PhD Thesis

Heavy-Tailed Lévy-Driven Moving Averages

Estimation and Limit Theorems

Mathias Mørck Ljungdahl

Department of Mathematics
Aarhus University
2020

Heavy-Tailed Lévy-Driven Moving Averages
Estimation and Limit Theorems

PhD thesis by
Mathias Mørck Ljungdahl
201205667

Department of Mathematics, Aarhus University
Ny Munkegade 118, 8000 Aarhus C, Denmark

Supervised by
Professor Mark Podolskij

Submitted to Graduate School of Natural Sciences, Aarhus, 31 August 2020

Monograph was typesat in British English
under the Oxford style using kpfonts with
pdfLaTeX and the memoir class

AARHUS
UNIVERSITY
DEPARTMENT OF MATHEMATICS

Preface

The thesis in hand is the result of my PhD studies from 1 September 2017 to 31 August 2020 under supervision of Professor Mark Podolskij at the Department of Mathematics, Aarhus University.

This treatise encompasses the following five papers which can be read independently, but several of these contains natural links to each other.

Paper A A limit theorem for a class of stationary increments Lévy moving average process [*sic*] with multiple singularities. Modern Stochastics: Theory and Applications 5(3), 297–316.

Paper B A minimal contrast estimator for the linear fractional stable motion. Statistical Inference for Stochastic Processes 23, 381–413.

Paper C A note on parametric estimation of Lévy moving average processes. Springer Proceedings in Mathematics & Statistics 294, 41–56.

Paper D Multi-dimensional normal approximation of heavy-tailed moving averages. Submitted.

Paper E Multi-dimensional parameter estimation of heavy-tailed moving averages. Submitted.

Besides minor corrections in terms of spelling and typographical adjustment, the Papers A–E agree with the submitted or accepted versions. While Paper A was technically finished during my part A studies it was not included in the progress report due to page constraints. Moreover, parts of the necessary arguments was included in my master thesis. Paper B was partially included in the aforementioned progress report but critical additions and corrections has been made before its completion and subsequent submission and acceptance. In extension of Paper B I would like to thank Dmitry Otryakhin for collaboration culminating in the inclusion of the estimation techniques from this paper into the R-package `rlfsm`[(*)]. Paper C was included in the aforementioned report almost as is. Papers D and E have been developed and completed in part B of my PhD studies.

⬩ ⬩ ⬩

The dissertation starts with an introductory part which partly serves to motivate the use of Lévy-driven moving averages and more specifically the heavy-tailed setup. Secondly, this part also raises classical and always relevant statistical question, and relates the questions to previous literature. Thirdly, a summary of each paper is given relating them not only to the previously mentioned questions, but also internally. Indeed, Papers B–E have a natural line of thought and development which hopefully will be come clear.

⬩ ⬩ ⬩

The three years of studies culminating in this treatise is at its end and my gratitude for the resulting journey cannot be expressed in a few words, but I shall try anyway.

(*) See https://gitlab.com/Dmitry_Otryakhin/Tools-for-parameter-estimation-of-the-linear-fractional-stable-motion

First and foremost, to my supervisor Mark Podolskij, my heartfelt thanks for plunging me into a place amidst active research, but also for your guidance; the magnitude, multitude and quality of which was appropriate and suitably for my slightly idiosyncratic way of study. I will not try to express my internal reaction the day you offered me a chance at a PhD, you can of course guess it.

I thank moreover Christoph Thäle from Ruhr University and Ehsan Azmoodeh from University of Liverpool for a highly efficient and fun collaboration. In a similar note, Danijel Grahovac deserves my gratitude for not only providing and commenting on data, see Figure 2, but also discussing possible model applications.

A thanks goes out to my colleagues at the Department of Mathematics at Aarhus University, including Andreas Basse-O'Connor for many mathematical discussions and his general interest. Thank you also Vytaute Pilipauskaite for your voluminous laughter and our conference trips together.

I would especially like to mention my officemates Thorbjørn Ø. B. Grønbæk for always being cheerful and Mikkel S. Nielsen for fruitful collaboration, discussions and your interminable stream of questions.

I thank of course Lars 'daleif' Madsen for providing expert LaTeX-commentary and specialized solutions.

Noblesse oblige compels me to mention my family and friends; you have shown me that the world is not driven by Lévy processes but by the moments of everyday life—whether or not this is a heavy-tailed distribution I leave uncontested.

Lastly, but certainly not least, an immense gratitude goes to my girlfriend Michelle Lovring—your support and understanding has helped more than you probably know, you have been the Tinúviel to my Beren, the window to my wall.

Mathias Mørck Ljungdahl
August, 2020

Contents

Summary

This thesis concerns continuous-time Lévy-driven moving averages with deterministic kernels. These are motivated as a noisy component of a more general model using intuitive concepts and then deduced and specified by more rigorous mathematical theorems. The driver will use a heavy-tailed distribution chiefly to accommodate certain real-world phenomena such as rare events being non-negligible in terms of a probabilistic analysis. The endeavour is to understand the possible dynamics of such driven moving averages—indeed, this is key to extracting or inferring central aspects of the chosen model from actual data. In more concrete terms, using, extending and proving novel limit theory for variational statistics, e.g. the power variation, of given stationary data in this heavy world we shall develop a statistical methodology which allows inference and in particular estimation in a large class of parametrized Lévy-driven moving averages for which the dynamics are allowed to be surprisingly preposterous. Even more concretely, we shall, among other tasks, derive first- and second-order asymptotics for an estimator which finds the optimal parametric model by comparing, under a suitable weighing, the variational statistic of the empirical distribution with the theoretical counterpart of a possible distribution provided by the model—a procedure aptly named the *minimal contrast method*.

Resumé

Denne afhandling omhandler Lévy-drevne glidende gennemsnit i kontinuert tid med deterministiske kerner. Disse gennemsnit vil blive motiveret som en støj-komponent af en mere generel model, ved hjælp af intuitivt klare koncepter og derefter deduceret og specificeret med stringente, matematiske sætninger. Den drivende kraft vil følge en tung-halet fordeling, hovedsageligt for at afstedkomme fænomener fra virkelighedens verden, såsom sjældne hændelser der ikke er negligerbare set fra sandsynlighedsteoretiske overvejelser. Formålet vil være at forstå de mulige dynamiske egenskaber ved disse sådan drevne glidende gennemsnit – dette er nemlig nøglen til at ekstrahere eller ekstrapolere centrale aspekter af den valgte model fra data. Konkret set vil vi i denne tunge verden bruge, udvide eller bevise ny grænseværditeori for variationer, som for eksempel kvadratisk variation, af given stationær data til at udvikle en statistisk metodologi som muliggør ekstrapolering og i særdeleshed estimation i en stor klasse af parametriserede Lévy-drevne glidende gennemsnit, hvis dynamiske egenskaber kan tillades at være helt henne i det absurde. I endnu mere konkrete termer, så skal vi, blandt andet, udlede første- og andenordens asymptotik for en estimator som bestemmer den optimale parametriske model ved at sammenligne, under passende vejning, variationen af den empiriske fordeling med den teoretiske pendant stillet tilrådighed af modellen – en procedure som passende kaldes *minimal-kontrast metoden*.

Introduction

This part offers a mostly informal introduction to the class of heavy-tailed Lévy-driven moving averages. We shall in particular deduce this class as a (noise) component of a general time series model. Then the aim of this treatise will be formulated and a very rough overall methodology to tackle this will be sketched. This methodology will then in turn be used as a motivation for each individual paper.

Stationarity and Heavy-Tailed Models

Any analysis of a time series entails the study of the underlying dynamics of the system at hand and any proper analysis will identify components which are fairly homogeneous in some regard. Indeed, the *classical decomposition model* of a time series $(X_t)_{t\in\mathbb{T}}$ reads as

$$X_t = T_t + S_t + Y_t, \qquad (t \in \mathbb{T}), \tag{1}$$

consisting of a trend $(T_t)_{t\in\mathbb{T}}$, a seasonal or periodic component $(S_t)_{t\in\mathbb{T}}$ and a noise process $(Y_t)_{t\in\mathbb{T}}$. Different standard methods exist for either estimating the trend and seasonality components or transforming (X_t) appropriately, such as taking (higher order) increments at possibly different rates, so we are left with the noise term (Y_t), see [14, Section 1.4]. A frequent assumption on the noise is some kind of statistical equilibrium, such as a symmetrical variation around a mean both of which are constant in time, or a mean reversion over time. This leads to the class of (weakly) stationary processes as a model for the noise component (Y_t) which we shall discuss in the coming text.

Regardless of the type of observations (e.g. high or low frequency) of our data set let us assume a continuous-time setup as such is reality, in short set $\mathbb{T} = \mathbb{R}$. Suppose for the time being that we may formulate the aforementioned equilibrium in the classical, rigorous terms of expectation $\mathbb{E}$ and (co)variance Cov; then homogeneity in time of these quantities could be defined as

$$\mathbb{E}[Y_t] \equiv m \in \mathbb{R} \qquad \text{and} \qquad \operatorname{Cov}(Y_{t+h}, Y_h) = \operatorname{Cov}(Y_t, Y_0) \qquad \text{for all } h, t \in \mathbb{R}, \tag{2}$$

allowing the covariance to be time-dependent but not the individual variance. Property (2) is known in the literature as weak stationarity or sometimes as second-order stationarity. Let us glance over the seemingly innocuous implication that (Y_t) now belongs to $\mathcal{L}^2(\mathbb{P})$ and hence the powerful tools of harmonic analysis and Hilbert space theory are at one's disposal. Under fairly mild regularity conditions on the auto-covariance function in (2) then Y_t, if centred, will admit a Wold-type decomposition (see [54, 29]):

$$Y_t = \int_{-\infty}^{t} g(t-s)\,\mathrm{d}L_s + \xi_t \tag{3}$$

consisting of a casual moving average and a term ξ_t which is essentially known in terms of the observations Y_t from the beginning of time $t = -\infty$; see the formulation in [2, Theorem 4.1]. Here (L_s) is a square integrable process with weakly stationary and orthogonal increments, and the integral should be understood as an $\mathcal{L}^2$-limit of simple functions. The kernel g is a square integrable Borel-measurable function, also known as the spectral density. Assuming additionally that the auto-covariance at (2) is an absolutely continuous function with respect to the Lebesgue measure the term (ξ_t) vanishes and (Y_t) reduces to a casual moving average.

While (3) provides a moving average structure for a large class of noisy components it necessarily does not provide much information besides the second-order properties of (Y_t) and additionally, for modelling certain dynamics of the given system, it would therefore be desirable to put extra properties on (L_t), see the introduction in [38]. Alluding instead to the discrete-time setup (such as MA(∞)-processes, see [14]) a natural noise component (Y_t) would be driven by i.i.d. observations (L_t). But unfortunately no such, even remotely, nice process can exists, even if we drop the assumption of square integrability. Indeed, if $(L_t)_{t\in\mathbb{R}}$ is any i.i.d. collection of random variables with a bi-measurable version, then at least one variable would be independent of itself and hence this and therefore all of the collection would be degenerate, see [26, 19] for more on this.

Thus quite naturally (3) inspires us to instead consider dynamics as in (3) with drivers having stationary and independent increments. Requiring independent increments instead of orthogonal allows a framework beyond square integrability for our driver. Assuming slightly more than simply bi-measurability a driver must necessarily be a (two-sided) Lévy process, see [51, 1] for an introduction to Lévy processes and infinitely divisible distributions. Now, in this generality the moving average at (3) can no longer be defined in $\mathcal{L}^2$-terms, but rest assured we shall return to this problem later. These driving processes are sufficiently flexible to warrant an already extensive and increasing literature on continuous-time Lévy-driven moving averages. To name simply a few contributions in this area, dependence properties such as the semi-martingale property [6] and path properties [46] has been studied. Mixing conditions [15, 23] exists and uniqueness of the kernel [47] is understood, both of which are crucial in the inference realm of study. The former is crucial in first-order limit theory and the latter in proper parametrization or specification of a moving average model.

It would also be appropriate to name a few common and popular classes of Lévy-driven moving averages. The first example is the Ornstein–Uhlenbeck process with kernel $g(s) = \mathrm{e}^{-\lambda s}$ and for a non-Gaussian driver the authors of [5] show that certain stylized facts from finance can be modelled by this class. One example generalizing the Ornstein–Uhlenbeck is the CARMA processes, whose kernel may be represented as sum of exponential terms, see [13]. Another example generalizing the Ornstein–Uhlenbeck process are the so-called stochastic delay differential equations, see, e.g. [9]. The common trait of these models are that they are defined through equations which govern their dynamics directly rather than specify the kernel g explicitly, which is obviously enticing from a modelling viewpoint.

On the other hand, certain kernel behaviours are known to be, at least on an intuitive level, reflected in the moving average process. For example a non-smooth behaviour at 0 leads to more erratic local behaviour and a slowly decaying tail of g results in longer memory of the process. This leads more or less directly to the class of fractional Lévy processes that has power law kernels. Foreshadowing, we shall also see that these two behaviours will be crucial in this treatise.

⋄ ⋄ ⋄

In systems where the underlying dynamics are propelled by gradual smaller changes, the assumption of light tails or even as little as square integrability is definitely satisfied. It is disastrous, even catastrophic, if light-tail modelling is applied to systems where the governing, sovereign unit is large movements. Indeed, in the heavy-tailed spectrum and in general the field of extreme-value theory rare events have a non-negligible probability and it would be ruinous to ignore such events. For example, value-at-risk estimation would be so stupendously off that it would be useless. These systems arise many places such as, but not exclusive to,

- Insurance claims, see Figure 1.
- Financial data, e.g. stock market returns, see for example [45, Section 1.3.2].
- Data networks, such as LAN packet traces. See also Figure 2.

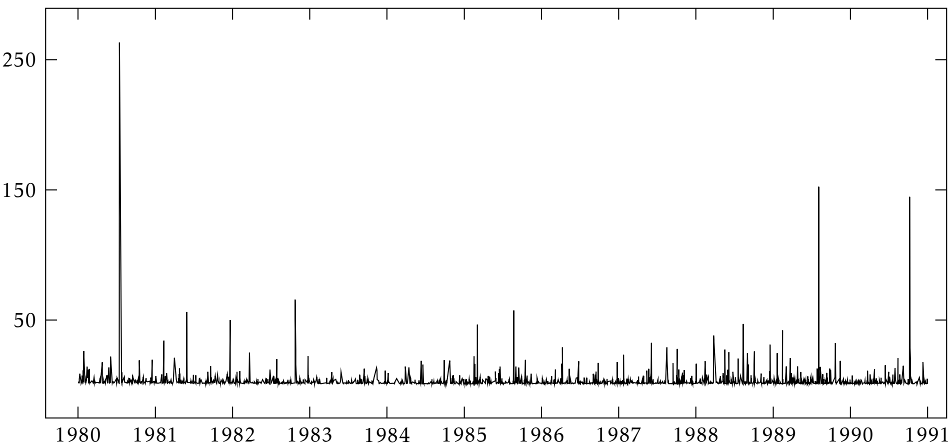

Figure 1. Fire loss insurance claims in units of 1 million DKK from the period of 1980 to 1990, adjusted for inflation.

Moreover, many common distributions have a fat tail, such as the log-normal and t-distribution, just to mention a one-sided and a two-sided example.

Judging whether or not the data is heavy-tailed or not and how heavy is obviously crucial, but let us remark that some standard methods of assessment exist, see again [45] and let us suppose from now on that this assessment has been answered in the affirmative. Of course in the heavy-tailed setup we do not necessarily have any Wold decomposition to motivate the choice of a moving average structure for the noisy component. Even more grievously, does any quantity resembling a convolutional structure between the Lévy process and the kernel exists? In layman's terms, we need to properly define Lévy-driven moving averages. Luckily, not only sufficient but also necessary conditions in terms of the characteristic triplet for L, as provided by the Lévy–Khintchine formula, are given in [44] which specify the class of deterministic kernels g for which the integral can be defined as a limit of simple functions—see also [23] for the multivariate matrix-valued case. For modelling or simply general interest, it is worth remarking that the theory of Lévy-driven moving averages can be extended to predictable, *random* kernels, see [30], or [25] for a brief summary of basics.

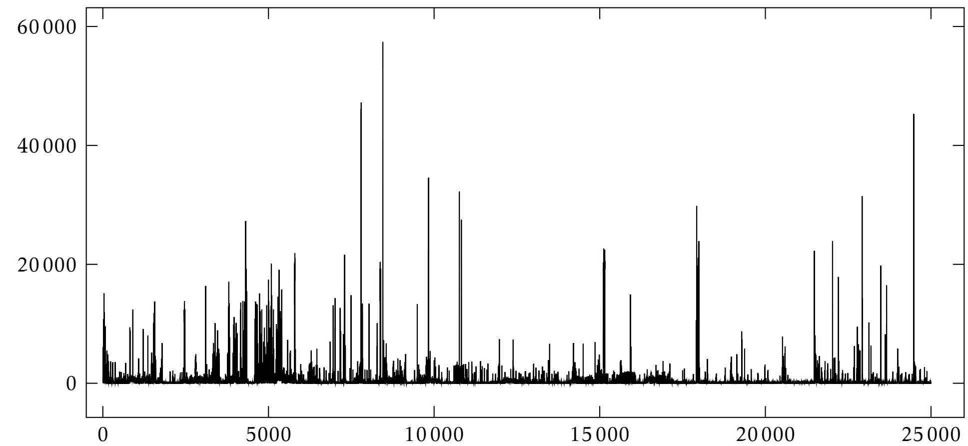

Figure 2. Ethernet trace recorded at the Bellcore Morristown Research and Engineering facility (BC-Oct89Ext). Packet arrival times in seconds against number of packets in bytes. The data has been aggregated to arrival times in blocks of 1 s and only the 25 000 first points are included.

Unfortunately our previous definition of statistical equilibrium through the covariance or expected value is no longer viable—consider, e.g. the extreme, but classical, Pareto distribution with tail index < 1. And therefore the definition of *weak* stationarity is not meaningful, and instead we must rely on the stronger requirement (when both concepts are meaningful) of stationarity, meaning that the finite dimensional distributions

$$(Y_{t_1+h}, \dots, Y_{t_n+h}), \qquad (t_1, \dots, t_n \in \mathbb{R}),$$

are invariant under the shifts $h \in \mathbb{R}$. If we still assume symmetry of our noise, then one natural candidate would be the class of symmetric α-stable (SαS) distributions. These are natural partly since they are infinitely divisible and closed under convolution and scaling, but maybe chiefly since they are the universal (central) limits of (symmetrically distributed) sums of i.i.d. random variables—with $\alpha = 2$ corresponding to the normal distribution and the classic central limit theorem.

This discussion puts our noise in the class of stationary SαS processes and while we no longer have a Wold-type decomposition, [48, Theorem 6.1] argues that

$$Y_t = X_t^1 + X_t^2 + X_t^3 \tag{4}$$

with equality in distribution. Here (X_t^1) is a so-called mixed moving average, which can be thought of as superposition of moving averages, (X_t^2) is a harmonizable process (see [50]) and (X_t^3) is a stationary SαS process which is neither a moving average or harmonizable process, but is, as of now, unsatisfactorily described. Since the components in (4) are mutually independent a study of Y_t could be split into an analysis of each component separately. As a side-note (4) shows that for SαS processes stationarity and harmonizability are not the same, which is in stark contrast to the Gaussian case, where almost all Gaussian processes have both a harmonizable and a moving average representation, see [22]. This speaks volumes of the flexibility of stationary SαS processes, at least compared to stationary Gaussian processes. But as might be apparent from the discussion so far, we will focus solely on the (non-mixed) moving average part in this dissertation.

Lastly, let us remark that there is nothing wrong with a Gaussian driver as our Lévy process. Indeed, quite interesting phenomenon such as turbulence, [20], could be modelled with such processes. The limit theory is quite developed for a Brownian process, even for random volatility kernel functions, known as Brownian semi-stationary processes, see [3, 24]. We will therefore focus on the processes without a Gaussian part and their corresponding novel and emerging limit theory.

Inference, Simulation and Limit Theory

Now that we have settled on the class of SαS moving averages as a potential model for Y the ensuing quest becomes manifold. It includes in particular the inference of the kernel function g and Lévy-driver. In the following we will discuss the overall elements in our study.

To better encompass the processes studied in the treatise we first generalize the SαS moving average processes and consider *stationary increments Lévy-driven moving averages* $Y = (Y_t)$ given by:

$$Y_t = \int_{-\infty}^{t} (g(t-s) - g_0(-s))\, \mathrm{d}L_s, \qquad (t \geq 0), \tag{5}$$

where $g, g_0 : \mathbb{R} \to \mathbb{R}$ are deterministic Borel-measurable functions both of which vanishes on $(-\infty, 0)$. Moreover, $L = (L_t)$ is a symmetric Lévy process with $L_0 = 0$ with no Gaussian part. The extra function g_0 compared to the moving average in, e.g. (3) does more than simply make the moving average 'two-sided'. Indeed, data (X_t) according to the general decomposition model at (1) may very well be transformed using, e.g. increments, which in turn will transform the Y at (5) into a proper (stationary) moving average. In other words, (5) allows for a 'noise' component which only after transformation of the data is stationary. Moreover, the case $g_0 = g$ is necessary for fractional processes and taking a g_0 different from g would be akin to the inclusion of some initial term.

The point of departure for inference in any (sufficiently complicated) model is in the frequentist world often asymptotic theory. This includes first-order limit theory, often associated with the 'law of large number'. Indeed, this first order describes what the quantity at hand fluctuates *around*—the common example being the mean. This is often enough for estimation of a particular characteristic of the underlying model, but we have not described *how* our quantity fluctuates. The 'central limit theorem' is used to describe how the empirical average fluctuates around it's 'true' value—in the classical case the description is via weak convergence to the normal distribution at some rate, often $\sqrt{n}$. If this second-order limit theory is obtained it allows the construction of asymptotic confidence regions paving the way for a more refined inference.

We describe now the quantities of interest in our particular setup, which are of the type:

$$V_n(Y; f) = \frac{1}{n} \sum_{i=1}^{n} f(\Delta_i Y), \qquad (n \in \mathbb{N}), \tag{6}$$

where $f : \mathbb{R} \to \mathbb{R}$ belongs to a suitably large class of functions and $\Delta_i Y = Y_i - Y_{i-1}$ denotes the increment. Note that the ergodic theorem dictates that $V_n(Y; f)$ fluctuates around the mean $\mathbb{E}[f(\Delta_1 Y)]$ if the latter is finite.

A classical example of f includes the power variation and it is incessantly studied in the setting of (Itô) semi-martingales, see for example [28, 4, 43]. Providing first- and second-order limit theory in this setting gives a direct inference-link to the integrated volatility of a financial model—a crucial and fundamental quantity.

For certain f we may in general extract important information of the underlying model of (Y_t) via the quantity $\mathbb{E}[f(\Delta_1 Y)]$, such as in the case of quadratic power variation. To be more concrete, suppose a parametric model $\{\mathbb{P}_\theta \,|\, \theta \in \Theta\}$ for (Y_t). Then it might be possible to extract any value θ from $\mathbb{E}_\theta[f(\Delta_1 Y)]$ using a particular f. Since only the empirical version, $V_n(Y;f)$, is available from our data it would be natural to consider a comparison map:

$$\theta \longmapsto \rho\big(V_n(Y;f), \mathbb{E}_\theta[f(\Delta_1 Y)]\big), \tag{7}$$

for some 'distance' ρ. Asymptotically it holds $V_n(Y;f) \approx \mathbb{E}_{\theta_0}[f(\Delta_1 Y)]$ for the true parameter θ_0, so the argument that minimises the distance ρ should be close to θ_0—under suitable injectivity assumptions of course. This idea lies at the heart of the minimal contrast approach. Setting $f_u(x) = \mathrm{e}^{\mathrm{i}ux}$, (7) then compares the empirical characteristic function with the theoretical one as a function of the model parameters. In other words, ρ compares the empirical distribution with theoretical distribution $\mathbb{P}_\theta$.

Considering f_u for only a fixed value $u \in \mathbb{R}$ seems arbitrary and instead we shall consider a minimal contrast estimator θ_n by comparing all function values:

$$\theta_n = \operatorname*{argmin}_{\theta\in\Theta} \int_0^\infty \Big|V_n(Y;f_u) - \mathbb{E}_\theta[f_u(\Delta_1 Y)]\Big|^2 \mu(\mathrm{d}u) =: \operatorname*{argmin}_{\theta\in\Theta} F(Y_n,\theta), \tag{8}$$

where μ is a (symmetric) probability measure on $\mathbb{R}$ which weighs the contrast suitably. θ_n is an M-type estimator and by differentiation it is turned into a Z-estimator, that is, θ_n is determined by solving the *estimating equation*:

$$\nabla_\theta F(Y_n,\theta) = 0.$$

General theory exists for M- and Z-estimators, see [53], and knowing this it should be apparent that limit theorems for quantities of the type V_n at (6) are crucial. Indeed, if we are to have any success, then the quantities that θ_n minimizes over should themselves converge. We remark that with these observations that a possible approach would be to deploy the machinery of empirical process theory as is often done for M- and Z-estimator, but the classical theory requires at least independence and for, e.g. stationary processes this field is still an active field of research. Instead we shall make use of the implicit function theorem for infinite dimensional spaces; a technique which does have some similarity with the so-called *linearization* argument, see again [53] or [27]. But instead of trying to fit a square peg into a round hole we formulate the methodology without mention of this type of argument, see, e.g. Section 4.2 in Paper C.

⋄ ⋄ ⋄

Lastly, for various inference procedures such as parametric bootstrap methods it is of fundamental import to be able to resample from the model $\mathbb{P}_\theta$. As of now, no exact procedure for simulating directly from (5) exists, but for specific subclasses it may be possible to find quite suitable methods. The general simulation method in the coming text comes as no surprise. Consider in the following an ordinary moving average Y at (5) with $g_0 \equiv 0$. For a fixed $t \in \mathbb{N}$ truncate first the integration region to $[-M, t]$ and then approximate the integral with a Riemann sum of mesh size $\frac{1}{m}$, $m \in \mathbb{N}$. Consider then a fixed element of the Riemann sum and write it as a trivial integral with respect to L. This can then be paired with a corresponding term in the truncated integral yielding an error term of the form:

$$\int_{k-1}^{k} (g(t - s/m) - g(t - k/m))\,\mathrm{d}L_{s/m}, \qquad (k \in \{-mM + 1, \dots, mt\}).$$

If L is an α-stable Lévy motion, then this is an α-stably distributed error term. Since any $\mathcal{L}^p$-moment, $p < \alpha$, of an α-stable random variable can be expressed as a constant times its scale parameter, see [50, Property 1.2.17], it is sufficient to provide bounds on the scale parameter. It is now clear that the error analysis boils down to two behaviours of the kernel: the one at 0 corresponding to the error from the Riemann sum approximation, and the tail decay corresponding to the truncation of the integral. For bounds in terms of concrete behaviours see [39] and note in particular a potential trade-off between the truncation parameter M and the mesh size m. For a whole sequence $Y_1, \dots, Y_n$ one would then need an efficient way of computing the many Riemann sums. This is possible using a Fast Fourier Transform algorithm based on a convolution form of the Riemann approximation—see [52] for the specific example of the linear fractional stable motion. We conclude that parametric bootstrap procedures are often possible in the general setup, see also Paper B.

Paper A

In Paper A we derive a first-order limit theorem for the power variation $f(x) = |x|^p$ at (6) for high frequency observations from the kth order increments of Y:

$$\Delta_{i,k}^n Y = \sum_{j=0}^{k} (-1)^j \binom{k}{j} Y_{(i-j)/n}, \qquad i \geq k. \tag{9}$$

Before proceeding with the specific investigation of the current paper we note that limit theorems of (9) for power variations have been investigated in [8] and later generalised in [7] to a larger class of functions which in particular includes bounded functions, such as the (real part) of the characteristic function f_u discussed in the previous section. In parallel, a generalisation for the power variation to semi-stationary case is given in [12], which constitutes a class with kernels modulated by a random volatility term.

The aforementioned limit theory shows that not only does the mode of convergence (weak or in probability) but also the type of limit (deterministic or random) depend in a non-trivial way on the interplay between the order of increments, k, the Blumenthal–Getoor index $\beta \in [0, 2]$ of L (see equation (1.4) in Paper A), the behaviour of the functional g at 0 and the type of functional at hand, where the later type in the power variation case references the specific power $p > 0$. The behaviour of g at the point 0 is defined in terms of a power law equivalence in the sense that

$$g(t) \sim t^{\alpha} \qquad \text{as } t \downarrow 0 \tag{10}$$

for $\alpha \in \mathbb{R}$, where the equivalence $f(t) \sim g(t)$ as $t \downarrow 0$ means that $f(t)/g(t) \to 1$ as $t \downarrow 0$. [8] then provides (under some additional conditions, see Assumption (A) and Assumption (A-log) in Paper A) three possible regimes for the 'law of large numbers' of the power variation:

(i) If $\alpha < k - 1/p$ and $p > \beta$ then $V_n(Y;p)$ converges *weakly* to a *random* limit at rate $n^{\alpha p}$.

(ii) If $\alpha < k - 1/p$ and $p < \beta$ then $V_n(Y;p)$ converges *in probability* to a *deterministic* limit at rate $n^{-1+p(\alpha+1/\beta)}$.

(iii) If $p \geq 1$ and $\alpha > k - 1/(\beta \vee p)$ then $V_n(Y;p)$ converges *in probability* to a *random* limit at rate n^{-1+pk}.

Note that in the regime (i) 0 is a singularity of g in the sense that the kth derivative $g^{(k)}$ explodes at 0 due to the behaviour (10). This regime is also strikingly different from the case of semi-martingales with p-summable jumps, see, e.g. [28]. The main purpose of Paper A is to investigate the situation of multiple singularity points $0 = \theta_0 < \cdots < \theta_l$:

$$g(t) \sim |t - \theta_z|^{\alpha_z} \qquad \text{as } t \to \theta_z \qquad (z = 1, \ldots, l).$$

We remark that a similar question is carried out in [24] in the case of Brownian semi-stationary processes. Paper A shows that each singularity θ_z propagates through to the limit variable in a similar manner to the point 0, except that the propagation depends directly on the fractional size of the resulting blow-up of the singularity θ_z. Indeed, we extend the result from (i) to multiple singularities for any subsequence $(n_j)_{n\in\mathbb{N}} \subseteq \mathbb{N}$ such that the fractional limit

$$\lim_{j\to\infty} \{n_j \theta_z\} =: \eta_z \qquad \text{for each } z \in \{0, 1 \ldots, l\} \tag{11}$$

exists. Hence the generalisation of (i) to multiple singularities is even more peculiar than the ordinary case since different subsequence (n_j) can lead to different η_z's at (11), so the power variation $(V_n(Y;p))_{n\in\mathbb{N}}$ is in general only tight and with multiple accumulation points.

We note that (i) has as a critical case: $\alpha = k - 1/p$. This case has been dealt with separately in article [10] and Paper A also studies this particular case, which for multiple singularity points corresponds to $\alpha_1 = \cdots = \alpha_l = k - 1/p$.

Before concluding it is worth noting that [8] also provides second-order limit theory which in particular displays the multitude of possible limits: central, non-central, Gaussian and non-Gaussian cases.

Paper B

This article studies the *linear fractional stable motion* (lfsm) which corresponds to Y at (5) with the polynomial type kernel: $g(s) = g_0(s) = s_+^{H-1/\alpha}$ where $s_+ = \max\{s, 0\}$, $H \in (0,1)$ and L is symmetric α-stable Lévy motion with scale parameter $\sigma > 0$. As we shall now discover, the lfsm can be motivated in several ways. First, the integral at (5) in this case is a direct analogue to the moving average representation of the *fractional Brownian motion* (fBM) B:

$$B_t = \int_{-\infty}^{t} (t-s)_+^{H-1/2} - (-s)_+^{H-1/2} \, \mathrm{d}W_s, \qquad (t \geq 0),$$

where $W = (W_t)$ is a zero-mean two-sided Brownian motion, cf. [36]. The lfsm is in this regard a generalisation which contains the fractional Brownian motion as the special case: $\alpha = 2$. The wide applicability of the fBM can then be transferred to the lfsm and actually much empirical data exhibits the *Joseph* effect, commonly known as long range dependence of increments—a trait the fBM is famous for. Additionally, data may also display what is known as the *Noah* effect, which in a nut shell is larger governing dynamics. This latter effect is not displayed by the light-tailed distribution of the Gaussian process, but it is by the SαS-distributed marginal distributions of the lfsm. Moreover, initially, for simplicity and lack of probabilistic tools, the two effects where modelled separately, see [35] for more information. The lfsm may therefore capture both effects. Of course, long range dependence is not understood in the classical sense of a slow decay of the auto-correlations but instead of a slow decay of the (incremented) kernel:

$$x^{H-1/\alpha} - (x-1)^{H-1/\alpha} \sim Cx^{H-1-1/\alpha} \qquad \text{as } x \to \infty$$

for some constant $C \in \mathbb{R}$. In a similar note the lfsm can be motivated using fractional calculus, cf., e.g. [42], since these tools are sometimes used to obtain long memory or dependence from a given process—in the lfsm case a SαS Lévy process. Such a procedure has been done in [21] for the aforementioned stochastic delay differential equations to obtain a semi-martingale, contrary to the lfsm, with long-range dependence.

So secondly, the lfsm is also motivated by real-world phenomena. Indeed, the lfsm is considered as a possible model for heavy network traffic, such as Figure 2, since this kind of data is considered to exhibit both self-similarity and heavy tails, see [34].

The fBm is self-similar with (Hurst) index $H \in (0,1)$, meaning that in distribution:

$$(B_{ct})_{t\geq 0} = (c^H B_t)_{t\geq 0},$$

and it is the only (up to a scale $\sigma > 0$) zero-mean Gaussian process with stationary increments with this property. While the lfsm is also self-similar with index H it is no longer uniquely determined by this property among the SαS distributions, cf. [16]. But the contrasting properties of the lfsm and the fBM does not stop here. The path properties of the lfsm are well-known, see [50], but contrary to the fBM which has locally Hölder continuous path of any order $< H$ the lfsm has only Hölder continuous paths up to order $H - 1/\alpha$ in the case $H - 1/\alpha > 0$, but if this exponent is negative the lfsm is unbounded on any open interval, see Figure 3.

Parameter estimation of (H, σ) for the fBM has been tackled successfully, see the references in Paper B. But the situation for estimation of the three-dimensional parameter (σ, α, H) for the lfsm has been much more partial in the sense that no estimation of the joint parameter (σ, α, H) together with accompanying second-order theory has been proposed—that is, until recently in [37]. Here the authors proposes a ratio-type estimator for the Hurst parameter H based on the power variation of the kth orders increments at (9) but at different rates. This is then combined with identities for the characteristic function of the kth order increments to obtain expressions for the scale σ and stability index α, see Section 4.1 in Paper B. Moreover, using [8] as a starting point the authors of [37] provide second-order limit theory consisting of a normal regime in the case $k > H - 1/\alpha$ and a stable regime when $k < H - 1/\alpha$.

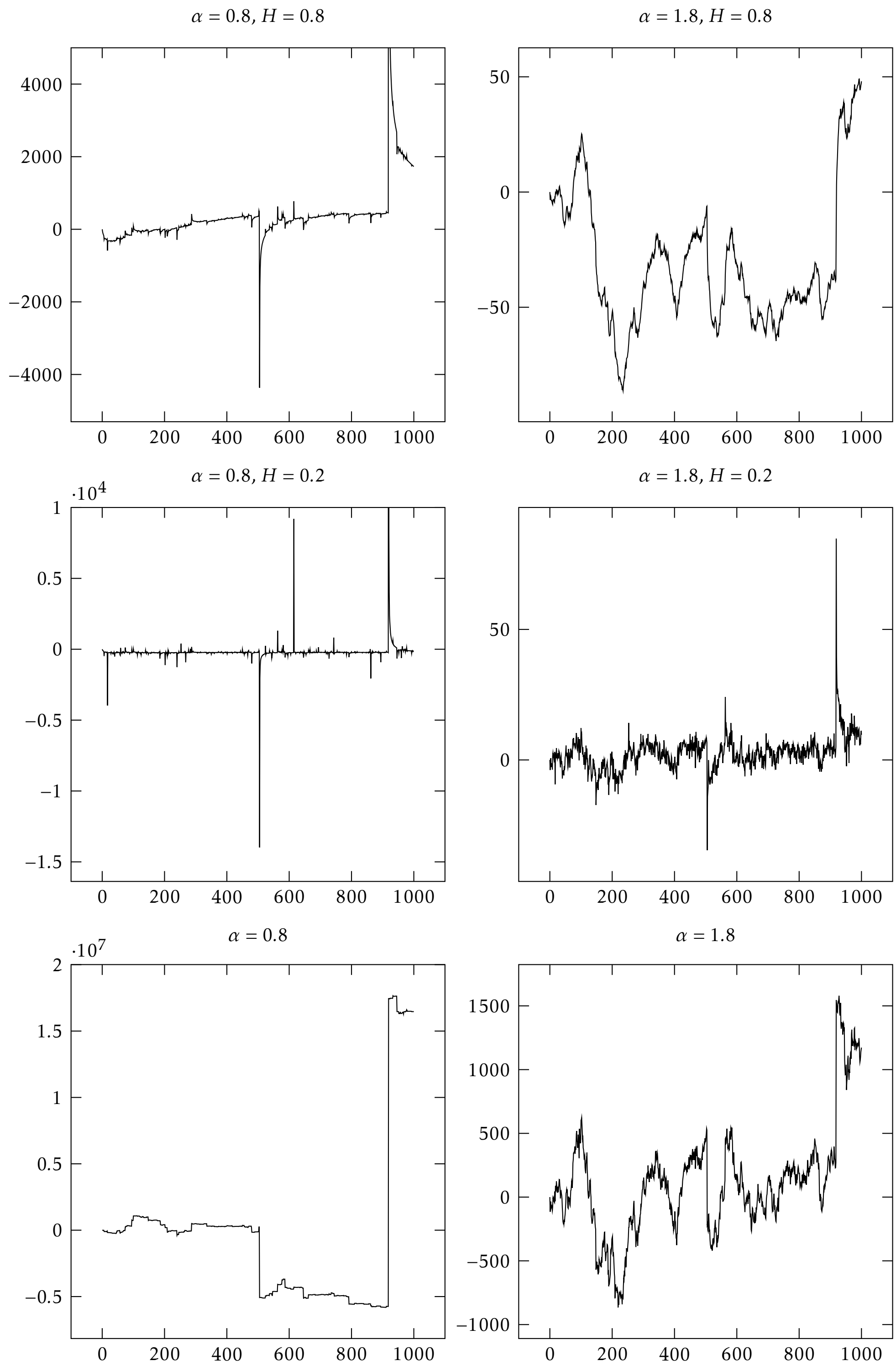

Figure 3. Top two rows are paths of the linear fractional stable motion and the bottom row is the driving α-stable Lévy motion.

To obtain the necessary identities for the scale and stability index (σ, α) only two values of the characteristic function are needed, concretely the values at $t_1 = 1$ and $t_2 = 2$ are used. Hence, as already mentioned, it would be natural to instead compare for all values $t \in \mathbb{R}$ and therefore consider the minimal contrast estimator θ_n at (8) with $\theta := (\sigma, \alpha, H) \in \Theta := (0,\infty) \times (0,2) \times (0,1)$. To obtain first- and second-order limit theory we will therefore naturally need limit theorems for the underlying integral functional F at (8). In Paper B we analyse, extend and develop the necessary arguments for our 'linearization' methodology to work. To obtain a weak limit theory for the centred object $n^r(\theta_n - \theta)$ for some rate r, three fundamental steps are required:

Step 1 Obtain finite dimensional weak convergence of the underlying processes $u \mapsto V_n(Y; f_u) - \mathbb{E}_\theta[f_u(\Delta_1 Y)]$.

Step 2 Analysis of the path properties of the limit process obtain at Step 1.

Step 3 Extend the convergence from Step 1 to convergence of the integral functional $F(Y_n, \theta)$.

Step 1 has been dealt with in [37] which is necessary for their limit theory, but it is not viewed in the sense of finite dimensional convergence of processes as it is for the minimal contrast estimator in Paper B.

Step 3 is related to the general question of whether finite dimensional convergence of a sequence of processes can be extended to integral functional of these. This has of course been studied before but the available results are not particularly satisfactory for our purpose. Instead we observe that certain moment bounds are sufficient for the extension in Step 3. This is also related to Step 2 since in the normal regime, $k > H + 1/\alpha$, we analyse the covariance of the functionals in Step 1 to obtain an expression for the covariance function of the Gaussian limit. This analysis also yields a Hölder continuous version of the limit Gaussian process which concludes Step 2 and here the tractability of the characteristic function of an SαS distributed variable is important since it is directly related to the covariance function of the limit, see Theorem 2.1(i) and the quantity at (2.4) in Paper B.

For the stable regime, $k < H + 1/\alpha$, covariance bounds are not available and instead we prove a Karamata type theorem to provide uniform bounds on the moments of the processes at Step 1, see Proposition 5.7 in Paper B. These moment bounds are not suitable for the analysis of the stable limit process, but this process is extremely simple as it is of the form: $(\kappa(u)S)_{u \geq 0}$, for some deterministic function $\kappa : \mathbb{R}_+ \to \mathbb{R}$ and some skewed α-stable variable S—making Step 2 a trivial matter in this regime.

After the second-order limit theory has been developed Paper B tackles the question of asymptotic confidence regions for the parameters. The most immediate problem is which regime we are in, the normal regime, $k > H + 1/\alpha$, or the stable regime, $k < H + 1/\alpha$. This is quite crucial since the rate in the stable regime is $r = 1 - 1/(1 + \alpha(k - H))$ while in the normal case it is $r = 1/2$. Since the parameters are unknown the regime and rate is a priori unknown. Moreover, $1/\alpha$ is unbounded in $\alpha \in (0,2)$ so we cannot simply pick a large increment $k \in \mathbb{N}$. However, if we are in the continuous case $H - 1/\alpha > 0$, then trivial algebra says that any $k \geq 2$ will place us in the normal regime. In this case Paper B uses a parametric bootstrap approach to build asymptotic confidence regions, see Section 4.2.

In the general case pre-estimation of α will provide an estimate of the necessary order $k \in \mathbb{N}$ such that we may obtain a known rate of $\sqrt{n}$-convergence. Unfortunately this yields four different regimes and the limiting type of distribution is no longer known. To overcome this Paper B then proposes in Section 4.3 a subsampling procedure to estimate the distribution functions of the limiting distribution and hence estimate the necessary quantiles for building the asymptotic confidence regions.

We conclude with an unfortunate problem for our general minimal contrast approach. The parameters (σ, α, H) are not completely determined from the (one-dimensional) characteristic function. This forces us to use a plug-in approach by inserting the same ratio estimator as in [37] for the Hurst parameter H in our minimal contrast estimator for (σ, α). In particular we really do need the more complicated limit theory of the variation $V_n(Y; f)$ for unbounded functionals f.

Paper C

In Paper C we take a step back from Paper B and realize that much of the overall approach, namely Steps 1–3 are generally applicable. Paper C then provides first- and second-order limit theory for the minimal contrast estimator for the class of parametric SαS-driven moving averages, that is, (Y_t) at (5) with $g_0 \equiv 0$ and $g = g_\theta$ for a one-dimensional parameter $\theta \in \Theta \subseteq \mathbb{R}$. Of course the underlying limit theory at Step 1 from Paper B is no longer available, but theory for general *bounded* functionals such as the characteristic function has luckily been developed in [41]. For later emphasis we remark here that the functionals in $V_n(Y; f)$ are of the type:

$$f : \mathbb{R} \longrightarrow \mathbb{R}^d. \tag{12}$$

I.e. the co-domain is multi-dimensional, and this is important as it will ensure finite dimensional convergence of our processes induces by the (empirical) characteristic functions. If $d = 1$ then we would only have weak convergence of a single fixed value of our characteristic functions.

As Papers B and C has taught us it would now be foolish to hope that the we can tackle a multi-parametric framework $\Theta \subseteq \mathbb{R}^m$ with the current methodology. The desire to envelop this framework leads us to the following two papers. The first paper, Paper D, relates directly to Step 1 above and the second, Paper E, completes the second-order limit theory in Steps 2 and 3 for the general multi-parametric framework.

Paper D

The main purpose of this paper is to generalise the limit theory for bounded functionals as in (12) of heavy-tailed Lévy-driven moving averages to multivariate functionals of the type:

$$f : \mathbb{R}^m \longrightarrow \mathbb{R}^d. \tag{13}$$

This was at the time an interesting question in itself and deserved a specific study, hence the separate, independent article. Our statistical motivation will become clear in the description of the next article, but already now it is intuitively clear that functionals for $m > 1$ can capture significantly more complicated behaviour, especially dependence in the stationary sequence $(\Delta_{i,k} Y)_{i \geq k}$.

