble mean squared displacement (EMSD) by

$$\bar{\mu}_2(h)_{\widehat{\text{EM}}} = \frac{1}{\nu} \sum_{l=1}^{\nu} \bar{\mu}_2(h)_l, \quad (3.12)$$

and one can then produce $\widehat{\text{MSD}}$ -based ensemble estimators

$$\widehat{\xi}_{\widehat{\text{EM}}} = (\widehat{\log \theta}_{\widehat{\text{EM}}}, \hat{\alpha}_{\widehat{\text{EM}}}) \quad (3.13)$$

by performing the pseudo-code of generating improved estimator $\widehat{\xi}$. To investigate the accuracy of $\Sigma_{\hat{\xi}}(\hat{\alpha}_{\widehat{\text{EM}}})$ as an estimator of $\Sigma_{\hat{\xi}}(\alpha)$, we generated 1000 independent paths of length 2^{10} . FIG. 3.4 shows Monte Carlo variances as well as their estimates $\Sigma_{\hat{\xi}}(\hat{\alpha}_{\widehat{\text{EM}}})$ for several values of the parameter α . The estimates nearly perfectly match the Monte Carlo results in the subdiffusive range. A slight deviation appears in the superdiffusive range, but still within an acceptable margin.

When $\alpha > 3/2$, under H_0 the estimators $\widehat{\xi}_i$, $i = 1, \dots, \nu$, converge in distribution to ν independent and identically distributed non-Gaussian (Rosenblatt-type, see Corollary 2.3.1 and [74, 80, 82, 84]) random vectors at rate $(\frac{n}{h})^{2-\alpha}$, and so do the estimators $\widehat{\xi}_i$, $i = 1, \dots, \nu$. In this case, the marginal distributions of the uncorrelated vector (S_1^2, S_2^2) do not approach chi-squared distributions. However, if ν is large

enough (heuristically, $\nu \geq 30$),

$$S_1^2 \overset{\text{asympt}}{\sim} N\left(1, \frac{\kappa_1 - 1}{\nu}\right), \quad S_2^2 \overset{\text{asympt}}{\sim} N\left(1, \frac{\kappa_2 - 1}{\nu}\right), \quad (3.14)$$

where the parameter $\kappa_1 := \mathbb{E}[\widehat{\zeta}_{\cdot,1} - \mathbb{E}\widehat{\zeta}_{\cdot,1}]^4$, $\kappa_2 := \mathbb{E}[\widehat{\zeta}_{\cdot,2} - \mathbb{E}\widehat{\zeta}_{\cdot,2}]^4$ can be numerically approximated (see Table 3.2 and Appendix B.3). Thus, at significance level ϵ , we reject the null hypothesis H_0 if

$$R_{\text{intra}} : \frac{S_1^2 - 1}{\sqrt{(\kappa_1 - 1)/\nu}} > z_{\epsilon/2} \text{ or } \frac{S_2^2 - 1}{\sqrt{(\kappa_2 - 1)/\nu}} > z_{\epsilon/2} \quad (3.15)$$

(see Table 3.3).

To check the finite sample size (significance level) of the test, we conducted a Monte Carlo study with 50 simulated paths of length 2^{12} and recorded whether or not the null hypothesis H_0 is rejected at 0.05 significance level. This procedure was repeated 500 times, where the rejection rate was expected to be around 0.05. Since each outcome is a Bernoulli trial (reject or not H_0), the simulation rejection rate follows a binomial distribution with $n = 500$ and $p = 0.05$. Thus, a normal approximation to the 95% confidence interval of the rejection rate gives (0.031, 0.069). As shown in FIG. 3.5(a), the observed simulation rejection rate was around 0.05 and within a 95% confidence interval, as expected.

Now consider the setting of interfluid homogeneity testing, namely, suppose ν_I and ν_{II} paths, $\nu_I, \nu_{II} \in \mathbb{N}$, are obtained from two physically homogeneous fluid samples I and II, respectively. We are interested in testing whether the samples I and II are homogeneous, namely, whether or not particle diffusion in the samples displays the same underlying parameter value ξ . These two possibilities, labeled H_0 and H_a , respectively, are described in the row “interfluid” in Table 3.1.

Since multiple particle paths are available for each fluid sample, we can produce $\widehat{\text{MSD}}$ -based ensemble estimators $\widehat{\xi}_{\text{EM},I}$ and $\widehat{\xi}_{\text{EM},II}$ (see (3.12) and (3.13)). Their finite

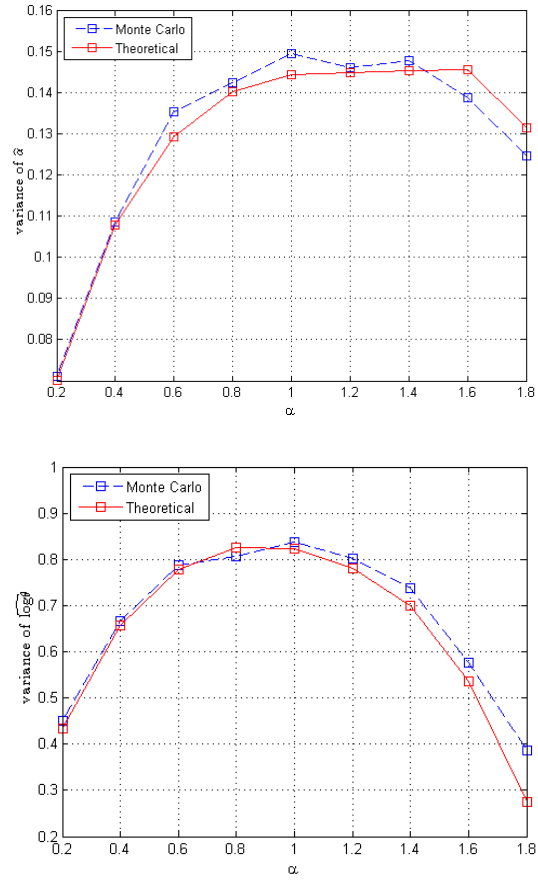


Figure 3.4: Monte Carlo standard error of the estimator $\hat{\alpha}$ (dashed line) and the estimator $\Sigma_{\hat{\alpha}}(\hat{\alpha}_{\text{EM}})$ (solid line) as a function of the diffusion exponent α . The latter closely matches the former, especially in the subdiffusive range $\alpha < 1$.

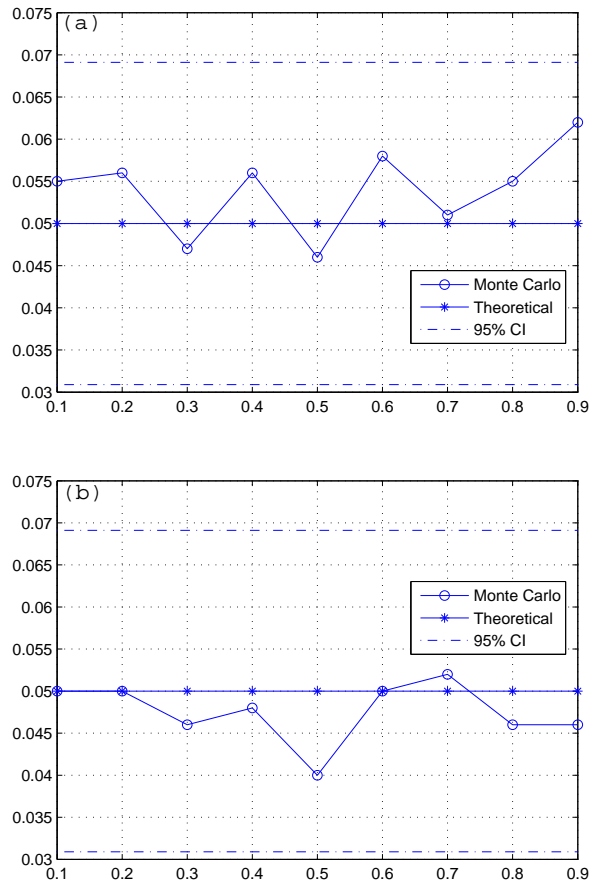


Figure 3.5: x -axis: $H = \alpha/2$. Rejection rate of (a) chi-square heterogeneity test of multiple particle paths from a fluid (b) z -test of multiple particle paths from two fluid samples, as a function of the diffusion exponent. For every value of α , each of 500 Monte Carlo runs consisted of generating 50 independent paths and conducting a test at 95% confidence level. The Monte Carlo rejection rate is very close to the theoretical value of 0.05.

sample covariance matrices are $\frac{1}{\nu_I} \Sigma_{\widehat{\boldsymbol{\xi}}_{\text{GLS}}}(\alpha_I)$ and $\frac{1}{\nu_{II}} \Sigma_{\widehat{\boldsymbol{\xi}}_{\text{GLS}}}(\alpha_{II})$, respectively. Define the standardized estimators

$$\widehat{\boldsymbol{\zeta}}_j = \begin{pmatrix} \widehat{\zeta}_{j,1} \\ \widehat{\zeta}_{j,2} \end{pmatrix} = \frac{1}{\nu_j} \Sigma_{\widehat{\boldsymbol{\xi}}_{\text{GLS}}}^{-1/2}(\widehat{\alpha}_{\text{EM},j}) \widehat{\boldsymbol{\xi}}_{\text{EM},j}, \quad j = \text{I, II}.$$

Let

$$T_1 = \frac{\widehat{\zeta}_{\text{I},1} - \widehat{\zeta}_{\text{II},1}}{\sqrt{2}}, \quad T_2 = \frac{\widehat{\zeta}_{\text{I},2} - \widehat{\zeta}_{\text{II},2}}{\sqrt{2}} \quad (3.16)$$

be the test statistics (see Table 3.2). This leads to the rejection region

$$R_{\text{inter}} : |T_1| > z_{\epsilon/4} \text{ or } |T_2| > z_{\epsilon/4} \quad (3.17)$$

(see Table 3.3). To check the test's accuracy, we produced a 500-run Monte Carlo study based on two sets of 50 paths with the same diffusion exponent, where tests were conducted at significance level 0.05. As shown in FIG. 3.5(b), the rejection rate was close to 0.05 and within a 95% confidence interval, as expected.

In FIG. 3.6, we investigate the interfluid test power as a function of the path lengths and number of paths. The x -axis represents the difference between the diffusion parameters from two fluids, namely, $\delta_\alpha = |\alpha_I - \alpha_{II}|$, whereas the y -axis is the test power at 0.05 significance level. In FIG. 3.6(a) and (b), the values of α are 0.2 and 1.0, respectively. It turns out that there is no visible difference between doubling the path lengths or the number of paths. On the other hand, in FIG. 3.6(c), α is 1.8, i.e., the underlying particles are strongly super-diffusive. In this case, doubling the number of paths increases the test power more than doubling the path lengths. This reflects the fact that, in this parameter range, the convergence rate of the $\widehat{\text{MSD}}$ -based estimators is slower than $\frac{1}{\sqrt{n}}$.

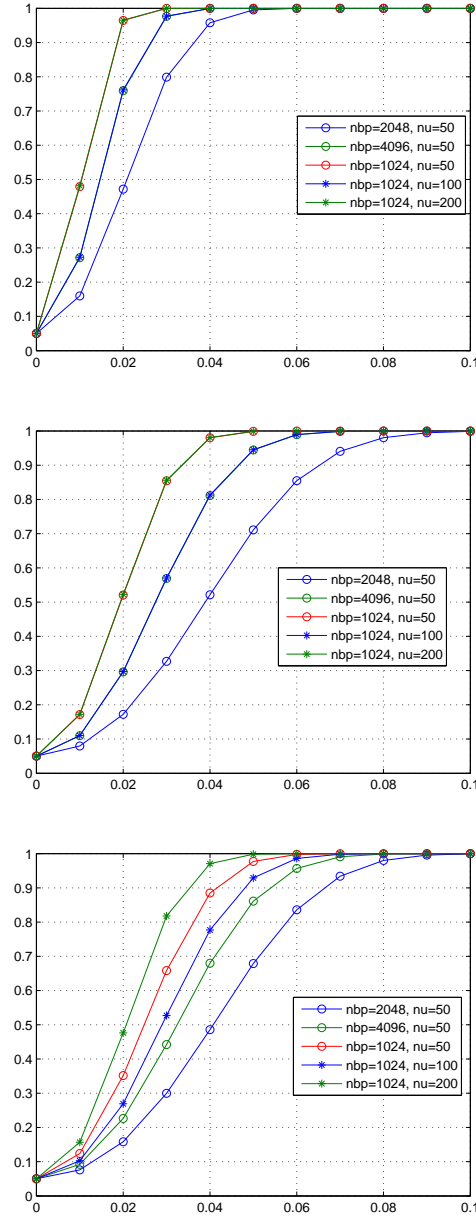


Figure 3.6: Interfluid test power v.s. $\delta_\alpha = |\alpha_I - \alpha_{II}|$. Top: $\alpha = 0.2$. Middle: $\alpha = 1.0$. Bottom: $\alpha = 1.8$

3.4 Ensemble estimation: more or longer paths?

The framework put forward in Sections 3.2 and 3.3 allows addressing the following issue. Suppose the fluid is locally homogeneous, and that data storage or cost restrictions are in place. Should experimentalists record the diffusion of a larger number of particles (ν) over a fixed period of time (hence, keeping constant the average sample path data length n), or should they record the same number of particles ν over a longer period of time (hence, yielding a larger average n)? We tackle this problem by investigating the performance of the ensemble estimator in (3.13) in terms of bias and variance. Recall that

$$\mathbb{E}\bar{\mu}_2(h)_{\widehat{\text{EM}}} = \mathbb{E}\bar{\mu}_2(h) = \mathbb{E}X(h)^2,$$

$$\text{Var}\bar{\mu}_2(h)_{\widehat{\text{EM}}} = \frac{1}{\nu}\text{Var}\bar{\mu}_2(h).$$

By a similar reasoning to the one leading to (3.2), the bias of $\log_2 \bar{\mu}_2(h)_{\widehat{\text{EM}}}$ as an estimator of $\alpha \log h + \log \theta$ is given by

$$O(h^{-\delta}) - \mathbb{E}\left[\frac{(\bar{\mu}_2(h)_{\widehat{\text{EM}}} - \mathbb{E}X^2(h))^2}{2\mathbb{E}X^2(h)^2}\right] = O(h^{-\delta}) - \frac{1}{\nu}\mathbb{E}\left[\frac{(\bar{\mu}_2(h) - \mathbb{E}X^2(h))^2}{2\mathbb{E}X^2(h)^2}\right],$$

and the variance of $\log \bar{\mu}_2(h)_{\widehat{\text{EM}}}$ is, approximately,

$$\frac{1}{\nu}\text{Var}(\log \bar{\mu}_2(h)).$$

Then, the question asked above boils down to: in order to improve the performance of the ensemble estimator in terms of bias, standard deviation and square root MSE, should we use larger ν (hereinafter Method I), or, larger n (hereinafter Method II)? To answer this question, we will compare these two methods under same total number of observations (i.e., the total number of data points $\nu \times n$ recorded). Therefore, we

use the following rule:

1. start at the same initial status: 16 paths of length 256 for each method, run 500 Monte Carlo simulations to get the bias, standard deviation and square root MSE of $\hat{\alpha}$ for Method I and II;
2. for Method I, fix the paths length, generate 4 times number of paths and redo the Monte Carlos experiments;
3. for Method II, fix the number of paths, get paths of 4 times length, meanwhile, multiply all lags by 2 and redo the Monte Carlos experiments;

For example, in step 1, both methods make use of a total of $16 \times 256 = 4096$ observations. In step 2, we generate $4 \times 16 = 64$ paths for Method I, which yields a total of $64 \times 256 = 16384$ observations. To keep the balance, we generate 16 paths of length $4 \times 256 = 1024$ for Method II. Therefore, Method II also draws upon $16 \times 1024 = 16384$ observations.

We compare the results in FIG. 3.7. In FIG. 3.7(a) and (b), the diffusion exponent is set to $\alpha = 0.6$ and 1.0 , respectively. Method II has smaller bias and square root MSE. The reason is that, when ν is large enough, $O(h^{-\delta})$ dominates the bias. Thus, increasing the number of paths ν does not reduce the bias. However, increasing the path length n means that the $\widehat{\text{MSD}}$ terms $\bar{\mu}_2(h)$ with larger lag values h can be used in the regression procedure. This implies a reduction in the magnitude of the term $O(h^{-\delta})$, and hence, smaller bias. Method I displays smaller standard deviation because a 4-fold increase in ν reduces the standard error by a factor of $1/2$. Meanwhile, for Method II, the standard deviation is proportional to $\sqrt{h/n}$. By multiplying n by 4 and h by 2, the standard error is reduced by a factor of $1/\sqrt{2}$.

For FIG. 3.7(c), the diffusion exponent is set to $\alpha = 1.8$. Method II still shows a smaller bias by comparison to Method I, as expected. However, since δ increases as a function of α , then $O(h^{-\delta})$ shrinks with α . Therefore, the component $O(h^{-\delta})$

carries less weight in the estimator's bias for the superdiffusive case than for the subdiffusive case. Since $\alpha > 3/2$, by (2.33), the standard deviation for Method II is proportional to $(h/n)^{2-\alpha} = (n/h)^{0.2}$. Thus, again assuming a 4-fold increase in n and a 2-fold increase in h , the standard error is reduced by a factor of $1/2^{0.2}$, which is much slower than the standard error reduction factor of $1/2$ for Method I. These two reasons explain why Method I displays smaller square root MSE than Method II.

3.5 Application in experimental data: heterogeneity of treated *P. aeruginosa* biofilms

The David B. Hill Lab (UNC-Chapel Hill) produced data from experiments on disruption and eradication of *P. aeruginosa* biofilms using nitric oxide-releasing chitosan oligosaccharides [72].

We now provide a brief description of the experiments. Cystic fibrosis (CF) lung disease is caused by defective chloride transport, resulting in thickened, dehydrated mucus. The latter restricts bacterial motility and promotes *P. aeruginosa* biofilm formation. Inhaled tobramycin is currently the only antibiotic recommended for the treatment of both initial and chronic *P. aeruginosa* infections in patients with CF. While inhaled tobramycin is effective at eradicating bacteria within biofilms, it fails to physically remove the structural remnants of the biofilm from the airways. This may cause biofilm regrowth and the development of antibiotic-resistant infections. Therefore, an ideal anti-biofilm therapeutic for CF would both eradicate bacteria and physically degrade the biofilm, facilitating clearance from the airway.

On the other hand, nitric oxide (NO) is an endogenously produced diatomic free radical with significant antibacterial activity against *P. aeruginosa* biofilms. Furthermore, atomic force microscopy has revealed that NO exposure causes structural damage to the membranes of planktonic Gram-negative bacteria, including *P. aerugi-*

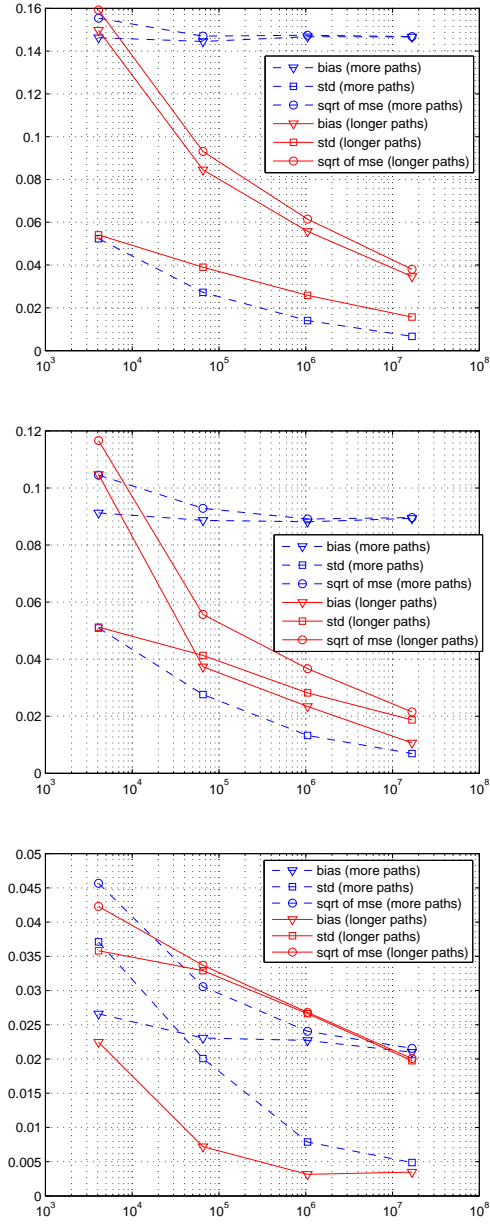


Figure 3.7: Bias, standard deviation, and MSE of two sampling strategies: more paths v.s. longer paths. x -axis: total number of data points. Top: $\alpha = 0.6$. Middle: $\alpha = 1.0$. Bottom: $\alpha = 1.8$.

nosa. The interest lies in the utility of NO-releasing chitosan oligosaccharides to both eradicate and physically alter *P. aeruginosa* biofilms and in comparing its effect with tobramycin.

Sample heterogeneity has been correlated with increased viscoelasticity for complex biological materials such as sputum [19]. In the experiments, the effect of antibacterial treatment on biofilm heterogeneity was thus evaluated at different concentrations. In Table 3.4, we use the data to test the intrafluid heterogeneity of *P. aeruginosa* biofilms after the tobramycin treatments at concentrations levels 25, 50, 100, 200, and 400 $\mu\text{g ml}^{-1}$. From each of these fluid samples, we randomly select 100 paths, where all paths have length $n = 1800$. An application of the intrafluid test (3.11) produces strong evidence (negligible p -values) of intrafluid heterogeneity in every sample. This conclusion matches those reported in [72]. Since no homogeneous fluid samples are detected from any of these five samples, we do not perform the interfluid heterogeneity test (3.17).

In Table 3.5, we apply (3.11) in the testing of intrafluid heterogeneity of *P. aeruginosa* biofilms after COS2-NO treatment at concentration levels 1, 2, 4, 8, and 16 mg ml^{-1} . As before, 100 paths of length 1800 were randomly selected for each concentration level. At concentrations 1 or 2 mg ml^{-1} , the p -values are still less than machine error, which indicates a strongly significant heterogeneity. As the concentration levels increases to 4 and 8 mg ml^{-1} , the p -values also increase. When the concentration is 16 mg ml^{-1} , the p -value is 0.18 and we fail to reject the null hypothesis of intrafluid homogeneity. This provides evidence that the COS2-NO treatment is effective in the eradication of *P. aeruginosa* biofilms. Once again, this confirms the conclusions reported in [72]. In Table 3.6, by applying (3.17), we test the interfluid heterogeneity of *P. aeruginosa* biofilms after COS2-NO treatment at concentration 16 mg ml^{-1} . From each fluid sample (A, B and C), we select 100 paths of length 1800 and conduct the test. As we can see, there is no evidence whatsoever that the fluid samples A, B and

Tobramycin ($\mu\text{g ml}^{-1}$)	p -value
25	<e-16 ***
50	<e-16 ***
100	<e-16 ***
200	<e-16 ***
400	<e-16 ***

Table 3.4: Intrafluid biofilm heterogeneity testing after treatment with tobramycin at concentration levels 25, 50, 100, 200, 400 $\mu\text{g ml}^{-1}$. 100 independent paths of length 1800 are randomly selected for each concentration level.

COS2-NO (mg ml^{-1})	p -value
1	<e-16 ***
2	<e-16 ***
4	2e-13 ***
8	3e-09 ***
16	0.18

Table 3.5: Intrafluid biofilm heterogeneity testing after treatment with COS2-NO at concentration levels 1, 2, 4, 8, 16 mg ml^{-1} . 100 independent paths of length 1800 are randomly selected for each concentration level.

