

Estimation and Numerical Simulation of the Linear Fractional Stable Motion and Related Processes

PhD Thesis

Dmitry Otryakhin

Supervisor: Mark Podolskij

Department of Mathematics
Aarhus University
August, 2019

Preface

The linear fractional stable motion (shortly, lfsm) is in the focus of the present thesis. This is a special type of a stochastic integral driven by α - stable Lévy process. It possesses the following properties: self-similarity, stationarity of increments, heavy tails and long-range dependence, some of which helped the process find its way in modeling of computer networks and solar activities [22, 46]. It is also a generalization of fractional Brownian motion as it has the same integral core. Apart from lfsms, the thesis considers another type of stochastic integrals- Lévy driven moving averages.

The thesis comprises four chapters. Chapter 1 presents some introduction to the field of Lévy processes as well as integrals based on them. It briefly summarizes well-understood notions and theorems from the literature to help to get acquainted with the field. Chapter 2 is a joint paper with professors Stepan Mazur and Mark Podolskij which is to appear in Bernoulli journal. It presents feasible estimation techniques for all parameters of lfsms in both low and high frequency settings. In the end, the paper contains a simulation study that shows finite sample performances of the estimators along with some comparisons of them. Chapter 3 is a paper written together with Stepan Mazur and is submitted. It is devoted to practical aspects of Monte-Carlo simulations and statistical procedures on lfsms. It describes R package **rlfsm** designed for those tasks, along with results of simulations and artifacts found during numerical experiments with the type of motions. A special attention is paid to our implementation of the FFT-based simulation method developed by Stoev and Taqqu [40]. Finally, Chapter 4 describes a hybrid method for simulation of multidimensional Lévy-driven moving averages, where the core of the integral shows explosive behavior near zero. In this method, simulation of the tail is powered by multidimensional convolution theorem and is a generalization of the idea presented in [40]. For the area near zero a modification of a method developed by Cohen et al. [17] is used, and it is based on shot-noise series approximations of the integral. The paper is an early stage project with Mark Podolskij.

My main contribution comprises designing and performing numerical experiments in the first two papers, architecturing and developing a large part of the software in **rlfsm**, redesigning the algorithm for simulation of lfsms, and creating most of the third paper except some large-scale ideas and strategies.

Resume

I denne afhandling er der fokus på den lineære fraktionale stabile bevægelse (kort, lfsb). Den er en speciel type stokastisk integral drevet af en alfastabil Lévy-proces. Den har følgende egenskaber: selv-lighed, stationære tilvækst, tunge haler og langtidshukommelse, nogle af disse egenskaber hjalp processen med at finde sin vej til brug ved modellering af computernetværk og solaktiviteter [22, 46]. Det er også en generalisering af fraktionale Brownsk bevægelse, da den har den samme integrerede kerne. Bortset fra lfsbs betragter tesen en anden type stokastiske integraler - Lévy-drevne glidende gennemsnit.

Tesen består af fire kapitler. Kapitel 1 præsenterer nogle introduktioner til om Lévy-processer og integraler baseret på dem. Den opsummerer vel kendte forestillinger og teoremer fra litteraturen for at hjælpe med at blive bekendt med området. Kapitel 2 er en fælles artikel fra professorer Stepan Mazur og Mark Podolskij, som kommer til at blive udgivet i Bernoulli-tidsskriftet. Den præsenterer mulige estimeringsteknikker for alle parametre af lfsbs ved både lav- og højfrekvensindstillinger. Til sidst viser artiklen en simuleringundersøgelse, der viser endelige prøveudførelser af estimererne sammen med nogle sammenligninger af dem. Kapitel 3 er en artikel skrevet sammen med Stepan Mazur, denne er indsendt til udgivelse. Den er dedikeret til praktiske aspekter af Monte-Carlo-simuleringer og statistiske procedurer på lfsbs. Den beskriver R-pakken **rlfsm** designet til disse opgaver sammen med resultater af simuleringer og artefakter fundet under numeriske eksperimenter med disse typer af bevægelser. Der lægges særlig vægt på vores implementering af den FFT-baserede simuleringsmetode udviklet af Stoev and Taqqu [40]. Endelig beskriver Kapitel 4 en hybrid metode til simulering af multidimensionel Lévy-drevet glidende gennemsnit, hvor kernen i integralet viser eksplosiv opførsel i nærheden af nul. I denne metode er simulering af halen drevet af multidimensionel konvolutions-sætning og er en generalisering af ideen præsenteret i [40]. For området nær nul bruges en modifikation af en metode udviklet af Cohen et al. [17], og den er baseret på korte støj-serier approksimationer af integralet. Projektet er i et tidligt stadium i samarbejde med Mark Podolskij.

Mit vigtigste bidrag omfatter design og udførelse af numeriske eksperimenter i de første to artikler, arkitektur og udvikling af en stor del af softwaren i **rlfsm**, designudvikling af algoritmen til simulering af lfsbs og udvikling af det meste af i det tredje artikel, undtagen nogle overordnede ideer og strategier.

Notation

Here the reader may find some specific and frequent notations and notions used in the thesis.

- $a \cdot b$ - scalar product of two vectors
- Minimum and maximum
 - $a \vee b = \max(a, b)$, maximum value of two numbers
 - $a \wedge b = \min(a, b)$ minimum value of two numbers
- $a * b$ - convolution of two functions or sequences
- $a * \cdots * b$ - d-dimensional convolution
- Products
 - $a \times b$ - regular multiplication
 - $A \times B$ cartesian product of sets
 - $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$
- a.s. - almost surely
- a.e. - almost every
- $a := b$ - a equals to b by definition
- ϕ^{-1} - inversion of function ϕ
- Γ - Gamma function
- $a_+ := \max(0, a)$
- $\mathbb{E}[x]$ - expected value of x
- $\mathbb{P}(A)$ - probability that set A is observed
- $\stackrel{d}{=}$ - equality in distribution
- $\xrightarrow{a.s.}$ - almost sure convergence
- \xrightarrow{P} - convergence in probability
- \xrightarrow{d} - convergence in distribution
- $\mathcal{N}(\mu, \sigma)$ - normally distributed random variable with mean vector μ and covariance matrix σ .
- sas - symmetric α -stable
- for a function $g : \mathbb{R} \rightarrow \mathbb{R}$, $\|g\|_\alpha^\alpha := \int_{\mathbb{R}} |g_s|^\alpha ds$

- $\|\xi\|_\alpha$ - the scale coefficient of a S α S random variable ξ . It appears in the characteristic function of ξ , $\mathbb{E}[e^{i\xi t}] = e^{-\|\xi\|_\alpha |t|^\alpha}$.
- \square - end of a proof, Q.E.D.
- $1_{x \in A}$ - indicator function, 1 if $x \in A$, 0 otherwise.
- \mathfrak{F} , FT - Fourier transform
- FFT - fast Fourier transform
- DFT - discrete Fourier transform
- DTFT - discrete time Fourier transform

Contents

Contents	5
1 Introduction	7
1.1 Introduction to mathematical tools	7
2 Estimation of the linear fractional stable motion	11
2.1 Overview and preliminaries	11
2.2 Introduction	12
2.3 First properties and some asymptotic results	14
2.4 Statistical inference in the continuous case $H - 1/\alpha > 0$	20
2.5 Statistical inference in the general case	23
2.6 A simulation study	29
2.7 Proofs	33
2.8 Appendix	44
3 Numerical techniques and results for the Linear Fractional Stable Motion	47
3.1 Overview of Chapter 3	47
3.2 Introduction	48
3.3 Basic R functions	48
3.4 Parameter Estimation of the linear fractional stable motion	60
3.5 S4 classes for Levy-driven motions	68
3.6 Appendix	72
4 Simulation of multidimensional stochastic integrals with convolving kernels driven by Lévy basis	75
4.1 The model	75
4.2 Shot-noise approximation of non-Gaussian fractional fields	76
4.3 Review of the multidimensional discrete Fourier transform	77
4.4 Algorithms based on multidimensional fast Fourier transforms.	77
4.5 stable drivers	79
Bibliography	83

Chapter 1

Introduction

1.1 Introduction to mathematical tools

Lévy processes

The core of this work relies heavily on Lévy processes as they will be used as drivers for all of the stochastic integrals under consideration. Rigorous details of these processes can be found in [39]. A more intuitive introduction to Lévy processes and their applications is given in [44]. The following definitions, propositions and theorems are adopted from these two books. We start with the definition of Lévy processes.

Definition 1.1.1. *A càdlàg stochastic process $(X_t)_{t \geq 0}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}}, \mathbb{P})$ with values in \mathbb{R}^d such that $X_0 = 0$ is called a Lévy process if it possesses the following properties:*

- *Independent increments: for every increasing sequence of times (t_0, \dots, t_n) , the random variables $X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}$ are independent.*
- *Stationary increments: the law of $X_{t+h} - X_t$ does not depend on t .*
- *Stochastic continuity:*

$$\lim_{h \rightarrow 0} \mathbb{P}(|X_{t+h} - X_t| \geq \epsilon) = 0 \quad \text{for all } \epsilon > 0.$$

The following theorem, the Lévy-Khinchin representation shows that any Lévy process is fully characterized by only three parameters which are called a characteristic triplet.

Theorem 1.1.2 (Lévy-Khinchin representation). *Let $(X_t)_{t \geq 0}$ be a Lévy process on \mathbb{R}^d . Then, there exists a characteristic triplet (A, ν, γ) , where A is a symmetric nonnegative-definite $d \times d$ matrix, ν is a measure on \mathbb{R}^d satisfying*

$$\int_{\mathbb{R}^d} (1 \wedge |x|^2) \nu(dx) < \infty; \quad \nu(\{0\}) = 0 \tag{1.1}$$

and $\gamma \in \mathbb{R}^d$. Then

$$\mathbb{E}[e^{iz \cdot X_t}] = e^{t\psi(z)}, \quad z \in \mathbb{R}^d \tag{1.2}$$

with

$$\psi(z) = -\frac{1}{2}z \cdot Az + i\gamma \cdot z + \int_{\mathbb{R}^d} (e^{iz \cdot x} - 1 - iz \cdot x 1_{|x| \leq 1}) \nu(dx). \tag{1.3}$$

A very important and intuitive result for understanding of Lévy processes is the Lévy-Itô decomposition. It interprets the motion as a sum of four independent and rather simple processes: deterministic drift, Brownian motion, a compound Poisson process and an infinite superposition of independent compensated Poisson processes. This fact is often used in simulation algorithms, and in fact, will echo in the last chapter, where integrals driven by Lévy bases will be represented as integrals on Brownian fields, and series with Poissonian nature.

Proposition 1.1.3 (Lévy-Itô decomposition). *Let $(X_t)_{t \geq 0}$ be a Lévy process on \mathbb{R}^d with a characteristic triplet (A, ν, γ) . Then:*

- *The jump measure of X , denoted by J_X , is a Poisson random measure on $[0, \infty[\times \mathbb{R}^d$ with intensity measure $\nu(dx)dt$*
- *There exist a d -dimensional Brownian motion $(B_t)_{t \geq 0}$ with covariance matrix A such that*

$$\begin{aligned} X_t &= \gamma t + B_t + X_t^l + \lim_{\epsilon \rightarrow 0} \tilde{X}_t^\epsilon, \text{ where} \\ X_t^l &= \int_{|x| \geq 1, s \in [0, t]} x J_X(ds \times dx) \text{ and} \\ \tilde{X}_t^\epsilon &= \int_{\epsilon \leq |x| < 1, s \in [0, t]} x (J_X(ds \times dx) - \nu(dx)ds) \end{aligned} \quad (1.4)$$

The terms in (1.4) are independent and the convergence in the last term is almost sure and uniform in t on $[0, T]$.

In this work a special attention is paid to a subclass of the Lévy processes- stable processes.

Definition 1.1.4. *A Lévy process X_t is said to be selfsimilar if*

$$\forall a > 0, \exists b(a) > 0 : \left(\frac{X_{at}}{b(a)} \right)_{t \geq 0} \stackrel{d}{=} (X_t)_{t \geq 0} \quad (1.5)$$

Definition 1.1.5. *A random variable $X \in \mathbb{R}^d$ is said to have stable distribution if for every $a > 0$ there exist $b(a) > 0$ and $c(a) \in \mathbb{R}^d$ such that*

$$\Phi_X(z)^a = \Phi_X(zb(a))e^{ic \cdot z} \quad \text{for all } z \in \mathbb{R}^d. \quad (1.6)$$

It is said to have a strictly stable distribution if

$$\Phi_X(z)^a = \Phi_X(zb(a)) \quad \text{for all } z \in \mathbb{R}^d. \quad (1.7)$$

The name stable comes from the following stability under addition property: if X has stable distribution and $X^{(1)}, \dots, X^{(n)}$ are independent copies of X then there exist a positive number c_n and a vector d such that

$$X^{(1)} + \dots + X^{(n)} \stackrel{d}{=} c_n X + d \quad (1.8)$$

It can be shown (see [37], corollary 2.1.3) that for every stable distribution there exists a constant $\alpha \in (0, 2]$ such that in Equation (1.6), $b(a) = a^{1/\alpha}$. This constant is called the

index of stability and stable distributions with index α are also referred to as α -stable distributions. A selfsimilar Lévy process has strictly stable distribution at all times. For this reason, such processes are also called strictly stable Lévy processes. A strictly α -stable Lévy process satisfies:

$$\forall a > 0, \left(\frac{X_{at}}{a^{1/\alpha}} \right)_{t \geq 0} \stackrel{d}{=} (X_t)_{t \geq 0}. \quad (1.9)$$

In the multidimensional setting there is an analog to a Lévy process- a Lévy basis. Let \mathcal{S} be a collection of sets, which is nonempty. Then \mathcal{S} is called a σ -ring if:

$$\begin{aligned} \forall n, A_n \in \mathcal{S} &\Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{S} \\ A \in \mathcal{S} \text{ and } B \in \mathcal{S} &\Rightarrow A \setminus B \in \mathcal{S} \end{aligned} \quad (1.10)$$

A Lévy basis $L = \{L(B) : B \in \mathcal{S}\}$, where \mathcal{S} is a δ -ring of an arbitrary non-empty set S such that there exists an increasing sequence of sets $(S_n) \subset \mathcal{S}$ with $\cup_{n \in \mathbb{N}} S_n = S$, is an independently scattered random measure with Lévy-Khinchin representation

$$\begin{aligned} \log \mathbb{E}[\exp(iuL(B))] &= iuv_1(B) - \frac{1}{2}u^2v_2(B) + \\ &+ \int_{\mathbb{R}} \left(\exp(iuy) - 1 - iuy1_{[-1,1]}(y) \right) v_3(dy, B). \end{aligned} \quad (1.11)$$

Here v_1 is a signed measure on \mathcal{S} , v_2 is a measure on \mathcal{S} and $v_3(\cdot, \cdot)$ is a generalised Lévy measure on $\mathbb{R} \times \mathcal{S}$ (see e.g. [34] for details).

Integrals with Lévy drivers

A classical theory of such integrals was introduced in [34] and summarized in [33]. The following is adopted from [33]. We study an integral of the form

$$\int_A f dL,$$

where L is a Lévy basis on a δ -ring \mathcal{S} , $f : (S, \sigma(\mathcal{S})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ a measurable real valued function and $A \in \sigma(\mathcal{S})$, which is the main object of a seminal paper [34]. We will briefly recall the most important results of this work. By definition of a Lévy basis the law of $(L(A_1), \dots, L(A_d))$, $A_i \in \mathcal{S}$, is infinitely divisible, the random variables $L(A_1), \dots, L(A_d)$ are independent when the sets A_1, \dots, A_d are disjoint, and

$$L(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} L(A_i) \quad \mathbb{P} - \text{almost surely}$$

for disjoint A_i 's with $\cup_{i=1}^{\infty} A_i \in \mathcal{S}$. Recalling the characteristic triplet (v_1, v_2, v_3) from the Lévy-Khinchin representation (1.11), the *control measure* λ of L is defined via

$$\lambda(A) := |v_1|(A) + v_2(A) + \int_{\mathbb{R}} \min(1, x^2) v_3(dx, A), \quad A \in \mathcal{S},$$

where $|v_1|$ denotes the total variation measure associated with v_1 . In this subsection we will use the truncation function

$$\tau(z) := \begin{cases} z & : \quad \|z\| \leq 1 \\ z/\|z\| & : \quad \|z\| > 1 \end{cases}$$

Now, for any *simple function*

$$f(x) = \sum_{i=1}^d a_i 1_{A_i}(x), \quad a_i \in \mathbb{R}, A_i \in \mathcal{S},$$

the stochastic integral is defined as

$$\int_A f dL := \sum_{i=1}^d a_i L(A \cap A_i), \quad A \in \sigma(\mathcal{S}).$$

The extension of this definition is as follows.

Definition 1.1.6. *A measurable function $f : (\mathcal{S}, \sigma(\mathcal{S})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is called L -integrable if there exists a sequence of simple functions $(f_n)_{n \geq 1}$ such that*

(i) $f_n \rightarrow f$ λ -almost surely.

(ii) For any $A \in \sigma(\mathcal{S})$ the sequence $(\int_A f_n dL)$ converges in probability.

In this case the stochastic integral is defined by

$$\int_A f dL := \mathbb{P} - \lim_{n \rightarrow \infty} \int_A f_n dL.$$

Although this definition is quite intuitive, it does not specify the class of L -integrable functions explicitly. The next theorem, which is one of the main results of [34], gives an explicit condition on the L -integrability of a function f .

Theorem 1.1.7. ([34, Theorem 2.7]) *Let $f : (\mathcal{S}, \sigma(\mathcal{S})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a measurable function. Then f is L -integrable if and only if the following conditions hold:*

$$\int_S U(f(s), s) \lambda(ds) < \infty, \quad \int_S f^2(s) v_2(s) \lambda(ds) < \infty, \quad \int_S V_0(f(s), s) \lambda(ds) < \infty,$$

where

$$U(u, s) := uv_1(s) + \int_{\mathbb{R}} (\tau(xu) - u\tau(x)) v_3(dx, s),$$

$$V_0(u, s) := \int_{\mathbb{R}} \min(1, |xu|^2) v_3(dx, s).$$

Furthermore, the real valued random variable $X = \int_S f dL$ is infinitely divisible with Lévy-Khinchin representation

$$\log \mathbb{E}[\exp(iuX)] = iuv_1(f) - \frac{1}{2}u^2v_2(f) + \int_{\mathbb{R}} (\exp(iuy) - 1 - iuy1_{[-1,1]}(y)) v_3^f(dy),$$

where

$$v_1(f) = \int_S U(f(s), s) \lambda(ds),$$

$$v_2(f) = \int_S f^2(s) v_2(s) \lambda(ds),$$

$$v_3^f(B) = v_3 \{(x, s) \in \mathbb{R} \times S : xf(s) \in B \setminus \{0\}\}, \quad B \in \mathcal{B}(\mathbb{R}).$$

In case when the integral is driven by an α -stable Lévy basis, the conditions presented in Theorem 1.1.7 are significantly simplified:

$$\int_{\mathbb{R}^d} |f(s)|^\alpha ds < \infty$$

Chapter 2

Estimation of the linear fractional stable motion

STEPAN MAZUR, DMITRY OTRYAKHIN, MARK PODOLSKIJ

2.1 Overview and preliminaries

In this paper we investigate the parametric inference for the linear fractional stable motion in high and low frequency setting. The symmetric linear fractional stable motion is a three-parameter family, which constitutes a natural non-Gaussian analogue of the scaled fractional Brownian motion. It is fully characterised by the scaling parameter $\sigma > 0$, the self-similarity parameter $H \in (0, 1)$ and the stability index $\alpha \in (0, 2)$ of the driving stable motion. The parametric estimation of the model is inspired by the limit theory for stationary increments Lévy moving average processes that has been recently studied in [7]. More specifically, we combine (negative) power variation statistics and empirical characteristic functions to obtain consistent estimates of (σ, α, H) . We present the law of large numbers and some fully feasible weak limit theorems.

Now we overview some notions and a theorem obtained by Pipiras et al. [31], which played one of the key roles in our estimation technique. The primary objects in [31] are α -stable moving average sequences $\{X_{j,n}\}_{n \geq 1}$, $j = 1, \dots, J$ of the form

$$X_{j,n} = \int_{\mathbb{R}} a_j(n - c_j x) M(dx), \quad (1.1)$$

where $\alpha \in (0, 2)$, $c_j > 0$, $a_j \in L^\alpha(\mathbb{R}, dx)$ and, in case when $\alpha = 1$, $a_j \log |a_j| \in L^1(\mathbb{R}, dx)$, and M is an α -stable random measure on \mathbb{R} with the Lebesgue control measure $m(dx) = dx$ and the skewness $\beta \in [-1, 1]$. The moving average $\{X_{j,n}\}_{n \geq 1}$ is called one-sided if $a_j(x) = 0$ for either $x < x_0$ or $x > x_0$, $x_0 \in \mathbb{R}$. Otherwise, it is called two-sided. Fix n_j , $j = 1, \dots, J$, and let $N_j > 0$ be integers satisfying

$$N_j \sim \frac{N}{n_j}, \quad N \rightarrow \infty. \quad (1.2)$$

For $j = 1, \dots, J$, and measurable functions K_j set

$$S_{j,N_j} := \sum_{n=1}^{N_j} (K_j(X_{j,n}) - \mathbb{E}K_j(X_{j,n})) \quad (1.3)$$

An important result of [31] which we used in this paper is Theorem 2.1 that states the following:

Theorem 2.1.1. *Let $\alpha \in (0, 2)$ and $\{X_{j,n}\}_{n \geq 1}$, $j = 1, \dots, J$, be α -stable moving averages defined by (1.1). Suppose that, for each $j = 1, \dots, J$, the kernel a_j in (1.1) satisfies the condition*

$$\sum_{m=-\infty}^{\infty} \left(\int_{m-1}^m |a_j(x)|^\alpha dx \right)^{1/2} < \infty. \quad (1.4)$$

Suppose also that, for each $j = 1, \dots, J$, the function K_j in (1.3) is bounded if $\{X_{j,n}\}_{n \geq 1}$ is one-sided, and is bounded and twice differentiable with bounded derivatives if $\{X_{j,n}\}_{n \geq 1}$ is two-sided. Then, for $j = 1, \dots, J$,

$$\left(N_j^{-1/2} S_{j, N_j} \right)_{j=1}^J \xrightarrow{d} \mathcal{N}(0, \Sigma^M)$$

where $\Sigma^M = (\Sigma_{jk}^M)_{j,k=1, \dots, J}$.

$$\Sigma_{jk}^M = \lim_{N \rightarrow \infty} \mathbb{E} \frac{S_{j, N_j} S_{k, N_k}}{N_j^{1/2} N_k^{1/2}} < \infty \quad (1.5)$$

2.2 Introduction

Since the pioneering work by Mandelbrot and van Ness [28] fractional Brownian motion (fBm) became one of the most prominent Gaussian processes in the probabilistic and statistical literature. As a building block in stochastic models it found various applications in natural and social sciences such as physics, biology or economics. Mathematically speaking, the scaled fBm is fully characterised by its scaling parameter $\sigma > 0$ and Hurst parameter $H \in (0, 1)$. More specifically, the scaled fBm $Z_t = \sigma B_t^H$ is a zero mean Gaussian process with covariance kernel determined by

$$\mathbb{E} \left[B_t^H B_s^H \right] = \frac{1}{2} \left(t^{2H} + s^{2H} - |t - s|^{2H} \right), \quad t, s \geq 0.$$

We recall that the (scaled) fBm with Hurst parameter $H \in (0, 1)$ is the unique Gaussian process with stationary increments and self-similarity index H , i.e. it holds that $(a^H Z_t)_{t \geq 0} = (Z_{at})_{t \geq 0}$ in distribution for any $a > 0$. Over the last forty years there has been a lot of progress in limit theorems and statistical inference for fBm's. The estimation of the Hurst parameter H and/or the scaling parameter σ has been investigated in numerous papers both in low and high frequency framework. We refer to [19] for efficient estimation of the Hurst parameter H in the low frequency setting and to [12, 16, 25] for the estimation of (σ, H) in the high frequency setting, among many others. In the low frequency framework the spectral density methods are usually applied and the optimal convergence rate for the estimation of (σ, H) is known to be \sqrt{n} . In the high frequency setting the estimation of the pair (σ, H) typically relies upon power variations and related statistics, and the optimal convergence rate is known to be $(\sqrt{n}/\log(n), \sqrt{n})$. More recently, the class of multifractional Brownian motions, which accounts for time varying Hurst parameter, has been introduced in the literature (see e.g. [5, 27, 42]). We refer to

the work [6, 26] for estimation techniques for the regularity of a multifractional Brownian motion.

If we drop the Gaussianity assumption the class of stationary increments self-similar processes becomes much larger. This is a consequence of the work by Pipiras and Taqqu [30], which in turn applies the decomposition results from the seminal paper by Rosiński [36] (see also [38]). The crucial theorem proved in [36] shows that each stationary stable process can be uniquely decomposed (in distribution) into three independent parts: the mixed moving average process, the harmonizable process and the “third kind” process described by a conservative nonsingular flow. The most prominent example of a non-Gaussian stationary increments self-similar process is the linear fractional stable motion (an element of the first class), which has been introduced in [13]. It is defined as follows: On a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}}, \mathbb{P})$, we introduce the process

$$X_t = \int_{\mathbb{R}} \left\{ (t-s)_+^{H-1/\alpha} - (-s)_+^{H-1/\alpha} \right\} dL_s, \quad x_+ := \max\{x, 0\}, \quad (2.2.1)$$

where L is a symmetric α -stable Lévy motion, $\alpha \in (0, 2)$, with scale parameter $\sigma > 0$ and $H \in (0, 1)$ (here we use the convention $x_+^a = 0$ for any $x \leq 0$ and $a \in \mathbb{R}$). In some sense the linear fractional stable motion is a non-Gaussian analogue of fBm. The process $(X_t)_{t \in \mathbb{R}}$ has symmetric α -stable marginals, stationary increments and it is self-similar with parameter H . Fractional stable motions are often used in natural sciences, e.g. in physics or internet traffic, where the process under consideration exhibits stationarity and self-similarity along with heavy tailed marginals (see e.g. [22] for the context of turbulence modelling). The probabilistic properties of linear fractional stable motions, such as integration concepts, path and variational properties, have been intensively studied in several papers, see for example [8, 10, 9] among many others. However, from the statistical point of view, very little is known about the inference for the parameter $\theta = (\sigma, \alpha, H) \in \mathbb{R}_+ \times (0, 2) \times (0, 1)$ in high or low frequency setting. The few existing papers mostly concentrate on estimation of the self-similarity parameter H . The work [4, 32] investigates the asymptotic theory for a wavelet-based estimator of H when $\alpha \in (1, 2)$. In [7, 41] the authors suggest to use power variation statistics to obtain an estimator of H , but this method also requires the a priori knowledge of the lower bound for the stability parameter α . Recently, the work [21] suggested to use negative power variations to get a consistent estimator of H , which applies for any $\alpha \in (0, 2)$, but this article does not contain a central limit theorem for this method. Finally, in [7, 22] the authors propose to use an empirical scale function to estimate the pair (α, H) . However, this approach only provides a $\log(n)$ -consistent estimator without any hope for a central limit theorem.

In this paper we will propose a new estimation procedure for the parameter $\theta = (\sigma, \alpha, H)$ in high and low frequency framework. Our methodology is based upon the use of power variation statistics, with possibly negative powers, and the empirical characteristic function. The probabilistic techniques originate from the recent article [7], which has developed the asymptotic theory for power variations of higher order differences of stationary increments Lévy moving averages (see also [31, 32] for related asymptotic theory). However, we will need to derive much more complex asymptotic results to obtain a complete distributional theory for the estimator of the parameter $\theta \in \mathbb{R}_+ \times (0, 2) \times (0, 1)$. We will obtain a fully feasible asymptotic theory for our estimator with convergence rates $(\sqrt{n}, \sqrt{n}, \sqrt{n})$ in the low frequency setting and $(\sqrt{n}/\log(n), \sqrt{n}/\log(n), \sqrt{n})$ in the high frequency setting.

The paper is structured as follows. Section 2.3 presents the basic properties of the linear fractional stable motion, the review of the probabilistic results from [7] and a multivariate limit theorem, which plays a key role for the statistical estimation. Section 2.4 is devoted to the statistical inference in the continuous case $H - 1/\alpha > 0$. The general case is treated in Section 2.5. Finally, Section 2.6 demonstrates some simulation results. All proofs are collected in Section 2.7.

2.3 First properties and some asymptotic results

Distributional and path properties

In this section we review some basic properties of the linear fractional stable motion. First of all, we recall that the symmetric α -stable process $(L_t)_{t \in \mathbb{R}}$ with scale parameter $\sigma > 0$ is uniquely determined by the characteristic function of L_1 , which is given by

$$\mathbb{E}[\exp(itL_1)] = \exp(-\sigma^\alpha |t|^\alpha), \quad t \in \mathbb{R}. \quad (2.3.1)$$

Following the theory of integration with respect to infinitely divisible processes investigated in [34], we know that for any deterministic function $g : \mathbb{R} \rightarrow \mathbb{R}$

$$X = \int_{\mathbb{R}} g_s dL_s < \infty \quad \text{almost surely} \quad \Leftrightarrow \quad \|g\|_\alpha^\alpha := \int_{\mathbb{R}} |g_s|^\alpha ds < \infty.$$

Furthermore, if $\|g\|_\alpha < \infty$ then X has a symmetric α -stable distribution with scale parameter $\sigma \|g\|_\alpha$. In particular, setting

$$X_t = \int_{\mathbb{R}} g_t(s) dL_s, \quad g_t(s) := \left\{ (t-s)_+^{H-1/\alpha} - (-s)_+^{H-1/\alpha} \right\}, \quad (2.3.2)$$

we see that $g_t \in L^\alpha(\mathbb{R})$ for any $t \in \mathbb{R}$, since $|g_t(s)| \leq C_t |s|^{H-1-1/\alpha}$ when $s \rightarrow -\infty$ and $H \in (0, 1)$. Hence, X_t is well defined for any $t \in \mathbb{R}$ and all finite dimensional distributions of the linear fractional stable motion $(X_t)_{t \in \mathbb{R}}$ are symmetric α -stable. It is easily seen that the linear fractional stable motion has stationary increments.

We recall that symmetric α -stable random variables with $\alpha \in (0, 2)$ do not exhibit finite second moments, and hence their dependence structure can't be measured via the classical covariance kernel. Instead it is often useful to consider the following measure of dependence. Let $X = \int_{\mathbb{R}} g_s dL_s$ and $Y = \int_{\mathbb{R}} h_s dL_s$ with $\|g\|_\alpha, \|h\|_\alpha < \infty$. Then we introduce the measure of dependence $U_{g,h} : \mathbb{R}^2 \rightarrow \mathbb{R}$ via

$$\begin{aligned} U_{g,h}(u, v) &:= \mathbb{E}[\exp(i(uX + vY))] - \mathbb{E}[\exp(iuX)]\mathbb{E}[\exp(ivY)] \\ &= \exp(-\sigma^\alpha \|ug + vh\|_\alpha^\alpha) - \exp(-\sigma^\alpha (\|ug\|_\alpha^\alpha + \|vh\|_\alpha^\alpha)). \end{aligned} \quad (2.3.3)$$

The quantity $U_{g,h}$ is extremely useful when computing covariances $\text{cov}(K_1(X), K_2(Y))$ for functions $K_1, K_2 \in L^1(\mathbb{R})$; see for instance [32]. Let \mathfrak{F} denote the Fourier transform and let \mathfrak{F}^{-1} be its inverse. Furthermore, let $p_{(X,Y)}$, p_X and p_Y denote the density of (X, Y) , X and Y , respectively. We recall that these densities are not available in a closed form

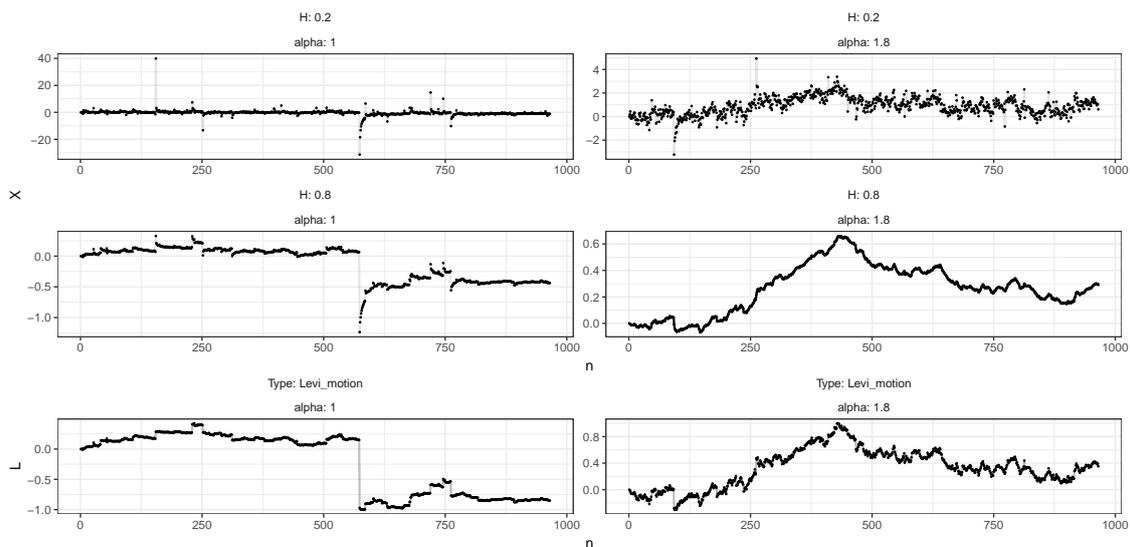


Figure 2.1: Left (from bottom to top): The driving symmetric stable Lévy process with $\alpha = 1$, linear fractional stable motions with parameters $\alpha = 1, H = 0.8$ and $\alpha = 1, H = 0.2$. Right (from bottom to top): The driving symmetric stable Lévy process with $\alpha = 1.8$, linear fractional stable motions with parameters $\alpha = 1.8, H = 0.8$ and $\alpha = 1.8, H = 0.2$.

except in some special cases. Using the duality relationship we obtain the identity

$$\begin{aligned}
 \text{cov}(K_1(X), K_2(Y)) &= \int_{\mathbb{R}^2} K_1(x)K_2(y) \left(p_{(X,Y)}(x, y) - p_X(x)p_Y(y) \right) dx dy \\
 &= \int_{\mathbb{R}^2} K_1(x)K_2(y) \mathfrak{F}^{-1} U_{g,h}(x, y) dx dy \\
 &= \int_{\mathbb{R}^2} \left(\mathfrak{F}^{-1} K_1(x) \right) \left(\mathfrak{F}^{-1} K_2(y) \right) U_{g,h}(x, y) dx dy.
 \end{aligned} \tag{2.3.4}$$

We remark that the latter provides an explicit formula for computation of covariances $\text{cov}(K_1(X), K_2(Y))$.