A very successful method in deriving Gaussian limit theorems has been the combination of Stein's method with Malliavin calculus. We will sketch the overall (univariate) approach in the following.

Consider a metric d on the space of Borel probability measures on $\mathbb{R}$ which metricizes weak convergence. For concreteness consider the Wasserstein distance defined as:

$$d(Y,N) := \sup_{h \in \mathrm{Lip}_1(\mathbb{R})} \left| \mathbb{E}[h(Y)] - \mathbb{E}[h(N)] \right|,$$

where Y and N are random variables on $\mathbb{R}$ and $\mathrm{Lip}_1(\mathbb{R})$ denotes the space of Lipschitz functions $h : \mathbb{R} \to \mathbb{R}$ with Lipschitz constant 1. Recall now Stein's Lemma which states that N has a standard normal distribution if and only if

$$\mathbb{E}[f'(N) - Nf(N)] = 0 \tag{14}$$

for all $f : \mathbb{R} \to \mathbb{R}$ in a class of sufficiently smooth functions. So if Y is supposed to be close to N in distribution then replacing N in (14) with Y should result in something small. Similarly, the expectations of $h(Y)$ and $h(N)$ should be roughly the same, or equivalently, the difference should be zero, for a large class of functions $h : \mathbb{R} \to \mathbb{R}$. Comparing the two differences then yields, on average,

$$f'(x) - xf(x) \approx h(x) - \mathbb{E}[h(N)].$$

Fixing h and replacing the approximation '$\approx$' with an equality yields a first-order differential equation in f known as *Stein's equation for normal approximation*. For some classes of h the functional solution f to Stein's equation is known to satisfy certain regularity conditions—e.g. for Lipschitz h as in the Wasserstein distance the solution f is a continuous differentiable function with absolutely continuous derivative. Hence we obtain the following bound:

$$d(Y,N) \leq \sup_{f \in \mathcal{H}} |\mathbb{E}[f'(Y) - Yf(Y)]|,$$

for a certain class of functions $\mathcal{H}$; we refer to [18] for more details. We have now reduced the problem to terms depending solely on Y. Suppose from now on that Y is a Poisson functional, i.e. a function of a Poisson (point) process (on some abstract space). This includes moving averages driven by a pure jump Lévy process, [49, Proposition 2.10], but is not exclusive to these. In this Poissonian framework the powerful tools of Malliavin calculus (on Poisson spaces) are available to us. We refer the reader to [33, 32] for an excellent introduction into this elegant field of mathematics. Indeed, using key formulas it is possible to obtain so-called second-order Poincaré inequalities, where 'second-order' simply refers to the fact that bounds on, e.g. the Wasserstein distance leads to second-order limit theorems, see [17]. Correspondingly, these inequalities may involve the analysis of (second-order) Malliavin derivatives.

Until recently the available second-order Poincaré inequalities in [31] where not suitable for heavy-tailed moving averages since the available bounds diverged in this case. This was rectified in [11] and a refined second-order Poincaré inequality was established by careful distinction between small and large values for certain Malliavin derivatives.

The method of [11] was also the departure for Paper D where the extension to functionals at (13) for a general $m > 1$ was possible due to the simple, but not simplistic, method. For the extension to (13) for a general $d > 1$ the classical Wasserstein was no longer suitable. Indeed, if we wish to provide bounds between a multivariate normal distribution and our Poisson functional the situation depends on the covariance matrix of the normal distribution—if we do not put any assumption of invertibility on this matrix, then we need to increase the smoothness of the class $\mathcal{H}$, see [40, Table 1] for a concrete comparison.

Luckily, for our statistical purposes the exact metric is not so important as long as it implies weak convergence and as a corollary of the refined Poincaré inequality in a multivariate setting, $m > 1$, the article provide bounds for a suitably large class of multi-dimensional moving averages. Let us elaborate slightly more on this, consider an m-dimensional random vector $(Y^1, \dots, Y^m)$ of moving averages with

$$Y_t^i = \int_{-\infty}^t g_i(t-s)\,\mathrm{d}L_s, \qquad (i \in \{1, \dots, m\}),$$

where the kernel $g_i : \mathbb{R} \to \mathbb{R}$ satisfy certain power law behaviours at 0 and at ∞ and L is a common SαS-driver. To draw parallels with previous limit theory suppose that

$$g_i(t) \sim t^{\kappa_i} \quad \text{as } t \downarrow 0 \quad \text{and} \quad g_i(t) \sim t^{-\beta_i} \quad \text{as } t \uparrow \infty$$

for $\beta_1, \dots, \beta_m > 0$ and $\kappa_1, \dots, \kappa_m \in \mathbb{R}$. Then if the kernel is not too 'explosive' at 0, i.e. $\kappa_i > -1/\alpha$, and the underlying common driver L is not too heavy-tailed combined with the memory being not too long, i.e. $\alpha\beta_i > 2$, then we obtain a joint central limit theorem for

$$V_n(Y; f) = \frac{1}{n} \sum_{i=1}^n f(Y_i^1, \dots, Y_i^m)$$

for bounded C^2-functions $f : \mathbb{R}^m \to \mathbb{R}^d$ with bounded derivatives. We conclude by remarking that this fits our intuition, based on previously established theory such as [7], quite well, and that Paper D generalises the framework of $f : \mathbb{R} \to \mathbb{R}$ in [11] to functionals as in (13).

Paper E

We now return to our previous defeat in Papers B and C. Viz, the methodology has so far only been able to tackle moving averages Y as in (5) with $g_0 \equiv 0$ and $g = g_\theta$ for a low-dimensional parameter $\theta \in \Theta$. To make this more precise, we consider the theoretical characteristic function of the marginal Y_1:

$$\phi_\theta(u) = \mathbb{E}_\theta[\exp(\mathrm{i}uY_1)] = \exp(-u^\alpha \, \|g_\theta\|_\alpha^\alpha), \qquad (u \in \mathbb{R}) \tag{15}$$

where $\|g_\theta\|_\alpha^\alpha = \int_{\mathbb{R}} |g_\theta(s)|^\alpha \,\mathrm{d}s$ denotes the ordinary $\mathcal{L}^\alpha(\mathbb{R})$-norm—this specific form for the characteristic function follows since Y_1 is a SαS-distributed, see [50, Chapter 3]. It seems clear that it is unreasonable to deduce high-dimensional parameters θ from (15), especially parameters retaining to path or dependence properties such as a self-similarity index. This of course relates to a discussion on uniqueness of the spectral representation of (SαS-) moving averages as mentioned previously; see also [47].

A natural step towards a solution would be to instead consider the characteristic function of the joint distribution $(Y_1, \dots, Y_m)$:

$$\varphi_\theta(u_1, \dots, u_m) = \mathbb{E}_\theta\big[e^{i \sum_{k=1}^m u_k Y_k}\big] = \exp\Big(-\Big\|\sum_{k=1}^m u_k g_\theta(\,\cdot\, + k)\Big\|_\alpha^\alpha\Big)$$

and it is empirical counterpart:

$$V_n(Y; f_u) = \frac{1}{n-m} \sum_{s=0}^{n-m} f_u(Y_{s+1}, \dots, Y_{s+m}), \qquad (u \in \mathbb{R}^m),$$

where $f_u(y_1, \dots, y_m) = \exp(i \sum_{k=1}^m u_k y_k)$. This immediately places us in the framework of $m > 1$ and $d = 1$ of Paper D. Moreover, as already mentioned we need to vary $u \in \mathbb{R}^m$ to obtain finite dimensional convergence of our (empirical) processes (which are now of course called fields) and therefore we require the full generality of our newly developed framework in Paper D: $d, m > 1$.

Paper E then discusses the assumptions on the kernel and (theoretical) characteristic function necessary for the statistical methodology to work. Indeed, these assumptions falls into two categories; one related to the parameter identification from the m-dimensional marginal distributions as we have just discussed, and a category related to our linearization type argument.

Of course Paper E also studies several important examples including Ornstein–Uhlenbeck type processes and we are finally able to defeat the lfsm and provide a full minimal contrast estimator for this process without relying on a plug-in method. It should be clear from the general formulation that the methodology could in principle handle a very large class of parametric moving average processes and provide Gaussian limit theorems in the case where this is a reasonable goal to pursue, i.e. under appropriate behaviour of the kernel g_θ at 0 and ∞—as we have alluded to at several times.

References

[1] David Applebaum (2009). »*Lévy processes and stochastic calculus*«. 2nd edition. Cambridge Studies in Advanced Mathematics 116. Cambridge University Press. ISBN: 978-0-511-80978-1. DOI: 10.1017/CBO9780511809781.

[2] Ole E. Barndoff-Nielsen and Andreas Basse-O'Connor (2011). »*Quasi Ornstein–Uhlenbeck processes*«. Bernoulli 17(3), 916–941. DOI: 10.3150/10-BEJ311.

[3] Ole E. Barndorff-Nielsen, José Manuel Corcuera and Mark Podolskij (2013). »*Limit theorems for functionals of higher order differences of Brownian semi-stationary processes*«. In: »*Prokhorov and contemporary probability theory*«. Edited by Albert N. Shiryaev, S. R. S. Varadhan and Ernst L. Presman. Springer Berlin Heidelberg, 69–96. ISBN: 978-3-642-33549-5.

[4] Ole E. Barndorff-Nielsen and Neil Shephard (2003). »*Realized power variation and stochastic volatility models*«. Bernoulli 9(2), 243–265. DOI: 10.3150/bj/1068128977.

[5] Ole E. Barndorff-Nielsen and Niel Shephard (2001). »*Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics*«. Journal of the Royal Statistical Society 63(2), 167–241. doi: 10.1111/1467-9868.00282.

[6] Andreas Basse and Jan Pedersen (2009). »*Lévy driven moving averages and semimartingales*«. Stochastic Processes and Their Applications 119(9), 2970–2991. doi: 10.1016/j.spa.2009.03.007.

[7] Andreas Basse-O'Connor, Claudio Heinrich and Mark Podolskij (2019). »*On limit theory for functionals of stationary increments Lévy driven moving averages*«. Electronic Journal of Probability 24(79), 42 pp. doi: 10.1214/19-EJP336.

[8] Andreas Basse-O'Connor, Raphaël Lachièze-Rey and Mark Podolskij (2017). »*Power variation for a class of stationary increments Lévy driven moving averages*«. The Annals of Probability 45(6B), 4477–4528. doi: 10.1214/16-AOP1170.

[9] Andreas Basse-O'Connor, Mikkel Slot Nielsen, Jan Pedersen and Victor Rohde (2020). »*Stochastic delay differential equations and related autoregressive models*«. Stochastics: An International Journal of Probability and Stochastic Processes 92(3), 454–477. doi: 10.1080/17442508.2019.1635601.

[10] Andreas Basse-O'Connor and Mark Podolskij (2017). »*On critical cases in limit theory for stationary increments Lévy driven moving averages*«. Stochastics 89(1), 360–383. doi: 10.1080/17442508.2016.1191493.

[11] Andreas Basse-O'Connor, Mark Podolskij and Christoph Thäle (2019). »*A Berry–Esseén theorem for partial sums of functionals of heavy-tailed moving averages*«. arXiv: 1904.06065.

[12] Andreas Basse-O'Connor, Claudio Heinrich and Mark Podolskij (2018). »*On limit theory for Lévy semi-stationary processes*«. Bernoulli 24(4A), 3117–3146. doi: 10.3150/17-BEJ956.

[13] Peter J. Brockwell (2004). »*Representations of continuous-time ARMA processes*«. Journal of Applied Probability 41, 375–382.

[14] Peter J. Brockwell and Richard A. Davis (2006). »*Time series: Theory and methods*«. 2nd edition. Springer Series in Statistics. Springer. isbn: 978-1-4419-0319-8. doi: 10.1007/978-1-4419-0320-4.

[15] Stamatis Cambanis, Clyde D. Jr. Hardin and Aleksander Weron (1987). »*Ergodic properties of stationary stable processes*«. Stochastic Processes and Their Applications 24(1), 1–18. doi: 10.1016/0304-4149(87)90024-X.

[16] Stamatis Cambanis and Makoto Maejima (1989). »*Two classes of self-similar stable processes with stationary increments*«. Stochastic Processes and their Applications 32(2), 305–329. issn: 0304-4149. doi: 10.1016/0304-4149(89)90082-3.

[17] Sourav Chatterjee (2009). »*Fluctuations of eigenvalues and second order Poincaré inequalities*«. Probability Theory and Related Fields 143, 1–40. doi: 10.1007/s00440-007-0118-6.

[18] Louis H. Y. Chen, Larry Goldstein and Qi-Man Shao (2011). »*Normal approximation by Stein's method*«. Probability and Its Applications. Springer-Verlag Berlin Heidelberg. ISBN: 978-3-642-15006-7. DOI: 10.1007/978-3-642-15007-4.

[19] Donald L. Cohn (1972). »*Measurable choice of limit points and the existence of separable and measurable processes*«. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 22, 161–165. DOI: 10.1007/BF00532735.

[20] José Manuel Corcuera, Emil Hedevang, Mikko S. Pakkanen and Mark Podolskij (2013). »*Asymptotic theory for Brownian semi-stationary processes with application to turbulence*«. Stochastic Processes and Their Applications 123(7), 2552–2574. DOI: 10.1016/j.spa.2013.03.011.

[21] Richard A. Davis, Mikkel Slot Nielsen and Victor Rohde (2020). »*Stochastic differential equations with a fractionally filtered delay: a semimartingale model for long-range dependent processes*«. Bernoulli 26(2), 799–827. DOI: 10.3150/18-BEJ1086.

[22] Joseph L. Doob (1990). »*Stochastic processes*«. John Wiley & Sons, Inc. ISBN: 0-471-52369-0.

[23] Florian Fuchs and Robert Stelzer (2013). »*Mixing conditions for multivariate infinitely divisible processes with an application to mixed moving averages and the supOU stochastic volatility model*«. ESAIM: PS 17, 455–471. DOI: 10.1051/ps/2011158.

[24] Kerstin Gärtner and Mark Podolskij (2015). »*On non-standard limits of Brownian semi-stationary processes*«. Stochastic Processes and their Applications 125(2), 653–677. DOI: 10.1016/j.spa.2014.09.019.

[25] Claudio Heinrich (2017). »*Fine scale properties of ambit fields: Limit theory and simulation*«. PhD thesis. URL: https://data.math.au.dk/publications/phd/2017/math-phd-2017-ch.pdf.

[26] Jørgen Hoffman-Jørgensen (1973). »*Existence of measurable modifications of stochastic processes*«. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 25, 205–207. DOI: 10.1007/BF00535892.

[27] Jørgen Hoffman-Jørgensen (1994). »*Probability with a view towards statistics*«. Volume II. Chapman & Hall. ISBN: 0-412-05231-8.

[28] Jean Jacod and Philip E. Protter (2012). »*Discretization of processes*«. Stochastic Modelling and Applied Probability 67. Springer-Verlag Berlin Heidelberg. ISBN: 978-3-642-24126-0. DOI: 10.1007/978-3-642-24127-7.

[29] Kari Karhunen (1950). »*Über die Struktur stationärer zufälliger Funktionen*«. Arkiv för Matematik 1(2), 141–160. DOI: 10.1007/BF02590624.

[30] Stanislaw Kwapién and Wojbor Woyczyński (1992). »*Random series and stochastic integrals: Single and multiple*«. Probability and Its Applications. Birkhäuser. ISBN: 978-0-8176-4198-6.

[31] Günter Last, Giovanni Peccati and Matthias Schulte (2016). »*Normal approximation on Poisson spaces: Mehler's formula, second order Poincaré inequalities and stabilization*«. Probability Theory and Related Fields 165(3), 667–723. doi: 10.1007/s00440-015-0643-7.

[32] Günter Last and Mathew D. Penrose (2011). »*Poisson process Fock space representation, chaos expansion and covariance inequalities*«. Probability Theory and Related Fields 150(3–4), 663–690. doi: 10.1007/s00440-010-0288-5.

[33] Günter Last and Mathew D. Penrose (2017). »*Lectures on the Poisson process*«. Institute of Mathematical Statistics Textbooks 7. Cambridge University Press. isbn: 978-1-316-10447-7. doi: 10.1017/9781316104477.

[34] Will E. Leland, Murad S. Taqqu, Walter Willinger and Daniel V. Wilson (1994). »*On the self-similar nature of the Ethernet traffic (extended version)*«. IEEE/ACM Transactions on Networking 2(1), 1–15. doi: 10.1109/90.282603.

[35] Benoit B. Mandelbrot (1997). »*Fractals and scaling in finance: Discontinuity, concentration, risk. Selecta volume E*«. Springer-Verlag New York. isbn: 978-0-387-98363-9. doi: 10.1007/978-1-4757-2763-0.

[36] Benoit B. Mandelbrot and John W. Van Ness (1968). »*Fractional Brownian motions, fractional noises and applications*«. SIAM Review 10(4), 422–437. doi: 10.1137/1010093.

[37] Stepan Mazur, Dmitry Otryakhin and Mark Podolskij (2020). »*Estimation of the linear fractional stable motion*«. Bernoulli 26(1), 226–252. doi: 10.3150/19-BEJ1124.

[38] Mikkel Slot Nielsen (2019). »*Continuous-time modeling with Lévy-driven moving averages: Representations, limit theorems and other properties*«. PhD thesis. url: https://data.math.au.dk/publications/phd/2019/math-phd-2019-msn.pdf.

[39] Dmitry Otryakhin (2019). »*Estimation and numerical simulation of the linear fractional stable motion and related processes*«. PhD thesis. url: https://data.math.au.dk/publications/phd/2019/math-phd-2019-do.pdf.

[40] Giovanni Peccati and Cengbo Zheng (2010). »*Multi-dimensional Gaussian fluctuations on the Poisson space*«. Electronic Journal of Probability 15(48), 1487–1527. doi: 10.1214/EJP.v15-813.

[41] Vladas Pipiras and Murad S. Taqqu (2003). »*Central limit theorems for partial sums of bounded functionals of infinite-variance moving averages*«. Bernoulli 9(5), 833–855. doi: 10.3150/bj/1066418880.

[42] Vladas Pipiras and Murad S. Taqqu (2017). »*Long-range dependence and self-similarity*«. Cambridge Series in Statistical and Probabilistic Mathematics 45. Cambridge University Press. isbn: 978-1-107-03946-9.

[43] Mark Podolskij and Mathias Vetter (2010). »*Understanding limit theorems for semimartingales: a short survey*«. Statistica Neerlandica 64(3), 329–351. doi: 10.1111/j.1467-9574.2010.00460.x.

[44] Balram S. Rajput and Jan Rosiński (1989). »*Spectral representations of infinitely divisible processes*«. Probability Theory and Related Fields 82(3), 451–487. doi: 10.1007/BF00339998.

[45] Sidney I. Resnick (2007). »*Heavy-tail phenomena: Probabilistic and statistical modeling*«. Springer Series in Operations Research and Financial Engineering. Springer. isbn: 978-0-387-24272-9.

[46] Jan Rosinski (1989). »*On path properties of certain infinitely divisible processes*«. Stochastic Processes and their Applications 33(1), 73–87. doi: 10.1016/0304-4149(89)90067-7.

[47] Jan Rosiński (1994). »*On uniquness of the spectral representation of stable processes*«. Journal of Theoretical Probability 7(3), 615–634. doi: 10.1007/BF02213572.

[48] Jan Rosiński (1995). »*On the structure of stationary stable processes*«. The Annals of Probability 23(3), 1163–1187. doi: 10.1214/aop/1176988178.

[49] Jan Rosiński (2018). »*Representations and isomorphism identities for infinitely divisible processes*«. The Annals of Probability 46(6), 3229–3274. doi: 10.1214/17-AOP1246.

[50] Gennady Samorodnitsky and Murad Taqqu (1994). »*Stable non-Gaussian random processes: Stochastic models with infinite variance*«. Stochastic modeling. Chapmann and Hall, New York. isbn: 978-0-412-05171-5.

[51] Ken-Iti Sato (1999). »*Lévy processes and infinitely divisible distributions*«. Cambridge Studies in Advanced Mathematics 68. Cambridge University Press. isbn: 978-0-521-55302-5.

[52] Stilian Stoev and Murad S. Taqqu (2004). »*Simulation methods for linear fractional stable motion and FARIMA using the Fast Fourier Transform*«. Fractals 12(1), 95–121. doi: 10.1142/S0218348X04002379.

[53] Aad W. van der Vaart and Jon A. Wellner (1996). »*Weak convergence and empirical processes: with applications to statistics*«. Springer. isbn: 978-1-4757-2547-6.

[54] Herman Wold (1953). »*A study in the analysis of stationary time series*«. 2nd edition. Almqvist & Wiksell.

A PAPER

A Limit Theorem for a Class of Stationary Increments Lévy Moving Average Processes with Multiple Singularities

Mathias Mørck Ljungdahl

Department of Mathematics, Aarhus University

Mark Podolskij

Department of Mathematics, Aarhus University

Abstract. In this paper we present some new limit theorems for power variations of stationary increments Lévy moving average processes. Recently, such asymptotic results have been investigated in [5, 6] under the assumption that the kernel function potentially exhibits a singular behaviour at 0. The aim of this work is to demonstrate how some of the results change when the kernel function has multiple singularity points. Our paper is also related to the article [8] that studied the same mathematical question for the class of Brownian semi-stationary models.

Key words: Lévy processes, limit theorems, moving averages, fractional processes, stable convergence, high frequency data

AMS 2010 subject classifications: Primary 60F05, 60F15, 60G22; secondary 60G48, 60H05

1 Introduction

In recent years limit theorems and statistical inference for high frequency observations of stochastic processes have received a great deal of attention. The most prominent class of high frequency statistics are power variations that have been proved to be of immense importance for the analysis of the fine structure of an underlying stochastic process. The asymptotic theory for power variations and related statistics has been intensively studied in the setting of Itô semimartingales, fractional Brownian motion and Brownian semi-stationary processes, to name just a few; see for example [2, 3, 4, 7, 9] among many others.

In the recent work [5, 6] power variations of stationary increments Lévy moving average processes have been investigated in details. These are continuous-time stochastic processes $(X_t)_{t\geq 0}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, that are given by

$$X_t = \int_{-\infty}^{t} \big(g(t-s) - g_0(-s)\big)\, \mathrm{d}L_s, \tag{1.1}$$

where $L = (L_t)_{t\in\mathbb{R}}$ is a symmetric Lévy process on $\mathbb{R}$ with $L_0 = 0$ and without Gaussian component. Moreover, $g, g_0 : \mathbb{R} \to \mathbb{R}$ are deterministic functions vanishing on $(-\infty, 0)$. The most prominent subclasses include Lévy moving average processes, which correspond to the setting $g_0 = 0$, and the linear fractional stable motion, which is obtained by taking $g(s) = g_0(s) = s_+^\alpha$ and L being a symmetric β-stable Lévy process with $\beta \in (0, 2)$. The latter is a self-similar process with index $H = \alpha + 1/\beta$; see [12].

We introduce the kth order increments $\Delta_{i,k}^n X$ of X, $k \in \mathbb{N}$, that are defined by

$$\Delta_{i,k}^n X := \sum_{j=0}^{k} (-1)^j \binom{k}{j} X_{(i-j)/n}, \qquad i \geq k. \tag{1.2}$$

For example, we have that $\Delta_{i,1}^n X = X_{\frac{i}{n}} - X_{\frac{i-1}{n}}$ and $\Delta_{i,2}^n X = X_{\frac{i}{n}} - 2X_{\frac{i-1}{n}} + X_{\frac{i-2}{n}}$. The main statistic of interest is the power variation computed on the basis of kth order increments:

$$V(X, p; k)_n := \sum_{i=k}^{n} |\Delta_{i,k}^n X|^p, \qquad p > 0. \tag{1.3}$$

A variety of asymptotic results has been shown for the statistic $V(X, p; k)_n$ in [5, 6]. The mode of convergence and possible limits heavily depend on the interplay between the power p, the form of the kernel function g and the Blumenthal–Getoor index of L. We recall that the Blumenthal–Getoor index is defined via

$$\beta := \inf\Big\{r \geq 0 : \int_{-1}^{1} |x|^r \, \nu(\mathrm{d}x) < \infty\Big\} \in [0, 2], \tag{1.4}$$

where ν denotes the Lévy measure of L. It is well-known that $\sum_{s\in[0,1]} |\Delta L_s|^p$ is finite when $p > \beta$, while it is infinite for $p < \beta$. Here $\Delta L_s = L_s - L_{s-}$ where $L_{s-} = \lim_{u\uparrow s, u<s} L_u$. To formulate the results of [5, 6], we introduce the following set of assumptions on g, g_0 and ν:

Assumption (A). The function $g : \mathbb{R} \to \mathbb{R}$ satisfies the condition

$$g(t) \sim c_0 t^\alpha \qquad \text{as } t \downarrow 0 \qquad \text{for some } \alpha > 0 \text{ and } c_0 \neq 0, \tag{1.5}$$

where $g(t) \sim f(t)$ as $t \downarrow 0$ means that $\lim_{t\downarrow 0} g(t)/f(t) = 1$. For some $w \in (0, 2]$ then $\limsup_{t\to\infty} \nu(\{x : |x| \geq t\}) t^w < \infty$ and $g - g_0$ is a bounded function in $\mathcal{L}^w(\mathbb{R}_+)$. Furthermore, g is k-times continuous differentiable on $(0, \infty)$ and there exists a $\delta > 0$ such that $|g^{(k)}(t)| \leq K t^{\alpha-k}$ for all $t \in (0, \delta)$, $|g^{(k)}|$ is decreasing on (δ, ∞) and $g^{(j)} \in \mathcal{L}^w((\delta, \infty))$ for $j \in \{1, k\}$.

Assumption (A-log). In addition to Assumption (A) suppose that

$$\int_{\delta}^{\infty} |g^{(k)}(s)|^w \, \big|\log(|g^{(k)}(s)|)\big| \, \mathrm{d}s < \infty.$$

Intuitively speaking, Assumption (A) says that $g^{(k)}$ may have a singularity at 0 when α is small, but it is smooth outside of 0. The theorem below has been proved in [5, 6]. We recall that a sequence of $\mathbb{R}^d$-valued random variables $(Y_n)_{n\in\mathbb{N}}$ is said to converge stably in law towards a random variable Y, defined on an extension of the original probability space $(\Omega, \mathcal{F}, \mathbb{P})$, whenever the joint convergence in distribution $(Y_n, Z) \xrightarrow{\mathcal{L}} (Y, Z)$ holds for any $\mathcal{F}$-measurable Z; in this case we use the notation $Y_n \xrightarrow{\mathcal{L}\text{-s}} Y$. We refer to [1, 11] for a detailed exposition of stable convergence.

Theorem 1.3 ([5, Theorem 1.1(i)] and [6, Theorem 1.2(i)]).
Suppose Assumption (A) holds, the Blumenthal–Getoor index satisfies $\beta < 2$ and $p > \beta$. If $w = 1$ assume that Assumption (A-log) holds. Then we obtain the following cases:

(i) *When $\alpha < k - 1/p$, then we have the stable convergence*

$$\begin{aligned} n^{\alpha p} V(X, p; k)_n &\xrightarrow{\mathcal{L}\text{-s}} |c_0|^p \sum_{m: T_m \in [0,1]} |\Delta L_{T_m}|^p V_m \\ \text{with} \qquad V_m &= \sum_{l=0}^{\infty} |h_k(l + U_m)|^p, \end{aligned} \tag{1.6}$$

where $(T_m)_{m\in\mathbb{N}}$ denotes the jump times of L, $(U_m)_{m\in\mathbb{N}}$ is an i.i.d. $\mathcal{U}(0,1)$-distributed sequence independent of L, and the function h_k is defined by

$$h_k(x) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} (x - j)_+^\alpha \qquad \text{with} \qquad y_+ = \max\{y, 0\}. \tag{1.7}$$

(ii) *When $\alpha = k - 1/p$ and additionally $1/p + 1/w > 1$, then we have*

$$\frac{n^{\alpha p}}{\log(n)} V(X, p; k)_n \xrightarrow{\mathbb{P}} |c_0 q_{k,\alpha}|^p \sum_{s \in (0,1]} |\Delta L_s|^p \qquad \text{with} \qquad q_{k,\alpha} := \prod_{i=0}^{k-1} (\alpha - i). \tag{1.8}$$

We remark that the first-order asymptotic theory of [5, Theorem 1.1] includes two more regimes: an ergodic type limit theorem in the setting $p < \beta$, $\alpha < k - 1/\beta$ and convergence in probability towards a random integral in the setting $p \geq 1$, $\alpha > k - 1/\max\{p, \beta\}$. However, in this paper we concentrate on the results of Theorem 1.3, which are quite non-standard in the literature. More specifically, our aim is to extend the theory of Theorem 1.3 to kernels g that exhibit multiple singularities. We call a point $x \in \mathbb{R}_+$ a singularity point when the kth derivative $g^{(k)}$ of g explodes at x. Note that under Assumption (A) and condition $\alpha \leq k - 1/p$ the function g has only one singularity point at $x = 0$. In practical applications a singularity point $x \in \mathbb{R}_+$ leads to a strong feedback effect stemming from the past jumps around the time $t - x$. Such effects has been discussed in the context of turbulence modelling in [8].

We will show that the limits in Theorem 1.3(i) and (ii) will be affected by the presence of multiple singularity points. More precisely, we will see that the increments $\Delta_{i,k}^n X$ can be heavily influenced by the jumps of L that happened in the past, and the time delay is determined by the singularity points of g. The obtained result is similar in spirit to the work [8] that studied quadratic variation of Brownian semi-stationary processes under multiple singularities of the kernel g. Furthermore, we will prove that in general the stable convergence in Theorem 1.3(i) only holds along a subsequence.

The paper is structured as follows. Section 2 presents the main results of the article. Proofs are collected in Section 3.

2 Main Results

We consider stationary increments Lévy moving average processes as defined at (1.1) and recall that the driving motion L is a pure jump Lévy process with Lévy measure ν. Now, we introduce the condition on the kernel function g:

Assumption (B). For some $w \in (0,2]$, $\limsup_{t\to\infty} \nu(\{x : |x| \geq t\})t^w < \infty$ and $g - g_0$ is a bounded function in $\mathcal{L}^w(\mathbb{R}_+)$. Furthermore, there exist points $0 = \theta_0 < \theta_1 < \cdots < \theta_l$ such that the following properties hold:

(i) $g(t) \sim c_0 t^{\alpha_0}$ as $t \downarrow 0$ for some $\alpha_0 > 0$ and $c_0 \neq 0$.

(ii) $g(t) \sim c_z |t-\theta_z|^{\alpha_z}$ as $t \to \theta_z$ for some $\alpha_z > 0$ and $c_z \neq 0$ and for all $z \in \{1,\dots,l\}$.

(iii) $g \in C^k(\mathbb{R}_+ \setminus \{\theta_0,\dots,\theta_l\})$.

(iv) There exist $\delta, K > 0$ such that $|g^{(k)}(t)| \leq K|t-\theta_z|^{\alpha_z - k}$ for all $t \in (\theta_z - \delta, \theta_z + \delta)\setminus\{\theta_z\}$, for any $z = 0,\dots,l$. Furthermore, there exists a $\delta' > 0$ such that $|g^{(k)}|$ is decreasing on $(\theta_l + \delta', \infty)$ and $g^{(j)} \in \mathcal{L}^w((\theta_l + \delta', \infty))$ for $j \in \{1,k\}$.

Let us give some remarks on Assumption (B). First of all, conditions (B)(i) and (B)(ii), which are direct extensions of (1.5), mean that for small powers $\alpha_z > 0$ the points θ_z are singularities of g in the sense that $g^{(k)}(\theta_z)$ does not exist. On the other hand, condition (B)(iii) states that there exist no further singularities. The parameter w is by no means unique. It simultaneously describes the tail behaviours of the Lévy measure ν and the integrability of the function $|g^{(k)}|$, which exhibit a trade-off. When L is β-stable we always take $w = \beta$. Furthermore, Assumption (B) guarantees the existence of X_t for all $t \geq 0$. Indeed, it follows from [10, Theorem 7] that the process X is well-defined if and only if for all $t \geq 0$,

$$\int_{-t}^{\infty} \int_{\mathbb{R}} \Big(|f_t(s)x|^2 \wedge 1 \Big) \nu(\mathrm{d}x)\,\mathrm{d}s < \infty, \tag{2.1}$$

where $f_t(s) = g(t+s) - g_0(s)$. By adding and subtracting g to f_t it follows by Assumption (B) and the mean value theorem that f_t is a bounded function in $\mathcal{L}^w(\mathbb{R}_+)$. For all $\epsilon > 0$, Assumption (B) implies that

$$\int_{\mathbb{R}} (|yx|^2 \wedge 1)\,\nu(\mathrm{d}x) \leq K\Big(\mathbb{1}_{\{|y|\leq 1\}} |y|^w + \mathbb{1}_{\{|y|>1\}} |y|^{\beta+\epsilon} \Big),$$

which shows (2.1) since f_t is a bounded function in $\mathcal{L}^w(\mathbb{R}_+)$.

Remark 2.2 (Toy example).
Recall the following well-known results about the power variation of a pure jump Lévy process L:

$$V(L,p;k)_n \xrightarrow{\mathbb{P}} \sum_{s\in[0,1]} |\Delta L_s|^p < \infty$$

for any $k \geq 1$ and any $p > \beta$. Let us now consider a simple stationary increments Lévy moving average process X with $g_0 = 0$ and $g(x) = \mathbb{1}_{[0,1]}(x)$. In this case we may call the points $\theta_0 = 0$ and $\theta_1 = 1$ the singularities of g, although they do not precisely correspond to conditions (B)(i) and (B)(ii), and we observe that $X_t = L_t - L_{t-1}$. Hence, we obtain the convergence in probability

$$V(X,p;k)_n \xrightarrow{\mathbb{P}} \sum_{s\in[0,1]} |\Delta L_s|^p + \sum_{s\in[-1,0]} |\Delta L_s|^p$$

for any $k \geq 1$ and any $p > \beta$. This result demonstrates that even in the simplest setting multiple singularities lead to a different limit. $\diamond$

It turns out that only the minimal powers among $\alpha_0, \dots, \alpha_l$ determine the asymptotic behaviour of the statistic $V(X,p;k)_n$. Thus, we define

$$\alpha := \min\{\alpha_0, \dots, \alpha_l\} \qquad \text{and} \qquad \mathcal{A} := \{z : \alpha_z = \alpha\}. \tag{2.2}$$

Furthermore, we introduce the notation $h_{k,0} := h_k$ and

$$h_{k,z}(x) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} |x-j|^{\alpha_z} \qquad \text{for } z \in \{1, \dots, l\}. \tag{2.3}$$

In the main result below we consider a subsequence $(n_j)_{j\in\mathbb{N}}$ such that the following condition holds:

$$\lim_{j\to\infty} \{n_j \theta_z\} = \eta_z \in [0,1] \qquad \text{for all } z \in \mathcal{A}, \tag{2.4}$$

where $\{x\}$ denotes the fractional part of $x \in \mathbb{R}$. Obviously, such a subsequence always exists since $(\{n\theta_z\})_{n\in\mathbb{N}}$ is a bounded sequence. Sometimes we will require a stronger condition, which is analogous to Assumption (A-log):

Assumption (B-log). Assumption (B) holds and we have that

$$\int_{\theta_l+\delta'}^{\infty} |g^{(k)}(t)|^w |\log(|g^{(k)}(t)|)| \mathrm{d}t < \infty.$$

The main result of the paper is the following theorem.

Theorem 2.4. *Suppose that Assumption (B) holds, the Blumenthal–Getoor index satisfies $\beta < 2$ and $p > \beta$. If $w = 1$ assume that Assumption (B-log) holds. Recall the notations* (2.2) *and* (2.3)*. Then we obtain the following cases:*

(i) *When $\max_{0\leq z\leq l} \alpha_z < k - 1/p$ and condition* (2.4) *holds, then we have the stable convergence*

$$\begin{aligned} n_j^{\alpha p} V(X,p;k)_{n_j} &\xrightarrow{\mathcal{L}\text{-s}} \sum_{z\in\mathcal{A}} |c_z|^p \sum_{m:T_m\in[-\theta_z,1-\theta_z]} |\Delta L_{T_m}|^p V_m^z \\ \text{with} \qquad V_m^z &= \sum_{r\in\mathbb{Z}} |h_{k,z}(r+1-\{U_m+\eta_z\})|^p. \end{aligned} \tag{2.5}$$

as $j \to \infty$, where $(U_m)_{m\in\mathbb{N}}$ is an i.i.d. $\mathcal{U}(0,1)$-distributed sequence independent of L.

(ii) *Let $\alpha = \alpha_0 = \cdots = \alpha_l = k - 1/p$. Assume that the functions $f_z : \mathbb{R}_+ \to \mathbb{R}$ defined by $f_z(x) = g(x)/|x-\theta_z|^\alpha$ are in $C^k((\theta_z - \delta, \theta_z + \delta))$ for all $\delta < \max_{1\leq j\leq l}(\theta_j - \theta_{j-1})$. If $1/p + 1/w > 1$, then we have*

$$\frac{n^{\alpha p}}{\log(n)} V(X,p;k)_n \xrightarrow{\mathbb{P}} |q_{k,\alpha}|^p \sum_{z=0}^{l} |c_z|^p (1 + \mathbb{1}_{\{z\geq 1\}}) \sum_{m:T_m\in[-\theta_z,1-\theta_z]} |\Delta L_{T_m}|^p, \tag{2.6}$$

where the constant $q_{k,\alpha}$ has been introduced in Theorem 1.3(ii).

We remark that the stable convergence in Theorem 2.4(i) only holds along the subsequence $(n_j)_{j\in\mathbb{N}}$, which is seen from the form of the limit in (2.5) that depends on (η_z). The original statistic $(n^{\alpha p}V(X,p;k)_n)_{n\in\mathbb{N}}$ is tight, but does not converge except when $\theta_z \in \mathbb{N}$ for all $z \in \mathcal{A}$. On the other hand, in Theorem 2.4(ii) we do not require to consider a subsequence.

Notice that the interval $[-\theta_z, 1-\theta_z]$, which appears in Theorem 2.4, is the set $[0,1]$ shifted by θ_z to the left. Given the discussion of Remark 2.2, such a shift in the limit is not really surprising. We recall that a similar phenomenon has been discovered in [8] in the context of Brownian semi-stationary processes. These are stochastic processes $(Y_t)_{t\geq 0}$ defined by

$$Y_t = \int_{-\infty}^{t} g(t-s)\sigma_s \,\mathrm{d}W_s,$$

where W is a two-sided Brownian motion and $(\sigma_t)_{t\in\mathbb{R}}$ is a càdlàg process. When the kernel function g satisfies conditions (B)(i) and (B)(ii) along with some further assumptions, which in particular ensure the existence of Y_t, the authors have shown the following convergence in probability (see [8, Theorem 3.2]):

$$\frac{1}{n\tau_n^2} V(Y,2;k)_n \xrightarrow{\mathbb{P}} \sum_{z\in\mathcal{A}} \pi_z \int_{-\theta_z}^{1-\theta_z} \sigma_s^2 \,\mathrm{d}s,$$

where $\tau_n^2 = \mathbb{E}[(\Delta_{k,k}^n G)^2]$ with $G_t = \int_{-\infty}^t g(t-s)\,\mathrm{d}W_s$, and the probability weights $(\pi_z)_{z\in\mathcal{A}}$ are given by

$$\pi_z = \frac{c_z^2 \,\|h_{k,z}\|^2_{\mathcal{L}^2(\mathbb{R})}}{\sum_{z\in\mathcal{A}} c_z^2 \,\|h_{k,z}\|^2_{\mathcal{L}^2(\mathbb{R})}}.$$

Hence, we observe the same shift phenomenon in the integration region as in Theorem 2.4.

3 Proofs

Throughout this section all positive constants are denoted by C although they may change from line to line. We will divide the proof of Theorem 2.4 into several steps. First, we will show the statements (2.5) and (2.6) for a compound Poisson process. In the second step we will decompose the jump measure of L into jumps that are bigger than ϵ and jumps that are smaller than ϵ. The big jumps form a compound Poisson process and hence the claim follows from the first step. Finally, we prove negligibility of small jumps when $\epsilon \to 0$.

We start with an important proposition.

Proposition 3.1. *Let $T = (T_1, \ldots, T_d)$ be a stochastic vector with a density $v : \mathbb{R}^d \to \mathbb{R}_+$. Suppose there exists an open convex set $A \subseteq \mathbb{R}^d$ such that v is continuously differentiable on A and vanishes outside of A. Then, under condition* (2.4)*, it holds that*

$$(\{n_j T + n_j\theta_z\})_{z\in\mathcal{A}} \xrightarrow{\mathcal{L}\text{-s}} (\{U + \eta_z\})_{z\in\mathcal{A}} \qquad \text{as } j\to\infty, \tag{3.1}$$

where $\{x\}$ denotes the fractional parts of the vector $x \in \mathbb{R}^d$ and $x+a$, $a \in \mathbb{R}$, is componentwise addition. Here $U = (U_1, \ldots, U_d)$ consists of i.i.d. $\mathcal{U}(0,1)$-distributed random variables defined on an extension of the space $(\Omega, \mathcal{F}, \mathbb{P})$ and being independent of $\mathcal{F}$.

Proof. We first show the stable convergence

$$\{nT\} \xrightarrow{\mathcal{L}\text{-s}} U. \tag{3.2}$$

This statement has already been shown in [5, Lemma 4.1], but we demonstrate its proof for completeness. Let $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a C^1-function, which vanishes outside some closed ball in $A \times \mathbb{R}^d$. We claim that there exists a finite constant $K > 0$ such that for all $\rho > 0$

$$D_\rho := \left| \int_{\mathbb{R}^d} f(x, \{x/\rho\}) v(x)\,\mathrm{d}x - \int_{\mathbb{R}^k} \left(\int_{[0,1]^d} f(x,u)\,\mathrm{d}u \right) v(x)\,\mathrm{d}x \right| \le K\rho. \tag{3.3}$$

By (3.3) used for $\rho = 1/n$ we obtain that

$$\mathbb{E}[f(T, \{nT\})] \to \mathbb{E}[f(T,U)] \qquad \text{as } n \to \infty. \tag{3.4}$$

Moreover, due to [1, Proposition 2(D")], (3.4) implies the stable convergence $\{nT\} \xrightarrow{\mathcal{L}\text{-s}} U$ as $n \to \infty$. Thus, we need to prove the inequality (3.3). Define $\phi(x,u) := f(x,u)v(x)$. Then it holds by substitution that

$$\int_{\mathbb{R}^d} f(x, \{x/\rho\}) v(x)\,\mathrm{d}x = \sum_{j \in \mathbb{Z}^d} \int_{(0,1]^d} \rho^d \phi(\rho j + \rho u, u)\,\mathrm{d}u$$

and

$$\int_{\mathbb{R}^d} \left(\int_{[0,1]^d} f(x,u)\,\mathrm{d}u \right) v(x)\,\mathrm{d}x = \sum_{j \in \mathbb{Z}^d} \int_{[0,1]^d} \left(\int_{(\rho j, \rho(j+1)]} \phi(x,u)\,\mathrm{d}x \right) \mathrm{d}u.$$

Hence, we conclude that

$$\begin{aligned} D_\rho &\le \sum_{j \in \mathbb{Z}^d} \int_{(0,1]^d} \left| \int_{(\rho j, \rho(j+1)]} \phi(x,u)\,\mathrm{d}x - \rho^d \phi(\rho j + \rho u, u) \right| \mathrm{d}u \\ &\le \sum_{j \in \mathbb{Z}^d} \int_{(0,1]^d} \int_{(\rho j, \rho(j+1)]} |\phi(x,u) - \phi(\rho j + \rho u, u)|\,\mathrm{d}x\,\mathrm{d}u. \end{aligned}$$

Using that A is convex and open, we deduce by the mean value theorem that there exist a positive constant K and a compact set $B \subseteq \mathbb{R}^d \times \mathbb{R}^d$ such that for all $j \in \mathbb{Z}^d$, $x \in (\rho j, \rho(j+1)]$ and $u \in (0,1]^d$ we have

$$|\phi(x,u) - \phi(\rho j + \rho u, u)| \le K\rho \mathbb{1}_B(x,u).$$

Thus, $D_\rho \le K\rho \int_{(0,1]^d} \int_{\mathbb{R}^d} \mathbb{1}_B(x,u)\,\mathrm{d}x\,\mathrm{d}u$, which shows (3.2).