C are heterogeneous.

3.6 Conclusion

In this chapter, we draw upon results from Chapter 2 on the asymptotic distribution of the mean squared displacement in single particle experiments to propose a framework for fluid heterogeneity testing. It is assumed that the observed anomalously diffusive particle follows a fractional Gaussian, stationary increment process. The assumptions on particle behavior cover a broad family of processes which includes fractional Brownian motion as well as processes with non-fractional high-frequency

COS2-NO 8 mg ml^{-1}	p -value
Group A v.s. Group B	0.9996
Group A v.s. Group C	0.9998
Group B v.s. Group C	0.9999

Table 3.6: Interfluid biofilm heterogeneity testing after treatment with COS2-NO at concentration level 16 mg ml^{-1} . 100 independent paths of length 1800 are randomly selected for each group.

behavior, such as the (integrated) fractional Ornstein-Uhlenbeck.

By building upon an MSD-based improved estimation in terms of bias and variance comparing to the common OLS estimation, we propose an encompassing methodology for the testing of intra- and inter-sample heterogeneity, namely, whether different regions of the fluid are heterogeneous, or samples from two fluids are heterogeneous. The proposed methodology covers the situations where just one or multiple observed paths are available, and its focus is on the estimation of the diffusivity coefficient and the diffusion exponent of the underlying anomalous diffusion. The testing methods allowed us to investigate and provide more accurate quantitative analysis of the experimental data from the David B. Hill Lab (UNC-Chapel Hill) and the results reported in [72] were generally confirmed.

Chapter 4

The asymptotic distribution of a mixing estimator

4.1 Introduction

The foundation of statistical physics is the ergodic hypothesis, i.e., the idea that the phase average of an physical quantity (the theoretical value) may be compared with its infinite time average (the experimental values of an long trajectory). Under this hypothesis, physicists can observe a long trajectory instead of an ensemble of independent realizations. It was L. Boltzmann who introduced the ergodic hypothesis and called the Boltzmann ergodic hypothesis (see [2]). In 1949, Khinchin proposed a new approach (see [40]) that related the ergodicity of a physical system to the irreversibility of its correlation function.

To test the ergodicity of a stochastic process, one can calculate the ensemble averages and compare them with the temporal average given multiple realizations. However, it maybe impossible to measure multiple trajectories of a stochastic process due to some unrepeatable characteristic of the experiment. Therefore, one needs to develop a testing method based on a single trajectory. Another way to test ergodicity is to test the stronger mixing properties [51]. So, consider a stochastic process

$Y = \{Y(n)\}_{n \in \mathbb{N}}$. Y can be represented as a probability measure \mathbb{P} on the space (Ω, \mathcal{B}) . Here, Ω is the phase space of all the functions $f : \mathbb{N} \rightarrow \mathbb{R}$ and \mathcal{B} is the σ -algebra of events [35]. The probability space $(\Omega, \mathcal{B}, \mathbb{P})$, together with the usual shift transformation $S : \Omega \rightarrow \Omega$, $S[f(n)] = f(n+1)$, is a standard dynamical system that fully describes the evolution in time of the process Y . We say that the dynamical system $(\Omega, \mathcal{B}, \mathbb{P}, S)$ is mixing, or equivalently, the process Y is mixing if

$$\lim_{n \rightarrow \infty} \mathbb{P}[A \cap S^n(B)] = P(A)P(B) \quad (4.1)$$

for all $A, B \in \mathcal{B}$. Here, S^n denotes n -fold composition of S . Thus, mixing can be interpreted as the asymptotic independence of the sets A and B under the transformation S .

In [50], an estimator for testing the mixing property of single trajectory is proposed. It is based on the statistic

$$\widehat{E}(n) = \frac{1}{N-n+1} \sum_{k=0}^{N-n} \exp\{i[Y(n+k) - Y(k)]\} - \left| \frac{1}{N+1} \sum_{k=0}^N \exp\{iY(k)\} \right|^2, \quad (4.2)$$

where $n \in \mathbb{Z}$ and $n < N$. Furthermore, if $Y(k)$ is a mixing stochastic process, the author argue that, when n is large enough,

$$\widehat{E}(n) \approx 0. \quad (4.3)$$

It should be emphasized that (4.3) is only a necessary and not a sufficient condition for mixing. This means that we can only disprove mixing if we have one trajectory of a random process. Note that $\widehat{E}(n)$ can be written as $\widehat{E}_1(n) + \widehat{E}_2(n)$, where

$$\widehat{E}_1(n) = \frac{1}{N-n+1} \sum_{k=0}^{N-n} \cos(Y(n+k) - Y(k)) + i \frac{1}{N-n+1} \sum_{k=0}^{N-n} \sin(Y(n+k) - Y(k)), \quad (4.4)$$

$$\widehat{E}_2(n) = \left| \frac{1}{N+1} \sum_{k=0}^N \cos(Y(k)) \right|^2 + \left| \frac{1}{N+1} \sum_{k=0}^N \sin(Y(k)) \right|^2. \quad (4.5)$$

In this chapter, our goal is to establish the asymptotic distribution of $\widehat{E}(n)$, which has remained an open problem in the biophysical literature. We assume that the underlying stochastic process is a stationary Gaussian process whose covariance function γ_Y satisfying a decay condition of the type

$$\gamma_Y(k) \sim Ck^{\alpha-2}, \quad k \rightarrow \infty,$$

for some constant C . This is a key step in the construction of an asymptotically valid test for the hypothesis of the mixing property. Our investigation is based on setting $G(x)$ to $\cos(x)$ or $\sin(x)$ and developing the limiting distribution of the partial sum

$$\frac{1}{n} \sum_{m=1}^n G(Y(m)) - \mathbb{E}G(Y(m)).$$

The difficulty of this problem, i.e., in establishing the asymptotic distribution of the partial sum, lies in the presence of fractional memory. The latter may give rise to unconventional convergence rates and non-Gaussian distributions.

The chapter is organized as follows. Section 4.2 contains a review of non-central and central limit theorems. In Section 4.3, we establish the limiting distributions of $\widehat{E}_1(n)$, $\widehat{E}_2(n)$ and $\widehat{E}(n)$. All proofs can be found in Appendix C.

4.2 Non-central and central limit theorems

In this section, we quote some results from [69] that will be used throughout the chapter.

Definition 4.2.1. The *Hermite polynomial* of order 0 is $H_0(x) = 1$. The *Hermite*

polynomial of order n , $n \in \mathbb{N}$, is defined by

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}, \quad x \in \mathbb{R}.$$

In particular,

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1.$$

Let

$$\phi(dx) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx$$

be the probability measure on \mathbb{R} associated with a standard normal variable X and let $L^2(\phi)$ be the space of measurable, square-integrable function with respect to $\phi(dx)$. Then, $G \in L^2(\mathbb{R}, \phi)$ if and only if $\mathbb{E}G^2(X) < \infty$. Furthermore, G has a series expansion in Hermite polynomials

$$G(x) = \sum_{n=0}^{\infty} g_n H_n(x),$$

where

$$g_k = \frac{1}{k!} (G, H_k)_{L^2(\mathbb{R}, \phi)} \quad (4.6)$$

(see [69]).

Definition 4.2.2. Let $G \in L^2(\mathbb{R}, \phi)$ and let g_k , $k \geq 0$, be the coefficients in its Hermite expansion. The *Hermite rank* r of G is defined as the smallest index $k \geq 1$ for which $g_k \neq 0$, that is,

$$r = \min\{k \geq 1 : g_k \neq 0\}.$$

Example 4.2.1. Note that

$$\int_{\mathbb{R}} \sin(x) H_1(x) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \frac{1}{\sqrt{e}},$$

$$\int_{\mathbb{R}} \cos(x) H_1(x) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = 0, \quad \frac{1}{2!} \int_{\mathbb{R}} \cos(x) H_2(x) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = -\frac{1}{2\sqrt{e}}, \quad (4.7)$$

where the first equality in $e : mixingcosh1h2$ follows from the fact that $\cos(x)$ is an even function. Consequently, the Hermite ranks of the functions $\sin(x)$ and $\cos(x)$ are 1 and 2 respectively. Hereinafter,

$$g_{\cos,n} = \frac{1}{n!} \int_{\mathbb{R}} \cos(x) H_n(x) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx, \quad g_{\sin,n} = \frac{1}{n!} \int_{\mathbb{R}} \sin(x) H_n(x) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx, \quad (4.8)$$

will denote the coefficients of the series expansions of $\cos(x)$ and $\sin(x)$, respectively, in Hermite polynomials.

Let $(X, Y)^T$ be a Gaussian vector. The following proposition allows us to calculate the covariance between $G_1(X)$ and $G_2(Y)$, where G_1 and G_2 are suitable transformations.

Proposition 4.2.1. *Let $(X, Y)^T$ be a Gaussian vector with $\mathbb{E}X = \mathbb{E}Y = 0$ and $\mathbb{E}X^2 = \mathbb{E}Y^2 = 1$. Suppose $G_1, G_2 \in L^2(\mathbb{R}, \phi)$ and let $g_{1,n}$ and $g_{2,n}$, $n \geq 0$, be the coefficients in the Hermite expansions of G_1 and G_2 , respectively, as in (4.6). Then,*

$$\mathbb{E}G_1(X)G_2(Y) = \sum_{n=0}^{\infty} g_{1,n}g_{2,n}n!(\mathbb{E}XY)^n$$

and

$$\text{Cov}(G_1(X), G_2(Y)) = \sum_{n=1}^{\infty} g_{1,n}g_{2,n}n!(\mathbb{E}XY)^n.$$

From now on, in this section, we will assume that $\{X_n\}_{n \in \mathbb{Z}}$ is stationary a Gaussian sequence with covariance function

$$\gamma_X(k) = L_2(k)k^{2d-1}, \quad k \in \{0\} \cup \mathbb{N}, \quad (4.9)$$

and satisfying $\mathbb{E}X_n = 0$, $\text{Var}X_n = 1$, $n \in \mathbb{N}$. In (4.9), L_2 is a *slowly varying function*

at infinity, namely, it is positive on $[c, \infty)$ with $c \geq 0$ and, for any $a > 0$,

$$\lim_{u \rightarrow \infty} \frac{L_2(au)}{L_2(u)} = 1. \quad (4.10)$$

Definition 4.2.3. Let m be a symmetric measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ in the sense that

$$m(A) = m(-A), \quad \text{for } A \in \mathcal{B}(\mathbb{R}),$$

where $-A = \{x \in \mathbb{R}, (-x) \in A\}$, and let

$$\mathcal{B}(\mathbb{R})_0 = \{A \in \mathcal{B}(\mathbb{R}) : m(A) < \infty\}.$$

An Hermitian Gaussian random measure Z on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a complex-valued Gaussian measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with the symmetric control measure m such that

$$\overline{Z(A)} = Z(-A), \quad A \in \mathcal{B}(\mathbb{R})_0.$$

Definition 4.2.4. Let $k \geq 1$ be an integer and

$$H \in \left(\frac{1}{2}, 1\right).$$

Set

$$H_0 = 1 - \frac{1-H}{k} \in \left(1 - \frac{1}{2k}, 1\right),$$

so that $H = 1 - k(1 - H_0)$. The *Hermite process* $\{Z_H^{(k)}(t)\}_{t \in \mathbb{R}}$ of order k is defined as

$$\{Z_H^{(k)}(t)\}_{t \in \mathbb{R}} \stackrel{fdd}{=} \left\{ a_{k, H_0} \int_{\mathbb{R}^k}' \left\{ \int_0^t \prod_{j=1}^k (s - u_j)_+^{H_0 - \frac{3}{2}} ds \right\} B(du_1) \cdots B(du_k) \right\}_{t \in \mathbb{R}},$$

where $\stackrel{fdd}{=}$ denotes equality of finite dimensional distributions, $B(du)$ is a Gaussian

random measure on \mathbb{R} with the Lebesgue control measure du , $u_+ = \max\{u, 0\}$ and a_{k,H_0} is a normalizing constant. The integral \int_0^t is interpreted as $-\int_t^0$ when $t < 0$. The Hermite process $\{Z_H^{(k)}(t)\}_{t \in \mathbb{R}}$ is called *standard* if $\mathbb{E}(Z_H^{(k)}(1))^2 = 1$.

In particular, the Hermite process of order $k = 1$ is a fractional Brownian motion $B_H(n)$ (see Example 2.2.1) with covariance function

$$\mathbb{E}B_H(s)B_H(t) = \frac{1}{2}\{|s|^{2H} + |t|^{2H} - |s - t|^{2H}\}. \quad (4.11)$$

The Hermite process of order $k = 2$ is called the Rosenblatt process [80]. The following proposition provides the spectral representation of Hermite processes. We write the latter using the parametrization H_0 .

Proposition 4.2.2. *The Hermite process of order k can be represented as*

$$\{Z_H^{(k)}(t)\}_{t \in \mathbb{R}} \stackrel{d}{=} \left\{ b_{k,H_0} \int_{\mathbb{R}^k}'' \frac{e^{it(x_1 + \dots + x_k)} - 1}{i(x_1 + \dots + x_k)} \prod_{j=1}^k |x_j|^{\frac{1}{2} - H_0} \widehat{B}(dx_1) \cdots \widehat{B}(dx_k) \right\}_{t \in \mathbb{R}}, \quad (4.12)$$

where $\widehat{B}(dx)$ is an Hermitian Gaussian random measure on \mathbb{R} with control measure dx .

Let $G : \mathbb{R} \rightarrow \mathbb{R}$ be a deterministic function such that $\mathbb{E}G(X_n)^2 < \infty$. The motivation behind the following results is to investigate the limit of the partial sum process

$$\sum_{n=1}^{[Nt]} G(X_n), \quad t \geq 0, \quad (4.13)$$

as $N \rightarrow \infty$. The following theorem establishes that this limit is an Hermite process, which is only Gaussian when the Hermite rank of G is $k = 1$.

Theorem 4.2.1. *Let G be a function of Hermite rank $k \geq 1$ and suppose that d as in (4.9) satisfies*

$$d \in \left(\frac{1}{2} \left(1 - \frac{1}{k} \right), \frac{1}{2} \right).$$

Then,

$$\frac{1}{(L_2(N))^{k/2} N^{k(d-1/2)+1}} \sum_{n=1}^{[Nt]} (G(X_n) - \mathbb{E}G(X_n)) \xrightarrow{fdd} g_k \beta_{k,H} Z_H^{(k)}(t), \quad t \geq 0, \quad (4.14)$$

where \xrightarrow{fdd} denotes the convergence of finite-dimensional distributions, g_k is the first non-zero coefficient in (4.6) and

$$\beta_{k,H} = \left(\frac{k!}{H(2H-1)} \right)^{1/2}. \quad (4.15)$$

In (4.14), the self-similarity parameter is given by

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

and $\{Z_H^{(k)}(t)\}_{t \in \mathbb{R}}$ is the Hermite process defined by (4.12).

The following result shows that, in some cases, the limit of partial sum processes in (4.13) is the usual Brownian motion.

Theorem 4.2.2. *Let γ_X be the autocovariance function of $\{X_n\}_{n \in \mathbb{Z}}$. Let G be a function with Hermite rank $k \geq 1$ in the sense of Definition 4.2.2. If*

$$\sum_{l=1}^{\infty} |\gamma_X(l)|^k < \infty, \quad (4.16)$$

then

$$\frac{1}{N^{1/2}} \sum_{n=1}^{[Nt]} (G(X_n) - \mathbb{E}G(X_n)) \xrightarrow{fdd} \sigma B(t), \quad t \geq 0, \quad (4.17)$$

where $\{B(t)\}_{t \geq 0}$ is a standard Brownian motion and

$$\sigma^2 = \sum_{m=k}^{\infty} g_m^2 m! \sum_{l=-\infty}^{\infty} (\gamma_X(l))^m.$$

Corollary 4.2.1. *If*

$$d \in \left(0, \frac{1}{2} \left(1 - \frac{1}{k}\right)\right)$$

in expression (4.9) for the covariance function $\gamma_X(\cdot)$ of X , then the convergence (4.17) to a Brownian motion holds.

Example 4.2.2. If $k = 1$ in (4.16), then the absolute summability of the autocovariance function leads to an ordinary Brownian limit. If $k = 2$, on the other hand, this limit only emerges when $d < \frac{1}{4}$.

We now turn to multivariate limit theorems. Consider the vector-valued random process

$$V_N(t) = \left(\frac{1}{A_r(N)} \sum_{n=1}^{[Nt]} (G_r(X_n) - \mathbb{E}G_r(X_n)) \right)_{r=1, \dots, R}, \quad t \geq 0, \quad (4.18)$$

where $G_r, r = 1, \dots, R$, are deterministic functions and $A_r(N), r = 1, \dots, R$, are appropriate normalization functions. The following theorem provides sufficient conditions for the process (4.18) to converge, as $N \rightarrow +\infty$, to a multivariate Gaussian process with dependent Brownian motion marginals.

Theorem 4.2.3. *Let $\{X_n\}_{n \in \mathbb{Z}}$ be a Gaussian stationary sequences with autocovariance function γ_X , where that $\mathbb{E}X_n = 0, \mathbb{E}X_n^2 = 1$. Let $G_r, r = 1, \dots, R$, be deterministic functions with respective Hermite ranks $k_r \geq 1, r = 1, \dots, R$. If*

$$\sum_{n=1}^{\infty} |\gamma_X(n)|^{k_r} < \infty, \quad r = 1, \dots, R,$$

then

$$V_N(t) \xrightarrow{fdd} V(t), \quad t \geq 0,$$

where $V_N(t)$ is given in (4.18) with $A_r(N) = N^{1/2}, r = 1, \dots, R$. The limit process can be expressed as

$$V(t) = (\sigma_1 B_1(t), \dots, \sigma_R B_R(t))^T, \quad (4.19)$$

where

$$\sigma_r^2 = \sum_{m=k_r}^{\infty} g_{r,m}^2 m! \sum_{l=-\infty}^{\infty} (\gamma_X(l))^m, \quad r = 1, \dots, R,$$

and $g_{r,m}$, $r = 1, \dots, R$, are the coefficients in the Hermite expansion (4.6) of G_r . In (4.19), $\{B_r(t)\}_{t \in \mathbb{R}}$, $r = 1, \dots, R$, are standard Brownian motions with cross-covariance

$$\mathbb{E}B_{r_1}(t_1)B_{r_2}(t_2) = (t_1 \wedge t_2) \frac{\sigma_{r_1, r_2}}{\sigma_{r_1} \sigma_{r_2}}, \quad t_1, t_2 \geq 0,$$

and

$$\sigma_{r_1, r_2} = \sum_{m=k_{r_1} \vee k_{r_2}}^{\infty} g_{r_1, m} g_{r_2, m} m! \sum_{n=-\infty}^{\infty} \gamma_X(n)^m$$

($a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$).

The next result concerns the general case where the resulting limit law for (4.18) is a multivariate process with dependent Hermite processes as marginals. The multivariate Hermite processes can be established as follows. Let

$$f_{H,k,t}(x_1, \dots, x_k) = B_{k,H} \frac{e^{it(x_1 + \dots + x_k)} - 1}{i(x_1 + \dots + x_k)} \prod_{j=1}^k |x_j|^{\frac{1-H}{k} - \frac{1}{2}}, \quad (4.20)$$

where

$$B_{k,H} = \left(\frac{H(2H-1)}{k! [2\Gamma(\frac{2(1-H)}{k}) \sin((\frac{1}{2} - \frac{1-H}{k})\pi)]^k} \right)^{1/2}, \quad j = 1, \dots, k.$$

Define

$$\widehat{I}_k(f) = \int_{\mathbb{R}^k}'' f(x_1, \dots, x_k) Z(dx_1) \cdots Z(dx_k). \quad (4.21)$$

Theorem 4.2.4. *Let $\{X_n\}_{n \in \mathbb{Z}}$ be a Gaussian stationary sequence with d as in (4.9). Suppose that $\mathbb{E}X_n = 0$, $\mathbb{E}X_n^2 = 1$. Let G_r , $r = 1, \dots, R$, be deterministic functions with respective Hermite ranks $k_r \geq 1$, $r = 1, \dots, R$. Suppose*

$$d \in \left(\frac{1}{2} \left(1 - \frac{1}{k_r} \right), \frac{1}{2} \right), \quad r = 1, \dots, R.$$

Consider the process $V_N(t)$ given by (4.18) with

$$A_r(N) = (L_2(N))^{k_r/2} N(k_r(d - 1/2) + 1), \quad r = 1, \dots, R.$$

Then

$$V_N(t) \xrightarrow{d} V^{(d)}(t), \quad t \geq 0,$$

where the limit process can be represented as

$$V^{(d)}(t) \stackrel{d}{=} \left(g_{r,k_r} \beta_{k_r, H_r} \widehat{I}_{k_r}(f_{H_r, k_r, t}) \right),$$

$$H_r = k_r \left(d - \frac{1}{2} \right) + 1 \in \left(\frac{1}{2}, 1 \right), \quad r = 1, \dots, R,$$

g_{r,k_r} is the first non-zero coefficient in the Hermite expansion of G_r in (4.6), $\beta_{k,H}$ is the constant given in (4.15), and $f_{H,k,t}$ is the kernel function defined in (4.20).

4.3 The asymptotic distribution of $\widehat{E}(n)$

In this section, we assume throughout that Y is a standard fractional Gaussian noise (fGn) with diffusion exponent $\alpha = 2H, 0 < \alpha < 2$. In other words, we can write

$$Y(n) = B_H(n+1) - B_H(n), \quad \mathbb{E}Y^2(n) = 1.$$

As a consequence of (4.11), the covariance function of Y is given by

$$\gamma_Y(k) = \frac{1}{2} \{ |k+1|^\alpha - 2|k|^\alpha + |k-1|^\alpha \}, \quad k \in \mathbb{Z}.$$

Furthermore, we can rewrite the autocovariance function as $\gamma_Y(k) = L_2(k)k^{\alpha-2}$, where

$$L_2(k) = \frac{k^2}{2} \left\{ \left| 1 + \frac{1}{k} \right|^\alpha - 2 + \left| 1 - \frac{1}{k} \right|^\alpha \right\}. \quad (4.22)$$

In fact, by a Taylor expansion of order 2 for $|1 \pm x|^{\alpha-2}$, it can be shown that $L_2(k)$ is a slowly varying function (see (4.10)). Therefore, Y satisfies condition (4.9) with $d = \frac{\alpha}{2} - \frac{1}{2}$, which will allow us to apply results from Section 4.2.

Fix $n \in \mathbb{N}$. Note that $\widehat{E}(n)$ is a complex-valued random variable, where

$$\Re \widehat{E}(n) = \Re \widehat{E}_1(n) - \widehat{E}_2(n), \quad \Im \widehat{E}(n) = \Im \widehat{E}_1(n),$$

and $\Re[z]$ and $\Im[z]$ denote the real and imaginary parts of the complex number z , respectively. We will investigate the limiting distributions of the real and imaginary parts of $\widehat{E}(n)$ separately, as $N \rightarrow \infty$. We first consider the limit law of $\widehat{E}_1(n)$. Define

$$\widetilde{Z}_n(k) = \sigma(n)[Y(n+k) - Y(k)], \quad (4.23)$$

where

$$\sigma(n) = \frac{1}{\sqrt{2 - |n-1|^\alpha + 2|n|^\alpha - |n+1|^\alpha}}.$$

In other words, the process $\{\widetilde{Z}_n(k)\}_{k \in \{0\} \cup \mathbb{Z}}$ consists of the normalized increments of the fGn Y . Then, $\widehat{E}_1(n)$ in (4.4) can be rewritten as

$$\widehat{E}_1(n) = \frac{1}{N-n+1} \sum_{k=0}^{N-n} \cos\left(\frac{\widetilde{Z}_n(k)}{\sigma(n)}\right) + i \frac{1}{N-n+1} \sum_{k=0}^{N-n} \sin\left(\frac{\widetilde{Z}_n(k)}{\sigma(n)}\right). \quad (4.24)$$

Let

$$G_{1,n}(x) = \cos(x/\sigma(n)), \quad G_{2,n}(x) = \sin(x/\sigma(n))^T. \quad (4.25)$$

We can use Theorem 4.2.2 to establish the limit law of the random vector $(\Re \widehat{E}_1(n), \Im \widehat{E}_1(n))$.