Finally, we recall that the path properties of a linear fractional stable motion strongly depend on the interplay between the parameters H and α . When $H - 1/\alpha > 0$ the process $(X_t)_{t \in \mathbb{R}}$ is Hölder continuous on compact intervals of any order smaller than $H - 1/\alpha$; we refer to [8] for more details on this property. If $H - 1/\alpha < 0$ the linear fractional stable motion explodes at jump times of the driving Lévy process L ; in particular, X has unbounded paths on compact intervals. We demonstrate some sample paths of the linear fractional stable motions in Figure 2.1. In the critical case $H - 1/\alpha = 0$ we obviously have the identity $X_t = L_t$. In this situation the parameter estimation has been investigated in [3].

Review of the limit theory

In this section we review some probabilistic results, which will be relevant for our estimation method. Due to stationarity of the increments and self-similarity of the process $(X_t)_{t \in \mathbb{R}}$, we can discuss the limit theory for the high and low frequency case simultaneously. We start by introducing higher order increments of X . We denote by $\Delta_{i,k}^{n,r} X$

($i, k, r, n \in \mathbb{N}$) the k th order increment of X at stage i/n and frequency r/n , i.e.

$$\Delta_{i,k}^{n,r} X := \sum_{j=0}^k (-1)^j \binom{k}{j} X_{(i-rj)/n}, \quad i \geq rk. \quad (2.3.5)$$

Note that for $r = k = 1$ we obtain the usual increments $\Delta_{i,1}^{n,1} X = X_{i/n} - X_{(i-1)/n}$. For the ease of notation we will often drop the index r (resp. k and n) in $\Delta_{i,k}^{n,r} X$ and other quantities when $r = 1$ (resp. $k = 1$ and $n = 1$). In particular, the low frequency k th order increments of X are denoted by

$$\Delta_{i,k}^r X := \sum_{j=0}^k (-1)^j \binom{k}{j} X_{i-rj}, \quad i \geq rk. \quad (2.3.6)$$

According to the self-similarity of the process $(X_t)_{t \in \mathbb{R}}$ we readily have that $(n^H \Delta_{i,k}^{n,r} X)_{i \geq rk} \stackrel{d}{=} (\Delta_{i,k}^r X)_{i \geq rk}$. Our main probabilistic tools will be statistics of the form

$$V_{\text{high}}(f; k, r)_n := \frac{1}{n} \sum_{i=rk}^n f(n^H \Delta_{i,k}^{n,r} X), \quad V_{\text{low}}(f; k, r)_n := \frac{1}{n} \sum_{i=rk}^n f(\Delta_{i,k}^r X), \quad (2.3.7)$$

where $f : \mathbb{R} \rightarrow \mathbb{R}$ is a measurable function. It is well known that the process $(X_t)_{t \in \mathbb{R}}$ is mixing, see e.g. [14]. Hence, Birkhoff's ergodic theorem implies the convergence $V_{\text{low}}(f; k, r)_n \rightarrow \mathbb{E}[f(\Delta_{rk,k}^r X)]$ almost surely whenever $\mathbb{E}[|f(\Delta_{rk,k}^r X)|] < \infty$. The same result holds in probability for the statistic $V_{\text{high}}(f; k, r)_n$ due to self-similarity of the process X . However, the weak limit theorems associated with the aforementioned law of large numbers and the framework of functions f with $\mathbb{E}[|f(\Delta_{rk,k}^r X)|] = \infty$ are not completely understood in the literature. To get an idea about possible limits that may appear we briefly demonstrate some recent theoretical developments from the paper [7], where the case $f_p(x) = |x|^p$ ($p > 0$) has been investigated. We remark that their results are obtained for a wider class of processes, namely stationary increments Lévy moving average processes, and we adapt them to the setting of linear fractional stable motions.

We need to introduce some more notation to describe the various limits. For $p \in (-1, 1) \setminus \{0\}$ we define the constant

$$a_p := \begin{cases} \int_{\mathbb{R}} (1 - \cos(y)) |y|^{-1-p} dy : & p \in (0, 1) \\ \sqrt{2\pi} \Gamma(-p/2) / 2^{p+1/2} \Gamma((p+1)/2) : & p \in (-1, 0) \end{cases}, \quad (2.3.8)$$

where Γ denotes the Gamma function. It is easy to see that $a_p > 0$ is indeed finite in all relevant cases. For any functions $g, h \in L^\alpha(\mathbb{R})$, we introduce the notation

$$\theta(g, h)_p = a_p^{-2} \int_{\mathbb{R}^2} |xy|^{-1-p} U_{g,h}(x, y) dx dy, \quad (2.3.9)$$

where $U_{g,h}$ is defined in (2.3.3), whenever the above double integral is finite. Furthermore, for $k, r \in \mathbb{N}$, we define the function $h_{k,r} : \mathbb{R} \rightarrow \mathbb{R}$ by

$$h_{k,r}(x) = \sum_{j=0}^k (-1)^j \binom{k}{j} (x - rj)_+^{H-1/\alpha}, \quad x \in \mathbb{R}. \quad (2.3.10)$$

Below $(U_m)_{m \geq 1}$ is an i.i.d. $\mathcal{U}(0, 1)$ -distributed sequence of random variables independent of L , $(T_m)_{m \geq 1}$ are jump times of L and $\Delta L_{T_m} := L_{T_m} - L_{T_m-}$ are jump sizes. The following result summarises the limit theory for the statistic $V_{\text{high}}(f_p; k)_n$ (i.e. $r = 1$) in the power variation setting.

Theorem 2.3.1. ([7, Theorems 1.1 and 1.2]) *We consider the function $f_p(x) = |x|^p$ ($p > 0$) and assume that $H - 1/\alpha > 0$.*

(i) *(First order asymptotics) If $p > \alpha$ we obtain convergence in law*

$$n^{1-p/\alpha} V_{\text{high}}(f_p; k)_n \xrightarrow{d} \sum_{m: T_m \in [0, 1]} |\Delta L_{T_m}|^p \left(\sum_{l=0}^{\infty} |h_k(l + U_m)|^p \right).$$

If $p < \alpha$ we deduce the law of large numbers

$$V_{\text{high}}(f_p; k)_n \xrightarrow{\mathbb{P}} m_{p,k} := \mathbb{E}[|\Delta_{k,k} X|^p].$$

(ii) *(Second order asymptotics) Assume that $p < \alpha/2$. If $H < k - 1/\alpha$ we obtain the central limit theorem*

$$\sqrt{n} (V_{\text{high}}(f_p; k)_n - m_{p,k}) \xrightarrow{d} \mathcal{N}(0, \eta^2), \quad \eta^2 = \theta(h_k, h_k)_p + 2 \sum_{j=1}^{\infty} \theta(h_k, h_k(\cdot + j))_p,$$

where the quantity $\theta(g, h)$ has been introduced at (2.3.9). If $H > k - 1/\alpha$ we deduce a non-central limit theorem

$$n^{1-1/(1+\alpha(k-H))} (V_{\text{high}}(f_p; k)_n - m_{p,k}) \xrightarrow{d} S,$$

where S is a totally right skewed $(1 + \alpha(k - H))$ -stable random variable with mean zero and scale parameter $\tilde{\sigma}$, which is defined in [7, Theorem 1.2].

We remark that the results of Theorem 2.3.1 remain valid for the low frequency statistic $V_{\text{low}}(f_p; k)_n$ due to self-similarity property of L . Apart from various critical cases Theorem 2.3.1 gives a rather complete understanding of the asymptotic behaviour of the power variation $V_{\text{high}}(f_p; k)_n$ in the setting $H - 1/\alpha > 0$. The strong law of large numbers in Theorem 2.3.1(i) will be useful for estimation of the parameter H . However, without an a priori knowledge about the stability parameter α , we can't insure that the condition $p < \alpha$ holds. Similarly, we would like to use the central limit theorem in Theorem 2.3.1(ii) whose convergence rate \sqrt{n} is faster than the rate $n^{1-1/(1+\alpha(k-H))}$ in the non-central limit theorem. But the conditions of Theorem 2.3.1(ii) rely again on an a priori knowledge about α .

There are some few related results in the literature. In [31] the authors have shown a central limit theorem a standardised version of the statistic $\sum_{i=1}^n f(Y_i)$, where f is a bounded function and $(Y_t)_{t \in \mathbb{R}}$ is a stable moving average process. In a later work [32] the result has been extended to a certain class of unbounded functions f under the additional assumption that $\alpha \in (1, 2)$. Similarly to Theorem 2.3.1 the sufficient conditions for the validity of the central limit theorems in [31, 32] depend on the interplay between the kernel function of the stable moving average process and the stability index α . We remark that extensions of these results in various directions will be necessary to obtain the full asymptotic theory for estimators of the parameter $\theta = (\sigma, \alpha, H)$.

A multivariate weak limit theorem

Although Theorem 2.3.1(ii) gives a rather complete picture of the weak limit theory in the power variation case, we will require a much stronger result for our statistical applications. We introduce the function $\psi_t : \mathbb{R} \rightarrow \mathbb{R}$ with $\psi_t(x) = \cos(tx)$ and define the statistics

$$\varphi_{\text{high}}(t; H, k)_n := V_{\text{high}}(\psi_t; k)_n \quad \text{and} \quad \varphi_{\text{low}}(t; k)_n := V_{\text{low}}(\psi_t; k)_n, \quad (2.3.11)$$

which correspond to $r = 1$. Notice that, in contrast to $\varphi_{\text{low}}(t; k)_n$, the high frequency statistic $\varphi_{\text{high}}(t; H, k)_n$ depends on the unknown self-similarity parameter H . In fact, this is the major difference between the high and low frequency settings, which will result in different rates of convergence later on. Applying again the strong law of large numbers we readily obtain the strong consistency

$$\varphi_{\text{low}}(t; k)_n \xrightarrow{\text{a.s.}} \varphi(t; k) := \exp(-|\sigma \|h_k\|_\alpha t|^\alpha). \quad (2.3.12)$$

Clearly, the same result holds in probability for the high frequency statistic $\varphi_{\text{high}}(t; H, k)_n$. Next, we introduce various types of statistics, which will play a major role in estimation of the unknown parameter θ . More specifically, we will extend the definition of power variation to certain negative powers and prove a multivariate limit theorem for power variations and empirical characteristic functions. We fix $d \in \mathbb{N}$ and define the statistics for any $1 \leq j \leq d$, $r_j \in \{1, 2\}$, $p \in (-1/2, 1/2) \setminus \{0\}$ and $t_j > 0$:

$$\left. \begin{aligned} W(n)_j^{(1)} &:= \sqrt{n} \left(V_{\text{low}}(f_p; k_j, r_j) - r_j^H m_{p, k_j} \right) \\ W(n)_j^{(2)} &:= \sqrt{n} \left(V_{\text{low}}(\psi_{t_j}; k_j)_n - \varphi(t_j; k_j) \right) \end{aligned} \right\} \quad \text{when } k_j > H + 1/\alpha \quad (2.3.13)$$

$$\left. \begin{aligned} S(n)_j^{(1)} &:= n^{1-1/(1+\alpha(k-H))} \left(V_{\text{low}}(f_p; k, r_j) - r_j^H m_{p, k} \right) \\ S(n)_j^{(2)} &:= n^{1-1/(1+\alpha(k-H))} \left(V_{\text{low}}(\psi_{t_j}; k)_n - \varphi(t_j; k) \right) \end{aligned} \right\} \quad \text{when } k < H + 1/\alpha$$

Note the identity $\mathbb{E}[|\Delta_{r_k, k}^r X|^p] = r^H m_{p, k}$, which explains the centring of the statistics $W(n)^{(1)}$ and $S(n)^{(1)}$. We remark that the functionals $W(n)^{(1)}$ and $W(n)^{(2)}$ are in the domain of attraction of the normal distribution (under appropriate assumption on the powers p) while the functionals $S(n)^{(1)}$ and $S(n)^{(2)}$ are in the domain of attraction of the $(1 + \alpha(k - H))$ -stable distribution. The latter fact is rather surprising since the statistic $S(n)_j^{(2)}$ exhibits finite moments of any order.

Before we proceed with the main result of this section we need to introduce some more notation. In the first step, for any $x \in \mathbb{R}$, we define the functions

$$\Phi_j^{(1)}(x) = \mathbb{E}[f_p(\Delta_{r_j k, k}^{r_j} X + x)] - \mathbb{E}[f_p(\Delta_{r_j k, k}^{r_j} X)], \quad (2.3.14)$$

$$\Phi_j^{(2)}(x) = \mathbb{E}[\psi_{t_j}(\Delta_{k, k} X + x)] - \mathbb{E}[\psi_{t_j}(\Delta_{k, k} X)].$$

Since the functions f_p and ψ_t are even we readily obtain that $\Phi_j^{(l)}(0) = \nabla \Phi_j^{(l)}(0) = 0$ for all l, j . Thus, using Lemma 2.7.5, we deduce the growth estimates

$$|\Phi_j^{(1)}(x)| \leq C \left(x^2 \wedge |x|^{\max\{p, 0\}} \right), \quad |\Phi_j^{(2)}(x)| \leq C \left(x^2 \wedge 1 \right), \quad (2.3.15)$$

for some positive constant C . Next, we introduce the functions

$$\bar{\Phi}_j^{(1)}(x) = \sum_{i=1}^{\infty} \Phi_j^{(1)}(h_{k,r_j}(i)x), \quad \bar{\Phi}_j^{(2)}(x) = \sum_{i=1}^{\infty} \Phi_j^{(2)}(h_k(i)x). \quad (2.3.16)$$

Note that these functions are indeed finite due to (2.3.15) and the estimate $|h_{k,r}(x)| \leq C|x|^{H-1/\alpha-k}$ for large x . Finally, we set $\bar{\Phi} = (\bar{\Phi}^{(1)}, \bar{\Phi}^{(2)}) = (\Phi_1^{(1)}, \dots, \Phi_d^{(1)}, \Phi_1^{(2)}, \dots, \Phi_d^{(2)})$. The main probabilistic result of this paper is the following theorem.

Theorem 2.3.2. *Assume that either $p \in (-1/2, 0)$ or $p \in (0, 1/2)$ and $p < \alpha/2$. Set $W(n)^{(i)} = (W(n)_1^{(i)}, \dots, W(n)_d^{(i)})$ and $S(n)^{(i)} = (S(n)_1^{(i)}, \dots, S(n)_d^{(i)})$ for $i = 1, 2$. Then we obtain weak convergence in law on \mathbb{R}^{4d} :*

$$\left(W(n)^{(1)}, W(n)^{(2)}, S(n)^{(1)}, S(n)^{(2)} \right) \xrightarrow{d} \left(W^{(1)}, W^{(2)}, S^{(1)}, S^{(2)} \right), \quad (2.3.17)$$

where $W = (W^{(1)}, W^{(2)})$ and $S = (S^{(1)}, S^{(2)})$ are independent, W is a centred $2d$ -dimensional normal distribution with covariance matrix determined by

$$\text{cov}\left(W_j^{(i)}, W_{j'}^{(i')}\right) = \lim_{n \rightarrow \infty} \text{cov}\left(W(n)_j^{(i)}, W(n)_{j'}^{(i')}\right) \quad 1 \leq j, j' \leq d, \quad i, i' = 1, 2,$$

and $S^{(1)}, S^{(2)}$ are independent d -dimensional $(1 + \alpha(k - H))$ -stable random variables. The law of $S^{(1)}$ (resp. $S^{(2)}$) is determined by the Lévy measure ν_1 (resp. ν_2) whose support is the cone $(\mathbb{R}_+)^d$ (resp. $(\mathbb{R}_-)^d$). More specifically, for any Borel sets $A_1 \in (\mathbb{R}_+)^d$, $A_2 \in (\mathbb{R}_-)^d$ bounded away from 0 the quantities $\nu_1(A_1), \nu_2(A_2)$ are determined by the identity

$$\nu_l(A_l) = \lim_{n \rightarrow \infty} n\mathbb{P}\left(n^{-1/(1+\alpha(k-H))}\bar{\Phi}^{(l)}(L_1) \in A_l\right), \quad l = 1, 2. \quad (2.3.18)$$

The probabilistic result of Theorem 2.3.2 is new in the literature; neither the negative power variations nor the (real part of) empirical characteristic function have been studied from the distributional perspective. We remark that the statistics $W(n)^{(1)}$ and $S(n)^{(1)}$ use the same powers p while the quantities $S(n)^{(1)}$ and $S(n)^{(2)}$ are based on the same order of increments k . The result of Theorem 2.3.2 does not really use these particular restrictions, but its statement is sufficient for the statistical application under investigation.

There exists an explicit expression for the covariance matrix of the limit W . We obtain the following representations:

$$\begin{aligned} \text{cov}\left(W_j^{(1)}, W_{j'}^{(1)}\right) &= \sum_{l \in \mathbb{Z}} \theta(h_{k_j, r_j}, h_{k_{j'}, r_{j'}}(\cdot + l))_p, \\ \text{cov}\left(W_j^{(2)}, W_{j'}^{(2)}\right) &= \frac{1}{2} \sum_{l \in \mathbb{Z}} \left(U_{h_{k_j}, h_{k_{j'}}(\cdot + l)}(t_j, t_{j'}) + U_{h_{k_j}, -h_{k_{j'}}(\cdot + l)}(t_j, t_{j'}) \right), \\ \text{cov}\left(W_j^{(1)}, W_{j'}^{(2)}\right) &= \sum_{l \in \mathbb{Z}} \bar{\theta}(l)_{jj'}, \end{aligned} \quad (2.3.19)$$

with

$$\bar{\theta}(l)_{jj'} = -a_p^{-1} \int_{\mathbb{R}} |y|^{-1-p} U_{h_{k_j}, h_{k_{j'}}(\cdot + l)}(y, t_{j'}) dy.$$

We will prove that $\text{cov}(W) < \infty$ in all relevant cases and the mapping $(\sigma, \alpha, H) \mapsto \text{cov}(W)$ is continuous (see Section 2.7). The latter allows us to estimate the covariance matrix

$\text{cov}(W) < \infty$ and thus obtain a feasible version of the central limit theorem in Theorem 2.3.2.

Similarly, the Lévy measures ν_l ($l = 1, 2$) can be determined explicitly. First of all, the representation (2.7.2) from Section 2.7 implies the identities

$$\begin{aligned}\Phi_j^{(1)}(x) &= a_p^{-1} \int_{\mathbb{R}} (1 - \cos(ux)) \exp\left(-|\sigma \|h_{k,r_j}\|_{\alpha} |u|^{\alpha}\right) |u|^{-1-p} du, \\ \Phi_j^{(2)}(x) &= (\cos(t_j x) - 1) \exp\left(-|\sigma \|h_k\|_{\alpha} t_j |x|^{\alpha}\right).\end{aligned}$$

In particular, it holds that $\Phi_j^{(1)}(x) \geq 0$ and $\Phi_j^{(2)}(x) \leq 0$. In the next step we need to determine the asymptotic behaviour of $\bar{\Phi}_j^{(1)}(x)$ (resp. $\bar{\Phi}_j^{(2)}(x)$) as $x \rightarrow \infty$ (resp. as $x \rightarrow -\infty$). By the substitution $u = (x/z)^{1/(k+1/\alpha-H)}$ we have that

$$\begin{aligned}x^{1/(H-k-1/\alpha)} \bar{\Phi}_j^{(1)}(x) &= x^{1/(H-k-1/\alpha)} \int_0^{\infty} \Phi_j^{(1)}\left(h_{k,r_j}(\lfloor u \rfloor + 1)x\right) du \\ &= (k + 1/\alpha - H)^{-1} \int_0^{\infty} \Phi_j^{(1)}\left(h_{k,r_j}(\lfloor (x/z)^{1/(k+1/\alpha-H)} \rfloor + 1)x\right) z^{-1+1/(H-k-1/\alpha)} dz \\ &\rightarrow c_j^{(1)} := (k + 1/\alpha - H)^{-1} \int_0^{\infty} \Phi_j^{(1)}\left(r_j^k \prod_{i=0}^{k-1} (H - 1/\alpha - i) \cdot z\right) z^{-1+1/(H-k-1/\alpha)} dz\end{aligned}\tag{2.3.20}$$

as $x \rightarrow \infty$. The convergence at (2.3.20) follows from the asymptotic behaviour $h_{k,r_j}(x) \sim r_j^k \prod_{i=0}^{k-1} (H - 1/\alpha - i) \cdot x^{H-1/\alpha-k}$ as $x \rightarrow \infty$. Applying the same technique we deduce that

$$\begin{aligned}|x|^{1/(H-k-1/\alpha)} \bar{\Phi}_j^{(2)}(x) &\rightarrow \\ c_j^{(2)} &:= (k + 1/\alpha - H)^{-1} \int_0^{\infty} \Phi_j^{(2)}\left(\prod_{i=0}^{k-1} (H - 1/\alpha - i) \cdot z\right) z^{-1+1/(H-k-1/\alpha)} dz\end{aligned}\tag{2.3.21}$$

as $x \rightarrow -\infty$. Now, both measures ν_1 and ν_2 from Theorem 2.3.2 can be related to the Lévy measure ν of L . We introduce the mappings $\tau_1 : \mathbb{R}_+ \rightarrow (\mathbb{R}_+)^d$ and $\tau_2 : \mathbb{R}_- \rightarrow (\mathbb{R}_-)^d$ via

$$\tau_1(x) = x^{1/(k+1/\alpha-H)} \left(c_1^{(1)}, \dots, c_d^{(1)}\right), \quad \tau_2(x) = |x|^{1/(k+1/\alpha-H)} \left(c_1^{(2)}, \dots, c_d^{(2)}\right).$$

Then, for Borel sets A_1, A_2 as defined in Theorem 2.3.2, we deduce the identity

$$\nu_l(A_l) = \lim_{n \rightarrow \infty} n \mathbb{P}\left(\tau_l(n^{-1/\alpha} L_1) \in A_l\right) = \nu\left(\tau_l^{-1}(A_l)\right), \quad l = 1, 2.\tag{2.3.22}$$

2.4 Statistical inference in the continuous case $H - 1/\alpha > 0$

We start with the continuous case $H - 1/\alpha > 0$, which turns out to be somewhat easier to treat compared to the general setting. Since $H \in (0, 1)$ and $\alpha \in (0, 2)$, condition $H - 1/\alpha > 0$ implies the restrictions

$$\alpha \in (1, 2) \quad \text{and} \quad H \in (1/2, 1).$$

It is the lower bound $\alpha > 1$ that enables us to use the law of large numbers in Theorem 2.3.1(i) whenever $p < 1$, and the central limit theorem in Theorem 2.3.1(ii) whenever $p < 1/2$ and $H < k - 1/\alpha$. The latter condition $H < k - 1/\alpha$ never holds for $k = 1$ since $0 < H - 1/\alpha < 1 - 2/\alpha < 0$ gives a contradiction, but it is always satisfied for any $k \geq 2$ since

$$H < 1 < k - 1/\alpha \quad \text{for any } k \geq 2,$$

because $\alpha > 1$.

Now, we introduce an estimator for the parameter $\theta = (\sigma, \alpha, H)$ in high and low frequency setting. We start with the statistical inference for the self-similarity parameter H , which is based upon a ratio statistic that compares power variations at two different frequencies. More specifically, we define the quantities

$$R_{\text{high}}(p, k)_n := \frac{\sum_{i=2k}^n |\Delta_{i,k}^{n,2} X|^p}{\sum_{i=k}^n |\Delta_{i,k}^{n,1} X|^p}, \quad R_{\text{low}}(p, k)_n := \frac{\sum_{i=2k}^n |\Delta_{i,k}^2 X|^p}{\sum_{i=k}^n |\Delta_{i,k}^1 X|^p}, \quad (2.4.1)$$

where the increments $\Delta_{i,k}^r X$ have been defined at (2.3.6). We obtain the convergence

$$R_{\text{high}}(p, k)_n \xrightarrow{\mathbb{P}} 2^{pH}, \quad R_{\text{low}}(p, k)_n \xrightarrow{\text{a.s.}} 2^{pH}$$

for any $p \in (0, 1)$ as an immediate consequence of Theorem 2.3.1(i). Consequently, defining the statistics

$$\hat{H}_{\text{high}}(p, k)_n := \frac{1}{p} \log_2 R_{\text{high}}(p, k)_n, \quad \hat{H}_{\text{low}}(p, k)_n := \frac{1}{p} \log_2 R_{\text{low}}(p, k)_n, \quad (2.4.2)$$

we deduce the consistency $\hat{H}_{\text{high}}(p, k)_n \xrightarrow{\mathbb{P}} H$, $\hat{H}_{\text{low}}(p, k)_n \xrightarrow{\text{a.s.}} H$ as $n \rightarrow \infty$ for any $k \geq 1$ and any $p \in (0, 1)$. We remark that this type of ratio statistics is commonly used in the framework of fBm's when estimating the Hurst parameter H (see e.g. [25] among many others). In the Gaussian setting, which corresponds to $\alpha = 2$, the central limit theorem for the quantity $\sqrt{n}(\hat{H}_{\text{high}}(p, k)_n - H)$ holds for all $k \geq 2$ and also for $k = 1$ if further $H \in (0, 3/4)$. As we indicated above, in the framework of pure jump α -stable driving motion L the central limit theorem never holds if $k = 1$. Hence, there is no smooth transition between the non-Gaussian and Gaussian setting when $\alpha \rightarrow 2$.

The estimation strategy for the parameter $\theta = (\sigma, \alpha, H)$ based on high frequency observations is now straightforward: Infer the self-similarity parameter H by (2.4.2) and use the plug-in estimator $\varphi_{\text{high}}(t; \hat{H}_{\text{high}}(p, k), k)_n$ for two different values of t to infer the scale parameter σ and the stability index α . For the latter step we consider $t_2 > t_1 > 0$ and observe the identities

$$\sigma = (-\log \varphi(t_1; k))^{1/\alpha} / t_1 \|h_k\|_\alpha, \quad \alpha = \frac{\log |\log \varphi(t_2; k)| - \log |\log \varphi(t_1; k)|}{\log t_2 - \log t_1}.$$

Recalling that the function h_k depends on α and H , we readily obtain a function G such that

$$(\sigma, \alpha) = G(\varphi(t_1; k), \varphi(t_2; k), H) \quad (2.4.3)$$

where we applied the above identities. Next, we present the estimator of the pair (σ, α) in high and low frequency setting, recalling that the estimators of the self-similarity parameter H have been defined at (2.4.2). We introduce the following estimators:

$$\begin{aligned} & (\widehat{\sigma}_{\text{high}}(k, t_1, t_2)_n, \widehat{\alpha}_{\text{high}}(k, t_1, t_2)_n) \\ &= G\left(\varphi_{\text{high}}(t_1; \widehat{H}_{\text{high}}(p, k)_n, k)_n, \varphi_{\text{high}}(t_2; \widehat{H}_{\text{high}}(p, k)_n, k)_n, \widehat{H}_{\text{high}}(p, k)_n\right), \\ & (\widehat{\sigma}_{\text{low}}(k, t_1, t_2)_n, \widehat{\alpha}_{\text{low}}(k, t_1, t_2)_n) = G\left(\varphi_{\text{low}}(t_1; k)_n, \varphi_{\text{low}}(t_2; k)_n, \widehat{H}_{\text{low}}(p, k)_n\right). \end{aligned} \quad (2.4.4)$$

Before we present the main result of this section we need to introduce more notation. We define the functions $v_p : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ and $F : \mathbb{R}_+^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^3$ by

$$v_p(x, y) = p^{-1}(\log_2 y - \log_2 x), \quad F(x, y, u, w) = (G(u, w, v_p(x, y)), v_p(x, y)), \quad (2.4.5)$$

and let JF denotes the Jacobian of F . For any matrix A we write A^* for its transpose. The asymptotic normality in the low and high frequency setting is summarised in the following theorem.

Theorem 2.4.1. *Consider the linear fractional stable motion $(X_t)_{t \in \mathbb{R}}$ introduced at (2.2.1). Let $k \geq 2$ and $t_2 > t_1 > 0$.*

(i) *(Low frequency case) Let $W = (W^{(1)}, W^{(2)})$ be the 4-dimensional normal limit defined in Theorem 2.3.2 associated with $d = 2$, $p \in (0, 1/2)$, $k_1 = k_2 = k$ and $r_j = j$. Then we obtain the central limit theorem*

$$\sqrt{n} \begin{pmatrix} \widehat{\sigma}_{\text{low}}(k, t_1, t_2)_n - \sigma \\ \widehat{\alpha}_{\text{low}}(k, t_1, t_2)_n - \alpha \\ \widehat{H}_{\text{low}}(p, k)_n - H \end{pmatrix} \xrightarrow{d} B_{\text{low}}^{\text{nor}}(p, k) = JF\left(m_{p,k}, 2^H m_{p,k}, \varphi(t_1; k), \varphi(t_2; k)\right) W^*.$$

(ii) *(High frequency case) We obtain the central limit theorem*

$$\begin{pmatrix} \sqrt{n}(\log n)^{-1} (\widehat{\sigma}_{\text{high}}(k, t_1, t_2)_n - \sigma) \\ \sqrt{n}(\log n)^{-1} (\widehat{\alpha}_{\text{high}}(k, t_1, t_2)_n - \alpha) \\ \sqrt{n} (\widehat{H}_{\text{high}}(p, k)_n - H) \end{pmatrix} \xrightarrow{d} B_{\text{high}}^{\text{nor}}(p, k) = \nabla v(m_{p,k}, 2^H m_{p,k})(W^{(1)})^* \times \begin{pmatrix} \nabla G_1(\varphi(t_1; k), \varphi(t_2; k), H) (t_1 \varphi'(t_1; k), t_2 \varphi'(t_2; k), 0)^* \\ \nabla G_2(\varphi(t_1; k), \varphi(t_2; k), H) (t_1 \varphi'(t_1; k), t_2 \varphi'(t_2; k), 0)^* \\ 1 \end{pmatrix}.$$

We remark that the central limit theorem of Theorem 2.4.1(i) is a simple consequence of Theorem 2.3.2 and the delta method. In contrast to the low frequency case Theorem 2.4.1(ii) is degenerate in the sense that the limit distribution is solely driven by the asymptotics of the term $\sqrt{n}(\widehat{H}_{\text{high}}(p, k)_n - H)$. Since the parameter H enters the quantity $\varphi_{\text{high}}(t; H, k)_n$ via n^H the additional term $(\log n)^{-1}$ appears in the convergence rate.

For a later use we need to extend the definition of the random variables $B_{\text{high}}^{\text{nor}}(p, k)$ and $B_{\text{low}}^{\text{nor}}(p, k)$ to various directions. First of all, we will allow for negative powers $-p$ with $p \in (0, 1/2)$. Secondly, we would like to define the same limiting variables but associated with the stable limit $S = (S^{(1)}, S^{(2)})$ from Theorem 2.3.2 rather than W . Thus, for $d = 2$,

$p \in (-1/2, 1/2) \setminus \{0\}$, $k_1 = k_2 = k$ and $r_j = j$, we set

$$B_{\text{low}}^{\text{sta}}(p, k) = JF \left(m_{p,k}, 2^H m_{p,k}, \varphi(t_1; k), \varphi(t_2; k) \right) S^*,$$

$$B_{\text{high}}^{\text{sta}}(p, k) =$$

$$\nabla v(m_{p,k}, 2^H m_{p,k})(S^{(1)})^* \times \begin{pmatrix} \nabla G_1(\varphi(t_1; k), \varphi(t_2; k), H) (t_1 \varphi'(t_1; k), t_2 \varphi'(t_2; k), 0)^* \\ \nabla G_2(\varphi(t_1; k), \varphi(t_2; k), H) (t_1 \varphi'(t_1; k), t_2 \varphi'(t_2; k), 0)^* \\ 1 \end{pmatrix}.$$

Remark 2.4.2. In Theorem 2.4.1 we use two values $t_1, t_2 \in \mathbb{R}_+^2$ and an estimator \hat{H} to infer the parameters (σ, α) . Applying basic statistical principles it is more natural to use all $t \in \mathbb{R}_+$ for the estimation procedure. For example, when considering the low frequency framework, we may estimate the parameters (σ, α) via a minimal contrast approach. Given a positive weight function $w \in L^1(\mathbb{R}_+)$ we obtain an estimator $(\tilde{\sigma}_n, \tilde{\alpha}_n)$ of (σ, α) by

$$(\tilde{\sigma}_n, \tilde{\alpha}_n) \in \operatorname{argmin}_{\theta \in \mathbb{R}_+ \times (0, 2)} \int_0^\infty (\varphi_{\text{low}}(t; k)_n - \hat{\varphi}(t; k))^2 w(t) dt,$$

where $\hat{\varphi}(t; k) = \varphi(\hat{H}_{\text{low}}(p, k)_n, t; k)$. In this setting we are likely to require tightness or a similar property of the stochastic process $\varphi_{\text{low}}(\cdot; k)_n$ to prove asymptotic normality of $(\tilde{\sigma}_n, \tilde{\alpha}_n)$. However, this seems to be a non-trivial problem, at least when using standard tightness criteria for the space $(C(\mathbb{R}_+), \|\cdot\|_\infty)$. We leave it for future research. \square

Remark 2.4.3. The described statistical methodology can be applied to more general processes than the mere linear fractional stable motion. In the paper [7] the authors investigated limit theorems for stochastic processes of the form

$$Y_t = \int_{\mathbb{R}} \{g(t-s) - g_0(-s)\} dL_s,$$

where g, g_0 are deterministic functions vanishing on \mathbb{R}_- with $g(x) = x^{H-1/\alpha} f(x)$ and $f(0) \neq 0$, and L is a symmetric α -stable Lévy motion. In the high frequency setting the process Y exhibits the tangent process $f(0)X$, i.e. we have that

$$\Delta_{i,k}^{n,r} Y \approx f(0) \Delta_{i,k}^{n,r} X.$$

In particular, under certain assumption on f (cf. [7]), the central limit theorem part of Theorem 2.3.2 holds for the more general class of processes Y . Hence, in this semi-parametric model it is possible to estimate the parameter $(|f(0)|\sigma, \alpha, H)$ via the same approach as presented in Theorem 2.4.1(ii). We remark that the function f can't be inferred from high frequency observations on a fixed time interval. \square

2.5 Statistical inference in the general case

In this section we treat the case of a general linear fractional stable motion as it has been introduced at (2.2.1). We recall that in the continuous setting the restriction $H - 1/\alpha > 0$ has led to the lower bound $\alpha > 1$, which is essential for obtaining the asymptotic results of Theorem 2.4.1. Without having an explicit lower bound for the stability parameter α statistical inference turns out to be more complex. As a consequence we will require a different estimation method for the self-similarity parameter H and a two-step procedure to choose the right order of increments k . Furthermore, in order to obtain fast rates of convergence we need different treatments for the low and high frequency frameworks.