Now, we are ready to prove the statement (3.1). By (3.2) and condition (2.4) we conclude that

$$\left(\{n_j T\}, \{n_j \theta_z\} \right)_{z \in \mathcal{A}} \xrightarrow{\mathcal{L}\text{-s}} (U, \eta_z)_{z \in \mathcal{A}} \qquad \text{as } j \to \infty.$$

Next, consider the map $f : \mathbb{R}^d \times \mathbb{R}^{l'} \to \mathbb{R}^{d \times l'}$, where l' denotes the cardinality of $\mathcal{A}$, given by

$$f(x, y_1, \dots, y_{l'}) = (\{x + y_1\}, \dots, \{x + y_{l'}\}).$$

This map is discontinuous exactly in those points $x, y_1, \dots, y_{l'}$ for which $x_j + y_i \in \mathbb{Z}$ for some $i \in \{1, \dots, l'\}$ and some $j \in \{1, \dots, d\}$. Note that the probability of the limiting variable $(U,\eta_z)_{z\in\mathcal{A}}$ lying in the latter set is 0. Hence, it follows from the continuous mapping theorem for stable convergence that

$$f\big(\{n_jT\},(\{n_j\theta_z\})_{z\in\mathcal{A}}\big)\xrightarrow{\mathcal{L}\text{-s}} f(U,(\eta_z)_{z\in\mathcal{A}})=(\{U+\eta_z\})_{z\in\mathcal{A}}$$

as $j\to\infty$. Since $x=\{x\}+\lfloor x\rfloor$ we have the identity $\{x+y\}=\{\{x\}+\{y\}\}$ and the left hand side becomes

$$f\big(\{n_jT\},(\{n_j\theta_z\})_{z\in\mathcal{A}}\big)=(\{n_jT+n_j\theta_z\})_{z\in\mathcal{A}},$$

which concludes the proof of Proposition 3.1. □

Now, we introduce the notation

$$g_{i,n}(x)=\sum_{j=0}^{k}(-1)^j\binom{k}{j}g((i-j)/n-x), \tag{3.5}$$

and observe the identity

$$\Delta_{i,k}^n X=\int_{\mathbb{R}} g_{i,n}(s)\,\mathrm{d}L_s.$$

The next lemma presents some estimates for the function $g_{i,n}$. Its proof is a straightforward consequence of Assumption (B) and Taylor expansion.

Lemma 3.2. *Suppose that Assumption (B) holds and let $z\in\{1,\dots,l\}$. Then there exists an $N\in\mathbb{N}$ such that for all $n\geq N$ and $i\in\{k,\dots,n\}$ the following holds:*

(i) $|g_{i,n}(x)|\leq C(|i/n-x-\theta_z|^{\alpha_z}+n^{-\alpha_z})$ *for all* $x\in[\frac{i-2k}{n}-\theta_z,\frac{i+2k}{n}-\theta_z]$.

(ii) $|g_{i,n}(x)|\leq Cn^{-k}|(i-k)/n-x-\theta_z|^{\alpha_z-k}$ *for all* $x\in(\frac{i}{n}-\delta-\theta_z,\frac{i-k}{n}-\theta_z)$ *if* $\alpha_z-k<0$.

(iii) $|g_{i,n}(x)|\leq Cn^{-k}|(i-k)/n-x-\theta_z|^{\alpha_z-k}$ *for all* $x\in(\frac{i+k}{n}-\theta_z,\frac{i-k}{n}+\delta-\theta_z)$ *if* $\alpha_z-k<0$.

(iv) $|h_{k,z}(x)|\leq|x-k|^{\alpha-k}$ *for all* $x\geq k+1$ *and* $|h_{k,z}(x)|\leq|x+k|^{\alpha-k}$ *for all* $x\leq -k-1$, *if* $\alpha_z-k<0$.

(v) *For each $\varepsilon>0$ it holds that*

$$n^k|g_{i,n}(s)|\mathbb{1}_{(-\infty,\frac{i}{n}-\varepsilon-\theta_l]}(s)\leq C_\varepsilon\Big(\mathbb{1}_{[-\theta_l-\delta',1-\theta_l]}(s)+\mathbb{1}_{(-\infty,-\theta_l-\delta')}(s)\,|g^{(k)}(-s)|\Big).$$

Furthermore, similar estimates hold for $z=0$ with obvious adjustments that account for the fact that g and $h_{k,0}$ are both vanishing on $(-\infty,0)$.

3.1 Proof of Theorem 2.4 in the Compound Poisson Case

In this subsection we assume that L is a compound Poisson process. Recall that $(T_m)_{m\in\mathbb{N}}$ denotes the jump times of L. Let $\varepsilon>0$ and consider $n_j\in\mathbb{N}$ such that $\varepsilon n_j>4k$. Define the set

$$\Omega_\varepsilon=\Big\{\omega\in\Omega:\begin{array}{ll}\text{(a)} & \text{for all } m\in\mathbb{N} \text{ with } T_m(\omega)\in[-\theta_l,1] \text{ then } |T_m(\omega)-T_i(\omega)|>2\varepsilon\\ & \text{and } T_m(\omega)+\theta_z-\theta_{z'}\notin[T_i(\omega)-2\varepsilon,T_i(\omega)+2\varepsilon] \text{ for all } i\neq m \text{ and}\\ & \text{for all } z,z'\in\{0,\dots,l\}\\ \text{(b)} & \Delta L_s(\omega)=0 \text{ for all } s\in[-\varepsilon-\theta_z,-\theta_z+\varepsilon]\cup[1-\varepsilon-\theta_z,1-\theta_z+\varepsilon]\\ & \text{and for all } z\in\{0,\dots,l\}\Big\}.\end{array}$$

Roughly speaking, on the set Ω_ε the jump times in $[-\theta_l, 1]$ are well separated, their increments are outside a small neighbourhood of $\theta_z - \theta_{z'}$ which in total corresponding statement (a), and there are no jumps around the fixed points $-\theta_z$ and $1-\theta_z$ according to (b). In particular, it obviously holds that $\mathbb{P}(\Omega_\varepsilon) \to 1$ as $\varepsilon \to 0$.

Throughout the proof we assume without loss of generality that $0 \in \mathcal{A}$. Now, we introduce a decomposition, which is central for the proof. Recalling the definition of $g_{i,n}$ at (3.5), we observe the identity

$$\Delta_{i,k}^n X = \sum_{z\in\mathcal{A}} M_{i,n,\varepsilon,z} + \sum_{z\in\mathcal{A}^c} M_{i,n,\varepsilon,z} + R_{i,n,\varepsilon}, \tag{3.6}$$

where

$$M_{i,n,\varepsilon,0} = \int_{\frac{i}{n}-\varepsilon}^{\frac{i}{n}} g_{i,n}(s)\,\mathrm{d}L_s, \qquad M_{i,n,\varepsilon,z} = \int_{\frac{i}{n}-\theta_z-\varepsilon}^{\frac{i}{n}-\theta_z+\frac{\lfloor n\varepsilon\rfloor}{n}} g_{i,n}(s)\,\mathrm{d}L_s \qquad \text{for } z \in \{1,\ldots,l\},$$

$$R_{i,n,\varepsilon} = \int_{-\infty}^{\frac{i}{n}-\theta_l-\varepsilon} g_{i,n}(s)\,\mathrm{d}L_s + \sum_{z=1}^{l} \int_{\frac{i}{n}-\theta_z+\frac{\lfloor n\varepsilon\rfloor}{n}}^{\frac{i}{n}-\theta_{z-1}-\varepsilon} g_{i,n}(s)\,\mathrm{d}L_s.$$

It turns out that the first term $\sum_{z\in\mathcal{A}} M_{i,n,\varepsilon,z}$ is dominating, while the other two are negligible.

3.1.1 Main Terms in Theorem 2.4(i)

In this subsection we consider the dominating term in the decomposition (3.6). We want to prove that, on Ω_ε,

$$n_j^{\alpha p} \sum_{i=k}^{n_j} \Big| \sum_{z\in\mathcal{A}} M_{i,n_j,\varepsilon,z} \Big|^p \xrightarrow{\mathcal{L}\text{-s}} \sum_{z\in\mathcal{A}} |c_z|^p \sum_{m:T_m\in[-\theta_z,1-\theta_z]} |\Delta L_{T_m}|^p V_m^z \qquad \text{as } j\to\infty, \tag{3.7}$$

where the limit has been introduced in (2.5). Let us fix an index $z \in \mathcal{A}$. Then, on Ω_ε, for each jump time $T_m \in (-\theta_z, 1-\theta_z]$ there exists a unique random variable $i_{m,z} \in \mathbb{N}$ such that

$$T_m \in \Big(\frac{i_{m,z}-1}{n} - \theta_z, \frac{i_{m,z}}{n} - \theta_z\Big].$$

We also observe the following implication, which follows directly from the definition of the set Ω_ε:

$$\text{On } \Omega_\varepsilon, \text{ if } M_{i,n,\varepsilon,z} \neq 0 \text{ for some } z \in \mathcal{A} \quad \Longrightarrow \quad M_{i,n,\varepsilon,z'} = 0 \text{ for any } z' \neq z \text{ in } \mathcal{A}.$$

Indeed, this is the consequence of the definition of the term $M_{i,n,\varepsilon,z}$ and the statement

$$T_m(\omega) + \theta_z - \theta_{z'} \notin [T_{m'}(\omega) - 2\varepsilon, T_{m'}(\omega) + 2\varepsilon] \qquad \forall m' \neq m \ \forall z,z' \in \{0,\ldots,l\},$$

which holds on Ω_ε. Hence, we conclude that

$$n^{\alpha p} \sum_{i=k}^{n} \Big| \sum_{z\in\mathcal{A}} M_{i,n,\varepsilon,z} \Big|^p = n^{\alpha p} \sum_{z\in\mathcal{A}} \sum_{i=k}^{n} |M_{i,n,\varepsilon,z}|^p$$

on Ω_ε, and we obtain the representation

$$\begin{aligned} n^{\alpha p} \sum_{i=k}^{n} |M_{i,n,\varepsilon,z}|^p &= V_{n,\varepsilon,z} \qquad \text{with} \\ V_{n,\varepsilon,z} &= n^{\alpha p} \sum_{m:T_m\in(-\theta_z,1-\theta_z]} |\Delta L_{T_m}|^p \sum_{u=-\lfloor n\varepsilon\rfloor}^{\lfloor n\varepsilon\rfloor + v_m^z} |g_{i_{m,z}+u,n}(T_m)|^p, \end{aligned} \tag{3.8}$$

where v_m^z are random variables taking values in $\{-2,-1,0\}$ that are measurable with respect to T_m. If $z=0$ then the sum above is one-sided, i.e. from $u=0$ to $\lfloor n\varepsilon \rfloor$, cf. [5, Eq. (4.2)]. Next, we observe the identity

$$\{nT_m + n\theta_z\} = nT_m + n\theta_z - \lfloor nT_m + n\theta_z \rfloor = nT_m + n\theta_z - (i_{m,z} - 1).$$

Due to Assumption (B), we can write $g(x) = c_z|x-\theta_z|^\alpha f(x)$ with $f(x)\to 1$ as $x\to\theta_z$, for any $z\in\mathcal{A}$ (for $\theta_0=0$ we need to replace $|x|^\alpha$ by x_+^α). This allows us to decompose

$$\begin{aligned}
n^\alpha g\Big(\frac{i_{m,z}+u-r}{n} - T_m\Big) &= c_z n^\alpha \left|\frac{i_{m,z}+u-r}{n} - T_m - \theta_z\right|^\alpha f\Big(\frac{i_{m,z}+u-r}{n} - T_m\Big)\\
&= c_z|u-r+i_{m,z}-nT_m-n\theta_z|^\alpha f\Big(\frac{u-r}{n} + n^{-1}(i_{m,z}-nT_m)\Big)\\
&= c_z|u-r+1-\{nT_m+n\theta_z\}|^\alpha f\Big(\frac{u-r}{n} + n^{-1}(n\theta_z+1-\{nT_m+n\theta_z\})\Big)\\
&= c_z|u-r+1-\{nT_m+n\theta_z\}|^\alpha f\Big(\frac{u-r+1-\{nT_m+n\theta_z\}}{n} + \theta_z\Big).
\end{aligned} \tag{3.9}$$

for any $m\in\mathbb{N}$, $0\le r\le k$ and $z\in\mathcal{A}$. Since $f(x)\to 1$ as $x\to\theta_z$, we find that for any $d\in\mathbb{N}$

$$\begin{aligned}
&\Big(n_j^\alpha g\Big(\frac{i_{m,z}+u-r}{n_j} - T_m\Big)\Big)_{|u|,m\le d,0\le r\le k,z\in\mathcal{A}}\\
&\qquad\xrightarrow{\mathcal{L}\text{-s}} (c_z|u-r+1-\{U_m+\eta_z\}|^\alpha)_{|u|,m\le d,0\le r\le k,z\in\mathcal{A}},
\end{aligned}$$

which holds due to condition (2.4), decomposition (3.9) and Proposition 3.1 (for $\theta_0=0$ we again need to replace $|x|^\alpha$ by x_+^α). Hence, by the continuous mapping theorem for stable convergence we deduce that

$$\big(n_j^\alpha g_{i_{m,z}+u,n_j}(T_m)\big)_{|u|,m\le d,z\in\mathcal{A}} \xrightarrow{\mathcal{L}\text{-s}} \big(c_z h_{k,z}(1+u-\{U_m+\eta_z\})\big)_{|u|,m\le d,z\in\mathcal{A}} \tag{3.10}$$

as $j\to\infty$, which is a key result of the proof. We now define a truncated version of $V_{n,\varepsilon,z}$ introduced in (3.8):

$$V_{n,\varepsilon,z,d} := n^{\alpha p} \sum_{\substack{m\le d:\\ T_m\in(-\theta_z,1-\theta_z]}} |\Delta L_{T_m}|^p \Bigg(\sum_{u=-\lfloor \varepsilon d\rfloor}^{\lfloor \varepsilon d\rfloor + v_m^z} |g_{i_{m,z}+u,n}(T_m)|^p\Bigg).$$

From (3.10) and properties of stable convergence we conclude that

$$(V_{n_j,\varepsilon,z,d})_{z\in\mathcal{A}} \xrightarrow{\mathcal{L}\text{-s}} (V_{\varepsilon,z,d})_{z\in\mathcal{A}} \qquad \text{as } j\to\infty, \tag{3.11}$$

where

$$V_{\varepsilon,z,d} = |c_z|^p \sum_{\substack{m\le d:\\ T_m\in(-\theta_z,1-\theta_z]}} |\Delta L_{T_m}|^p \Bigg(\sum_{u=-\lfloor \varepsilon d\rfloor}^{\lfloor \varepsilon d\rfloor + v_m^z} |h_{k,z}(1+u-\{U_m+\eta_z\})|^p\Bigg).$$

Applying a monotone convergence argument, we deduce the almost sure convergence

$$V_{\varepsilon,z,d} \uparrow V_z = |c_z|^p \sum_{T_m\in(-\theta_z,1-\theta_z]} |\Delta L_{T_m}|^p \Bigg(\sum_{u\in\mathbb{Z}} |h_{k,z}(1+u-\{U_m+\eta_z\})|^p\Bigg) \tag{3.12}$$

as $d \to \infty$, where the second sum on the right hand side is finite, since $|h_{k,z}(x)| \le C|x|^{\alpha-k}$ for large enough $|x|$ and all $z \in \mathcal{A}$, and $\alpha < k-1/p$. In view of (3.11) and (3.12), we are left to proving the convergence

$$\lim_{d\to\infty}\limsup_{n\to\infty}|V_{n,\varepsilon,z,d} - V_{n,\varepsilon,z}| = 0$$

on Ω_ε. Set $K_d = \sum_{m>d:T_m\in(-\theta_z,1-\theta_z]}|\Delta L_{T_m}|^p$ and observe that $K_d \to 0$ as $d \to \infty$, since L is a compound Poisson process. Due to Lemma 3.2 we conclude that $|n^\alpha g_{i,n}(x)| \le C\min\{1, |i/n - x|^{\alpha-k}\}$ and thus

$$|V_{n,\varepsilon,z,d} - V_{n,\varepsilon,z}| \le C\Big(K_d + \sum_{|u|>\lfloor \varepsilon d\rfloor}|u|^{p(\alpha-k)}\Big) \qquad \text{for all } z \in \mathcal{A},$$

and the latter converges to 0 almost surely as $d \to \infty$ because $\alpha < k-1/p$. Consequently, we have shown (3.7). □

3.1.2 Main Terms in Theorem 2.4(ii)

We start with a simple lemma.

Lemma 3.3. *Let $(a_i)_{i\in\mathbb{N}}$ be a sequence of positive real numbers such that $\lim_{i\to\infty} ia_i = 1$. Then it holds that*

$$\lim_{n\to\infty}\frac{1}{\log(n)}\sum_{i=1}^{cn}a_i = 1$$

for any fixed $c \in \mathbb{N}$.

Proof. Due to the assumption of the lemma, we have that $(a_i)_{i\in\mathbb{N}}$ is a bounded sequence and for each $\epsilon > 0$ there exists an $N = N(\epsilon)$ with

$$|a_i - i^{-1}| \le \epsilon i^{-1} \qquad \text{for all } i \ge N.$$

It obviously holds that $\lim_{n\to\infty}\sum_{i=1}^{cn} i^{-1}/\log(n) = 1$. On the other hand, we obtain that

$$\limsup_{n\to\infty}\frac{1}{\log(n)}\sum_{i=N}^{cn}|a_i - i^{-1}| \le \epsilon\limsup_{n\to\infty}\frac{1}{\log(n)}\sum_{i=1}^{cn}i^{-1} = \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we conclude the statement of Lemma 3.3. □

Now, we will again use the decomposition (3.8), which holds on Ω_ε, and treat each term $V_{n,\varepsilon,z}$ separately. We consider $z \ge 1$ and we will show that

$$\frac{1}{\log(n)}\sum_{u=-\lfloor n\varepsilon\rfloor}^{\lfloor n\varepsilon\rfloor+v_m^z}|n^\alpha g_{i_{m,z}+u,n}(T_m) - c_z h_{k,z}(u+1-\{nT_m + n\theta_z\})|^p \longrightarrow 0 \tag{3.13}$$

as $n \to \infty$, for any $m \in \mathbb{N}$. Let us first consider the case $|u| \ge k$. Recall that we have assumed that $f_z(x) = g(x)/|x-\theta_z|^\alpha$ is in $C^k((\theta_z-\delta, \theta_z+\delta))$ for any $\delta < \max_{1\le j\le l}(\theta_j - \theta_{j-1})$. Now, due to identity (3.9) and Taylor expansion of order k, we obtain the bound (cf. [6, Eqs. (4.8) and (4.9)])

$$\sum_{u=-\lfloor n\varepsilon\rfloor}^{\lfloor n\varepsilon\rfloor+v_m^z}\Big|n^\alpha g_{i_{m,z}+u,n}(T_m) - c_z h_{k,z}(u+1-\{nT_m+n\theta_z\})\Big|^p \mathbb{1}_{\{|u|\ge k\}} \le C,$$

for any $\varepsilon < \max_{1\le j\le l}(\theta_j - \theta_{j-1})$. Since $|n^\alpha g_{i_{m,z}+u,n}(T_m)|$ is bounded for any $|u| < k$ due to Lemma 3.2, we deduce the convergence in (3.13).

Next, for large enough $|u|$ we observe the bounds

$$|q_{k,\alpha}|^p a_u \le \left|h_{k,z}(u+1-\{nT_m + n\theta_z\})\right|^p \le |q_{k,\alpha}|^p a_{u-k-1} \qquad \text{where } a_u = |u|^{-1}.$$

Hence, by Lemma 3.3, we conclude the convergence

$$\frac{1}{\log(n)} \sum_{u=-\lfloor n\varepsilon \rfloor}^{\lfloor n\varepsilon \rfloor + v_m^z} |h_{k,z}(u+1-\{nT_m + n\theta_z\})|^p \to 2\,|q_{k,\alpha}|^p \qquad \text{as } n\to\infty. \tag{3.14}$$

The same statement holds for $z=0$, but the limit becomes $|q_{k,\alpha}|^p$, since in this setting the sum is one-sided. We set $\|x\|_p^p = \sum_{i=1}^m |x_i|^p$ for any $x\in\mathbb{R}^m$ and $p>0$, and recall that $\|x\|_p$ is a norm for $p\ge 1$. It holds that

$$\begin{aligned} \left|\|x\|_p^p - \|y\|_p^p\right| &\le \|x-y\|_p^p \qquad \text{when } p\in(0,1], \\ \left|\|x\|_p - \|y\|_p\right| &\le \|x-y\|_p \qquad \text{when } p>1. \end{aligned} \tag{3.15}$$

By (3.13), (3.14) and (3.15), and taking into account the definition of $V_{n,\varepsilon,z}$ at (3.8), we readily deduce the convergence

$$\frac{V_{n,\varepsilon,z}}{\log(n)} \xrightarrow{\mathbb{P}} |q_{k,\alpha} c_z|^p (1 + \mathbb{1}_{\{z\ge 1\}}) \sum_{m: T_m \in [-\theta_z, 1-\theta_z]} |\Delta L_{T_m}|^p$$

as $n\to\infty$, and hence

$$n^{\alpha p} \sum_{i=k}^n \left|\sum_{z=0}^l M_{i,n,\varepsilon,z}\right|^p \xrightarrow{\mathbb{P}} |q_{k,\alpha}|^p \sum_{z=0}^l |c_z|^p (1 + \mathbb{1}_{\{z\ge 1\}}) \sum_{m: T_m \in [-\theta_z, 1-\theta_z]} |\Delta L_{T_m}|^p$$

as $n\to\infty$, on Ω_ε. □

3.1.3 Negligible Terms

Due to the inequalities at (3.15), it suffices to show that on Ω_ε

$$a_n \sum_{i=k}^n |R_{i,n,\varepsilon}|^p \xrightarrow{\mathbb{P}} 0 \qquad \text{and} \qquad a_n \sum_{i=k}^n |M_{i,n,\varepsilon,z}|^p \xrightarrow{\mathbb{P}} 0 \qquad \text{for } z\in\mathcal{A}^c, \tag{3.16}$$

as $n\to\infty$, where $a_n = n^{\alpha p}$ in Theorem 2.4(i) and $a_n = n^{\alpha p}/\log(n)$ in Theorem 2.4(ii), to prove that these terms do not affect the limits in Theorem 2.4. At this stage we notice that outside the singularity points the kernel function g satisfies the same properties under Assumption (B) (resp. Assumption (B-log)) as under Assumption (A) (resp. Assumption (A-log)). Consequently, we can apply the estimates for the term $R_{i,n,\varepsilon}$ derived in [5, Eqs. (4.8), (4.12)] and [6, Section 4] under assumptions (A) and (A-log)

$$\sup_{n\in\mathbb{N}, i\in\{k,\dots,n\}} n^k |R_{i,n,\varepsilon}| < \infty \qquad \text{almost surely if } w\in(0,1],$$

$$\sup_{n\in\mathbb{N}, i\in\{k,\dots,n\}} \frac{n^k |R_{i,n,\varepsilon}|}{(\log(n))^q} < \infty \qquad \text{almost surely if } w\in(1,2],$$

where q is determined via $1/q + 1/w = 1$, since $R_{i,n,\varepsilon}$ is only affected by the function g outside the singularity points θ_z. We readily conclude the first convergence at (3.16) in the setting of Theorem 2.4(i), because $\alpha < k - 1/p$. It also holds in the setting of Theorem 2.4(ii), where for $w \in (1, 2]$ we use the assumption that $1/p + 1/w > 1$.

Now, we show the second statement of (3.16), which is only relevant in the setting of Theorem 2.4(i). Since $\alpha_z < k - 1/p$ for all z, we can apply to $\sum_{i=k}^{n} |M_{i,n,\varepsilon,z}|^p$, $z \in \mathcal{A}^c$, the same techniques as for $\sum_{i=k}^{n} |M_{i,n,\varepsilon,z}|^p$, $z \in \mathcal{A}$. Hence, using the same methods as in Section 3.1.1, we conclude that on Ω_ε

$$n^{\alpha p} \sum_{i=k}^{n} |M_{i,n,\varepsilon,z}|^p = O_{\mathbb{P}}\Big(n^{p(\alpha - \alpha_z)}\Big) \qquad \text{for all } z \in \mathcal{A}^c,$$

where the notation $Y_n = O_{\mathbb{P}}(a_n)$ means that the sequence $a_n^{-1} Y_n$ is tight. Since $\alpha_z > \alpha$ for all $z \in \mathcal{A}^c$, we obtain the second statement of (3.16). The results of Sections 3.1.1–3.1.3 and the fact that $\mathbb{P}(\Omega_\varepsilon) \uparrow 1$ as $\varepsilon \to 0$ imply the assertion of Theorem 2.4 in the compound Poisson case. □

3.2 Proof of Theorem 2.4 in the General Case

Let now $(L_t)_{t \in \mathbb{R}}$ be a general symmetric pure jump Lévy process with Blumenthal–Getoor index β. We denote by N the corresponding Poisson random measure defined by $N(A) := \#\{t \in \mathbb{R} : (t, \Delta L_t) \in A\}$ for all measurable $A \subseteq \mathbb{R} \times (\mathbb{R} \setminus \{0\})$. Next, we introduce the process

$$X_t(m) = \int_{(-\infty, t] \times [-\frac{1}{m}, \frac{1}{m}]} x\big(g(t-s) - g_0(-s)\big) N(\mathrm{d}s, \mathrm{d}x),$$

which only involves small jumps of L. We will prove that

$$\lim_{m \to \infty} \limsup_{n \to \infty} \mathbb{P}(a_n V(X(m), p; k)_n > \epsilon) = 0 \qquad \text{for any } \epsilon > 0, \tag{3.17}$$

where $a_n = n^{\alpha p}$ in Theorem 2.4(i) and $a_n = n^{\alpha p}/\log(n)$ in Theorem 2.4(ii). First, due to Markov's inequality and the stationary increments of $X_t(m)$, it follows that

$$\mathbb{P}(a_n V(X(m), p; k)_n > \epsilon) \le \epsilon^{-1} a_n \sum_{i=k}^{n} \mathbb{E}[|\Delta_{i,k}^n X(m)|^p] \le \epsilon^{-1} b_n \mathbb{E}[|\Delta_{k,k}^n X(m)|^p],$$

where $b_n = n a_n$. Hence, it is enough to prove that

$$\lim_{m \to \infty} \limsup_{n \to \infty} \mathbb{E}[|Y_{n,m}|^p] = 0 \qquad \text{where } Y_{n,m} = b_n^{1/p} \Delta_{k,k}^n X(m). \tag{3.18}$$

Notice the representation

$$Y_{n,m} = \int_{(-\infty, \frac{k}{n}] \times [-\frac{1}{m}, \frac{1}{m}]} \big(b_n^{1/p} g_{k,n}(s)\big) x \, N(\mathrm{d}s, \mathrm{d}x).$$

Using this together with [10, Theorem 3.3], (3.18) will follow if

$$\lim_{m \to \infty} \limsup_{n \to \infty} \xi_{n,m} = 0 \qquad \text{where} \qquad \xi_{n,m} = \int_{-\frac{1}{m}}^{\frac{1}{m}} \chi_n(x) \, \nu(\mathrm{d}x) \qquad \text{and}$$

$$\chi_n(x) = \int_{-\infty}^{\frac{k}{n}} \Big(|b_n^{1/p} g_{k,n}(s) x|^p \mathbb{1}_{\{|b_n^{1/p} g_{k,n}(s) x| \ge 1\}} + |b_n^{1/p} g_{k,n}(s) x|^2 \mathbb{1}_{\{|b_n^{1/p} g_{k,n}(s) x| < 1\}} \Big) \mathrm{d}s.$$

Suppose there exists a constant $K \geq 0$ such that for all large $n \in \mathbb{N}$

$$\chi_n(x) \leq K(|x|^p + |x|^2) \qquad \text{for all } x \in [-1,1], \tag{3.19}$$

then the dominated convergence theorem implies that

$$\limsup_{m\to\infty}\Big[\limsup_{n\to\infty} \xi_{n,m}\Big] \leq K \limsup_{m\to\infty} \int_{-\frac{1}{m}}^{\frac{1}{m}} (|x|^p + |x|^2)\, \nu(\mathrm{d}x) = 0,$$

using the assumption that $p > \beta$. We consider only (3.19) in the case of Theorem 2.4(i) as (ii) is very similar, see [6]. In the case of (i) then $b_n^{1/p} = n^{\alpha+1/p}$. For short notation define $\Phi_p : \mathbb{R} \to \mathbb{R}_+$ as the function

$$\Phi_p(y) = |y|^2 \mathbb{1}_{\{|y|\leq 1\}} + |y|^p \mathbb{1}_{\{|y|>1\}}, \qquad (y \in \mathbb{R}).$$

Note that Φ_p is of modular growth, i.e. there exists a constant $K_p > 0$ depending only on p such that for any $x, y \in \mathbb{R}$ then $\Phi_p(x+y) \leq K_p(\Phi_p(x) + \Phi_p(y))$. We consider the following decomposition

$$\begin{aligned}
\chi_n(x) = {} & \int_{\frac{k}{n}-\frac{1}{n}}^{\frac{k}{n}} \Phi_p(n^{\alpha+1/p} g_{k,n}(s)x)\,\mathrm{d}s + \sum_{z=1}^{l} \int_{\frac{k}{n}-\theta_z-\frac{1}{n}}^{\frac{k}{n}-\theta_z+\frac{1}{n}} \Phi_p(n^{\alpha+1/p} g_{k,n}(s)x)\,\mathrm{d}s \\
& + \sum_{z=1}^{l} \int_{\frac{k}{n}-\theta_z+\frac{1}{n}}^{\frac{k}{n}-\theta_{z-1}-\frac{1}{n}} \Phi_p(n^{\alpha+1/p} g_{k,n}(s)x)\,\mathrm{d}s + \int_{\frac{k}{n}-\theta_l-\delta}^{\frac{k}{n}-\theta_l-\frac{1}{n}} \Phi_p(n^{\alpha+1/p} g_{k,n}(s)x)\,\mathrm{d}s \\
& + \int_{-\infty}^{\frac{k}{n}-\theta_l-\delta} \Phi_p(n^{\alpha+1/p} g_{k,n}(s)x)\,\mathrm{d}s \\
=: {} & I_0(x) + \sum_{z=1}^{l} I_{1,z}(x) + \sum_{z=1}^{l} I_{2,z}(x) + I_3(x) + I_4(x).
\end{aligned}$$

We treat the five types of terms separately.

Estimation of I_0: By Lemma 3.2

$$|g_{k,n}(x)| \leq K(|\tfrac{k}{n} - s|^{\alpha_0}) \qquad \text{for all } s \in [\tfrac{k}{n} - \tfrac{1}{n}, \tfrac{k}{n}].$$

Since Φ_p is increasing on $\mathbb{R}_+$ and $\alpha \leq \alpha_0$ it follows that

$$I_0(x) \leq K \int_0^{\frac{1}{n}} \Phi_p(x n^{\alpha+1/p} s^{\alpha_0})\,\mathrm{d}s \leq K \int_0^{\frac{1}{n}} \Phi_p(x n^{\alpha+1/p} s^{\alpha})\,\mathrm{d}s.$$

By elementary integration it follows that

$$\begin{aligned}
&\int_0^{\frac{1}{n}} |x n^{\alpha+1/p} s^{\alpha}|^2 \mathbb{1}_{\{|x n^{\alpha+1/p} s^{\alpha}|\leq 1\}}\,\mathrm{d}s \\
&\qquad \leq K(x^2 \mathbb{1}_{\{|x|\leq n^{-1/p}\}} n^{2/p-1} + \mathbb{1}_{\{|x|>n^{-1/p}\}} |x|^{-1/\alpha} n^{-1-1/(\alpha p)}) \\
&\qquad \leq K(x^2 + |x|^p).
\end{aligned}$$

The second term in Φ_p is dealt with as follows

$$\int_0^{\frac{1}{n}} |x n^{\alpha+1/p} s^{\alpha}|^p \mathbb{1}_{\{|x n^{\alpha+1/p} s^{\alpha}|>1\}}\,\mathrm{d}s \leq |x|^p n^{\alpha p+1} \int_0^{\frac{1}{n}} s^{\alpha p}\,\mathrm{d}s = \frac{|x|^p}{\alpha p + 1}.$$

Combining the two estimates above it follows that $I_0(x) \leq K(|x|^2 + |x|^p)$.

Estimation of $I_{1,z}$: Similarly as for I_0 we have, using arguments as in Lemma 3.2(i), that

$$|g_{k,n}(s)| \leq K \sum_{j=0}^{k} |\tfrac{k-j}{n} - s - \theta_z|^{\alpha_z} \qquad \text{for all } s \in [\tfrac{k}{n} - \theta_z - \tfrac{1}{n}, \tfrac{k}{n} - \theta_z + \tfrac{1}{n}].$$

Using the modular growth of Φ_p it follows that

$$\begin{aligned}
\int_{\frac{k}{n}-\theta_z-\frac{1}{n}}^{\frac{k}{n}-\theta_z+\frac{1}{n}} \Phi_p(n^{\alpha+1/p} g_{k,n}(s)x)\,\mathrm{d}s &\leq K_p \sum_{j=0}^{k} \int_{\frac{k}{n}-\theta_z-\frac{1}{n}}^{\frac{k}{n}-\theta_z+\frac{1}{n}} \Phi_p(n^{\alpha+1/p} |\tfrac{k-j}{n} - s - \theta_z|^{\alpha_z} x)\,\mathrm{d}s \\
&= K_p \sum_{j=0}^{k} \int_{-\frac{j}{n}-\frac{1}{n}}^{-\frac{j}{n}+\frac{1}{n}} \Phi_p(n^{\alpha+1/p} |s|^{\alpha_z} x)\,\mathrm{d}s \\
&\leq K_p \int_{-\frac{k+1}{n}}^{\frac{k+1}{n}} \Phi_p(n^{\alpha+1/p} |s|^{\alpha} x)\,\mathrm{d}s \\
&= K_p \int_{0}^{\frac{k+1}{n}} \Phi_p(n^{\alpha+1/p} |s|^{\alpha} x)\,\mathrm{d}s.
\end{aligned}$$

As for I_0 we get $I_{1,z}(x) \leq K(|x|^2 + |x|^p)$.

Estimation of $I_{2,z}$: We decompose $I_{2,z}$ into three terms corresponding to whether we are close to the singularity θ_z from the right or close to the singularity θ_{z-1} from the left or in between them, but bounded away from both. More specifically, we decompose as

$$\begin{aligned}
I_{2,z}(x) = &\int_{\frac{k}{n}-\theta_z+\frac{1}{n}}^{\frac{k}{n}-\theta_z+\delta} \Phi_p(n^{\alpha+1/p} g_{k,n}(s)x)\,\mathrm{d}s + \int_{\frac{k}{n}-\theta_z+\delta}^{\frac{k}{n}-\theta_{z-1}-\delta} \Phi_p(n^{\alpha+1/p} g_{k,n}(s)x)\,\mathrm{d}s \\
&+ \int_{\frac{k}{n}-\theta_{z-1}-\delta}^{\frac{k}{n}-\theta_{z-1}-\frac{1}{n}} \Phi_p(n^{\alpha+1/p} g_{k,n}(s)x)\,\mathrm{d}s =: I_{2,z}^{l}(x) + I_{2,z}^{b}(x) + I_{2,z}^{r}(x).
\end{aligned}$$

First we note that arguments similar to Lemma 3.2(iii) implies that

$$|g_{k,n}(s)| \leq K n^{-k} |\tfrac{k}{n} - s - \theta_z|^{\alpha_z - k} \qquad \text{for all } s \in [\tfrac{k}{n} - \theta_z + \tfrac{1}{n}, \tfrac{k}{n} - \theta_z + \delta].$$

Using again that Φ_p is decreasing on $\mathbb{R}_+$ it follows that

$$\begin{aligned}
I_{2,z}^{l}(x) &\leq K \int_{\frac{k}{n}-\theta_z+\frac{1}{n}}^{\frac{k}{n}-\theta_z+\delta} \Phi_p(n^{\alpha+1/p-k} |\tfrac{k}{n} - s - \theta_z|^{\alpha_z-k} x)\,\mathrm{d}s \\
&\leq K \int_{\frac{1}{n}}^{\delta} \Phi_p(n^{\alpha+1/p-k} |s|^{\alpha_z-k} x)\,\mathrm{d}s.
\end{aligned}$$

If $\alpha_z = k - 1/2$ then

$$\int_{\frac{1}{n}}^{\delta} |x n^{\alpha+1/p-k} s^{\alpha_z-k}|^2 \mathbb{1}_{\{|x^2 n^{\alpha+1/p-k} s^{\alpha_z-k}| \leq 1\}}\,\mathrm{d}s \leq x^2 n^{2(\alpha+1/p-k)} \int_{\frac{1}{n}}^{\delta} s^{-1}\,\mathrm{d}s \leq K x^2,$$

where we used that $\alpha < k - 1/p$. For $\alpha_z \neq k - 1/2$ we have that

$$\begin{aligned}
&\int_{\frac{1}{n}}^{\delta} |x n^{\alpha+1/p-k} s^{\alpha_z-k}|^2 \mathbb{1}_{\{|x n^{\alpha+1/p-k} s^{\alpha_z-k}| \leq 1\}}\,\mathrm{d}s \\
&\quad\leq K\Big(|x|^2 n^{2(\alpha+1/p-k)} + |x|^2 n^{2(\alpha-\alpha_z)+2/p-1} \mathbb{1}_{\{|x| \leq n^{-1/p}\}} + |x|^{\frac{1}{k-\alpha_z}} n^{\frac{\alpha+1/p-k}{k-\alpha_z}} \mathbb{1}_{\{|x| > n^{-1/p}\}}\Big) \\
&\quad\leq K(x^2 + |x|^p),
\end{aligned}$$

where we used that $\alpha \le \alpha_z < k - 1/p$. Moreover,

$$\int_{\frac{1}{n}}^{\delta} |xn^{\alpha+1/p-k} s^{\alpha_z-k}|^p \mathbb{1}_{\{|xn^{\alpha+1/p-k}s^{\alpha_z-k}|>1\}} \, \mathrm{d}s \le K|x|^p.$$

The term $I_{2,z}^r$ is handled similarly. For the last term $I_{2,z}^b$ we note that, since we are bounded away from both θ_{z-1} and θ_z, there exists a constant $K > 0$ such that

$$|g_{k,n}(s)| \le Kn^{-k} \qquad \text{for all } s \in [\tfrac{k}{n} - \theta_z + \delta, \tfrac{k}{n} - \theta_{z-1} - \delta].$$

This readily implies the bound $I_{2,z}^b(x) \le K(x^2 + |x|^p)$.

Estimation of I_3: Arguments as in Lemma 3.2 imply that

$$|g_{k,n}(s)| \le Kn^{-k} |\tfrac{k}{n} - s - \theta_z|^{\alpha_l - k} \qquad \text{for all } s \in [\tfrac{k}{n} - \theta_l - \delta, \tfrac{k}{n} - \theta_l - \tfrac{1}{n}].$$

One may then proceed as for the term $I_{2,z}^l$ above to conclude that $I_3(x) \le K(x^2 + |x|^p)$.

Estimation of I_4: First we decompose into two regions:

$$\begin{aligned}\int_{-\infty}^{\frac{k}{n}-\theta_l-\delta} \Phi_p(n^{\alpha+1/p} g_{k,n}(s)x) \, \mathrm{d}s = &\int_{-\delta'-\theta_l}^{\frac{k}{n}-\delta-\theta_l} \Phi_p(n^{\alpha+1/p} g_{k,n}(s)x) \, \mathrm{d}s \\ &+ \int_{-\infty}^{-\delta'-\theta_l} \Phi_p(n^{\alpha+1/p} g_{k,n}(s)x) \, \mathrm{d}s.\end{aligned}$$

In the first integral we are bounded away from θ_l, hence $|g_{k,n}(s)| \le Kn^{-k}$ for all s in the interval $[-\delta' - \theta_l, \frac{k}{n} - \delta - \theta_l]$. For the latter integral note first that by Lemma 3.2(v)

$$\int_{-\infty}^{-\delta'-\theta_l} \Phi_p(n^{\alpha+1/p} g_{k,n}(s)x) \, \mathrm{d}s \le \int_{-\infty}^{-\delta'-\theta_l} \Phi_p(n^{\alpha+1/p-k} |g^{(k)}(-s)|x) \, \mathrm{d}s.$$

Now

$$\int_{\delta'+\theta_l}^{\infty} |xn^{\alpha+1/p-k} g^{(k)}(s)|^2 \mathbb{1}_{\{|xn^{\alpha+1/p-k}g^{(k)}(s)|\le 1\}} \, \mathrm{d}s \le |xn^{\alpha+1/p-k}|^2 \int_{\delta'+\theta_l}^{\infty} |g^{(k)}(s)|^2 \, \mathrm{d}s.$$

Since $|g^{(k)}|$ is decreasing on $(\theta_l + \delta', \infty)$ and $g^{(k)} \in \mathcal{L}^w((\theta_l + \delta', \infty))$ for some $w \le 2$ it follows that the last integral is finite. Lastly, we find for $x \in [-1, 1]$ that

$$\begin{aligned}&\int_{\theta_l+\delta'}^{\infty} |xn^{\alpha+1/p-k} g^{(k)}(s)|^p \mathbb{1}_{\{|xn^{\alpha+1/p-k}g^{(k)}(s)|>1\}} \, \mathrm{d}s \\ &\qquad \le |x|^p n^{p(\alpha+1/p-k)} \int_{\theta_l+\delta'}^{\infty} |g^{(k)}(s)|^p \mathbb{1}_{\{|g^{(k)}(s)|>1\}} \, \mathrm{d}s.\end{aligned}$$

By our assumptions the last integral is finite, indeed

$$\int_{\delta'+\theta_l}^{\infty} |g^{(k)}(s)|^p \mathbb{1}_{\{|g^{(k)}(s)|>1\}} \, \mathrm{d}s \le K_p \, \|g^{(k)}\|_{\mathcal{L}^w((\delta'+\theta,\infty))}^w < \infty.$$

3.2.1 Negligibility of Small Jumps

Now, we note that $X_t - X_t(m)$ is the integral (1.1), where the integrator is a compound Poisson process that corresponds to big jumps of L. Hence, we obtain the results of Theorem 2.4 for the process $X - X(m)$ as in Section 3.1. More specifically, under assumptions of Theorem 2.4(i) it holds that

$$n_j^{\alpha p} V(X - X(m), p; k)_{n_j} \xrightarrow{\mathcal{L}\text{-s}} \sum_{z \in \mathcal{A}} |c_z|^p \sum_{r: T_r \in [-\theta_z, 1-\theta_z]} |\Delta L_{T_r}|^p \mathbb{1}_{\{|\Delta L_{T_r} > 1/m|\}} V_r^z$$

where V_r^z has been defined at (2.5). The term on the right hand side convergences to the limit of Theorem 2.4(i) as $m \to \infty$, since $\sum_{r: T_r \in [-\theta_z, 1-\theta_z]} |\Delta L_{T_r}|^p < \infty$ for any $p > \beta$. Finally, using the decomposition $X = (X - X(m)) + X(m)$ and letting first $n_j \to \infty$ and then $m \to \infty$, we deduce the statement of Theorem 2.4 by (3.17) and the inequalities (3.15). This completes the proof. □

Acknowledgment

The authors acknowledge financial support from the project 'Ambit fields: probabilistic properties and statistical inference' funded by Villum Fonden.

References

[1] David J. Aldous and G. K. Eagleson (1978). »*On mixing and stability of limit theorems*«. The Annals of Probability 6(2), 325–331. doi: 10.1214/aop/1176995577.

[2] Ole E. Barndorff-Nielsen, José Manuel Corcuera and Mark Podolskij (2011). »*Multipower variation for Brownian semistationary processes*«. Bernoulli 17(4), 1159–1194. doi: 10.3150/10-BEJ316.

[3] Ole E. Barndorff-Nielsen, José Manuel Corcuera and Mark Podolskij (2013). »*Limit theorems for functionals of higher order differences of Brownian semi-stationary processes*«. In: »*Prokhorov and contemporary probability theory*«. Edited by Albert N. Shiryaev, S. R. S. Varadhan and Ernst L. Presman. Springer Berlin Heidelberg, 69–96. isbn: 978-3-642-33549-5.

[4] Ole E. Barndorff-Nielsen, Svend Erik Graversen, Jean Jacod, Mark Podolskij and Neil Shephard (2006). »*From Stochastic Calculus to Mathematical Finance. Festschrift in Honour of A. N. Shiryaev*«. Edited by Yuri Kabanov, Robert Liptser and Jordan Stoyanov. Springer, Heidelberg. Chapter: »*A central limit theorem for realised power and bipower variations of continuous semimartingales*«, 33–68. isbn: 978-3-540-30788-4. doi: 10.1007/978-3-540-30788-4.