Proposition 4.3.1. *Let $\{\widetilde{Z}_n(k)\}_{k \in \mathbb{Z}}$ be the random sequence defined by (4.23) and let $\gamma_{\widetilde{Z}_n}$ be its autocovariance function. Also, let $\widehat{E}_1(n)$ be its associated statistic (4.24). Then, as $N \rightarrow \infty$,*

(i) the real part of $\widehat{E}_1(n)$ satisfies

$$\sqrt{N-n+1} \left(\Re \widehat{E}_1(n) - \mathbb{E} \cos \left(\frac{\widetilde{Z}_n(0)}{\sigma(n)} \right) \right) \xrightarrow{d} \sigma_{1,n} B_1(1); \quad (4.26)$$

(ii) the imaginary part of $\widehat{E}_1(n)$ satisfies

$$\sqrt{N-n+1} \Im \widehat{E}_1(n) \xrightarrow{d} \sigma_{2,n} B_2(1). \quad (4.27)$$

In (4.26) and (4.27),

$$\sigma_{r,n}^2 = \sum_{m=1}^{\infty} g_{r,n,m}^2 m! \sum_{l=-\infty}^{\infty} (\gamma_{\widetilde{Z}_n}(l))^m, \quad r = 1, 2, \quad (4.28)$$

$g_{1,n,m}$ and $g_{2,n,m}$ are the coefficients in the Hermite expansion (4.6) of the functions $G_{1,n}$ and $G_{2,n}$ in (4.25), respectively, and $B_1(1)$ and $B_2(1)$ are standard normal random variables.

Remark 4.3.1. Note that $\cos(x)$ is a even function while $\sin(x)$ is a odd function. Thus, $g_{1,n,2m+1} = g_{2,n,2m} = 0$, $m \in \{0\} \cup \mathbb{N}$. Then, by Proposition 4.2.1, $B_1(1)$ and $B_2(1)$ in Proposition 4.3.1 are two independent standard normal variables,

We now turn to $\widehat{E}_2(n)$ in (4.5). Recall that $\widehat{E}_2(n)$ is the sum of the squares of the sample averages of $\cos(Y(n))$ and $\sin(Y(n))$. Therefore, in order to describe the asymptotic distribution of $\widehat{E}_2(n)$, we should obtain the limiting joint distribution of $\frac{1}{N+1} \sum_{k=0}^N \cos(Y(k))$ and $\frac{1}{N+1} \sum_{k=0}^N \sin(Y(k))$. Recall that, as discussed in Section 4.2, the convergence rate of (4.13) is a function that depends on both the Hermite rank of G and $d = \alpha/2 - 1/2$. Thus, the convergence rates of $\frac{1}{N+1} \sum_{k=0}^N \cos(Y(k))$ and $\frac{1}{N+1} \sum_{k=0}^N \sin(Y(k))$ may be distinct, in which case one dominates the other. Based on Theorems 4.2.3 and 4.2.4, we can establish the following proposition.

Proposition 4.3.2. *Let Y be a fGn with $\mathbb{E}Y^2(0) = 1$ and $0 < \alpha < 2$, and let γ_Y*

be its autocovariance function. Let $\widehat{E}_2(n)$ be its associated statistic (4.5). Then, as $N \rightarrow \infty$,

(i) when $0 < \alpha < 3/2$,

$$\sqrt{N+1}[\widehat{E}_2(n) - |\mathbb{E} \cos(Y(0))|^2] \xrightarrow{d} 2\mathbb{E} \cos(Y(0)) \sigma_3 B_3(1), \quad (4.29)$$

where

$$\sigma_3^2 = \sum_{m=2}^{\infty} g_{\cos,m}^2 m! \sum_{k=-\infty}^{\infty} (\gamma_Y(k))^m, \quad (4.30)$$

and $g_{\cos,m}$ is given by (4.8);

(ii) when $3/2 < \alpha < 2$,

$$\frac{(N+1)^{2-\alpha}}{L_2(N+1)} [\widehat{E}_2(n) - |\mathbb{E} \cos(Y(0))|^2] \xrightarrow{d} g_{\cos,2} \beta_{2,\alpha-1} \widehat{I}_2(f_{\alpha-1,2,1}) + g_{\sin,1}^2 \beta_{1,\alpha/2}^2 \widehat{I}_1^2(f_{\alpha/2,1,1}). \quad (4.31)$$

where $g_{\cos,2}$ and $g_{\sin,1}$ are given by (4.8), $\beta_{2,\alpha-1}$ and $\beta_{1,\alpha/2}$ are defined in (4.15), $f_{H,k,t}$ is the kernel function defined in (4.20), $\widehat{I}_k(f)$ is defined in (4.21), $L_2(N+1)$ defined in (4.22).

Now we address the issue of the limiting distribution of $\widehat{E}(n)$ as $N \rightarrow \infty$. Since $\widehat{E}_2(n)$ is a real-valued, the asymptotic distribution of the imaginary part of $\widehat{E}(n)$ is same as that in (4.27). On the other hand, keeping in mind that $\Re \widehat{E}(n) = \Re \widehat{E}_1(n) - \widehat{E}_2(n)$ and the stochastic processes involved in the expressions for $\Re \widehat{E}_1(n)$ and $\widehat{E}_2(n)$ are distinct, i.e., Y and \widetilde{Z}_n . Therefore, we cannot directly apply Theorem 4.2.3 or Theorem 4.2.4 to obtain the joint distribution of $\Re \widehat{E}_1(n)$ and $\widehat{E}_2(n)$.

However, as shown in Proposition 4.3.2, when $3/2 < \alpha < 2$, the convergence rate of $\widehat{E}_2(n)$ is approximately $N^{2-\alpha}$, which is much slower than that of $\Re \widehat{E}_1(n)$. Thus, by Slutsky's theorem, the limiting distribution of $\Re \widehat{E}(n)$ is still given by (4.31). When $0 < \alpha < 3/2$, $\Re \widehat{E}_1(n)$ and $\widehat{E}_2(n)$ are both normally distributed. In this case, their

joint distribution is determined by their covariance matrix. By Proposition 4.2.1, this problem boils down to establishing the cross-covariance function of $\tilde{Z}_n(k+j)$ and $Y(j)$, which will be denote by $\{\gamma_{\tilde{Z}_n, Y}(k)\}_{k \in \mathbb{Z}}$. The following theorem encapsulates the above discussions.

Theorem 4.3.1. *Let $\{Y(k)\}_{k \in \mathbb{Z}}$ be a fGn with $\mathbb{E}Y^2(0) = 1$, and let $\{\tilde{Z}_n(k)\}_{k \in \{0\} \cup \mathbb{Z}}$ be the process defined by (4.23). Then, as $N \rightarrow \infty$,*

(i) *when $0 < \alpha < 3/2$,*

$$\begin{aligned} & \sqrt{N+1} \left(\Re \hat{E}(n) - \mathbb{E} \cos \left(\frac{\tilde{Z}_n(0)}{\sigma(n)} \right) + |\mathbb{E} \cos(Y(0))|^2 \right) \\ & \xrightarrow{d} \sigma_{1,n} B_1(1) + 2\mathbb{E} \cos(Y(0)) \sigma_3 B_3(1), \end{aligned}$$

where $\sigma_{1,n}$ is defined in (4.28), σ_3 is defined in (4.30), $B_1(1)$ and $B_3(1)$ are two standard normal distributions with correlation

$$\text{Corr}(B_1(1), B_3(1)) = \frac{\sum_{m=2}^{\infty} g_{1,n,m} g_{\cos,m} m! \sum_{k=-\infty}^{\infty} (\gamma_{\tilde{Z}_n, Y}(k))^m}{\sigma_{1,n} \sigma_3}. \quad (4.32)$$

In (4.32), $g_{1,n,m}$ and $g_{\cos,m}$ are the coefficients in the Hermite expansion (4.6) of $\cos(x/\sigma(n))$ and $\cos(x)$, respectively, and $\gamma_{\tilde{Z}_n, Y}$ is the cross-covariance function of $\tilde{Z}_n(k+j)$ and $Y(j)$;

(ii) *when $3/2 < \alpha < 2$,*

$$\begin{aligned} & \frac{(N+1)^{2-\alpha}}{L_2(N+1)} \left(\Re \hat{E}(n) - \mathbb{E} \cos(Y(n) - Y(0)) + |\mathbb{E} \cos(Y(0))|^2 \right) \\ & \xrightarrow{d} g_{\cos,2} \beta_{2,\alpha-1} \hat{I}_2(f_{\alpha-1,2,1}) + g_{\sin,1}^2 \beta_{1,\alpha/2}^2 \hat{I}_1^2(f_{\alpha/2,1,1}), \end{aligned}$$

where $g_{\cos,2}$ and $g_{\sin,1}$ are given in (4.8), $\beta_{2,\alpha-1}$ and $\beta_{1,\alpha/2}$ are defined in (4.15), $f_{H,k,t}$ is the kernel function defined in (4.20), $\hat{I}_k(f)$ is defined in (4.21), $L_2(N+1)$ is defined in (4.22);

(iii) for $0 < \alpha < 2$,

$$\sqrt{N+1} \Im \hat{E}(n) \xrightarrow{d} \sigma_{2,n} B_2(1),$$

where $\sigma_{2,n}$ is defined in (4.28), and $B_2(1)$ is a standard normal distribution.

In summary, Theorem 4.3.1 states that:

- irrespective of the value of α , the asymptotic distribution of $\Im \hat{E}(n)$ is Gaussian and the convergence rate is standard;
- the asymptotic distribution of $\Re \hat{E}(n)$ depends on the value of α . When $0 < \alpha < 3/2$, the convergence rate is $\sqrt{N+1}$ and the limiting distribution is Gaussian. When $3/2 < \alpha < 2$, the convergence rate is approximately $(N+1)^{2-\alpha}$ and the limiting distribution is non-Gaussian.

To illustrate Theorem 4.3.1, we conduct a Monte Carlo study. Figure 4.1 shows histograms of the finite sample distribution of $\hat{E}(n)$, where the red normal curve uses the sample mean and sample variance. The three plots on the left-hand side display histograms of the real part of $\hat{E}(n)$, while the plots on the right-hand side are of the imaginary part. In the simulation, the value of α for the plots on the top, middle and bottom are 0.6, 1, 1.8 respectively. The parameter n is set to 30, and the path length is $2^{10} = 1024$. As expected, all the plots show that they are normally distributed except the plot of the real part when $\alpha = 1.8 > 3/2$. The latter is right skewed as a consequence of the fact that the distribution is a linear combination of a Rosenblatt-type and a chi-squared distribution, both of which are right skewed.

Figure 4.2 depicts a Monte Carlo study of the convergence rate of $\hat{E}(n)$. In the simulations, $n = 30$, $\alpha = 0.6, 1, 1.8$, $N = 2^9, 2^{10}, 2^{11}, 2^{12}, 2^{13}, 2^{14}$. As expected, in the right plot, the convergence rate of $\Im \hat{E}(n)$ for different values of α is the same. In the left plot, the slope of $\alpha = 1.8$ is greater than that of $\alpha = 1$ and $\alpha = 0.6$ (these two are the same). This illustrates the fact that $\Re \hat{E}(n)$ has an unconventional convergence rate when $\alpha = 1.8$.

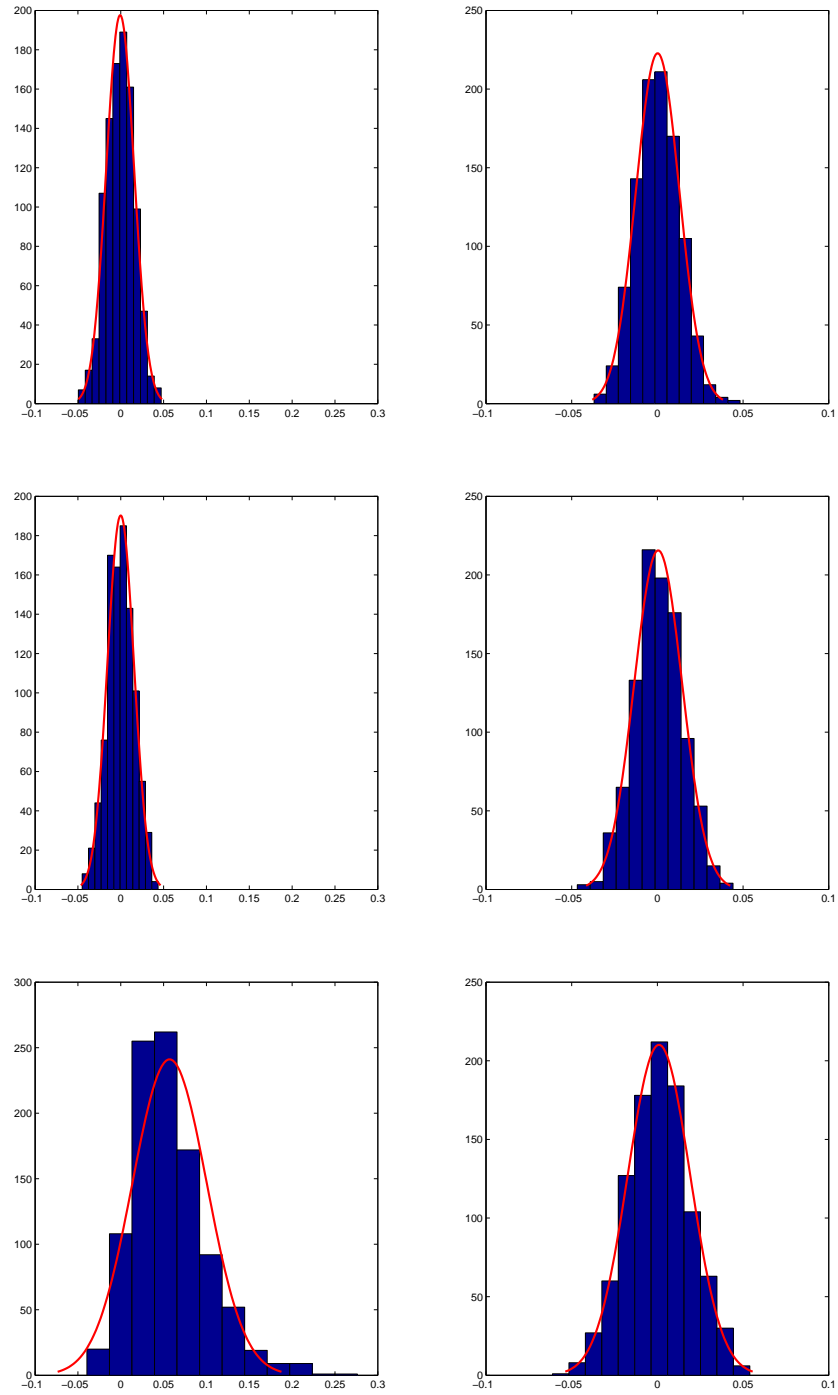


Figure 4.1: Histogram of $\widehat{E}(n)$ with fitted normal curve. Left: real part. Right: imaginary part. Top: $\alpha = 0.6$. Middle: $\alpha = 1.0$. Bottom: $\alpha = 1.8$.

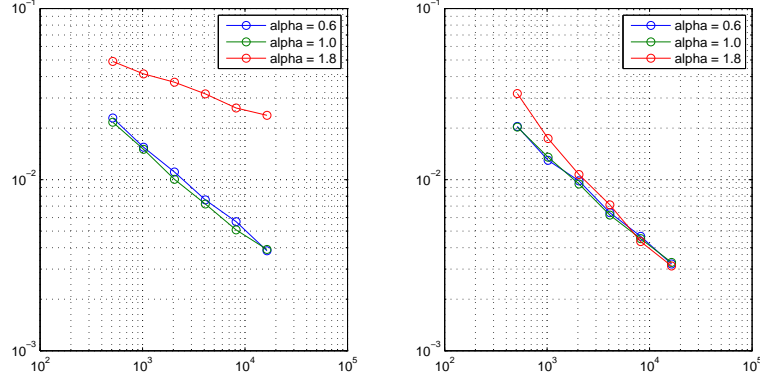


Figure 4.2: x -axis: $\log(N+1)$. y -axis: \log of the sample standard deviation. Left: real part of $\hat{E}(n)$. Right: imaginary part of $\hat{E}(n)$

4.4 Conclusion

In this chapter, we establish the asymptotic distribution of the mixing estimator proposed in [50]. We assume the stochastic process Y is a fractional Gaussian noise and only one trajectory is available. Depending on the diffusion exponent of the underlying process, namely, α , the real part of the mixing estimator has Gaussian or non-Gaussian limiting distribution, as well as different convergence rates. On the other hand, the imaginary part of the mixing estimator is always Gaussian and its convergence rate does not change with respect to α . Further work includes the construction of a hypothesis test for the mixing property starting from an observed trajectory X_1, \dots, X_N . Open problems include establishing the asymptotics distribution of the related ergodicity estimator defined in [50] as well as that of the mixing and ergodicity estimators starting from different classes of fractional processes such as continuous time random walks.

Appendix A

Proofs for Chapter 2

A.1 Proofs for Section 2.2

We first show a lemma that will be used in the proof of Proposition 2.2.1.

Lemma A.1.1. *Suppose assumptions (A1) and (A2) hold. Then, for $k_1, k_2 = 1, \dots, m$ and h_k as in (2.14),*

$$\left| \frac{\gamma_h(z)_{k_1, k_2}}{h^\alpha} - \frac{C_\alpha^2}{h^\alpha} \int_{\mathbb{R}} e^{ixz} \frac{(e^{ih_{k_1}x} - 1)(e^{-ih_{k_2}x} - 1)}{|x|^{\alpha+1}} dx \right| \leq Ch^{-\delta}, \quad (\text{A.1})$$

where $h, z \in \mathbb{Z}$, $h \geq \varepsilon_0^{-2}$, and $\delta = \min(\alpha/2, \delta_0/2) > 0$ (see (2.7) and (2.8)).

Proof. By (2.7), we obtain a harmonizable representation for the size h increment process $Y_z(h)$, namely,

$$Y_z(h) = X(z+h) - X(z) = C_\alpha \int_{\mathbb{R}} e^{izx} \frac{e^{ihx} - 1}{ix} \frac{s(x)}{|x|^{\alpha/2-1/2}} \tilde{B}(dx), \quad z \in \mathbb{Z}. \quad (\text{A.2})$$

Fix k_1 and k_2 . For notational simplicity, we will use the indices $k_1 = 1$, $k_2 = 2$. From (A.2), after the change of variables $xh = y$, we can write the covariance between the

increments $Y_z(h_{k_1})$ and $Y_z(h_{k_2})$ as

$$\gamma_h(z)_{1,2} = h^\alpha C_\alpha^2 \int_{\mathbb{R}} e^{iyz/h} \frac{(e^{iw_1 y} - 1)(e^{-iw_2 y} - 1)}{|y|^{\alpha+1}} |s(y/h)|^2 dy.$$

Now break up the left-hand side of the expression (A.1) into the sum

$$\left| C_\alpha^2 \left\{ \int_{|y| \leq \sqrt{h}} + \int_{|y| > \sqrt{h}} \right\} e^{iyz/h} \frac{(e^{iw_1 y} - 1)(e^{-iw_2 y} - 1)}{|y|^{\alpha+1}} (|s(y/h)|^2 - 1) dy \right| = |I_1 + I_2|,$$

where I_1 and I_2 denote the integrals over the domains $(-\sqrt{h}, \sqrt{h})$ and $\mathbb{R} \setminus (-\sqrt{h}, \sqrt{h})$, respectively. Then, for $h \geq \varepsilon_0^{-2}$,

$$\begin{aligned} |I_1| &\leq C \int_{|y| \leq \sqrt{h}} \frac{|e^{iw_1 y} - 1| |e^{-iw_2 y} - 1|}{|y|^{\alpha+1}} ||s(y/h)|^2 - 1| dy \\ &\leq C \int_{|y| \leq \sqrt{h}} \frac{|e^{iw_1 y} - 1| |e^{-iw_2 y} - 1|}{|y|^{\alpha+1}} |y/h|^{\delta_0} dy \\ &\leq Ch^{-\delta_0/2} \int_{\mathbb{R}} \frac{|e^{iw_1 y} - 1| |e^{-iw_2 y} - 1|}{|y|^{\alpha+1}} dy = Ch^{-\delta_0/2}. \end{aligned} \quad (\text{A.3})$$

Moreover, since $s(x)$ is bounded,

$$|I_2| \leq C \int_{|y| > \sqrt{h}} \frac{4M}{|y|^{\alpha+1}} dy \leq Ch^{-\alpha/2}. \quad (\text{A.4})$$

The expressions (A.3) and (A.4) yield (A.1). \square

\square

PROOF OF PROPOSITION 2.2.1: Fix k_1 and k_2 . For notational simplicity, we will use the indices $k_1 = 1$ and $k_2 = 2$. To show (i), set $z = 0$ and $h_{k_1} = h_{k_2} = h$ in (A.1).

Then,

$$\left| \frac{EX^2(h)}{\theta h^\alpha} - 1 \right| = \left| \frac{\gamma_h(0)_{1,1}}{\theta h^\alpha} - 1 \right| \leq Ch^{-\delta}, \quad h \rightarrow \infty,$$

where $\theta = C_\alpha^2 \int_{\mathbb{R}} |e^{iy} - 1|^2 |y|^{-(\alpha+1)} dy$.