Low frequency setting

We note that the basic idea behind the ratio statistic $R_{\text{low}}(p, k)_n$ introduced in (2.4.1) is the homogeneity of the function $f_p(x) = |x|^p$ and the fact that $m_{p,k} < \infty$ which is a consequence of $p < \alpha$ (for the associated central limit theorem we need the stronger condition $p < \alpha/2$). In order to keep both properties we may instead consider the negative power variation, which corresponds to the function $f_{-p}(x) = |x|^{-p}$, and we assume throughout this section that $p \in (0, 1/2)$. This approach has been originally proposed in [21], although central limit theorems have not been investigated in this setting. Note that the function f_{-p} is still homogenous and $m_{-2p,k} < \infty$, which is due to the fact that for any random variable Y with bounded density near 0 it holds that $\mathbb{E}[|Y|^a] < \infty$ for all $a \in (-1, 0)$. Thus, $\widehat{H}_{\text{low}}(-p, k)_n$ is a strongly consistent estimator of the parameter H for any $p \in (0, 1/2)$.

In the next step we need to ensure that we end up in the domain of attraction of the central limit theorem in Theorem 2.3.1(ii), which requires that $k > H + 1/\alpha$. To guarantee this we need a preliminary estimator of the parameter α . They are obtained as in (2.4.4) using the function f_{-p} and $k = 1$:

$$\widehat{\alpha}_{\text{low}}^0(t_1, t_2)_n = G_2 \left(\varphi_{\text{low}}(t_1)_n, \varphi_{\text{low}}(t_2)_n, \widehat{H}_{\text{low}}(-p)_n \right), \quad (2.5.1)$$

where $G = (G_1, G_2)$. Notice that this estimator is consistent, but we do not know if it is in the domain of attraction of a normal distribution or not. Now, we define

$$\widehat{k}_{\text{low}}(t_1, t_2)_n = 2 + \lfloor \widehat{\alpha}_{\text{low}}^0(t_1, t_2)_n^{-1} \rfloor. \quad (2.5.2)$$

For the sake of brevity we write $\widehat{k}_{\text{low}} = \widehat{k}_{\text{low}}(t_1, t_2)_n$. In the second step we estimate the parameter $\theta = (\sigma, \alpha, H)$ using \widehat{k}_{low} . The self-similarity parameter H is thus estimated by $\widehat{H}_{\text{low}}(-p, \widehat{k}_{\text{low}})_n$. Next, similarly to definitions at (2.4.4), we introduce the estimators

$$\begin{aligned} & \left(\widetilde{\sigma}_{\text{low}}(\widehat{k}_{\text{low}}, t_1, t_2)_n, \widetilde{\alpha}_{\text{low}}(\widehat{k}_{\text{low}}, t_1, t_2)_n \right) \\ &= G \left(\varphi_{\text{low}}(t_1; \widehat{k}_{\text{low}})_n, \varphi_{\text{low}}(t_2; \widehat{k}_{\text{low}})_n, \widehat{H}_{\text{low}}(-p, \widehat{k}_{\text{low}})_n \right). \end{aligned} \quad (2.5.3)$$

In order to determine the asymptotic distribution of the proposed estimators we will need the full force of Theorem 2.3.2. Due to definition (2.5.2) we also require a separate treatment of the cases $\alpha^{-1} \notin \mathbb{N}$ and $\alpha^{-1} \in \mathbb{N}$. In the first case $\widehat{k}_{\text{low}} \xrightarrow{\text{a.s.}} 2 + \lfloor \alpha^{-1} \rfloor$ while in the second case we will have

$$\mathbb{P} \left(\widehat{k}_{\text{low}} = 2 + \alpha^{-1} \right) \rightarrow \lambda \quad \text{and} \quad \mathbb{P} \left(\widehat{k}_{\text{low}} = 1 + \alpha^{-1} \right) \rightarrow 1 - \lambda$$

for a certain constant $\lambda \in (0, 1)$. In the first setting, which is easier to treat, we obtain the following result.

Theorem 2.5.1. *Let X be the linear fractional stable motion defined at (2.2.1). Assume that $p \in (0, 1/2)$ and $\alpha^{-1} \notin \mathbb{N}$. We obtain the central limit theorem*

$$\sqrt{n} \begin{pmatrix} \widetilde{\sigma}_{\text{low}}(\widehat{k}_{\text{low}}, t_1, t_2)_n - \sigma \\ \widetilde{\alpha}_{\text{low}}(\widehat{k}_{\text{low}}, t_1, t_2)_n - \alpha \\ \widehat{H}_{\text{low}}(-p, \widehat{k}_{\text{low}})_n - H \end{pmatrix} \xrightarrow{d} B_{\text{low}}^{\text{nor}} \left(-p, 2 + \lfloor \alpha^{-1} \rfloor \right).$$

In the framework $\alpha^{-1} \in \mathbb{N}$ we distinguish two further cases, that determine the asymptotic behaviour the preliminary estimate $\hat{\alpha}_{\text{low}}^0$, which is constructed using $k = 1$. According to Theorem 2.3.2 we are in the domain of the validity of a central limit theorem when $H < 1 - 1/\alpha$ while a non-central limit theorem holds if $H > 1 - 1/\alpha$.

Proposition 2.5.2. *Let X be the linear fractional stable motion defined at (2.2.1). Assume that $p \in (0, 1/2)$.*

(i) (Normal case) *Assume that $H < 1 - 1/\alpha$. Then we obtain the central limit theorem*

$$\sqrt{n} \left(\hat{\alpha}_{\text{low}}^0(t_1, t_2)_n - \alpha \right) \xrightarrow{d} B_{\text{low}}^{\text{nor}}(-p, 1)_2.$$

(ii) (Stable case) *Assume that $H > 1 - 1/\alpha$. Then we obtain the weak limit theorem*

$$n^{1-1/(1+\alpha(1-H))} \left(\hat{\alpha}_{\text{low}}^0(t_1, t_2)_n - \alpha \right) \xrightarrow{d} B_{\text{low}}^{\text{sta}}(-p, 1)_2.$$

We note that the result of Proposition 2.5.2(ii) is essentially the same as in the asymptotically normal regime except that the convergence rate is now $n^{1-1/(1+\alpha(1-H))}$ and the normal limit W is replaced by S .

The next theorem presents the statistical behaviour of the estimator $(\tilde{\sigma}_{\text{low}}, \tilde{\alpha}_{\text{low}}, \hat{H}_{\text{low}}(-p, \hat{k}_{\text{low}})_n)$ in the case $\alpha^{-1} \in \mathbb{N}$.

Theorem 2.5.3. *Let X be the linear fractional stable motion defined at (2.2.1). Assume that $p \in (0, 1/2)$ and $\alpha^{-1} \in \mathbb{N}$.*

(i) (Case $H < 1 - 1/\alpha$) *Assume that $H < 1 - 1/\alpha$. Then we obtain*

$$\sqrt{n} \begin{pmatrix} \tilde{\sigma}_{\text{low}}(\hat{k}_{\text{low}}, t_1, t_2)_n - \sigma \\ \tilde{\alpha}_{\text{low}}(\hat{k}_{\text{low}}, t_1, t_2)_n - \alpha \\ \hat{H}_{\text{low}}(-p, \hat{k}_{\text{low}})_n - H \end{pmatrix} \xrightarrow{d} D_{\text{low}}^{\text{nor}},$$

where the probability distribution $D_{\text{low}}^{\text{nor}}$ on \mathbb{R}^3 is given by

$$D_{\text{low}}^{\text{nor}}(\cdot) = \mathbb{P}(\{B_{\text{low}}^{\text{nor}}(-p, 2 + \alpha^{-1}) \in \cdot\} \cap \{B_{\text{low}}^{\text{nor}}(-p, 1)_2 < 0\}) \\ + \mathbb{P}(\{B_{\text{low}}^{\text{nor}}(-p, 1 + \alpha^{-1}) \in \cdot\} \cap \{B_{\text{low}}^{\text{nor}}(-p, 1)_2 > 0\}).$$

(ii) (Case $H > 1 - 1/\alpha$) *Assume that $H > 1 - 1/\alpha$. Then we obtain*

$$\sqrt{n} \begin{pmatrix} \tilde{\sigma}_{\text{low}}(\hat{k}_{\text{low}}, t_1, t_2)_n - \sigma \\ \tilde{\alpha}_{\text{low}}(\hat{k}_{\text{low}}, t_1, t_2)_n - \alpha \\ \hat{H}_{\text{low}}(-p, \hat{k}_{\text{low}})_n - H \end{pmatrix} \xrightarrow{d} D_{\text{low}}^{\text{sta}},$$

where the probability distribution $D_{\text{low}}^{\text{sta}}$ on \mathbb{R}^3 is given by

$$D_{\text{low}}^{\text{sta}}(\cdot) = \mathbb{P}(B_{\text{low}}^{\text{sta}}(-p, 1)_2 < 0) \mathbb{P}(B_{\text{low}}^{\text{nor}}(-p, 2 + \alpha^{-1}) \in \cdot) \\ + \mathbb{P}(B_{\text{low}}^{\text{sta}}(-p, 1)_2 > 0) \mathbb{P}(B_{\text{low}}^{\text{nor}}(-p, 1 + \alpha^{-1}) \in \cdot).$$

According to Theorem 2.3.2 the statistic $(B_{\text{low}}^{\text{nor}}(-p, k), B_{\text{low}}^{\text{nor}}(-p, 1))$ is jointly normal for $k \in \{1 + \alpha^{-1}, 2 + \alpha^{-1}\}$. Thus, the probability distribution $D_{\text{low}}^{\text{nor}}$ can be easily computed using conditioning rules for normal distribution.

Note however that it is problematic to use Theorem 2.5.3 for constructing confidence regions since we do not know a priori whether part (i) or part (ii) applies. We now introduce

a decision rule that helps us to solve this problem. Let $t_4 > t_3 > t_2 > t_1 > 0$ be given real numbers and let $\hat{\alpha}_{\text{low}}^0(t_1, t_2)_n, \hat{\alpha}_{\text{low}}^0(t_3, t_4)_n$ be two estimators of parameter $\alpha \in (0, 2)$ defined at (2.5.1). Then, similarly to Proposition 2.5.2, we deduce that

$$a_n \left(\hat{\alpha}_{\text{low}}^0(t_3, t_4)_n - \hat{\alpha}_{\text{low}}^0(t_1, t_2)_n \right) \quad \text{converges in law,}$$

where $a_n = \sqrt{n}$ if $H < 1 - 1/\alpha$ and $a_n = n^{1-1/(1+\alpha(1-H))}$ if $H > 1 - 1/\alpha$. Hence, we immediately conclude the convergence

$$d_n := -\frac{\log |\hat{\alpha}_{\text{low}}^0(t_3, t_4)_n - \hat{\alpha}_{\text{low}}^0(t_1, t_2)_n|}{\log(n)} \xrightarrow{\mathbb{P}} \begin{cases} 1/2 : & \text{if } H < 1 - 1/\alpha \\ 1 - 1/(1 + \alpha(1 - H)) : & \text{if } H > 1 - 1/\alpha \end{cases}$$

In other word, the statistic d_n helps us to identify the rate of convergence, but it has a bias of order $1/\log(n)$. Our decision rule is now as follows: Use Theorem 2.5.3(i) to perform statistical inference if

$$d_n > 1/2 - (\log(n))^{-1+\epsilon}$$

for some small chosen $\epsilon > 0$; otherwise use Theorem 2.5.3(ii).

Remark 2.5.4. *While we can obtain fully feasible asymptotic theory if we know whether $\alpha^{-1} \in \mathbb{N}$ or not, we are not yet able to deduce a complete statistical method without this a priori knowledge. Possibly subsampling procedures are required to obtain empirical confidence regions that automatically adapt to a given setting. \square*

High frequency setting

In the framework of high frequency observations the application of the empirical characteristic function might lead to suboptimal convergence rates for the estimator of (σ, α) . This comes from the following observation. Assume that $\alpha < 1$. Using the inequality $|\cos(x) - \cos(y)| \leq |x - y|^{\alpha'}$ for any $\alpha' < \alpha$ we obtain the upper bound

$$\begin{aligned} & |\varphi_{\text{high}}(t; \hat{H}_{\text{high}}(p, k)_n, k)_n - \varphi_{\text{high}}(t; H, k)_n| \\ & \leq \frac{t^{\alpha'} (n^{\hat{H}_{\text{high}}(p, k)_n - H} - 1)^{\alpha'}}{n} \sum_{i=k}^n |n^H \Delta_{i, k}^n X|^{\alpha'} = O_{\mathbb{P}} \left((n^{-1/2} \log n)^{-\alpha'/2} \right), \end{aligned}$$

where the last statement follows from $\mathbb{E}[|\Delta_{k, k} X|^{\alpha'}] < \infty$ and the ergodic theorem. Since the above expression is predominant in the asymptotic theory and it seems hard to improve it, we obtain slow rates of convergence for the parameters σ and α if we apply the same estimation procedure as in the previous section. For this reason we require a different approach in the high frequency setting.

First of all, we give an explicit formula for the constant $m_{-p, k} = \mathbb{E}[|\Delta_{k, k} X|^{-p}]$, $p \in (0, 1/2)$, which has been introduced in Theorem 2.3.1. We recall that the random variable $\Delta_{k, k} X$ is symmetric α -stable with scale parameter $\sigma \|h_k\|_{\alpha}$. Consequently, applying the identity [21, Eq. (18)] we conclude that

$$m_{-p, k} = \frac{(\sigma \|h_k\|_{\alpha})^{-p}}{a_{-p}} \int_{\mathbb{R}} \exp(-|y|^{\alpha}) |y|^{-1+p} dy = \frac{2(\sigma \|h_k\|_{\alpha})^{-p}}{\alpha a_{-p}} \Gamma(p/\alpha),$$

where the last equality follows by substitution $z = y^\alpha$ for $y > 0$. Now, we use the idea that has been originally proposed in [21] to identify the parameter α via power variation statistics. We consider $p, p' \in (0, 1/2)$, $p \neq p'$, and observe that

$$\frac{m_{-p',k}^p}{m_{-p,k}^{p'}} = \frac{(2/\alpha)^{p-p'} a_{-p}^{p'} \Gamma(p'/\alpha)^p}{a_{-p'}^p \Gamma(p/\alpha)^{p'}} =: \phi_{p,p'}(\alpha). \quad (2.5.4)$$

It has been shown in [21] that the mapping $\alpha \mapsto \phi_{p,p'}(\alpha)$ is invertible for any $p \neq p'$. Hence, we have $\alpha = \phi_{p,p'}^{-1}(m_{-p',k}^p/m_{-p,k}^{p'})$. Now, assuming that we know α and H (recall that the norm $\|h_k\|_\alpha$ depends on these parameters), we can recover the scale parameter σ via

$$\sigma = \left(\frac{\alpha a_{-p} m_{-p,k}}{2\Gamma(p/\alpha)} \right)^{-\frac{1}{p}} / \|h_k\|_\alpha.$$

Summarising the above identities we obtain the function $\bar{G} : (\mathbb{R}_+)^2 \times (0, 1) \rightarrow \mathbb{R}^2$ such that

$$(\sigma, \alpha) = \bar{G}(m_{-p,k}, m_{-p',k}, H). \quad (2.5.5)$$

Next, we follow the same two-stage routine as in the previous section. We first compute $\hat{H}_{\text{high}}(-p)_n = \hat{H}_{\text{high}}(-p, 1)_n$ with $p \in (0, 1/2)$ and define the preliminary estimator of α by

$$\hat{\alpha}_{\text{high}}^0(-p, -p')_n = \bar{G}_2 \left(V_{\text{high}}(f_{-p}, \hat{H}_{\text{high}}(-p)_n)_n, V_{\text{high}}(f_{-p'}, \hat{H}_{\text{high}}(-p)_n)_n, \hat{H}_{\text{high}}(-p)_n \right), \quad (2.5.6)$$

where the statistic $V_{\text{high}}(f_{-p}, \hat{H}_{\text{high}}(-p)_n)_n$ refers to power variation introduced in (2.3.7) with $k = 1$ and with H replaced by $\hat{H}_{\text{high}}(-p)_n$. Now, we define

$$\hat{k}_{\text{high}} = \hat{k}_{\text{high}}(-p, -p')_n = 2 + \lfloor \hat{\alpha}_{\text{high}}^0(-p, -p')_n^{-1} \rfloor \quad (2.5.7)$$

and introduce the estimator

$$\begin{aligned} & \left(\tilde{\sigma}_{\text{high}}(\hat{k}_{\text{high}}, -p, -p')_n, \tilde{\alpha}_{\text{high}}(\hat{k}_{\text{high}}, -p, -p')_n \right) \\ &= \bar{G} \left(V_{\text{high}}(f_{-p}, \hat{H}_{\text{high}}(-p, \hat{k}_{\text{high}})_n; \hat{k}_{\text{high}})_n, \right. \\ & \quad \left. V_{\text{high}}(f_{-p'}, \hat{H}_{\text{high}}(-p, \hat{k}_{\text{high}})_n; \hat{k}_{\text{high}})_n, \hat{H}_{\text{high}}(-p, \hat{k}_{\text{high}})_n \right). \end{aligned}$$

We again require a separate treatment of the cases $\alpha^{-1} \notin \mathbb{N}$ and $\alpha^{-1} \in \mathbb{N}$. We start with the first setting. When $H < k - 1/\alpha$ we consider the statistic $W(n)^{(1)} = (W(n)_1^{(1)}, W(n)_2^{(1)})$ associated with the power $-p$ and

$$k_1 = \hat{k}_{\text{high}}, r_1 = 1 \quad \text{and} \quad k_2 = \hat{k}_{\text{high}}, r_2 = 2.$$

Recall that $W(n)^{(1)} \xrightarrow{d} W^{(1)}$ according to Theorem 2.3.1. Now, similarly to Theorem 2.4.1, we define

$$\begin{aligned} \bar{B}_{\text{high}}^{\text{nor}}(-p, -p', k) &:= \nabla v_p(m_{-p,k}, 2^H m_{-p,k})(W^{(1)})^\star \\ &\times \left(\begin{array}{c} \nabla \bar{G}_1(m_{-p,k}, m_{-p',k}, H) (-pm_{-p,k}, -p'm_{-p',k}, H)^\star \\ \nabla \bar{G}_1(m_{-p,k}, m_{-p',k}, H) (-pm_{-p,k}, -p'm_{-p',k}, H)^\star \\ 1 \end{array} \right), \end{aligned} \quad (2.5.8)$$

where the function v_p has been introduced at (2.4.5). Our first result is the following theorem.

Theorem 2.5.5. *Let X be the linear fractional stable motion defined at (2.2.1). Assume that $p, p' \in (0, 1/2)$ and $\alpha^{-1} \notin \mathbb{N}$. Then we obtain the central limit theorem*

$$\begin{pmatrix} \sqrt{n}(\log n)^{-1} \left(\tilde{\sigma}_{high}(\hat{k}_{high}, -p, -p')_n - \sigma \right) \\ \sqrt{n}(\log n)^{-1} \left(\tilde{\alpha}_{high}(\hat{k}_{high}, -p, -p')_n - \alpha \right) \\ \sqrt{n} \left(\hat{H}_{high}(-p, \hat{k}_{high})_n - H \right) \end{pmatrix} \xrightarrow{d} \bar{B}_{high}^{nor} \left(-p, -p', 2 + \lfloor \alpha^{-1} \rfloor \right).$$

Next, we treat the case $\alpha^{-1} \in \mathbb{N}$. For this purpose, whenever $H > k - 1/\alpha$, we introduce the notation $\bar{B}_{high}^{sta}(-p, -p', k)$ to denote the random variable at (2.5.8) where $W^{(1)}$ is replaced by $S^{(1)}$. We deduce the following result, which is the analogue of Theorem 2.5.3.

Theorem 2.5.6. *Let X be the linear fractional stable motion defined at (2.2.1). Assume that $p, p' \in (0, 1/2)$ and $\alpha^{-1} \in \mathbb{N}$.*

(i) *(Case $H < 1 - 1/\alpha$) Assume that $H < 1 - 1/\alpha$. Then we obtain*

$$\begin{pmatrix} \sqrt{n}(\log n)^{-1} \left(\tilde{\sigma}_{high}(\hat{k}_{high}, -p, -p')_n - \sigma \right) \\ \sqrt{n}(\log n)^{-1} \left(\tilde{\alpha}_{high}(\hat{k}_{high}, -p, -p')_n - \alpha \right) \\ \sqrt{n} \left(\hat{H}_{high}(-p, \hat{k}_{high})_n - H \right) \end{pmatrix} \xrightarrow{d} D_{high}^{nor},$$

where the probability distribution D_{high}^{nor} on \mathbb{R}^3 is given by

$$\begin{aligned} D_{high}^{nor}(\cdot) = & \mathbb{P}(\{\bar{B}_{high}^{nor}(-p, -p', 2 + \alpha^{-1}) \in \cdot\} \cap \{\bar{B}_{high}^{nor}(-p, -p', 1)_2 < 0\}) \\ & + \mathbb{P}(\{\bar{B}_{high}^{nor}(-p, -p', 1 + \alpha^{-1}) \in \cdot\} \cap \{\bar{B}_{high}^{nor}(-p, -p', 1)_2 > 0\}). \end{aligned}$$

(ii) *(Case $H > 1 - 1/\alpha$) Assume that $H > 1 - 1/\alpha$. Then we obtain*

$$\begin{pmatrix} \sqrt{n}(\log n)^{-1} \left(\tilde{\sigma}_{high}(\hat{k}_{high}, -p, -p')_n - \sigma \right) \\ \sqrt{n}(\log n)^{-1} \left(\tilde{\alpha}_{high}(\hat{k}_{high}, -p, -p')_n - \alpha \right) \\ \sqrt{n} \left(\hat{H}_{high}(-p, \hat{k}_{high})_n - H \right) \end{pmatrix} \xrightarrow{d} D_{high}^{sta},$$

where the probability distribution D_{high}^{sta} on \mathbb{R}^3 is given by

$$\begin{aligned} D_{high}^{sta}(\cdot) = & \mathbb{P}(\bar{B}_{high}^{sta}(-p, -p', 1)_2 < 0) \mathbb{P}(\bar{B}_{high}^{nor}(-p, -p', 2 + \alpha^{-1}) \in \cdot) \\ & + \mathbb{P}(\bar{B}_{high}^{sta}(-p, -p', 1)_2 > 0) \mathbb{P}(\bar{B}_{high}^{nor}(-p, -p', 1 + \alpha^{-1}) \in \cdot). \end{aligned}$$

Remark 2.5.7. *We may use a similar decision rule as proposed in Section 2.5 to figure out whether part (i) or (ii) of Theorem 2.5.6 is applicable. Let $p_1, \dots, p_4 \in (0, 1/2)$ be distinct real numbers. As in the previous subsection we have that*

$$\begin{aligned} \bar{d}_n := & - \frac{\log \left| \hat{\alpha}_{high}^0(-p_1, -p_2)_n - \hat{\alpha}_{high}^0(-p_3, -p_4)_n \right|}{\log(n)} \xrightarrow{\mathbb{P}} \\ & \begin{cases} 1/2 : & \text{if } H < 1 - 1/\alpha \\ 1 - 1/(1 + \alpha(1 - H)) : & \text{if } H > 1 - 1/\alpha \end{cases} \end{aligned}$$

We thus use Theorem 2.5.6(i) to perform statistical inference when

$$\bar{d}_n > 1/2 - (\log(n))^{-1+\epsilon}.$$

□

2.6 A simulation study

In this section we demonstrate the finite sample performance of our estimators based upon the theoretical results of Theorems 2.4.1, 2.5.1 and 2.5.5, where the latter two correspond to the setting $\alpha^{-1} \notin \mathbb{N}$ (we dispense with the numerical analysis associated with Theorems 2.5.3 and 2.5.6). We simulate high and low frequency observations of the linear fractional stable motion defined at (2.2.1) for $n = 100, 1.000$ and 10.000 . Whenever we use the statistics $V_{\text{high}}(f; k, r)_n$ and $V_{\text{low}}(f; k, r)_n$ introduced in (2.3.7), we multiply them by $(n - rk + 1)/n$ to account for the actual number of summands. Throughout the section we set $t_1 = 1$ and $t_2 = 2$. We use 5000 repetitions to uncover the finite sample properties of our estimators. The asymptotic variances appearing in central limit theorems are rather hard to compute numerically, so we perform Monte Carlo simulations to estimate them. We generate the number of sample paths mentioned above and compute $(\hat{\sigma}, \hat{\alpha}, \hat{H})$ for each of them. Basing on the estimator values, we calculate sample mean and standard deviation, which are also used to construct empirical distribution functions, analogs of the functions on the right-hand side of the corresponding limit theorems.

Table 2.1: Bias/standard deviation of the estimators $(\hat{\sigma}_{\text{low}}, \hat{\alpha}_{\text{low}}, \hat{H}_{\text{low}})$ and $(\hat{\sigma}_{\text{high}}, \hat{\alpha}_{\text{high}}, \hat{H}_{\text{high}})$. We use $p = 0.4$ and $k = 2$, and the true parameter is $(\sigma, \alpha, H) = (0.3, 1.8, 0.8)$.

n	$\hat{\sigma}_{\text{low}}$	$\hat{\alpha}_{\text{low}}$	\hat{H}_{low}	$\hat{\sigma}_{\text{high}}$	$\hat{\alpha}_{\text{high}}$	\hat{H}_{high}
100	-0.024/0.06	-0.038/0.18	-0.05/0.12	0.06/0.18	-0.07/0.2	0.02/0.10
1000	-0.0008/0.02	0.012/0.068	-0.012/0.05	-0.001/0.12	0.015/0.07	-0.009/0.05
10000	0.00014/0.006	0.0005/0.022	-0.005/0.016	-0.010/0.05	0.001/0.022	-0.005/0.016

We begin with the discussion of Theorem 2.4.1. Table 2.1 reports the bias and the standard deviation of the estimator of $(\sigma, \alpha, H) = (0.3, 1.8, 0.8)$ in high and low frequency settings, where we use the power $p = 0.4$ and the order $k = 2$. We observe that our estimators exhibit a rather convincing finite sample performance in both settings. As expected from the theoretical statements of Theorem 2.4.1, the estimators of the self-similarity parameter H exhibit similar finite sample properties in high and low frequency settings, while the performance of the low frequency estimators for the parameters σ and α is better than in the high frequency case. This is obviously a consequence of a slightly slower convergence rate in the high frequency setting. Figure 2.2 plots the empirical densities of the standardised estimators from Theorem 2.4.1 in comparison to the density of the standard normal distribution. As mentioned earlier we use Monte Carlo simulations to estimate the theoretical variances. We again observe a very good performance of estimators of the parameter H , while the numerical results for the estimators of σ and α are better in the low frequency case.

Another approach to estimation of the self-similarity parameter H is the log-log regression, which is a generalisation of our approach. The key idea is the observation that the convergence $V_{\text{low}}(f_p; k, r) \xrightarrow{\text{a.s.}} r^p H m_{p,k}$ for $p \in (0, \alpha)$ or $p \in (-1, 0)$ leads to the

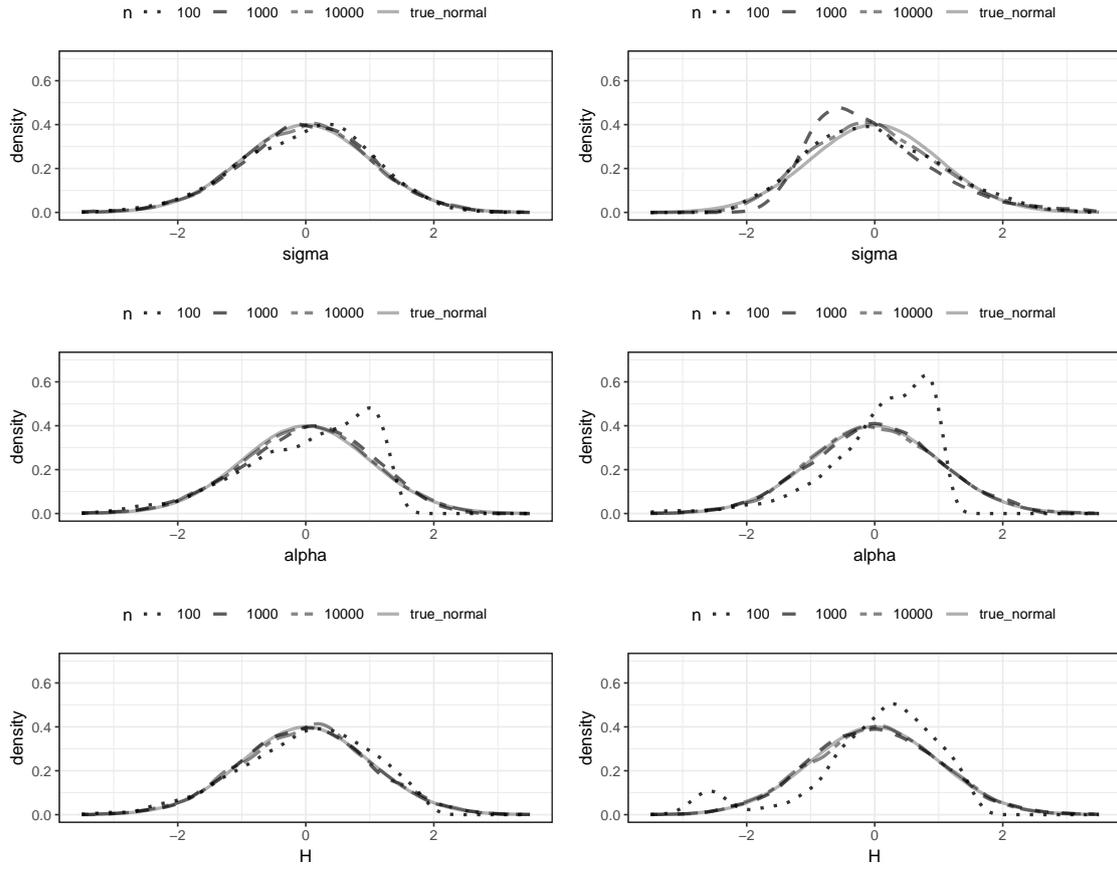


Figure 2.2: Empirical pdfs of $(\hat{\sigma}, \hat{\alpha}, \hat{H})$ in high and low frequency settings. The right column corresponds to the high frequency case and the left one to the low frequency case. The true parameter is $(\sigma, \alpha, H) = (0.3, 1.8, 0.8)$, $k = 2$, $p = 0.4$.

approximative identity

$$\log(V_{\text{low}}(f_p; k, r)_n) \approx \log(m_{p,k}) + pH \log(r), \quad r = 1, 2, \dots, \bar{r}.$$

Note that the latter is a linear regression and the slope identifies the parameter H . Indeed, H can be estimated from low frequency data via

$$\hat{H}_{\text{low}}^{\log} = \frac{\sum_{r=1}^{\bar{r}} (x_r - \bar{x})(y_r - \bar{y})}{p \sum_{r=1}^{\bar{r}} (x_r - \bar{x})^2},$$

where $x_r = \log(r)$, $y_r = \log(V_{\text{low}}(f_p; k, r)_n)$ and \bar{x} (resp. \bar{y}) denotes the empirical mean of x_r 's (resp. y_r 's). Obviously, the asymptotic theory for the estimator $\hat{H}_{\text{low}}^{\log}$ can be directly deduced from Theorem 2.3.2; we leave the details to the reader. Instead we restrict our attention to the empirical performance of $\hat{H}_{\text{low}}^{\log}$. The next table demonstrates the finite sample bias/standard deviation of the estimator $\hat{H}_{\text{low}}^{\log}$ in the setting of Theorem 2.4.1 with $p = 0.4$, $k = 2$ and $\bar{r} = \lfloor \log(n) \rfloor$.

Comparing Tables 2.1 and 2.2, we observe that the standard deviations of $\hat{H}_{\text{low}}^{\log}$ and \hat{H}_{low} are quite similar in all scenarios, but $\hat{H}_{\text{low}}^{\log}$ has a much lower bias.

Table 2.2: Bias/standard deviation of the regression-based estimator for H . Low frequency case. Here $p = -0.4$, $k = 2$ and $(\sigma, \alpha, H) = (0.3, 1.8, 0.8)$.

n	\bar{r}	$\widehat{H}_{\text{low}}^{\log}$
100	4	$-5.8 \times 10^{-3}/0.13$
1000	6	$-2.9 \times 10^{-4}/0.04$
10000	9	$-1.5 \times 10^{-4}/0.013$

Now, we turn our attention to the low frequency estimation discussed in Theorem 2.5.1. We use the power $p = -0.4$ and consider the true parameter $(\sigma, \alpha, H) = (0.3, 1.8, 0.8)$ and $(\sigma, \alpha, H) = (0.3, 0.8, 0.8)$. Observe that the first case corresponds to the setting of Theorem 2.4.1 and the second parameter corresponds to the discontinuous setting. The estimated order \widehat{k}_{low} is computed via (2.5.2). Table 2.3 displays the bias and standard deviation in the case $(\sigma, \alpha, H) = (0.3, 1.8, 0.8)$, while Table 2.4 demonstrates the numerical results in the case $(\sigma, \alpha, H) = (0.3, 0.8, 0.8)$.