[5] Andreas Basse-O'Connor, Raphaël Lachièze-Rey and Mark Podolskij (2017). »*Power variation for a class of stationary increments Lévy driven moving averages*«. The Annals of Probability 45(6B), 4477–4528. doi: 10.1214/16-AOP1170.

[6] Andreas Basse-O'Connor and Mark Podolskij (2017). »*On critical cases in limit theory for stationary increments Lévy driven moving averages*«. Stochastics 89(1), 360–383. doi: 10.1080/17442508.2016.1191493.

[7] Jean-François Coeurjolly (2001). »*Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths*«. Statistical Inference for Stochastic Processes 4(2), 199–227. doi: 10.1023/A:1017507306245.

[8] Kerstin Gärtner and Mark Podolskij (2015). »*On non-standard limits of Brownian semi-stationary processes*«. Stochastic Processes and their Applications 125(2), 653–677. doi: 10.1016/j.spa.2014.09.019.

[9] Jean Jacod and Philip E. Protter (2012). »*Discretization of processes*«. Stochastic Modelling and Applied Probability 67. Springer-Verlag Berlin Heidelberg. isbn: 978-3-642-24126-0. doi: 10.1007/978-3-642-24127-7.

[10] Balram S. Rajput and Jan Rosiński (1989). »*Spectral representations of infinitely divisible processes*«. Probability Theory and Related Fields 82(3), 451–487. doi: 10.1007/BF00339998.

[11] Alfréd Rényi (1963). »*On stable sequences of events*«. Sankhyā: The Indian Journal of Statistics. Series A 25(3), 293–302.

[12] Gennady Samorodnitsky and Murad Taqqu (1994). »*Stable non-Gaussian random processes: Stochastic models with infinite variance*«. Stochastic modeling. Chapmann and Hall, New York. isbn: 978-0-412-05171-5.

B PAPER

A Minimal Contrast Estimator for the Linear Fractional Stable Motion

Mathias Mørck Ljungdahl

Department of Mathematics, Aarhus University

Mark Podolskij

Department of Mathematics, Aarhus University

Abstract. In this paper we present an estimator for the three-dimensional parameter (σ, α, H) of the linear fractional stable motion, where H represents the self-similarity parameter, and (σ, α) are the scaling and stability parameters of the driving symmetric Lévy process L. Our approach is based upon a minimal contrast method associated with the empirical characteristic function combined with a ratio type estimator for the self-similarity parameter H. The main result investigates the strong consistency and weak limit theorems for the resulting estimator. Furthermore, we propose several ideas to obtain feasible confidence regions in various parameter settings. Our work is mainly related to [17, 19], in which parameter estimation for the linear fractional stable motion and related Lévy moving average processes has been studied.

Key words: linear fractional processes, Lévy processes, limit theorems, parametric estimation, bootstrap, subsampling, self-similarity, low frequency

AMS 2010 subject classifications: Primary 60G22, 62F12, 62E20; secondary 60E07, 60F05, 60G10

1 Introduction

During the last sixty years fractional stochastic processes have received a great deal of attention in probability, statistics and integration theory. One of the most prominent examples of a fractional model is the (scaled) fractional Brownian motion (fBm), which gained a lot of popularity in science since the pioneering work of Mandelbrot and van Ness [18]. The scaled fBm is the unique zero mean Gaussian process with stationary increments and self-similarity property. As a building block in stochastic models it found numerous applications in natural and social sciences such as physics, biology and economics. From the statistical perspective the scaled fBm is fully determined by its scale parameter $\sigma > 0$ and the self-similarity parameter (or Hurst index) $H \in (0,1)$. Nowadays, the estimation of (σ, H) is a well understood problem. We refer to [11] for efficient estimation of the Hurst parameter H in the low frequency setting, and to [6, 9, 15] for the estimation of (σ, H) in the high frequency setting, among many others. In more recent papers [2, 16] statistical inference for the multifractional Brownian motion has been investigated, which accounts for the time varying nature of the Hurst parameter.

This paper focuses on another extension of the fBm, the *linear fractional stable motion* (lfsm). The lfsm $(X_t)_{t\geq 0}$ is a three-parameter statistical model defined by

$$X_t = \int_{\mathbb{R}} \{(t-s)_+^{H-1/\alpha} - (-s)_+^{H-1/\alpha}\} \, \mathrm{d}L_s, \tag{1.1}$$

where $x_+ = x \vee 0$ denotes the positive part and we set $x_+^a = 0$ for all $a \in \mathbb{R}$, $x \leq 0$. Here $(L_t)_{t\in\mathbb{R}}$ is a symmetric α-stable Lévy process with $\alpha \in (0,2)$ and scale parameter $\sigma > 0$, and $H \in (0,1)$ represents the Hurst parameter. In some sense the lfsm is a non-Gaussian analogue of fBm. The process $(X_t)_{t\geq 0}$ has symmetric α-stable marginals, stationary increments and it is self-similar with parameter H. It is well known that the process X has continuous paths when $H - 1/\alpha > 0$, see, e.g. [5]. We remark that the class of stationary increments self-similar processes becomes much larger if we drop the Gaussianity assumption, cf. [21, 24], but the lfsm is one of its most famous representatives due to the ergodicity property. Linear 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. [13] for the context of Ethernet and solar flare modelling.

The limit theory for statistics of lfsm, which is indispensable for the estimation of the parameter $\xi = (\sigma, \alpha, H)$, turns out to be of a quite complex nature. First central limit theorems for partial sums of bounded functions of Lévy moving average processes, which in particular include the lfsm, have been discussed in [22] and later extended in [23] to certain unbounded functions. In a more recent work [4] the authors presented a rather complete asymptotic theory for power variations of stationary increments Lévy moving average processes. Finally, the results of [4] have been extended to general functions in [3], who demonstrated that the weak limit theory crucially depends on the *Appell rank* of the given function and the parameters of the model (all functions considered in this paper have Appell rank 2). More specifically, they obtained three different asymptotic regimes, a normal and two stable ones, depending on the particular setting. It is this phase transition that depends on the parameter (α, H) which makes the statistical inference for lfsm a rather complicated matter.

Since the probabilistic theory for functionals of lfsm was not well understood until the recent work [3, 4], the statistical literature on estimation of lfsm is rather scarce. The articles [1, 23] investigate the asymptotic theory for a wavelet-based estimator of H when $\alpha \in (1,2)$. In [4, 27] the authors 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. The work [12] 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. The paper [19] was the first instance, where estimation of the full parameter $\xi = (\sigma, \alpha, H)$

has been studied in low and high frequency settings. Their idea is based upon the identification of ξ through power variation statistics and the empirical characteristic function evaluated at two different values.

In this paper we aim at extending the approach of [19] by determining the asymptotic theory for the minimal contrast estimator of the parameter $\xi = (\sigma, \alpha, H)$, which is based upon the comparison of the empirical characteristic function with its theoretical counterpart under low frequency sampling. Indeed, the choice of the two evaluation points for the empirical characteristic function in [19] is rather ad hoc and we will show in the empirical study that the minimal contrast estimator exhibits better finite sample properties and robustness in various settings. Similarly to [19], we will show that the weak limit theory for our estimator has a normal and a stable regime, and the asymptotic distribution depends on the interplay between the parameters α and H. At this stage we remark that the minimal contrast approach has been investigated in [17] in the context of certain Lévy moving average models, which do not include the lfsm or its associated noise process, but only in the asymptotically normal regime. Another important contribution of our paper is the subsampling procedure, which provides confidence regions for the parameters of the model irrespectively of the unknown asymptotic regime.

The article is organized as follows. In Section 2 we introduce the necessary notation and formulate a new weak limit result related to [19], which is central to their parameter estimation for the linear fractional stable motion. The aforementioned theorem will be our starting point, where the aim is to extend the convergence of the finite dimensional distributions to convergence of integral functionals appearing in the minimal contrast method. Section 3 introduces the estimator and presents the main results of strong consistency and asymptotic distribution. Section 4 is devoted to a simulation study, which tests the finite sample performance of the minimal contrast estimator. We also discuss the parametric bootstrap and the subsampling method that are used to construct feasible confidence regions for the true parameters of the model. All proofs are collected in Section 5 and all larger tables are in Section 6.

2 Notation and Recent Results

We start out with introducing the main notation and statistics of interest. We consider low frequency observations $X_1, X_2, \ldots, X_n$ from the lfsm $(X_t)_{t\geq 0}$ introduced in (1.1). We denote by $\Delta_{i,k}^r X$ $(i,k,r \in \mathbb{N})$ the kth order increment of X at stage i and rate r, i.e.

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

The order k plays a crucial role in determining the asymptotic regime for statistics that we introduce below. We let the function $h_{k,r} : \mathbb{R} \to \mathbb{R}$ denote the kth order increment at rate r of the kernel in (1.1), specifically

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

We note that $\Delta_{i,k}^r X = \int_{\mathbb{R}} h_{k,r}(i-s)\,\mathrm{d}L_s$. For less cumbersome notation we drop the index r if $r = 1$, so $\Delta_{i,k} X := \Delta_{i,k}^1 X$ and $h_k = h_{k,1}$. Throughout this paper we write $\theta = (\sigma, \alpha)$

and $\xi = (\sigma, \alpha, H)$. The main probabilistic tools are statistics of the type

$$V_n(f,k,r) = \frac{1}{n}\sum_{i=rk}^{n} f(\Delta_{i,k}^r X),$$

where $f : \mathbb{R} \to \mathbb{R}$ is a Borel function satisfying $\mathbb{E}[|f(\Delta_{rk,k}^r X)|] < \infty$. We will specifically focus on two classes of functions, namely $f_p(x) = |x|^p$ for $p \in (-1,1)$ and $\delta_t(x) = \cos(tx)$ for $t \geq 0$. They correspond to power variation statistics and the real part of the empirical characteristic function respectively, and we use the notation

$$\varphi_n(t) = V_n(\delta_t, k, 1) \qquad \text{and} \qquad \psi_n(r) = V_n(f_p, k, r). \tag{2.1}$$

We note that Birkhoff's ergodic theorem implies the almost sure convergence

$$\varphi_n(t) \xrightarrow{\text{a.s.}} \varphi_\xi(t) := \exp(-|\sigma\,\|h_k\|_\alpha t|^\alpha) \qquad \text{where} \qquad \|h_k\|_\alpha^\alpha := \int_{\mathbb{R}} |h_k(x)|^\alpha \,\mathrm{d}x. \tag{2.2}$$

An important coefficient in our context is

$$\beta = 1 + \alpha(k - H). \tag{2.3}$$

The rate of convergence and the asymptotic distribution of statistics defined at (2.1) crucially depend on whether the condition $k > H + 1/\alpha$ is satisfied or not. Hence, we define the normalized versions of our statistics as

$$\left.\begin{aligned} W_n^1(r) &= \sqrt{n}(\psi_n(r) - r^H m_{p,k}), \\ W_n^2(t) &= \sqrt{n}(\varphi_n(t) - \varphi_\xi(t)) \end{aligned}\right. \qquad \text{when } k > H + 1/\alpha$$

and

$$\left.\begin{aligned} S_n^1(r) &= n^{1-1/\beta}(\psi_n(r) - r^H m_{p,k}), \\ S_n^2(t) &= n^{1-1/\beta}(\varphi_n(t) - \varphi_\xi(t)) \end{aligned}\right. \qquad \text{when } k < H + 1/\alpha.$$

Here $m_{p,k} = \mathbb{E}[|\Delta_{k,k} X|^p]$, which is finite for any $p \in (-1,\alpha)$, and φ_ξ is given at (2.2). Note that $\mathbb{E}[|\Delta_{rk,k}^r X|^p] = r^H m_{p,k}$, which explains the centring of W^1 and S^1.

It turns out that the finite dimensional limit of the statistics (W_n^1, W_n^2) is Gaussian while the corresponding limit of (S_n^1, S_n^2) is β-stable (see Theorem 2.1 below). We now introduce several notations to describe the limiting distribution. We start with the Gaussian case. For random variables $X = \int_{\mathbb{R}} g(s)\,\mathrm{d}L_s$ and $Y = \int_{\mathbb{R}} h(s)\,\mathrm{d}L_s$ with $\|g\|_\alpha, \|h\|_\alpha < \infty$ we define a dependence measure $U_{g,h} : \mathbb{R}^2 \to \mathbb{R}$ as

$$\begin{aligned} U_{g,h}(u,v) &= \mathbb{E}[\exp(\mathrm{i}(uX + vY))] - \mathbb{E}[\exp(\mathrm{i}uX)]\mathbb{E}[\exp(\mathrm{i}vY)] \\ &= \exp(-\sigma^\alpha \|ug + vh\|_\alpha^\alpha) - \exp\big(-\sigma^\alpha(\|ug\|_\alpha^\alpha + \|vh\|_\alpha^\alpha)\big). \end{aligned} \tag{2.4}$$

Next, for $p \in (-1,1) \setminus \{0\}$, we introduce the constant

$$a_p := \begin{cases} \displaystyle\int_{\mathbb{R}} (1 - \cos(y))|y|^{-1-p}\,\mathrm{d}y, & p \in (0,1), \\ \sqrt{2\pi}\,\Gamma(-p/2)/2^{p+1/2}\,\Gamma((p+1)/2), & p \in (-1,0), \end{cases} \tag{2.5}$$

where Γ denotes the Gamma-function. Now, for each $t \in \mathbb{R}_+$, we set

$$\begin{aligned} \Sigma_{11}(g,h) &= a_p^{-2} \int_{\mathbb{R}^2} |xy|^{-1-p} U_{g,h}(x,y)\,\mathrm{d}x\,\mathrm{d}y, \\ \Sigma_{12}(g,h;t) &= a_p^{-1} \int_{\mathbb{R}^2} |y|^{-1-p} U_{g,h}(x,t)\,\mathrm{d}x \end{aligned} \tag{2.6}$$

whenever the above integrals are finite. As it has been shown in [19], the following identities hold for any $p \in (-1/2, 1/2) \setminus \{0\}$ with $p < \alpha/2$ and $t \in \mathbb{R}_+$:

$$\mathrm{Cov}(|X|^p, |Y|^p) = \Sigma_{11}(g,h) \qquad \text{and} \qquad \mathrm{Cov}(|X|^p, \exp(\mathrm{i}tY)) = \Sigma_{12}(g,h;t).$$

Obviously, the quantities $\Sigma_{11}(g,h)$ and $\Sigma_{12}(g,h;t)$ will appear in the asymptotic covariance kernel of the vector $(\psi_n(r), \varphi_n(t))$ in the normal regime.

Now, we introduce the necessary notations for the stable case. First, we define the functions

$$\begin{aligned} \Phi_r^1(x) &:= a_p^{-1} \int_{\mathbb{R}} (1 - \cos(ux)) \exp(-|\sigma \, \|h_{k,r}\|_\alpha u|^\alpha) |u|^{-1-p} \, \mathrm{d}u, \qquad r \in \mathbb{N}, \\ \Phi_t^2(x) &:= (\cos(tx) - 1) \exp(-|\sigma \, \|h_k\|_\alpha t|^\alpha), \qquad t \geq 0, \end{aligned} \tag{2.7}$$

and set

$$q_{H,\alpha,k} := \prod_{i=0}^{k-1} (H - 1/\alpha - i).$$

Next, we introduce the functions $\kappa_1 : \mathbb{N} \to \mathbb{R}_+$ and $\kappa_2 : \mathbb{R} \to \mathbb{R}_-$ via

$$\begin{aligned} \kappa_1(r) &:= \tfrac{\alpha}{\beta} \int_0^\infty \Phi_r^1(q_{H,\alpha,k} z) z^{-1-\alpha/\beta} \, \mathrm{d}z, \\ \kappa_2(t) &:= \tfrac{\alpha}{\beta} \int_0^\infty \Phi_t^2(q_{H,\alpha,k} z) z^{-1-\alpha/\beta} \, \mathrm{d}z. \end{aligned} \tag{2.8}$$

In the final step we will need to define two Lévy measures ν_1 on $(\mathbb{R}_+)^2$ and ν_2 on $\mathbb{R}_+$ that are necessary to determine the asymptotic distribution of (S_n^1, S_n^2). Let us denote by ν the Lévy measure of the symmetric α-stable Lévy motion L (i.e. $\nu(\mathrm{d}x) = c(\sigma)|x|^{-1-\alpha} \, \mathrm{d}x$) and define the mappings $\tau_1 : \mathbb{R} \to (\mathbb{R}_+)^2$ and $\tau_2 : \mathbb{R} \to \mathbb{R}_+$ via

$$\tau_1(x) = |x|^{\alpha/\beta}(\kappa_1(1), \kappa_1(2)), \qquad \tau_2(x) = |x|^{\alpha/\beta}.$$

Then, for any Borel sets $A_1 \subseteq (\mathbb{R}_+)^2$ and $A_2 \subseteq \mathbb{R}_+$ bounded away from $(0,0)$ and 0, respectively, we introduce

$$\nu_l(A_l) := \nu(\tau_l^{-1}(A_l)), \qquad l = 1, 2. \tag{2.9}$$

In the weak limit theorem below we write $Z^n \xrightarrow{\mathcal{L}\text{-f}} Z$ to denote the convergence of finite dimensional distributions, i.e. the convergence in distribution

$$(Z_{t_1}^n, \ldots, Z_{t_d}^n) \xrightarrow{\mathcal{L}} (Z_{t_1}, \ldots, Z_{t_d})$$

for any $d \in \mathbb{N}$ and $t_i \in \mathbb{R}_+$. The following theorem is key for statistical applications.

Theorem 2.1. *Assume that either $p \in (-1/2, 0)$ or $p \in (0, 1/2)$ together with $p < \alpha/2$.*

(i) *If $k > H + 1/\alpha$ then as $n \to \infty$*

$$(W_n^1(1), W_n^1(2), W_n^2(t)) \xrightarrow{\mathcal{L}\text{-f}} (W_1^1, W_2^1, W_t),$$

where $W^1 = (W_1^1, W_2^1)$ *is a centred 2-dimensional normal distribution and* $(W_t)_{t\geq 0}$ *is a centred Gaussian process with*

$$\begin{aligned} \mathrm{Cov}(W_j^1, W_{j'}^1) &= \sum_{l\in\mathbb{Z}} \Sigma_{11}(h_{k,j}, h_{k,j'}(\,\cdot\,+l)) && j,j'=1,2,\\ \mathrm{Cov}(W_j^1, W_t) &= \sum_{l\in\mathbb{Z}} \Sigma_{12}(h_{k,j}, h_k(\,\cdot\,+l);t) && j=1,2,\ t\in\mathbb{R}_+,\\ \mathrm{Cov}(W_s, W_t) &= \frac{1}{2}\sum_{l\in\mathbb{Z}}\Big(U_{h_k,h_k(\cdot+l)}(s,t) + U_{h_k,-h_k(\cdot+l)}(s,t)\Big) && s,t\in\mathbb{R}, \end{aligned}$$

where the quantity Σ_{ij} *has been introduce at* (2.6)*. Moreover, the Gaussian process* W *exhibits a modification (denoted again by* W*), which is locally Hölder continuous of any order smaller than* $\alpha/4$.

(ii) *If* $k < H + 1/\alpha$ *then as* $n\to\infty$

$$(S_n^1(1), S_n^1(2), S_n^2(t)) \xrightarrow{\mathcal{L}\text{-f}} (S_1^1, S_2^1, \kappa_2(t)S).$$

where $S^1 = (S_1^1, S_2^1)$ *is a* β*-stable random vector with Lévy measure* ν_1 *independent of the totally right skewed* β*-stable random variable* S *with Lévy measure* ν_2*, and* ν_1, ν_2 *have been defined in* (2.9).

The finite dimensional asymptotic distribution demonstrated in Theorem 2.1 is a direct consequence of [19, Theorem 2.2], which even contains a more general multivariate result. However, the smoothness property of the limiting Gaussian process W and the particular form of the limit of S_n^2 have not been investigated in [19]. Both properties are crucial for the statistical analysis of the minimal contrast estimator.

We observe that from a statistical perspective it is more favourable to use Theorem 2.1(i) to estimate the parameter $\xi = (\sigma, \alpha, H)$, since the convergence rate $\sqrt{n}$ in (i) is faster than the rate $n^{1-1/\beta}$ in (ii). However, the phase transition happens at the point $k = H + 1/\alpha$, which depends on unknown parameters α and H. This poses major difficulties in statistical applications and we will address this issue in the forthcoming discussion.

3 Main Results

In this section we describe our minimal contrast approach and present the corresponding asymptotic theory. Before stating our the main result we define a power variation based estimator of the parameter $H \in (0,1)$. Since the increments of the process $(X_t)_{t\geq 0}$ are strongly ergodic (cf. [8]), we deduce by Birkhoff's ergodic theorem the almost sure convergence

$$\psi_n(r) = \frac{1}{n}\sum_{i=rk}^{n} |\Delta_{i,k}^r X|^p \xrightarrow{\text{a.s.}} \mathbb{E}[|\Delta_{rk,k}^r X|^p] = r^{pH} m_{p,k}$$

for any $p \in (-1, \alpha)$. In particular, we have that

$$R_n(p,k) := \frac{\psi_n(2)}{\psi_n(1)} \xrightarrow{\text{a.s.}} 2^{pH},$$

consequently yielding a consistent estimator $H_n(p,k)$ of H as

$$H_n(p,k) = \frac{1}{p}\log_2(R_n(p,k)) \xrightarrow{\text{a.s.}} H \tag{3.1}$$

for any $p \in (-1,\alpha)$. The idea of using negative powers $p \in (-1,0)$ to estimate H, which has been proposed in [12] and applied in [19], has the obvious advantage that it does not require knowledge of the parameter α. From Theorem 2.1(i) and the δ-method applied to the function $v_p(x,y) = \frac{1}{p}(\log_2(x) - \log_2(y))$ we immediately deduce the convergence

$$\left(\sqrt{n}(H_n(p,k) - H), W_n^2(t)\right) \xrightarrow{\mathcal{L}\text{-f}} (M_1, W_t) \tag{3.2}$$

for $k > H + 1/\alpha$, where M_1 is a centred Gaussian random variable. Similarly, when $k < H + 1/\alpha$, we deduce the convergence

$$\left(n^{1-1/\beta}(H_n(p,k) - H), S_n^2(t)\right) \xrightarrow{\mathcal{L}\text{-f}} (M_2, \kappa_2(t)S) \tag{3.3}$$

from Theorem 2.1(ii).

We will now introduce the minimal contrast estimator of the parameter $\theta = (\sigma,\alpha)$. Let $w \in \mathcal{L}^1(\mathbb{R}_+)$ denote a positive weight function. Define for $r > 1$ the norm

$$\|h\|_{w,r} = \left(\int_0^\infty |h(t)|^r w(t)\,\mathrm{d}t\right)^{1/r},$$

where $h : \mathbb{R}_+ \to \mathbb{R}$ is a Borel function. Denote by $\mathcal{L}_w^r(\mathbb{R}_+)$ the space of functions h with $\|h\|_{w,r} < \infty$. Let $(\theta_0, H_0) = (\sigma_0, \alpha_0, H_0) \in (0,\infty) \times (0,2) \times (0,1)$ be the true parameter of the model (1.1) and consider an open neighbourhood $\Theta_0 \subseteq (0,\infty) \times (0,2)$ around θ_0 bounded away from $(0,0)$. Define the map $F : \mathcal{L}_w^2(\mathbb{R}_+) \times (0,1) \times \Theta_0 \to \mathbb{R}$ as

$$F(\varphi, H, \theta) = \|\varphi - \varphi_{\theta,H}\|_{w,2}^2, \tag{3.4}$$

where $\varphi_{\theta,H}$ is the limit introduced at (2.2). We define the minimal contrast estimator θ_n of θ_0 as

$$\theta_n \in \operatorname*{argmin}_{\theta \in \Theta_0} F(\varphi_n, H_n(p,k), \theta). \tag{3.5}$$

We remark that by, e.g. [26, Theorem 2.17] it is possible to choose θ_n universally measurable with respect to the underlying probability space. The joint estimator of (θ_0, H_0) is then given as

$$\xi_n = (\theta_n, H_n(p,k))^\top.$$

Below we denote by ∇_θ the gradient with respect to the parameter θ and similarly by ∇_θ^2 the Hessian. We further write $\partial_z f$ for the partial derivative of f with respect to $z \in \{\sigma, \alpha, H\}$. The main theoretical result of this paper is the following theorem.

Theorem 3.1. *Suppose $\xi_0 = (\theta_0, H_0)$ is the true parameter of the linear fractional stable motion $(X_t)_{t\geq 0}$ at* (1.1) *and that the weight function $w \in \mathcal{L}^1(\mathbb{R}_+)$ is continuous.*

(i) $\xi_n \xrightarrow{\text{a.s.}} \xi_0$ *as $n \to \infty$.*

(ii) *If $k > H + 1/\alpha$ then*

$$\sqrt{n}(\xi_n - \xi_0) \xrightarrow{\mathcal{L}} \Big[-2\nabla_\theta^2 F(\varphi_{\xi_0}, \xi_0)^{-1}\Big(\int_0^\infty W_t \nabla_\theta \varphi_{\xi_0}(t) w(t)\,\mathrm{d}t + \partial_H \nabla_\theta F(\varphi_{\xi_0}, \xi_0) M_1\Big), M_1\Big]^\top.$$

(iii) *If $k < H + 1/\alpha$ then*

$$n^{1-1/\beta}(\xi_n - \xi_0) \xrightarrow{\mathcal{L}} \Big[-2\nabla_\theta^2 F(\varphi_{\xi_0}, \xi_0)^{-1}\Big(S\int_0^\infty \kappa_2(t)\nabla_\theta \varphi_{\xi_0}(t)w(t)\,\mathrm{d}t + \partial_H \nabla_\theta F(\varphi_{\xi_0}, \xi_0)M_2\Big), M_2\Big]^\top.$$

In principle, the statement of Theorem 3.1 follows from (3.2), (3.3) and an application of the implicit function theorem. For general infinite dimensional functionals of our statistics we would usually need to show tightness of the process $(W_n^2(t))_{t\geq 0}$ (or $(S_n^2(t))_{t\geq 0}$), which is by far not a trivial issue. However, in the particular setting of integral functionals, it suffices to show a weaker condition that is displayed in Proposition 5.6. Indeed, this is the key step of the proof.

4 Simulations, Parametric Bootstrap and Subsampling

The theoretical results of Theorem 3.1 are far from easy to apply in practice. There are a number of issues, which need to be addressed. First of all, since the parameters H_0 and α_0 are unknown, we do not know whether we are in the regime of Theorem 3.1(ii) or (iii). Furthermore, even if we could determine whether the condition $k > H_0 + 1/\alpha_0$ holds or not, the exact computation or a reliable numerical simulation of the quantities defined in (2.6) and (2.7) seems to be out of reach. Below we will propose two methods to overcome these problems. In the setting where the lfsm $(X_t)_{t\geq 0}$ is continuous, which corresponds to the condition $H_0 - 1/\alpha_0 > 0$, we will see that it suffices to choose $k = 2$ to end up in the normal regime of Theorem 3.1(ii). The confidence regions are then constructed using the parametric bootstrap approach. In the general setting we propose a novel subsampling method which, in some sense, automatically adapts to the unknown limiting distribution.

For comparison reasons we include the estimation of the parameter H_0 using $H_n(p,k)$ defined at (3.1), even though its properties have already been studied in [19]. Moreover, we pick our weight function in the class of Gaussian kernels:

$$w_\nu(t) = \exp(-\tfrac{t^2}{2\nu^2}) \qquad (t, \nu > 0).$$

In particular we can use Gauss–Hermite quadrature, see [25], to estimate the integral

$$\|\varphi_n - \varphi_\xi\|_{w,2}^2 = \int_0^\infty (\varphi_n(t) - \varphi_\xi(t))^2 w_\nu(t)\,\mathrm{d}t.$$

This procedure is based on a number of weights, unless otherwise stated we pick 12 weights. We mentioned that it is possible to choose other weight functions and standard numerical procedures exist for these. We restrict our simulation study to three different values of $\nu \in \{0.05, 0.1, 1\}$ for the bootstrap method in Section 4.2. Additionally, while the theoretical characteristic function φ_ξ has an explicit form it depends on the norm $\|h_k\|_\alpha$ which is not readily computable, hence needs to be approximated.

To produce the simulation study we generate observations from the lfsm using the recent R-package `rlfsm`, which implements an algorithm based on [28]; this package already includes an implementation of the minimal contrast estimator. To compute the estimator a minimization

$$\operatorname*{argmin}_{\sigma,\alpha} \int_0^\infty (\varphi_n(t) - \varphi_{\sigma,\alpha,H_n(p,k)}(t))^2 w_\nu(t)\,\mathrm{d}t$$

has to be carried out. For this purpose we use [20], which in particular entails picking a starting point for the algorithm. For σ no immediate choice exists, so we simply pick $\sigma = 2$, while $\alpha = 1$ seems obvious.

4.1 Empirical Bias and Variance

In this section we will check the bias and variance performance of our minimal contrast estimator. First, we consider the empirical bias and standard deviation, which are summarized in Tables B.5 and B.6 for $n = 1000$ and Tables B.7 and B.8 for $n = 10\,000$. These are based on Monte Carlo simulation with at least 1000 repetitions. We fix the parameters $k = 2$, $p = -0.4$, $\nu = 0.1$ and $\sigma_0 = 0.3$ and perform the estimation procedure for various values of α and H.

At this stage we recall that due to Theorem 3.1(iii) we obtain a slower rate of convergence when $H_0 + 1/\alpha_0 > 2$; the stable regime is indicated in bold in Tables B.5–B.8. This explains a rather bad estimation performance for $\alpha_0 = 0.4$. The effect is specifically pronounced for the parameter σ, which has the worst performance when $\alpha_0 = 0.4$. This observation is in line with the findings of [19], who concluded that the scale parameter σ is the hardest to estimate in practice. Also the starting point $\sigma = 2$ of the minimization algorithm, which is not close to $\sigma_0 = 0.3$, might have a negative effect on the performance. The estimation performance for values $\alpha_0 > 0.4$ is quite satisfactory for all parameters, improving from $n = 1000$ to $n = 10\,000$. We remark the superior performance of our method around the value $\alpha = 1$, which is explained by the fact that $\alpha = 1$ is the starting point of the minimisation procedure.

For comparison, we display the bias and standard deviation of our estimator for $k = 1$ based on $n = 10\,000$ observations in Tables B.9 and B.10, where the stable regime is again highlighted in bold. We see a better finite sample performance for $\alpha_0 = 0.4$, but in most other cases we observe a larger bias and standard deviation compared to $k = 2$. This is explained by slower rates of convergence in the setting of the stable regime and $k = 1$.

We will now compare the minimal contrast estimator with the estimator proposed in [19]. To recall the latter estimator we observe the following identities due to (2.2):

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

for fixed values $0 < t_1 < t_2$. Since h_k depends on α and H we immediately obtain a function $G : \mathbb{R}^3 \to \mathbb{R}^2$ such that

$$(\sigma, \alpha) = G(\varphi_\xi(t_1;k), \varphi_\xi(t_2;k), H), \qquad \xi = (\sigma, \alpha, H).$$

Hence, an estimator for (σ, α, H) is obtained by insertion of the empirical characteristic function and $H_n(p,k)$ from (3.1):

$$(\widetilde{\sigma}_{\text{low}}, \widetilde{\alpha}_{\text{low}}, H_n(p,k)) = G\big(\varphi_n(t_1;k), \varphi_n(t_2;k), H_n(p,k)\big). \tag{4.1}$$

The first comparison of the estimators is between Tables B.7 and B.8 for the minimal contrast estimator with Tables B.11 and B.12, all of which are based on at least

1000 Monte Carlo repetitions and the same parameter choices. We see that the minimal contrast estimator outperforms the old estimator for values $\alpha \leq 0.8$, where the latter is completely unreliable in most cases. But for larger values of α the old estimator might be slightly better. However, the results for the estimation of the scale parameter σ in the case $\alpha = 0.4$ and $H_0 \in \{0.2, 0.4\}$ are hard to interpret, since $\widetilde{\sigma}_{\text{low}}$ often delivers values that are indistinguishable from 0 yielding a small variance and relatively small bias for $\sigma_0 = 0.3$.

Another point is the instability of the old estimator. Table B.13, which is based on at least 1200 simulations, shows the rate at which the estimators fail to return a value $\alpha \in (0,2)$. In this regard the minimal contrast estimator is far superior in most cases and actually in the case $k = 1$ the estimator almost never fails, although we dispense with the simulation results. We remark that in theory the minimal contrast estimator should never return α's not in the interval $[0,2]$. However, we apply the minimization procedure from [20] that does not allow constrained optimization. One could instead use, e.g. the procedure from [7]; we choose the first method as it is hailed as very robust.

An additional advantage of the minimal contrast estimator is that it allows incorporation of a priori knowledge of the parameters σ and α, using the weight function but also the starting point for the minimization algorithm.

4.2 Bootstrap Inference in the Continuous Case

In this section we only consider the continuous case, which corresponds to the setting $H_0 - 1/\alpha_0 > 0$. In this setup $H_0 \in (1/2, 1)$ and $\alpha_0 \in (1,2)$ must hold. We can in particular choose $k = 2$ to ensure that Theorem 3.1(ii) applies, thus yielding the faster convergence rate $\sqrt{n}$. We are interested in obtaining feasible confidence regions for all parameters, but, as we mentioned earlier, the computation or reliable numerical approximation of the asymptotic variance in Theorem 3.1(ii) is out of reach. Instead we propose the following parametric bootstrap procedure to estimate the confidence regions for the true parameters.

(1) Compute the minimal contrast ξ_n estimator for given observations $X_1, \dots, X_n$.

(2) Generate new samples $X_1^i, \dots, X_n^i$ for $i = 1, \dots, N$ using the parameter ξ_n.

(3) Compute new estimators ξ_n^i from the samples generated in (2) for each $i = 1, \dots, N$.

(4) Calculate the empirical variance $\hat{\Sigma}_n$ of ξ_n based on the estimators $\xi_n^1, \dots, \xi_n^N$.

(5) For each parameter construct 95 %-confidence regions based on the relation

$$\sqrt{n}(\xi_n - \xi_0) \approx \mathcal{N}(0, n\hat{\Sigma}_n).$$

To test this we repeated the above procedure for 200 Monte Carlo simulations with $N = 200$. Tables B.1 and B.2 report acceptance rates for $\sqrt{n}(\xi_n - \xi_0)$ for an approximate 95 %-confidence interval. We observe a good performance for all estimators with the exception of $n = 1000$ for σ with $\nu = 0.05, 0.1$ and α for $\nu = 0.05$ in Table B.1. The estimator is fairly stable under changes in ν in this parameter regime, but it should be mentioned that a smaller ν-value does lead to a larger failure rate, we dispense with the numerics.

Table B.1. Acceptance rates for the true parameter $(\sigma_0, \alpha_0, H_0) = (0.3, 1.8, 0.8)$ and power $p = 0.4$.

$\nu = 0.1$

n	σ	α	H
1000	69.2	95.8	96.3
2500	96.6	98.5	95.6
5000	99.0	99.5	96.1

$\nu = 1$

n	σ	α	H
1000	100	95.5	94.8
2500	100	96.3	93.1
5000	100	93.5	93.5

$\nu = 0.05$

n	σ	α	H
1000	81.7	44.2	99.5
2500	99.6	97.9	97.9
5000	100	99.5	96.8

Table B.2. Acceptance rates for the true parameter $(\sigma_0, \alpha_0, H_0) = (0.3, 1.3, 0.8)$ and power $p = 0.4$.

$\nu = 0.1$

n	σ	α	H
1000	87.1	95.4	93.1
2500	90.1	91.6	97.5
5000	91.6	91.6	95.1

$\nu = 1$

n	σ	α	H
1000	93.9	95.7	93.9
2500	95.7	96.7	94.3
5000	93.5	93.1	93.5

$\nu = 0.05$

n	σ	α	H
1000	91.9	98.6	95.9
2500	89.8	96.6	95.1
5000	91.4	94.3	97.1

4.3 Subsampling Method in the General Case

In contrast to the continuous setting $H_0 - 1/\alpha_0 > 0$, there exists no a priori choice of k in the general case, which ensures the asymptotically normal regime of Theorem 3.1(ii). This problem was tackled in [19] via the following two stage approach. In the first step they obtained a preliminary estimator $\alpha_n^0(t_1, t_2)$ of α_0 using (4.1) for $k = 1$, $t_1 = 1$ and $t_2 = 2$. In the second step they defined the random number

$$\hat{k} = 2 + \lfloor \alpha_n^0(t_1, t_2)^{-1} \rfloor, \tag{4.2}$$

and computed the estimator $(\widetilde{\sigma}_{\text{low}}, \widetilde{\alpha}_{\text{low}}, H_n(p,k))$ based on $k = \hat{k}$. They showed that the resulting estimator is $\sqrt{n}$-consistent and derived the associated weak limit theory. However, this approach does not completely solve the original problem, since they obtained four different convergence regimes according to whether $1 > H_0 + 1/\alpha_0$ or not, and whether $\alpha_0^{-1} \in \mathbb{N}$ or not.

Nevertheless, we apply their idea to propose a new subsampling method to determine feasible confidence regions for the parameters of the model. For our procedure it is crucial that the convergence rate is known explicitly and the weak convergence of the involved statistics is insured. We proceed as follows:

(1) Given observations $X_1,\dots,X_n$ compute $\hat{k}$ from (4.2) and construct the minimal contrast estimator $\xi_n = (\sigma_n, \alpha_n, H_n(p,\hat{k}))$.

(2) Split $X_1,\dots,X_n$ into L groups where the lth group contains $(X_{(l-1)n/L+i})_{i=1}^{n/L}$ (n/L is assumed to be an integer). For each $l = 1,\dots,L$ calculate $\hat{k}_l$ from (4.2).

(3) For each $l = 1,2,\dots,L$ construct the minimal contrast estimators (σ_n^l, α_n^l) and $H_n^l(p,\hat{k}_l)$ based on the lth group. For the estimation of (σ,α) use $H_n(p,\hat{k})$ from (1) as plug-in.

(4) Compute the 97.5% and 2.5% quantiles for each of the distribution functions

$$\frac{1}{L}\sum_{l=1}^{L} \mathbb{1}_{\left\{\sqrt{\frac{n}{L}}(\sigma_n^l-\sigma_n)\leq x\right\}}, \qquad \frac{1}{L}\sum_{l=1}^{L} \mathbb{1}_{\left\{\sqrt{\frac{n}{L}}(\alpha_n^l-\alpha_n)\leq x\right\}}, \qquad \frac{1}{L}\sum_{l=1}^{L} \mathbb{1}_{\left\{\sqrt{\frac{n}{L}}(H_n^l(p,\hat{k}_l)-H_n(p,\hat{k}))\leq x\right\}}.$$

Let us explain the intuition behind the proposed subsampling procedure. First of all, similarly to the theory developed in [19], the minimal contrast estimator ξ_n obtained through a two step method described in the beginning of the section leads to four different limit regimes for $\sqrt{n}(\xi_n-\xi_0)$ (although we leave out the theoretical derivation here). Using this knowledge we may conclude that, for each $l = 1,\dots,L$, $\sqrt{n/L}(\xi_n^l - \xi_0)$ has the same (unknown) asymptotic distribution as the statistic $\sqrt{n}(\xi_n - \xi_0)$ as long as $n/L \to \infty$. Since the true parameter ξ_0 is unknown, we use its approximation ξ_n, which has a much faster rate of convergence than $\sqrt{n/L}$ when $L \to \infty$. Finally, the statistics constructed on different blocks are asymptotically independent, which follows along the lines of the proofs in [19]. Hence, the law of large numbers implies that the proposed subsampling statistics converge to the unknown true asymptotic distributions when $L \to \infty$ and $n/L \to \infty$.

In Tables B.3 and B.4 we report the empirical 95%-confidence regions for the parameters of the model using the subsampling approach. We perform 500 Monte Carlo simulations and choose $n = 12.5 \times L^2$.

Table B.3. Acceptance rates (%) for the true parameter $(\sigma_0, \alpha_0, H_0) = (0.3, 0.8, 0.8)$. Here $p = -0.4$ and $\nu = 0.1$.

L	n/L	σ	α	H
80	1000	90.65	94.39	89.72
100	1250	87.72	94.24	89.25

Table B.4. Acceptance rates (%) for the true parameter $(\sigma_0, \alpha_0, H_0) = (0.3, 1.8, 0.8)$. Here $p = -0.4$ and $\nu = 0.1$.

L	n/L	σ	α	H
80	1000	60.70	67.31	92.19
100	1250	68.88	74.92	94.56

Table B.3 shows a satisfactory performance for all estimators, while the results of Table B.4 are quite unreliable for the parameters σ and α. The reason for the latter finding is the suboptimal finite sample performance of the estimators in the case of $(\alpha, H) = (1.8, 0.8)$, which is displayed in Tables B.5 and B.6.

We conclude this section by remarking the rather satisfactory performance of our estimator in the continuous setting $H_0 - 1/\alpha_0 > 0$. On the other hand, when using the subsampling method in the general setting, a further careful tuning seems to be required. In particular, the choice of the weight function w and the group number L plays an important role in estimator's performance. We leave this study for future research.

5 Proofs

We denote by C a finite, positive constant which may differ from line to line. Moreover, any important dependence on other constants warrants a subscript. To simplify notations we set $H_n = H_n(p,k)$.

5.1 Proof of Theorem 2.1(i)

As we mentioned earlier, the convergence of finite dimensional distributions has been shown in [19, Theorem 2.2], and thus we only need to prove the smoothness property of the limit W. We recall the definition of the quantity $U_{g,h}$ at (2.4) and start with the following lemma.

Lemma 5.1 ([23, Eqs. (3.4)–(3.6)]).
Let $g, h \in \mathcal{L}^\alpha(\mathbb{R}_+)$. Then for any $u, v \in \mathbb{R}$

$$\begin{aligned}|U_{g,h}(u,v)| &\le 2|uv|^{\alpha/2} \int_0^\infty |g(x)h(x)|^{\alpha/2}\,\mathrm{d}x \\ &\quad \times \exp\Big(-2|uv|^{\alpha/2}\Big(\|g\|_\alpha^\alpha \|h\|_\alpha^\alpha - \int_0^\infty |g(x)h(x)|^{\alpha/2}\,\mathrm{d}x\Big)\Big), \\ |U_{g,h}(u,v)| &\le 2|uv|^{\alpha/2} \int_0^\infty |g(x)h(x)|^{\alpha/2}\,\mathrm{d}x \\ &\quad \times \exp\Big(-\big(\|ug\|_\alpha^{\alpha/2} - \|vh\|_\alpha^{\alpha/2}\big)^2\Big).\end{aligned}$$

In particular, it holds that $|U_{g,h}(u,v)| \le 2|uv|^{\alpha/2} \int_0^\infty |g(x)h(x)|^{\alpha/2}\,\mathrm{d}x$.

Next, we define for each $l \in \mathbb{Z}$

$$\rho_l = \int_0^\infty |h_k(x)h_k(x+l)|^{\alpha/2}\,\mathrm{d}x$$

and recall the following lemma from [19].