We now show (ii). The proof draws upon conveniently rewriting the integral term in (A.1) based on the closed form expression for the covariance of a (standard) fBm. In fact, recall that $\tau = (\frac{C_\alpha}{C_H})^2 \frac{\alpha(\alpha-1)}{2}$, by (2.22). Then,

$$\begin{aligned} & \left| \frac{\gamma_h(z)_{1,2}}{h^\alpha} - w_1 w_2 \tau \left(\frac{z}{h} \right)^{\alpha-2} \right| \\ & \leq \left| \frac{\gamma_h(z)_{1,2}}{h^\alpha} - \frac{C_\alpha^2}{h^\alpha} \int_{\mathbb{R}} e^{ixz} \frac{(e^{ih_1x} - 1)(e^{-ih_2x} - 1)}{|x|^{\alpha+1}} dx \right| \\ & + \left| \left(\frac{C_\alpha}{C_H} \right)^2 \left\{ \frac{C_H^2}{h^\alpha} \int_{\mathbb{R}} e^{ixz} \frac{(e^{ih_1x} - 1)(e^{-ih_2x} - 1)}{|x|^{\alpha+1}} dx \right\} - w_1 w_2 \tau \left(\frac{z}{h} \right)^{\alpha-2} \right|. \end{aligned} \quad (\text{A.5})$$

Note that

$$\frac{C_H^2}{h^\alpha} \int_{\mathbb{R}} e^{ixz} \frac{(e^{ih_1x} - 1)(e^{-ih_2x} - 1)}{|x|^{\alpha+1}} dx \quad (\text{A.6})$$

is the expression for the covariance $\gamma_h(z)_{1,2}$ of a size h increment process $Y_z(h)$ (see (A.2)) formed from a standard fBm B_H . Pick $|z| \geq h_m + 1$. If $\alpha = 1$, the integral (A.6) is identically zero, by the independence of non-overlapping increments. Alternatively, when $\alpha \neq 1$, the closed form (2.10) with $\sigma^2 = 1$ allows us to rewrite (A.6) as

$$\begin{aligned} & \frac{1}{h^\alpha} [EB_H(z+h_1)B_H(h_2) - EB_H(z)B_H(h_2)] = \frac{1}{2h^\alpha} \{|z+h_1|^\alpha - |z+h_1-h_2|^\alpha - |z|^\alpha + |z-h_2|^\alpha\} \\ & = \frac{|z|^\alpha}{2h^\alpha} \left\{ \left(\left| 1 + w_1 \frac{h}{z} \right|^\alpha - 1 \right) - \left(\left| 1 + (w_1 - w_2) \frac{h}{z} \right|^\alpha - 1 \right) + \left(\left| 1 - w_2 \frac{h}{z} \right|^\alpha - 1 \right) \right\}. \end{aligned} \quad (\text{A.7})$$

Let $f(x) = x^\alpha$. Based on second order Taylor expansions of f around 1, we can recast the expression (A.7) as

$$\begin{aligned} & \frac{|z|^\alpha}{2h^\alpha} \left\{ \frac{\alpha(\alpha-1)}{2} \left(\frac{h}{z} \right)^2 [w_1^2 - (w_1 - w_2)^2 + w_2^2] + O\left[\left(\frac{h}{z} \right)^3 \right] \right\} \\ & = \left(\frac{|z|}{h} \right)^{\alpha-2} \frac{\alpha(\alpha-1)w_1w_2}{2} \left(1 + O\left(\frac{h}{z} \right) \right). \end{aligned} \quad (\text{A.8})$$

Therefore, based on (A.8) (which also encompasses the case $\alpha = 1$) and (A.1), the

expression (A.5) can be further bounded by $C[(\frac{|z|}{h})^{\alpha-3} + h^{-\delta}]$. As a consequence, we arrive at

$$\left| \frac{\gamma_h(z)_{1,2}}{|z|^{\alpha-2}h_1h_2} - \tau \right| \leq C \left(\frac{h}{|z|} + \left(\frac{h}{|z|} \right)^\alpha \frac{z^2}{h^{2+\delta}} \right) \leq C \left(\frac{h}{|z|} \right)^{\min(1,\alpha)} \leq C \left(\frac{h}{|z|} \right)^\delta,$$

where the last two inequalities result from (2.17) and (2.21). Setting $g(z, h) = \frac{\gamma_h(z)_{1,2}}{|z|^{\alpha-2}h_1h_2} - \tau$ yields (2.20). \square

A.2 Proofs for Section 2.3

PROOF OF THEOREM 2.3.1: In view of (2.17), the claim is equivalent to

$$\left(\eta^{-1}(n) \zeta^{-1}(h_k) \sum_{i_k=1}^n \{Y_{i_k}^2(h_k) - EX^2(h_k)\} \right)_{k=1, \dots, m} \xrightarrow{d} \mathbf{Z}.$$

Consider the vector of increments

$$\mathbf{Y} = (Y_1(h_1), \dots, Y_n(h_1); Y_1(h_2), \dots, Y_n(h_2); \dots; Y_1(h_m), \dots, Y_n(h_m))^T,$$

The covariance matrix of \mathbf{Y} can be written as $R_m(n) = (R_{k_1, k_2}(n))_{k_1, k_2=1, \dots, m}$, where

$$R_{k_1, k_2}(n) := \begin{pmatrix} \gamma_h(0)_{k_1, k_2} & \cdots & \gamma_h(1-n)_{k_1, k_2} \\ \vdots & \ddots & \vdots \\ \gamma_h(n-1)_{k_1, k_2} & \cdots & \gamma_h(0)_{k_1, k_2} \end{pmatrix} \in \mathbb{R}^{n^2}.$$

Let

$$\mathbf{Z}_n = (Z_n(h_1), \dots, Z_n(h_m))^T \tag{A.9}$$

be the centered statistic defined by

$$Z_n(h_k) = \eta^{-1}(n)\zeta^{-1}(h_k) \sum_{i_k=1}^n \{Y_{i_k}^2(h_k) - EX^2(h_k)\}, \quad k = 1, \dots, m.$$

Also, let

$$D_m(n) := \text{diag}\left(D_{1,1}(n), D_{2,2}(n), \dots, D_{m,m}(n)\right), \quad D_{i,i}(n) := \frac{t_i}{\zeta(h_i)} I_n, \quad i = 1, \dots, m,$$

where I_n denotes an $n \times n$ identity matrix. The weak limit (2.32) can be established via characteristic functions. The initial manipulation of the characteristic function is very similar to that in Rosenblatt [74]. First note that

$$\int_{\mathbb{R}^n} \exp\left\{-\frac{1}{2}\mathbf{y}^T(R_m^{-1}(n) - icD_m(n))\mathbf{y}\right\} d\mathbf{y} = (2\pi)^{n/2} \det(R_m^{-1}(n) - icD_m(n))^{-1/2}, \quad c \in \mathbb{R}.$$

By a similar computation to that in Taqqu [82], pp.42–43,

$$\begin{aligned} \phi_{\mathbf{Z}_n}(\mathbf{t}) &= E(e^{i\mathbf{t}^T \mathbf{Z}_n}) = E e^{i\mathbf{t}^T (Z_n(h_1), Z_n(h_2), \dots, Z_n(h_m))^T} \\ &= \int_{\mathbb{R}^{mn}} \exp\left\{i \sum_{k=1}^m t_k z_n(h_k)\right\} \frac{1}{\sqrt{\det(2\pi R_m(n))}} \exp\left\{-\frac{1}{2}\mathbf{y}^T R_m^{-1}(n)\mathbf{y}\right\} d\mathbf{y} \\ &= \exp\left\{\frac{1}{2}\left[-2i\eta^{-1}(n) \sum_{k=1}^m n t_k \zeta^{-1}(h_k) \gamma_h(0)_{k,k} - \sum_{l=1}^{mn} \log(1 - 2i\eta^{-1}(n) \lambda_{l,mn})\right]\right\}. \quad (\text{A.10}) \end{aligned}$$

The scalars $\lambda_{l,mn}$, $l = 1, \dots, mn$, denote the eigenvalues (characteristic roots) of $R_m(n)D_m(n) = PJP^{-1}$, where $P \in GL(mn, \mathbb{C})$, and J is in Jordan form. By the analytic expansion of $\log(1 - 2i\eta^{-1}(n)\lambda_{l,mn})$,

$$-\sum_{l=1}^{mn} \log(1 - 2i\eta^{-1}(n)\lambda_{l,mn}) = 2i\eta^{-1}(n) \sum_{l=1}^{mn} \lambda_{l,mn} + \sum_{s=2}^{\infty} \frac{(2i)^s}{s} \eta^{-s}(n) \sum_{l=1}^{mn} \lambda_{l,mn}^s. \quad (\text{A.11})$$

However, $\sum_{l=1}^{mn} \lambda_{l,mn} = \text{tr}(R_m(n)D_m(n)) = \sum_{k=1}^m n t_k \zeta^{-1}(h_k) \gamma_h(0)_{k,k}$. Thus, by (A.11)

we can rewrite (A.10) as

$$Ee^{it^T \mathbf{Z}_n} = \exp \left\{ \frac{1}{2} \sum_{s=2}^{\infty} \frac{(2i)^s}{s} \eta^{-s}(n) \sum_{l=1}^{mn} \lambda_{l,mn}^s \right\}. \quad (\text{A.12})$$

Moreover,

$$\begin{aligned} \eta^{-s}(n) \sum_{l=1}^{mn} \lambda_{l,mn}^s &= \eta^{-s}(n) \text{tr}[(R_m(n)D_m(n))^s] \\ &= \eta^{-s}(n) \sum_{k_1, \dots, k_s=1}^m \left\{ \text{tr} \left[R_{k_1, k_2}(n) D_{k_2, k_2}(n) R_{k_2, k_3}(n) D_{k_3, k_3}(n) \dots R_{k_s, k_1}(n) D_{k_1, k_1}(n) \right] \right\} \\ &= \sum_{k_1, \dots, k_s=1}^m \left\{ t_{k_1} t_{k_2} \dots t_{k_s} \zeta^{-1}(h_{k_1}) \zeta^{-1}(h_{k_2}) \dots \zeta^{-1}(h_{k_s}) \right. \\ &\quad \times \eta^{-s}(n) \sum_{i_1, \dots, i_s=1}^n \gamma_h(i_1 - i_2)_{k_1, k_2} \gamma_h(i_2 - i_3)_{k_2, k_3} \dots \gamma_h(i_s - i_1)_{k_s, k_1} \left. \right\}. \end{aligned} \quad (\text{A.13})$$

The weak limits (2.34), (2.36) and (2.37) are a consequence of Propositions A.3.1 and A.3.2. \square

PROOF OF COROLLARY 2.3.1 We first show that

$$\frac{nh^\alpha}{\eta(n)\zeta(h)} \left(\frac{\bar{\mu}_2(h_k)}{\theta h_k^\alpha} - 1 \right)_{k=1, \dots, m} \xrightarrow{d} A\mathbf{Z}, \quad (\text{A.14})$$

where

$$A = A(\theta, \alpha) = \text{diag}(\zeta(w_1)/(\theta w_1^\alpha), \dots, \zeta(w_m)/(\theta w_m^\alpha)) \quad (\text{A.15})$$

and \mathbf{Z} is as in Theorem 2.3.1. Based on (2.30), rewrite the left-hand side of the expression (2.32) as

$$\left(\frac{N_k \theta h_k^\alpha}{\eta(N_k) \zeta(h_k)} \left(\frac{\bar{\mu}_2(h_k)}{\theta h_k^\alpha} - \frac{EX^2(h_k)}{\theta h_k^\alpha} \right) \right)_{k=1, \dots, m}.$$

This random vector has the same asymptotic distribution as

$$\frac{nh^\alpha}{\eta(n)\zeta(h)} \left(\frac{\theta w_k^\alpha}{\zeta(w_k)} \left(\frac{\bar{\mu}_2(h_k)}{\theta h_k^\alpha} - 1 \right) \right)_{k=1,\dots,m} - \frac{nh^\alpha}{\eta(n)\zeta(h)} \left(\frac{\theta w_k^\alpha}{\zeta(w_k)} \left(\frac{EX^2(h_k)}{\theta h_k^\alpha} - 1 \right) \right)_{k=1,\dots,m}. \quad (\text{A.16})$$

However, the bound (2.19) yields

$$\frac{nh^\alpha}{\eta(n)\zeta(h)} \frac{\theta w_k^\alpha}{\zeta(w_k)} \left| \frac{EX^2(h_k)}{\theta h_k^\alpha} - 1 \right| \leq \frac{nh^\alpha}{\eta(n)\zeta(h)} h^{-\delta} \rightarrow 0, \quad k = 1, \dots, m, \quad (\text{A.17})$$

where the zero limit is a consequence of (2.17) and (2.33). The expression (A.14) is now a consequence of (A.16), (A.17) and (2.32).

To show (2.43), rewrite

$$\widehat{\beta}_n - \beta = (M_n^T M_n)^{-1} M_n^T (Q_n - M_n \beta). \quad (\text{A.18})$$

By entrywise first order Taylor expansions,

$$Q_n - M_n \beta = \left(\log \left(\frac{\bar{\mu}_2(h_k)}{\theta h_k^\alpha} \right) \right)_{k=1,\dots,m} = \left(\frac{\bar{\mu}_2(h_k)}{\theta h_k^\alpha} - 1 \right)_{k=1,\dots,m} + O \left(\frac{\bar{\mu}_2(h_k)}{\theta h_k^\alpha} - 1 \right)_{k=1,\dots,m}^2 \quad (\text{A.19})$$

On the other hand, note that $\det(M_n^T M_n) = c_w$ (see (2.45)) is a constant with respect to n . Thus,

$$\begin{aligned} (M_n^T M_n)^{-1} M_n^T &= \frac{1}{c_w} \begin{pmatrix} \sum_{k=1}^m \log^2 h_k & -\sum_{k=1}^m \log h_k \\ -\sum_{k=1}^m \log h_k & m \end{pmatrix} \begin{pmatrix} 1 & \dots & 1 \\ \log h_1 & \dots & \log h_m \end{pmatrix} \\ &= \frac{1}{c_w} \begin{pmatrix} \sum_{k=1}^m \log^2 h_k - \log h_1 \sum_{k=1}^m \log h_k & \dots & \sum_{k=1}^m \log^2 h_k - \log h_m \sum_{k=1}^m \log h_k \\ m \log h_1 - \sum_{k=1}^m \log h_k & \dots & m \log h_m - \sum_{k=1}^m \log h_k \end{pmatrix}. \end{aligned}$$

Moreover, for $j = 1, \dots, m$,

$$\sum_{k=1}^m \log^2 h_k - \log h_j \sum_{k=1}^m \log h_k = \log h \sum_{k=1}^m \log(w_k/w_j) + \sum_{k=1}^m \log w_k \log(w_k/w_j)$$

and $m \log h_j - \sum_{k=1}^m \log h_k = \sum_{k=1}^m \log(w_j/w_k)$. Therefore, by (2.17), we obtain the entrywise asymptotic equivalence

$$(M_n^T M_n)^{-1} M_n^T \sim \frac{1}{c_w} \begin{pmatrix} \log h & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sum_{k=1}^m \log(w_k/w_1) & \dots & \sum_{k=1}^m \log(w_k/w_m) \\ \sum_{k=1}^m \log(w_1/w_k) & \dots & \sum_{k=1}^m \log(w_m/w_k) \end{pmatrix}. \quad (\text{A.20})$$

By (A.18), (A.19), (A.20), (A.14), and (2.44), we arrive at (2.43). \square

A.3 Auxiliary results

Lemmas A.3.1-A.3.4, stated below, are used in the proofs in Propositions A.3.1 and A.3.2. The proofs of the lemmas can be found in the supporting information paper Didier and Zhang [24].

Lemma A.3.1. *Consider $3/2 < \alpha < 2$ and $s \geq 2$, and suppose the assumptions (A1) and (A2) hold. Then, as $n \rightarrow \infty$,*

$$\zeta^{-1}(h_{k_1}) \cdots \zeta^{-1}(h_{k_s}) \eta^{-s}(n) \sum_{\substack{i_1, \dots, i_s=1 \\ |i_1 - i_2| \leq h \cup \dots \cup |i_s - i_1| \leq h}}^n \gamma_h(i_1 - i_2)_{k_1, k_2} \cdots \gamma_h(i_s - i_1)_{k_s, k_1} \rightarrow 0. \quad (\text{A.21})$$

Lemma A.3.2. *Consider $3/2 < \alpha < 2$ and $s \geq 2$, and suppose the assumptions (A1) and (A2) hold. Then, as $n \rightarrow \infty$,*

$$\zeta^{-1}(h_{k_1}) \cdots \zeta^{-1}(h_{k_s}) \eta^{-s}(n) \sum_{\substack{i_1, \dots, i_s=1 \\ |i_1 - i_2| \geq h+1, \dots, |i_s - i_1| \geq h+1}}^n \gamma_h(i_1 - i_2)_{k_1, k_2} \cdots \gamma_h(i_s - i_1)_{k_s, k_1}$$

$$\rightarrow \tau^s \int_0^1 \dots \int_0^1 |x_1 - x_2|^{\alpha-2} \dots |x_s - x_1|^{\alpha-2} dx_1 \dots dx_s. \quad (\text{A.22})$$

Lemma A.3.3. Consider $0 < \alpha \leq 3/2$ and $s \geq 3$, and suppose the assumptions (A1) and (A2) hold. Then, as $n \rightarrow \infty$,

$$\zeta^{-1}(h_{k_1}) \dots \zeta^{-1}(h_{k_s}) \eta^{-s}(n) \sum_{i_1, \dots, i_s=1}^n \gamma_h(i_1 - i_2)_{k_1, k_2} \dots \gamma_h(i_s - i_1)_{k_s, k_1} \rightarrow 0. \quad (\text{A.23})$$

Lemma A.3.4. Suppose the assumptions (A1) and (A2) hold. Then, as $n \rightarrow \infty$,

(i) in the parameter range $0 < \alpha < 3/2$,

$$\begin{aligned} & \eta^{-2}(n) \zeta^{-1}(h_{k_1}) \zeta^{-1}(h_{k_2}) \sum_{i_1, i_2=1}^n \gamma_h^2(i_1 - i_2)_{k_1, k_2} \\ & \rightarrow w_{k_1}^{-(\alpha+1/2)} w_{k_2}^{-(\alpha+1/2)} \left(\frac{C_\alpha}{C_H} \right)^4 \|\widehat{G}(y; w_{k_1}, w_{k_2})\|_{L^2(\mathbb{R})}^2, \end{aligned} \quad (\text{A.24})$$

where $\widehat{G}(y; w_{k_1}, w_{k_2})$, C_α and C_H are defined by (2.35), (2.7) and (2.13), respectively;

(ii) when $\alpha = 3/2$,

$$\eta^{-2}(n) \zeta^{-1}(h_{k_1}) \zeta^{-1}(h_{k_2}) \sum_{i_1, i_2=1}^n \gamma_h^2(i_1 - i_2)_{k_1, k_2} \rightarrow 2\tau^2, \quad (\text{A.25})$$

where τ is given by (2.22).

Proposition A.3.1. Consider the parameter range $3/2 < \alpha < 2$ and suppose the assumptions (A1)–(A2) hold. Then, as $n \rightarrow \infty$, the vector $\mathbf{Z}_n = (Z_n(h_1), Z_n(h_2), \dots, Z_n(h_m))^T$ in (A.9) converges in law to a Rosenblatt-like distribution whose characteristic function is given by

$$\phi_{\mathbf{Z}}(\mathbf{t}) = \exp \left\{ \frac{1}{2} \sum_{s=2}^{\infty} \frac{(2i\tau \sum_{k=1}^m t_k)^s}{s} \int_0^1 \dots \int_0^1 |x_1 - x_2|^{\alpha-2} \dots |x_s - x_1|^{\alpha-2} dx_1 \dots dx_s \right\}. \quad (\text{A.26})$$

Proof. Let $s \geq 2$ and consider the expression (A.13). By Lemmas A.3.1 and A.3.2, as $n \rightarrow \infty$ the right-hand side of the latter converges to

$$\begin{aligned} & \sum_{k_1, \dots, k_s=1}^m t_{k_1} t_{k_2} \cdots t_{k_s} \tau^s \int_0^1 \cdots \int_0^1 |x_1 - x_2|^{\alpha-2} \cdots |x_s - x_1|^{\alpha-2} dx_1 \cdots dx_s \\ &= \left(\sum_{k=1}^m t_k \right)^s \tau^s \int_0^1 \cdots \int_0^1 |x_1 - x_2|^{\alpha-2} \cdots |x_s - x_1|^{\alpha-2} dx_1 \cdots dx_s. \end{aligned}$$

Therefore, the characteristic function (A.12) converges to (A.26), as claimed. \square

\square

Proposition A.3.2. *For $0 < \alpha \leq 3/2$, suppose the assumptions (A1)–(A2) hold. Let $\mathbf{Z}_n = (Z_n(h_1), Z_n(h_2), \dots, Z_n(h_m))^T$ be the random vector in (A.9). Then, as $n \rightarrow \infty$, $\mathbf{Z}_n \xrightarrow{d} N(\mathbf{0}, \Sigma)$, where Σ is a $m \times m$ matrix with components*

$$\Sigma_{k_1, k_2} = \begin{cases} 2w_{k_1}^{-\alpha-1/2} w_{k_2}^{-\alpha-1/2} \left(\frac{C_\alpha}{C_H}\right)^4 \|\widehat{G}(y; w_{k_1}, w_{k_2})\|_{L^2(\mathbb{R})}^2, & 0 < \alpha < 3/2; \\ 4\tau^2, & \alpha = 3/2, \end{cases} \quad (\text{A.27})$$

and $\widehat{G}(y; w_{k_1}, w_{k_2})$ is defined by (2.35).

Proof. When $0 < \alpha \leq 3/2$, by Lemma A.3.3 it suffices to consider the term (A.13) of order $s = 2$. Therefore, by Lemma A.3.4, as $n \rightarrow \infty$ the characteristic function (A.12) converges to that of a multivariate normal distribution with covariance matrix $\Sigma = (\Sigma_{k_1, k_2})_{k_1, k_2=1, \dots, m}$ as in (A.27). \square

\square

A.4 Additional proofs

This section contains the proofs of Lemmas A.3.1–A.3.4 in Didier and Zhang [23].

For $h_m + 1 \leq |z| \leq n$, recall that conditions (2.20) and (2.23) can be jointly expressed as

$$\gamma_h(z)_{k_1, k_2} = w_{k_1} w_{k_2} |z|^{\alpha-2} h^2 \{\tau + g(z, h)_{k_1, k_2}\}, \quad |g(z, h)_{k_1, k_2}| \leq C \left(\frac{h}{|z|} \right)^\delta, \quad (\text{A.28})$$

for a general pair of indices $k_1, k_2 = 1, \dots, m$ representing shifting lag values. Moreover, by the Cauchy-Schwarz inequality and (2.19),

$$|\gamma_h(z)_{k_1, k_2}| \leq C h^\alpha, \quad h, z \in \mathbb{Z}, \quad (\text{A.29})$$

where $C > 0$ does not depend on k_1, k_2 . In particular, for a single shifting lag value

$$h_{k_1} = h_{k_2} = h(n) =: h, \quad (\text{A.30})$$

the expressions (A.28) and (2.19) imply that

$$\gamma_h(z) := \gamma_h(z)_{1,1} = |z|^{\alpha-2} h^2 \{\tau + g(z, h)\}, \quad |g(z, h)| \leq C \left(\frac{h}{|z|} \right)^\delta. \quad (\text{A.31})$$

Thus, in the proofs of Lemmas A.3.1–A.3.4 below, we will first establish the statements for a single index (shifting lag value) $m = 1$ and (A.30), and then adjust the constants to obtain the general statements for $m > 1$. In particular, it will be implicit that when a multiple summation is taken over index ranges of the form $|i_1 - i_2| \geq h + 1$ under $m = 1$, in the general case one should substitute h_m for h under $m > 1$.