Table 2.3: Bias/standard deviation of the estimator $(\widetilde{\sigma}_{\text{low}}, \widetilde{\alpha}_{\text{low}}, \widetilde{H}_{\text{low}})$. Here $p = -0.4$, \widehat{k}_{low} is computed from (2.5.2) and $(\sigma, \alpha, H) = (0.3, 1.8, 0.8)$.

n	$\widetilde{\sigma}_{\text{low}}$	$\widetilde{\alpha}_{\text{low}}$	$\widetilde{H}_{\text{low}}$
100	$-0.05/0.09$	$-0.031/0.18$	$-0.12/0.23$
1000	$-0.004/0.04$	$0.01/0.068$	$-0.018/0.12$
10000	$0.0003/0.015$	$0.001/0.022$	$-0.003/0.05$

Table 2.4: Bias/standard deviation of the estimator $(\widetilde{\sigma}_{\text{low}}, \widetilde{\alpha}_{\text{low}}, \widetilde{H}_{\text{low}})$. Here $p = -0.4$, \widehat{k}_{low} is computed from (2.5.2) and $(\sigma, \alpha, H) = (0.3, 0.8, 0.8)$.

n	$\widetilde{\sigma}_{\text{low}}$	$\widetilde{\alpha}_{\text{low}}$	$\widetilde{H}_{\text{low}}$
100	$-0.06/0.31$	$-0.003/0.41$	$-0.15/0.24$
1000	$-0.05/0.27$	$-0.08/0.31$	$0.003/0.13$
10000	$0.03/0.26$	$0.008/0.27$	$0.04/0.05$

Comparing the simulation results of Theorems 2.4.1 and 2.5.1, we see that the finite sample performance of estimators σ and H in Theorem 2.5.1 is inferior. This is not really surprising, since the methodology of Theorem 2.5.1 requires preliminary estimation of α and k , and hence leads to an accumulation of errors. In turn, alpha estimator is not as sensitive to errors because of the double logarithm. Furthermore, in the setting of a fractional Brownian motion it is well known that low values of the parameter k give more efficient estimators. We conjecture that a similar effect appears for linear fractional stable motions. This would explain the superiority of the results in Table 2.3 compared to those in Table 2.4, since $\lfloor \alpha^{-1} \rfloor + 2 = 2$ in the first setting while $\lfloor \alpha^{-1} \rfloor + 2 = 3$ in the second setting. Figures 2.3 and 2.4 show the empirical density functions, where the theoretical variances have been estimated via a Monte Carlo simulations. They confirm the better performance of the estimators in the continuous setting $(\sigma, \alpha, H) = (0.3, 1.8, 0.8)$. We also observe that the estimator of the parameter σ exhibits the worst finite sample properties in the setting $(\sigma, \alpha, H) = (0.3, 0.8, 0.8)$.

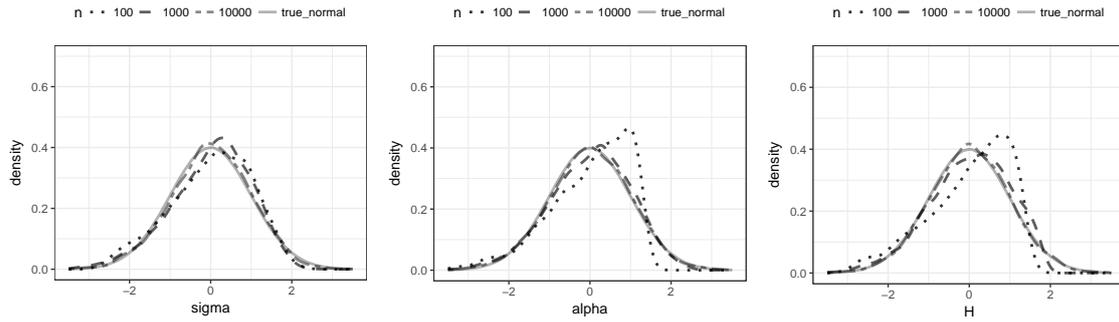


Figure 2.3: Empirical pdfs of $(\tilde{\sigma}_{10w}, \tilde{\alpha}_{10w}, \tilde{H}_{10w})$. Here $(\sigma, \alpha, H) = (0.3, 1.8, 0.8)$ and $p = -0.4$.

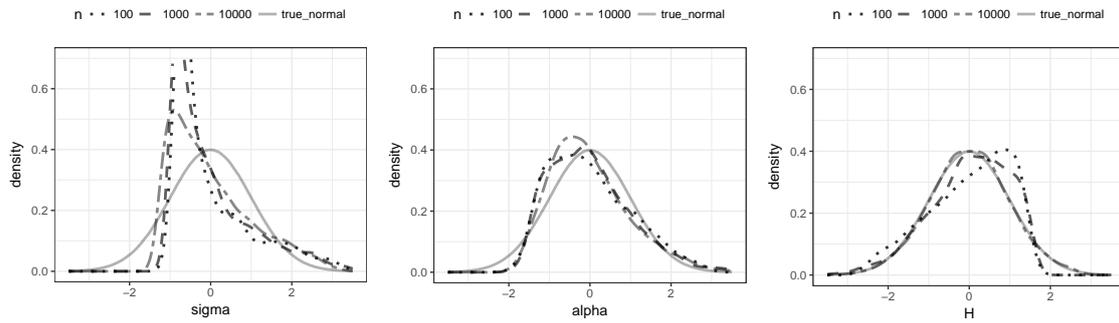


Figure 2.4: Empirical pdfs of $(\tilde{\sigma}_{10w}, \tilde{\alpha}_{10w}, \tilde{H}_{10w})$. Here $(\sigma, \alpha, H) = (0.3, 0.8, 0.8)$ and $p = -0.4$.

Finally, let us discuss the finite sample performance of the high frequency estimators from Theorem 2.5.5. We again consider two parameter settings $(\sigma, \alpha, H) = (0.3, 1.8, 0.8)$ and $(\sigma, \alpha, H) = (0.3, 0.8, 0.8)$, and we use $p = -0.4$ and $p' = -0.2$. The estimated order \hat{k}_{high} is computed via (2.5.7). Tables 2.5 and 2.6 display the biases and standard deviations in both parameter settings. We observe that the estimators of the parameter σ have the worst performance and we only obtain reasonable results for $n = 10.000$. Similar conclusions can be drawn from Figures 2.5 and 2.6 that plot the empirical density functions. The bad performance of the estimator of σ in Theorem 2.5.5 is explained by the fact that we not only require a preliminary estimation step for our procedure, but we also need to estimate the parameters H and α first to obtain an estimator of σ . This leads to accumulation of finite sample errors, which results in large bias and variance for small n . To further highlight this issue, we have plotted the empirical densities for the estimators of σ from Theorems 2.5.1 and 2.5.5 in Figure 2.7 in the setting $(\sigma, \alpha, H) = (0.3, 0.8, 0.8)$ where the parameter (α, H) is assumed to be known. We observe a much better finite sample performance, which confirms that the bad finite sample properties of the estimator of σ are largely due to preliminary estimation of (α, H) .

Acknowledgment

The authors acknowledge financial support from the project ‘‘Ambit fields: probabilistic properties and statistical inference’’ funded by Villum Fonden and from CREATES funded by the Danish National Research Foundation.

Table 2.5: Bias/standard deviation of the estimator $(\tilde{\sigma}_{\text{high}}, \tilde{\alpha}_{\text{high}}, \tilde{H}_{\text{high}})$. Here $p = -0.4, p' = -0.2$ and $(\sigma, \alpha, H) = (0.3, 1.8, 0.8)$.

n	$\tilde{\sigma}_{\text{high}}$	$\tilde{\alpha}_{\text{high}}$	\tilde{H}_{high}
100	60/1443	-0.02/0.77	0.23/0.33
1000	0.18/0.82	0.19/0.67	0.02/0.13
10000	-0.003/0.17	0.052/0.26	-0.003/0.05

Table 2.6: Bias/standard deviation of the estimator $(\tilde{\sigma}_{\text{high}}, \tilde{\alpha}_{\text{high}}, \tilde{H}_{\text{high}})$. Here $p = -0.4, p' = -0.2$ and $(\sigma, \alpha, H) = (0.3, 0.8, 0.8)$.

n	$\tilde{\sigma}_{\text{high}}$	$\tilde{\alpha}_{\text{high}}$	\tilde{H}_{high}
100	16/341	0.19/0.37	0.13/0.4
1000	0.103/1	0.02/0.09	0.06/0.16
10000	-0.11/0.12	0.003/0.04	0.04/0.06

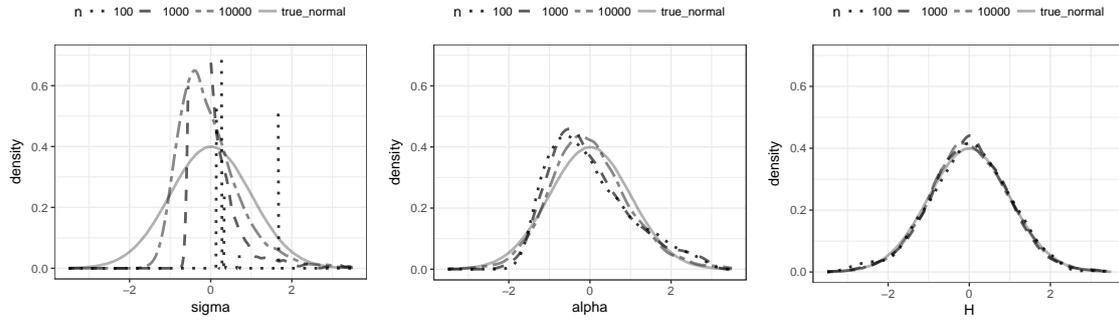


Figure 2.5: Empirical pdfs of $(\tilde{\sigma}_{\text{high}}, \tilde{\alpha}_{\text{high}}, \tilde{H}_{\text{high}})$. Here $p = -0.4, p' = -0.2$ and $(\sigma, \alpha, H) = (0.3, 1.8, 0.8)$.

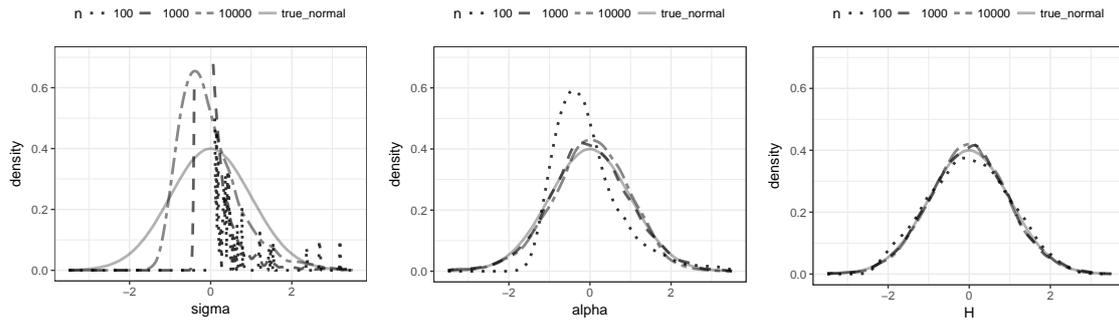


Figure 2.6: Empirical pdfs of $(\tilde{\sigma}_{\text{high}}, \tilde{\alpha}_{\text{high}}, \tilde{H}_{\text{high}})$. Here $p = -0.4, p' = -0.2$ and $(\sigma, \alpha, H) = (0.3, 0.8, 0.8)$.

2.7 Proofs

In this section we denote all positive constants by C although they may change from line to line.

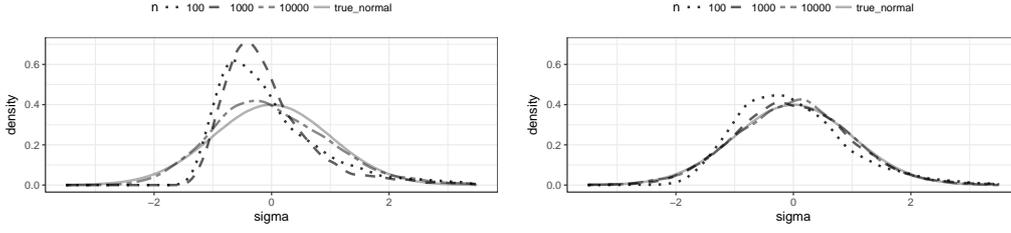


Figure 2.7: Empirical pdfs for σ (Left=Theorem 2.5.1, Right=Theorem 2.5.5) when the parameter $(\alpha, H) = (0.8, 0.8)$ is known. Here $\sigma = 0.3$, $p = -0.4$, $p' = -0.2$ and $k = 3$.

Preliminaries

Here we will show some technical results, which are necessary to prove the main theorems. We start with the following lemma that is a straightforward consequence of Taylor expansion.

Lemma 2.7.1. *Let $h_{k,r}$ be defined as in (2.3.10). Then it holds that*

$$|h_{k,r}(x)| \leq C \left(x^{H-1/\alpha} \mathbf{1}_{(0, rk+1]}(x) + x^{H-k-1/\alpha} \mathbf{1}_{(rk+1, \infty)}(x) \right).$$

Furthermore, the function $|h_{k,r}|$ is strictly decreasing on $(rk+1, \infty)$.

An important quantity when considering various asymptotic covariances is the following object:

$$\rho_l := \int_0^\infty |h_{k,r}(x)h_{k,r}(x+l)|^{\alpha/2} dx. \quad (2.7.1)$$

The next lemma determines the asymptotic behaviour of ρ_l when $l \rightarrow \infty$.

Lemma 2.7.2. *For $l > rk$ it holds that*

$$\rho_l \leq C \begin{cases} l^{(\alpha(H-k)-1)/2} & \text{when } k > H + 1/\alpha \\ l^{\alpha(H-k)} & \text{when } k < H + 1/\alpha \end{cases}$$

Proof. Assume that $l > rk$. Applying Lemma 2.7.1 we obtain the inequality

$$\begin{aligned} \int_0^l |h_{k,r}(x)h_{k,r}(x+l)|^{\alpha/2} dx &\leq C l^{(\alpha(H-k)-1)/2} \int_0^l |h_{k,r}(x)|^{\alpha/2} dx \\ &\leq C \begin{cases} l^{(\alpha(H-k)-1)/2} \int_0^\infty |h_{k,r}(x)|^{\alpha/2} dx & k > H + 1/\alpha \\ l^{\alpha(H-k)} & k < H + 1/\alpha \end{cases} \end{aligned}$$

When $k > H + 1/\alpha$ we have $\int_0^\infty |h_{k,r}(x)|^{\alpha/2} dx < \infty$, which is due to Lemma 2.7.1; on the other hand, for $k < H + 1/\alpha$ we deduce that $\int_0^l |h_{k,r}(x)|^{\alpha/2} dx \leq C l^{1+(\alpha(H-k)-1)/2}$. Applying Lemma 2.7.1 once again and using the substitution $x = ly$ we deduce the inequality

$$\begin{aligned} \int_l^\infty |h_{k,r}(x)h_{k,r}(x+l)|^{\alpha/2} dx &\leq C \int_l^\infty |x(x+l)|^{(\alpha(H-k)-1)/2} dx \\ &= C l^{\alpha(H-k)} \int_1^\infty |y(y+1)|^{(\alpha(H-k)-1)/2} dy. \end{aligned}$$

Indeed, the last integral is finite since $H < 1 \leq k$. Hence, the statement of Lemma 2.7.2 is proved. \square

In the next step we will determine the behaviour of the function $U_{g,h}$ defined at (2.3.3). The following result is the statement of inequalities (3.4)-(3.6) from [32].

Lemma 2.7.3. *For any $u, v \in \mathbb{R}$ it holds that*

$$\begin{aligned} |U_{g,h}(u, v)| &\leq 2|uv|^{\alpha/2} \int_0^\infty |g(x)h(x)|^{\alpha/2} dx \\ &\quad \times \exp\left(-2|uv|^{\alpha/2} \left(\|g\|_\alpha^{\alpha/2} \|h\|_\alpha^{\alpha/2} - \int_0^\infty |g(x)h(x)|^{\alpha/2} dx\right)\right), \\ |U_{g,h}(u, v)| &\leq 2|uv|^{\alpha/2} \int_0^\infty |g(x)h(x)|^{\alpha/2} dx \\ &\quad \times \exp\left(-\left(\|ug\|_\alpha^{\alpha/2} - \|vh\|_\alpha^{\alpha/2}\right)^2\right). \end{aligned}$$

In particular, we have that $|U_{g,h}(u, v)| \leq 2|uv|^{\alpha/2} \int_0^\infty |g(x)h(x)|^{\alpha/2} dx$.

Now we turn our attention to formula (2.3.19), which presents an explicit expression for the asymptotic covariance matrix $\text{cov}(W)$. In the following we will prove this identity. For the sake of brevity we will only show formula (2.3.19) for $d = 1$ and only for the component $\text{var}(W_1^{(1)})$ with $k_1 = k$ and $r_1 = r$. All other identities are in fact easier to prove and we leave them to the reader.

The expression for $\text{var}(W_1^{(1)})$ for $p \in (-1/2, 0)$ and its finiteness have been shown in [21, Corollary 3.3 and Theorem 4.2] using methods from distribution theory, so we concentrate on the case $p \in (0, 1/2)$. For $p \in (0, 1)$, we have the relationship

$$|x|^p = a_p^{-1} \int_{\mathbb{R}} (1 - \exp(ixy)) |y|^{-1-p} dy. \quad (2.7.2)$$

which can be shown by substitution $xy = z$ (recall the definition of a_p at (2.3.8)). Note that similarly to (2.3.4) the latter connects power functions with characteristic functions, which are explicit in the α -stable case. Applying this formula and using stationarity of the increments $\Delta_{i,k}^r X$ we conclude that

$$\text{cov}\left(f_p\left(\Delta_{i,k}^r X\right), f_p\left(\Delta_{i+l,k}^r X\right)\right) = \theta(h_{k,r}, h_{k,r}(\cdot + l))_p,$$

where the quantity $\theta(g, h)_p$ has been introduced at (2.3.9). Since $W(n)_1^{(1)}$ is a sum of stationary random variables it remains to prove that $\theta(h_{k,r}, h_{k,r}(\cdot + l))_p$ is absolutely summable in l to show the identity (2.3.19). This is the statement of the next lemma.

Lemma 2.7.4. *For $p \in (0, 1/2)$ with $p < \alpha/2$ it holds that*

$$|\theta(h_{k,r}, h_{k,r}(\cdot + l))_p| \leq C\rho_l.$$

In particular, if $k > H + 1/\alpha$ we obtain $\sum_{l=1}^\infty |\theta(h_{k,r}, h_{k,r}(\cdot + l))_p| < \infty$.

Proof. The second part of the statement follows directly from Lemma 2.7.2 and the fact that $(\alpha(H - k) - 1)/2 < -1$ when $k > H + 1/\alpha$. To show the first part of the statement we will use the inequalities of Lemma 2.7.3. Recalling the definition of $\theta(h_{k,r}, h_{k,r}(\cdot + l))_p$ it is sufficient to compute the double integral over the set $(0, \infty)^2$ (instead of \mathbb{R}^2), which

is due to symmetry. The domain $(0, \infty)^2$ is further decomposed into the regions $(0, 1)^2$, $(0, 1) \times [1, \infty)$, $[1, \infty) \times (0, 1)$ and $[1, \infty)^2$, and we denote the corresponding integrals by I_1, I_2, I_3 and I_4 , respectively.

For the integral I_1 we use the inequality $|U_{g,h}(u, v)| \leq 2|uv|^{\alpha/2} \int_0^\infty |g(x)h(x)|^{\alpha/2} dx$ of Lemma 2.7.3 to deduce that

$$|I_1| \leq a_p^{-2} \int_{(0,1)^2} (xy)^{-1-p} |U_{h_{k,r}, h_{k,r}(\cdot+l)}(x, y)| dx dy \leq C \rho_l \int_{(0,1)^2} (xy)^{-1-p+\alpha/2} dx dy,$$

where the last integral is finite because $p < \alpha/2$. Applying the main statement of Lemma 2.7.3 we also conclude the inequality

$$|I_4| \leq C \rho_l \int_{[1,\infty)^2} (xy)^{-1-p+\alpha/2} \exp\left(-2(xy)^{\alpha/2} (\|h_{k,r}\|_\alpha^\alpha - \rho_l)\right) dx dy,$$

By Cauchy-Schwarz inequality we have that $\rho_l < \|h_{k,r}\|_\alpha^\alpha$. Furthermore, $\lim_{l \rightarrow \infty} \rho_l = 0$ by Lemma 2.7.2 and thus, for a given $\epsilon \in (0, 1)$, $\rho_l < \epsilon$ for almost all $l \in \mathbb{N}$. Hence, there exists a constant $C > 0$ such that

$$|I_4| \leq C \rho_l \int_{[1,\infty)^2} (xy)^{-1-p+\alpha/2} \exp\left(-2C(xy)^{\alpha/2}\right) dx dy,$$

where the latter integral is obviously finite. For the integral I_2 we apply Lemma 2.7.3 once more to obtain

$$\begin{aligned} |I_2| &\leq C \rho_l \int_{(0,1) \times [1,\infty)} (xy)^{-1-p+\alpha/2} \exp\left(-\|h_{k,r}\|_\alpha^{\alpha/2} (y^{\alpha/2} - x^{\alpha/2})^2\right) dx dy \\ &\leq C \rho_l \int_{(0,1) \times [1,\infty)} (xy)^{-1-p+\alpha/2} \exp\left(-\|h_{k,r}\|_\alpha^{\alpha/2} (y^{\alpha/2} - 1)^2\right) dx dy \end{aligned}$$

and the last integral is again finite since $p < \alpha/2$. The term I_3 is treated exactly the same way as I_2 and we are done. \square

At the end of this subsection we remark that the covariance matrix $\text{cov}(W)$ is a continuous function in $(\sigma, \alpha, H) \in \mathbb{R}_+ \times (0, 2) \times (0, 1)$, which follows by Lemma 2.7.4 and a dominated convergence theorem.

Proof of Theorem 2.3.2

The proof of Theorem 2.3.2 will be divided into several steps. Some parts of the proof will rely upon asymptotic expansions investigated in [7, 31].

Asymptotic decomposition of the statistic $(W(n)^{(1)}, W(n)^{(2)})$

In this section we introduce several approximations of the statistic appearing in Theorem 2.3.2. We start with the asymptotically normal part $(W(n)^{(1)}, W(n)^{(2)})$. Recalling the notation (2.3.10) we observe the identity

$$\Delta_{i,k}^r X = \int_{\mathbb{R}} h_{k,r}(i-s) dL_s. \quad (2.7.3)$$

In the first step we introduce the short memory approximation of $\Delta_{i,k}^r X$ by truncating the integration region:

$$\Delta_{i,k}^r X(m) := \int_{i-m}^{i+m} h_{k,r}(i-s) dL_s. \quad (2.7.4)$$

Note that the random variables $(\Delta_{i,k}^r X(m))_{i \geq rk}$ are stationary and $2m$ -dependent, i.e. $\Delta_{i,k}^r X(m)$ and $\Delta_{j,k}^r X(m)$ are independent if $|i-j| \geq 2m$. For $f(x) = |x|^p$ with $p \in (0, 1/2)$ and $p < \alpha/2$, or $f(x) = \cos(tx)$ we introduce the notation

$$W(n, m)_j^{(1)} := \frac{1}{\sqrt{n}} \sum_{i=r_j k_j}^n \left\{ f_p \left(\Delta_{i,k_j}^{r_j} X(m) \right) - \mathbb{E} \left[f_p \left(\Delta_{i,k_j}^{r_j} X(m) \right) \right] \right\} \quad (2.7.5)$$

$$W(n, m)_j^{(2)} := \frac{1}{\sqrt{n}} \sum_{i=k}^n \left\{ \psi_{t_j} \left(\Delta_{i,k_j} X(m) \right) - \mathbb{E} \left[\psi_{t_j} \left(\Delta_{i,k_j} X(m) \right) \right] \right\}$$

For the function $f_{-p}(x) = |x|^{-p}$ with $p \in (0, 1/2)$ we set $f_{-p}^\epsilon(x) = |x|^{-p} 1_{\{|x| > \epsilon\}}$ and note that the latter is a bounded function. In this setting we define

$$W(n, m, \epsilon)_j^{(1)} := \frac{1}{\sqrt{n}} \sum_{i=r_k}^n \left\{ f_{-p}^\epsilon \left(\Delta_{i,k}^r X(m) \right) - \mathbb{E} \left[f_{-p}^\epsilon \left(\Delta_{i,k}^r X(m) \right) \right] \right\}. \quad (2.7.6)$$

In [7, Section 5.4] it has been shown that the convergence

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} \left[\left(W(n, m)_j^{(1)} - W(n)_j^{(1)} \right)^2 \right] = 0 \quad (2.7.7)$$

holds. On the other hand, since the functions ψ_{t_j} and f_{-p}^ϵ are bounded, we obtain the convergence

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} \left[\left(W(n, m)_j^{(2)} - W(n)_j^{(2)} \right)^2 \right] = 0, \quad (2.7.8)$$

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} \left[\left(W(n, m, \epsilon)_j^{(1)} - W(n, \epsilon)_j^{(1)} \right)^2 \right] = 0$$

from [31]. Here $W(n, \epsilon)_j^{(1)}$ is the original statistic defined at (2.3.13) associated with the function f_{-p}^ϵ .

Asymptotic decomposition of the statistic $(S(n)^{(1)}, S(n)^{(2)})$

In this subsection we derive an asymptotic expansion for the statistic $(S(n)^{(1)}, S(n)^{(2)})$. The main ideas originate from the work [7] and we will adapt their principles to our setting. The following estimates and decomposition have been treated in the case of power variation with $p \in (0, 1/2)$, $p < \alpha/2$, in [7], so we will rather concentrate on the functions f_{-p} , $p \in (0, 1/2)$, and ψ_t .

All expansions are valid componentwise, so we may assume that $d = 1$. We recall the notation introduced at (2.3.14). For a symmetric α -stable random variable Y with scaling parameter $\rho > 0$ and a measurable function $f : \mathbb{R} \rightarrow \mathbb{R}$, we introduce the function

$$\Phi_\rho(f)(x) := \mathbb{E}[f(Y+x)] - \mathbb{E}[f(Y)], \quad x \in \mathbb{R}, \quad (2.7.9)$$

whenever the latter is finite. In the following we will derive various estimates for $\Phi_\rho(f_{-p})(x)$ with $p \in (0, 1/2)$. First of all, using the identity [21, Eq. (18)] we obtain the representation

$$\Phi_\rho(f_{-p})(x) = a_p^{-1} \int_{\mathbb{R}} (1 - \cos(xy)) \exp(-|\rho y|^\alpha) |y|^{-1+p} dy. \quad (2.7.10)$$

This identity implies the following result.

Lemma 2.7.5. *Assume that $\rho, \rho_1, \rho_2 > \epsilon > 0$. Then there exists a constant $C_\epsilon > 0$ such that the following inequalities hold:*

$$\begin{aligned} |\Phi_\rho(f_{-p})(x)| &\leq C_\epsilon(1 \wedge x^2), & |\Phi_\rho(f_{-p})^{(v)}(x)| &\leq C_\epsilon \quad \text{for } v = 1, 2, \\ |\Phi_\rho(f_{-p})(x) - \Phi_\rho(f_{-p})(y)| &\leq C_\epsilon \left((1 \wedge |x| + 1 \wedge |y|) |x - y| 1_{\{|x-y| \leq 1\}} + 1_{\{|x-y| > 1\}} \right), \\ |\Phi_{\rho_1}(f_{-p})(x) - \Phi_{\rho_2}(f_{-p})(x)| &\leq C_\epsilon |\rho_2^\alpha - \rho_1^\alpha|, \\ \int_0^x \int_0^y \Phi_\rho(f_{-p})(a + z + w) dz dw &\leq C_\epsilon(1 \wedge x)(1 \wedge y) \quad \text{for any } x, y > 0, a \in \mathbb{R}, \end{aligned}$$

where $\Phi_\rho(f_{-p})^{(v)}$ denotes the v th derivative of $\Phi_\rho(f_{-p})$.

Proof. Note that the function f_{-p} is even and hence $\Phi_\rho(f_{-p})(0) = \Phi_\rho(f_{-p})^{(1)}(0) = 0$. Using the identity (2.7.10) we immediately see that $|\Phi_\rho(f_{-p})^{(v)}(x)| \leq C_\epsilon$ for $v = 0, 1, 2$. Thus, we obtain the first two inequalities. By the same arguments we get $|\Phi_\rho(f_{-p})^{(1)}(x)| \leq C_\epsilon(1 \wedge |x|)$. Observing the identity

$$|\Phi_\rho(f_{-p})(x) - \Phi_\rho(f_{-p})(y)| = \left| \int_y^x \Phi_\rho(f_{-p})^{(1)}(u) du \right|$$

we readily deduce the third inequality. The fourth inequality follows immediately from (2.7.10) and the mean value theorem. The last statement is a straightforward consequence of the first three inequalities of Lemma 2.7.5. \square

It is important to note that the result of Lemma 2.7.5 remains valid for the function $\Phi_\rho(\psi_t)$. In this case it is a consequence of the fact the ψ_t is a bounded and even function.

In the next step we present some decompositions, which have been investigated in [7]. For any fixed r and k , and the function $f = f_p, f_{-p}$, $p \in (0, 1/2)$, or ψ_t , we define the random variable

$$S(f)_n = n^{-1/(1+\alpha(k-H))} \sum_{i=rk}^n \left\{ f \left(\Delta_{i,k}^r X \right) - \mathbb{E} \left[f \left(\Delta_{i,k}^r X \right) \right] \right\} =: \sum_{i=rk}^n V_i^n.$$

We also introduce the σ -algebras

$$\mathcal{G}_s := \sigma(L_v - L_u : v, u \leq s), \quad \mathcal{G}_s^1 := \sigma(L_v - L_u : s \leq v, u \leq s + 1),$$

and note that $(\mathcal{G}_s^1)_{s \in \mathbb{R}}$ is not a filtration. Now, we introduce the notation

$$R_i^n := \sum_{j=1}^n \zeta_{i,j}^n, \quad Q_i^n := \sum_{j=1}^n \mathbb{E}[V_i^n | \mathcal{G}_{i-j}^1],$$

$$\text{where } \zeta_{i,j}^n := \mathbb{E}[V_i^n | \mathcal{G}_{i-j+1}] - \mathbb{E}[V_i^n | \mathcal{G}_{i-j}] - \mathbb{E}[V_i^n | \mathcal{G}_{i-j}^1].$$

Finally, we observe the decomposition

$$S(f)_n = \sum_{i=rk}^n R_i^n + \left(-\bar{S}(f)_n + \sum_{i=rk}^n Q_i^n \right) + \bar{S}(f)_n, \quad (2.7.11)$$

$$\bar{S}(f)_n := n^{-1/(1+\alpha(k-H))} \sum_{i=rk}^n \left\{ \bar{\Phi}(f)(L_i - L_{i-1}) - \mathbb{E}[\bar{\Phi}(f)(L_i - L_{i-1})] \right\},$$

where $\bar{\Phi}(f)(x) := \sum_{j=1}^{\infty} \Phi_{\rho}(f)(h_{k,r}(j)x)$ with $\rho = \sigma \|h_{k,r}\|_{\alpha}$. Note that $\bar{S}(f)_n$ is a sum of i.i.d random variables. For $f = f_p$ with $p \in (0, 1/2)$, $p < \alpha/2$ and under assumptions of Theorem 2.3.2, the convergence

$$\sum_{i=rk}^n R_i^n \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad -\bar{S}(f)_n + \sum_{i=rk}^n Q_i^n \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \rightarrow \infty \quad (2.7.12)$$

has been shown in [7] (cf. eqs. (5.30), (5.31) and (5.38) therein). The proof of these convergence results follows from a number of estimates on the function $\Phi_{\rho}(f_p)$, $p \in (0, 1/2)$, which are stated in [7, eqs. (5.14)-(5.18) and Lemma 5.8]. But according to Lemma 2.7.5 the same estimates hold also for $\Phi_{\rho}(f_{-p})$, $p \in (0, 1/2)$, and $\Phi_{\rho}(\psi_t)$ (in fact, the latter estimates are stronger). Consequently, the convergence at (2.7.12) also holds for the cases $f = f_{-p}$ and $f = \psi_t$ and we deduce that

$$S(f)_n - \bar{S}(f)_n \xrightarrow{\mathbb{P}} 0 \quad \text{for } f = f_p, f_{-p} \text{ or } \psi_t. \quad (2.7.13)$$

A limit theorem for the approximations

Recalling the notation introduced in (2.3.14) and (2.3.16) we obtain the identities

$$\begin{aligned} \bar{S}(n)_j^{(1)} &:= \bar{S}(f_p)_{n,j} = n^{-1/(1+\alpha(k-H))} \sum_{i=r_j k}^n \left\{ \bar{\Phi}_j^{(1)}(L_i - L_{i-1}) - \mathbb{E}[\bar{\Phi}_j^{(1)}(L_i - L_{i-1})] \right\}, \\ \bar{S}(n)_j^{(2)} &:= \bar{S}(\psi_{t_j})_n = n^{-1/(1+\alpha(k-H))} \sum_{i=r_j k}^n \left\{ \bar{\Phi}_j^{(2)}(L_i - L_{i-1}) - \mathbb{E}[\bar{\Phi}_j^{(2)}(L_i - L_{i-1})] \right\}, \end{aligned} \quad (2.7.14)$$

where $p \in (-1/2, 1/2) \setminus \{0\}$ and the statistic $\bar{S}(f_p)_{n,j}$ is defined as in (2.7.11) using the parameters r_j and k . As a consequence of (2.7.7), (2.7.8) and (2.7.13) it is now sufficient to show a weak limit theorem for the statistic

$$(W(n, m)^{(1)}, W(n, m)^{(2)}, \bar{S}(n)^{(1)}, \bar{S}(n)^{(2)})$$

(resp. $(W(n, m, \epsilon)^{(1)}, W(n, m)^{(2)}, \bar{S}(n)^{(1)}, \bar{S}(n)^{(2)})$) when $p \in (0, 1/2)$ and $p < \alpha/2$ (resp. $-p \in (0, 1/2)$) as $n \rightarrow \infty$ and then $m \rightarrow \infty$.