Lemma 5.2 ([19, Lemma 6.2]).
If $k > H + 1/\alpha$ and $l > k$ then

$$\rho_l \le l^{(\alpha(H-k)-1)/2}.$$

To prove that the process W is locally Hölder continuous of any order smaller than $\alpha/4$, we use Kolmogorov's criterion. Since W is a Gaussian process it suffices to prove that for each $T > 0$ there exists a constant $C_T \geq 0$ such that

$$\mathbb{E}[(W_t - W_s)^2] \leq C_T\,|t-s|^{\alpha/2} \qquad \text{for all } s,t \in [0,T]. \tag{5.1}$$

This is performed in a similar fashion as in [17, Section 4.1]. First, we reduce the problem. Using $\cos(tx) = (\exp(\mathrm{i}tx)+\exp(-\mathrm{i}tx))/2$ and the symmetry of the distribution of X, we observe the identity

$$\operatorname{Cov}(\cos(t\Delta_{i,k}X), \cos(s\Delta_{i+l,k}X)) = \tfrac{1}{2}\big(U_{h_k,-h_k(l+\cdot)}(t,s) + U_{h_k,h_k(l+\cdot)}(t,s)\big).$$

In the following we focus on the first term in the above decomposition (the second term is treated similarly). More specifically, we will show the inequality (5.1) for the quantity $\overline{r}(t,s)$, which is given as

$$\begin{aligned}
\overline{r}(t,s) &= \sum_{l\in\mathbb{Z}} \overline{r}_l(t,s) \qquad \text{where} \\
\overline{r}_l(t,s) &= U_{h_k,-h_k(l+\cdot)}(t,s) \\
&= \exp(-\|th_k - sh_k(l+\cdot)\|_\alpha^\alpha) - \exp(-(t^\alpha + s^\alpha)\,\|h_k\|_\alpha^\alpha).
\end{aligned}$$

Moreover, since $\overline{r}(t,t)+\overline{r}(s,s)-2\overline{r}(t,s) \leq |\overline{r}(t,t)-\overline{r}(t,s)|+|\overline{r}(s,s)-\overline{r}(t,s)|$ it is by symmetry enough to prove that

$$|\overline{r}(t,t) - \overline{r}(t,s)| \leq C_T\,|t-s|^{\alpha/2} \qquad \text{for all } s,t \in [0,T].$$

For $l \in \mathbb{Z}$ decompose now as follows:

$$\begin{aligned}
&\overline{r}_l(t,t) - \overline{r}_l(t,s) \\
&\quad = \exp(-2t^\alpha\,\|h_k\|_\alpha^\alpha)\Big[\exp(-\|t(h_k - h_k(\,\cdot\,+l))\|_\alpha^\alpha + 2t^\alpha\,\|h_k\|_\alpha^\alpha) - 1\Big] \\
&\qquad - \exp(-(t^\alpha+s^\alpha)\,\|h_k\|_\alpha^\alpha)\Big[\exp(-\|th_k - sh_k(l+\cdot)\|_\alpha^\alpha + (t^\alpha+s^\alpha)\,\|h_k\|_\alpha^\alpha) - 1\Big] \\
&\quad = \Big[\exp(-2t^\alpha\,\|h_k\|_\alpha^\alpha) - \exp(-(t^\alpha+s^\alpha)\,\|h_k\|_\alpha^\alpha)\Big] \\
&\qquad \times \Big[\exp(-\|t(h_k - h_k(l+\cdot))\|_\alpha^\alpha + 2t^\alpha\,\|h_k\|_\alpha^\alpha) - 1\Big] + \exp(-(t^\alpha+s^\alpha)\,\|h_k\|_\alpha^\alpha) \\
&\qquad \times \Big[\exp(-\|t(h_k - h_k(l+\cdot))\|_\alpha^\alpha + 2t^\alpha\,\|h_k\|_\alpha^\alpha) \\
&\qquad\quad - \exp(-\|th_k - sh_k(l+\cdot)\|_\alpha^\alpha + (t^\alpha+s^\alpha)\,\|h_k\|_\alpha^\alpha)\Big] \\
&\quad =: \overline{r}_l^{(1)}(t,s) + \overline{r}_l^{(2)}(t,s).
\end{aligned}$$

Applying the second inequality of Lemma 5.1 and the mean value theorem we deduce the estimate:

$$|\overline{r}^{(1)}(t,s)| \leq C_T\rho_l\,|t^\alpha - s^\alpha| \leq C_T\rho_l\,|t-s|^{\alpha/2} \qquad \text{for all } s,t \in [0,T]. \tag{5.2}$$

Again by the mean value theorem we find that

$$|\overline{r}^{(2)}(t,s)| \leq C_T\,\Big|\|th_k - sh_k(l+\cdot)\|_\alpha^\alpha - \|t(h_k - h_k(l+\cdot))\|_\alpha^\alpha + (t^\alpha - s^\alpha)\,\|h_k\|_\alpha^\alpha\Big|.$$

The last term can be rewritten as

$$\begin{aligned}
&\|th_k - sh_k(l+\cdot)\|_\alpha^\alpha - \|t(h_k - h_k(l+\cdot))\|_\alpha^\alpha + (t^\alpha - s^\alpha)\,\|h_k\|_\alpha^\alpha \\
&\qquad = \int_0^\infty |th_k(x) - sh_k(x+l)|^\alpha - |t(h_k(x) - h_k(l+x))|^\alpha + (t^\alpha - s^\alpha)\,|h_k(x+l)|^\alpha\,\mathrm{d}x.
\end{aligned}$$

As $\alpha \in (0,2)$ we have $|x^\alpha - y^\alpha| \le |x^2 - y^2|^{\alpha/2}$ for all $x, y \ge 0$. In particular

$$\begin{aligned}&\Big||th_k(x) - sh_k(x+l)|^\alpha - |t(h_k(x) - h_k(x+l))|^\alpha\Big| \\ &\qquad \le C_T\,|t-s|^{\alpha/2}\big(|h_k(x+l)|^\alpha + |h_k(x)h_k(x+l)|^{\alpha/2}\big).\end{aligned}$$

It then follows that

$$|\overline{r}_l^{(2)}(t,s)| \le C_T\,|t-s|^{\alpha/2}(\rho_l + \mu_l) \qquad \text{for all } s,t \in [0,T], \tag{5.3}$$

where μ_l is the quantity defined as

$$\mu_l = \int_0^\infty |h_k(x+l)|^\alpha \,\mathrm{d}x.$$

It remains to prove that

$$\sum_{l\in\mathbb{Z}} \rho_l < \infty \qquad \text{and} \qquad \sum_{l\in\mathbb{Z}} \mu_l < \infty. \tag{5.4}$$

The first claim is a direct consequence of Lemma 5.2. The second convergence can equivalently be formulated as

$$\sum_{l=1}^\infty l \int_l^{l+1} |h_k(x)|^\alpha \,\mathrm{d}x < \infty.$$

Recall that $|h_k(x)| \le C\,|x|^{H-1/\alpha-k}$ for large x, hence

$$\sum_{l=1}^\infty l \int_l^{l+1} |h_k(x)|^\alpha \,\mathrm{d}x \le C \sum_{l=1}^\infty l \int_l^{l+1} x^{\alpha(H-k)-1} \,\mathrm{d}x \le C \sum_{l=1}^\infty l^{\alpha(H-k)} < \infty,$$

where we used the assumption $k > H + 1/\alpha$. Combining (5.2) and (5.3) with (5.4) we can conclude (5.1), and hence the proof of Theorem 2.1(i) is complete.

5.2 Proof of Theorem 2.1(ii)

We recall that the asymptotic distribution of the vector $(S_n^1(1), S_n^1(2))$ and its asymptotic independence of $S_n^2(t)$ have been shown in [19, Theorem 2.2]. Hence, we only need to determine the functional form of the limit of the statistic $S_n^2(t)$.

In the following we will recall a number of estimates and decompositions from [19, Theorem 2.2], which will be also helpful in the proof of Theorem 3.1(iii). We start out with a series of estimates on the function Φ_t^2 given at (2.7), but for a general scale parameter $\eta > 0$. Let $\Phi_{t,\eta}$ denote the function

$$\Phi_{t,\eta}(x) = \mathbb{E}[\cos(t(Y+x))] - \mathbb{E}[\cos(tY)] \qquad x \in \mathbb{R},$$

where Y is an SαS distributed random variable with scale parameter η. We obviously have the representation

$$\Phi_{t,\eta}(x) = (\cos(xt) - 1)\exp(-|\eta t|^\alpha). \tag{5.5}$$

The next lemma gives some estimates on the function $\Phi_{t,\eta}$.

Lemma 5.3. *For $\eta > 0$ set $g_\eta(t) = \exp(-|\eta t|^\alpha)$ and let $\Phi_{t,\eta}^{(v)}(x)$ denote the vth derivative at $x \in \mathbb{R}$. Then there exists a constant $C > 0$ such that for all $t \ge 0$ it holds that*

(i) $|\Phi^{(v)}_{t,\eta}(x)| \le Ct^v g_\eta(t)$ *for all* $x \in \mathbb{R}$ *and* $v \in \{0,1,2\}$.

(ii) $|\Phi_{t,\eta}(x)| \le g_\eta(t)(1 \wedge |xt|^2)$.

(iii) $|\Phi_{t,\eta}(x) - \Phi_{t,\eta}(y)| \le t^2 g_\eta(t)((1 \wedge |x| + 1 \wedge |y|)|x-y|\mathbb{1}_{\{|x-y|\le 1\}} + \mathbb{1}_{\{|x-y|>1\}})$.

(iv) *For any* $x, y > 0$ *and* $a \in \mathbb{R}$ *then*

$$F(a,x,y) := \left| \int_0^y \int_0^x \Phi^{(v)}_{t,\eta}(a+u+v)\,\mathrm{d}u\,\mathrm{d}v \right| \le Cg_\eta(t)(t+1)^2(1 \wedge x)(1 \wedge y).$$

Proof. (i): This follows directly from (5.5).

(ii): This is straightforward using the standard inequality $1 - \cos(y) \le y^2$.

(iii): (i) implies that $|\Phi^{(1)}_{t,\eta}(x)| \le t^2 g_\eta(t)(1 \wedge |x|)$ and note that

$$|\Phi_{t,\eta}(x) - \Phi_{t,\eta}(y)| = \left| \int_y^x \Phi^{(1)}_{t,\eta}(u)\,\mathrm{d}u \right|.$$

If $|x-y| > 1$ we simply bound the latter by $t^2 g_\eta(t)$. If $|x-y| \le 1$, then by the mean value theorem there exists a number s with $|x-s| \le |x-y|$ such that

$$\left| \int_y^x \Phi^{(1)}_{t,\eta}(u)\,\mathrm{d}u \right| = |\Phi^{(1)}_{t,\eta}(s)||x-y|.$$

Observe then

$$|\Phi^{(1)}_{t,\eta}(s)| \le t^2 g_\eta(t)(1 \wedge |s|) \le t^2 g_\eta(t)\big(1 \wedge (|x| + |y|)\big).$$

This completes the proof of the inequality.

(iv): Let a, x and y be given. Observe that

$$\begin{aligned}\int_0^y \int_0^x \Phi^{(2)}_{t,\eta}(a+u+v)\,\mathrm{d}u\,\mathrm{d}v &= \int_0^y \Phi^{(1)}_{t,\eta}(a+x+v) - \Phi^{(1)}_{t,\eta}(a+v)\,\mathrm{d}v \\ &= \Phi_{t,\eta}(a+x+y) - \Phi_{t,\eta}(a+y) - (\Phi_{t,\eta}(a+x) - \Phi_{t,\eta}(a)).\end{aligned}$$

The last equality implies that $F(a,x,y) \le Cg_\eta(t)$. The first equality implies that $F(a,x,y) \le Cg_\eta(t)ty$. Reversing the order of integration we get a similar expression as the first equality with x replaced by y. Hence, $F(a,x,y) \le Cg_\eta(t)tx$. Lastly, using (i) on the first integral yields $F(a,x,y) \le Cg_\eta(t)t^2xy$. Splitting into the four cases completes the proof. □

We will consider the asymptotic decomposition of the statistic $S^2_n(t)$ given in [3, Section 5] (see also [4, 19]). We set

$$S^2_n(t) = n^{-1/\beta} \sum_{i=k}^{n} \big(\cos(t\Delta_{i,k}X) - \varphi_\xi(t)\big) =: n^{-1/\beta} \sum_{i=k}^{n} V_i(t).$$

Define for each $s \ge 0$ we define the σ-algebras

$$\mathcal{G}_s = \sigma(L_v - L_u : v, u \le s) \qquad \text{and} \qquad \mathcal{G}^1_s = \sigma(L_v - L_u : s \le v, u \le s+1).$$

We also set for all $n \geq k$, $i \in \{k,\dots,n\}$ and $t \geq 0$

$$R_i(t) = \sum_{j=1}^{\infty} \zeta_{i,j}(t) \qquad \text{and} \qquad Q_i(t) = \sum_{j=1}^{\infty} \mathbb{E}[V_i(t) \,|\, \mathcal{G}^1_{i-j}],$$

where

$$\zeta_{i,j}(t) = \mathbb{E}[V_i(t) \,|\, \mathcal{G}_{i-j+1}] - \mathbb{E}[V_i(t) \,|\, \mathcal{G}_{i-j}] - \mathbb{E}[V_i(t) \,|\, \mathcal{G}^1_{i-j}].$$

Then the following decomposition holds:

$$S_n^2(t) = n^{-1/\beta} \sum_{i=k}^{n} R_i(t) + \Big(n^{-1/\beta} \sum_{i=k}^{n} Q_i(t) - \overline{S}_n(t) \Big) + \overline{S}_n(t), \tag{5.6}$$

where

$$\overline{S}_n(t) = n^{-1/\beta} \sum_{i=k}^{n} \big(\overline{\Phi}_t(L_{i+1} - L_i) - \mathbb{E}[\overline{\Phi}_t(L_{i+1} - L_i)] \big),$$

$$\overline{\Phi}_t(x) := \sum_{i=1}^{\infty} \Phi_t^2(h_k(i)x). \tag{5.7}$$

It turns out that the first two terms in (5.6) are negligible while $\overline{S}_n$ is the dominating term. More specifically, we can use similar arguments as in [4, Eq. (5.22)] and deduce the following proposition from Lemma 5.3.

Proposition 5.4. *For any $\varepsilon > 0$ there exists a constant $C > 0$ such that for all $n \in \mathbb{N}$*

$$\sup_{t \geq 0} \mathbb{E}\Big[\Big(n^{-1/\beta} \sum_{i=k}^{n} R_i(t) \Big)^2 \Big] \leq C n^{2(2-\beta-1/\beta)+\varepsilon}.$$

Using the inequality $2 - x - 1/x < 0$ for all $x > 1$ on $\beta > 1$ it follows by picking $\varepsilon > 0$ small enough that the first term in (5.6) is asymptotically negligible. Decomposing the second term and using arguments as in the equations (5.30), (5.31) and (5.38) in [4] we obtain the following result.

Proposition 5.5. *For any $\varepsilon > 0$ there exist an $r > 1$, an $r' > \beta \vee r$ and a constant $C > 0$ such that for all n in $\mathbb{N}$*

$$\sup_{t \geq 0} \mathbb{E}\Big[\Big| n^{-1/\beta} \sum_{i=k}^{n} Q_i(t) - \overline{S}_n(t) \Big|^r \Big] \leq C\Big(n^{r(\varepsilon+2-\beta-1/\beta)} + n^{\frac{r}{r'}(1-r'/\beta)} \Big).$$

Using again the inequality $2 - x - 1/x < 0$ for all $x > 1$, it follows immediately that the second term in (5.6) is asymptotically negligible. Hence, $\overline{S}_n(t)$ is asymptotically equivalent to the statistic $S_n^2(t)$, and it suffices to analyse its finite dimensional distribution.

Consider $t_1, \dots, t_d \in \mathbb{R}_+$. We will now recall the limiting distribution of the vector $(\overline{S}_n(t_1), \dots, \overline{S}_n(t_d))$. Observing the definition (5.7), we deduce the uniform convergence

$$\sup_{t \geq 0} \Big| |x|^{-\alpha/\beta} \overline{\Phi}_t(x) - \kappa_2(t) \Big| \to 0 \qquad \text{as } x \to -\infty, \tag{5.8}$$

where κ_2 has been introduced at (2.8). Indeed by substituting $u = (|x|/z)^{\alpha/\beta}$ we have that

$$\begin{aligned}
&\sup_{t\geq 0}\Big||x|^{-\alpha/\beta}\overline{\Phi}_t(x) - \kappa_2(t)\Big| \\
&\quad = \sup_{t\geq 0}\Big||x|^{-\alpha/\beta}\int_0^\infty \Phi_t^2\big(h_k(\lfloor u\rfloor + 1)|x|\big)\,\mathrm{d}u - \int_0^\infty \Phi_t^2(q_{H,\alpha,k}z)z^{-1-\alpha/\beta}\,\mathrm{d}u\Big| \\
&\quad = \frac{\alpha}{\beta}\sup_{t\geq 0}\Big|\int_0^\infty \Phi_t^2\Big(h_k\big(\lfloor(|x|/z)^{\alpha/\beta}\rfloor + 1\big)|x|\Big)z^{-1-\alpha/\beta}\,\mathrm{d}u - \int_0^\infty \Phi_t^2(q_{H,\alpha,k}z)z^{-1-\alpha/\beta}\,\mathrm{d}u\Big| \\
&\quad \leq \frac{\alpha}{\beta}\int_0^\infty \sup_{t\geq 0}\Big|\Phi_t^2\Big(h_k\big(\lfloor(|x|/z)^{\alpha/\beta}\rfloor + 1\big)x\Big) - \Phi_t^2(q_{H,\alpha,k}z)\Big|z^{-1-\alpha/\beta}\,\mathrm{d}u.
\end{aligned}$$

By Lemma 5.3(iii) the integrand vanishes pointwise in z as $x \to -\infty$ due to the asymptotics

$$h_k(x) \sim q_{H,\alpha,k}x^{-\beta/\alpha} \qquad \text{as } x \to \infty. \tag{5.9}$$

Due to Lebesgue's dominated convergence theorem it is enough to bound the integrand uniformly in $x < -1$. By the triangle inequality it is enough to treat each Φ_t^2-term separately. For the first term Lemma 5.3(ii) implies that

$$\sup_{t\geq 0}\Big|\Phi_t^2\Big(h_k\big(\lfloor(|x|/z)^{\alpha/\beta}\rfloor + 1\big)|x|\Big)\Big|z^{-1-\alpha/\beta} \leq C\Big(1 \wedge \Big|h_k\big(\lfloor(|x|/z)^{\alpha/\beta}\rfloor + 1\big)x\Big|^2\Big)z^{-1-\alpha/\beta}.$$

For large z, say $z > 1$, the latter is bounded by the integrable function $z^{-1-\alpha/\beta}\mathbb{1}_{\{z>1\}}$. For $z \in (0,1]$ we deduce by (5.9)

$$\Big|h_k\big(\lfloor(|x|/z)^{\alpha/\beta}\rfloor + 1\big)x\Big|^2 z^{-1-\alpha/\beta} \leq Cx^2\big(\lfloor(|x|/z)^{\alpha/\beta}\rfloor + 1\big)^{-2\beta/\alpha}z^{-1-\alpha/\beta} \leq Cz^{1-\alpha/\beta},$$

where we used that $(|x|/z)^{\alpha/\beta} \leq \lfloor(|x|/z)^{\alpha/\beta}\rfloor + 1$. Recalling (2.3), we deduce that $\alpha/\beta < 2$ and an integrable bound is obtained. The second Φ_t^2-term is treated similarly. Hence, we have (5.8).

Define the map $\tau_{t_1,\dots,t_d} : \mathbb{R} \to \mathbb{R}_-^d$ as

$$\tau_{t_1,\dots,t_d}(x) = |x|^{\alpha/\beta}(\kappa_2(t_1),\dots,\kappa_2(t_d)).$$

Following [3, Lemma 6.6] the limit of the vector $(\overline{S}_n(t_1),\dots,\overline{S}_n(t_d))$ is determined by the Lévy measure

$$\nu_{t_1,\dots,t_d}(A) := \nu(\tau_{t_1,\dots,t_d}^{-1}(A)),$$

where $A \subseteq \mathbb{R}_-^d$ is a Borel set and ν is the Lévy measure of L. But $\nu_{t_1,\dots,t_d}$ is also the Lévy measure of the vector $(\kappa_2(t_1),\dots,\kappa_2(t_d))S$, where the random variable S has been introduced in Theorem 2.1(ii). This implies the desired result.

5.3 Finite Dimensional Convergence and Integral Functionals

Let $(Y^n)_{n\geq 1}$ and Y be stochastic processes index by $\mathbb{R}_+$ with paths in $\mathcal{L}^1(\mathbb{R}_+)$. We will give simple sufficient conditions for when the implication

$$Y^n \xrightarrow{\mathcal{L}\text{-f}} Y \quad \implies \quad \int_0^\infty Y_u^n\,\mathrm{d}u \xrightarrow{\mathcal{L}} \int_0^\infty Y_u\,\mathrm{d}u \tag{5.10}$$

holds true. Such a result is obviously required to obtain Theorem 3.1 from Theorem 2.1. Before we state these conditions we remark that the question has already been studied in the literature. As an example Theorem 22 in Appendix I of [14] gives two sufficient conditions for (5.10) to hold, but the second condition is a Hölder type criteria, which is not easily verifiable in our setting. Moreover, the theorem only deals with integration over bounded sets. The article [10] studies this question in general, but the conditions of e.g. Lemma 1 therein are too abstract even though we are in the case of a finite measure (the one induced by the weight function w). What can be deduced from [10] is that some kind of uniform integrability (with respect to the product measure) is sufficient for (5.10).

To formulate the lemma, define for each $n,m,l \in \mathbb{N}$ the intermediate random variables

$$X_{n,m,l} = \int_0^l Y^n_{\lfloor um \rfloor/m}\,\mathrm{d}u \qquad \text{and} \qquad X_{n,l} = \int_0^l Y^n_u\,\mathrm{d}u.$$

Proposition 5.6. *Suppose that $(Y^n)_{n\geq 1}$ and Y are continuous stochastic processes and assume that the following conditions hold:*

$$\lim_{l\to\infty}\limsup_{n\to\infty}\int_l^\infty \mathbb{E}[|Y^n_u|]\,\mathrm{d}u = 0, \qquad \lim_{m\to\infty}\limsup_{n\to\infty}\mathbb{P}(|X_{n,m,l}-X_{n,l}|\geq\varepsilon) = 0, \tag{5.11}$$

for all $l,\varepsilon > 0$. Then (5.10) *holds.*

Proof. Note that for each $n,m,l \in \mathbb{N}$ we have the decomposition

$$\int_0^\infty Y^n_u\,\mathrm{d}u = X_{n,m,l} + (X_{n,l} - X_{n,m,l}) + \int_l^\infty Y^n_u\,\mathrm{d}u.$$

As $Y^n \xrightarrow{\mathcal{L}\text{-f}} Y$ we deduce the weak convergence

$$X_{n,m,l} \xrightarrow[n\to\infty]{\mathcal{L}} Y_{m,l} := \int_0^l Y_{\lfloor um \rfloor/m}\,\mathrm{d}u \qquad \text{for each } m \in \mathbb{N}.$$

The continuity of Y implies immediately that $Y_{m,l} \xrightarrow{\text{a.s.}} \int_0^l Y_u\,\mathrm{d}u$ as $m \to \infty$. The assumptions in (5.11) then imply the convergence $\int_0^\infty Y^n_u\,\mathrm{d}u \xrightarrow{\mathcal{L}} \int_0^\infty Y_u\,\mathrm{d}u$.

□

5.4 Proof of Theorem 3.1(i) and (ii)

The strong consistency result of Theorem 3.1(i) is an immediate consequence of (3.1) and $\|\varphi_n - \varphi_{\xi_0}\|_{w,2} \xrightarrow{\text{a.s.}} 0$, where the latter follows from (2.2) and the dominated convergence theorem. Hence, we are left to proving Theorem 3.1(ii).

Recall the definition of the function $F : \mathcal{L}^2_w(\mathbb{R}_+)\times(0,1)\times\Theta_0 \to \mathbb{R}$ at (3.4), where $\Theta_0 \subseteq (0,\infty)\times(0,2)$ is an open neighbourhood of (σ_0,α_0) bounded away from $(0,0)$. Now, the minimal contrast estimator at (3.5) can be obtained using the criteria

$$\nabla_\theta F(\psi,H,\theta) = 0,$$

which is satisfied at (φ_{ξ_0},ξ_0). Denote by $\zeta(\psi,H)$ an element of $(0,1)\times\Theta_0$ such that

$$\nabla_\theta F(\psi,H,\zeta(\psi,H)) = 0.$$

To determine the derivative of ζ we will need the infinite dimensional implicit function theorem, which we briefly repeat.

Consider three Banach spaces $(E_i, \|\cdot\|_i)$, $i = 1,2,3$, and open subsets $U_i \subseteq E_i$, $i = 1,2$. Let $f : U_1 \times U_2 \to E_3$ be a Fréchet differentiable map. For a point $(e_1, e_2) \in U_1 \times U_2$ and a direction $(h_1, h_2) \in E_1 \times E_2$ we denote by $D^k_{e_1,e_2} f(h_k)$, $k = 1,2$, the Fréchet derivative of f at the point (e_1, e_2) in the direction $h_k \in E_k$. Assume that $(e_1^0, e_2^0) \in U_1 \times U_2$ satisfies the equation $f(e_1^0, e_2^0) = 0$ and that the map $D^2_{e_1^0,e_2^0} f : E_2 \to E_3$ is continuous and invertible. Then there exists open sets $V_1 \subseteq U_1$ and $V_2 \subseteq U_2$ with $(e_1^0, e_2^0) \in V_1 \times V_2$ and a bijective function $G : V_1 \to V_2$ such that

$$f(e_1, e_2) = 0 \quad \Longleftrightarrow \quad G(e_1) = e_2.$$

Moreover, G is Fréchet differentiable with derivative

$$D_{e_1} G(h) = -(D^2_{e_1,G(e_1)} f)^{-1}(D^1_{e_1,G(e_1)} f(h)). \tag{5.12}$$

We will adapt this to our setup, which corresponds to $U_1 = \mathcal{L}^r_w(\mathbb{R}_+) \times (0,1)$, for a $r > 1$, $E_2 = \Theta_0$, $E_3 = \mathbb{R}$ and $f = \nabla_\theta F$. A straightforward calculation shows that the map $\theta \mapsto \nabla^2_\theta F(\varphi, H, \theta)$ is differentiable with derivative at (φ_{ξ_0}, ξ_0) represented by the Hessian

$$D^2_{\varphi_{\xi_0},\xi_0} \nabla_\theta F = \nabla^2_\theta F(\varphi_{\xi_0}, \xi_0) = 2\left(\int_0^\infty \partial_{\theta_i}\varphi_{\xi_0}(t)\partial_{\theta_j}\varphi_{\xi_0}(t) w(t)\,\mathrm{d}t\right)_{i,j=1,2}.$$

The linear independence of the maps $\partial_{\theta_1}\varphi_{\xi_0}$ and $\partial_{\theta_2}\varphi_{\xi_0}$ immediately shows the invertibility of the Hessian. Moreover, standard theory for convergence in $\mathcal{L}^r(\mathbb{R}_+)$, $r > 1$, shows that the map $(\varphi, H) \mapsto \nabla^2_\theta F(\varphi, H, \cdot)$ is continuous, which is needed to assert that $\nabla_\theta F$ is C^1.

The determination of the remaining derivative $D^1_{\varphi,\xi}\nabla_\theta F$ for a point $(\varphi, \xi) \in \mathcal{L}^r_w(\mathbb{R}_+) \times (0,1) \times \Theta_0$ is slightly more involved. It is given by its two components $D^1 = (D^{1,1}, D^{1,2})$ corresponding to the partial derivatives. Indeed, $D^{1,1}_{\varphi,\xi}\nabla_\theta F$ is the derivative with respect to the functional coordinate $\varphi \in \mathcal{L}^r_w(\mathbb{R}_+)$ and $D^{1,2}_{\varphi,\xi}\nabla_\theta F$ the derivative with respect to the Hurst parameter $H \in (0,1)$, where $\xi = (H, \alpha, \sigma)$. It is easily seen that

$$D^{1,1}_{\varphi,\xi}\nabla_\theta F(h) = D^{1,1}_\xi \nabla_\theta F(h) = -2\int_0^\infty h(t)\nabla_\theta\varphi_\xi(t) w(t)\,\mathrm{d}t, \qquad h \in \mathcal{L}^r_w(\mathbb{R}_+).$$

An application of Hölder's inequality proves the continuity of the linear map $\xi \mapsto D^{1,1}_\xi \nabla_\theta F$. The second partial derivative at (φ, ξ) is the linear map represented by the two dimensional vector

$$D^{1,2}_{\varphi,\xi}\nabla_\theta F = 2\int_0^\infty \partial_H\varphi_\xi(t)\nabla_\theta\varphi_\xi(t) w(t)\,\mathrm{d}t - 2\int_0^\infty (\varphi(t) - \varphi_\xi(t))\partial_H\nabla_\theta\varphi_\xi(t) w(t)\,\mathrm{d}t.$$

Evaluated at the point $(\varphi_{\xi_0}, \xi_0) = (\varphi_{\xi_0}, G(\varphi_{\xi_0}, H_0))$ yields the simpler expression:

$$D^{1,2}_{\varphi_{\xi_0},\xi_0}\nabla_\theta F = 2\int_0^\infty \partial_H\varphi_{\xi_0}(t)\nabla_\theta\varphi_{\xi_0}(t) w(t)\,\mathrm{d}t.$$

Suppose we are in the case $k > H + 1/\alpha$ then we may pick $r = 2$ in the discussion above. By Fréchet differentiability it follows that

$$\begin{aligned} \sqrt{n}(\xi_n - \xi_0) &= \sqrt{n}\Big(G(\varphi_n, H_n) - G(\varphi_{\xi_0}, H_0)\Big) \\ &= D_{\varphi_{\xi_0},\xi_0} G\Big(\sqrt{n}(\varphi_n - \varphi_{\xi_0}), \sqrt{n}(H_n - H_0)\Big) \\ &\quad + \sqrt{n}\Big(\|\varphi_n - \varphi_{\xi_0}\|_{w,2} + |H_n - H_0|\Big) R(\varphi_n - \varphi_{\xi_0}, H_n - H_0), \end{aligned} \tag{5.13}$$

where the remainder term R satisfies that $R(\varphi_n-\varphi_{\xi_0},H_n-H_0)\xrightarrow{\text{a.s.}}0$ as $\|\varphi_n-\varphi_{\xi_0}\|_{w,2}+|H_n-H_0|\xrightarrow{\text{a.s.}}0$. Recall now the derivative of G at (5.12). In order to show Theorem 3.1(ii) it suffices to prove the convergences

$$\begin{aligned}\sqrt{n}(\|\varphi_n-\varphi_{\xi_0}\|_{w,2}+|H_n-H_0|)&\xrightarrow{\mathcal{L}}\|W\|_{w,2}+|M_1|,\\ \sqrt{n}\int_0^\infty(\varphi_n(t)-\varphi_{\xi_0}(t))\nabla_\theta\varphi_{\xi_0}(t)w(t)\,\mathrm{d}t&\xrightarrow{\mathcal{L}}\int_0^\infty W_t\nabla_\theta\varphi_{\xi_0}(t)w(t)\,\mathrm{d}t,\end{aligned}\tag{5.14}$$

where $W=(W_t)_{t\geq 0}$ has been introduced in Theorem 2.1(i).

We will only consider the second convergence at (5.14) since the first is shown similarly (see also [17, page 14]). For the conditions (5.11) it suffices to find a constant $C>0$ such that

$$\sup_{n\in\mathbb{N},t\geq 0}\operatorname{Var}(W_n^2(t))\leq C<\infty.\tag{5.15}$$

The identity $\Delta_{i,k}X=\int_{\mathbb{R}}h_k(i-s)\,\mathrm{d}L_s$ together with stationarity of the increments $\{\Delta_{i,k}X\mid i\geq k\}$ shows that

$$|\operatorname{Cov}(W_n^2(s),W_n^2(t))|\leq\frac{1}{2}\sum_{l\in\mathbb{Z}}|U_{h_k,h_k(l+\cdot)}(s,t)+U_{h_k,-h_k(l+\cdot)}(s,t)|.\tag{5.16}$$

Indeed, split the series at (5.16) into three terms. For $l=0$ it follows from (2.4) that

$$U_{h_k,h_k}(t,t)+U_{h_k,-h_k}(t,t)=1+2\exp(-|2t\sigma\|h_k\|_\alpha|^\alpha)-2\exp(-2|\sigma t\|h_k\|_\alpha|^\alpha),\tag{5.17}$$

which is obviously uniformly bounded in $t\geq 0$. For $l\neq 0$ with $|l|\leq k$ the first inequality of Lemma 5.1 implies that

$$\begin{aligned}\sum_{l\in\mathbb{Z}:|l|\leq k}|U_{h_k,h_k(l+\cdot)}(t,t)+U_{h_k,-h_k(l+\cdot)}(t,t)|&\leq 2t^\alpha\sum_{l\in\mathbb{Z}:|l|\leq k}\rho_l\exp\big(-2t^\alpha(\|h_k\|_\alpha^\alpha-\rho_l)\big)\\&\leq Ct^\alpha\exp\big(-2t^\alpha(\|h_k\|_\alpha^\alpha-\max_{|l|\leq k}\rho_l)\big).\end{aligned}\tag{5.18}$$

Now by Cauchy–Schwarz inequality $\rho_l<\|h_k\|_\alpha^\alpha$ for all l, and we obtain a uniform bound in $t\geq 0$. By Lemmas 5.1 and 5.2 there exist constants $C,K>0$ such that

$$\begin{aligned}&\sum_{l\in\mathbb{Z}:|l|>k}|U_{h_k,h_k(l+\cdot)}(t,t)+U_{h_k,-h_k(l+\cdot)}(t,t)|\\&\quad\leq 2^\alpha t^\alpha\exp\Big(-2t^\alpha(\|h_k\|_\alpha^\alpha-\sup_{|l|>k}\rho_l)\Big)\sum_{l\in\mathbb{Z}:|l|>k}|l|^{(\alpha(H-k)-1)/2}\\&\quad\leq Ct^\alpha\exp(-Kt^\alpha),\end{aligned}\tag{5.19}$$

where we used the assumption $k>H+1/\alpha$ and that $\rho_l\to 0$ for $|l|\to\infty$ by Lemma 5.2. Combining (5.17), (5.18) and (5.19) we can conclude (5.15). This completes the proof of Theorem 3.1(ii).

5.5 Proof of Theorem 3.1(iii)

As in the proof of Theorem 3.1(ii) we obtain the decomposition (5.13), where the convergence rate $\sqrt{n}$ is replaced by $n^{1-1/\beta}$. Furthermore, as in (5.14), it suffices to prove that for some $r\in(1,2)$ then as $n\to\infty$

$$\begin{aligned}n^{1-1/\beta}\big(\|\varphi_n-\varphi_{\xi_0}\|_{w,r}+|H_n-H_0|\big)&\xrightarrow{\mathcal{L}}\|\kappa_2S\|_{w,r}+|M_2|,\\ n^{1-1/\beta}\int_0^\infty\big(\varphi_n(t)-\varphi_{\xi_0}(t)\big)\nabla_\theta\varphi_{\xi_0}(t)w(t)\,\mathrm{d}t&\xrightarrow{\mathcal{L}}S\int_0^\infty\kappa_2(t)\nabla_\theta\varphi_{\xi_0}(t)w(t)\,\mathrm{d}t.\end{aligned}$$

As before in the Gaussian case it is enough to provide uniform bounds (in n and t) on the moments in order to use Proposition 5.6.

Recall that the dominating term in (5.6) is given by

$$\overline{S}_n(t) = n^{-1/\beta} \sum_{i=k}^{n} \big(\overline{\Phi}_t(L_{i+1} - L_i) - \mathbb{E}[\overline{\Phi}_t(L_{i+1} - L_i)]\big).$$

Inspired by the classical case of i.i.d. random variables, each in the domain of attraction of a stable distribution, we shall prove the following result.

Proposition 5.7. *For any $r \in (0, \beta)$ we have that*

$$\sup_{n\in\mathbb{N}, t\geq 0} \mathbb{E}[|\overline{S}_n(t)|^r] < \infty.$$

Proof. By Jensen's inequality it suffices to consider $r > 1$ (indeed $\beta \in (1,2)$). Recall the relation $\Phi_{t,\sigma\|h_k\|_\alpha}(x) = \Phi_t^2(x) = \exp(-|\sigma t \|h_k\|_\alpha|^\alpha)(\cos(tx) - 1)$ together with

$$\overline{\Phi}_t(x) = \sum_{i=1}^{\infty} \Phi_t^2(h_k(i)x).$$

Note that for all x in some bounded set by Lemma 5.3(ii):

$$\begin{aligned}
\sup_{t\geq 0} |\overline{\Phi}_t(x)| &\leq \sup_{t\geq 0} \exp(-|\sigma \|h_k\|_\alpha t|^\alpha) \sum_{i=1}^{\infty} (1 \wedge (|x t h_k(i)|)^2) \\
&\leq \sup_{t\geq 0} \exp(-|\sigma \|h_k\|_\alpha t|^\alpha)(t+1)^2(|x|+1)^2 \sum_{i=1}^{\infty} (1 \wedge |h_k(i)|^2) \leq C < \infty.
\end{aligned}$$

By (5.8) there exists $x_0 < -1$ such that for all $t > 0$

$$\left| \frac{|\overline{\Phi}_t(x)|}{|x|^q} - |\kappa_2(t)| \right| \leq 1 \qquad \text{for all } x < x_0, \tag{5.20}$$

where $q = \alpha/\beta$, $\beta = 1 + \alpha(k - H) \in (1,2)$ and $\kappa_2(t) = K t^q \exp(-|t\|h_k\|_\alpha \sigma|^\alpha)$ with $K < 0$. For shorter notation we write $D_i = L_{i+1} - L_i$ to denote the ith increment of L. Since $\mathbb{E}[\overline{\Phi}_t(D_1)]$ is bounded in t we may replace $\overline{\Phi}_t(x)$ with $\overline{\Phi}_t(x) - \mathbb{E}[\overline{\Phi}_t(D_1)]$ in (5.20) if x_0 is chosen large enough. Define for each $t \geq 0$, $n \in \mathbb{N}$ and $i \in \{k, \ldots, n\}$

$$\begin{aligned}
Y_{n,t,i} &= \big(\overline{\Phi}_t(D_i) - \mathbb{E}[\overline{\Phi}_t(D_i)]\big)\mathbb{1}_{\{|\overline{\Phi}_t(D_i) - \mathbb{E}[\overline{\Phi}_t(D_i)]| \leq n^{1/\beta}\}}, \\
Z_{n,t,i} &= \big(\overline{\Phi}_t(D_i) - \mathbb{E}[\overline{\Phi}_t(D_i)]\big)\mathbb{1}_{\{|\overline{\Phi}_t(D_i) - \mathbb{E}[\overline{\Phi}_t(D_i)]| > n^{1/\beta}\}}.
\end{aligned}$$

We have the decomposition

$$\begin{aligned}
T_{n,t} &:= \sum_{i=k}^{n} \big(\overline{\Phi}_t(D_i) - \mathbb{E}[\overline{\Phi}_t(D_i)]\big) = \sum_{i=k}^{n} (Y_{n,t,i} - \mathbb{E}[Y_{n,t,i}]) + \sum_{i=k}^{n} (Z_{n,t,i} - \mathbb{E}[Z_{n,t,i}]) \\
&=: T_{n,t,1} + T_{n,t,2}.
\end{aligned}$$

The proposition then asserts that

$$\sup_{n\in\mathbb{N}, t\geq 0} \mathbb{E}[|n^{-1/\beta} T_{n,t}|^r] < \infty \qquad \text{for all } r \in (1, \beta).$$

To prove this we observe that

$$\mathbb{E}[|n^{-1/\beta}T_{n,t}|^r] \le C_r\Big(\mathbb{E}[|n^{-1/\beta}T_{n,t,1}|^r] + \mathbb{E}[|n^{-1/\beta}T_{n,t,2}|^r]\Big).$$

For the first term we obtain the inequality

$$\mathbb{E}[|n^{-1/\beta}T_{n,t,1}|^r] \le \mathbb{E}\big[|n^{-1/\beta}T_{n,t,1}|^2\big]^{r/2} \le C_r\big(n^{1-2/\beta}\mathbb{E}[|Y_{n,t,k}|^2]\big)^{r/2}.$$

For short notation let $E_t = \mathbb{E}[\overline{\Phi}_t(D_1)]$, which is uniformly bounded in $t \ge 0$. Additionally let p_α denote the density of an SαS distribution and recall that $p_\alpha(x) \le C(1+|x|)^{-1-\alpha}$ for all $x \in \mathbb{R}$, cf. [29, Theorem 1.1]. We decompose $\mathbb{E}[|Y_{n,t,k}|^2]$ into two regions corresponding to (5.20):

$$\begin{aligned} n^{1-2/\beta}\mathbb{E}[|Y_{n,t,k}|^2] &= 2n^{1-2/\beta}\int_{x_0}^{0} |\overline{\Phi}_t(x) - E_t|^2 \mathbb{1}_{\{|\overline{\Phi}_t(x)-E_t| \le n^{1/\beta}\}} p_\alpha(x)\,\mathrm{d}x \\ &\quad + 2n^{1-2/\beta}\int_{-\infty}^{x_0} |\overline{\Phi}_t(x) - E_t|^2 \mathbb{1}_{\{|\overline{\Phi}_t(x)-E_t| \le n^{1/\beta}\}} p_\alpha(x)\,\mathrm{d}x. \end{aligned}$$

The first term vanishes as $n \to \infty$ since $2/\beta > 1$ and the fact that $|\overline{\Phi}_t(x) - E_t|$ is bounded for all t and $x \in (x_0, 0)$. The second term is further split into two terms:

$$\begin{aligned} &n^{1-2/\beta}\int_{-\infty}^{x_0} |\overline{\Phi}_t(x) - E_t|^2 \mathbb{1}_{\{|\overline{\Phi}_t(x)-E_t| \le n^{1/\beta}\}} p_\alpha(x)\,\mathrm{d}x \\ &\quad = n^{1-2/\beta}\int_{-\infty}^{x_0} |\overline{\Phi}_t(x) - E_t|^2 \mathbb{1}_{\{|x|^q < |\overline{\Phi}_t(x)-E_t| \le n^{1/\beta}\}} p_\alpha(x)\,\mathrm{d}x \\ &\quad\quad + n^{1-2/\beta}\int_{-\infty}^{x_0} |\overline{\Phi}_t(x) - E_t|^2 \mathbb{1}_{\{|\overline{\Phi}_t(x)-E_t| \le n^{1/\beta} \wedge |x|^q\}} p_\alpha(x)\,\mathrm{d}x. \end{aligned}$$

Using (5.20) and the boundedness of κ_2 on the first term we have that

$$\begin{aligned} &n^{1-2/\beta}\int_{-\infty}^{x_0} |\overline{\Phi}_t(x) - E_t|^2 \mathbb{1}_{\{|x|^q < |\overline{\Phi}_t(x)-E_t| \le n^{1/\beta}\}} p_\alpha(x)\,\mathrm{d}x \\ &\quad \le n^{1-2/\beta}\int_{-x_0}^{\infty} (|\kappa_2(t)|+1)^2 x^{2q} \mathbb{1}_{\{x^q \le n^{1/\beta}\}} p_\alpha(x)\,\mathrm{d}x \\ &\quad \le C_q n^{1-2/\beta}\int_{-x_0}^{n^{1/q\beta}} x^{2q-1-\alpha}\,\mathrm{d}x = C_q n^{1-2/\beta}(1 + n^{(2q-\alpha)/q\beta}) \le C_q. \end{aligned}$$

The second term contains a similar consideration, indeed

$$\begin{aligned} n^{1-2/\beta}\int_{-\infty}^{x_0} |\overline{\Phi}_t(x) - E_t|^2 \mathbb{1}_{\{|\overline{\Phi}_t(x)-E_t| \le n^{1/\beta} \wedge |x|^q\}} p_\alpha(x)\,\mathrm{d}x &\le \int_{-x_0}^{\infty} (n^{1/\beta} \wedge x^q)^2 p_\alpha(x)\,\mathrm{d}x \\ = n^{1-2/\beta}\int_{-x_0}^{n^{1/q\beta}} x^{2q} p_\alpha(x)\,\mathrm{d}x + n^{1-2/\beta}\int_{n^{1/q\beta}}^{\infty} n^{2/\beta} p_\alpha(x)\,\mathrm{d}x &\le C_q. \end{aligned}$$

In the next step we treat the term $T_{n,t,2}$. Note first that by the von Bahr-Esseen inequality we obtain

$$\mathbb{E}[|n^{-1/\beta}T_{n,t,2}|^r] \le C_r n^{1-r/\beta}\mathbb{E}[|Z_{n,t,k}|^r].$$

Decomposing as above we have that

$$\begin{aligned} n^{1-r/\beta}\mathbb{E}[|Z_{n,t,k}|^r] &= 2n^{1-r/\beta}\int_{x_0}^{0} |\overline{\Phi}_t(x) - E_t|^r \mathbb{1}_{\{|\overline{\Phi}_t(x)-E_t| > n^{1/\beta}\}} p_\alpha(x)\,\mathrm{d}x \\ &\quad + 2n^{1-r/\beta}\int_{-\infty}^{x_0} |\overline{\Phi}_t(x) - E_t|^r \mathbb{1}_{\{|\overline{\Phi}_t(x)-E_t| > n^{1/\beta}\}} p_\alpha(x)\,\mathrm{d}x. \end{aligned}$$

For the first term we recall that $\overline{\Phi}_t(x)$ is bounded uniformly in t when x lies in a bounded set, hence $n^{1/\beta} > |\overline{\Phi}_t(x) - E_t|$ for all sufficiently large n, independent of $x \in (0, x_0)$ and $t \geq 0$, so the first term is zero for sufficiently large n.