PROOF OF LEMMA A.3.1 First assume $m = 1$. We only look at the subcase where the summation is taken over the index set

$$\{|i_1 - i_2| \leq h\} \cap \{|i_2 - i_3| \geq h + 1\} \cap \dots \cap \{|i_s - i_1| \geq h + 1\}, \quad (\text{A.32})$$

since the remaining $2^s - 2$ subcases can be tackled in a similar fashion. By (A.31), we can rewrite the expression of interest as

$$\frac{h^{2(s-1)}}{\eta^s(n)\zeta^s(h)} \sum_{\substack{i_1, i_2, \dots, i_s=1 \\ |i_1-i_2| \leq h, |i_2-i_3| \geq h+1, \dots, |i_s-i_1| \geq h+1}}^n \gamma_h(i_1 - i_2) \\ |i_2 - i_3|^{\alpha-2} \dots |i_s - i_1|^{\alpha-2} \{\tau + g(i_2 - i_3, h)\} \dots \{\tau + g(i_s - i_1, h)\}, \quad (\text{A.33})$$

where, under the summation sign, the terms of the form $\tau + g(\cdot, \cdot)$ can be uniformly bounded by a constant, and $\gamma_h(i_1 - i_2)$ is bounded by Ch^α (see (A.29)). Thus, the absolute value of (A.33) is bounded by

$$\begin{aligned} & C n^{s(1-\alpha)} h^{-2} n^{(s-1)(\alpha-1)} \sum_{\substack{i_1, \dots, i_s=1, i_1 \neq i_2 \\ |i_1-i_2| \leq h, |i_2-i_3| \geq h+1, \dots, |i_s-i_1| \geq h+1}}^n h^\alpha \left| \frac{i_2 - i_3}{n} \right|^{\alpha-2} \dots \left| \frac{i_s - i_1}{n} \right|^{\alpha-2} \frac{1}{n^{s-1}} \\ &= C \frac{h^{\alpha-2}}{n^{\alpha-1}} \sum_{z=-h}^h \sum_{\substack{i_2, \dots, i_s=1 \\ |i_2-i_3| \geq h+1, \dots, |i_s-i_2+z| \geq h+1}}^n \left| \frac{i_2 - i_3}{n} \right|^{\alpha-2} \dots \left| \frac{i_s - i_2 + z}{n} \right|^{\alpha-2} \frac{1}{n^{s-1}} \\ &\leq C \left(\frac{h}{n} \right)^{\alpha-1} \sum_{\substack{i_2, \dots, i_s=1 \\ |i_2-i_3| \geq h+1, \dots, |i_s-i_2+z| \geq h+1}}^n \left| \frac{i_2 - i_3}{n} \right|^{\alpha-2} \dots \left| \frac{i_s - i_2 - \text{sign}(i_s - i_2)h}{n} \right|^{\alpha-2} \frac{1}{n^{s-1}} \\ &\sim C \left(\frac{h}{n} \right)^{\alpha-1} \int_0^1 \dots \int_0^1 |x_2 - x_3|^{\alpha-2} \dots |x_s - x_2|^{\alpha-2} dx_2 \dots dx_s, \end{aligned}$$

which goes to zero as $n \rightarrow \infty$, since $\alpha > 3/2$ and by (2.17). This shows (A.21) for $m = 1$. In addition, adjusting for the constants w_{k_1}, w_{k_2} from (A.28) does not alter the zero limit. Hence, (A.21) also holds for $m > 1$. \square

PROOF OF LEMMA A.3.2 First assume $m = 1$. We start out by establishing that

$$\sum_{\substack{i_1, \dots, i_s=1 \\ |i_1-i_2| \geq h+1, \dots, |i_s-i_1| \geq h+1}}^n \left| \frac{i_1 - i_2}{n} \right|^{\alpha-2} \left| \frac{i_2 - i_3}{n} \right|^{\alpha-2} \dots \left| \frac{i_s - i_1}{n} \right|^{\alpha-2} \frac{1}{n^s}$$

$$\rightarrow \int_0^1 \int_0^1 \dots \int_0^1 |x_1 - x_2|^{\alpha-2} |x_2 - x_3|^{\alpha-2} \dots |x_s - x_1|^{\alpha-2} dx_1 dx_2 \dots dx_s, \quad n \rightarrow \infty. \quad (\text{A.34})$$

Indeed, since

$$\sum_{\substack{i_1, \dots, i_s=1 \\ i_1 \neq i_2, i_2 \neq i_3, \dots, i_s \neq i_1}}^n \left| \frac{i_1 - i_2}{n} \right|^{\alpha-2} \left| \frac{i_2 - i_3}{n} \right|^{\alpha-2} \dots \left| \frac{i_s - i_1}{n} \right|^{\alpha-2} \frac{1}{n^s} \\ \rightarrow \int_0^1 \int_0^1 \dots \int_0^1 |x_1 - x_2|^{\alpha-2} |x_2 - x_3|^{\alpha-2} \dots |x_s - x_1|^{\alpha-2} dx_1 dx_2 \dots dx_s, \quad n \rightarrow \infty, \quad (\text{A.35})$$

and the sum on the left-hand side of (A.35) can be broken up into

$$\left\{ \sum_{\substack{i_1, \dots, i_s=1 \\ |i_1 - i_2| \geq h+1, \dots, |i_s - i_1| \geq h+1}}^n + \sum_{\substack{i_1, \dots, i_s=1 \\ \{|i_1 - i_2| \geq h+1, \dots, |i_s - i_1| \geq h+1\}^c}}^n \right\} \left| \frac{i_1 - i_2}{n} \right|^{\alpha-2} \left| \frac{i_2 - i_3}{n} \right|^{\alpha-2} \dots \left| \frac{i_s - i_1}{n} \right|^{\alpha-2} \frac{1}{n^s}, \quad (\text{A.36})$$

then it suffices to show that the second summation term in (A.36) goes to zero.

However, the latter can be established by a similar argument to that in the proof of Lemma A.3.1. Thus, (A.34) holds.

Based on (A.31), recast the left-hand side of (A.22) as

$$\frac{h^{2s}}{\eta^s(n) \zeta^s(h)} \sum_{\substack{i_1, i_2, \dots, i_s=1 \\ |i_1 - i_2| \geq h+1, \dots, |i_s - i_1| \geq h+1}}^n |i_1 - i_2|^{\alpha-2} \{\tau + g(i_1 - i_2, h)\} \dots |i_s - i_1|^{\alpha-2} \{\tau + g(i_s - i_1, h)\}. \quad (\text{A.37})$$

In view of (A.34), we only need to show that the remaining terms involving at least one residual function g in (A.37) go to zero. Pick a number ρ in the interval $(0, \min(\delta, \alpha - 3/2))$. By (A.31), $|g(h/z)| \leq C(h/z)^\delta \leq C(h/z)^\rho$, $z \geq h + 1$. Therefore,

$$\frac{1}{\eta^{s(\alpha-1)}} \sum_{\substack{i_1, i_2, \dots, i_s=1 \\ |i_1 - i_2| \geq h+1, \dots, |i_s - i_1| \geq h+1}}^n |i_1 - i_2|^{\alpha-2} \dots |i_s - i_1|^{\alpha-2} g(i_s - i_1, h)$$

$$\begin{aligned}
&\leq \frac{C}{n^{s(\alpha-1)}} \sum_{\substack{i_1, i_2, \dots, i_s=1 \\ |i_1-i_2| \geq h+1, \dots, |i_s-i_1| \geq h+1}}^n |i_1 - i_2|^{\alpha-2} \dots |i_s - i_1|^{\alpha-2} \left| \frac{h}{i_s - i_1} \right|^\rho \\
&= C \left(\frac{h}{n} \right)^\rho \sum_{\substack{i_1, i_2, \dots, i_s=1 \\ |i_1-i_2| \geq h+1, \dots, |i_s-i_1| \geq h+1}}^n \left| \frac{i_1 - i_2}{n} \right|^{\alpha-2} \dots \left| \frac{i_{s-1} - i_s}{n} \right|^{\alpha-2} \left| \frac{i_s - i_1}{n} \right|^{\alpha-2-\rho} \frac{1}{n^s} \\
&\sim C \left(\frac{h}{n} \right)^\rho \int_0^1 \dots \int_0^1 |x_1 - x_2|^{\alpha-2} \dots |x_{s-1} - x_s|^{\alpha-2} |x_s - x_1|^{\alpha-2-\rho} dx_1 \dots dx_s \rightarrow 0,
\end{aligned} \tag{A.38}$$

as $n \rightarrow \infty$. The limit in (A.38) is a consequence of (2.17) and of the fact that the multiple integral is finite by the same argument as in Remark 2.3.1. This establishes (A.22) under (A.31).

For $m > 1$, by (2.20) and (2.33) the constants w_k , $k = 1, \dots, m$, in (A.37) cancel out. Moreover, by (A.28) and (A.29), the zero limit in (A.38) still holds; consequently, so does the limit (A.22). \square

PROOF OF LEMMA A.3.3 For $m = 1$, rewrite the sum in (A.23) as

$$\eta^{-s}(n) \zeta^{-s}(h) \left\{ \sum_{\substack{i_1, \dots, i_s=1 \\ |i_1-i_2| \leq h \cup \dots \cup |i_s-i_1| \leq h}}^n + \sum_{\substack{i_1, \dots, i_s=1 \\ |i_1-i_2| \geq h+1, \dots, |i_s-i_1| \geq h+1}}^n \right\} \gamma_h(i_1 - i_2) \dots \gamma_h(i_s - i_1). \tag{A.39}$$

We will show that both multiple summation terms go to zero. We first show this over the index range $|i_1 - i_2| \leq h \cup \dots \cup |i_s - i_1| \leq h$; moreover, as in the proof of Lemma A.3.1, we will only consider the index set (A.32).

Fix the parameter range $0 < \alpha < 3/2$. By (A.31), (A.29) and the Cauchy-Schwarz inequality, the expression (A.33) is bounded in absolute value by

$$C \frac{h^{2(s-1)}}{n^{s/2} h^{s(\alpha+1/2)}} \sum_{\substack{i_1, \dots, i_s=1, i_1 \neq i_2 \\ |i_1-i_2| \leq h, |i_2-i_3| \geq h+1, \dots, |i_s-i_1| \geq h+1}}^n h^\alpha |i_2 - i_3|^{\alpha-2} \dots |i_s - i_1|^{\alpha-2}$$

$$\begin{aligned}
& \leq C \frac{h^{(2-\alpha)(s-1)-s/2}}{n^{s/2}} \sum_{\substack{i_1, \dots, i_{s-1}=1, i_1 \neq i_2 \\ |i_1-i_2| \leq h, |i_2-i_3| \geq h+1, \dots, |i_{s-2}-i_{s-1}| \geq h+1}}^n |i_2 - i_3|^{\alpha-2} \dots |i_{s-2} - i_{s-1}|^{\alpha-2} \\
& \quad \left(\sum_{\substack{i_s=1 \\ |i_{s-1}-i_s| \geq h+1}}^n |i_{s-1} - i_s|^{2(\alpha-2)} \right)^{1/2} \left(\sum_{\substack{i_s=1 \\ |i_s-i_1| \geq h+1}}^n |i_s - i_1|^{2(\alpha-2)} \right)^{1/2} \\
& \leq C \frac{h^{(2-\alpha)(s-1)-s/2}}{n^{s/2}} \left(\sum_{z=h}^n z^{2\alpha-4} \right) \sum_{\substack{i_1, \dots, i_{s-1}=1, i_1 \neq i_2 \\ |i_1-i_2| \leq h, |i_2-i_3| \geq h+1, \dots, |i_{s-2}-i_{s-1}| \geq h+1}}^n |i_2 - i_3|^{\alpha-2} \dots |i_{s-2} - i_{s-1}|^{\alpha-2} \\
& \leq C \frac{h^{(2-\alpha)(s-1)-s/2}}{n^{s/2}} h^{2\alpha-3} \left(\sum_{z=h}^n z^{\alpha-2} \right)^{s-3} (nh) \leq C \frac{h^{(s-2)(3/2-\alpha)+(\alpha-1)}}{n^{s/2-1}} \left(\sum_{z=h}^n z^{\alpha-2} \right)^{s-3}.
\end{aligned} \tag{A.40}$$

In the subranges $0 < \alpha < 1$, $\alpha = 1$, $1 < \alpha < 3/2$, (A.40) is bounded, respectively, by the expressions $C(\frac{h}{n})^{s/2-1}$,

$$C\left(\frac{h}{n}\right)^{s/2-1} \log^{s-3}(n) = \left(\frac{h \log^2(n)}{n}\right)^{s/2-1} \frac{1}{\log(n)},$$

and $C(\frac{h}{n})^{(s-2)(3/2-\alpha)+(\alpha-1)}$, all of which converge to zero as $n \rightarrow \infty$ under (2.17) for $s \geq 3$.

Next consider the case $\alpha = 3/2$. By a simple adaptation of the procedure leading to (A.40), we arrive at the bound

$$C \frac{h^{1/2}}{n^{s/2-1} \log^{s/2}(n)} \left(\sum_{z=h}^n z^{-1} \right) \left(\sum_{z=h}^n z^{-1/2} \right)^{s-3} \leq C \left(\frac{h}{n} \right)^{1/2} \frac{1}{\log^{s/2-1}(n)} \rightarrow 0, \quad n \rightarrow \infty.$$

Therefore, in the parameter range $0 < \alpha \leq 3/2$, by extending the conclusion to the whole summation range of interest,

$$\eta^{-s}(n) \zeta^{-s}(h) \sum_{\substack{i_1, \dots, i_s=1 \\ |i_1-i_2| \leq h \cup \dots \cup |i_s-i_1| \leq h}}^n \gamma_h(i_1 - i_2) \dots \gamma_h(i_s - i_1) \rightarrow 0, \quad n \rightarrow \infty. \tag{A.41}$$

We now show that the multiple summation over the index range $|i_1 - i_2| \geq$

$h, \dots, |i_s - i_1| \geq h$ in (A.39) also goes to zero. Starting from the expression (A.37), by the same argument with the residual function g in the proof of Lemma A.3.2, it suffices to consider

$$\left(\frac{h^2}{\eta(n)\zeta(h)}\right)^s \sum_{\substack{i_1, i_2, \dots, i_s=1 \\ |i_1 - i_2| \geq h+1, |i_2 - i_3| \geq h+1, \dots, |i_s - i_1| \geq h+1}}^n |i_1 - i_2|^{\alpha-2} |i_2 - i_3|^{\alpha-2} \dots |i_s - i_1|^{\alpha-2}.$$

By Cauchy-Schwarz, this expression is bounded from above by

$$\begin{aligned} & \frac{h^{2s}}{\eta^s(n)\zeta^s(h)} \sum_{\substack{i_1, i_2, \dots, i_{s-1}=1 \\ |i_1 - i_2| \geq h+1, \dots, |i_{s-2} - i_{s-1}| \geq h+1}}^n |i_1 - i_2|^{\alpha-2} |i_2 - i_3|^{\alpha-2} \dots |i_{s-2} - i_{s-1}|^{\alpha-2} \\ & \left(\sum_{\substack{i_s=1 \\ |i_{s-1} - i_s| \geq h+1}}^n |i_{s-1} - i_s|^{2\alpha-4} \right)^{1/2} \left(\sum_{\substack{i_s=1 \\ |i_s - i_1| \geq h+1}}^n |i_s - i_1|^{2\alpha-4} \right)^{1/2} \\ & \leq \frac{Ch^{2s}}{\eta^s(n)\zeta^s(h)} \left(\sum_{z=h+1}^n z^{2\alpha-4} \right) \sum_{\substack{i_1, i_2, \dots, i_{s-1}=1 \\ |i_1 - i_2| \geq h+1, \dots, |i_{s-2} - i_{s-1}| \geq h+1}}^n |i_1 - i_2|^{\alpha-2} |i_2 - i_3|^{\alpha-2} \dots |i_{s-2} - i_{s-1}|^{\alpha-2}. \end{aligned} \quad (\text{A.42})$$

However, the multiple summation term in (A.42) is bounded by

$$\begin{aligned} & C \left\{ \sum_{z=h+1}^n z^{\alpha-2} \right\}^{s-3} \sum_{\substack{i_1, i_2=1 \\ |i_1 - i_2| \geq h+1}}^n |i_1 - i_2|^{\alpha-2} \\ & \leq C' \left\{ \sum_{z=h+1}^n z^{\alpha-2} \right\}^{s-3} n \sum_{z=h+1}^n \left(1 - \frac{z}{n}\right) z^{\alpha-2} \leq Cn \left\{ \sum_{z=h+1}^n z^{\alpha-2} \right\}^{s-2}. \end{aligned} \quad (\text{A.43})$$

Therefore, when $\alpha = 3/2$, by (A.43) the expression (A.42) can be bounded by

$$C \frac{\log(n) n n^{(s-2)/2}}{n^{s/2} \log^{s/2}(n)} = C \frac{1}{\log^{s/2-1}(n)} \rightarrow 0, \quad n \rightarrow \infty,$$

as $n \rightarrow \infty$, since $s \geq 3$ and by (2.17).

On the other hand, when $0 < \alpha < 1$, $\alpha = 1$ and $1 < \alpha < 3/2$, the bound for

(A.42) becomes, respectively,

$$C \frac{h^{2s}}{h^{s(\alpha+1/2)}} \frac{n}{n^{s/2}} h^{(\alpha-1)(s-2)} h^{2\alpha-3} = C \left(\frac{h}{n} \right)^{s/2-1} \rightarrow 0, \quad (\text{A.44})$$

$$C \frac{h^{2s}}{h^{3s/2}} \frac{n}{n^{s/2}} \log^{s-2}(n) h^{-1} = C \left(\frac{h \log^2(n)}{n} \right)^{s/2-1} \rightarrow 0, \quad (\text{A.45})$$

and

$$C \frac{h^{2s}}{h^{s(\alpha+1/2)}} \frac{n}{n^{s/2}} n^{(\alpha-1)(s-2)} h^{2\alpha-3} = C \left(\frac{h}{n} \right)^{(s-2)(3/2-\alpha)} \rightarrow 0, \quad (\text{A.46})$$

as $n \rightarrow \infty$. These three limits hold because $s \geq 3$ and by (2.17). Thus, the expressions (A.41), (A.44), (A.45), and (A.46) yield (A.23) for $m = 1$.

For $m > 1$, by (A.28) and (A.29) the zero limits in (A.41), (A.44), (A.45), and (A.46) still hold; consequently, so does (A.23). \square

PROOF OF LEMMA A.3.4 We begin by showing (i) for $m = 1$. Rewrite

$$\eta^{-2}(n) \zeta^{-2}(h) \sum_{i_1, i_2=1}^n \gamma_h^2(i_1 - i_2) = \sum_{z=-n+1}^{n-1} \left(1 - \frac{|z|}{n} \right) (h^{-\alpha} \gamma_h(z))^2 \frac{1}{h}. \quad (\text{A.47})$$

As $n \rightarrow \infty$, the summand in (A.47) goes to, and is also bounded by, $(h^{-\alpha} \gamma_h(z))^2 \frac{1}{h}$.

Therefore, if we can show that

$$\sum_{z=-n+1}^{n-1} (h^{-\alpha} \gamma_h(z))^2 \frac{1}{h} \rightarrow \left(\frac{C_\alpha}{C_H} \right)^4 \|\widehat{G}(x)\|_{L^2(\mathbb{R})}^2, \quad n \rightarrow \infty, \quad (\text{A.48})$$

then (A.24) is obtained as a consequence of the dominated convergence theorem.

Indeed, by setting $w_k = w_l = 1$ and making the change of variables $hx = y$ in the relation (A.1),

$$\left| \frac{\gamma_h(z)}{h^\alpha} - \left(\frac{C_\alpha}{C_H} \right)^2 C_H^2 \int_{\mathbb{R}} e^{iyz/h} \frac{|e^{iy} - 1|^2}{|y|^{\alpha+1}} dy \right| \leq Ch^{-\delta}, \quad h, z \in \mathbb{Z}, \quad h \geq \varepsilon_0^{-2}. \quad (\text{A.49})$$

Therefore,

$$h^{-\alpha}\gamma_h(z) = \left(\frac{C_\alpha}{C_H}\right)^2 G_H\left(\frac{z}{h}\right) + O(h^{-\delta}), \quad (\text{A.50})$$

where G_H denotes the covariance function of a standard fractional Gaussian noise (fGn) $Y(t) = B_H(t) - B_H(t-1)$, $t \in \mathbb{R}$, i.e.,

$$G_H(z) := EY(t)Y(t+z) = \frac{|1+z|^{2H} - 2|z|^{2H} + |1-z|^{2H}}{2}, \quad z \in \mathbb{R}.$$

So, recast the expression on the left-hand side of (A.48) as

$$\left(\frac{C_\alpha}{C_H}\right)^4 \sum_{z=-n+1}^{n-1} G_H^2\left(\frac{z}{h}\right) \frac{1}{h} + CO\left(\frac{1}{h^{1+\delta}}\right) \sum_{z=-n+1}^{n-1} G_H\left(\frac{z}{h}\right) + o(1), \quad (\text{A.51})$$

where the vanishing term $o(1)$ is a consequence of (2.17). Since $G_H(z) \in L^2(\mathbb{R})$ for $0 < H < 3/4$ ($0 < \alpha < 3/2$; see (2.11)), the first summation on the right-hand side of (A.51) converges to

$$\left(\frac{C_\alpha}{C_H}\right)^4 \int_{\mathbb{R}} G_H^2(z) dz = \left(\frac{C_\alpha}{C_H}\right)^4 \int_{\mathbb{R}} |\widehat{G}(x)|^2 dx. \quad (\text{A.52})$$

The equality in (A.52) is a consequence of Parseval's theorem based on the inverse Fourier transform $f(z) = (2\pi)^{-1/2} \int_{\mathbb{R}} e^{izx} \widehat{f}(x) dx$, $f \in L^2(\mathbb{R})$. Moreover,

$$O\left(\frac{1}{h^{1+\delta}}\right) \sum_{z=-n+1}^{n-1} G_H\left(\frac{z}{h}\right) \rightarrow 0, \quad n \rightarrow \infty, \quad (\text{A.53})$$

since the function $G_H(\cdot)$ is bounded and by (2.17). So, by the expressions (A.51), (A.52) and (A.53), we obtain (A.48), and hence (A.24), for $m = 1$.

For $m > 1$, essentially the same argument can be used, and we simply indicate the minor changes. The expression (A.47) must be replaced by

$$\eta^{-2}(n)\zeta^{-1}(h_{k_1})\zeta^{-1}(h_{k_2}) \sum_{i_1, i_2=1}^n \gamma_h^2(i_1-i_2)_{k_1, k_2} = \frac{1}{(w_{k_1} w_{k_2})^{\alpha+1/2}} \sum_{z=-n+1}^{n-1} \left(1 - \frac{|z|}{n}\right) (h^{-\alpha}\gamma_h(z)_{k_1, k_2})^2 \frac{1}{h}.$$

In addition, in expression (A.49) one should substitute $C_H^2 \int_{\mathbb{R}} e^{iyz/h} (e^{iw_{k_1}y} - 1)(e^{iw_{k_2}y} - 1)|y|^{-(\alpha+1)} dy$ for the integral $C_H^2 \int_{\mathbb{R}} e^{iyz/h} |e^{iy} - 1|^2 |y|^{-(\alpha+1)} dy$, where the former can be reinterpreted as the covariance between the increments $B_H(t) - B_H(t - w_{k_1})$ and $B_H(t') - B_H(t' - w_{k_2})$, $t - t' = \frac{z}{h}$, of a standard fBm B_H . The rest of the argument can be applied in the same way to eventually arrive at the limit (A.48) with $\|\widehat{G}(y; w_{k_1}, w_{k_2})\|_{L^2(\mathbb{R})}^2$ in place of $\|\widehat{G}(x)\|_{L^2(\mathbb{R})}^2$. Thus, (A.24) holds also for $m > 1$.

To show (ii) for $m = 1$, note that we can apply (A.29) with $\alpha = 3/2$ in the summation range $|i_1 - i_2| \leq h$ to obtain

$$n^{-1} \log^{-1}(n) h^{-4} \sum_{\substack{i_1, i_2=1 \\ |i_1 - i_2| \leq h}}^n \gamma_h^2(i_1 - i_2) \leq \frac{C}{\log(n)} \rightarrow 0, \quad (\text{A.54})$$

by (2.17). Alternatively, in the summation range $|i_1 - i_2| \geq h + 1$, by (2.20) we have

$$\begin{aligned} \zeta^{-2}(h) \eta^{-2}(n) \sum_{\substack{i_1, i_2=1 \\ |i_1 - i_2| \geq h+1}}^n \gamma_h^2(i_1 - i_2) &= 2n^{-1} \log^{-1}(n) h^{-4} \sum_{z=h+1}^{n-1} n \left(1 - \frac{z}{n}\right) z^{-1} h^4 \{\tau + g(z, h)\}^2 \\ &\sim 2 \log^{-1}(n) \sum_{z=h+1}^{n-1} z^{-1} \{\tau + g(z, h)\}^2 \\ &= 2 \log^{-1}(n) \left\{ \sum_{z=h+1}^{n-1} z^{-1} \tau^2 + \sum_{z=h+1}^{n-1} z^{-1} g^2(z, h) + 2 \sum_{z=h+1}^{n-1} z^{-1} \tau g(z, h) \right\}. \end{aligned} \quad (\text{A.55})$$

Note that for $\beta > 0$ and large enough n , $\int_{h+1}^n z^{-\beta} dz \leq \sum_{z=h+1}^{n-1} z^{-\beta} \leq \int_{h+1}^n (z-1)^{-\beta} dz$.