In order to prove this convergence we recall the results of [35] adapted to our setting. Let $(Y_i^{(1)})_{i \geq 1}$ and $(Y_i^{(2)})_{i \geq 1}$ be i.i.d sequences of centred random variables of dimensions d_1 and d_2 respectively, which are not necessarily independent. Define the statistics

$$Z_n^{(1)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i^{(1)}, \quad Z_n^{(2)} = n^{-1/\beta} \sum_{i=1}^n Y_i^{(2)} \quad \text{with } \beta \in (1, 2).$$

Assume now that $Z_n^{(1)} \xrightarrow{d} Z^{(1)}$ where $Z^{(1)}$ is a d_1 -dimensional centred normal distribution and assume that each coordinate $Y_{1,j}^{(2)}$, $1 \leq j \leq d_2$, is in the domain of attraction of a β -stable random variables, i.e.

$$\lim_{x \rightarrow +\infty} x^\beta \mathbb{P}(Y_{1,j}^{(2)} > x) = b_j^+ \quad \text{and} \quad \lim_{x \rightarrow -\infty} |x|^\beta \mathbb{P}(Y_{1,j}^{(2)} < x) = b_j^-.$$

Assume moreover that there exists a measure $\bar{\nu}$ such that for all sets $A \in \mathcal{B}(\mathbb{R}^{d_2})$ bounded away from 0 with $\bar{\nu}(\partial A) = 0$ it holds:

$$\lim_{n \rightarrow \infty} n \mathbb{P}(n^{-1/\beta} Y_i^{(2)} \in A) = \bar{\nu}(A).$$

Then we obtain the joint convergence

$$\left(Z_n^{(1)}, Z_n^{(2)} \right) \xrightarrow{d} \left(Z^{(1)}, Z^{(2)} \right), \quad (2.7.15)$$

where $Z^{(1)}$ and $Z^{(2)}$ are necessarily independent, and the law of $Z^{(2)}$ is determined by the Lévy measure $\bar{\nu}$. Indeed this result is a direct consequence of [35, Theorems 3 and 4] and their direct extension from bivariate to $(d_1 + d_2)$ -dimensional setting.

Next, we apply the weak convergence at (2.7.15) to our framework. Notice first that the statistics $W(n, m)^{(1)}$, $W(n, m)^{(2)}$ and $W(n, m, \epsilon)^{(1)}$ are sums of $2m$ -dependent random variables, but this setting can be reduced to sums of i.i.d random variables by the classical Bernstein's blocking technique. Hence, the theory of [35] also applies in this case.

For the sake of brevity we apply the convergence at (2.7.15) only for the statistic $(W(n, m)^{(1)}, W(n, m)^{(2)}, \bar{S}(n)^{(1)}, \bar{S}(n)^{(2)})$. We set

$$Z_{n,m}^{(1)} = \left(W(n, m)^{(1)}, W(n, m)^{(2)} \right) \quad \text{and} \quad Z_n^{(2)} = \left(\bar{S}(n)^{(1)}, \bar{S}(n)^{(2)} \right),$$

and define $\beta = (1 + \alpha(k - H))$. By the standard central limit theorem for sums of stationary $2m$ -dependent random variables we deduce the convergence

$$Z_{n,m}^{(1)} \xrightarrow{d} Z_m^{(1)} \sim \mathcal{N}_{2d}(0, \Sigma_m) \quad \text{as } n \rightarrow \infty,$$

where the asymptotic covariance matrix Σ_m is defined by

$$\begin{aligned} \Sigma_m^{ij} &= \sum_{l=-2m+1}^{2m-1} \text{cov} \left(f_p \left(\Delta_{r_i k_i, k_i}^{r_i} X(m) \right), f_p \left(\Delta_{r_i k_i + l, k_j}^{r_j} X(m) \right) \right), \quad 1 \leq i, j \leq d, \\ \Sigma_m^{ij} &= \sum_{l=-2m+1}^{2m-1} \text{cov} \left(f_p \left(\Delta_{r_i k_i, k_i}^{r_i} X(m) \right), \psi_{t_j} \left(\Delta_{r_i k_i + l, k_j} X(m) \right) \right), \quad d+1 \leq i+d, j \leq 2d, \\ \Sigma_m^{ij} &= \sum_{l=-2m+1}^{2m-1} \text{cov} \left(\psi_{t_i} \left(\Delta_{k_i, k_i} X(m) \right), \psi_{t_j} \left(\Delta_{k_i + l, k_j} X(m) \right) \right), \quad d+1 \leq i, j \leq 2d. \end{aligned}$$

In the next step we treat the statistic $Z_n^{(2)}$. Recalling the definition at (2.7.14), and the tail convergence of (2.3.20) and (2.3.21), we conclude that the limits of $\bar{S}(n)^{(1)}$ and $\bar{S}(n)^{(2)}$ must be independent since $b_j^- = 0$ for $1 \leq j \leq d$ and $b_j^+ = 0$ for $d+1 \leq j \leq 2d$. Furthermore, (2.3.22) readily implies the convergence

$$Z_n^{(2)} \xrightarrow{d} \left(S^{(1)}, S^{(2)} \right) \quad \text{as } n \rightarrow \infty,$$

where the vector $(S^{(1)}, S^{(2)})$ has been introduced in Theorem 2.3.2.

Finally, we will prove that the covariance matrix Σ_m converges as $m \rightarrow \infty$. In the following we write $\|Y\|_{\mathbb{L}^2}$ for $\mathbb{E}[Y^2]^{1/2}$ for any square integrable random variable Y . For $m_1, m_2 \in \mathbb{N}$ and $1 \leq j \leq d$ observe the decomposition

$$\begin{aligned} |(\Sigma_{m_1}^{jj})^{1/2} - (\Sigma_{m_2}^{jj})^{1/2}| &= \lim_{n \rightarrow \infty} \left| \|W(n, m_1)_j^{(1)}\|_{\mathbb{L}^2} - \|W(n, m_2)_j^{(1)}\|_{\mathbb{L}^2} \right| \\ &\leq \limsup_{n \rightarrow \infty} \left(\|W(n, m_1)_j^{(1)} - W(n)_j^{(1)}\|_{\mathbb{L}^2} + \|W(n, m_2)_j^{(1)} - W(n)_j^{(1)}\|_{\mathbb{L}^2} \right) \end{aligned} \quad (2.7.16)$$

and the latter converges to 0 as $m_1, m_2 \rightarrow \infty$ due to (2.7.7). Hence, $(\Sigma_m^{jj})_{m \geq 1}$ is a Cauchy sequence and thus it converges. Since $\text{var}(W(n)_j^{(1)}) \rightarrow \text{var}(W_j^{(1)})$ we must have that

$$\lim_{m \rightarrow \infty} \Sigma_m^{jj} = \text{var}(W_j^{(1)}).$$

The same argument applies to Σ_m^{jj} for $d+1 \leq j \leq 2d$ and also to covariances Σ_m^{ij} due to polarisation identity.

Summarising the results of Sections 2.7-2.7 we obtain the weak limit theorem

$$(W(n)^{(1)}, W(n)^{(2)}, S(n)^{(1)}, S(n)^{(2)}) \xrightarrow{d} (W^{(1)}, W^{(2)}, S^{(1)}, S^{(2)}),$$

for $p \in (0, 1/2)$ and $p < \alpha/2$, as claimed in (2.3.17). Similarly, for $-p \in (0, 1/2)$ we have also obtained the convergence

$$(W(n, \epsilon)^{(1)}, W(n)^{(2)}, S(n)^{(1)}, S(n)^{(2)}) \xrightarrow{d} (W(\epsilon)^{(1)}, W^{(2)}, S^{(1)}, S^{(2)}),$$

for any $\epsilon > 0$. Here the limit $(W(\epsilon)^{(1)}, W^{(2)}, S^{(1)}, S^{(2)})$ is defined as in Theorem 2.3.2, where the function f_{-p} is replaced by f_{-p}^ϵ . In order to prove the original theorem for $-p \in (0, 1/2)$ we need to let $\epsilon \rightarrow 0$, which is the subject of the next subsection.

Letting $\epsilon \rightarrow 0$

For simplicity we may assume that $d = 1$ and $r_1 = r$, $k_1 = k$. In the first step we will show that

$$\lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{E} \left[\left(W(n, \epsilon)^{(1)} - W(n)^{(1)} \right)^2 \right] = 0.$$

We define the function $\bar{f}_{-p}^\epsilon = f_{-p} - f_{-p}^\epsilon$, $p \in (0, 1/2)$. Notice that $\text{supp}(\bar{f}_{-p}^\epsilon) = [-\epsilon, \epsilon]$ and $|\mathfrak{F}^{-1}(f_{-p}^\epsilon)| \leq C\epsilon$ for all $x \in \mathbb{R}$. Applying the formula (2.3.4) we conclude that

$$\left| \text{cov} \left(\bar{f}_{-p}^\epsilon \left(\Delta_{i,k}^r X \right), \bar{f}_{-p}^\epsilon \left(\Delta_{i+l,k}^r X \right) \right) \right| \leq C\epsilon^2 \int_{\mathbb{R}^2} |U_{h_k, r, h_k, r(\cdot+l)}(x, y)| dx dy.$$

In [32, Lemma 3.4] it has been proved that the inequality

$$\int_{\mathbb{R}^2} |U_{h_k, r, h_k, r(\cdot+l)}(x, y)| dx dy \leq C\rho_l$$

holds (in fact, the proof is the same as for Lemma 2.7.4). Hence, we conclude by Lemma 2.7.2 and the condition $k > H + 1/\alpha$

$$\left| \text{cov} \left(\bar{f}_{-p}^\epsilon \left(\Delta_{i,k}^r X \right), \bar{f}_{-p}^\epsilon \left(\Delta_{i+l,k}^r X \right) \right) \right| \leq C\epsilon^{2l(\alpha(H-k)-1)/2} \quad (2.7.17)$$

Since $(\alpha(H - k) - 1)/2 < -1$ when $k > H + 1/\alpha$ we readily deduce the estimate

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left[\left(W(n, \epsilon)^{(1)} - W(n)^{(1)} \right)^2 \right] \leq C\epsilon^2$$

and the first statement follows.

Now, we are left to proving weak convergence for the vector $(W(\epsilon)^{(1)}, W^{(2)})$ as $\epsilon \rightarrow 0$. This random variable is bivariate normal with mean 0. Hence, it suffices to show that the covariance matrix converges. But this follows by setting $\epsilon = 1/N$ and applying a Cauchy sequence argument as presented in (2.7.16). Thus, the proof of Theorem 2.3.2 is complete. \square

Proof of Theorem 2.4.1

Part (i) of Theorem 2.4.1 follows from Theorem 2.3.2 applied to the setting $d = 2$, $p \in (0, 1/2)$, $k_j = k \geq 2$ (and hence $k > H + 1/\alpha$), and the classical delta method. In fact, we only use the central limit theorem part of Theorem 2.3.2.

Part (ii) of Theorem 2.4.1 is slightly more involved. We start with the identity ($t > 0$)

$$\varphi_{\text{high}}(t; \widehat{H}, k)_n - \varphi_{\text{high}}(t; H, k)_n = \frac{1}{n} \sum_{i=k}^n \left\{ \cos(tn^{\widehat{H}} \Delta_{i,k}^n X) - \cos(tn^H \Delta_{i,k}^n X) \right\},$$

where we use the short notation $\widehat{H} = \widehat{H}_{\text{high}}(p, k)_n$. Setting $M_n = n^{\widehat{H}-H}$ and using the inequality $|\cos(y) - \cos(x) + (y - x) \sin(x)| \leq C|y - x|^{\alpha'}$ for some $\alpha' \in (1, \alpha)$, we conclude that

$$\varphi_{\text{high}}(t; \widehat{H}, k)_n - \varphi_{\text{high}}(t; H, k)_n = -\frac{t(M_n - 1)}{n} \sum_{i=k}^n (n^H \Delta_{i,k}^n X) \sin(tn^H \Delta_{i,k}^n X) + R_n,$$

$$\text{where } |R_n| \leq C \frac{|M_n - 1|^{\alpha'}}{n} \sum_{i=k}^n |n^H \Delta_{i,k}^n X|^{\alpha'}.$$

We observe that $\sqrt{n}(\widehat{H} - H)$ is asymptotically normal, which follows by a delta method from Theorem 2.3.2 (take $d = 2$ and use the convergence in distribution $W(n)^{(1)} \xrightarrow{d} W^{(1)}$). By the mean value theorem we obtain that

$$\sqrt{n}(\log n)^{-1}(M_n - 1) = \sqrt{n}(\widehat{H} - H) + o_{\mathbb{P}}(1).$$

Hence, recalling that $\alpha' \in (1, \alpha)$, we deduce by Birkhoff's ergodic theorem

$$\sqrt{n}(\log n)^{-1} \left(\varphi_{\text{high}}(t; \widehat{H}, k)_n - \varphi_{\text{high}}(t; H, k)_n \right) = \sqrt{n}(\widehat{H} - H)t\varphi'(t; k) + o_{\mathbb{P}}(1), \quad (2.7.18)$$

where we used the identity $t\varphi'(t; k) = -\mathbb{E}[n^H \Delta_{i,k}^n X \sin(tn^H \Delta_{i,k}^n X)]$. Finally, we note that

$$\varphi_{\text{high}}(t; H, k)_n - \varphi(t; k) = O_{\mathbb{P}}(n^{-1/2}),$$

which follows from Theorem 2.3.2. Hence, observing the identities (2.4.3) and (2.4.4), we obtain the statement of Theorem 2.4.1(ii) by applying the delta method to Theorem 2.3.2. \square

Proof of Theorem 2.5.1

First of all, we note that $\delta := \alpha^{-1} - \lfloor \alpha^{-1} \rfloor \in (0, 1)$ since $\alpha^{-1} \notin \mathbb{N}$. Setting $\delta' := \min\{\delta, 1 - \delta\}/2 > 0$ we conclude that

$$\mathbb{P}\left(\widehat{k}_{\text{low}} \neq 2 + \lfloor \alpha^{-1} \rfloor\right) \leq \mathbb{P}\left(|\widehat{\alpha}_{\text{low}}^0(t_1, t_2)_n^{-1} - \alpha^{-1}| > \delta'\right) \rightarrow 0,$$

because $\widehat{\alpha}_{\text{low}}^0(t_1, t_2)_n \xrightarrow{\mathbb{P}} \alpha$ and $\alpha > 0$. This implies the convergence $\widehat{k}_{\text{low}} \xrightarrow{\text{a.s.}} 2 + \lfloor \alpha^{-1} \rfloor$. Thus, it suffices to prove the asymptotic results of Theorem 2.5.1 when \widehat{k}_{low} is replaced by $2 + \lfloor \alpha^{-1} \rfloor$. Now, notice that $k = 2 + \lfloor \alpha^{-1} \rfloor$ automatically satisfies the condition $k > H + 1/\alpha$ since $H \in (0, 1)$. This guarantees that the statistic $(W(n)^{(1)}, W(n)^{(2)})$ defined at (2.3.13) is in the domain of attraction of the central limit theorem. Hence, Theorem 2.5.1(i) follows directly by the delta method from Theorem 2.3.2 (cf. proof of Theorem 2.4.1(i)). \square

Proof of Proposition 2.5.2

Proposition 2.5.2 is shown by exactly the same arguments as Theorem 2.4.1. \square

Proof of Theorem 2.5.3

Recall that $\alpha^{-1} \in \mathbb{N}$. Hence, we have

$$\mathbb{P}\left(\widehat{k}_{\text{low}} \notin \{1 + \alpha^{-1}, 2 + \alpha^{-1}\}\right) \leq \mathbb{P}\left(|\widehat{\alpha}_{\text{low}}^0(t_1, t_2)_n^{-1} - \alpha^{-1}| > 1\right) \rightarrow 0,$$

because $\widehat{\alpha}_{\text{low}}^0(t_1, t_2)_n \xrightarrow{\mathbb{P}} \alpha$ and $\alpha > 0$. Note that $k \in \{1 + \alpha^{-1}, 2 + \alpha^{-1}\}$ satisfies the condition $k > H + \alpha^{-1}$, which guarantees the validity of a central limit theorem for the statistic $(W(n)^{(1)}, W(n)^{(2)})$ defined at (2.3.13).

We introduce the notation

$$T_{\text{low}}(\widehat{k}_{\text{low}}, n) := \sqrt{n} \begin{pmatrix} \widetilde{\sigma}_{\text{low}}(\widehat{k}_{\text{low}}, t_1, t_2)_n - \sigma \\ \widetilde{\alpha}_{\text{low}}(\widehat{k}_{\text{low}}, t_1, t_2)_n - \alpha \\ \widehat{H}_{\text{low}}(-p, \widehat{k}_{\text{low}})_n - H \end{pmatrix}$$

and

$$a_n := \begin{cases} \sqrt{n} : & \text{if } H < 1 - \alpha^{-1} \\ n^{1-1/(1+\alpha(1-H))} : & \text{if } H > 1 - \alpha^{-1} \end{cases}$$

We set $U_n = a_n(\widehat{\alpha}_{\text{low}}^0(t_1, t_2)_n - \alpha)$, $A = (a_1, b_1) \times (a_2, b_2) \times (a_3, b_3)$ and observe the decomposition

$$\begin{aligned} \mathbb{P}(T_{\text{low}}(\widehat{k}_{\text{low}}, n) \in A) &= \mathbb{P}\left(T_{\text{low}}(1 + \alpha^{-1}, n) \in A, \widehat{\alpha}_{\text{low}}^0(t_1, t_2)_n^{-1} - \alpha^{-1} < 0\right) \\ &\quad + \mathbb{P}\left(T_{\text{low}}(2 + \alpha^{-1}, n) \in A, \widehat{\alpha}_{\text{low}}^0(t_1, t_2)_n^{-1} - \alpha^{-1} \geq 0\right) + o(1) \\ &= \mathbb{P}\left(T_{\text{low}}(1 + \alpha^{-1}, n) \in A, U_n > 0\right) \\ &\quad + \mathbb{P}\left(T_{\text{low}}(2 + \alpha^{-1}, n) \in A, U_n \leq 0\right) + o(1). \end{aligned}$$

Applying Theorem 2.3.2 and Proposition 2.5.2, and using the same arguments as in the proof of Theorem 2.4.1, we thus conclude the convergence

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(T_{\text{low}}(\widehat{k}_{\text{low}}, n) \in A) &= \mathbb{P}\left(B_{\text{low}}^{\text{nor}}(-p, 1 + \alpha^{-1}) \in A, B_{\text{low}}(-p, 1)_2 > 0\right) \\ &\quad + \mathbb{P}\left(B_{\text{low}}^{\text{nor}}(-p, 2 + \alpha^{-1}) \in A, B_{\text{low}}(-p, 1)_2 < 0\right), \end{aligned}$$

where $B_{\text{low}}(-p, 1) = B_{\text{low}}^{\text{nor}}(-p, 1)$ if $H < 1 - \alpha^{-1}$, and $B_{\text{low}}(-p, 1) = B_{\text{low}}^{\text{sta}}(-p, 1)$ if $H > 1 - \alpha^{-1}$. Hence, we immediately obtain the assertion of Theorem 2.5.3. \square

Proof of Theorem 2.5.5

As in the proof of Theorem 2.5.1 we conclude that $\widehat{k}_{\text{high}} \xrightarrow{\text{a.s.}} 2 + \lfloor \alpha^{-1} \rfloor$. On the other hand, similarly to (2.7.18), we obtain the asymptotic expansion

$$\sqrt{n}(\log n)^{-1} \left(V_{\text{high}}(f_{-p}, \widehat{H}_{\text{high}}(-p)_n)_n - V_{\text{high}}(f_{-p})_n \right) = -\sqrt{n}(\widehat{H} - H)pm_{p,k} + o_{\mathbb{P}}(1),$$

for any $p \in (0, 1/2)$. Hence, the assertion of Theorem 2.5.5 follows the delta method and Theorem 2.3.2 (cf. the proof of Theorem 2.4.1). \square

Proof of Theorem 2.5.6

The results of Theorem 2.5.6 follow by the same methods as presented in the proof of Theorem 2.5.3. \square

2.8 Appendix

Here the codes for numerical experiments from Section 2.6 are presented.

```
#### Set of global parameters ####
library(rlfsm)
registerDoParallel()
library(gridExtra)

m<-25#256
M<-60#600

p<-.4; p_prime<-.2
t1<-1; t2<-2; k<-2

NmonteC<-5e2 #2e3
LofF<-NULL
```

The following chunk corresponds to Theorem 2.4.1, table 2.1 and Figure 2.2.

```
#####
S<-c(1e2, 1e3, 1e4)

alpha<-1.8
H<-0.8
sigma<-0.3

theor_3_1_H_clt<-CLT(s=S, fr='H', Nmc=NmonteC, m=m, M=M, alpha=alpha,
                    H=H, sigma=sigma, ContinEstim, t1=t1, t2=t2, p=p, k=k)
```

```

theor_3_1_L_clt<-CLT(s=S,fr='L',Nmc=NmonteC, m=m, M=M, alpha=alpha,
                    H=H, sigma=sigma, ContinEstim, t1=t1, t2=t2,p=p, k=k)

### Panel plot for LH CLTs
l_plot_H<-Plot_dens(par_vec=c('sigma', 'alpha', 'H'),
                  CLT_data=theor_3_1_H_clt, Nnorm=1e7)
l_plot_L<-Plot_dens(par_vec=c('sigma', 'alpha', 'H'),
                  CLT_data=theor_3_1_L_clt, Nnorm=1e7)

ggg<-grid.arrange(l_plot_L[[1]], l_plot_H[[1]], l_plot_L[[2]],
                 l_plot_H[[2]], l_plot_L[[3]], l_plot_H[[3]],
                 nrow=3, ncol=2)

```

The next chunk corresponds accordingly to table 2.3, Figure 2.3 and table 2.5, Figure 2.5.

```

#### Theor 4.1 and 4.5 contin #####

alpha<-1.8
H<-0.8
sigma<-0.3

theor_4_1_clt<-CLT(s=S,fr='L', Nmc=NmonteC, m=m, M=M, alpha=alpha,
                  H=H, sigma=sigma, GenLowEstim, t1=t1, t2=t2, p=p)

### Panel plot
l_plot<-Plot_dens(par_vec=c('sigma', 'alpha', 'H'),
                  CLT_data=theor_4_1_clt, Nnorm=1e7)
ggg<-grid.arrange(l_plot[[1]],l_plot[[2]],l_plot[[3]],nrow=1,ncol=3)

alpha<-1.8
H<-0.8
sigma<-0.3

theor_4_5_clt<-CLT(s=S, fr='H', Nmc=NmonteC, m=m, M=M, alpha=alpha,
                  H=H, sigma=sigma, GenHighEstim, p=p, p_prime=p_prime)

### Panel plot
l_plot<-Plot_dens(par_vec=c('sigma', 'alpha', 'H'),
                  CLT_data=theor_4_5_clt, Nnorm=1e7)
ggg<-grid.arrange(l_plot[[1]], l_plot[[2]], l_plot[[3]], nrow=1, ncol=3)

```

The next chunk corresponds accordingly to table 2.4, Figure 2.4 and table 2.6, Figure 2.6.

```

#### Theor 4.1 and 4.5 discontin #####

#### H-1/alpha<0 ####
#S<-c(1e3,1e4)
alpha<-.8
H<-0.8
sigma<-0.3

```

```
theor_4_1_clt_new<-CLT(s=S, fr='L', Nmc=NmonteC, m=m, M=M, alpha=alpha,
                      H=H, sigma=sigma, GenLowEstim, t1=t1, t2=t2, p=p)

### Panel plot
l_plot<-Plot_dens(par_vec=c('sigma','alpha','H'),
                 CLT_data=theor_4_1_clt_new, Nnorm=1e7)
ggg<-grid.arrange(l_plot[[1]],l_plot[[2]],l_plot[[3]],nrow=1,ncol=3)

alpha<-0.8
H<-0.8
sigma<-0.3

theor_4_5_clt_new<-CLT(s=S, fr='H', Nmc=NmonteC, m=m, M=M, alpha=alpha,
                      H=H, sigma=sigma, GenHighEstim, p=p, p_prime=p_prime)

l_plot<-Plot_dens(par_vec=c('sigma','alpha','H'),
                 CLT_data=theor_4_5_clt_new, Nnorm=1e7)
ggg<-grid.arrange(l_plot[[1]],l_plot[[2]],l_plot[[3]],nrow=1,ncol=3)

#####
```

Chapter 3

Numerical techniques and results for the Linear Fractional Stable Motion

STEPAN MAZUR, DMITRY OTRYAKHIN

3.1 Overview of Chapter 3

Chapter 3 is a natural continuation of Chapter 2. It introduces **rlfsm** package that was used to perform the numerical studies in Chapter 2, brings insight on how this software was designed and optimized, and shows through new numerical experiments how the estimators from Chapter 2 work in practice.

rlfsm is a package for R programming language which I initially wrote with Dr. Stepan Mazur specifically to study numerical properties of estimators related to the linear fractional stable motion. It is open-source software licensed under GPL-3, and is available on CRAN, The Comprehensive R Archive Network [1], which is today's largest repository for checked R packages. The latest version of **rlfsm** on CRAN 0.3.1 contains approximately 1770 lines of code and includes source code, in-line documentation, comments, examples and 37 unit tests. There is a gitlab developer repository [2] for this project, where you can find the freshest bleeding-edge version of this software, and also all the codes for numerical experiments we have performed regarding LFSMs. Packaging and releasing was done by means of **devtools** package [52]. Automated testing employs **testthat** [47] package which provides integration with standard R package checks (`R CMD check`) and a very formative and parsimonious output for test results. Testing on multiple platforms was done with the help of R-hub project and **rhub** package [18]. Documentation was written using **Roxigen2** [51] package. Use of the latter brings several advantages, e.g. some pieces of documentation are automatically created, so each time the code gets changed, the documentation updates accordingly. Another instance is that **Roxigen2** works together with **devtools**, thus documentation is renewed and vignettes are printed every build. All these tools are well described in [49].

I would like to mention the contribution of Mark Podolskij, whose intuition and experience guided me and Stepan during the very early stage of developing and debugging of the package, and also that of my peer Mathias Ljungdahl, who pointed at significant performance issues and brought a new estimator to the package.

The paper is organized as follows. Section 3.2 gives a brief introduction. In Section 3.3 we present the simulation method for sample paths of lfsm and its implementation in our `path` function. Then, we present functions for finite sample studies of statistical estimators, and some other functions. Section 3.4 describes implementations of the high- and the low-frequency parameter estimators and discusses reasons behind their numerical behavior. Finally, in Section 3.5 we suggest an object oriented system that simplifies software programming of Lévy-driven integrals.

3.2 Introduction

The linear fractional stable motion (shortly, lfsm) $(X_t)_{t \in \mathbb{R}}$ on a filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}}, \mathbb{P})$ is defined via

$$X_t = \int_{\mathbb{R}} \left\{ (t-s)_+^{H-1/\alpha} - (-s)_+^{H-1/\alpha} \right\} dL_s, \quad x_+ := \max\{x, 0\}, \quad (3.2.1)$$

where L_s is a symmetric α -stable Lévy motion, $\alpha \in (0, 2)$, with the scaling parameter $\sigma > 0$ and the self-similarity parameter $H \in (0, 1)$. The lfsm is heavy-tailed process with infinite variance and long-range dependence. A good overview on the role which this process plays in natural sciences is done by Watkins et al. [45]. One could also find a review of stochastic properties of lfsm in [29].

We proceed with introduction to existing software, with interest towards study of numerical properties of statistical estimators for lfsm as the main motivation. So far, there is no standard approach for software development to operating the general class of stochastic processes driven by Lévy processes. Moreover, there was no systematic indexed and peer-reviewed software for simulating sample paths of lfsm and related estimators prior to `rlfsm`. There is a particularly simple and useful numerical algorithm for simulating lfsms developed by Stoev and Taqqu [40]. The paper contains a minimalistic implementation of lfsm generator as a MATLAB function. However, some useful packages, that could be used in numerical routines with Lévy-driven processes (e.g. to create lfsm generator and perform unit testing), exist and have been implemented in R. For instance, R package `somebm` [24] contains functions for generation of fractional Brownian motion (fBM). `dvfBm` [15] has routines for generation of fBm and estimator of the Hurst parameter of the latter. `stabledist` [53] and `stable` [43] contain different functions for stable distributions and random variables. A generator of random variables of the kind has been also implemented in MATLAB (see the code in Chapter 1.7 in [37]).

3.3 Basic R functions

Types of data we use

This version of the package suggests that we work with two types of sample paths. In the low-frequency setting we only use points spaced 1 temporal index apart from each other, X_1, X_2, \dots, X_n . In the case of high-frequency, we use points with discretization equal to the length of the path vector, $X_{1/n}, X_{2/n}, \dots, X_1$. This division is dictated by two issues: 1) the same division in the setting of limit theorems obtained by Mazur et al. [29], and 2) the fact that there is no inference technique for an arbitrary mixture of the two frequencies. Consequently, temporal coordinates of low-frequency lfsm coincide with point

index (compare `coordinates` and `point_num` in the example in Section 3.3) which varies from 0 to N . Analogously, in case of high-frequency scheme, temporal coordinates equal to point indexes divided by the total number of sampled points. When after sampling the index set is different from either $(1, 2, \dots, N)$ or $(1/n, 2/n, \dots, 1)$, rescaling in time should be performed using the equality $(a^H X_t)_{t \geq 0} \stackrel{d}{=} (X_{at})_{t \geq 0}$ with $a > 0$ provided that H is known or obtained via preliminary estimation.

Simulation method for the linear fractional stable motion

In this section, we start with a discussion on the simulation method of the lfsm proposed by Stoev and Taqqu [40] which is implemented in R by us. In particular, simulation of sample paths is done via Riemann-sum approximations of its symmetric α -stable stochastic integral representation while Riemann-sums are computed efficiently by using the Fast Fourier Transform algorithm. In R, we introduce `path` function that creates sample paths of the lfsm. The idea underlying this sample path generator is that it should be always possible not only to obtain lfsm path, but also the underlying Lévy motion, generated during the procedure, and since the core function of lfsm is deterministic it should allow for lfsm path generation based on a given Lévy motion, and, in theory, otherwise (not always). For this reason generators of both processes were separated into independent parts (see Figure 3.1).

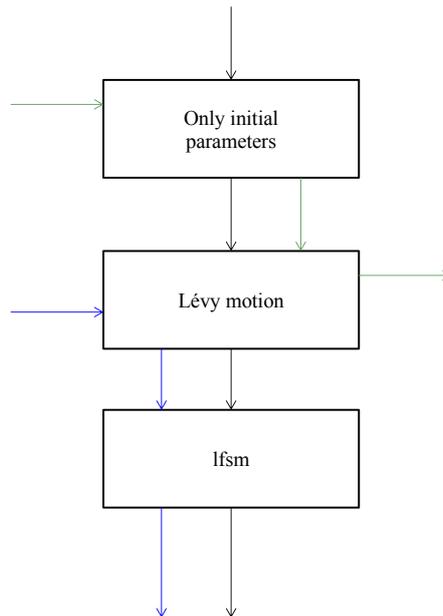


Figure 3.1: Scheme of generating Lévy motion and lfsm by `path`. Black arrows: when the algorithm initially is given the parameters, it generates Lévy motion, and then lfsm. Green arrows: when Lévy motion is needed without lfsm in order to save processing time, the algorithm bypasses computing of the later. Blue arrows: given a Lévy motion and some parameters, the generator computes the corresponding lfsm.

The function `path` can be used by

```
path(N,m,M,alpha,H,sigma,freq='L',disable_X=FALSE,
     levy_increments=NULL,seed=NA)
```

Parameters N , m , M regard to the index of the process, or time, if applicable. m and M are the only means to control precision of the integral computation. N is a number of points of the lfsm to generate. m is a discretization parameter that corresponds to the number of points where Lévy motion is sampled between two nearby indexes (e.g. N and $N - 1$). M is the truncation parameter, i.e. number of points after which the integrated function is set to zero; **freq** stands for the frequency of the motion which can take two values: "H" for high-frequency and "L" for the low-frequency setting. This is the switch between the two data types. **disable_X** is needed to disable computation of X , the default value is FALSE, when it is TRUE, only a Lévy motion is returned, which in turn reduces the computation time. **seed** is a parameter that performs seeding of the lfsm generator. Technically, in the **path** the seed is set just before Lévy increments are generated. The **path** function returns a list containing the lfsm, the underlying Lévy motion, the point number of the motions from 0 to N (**point_num**) and the corresponding coordinate which depends on the frequency, the parameters (σ, α, H) that were used to generate the lfsm, and the predefined frequency.

Generation of symmetric α -stable (sas) random variables is powered by function **rstable** from package **stabledist** [53] with S0 parametrization based on the Zolotarev's representation for an α -stable distribution with some modifications. S0 is used in order to make **sigma** a scale parameter of the motion and to get exempt from computing the normalization constant $C_{H,\alpha}$ presented in [40] and is given by

$$C_{H,\alpha} := \left(\int_{\mathbb{R}} |(1-s)_+^{H-1/\alpha} - (-s)_+^{H-1/\alpha}|^\alpha ds \right)^{1/\alpha}.$$

The discrete convolution based algorithm and particularities of indexing

As it was mentioned in the beginning of Section 3.3, one of the features of **path** is the ability to operate on a pair lfsm - Lévy motion and to switch between them. We recall that direct computation of the sum approximating the integral in the definition of lfsm (3.2.1) would involve number of operations proportional to NMm , which makes the method slow. Instead, the original algorithm by Stoev and Taqqu [40] suggests computing increments of lfsm with the help of

$$W(n) := \sum_{j=1}^{mM} a_{H,m}(j) Z_\alpha(n-j), \quad (3.3.2)$$

where $W(mk)$ is a discretized and truncated version of the increments of the lfsm, and in the limit has the same distribution as them

$$\{W(mk), k = 1, \dots, N\} \xrightarrow[m \rightarrow \infty; M \rightarrow \infty]{d} \{X(k) - X(k-1), k = 1, \dots, N\};$$

$Z_\alpha(k)$ are i.i.d. sas random variables that have indexes $-mM, \dots, mN - 1$ and scaling parameter equal to 1, and

$$a_{H,m}(j) := C_{H,\alpha}^{-1}(m, M) \left((j/m)^{H-1/\alpha} - (j/m - 1)_+^{H-1/\alpha} \right) m^{-1/\alpha}, j \in \mathbb{N}$$

with

$$C_{H,\alpha}(m, M) := m^{-1} \left(\sum_{j=1}^{mM} \left| (j/m)^{H-1/\alpha} - (j/m - 1)_+^{H-1/\alpha} \right|^\alpha \right)^{1/\alpha}.$$

	-2	-1	0	1	2	3	4	5	index
	7	9	1	3	8	2	6	8	Z
			4	9	2				a
			99	39	61	86	58	90	W(n)

Figure 3.2: Example of direct computation of sum of the form (3.3.2) for 2 vectors. a corresponds to the kernel and Z - to the Lévy motion.