The last term requires more computations. Due to (5.20) and the fact that κ_2 is bounded it follows that

$$
\begin{aligned}
&n^{1-r/\beta} \int_{-\infty}^{x_0} |\overline{\Phi}_t(x) - E_t|^r \mathbb{1}_{\{|\overline{\Phi}_t(x) - E_t| > n^{1/\beta}\}} p_\alpha(x)\,\mathrm{d}x \\
&\quad \leq C n^{1-r/\beta} \int_{-\infty}^{x_0} (|\kappa_2(t)| + 1)^r\, |x|^{rq} \mathbb{1}_{\{(|\kappa_2(t)|+1)\,|x|^q > n^{1/\beta}\}} p_\alpha(x)\,\mathrm{d}x \\
&\quad \leq C n^{1-r/\beta} \int_{-\infty}^{x_0} |x|^{rq-1-\alpha} \mathbb{1}_{\{|x| > n^{1/q\beta}/K\}}\,\mathrm{d}x \\
&\quad = C n^{1-r/\beta} \int_{n^{1/\alpha}/K}^{\infty} x^{rq-1-\alpha}\,\mathrm{d}x \leq C,
\end{aligned}
$$

where we used that $rq - \alpha < 0$ since $r < \beta$. □

Combining Propositions 5.4–5.7 we finally complete the proof of Theorem 3.1(iii).

6 Tables

Table B.5. Absolute value of the bias based on $n = 1000$, $k = 2$, $p = -0.4$, $\nu = 0.1$ and $\sigma_0 = 0.3$ for the minimal contrast estimator.

H_0	α_0	σ_n	α_n	$H_n(p,k)$
0.2	**0.4**	6.1153	0.4078	0.2015
	0.6	0.0998	0.0076	0.1818
	0.8	0.0417	0.0018	0.1346
	1	0.0540	0.0000	0.1137
	1.2	0.0451	0.0008	0.0881
	1.4	0.0367	0.0045	0.0702
	1.6	0.0309	0.0207	0.0706
	1.8	0.0117	0.0535	0.0643
0.4	**0.4**	5.4236	0.3281	0.1040
	0.6	0.0774	0.0035	0.1237
	0.8	0.0472	0.0051	0.0748
	1	0.0322	0.0003	0.0501
	1.2	0.0276	0.0010	0.0351
	1.4	0.0115	0.0090	0.0189
	1.6	0.0072	0.0487	0.0261
	1.8	0.0496	0.1229	0.0193
0.6	**0.4**	3.6173	0.2007	0.0638
	0.6	0.0493	0.0058	0.0608
	0.8	0.0375	0.0036	0.0402
	1	0.0328	0.0020	0.0235
	1.2	0.0115	0.0161	0.0187
	1.4	0.0067	0.0327	0.0065
	1.6	0.0366	0.0933	0.0112
	1.8	0.1129	0.2501	0.0089
0.8	**0.4**	1.6906	0.0866	0.0204
	0.6	0.0483	0.0027	0.0404
	0.8	0.0514	0.0025	0.0303
	1	0.0345	0.0027	0.0167
	1.2	0.0053	0.0215	0.0107
	1.4	0.0160	0.0511	0.0057
	1.6	0.0702	0.1425	0.0014
	1.8	0.1724	0.3915	0.0016

Table B.6. Standard deviation based on $n = 1000$, $k = 2$, $p = -0.4$, $\nu = 0.1$ and $\sigma_0 = 0.3$ for the minimal contrast estimator.

H_0	α_0	σ_n	α_n	$H_n(p,k)$
0.2	**0.4**	7.4104	0.4585	0.1609
	0.6	0.4289	0.0925	0.1622
	0.8	0.2445	0.0681	0.1468
	1	0.2047	0.0773	0.1355
	1.2	0.1831	0.0923	0.1283
	1.4	0.1733	0.1118	0.1244
	1.6	0.1551	0.1376	0.1216
	1.8	0.1415	0.1461	0.1210
0.4	**0.4**	7.9313	0.4407	0.1664
	0.6	0.3619	0.0717	0.1723
	0.8	0.2295	0.0686	0.1546
	1	0.1864	0.0869	0.1503
	1.2	0.1698	0.1151	0.1448
	1.4	0.1590	0.1583	0.1384
	1.6	0.1527	0.1964	0.1447
	1.8	0.1255	0.1959	0.1397
0.6	**0.4**	7.3589	0.3497	0.1923
	0.6	0.2561	0.0652	0.1729
	0.8	0.1948	0.0676	0.1579
	1	0.1817	0.1030	0.1538
	1.2	0.1683	0.1411	0.1406
	1.4	0.1651	0.1974	0.1498
	1.6	0.1559	0.2431	0.1465
	1.8	0.1171	0.2271	0.1410
0.8	**0.4**	5.0779	0.2183	0.1848
	0.6	0.2539	0.0632	0.1758
	0.8	0.2014	0.0748	0.1523
	1	0.1864	0.1140	0.1546
	1.2	0.1843	0.1765	0.1533
	1.4	0.1749	0.2353	0.1424
	1.6	0.1402	0.2413	0.1417
	1.8	0.1156	0.2342	0.1361

Table B.7. Absolute value of the bias based on $n = 10\,000$, $k = 2$, $p = -0.4$, $\nu = 0.1$ and $\sigma_0 = 0.3$ for the minimal contrast estimator.

H_0	α_0	σ_n	α_n	$H_n(p,k)$
0.2	**0.4**	2.2799	0.1415	0.2563
	0.6	0.0106	0.0018	0.1652
	0.8	0.0032	0.0025	0.1134
	1	0.0046	0.0005	0.0848
	1.2	0.0078	0.0012	0.0636
	1.4	0.0110	0.0015	0.0516
	1.6	0.0107	0.0012	0.0413
	1.8	0.0077	0.0010	0.0318
0.4	**0.4**	0.6629	0.0458	0.1766
	0.6	0.0081	0.0003	0.1022
	0.8	0.0050	0.0017	0.0600
	1	0.0018	0.0010	0.0366
	1.2	0.0052	0.0018	0.0234
	1.4	0.0068	0.0019	0.0170
	1.6	0.0097	0.0065	0.0103
	1.8	0.0124	0.0145	0.0041
0.6	**0.4**	0.1517	0.0171	0.1130
	0.6	0.0021	0.0010	0.0588
	0.8	0.0067	0.0015	0.0314
	1	0.0057	0.0011	0.0156
	1.2	0.0025	0.0008	0.0070
	1.4	0.0032	0.0010	0.0034
	1.6	0.0081	0.0050	0.0008
	1.8	0.0143	0.0126	0.0017
0.8	**0.4**	0.0575	0.0060	0.0793
	0.6	0.0053	0.0013	0.0335
	0.8	0.0033	0.0002	0.0131
	1	0.0049	0.0004	0.0065
	1.2	0.0000	0.0006	0.0014
	1.4	0.0030	0.0014	0.0004
	1.6	0.0095	0.0045	0.0016
	1.8	0.0039	0.0035	0.0002

Table B.8. Standard deviation based on $n = 10\,000$, $k = 2$, $p = -0.4$, $\nu = 0.1$ and $\sigma_0 = 0.3$ for the minimal contrast estimator.

H_0	α_0	σ_n	α_n	$H_n(p,k)$
0.2	**0.4**	4.4680	0.2759	0.0693
	0.6	0.0983	0.0292	0.0560
	0.8	0.1062	0.0313	0.0510
	1	0.0642	0.0236	0.0499
	1.2	0.0684	0.0354	0.0484
	1.4	0.0699	0.0506	0.0489
	1.6	0.0711	0.0666	0.0479
	1.8	0.0654	0.0678	0.0476
0.4	**0.4**	2.2370	0.1458	0.0704
	0.6	0.0826	0.0230	0.0628
	0.8	0.0933	0.0305	0.0509
	1	0.0577	0.0309	0.0504
	1.2	0.0574	0.0393	0.0494
	1.4	0.0647	0.0620	0.0465
	1.6	0.0812	0.0959	0.0484
	1.8	0.0825	0.1168	0.0479
0.6	**0.4**	0.6954	0.0689	0.0713
	0.6	0.1014	0.0226	0.0576
	0.8	0.0756	0.0290	0.0506
	1	0.0589	0.0327	0.0481
	1.2	0.0546	0.0446	0.0480
	1.4	0.0710	0.0749	0.0479
	1.6	0.0955	0.1238	0.0462
	1.8	0.0957	0.1469	0.0469
0.8	**0.4**	0.3571	0.0510	0.0749
	0.6	0.1290	0.0256	0.0596
	0.8	0.0524	0.0245	0.0515
	1	0.0645	0.0388	0.0497
	1.2	0.0670	0.0593	0.0475
	1.4	0.0845	0.0951	0.0470
	1.6	0.1119	0.1512	0.0482
	1.8	0.0987	0.1621	0.0508

Table B.9. Absolute value of bias based on $n = 10\,000$, $k = 1$, $p = -0.4$, $\nu = 0.1$ and $\sigma_0 = 0.3$.

H_0	α_0	σ_n	α_n	$H_n(p,k)$
0.2	**0.4**	0.0202	0.0177	0.1873
	0.6	0.0671	0.0001	0.1329
	0.8	0.0256	0.0015	0.0966
	1	0.0044	0.0028	0.0745
	1.2	0.0026	0.0018	0.0589
	1.4	0.0097	0.0065	0.0465
	1.6	0.0161	0.0135	0.0376
	1.8	0.0216	0.0270	0.0286
0.4	**0.4**	0.0656	0.0090	0.0824
	0.6	0.0623	0.0017	0.0564
	0.8	0.0201	0.0020	0.0394
	1	0.0017	0.0037	0.0278
	1.2	0.0076	0.0052	0.0202
	1.4	0.0109	0.0089	0.0128
	1.6	0.0229	0.0271	0.0066
	1.8	0.0177	0.0288	0.0048
0.6	**0.4**	0.0659	0.0110	0.0109
	0.6	0.0724	0.0018	0.0114
	0.8	0.0254	0.0022	0.0025
	1	0.0050	0.0101	0.0009
	1.2	0.0131	0.0127	0.0002
	1.4	0.0189	0.0194	0.0027
	1.6	0.0083	0.0167	0.0004
	1.8	0.0241	0.0456	0.0014
0.8	**0.4**	0.0813	0.0116	0.1184
	0.6	0.0937	0.0039	0.0825
	0.8	0.0343	0.0117	0.0499
	1	0.0033	0.0194	0.0258
	1.2	0.0035	0.0120	0.0005
	1.4	0.0012	0.0197	0.0034
	1.6	0.0162	0.0439	0.0054
	1.8	0.0214	0.0571	0.0034

Table B.10. Standard deviation based for $n = 10\,000$, $k = 1$, $p = -0.4$, $\nu = 0.1$ and $\sigma_0 = 0.3$.

H_0	α_0	σ_n	α_n	$H_n(p,k)$
0.2	**0.4**	0.6538	0.0803	0.0655
	0.6	0.0681	0.0247	0.0548
	0.8	0.0599	0.0239	0.0501
	1	0.0693	0.0324	0.0474
	1.2	0.0633	0.0422	0.0459
	1.4	0.0711	0.0598	0.0461
	1.6	0.0835	0.0901	0.0458
	1.8	0.0841	0.1071	0.0437
0.4	**0.4**	0.3547	0.0625	0.0638
	0.6	0.0769	0.0271	0.0562
	0.8	0.0754	0.0355	0.0494
	1	0.0549	0.0409	0.0473
	1.2	0.0636	0.0560	0.0478
	1.4	0.0751	0.0777	0.0452
	1.6	0.1003	0.1242	0.0459
	1.8	0.0685	0.1096	0.0441
0.6	**0.4**	0.8005	0.0619	0.0637
	0.6	0.0668	0.0342	0.0544
	0.8	0.0722	0.0512	0.0521
	1	0.0695	0.0575	0.0494
	1.2	0.0772	0.0717	0.0453
	1.4	0.1065	0.1194	0.0458
	1.6	0.0777	0.1138	0.0441
	1.8	0.0648	0.1238	0.0449
0.8	**0.4**	0.7398	0.0648	0.0660
	0.6	0.1042	0.0448	0.0550
	0.8	0.1182	0.0728	0.0506
	1	0.0845	0.0830	0.0460
	1.2	0.0764	0.0747	0.0444
	1.4	0.0630	0.1042	0.0464
	1.6	0.0700	0.1390	0.0452
	1.8	0.0602	0.1412	0.0419

Table B.11. Absolute value of bias for $(\widetilde{\sigma}_{\text{low}}, \widetilde{\alpha}_{\text{low}})$ for $n = 10\,000$, $p = -0.4$ and $\sigma_0 = 0.3$.

H_0	α_0	$\widetilde{\sigma}_{\text{low}}$	$\widetilde{\alpha}_{\text{low}}$
0.2	0.4	0.2872	0.0371
	0.6	0.2860	0.1737
	0.8	0.2725	0.3817
	1	0.2390	0.4889
	1.2	0.0888	0.1818
	1.4	0.0065	0.0044
	1.6	0.0077	0.0080
	1.8	0.0047	0.0013
0.4	0.4	0.2805	0.0364
	0.6	0.2693	0.1606
	0.8	0.2053	0.2938
	1	0.0543	0.1374
	1.2	0.0052	0.0043
	1.4	0.0012	0.0011
	1.6	0.0005	0.0010
	1.8	0.0006	0.0007
0.6	0.4	0.2764	0.0460
	0.6	0.2285	0.1362
	0.8	0.1752	0.2583
	1	0.0077	0.0135
	1.2	0.0020	0.0023
	1.4	0.0001	0.0001
	1.6	0.0002	0.0005
	1.8	0.0006	0.0000
0.8	0.4	0.2688	0.0713
	0.6	0.1389	0.0917
	0.8	0.0096	0.0286
	1	0.0062	0.0073
	1.2	0.0013	0.0000
	1.4	0.0003	0.0023
	1.6	0.0001	0.0010
	1.8	0.0000	0.0020

Table B.12. Standard deviation for $(\widetilde{\sigma}_{\text{low}}, \widetilde{\alpha}_{\text{low}})$ for $n = 10\,000$, $p = -0.4$ and $\sigma_0 = 0.3$.

H_0	α_0	$\widetilde{\sigma}_{\text{low}}$	$\widetilde{\alpha}_{\text{low}}$
0.2	0.4	0.0514	0.2249
	0.6	0.0420	0.2195
	0.8	0.0676	0.2181
	1	0.1036	0.2549
	1.2	0.1322	0.2632
	1.4	0.1000	0.2301
	1.6	0.0510	0.1362
	1.8	0.0264	0.0765
0.4	0.4	0.0680	0.2127
	0.6	0.0741	0.2198
	0.8	0.1525	0.2459
	1	0.2039	0.3633
	1.2	0.0528	0.0875
	1.4	0.0236	0.0451
	1.6	0.0151	0.0308
	1.8	0.0112	0.0222
0.6	0.4	0.0734	0.2108
	0.6	0.1522	0.2142
	0.8	0.1790	0.2618
	1	0.1245	0.1896
	1.2	0.0259	0.0371
	1.4	0.0149	0.0283
	1.6	0.0106	0.0252
	1.8	0.0078	0.0204
0.8	0.4	0.0853	0.2041
	0.6	0.2856	0.2329
	0.8	0.2399	0.2664
	1	0.0659	0.0793
	1.2	0.0201	0.0306
	1.4	0.0118	0.0281
	1.6	0.0084	0.0255
	1.8	0.0064	0.0208

Table B.13. Failure rates for $n = 10\,000$, $p = -0.4$, $k = 2$, $\nu = 0.1$ and $\sigma_0 = 0.3$.

		Failure rate (%)	
H_0	α_0	$(\widetilde{\sigma}_{\text{low}}, \widetilde{\alpha}_{\text{low}}, \widetilde{H}_{\text{low}})$	ξ_n
0.2	0.4	88.17	14.83
	0.6	84.08	0
	0.8	69.67	0
	1	38.58	0
	1.2	0.42	0
	1.4	3.25	0.17
	1.6	3.83	1
	1.8	1.58	2.25
0.4	0.4	86.08	20.17
	0.6	76.75	0.17
	0.8	62.58	0
	1	19.17	0.18
	1.2	0	0
	1.4	0	0.33
	1.6	0	1.42
	1.8	0	7.17
0.6	0.4	87	28.67
	0.6	72.67	0
	0.8	41.75	0
	1	0.08	0
	1.2	0	0
	1.4	0	0.08
	1.6	0	2.42
	1.8	0	10.25
0.8	0.4	89.50	26.33
	0.6	69.33	0.33
	0.8	2.17	0
	1	0.08	0.25
	1.2	0	0
	1.4	0	0.17
	1.6	0	3.92
	1.8	0	11

References

[1] Antoine Ayachea and Julien Hamoniera (2012). »*Linear fractional stable motion: A wavelet estimator of the α parameter*«. Statistics and Probability Letters 82(8), 1569–1575. doi: 10.1016/j.spl.2012.04.005.

[2] Jean-Marc Bardet and Donatas Surgailis (2013). »*Nonparametric estimation of the local Hurst function of multifractional Gaussian processes*«. Stochastic Processes and Their Applications 123(3), 1004–1045. doi: 10.1016/j.spa.2012.11.009.

[3] Andreas Basse-O'Connor, Claudio Heinrich and Mark Podolskij (2019). »*On limit theory for functionals of stationary increments Lévy driven moving averages*«. Electronic Journal of Probability 24(79), 42 pp. doi: 10.1214/19-EJP336.

[4] Andreas Basse-O'Connor, Raphaël Lachièze-Rey and Mark Podolskij (2017). »*Power variation for a class of stationary increments Lévy driven moving averages*«. The Annals of Probability 45(6B), 4477–4528. doi: 10.1214/16-AOP1170.

[5] Albert Benassi, Serge Cohen and Jacques Istas (2004). »*On roughness indices for fractional fields*«. Bernoulli 10(2), 357–373. doi: 10.3150/bj/1082380223.

[6] Alexandre Brouste and Masaaki Fukasawa (2017). »*Local asymptotic normality for fractional Gaussian noise under high-frequency observations*«. The Annals of Statistics 46(5), 2045–2061. doi: 10.1214/17-AOS1611.

[7] Richard H. Byrd, Peihuang Lu, Jorge Nocedal and Ciyon Zhu (1995). »*A limited memory algorithm for bound constrained optimization*«. SIAM Journal of Scientific Computing 16(5), 1190–1208. doi: 10.1137/0916069.

[8] Stamatis Cambanis, Clyde D. Jr. Hardin and Aleksander Weron (1987). »*Ergodic properties of stationary stable processes*«. Stochastic Processes and Their Applications 24(1), 1–18. doi: 10.1016/0304-4149(87)90024-X.

[9] Jean-François Coeurjolly and Jacques Istas (2001). »*Cramèr-Rao bounds for fractional Brownian motions*«. Statistics and Probability Letters 53(4), 435–447. doi: 10.1016/S0167-7152(00)00197-8.

[10] Heinz Cremers and Dieter Kadelka (1986). »*On weak convergence of integral functionals of stochastic processes with applications to processes taking paths in L_p^e*«. Stochastic Processes and their Applications 21(2), 305–317. doi: 10.1016/0304-4149(86)90102-X.

[11] Rainer Dahlhaus (1989). »*Efficient parameter estimation for self-similar processes*«. The Annals of Statistics 17(4), 1749–1766. doi: 10.1214/aos/1176347393.

[12] Thi To Nhu Dang and Jacques Istas (2017). »*Estimation of the Hurst and the stability indices of a H-self-similar stable process*«. Electronic Journal of Statistics 11(2), 4103–4150. doi: 10.1214/17-EJS1357.

[13] Danijel Grahovac, Nikolai N. Leonenko and Murad S. Taqqu (2015). »*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. doi: 10.1007/s10955-014-1126-4.

[14] I. A. Ibragimov and R. Z. Has'minskii (1981). »*Statistical estimation: Asymptotic theory*«. Translated by S. Kotz. Stochastic Modelling and Applied Probability 16. Springer-Verlag New York. doi: 10.1007/978-1-4899-0027-2.

[15] Jacques Istas and Gabriel Lang (1997). »*Quadratic variations and estimation of the local Hölder index of a Gaussian process*«. Annales de l'I.H.P. Probabilités et statistiques 33(4), 407–436.

[16] Joachim Lebovits and Mark Podolskij (2017). »*Estimation of the global regularity of a multifractional Brownian motion*«. Electronic Journal of Statistics 11(1), 78–98. doi: 10.1214/16-EJS1221.

[17] Mathias Mørck Ljungdahl and Mark Podolskij (2019). »*A note on parametric estimation of Lévy moving average processes*«. Springer Proceedings in Mathematics & Statistics 294, 41–56. doi: 10.1007/978-3-030-28665-1_3.

[18] Benoit B. Mandelbrot and John W. Van Ness (1968). »*Fractional Brownian motions, fractional noises and applications*«. SIAM Review 10(4), 422–437. doi: 10.1137/1010093.

[19] Stepan Mazur, Dmitry Otryakhin and Mark Podolskij (2020). »*Estimation of the linear fractional stable motion*«. Bernoulli 26(1), 226–252. doi: 10.3150/19-BEJ1124.

[20] J. A. Nelder and R. Mead (1965). »*A simplex method for function minimization*«. The Computer Journal 7(4), 308–313. doi: 10.1093/comjnl/7.4.308.

[21] Vladas Pipiras and Murad S. Taqqu (2002). »*The structure of self-similar stable mixed moving averages*«. The Annals of Probability 30(2), 898–932. doi: 10.1214/aop/1023481011.

[22] Vladas Pipiras and Murad S. Taqqu (2003). »*Central limit theorems for partial sums of bounded functionals of infinite-variance moving averages*«. Bernoulli 9(5), 833–855. doi: 10.3150/bj/1066418880.

[23] Vladas Pipiras, Murad S. Taqqu and Patrice Abry (2007). »*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. doi: 10.3150/07-BEJ6143.

[24] Jan Rosiński (1995). »*On the structure of stationary stable processes*«. The Annals of Probability 23(3), 1163–1187. doi: 10.1214/aop/1176988178.

[25] N. M. Steen, G. D. Byrne and E. M. Gelbard (1969). »*Gaussian quadratures for the integrals* $\int_0^\infty \exp(-x^2)f(x)dx$ *and* $\int_0^b \exp(-x^2)f(x)dx$«. Mathematics of Computation 23(107), 661–671. doi: 10.2307/2004395.

[26] Maxwell B. Stinchcombe and Halbert White (1992). »*Some measurability results for extrema of random functions over random sets*«. The Review of Economic Studies 59(3), 495–514. doi: 10.2307/2297861.

[27] Stilian Stoev, Vladas Pipiras and Murad S. Taqqu (2002). »*Estimation of the self-similarity parameter in linear fractional stable motion*«. Signal Processing 82(12), 1873–1901. doi: 10.1016/S0165-1684(02)00317-1.

[28] Stilian Stoev and Murad S. Taqqu (2004). »*Simulation methods for linear fractional stable motion and FARIMA using the Fast Fourier Transform*«. Fractals 12(1), 95–121. doi: 10.1142/S0218348X04002379.

[29] T. Watanabe (2007). »*Asymptotic estimates of multi-dimensional stable densities and their applications*«. Transactions of the American Mathematical Society 358(6), 2851–2879.

Paper C

A Note on Parametric Estimation of Lévy Moving Average Processes

Mathias Mørck Ljungdahl

Department of Mathematics, Aarhus University

Mark Podolskij

Department of Mathematics, Aarhus University

Abstract. In this paper we present a new parametric estimation method for a Lévy moving average process driven by a symmetric α-stable Lévy motion L, $\alpha \in (0,2)$. More specifically, we consider a parametric family of kernel functions g_θ with $\theta \in \Theta \subseteq \mathbb{R}$ and propose an asymptotically normal estimator of the pair (α, θ). The estimation idea is based upon the minimal contrast approach, which compares the empirical characteristic function of the Lévy moving average process with its theoretical counterpart. Our work is related to recent papers [14, 16] that are studying parametric estimation of a linear fractional stable motion.

1 Introduction

During the past decades a lot of progress has been achieved in the probabilistic and statistical analysis of fractional type processes. Since the pioneering work of Mandelbrot and van Ness [15] the fractional Brownian motion (fBm) has received a great deal of attention. The (scaled) fBm is the unique zero-mean self-similar Gaussian process with stationary increments. A variety of statistical estimation methods for fBm has been developed in low and high frequency setting. We refer to, e.g. [9, 6, 8, 12] for the statistical analysis of parametric estimators in both frameworks. More recently, researchers started to investigate the mathematical properties of the *linear fractional stable motion*, which constitutes a particular extension of fBm that drops the Gaussianity assumption and allows for general α-stable marginal distributions. The asymptotic theory for statistics of linear fractional stable motions turns out to be much more complex as it has been shown in the papers [2, 3, 4, 13], which consider a larger class of stationary increments Lévy moving average processes. First parametric estimation methods have been proposed in [1, 5, 10, 11, 18]. However, the complete asymptotic theory has been investigated only in recent papers [14, 16].

In this article we extend the statistical analysis of [14] to parametric Lévy moving average processes. We consider the model

$$X_t = \int_{-\infty}^{t} g_\theta(t-s)\,\mathrm{d}L_s, \qquad t \geq 0, \tag{1.1}$$

where $g_\theta : [0,\infty) \to \mathbb{R}$ is a deterministic kernel function parametrized by $\theta \in \Theta$, where $\Theta \subseteq \mathbb{R}$ is an open set, and L is a symmetric α-stable Lévy motion with $\alpha \in (0,2)$ and scale parameter 1; we use the abbreviation $L \sim S\alpha S(1)$. We assume that

$$\|g_\theta\|_\alpha^\alpha := \int_{\mathbb{R}} |g_\theta(x)|^\alpha\,\mathrm{d}x < \infty \tag{1.2}$$

for all $\theta \in \Theta$ and $\alpha \in (0,2)$, which guarantees the existence of the integral in (1.1), cf. [19] (we extend g_θ to the whole real line by setting $g_\theta(x) = 0$ for all $x < 0$). The aim of our paper is to construct an asymptotically normal estimator for the pair (α,θ) given low frequency observations $X_1,\dots,X_n$ of the model (1.1). Our approach is based upon the (real part of) empirical characteristic function defined as

$$\phi_n(u) := \frac{1}{n}\sum_{i=1}^{n} \cos(uX_i), \qquad u \geq 0. \tag{1.3}$$

Since the process X is strongly ergodic (cf. [7]), we obtain the strong consistency result

$$\phi_n(u) \xrightarrow{\text{a.s.}} \phi_{\alpha,\theta}(u) := \mathbb{E}[\cos(uX_1)] = \exp(-u^\alpha \|g_\theta\|_\alpha^\alpha) \qquad \text{for all } u \geq 0. \tag{1.4}$$

From the latter convergence result it becomes obvious that the parameter $\theta \in \Theta$ is identifiable through the function $\phi_{\alpha,\theta}$ if and only if the map $\theta \mapsto \|g_\theta\|_\alpha$ is bijective, which we assume in the following. We remark that the characteristic function $\phi_{\alpha,\theta}$ only takes into account the *marginal* distribution of X. In the articles [14, 16], which investigate parametric estimation for the linear fractional stable motion, which is a three-parameter family, the empirical characteristic function ϕ_n is combined with other statistics to obtain an estimator for the whole set of parameters of the model.

The goal of this paper is to study the asymptotic properties of the minimal contrast estimator associated with the empirical characteristic function ϕ_n. More specifically, we consider a positive bounded C^1-function w with $w \in \mathcal{L}^1(\mathbb{R}_+)$ and introduce the estimator $(\widehat{\alpha}_n, \widehat{\theta}_n)$ of the unknown parameter (α_0,θ_0) via

$$(\widehat{\alpha}_n, \widehat{\theta}_n) = \underset{(\alpha,\theta)\in(0,2)\times\Theta}{\operatorname{argmin}} \|\phi_n - \phi_{\alpha,\theta}\|_{\mathcal{L}_w^2}^2 \qquad \text{with} \qquad \|f\|_{\mathcal{L}_w^2}^2 := \int_{\mathbb{R}_+} f^2(x)w(x)\,\mathrm{d}x. \tag{1.5}$$

We remark that a similar minimal contrast approach has been studied in [14] in the setting of the linear fractional stable noise, which corresponds to the kernel function $g(x) = \theta\{(x+1)^\beta - x^\beta\}\mathbb{1}_{\{x>0\}}$. Hence, the present work extends the concepts of [14] to more general parametric classes $\{g_\theta : \theta \in \Theta\}$ under the bijectivity condition on the map $\theta \mapsto \|g_\theta\|_\alpha$ (on the other hand, in contrast to the present work, [14] does not assume that the parameter β is known). We will show that the minimal contrast estimator $(\widehat{\alpha}_n, \widehat{\theta}_n)$ defined at (1.5) is strongly consistent and, under further assumptions on the model, asymptotically normal.

The paper is structured as follows. Section 2 presents the model assumption and a short review of relevant results. Section 3 demonstrates the asymptotic theory for the minimal contrast estimator $(\widehat{\alpha}_n, \widehat{\theta}_n)$ including the strong consistency and asymptotic normality. The proofs of the main statements are collected in Section 4.

2 Model Assumptions and Literature Review

2.1 Assumptions, Remarks and Examples

We start with a set of assumptions on the functions g_θ and $\phi_{\alpha,\theta}$. Below we write $\partial^k_{z_1,\ldots,z_k} f$ to denote the kth derivative of f_θ with respect to $z_1,\ldots,z_k \in \{\alpha,\theta\}$.

Assmuption (A). Additionally to condition (1.2) we assume that

$$\sum_{i\geq 1}\left(\int_{i-1}^{i} |g_\theta(x)|^\alpha \,\mathrm{d}x\right)^{1/2} < \infty \tag{2.1}$$

for all $\theta \in \Theta$ and $\alpha \in (0,2)$. Furthermore, we assume that the map $(\theta,\alpha) \mapsto \|g_\theta\|_\alpha^\alpha$ is C^2.

Condition (2.1) is required to prove the central limit theorem for the sequence of statistics $\sqrt{n}(\phi_n(u) - \phi_{\alpha,\theta}(u))$. As we will see below, the standardized version of $\phi_n(u)$ may have a different asymptotic distribution when (2.1) is violated. On the other hand, the differentiability condition of Assmuption (A) ensures that the map $(\theta,\alpha) \mapsto \phi_{\theta,\alpha}(u)$ is differentiable for all $u \in \mathbb{R}_+$ and that we may interchange differentiation and integration when computing the asymptotic covariance matrix associated with the estimator $(\widehat{\alpha}_n, \widehat{\theta}_n)$.

Example 2.2 (Conditions on the kernel function g).
Let us first consider the exponential type family of kernels defined by $g_\theta(x) = \exp(-\theta x)$ with $\theta \in \Theta = (0,\infty)$. Then condition (1.2) is obviously satisfied for any $\theta \in \Theta$ and $\alpha \in (0,2)$. We also have that

$$\int_{i-1}^{i} |g_\theta(x)|^\alpha \,\mathrm{d}x \leq \exp(-\theta\alpha(i-1))$$

and thus condition (2.1) is satisfied.

Now, let us study the kernel function $g_\theta(x) = \theta((x+1)^\beta - x^\beta)\mathbb{1}_{\{x>0\}}$, where $\beta \in \mathbb{R}$ is a known constant and $\theta \in \Theta = (0,\infty)$, which turns out to be more complicated to treat. As remarked earlier this kernel stems from the linear fractional stable noise. In this setting, recalling that we require integrability of $|g_\theta|^\alpha$ near 0, condition (1.2) is equivalent to the statement

$$\beta \in (-1/\alpha, -1/\alpha + 1)$$

and this can never be satisfied for all $\alpha \in (0,2)$. Assume now that we know the bounds $\underline{\alpha} < \overline{\alpha}$, $\underline{\alpha}, \overline{\alpha} \in (0,2)$, such that $\alpha \in (\underline{\alpha}, \overline{\alpha})$ and

$$\frac{\underline{\alpha}}{2-\underline{\alpha}} > \overline{\alpha}.$$

Clearly, the latter condition gives a restriction on the numbers $\underline{\alpha}$ and $\overline{\alpha}$. In this restrictive setting, the condition $\beta \in (-1/\alpha, -1/\alpha + 1)$ is satisfied for all $\alpha \in (\underline{\alpha}, \overline{\alpha})$ whenever $\beta \in (-1/\overline{\alpha}, -1/\underline{\alpha} + 1)$. Furthermore, we have the inequality

$$\int_{i-1}^{i} |g_\theta(x)|^\alpha \,\mathrm{d}x \leq C_\theta (i-1)^{(\beta-1)\alpha} \qquad \text{for } i > 1.$$

Hence, in the setting $\alpha \in (\underline{\alpha}, \overline{\alpha})$ condition (2.1) is satisfied when $\beta \in (-1/\overline{\alpha}, -2/\underline{\alpha} + 1)$. ○

Remark 2.3 (Higher order filters).
Let us further study the polynomial family of kernels $g_\theta(x) = \theta((x+1)^\beta - x^\beta)\mathbb{1}_{\{x>0\}}$, where β is known and $\alpha \in (\underline{\alpha}, \overline{\alpha}) \subseteq (0,2)$. Suppose that condition $\beta > -1/\overline{\alpha}$ is satisfied, but $\beta \geq -2/\underline{\alpha}+1$ and hence (2.1) does not hold. In this setting we may apply the higher order filter to the observations $X_1, \dots, X_n$ to solve the problem. For a $k \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ we introduce the operator

$$\Delta^k X_i := \sum_{j=0}^{k} (-1)^j \binom{k}{j} X_{i-j} \qquad \text{and} \qquad \Delta^k g_\theta(x) := \sum_{j=0}^{k} (-1)^j \binom{k}{j} g_\theta(x-j),$$

for any $i > k$ and $x > k$. Defining the statistic

$$\phi_n(u;k) := \frac{1}{n-k} \sum_{i=k+1}^{n} \cos(u \Delta^k X_i), \tag{2.2}$$

we deduce the strong consistency result

$$\phi_n(u;k) \xrightarrow{\text{a.s.}} \phi_{\alpha,\theta}(u;k) = \exp\big(-u^\alpha \|\Delta^k g_\theta\|_\alpha^\alpha\big) \qquad \text{for all } u \geq 0.$$

We conclude that condition (2.1) holds for $\Delta^k g_\theta$ when

$$\beta < k + 1 - 2/\underline{\alpha}.$$

Consequently, when β and $\underline{\alpha}$, $\overline{\alpha}$ are known, we may proceed as follows: Choose the minimal $k \in \mathbb{N}_0$ such that $\beta < k + 1 - 2/\underline{\alpha}$ holds and construct the estimator of the unknown parameter (α_0, θ_0) via the minimal contrast approach introduced at (1.5), where $\phi_n(u)$ (resp. $\phi_{\alpha,\theta}(u)$) is replaced by $\phi_n(u;k)$ (resp. $\phi_{\alpha,\theta}(u;k)$) and the minimization problem is restricted to the set $(\underline{\alpha}, \overline{\alpha}) \times \Theta$. This procedure, which can be easily extended to other classes of kernel functions with polynomial decay at infinity, would result in the same asymptotic theory as presented below (up to obvious replacement of g_θ by $\Delta^k g_\theta$ at relevant places). ◇

2.2 Some Probabilistic and Statistical Results

In this section we demonstrate some limit theorems and statistical methods, which are related to our statistical problem. We start with the multivariate central limit theorem for bounded functionals of Lévy moving average processes, which has been proved in [17]; see also [18] for further extensions.

Theorem 2.4 ([17, Theorem 2.1]).
Let $(X_t^j)_{t \geq 0}$, $j = 1, \dots, d$, be Lévy moving average processes of the form

$$X_t^j = \int_{-\infty}^{t} g_j(t-s)\, \mathrm{d}L_s,$$

where $L \sim S\alpha S(1)$ and the kernels g_j satisfy $\|g_j\|_\alpha < \infty$. Define the statistics

$$V(f_j)_n^j = \frac{1}{n} \sum_{i=1}^{n} f_j(X_i), \qquad j = 1, \dots, d,$$

where f_j are measurable bounded functions. If each kernel g_j satisfies the condition (2.1)*, we obtain the central limit theorem*

$$\sqrt{n}\Big(V(f_j)_n^j - \mathbb{E}[f_j(X_1^j)]\Big)_{1\le j\le d} \xrightarrow{\mathcal{L}} \mathcal{N}_d(0,\Sigma)$$

where $\Sigma \in \mathbb{R}^{d\times d}$ is a finite matrix defined via $\Sigma_{jk} = \sum_{r\in\mathbb{Z}} \mathrm{Cov}(f_j(X_1^j), f_k(X_{1+r}^k))$.

Theorem 2.4 directly applies to the multivariate statistic $\sqrt{n}(\phi_n(u_j) - \phi_{\alpha,\theta}(u_j))_{1\le j\le d}$ with $u_1,\dots,u_d \in \mathbb{R}_+$ by setting, e.g. $f_j(x) = \cos(u_j x)$ and $g_j = g_\theta$ for all $j = 1,\dots,d$, a result which will be useful later. The recent paper [2] gives a more complete probabilistic picture when the condition (2.1) is violated. Their results have been formulated in the high frequency regime, so they do not apply to the low frequency setting without further modification. However, in the case of the linear fractional stable motion the results directly translate to the low frequency regime. To demonstrate ideas we consider a linear fractional stable noise defined by

$$Z_t = \int_{-\infty}^{t} \theta\Big\{(t+1-s)^{H-1/\alpha} - (t-s)^{H-1/\alpha}\Big\}\mathrm{d}L_s, \tag{2.3}$$

where $H \in (0,1)$ denotes the self-similarity parameter, i.e. $(Z_{at})_{t\ge0} = (a^H Z_t)_{t\ge0}$ in distribution for any $a > 0$. For simplicity of exposition let us consider the empirical characteristic function $\phi_n(Z,u;k)$ defined at (2.2) associated with $\Delta^k Z_i$. We already know from Remark 2.3 and Theorem 2.4 that the standardized version of the statistic $\phi_n(Z,u;k)$ is asymptotically normal when $k+1 > H + 1/\alpha$. Now, we present the limit distribution in the setting $k+1 < H+1/\alpha$.

Theorem 2.5 ([16, Theorem 2.2] and [2, Theorem 2.6]).
Consider the linear fractional stable noise defined at (2.3)*. Let $k \in \mathbb{N}_0$ be such that $k+1 < H + 1/\alpha$. Then we obtain the convergence in distribution*

$$n^{1-1/(1+\alpha(k+1-H))}(\phi_n(Z,u;k) - \phi(Z,u;k)) \xrightarrow{\mathcal{L}} \mathcal{S}(1+\alpha(k+1-H), 0, \rho, \eta),$$

where $\mathcal{S}(1+\alpha(k+1-H),0,\rho,\eta)$ denotes the $(1+\alpha(k+1-H))$-stable distribution with location parameter 0, scale parameter ρ and skewness parameter η; we refer to [2, Theorem 2.6] for the explicit definition of ρ and η.

In contrast to our model (1.1), the linear fractional stable noise Z is a three-parameter family. This has several consequences for the statistical analysis. First of all, while the function g_θ has a more general form, the power $\beta = H - 1/\alpha$ is unknown in the linear fractional stable noise setting. This means that in the latter case it is not known whether the empirical characteristic function $\phi_n(Z,u;k)$ is in the domain of attraction of the normal or the stable distribution. The paper [16] suggests a statistical method to overcome this problem and to obtain a feasible limit theorem for the parameter (θ,α,H).

3 Main Results

We start this section with the central limit theorem for the quantity $\sqrt{n}(\phi_n(u) - \phi_{\alpha,\theta}(u))$. In the following we use the notation $Y^n \xrightarrow{\mathcal{L}\text{-f}} Y$ for stochastic processes $(Y_t^n)_{t\ge0}$ and $(Y_t)_{t\ge0}$ to denote the finite dimensional convergence $(Y_{t_1}^n,\dots,Y_{t_k}^n) \xrightarrow{\mathcal{L}} (Y_{t_1},\dots,Y_{t_k})$ for any $k\in\mathbb{N}$ and $t_i \in \mathbb{R}_+$.

Proposition 3.1. *Suppose that Assmuption (A) holds. Then we obtain the convergence*

$$\sqrt{n}(\phi_n(u) - \phi_{\alpha,\theta}(u)) \xrightarrow{\mathcal{L}\text{-f}} G_u, \tag{3.1}$$

where $(G_t)_{t\geq 0}$ is a zero mean Gaussian process with covariance kernel defined by

$$\begin{aligned}\mathbb{E}[G_u G_v] = \frac{1}{2}\sum_{l\in\mathbb{Z}}\Big(&\exp(-\|ug_\theta + vg_\theta(\,\cdot\,+l)\|_\alpha^\alpha) + \exp(-\|ug_\theta - vg_\theta(\cdot+l)\|_\alpha^\alpha) \\ &- 2\exp(-(u^\alpha + v^\alpha)\|g_\theta\|_\alpha^\alpha)\Big), \qquad u,v\in\mathbb{R}_+.\end{aligned} \tag{3.2}$$

In particular, there exists a constant $C>0$ such that

$$\mathbb{E}[G_u^2] \leq Cu^\alpha \exp(-u^\alpha/C), \qquad u\in\mathbb{R}_+. \tag{3.3}$$

Furthermore, if

$$\sum_{l\geq 1}\mu_l < \infty \qquad \text{where } \mu_l := \int_{\mathbb{R}_+} |g_\theta(x+l)|^\alpha \,\mathrm{d}x, \tag{3.4}$$

there exists a constant $C_T > 0$ such that

$$\mathbb{E}[(G_u - G_v)^2] \leq C_T|u-v|^{\alpha/2}, \qquad u,v\in[0,T]. \tag{3.5}$$

Remark 3.2. Note that condition (3.4) can be equivalently written as

$$\sum_{l\geq 1} l\int_l^{l+1} |g_\theta(x)|^\alpha \,\mathrm{d}x < \infty,$$

and thus it does not follow from (2.1) in general. However, the two conditions are equivalent in the context of exponential and polynomial kernels discussed in Example 2.2. Furthermore, when $|g(x)|$, $x\in\mathbb{R}_+$, is a decreasing function, it holds that $\mu_l \leq \rho_l$ where ρ_l is defined at (4.1). In this case condition (3.4) does follow from (2.1) by Lemma 4.1. ◇

Note that condition (3.5) implies that the stochastic process $(G_t)_{t\geq 0}$ admits a continuous modification. The tightness result associated with convergence at (3.1) is a much more delicate problem. Indeed, it appears to be difficult to prove tightness by standard criteria and, in fact, we are not sure whether tightness holds. However, since our estimation functional defined at (1.5) is of the integral form, tightness is not required to deduce the asymptotic normality of the estimator $(\widehat{\alpha}_n, \widehat{\theta}_n)$ via Proposition 3.1. To be more specific, we demonstrate the following lemma.

Lemma 3.3. *Let $(Y_u^n)_{u\geq 0}$ and $(Y_u)_{u\geq 0}$ be continuous stochastic processes with $Y^n \xrightarrow{\mathcal{L}\text{-f}} Y$. Suppose that $\int_{\mathbb{R}_+}\mathbb{E}[|Y_u^n|]\,\mathrm{d}u < \infty$ and $\int_{\mathbb{R}_+}\mathbb{E}[|Y_u|]\,\mathrm{d}u<\infty$, and define*

$$X_{n,m,l} := \int_0^l Y^n_{\lfloor um\rfloor/m}\,\mathrm{d}u, \qquad X_{n,l} := \int_0^l Y_u^n\,\mathrm{d}u.$$

Assume that the following conditions hold:

$$\lim_{l\to\infty}\limsup_{n\to\infty}\int_l^\infty \mathbb{E}[|Y_u^n|]\,\mathrm{d}u = 0, \qquad \lim_{m\to\infty}\limsup_{n\to\infty}\mathbb{P}(|X_{n,m,l} - X_{n,l}| > \epsilon) = 0, \tag{3.6}$$

where the latter convergence holds for all $l,\epsilon>0$. Then we obtain the convergence in distribution

$$\int_{\mathbb{R}_+} Y_u^n\,\mathrm{d}u \xrightarrow{\mathcal{L}} \int_{\mathbb{R}_+} Y_u\,\mathrm{d}u \qquad \text{as } n\to\infty.$$

Proof. Observe the decomposition

$$\int_{\mathbb{R}_+} Y_u^n \,\mathrm{d}u = X_{n,m,l} + (X_{n,l} - X_{n,m,l}) + \int_l^\infty Y_u^n \,\mathrm{d}u.$$

For any fixed $m \in \mathbb{N}$, it holds that

$$X_{n,m,l} \xrightarrow{\mathcal{L}} X_{m,l} := \int_0^l Y_{\lfloor um \rfloor / m} \,\mathrm{d}u \qquad \text{as } n \to \infty,$$

since $Y^n \xrightarrow{\mathcal{L}\text{-f}} Y$. Due to continuity of the process Y we also have that $X_{m,l} \xrightarrow{\text{a.s.}} \int_0^l Y_u \,\mathrm{d}u$ as $m \to \infty$. By conditions (3.6) we obtain the assertion of Lemma 3.3. □

Note that the conditions stated in (3.6) are usually easier to check in practical applications than tightness of the process $(Y_t^n)_{t \geq 0}$.