Consequently, if $\beta = 1$,

$$\frac{\log(n) - \log(h+1)}{\log(n)} \leq \frac{\sum_{z=h+1}^{n-1} z^{-1}}{\log(n)} \leq \frac{\log(n-1) - \log(h)}{\log(n)}.$$

Thus, the left summation term in (A.55) goes to $2\tau^2$ as $n \rightarrow \infty$. We now show that the remaining two terms in (A.55) go to zero with n . It also suffices to look at the

third term in (A.55), because a similar approach can be used with the second term.

Indeed, the former can be bounded by

$$\left| \frac{C}{\log(n)} \sum_{z=h+1}^{n-1} z^{-1} \tau g(z, h) \right| \leq \frac{C'}{\log(n)} \sum_{z=h+1}^{n-1} \left(\frac{z}{h} \right)^{-(1+\delta)} \frac{1}{h} \leq \frac{C'}{\log(n)} \int_1^\infty x^{-(1+\delta)} dx \rightarrow 0,$$

as $n \rightarrow \infty$. Together with (A.54), this establishes (A.25) for $m = 1$.

For $m > 1$, by (2.20) and (2.33) the constants w_k , $k = 1, \dots, m$, in (A.55) cancel out. Moreover, by (A.28) and (A.29), the zero limits in (A.54) and in (A.55) (for the second and third terms) still hold; consequently, so does the limit (A.25). \square

Appendix B

Proofs for Chapter 3

This appendix section comprises three parts. In Section B.1, we present some lemmas that will be used in Section B.2 for the proofs of Theorem 3.2.1 and Theorem 3.2.2. In Section B.3, we develop a technique to approximate the fourth moment of $\widehat{\zeta} - \zeta$ numerically.

As a consequence of Theorem 2.3.1, we standardize $\bar{\mu}_2(h)$ as

$$\varpi(h) = \frac{\bar{\mu}_2(h)}{\mathbb{E}X^2(h)} \xrightarrow{p} 1, \quad (\text{B.1})$$

so that a Taylor expansion can be applied to $\log \varpi(h)$ around 1. Meanwhile, we define the standardized increment

$$W_j(h) = \frac{X(j+h) - X(j)}{\sqrt{\mathbb{E}X^2(h)}}. \quad (\text{B.2})$$

We will use the following results in our proofs. The first one is the classical Isserlis theorem, which reduces the higher moments of a multivariate normal vector to its second moments. The second one is a concentration inequality that will allow us to establish sharp bounds on the tails of centered quadratic forms.

Theorem B.0.1 (Isserlis, [34]). *If $(Z_1, Z_2, \dots, Z_{2n})$ is a zero mean multivariate normal*

random vector, then

$$\mathbb{E}Z_1Z_2\cdots Z_{2n} = \sum \prod \mathbb{E}Z_iZ_j,$$

where the notation $\sum \prod$ means summing over all distinct ways of partitioning Z_1, \dots, Z_{2n} into pairs Z_i, Z_j and each summand is the product of the n pairs.

Theorem B.0.2 ([45]). Let $Z_1, \dots, Z_n \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$. and $\eta_1, \dots, \eta_n \geq 0$, not all zero. Let $\|\boldsymbol{\eta}\|_2$ and $\|\boldsymbol{\eta}\|_\infty$ be the Euclidean square and sup norms of the vector $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$. Also, define the random variable $X = \sum_{i=1}^n \eta_{i,n}(Z_i^2 - 1)$. Then, for every $x > 0$,

$$P(X \geq 2\|\boldsymbol{\eta}\|_2\sqrt{x} + 2\|\boldsymbol{\eta}\|_\infty x) \leq \exp(-x),$$

$$P(X \leq -2\|\boldsymbol{\eta}\|_2\sqrt{x}) \leq \exp(-x).$$

Throughout this section, we will make use of the function $\varrho(h, n, \alpha)$, which was defined in (3.3).

B.1 Some lemmas

Lemma B.1.1. Let $\kappa \in \mathbb{N} \cup \{0\}$, $\kappa \geq 2$, and fix $0 < r < 1/2$. Then, as $n \rightarrow \infty$,

$$\mathbb{E}[(\varpi(h_1) - 1)^\kappa] = O(\varrho^\kappa(h, n, \alpha)). \quad (\text{B.3})$$

Proof. We will only establish (B.3) for $\kappa = 2$, since the remaining cases can be tackled by a similar argument. The left-hand side of (B.3) can be rewritten as

$$\begin{aligned} & \frac{1}{n^2} \mathbb{E} \left[\sum_{k_1=1}^n \sum_{k_2=1}^n \left(\frac{(X(h+k_1) - X(k_1))^2}{\mathbb{E}X^2(h)} - 1 \right) \left(\frac{(X(h+k_1) - X(k_2))^2}{\mathbb{E}X^2(h)} - 1 \right) \right]. \\ & \frac{1}{n^2} \mathbb{E} \sum_{k_1=1}^n \sum_{k_2=1}^n (W_{k_1}^2(h) - 1)(W_{k_2}^2(h) - 1). \end{aligned} \quad (\text{B.4})$$

By applying the Isserlis theorem,

$$\begin{aligned}\mathbb{E}[(W_{k_1}^2(h) - 1)(W_{k_2}^2(h) - 1)] &= \mathbb{E}W_{k_1}^2(h)W_{k_2}^2(h) - \mathbb{E}W_{k_1}^2(h) - \mathbb{E}W_{k_2}^2(h) + 1 \\ &= 2(\mathbb{E}W_{k_1}(h)W_{k_2}(h))^2 = \frac{2}{(\mathbb{E}X^2(h))^2}\gamma_h^2(k_1 - k_2),\end{aligned}$$

where $\gamma_h(k_1 - k_2)$ is defined by (2.16). Thus, (B.4) can be recast as

$$\frac{2}{n^2(\mathbb{E}X^2(h))^2} \sum_{k_1, k_2=1}^n \gamma_h^2(k_1 - k_2), \quad (\text{B.5})$$

Note that

$$\frac{2}{n^2(\mathbb{E}X^2(h))^2} = O\left[\frac{1}{n^2 h^{2\alpha}}\right] = O[\varrho^2(h, n, \alpha)\zeta^{-2}(h)\eta^{-2}(n)],$$

where $\zeta(h), \eta(n)$ are defined by (2.33). Then, by (B.5), Lemmas A.3.1-A.3.4, expression (B.4) is of the order $O(\varrho^2(h, n, \alpha))$ as claimed. \square

Lemma B.1.2. *Fix $-\infty < r < 1/2$. Then, there is $C > 0$, such that*

$$P(\varpi(h) \leq r) \leq \exp\{-C\varrho^{-2}(h, n, \alpha)\}. \quad (\text{B.6})$$

Proof. Let $W_j(h), j = 1, \dots, n$ be as in (B.2). Then for $\varpi(h)$ as in (B.1), we can write

$$\varpi(h) = \frac{1}{n} \sum_{j=1}^n W_j^2(h) = \frac{1}{n} \mathbf{W}_n^T \mathbf{W}_n,$$

where $\mathbf{W}_n = (W_1(h), \dots, W_n(h))^T$ is a multivariate normal vector with covariance matrix Γ . Let Q be an $n \times n$ orthogonal matrix, and let $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_n\}$ be an $n \times n$ diagonal matrix such that $Q\Lambda Q^T = \Gamma$. Then, $\mathbf{W}_n \stackrel{d}{=} Q\Lambda^{1/2}\mathbf{Z}_n$, where $\mathbf{Z}_n = (Z_1, \dots, Z_n)^T \sim N(0, I_n)$, I_n is the $n \times n$ identity matrix, and $\stackrel{d}{=}$ denotes equality

in distribution. Therefore,

$$\varpi(h) \stackrel{d}{=} \frac{1}{n} (Q\Lambda^{1/2}\mathbf{Z}_n)^T Q\Lambda^{1/2}\mathbf{Z}_n = \frac{1}{n} \mathbf{Z}_n^T \Lambda \mathbf{Z}_n = \sum_{j=1}^n \eta_{j,n} Z_j^2, \quad (\text{B.7})$$

where $\eta_{j,n} = \frac{\lambda_j}{n}$. Denote $\boldsymbol{\eta}_n = (\eta_{i,n})_{i=1,\dots,n}$. By Theorem B.0.2 and the same argument as in the proof of Lemma C.3 in [87],

$$P(\varpi(h) \leq r) \leq \exp \left\{ - \frac{C}{\|\boldsymbol{\eta}_n\|_2^2} \right\}$$

for some $C > 0$. By Lemma B.1.1,

$$\|\boldsymbol{\eta}_n\|_2^2 = \text{Var } \varpi(h) = \mathbb{E}[(\varpi(h) - 1)^2] = O(\varrho^2(h, n, \alpha)).$$

Thus, (B.6) follows. □

Corollary B.1.1. *Fix $-\infty < r < 1/2$. For large enough $n \in \mathbb{N}$,*

$$P(\varpi(h) \leq r) = o(\varrho^2(h, n, \alpha)).$$

Lemma B.1.3. *Let $p \geq 1$, there is a constant K_p only depending on p such that*

$$\mathbb{E} |\log \varpi(h)|^p \leq K_p$$

Proof. By (B.7), $\varpi(h)$ is a non-negative weighted sum of independent central chi-squares and all the weights do not vanish. By a similar argument in [63] (see page 184, expression (96)), one can easily prove Lemma B.1.3. □

Lemma B.1.4. *Let $\kappa \in \mathbb{N} \cup \{0\}$, $\kappa \geq 2$, and fix $0 < r < 1/2$. Then, as $n \rightarrow \infty$,*

$$\mathbb{E}[(\varpi(h) - 1)^\kappa 1_{\{\varpi(h) > r\}}] = O(\varrho^\kappa(h, n, \alpha)). \quad (\text{B.8})$$

Proof. By Lemma B.1.1 and Lemma B.1.2, the following expression

$$\begin{aligned}
& \mathbb{E}[(\varpi(h) - 1)^2] - \mathbb{E}[(\varpi(h) - 1)^2 1_{\{\varpi(h) > r\}}] = \mathbb{E}[(\varpi(h) - 1)^2 1_{\{\varpi(h) \leq r\}}] \\
& \leq \sqrt{\mathbb{E}[(\varpi(h) - 1)^4]} \sqrt{\mathbb{E}1_{\{\varpi(h) \leq r\}}} \leq O(\varrho^2(h, n, \alpha)) \sqrt{P(\varpi(h) \leq r)} \\
& \leq O(\varrho^2(h, n, \alpha)) \exp \left\{ -C \varrho^{-2}(h, n, \alpha) \right\} = o(\varrho(h, n, \alpha)^2).
\end{aligned}$$

Therefore, (B.8) holds. \square

Corollary B.1.2. *Let $\kappa_1, \kappa_2 \in \mathbb{N} \cup \{0\}$, $\kappa_1 + \kappa_2 \geq 2$, and fix $0 < r < 1/2$. Then, as $n \rightarrow \infty$,*

$$\mathbb{E}[(\varpi(h_1) - 1)^{\kappa_1} (\varpi(h_2) - 1)^{\kappa_2}] = O(\varrho^{\kappa_1 + \kappa_2}(h, n, \alpha)).$$

$$\mathbb{E}[(\varpi(h_1) - 1)^{\kappa_1} (\varpi(h_2) - 1)^{\kappa_2} 1_{\{\min\{\varpi(h_1), \varpi(h_2)\} > r\}}] = O(\varrho^{\kappa_1 + \kappa_2}(h, n, \alpha)). \quad (\text{B.9})$$

Proof. By Lemma B.1.1, B.1.4 and an application of the Cauchy-Schwarz inequality, (B.9) holds. \square

Lemma B.1.5.

$$\begin{aligned}
& \mathbb{E}(\varpi(h_{k_1}) - 1)(\varpi(h_{k_2}) - 1) = \frac{1}{2n} \sum_{i=-n+1}^{n-1} \left(1 - \frac{|i|}{n}\right) \times \\
& \times \left\{ \left| \frac{i}{\sqrt{h_{k_1} h_{k_2}}} + \sqrt{\frac{h_{k_1}}{h_{k_2}}} \right|^\alpha - \left| \frac{i}{\sqrt{h_{k_1} h_{k_2}}} + \sqrt{\frac{h_{k_1}}{h_{k_2}}} - \sqrt{\frac{h_{k_2}}{h_{k_1}}} \right|^\alpha - \right. \\
& \left. - \left| \frac{i}{\sqrt{h_{k_1} h_{k_2}}} \right|^\alpha + \left| \frac{i}{\sqrt{h_{k_1} h_{k_2}}} - \sqrt{\frac{h_{k_2}}{h_{k_1}}} \right|^\alpha \right\}^2 (1 + O(h^{-\delta})). \quad (\text{B.10})
\end{aligned}$$

Proof. For notational simplicity, assume $k_1 = 1$ and $k_2 = 2$. By (B.2), the left-hand side of (B.10) can be rewritten as

$$\frac{1}{n^2} \sum_{j_1, j_2=1}^n \mathbb{E}(W_{j_1}^2(h_1) - 1)(W_{j_2}^2(h_2) - 1) = \frac{1}{n^2} \sum_{j_1, j_2=1}^n \mathbb{E}W_{j_1}^2(h_1)W_{j_2}^2(h_2) - 1. \quad (\text{B.11})$$

By the Isserlis theorem,

$$\begin{aligned}\mathbb{E}W_{j_1}^2(h_1)W_{j_2}^2(h_2) &= \mathbb{E}W_{j_1}^2(h_1)\mathbb{E}W_{j_2}^2(h_2) + 2\left(\mathbb{E}W_{j_1}(h_1)W_{j_2}(h_2)\right)^2 \\ &= 1 + 2\left(\mathbb{E}W_{j_1}(h_1)W_{j_2}(h_2)\right)^2.\end{aligned}\tag{B.12}$$

By (A.1),

$$\begin{aligned}&\mathbb{E}\left[\frac{(X(j_1+h_1)-X(j_1))(X(j_2+h_2)-X(j_2))}{\sqrt{\mathbb{E}X^2(h_1)}\sqrt{\mathbb{E}X^2(h_2)}}\right] \\ &= \mathbb{E}\left[\frac{(B_H(j_1+h_1)-B_H(j_1))(B_H(j_2+h_2)-B_H(j_2))}{\sqrt{\mathbb{E}B_H^2(h_1)}\sqrt{\mathbb{E}B_H^2(h_2)}}\right](1+O(h^{-\delta})),\end{aligned}\tag{B.13}$$

where B_H is a standard fBm with Hurst parameter $H = \alpha/2$. By (B.2), (B.13) and the explicit expression (2.10),

$$\begin{aligned}\mathbb{E}W_{j_1}(h_1)W_{j_2}(h_2) &= \frac{1}{2}\left(\left|\frac{j_1-j_2}{\sqrt{h_1h_2}} + \sqrt{\frac{h_1}{h_2}}\right|^\alpha - \left|\frac{j_1-j_2}{\sqrt{h_1h_2}} + \sqrt{\frac{h_1}{h_2}} - \sqrt{\frac{h_2}{h_1}}\right|^\alpha - \right. \\ &\quad \left. - \left|\frac{j_1-j_2}{\sqrt{h_1h_2}}\right|^\alpha + \left|\frac{j_1-j_2}{\sqrt{h_1h_2}} - \sqrt{\frac{h_2}{h_1}}\right|^\alpha\right)(1+O(h^{-\delta})).\end{aligned}\tag{B.14}$$

Note that $\mathbb{E}W_{j_1}(h_1)W_{j_2}(h_2) = \mathbb{E}W_{j_1+k}(h_1)W_{j_2+k}(h_2)$ and (B.12), we thus can rewrite (B.11) as

$$\frac{1}{2n} \sum_{i=-n+1}^{n-1} \frac{1}{n} \sum_{j_1-j_2=i, j_1, j_2=1}^n (2\mathbb{E}W_{j_1}(h_1)W_{j_2}(h_2))^2.\tag{B.15}$$

Relation (B.10) is now a consequence of (B.15), (B.12) and (B.14). \square

Lemma B.1.6.

$$\mathbb{E}\log(\varpi(h_{k_1}))\log(\varpi(h_{k_2})) = \mathbb{E}(\varpi(h_{k_1})-1)(\varpi(h_{k_2})-1) + o(\varrho^2(h, n, \alpha)).\tag{B.16}$$

Proof. For notational simplicity, assume $k_1 = 1$ and $k_2 = 2$. Let

$$S_1 = \mathbb{E}\log(\varpi(h_1))\log(\varpi(h_2)) - \mathbb{E}\log(\varpi(h_1))\log(\varpi(h_2))1_{\{\min\{\varpi(h_1), \varpi(h_2)\} > r\}},$$

$$S_2 = \mathbb{E} \log(\varpi(h_1)) \log(\varpi(h_2)) 1_{\{\min\{\varpi(h_1), \varpi(h_2)\} > r\}} -$$

$$- \mathbb{E}(\varpi(h_1) - 1)(\varpi(h_2) - 1) 1_{\{\min\{\varpi(h_1), \varpi(h_2)\} > r\}},$$

$$S_3 = \mathbb{E}(\varpi(h_1) - 1)(\varpi(h_2) - 1) 1_{\{\min\{\varpi(h_1), \varpi(h_2)\} > r\}} - \mathbb{E}(\varpi(h_1) - 1)(\varpi(h_2) - 1).$$

Note that

$$\mathbb{E} \log(\varpi(h_1)) \log(\varpi(h_2)) = S_1 + S_2 + S_3.$$

Therefore, establishing (B.16) is equivalent to showing that $S_1 + S_2 + S_3 = o(\varrho^2(h, n, \alpha))$.

It suffices to show that

$$S_1 = o(\varrho^2(h, n, \alpha)), \quad S_2 = o(\varrho^2(h, n, \alpha)), \quad S_3 = o(\varrho^2(h, n, \alpha)).$$

Let $0 < r < 1/2$. We start with S_2 by writing out the almost sure Taylor expansion

$$\log \varpi(h) 1_{\{\varpi(h) > r\}} = \left\{ (\varpi(h) - 1) - \frac{1}{2} \left(\frac{\varpi(h) - 1}{\theta_+(\varpi(h))} \right)^2 \right\} 1_{\{\varpi(h) > r\}}, \quad (\text{B.17})$$

where $\theta_+(\varpi(h)) \in [\min\{\varpi(h), 1\}, \max\{\varpi(h), 1\}]$. Then,

$$\begin{aligned} & \mathbb{E}[\log \varpi(h_1) \log \varpi(h_2) 1_{\{\min\{\varpi(h_1), \varpi(h_2)\} > r\}}] \\ &= \mathbb{E}[(\varpi(h_1) - 1)(\varpi(h_2) - 1) 1_{\{\min\{\varpi(h_1), \varpi(h_2)\} > r\}}] \\ & \quad - \frac{1}{2} \mathbb{E} \left[(\varpi(h_1) - 1) \left(\frac{\varpi(h_2) - 1}{\theta_+(\varpi(h_2))} \right)^2 1_{\{\min\{\varpi(h_1), \varpi(h_2)\} > r\}} \right] \\ & \quad - \frac{1}{2} \mathbb{E} \left[\left(\frac{\varpi(h_1) - 1}{\theta_+(\varpi(h_1))} \right)^2 (\varpi(h_2) - 1) 1_{\{\min\{\varpi(h_1), \varpi(h_2)\} > r\}} \right] \\ & \quad + \frac{1}{4} \mathbb{E} \left[\left(\frac{\varpi(h_1) - 1}{\theta_+(\varpi(h_1))} \right)^2 \left(\frac{\varpi(h_2) - 1}{\theta_+(\varpi(h_2))} \right)^2 1_{\{\min\{\varpi(h_1), \varpi(h_2)\} > r\}} \right]. \end{aligned} \quad (\text{B.18})$$

The second, third and fourth terms can be bounded by a similar argument, so we

only develop the latter. Recast

$$\begin{aligned} \left(\frac{\varpi(h) - 1}{\theta_+(\varpi(h))} \right)^2 1_{\{\varpi(h) > r\}} &= \left(\frac{\varpi(h) - 1}{\theta_+(\varpi(h))} \right)^2 \left(1_{\{1/2 > \varpi(h) > r\}} + 1_{\{\varpi(h) \geq 1/2\}} \right) \\ &\leq \left(\frac{\varpi(h) - 1}{r} \right)^2 1_{\{1/2 > \varpi(h) > r\}} + \left(\frac{\varpi(h) - 1}{1/2} \right)^2 1_{\{\varpi(h) \geq 1/2\}}. \end{aligned} \quad (\text{B.19})$$

Therefore, we can rewrite the fourth term in (B.18) as

$$\begin{aligned} &\mathbb{E} \left[\left(\frac{\varpi(h_1) - 1}{\theta_+(\varpi(h_1))} \right)^2 \left(\frac{\varpi(h_2) - 1}{\theta_+(\varpi(h_2))} \right)^2 1_{\{\min\{\varpi(h_1), \varpi(h_2)\} > r\}} \right] \\ &\leq \frac{1}{r^4} \mathbb{E}[(\varpi(h_1) - 1)^2 1_{\{1/2 > \varpi(h_1) > r\}} (\varpi(h_2) - 1)^2 1_{\{1/2 > \varpi(h_2) > r\}}] \\ &\quad + \frac{1}{(r/2)^2} \mathbb{E}[(\varpi(h_1) - 1)^2 1_{\{\varpi(h_1) \geq 1/2\}} (\varpi(h_2) - 1)^2 1_{\{1/2 > \varpi(h_2) > r\}}] \\ &\quad + \frac{1}{(r/2)^2} \mathbb{E}[(\varpi(h_1) - 1)^2 1_{\{1/2 > \varpi(h_2) > r\}} (\varpi(h_2) - 1)^2 1_{\{\varpi(h_2) \geq 1/2\}}] \\ &\quad + \frac{1}{(1/2)^4} \mathbb{E}[(\varpi(h_1) - 1)^2 1_{\{\varpi(h_2) \geq 1/2\}} (\varpi(h_2) - 1)^2 1_{\{\varpi(h_2) \geq 1/2\}}] \end{aligned} \quad (\text{B.20})$$

By (B.9), the fourth term in (B.20) is bounded by

$$O(\varrho^4(h, n, \alpha)). \quad (\text{B.21})$$

By the Cauchy-Schwarz inequality, (B.6) and (B.9), the first term in the sum (B.20) is bounded by

$$\begin{aligned} &\frac{1}{r^4} \sqrt{\mathbb{E}[(\varpi(h_1) - 1)^4 (\varpi(h_1) - 1)^4]} \sqrt{\mathbb{E}[1_{\{1/2 > \varpi(h_1) > r\}} 1_{\{1/2 > \varpi(h_2) > r\}}]} \\ &\leq \frac{1}{r^4} O(\varrho^4(h, n, \alpha)) \sqrt{P(1/2 > \varpi(h_1) > r) P(1/2 > \varpi(h_2) > r)} \\ &\leq \frac{1}{r^4} O(\varrho^4(h, n, \alpha)) \exp\{-C\varrho^{-2}(h, n, \alpha)\} = o(\varrho^2(h, n, \alpha)). \end{aligned} \quad (\text{B.22})$$