Let us consider an example which will recur and evolve throughout this section. Consider computing sum (3.3.2) where $m = 1$, $M = 3$, and $N = 6$ (see Figure 3.2). The two rightmost cells for $W(n)$ are left empty because there is no sense in computing them without truncation of a .

A method based on the discrete convolution theorem is used to obtain $W(mk)$. The theorem relies on Discrete Fourier Transform (DFT), which needs to perform a number of operations proportional to $(mN + mM) \log(mN + mM)$ instead of NMm . In order to understand how this method works, we review several definitions and theorems.

Definition 3.3.1. For any sequence $x_n, n \in \mathbb{N}$, Discrete-Time Fourier Transform (DTFT) is defined as

$$X_\omega = \text{DTFT}\{x_n\}(\omega) = \sum_{n=-\infty}^{\infty} x_n \exp(-2\pi i n \omega).$$

The reverse transform, IDTFT, is defined as

$$x_n = \text{IDTFT}\{X\} = \frac{1}{2\pi} \int_0^{2\pi} X_{2\pi}(\omega) e^{i\omega n} d\omega.$$

Definition 3.3.2. Discrete convolution of two infinite sequences $\{A_n\}_{n \in \mathbb{N}}$ and $\{B_n\}_{n \in \mathbb{N}}$ is

$$(A * B)[n] := \sum_{m=-\infty}^{\infty} A[m]B[n - m].$$

There is a convolution theorem for discrete sequences which says that the discrete convolution of two sequences is equal to the Inverse Discrete Fourier Transform (IDFT) of the multiplication of the direct transforms of the sequences:

Theorem 3.3.3. For any discrete sequences x_n and $y_n, n \in \mathbb{N}$, it holds that

$$(x * y)[n] = \text{IDTFT}[\text{DTFT}\{x_n\}(\cdot) \times \text{DTFT}\{y_n\}(\cdot)].$$

Definition 3.3.4. Let $x_n, n \in \mathbb{N}$ be a sequence. Then $\{x_N\}[n], n \in \mathbb{N}$ is called N -periodic summation of the sequence:

$$\{x_N\}[n] := \sum_{k \in \mathbb{N}} x[n + kN].$$

It is straightforward, that the periodic summation in the definition above has period N . In our case, the latter theorem is applicable even though we will be interested in a finite sequence of length \tilde{N} . The sequence is padded with zeros to form an infinite one, and a periodic summation of a the length \tilde{N} is just a periodic extension of it.



Figure 3.3: Example of periodic summation of a zero-padded finite sequence where the period equals to the sequence length ($N = \tilde{N}$).

DTFT is not directly useful for simulation purpose, that is why we need a special case of Theorem 3.3.3, Circular Convolution Theorem which reduces DTFT to DFT.

Definition 3.3.5. *The DFT of a finite sequence x_n of length N is defined as*

$$X_k = \text{DFT}_k(x_n) := \sum_{n=0}^{N-1} x_n \exp(-2\pi i k n / N).$$

The IDFT is

$$x_n := \frac{1}{N} \sum_{k=0}^{N-1} X_k \exp(2\pi i k n / N).$$

Theorem 3.3.6.

$$(x_N * y)[n] = \text{IDFT}\{\text{DFT}(x_N)\text{DFT}(y_N)\}$$

Returning to the task of computing the sum in (3.3.2), we consider two vectors a of length mM and Z of length $m(M + N)$. Here, we again index vectors starting with zero, not one. If we extend Z periodically, pad a with zeros to make an infinite sequence, and compute $(a * Z_{m(N+M)})[n]$, values with indexes $[mM; m(N + M) - 1]$ would coincide with the result of a convolution of a and Z . The first mM values would be meaningless. This gives an idea how to use Circular Convolution Theorem for computation of (3.3.2): instead of $a * Z$ we compute one period of $(a * Z_{m(N+M)})[n]$ through the left part of 3.3.6 and leave only meaningful values. Figure 3.4 illustrates the use of Circular Convolution Theorem with periodic extensions of Z and padded a to compute (3.3.2). In this case results with indexes -1 and -2 are meaningless and should be discarded.

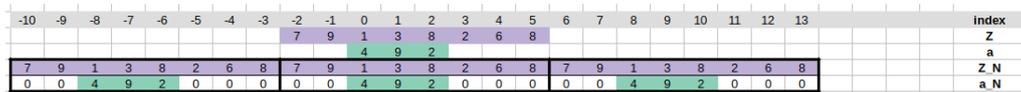
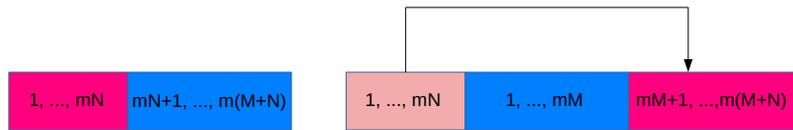
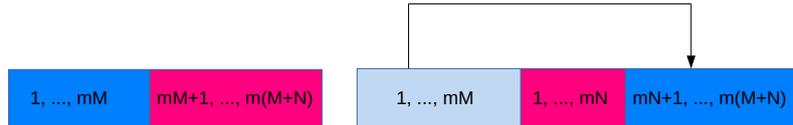


Figure 3.4: Example of transformation of vectors a and Z into sequences before computing their convolution.

Although the setup of the example as is on Figure 3.4 is the fastest, it is impossible to use it directly, because in some situations truncation parameter M is larger than N , the number of points of lfsm sample path that is needed to be simulated. In this case `path` function performs an index shift using the following property:



(a) Indexing of Lévy increments when `path` is to generate them. Left: Indexes before generation of `lfsm` vector. Right: vector modification prior to passing Lévy motion to the output.



(b) Indexing of Lévy increments when `path` is to simulate `lfsm` based on them. Left: initial indexes. Right: vector modification prior to `lfsm` generation.

Figure 3.6: Index change in path function

```
R> # Normalized paths
R> Norm_lfsm<-List[['lfsm']]/max(abs(List[['lfsm']]))
R> Norm_oLm<-List[['levy_motion']]/max(abs(List[['levy_motion']]))

R> # Visualization of the paths
R> plot(Norm_lfsm, col=2, type="l", ylab="coordinate")
R> lines(Norm_oLm, col=3)
R> leg.txt <- c("lfsm", "oLm")
R> legend("topright", legend = leg.txt, col =c(2,3), pch=1)
```

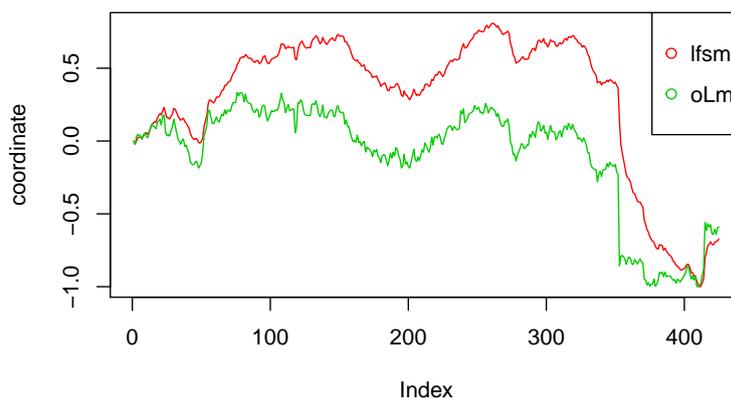


Figure 3.7: Plot of sample path and Lévy motion with `seed=2`

The result of the chart rendering is shown on Figure 3.7. The following example shows how to switch `path` function in order to alter between simulation of `lfsm` from scratch and

computing based on existing sample path of the Lévy motion.

```
R> m<-256; M<-600; N<-2^12-M
R> alpha<-1.8; H<-0.8; sigma<-1.8
R> seed<-2

R> # Creating Levy motion
R> levyIncrems<-path(N=N, m=m, M=M, alpha, H, sigma, freq='L',
                    disable_X=T, levy_increments=NULL, seed=seed)

R> # Creating lfsm based on the levy motion
R> lfsm_full<-path(m=m, M=M, alpha=alpha,
                  H=H, sigma=sigma, freq='L',
                  disable_X=F,
                  levy_increments=levyIncrems$levy_increments,
                  seed=seed)

R> sum(levyIncrems$levy_increments==
      lfsm_full$levy_increments)==length(lfsm_full$levy_increments)

[1] TRUE
```

In the example the Lévy motion is generated without computing the lfsm, which was done by setting `disable_X=T`, and saved to variable `levyIncrems`. After that, `path` was given the obtained Lévy increments and, basing on them, generated an lfsm path. As one can observe, the Lévy increments from the both objects produced by `path` are identical. The same holds when we obtain an lfsm path from the above procedure and one-step simulation of lfsm with seeding. These two facts are used in automated tests provided for `rlfsm` package.

CLT and numerical properties of statistical estimators

In order to study numerical properties of the estimation procedures developed in [29], we created a technique, that could be used in solving this problem for any pair stochastic process and an estimator. The approach was implemented in `CLT` function (Figure 3.10). The main motivation here is that for some estimators we have limit theorems, but we do not have theory which describes estimator behavior when the length of a path is relatively small, and thus, for instance, we cannot use closed-form expressions to obtain confidential intervals. In the following examples we show how to use functions `CLT`, `Plot_vb`, and `Plot_dens` for studying empirical variance, bias and a density function of an estimator. In the first example, we study `GenLowEstim` estimator, and its bias and variance dependencies on the length of the sample paths. In particular, one would be able to determine starting from which path length the estimator loses bias influence.

```
R> library(rlfsm)
R> library(gridExtra)
R> registerDoParallel()

R> m<-25; M<-55
R> p<-0.4; p_prime<-0.2
R> t1<-1; t2<-2
R> k<-2

R> NmonteC<-5e2
R> alpha<-1.8; H<-0.8; sigma<-0.3
```

```
R> S<-seq(from = 100, to = 2e3, by =50)
R> tilda_estst<-CLT(s=S, fr='L', Nmc=NmonteC, m=m, M=M,
                    alpha=alpha,H=H,sigma=sigma,
                    GenLowEstim,t1=t1,t2=t2,p=p)

# Structure of tilda_estst is
R> str(tilda_estst)

List of 7
 $ CLT_dataset:'data.frame':      1126 obs. of  4 variables:
 ..$ s      : num [1:1126] 100 100 100 100 100 100 100 100 100 100 ...
 ..$ H      : num [1:1126] 0.061 -2.3734 -0.4798 0.0613 0.7941 ...
 ..$ alpha: num [1:1126] -0.928 -0.129 0.655 -3.099 0.287 ...
 ..$ sigma: num [1:1126] -0.6235 -2.1252 0.0607 -1.8552 1.2099 ...
 $ BSdM     : num [1:3, 1:10] 100 1000 10000 1.76 1.82 ...
 ..- attr(*, "dimnames")=List of 2
 .. ..$ : NULL
 .. ..$ : chr [1:10] "s" "alpha_Mean" "alpha_Sd" "alpha_b" ...
 $ Inference :function (t1, t2, p, path, freq)
 ..- attr(*, "srcref")=Class 'srcref' atomic [1:8] 78 14 109 1 14 1 1000 1031
 .. ..- attr(*, "srcfile")=Classes 'srcfilealias', 'srcfile' <environment: 0x6837048>
 $ alpha    : num 1.8
 $ H        : num 0.8
 $ sigma    : num 0.3
 $ freq     : chr "L"

# Structure of BSdM is as follows
R> head(round(tilda_estst$BSdM,2))
```

	s	alpha_Mean	alpha_Sd	alpha_b	H_Mean	H_Sd	H_b	sigma_Mean	sigma_Sd	sigma_b
[1,]	100	1.78	0.18	-0.02	0.66	0.25	-0.14	0.25	0.10	-0.05
[2,]	150	1.81	0.15	0.01	0.68	0.22	-0.12	0.26	0.08	-0.04
[3,]	200	1.82	0.13	0.02	0.71	0.21	-0.09	0.27	0.08	-0.03
[4,]	250	1.84	0.11	0.04	0.74	0.18	-0.06	0.29	0.07	-0.01
[5,]	300	1.82	0.11	0.02	0.73	0.18	-0.07	0.28	0.07	-0.02
[6,]	350	1.82	0.11	0.02	0.75	0.16	-0.05	0.29	0.06	-0.01

```
R> Plot_vb(tilda_estst$BSdM)
```

Figure 3.8 shows that when $(\sigma, \alpha, H) = (0.3, 1.8, 0.8)$, estimator `GenLowEstim` is unbiased starting approximately from 1000 points.

The second example compares empirical standardized densities of estimates, obtained by `GenLowEstim` with the limiting standard normal ones.

```
R> S<-c(1e2,1e3,1e4)
R> tilda_estst<-CLT(s=S, fr='L', Nmc=NmonteC ,m=m, M=M,
                    alpha=alpha, H=H, sigma=sigma,
                    GenLowEstim,t1=t1,t2=t2,p=p)

R> l_plot<-Plot_dens(par_vec=c('sigma','alpha','H'), CLT_data=tilda_estst, Nnorm=1e7)
R> ggg<-grid.arrange(l_plot[[1]],l_plot[[2]],l_plot[[3]],nrow=1,ncol=3)
```

In short, in these examples for different path lengths s , `NmonteC` lfsm paths are simulated. To each path we apply tilde-statistic (see Section 3.4), therefore obtaining `NmonteC` estimates $(\tilde{\sigma}_{low}, \tilde{\alpha}_{low}, \tilde{H}_{low})$ for every s , which in turn, are used to calculate biases, standard deviations, and density functions (also, for each s separately).

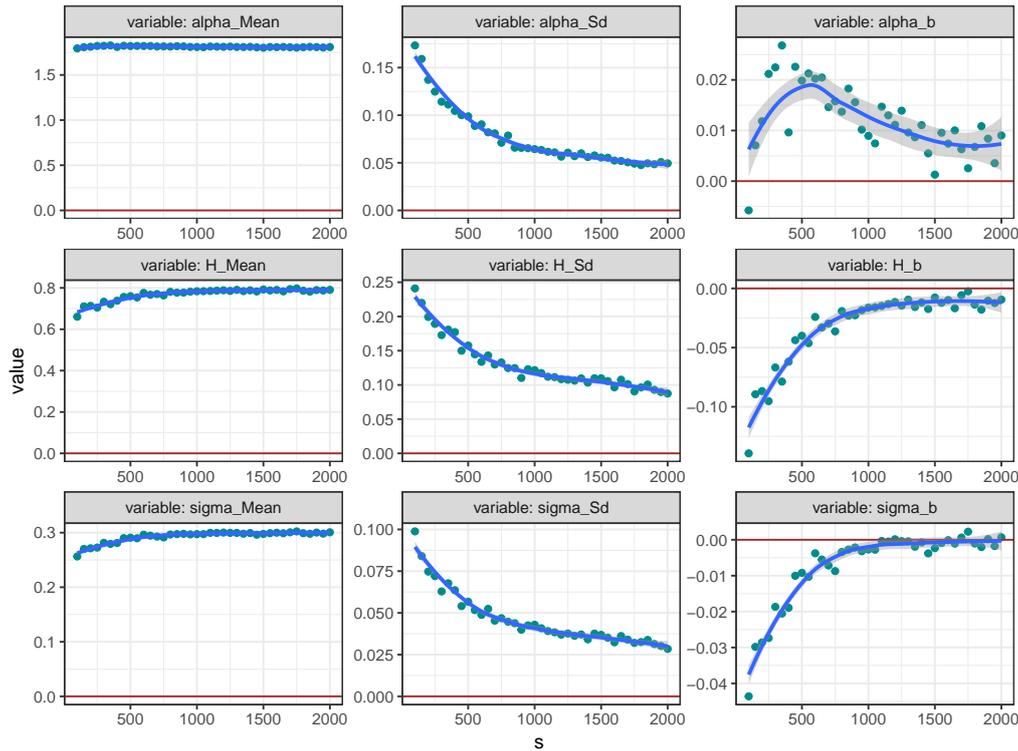


Figure 3.8: Variance and bias dependence on path length of tilde- estimators, described in Section 3.4.

CLT architecture and optimization

It is important to notice that generation of `lfsm` is numerically heavy routine and also a large number of estimates is needed to compare their empirical distributions with the limiting ones. The latter task gave CLT its name. Thus, in order to make computations feasible in terms of time and memory use, the architecture of CLT must be well-optimized. Apparently, a multi-core setup is crucial for dealing with the task.

Having fixed a path length, the whole procedure behind CLT could be split in two parts. First, we need to obtain samples for each estimator. Second, we obtain statistics of these samples (see Figure 3.10). Once finished, CLT proceeds to the next length value until reaches the end of the vector of lengths.

In the first part, we generate $N_{\text{Monte Carlo}}$ `lfsm` paths of the length `s[i]` via `path_fast` function. To each of the paths we apply all the estimators to obtain H , α , and σ estimates. During this stage, we use a `foreach`-based parallel loop, where each node simulates a path, computes and returns the statistics removing the path from memory. `path_fast` is an unavailable for users version of `path` with significantly reduced functionality for the sake of saving execution time. The further desired enlargement of the node task by adding generation of the whole set of paths instead of just one, making the loop over `s[i]` parallel, leads to extreme memory consumption as well as unequal distribution of load among nodes. The number of numeric values in the set of paths equals to $N_{\text{Monte Carlo}} \times s[i]$. Simulations, performed by Mazur et al. [29] showed that normal distribution is attained by estimators at $s = 10^3$. Given the fact that we need at least 10^5 Monte Carlo trials for a neat histogram of a distribution, one can obtain the amount of memory required to store a matrix of size

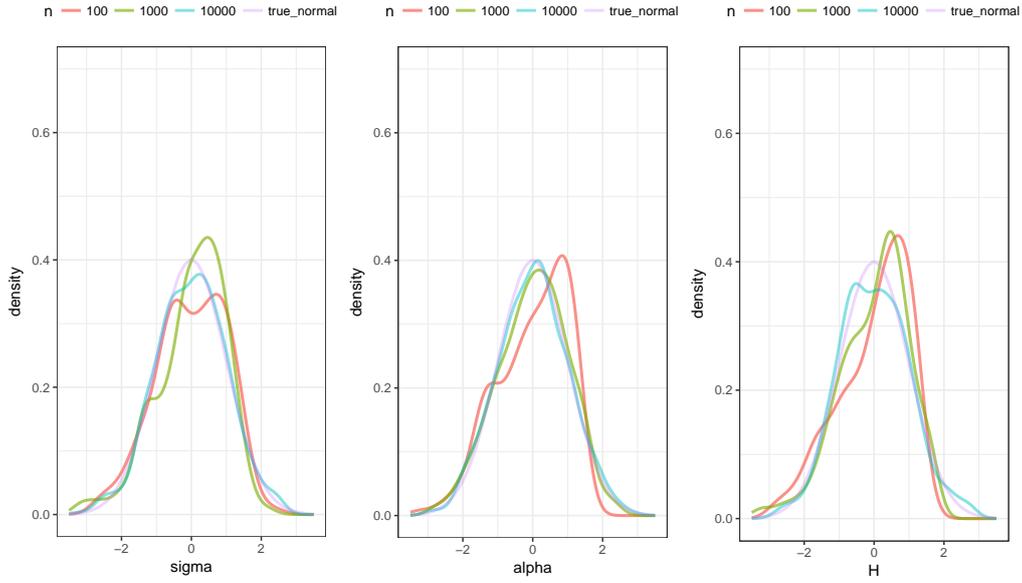


Figure 3.9: Empirical distributions of tilde- estimates, described in Section 3.4.

$N_{\text{Monte Carlo}} \times s[i]$, which makes 763Mb, while some estimators require 80Gb per node. That is the reason why in the current version of CLT the loop over \mathbf{s} is sequential, and the one over N_{monteC} is parallel.

During the second part, averages and standard deviations of the samples are computed, and subsequently used to compute the standardized empirical distributions. So that, the three characteristics naturally come together within the same numerical procedure. So far there is no empirical evidence that parallel execution in this section makes CLT more efficient.

Such architecture is of great use when the number of nodes available for computations exceeds the number of path length, and the length $s[i]$ differs significantly from $s[j]$ when $i \neq j$.

On some of the other basic functions

In this part, we will describe aspects of some of the other R functions implemented in the package.

Higher-order increments

These increments are the main building block for all statistics we use (see Section 3.4). They are defined as k -th iterated increments of step r of a sample path. In particular, $\Delta_{i,1}^{n,1} X := X_{\frac{i}{n}} - X_{\frac{i-1}{n}}$, and $\Delta_{i,2}^{n,1} X := X_{\frac{i}{n}} - 2X_{\frac{i-1}{n}} + X_{\frac{i-2}{n}}$. In **rlfsm**, we built two functions for computation of n objects of this class- `increment()` and `increments()`. The former accepts a vector of points at which a user wants to evaluate higher-order increments, and computes them using formula

$$\Delta_{i,k}^{n,r} X := \sum_{j=0}^k (-1)^j \binom{k}{j} X_{(i-rj)/n}. \quad (3.3.4)$$

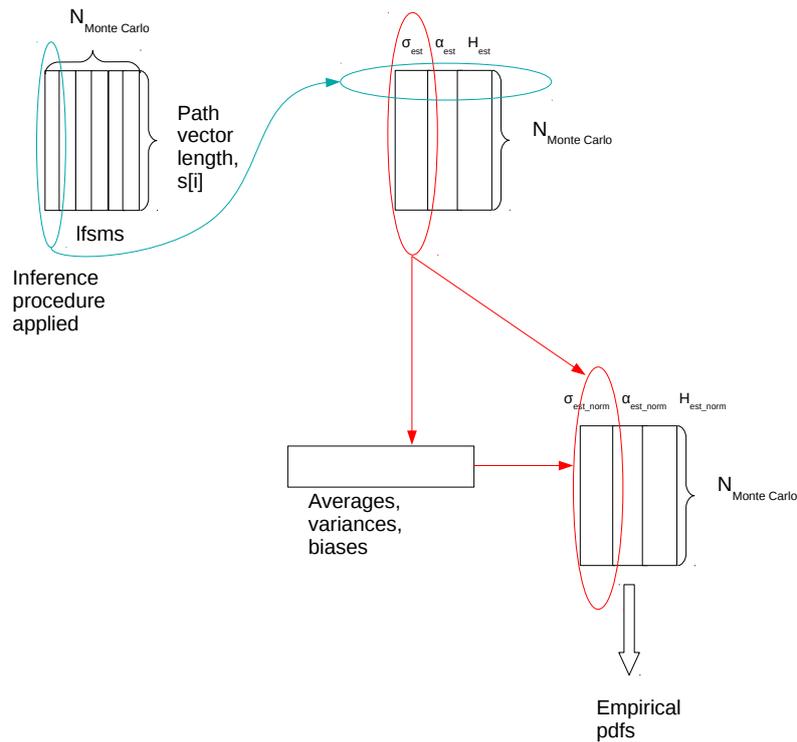


Figure 3.10: Scheme of extracting estimator statistics by function CLT for a chosen path length.

Before evaluation of (3.3.4), the function checks the condition $i < kr$. Evaluation of the increments on a sample path of length N takes $(k+1)(N-kr)$ operations- $k+1$ sums for $N-kr$ points. `increments()` computes increments iteratively on the whole set of path points. The first iteration gives $N-r$ increments, the second- $N-2r$ and so on. Thus, the total number of performed operations is

$$\sum_{j=1}^k (N-jr) = kN - r(k+1)k/2.$$

It is clear that `increments()` is faster on sample paths with large number of points, but slower when the increment order is high. As we will show later, orders greater than ~ 10 are not usable for statistical inference. That is the reason why in all statistics we use either `increments()` or its hidden “relatives”.

A visualization method for sample paths

We introduce a pair of functions which makes a panel plot of sample paths produced by processes with different parameters. `Path_array` takes a set of α - H values, generates a path for each combination, and stacks the paths together in a data frame. In the produced data frame all the paths are tagged with α and H values.

`head(arr)`

```

  n      X alpha  H freq
1 1  0.0000000  0.5 0.2  H
2 2  0.2329891  0.5 0.2  H
3 3  1.1218238  0.5 0.2  H
4 4 -6.1284620  0.5 0.2  H
5 5 -2.2450357  0.5 0.2  H
6 6  3.4979978  0.5 0.2  H

str(arr)

'data.frame':      2709 obs. of  5 variables:
 $ n      : num  1 2 3 4 5 6 7 8 9 10 ...
 $ X      : num  0 0.233 1.122 -6.128 -2.245 ...
 $ alpha  : Factor w/ 3 levels "0.5","1","1.5": 1 1 1 1 1 1 1 1 1 ...
 $ H      : Factor w/ 3 levels "0.2","0.5","0.8": 1 1 1 1 1 1 1 1 1 ...
 $ freq   : Factor w/ 1 level "H": 1 1 1 1 1 1 1 1 1 ...

```

`Plot_list_paths()` takes the data frame as an argument and plots the sample paths on different panels based on their (α, H) values. This functionality is powered by `facet_wrap()` from `ggplot2` [50]. For discontinuous paths `Plot_list_paths()` draws an overlapping semi-transparent line joining neighbouring points in order to highlight jumps.

```

l=list(H=c(0.2,0.5,0.8),alpha=c(0.5,1,1.5), freq="H")
arr<-Path_array(N=300,m=30,M=100,l=1,sigma=0.3)

```

```
Plot_list_paths(arr)
```

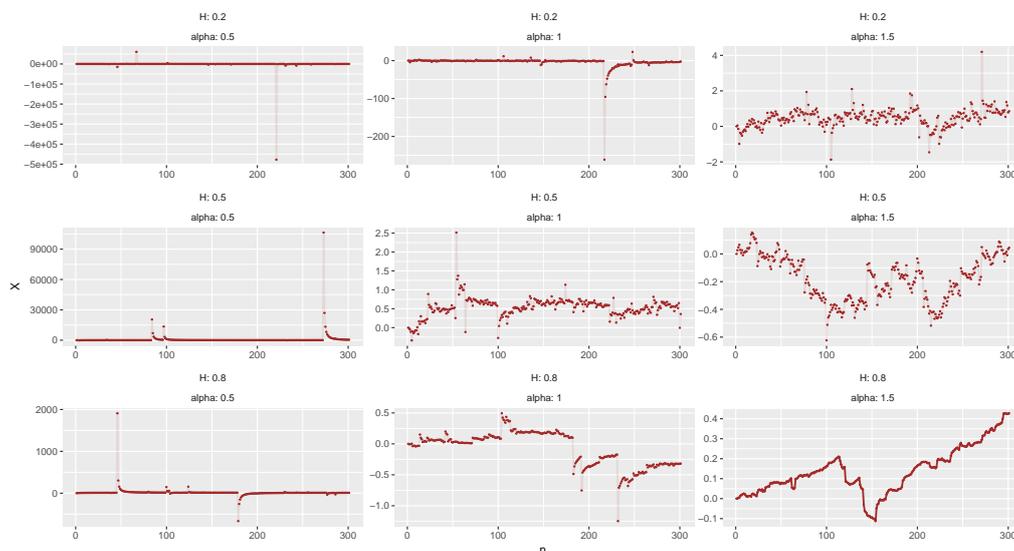


Figure 3.11: Graph rendered by `Plot_list_paths`

3.4 Parameter Estimation of the linear fractional stable motion

In this section, we describe estimators for the parameters H , α , and σ that are obtained in the recent paper by Mazur et al. [29], and their implementation in R.

Parameter estimation in the continuous case

First, we consider the case $H - 1/\alpha > 0$ which leads us to the important property that the lfsm $(X_t)_{t \in R}$ is locally Hölder continuous of any order up to $H - 1/\alpha$. Moreover, this condition implies the following restrictions

$$\alpha \in (1, 2) \quad \text{and} \quad H \in (1/2, 1)$$

that allow us to use the law of large numbers in Theorem 1.1 of [7] when $p < 1$, and the central limit theorem in Theorem 1.2 of [7] when $p < 1/2$, $k \geq 2$ and $H < k - 1/\alpha$.

Now, we consider consistent estimators for the self-similarity parameter H in high- and low-frequency setting, defined by

$$\begin{aligned} \widehat{H}_{high}(p, k)_n &:= \frac{1}{p} \log_2 \left(\frac{\sum_{i=2k}^n |\Delta_{i,k}^{n,2} X|^p}{\sum_{i=2k}^n |\Delta_{i,k}^{n,1} X|^p} \right), \\ \widehat{H}_{low}(p, k)_n &:= \frac{1}{p} \log_2 \left(\frac{\sum_{i=2k}^n |\Delta_{i,k}^2 X|^p}{\sum_{i=2k}^n |\Delta_{i,k}^1 X|^p} \right). \end{aligned}$$

Both estimators for H are based upon a ratio statistic that compares power variations at two different frequencies.

Let us define the following two statistics

$$V_{high}(f; k, r)_n := \frac{1}{n} \sum_{i=rk}^n f \left(n^H \Delta_{i,k}^{n,r} X \right), \quad V_{low}(f; k, r)_n := \frac{1}{n} \sum_{i=rk}^n f \left(n^H \Delta_{i,k}^r X \right), \quad (3.4.5)$$

where $f : \mathbb{R} \rightarrow \mathbb{R}$ is a measurable function. Estimators for the stability index α of the driving stable motion in high and low frequency setting are based on the empirical characteristic functions given by

$$\varphi_{high}(t; H, k)_n := V_{high}(\psi_t; k)_n \quad \text{and} \quad \varphi_{low}(t; k)_n := V_{low}(\psi_t; k)_n$$

with $\psi_t(x) := \cos(tx)$, for two different values t_1 and t_2 such that $t_2 > t_1 > 0$. Let us note that the empirical characteristic function $\varphi_{high}(t; H, k)_n$ depends on the parameter H while $\varphi_{low}(t; k)_n$ does not. Thus, we should infer the self-similarity parameter H by $\widehat{H}_{high}(p, k)_n$ and then we should use the plug-in estimator $\varphi_{high}(t; \widehat{H}_{high}(p, k)_n, k)_n$ to infer the stability index α in high-frequency setting. Estimators for the parameter α are given by

$$\begin{aligned} \widehat{\alpha}_{high} &:= \frac{\log |\log \varphi_{high}(t_2; \widehat{H}_{high}(p, k)_n, k)_n| - \log |\log \varphi_{high}(t_1; \widehat{H}_{high}(p, k)_n, k)_n|}{\log t_2 - \log t_1}, \\ \widehat{\alpha}_{low} &:= \frac{\log |\log \varphi_{low}(t_2; k)_n| - \log |\log \varphi_{low}(t_1; k)_n|}{\log t_2 - \log t_1}. \end{aligned}$$

Estimators for the scale parameter σ in high- and low-frequency are also based on the empirical characteristic functions which are defined for one value of $t > 0$. Further, we define a function $h_{k,r} : R \rightarrow R$ as follows:

$$h_{k,r}(x) = \sum_{j=0}^k (-1)^j \binom{k}{j} (x - rj)_+^{H-1/\alpha}, \quad x \in R,$$

where $k, r \in N$, and let $\|h_{k,r}\|_\alpha^\alpha := \int_R |h_{k,r}(s)|^\alpha ds$. Let us note that the function $h_{k,r}$ depends on two parameters α and H which need to be pre-estimated. Estimators for the parameter σ are expressed as

$$\begin{aligned}\hat{\sigma}_{high} &:= \left(-\log \varphi_{high}(t_1; \hat{H}_{high}(p, k)_n, k) \right)^{1/\hat{\alpha}_{high}} / t_1 \|h_{k,1}\|_{\hat{\alpha}_{high}}, \\ \hat{\sigma}_{low} &:= \left(-\log \varphi_{low}(t_1; k) \right)^{1/\hat{\alpha}_{low}} / t_1 \|h_{k,1}\|_{\hat{\alpha}_{low}}.\end{aligned}$$

Parameter estimation in the general case

Here, we consider general case when an explicit lower bound for α is unknown. First, we consider estimators which are obtained in low frequency setting. Consistent estimator for parameter H for any $p \in (1, 1/2)$ is obtained by

$$\hat{H}_{low}(-p, k)_n := \frac{1}{p} \log_2 \left(\frac{\sum_{i=2k}^n |\Delta_{i,k}^2 X|^{-p}}{\sum_{i=2k}^n |\Delta_{i,k}^1 X|^{-p}} \right).$$

Next, we consider two-step procedure to choose the order of increments k , since we should be in the domain of attraction of Theorem 1.2 of [7] that requires $k > H + 1/\alpha$. That's why we consider the preliminary estimator of α with $k = 1$ that is consistent given by

$$\hat{\alpha}_{low}^0(t_1, t_2)_n = \frac{\log |\log \varphi_{low}(t_2; 1)_n| - \log |\log \varphi_{low}(t_1; 1)_n|}{\log t_2 - \log t_1}.$$

Since we do not know if $\hat{\alpha}_{low}^0(t_1, t_2)_n$ is in the domain of attraction, we define the estimator of the parameter k as

$$\hat{k}_{low}(t_1, t_2)_n := 2 + \left\lfloor \hat{\alpha}_{low}^0(t_1, t_2)_n^{-1} \right\rfloor.$$

In the second step we use estimator $\hat{k}_{low} := \hat{k}_{low}(t_1, t_2)_n$ for the estimation of parameters H , α and σ . In particular, we get the following consistent estimators

$$\begin{aligned}\hat{H}_{low}(-p, \hat{k}_{low})_n &= \frac{1}{p} \log_2 \left(\frac{\sum_{i=2\hat{k}_{low}}^n |\Delta_{i,\hat{k}_{low}}^2 X|^{-p}}{\sum_{i=2\hat{k}_{low}}^n |\Delta_{i,\hat{k}_{low}}^1 X|^{-p}} \right), \\ \hat{\alpha}_{low}(\hat{k}_{low}; t_1, t_2)_n &= \frac{\log |\log \varphi_{low}(t_2; \hat{k}_{low})_n| - \log |\log \varphi_{low}(t_1; \hat{k}_{low})_n|}{\log t_2 - \log t_1}, \\ \tilde{\sigma}_{low}(\hat{k}_{low}; t_1, t_2)_n &= \left(-\log \varphi_{low}(t_1; \hat{k}_{low}) \right)^{1/\hat{\alpha}_{low}} / t_1 \|h_{\hat{k}_{low},1}\|_{\hat{\alpha}_{low}}.\end{aligned}$$

Next, we consider two-stage estimation procedure in the general case in high-frequency setting which is the same as in the low-frequency setting. For $p \in (0, 1/2)$ we compute $\hat{H}_{high}(-p)_n = \hat{H}_{high}(-p, 1)_n$ and, therefore, we can define the preliminary estimator of α by

$$\hat{\alpha}_{high}^0(p, p')_n = \phi^{-1} \left(\frac{V_{high}(f_{-p'}, \hat{H}_{high}(-p)_n)_n^p}{V_{high}(f_{-p}, \hat{H}_{high}(-p)_n)_n^{p'}} \right)$$

with

$$\phi(\hat{\alpha}_{high}^0(p, p')_n) := \frac{\left(2/\hat{\alpha}_{high}^0(p, p')_n \right)^{p-p'} a_{-p'}^{p'} \Gamma(p'/\hat{\alpha}_{high}^0(p, p')_n)^p}{a_{-p'}^p \Gamma(p/\hat{\alpha}_{high}^0(p, p')_n)^{p'}}$$

where $p, p' \in (0, 1/2)$ such that $p \neq p'$, and $V_{high}(f_{-p}, \hat{H}_{high}(-p)_n)_n$ is given in formula (3.4.5) with $k = 1$, $f_{-p}(x) = |x|^{-p}$ and preliminary estimator $\hat{H}_{high}(-p)_n$ for the parameter H . It is remarkable that $\phi(\cdot)$ is always invertible for all $p \neq p'$ (see Dang and Istas [21]). Consequentially, we can define the estimator of k in high-frequency setting by

$$\hat{k}_{high} := \hat{k}_{high}(p, p')_n = 2 + \left\lfloor \hat{\alpha}_{high}^0(p, p')_n^{-1} \right\rfloor.$$

Thus, consistent estimators of H , α and σ , in high-frequency setting are given by

$$\begin{aligned} \hat{H}_{high}(-p, \hat{k}_{high})_n &= \frac{1}{p} \log_2 \left(\frac{\sum_{i=2\hat{k}_{high}}^n \left| \Delta_{i, \hat{k}_{high}}^{n, 2} X \right|^{-p}}{\sum_{i=2\hat{k}_{high}}^n \left| \Delta_{i, \hat{k}_{high}}^{n, 1} X \right|^{-p}} \right), \\ \tilde{\alpha}_{high}(\hat{k}_{high}; t_1, t_2)_n &= \phi^{-1} \left(\frac{V_{high}(f_{-p'}, \hat{H}_{high}(-p, \hat{k}_{high})_n; \hat{k}_{high})_n^p}{V_{high}(f_{-p}, \hat{H}_{high}(-p, \hat{k}_{high})_n; \hat{k}_{high})_n^{p'}} \right), \\ \tilde{\sigma}_{high}(\hat{k}_{high}; p, p')_n &= \left(\frac{\tilde{\alpha}_{high}^{\alpha-p} V_{high}(f_{-p}, \hat{H}_{high}(-p)_n)_n}{2\Gamma(p/\tilde{\alpha}_{high})} \right)^{-\frac{1}{p}} / \|h_{\hat{k}_{high}, 1}\|_{\tilde{\alpha}_{high}}. \end{aligned}$$

Implementation in R

We introduce function `ContinEstim` for performing statistical inference according to Section 3.4 when $H - 1/\alpha > 0$.

```
ContinEstim(t1, t2, p, k, path, freq)
```

The function is basically comprised by simpler functions `alpha_hat`, `H_hat` and `sigma_hat` responsible for retrieving the corresponding parameters. `sigma_hat` is called using `tryCatch` as the former may return an error due to numerical integration in `Norm_alpha`.