To formulate the main result of the paper we need to introduce some more notations. We set $\xi = (\alpha, \theta) \in (0,2) \times \Theta$ and denote by ξ_0 the true parameter of the model. We further define the function

$$F(\psi, \xi) := \|\psi - \phi_\xi\|^2_{\mathcal{L}^2_w}, \qquad \psi \in \mathcal{L}^2_w. \tag{3.7}$$

We now transform the M-estimator at (1.5) into a Z-estimator by using the criterium

$$\nabla_\xi F(\psi, \xi) = 0, \tag{3.8}$$

which is satisfied at (ϕ_{ξ_0}, ξ_0). For each $\xi \in (0,2) \times \Theta$ and $\psi \in \mathcal{L}^2_w$ we denote by $\Phi(\psi)$ an element of $(0,2) \times \Theta$ such that $\nabla_\xi F(\psi, \Phi(\psi)) = 0$ (if such an element exists). To compute the derivative of Φ we recall the implicit function theorem on Banach spaces. Let $(E_j, \|\cdot\|_{E_j})$, $j = 1,2,3$, be some Banach spaces and let $\mathcal{E}_j \subseteq E_j$, $j = 1,2$, be open sets. Consider a Fréchet differentiable function $f : \mathcal{E}_1 \times \mathcal{E}_2 \to E_3$. For $(e_1, e_2) \in \mathcal{E}_1 \times \mathcal{E}_2$ and $(h_1, h_2) \in E_1 \times E_2$ we denote by $D^k_{h_k} f(e_1, e_2)$, $k = 1,2$, the Fréchet derivative of f in the direction $h_k \in E_k$. The implicit function theorem is then formulated as follows. Assume that an element $(e_1^0, e_2^0) \in \mathcal{E}_1 \times \mathcal{E}_2$ satisfies $f(e_1^0, e_2^0) = 0$ and the map $D^2_\cdot f(e_1^0, e_2^0) : E_2 \to E_3$ is continuous and invertible. Then there exist open sets $U \subseteq \mathcal{E}_1$ and $V \subseteq \mathcal{E}_2$ such that $(e_1^0, e_2^0) \in U \times V$ and a bijective map $p : U \to V$ with

$$f(e_1, e_2) = 0 \qquad \text{if and only if} \qquad p(e_1) = e_2.$$

Furthermore, the function p is Fréchet differentiable and the derivative $D_h p : U \to V$ is given by

$$D_h p(e_1) = -\Big(D^2_\cdot f(e_1, p(e_1))\Big)^{-1} \Big(D^1_h f(e_1, p(e_1))\Big).$$

We now apply this statement to our setting. Here $\mathcal{E}_1 = \mathcal{L}^2_w$, $\mathcal{E}_2 = (0,2) \times \Theta \subseteq \mathbb{R}^2$, $f = \nabla_\xi F$ and $p = \Phi$. Using the differentiability condition of Assmuption (A) we can conclude the existence of $D^2 \nabla_\xi F(\psi, \xi)$ and $D^1_h \nabla_\xi F(\psi, \xi)$. In particular, applying the representation $F(\psi, \xi) = \langle \psi - \phi_\xi, \psi - \phi_\xi \rangle_{\mathcal{L}^2_w}$, we obtain the formulae

$$\begin{aligned} D^2 \nabla_\xi F(\phi_{\xi_0}, \xi_0) &= \nabla^2_\xi F(\phi_{\xi_0}, \xi_0) = 2\big(\langle \partial_{\xi_i} \phi_{\xi_0}, \partial_{\xi_j} \phi_{\xi_0} \rangle_{\mathcal{L}^2_w}\big)_{i,j=1,2}, \\ D^1_h \nabla_\xi F(\phi_{\xi_0}, \xi_0) &= -2\big(\langle \partial_{\xi_i} \phi_{\xi_0}, h \rangle_{\mathcal{L}^2_w}\big)_{i=1,2}, \qquad h \in L^2_w. \end{aligned} \tag{3.9}$$

The matrix $\nabla_\xi^2 F(\phi_{\xi_0}, \xi_0)$ is positive definite if and only if the functions $\partial_\theta \phi_{\xi_0}$ and $\partial_\alpha \phi_{\xi_0}$ are linearly independent, which is obviously true since

$$\begin{aligned} \partial_\theta \phi_{\xi_0}(u) &= -u^{\alpha_0} \phi_{\xi_0}(u) \partial_\theta \big(\|g_\theta\|_{\alpha_0}^{\alpha_0} \big)_{\theta=\theta_0} \\ \partial_\alpha \phi_{\xi_0}(u) &= -\phi_{\xi_0}(u) u^{\alpha_0} \Big(\log(u) \|g_{\theta_0}\|_{\alpha_0}^{\alpha_0} + \partial_\alpha \big(\|g_{\theta_0}\|_\alpha^\alpha \big)_{\alpha=\alpha_0} \Big). \end{aligned} \tag{3.10}$$

The following result presents the asymptotic properties of the minimal contrast estimator $\xi_n := (\widehat{\alpha}_n, \widehat{\theta}_n)$ defined at (1.5).

Theorem 3.4. *Let $\xi_0 \in (0,2) \times \Theta$ denote the true parameter of the model* (1.1)*. Then we obtain the strong consistency*

$$\xi_n \xrightarrow{\text{a.s.}} \xi_0 \qquad \text{as } n \to \infty.$$

Furthermore, if Assmuption (A) and (3.4) *hold, we have*

$$\sqrt{n}(\xi_n - \xi_0) \xrightarrow{\mathcal{L}} \big(\langle \partial_{\xi_i} \phi_{\xi_0}, \partial_{\xi_j} \phi_{\xi_0} \rangle_{\mathcal{L}_w^2} \big)_{i,j=1,2}^{-1} \big(\langle \partial_{\xi_i} \phi_{\xi_0}, G \rangle_{\mathcal{L}_w^2} \big)_{i=1,2}, \tag{3.11}$$

where the Gaussian process $(G_t)_{t\ge 0}$ has been introduced in Proposition 3.1. In particular, the above asymptotic distribution is a two-dimensional Gaussian with mean 0*.*

In theory we may obtain confidence regions for the unknown parameter $\xi_0 \in (0,2) \times \Theta$ by estimating the asymptotic covariance matrix of the bivariate normal limit appearing in Theorem 3.4. Indeed, by continuity and dominated convergence we have that

$$\big(\langle \partial_{\xi_i} \phi_{\xi_n}, \partial_{\xi_j} \phi_{\xi_n} \rangle_{\mathcal{L}_w^2} \big)_{i,j=1,2}^{-1} \xrightarrow{\text{a.s.}} \big(\langle \partial_{\xi_i} \phi_{\xi_0}, \partial_{\xi_j} \phi_{\xi_0} \rangle_{\mathcal{L}_w^2} \big)_{i,j=1,2}^{-1},$$

and similarly we can consistently estimate the covariance kernel of G via replacing ξ_0 by ξ_n at (3.2). However, this procedure is extremely involved from the numerical point of view, since the asymptotic covariance matrix contains improper integrals and an infinite sum with a potentially slow rate of convergence. Even in the setting of the exponential family g_θ discussed in Example 2.2, the asymptotic covariance can not be computed explicitly.

To overcome this problem we propose an alternative numerical procedure, which provides an estimator of the asymptotic covariance matrix in the central limit theorem (3.11). Our method is based upon the following steps:

1. Compute the minimal contrast estimator ξ_n of ξ_0 from the data $X_1, \ldots, X_n$.
2. For $k = 1, \ldots, N$ generate new independent samples $X_1^{(k)}, \ldots, X_n^{(k)}$ from model (1.1) with parameter ξ_n.
3. For $k = 1, \ldots, N$ compute estimators $\xi_n^{(k)}$ from the data $X_1^{(k)}, \ldots, X_n^{(k)}$ via (1.5).
4. Compute the empirical covariance matrix $V_{n,N}$ of $\xi_n^{(k)}$, $k = 1, \ldots, N$.

The estimator at (1.5) can be obtained by solving the equation

$$\nabla_\xi F(\phi_n, \xi) = 0,$$

which is a standard numerical problem. Hence, the estimator $V_{n,N}$ for large n and N is likely to be a better proxy for the unknown asymptotic covariance matrix in (3.11) than its direct numerical approximation.

4 Proofs

Throughout the proof we denote all positive constants by C (or by C_p if they depend on the external parameter p) although they may change from line to line.

We start with some preliminary results. Let us introduce the quantity

$$\rho_l := \int_{\mathbb{R}} |g_\theta(x) g_\theta(x+l)|^{\alpha/2} \, \mathrm{d}x, \qquad l \in \mathbb{Z}. \tag{4.1}$$

Our first result, which has been shown in [18], concerns the summability of the coefficients ρ_l.

Lemma 4.1. *Assume that conditions* (2.1) *holds. Then we have*

$$\sum_{l\in\mathbb{Z}} \rho_l < \infty.$$

Proof. We obtain by Cauchy–Schwarz inequality that

$$\begin{aligned}\sum_{l\in\mathbb{Z}} \rho_k &= \sum_{l,m\in\mathbb{Z}} \int_{m-1}^{m} |g_\theta(x) g_\theta(x+l)|^{\alpha/2} \, \mathrm{d}x \\ &\le \sum_{l,m\in\mathbb{Z}} \left(\int_{m-1}^{m} |g_\theta(x)|^\alpha \, \mathrm{d}x \right)^{1/2} \left(\int_{m-1}^{m} |g_\theta(x+l)|^\alpha \, \mathrm{d}x \right)^{1/2} \\ &= \sum_{m\in\mathbb{Z}} \left(\int_{m-1}^{m} |g_\theta(x)|^\alpha \, \mathrm{d}x \right)^{1/2} \sum_{l\in\mathbb{Z}} \left(\int_{m+l-1}^{m+l} |g_\theta(x)|^\alpha \, \mathrm{d}x \right)^{1/2}.\end{aligned}$$

Hence, the assertion follows from (2.1). □

Another important ingredient is the following measure of dependence. Let $X = \int_{\mathbb{R}} g_s \, \mathrm{d}L_s$ and $Y = \int_{\mathbb{R}} h_s \, \mathrm{d}L_s$ with $\|g\|_\alpha, \|h\|_\alpha < \infty$. Then we introduce the measure of dependence $U_{g,h} : \mathbb{R}^2 \to \mathbb{R}$ via

$$\begin{aligned} U_{g,h}(u,v) &:= \mathbb{E}[\exp(\mathrm{i}(uX - vY))] - \mathbb{E}[\exp(\mathrm{i}uX)]\mathbb{E}[\exp(-\mathrm{i}vY)] \\ &= \exp(-\|ug - vh\|_\alpha^\alpha) - \exp(-\|ug\|_\alpha^\alpha - \|vh\|_\alpha^\alpha). \end{aligned} \tag{4.2}$$

The following result is the statement of inequalities (3.4)–(3.6) from [18].

Lemma 4.2. *For any $u, v \in \mathbb{R}$ it holds that*

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

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

4.1 Proof of Proposition 3.1

The finite dimensional convergence in (3.1) is a direct consequence of Theorem 2.4. To identify the covariance kernel of $(G_t)_{t\geq 0}$, note that for $Y = \int_{\mathbb{R}} g_s \, \mathrm{d}L_s$ with $\|g\|_\alpha < \infty$ it holds that

$$\mathbb{E}[\exp(\mathrm{i}uY)] = \exp(-\|ug\|_\alpha^\alpha).$$

Applying the identity $\cos(ux) = (\exp(\mathrm{i}ux) + \exp(-\mathrm{i}ux))/2$ we deduce the formula for $\mathbb{E}[G_u G_v]$ through a straightforward computation.

To show the remaining statements we introduce the definition

$$r(u,v) := \mathbb{E}[G_u G_v] = \sum_{l\in\mathbb{Z}} r_l(u,v)$$

with

$$\begin{aligned} r_l(u,v) := \tfrac{1}{2}\Big(&\exp(-\|ug_\theta + vg_\theta(\,\cdot\,+l)\|_\alpha^\alpha) + \exp(-\|ug_\theta - vg_\theta(\,\cdot\,+l)\|_\alpha^\alpha) \\ &- 2\exp(-(u^\alpha + v^\alpha)\|g_\theta\|_\alpha^\alpha)\Big). \end{aligned} \tag{4.3}$$

Applying Lemma 4.2 to the functions $g = g_\theta$, $h = g_\theta(\,\cdot\,+l)$ and $u \geq 0$, we deduce that

$$|r_l(u,u)| \leq C\rho_l u^\alpha \exp\!\big(-2u^\alpha(\|g_\theta\|_\alpha^\alpha - \rho_l)\big).$$

By Cauchy–Schwarz inequality we conclude that $\sup_{l\in\mathbb{Z}\setminus\{0\}} \rho_l < \|g_\theta\|_\alpha^\alpha$ and hence

$$|r_l(u,u)| \leq C\rho_l u^\alpha \exp(-u^\alpha/C). \tag{4.4}$$

Now, Lemma 4.1 implies the statement (3.3).

Next, we turn our attention to the proof of (3.5). We first start with a simplification. Since $\cos(ux) = (\exp(\mathrm{i}ux) + \exp(-\mathrm{i}ux))/2$ it suffices to show (3.5) for the asymptotic covariance kernel that corresponds to the function $\exp(\mathrm{i}ux)$, i.e.

$$\begin{aligned} \overline{r}(u,v) = \sum_{l\in\mathbb{Z}} \overline{r}_l(u,v), \quad \overline{r}_l(u,v) \\ = \exp(-\|ug_\theta - vg_\theta(\,\cdot\,+l)\|_\alpha^\alpha) - \exp(-(u^\alpha+v^\alpha)\|g_\theta\|_\alpha^\alpha). \end{aligned}$$

Furthermore, due to $\overline{r}(u,u) + \overline{r}(v,v) - 2\overline{r}(u,v) \leq |\overline{r}(u,u) - \overline{r}(u,v)| + |\overline{r}(v,v) - \overline{r}(u,v)|$, we will only prove that

$$|\overline{r}(u,u) - \overline{r}(u,v)| \leq C_T |u-v|^{\alpha/2}, \quad u,v \in [0,T].$$

Observe the identity

$$\begin{aligned} &\overline{r}_l(u,u) - \overline{r}_l(u,v) \\ &\quad= \exp(-2u^\alpha\|g_\theta\|_\alpha^\alpha)\Big\{\exp(-\|u(g_\theta - g_\theta(\,\cdot\,+l))\|_\alpha^\alpha + 2u^\alpha\|g_\theta\|_\alpha^\alpha) - 1\Big\} \\ &\qquad- \exp(-(u^\alpha+v^\alpha)\|g_\theta\|_\alpha^\alpha)\Big\{\exp(-\|ug_\theta - vg_\theta(\,\cdot\,+l)\|_\alpha^\alpha + (u^\alpha+v^\alpha)\|g_\theta\|_\alpha^\alpha) - 1\Big\} \\ &\quad= \Big\{\exp(-2u^\alpha\|g_\theta\|_\alpha^\alpha) - \exp(-(u^\alpha+v^\alpha)\|g_\theta\|_\alpha^\alpha)\Big\} \\ &\qquad\times\Big\{\exp(-\|u(g_\theta - g_\theta(\,\cdot\,+l))\|_\alpha^\alpha + 2u^\alpha\|g_\theta\|_\alpha^\alpha) - 1\Big\} + \exp(-(u^\alpha+v^\alpha)\|g_\theta\|_\alpha^\alpha) \\ &\qquad\times\Big\{\exp(-\|u(g_\theta - g_\theta(\,\cdot\,+l))\|_\alpha^\alpha + 2u^\alpha\|g_\theta\|_\alpha^\alpha) \\ &\qquad- \exp(-\|ug_\theta - vg_\theta(\,\cdot\,+l)\|_\alpha^\alpha + (u^\alpha+v^\alpha)\|g_\theta\|_\alpha^\alpha)\Big\} \\ &\quad=: \overline{r}_l^{(1)}(u,v) + \overline{r}_l^{(2)}(u,v). \end{aligned}$$

We apply the second inequality of Lemma 4.2 and the mean value theorem to conclude the estimate

$$|\overline{r}_l^{(1)}(u,v)| \le C_T \rho_l |u^\alpha - v^\alpha| \le C_T \rho_l |u-v|^{\alpha/2} \qquad \text{for } u,v \in [0,T]. \tag{4.5}$$

Applying the mean value theorem once again we deduce that

$$|\overline{r}_l^{(2)}(u,v)| \le C_T \left| \|u g_\theta - v g_\theta(\,\cdot\,+l)\|_\alpha^\alpha - \|u(g_\theta - g_\theta(\,\cdot\,+l))\|_\alpha^\alpha + (u^\alpha - v^\alpha)\|g_\theta\|_\alpha^\alpha \right|.$$

In the next step we write

$$\begin{aligned} &\|u g_\theta - v g_\theta(\,\cdot\,+l)\|_\alpha^\alpha - \|u(g_\theta - g_\theta(\,\cdot\,+l))\|_\alpha^\alpha + (u^\alpha - v^\alpha)\|g_\theta\|_\alpha^\alpha \\ &\quad = \int_{\mathbb{R}_+} |u g_\theta(x) - v g_\theta(x+l)|^\alpha - |u(g_\theta(x) - g_\theta(x+l))|^\alpha + (u^\alpha - v^\alpha)|g_\theta(x+l)|^\alpha \,\mathrm{d}x. \end{aligned}$$

Since $\alpha \in (0,2)$ we have the inequality $|x^\alpha - y^\alpha| \le |x^2 - y^2|^{\alpha/2}$ for any $x,y \in \mathbb{R}_+$. Hence, we conclude that

$$\begin{aligned} &\left| |u g_\theta(x) - v g_\theta(x+l)|^\alpha - |u(g_\theta(x) - g_\theta(x+l))|^\alpha \right| \\ &\qquad \le C_T |u-v|^{\alpha/2} \times \left(|g_\theta(x+l)|^\alpha + |g_\theta(x) g_\theta(x+l)|^{\alpha/2} \right) \end{aligned}$$

for any $u,v \in [0,T]$. Consequently, it holds that

$$|\overline{r}_l^{(2)}(u,v)| \le C_T |u-v||^{\alpha/2}(\rho_l + \mu_l) \qquad \text{for } u,v \in [0,T], \tag{4.6}$$

where the quantity μ_l has been introduced in (3.4). Finally, by Lemma 4.1, condition (3.4) and inequalities (4.5), (4.6), we obtain the assertion

$$\mathbb{E}[(G_u - G_v)^2] \le C_T |u-v|^{\alpha/2}, \qquad u,v \in [0,T],$$

which finishes the proof of Proposition 3.1. □

4.2 Proof of Theorem 3.4

We recall the notation $\xi = (\alpha, \theta)$, $\xi_n = (\widehat{\alpha}_n, \widehat{\theta}_n)$ and $\xi_0 = (\alpha_0, \theta_0)$. The strong consistency result of the estimator ξ_n follows from standard results for M-estimators, which we state for completeness. Since the map $\xi \mapsto \phi_\xi$ is bijective and continuous it suffices to show that

$$\|\phi_{\xi_n} - \phi_{\xi_0}\|_{\mathcal{L}_w^2} \xrightarrow{\text{a.s.}} 0$$

to prove $\xi_n \xrightarrow{\text{a.s.}} \xi_0$. We deduce the inequality

$$\begin{aligned} \|\phi_{\xi_n} - \phi_{\xi_0}\|_{\mathcal{L}_w^2} &\le \|\phi_n - \phi_{\xi_0}\|_{L_w^2} + \|\phi_n - \phi_{\xi_n}\|_{\mathcal{L}_w^2} \\ &\le 2\|\phi_n - \phi_{\xi_0}\|_{\mathcal{L}_w^2}. \end{aligned}$$

Since $\phi_n(u) \xrightarrow{\text{a.s.}} \phi_{\xi_0}(u)$ for all $u \in \mathbb{R}_+$, we conclude by a standard argument and dominated convergence that

$$\|\phi_{\xi_n} - \phi_{\xi_0}\|_{\mathcal{L}_w^2} \xrightarrow{\text{a.s.}} 0,$$

which shows the strong consistency of our minimal contrast estimator.

Now, we prove the central limit theorem of Theorem 3.4. First of all, note that

$$\Phi(\phi_n) = \xi_n \qquad \text{and} \qquad \Phi(\phi_{\xi_0}) = \xi_0.$$

Since the function Φ is Fréchet differentiable, we obtain the decomposition

$$\begin{aligned}\sqrt{n}(\xi_n - \xi_0) &= \sqrt{n}\big(\Phi(\phi_n) - \Phi(\phi_{\xi_0})\big) \\ &= \big(\nabla_\xi^2 F(\xi_0)\big)^{-1} D^1_{\sqrt{n}(\phi_n - \phi_{\xi_0})} \nabla_\xi F(\xi_0) + \sqrt{n}\|\phi_n - \phi_{\xi_0}\|_{\mathcal{L}^2_w} R(\phi_n - \phi_{\xi_0}),\end{aligned} \tag{4.7}$$

where $R(\phi_n - \phi_{\xi_0}) \xrightarrow{\text{a.s.}} 0$ as $\|\phi_n - \phi_{\xi_0}\|_{L^2_w} \xrightarrow{\text{a.s.}} 0$. Due to (3.9), we deduce that

$$\begin{aligned}\big(\nabla_\xi^2 F(\xi_0)\big)^{-1} D^1_{\sqrt{n}(\phi_n - \phi_{\xi_0})} \nabla_\xi F(\xi_0) &= \big(\langle \partial_{\xi_i}\phi_{\xi_0}, \partial_{\xi_j}\phi_{\xi_0}\rangle_{\mathcal{L}^2_w}\big)^{-1}_{i,j=1,2} \\ &\quad \times \big(\langle \partial_{\xi_i}\phi_{\xi_0}, \sqrt{n}(\phi_n - \phi_{\xi_0})\rangle_{\mathcal{L}^2_w}\big)_{i=1,2}.\end{aligned}$$

In view of the decomposition (4.7) it suffices to show the convergence results

$$\begin{gathered}\big(\langle \partial_{\xi_i}\phi_{\xi_0}, \sqrt{n}(\phi_n - \phi_{\xi_0})\rangle_{\mathcal{L}^2_w}\big)_{i=1,2} \xrightarrow{\mathcal{L}} \big(\langle \partial_{\xi_i}\phi_{\xi_0}, G\rangle_{\mathcal{L}^2_w}\big)_{i=1,2}, \\ \sqrt{n}\|\phi_n - \phi_{\xi_0}\|_{\mathcal{L}^2_w} \xrightarrow{\mathcal{L}} \|G\|_{\mathcal{L}^2_w}\end{gathered}$$

to prove the central limit theorem in Theorem 3.4. We concentrate on the first convergence, since the second one follows by the same arguments. We apply Lemma 3.3 and note that conditions (3.6) can be checked for each component separately. Hence, we set for a fixed $i = 1, 2$

$$Y_u^n = \sqrt{n}(\phi_n(u) - \phi_{\xi_0}(u))\partial_{\xi_i}\phi_{\xi_0}(u)w(u), \qquad Y_u = G_u \partial_{\xi_i}\phi_{\xi_0}(u)w(u).$$

We obviously have that $Y^n \xrightarrow{\mathcal{L}\text{-f}} Y$ by Proposition 3.1. Observe that

$$\mathbb{E}[\leq](\partial_{\xi_i}\phi_{\xi_0}(u)w(u))^2 \sum_{l\in\mathbb{Z}} |r_l(u,u)|,$$

where $r_l(u,v)$ has been introduced at (4.3). Applying the inequality (4.4) we obtain that

$$\mathbb{E}[|Y_u^n|] \leq C\,|\partial_{\xi_i}\phi_{\xi_0}(u)w(u)| u^{\alpha/2} \exp(-u^\alpha/C),$$

which shows the first condition of (3.6). Similarly, we have the inequality

$$\mathbb{E}[(Y_u^n - Y_v^n)^2] \leq C_T\Big(|u-v|^{\min(\alpha,1)} + \sum_{l\in\mathbb{Z}} |r_l(u,u) + r_l(v,v) - 2r_l(u,v)|\Big)$$

for any $u, v \in [0, T]$, thanks to (3.10) and $w \in C^1(\mathbb{R}_+)$. Thus, using the arguments from the previous section we conclude that

$$\mathbb{E}[(Y_u^n - Y_v^n)^2] \leq C_T\,|u-v|^{\alpha/2}, \qquad u, v \in [0, T].$$

The latter estimate and the Markov inequality imply the second condition of (3.6). This completes the proof of Theorem 3.4.

Acknowledgement

The authors acknowledge financial support from the project “Ambit fields: probabilistic properties and statistical inference” funded by Villum Fonden.

References

[1] Antoine Ayachea and Julien Hamoniera (2012). »*Linear fractional stable motion: A wavelet estimator of the α parameter*«. Statistics and Probability Letters 82(8), 1569–1575. doi: 10.1016/j.spl.2012.04.005.

[2] Andreas Basse-O'Connor, Claudio Heinrich and Mark Podolskij (2019). »*On limit theory for functionals of stationary increments Lévy driven moving averages*«. Electronic Journal of Probability 24(79), 42 pp. doi: 10.1214/19-EJP336.

[3] Andreas Basse-O'Connor, Raphaël Lachièze-Rey and Mark Podolskij (2017). »*Power variation for a class of stationary increments Lévy driven moving averages*«. The Annals of Probability 45(6B), 4477–4528. doi: 10.1214/16-AOP1170.

[4] Andreas Basse-O'Connor and Mark Podolskij (2017). »*On critical cases in limit theory for stationary increments Lévy driven moving averages*«. Stochastics 89(1), 360–383. doi: 10.1080/17442508.2016.1191493.

[5] Albert Benassi, Serge Cohen and Jacques Istas (2004). »*On roughness indices for fractional fields*«. Bernoulli 10(2), 357–373. doi: 10.3150/bj/1082380223.

[6] Alexandre Brouste and Masaaki Fukasawa (2017). »*Local asymptotic normality for fractional Gaussian noise under high-frequency observations*«. The Annals of Statistics 46(5), 2045–2061. doi: 10.1214/17-AOS1611.

[7] Stamatis Cambanis, Clyde D. Jr. Hardin and Aleksander Weron (1987). »*Ergodic properties of stationary stable processes*«. Stochastic Processes and Their Applications 24(1), 1–18. doi: 10.1016/0304-4149(87)90024-X.

[8] Jean-François Coeurjolly and Jacques Istas (2001). »*Cramèr-Rao bounds for fractional Brownian motions*«. Statistics and Probability Letters 53(4), 435–447. doi: 10.1016/S0167-7152(00)00197-8.

[9] Rainer Dahlhaus (1989). »*Efficient parameter estimation for self-similar processes*«. The Annals of Statistics 17(4), 1749–1766. doi: 10.1214/aos/1176347393.

[10] Thi To Nhu Dang and Jacques Istas (2017). »*Estimation of the Hurst and the stability indices of a H-self-similar stable process*«. Electronic Journal of Statistics 11(2), 4103–4150. doi: 10.1214/17-EJS1357.

[11] Danijel Grahovac, Nikolai N. Leonenko and Murad S. Taqqu (2015). »*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. doi: 10.1007/s10955-014-1126-4.

[12] Jacques Istas and Gabriel Lang (1997). »*Quadratic variations and estimation of the local Hölder index of a Gaussian process*«. Annales de l'I.H.P. Probabilités et statistiques 33(4), 407–436.

[13] Mathias Mørck Ljungdahl and Mark Podolskij (2018). »*A limit theorem for a class of stationary increments Lévy moving averages processes with multiple singularities*«. Modern Stochastics: Theory and Applications 5(3), 297–316. doi: 10.15559/18-VMSTA111.

[14] Mathias Mørck Ljungdahl and Mark Podolskij (2020). »*A minimal contrast estimator for the linear fractional motion*«. Statistical Inference for Stochastic Processes 23, 381–413. doi: 10.1007/s11203-020-09216-2.

[15] Benoit B. Mandelbrot and John W. Van Ness (1968). »*Fractional Brownian motions, fractional noises and applications*«. SIAM Review 10(4), 422–437. doi: 10.1137/1010093.

[16] Stepan Mazur, Dmitry Otryakhin and Mark Podolskij (2020). »*Estimation of the linear fractional stable motion*«. Bernoulli 26(1), 226–252. doi: 10.3150/19-BEJ1124.

[17] Vladas Pipiras and Murad S. Taqqu (2003). »*Central limit theorems for partial sums of bounded functionals of infinite-variance moving averages*«. Bernoulli 9(5), 833–855. doi: 10.3150/bj/1066418880.

[18] Vladas Pipiras, Murad S. Taqqu and Patrice Abry (2007). »*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. doi: 10.3150/07-BEJ6143.

[19] Balram S. Rajput and Jan Rosiński (1989). »*Spectral representations of infinitely divisible processes*«. Probability Theory and Related Fields 82(3), 451–487. doi: 10.1007/BF00339998.

D

PAPER

Multi-Dimensional Normal Approximation of Heavy-Tailed Moving Averages

Ehsan Azmoodeh

Department of Mathematical Sciences, University of Liverpool

Mathias Mørck Ljungdahl

Department of Mathematics, Aarhus University

Christoph Thäle

Faculty of Mathematics, Ruhr University Bochum

Abstract. In this paper we extend the refined second-order Poincaré inequality in [2] from a one-dimensional to a multi-dimensional setting. Its proof is based on a multivariate version of the Malliavin–Stein method for normal approximation on Poisson spaces. We also present an application to partial sums of vector-valued functionals of heavy-tailed moving averages. The extension we develop is not only in the co-domain of the functional, but also in its domain. Such a set-up has previously not been explored in the framework of stable moving average processes. It can potentially capture probabilistic properties which cannot be described solely by the one-dimensional marginals, but instead require the joint distribution.

Key words: Central limit theorem, heavy-tailed moving average, Lévy process, Malliavin–Stein method, Poisson random measure, second-order Poincaré inequality

AMS 2010 subject classifications: 60F05, 60G10, 60G15, 60G52, 60G55, 60H07

1 Introduction

In recent decades the combination of Malliavin calculus and Stein's method for normal approximation has led to a plethora of Gaussian limit theorems in fields ranging from stochastic geometry, over cosmology to statistics. Classically, the assumptions require third or fourth moment conditions which makes the Malliavin–Stein method unsuitable for distributions with heavier tails. However, in [2] a careful differentiation between small and large values has led to a refined so-called second-order Poincaré inequality for Poisson functionals, which allows to circumvent these difficulties to a certain extent. Based on the approach in [13] the principal goal of this paper is to obtain a multivariate extension of the central results in [2]. This opens the possibility to capture properties of the underlying process not accessible solely by the one-dimensional marginal distributions.

We shall now define the heavy-tailed moving average model to which we are going to apply our general multivariate central limit theorem. Let $L = (L_t)_{t\in\mathbb{R}}$ be a two-sided Lévy process with no Gaussian component and Lévy measure ν. We assume that the latter admits a Lebesgue density $w : \mathbb{R} \to \mathbb{R}$ such that

$$|w(x)| \le C\,|x|^{-1-\beta} \tag{1.1}$$

for all $x \neq 0$, some $\beta \in (0,2)$ and a constant $C > 0$. Hence, the distribution of L_1 exhibits β-stable tails, see [19]. Consider then for each $i \in \{1,\dots,m\}$, $m \in \mathbb{N}$, the process

$$X_t^i := \int_{\mathbb{R}} g_i(t-s)\,\mathrm{d}L_s, \qquad t \in \mathbb{R}, \tag{1.2}$$

for some measurable function $g_i : \mathbb{R} \to \mathbb{R}$. Necessary and sufficient conditions for the integral to exists are given in [15] and if L is symmetric around zero, i.e. if $-L_1$ and L_1 are identically distributed, then we mention that a sufficient condition is $\int_{\mathbb{R}} |g_i(s)|^\beta\,\mathrm{d}s < \infty$.

The main examples of kernels g_i we consider satisfy a power law behaviour around zero and at infinity. Henceforth we shall assume for all $i \in \{1,\dots,m\}$ the existence of a constant $K > 0$ together with exponents $\alpha_i > 0$ and $\kappa_i \in \mathbb{R}$ such that

$$|g_i(x)| \le K(x^{\kappa_i}\mathbb{1}_{[0,a_i)}(x) + x^{-\alpha_i}\mathbb{1}_{[a_i,\infty)}(x)) \tag{1.3}$$

for all $x \in \mathbb{R}$, where $a_i > 0$ are suitable splitting points, which may alter the constant K. Without loss of generality we choose $a_i = 1$ for all $i \in \{1,\dots,m\}$ and let K stand for the corresponding constant. Note that this in particular implies that $g_i(x) = 0$ for $x < 0$, consequently we will only consider *casual* moving averages.

The main objects of interest in this paper are rescaled partial sums of multi-dimensional functionals of the joint distribution $X_s = (X_s^1,\dots,X_s^m)$, namely

$$V_n(X;f) = \frac{1}{\sqrt{n}}\sum_{s=1}^{n}\Big(f(X_s^1,\dots,X_s^m) - \mathbb{E}[f(X_0^1,\dots,X_0^m)]\Big), \tag{1.4}$$

where $f : \mathbb{R}^m \to \mathbb{R}^d$ is a suitable Borel-measurable function, with d being some positive integer. Observe that $V_n(X;f)$ is a d-dimensional random vector and for convenience we shall denote by $V_n^i(X;f)$ its ith coordinate. We remark that in the one-dimensional case $d = m = 1$ the distributional convergence of $V_n(X;f)$, as $n \to \infty$, is studied for general functions f in [1] and here the so-called Appell rank of f is seen to play an important role. The results in that paper also imply that one cannot in general expect convergence in distribution after rescaling with the factor $\sqrt{n}$ as in (1.4) or a Gaussian limiting distribution if the memory of the processes are too long, i.e. if the α_i are too close to 0. We shall see that if the tails are not too heavy and the memory is not too long, which in our case means that $\alpha_i\beta > 2$, we do in fact have convergence in distribution of $V_n(X;f)$ to a Gaussian random variable and we shall discuss the *speed* of this convergence by considering an appropriate metric on the space of probability laws on $\mathbb{R}^d$, see Section 2 below. To conclude such a result, we could also in principle rely on a multivariate second-order Poincaré inequality for random vectors of Poisson functionals in [18]. But as already observed in the one-dimensional case, the existing bounds are not suitable for the application to Lévy driven moving averages just described. In fact, in this specific situation the bounds in [18] do not even tend to zero, as n increases. Against this background we will develop in this paper a refined multivariate second-order Poincaré inequality for general random vectors of Poisson functionals, which is more adapted to our situation and allows us to distinguish carefully between small and large values. We believe that this result is of independent interest as well. This eventually paves the way to the central limit theory for the random vectors $V_n(X;f)$.

One motivation for the extension of the theory from [2] to a multivariate set-up is the fact that important properties of random processes, such as self-similarity, are determined by the finite dimensional distributions of X, but not by the one-dimensional marginals. The one-dimensional theory, i.e. the case $d = m = 1$, could so far capture only probabilistic properties of the distribution of X_1. In the case of the linear fractional stable motion this made the estimation of the Hurst parameter problematic if jointly estimated with the scale parameter and the stability index β, see [9]. We will come back to such applications of the results we develop in this paper in a separate work. However, finally we would like to mention that the case $m = 1$ and general d has been considered in the seminal paper [14].

2 Main Results

2.1 A Refined Multivariate Second-Order Poincaré Inequality

Consider a measurable space $(S,\mathcal{S})$ equipped with a σ-finite measure μ. Let η be a Poisson process on $(S,\mathcal{S})$ with intensity measure μ. This means that η is a collection of random variables of the form $\eta(B)$, $B \in \mathcal{S}$, with the properties that

(i) for each $B \in \mathcal{S}$ with $\mu(B) < \infty$ the random variable $\eta(B)$ is Poisson distributed with mean $\mu(B)$,

(ii) for $m \in \mathbb{N}$ and pairwise disjoint $B_1,\dots,B_m \in \mathcal{S}$ with $\mu(B_1),\dots,\mu(B_m) < \infty$ the random variables $\eta(B_1),\dots,\eta(B_m)$ are independent.

We can and will regard η as a random function from an underlying probability space $(\Omega,\mathcal{F},\mathbb{P})$ to $\mathcal{N}$, the space of all integer-valued σ-finite measures on $(S,\mathcal{S})$. The set $\mathcal{N}$ is equipped with the evaluation σ-algebra, i.e. the σ-algebra generated by the evaluation mappings $\mu \mapsto \mu(A)$, $A \in \mathcal{S}$.

To each Poisson process η we associate the Hilbert space $\mathcal{L}^2_\eta(\mathbb{P})$ consisting of all square integrable Poisson functionals F, i.e. those random variables for which there exists a function $\phi : \mathcal{N} \to \mathbb{R}$ such that almost surely $F = \phi(\eta) \in \mathcal{L}^2(\mathbb{P})$. Finally, we introduce the notion of the Malliavin derivative in a Poisson setting, which is also known as the add-one-cost operator. For each $z \in S$ and $F = \phi(\eta) \in \mathcal{L}^2_\eta(\mathbb{P})$ we define D_zF as

$$D_zF := \phi(\eta + \delta_z) - \phi(\eta)$$

and note that DF is a bi-measurable map from $\Omega \times S$ to $\mathbb{R}$. In a straightforward way this definition extends to vector-valued Poisson functionals. Indeed, consider $F = (F_1, \dots, F_d)$ where each F_i lies in $\mathcal{L}^2_\eta(\mathbb{P})$, then the Malliavin derivative $D_z F$ at $z \in S$ is given by

$$D_z F = (D_z F_1, \dots, D_z F_d).$$

Similarly to $D_z F$ we may introduce the iterated Malliavin derivative $D^2 F$ of F by putting

$$D^2_{z_1, z_2} F := D_{z_1}(D_{z_2} F) = D_{z_2}(D_{z_1} F), \qquad z_1, z_2 \in S.$$

For further background material on Poisson processes we refer to the treatments in [6, 5, 12], for the Malliavin formalism on Poisson spaces we refer to Section 3.1 below.

To measure the distance between (the laws of) two random vectors X and Y taking values in $\mathbb{R}^d$ we use the so-called d_3-distance, see [13]. To introduce it, assume that $\mathbb{E}[\|X\|^2_{\mathbb{R}^d}], \mathbb{E}[\|Y\|^2_{\mathbb{R}^d}] < \infty$, where $\|\cdot\|_{\mathbb{R}^d}$ stands for the Euclidean norm in $\mathbb{R}^d$. The d_3-distance between (the laws of) random vectors X and Y, denoted by $d_3(X, Y)$, is given by

$$d_3(X, Y) := \sup_{\varphi \in \mathcal{H}_3} \big| \mathbb{E}[\varphi(X)] - \mathbb{E}[\varphi(Y)] \big|,$$

where the class $\mathcal{H}_3$ of test functions indicates the collection of all thrice differentiable functions $\varphi : \mathbb{R}^d \to \mathbb{R}$ (i.e. $\varphi \in \mathcal{C}^3(\mathbb{R}^d, \mathbb{R})$) such that $\|\varphi''\|_\infty \le 1$ and $\|\varphi'''\|_\infty \le 1$, where

$$\|\varphi''\|_\infty := \max_{1 \le i,j \le d} \sup_{x \in \mathbb{R}^d} \Big| \frac{\partial^2}{\partial x_i \partial x_j} \varphi(x) \Big|, \qquad \|\varphi'''\|_\infty := \max_{1 \le i,j,k \le d} \sup_{x \in \mathbb{R}^d} \Big| \frac{\partial^3}{\partial x_i \partial x_j \partial x_k} \varphi(x) \Big|.$$

We can now formulate our multivariate second-order Poincaré inequality, which generalizes [2, Theorem 3.1] and refines [18, Theorem 1.1]. Its proof, which is given in Section 4 below, is based on the Malliavin–Stein technique for normal approximation of random vectors of Poisson functionals. For two Poisson functionals $F, G \in \mathcal{L}^2_\eta(\mathbb{P})$ we define the quantities

$$\gamma_1^2(F, G) := 3 \int_{S^3} \mathbb{E}\big[(D^2_{z_1, z_3} F)^2 (D^2_{z_2, z_3} F)^2\big]^{1/2} \mathbb{E}\big[(D_{z_1} G)^2 (D_{z_2} G)^2\big]^{1/2} \mu^3(\mathrm{d}z_1, \mathrm{d}z_2, \mathrm{d}z_3),$$
$$\gamma_2^2(F, G) := \int_{S^3} \mathbb{E}\big[(D^2_{z_1, z_3} F)^2 (D^2_{z_2, z_3} F)^2\big]^{1/2} \mathbb{E}\big[(D^2_{z_1, z_3} G)^2 (D^2_{z_2, z_3} G)^2\big]^{1/2} \mu^3(\mathrm{d}z_1, \mathrm{d}z_2, \mathrm{d}z_3).$$

Moreover, for $x, y \in \mathbb{R}$ we denote by $x \wedge y = \min\{x, y\}$ the minimum of x and y.

Theorem 2.1. *Let $d \ge 1$ and assume that $F_1, \dots, F_d \in \mathcal{L}^2_\eta(\mathbb{P})$ satisfy $DF_i \in \mathcal{L}^2(\mathbb{P} \otimes \mu)$ and $\mathbb{E}[F_i] = 0$ for all $i \in \{1, \dots, d\}$. Let $\sigma_{ik} := \mathbb{E}[F_i F_k]$ and define the covariance matrix $\Sigma^2 = (\sigma_{ik})_{i,k=1}^d$. Let $Y \sim N_d(0, \Sigma^2)$ be a centred Gaussian random vector with covariance matrix Σ^2 and put $F := (F_1, \dots, F_d)$. Then*

$$d_3(F, Y) \le \sum_{i,k=1}^d (\gamma_1(F_i, F_k) + \gamma_2(F_i, F_k)) + \gamma_3,$$

where the term γ_3 is defined as

$$\gamma_3 := \sum_{i,j,k=1}^d \int_S \mathbb{E}\big[|D_z F_j D_z F_k|^{3/2} \wedge \|D_z F\|^{3/2}_{\mathbb{R}^d}\big]^{2/3} \mathbb{E}\big[|D_z F_i|^3\big]^{1/3} \mu(\mathrm{d}z). \tag{2.1}$$

Remark 2.2.

(i) The difference between Theorem 2.1 and [18, Theorem 1.1] lies in the term γ_3. We emphasize that the bound in [18] does not lead to a meaningful error bound in the application to heavy-tailed moving averages we consider in the next section as the corresponding γ_3-term in [18] would diverge. Similarly to the univariate case, the bound provided by Theorem 2.1 is much more suitable for our purposes as it leads to a reasonable error bound, which tends to zero, as the number of observations n there tends to infinity.

(ii) It is in principal possible to derive error bounds as in Theorem 2.1 for probability metrics different from the d_3-metric. Namely, assuming in addition that the covariance matrix Σ^2 is *positive definite*, one can deal with the d_2-distance used in [13] and even with the convex distance introduced and studied in [18]. Since the corresponding error bounds for these notions of distance become rather long and technical, we refrain from presenting results in this direction. Moreover, in our application in the next section it seems in general rather difficult to check whether or not the covariance matrix is positive definite. This is another reason for us considering only the d_3-distance.

(iii) We would like to point out that quantitative central limit theorems for random vectors of Poisson functionals having a *finite* Wiener–Itô chaos expansion with respect to the d_3-distance were obtained [7]. Specifically, random vectors of so-called Poisson U-statistics were considered in [7] together with applications in stochastic geometry to Poisson process of k-dimensional flat in $\mathbb{R}^n$. ◇

2.2 Asymptotic Normality of Multivariate Heavy-Tailed Moving Averages

Here, we present our application of the refined multivariate second-order Poincaré inequality formulated in the previous section. For this recall the set-up described in the introduction. Especially recall the definition of the random processes $(X_t^i)_{t\in\mathbb{R}}$, $i\in\{1,\dots,m\}$ from (1.2). Recall also that the exponents α_i control the memory of the processes X^i. Given the limit theory for heavy-tailed moving averages as developed in [8] it comes as no surprise that the smallest such α_i will be of dominating importance. Hence, we define

$$\underline{\alpha}=\min\{\alpha_1,\dots,\alpha_m\}.$$

Finally, by $C_b^2(\mathbb{R}^m,\mathbb{R}^d)$ we denote the space of bounded functions $f:\mathbb{R}^m\to\mathbb{R}^d$ which are twice continuously differentiable and have all partial derivatives up to order two bounded by some constant.