By the Cauchy-Schwarz inequality, (B.6) and (B.9), the second term in the sum (B.20)

is bounded by

$$\begin{aligned}
& \frac{4}{r^2} \sqrt{\mathbb{E}[(\varpi(h_1) - 1)^4(\varpi(h_1) - 1)^4]} \sqrt{\mathbb{E}[1_{\{\varpi(h_1) \geq 1/2\}} 1_{\{1/2 > \varpi(h_2) > r\}}]} \\
& \leq \frac{4}{r^2} O(\varrho^4(h, n, \alpha)) \sqrt{P(\varpi(h_1) \geq 1/2) P(1/2 > \varpi(h_2) > r)} \\
& \leq \frac{4}{r^2} O(\varrho^4(h, n, \alpha)) \exp\{-C\varrho^{-2}(h, n, \alpha)\} = o(\varrho^2(h, n, \alpha)). \tag{B.23}
\end{aligned}$$

An analogous bound holds for the third term in the sum (B.20). Therefore,

$$S_2 = o(\varrho^2(h, n, \alpha)).$$

Now we tackle S_3 , which can be rewritten as

$$\begin{aligned}
& -\mathbb{E}[(\varpi(h_1) - 1)(\varpi(h_2) - 1)(1_{\{\varpi(h_1) > r\}} 1_{\{\varpi(h_2) \leq r\}} \\
& + 1_{\{\varpi(h_1) \leq r\}} 1_{\{\varpi(h_2) > r\}} + 1_{\{\varpi(h_1) \leq r\}} 1_{\{\varpi(h_2) \leq r\}})]. \tag{B.24}
\end{aligned}$$

By the Cauchy-Schwarz inequality, (B.6) and (B.9), the first term on the right-hand side of (B.24) is bounded by

$$\begin{aligned}
& \sqrt{\mathbb{E}[(\varpi(h_1) - 1)^2(\varpi(h_2) - 1)^2]} \sqrt{P(\varpi(h_2) \leq r)} \\
& \leq O(\varrho^2(h, n, \alpha)) \exp\{-C\varrho^{-2}(h, n, \alpha)\} = o(\varrho^2(h, n, \alpha)).
\end{aligned}$$

Similar bounds hold for the remaining terms on the right-hand side of (B.24). Therefore,

$$S_3 = o(\varrho^2(h, n, \alpha)).$$

Now we deal with S_1 , which can be rewritten as

$$\mathbb{E}[\log \varpi(h_1) \log \varpi(h_2)(1_{\{\varpi(h_1) > r\}} 1_{\{\varpi(h_2) \leq r\}} + 1_{\{\varpi(h_1) \leq r\}} 1_{\{\varpi(h_2) > r\}} + 1_{\{\varpi(h_1) \leq r\}} 1_{\{\varpi(h_2) \leq r\}})]. \quad (\text{B.25})$$

Note that, by Lemma B.1.3, $\mathbb{E}[\log^4 \varpi(h)]$ is bounded. Then, by applying the Cauchy-Schwarz inequality twice, the first term on the right-hand side of (B.25) is bounded by

$$\begin{aligned} & \sqrt{\mathbb{E}[\log^2 \varpi(h_1) \log^2 \varpi(h_2)]} \sqrt{P(\varpi(h_2) \leq r)} \\ & \leq \left(\mathbb{E}[\log^4 \varpi(h_1)] \mathbb{E}[\log^4 \varpi(h_2)] \right)^{1/4} \exp\{-C\varrho^{-2}(h, n, \alpha)\} = o(\varrho^2(h, n, \alpha)). \end{aligned}$$

Similar bounds hold for the remaining terms on the right-hand side of (B.25). Therefore,

$$S_1 = o(\varrho^2(h, n, \alpha)).$$

Thus, (B.16) follows. \square

Lemma B.1.7.

$$\mathbb{E} \log(\varpi(h)) + \frac{1}{2} \mathbb{E}(\varpi(h) - 1)^2 = o(\varrho^2(h, n, \alpha)). \quad (\text{B.26})$$

Proof. Fix $0 < r < 1/2$. Let

$$\begin{aligned} T_1 &= \mathbb{E} \log(\varpi(h)) - \mathbb{E} \log(\varpi(h)) 1_{\{\varpi(h) > r\}}, \\ T_2 &= \mathbb{E} \log(\varpi(h)) 1_{\{\varpi(h) > r\}} + \frac{1}{2} \mathbb{E}(\varpi(h) - 1)^2 1_{\{\varpi(h) > r\}}, \\ T_3 &= \frac{1}{2} \mathbb{E}(\varpi(h) - 1)^2 - \frac{1}{2} \mathbb{E}(\varpi(h) - 1)^2 1_{\{\varpi(h) > r\}}. \end{aligned}$$

Note that, by Lemma B.1.3, $\mathbb{E} \log^2(\varpi(h))$ is bounded, by the Cauchy-Schwarz in-

equality and Lemma B.1.2,

$$\begin{aligned} T_1 &= \mathbb{E} \log(\varpi(h)) 1_{\{\varpi(h) \leq r\}} \leq \sqrt{\mathbb{E} \log^2(\varpi(h))} \sqrt{P(\varpi(h) \leq r)} \\ &\leq \sqrt{\mathbb{E} \log^2(\varpi(h))} \exp\{-C\varrho^{-2}(h, n, \alpha)\} = o(\varrho^2(h, n, \alpha)). \end{aligned} \quad (\text{B.27})$$

By a similar reasoning, we can further prove that

$$T_3 = o(\varrho^2(h, n, \alpha)). \quad (\text{B.28})$$

Now, we turn to T_2 . By an almost sure Taylor expansion,

$$\log \varpi(h) 1_{\{\varpi(h) > r\}} = \left\{ (\varpi(h) - 1) - \frac{1}{2}(\varpi(h) - 1)^2 + \frac{1}{3} \left(\frac{\varpi - 1}{\theta_+(\varpi)} \right)^3 \right\} 1_{\{\varpi(h) > r\}},$$

where $\theta_+(\varpi(h)) \in [\min\{\varpi(h), 1\}, \max\{\varpi(h), 1\}]$. Then, T_2 is bounded by

$$\left| \mathbb{E}(\varpi(h) - 1) 1_{\{\varpi(h) > r\}} \right| + \left| \mathbb{E} \frac{1}{3} \left(\frac{\varpi - 1}{\theta_+(\varpi)} \right)^3 1_{\{\varpi(h) > r\}} \right|. \quad (\text{B.29})$$

Since $\mathbb{E}(\varpi(h) - 1) = 0$, by the Cauchy-Schwarz inequality, (B.9) and Corollary B.1.1, the first term in (B.29) can be rewritten as

$$\begin{aligned} \left| \mathbb{E}(\varpi(h) - 1) 1_{\{\varpi(h) > r\}} \right| &= \left| \mathbb{E}(\varpi(h) - 1) 1_{\{\varpi(h) > r\}} - \mathbb{E}(\varpi(h) - 1) \right| \\ &= \left| \mathbb{E}(\varpi(h) - 1) 1_{\{\varpi(h) \leq r\}} \right| \leq \sqrt{\mathbb{E}(\varpi(h) - 1)^2} \sqrt{P(\varpi(h) \leq r)} \\ &= O(\varrho(h, n, \alpha)) o(\varrho(h, n, \alpha)) = o(\varrho^2(h, n, \alpha)). \end{aligned}$$

Meanwhile, by Lemma B.1.1 and the Cauchy-Schwarz inequality, the second term in (B.29) is bounded by

$$\frac{1}{3r^3} \left| \mathbb{E}(\varpi - 1)^3 1_{\{1/2 > \varpi(h) > r\}} \right| + \frac{1}{3(1/2)^3} \left| \mathbb{E}(\varpi - 1)^3 1_{\{\varpi(h) \geq 1/2\}} \right|$$

$$\begin{aligned}
&\leq \frac{1}{3r^3} \sqrt{\mathbb{E}(\varpi - 1)^6} P(1/2 > \varpi(h) > r) + \frac{1}{3(1/2)^3} \sqrt{\mathbb{E}(\varpi - 1)^6} P(\varpi(h) \geq 1/2) \\
&\leq \frac{1}{3r^3} O(\varrho^3(h, n, \alpha)) \exp\{-C\varrho^{-2}(h, n, \alpha)\} + \frac{1}{3(1/2)^3} O(\varrho^3(h, n, \alpha)) = o(\varrho^2(h, n, \alpha)).
\end{aligned} \tag{B.30}$$

Thus,

$$T_2 = o(\varrho^2(h, n, \alpha)). \tag{B.31}$$

Relations (B.27), (B.28) and (B.31) imply (B.26). \square

B.2 Proofs of Theorem 3.2.1 and Theorem 3.2.2

PROOF OF THEOREM 3.2.2: For $k_1, k_2 = 1, \dots, m$, rewrite

$$\begin{aligned}
\tilde{\sigma}_{k_1, k_2} &= \text{Cov}(\log \bar{\mu}_2(h_{k_1}), \log \bar{\mu}_2(h_{k_2})) \\
&= \mathbb{E}[\log \bar{\mu}_2(h_{k_1}) - \mathbb{E} \log \bar{\mu}_2(h_{k_1})][\log \bar{\mu}_2(h_{k_2}) - \mathbb{E} \log \bar{\mu}_2(h_{k_2})] \\
&= \mathbb{E}[\log \varpi(h_{k_1}) - \mathbb{E} \log \varpi(h_{k_1})][\log \varpi(h_{k_2}) - \mathbb{E} \log \varpi(h_{k_2})] \\
&= \mathbb{E}[\log \varpi(h_{k_1}) \log \varpi(h_{k_2})] - \mathbb{E} \log \varpi(h_{k_1}) \mathbb{E} \log \varpi(h_{k_2}).
\end{aligned} \tag{B.32}$$

By Lemmas B.1.1 and B.1.7,

$$\mathbb{E} \log \varpi(h_{k_1}) = O(\varrho^2(h, n, \alpha)).$$

Therefore, (B.32) can be reexpressed as

$$\mathbb{E}[\log \varpi(h_{k_1}) \log \varpi(h_{k_2})] + o(\varrho^2(h, n, \alpha)). \tag{B.33}$$

By Lemmas B.1.5 and Lemma B.1.6, expression (3.7) holds. \square

PROOF OF THEOREM 3.2.1: The left-hand side of (3.2) can be rewritten as

$$\mathbb{E} \log \frac{\bar{\mu}_2(h_k)}{\mathbb{E}X^2(h_k)} + \log \frac{\mathbb{E}X^2(h_k)}{\theta h_k^\alpha} = \mathbb{E} \log \varpi(h_k) + \log \frac{\mathbb{E}X^2(h_k)}{\theta h_k^\alpha}. \quad (\text{B.34})$$

As shown in [23], as $n \rightarrow +\infty$, the second sum term on the right-hand side of (B.34) behaves like

$$\log(1 + O(h^{-\delta})) = O(h^{-\delta}).$$

By Lemmas B.1.7 and B.1.5, we can recast the first sum term on the right-hand side of (B.34) as

$$\begin{aligned} & -\frac{1}{2} \mathbb{E}(\varpi(h) - 1)^2 + o(\varrho^2(h, n, \alpha)) \\ & -\frac{1}{4n} \sum_{i=-n+1}^{n-1} \left(1 - \frac{|i|}{n}\right) \left\{ \left| \frac{i}{h_k} + 1 \right|^\alpha - 2 \left| \frac{i}{h_k} \right|^\alpha + \left| \frac{i}{h_k} - 1 \right|^\alpha \right\}^2 + O(h^{-\delta}) + o(\varrho^2(h, n, \alpha)). \end{aligned}$$

Thus, (3.2) follows. \square

B.3 A numerical approximation of κ_1 and κ_2

We only develop the numerical approximation of κ_1 , since that of κ_2 can be obtained by a similar argument. Recall that $\kappa_1 = \mathbb{E}[\widehat{\zeta}_{\cdot,1} - \mathbb{E}\widehat{\zeta}_{\cdot,1}]^4$, where $\widehat{\zeta}_{\cdot,1}$ is a linear combination of $\widehat{\alpha}$ and $\widehat{\log \theta}$, which can be further written as a linear combination of $\log \bar{\mu}_2(h_k)$, $k = 1, \dots, m$. In fact, let

$$(c_1, \dots, c_m)^T = (1, 0)(X^T \widetilde{\Sigma}^{-1} X)^{-1/2} X^T \widetilde{\Sigma}^{-1}.$$

Then, by (3.4), (3.5) and (3.9),

$$\widehat{\zeta}_{\cdot,1} = (c_1, \dots, c_m)(\log \bar{\mu}_2(h_1), \dots, \log \bar{\mu}_2(h_m))^T.$$

Thus,

$$\kappa_1 = \mathbb{E} \left(\sum_{k=1}^m c_k (\log \bar{\mu}_2(h_k) - \mathbb{E} \log \bar{\mu}_2(h_k)) \right)^4 \quad (\text{B.35})$$

By Theorem 3.2.1, we can approximate (B.35) by

$$\mathbb{E} \left(\sum_{k=1}^m c_k \left(\log \varpi(h_k) + \log \frac{\mathbb{E} X^2(h_k)}{\theta h_k^\alpha} \right) \right)^4 = \mathbb{E} \left(\sum_{k=1}^m c_k (\log \varpi(h_k) + O(h^{-\delta})) \right)^4. \quad (\text{B.36})$$

By Lemma B.1.3, $\mathbb{E} \log \varpi^r(h_k)$, $r = 1, 2, 3, 4$ is bounded in absolute value. Thus, (B.36) can be rewritten as

$$O(h^{-\delta}) + \sum_{k_1, k_2, k_3, k_4=1}^m c_{k_1} c_{k_2} c_{k_3} c_{k_4} \mathbb{E} \log \varpi(h_{k_1}) \log \varpi(h_{k_2}) \log \varpi(h_{k_3}) \log \varpi(h_{k_4}). \quad (\text{B.37})$$

By a similar argument to the proof of Lemma B.1.6, it can be shown that

$$\begin{aligned} & \mathbb{E} \log \varpi(h_{k_1}) \log \varpi(h_{k_2}) \log \varpi(h_{k_3}) \log \varpi(h_{k_4}) \\ &= \mathbb{E} (\varpi(h_{k_1}) - 1)(\varpi(h_{k_2}) - 1)(\varpi(h_{k_3}) - 1)(\varpi(h_{k_4}) - 1) + o(\varrho^4(h, n, \alpha)). \end{aligned} \quad (\text{B.38})$$

For notational simplicity, assume $k_i = i$, $i = 1, 2, 3, 4$. The first term in (B.38) is

$$\begin{aligned} & \frac{1}{n^4} \sum_{j_1, j_2, j_3, j_4=1}^n \mathbb{E} (W_{j_1}^2(h_1) - 1)(W_{j_2}^2(h_2) - 1)(W_{j_3}^2(h_3) - 1)(W_{j_4}^2(h_4) - 1) \\ &= \frac{1}{n^4} \sum_{j_1, j_2, j_3, j_4=1}^n \mathbb{E} W_{j_1}^2(h_1) W_{j_2}^2(h_2) W_{j_3}^2(h_3) W_{j_4}^2(h_4) \\ & - \frac{1}{n^3} \sum_{(i_1, i_2, i_3, i_4) \in S(4,1)} \sum_{j_{i_1}, j_{i_2}, j_{i_3}=1}^n \mathbb{E} W_{j_{i_1}}^2(h_{i_1}) W_{j_{i_2}}^2(h_{i_2}) W_{j_{i_3}}^2(h_{i_3}) \\ & + \frac{1}{n^2} \sum_{(i_1, i_2, i_3, i_4) \in S(4,2)} \sum_{j_{i_1}, j_{i_2}=1}^n \mathbb{E} W_{j_{i_1}}^2(h_{i_1}) W_{j_{i_2}}^2(h_{i_2}) - 3, \end{aligned} \quad (\text{B.39})$$

where

$$S(4, 1) = \{(1, 2, 3, 4), (2, 3, 4, 1), (3, 4, 1, 2), (4, 1, 2, 3)\},$$

$$S(4, 2) = \{(1, 2, 3, 4), (1, 3, 2, 4), (2, 3, 1, 4), (1, 4, 2, 3), (2, 4, 1, 3), (3, 4, 1, 2)\}.$$

Since $W_{j_i}(h_i), i = 1, 2, 3, 4$ are mean zero normal random variables, by the Isserlis theorem, we can expand each moment

$$\mathbb{E}[W_{j_1}^2(h_1)W_{j_2}^2(h_2)W_{j_3}^2(h_3)W_{j_4}^2(h_4)], \mathbb{E}[W_{j_1}^2(h_{i_1})W_{j_2}^2(h_{i_2})W_{j_3}^2(h_{i_3})], \mathbb{E}[W_{j_1}^2(h_{i_1})W_{j_2}^2(h_{i_2})],$$

into summation of products of terms of the form $\mathbb{E}W_{j_{i_1}}(h_{i_1})W_{j_{i_2}}(h_{i_2})$. Therefore, the numerical approximation of κ_1 boils down to the numerical approximation of $\mathbb{E}W_{j_{i_1}}(h_{i_1})W_{j_{i_2}}(h_{i_2})$, where, by (B.14), the latter can be approximated as

$$\begin{aligned} & \frac{1}{2} \left(\left| \frac{j_1 - j_2}{\sqrt{h_1 h_2}} + \sqrt{\frac{h_1}{h_2}} \right|^\alpha - \left| \frac{j_1 - j_2}{\sqrt{h_1 h_2}} + \sqrt{\frac{h_1}{h_2}} - \sqrt{\frac{h_2}{h_1}} \right|^\alpha - \right. \\ & \quad \left. - \left| \frac{j_1 - j_2}{\sqrt{h_1 h_2}} \right|^\alpha + \left| \frac{j_1 - j_2}{\sqrt{h_1 h_2}} - \sqrt{\frac{h_2}{h_1}} \right|^\alpha \right). \end{aligned}$$

Appendix C

Proofs for Chapter 4

Lemma C.0.1. *Let $\{\gamma_{\tilde{Z}_n}(k)\}_{k \in \mathbb{Z}}$ be the autocovariance function of the sequence $\{\tilde{Z}_n(k)\}_{k \in \{0\} \cup \mathbb{Z}}$ (see (4.23)). Then,*

$$\sum_{k=1}^{\infty} |\gamma_{\tilde{Z}_n}(k)| < \infty, \quad \sum_{k=1}^{\infty} |\gamma_{\tilde{Z}_n}(k)|^2 < \infty. \quad (\text{C.1})$$

Proof. We start with the autocovariance function of $\{Y(k)\}_{k \in \{0\} \cup \mathbb{Z}}$. By a Taylor expansion of order 2,

$$\begin{aligned} \gamma_Y(k) &= \frac{1}{2} |k|^\alpha \left(\left| 1 - \frac{1}{k} \right|^\alpha - 2 + \left| 1 + \frac{1}{k} \right|^\alpha \right) \\ &= \frac{1}{2} |k|^\alpha \left(1 - \alpha k^{-1} + \frac{\alpha(\alpha-1)}{2} k^{-2} + O(k^{-3}) - 2 + 1 + \alpha k^{-1} + \frac{\alpha(\alpha-1)}{2} k^{-2} + O(k^{-3}) \right) \\ &= \frac{\alpha(\alpha-1)}{2} k^{\alpha-2} + O(k^{\alpha-3}), \end{aligned} \quad (\text{C.2})$$

as $k \rightarrow \infty$. Consider the standardized increments of fGn, $\tilde{Z}_n(k) = \sigma^2(n)[Y(n+k) - Y(k)]$. By a Taylor expansion of order 1, for $k \in \mathbb{Z}$,

$$\gamma_{\tilde{Z}_n}(k) = \sigma^2(n) \mathbb{E}(Y(n+k+j) - Y(k+j))(Y(n+j) - Y(j))$$

$$\begin{aligned}
&= -\sigma^2(n)[\gamma_Y(k+n) - 2\gamma_Y(k) + \gamma_Y(k-n)] \\
&= -\frac{\sigma^2(n)\alpha(\alpha-1)}{2}(|k+n|^{\alpha-2} - 2|k|^{\alpha-2} + |k-n|^{\alpha-2}) + O(k^{\alpha-3}) \\
&= \frac{\sigma^2(n)\alpha(\alpha-1)}{2}|k|^{\alpha-2} \left(2 - \left| 1 - \frac{n}{k} \right|^{\alpha-2} - \left| 1 + \frac{n}{k} \right|^{\alpha-2} \right) + O(k^{\alpha-3}) \\
&= \frac{\sigma^2(n)\alpha(\alpha-1)}{2}|k|^{\alpha-2} O(k^{-1}) + O(k^{\alpha-3}) = O(k^{\alpha-3}),
\end{aligned}$$

as $k \rightarrow \infty$. Therefore, (C.1) holds. \square

PROOF OF PROPOSITION 4.3.1:

Proof. Note that, for $N > n - 1$,

$$\begin{aligned}
&\sqrt{N-n+1}(\mathfrak{R}\widehat{E}_1(n) - \mathbb{E} \cos(Y(n) - Y(0))) \\
&= \frac{1}{\sqrt{N-n+1}} \sum_{k=0}^{N-n} \cos(Y(n+k) - Y(k)) - \mathbb{E} \cos(Y(n) - Y(0)) \\
&= \frac{1}{\sqrt{N-n+1}} \sum_{k=0}^{N-n} \left[\cos\left(\frac{\widetilde{Z}_n(k)}{\sigma(n)}\right) - \mathbb{E} \cos\left(\frac{\widetilde{Z}_n(k)}{\sigma(n)}\right) \right].
\end{aligned}$$

Then, by Lemma C.0.1 and Theorem 4.2.2, expression (4.26) holds. An analogous reasoning further establishes (4.27). \square

PROOF OF PROPOSITION 4.3.2:

Proof. We prove (i) first. By expanding $\widehat{E}_2(n)$, we can rewrite the left-hand side of (4.29) as

$$\begin{aligned}
&\sqrt{N+1} \left\{ \left| \mathbb{E} \cos(Y(0)) + \frac{1}{N+1} \sum_{k=0}^N \left(\cos(Y(k)) - \mathbb{E} \cos(Y(k)) \right) \right|^2 \right. \\
&\quad \left. + \left| \frac{1}{N+1} \sum_{k=0}^N \sin(Y(k)) \right|^2 - |\mathbb{E} \cos(Y(0))|^2 \right\}
\end{aligned}$$

$$\begin{aligned}
&= 2\mathbb{E} \cos(Y(0)) \frac{1}{\sqrt{N+1}} \sum_{k=0}^N \left(\cos(Y(k)) - \mathbb{E} \cos(Y(k)) \right) \\
&+ \frac{1}{\sqrt{N+1}} \left(\frac{1}{\sqrt{N+1}} \sum_{k=0}^N \left(\cos(Y(k)) - \mathbb{E} \cos(Y(k)) \right) \right)^2 \\
&+ \left(\frac{1}{(N+1)^{3/4}} \sum_{k=0}^N \sin(Y(k)) \right)^2. \tag{C.3}
\end{aligned}$$

We will show that, as $N \rightarrow \infty$, the first term in the sum (C.3) converges to a non-degenerate random variable in distribution, and that the second and third terms in (C.3) converges to zero in probability. Note that, when $0 < \alpha < 3/2$,

$$\sum_{k=1}^{\infty} |\gamma_Y(k)|^2 < \infty.$$

By Theorem 4.2.2,

$$\frac{1}{\sqrt{N+1}} \sum_{k=0}^N \left(\cos(Y(k)) - \mathbb{E} \cos(Y(0)) \right) \xrightarrow{d} \sigma_3 B_3(1), \tag{C.4}$$

where

$$\sigma_3^2 = \sum_{m=2}^{\infty} g_{\cos,m}^2 m! \sum_{k=-\infty}^{\infty} (\gamma_Y(k))^m.$$

Since $\frac{1}{\sqrt{N+1}} \rightarrow 0$, by (C.4) and Slutsky's theorem, the second term in the sum (C.3) convergence to zero in probability. As for the third term in the sum (C.3), when $0 < \alpha \leq 1$,

$$\sum_{k=1}^{\infty} |\gamma_Y(k)| < \infty.$$

Thus, by Theorem 4.2.2,

$$\frac{1}{\sqrt{N+1}} \sum_{k=0}^N \sin(Y(k)) \xrightarrow{d} \sigma_4 B_4(1). \tag{C.5}$$

where

$$\sigma_4^2 = \sum_{m=1}^{\infty} g_{\sin,m}^2 m! \sum_{k=-\infty}^{\infty} (\gamma_Y(k))^m.$$

Then, the third term in the sum (C.3) can be written as

$$\frac{1}{\sqrt{N+1}} \left(\frac{1}{\sqrt{N+1}} \sum_{k=0}^N \sin(Y(k)) \right)^2 \xrightarrow{P} 0, \quad N \rightarrow \infty,$$

which is a consequence of (C.5) and Slutsky's theorem. On the other hand, when $1 < \alpha < 3/2$, $d \in (0, \frac{1}{2})$. Thus, by Theorem 4.2.1,

$$\frac{1}{\sqrt{L_2(N+1)}(N+1)^{\alpha/2}} \sum_{k=0}^N \sin(Y(k)) \xrightarrow{d} g_1 \beta_{1,H} Z_H^{(1)}(1). \quad (\text{C.6})$$

Then, the third term in the sum (C.3) can be written as

$$\frac{L_2(N+1)}{(N+1)^{3/2-\alpha}} \left(\frac{1}{\sqrt{L_2(N+1)}(N+1)^{\alpha/2}} \sum_{k=0}^N \sin(Y(k)) \right)^2 \xrightarrow{P} 0, \quad N \rightarrow \infty,$$

which results from (C.6) and Slutsky's theorem. Thus, the expression (4.29) follows.