General low-frequency estimation technique, described in Section 3.4 is implemented in `GenLowEstim`.

```
GenLowEstim(t1, t2, p, path, freq)
```

This estimator first sets a preliminary k to be equal to 1, and uses it to compute preliminary parameters H_0 and α_0 . Using these H_0 and α_0 , a new k is obtained through `2+floor(alpha.0^(-1))`, and then the new k is used for the same estimation procedure as in `ContinEstim`. This approach induces an effect, which does not exist in the case when `ContinEstim` is applied. When α is smaller than, or close to $2/N$, where N is the observed lfsm path length, the computational errors are more frequent. These extra errors occur when the preliminary estimation of k appears to exceed $N/2$, making it impossible to compute $\Delta_{i, \hat{k}_{low}}^2 X$ in statistic $\hat{H}_{low}(-p, \hat{k}_{low})_N$. In case of other sample path realizations $k < H + 1/\alpha$, and it is still possible to obtain the estimates which happen to converge to the true value $(\hat{H}, \hat{\alpha}, \hat{\sigma})$, because in this case one would be in the domain of attraction of Theorem 2.2 of [29]. Though, the limiting distribution is not stable anymore, and the rate of convergence depends on α and H . Real distributions of estimates in this case are left unexplored.

High-frequency estimator from the same section was implemented in `GenHighEstim`.

```
GenHighEstim<-function(p, p_prime, path, freq, low_bound=0.01, up_bound=4)
```

Estimate deterioration

Although the general high- and low-frequency estimators presented in Section 3.4 have important advantages, namely closed form expressions for distribution functions and non-suboptimal convergence rates, they also reveal two drawbacks in performance. Due to condition and error handling, the time performances of the general estimators are much worse than those of the continuous ones. On top of that, the plug-in estimators (because of their nature) have much less probability of obtaining an estimate at all. The main idea is as follows: the more statistics are used in a plug-in estimator, the higher the probability to stumble upon a numerical error during the estimation procedure. We illustrate this effect by the following experiment, wherein the general high- and low-frequency estimators are compared to the corresponding continuous ones. For each pair from a set of parameters (H, α) , `NmonteC` sample paths of the both frequencies were generated, and to each of them the relevant procedures `ContinEstim`, `GenLowEstim` and `GenHighEstim` were applied (see the code below). Then, the rates of successful computation results were computed. The result of estimation was considered “successful” if during the procedure all three parameters were obtained, no error occurred, and the estimates are meaningful, namely $(\hat{H}, \hat{\alpha}) \in (0, 1) \times (0, 2)$.

```
R> ##### Set of global parameters #####
R> library(rlfsm)

R> m<-45; M<-60; N<-200
R> p<-0.4; p_prime<-0.2
R> t1<-1; t2<-2; k<-2

R> NmonteC<-3e2
R> sigma<-0.3

R> ### Grid for alphas and Hs continuous case

R> by_hs<-0.05; by_als<-0.1
R> hs<-seq(0.5+by_hs, 1-by_hs, by=by_hs)
R> als<-seq(1+by_als, 2-by_als, by=by_als)
R> #####

R> test<-list()
R> dimns <-list(hs, als)

R> mtrx_contin_l<-matrix(data = NA, nrow = length(hs),
                        ncol = length(als), dimnames=dimns)
R> mtrx_contin_h<-matrix(data = NA, nrow = length(hs),
                        ncol = length(als), dimnames=dimns)
R> mtrx_gen_l<-matrix(data = NA, nrow = length(hs),
                     ncol = length(als), dimnames=dimns)
R> mtrx_gen_h<-matrix(data = NA, nrow = length(hs),
                     ncol = length(als), dimnames=dimns)

R> ##### A function for NA/ NaN / error filtering
R> # returns 1 if everything is OK
R> Errfilter<-function(res){
  b1<-ifelse(is.character(res), 0, 1)
  b2<-ifelse(length(grep('NA', res))>0, 0, 1)
  b3<-ifelse(length(grep('NaN', res))>0, 0, 1)
  b1*b2*b3
R> }
R> #####
```

```

R> ##### The experiment #####
R> for(ind_hs in (1:length(hs))) {
  R> for(ind_als in (1:length(als))) {
    R> res<-data.frame()

    R> res<-foreach (j_ind = 1:NmonteC, .combine = rbind, .packages='stabledist',
      .inorder=FALSE) %dopar% {

      R> pathL <- path(N=N,m=m,M=M,alpha=als[ind_als],H=hs[ind_hs],
        sigma=sigma,freq='L')
      R> pathH <- path(N=N,m=m,M=M,alpha=als[ind_als],H=hs[ind_hs],
        sigma=sigma,freq='H')

      R> ConEstLow<-ContinEstim(t1=t1,t2=t2,p=p,k=2,path=pathL$lfsm,freq='L')
      R> GenEstLow<-GenLowEstim(t1=t1,t2=t2,p=p,path=pathL$lfsm,freq='L')

      R> ConEstHigh<-ContinEstim(t1=t1,t2=t2,p=p,k=2,path=pathH$lfsm,freq='H')
      R> GenEstHigh<-GenHighEstim(p=p,p_prime=p_prime,path=pathH$lfsm,freq='H',
        low_bound=0.01,up_bound=2)

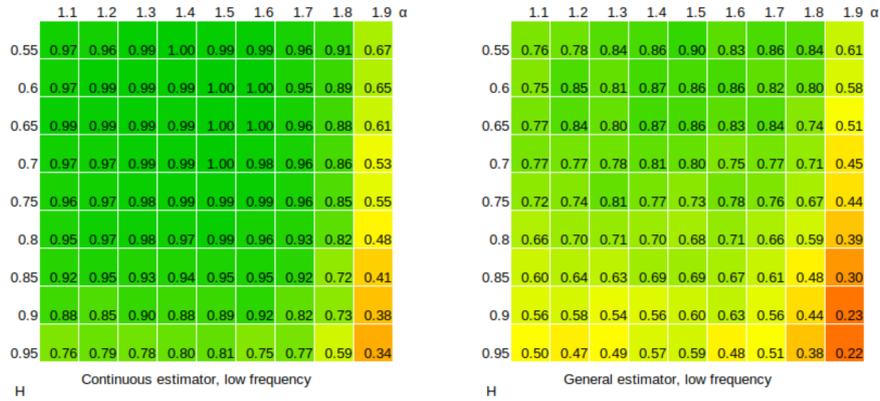
      R> rcol<-cbind(CEL=Errfilter(ConEstLow),
        GEL=Errfilter(GenEstLow),
        CEH=Errfilter(ConEstHigh),
        GEH=Errfilter(GenEstHigh))

      R> rcol
    }

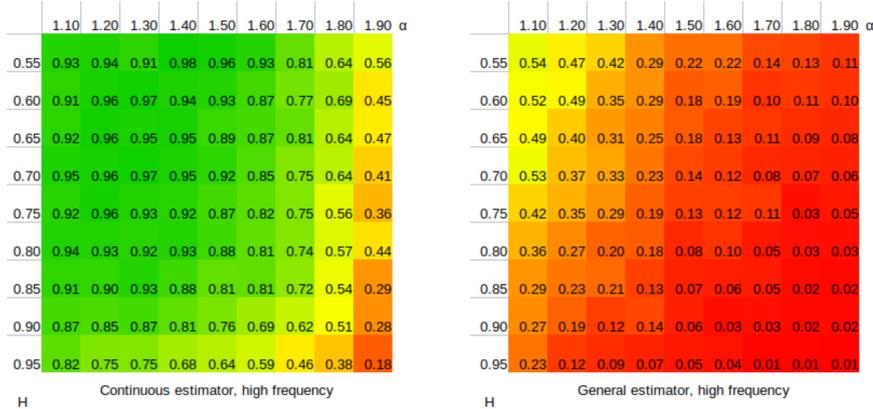
    R> suc_rate<-colSums(res)/NmonteC

    R> mtrx_cont_in_l[ind_hs,ind_als]<-suc_rate['CEL']
    R> mtrx_gen_in_l[ind_hs,ind_als]<-suc_rate['GEL']
    R> mtrx_cont_in_h[ind_hs,ind_als]<-suc_rate['CEH']
    R> mtrx_gen_in_h[ind_hs,ind_als]<-suc_rate['GEH']
  }
}

```



(a) Comparison of success rates for `ContinEstim` and `GenLowEstim`. Low frequency case. Path length $N=200$, number of sample paths $N_{\text{monteC}}=300$.



(b) Comparison of success rates for `ContinEstim` and `GenHighEstim`. High frequency case. Path length $N=200$, number of sample paths $N_{\text{monteC}}=300$.

Figure 3.12: Comparison of success rates of estimators

This experiment shows that in both high- and low-frequency cases `ContinEstim` gives much better precision than the corresponding general estimator. The outcome is rigorous in low-frequency technique since `ContinEstim` and `GenLowEstim` have the same set of tuning parameters. On the other hand, the high-frequency estimators have non-coinciding parameter sets, and thus, without fine tuning, the result is merely intuitive. One could

observe (Figures 3.12a and 3.12b) that in general estimation near the boundaries of the interval $(\hat{H}, \hat{\alpha}) \in (0, 1) \times (0, 2)$ produces more errors, which is partly due to the fact that near the boundaries it is easier to obtain an estimate outside the interval. Such an estimate is removed by `Errfilter` function in the experiment.

Zones with different convergence regimes in the low-frequency case

In order to show how the general low-frequency estimation works in practice, we perform a numerical experiment. We set a constant σ and choose two sets of parameters- one for α and one for H . Then, for each combination of them a number $N_{mc} = 500$ of sample paths is created. All path lengths are set to a constant $N = 1000$. To each path we apply several statistics. One of them is `k_new<-2+floor(alpha_0*(-1))` where `alpha_0` is obtained via `alpha_hat` with parameters `k=1,freq='L'` plugged-in. This provides us simulated distribution of \hat{k}_{low} (Figure 3.13). Also, we fix a set `k_ind = seq(1,8,by=1)` and, given a path, for each of these `k`'s extract statistics $\varphi_{low}(t, k = k_{ind})_n$ and $\hat{\alpha}_{low}(t_1, t_2; k = k_{ind})_n$, see Figures 3.14 and 3.15.

Three regimes of performance of `GenLowEstim` (read, the general low-frequency estimator $\hat{\alpha}_{low}(k, t_1, t_2)_n$) are observed. To a large extent, only parameter α determines which regime is in presence.

Due to small variance of $\hat{\alpha}_{low}^0(t_1, t_2)_n$ (Figure 3.15), when $\alpha \in (1, 2)$ the estimation $\hat{k}_{low}(t_1 = 1, t_2 = 2)_n$ returns 2 except from the boundaries, where edge effects are observed. This results in the fact that in cases when statistics $\hat{k}_{low}(1, 2)_n$ can be computed without stumbling on numerical errors performances of `GenLowEstim` and low frequency `ContinEstim` are the same. At the same time, statistic $\hat{\alpha}_{low}(k, t_1, t_2)_n$ is not far from its limit value for $k < 3$, that's why the parameter estimation of the LFSM is technically possible by `ContinEstim` and `GenLowEstim` at such length of the sample path.

When α is near 1 there is a transition between the regime with values of $\hat{k}_{low}(1, 2)_n$ concentrated at point $k = 2$, and the regime where $\hat{k}_{low}(1, 2)_n$ is highly dispersed. This shift is characterized by only two values of $\hat{k}_{low}(1, 2)_n$: 2 and 3. Such behavior of the estimated order of increments is due to the fact that when $\alpha^{-1} \in N$

$$\mathbb{P}(\hat{k}_{low} = 2 + \alpha^{-1}) \rightarrow \lambda \quad \text{and} \quad \mathbb{P}(\hat{k}_{low} = 1 + \alpha^{-1}) \rightarrow 1 - \lambda$$

for some constant $\lambda \in (0, 1)$, see [29], Section 4.1. Surprisingly, λ is close to 0.5 throughout the whole set of H 's (Figure 3.13). There are no $\hat{k}_{low}(1, 2)_n$ higher than 3 observed because the preliminary estimation of α is still quite precise as one can see from the middle row on Figure 3.15. After obtaining $\hat{k}_{low}(1, 2)_n$ equal to either 2 or 3, $\hat{\alpha}_{low}(\hat{k}_{low}(1, 2)_n, t_1, t_2)_n$ is computed again quite precisely, but worse than in the continuous case.

At $\alpha < 1$ $\hat{\alpha}_{low}(k, t_1, t_2)_n$ has high variance regardless of what `k` is chosen, therefore different values are obtained when computing $\hat{k}_{low}(1, 2)_n$. These values plugged-in to $\hat{\alpha}_{low}(k, t_1, t_2)_n$ produce again very dispersed estimates of parameter α . This mechanism explains why $\tilde{\alpha}_{low}$ has higher variance in discontinuous case ($H - 1/\alpha < 0$) than in continuous (see the numerical study in Section 5 in [29]).

The way $\tilde{\alpha}_{low}$ behaves could be explained using pic.(3.14), where φ_n and $V_{low}(\psi_t, k)_n$ are plotted. Cases wherein $\hat{\alpha}_{low}$ performs poorly coincide with ones wherein φ_n and $V_{low}(\psi_t, k)_n$ are significantly distant from each other, so convergence $V_{low}(\psi_t, k)_n \xrightarrow{a.s.} \varphi_n(t; k)$ isn't observed at the given length of sample paths, which ruins the whole idea of (σ, α) estimation. Of course, this effect doesn't affect H -estimation because it is based on ratio statistic, which has a different form.

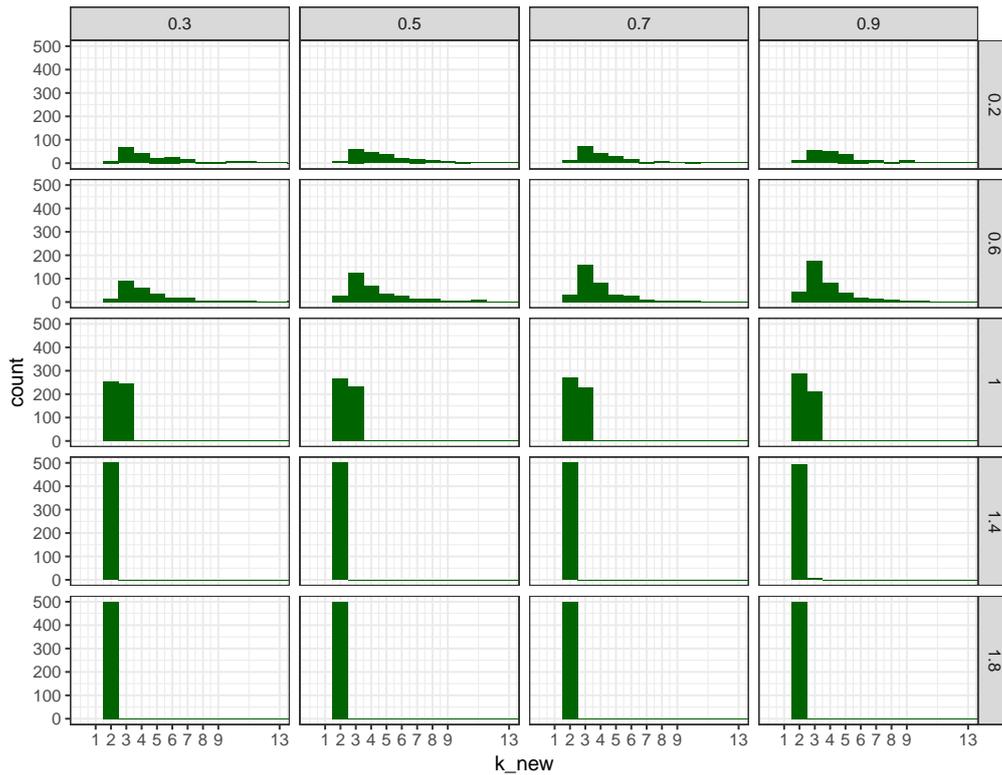


Figure 3.13: Histograms of preliminary estimations of k , $\hat{k}_{low}(1, 2)_n$. α 's are on vertical labels, H 's- on horizontal.

3.5 S4 classes for Lèvy-driven motions

Here we describe a simple S4 system (a short introduction to S4 classes is given in [48], Chapter OO field guide) that could be used to simplify manipulations with the two types of observations of the linear fractional stable motion. Additionally, we present a possible way to extend the system so that it encompasses more general stochastic processes. The system aims to be helpful in

- passing "attributes" (frequency, σ , α , H) from objects to functions automatically (without additional developer's efforts).
- hiding complicated details of interfaces from users.
- using generics to protract functions on different objects by means of inheritance. For instance, plotting function written for `lfsm` could be used for other types of stochastic integral.

Classes for simulated lfsm

Here we describe the least general classes- `SimulatedLfsmLow` and `SimulatedLfsmHigh`, objects of which are obtained by simulating low- and high-frequency linear fractional stable motions. Figure 3.16 shows their internal structure. Roughly speaking, these classes were designed to contain minimum information that could fully describe a simulated LFSM

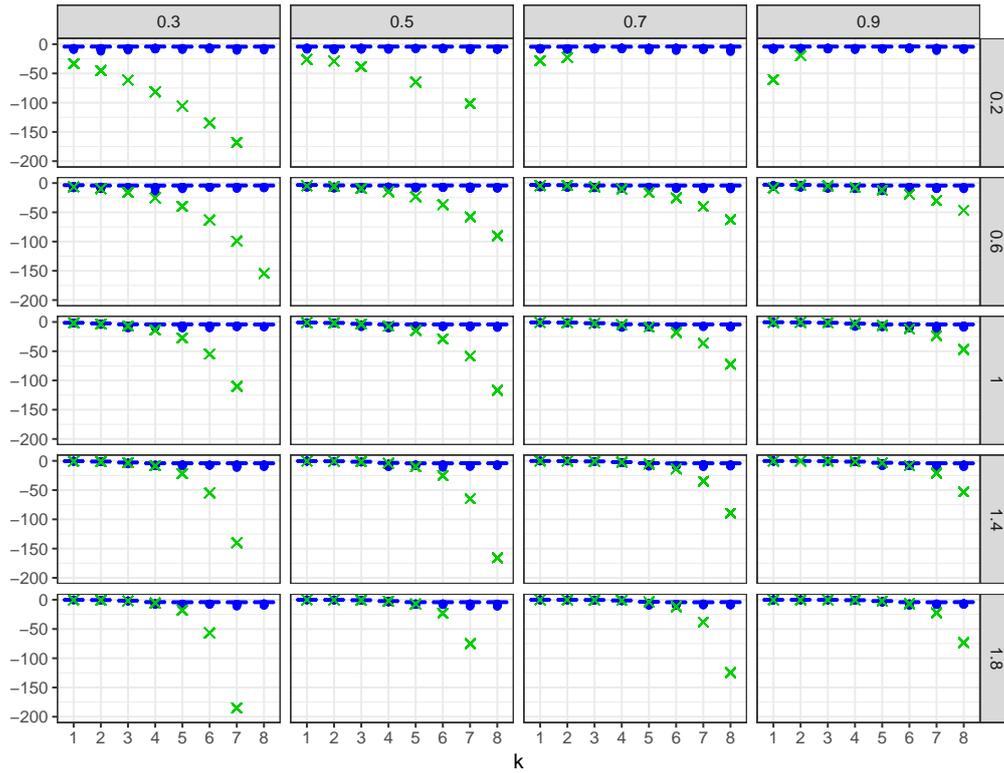


Figure 3.14: Comparison of the real $\varphi_n(t = 1; k)$ and the one estimated via $V_{low}(\psi_{t=1}, k)_n$ on the logarithmic scale. α 's are on vertical labels, H 's- on horizontal. The lower and upper box sides correspond to the 25th and 75th percentiles.

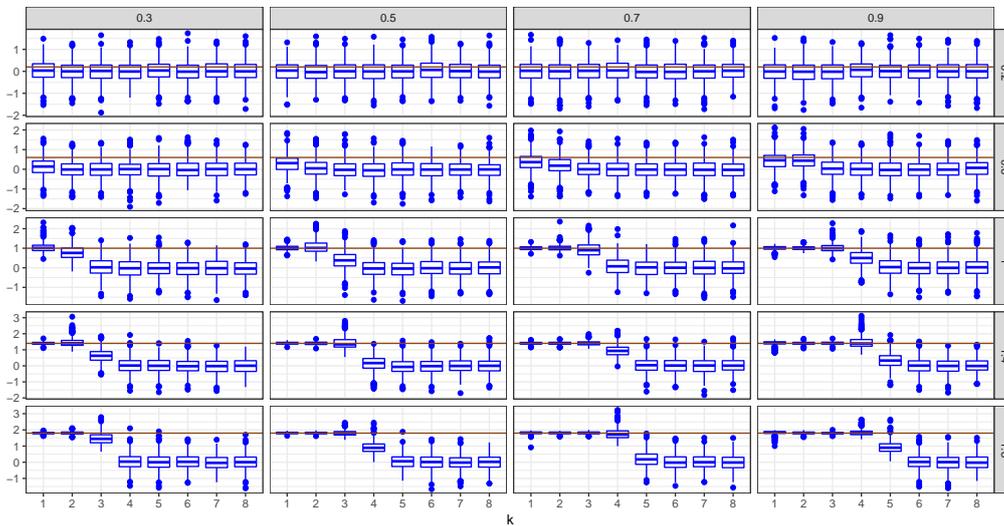


Figure 3.15: Convergence of $\hat{\alpha}_{low}(k, t_1, t_2)_n$ to the real α (red line) for different k . α 's are on vertical labels, H 's- on horizontal. The lower and upper box sides correspond to the 25th and 75th percentiles.

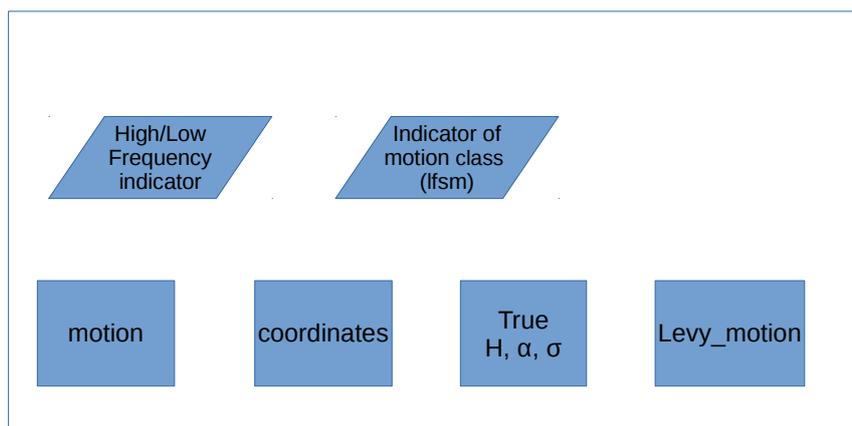


Figure 3.16: Structure of the classes of simulated lfsm. Frequency indicator and indicator of process type are included in the class name, whilst motion, coordinates, parameters for which the path was simulated and the Lévy motion are written in the slots.

path. Indicators of frequency and a process type are included in the name of a class, which is supposed to make a method dispatch more straightforward, without additional condition blocks. Moreover, all generic functions distinct high- and low-frequency schemes of all types with the help of class names. The same holds for motion types. Parameters H, α, σ , as well as Lévy motion, coordinates and the lfsm itself are written in corresponding slots.

Examples

In the following example we see how an instance of class `SimulatedLfsmLow` is created and then plotting and inference is performed using generic functions `plot` and `ContinInfer`.

```

N<-3000; m<-65; M<-300
sigma<-0.3; alpha<-1.8; H<-0.8
p<-0.4; t1<-1; t2<-2; k<-2

# Make an object of S4 class SimulatedLfsmLow
R> List <- path(N,m,M,alpha,H,sigma,freq='L',disable_X=FALSE,seed=3)

# Make an object of parameters
R> prmts<-new("AlpaHSigma",alpha=List$pars[['alpha']],
R> H=List$pars[['H']],sigma=List$pars[['sigma']])

R> X_sim <- new("SimulatedLfsmLow", Process = List$lfsm,
               coordinates = List$coordinates, pars = prmts,
               levy_motion = List$levy_motion)

# plot the motion
R> plot(X_sim)

# structure of the instance
R> str(X_sim)

Formal class 'SimulatedLfsmLow' [package ".GlobalEnv"] with 4 slots
 ..@ pars      :Formal class 'AlpaHSigma' [package ".GlobalEnv"] with 3 slots
 .. .. ..@ alpha: num 1.8
 .. .. ..@ H    : num 1.8

```

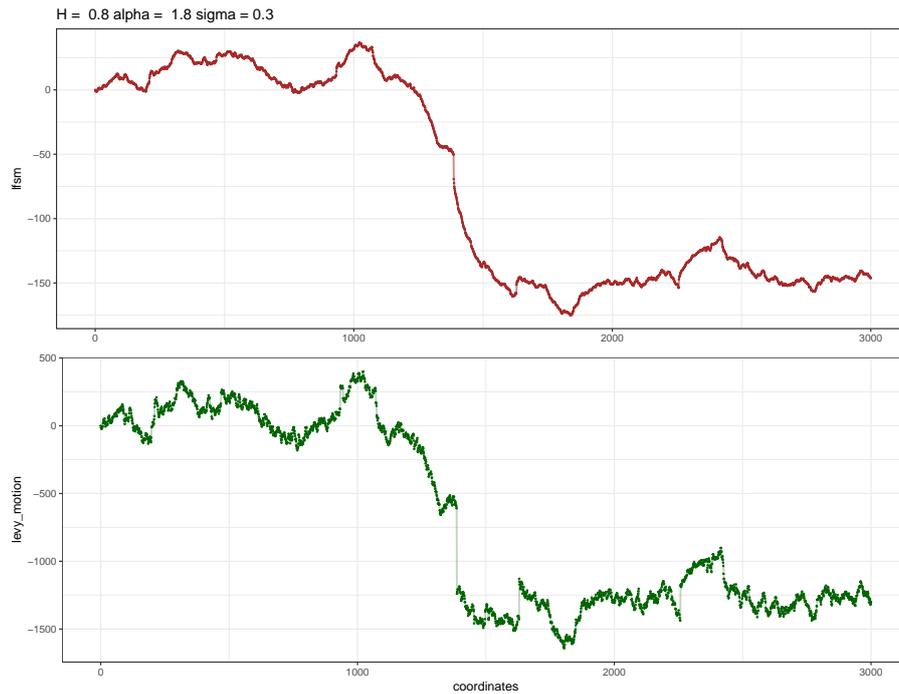


Figure 3.17: Output of plot method for simulated lfsm

```

.. .. ..@ sigma: num 1.8
..@ levy_motion: num [1:3497] 0 -15 -19.8 -21.2 -24.1 ...
..@ Process      : num [1:3497] 0 -0.542 -0.912 -1.12 -1.276 ...
..@ coordinates: int [1:3497] 0 1 2 3 4 5 6 7 8 9 ...

R> ContinInfer(x=X_sim,t1=t1,t2=t2,k=k,p=p)

$alpha
[1] 1.870217

$H
[1] 0.8314528

$sigma
[1] 0.3227219

```

In this example, the plot function takes almost no effort, compared to the similar one from Section 3.3, which is due to the fact, that there has been a method defined for generic `plot` and object `SimulatedLfsmLow`. The last function, `ContinInfer`, is a generic which has a registered method for class `StochasticProcLow`, general stochastic processes in low-frequency setting. Since `SimulatedLfsmLow` inherits from `StochasticProcLow`, the generic dispatched this method and performed statistical inference. `ContinInfer` was designed to perform inference according to Theorem 3.1 from [29] and is based on R function `ContinInfer`. One can see that `plot` (and, less obviously, `ContinInfer`) used 'Low' from the name of the class to perform computations.

Computational details

The present article corresponds to version 0.3.1 of package `rlfsm` which is available on CRAN.

Acknowledgments

The authors acknowledge financial support from the project "Ambit fields: probabilistic properties and statistical inference" funded by Villum Fonden. Stepan Mazur acknowledges financial support from the internal research grants at Örebro University, and from the project "Models for macro and financial economics after the financial crisis" (Dnr: P18-0201) funded by Jan Wallander and Tom Hedelius Foundation. The authors would also like to thank Prof. Mark Podolskij for valuable discussions.

3.6 Appendix

Here you can see the code for the experiment in (3.4).

```
#### Set of global parameters ####
library(rlfsm)
library(doParallel)
library(foreach)
library(gridExtra)
library(ggplot2)

registerDoParallel()

m<-45; M<-90; N<-1e3
p<-0.4; p_prime<-0.2
t<-1; t1<-1; t2<-2
fr<-'L'; NmonteC<-5e2

sigma<-0.3

## %%%%%%%%%%%%%%% ##

# Grid for (alpha, H)
hs<-c(0.3,0.5,0.7,0.9)
by_als<-0.4
als<-seq(0.2,1.8,by=by_als)

# Sample path creation

path_set<-list()

for(ind_hs in (1:length(hs))) {

  path_set_als<-list()
  for(ind_als in (1:length(als))) {

    S_Path<-paths(N_var=NmonteC,N=N,m=m,M=M,alpha=als[ind_als],
                 H=hs[ind_hs],sigma=sigma,freq=fr,disable_X=FALSE,
                 levy_increments=NULL,parallel = TRUE)

    ll<-list(H=hs[ind_hs], alpha=als[ind_als], S_Path=S_Path)
    path_set_als[[ind_als]]<-ll
  }
  path_set[[ind_hs]]<-path_set_als
}

# Phi stats vs exp
phis_on_k<-function(p_data){
```

```

S_paths<-p_data$S_Path
alpha<-p_data$alpha
H<-p_data$H

data<-foreach(SP_ind = (1:NmonteC), .combine = rbind) %dopar% {

  S_Path<-S_paths[,SP_ind]
  d<-foreach(k_ind = seq(1,8,by=1), .combine = rbind) %do% {

    phi_vect<-phi(t=t1,k=k_ind,path=S_Path,H=H,freq='L')
    exp_vect<-tryCatch(
      exp(-(abs(sigma*Norm_alpha(h_kr,alpha=alpha,
        k=k_ind,r=1,H=H,l=0)$result)^alpha)),
      error=function(c) NA)
    al_vect<-alpha_hat(t1=1,t2=2,k=k_ind,path=S_Path,H=H,freq='L')

    c(sample_num=SP_ind, k=k_ind,
      phi_exper=phi_vect, phi_theor=exp_vect,
      alpha=alpha, H=H, alpha_est=al_vect)
  }
  d
}

data

}

# k_new estimates
Monte_k<-function(p_data){

  S_paths<-p_data$S_Path
  alpha<-p_data$alpha
  H<-p_data$H

  data<-foreach(SP_ind = (1:NmonteC), .combine = rbind) %dopar% {

    alpha_0<-alpha_hat(t1=t1,t2=t2,k=1,path=S_paths[,SP_ind],H=NULL,freq='L')
    if(alpha_0<=0) k_new<-NA else k_new<-2+floor(alpha_0^(-1))
    c(k_new=k_new, alpha=alpha, H=H)
  }
  data
}

# Computing on the sample paths
pl<-data.frame()
pk<-data.frame()

for(ind_hs in (1:length(hs))) {
  for(ind_als in (1:length(als))) {

    pl<-rbind(pl,phis_on_k(path_set[[ind_hs]][[ind_als]]))
    pk<-rbind(pk,Monte_k(path_set[[ind_hs]][[ind_als]]))
  }
}

# Plotting Phi on k

pl_log<-pl
pl_log$phi_theor<-log(pl_log$phi_theor)
pl_log$phi_exper<-log(pl_log$phi_exper)

pl_log$k<-factor(pl_log$k)
ggp <- ggplot(pl_log) +
  geom_boxplot(aes(x = k, y = phi_exper), color='blue') +
  geom_point(aes(x = k, y = phi_theor), inherit.aes = FALSE,

```

```

      size = 2, shape = 4, color='green3') +
  scale_y_continuous(limits = c(-200,0)) +
  facet_grid(alpha~H, scales = "free") +
  ggplot2::theme_bw() + labs(x = "k", y=NULL) #labs(x = "k", y = "phi")
ggp
ggsave(filename='phi_on_k.pdf', plot = ggp)

# Plotting alpha_hat on k

pl$k<-factor(pl$k)
ggp <- ggplot(pl) +
  geom_boxplot(aes(x = k, y = alpha_est), color='blue') +
  geom_hline(aes(yintercept = alpha), colour = "chocolate4") +
  facet_grid(alpha~H, scales = "free") +
  ggplot2::theme_bw() + labs(x = "k", y=NULL)
ggp
ggsave(filename='alpha_est_on_k.pdf', plot = ggp)

# Plotting histograms of k_new

ggk<-ggplot(data=pk, aes(k_new)) +
  geom_histogram(binwidth = 1, fill = 'darkgreen', center=0) +
  facet_grid(alpha~H) + coord_cartesian(xlim = c(0, 13)) +
  scale_x_continuous(breaks=c(1:9,13)) +
  theme_bw()
ggk
ggsave(filename='hist_of_k.pdf', plot = ggk)

```

Chapter 4

Simulation of multidimensional stochastic integrals with convolving kernels driven by Lévy basis

DMITRY OTRYAKHIN, MARK PODOLSKIJ

4.1 The model

In this Chapter, we are going to discuss methods for simulation of a certain type of stochastic integrals

$$Y_t = \int_{\mathbb{R}^d} g(t-s)\Lambda(ds), \text{ where} \quad (4.1.1)$$

- $d \in \mathbb{N}$.
- $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is an integrable function in the sense of Definition 1.1.6.
- Λ is a Lévy basis on \mathbb{R}^d , see (1.11).

We are interested in kernel functions having two specific characteristics- singularity near zero and a long tail at infinity. The function g can have one-dimensional values without loss of generality, because multidimensional functions might be split by components and integrated separately. We suggest using a combination of two different methods for computing integrals of type (4.1.1): the one developed by Cohen et al. [17] should be used near singularity points, and the method based on the convolution theorem- outside that area. At this stage, the two algorithms are not combined together due to lack of knowledge of some joint distributions. The idea to mix several techniques while computing a stochastic integral comes from earlier papers by Bennedsen et al. [11] and Heinrich et al. [23] where the authors applied hybrid schemes to Gaussian fields.

In this chapter, we consider Lévy drivers, although we remark that the technique based on the convolution theorem allows more general random measures and even different types of integrals. The choice of the driving measure determines the way numerical errors are described. Throughout the chapter the dimensionality is set $d = 2$ for simplicity of notation, but similar results hold for any d .

4.2 Shot-noise approximation of non-Gaussian fractional fields

This section was derived from [17] until the last paragraph.

Notation. Let $v \in \mathbb{N}$, $(V_v)_{v \geq 1}$ and $(U_v)_{v \geq 1}$ be independent sequences of random variables. We assume that $(U_v, V_v)_{v \geq 1}$ is independent of $(T_v)_{v \geq 1}$.

- $(V_v)_{v \geq 1}$ is a sequence of i.i.d. random variables with common law $\nu(dv)/\nu(\mathbb{R})$.
- $(U_v)_{v \geq 1}$ is a sequence of i.i.d. random variables such that U_1 is uniformly distributed on the unit sphere S^{d-1} of the Euclidean space \mathbb{R}^d .
- c_d is the volume of the unit ball of \mathbb{R}^d .

When $\nu(\mathbb{R}) < \infty$, the integral (4.1.1) can be represented by a shot noise series.

Proposition 4.2.1 (Proposition 4.1, [17]). *Under some assumptions, for every $t \in \mathbb{R}^d$, the series*

$$Y_t^{ser} = \sum_{v=1}^{+\infty} g\left(t, \left(\frac{T_v}{c_d \nu(\mathbb{R})}\right)^{1/d} U_v\right) V_v \quad (4.2.2)$$

converges almost surely. Furthermore,

$$\{Y_t : t \in \mathbb{R}^d\} \stackrel{d}{=} \{Y_t^{ser} : t \in \mathbb{R}^d\} \quad (4.2.3)$$

In simulations a truncated version of sum (4.2.2) is used:

$$Y_{t, \Upsilon}^{ser} = \sum_{n=1}^{\Upsilon} g\left(t, \left(\frac{T_v}{c_d \nu(\mathbb{R})}\right)^{1/d} U_v\right) V_v, \quad (4.2.4)$$

For this truncation, the next theorem determines the error rate with the help of Markov's inequality:

Theorem 4.2.2. *Let $t \in \mathbb{R}^d$. Assume that*

$$\forall \xi \neq 0, \quad |g(t, \xi)| \leq \frac{C}{\|\xi\|^\beta} \quad (4.2.5)$$

where $\beta > d/2$ and $C > 0$. Furthermore, assume there exists $r \in (d/\beta, 2]$ such that $\mathbb{E}(|V_1|^r) < +\infty$. Then, for every $\epsilon \in (0, \beta/d - 1/r)$, almost surely,

$$\sup_{\Upsilon \geq 1} \Upsilon^\epsilon |Y_t^{ser} - Y_{t, \Upsilon}^{ser}| < +\infty \quad (4.2.6)$$

Moreover, for every $\Upsilon > r\beta/d$,

$$\mathbb{E}(|Y_t^{ser} - Y_{t, \Upsilon}^{ser}|^r) \leq C(r, \beta) \frac{D(\Upsilon, r, \beta)}{\Upsilon^{r\beta/d - 1}} \quad (4.2.7)$$

where

$$D(\Upsilon, r, \beta) = \frac{\Gamma(\Upsilon + 1 - r\beta/d)(\Upsilon + 1)^{r\beta/d}}{\Gamma(\Upsilon + 1)} \quad (4.2.8)$$

and

$$C(r, \beta) = \frac{dC^r(c_d \nu(\mathbb{R})^{r\beta/d} \mathbb{E}(|V_1|^r))}{r\beta - d} \quad (4.2.9)$$

When $\nu(\mathbb{R}) = \infty$, the integral (4.1.1) is approximated by a superposition of a shot-noise series and a Gaussian field. Note that the computational complexity of this algorithm is proportional to $N \times \Upsilon$, where N is the number of values of (4.1.1) to compute, and Υ is the number of jumps. On a rectangular grid, $N = \prod_{j=1}^d N_j$, where N_j denotes the number of projections of the points. There are two reasons why it is not always possible to use the algorithm introduced by Cohen et al. [17]: first, when the process has too many jumps, the computational complexity is too high to be used in practice, and second, it comes with relatively strong assumptions.

4.3 Review of the multidimensional discrete Fourier transform

A meticulous review on multidimensional Fourier transforms could be found in a book by Dan E. Dudgeon [20]. Here, only the key results are given. Let $\tilde{x}(n_1, \dots, n_d)$ be a d -dimensional rectangular periodic sequence with periodicity (N_1, \dots, N_d) , i.e.

$$\tilde{x}(n_1, \dots, n_d) = \tilde{x}(n_1 + N_1, \dots, n_d + N_d) \quad (4.3.10)$$

with values in \mathbb{R} . For such a sequence discrete Fourier transform is defined as

$$\tilde{X}(k_1, \dots, k_d) = \sum_{n_1=0}^{N_1-1} \cdots \sum_{n_d=0}^{N_d-1} \tilde{x}(n_1, \dots, n_d) \cdot \exp\left(-\sum_{r=1}^d \frac{2\pi i k_r n_r}{N_r}\right), \quad (4.3.11)$$

for all integer k_j . Note, that $\tilde{X}(k_1, \dots, k_d)$ is periodic in k_j with a period N_j . \tilde{X} can be interpreted as a superposition of $2 \times d$ -dimensional waves (2 due to complex space). The inverse transform is

$$\tilde{x}(n_1, \dots, n_d) = \frac{1}{\prod_{r=1}^d N_r} \sum_{k_1=0}^{N_1-1} \cdots \sum_{k_d=0}^{N_d-1} \tilde{X}(k_1, \dots, k_d) \cdot \exp\left(\sum_{j=1}^d \frac{2\pi i k_j n_j}{N_j}\right). \quad (4.3.12)$$

The convolution theorem states that for 2 periodic sequences $\tilde{x}(n_1, \dots, n_d)$ and $\tilde{a}(n_1, \dots, n_d)$ with discrete Fourier transforms $\tilde{X}(k_1, \dots, k_d)$ and $\tilde{A}(k_1, \dots, k_d)$ it holds that their linear convolution

$$\tilde{y}(n_1, \dots, n_d) := \tilde{x} * \cdots * \tilde{a} = \sum_{m_1=0}^{N_1-1} \cdots \sum_{m_d=0}^{N_d-1} \tilde{a}(m_1, \dots, m_d) \cdot \tilde{x}(n_1 - m_1, \dots, n_d - m_d) \quad (4.3.13)$$

is equal to the inverse Fourier transform of the $\tilde{Y}(k_1, \dots, k_d) = \tilde{A}(k_1, \dots, k_d) \times \tilde{X}(k_1, \dots, k_d)$. Computational complexity of multidimensional fast Fourier transform is

$$O\left(\prod_{k=1}^d N_k \log \left[\prod_{j=1}^d N_j \right]\right) \quad (4.3.14)$$

4.4 Algorithms based on multidimensional fast Fourier transforms.

The idea in its one-dimensional form was used in [40]. Here we proceed as follows: the integral is approximated by a sum, and the error is computed accordingly to this approximation. Then, fast Fourier transform in conjunction with multidimensional convolution

theorem is used to fasten the sum computation. In order to capture two features of the kernel, described in Section 4.1, we suggest splitting of the kernel domain into three pieces, subsequent computing of (4.1.1) separately on the first two, and truncating the third one, Figure 4.1. We choose rectangular grid for sampling the Lévy basis, then the truncated

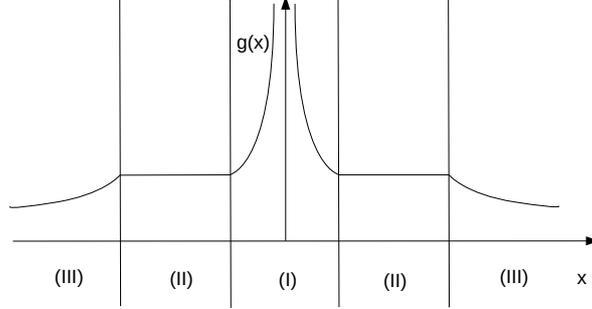


Figure 4.1: Example of splitting the domain of the kernel function into 3 zones. $x \in \mathbb{R}^1$. The function $g(x) \rightarrow \infty$ as $x \rightarrow 0$, zone (I). Zone (II) contains its Hölder continuous part. In zone (III) $g(x) \rightarrow 0$ as $x \rightarrow \infty$.

integral (4.1.1) is written as

$$Y_t^{trunc} = \int_{t_1-M_1}^{t_1+M_1} \int_{t_2-M_2}^{t_2+M_2} g(t-s) \Lambda(ds), \quad M_i \text{ are truncation parameters.} \quad (4.4.15)$$

We approximate it as follows:

$$\Sigma_{M_1, M_2, m}(t) = \sum_{k_1=t_1+1-mM_1}^{t_1+mM_1} \sum_{k_2=t_2+1-mM_2}^{t_2+mM_2} g\left(t - \left[\frac{k_1}{m}, \frac{k_2}{m}\right]\right) \times \Delta\Lambda\left(\frac{k_1}{m}, \frac{k_2}{m}\right) \quad (4.4.16)$$

and define

$$I^{k_1, k_2}(t) = \int_{k_1-1}^{k_1} \int_{k_2-1}^{k_2} g(t - \xi/m) d\Lambda\left(\frac{\xi_1}{m}; \frac{\xi_2}{m}\right) \quad (4.4.17)$$

$$\Sigma_{M_1, M_2, m}^{k_1, k_2}(t) = g\left(t - \left[\frac{k_1}{m}, \frac{k_2}{m}\right]\right) \times \Delta\Lambda\left(\frac{k_1}{m}, \frac{k_2}{m}\right),$$

where the former is the restriction of the integral (4.4.15) to the segment $(k_1 - 1, k_1) \times (k_2 - 1, k_2)$:

$$Y_t^{trunc} = \sum_{k_1=t_1+1-mM_1}^{t_1+mM_1} \sum_{k_2=t_2+1-mM_2}^{t_2+mM_2} I^{k_1, k_2}(t), \quad (4.4.18)$$

where m is a number of observations between neighboring indexes on one axis, e.g. M_1 and $M_1 - 1$. This parameter controls precision. On one rectangular segment of the area $(k_1 - 1, k_1) \times (k_2 - 1, k_2)$ the approximation error is

$$I^{k_1, k_2}(t) - \Sigma_{M_1, M_2, m}^{k_1, k_2}(t) = \int_{k_1-1}^{k_1} \int_{k_2-1}^{k_2} \left(g(t - \xi/m) - g\left[t - \left(\frac{k_1}{m}; \frac{k_2}{m}\right)\right] \right) d\Lambda\left(\frac{\xi_1}{m}; \frac{\xi_2}{m}\right), \quad (4.4.19)$$

After a change of variables $t - k = \tau$, (4.4.16) can be expressed as

$$\Sigma_{M_1, M_2, m}(t) = \sum_{\tau_1 = -mM_1}^{mM_1-1} \sum_{\tau_2 = -mM_2}^{mM_2-1} g(\tau) \times \Delta\Lambda(t - \tau) \quad (4.4.20)$$

and after another change of variables, $\tau = \hat{\tau} - \tau'$, this expression becomes

$$\begin{aligned} \Sigma_{M_1, M_2, m}(t) &= \sum_{\hat{\tau}_1=0}^{(2mM_1-1)} \sum_{\hat{\tau}_2=0}^{(2mM_2-1)} g(\hat{\tau} - \tau') \times \Delta\Lambda(t - \hat{\tau} + \tau') = \\ &= \sum_{\hat{\tau}_1=0}^{(2mM_1-1)} \sum_{\hat{\tau}_2=0}^{(2mM_2-1)} \tilde{g}(\hat{\tau}) \times \Delta\tilde{\Lambda}(t - \hat{\tau}) \end{aligned} \quad (4.4.21)$$

One can see that integral approximation (4.4.21) has the form of discrete convolution. Thus, we can apply the convolution theorem and DFT transforms.

FFT-based algorithm for simulations

- Generate $m^2(2M_1 + N_1)(2M_2 + N_2)$ independent Lévy random variables $\Lambda(k_1; k_2)$, where $k_j \in [0, 2m(M_j + N_j) - 1]$
- Compute $a_{k_1, k_2} = \tilde{g}\left(\frac{k_1}{m}, \frac{k_2}{m}\right)$ for $k_j \in [0, 2mM_j - 1]$ and pad the sequence a_{k_1, k_2} with zeros to make it of the same size as $\Lambda(k_1; k_2)$
- Use DFT to compute $DFT(a)$ and $DFT(\Lambda)$
- Compute inverse DFT of $DFT(a) \times DFT(\Lambda)$
- keep only elements with indexes $[mM_1, mM_1 + mN_1] \times [mM_2, mM_2 + mN_2]$

4.5 α -stable drivers

Although the method based on the convolution theorem doesn't restrict us to use any specific sort of the driver in the integral (4.1.1), the type of the driver still affects the error produced in the approximation. A very simple in terms of error computation class of drivers is α -stable Lévy processes, which is going to be discussed in this section. The value of error (4.4.19) is $S\alpha S$ distributed with scaling parameter

$$\|I^{k_1, k_2}(t) - \Sigma_{M_1, M_2, m}^{k_1, k_2}(t)\|_\alpha^\alpha = \int_{k_1-1}^{k_1} \int_{k_2-1}^{k_2} \left| g(t - \xi/m) - g\left[t - \left(\frac{k_1}{m}; \frac{k_2}{m}\right)\right] \right|^\alpha ds \quad (4.5.22)$$

Points of singularity in case of α -stable drivers

Let us consider (4.1.1) with a one-dimensional Lévy basis (multidimensional setting is similar). When the driving basis is $S\alpha S$ the integral $\int g(t-s)\Lambda_{s\alpha S}(ds)$ is defined if $\int_0^\infty |g(t-s)|^\alpha ds < \infty$, which implies that $\int_{Const}^t (t-s)^{\alpha\phi} ds < \infty$, which, in turn entails the condition $\phi > -1/\alpha$ in the case $g(x) \sim x^\phi$ near zero, $\phi < 0$.

If we are to compute the integral (4.1.1) value at a single point, and truncate the kernel function near zero, it will yield an error which is distributed according to the symmetric α -stable law:

$$Y_t - Y_t^{tr} \stackrel{d}{=} \|g^{tr}\|_\alpha^\alpha \times S\alpha S(1), \quad (4.5.23)$$

where $S\alpha S(1)$ stands for a symmetric α -stable random variable with the scale parameter equal to 1.

$$\|g^{tr\delta}\|_\alpha^\alpha = \int_{t-\delta}^t |g(t-s)|^\alpha ds \sim \int_0^\delta s^{\phi\alpha} ds \quad (4.5.24)$$

up to multiplying by a constant. Therefore, if the kernel parameter ϕ is close to $-1/\alpha$, then the integral (4.5.24) has a high value (up to infinity), yielding a high value of error (4.5.23).

Hölder continuous function

Functions of this type cannot have a point on a compact where they converge to infinity, but they are useful to describe long tails. On the other hand, the series representation requires a power-law behavior, which might not be the case, or might produce a large error. The following theorem is proposed to deal with this situation.

Proposition 4.5.1. *For a κ -Hölder continuous kernel function $g(s)$ the scaling of the error on a square segment*

$$\|I^{k_1, k_2}(t) - \Sigma_{M_1, M_2, m}^{k_1, k_2}(t)\|_\alpha^\alpha \leq 2^{\frac{\alpha\kappa}{2}} C m^{-(2+\alpha\kappa)} \quad (4.5.25)$$

Also, the scaling of the total error

$$\|Y_t^{trunc} - \Sigma_{M_1, M_2, m}(t)\|_\alpha^\alpha \leq 2^{2+\frac{\alpha\kappa}{2}} C M_1 M_2 m^{-\alpha\kappa}. \quad (4.5.26)$$

Proof.

$$\begin{aligned} \|I^{k_1, k_2}(t) - \Sigma_{M_1, M_2, m}^{k_1, k_2}(t)\|_\alpha^\alpha &\leq C \int_{k_1-1}^{k_1} \int_{k_2-1}^{k_2} \left\| (t - \xi/m) - \left[t - \left(\frac{k_1}{m}; \frac{k_2}{m} \right) \right] \right\|^{\alpha\kappa} d\xi = \\ &C \int_{k_1-1}^{k_1} \int_{k_2-1}^{k_2} \left\| \left(\frac{k_1}{m}; \frac{k_2}{m} \right) - \xi/m \right\|^{\alpha\kappa} d\xi \end{aligned} \quad (4.5.27)$$

The norm can be bounded by a maximal distance on the square, $\sqrt{(1/m)^2 + (1/m)^2}$. Thus, on a single square the error is

$$\|I^{k_1, k_2}(t) - \Sigma_{M_1, M_2, m}^{k_1, k_2}(t)\|_\alpha^\alpha \leq C 2^{\alpha\kappa/2} m^{-(2+\alpha\kappa)}. \quad (4.5.28)$$

Summation over all $2mM_1 \times 2mM_2$ squares produces the following bound

$$\begin{aligned} &\sum_{k_1=t_1-mM_1+1}^{t_1+mM_1} \sum_{k_2=t_2-mM_2+1}^{t_2+mM_2} \|I^{k_1, k_2}(t) - \Sigma_{M_1, M_2, m}^{k_1, k_2}(t)\|_\alpha^\alpha \\ &\leq C \sum_{k_1=-mM_1+1}^{mM_1} \sum_{k_2=-mM_2+1}^{mM_2} 2^{\alpha\kappa/2} m^{-(2+\alpha\kappa)} = \\ &\quad 2^{2+\alpha\kappa/2} C M_1 M_2 m^{-\alpha\kappa} \end{aligned} \quad (4.5.29)$$

□

Remark 4.5.2. *Linearity of the error in M_1 and M_2 is a result of Hölder continuity of the kernel. This property prohibits using the error bound for long decreasing tails of kernels because smaller values of the function will be assigned the same error bounds.*

Truncation of kernel function

Proposition 4.5.3. *Truncation of exponential tail.*

If the kernel satisfies the condition

$$|g(\xi_1, \xi_2)| \leq C \exp(-\theta_1|\xi_1| - \theta_2|\xi_2|) \quad (4.5.30)$$

then truncation (4.4.15) from $\mathbb{R} \times \mathbb{R}$ to $[-M_1, M_1] \times [-M_2, M_2]$ yields error which is distributed as $s\alpha s$ random variable with scaling parameter

$$\frac{4C}{\alpha^2\theta_1\theta_2} \times \exp(-\alpha(\theta_1M_1 + \theta_2M_2)) \quad (4.5.31)$$

Proof.

$$\begin{aligned} & \left\| \iint_{\mathbb{R} \times \mathbb{R} / [-M_1, M_1] \times [-M_2, M_2]} g(\xi_1; \xi_2) dL(\xi_1, \xi_2) \right\|_{\alpha}^{\alpha} \leq \\ & C \iint_{\mathbb{R} \times \mathbb{R} / [-M_1, M_1] \times [-M_2, M_2]} (\exp(-\theta_1\xi_1))^{\alpha} (\exp(-\theta_2\xi_2))^{\alpha} d\xi_1 d\xi_2 = \\ & \frac{C}{\alpha^2\theta_1\theta_2} \iint_{\mathbb{R} \times \mathbb{R} / [-M_1, M_1] \times [-M_2, M_2]} e^{-\alpha\theta_1\xi_1} e^{-\alpha\theta_2\xi_2} d(\alpha\theta_2\xi_2) d(\alpha\theta_1\xi_1) = \\ & \frac{4C}{\alpha^2\theta_1\theta_2} e^{-\alpha(\theta_1M_1 + \theta_2M_2)} \end{aligned} \quad (4.5.32)$$

□

An analogous theorem holds when the kernel function decays with a power of the distance.

Proposition 4.5.4. *Truncation of tail decaying as a power function.*

If the kernel satisfies the condition

$$|g(\xi_1, \xi_2)| \leq \frac{C}{(\xi_1)^{r_1} (\xi_2)^{r_2}} \quad (4.5.33)$$

for some $r_1\alpha > 1$ and $r_2\alpha > 1$ then truncation from $\mathbb{R} \times \mathbb{R}$ to $[-M_1, M_1] \times [-M_2, M_2]$ yields error which is distributed as $s\alpha s$ random variable with scaling parameter

$$4CM_1^{(1-\alpha r_1)} M_2^{(1-\alpha r_2)} \quad (4.5.34)$$

Proof.

$$\begin{aligned} & \left\| \iint_{\mathbb{R} \times \mathbb{R} / [-M_1, M_1] \times [-M_2, M_2]} g(\xi_1; \xi_2) dL(\xi_1, \xi_2) \right\|_{\alpha}^{\alpha} \leq \\ & 4C \iint_{\mathbb{R} \times \mathbb{R} / [-M_1, M_1] \times [-M_2, M_2]} (\xi_1)^{-r_1\alpha} (\xi_2)^{-r_2\alpha} d\xi_1 d\xi_2 = \\ & 4CM_1^{(1-\alpha r_1)} M_2^{(1-\alpha r_2)} \end{aligned} \quad (4.5.35)$$

□

Bibliography

- [1] The comprehensive r archive network. <https://cran.r-project.org/>.
- [2] Tools for parameter estimation of the linear fractional stable motion. https://gitlab.com/Dmitry_Otryakhin/Tools-for-parameter-estimation-of-the-linear-fractional-stable-motion.
- [3] Y. Aït-Sahalia, J. Jacod, et al. Volatility estimators for discretely sampled lévy processes. *The Annals of Statistics*, 35(1):355–392, 2007.
- [4] A. Ayache and J. Hamonier. Linear fractional stable motion: a wavelet estimator of the α parameter. *Statistics & Probability Letters*, 82(8):1569–1575, 2012.
- [5] A. Ayache, S. Cohen, and J. L. Véhel. The covariance structure of multifractional brownian motion, with application to long range dependence. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100)*, volume 6, pages 3810–3813. IEEE, 2000.
- [6] J.-M. Bardet and D. Surgailis. Nonparametric estimation of the local hurst function of multifractional gaussian processes. *Stochastic processes and their applications*, 123(3):1004–1045, 2013.
- [7] A. Basse-O’Connor, R. Lachièze-Rey, and M. Podolskij. Power variation for a class of stationary increments lévy driven moving averages. *The Annals of Probability*, 45(6B):4477–4528, 2017.
- [8] A. Benassi, S. Cohen, J. Istas, et al. On roughness indices for fractional fields. *Bernoulli*, 10(2):357–373, 2004.
- [9] C. Bender, T. Marquardt, et al. Stochastic calculus for convoluted lévy processes. *Bernoulli*, 14(2):499–518, 2008.
- [10] C. Bender, A. Lindner, and M. Schicks. Finite variation of fractional lévy processes. *Journal of Theoretical Probability*, 25(2):594–612, 2012.
- [11] M. Bennedsen, A. Lunde, and M. S. Pakkanen. Hybrid scheme for brownian semistationary processes. *Finance and Stochastics*, 21(4):931–965, 2017.
- [12] A. Brouste, M. Fukasawa, et al. Local asymptotic normality property for fractional gaussian noise under high-frequency observations. *The Annals of Statistics*, 46(5):2045–2061, 2018.
- [13] S. Cambanis and M. Maejima. Two classes of self-similar stable processes with stationary increments. *Stochastic Processes and their Applications*, 32(2):305–329, 1989.

- [14] S. Cambanis, C. D. Hardin Jr, and A. Weron. Ergodic properties of stationary stable processes. *Stochastic Processes and their Applications*, 24(1):1–18, 1987.
- [15] J.-F. Coeurjolly. *dvfBm: Discrete variations of a fractional Brownian motion*, 2009. URL <https://CRAN.R-project.org/package=dvfBm>. R package version 1.0.
- [16] J.-F. Coeurjolly and J. Istas. Cramér–rao bounds for fractional brownian motions. *Statistics & probability letters*, 53(4):435–447, 2001.
- [17] S. Cohen, C. Lacaux, and M. Ledoux. A general framework for simulation of fractional fields. *Stochastic processes and their Applications*, 118(9):1489–1517, 2008.
- [18] G. Csárdi and M. Salmon. *rhub: Connect to 'R-hub'*, 2019. URL <https://CRAN.R-project.org/package=rhub>. R package version 1.1.1.
- [19] R. Dahlhaus. Efficient parameter estimation for self-similar processes. *The annals of Statistics*, pages 1749–1766, 1989.
- [20] R. M. M. Dan E. Dudgeon. *Multidimensional Digital Signal Processing*, volume 1. Prentice Hall, 1983.
- [21] T. Dang and J. Istas. Estimation of the hurst and the stability indices of a h -self-similar stable process. *Electronic Journal of Statistics*, 11(2):4103–4150, 2017.
- [22] D. Grahovac, N. N. Leonenko, and M. S. Taqqu. Scaling properties of the empirical structure function of linear fractional stable motion and estimation of its parameters. *Journal of statistical physics*, 158(1):105–119, 2015.
- [23] C. Heinrich, M. S. Pakkanen, and A. E. Veraart. Hybrid simulation scheme for volatility modulated moving average fields. *Mathematics and Computers in Simulation*, 2019.
- [24] J. Huang. *somebm: some Brownian motions simulation functions*, 2013. URL <https://CRAN.R-project.org/package=somebm>. R package version 0.1.
- [25] J. Istas and G. Lang. Quadratic variations and estimation of the local hölder index of a gaussian process. In *Annales de l'Institut Henri Poincaré (B) Probability and Statistics*, volume 33, pages 407–436. Elsevier, 1997.
- [26] J. Lebovits, M. Podolskij, et al. Estimation of the global regularity of a multifractional brownian motion. *Electronic Journal of Statistics*, 11(1):78–98, 2017.
- [27] J. Lévy-Véhel and R. Peltier. Multifractional brownian motion: definition and preliminary results. *Rapport de recherche de l'INRIA*, 2645, 1995.
- [28] B. B. Mandelbrot and J. W. Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.
- [29] S. Mazur, D. Otryakhin, and M. Podolskij. Estimation of the linear fractional stable motion. *Bernoulli*, 2019+.
- [30] V. Pipiras, M. S. Taqqu, et al. The structure of self-similar stable mixed moving averages. *The Annals of Probability*, 30(2):898–932, 2002.

- [31] V. Pipiras, M. S. Taqqu, et al. Central limit theorems for partial sums of bounded functionals of infinite-variance moving averages. *Bernoulli*, 9(5):833–855, 2003.
- [32] V. Pipiras, M. S. Taqqu, P. Abry, et al. Bounds for the covariance of functions of infinite variance stable random variables with applications to central limit theorems and wavelet-based estimation. *Bernoulli*, 13(4):1091–1123, 2007.
- [33] M. Podolskij. Ambit fields: survey and new challenges. In *XI Symposium on Probability and Stochastic Processes*, pages 241–279. Springer, 2015.
- [34] B. S. Rajput and J. Rosinski. Spectral representations of infinitely divisible processes. *Probability Theory and Related Fields*, 82(3):451–487, 1989.
- [35] S. Resnick and P. Greenwood. A bivariate stable characterization and domains of attraction. *Journal of Multivariate Analysis*, 9(2):206–221, 1979.
- [36] J. Rosinski et al. On the structure of stationary stable processes. *The Annals of Probability*, 23(3):1163–1187, 1995.
- [37] G. Samorodnitsky and M. S. Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance*, volume 1. CRC press, 1994.
- [38] G. Samorodnitsky et al. Null flows, positive flows and the structure of stationary symmetric stable processes. *The Annals of Probability*, 33(5):1782–1803, 2005.
- [39] K.-i. Sato. *Lévy processes and infinitely divisible distributions*. Cambridge university press, 1999.
- [40] S. Stoev and M. Taqqu. Simulation methods for linear fractional stable motion and farima using the fast fourier transform. *Fractals*, 95(1):95–121, 2004.
- [41] S. Stoev, V. Pipiras, and M. S. Taqqu. Estimation of the self-similarity parameter in linear fractional stable motion. *Signal Processing*, 82(12):1873–1901, 2002.
- [42] S. A. Stoev and M. S. Taqqu. How rich is the class of multifractional brownian motions? *Stochastic Processes and their Applications*, 116(2):200–221, 2006.
- [43] B. Swihart, J. Lindsey, and P. Lambert. *stable: Probability Functions and Generalized Regression Models for Stable Distributions*, 2017. URL <https://CRAN.R-project.org/package=stable>. R package version 1.1.2.
- [44] P. Tankov. *Financial modelling with jump processes*. Chapman and Hall/CRC, 2003.
- [45] N. W. Watkins, D. Credgington, R. Sanchez, and S. C. Chapman. A kinetic equation for linear fractional stable motion with applications to space plasma physics. *ArXiv e-prints*, Mar. 2008.
- [46] N. W. Watkins, D. Credgington, R. Sanchez, S. Rosenberg, and S. Chapman. Kinetic equation of linear fractional stable motion and applications to modeling the scaling of intermittent bursts. *Physical Review E*, 79(4):041124, 2009.
- [47] H. Wickham. testthat: Get started with testing. *The R Journal*, 3:5–10, 2011. URL <https://journal.r-project.org/archive/2011-1/RJournal.2011-1.Wickham.pdf>.

- [48] H. Wickham. *Advanced r*. CRC Press, 2014.
- [49] H. Wickham. *R packages: organize, test, document, and share your code*. O’Reilly Media, Inc., 2015.
- [50] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL <https://ggplot2.tidyverse.org>.
- [51] H. Wickham, P. Danenberg, and M. Eugster. *roxygen2: In-Line Documentation for R*, 2018. URL <https://CRAN.R-project.org/package=roxygen2>. R package version 6.1.1.
- [52] H. Wickham, J. Hester, and W. Chang. *devtools: Tools to Make Developing R Packages Easier*, 2019. URL <https://CRAN.R-project.org/package=devtools>. R package version 2.1.0.
- [53] D. Wuertz, M. Maechler, and R. core team members. *stabledist: Stable Distribution Functions*, 2016. URL <https://CRAN.R-project.org/package=stabledist>. R package version 0.7-1.