We now prove (ii). Rewrite the left hand side of (4.31) as

$$\begin{aligned} & \frac{(N+1)^{2-\alpha}}{L_2(N+1)} \left\{ \left| \mathbb{E} \cos(Y(0)) + \frac{1}{N+1} \sum_{k=0}^N \left(\cos(Y(k)) - \mathbb{E} \cos(Y(k)) \right) \right|^2 \right. \\ & \quad \left. + \left| \frac{1}{N+1} \sum_{k=0}^N \sin(Y(k)) \right|^2 - |\mathbb{E} \cos(Y(0))|^2 \right\} \\ & = 2\mathbb{E} \cos(Y(0)) \frac{1}{L_2(N+1)(N+1)^{\alpha-1}} \sum_{k=0}^N \left(\cos(Y(k)) - \mathbb{E} \cos(Y(k)) \right) \\ & \quad + \frac{L_2(N+1)}{(N+1)^{2-\alpha}} \left(\frac{1}{L_2(N+1)(N+1)^{\alpha-1}} \sum_{k=0}^N \left(\cos(Y(k)) - \mathbb{E} \cos(Y(k)) \right) \right)^2 \end{aligned}$$

$$+ \left(\frac{1}{\sqrt{L_2(N+1)}(N+1)^{\alpha/2}} \sum_{k=0}^N \sin(Y(k)) \right)^2. \quad (\text{C.7})$$

When $3/2 < \alpha < 2$, by Theorem 4.2.4,

$$\begin{aligned} & \left(\frac{1}{L_2(N+1)(N+1)^{\alpha-1}} \sum_{k=0}^N \cos(Y(k)) - \mathbb{E} \cos(Y(k)), \frac{1}{\sqrt{L_2(N+1)}(N+1)^{\alpha/2}} \sum_{k=0}^N \sin Y(k) \right)^T \\ & \xrightarrow{d} \left(g_{1,2} \beta_{2,\alpha-1} \widehat{I}_2(f_{\alpha-1,2,1}), g_{2,1} \beta_{1,\alpha/2} \widehat{I}_1(f_{\alpha/2,1,1}) \right)^T. \end{aligned} \quad (\text{C.8})$$

By Slutsky's theorem and (C.8), the second term in the sum (C.7) converges to zero in probability. Then, by (C.8), relation (4.31) holds. \square

PROOF OF THEOREM 4.3.1:

Proof. We prove (i) firstly. Note that, when $0 < \alpha < 3/2$, by Propositions 4.3.1 and 4.3.2 and Slutsky theorem,

$$\begin{aligned} & \sqrt{N+1} [\Re \widehat{E}(n) - \mathbb{E} \cos(Y(n) - Y(0)) + |\mathbb{E} \cos(Y(0))|^2] \\ &= \frac{\sqrt{N+1}}{\sqrt{N-n+1}} \sqrt{N-n+1} [\Re \widehat{E}_1(n) - \mathbb{E} \cos(Y(n) - Y(0))] \\ & \quad - \sqrt{N+1} [\widehat{E}_2(n) - |\mathbb{E} \cos(Y(0))|^2] \\ & \xrightarrow{d} \sigma_1 B_1(1) + 2 \mathbb{E} \cos(Y(0)) \sigma_3 B_3(1), \end{aligned}$$

where $(B_1(1), B_3(1))^T$ is a Gaussian vector since (Y, \widetilde{Z}_n) is a Gaussian vector. Let $G_1(x) = \cos(x/\sigma(n))$, $G_2(x) = \cos(x)$. By Proposition 4.2.1,

$$\text{Cov} \left(\cos(\widetilde{Z}_n(k+j)), \cos(Y(j)) \right) = \sum_{m=2}^{\infty} g_{1,n,m} g_{\cos,m} m! (\gamma_{\widetilde{Z}_n, Y}(k))^m,$$

since $g_{1,n,1} = g_{\cos,1} = 0$. Therefore, expression (4.32) holds. This establishes (i).

Now we show (ii). When $3/2 < \alpha < 2$, by Propositions 4.3.1 and 4.3.2, and

Slutsky's theorem,

$$\begin{aligned}
& \frac{(N+1)^{2-\alpha}}{L_2(N+1)} [\Re \widehat{E}(n) - \mathbb{E} \cos(Y(n) - Y(0)) + |\mathbb{E} \cos(Y(0))|^2] \\
&= \frac{(N+1)^{2-\alpha}}{L_2(N+1) \sqrt{N-n+1}} \sqrt{N-n+1} [\Re \widehat{E}_1(n) - \mathbb{E} \cos(Y(n) - Y(0))] \\
&\quad - \frac{(N+1)^{2-\alpha}}{L_2(N+1)} [\widehat{E}_2(n) - |\mathbb{E} \cos(Y(0))|^2] \\
&\xrightarrow{d} g_{1,2} \beta_{2,\alpha-1} \widehat{I}_2(f_{\alpha-1,2,1}) + g_{2,1}^2 \beta_{1,\alpha/2}^2 \widehat{I}_1^2(f_{\alpha/2,1,1}).
\end{aligned}$$

Thus, (ii) holds.

Statement (iii) is a consequence of the fact that $\Im \widehat{E}(n) = \Im \widehat{E}_1(n)$, and of Proposition 4.3.1 and Slutsky's theorem. \square

Bibliography

- [1] A. Andreanov and D. Grebenkov. Time-averaged MSD of Brownian motion. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(07):P07001, 2012.
- [2] M. Badino. The foundational role of ergodic theory. *Foundations of Science*, 11(4):323–347, December 2006.
- [3] J.-M. Bardet. Testing for the presence of self-similarity of Gaussian time series having stationary increments. *Journal of Time Series Analysis*, 21(5):497–515, 2000.
- [4] E. Barkai, Y. Garini, and R. Metzler. Strange kinetics of single molecules in living cells. *Physics Today*, 65(8):29–35, 2012.
- [5] A. J. Berglund. Statistics of camera-based single-particle tracking. *Physical Review E*, 82(1):011917, 2010.
- [6] S. Berning, K. I. Willig, H. Steffens, P. Diba, and S. W. Hell. Nanoscopy in a living mouse brain. *Science*, 335(6068):551–551, 2012.
- [7] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacio, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645, 2006.
- [8] D. Boyer, D. S. Dean, C. Mejía-Monasterio, and G. Oshanin. Optimal estimates of the diffusion coefficient of a single Brownian trajectory. *Physical Review E*, 85(3):031136, 2012.
- [9] D. Boyer, D. S. Dean, C. Mejía-Monasterio, and G. Oshanin. Distribution of the least-squares estimators of a single Brownian trajectory diffusion coefficient. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(04):P04017, 2013.
- [10] X. Brokman, J.-P. Hermier, G. Messin, P. Desbiolles, J.-P. Bouchaud, and M. Dahan. Statistical aging and nonergodicity in the fluorescence of single nanocrystals. *Physical Review Letters*, 90(12):120601, 2003.
- [11] B. Buchmann and N. H. Chan. Integrated functionals of normal and fractional processes. *Annals of Applied Probability*, 19(1):49–70, 2009.

- [12] K. Burnecki. FARIMA processes with application to biophysical data. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(05):P05015, 2012.
- [13] K. Burnecki, E. Kepten, Y. Garini, G. Sikora, and A. Weron. Estimating the anomalous diffusion exponent for single particle tracking data with measurement errors-an alternative approach. *Scientific Reports*, 5(11306):1–11, 2015.
- [14] K. Burnecki, E. Kepten, J. Janczura, I. Bronshtein, Y. Garini, and A. Weron. Universal algorithm for identification of fractional brownian motion. a case of telomere subdiffusion. *Biophysical Journal*, 103(9):1839–1847, 2012.
- [15] S. Burov, J.-H. Jeon, R. Metzler, and E. Barkai. Single particle tracking in systems showing anomalous diffusion: the role of weak ergodicity breaking. *Physical Chemistry Chemical Physics*, 13(5):1800–1812, 2011.
- [16] P. Cheridito, H. Kawaguchi, and M. Maejima. Fractional Ornstein-Uhlenbeck processes. *Electronic Journal of Probability*, 8(3):1–14, 2003.
- [17] R. Christensen. *Plane answers to complex questions: the theory of linear models*. Springer Science & Business Media, 2011.
- [18] H. Cramér and M. R. Leadbetter. *Stationary and related stochastic processes: sample function properties and their applications*. Courier Dover Publications, 1967.
- [19] M. Dawson, D. Wirtz, and J. Hanes. Enhanced viscoelasticity of human cystic fibrotic sputum correlates with increasing microheterogeneity in particle transport. *Journal of Biological Chemistry*, 278(50):50393–50401, 2003.
- [20] W. Deng and E. Barkai. Ergodic properties of fractional Brownian-Langevin motion. *Physical Review E*, 79(1):011112, 2009.
- [21] G. Didier and J. Fricks. On the wavelet-based simulation of anomalous diffusion. *Journal of Statistical Computation and Simulation*, 84(4):697–723, 2014.
- [22] G. Didier, S.A. McKinley, D.B. Hill, and J. Fricks. Statistical challenges in microrheology. *Journal of Time Series Analysis*, 33(55):724–743, September 2012.
- [23] G. Didier and K. Zhang. The asymptotic distribution of the pathwise mean-square displacement in single-particle tracking experiments. *To appear in Journal of Time Series Analysis*, 2016.
- [24] G. Didier and K. Zhang. The asymptotic distribution of the pathwise mean-square displacement in single-particle tracking experiments: supporting information. *To appear in Journal of Time Series Analysis*, 2016.
- [25] D. Grebenkov. Probability distribution of the time-averaged mean-square displacement of a Gaussian process. *Physical Review E*, 84(3):031124, 2011.

- [26] D. Grebenkov. Time-averaged quadratic functionals of a Gaussian process. *Physical Review E*, 83(6):061117, 2011.
- [27] Xavier Guyon and José León. Convergence en loi des H -variations d'un processus Gaussien stationnaire sur R . 25(3):265–282, 1989.
- [28] S. Havlin and D. Ben-Avraham. Diffusion in disordered media. *Advances in Physics*, 36(6):695–798, 1987.
- [29] S. W. Hell. Toward fluorescence nanoscopy. *Nature Biotechnology*, 21(11):1347–1355, 2003.
- [30] S. W. Hell. Microscopy and its focal switch. *Nature Methods*, 6(1):24–32, 2008.
- [31] S. T. Hess, T.P.K. Girirajan, and M. D. Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical Journal*, 91(11):4258–4272, 2006.
- [32] 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.
- [33] F. Huang, T.M.P. Hartwich, F. E. Rivera-Molina, Y. Lin, W. C. Duim, J. J. Long, P. D. Uchil, J. R. Myers, M. A. Baird, W. Mothes, et al. Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms. *Nature Methods*, 10(7):653–658, 2013.
- [34] L. Isserlis. On certain probable errors and correlation coefficients of multiple frequency distributions with skew regression. *Biometrika*, 11:185C190, 1916.
- [35] A. Janicki and A. Weron. *Simulation and Chaotic Behavior of α -stable Stochastic Processes*. CRC Press, 1994.
- [36] J.-H. Jeon, E. Barkai, and R. Metzler. Noisy continuous time random walks. *Journal of Chemical Physics*, 139(12):121916, 2013.
- [37] J.-H. Jeon and R. Metzler. Analysis of short subdiffusive time series: scatter of the time-averaged mean-squared displacement. *Journal of Physics A: Mathematical and Theoretical*, 43(25):252001, 2010.
- [38] S. A. Jones, S.-H. Shim, J. He, and X. Zhuang. Fast, three-dimensional super-resolution imaging of live cells. *Nature Methods*, 8(6):499–505, 2011.
- [39] E. Kepten, I. Bronshtein, and Y. Garini. Improved estimation of anomalous diffusion exponents in single-particle tracking experiments. *Physical Review E*, 87(5):052713, 2013.
- [40] A.I. Khinchin. *Mathematical Foundations of Statistical Mechanics*. Dover Books on Mathematics. Dover Publications, 1949.

- [41] S. C. Kou. Stochastic modeling in nanoscale biophysics: subdiffusion within proteins. *Annals of Applied Statistics*, 2(2):501–535, 2008.
- [42] S. C. Kou and X. S. Xie. Generalized Langevin equation with fractional Gaussian noise: subdiffusion within a single protein molecule. *Physical Review Letters*, 93(18):180603, 2004.
- [43] K. Kremer and G.S. Grest. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *The Journal of Chemical Physics*, 92:5057, 1990.
- [44] S.K. Lai, Y.Y. Wang, R. Cone, D. Wirtz, and J. Hanes. Altering mucus rheology to solidify human mucus at the nanoscale. *PLoS One*, 4(1):e4294, 2009.
- [45] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, pages 1302–1338, 2000.
- [46] O. Lieleg, I. Vladescu, and K. Ribbeck. Characterization of particle translocation through mucin hydrogels. *Biophysical Journal*, 98(9):1782, 2010.
- [47] Eugene Lukacs. *Characteristic Functions*. Charles Griffin & Company Limited, London, 2nd edition, 1970.
- [48] M. Lysy, N. Pillai, D. B. Hill, M. G. Forest, J. Mellnik, P. Vasquez, and S. A. McKinley. Model comparison for single particle tracking in biological fluids. *arXiv:1407.5962*, pages 1–38, 2014.
- [49] M. Lysy, N. Pillai, D. B. Hill, M. G. Forest, J. Mellnik, P. Vasquez, and S. A. McKinley. Model comparison for single particle tracking in biological fluids. *To appear in Journal of the American Statistical Association*, pages 1–44, 2016.
- [50] M. Magdziarz and A. Weron. Anomalous diffusions: Testing ergodicity breaking in experimental data. *Physical Review E*, 84:051138, 2011.
- [51] M. Magdziarz and A. Weron. Ergodic properties of anomalous diffusion processes. *Annals of Physics*, 326(9):2431C2443, September 2011.
- [52] G. Margolin and E. Barkai. Nonergodicity of blinking nanocrystals and other Lévy-walk processes. *Physical Review Letters*, 94(8):080601, 2005.
- [53] T.G. Mason and D.A. Weitz. Optical measurements of the linear viscoelastic moduli of complex fluids. *Physical Review Letters*, 74:1250–1253, 1995.
- [54] H. Matsui, V.E. Wagner, D.B. Hill, U.E. Schwab, T.D. Rogers, B. Button, R.M. Taylor, R. Superfine, M. Rubinstein, B.H. Iglewski, and R.C. Boucher. A physical linkage between cystic fibrosis airway surface dehydration and *Pseudomonas aeruginosa* biofilms. *Proceedings of the National Academy of Sciences*, 103(48):18131, 2006.

- [55] M. Meerschaert and H.-P. Scheffler. Limit theorems for continuous-time random walks with infinite mean waiting times. *Journal of Applied Probability*, 41:623–638, 2004.
- [56] J. W. R. Mellnik, M. Lysy, P. A. Vasquez, N. S. Pillai, D. B. Hill, J. Cribb, S. A. McKinley, and M. G. Forest. Maximum likelihood estimation for single particle, passive microrheology data with drift. *Journal of Rheology*, 60(3):379–392, 2016.
- [57] Y. Meroz and I. M. Sokolov. A toolbox for determining subdiffusive mechanisms. *Physics Reports*, 573:1–29, 2015.
- [58] R. Metzler and J. Klafter. The random walk’s guide to anomalous diffusion: a fractional dynamics approach. *Physics Reports*, 339(1):1–77, 2000.
- [59] R. Metzler, V. Tejedor, J.H. Jeon, Y. He, W.H. Deng, S. Burov, and E. Barkai. Analysis of single particle trajectories: from normal to anomalous diffusion. *Acta Physica Polonica B*, 40(5):1315–1331, 2009.
- [60] X. Michalet and A. J. Berglund. Optimal diffusion coefficient estimation in single-particle tracking. *Physical Review E*, 85(6):061916, 2012.
- [61] N. Monnier, S.-M. Guo, M. Mori, J. He, P. Lénárt, and M. Bathe. Bayesian approach to MSD-based analysis of particle motion in live cells. *Biophysical Journal*, 103(3):616–626, 2012.
- [62] E. Moulines, F. Roueff, and M. S. Taqqu. Central limit theorem for the log-regression wavelet estimation of the memory parameter in the Gaussian semi-parametric context. *Fractals*, 15(04):301–313, 2007.
- [63] E. Moulines, F. Roueff, and M.S. Taqqu. On the spectral density of the wavelet coefficients of long-memory time series with application to the log-regression estimation of the memory parameter. *Journal of Time Series Analysis*, 28(2):155–187, 2007.
- [64] E. Moulines, F. Roueff, and M.S. Taqqu. A wavelet Whittle estimator of the memory parameter of a nonstationary Gaussian time series. *Annals of Statistics*, pages 1925–1956, 2008.
- [65] A. Nandi, D. Heinrich, and B. Lindner. Distributions of diffusion measures from a local mean-square displacement analysis. *Physical Review E*, 86(2):021926, 2012.
- [66] I. Nourdin, D. Nualart, and C. Tudor. Central and non-central limit theorems for weighted power variations of fractional Brownian motion. *Annales de l’Institut Henri Poincaré (B) Probabilités et Statistiques*, 46(4):1055–1079, 2010.
- [67] D. O’Malley and J. Cushman. A renormalization group classification of nonstationary and/or infinite second moment diffusive processes. *Journal of Statistical Physics*, 146(5):989–1000, 2012.

- [68] R. F. Peltier and J. L. Véhel. A new method for estimating the parameter of fractional Brownian motion. *Rapport de recherche - Institut national de recherche en informatique et en automatique*, pages 1–27, 1994.
- [69] V. Pipiras and M.S. Taqqu. *Long-Range Dependence and Self-Similarity*. 2017.
- [70] B.L.S. Prakasa Rao. *Statistical Inference for Fractional Diffusion Processes*. Wiley Series in Probability and Statistics, 2010.
- [71] H. Qian, M. Sheetz, and E. Elson. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophysical Journal*, 60(4):910–921, 1991.
- [72] K. P. Reighard, D. B. Hill, G. A. Dixon, B. V. Worley, and M. H. Schoenfisch. Disruption and eradication of *P. aeruginosa* biofilms using nitric oxide-releasing chitosan oligosaccharides. *Biofouling*, 31(9-10):775–787, 2015.
- [73] A. Réveillac, M. Stauch, and C. Tudor. Hermite variations of the fractional Brownian sheet. *Stochastics and Dynamics*, 12(3), 2012.
- [74] M. Rosenblatt. Independence and dependence. In *Proceedings of the 4th Berkeley symposium on mathematical statistics and probability*, volume 2, pages 431–443, 1961.
- [75] M. Rubinstein and R.H. Colby. *Polymer Physics*, volume 105. Oxford University Press New York, 2003.
- [76] M. J. Rust, M. Bates, and X. Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 3(10):793–796, 2006.
- [77] T. Sandev, R. Metzler, and Ž. Tomovski. Velocity and displacement correlation functions for fractional generalized Langevin equations. *Fractional Calculus and Applied Analysis*, 15(3):426–450, 2012.
- [78] I. M. Sokolov. Statistics and the single molecule. *Physics*, 1:8, 2008.
- [79] J. Suh, M. Dawson, and J. Hanes. Real-time multiple-particle tracking: applications to drug and gene delivery. *Advanced Drug Delivery Reviews*, 57:63–78, 2005.
- [80] M. S. Taqqu. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Probability Theory and Related Fields*, 31(4):287–302, 1975.
- [81] M. S. Taqqu. Fractional Brownian motion and long range dependence. In *Theory and Applications of Long-Range Dependence (P. Doukhan, G. Oppenheim and M. S. Taqqu, eds.)*, pages 5–38. Birkhäuser, Boston, 2003.
- [82] M. S. Taqqu. The Rosenblatt process. In *The selected works of Murray Rosenblatt (Davis, R. A. and Lii, K.-S. and Politis, D. N., eds.)*, pages 29–45. Springer, 2011.

- [83] M. Valentine, P. Kaplan, D. Thota, J. Crocker, T. Gisler, R. Prudhomme, M. Beck, and D. A. Weitz. Investigating the microenvironments of inhomogeneous soft materials with multiple particle tracking. *Physical Review E*, 64(6):061506, 2001.
- [84] M. Veillette and M. S. Taqqu. Properties and numerical evaluation of the rosenblatt distribution. *Bernoulli*, 19(3):982–1005, 2013.
- [85] D. Veitch and P. Abry. A wavelet-based joint estimator of the parameters of long-range dependence. *IEEE Transactions on Information Theory*, 45(3):878–897, 1999.
- [86] C. L. Vestergaard, P. C. Blainey, and H. Flyvbjerg. Optimal estimation of diffusion coefficients from single-particle trajectories. *Physical Review E*, 89(2):022726, 2014.
- [87] H. Wendt, G. Didier, S. Combrexelle, and P. Abry. Multivariate Hadamard self-similarity: testing fractal connectivity. *Under review*, 2016.
- [88] V. Westphal, S. O. Rizzoli, M. A. Lauterbach, D. Kamin, R. Jahn, and S. W. Hell. Video-rate far-field optical nanoscopy dissects synaptic vesicle movement. *Science*, 320(5873):246–249, 2008.
- [89] Denis Wirtz. Particle-tracking microrheology of living cells: principles and applications. *Annual Review of Biophysics*, 38:301–326, 2009.

Biography

Kui Zhang got both his Bachelor and Master degree of Science in University of Science and Technology of China under the supervision of Professor Yu Ye. He then became a graduate student of Tulane in August 2012, under the supervision of Professor Gustavo Didier, eventually completing the program in May 2017.