Theorem 2.3. *Fix $d,m\geq 1$. Let (X_t^i), $i=1,\dots,m$, be moving averages as in* (1.2) *with Lévy measure having density w satisfying* (1.1) *for some $\beta\in(0,2)$ and kernels g_i which satisfy* (1.3) *with $\alpha_i\beta>2$ and $\kappa_i>-1/\beta$. Let a function $f=(f_1,\dots,f_d)\in C_b^2(\mathbb{R}^m,\mathbb{R}^d)$ be given and consider $V_n(X;f)$ as at* (1.4) *based on f and $X=(X^1,\dots,X^m)$. Let $\Sigma_n=\mathrm{Cov}(V_n(X;f))^{1/2}$ denote a non-negative definite square root of the covariance matrix $\mathrm{Cov}(V_n(X;f))$ of the d-dimensional random vector $V_n(X;f)$. Then $\Sigma_n\to\Sigma=(\Sigma_{i,j})_{i,j=1}^d$, as $n\to\infty$, where, for*

$i,j \in \{1,\dots,d\}$,

$$\begin{aligned}\Sigma_{i,j}^2 = &\sum_{s=0}^{\infty} \mathrm{Cov}(f_i(X_s^1,\dots,X_s^m), f_j(X_0^1,\dots,X_0^m)) \\ &+ \sum_{s=1}^{\infty} \mathrm{Cov}(f_i(X_0^1,\dots,X_0^m), f_j(X_s^1,\dots,X_s^m)).\end{aligned} \tag{2.2}$$

Moreover, $V_n(X;f)$ converges in distribution, as $n \to \infty$, to a d-dimensional centred Gaussian random vector $Y \sim N_d(0,\Sigma^2)$ with covariance matrix Σ^2. More precisely, there exists a constant $C > 0$ such that

$$d_3(V_n(X;f), Y) \le C \begin{cases} n^{-1/2}, & \text{if } \underline{\alpha}\beta > 3, \\ n^{-1/2}\log(n), & \text{if } \underline{\alpha}\beta = 3, \\ n^{(2-\underline{\alpha}\beta)/2}, & \text{if } 2 < \underline{\alpha}\beta < 3. \end{cases}$$

Remark 2.4.

(i) We remark that in the special case $d = m = 1$ the order for the d_3-distance provided by Theorem 2.1 is precisely the same as that for the Wasserstein distance in [2].

(ii) Even for particular functions $f = (f_1,\dots,f_d)$, such as trigonometric functions, it seems to be a rather demanding task to check whether the covariance matrix Σ^2 is positive definite or not. Note in this context that even in the one-dimensional case $d = m = 1$ the question of whether the asymptotic variance constant is strictly positive or not is generally difficult. This is the reason why we are working with the d_3-distance in this paper, since more refined probability metric usually require positive definiteness of the covariance matrix, see Remark 2.2.

(iii) It is straightforward to modify the proof of Theorem 2.3 to the situation where $X = (X_1,\dots,X_m)$ for some fixed moving average $(X_t)_{t\in\mathbb{R}}$ as in (1.2) and where the kernel g satisfy

$$|g(x)| \le K(x^{\kappa}\mathbb{1}_{[0,a)}(x) + x^{-\alpha}\mathbb{1}_{[a,\infty)}(x))$$

for some constants $a,\alpha,K > 0$ and $\kappa \in \mathbb{R}$ such that $\alpha\beta > 2$ and $\kappa > -1/\beta$. In this case the kernel of $X^i = X_i$ is simply $g_i = g(i+\cdot)$. Choosing an appropriate functional f in $V_n(X;f)$, such as the empirical characteristic function of X, opens up the possibility of inference on $(X_t)_{t\in\mathbb{R}}$ based on not only the marginal distribution X_1 as in much of the previous literature, but also on the joint distribution $(X_1,\dots,X_m)$. ◇

As in [2], Theorem 2.1 can be applied to particular processes (X_t^i). We mention here the linear fractional stable noises, which may be regarded as heavy-tailed extensions of a fractional Brownian motion. Let L be a β-stable Lévy process with $\beta \in (0,2)$ and put

$$X_t^i := Y_t^i - Y_{t-1}^i \qquad \text{for } Y_t^i := \int_{-\infty}^{t} \left[(t-s)_+^{H_i-1/\beta} - (-s)_+^{H_i-1/\beta}\right] \mathrm{d}L_s,$$

where $H_1,\dots,H_m \in (0,1)$ (if $\beta = 1$ we additionally suppose that L is symmetric). In this case, $\alpha_i = 1 - H_i + 1/\beta$ for all $i \in \{1,\dots,m\}$ and the condition $\underline{\alpha}\beta > 2$ translates into

$\beta \in (1,2)$ and $\max\{H_1,\dots,H_m\} < 1 - \frac{1}{\beta}$. Note that since $\beta > 1$ we automatically have that $\underline{\alpha}\beta < 3$. In this set-up the bound in Theorem 2.1 reads as follows:

$$d_3(V_n(X;f),Y) \le C n^{1/2-\beta(1-\max\{H_1,\dots,H_m\})/2}.$$

As a second application we mention a stable Ornstein–Uhlenbeck process. Again, for a β-stable Lévy process L with $\beta \in (0,2)$ define for $i \in \{1,\dots,m\}$,

$$X_t^i := \int_{-\infty}^t \mathrm{e}^{-\lambda_i(t-s)}\,\mathrm{d}L_s,$$

where $\lambda_1,\dots,\lambda_m > 0$. In this case, the parameters $\alpha_1,\dots,\alpha_m$ may be arbitrary and the error bound in Theorem 2.1 reduces to

$$d_3(V_n(X;f),Y) \le C\, n^{-1/2}.$$

In a similar spirit, one may consider multivariate quantitative central limit theorems for functionals of linear fractional Lévy noises or of stable fractional ARIMA processes, see [2] for the corresponding one-dimensional situations.

3 Background Material

3.1 Malliavin Calculus on Poisson Spaces

To take advantage of the powerful Malliavin–Stein method we need to recall some background material regarding the Malliavin formalism on Poisson spaces. For further details we refer to [6, 5, 11].

Throughout this section η denotes a Poisson process with intensity measure μ defined on some measurable space $(S,\mathcal{S})$ and over some probability space $(\Omega,\mathcal{F},\mathbb{P})$. We start by recalling that any $F \in \mathcal{L}^2_\eta(\mathbb{P})$ admits a chaos expansion (with convergence in $\mathcal{L}^2(\mathbb{P})$). That is,

$$F = \sum_{n=0}^{\infty} I_n(f_n), \tag{3.1}$$

where I_n denotes the nth order Wiener–Itô integral with respect to the compensated Poisson process $\eta - \mu$ and the kernels $f_n \in \mathcal{L}^2(\mu^n)$ are symmetric functions (i.e. they are invariant under permutations of its variables). Especially, $I_0(c) = c$ for all $c \in \mathbb{R}$.

The Kabanov–Skorohod integral δ is defined for a subclass of random processes $u \in \mathcal{L}^2(\mathbb{P} \otimes \mu)$ having chaotic decomposition

$$u(z) = \sum_{n=0}^{\infty} I_n(h_n(\,\cdot\,,z)),$$

where for each $z \in S$ the function $h_n(\cdot,z)$ is symmetric and belongs to $\mathcal{L}^2(\mu^n)$. Denoting by $\tilde{h}$ the canonical symmetrisation of a function $h : S^n \to \mathbb{R}$, i.e.

$$\tilde{h}(z_1,\dots,z_n) = \frac{1}{n!}\sum_{\sigma \in S_n} h(z_{\sigma(1)},\dots,z_{\sigma(n)}),$$

with S_n being the group of all permutations of $\{1,\dots,n\}$, we put

$$\delta(u) := \sum_{n=0}^{\infty} I_{n+1}(\tilde{h}_n),$$

whenever $\sum_{n=0}^{\infty}(n+1)\|\tilde{h}_n\|^2_{\mathcal{L}^2(\mu^{n+1})} < \infty$ (we indicate this by writing $u \in \operatorname{dom}\delta$), where $\|\cdot\|_{\mathcal{L}^2(\mu^{n+1})}$ denotes the usual $\mathcal{L}^2$-norm with respect to μ^{n+1}.

Next, we shall define the two operators $L : \operatorname{dom} L \to \mathcal{L}^2_\eta(\mathbb{P})$ and $L^{-1} : \mathcal{L}^2_\eta(\mathbb{P}) \to \mathcal{L}^2_\eta(\mathbb{P})$, where $\operatorname{dom} L$ denotes the class of Poisson functionals $F \in \mathcal{L}^2_\eta(\mathbb{P})$ with chaos expansion as in (3.1) satisfying $\sum_{n=1}^{\infty} n^2 n! \|f_n\|^2_{\mathcal{L}^2(\mu^n)} < \infty$. Then, we define

$$LF := -\sum_{n=1}^{\infty} n I_n(f_n).$$

Similarly, the pseudo-inverse L^{-1} of L acts on centred $F \in \mathcal{L}^2_\eta(\mathbb{P})$ with chaotic expansion (3.1) as follows:

$$L^{-1}F := \sum_{n=1}^{\infty} \frac{1}{n} I_n(f_n).$$

Finally, we recall that for $F \in \mathcal{L}^2_\eta(\mathbb{P})$ with chaotic expansion (3.1) satisfying $\sum_{n=0}^{\infty}(n+1)!\|f_n\|^2_{\mathcal{L}^2(\mu^n)} < \infty$ the Malliavin derivative admits the representation

$$D_z F = \sum_{n=1}^{\infty} n I_{n-1}(f_n(\,\cdot\,, z)), \qquad z \in S.$$

Using these definitions and representations, one may prove the following crucial formulas and relationships of Malliavin calculus, which also play a prominent role in our approach:

(i) $LL^{-1}F = F$ if F is centred.

(ii) $LF = -\delta D F$ for $F \in \operatorname{dom} L$.

(iii) $\mathbb{E}[F\delta(u)] = \mathbb{E}[\int_S (D_z F) u(z)\,\mu(\mathrm{d}z)]$, when $u \in \operatorname{dom}\delta$.

3.2 Multivariate Normal Approximation by Stein's Method

Stein's method for multivariate normal approximation is a powerful device to prove quantitative multivariate central limit theorems. The proof of Theorem 2.1 is based on the following result, which is known as Stein's Lemma (see [10, Lemma 4.1.3]). To present it, let us recall that the Hilbert–Schmidt inner product between two $d \times d$ matrices $A = (a_{ik})$ and $B = (b_{ik})$ is defined as

$$\langle A, B\rangle_{\mathrm{HS}} = \mathrm{Tr}(B^\top A) = \sum_{i,k=1}^{d} b_{ki} a_{ki}.$$

Moreover, for a differentiable function $\varphi : \mathbb{R}^d \to \mathbb{R}$ we shall write $\nabla\varphi$ for the gradient and $\nabla^2\varphi$ for the Hessian of φ. Also, we let $\langle\cdot,\cdot\rangle_{\mathbb{R}^d}$ denote the Euclidean scalar product in $\mathbb{R}^d$.

Lemma 3.1 (Stein's Lemma).
Let $\Sigma^2 \in \mathbb{R}^{d\times d}$ be a positive semi-definite matrix and Y be a d-dimensional random vector. Then $Y \sim N_d(0, \Sigma^2)$ if and only if it for all twice continuously differentiable functions $\varphi : \mathbb{R}^d \to \mathbb{R}$ with bounded derivatives one has that

$$\mathbb{E}[\langle Y, \nabla\varphi(Y)\rangle_{\mathbb{R}^d} - \langle \Sigma^2, \nabla^2\varphi(Y)\rangle_{\mathrm{HS}}] = 0.$$

4 Proof of Theorem 2.1

By definition of the d_3-distance we need to prove that

$$|\mathbb{E}[\varphi(Y)] - \mathbb{E}[\varphi(F)]| \leq \sum_{i,k=1}^{d} (\gamma_1(F_i, F_k) + \gamma_2(F_i, F_k)) + \gamma_3$$

for every function $\varphi \in \mathcal{H}_3$. For this, we may assume that Y and F are independent. We start out by applying the interpolation technique already demonstrated in [13]. Consider the function $\Psi : [0,1] \to \mathbb{R}$ given by

$$\Psi(t) := \mathbb{E}[\varphi(\sqrt{1-t}F + \sqrt{t}Y)], \qquad t \in [0,1].$$

Note that from the mean value theorem it follows that

$$|\mathbb{E}[\varphi(Y)] - \mathbb{E}[\varphi(F)]| = |\Psi(1) - \Psi(0)| \leq \sup_{t\in(0,1)} |\Psi'(t)|.$$

Hence it is enough to consider Ψ', which is given by

$$\Psi'(t) = \mathbb{E}\Big[\Big\langle \nabla\varphi(\sqrt{1-t}F + \sqrt{t}Y), \tfrac{1}{2\sqrt{t}}Y - \tfrac{1}{2\sqrt{1-t}}F\Big\rangle_{\mathbb{R}^d}\Big] =: \tfrac{1}{2\sqrt{t}}T_1 - \tfrac{1}{2\sqrt{1-t}}T_2.$$

We consider the two terms T_1 and T_2 separately. For T_1 it follows first by independence of F and Y and Stein's Lemma (used on the function $y \mapsto \varphi(\sqrt{1-t}a + \sqrt{t}y)$ and then dividing by $\sqrt{t}$) that

$$\begin{aligned} T_1 &= \mathbb{E}\big[\langle \nabla\varphi(\sqrt{1-t}F + \sqrt{t}Y), Y\rangle_{\mathbb{R}^d}\big] \\ &= \mathbb{E}\Big[\mathbb{E}[\langle \nabla\varphi(\sqrt{1-t}a + \sqrt{t}Y), Y\rangle_{\mathbb{R}^d}]\,|_{a=F}\Big] \\ &= \sqrt{t}\,\mathbb{E}\Big[\mathbb{E}[\langle \Sigma^2, \nabla^2\varphi(\sqrt{1-t}a + \sqrt{t}Y)\rangle_{\mathrm{HS}}]\,|_{a=F}\Big]. \end{aligned}$$

Let $\partial_i f$ denote the derivative of f in the ith coordinate. We have by independence of F and Y and the Malliavin rules (i)–(iii) phrased at the end of Section 3.1 that

$$\begin{aligned} T_2 &= \mathbb{E}\big[\langle \nabla\varphi(\sqrt{1-t}F + \sqrt{t}Y), F\rangle_{\mathbb{R}^d}\big] = \sum_{i=1}^{d} \mathbb{E}\Big[\mathbb{E}[\partial_i\varphi(\sqrt{1-t}F + \sqrt{t}a)F_i]\,|_{a=Y}\Big] \\ &= \sum_{i=1}^{d} \mathbb{E}\Big[\mathbb{E}[\partial_i\varphi(\sqrt{1-t}F + \sqrt{t}a)L(L^{-1}F_i)]\,|_{a=Y}\Big] \\ &= -\sum_{i=1}^{d} \mathbb{E}\Big[\mathbb{E}[\partial_i\varphi(\sqrt{1-t}F + \sqrt{t}a)\delta(DL^{-1}F_i)]\,|_{a=Y}\Big] \\ &= \sum_{i=1}^{d} \mathbb{E}\Big[\mathbb{E}[\langle D\partial_i\varphi(\sqrt{1-t}F + \sqrt{t}a), -DL^{-1}(F_i)\rangle_{\mathcal{L}^2(\mu)}]\,|_{a=Y}\Big]. \end{aligned}$$

Consider now the function $\varphi_i^{t,a} : \mathbb{R}^d \to \mathbb{R}$ defined by

$$\varphi_i^{t,a}(x) := \partial_i\varphi(\sqrt{1-t}x + \sqrt{t}a).$$

By Taylor expansion we can write

$$D_z\varphi_i^{t,a}(F) = \sum_{k=1}^{d} \partial_k\varphi_i^{t,a}(F)(D_zF_k) + R_i^a(D_zF)$$

for any $z \in \mathbb{R}^d$, where the remainder term $R_i^a(D_zF) = \sum_{j,k=1}^d R_{i,j,k}^a(D_zF_k, D_zF_j)$ satisfies the estimate

$$\begin{aligned}|R_{i,j,k}^a(x,y)| &\le \tfrac{1}{2}|xy| \max_{k,l} \sup_{x\in\mathbb{R}^d} \left|\partial_{k,l}\varphi_i^{t,a}(x)\right| \\ &\le \tfrac{1}{2}|xy|(1-t)\max_{k,l}\sup_{x\in\mathbb{R}^d}\left|\partial_{i,k,l}\varphi(\sqrt{1-t}x+\sqrt{t}a)\right| \\ &\le \tfrac{1}{2}(1-t)|xy|.\end{aligned} \tag{4.1}$$

Here, we have used the definition of the class $\mathcal{H}_3$. On the other hand, the remainder term also satisfies the inequality

$$\begin{aligned}\left|D_z\varphi_i^{t,a}(F) - \sum_{k=1}^d \partial_k\varphi_i^{t,a}(F)(D_zF_k)\right| &\le |D_z\varphi_i^{t,a}(F)| + |\langle\nabla\varphi_i^{t,a}(F), D_zF\rangle_{\mathbb{R}^d}| \\ &\le 2\,\|\nabla\varphi_i^{t,a}(F)\|_{\mathbb{R}^d}\,\|D_zF\|_{\mathbb{R}^d} \\ &\le 2\sqrt{1-t}\,\|D_zF\|_{\mathbb{R}^d},\end{aligned} \tag{4.2}$$

where we used again the mean value theorem and the Cauchy–Schwarz inequality. We may thus rewrite T_2 as

$$\begin{aligned}T_2 &= \sum_{i,k=1}^d \mathbb{E}\big[\mathbb{E}[\langle\partial_k\varphi_i^{t,a}(F)(DF_k), -DL^{-1}(F_i)\rangle_{\mathcal{L}^2(\mu)}]\,|_{a=Y}\big] \\ &\quad + \sum_{i=1}^d \mathbb{E}\big[\mathbb{E}[\langle R_i^a(DF), -DL^{-1}(F_i)\rangle_{\mathcal{L}^2(\mu)}]\,|_{a=Y}\big] \\ &= \sqrt{1-t}\sum_{i,k=1}^d \mathbb{E}\big[\partial_{k,i}\varphi(\sqrt{1-t}F+\sqrt{t}Y)\langle DF_k, -DL^{-1}(F_i)\rangle_{\mathcal{L}^2(\mu)}\big] \\ &\quad + \sum_{i=1}^d \mathbb{E}\big[\mathbb{E}[\langle R_i^a(DF), -DL^{-1}(F_i)\rangle_{\mathcal{L}^2(\mu)}]\,|_{a=Y}\big].\end{aligned}$$

From this together with the Cauchy–Schwarz inequality and the bounds (4.1) and (4.2) it follows that

$$\begin{aligned}&|\mathbb{E}[\varphi(Y)] - \mathbb{E}[\varphi(F)]| \le \sup_{t\in(0,1)} |\Psi'(t)| \\ &\le \sup_{t\in(0,1)} \frac{1}{2}\sum_{i,k=1}^d \mathbb{E}\Big[\big|\partial_{i,k}\varphi(\sqrt{1-t}F+\sqrt{t}X)\big|\,\big|\sigma_{ik} - \langle DF_k, -DL^{-1}(F_i)\rangle_{\mathcal{L}^2(\mu)}\big|\Big] \\ &\quad + \sup_{t\in(0,1)} \frac{1}{2\sqrt{1-t}}\sum_{i=1}^d \mathbb{E}\big[|\langle R_i^a(DF), -DL^{-1}(F_i)\rangle_{\mathcal{L}^2(\mu)}|\,|_{a=Y}\big] \\ &\le \frac{1}{2}\sum_{i,k=1}^d \mathbb{E}\big[|\sigma_{ik} - \langle DF_k, -DL^{-1}F_i\rangle_{\mathcal{L}^2(\mu)}|\big] \\ &\quad + \sum_{i,j,k=1}^d \int_S \mathbb{E}\big[(|D_zF_jD_zF_k| \wedge \|D_zF\|_{\mathbb{R}^d})\,|D_zL^{-1}F_i|\big]\,\mu(\mathrm{d}z).\end{aligned}$$

Applying now Proposition 4.1 in [4] to the first of these terms yields the inequality

$$\sum_{i,k=1}^d \mathbb{E}\Big[\big|\sigma_{ik} - \langle DF_k, -DL^{-1}F_i\rangle_{\mathcal{L}^2(\mu)}\big|\Big] \le 2\sum_{i,k=1}^d (\gamma_{1,i,k} + \gamma_{2,i,k}).$$

For the remainder term we deduce by Hölder's inequality with exponents 3 and 3/2 that

$$
\begin{aligned}
&\int_S \mathbb{E}\big[\big(|D_zF_jD_zF_k| \wedge \|D_zF\|_{\mathbb{R}^d}\big)|D_zL^{-1}F_i|\big]\,\mu(\mathrm{d}z) \\
&\quad\le \int_S \mathbb{E}\big[\big(|D_zF_jD_zF_k| \wedge \|D_zF\|_{\mathbb{R}^d}\big)^{3/2}\big]^{2/3}\mathbb{E}\big[|D_zL^{-1}F_i|^3\big]^{1/3}\,\mu(\mathrm{d}z) \\
&\quad\le \int_S \mathbb{E}\big[|D_zF_jD_zF_k|^{3/2} \wedge \|D_zF\|_{\mathbb{R}^d}^{3/2}\big]^{2/3}\mathbb{E}\big[|D_zF_i|^3\big]^{1/3}\,\mu(\mathrm{d}z),
\end{aligned}
$$

where we also used the contraction inequality $\mathbb{E}[|D_zL^{-1}F_i|^p] \le \mathbb{E}[|D_zF_i|^p]$ from [4, Lemma 3.4], which is valid for all $p \ge 1$ and $z \in \mathbb{R}^d$. This completes the proof of Theorem 2.1. □

5 Proof of Theorem 2.3

In order to apply Theorem 2.1 we need to ensure that the processes (X_t^i) can be represented in terms of a Poisson process. Indeed, following [16] and [2] we can represent X^i as the integral

$$X_t^i = \int_{\mathbb{R}^2} g_i(t-s)x\Big(\eta(\mathrm{d}s,\mathrm{d}x) - \tau(g_i(t-s)x)\,\mathrm{d}s\,\nu(\mathrm{d}x)\Big) + \tilde{b}_i,$$

with

$$\tilde{b}_i := \int_{\mathbb{R}}\Big(g_i(s)b + \int_{\mathbb{R}}(\tau(xg_i(s)) - g_i(s)\tau(x))\,\nu(\mathrm{d}x)\Big)\mathrm{d}s,$$

and where η is a Poisson process on $\mathbb{R}^2$ with intensity measure $\mu(\mathrm{d}s,\mathrm{d}x) := \mathrm{d}s\,\nu(\mathrm{d}x)$. Here ν is the Lévy measure of L, b the shift parameter in the characteristic triple for L_1 and τ is a truncation function, cf. (8.3)–(8.4) in [17].

In what follows, C will denote a strictly positive constant whose value might change from occasion to occasion.

5.1 Estimating the Malliavin Derivative

We start out by deriving simple estimates on the Malliavin derivative. By definition of the terms γ_1, γ_2, γ_3 introduced in Section 2.1 it is sufficient to consider the Malliavin derivatives of each of the coordinates of $f = (f_1,\dots,f_d)$ separately. So, let $i \in \{1,\dots,d\}$ and $z_j = (x_j,t_j) \in \mathbb{R}^2$ for $j \in \{1,2\}$ be given. Define for $z = (x,t) \in \mathbb{R}^2$ the vector $\delta_s(z)$, for $s \in \mathbb{R}$, as

$$\delta_s(z) := x(g_1(s-t),\dots,g_m(s-t)) \in \mathbb{R}^m. \tag{5.1}$$

The mean value theorem together with the Cauchy–Schwarz inequality and the assumption that $f_i \in \mathcal{C}_b^2(\mathbb{R}^m,\mathbb{R})$ then yield the existence of a constant $C > 0$ such that

$$
\begin{aligned}
|D_{z_1}f_i(X_s^1,\dots,X_s^m)| &= |f_i((X_s^1,\dots,X_s^m) + \delta_s(z_1)) - f_i(X_s^1,\dots,X_s^m)| \\
&\le C(1 \wedge \|\delta_s(z_1)\|_{\mathbb{R}^m}).
\end{aligned} \tag{5.2}
$$

Similarly, we deduce again by the mean value theorem and boundedness of f_i and its derivatives the following inequality for the iterated Malliavin derivative:

$$\begin{aligned}|D^2_{z_1,z_2} f_i(X_s^1,\dots,X_s^m)| = \Big| & f_i((X_s^1,\dots,X_s^m) + \delta_s(z_1) + \delta_s(z_2)) \\ & - f_i((X_s^1,\dots,X_s^m) + \delta_s(z_1)) \\ & - f_i((X_s^1,\dots,X_s^m) + \delta_s(z_2)) + f_i(X_s^1,\dots,X_s^m)\Big| \\ \le C(1\wedge \|\delta_s(z_1)\|_{\mathbb{R}^m}) & (1\wedge\|\delta_s(z_2)\|_{\mathbb{R}^m}).\end{aligned} \tag{5.3}$$

Note that the estimates (5.2) and (5.3) are purely deterministic and allow us to replace stochastic terms by deterministic estimates of the underlying kernels. This confirms in another context that many properties of moving averages can be deduced solely from the driving spectral density, see, for example, [3].

5.2 Analysing the Asymptotic Covariance Matrix

Define for each $k \in \mathbb{Z}$ and $i, j \in \{1,\dots,m\}$ the integral

$$\rho_{i,j,k} := \int_{\mathbb{R}} |g_i(x) g_j(x+k)|^{\beta/2}\,\mathrm{d}x \tag{5.4}$$

and observe that $\rho_{i,j,k} = \rho_{j,i,-k}$. Now, $\rho_{i,j,k}$ is closely related to the asymptotic covariances, which motivates the following technical lemma, which in turn leads to our assumption that $\alpha_i\beta > 2$ for any $i \in \{1,\dots,m\}$. In what follows we write $x \vee y := \max\{x,y\}$ for the maximum of $x, y \in \mathbb{R}$.

Lemma 5.1. *Let $k \in \mathbb{N}$ and $i, j \in \{1,\dots,m\}$. Then there is a constant $C > 0$ such that*

$$\rho_{i,j,k} \le C k^{-(\alpha_i\wedge\alpha_j)\beta/2}.$$

Proof. The same technique as in the proof of [2, Lemma 4.1] yields the bound $\rho_{i,j,k} \le Ck^{-\alpha_j\beta/2}$ and to obtain a bound symmetric in i and j observe that obviously

$$\rho_{i,j,k} \le C(k^{-\alpha_i\beta/2} \vee k^{-\alpha_j\beta/2}) = Ck^{-(\alpha_i\wedge\alpha_j)\beta/2}.$$

This completes the argument. □

Proposition 5.2. *The series defining $\Sigma^2_{i,j}$ in (2.2) is absolutely convergent and we have that $\Sigma_n^2 \to \Sigma^2$, as $n\to\infty$. In particular, $\Sigma_n \to \Sigma$.*

Proof. First, we prove that the series in (2.2) converges absolutely. By symmetry it is enough to show that

$$\sum_{s=1}^{\infty} \Big|\mathrm{Cov}(f_i(X_s^1,\dots,X_s^m), f_j(X_0^1,\dots,X_0^m))\Big| < \infty \qquad \text{for all } i,j \in \{1,\dots,d\}.$$

To this end, we let $\widetilde{\eta}$ be a Poisson process on $[0,1]\times\mathbb{R}^2$ with intensity measure $\mathrm{d}u\,\mu(\mathrm{d}z)$ with the property that $\eta = \widetilde{\eta}([0,1]\times\cdot)$. Using now the covariance identity for Poisson functionals from [5, Theorem 5.1] we conclude that

$$\begin{aligned}&\mathrm{Cov}(f_i(X_s^1,\dots,X_s^m), f_j(X_0^1,\dots,X_0^m)) \\ &\quad = \mathbb{E}\Big[\int_0^1 \Big(\int_{\mathbb{R}} \mathbb{E}[D_z f_i(X_s^1,\dots,X_s^m)\,|\,\mathcal{G}_u]\mathbb{E}[D_z f_j(X_0^1,\dots,X_0^m)\,|\,\mathcal{G}_u]\,\mu(\mathrm{d}z)\Big)\mathrm{d}u\Big],\end{aligned}$$

where $\mathcal{G}_u$ is the σ-algebra generated by the restriction of the Poisson process $\widetilde{\eta}$ to $[0,u]\times\mathbb{R}^2$. Applying the Cauchy–Schwarz inequality, our assumption (1.1) on ν and (5.2) implies that

$$
\begin{aligned}
&\Big|\mathrm{Cov}(f_i(X_s^1,\dots,X_s^m), f_j(X_0^1,\dots,X_0^m))\Big| \\
&\quad \le \int_0^1\Big(\int_{\mathbb{R}^2}\mathbb{E}\Big[\big|\mathbb{E}[D_z f_i(X_s^1,\dots,X_s^m)\,|\,\mathcal{G}_u]\mathbb{E}[D_z f_j(X_0^1,\dots,X_0^m)\,|\,\mathcal{G}_u]\big|\Big]\mu(\mathrm{d}z)\Big)\mathrm{d}u \\
&\quad \le \int_{\mathbb{R}^2}\mathbb{E}\big[|D_z f_i(X_s^1,\dots,X_s^m)|^2\big]^{1/2}\mathbb{E}\big[|D_z f_j(X_0^1,\dots,X_0^m)|^2\big]^{1/2}\mu(\mathrm{d}z) \\
&\quad \le C\int_{\mathbb{R}}\Big(\int_{\mathbb{R}}\Big(1\wedge|x|^2\,\|(g_\ell(s-t))_{\ell=1}^m\|_{\mathbb{R}^m}\,\|(g_\ell(-t))_{\ell=1}^m\|_{\mathbb{R}^m}\Big)|x|^{-1-\beta}\,\mathrm{d}x\Big)\mathrm{d}t \\
&\quad = C\int_{\mathbb{R}}\|(g_\ell(s-t))_{\ell=1}^m\|_{\mathbb{R}^m}^{\beta/2}\,\|(g_\ell(-t))_{\ell=1}^m\|_{\mathbb{R}^m}^{\beta/2}\,\mathrm{d}t \\
&\quad \le C\sum_{k,\ell=1}^m\int_{\mathbb{R}}|g_\ell(s-t)g_k(-t)|^{\beta/2}\,\mathrm{d}t \\
&\quad = C\sum_{k,\ell=1}^m\rho_{k,\ell,s} \\
&\quad \le Cs^{-\underline{\alpha}\beta/2},
\end{aligned}
$$

where the last inequality follows from Lemma 5.1. Since $\underline{\alpha}\beta>2$ by assumption the series in (2.2) converges absolutely. To deduce the convergence $\Sigma_n^2\to\Sigma^2$ we use the stationarity of the sequence $(X_t^1,\dots,X_t^m)$, $t\in\mathbb{R}$, to see that, for any $i,j\in\{1,\dots,d\}$,

$$
\begin{aligned}
&\mathrm{Cov}(V_n^i(X;f),V_n^j(X;f)) \\
&\quad = n^{-1}\sum_{s,t=1}^n\mathrm{Cov}(f_i(X_s^1,\dots,X_s^m),f_j(X_t^1,\dots,X_t^m)) \\
&\quad = n^{-1}\sum_{\substack{s,t=1\\ s\ge t}}^n\mathrm{Cov}(f_i(X_{s-t}^1,\dots,X_{s-t}^m),f_j(X_0^1,\dots,X_0^m)) \\
&\qquad + n^{-1}\sum_{\substack{s,t=1\\ s<t}}^n\mathrm{Cov}(f_i(X_0^1,\dots,X_0^m),f_j(X_{t-s}^1,\dots,X_{t-s}^m)) \\
&\quad = \sum_{k=0}^{n-1}(1-\tfrac{k}{n})\mathrm{Cov}(f_i(X_k^1,\dots,X_k^m),f_j(X_0^1,\dots,X_0^m)) \\
&\qquad + \sum_{k=1}^{n-1}(1-\tfrac{k}{n})\mathrm{Cov}(f_i(X_0^1,\dots,X_0^m),f_j(X_k^1,\dots,X_k^m)) \\
&\quad \longrightarrow \Sigma_{i,j}^2,
\end{aligned}
$$

as $n\to\infty$, where the convergence follows by Lebesgue's dominated convergence theorem together with the absolute convergence of the series defining the limit $\Sigma_{i,j}^2$. Finally, the last claim simply follows by continuity of the square root. □

5.3 Bounding $d_3(V_n,Y)$

Recall for $i,k\in\{1,\dots,m\}$ the definition of the quantities $\gamma_1(F_i,F_k)$ and $\gamma_2(F_i,F_k)$ from Section 2.1, which are applied with $F_i=V_n^i(X;f)$ and $F_k=V_n^k(X;f)$. According to

Theorem 2.1 we have that for any $n \in \mathbb{N}$,

$$d_3(V_n(X;f),Y) \le \sum_{i,k=1}^{d} (\gamma_1(F_i,F_k) + \gamma_2(F_i,F_k)) + \gamma_3,$$

where γ_3 is defined at (2.1). We consider each of these terms separately in the following three lemmas. Let us point to the fact that the sum will converge at a speed of order $1/\sqrt{n}$, whereas the γ_3-term will generally converge at a lower speed, depending on the parameters $\underline{\alpha}$ and β. It is also this last term that requires the stronger assumption (1.3) rather than just $\sum_{u=0}^{\infty} \rho_{i,j,u} < \infty$ for all $i,j \in \{1,\dots,m\}$. Indeed, as a product, in γ_3 we carefully have to distinguish between small and large values, where the latter are non-negligible for heavy-tailed moving averages.

Lemma 5.3. *There exists a constant $C > 0$ such that $\gamma_1(F_i,F_k) \le Cn^{-1/2}$ for any $i,k \in \{1,\dots,m\}$.*

Proof. To simplify the notation put $V_n^i := V_n^i(X;f)$ and recall that

$$\gamma_1^2(F_i,F_k) = 3\int_{(\mathbb{R}^2)^3} \mathbb{E}\big[(D^2_{z_1,z_3} V_n^i)^2 (D^2_{z_2,z_3} V_n^i)^2\big]^{1/2} \times \mathbb{E}\big[(D_{z_1} V_n^k)^2 (D_{z_2} V_n^k)^2\big]^{1/2} \mu^3(\mathrm{d}z_1,\mathrm{d}z_2,\mathrm{d}z_3).$$

If $z_i = (x_i,t_i) \in \mathbb{R}^2$ for $i \in \{1,2,3\}$, the integrand can be bounded using (5.2) and (5.3) as follows:

$$\begin{aligned}
&\mathbb{E}\big[(D^2_{z_1,z_3} V_n^i)^2 (D^2_{z_2,z_3} V_n^i)^2\big]^{1/2} \mathbb{E}\big[(D_{z_1} V_n^k)^2 (D_{z_2} V_n^k)^2\big]^{1/2} \\
&\quad \le \frac{C}{n^2}\Big(\sum_{s_1=1}^{n} \big(1 \wedge \|\delta_{s_1}(z_1)\|_{\mathbb{R}^m}\big)\big(1 \wedge \|\delta_{s_1}(z_3)\|_{\mathbb{R}^m}\big)\Big) \\
&\qquad \times \Big(\sum_{s_2=1}^{n} \big(1 \wedge \|\delta_{s_2}(z_2)\|_{\mathbb{R}^m}\big)\big(1 \wedge \|\delta_{s_2}(z_3)\|_{\mathbb{R}^m}\big)\Big) \\
&\qquad \times \Big(\sum_{s_3=1}^{n} \big(1 \wedge \|\delta_{s_3}(z_1)\|_{\mathbb{R}^m}\big)\Big)\Big(\sum_{s_4=1}^{n} \big(1 \wedge \|\delta_{s_4}(z_2)\|_{\mathbb{R}^m}\big)\Big) \\
&\quad \le \frac{C}{n^2} \sum_{s_1,\dots,s_4=1}^{n} \Big[\big(1 \wedge \|\delta_{s_1}(z_1)\|_{\mathbb{R}^m} \|\delta_{s_3}(z_1)\|_{\mathbb{R}^m}\big)\big(1 \wedge \|\delta_{s_2}(z_2)\|_{\mathbb{R}^m} \|\delta_{s_4}(z_2)\|_{\mathbb{R}^m}\big) \\
&\qquad \times \big(1 \wedge \|\delta_{s_1}(z_3)\|_{\mathbb{R}^m} \|\delta_{s_2}(z_3)\|_{\mathbb{R}^m}\big)\Big] \\
&\quad \le \frac{C}{n^2} \sum_{s_1,\dots,s_4=1}^{n} \sum_{j_1,\dots,j_6=1}^{m} \big(1 \wedge x_1^2 |g_{j_1}(s_1-t_1) g_{j_2}(s_3-t_1)|\big) \\
&\qquad \times \big(1 \wedge x_2^2 |g_{j_3}(s_2-t_2) g_{j_4}(s_4-t_2)|\big)\big(1 \wedge x_3^2 |g_{j_5}(s_1-t_3) g_{j_6}(s_2-t_3)|\big).
\end{aligned}$$

Using the substitution $u_i = x_i^2 y_i$ for $y_i > 0$, $i \in \{1,2,3\}$, one easily verifies the relation

$$\int_{\mathbb{R}^3} (1 \wedge x_1^2 y_1)(1 \wedge x_2^2 y_2)(1 \wedge x_3^2 y_3) |x_1 x_2 x_3|^{-1-\beta} \,\mathrm{d}x_1 \,\mathrm{d}x_2 \,\mathrm{d}x_3 = C y_1^{\beta/2} y_2^{\beta/2} y_3^{\beta/2},$$

for $\beta \in (0,2)$. This yields the bound

$$
\begin{aligned}
\gamma_1^2(F_i,F_k) &\le \frac{C}{n^2} \sum_{j_1,\dots,j_6=1}^{m} \sum_{s_1,\dots,s_4=1}^{n} \Big(\int_{\mathbb{R}} |g_{j_1}(s_1-t_1)g_{j_2}(s_3-t_1)|^{\beta/2}\,\mathrm{d}t_1 \\
&\qquad \times \int_{\mathbb{R}} |g_{j_3}(s_2-t_2)g_{j_4}(s_4-t_2)|^{\beta/2}\,\mathrm{d}t_2 \int_{\mathbb{R}} |g_{j_5}(s_1-t_3)g_{j_6}(s_2-t_3)|^{\beta/2}\,\mathrm{d}t_3 \Big) \\
&= \frac{C}{n^2} \sum_{j_1,\dots,j_6=1}^{m} \sum_{s_1,\dots,s_4=1}^{n} \rho_{j_1,j_2,s_3-s_1}\rho_{j_3,j_4,s_4-s_2}\rho_{j_5,j_6,s_2-s_1} \\
&\le \frac{C}{n} \sum_{j_1,\dots,j_6=1}^{m} \Big(\sum_{u=-n}^{n} \rho_{j_1,j_2,u} \Big)\Big(\sum_{u=-n}^{n} \rho_{j_3,j_4,u} \Big)\Big(\sum_{u=-n}^{n} \rho_{j_5,j_6,u} \Big) \le \frac{C}{n},
\end{aligned}
$$

where the penultimate inequality follows by substitution and the last inequality is due to Lemma 5.1, and where we used that $\sum_{u=0}^{\infty} \rho_{j,\ell,u} < \infty$ for all $j,\ell \in \{1,\dots,m\}$. □

Lemma 5.4. *There exists a constant $C > 0$ such that $\gamma_2(F_i,F_k) \le Cn^{-1/2}$ for all $i,k \in \{1,\dots,m\}$ and $n \in \mathbb{N}$.*

Proof. Using (5.3) we conclude that the integrand in the definition of $\gamma_2(F_i,F_k)$ is bounded as follows:

$$
\begin{aligned}
&\mathbb{E}\big[(D^2_{z_1,z_3}V_n^i)(D^2_{z_2,z_3}V_n^i)\big]^{1/2}\mathbb{E}\big[(D^2_{z_1,z_3}V_n^k)(D^2_{z_2,z_3}V_n^k)\big]^{1/2} \\
&\quad \le \frac{C}{n^2} \sum_{s_1,\dots,s_4=1}^{n} \Big(1 \wedge \|\delta_{s_1}(z_1)\|_{\mathbb{R}^m}\|\delta_{s_3}(z_1)\|_{\mathbb{R}^m}\Big)\Big(1 \wedge \|\delta_{s_2}(z_2)\|_{\mathbb{R}^m}\|\delta_{s_4}(z_2)\|_{\mathbb{R}^m}\Big) \\
&\qquad\qquad \times \Big(1 \wedge \|\delta_{s_1}(z_3)\|_{\mathbb{R}^m}\|\delta_{s_2}(z_3)\|_{\mathbb{R}^m}\|\delta_{s_3}(z_3)\|_{\mathbb{R}^m}\|\delta_{s_4}(z_3)\|_{\mathbb{R}^m}\Big) \\
&\quad \le \frac{C}{n^2} \sum_{s_1,\dots,s_4=1}^{n} \sum_{j_1,\dots,j_8=1}^{m} \Big(1 \wedge x_1^2 |g_{j_1}(s_1-t_1)g_{j_2}(s_3-t_1)|\Big) \\
&\qquad\qquad \times \Big(1 \wedge x_2^2 |g_{j_3}(s_2-t_2)g_{j_4}(s_4-t_2)|\Big) \\
&\qquad\qquad \times \Big(1 \wedge x_3^4 |g_{j_5}(s_1-t_3)g_{j_6}(s_2-t_3)g_{j_7}(s_3-t_3)g_{j_8}(s_4-t_3)|\Big).
\end{aligned}
$$

Moreover, as in the proof of the previous lemma we have that

$$
\int_{\mathbb{R}^3} (1 \wedge x_1^2 y_1)(1 \wedge x_2^2 y_2)(1 \wedge x_3^4 y_3)|x_1x_2x_3|^{-1-\beta}\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3 = C y_1^{\beta/2} y_2^{\beta/2} y_3^{\beta/4}
$$

for $\beta \in (0,2)$ and real numbers $y_1,y_2,y_3 > 0$. This implies that

$$
\begin{aligned}
\gamma_2^2(F_i,F_k) &\le \frac{C}{n^2} \sum_{j_1,\dots,j_8=1}^{m} \sum_{s_1,\dots,s_4=1}^{n} \int_{\mathbb{R}} |g_{j_1}(s_1-t_1)g_{j_2}(s_3-t_1)|^{\beta/2}\,\mathrm{d}t_1 \\
&\qquad \times \int_{\mathbb{R}} |g_{j_3}(s_2-t_2)g_{j_4}(s_4-t_2)|^{\beta/2}\,\mathrm{d}t_2 \\
&\qquad \times \int_{\mathbb{R}} |g_{j_5}(s_1-t_3)g_{j_6}(s_2-t_3)g_{j_7}(s_3-t_3)g_{j_8}(s_4-t_3)|^{\beta/4}\,\mathrm{d}t_3 \\
&\le \frac{C}{n^2} \sum_{j_1,\dots,j_8=1}^{m} \sum_{s_1,\dots,s_4=1}^{n} \rho_{j_1,j_2,s_3-s_1}\rho_{j_3,j_4,s_4-s_2}(\rho_{j_5,j_6,s_2-s_1} + \rho_{j_7,j_8,s_4-s_3}) \\
&\le \frac{C}{n},
\end{aligned}
$$

where the last inequality follows as in Lemma 5.3 and the penultimate inequality follows immediately from the fact that $|xy| \le x^2 + y^2$ for all $x,y \in \mathbb{R}$. □

Finally, we consider the crucial term γ_3.

Lemma 5.5. *There exists a constant $C > 0$ such that, for all $n \in \mathbb{N}$,*

$$\gamma_3 \leq C \begin{cases} n^{-1/2}, & \text{if } \underline{\alpha}\beta > 3, \\ n^{-1/2}\log(n), & \text{if } \underline{\alpha}\beta = 3, \\ n^{(2-\underline{\alpha}\beta)/2}, & \text{if } 2 < \underline{\alpha}\beta < 3. \end{cases}$$

Proof. Recall the definition of $\delta_s(z) = (\delta_s^1(z), \dots, \delta_s^m(z))$ from (5.1) and define for $i \in \{1, \dots, m\}$,

$$A_n^i(z) := \frac{1}{\sqrt{n}} \sum_{s=1}^n \Big(1 \wedge |\delta_s^i(z)|\Big).$$

By (5.2) and the sub-additivity of the minimum it follows that

$$\begin{aligned} \gamma_3 &\leq C \int_{\mathbb{R}^2} \Big(n^{-1/2} \sum_{s=1}^n 1 \wedge \|\delta_s(z)\|_{\mathbb{R}^m}\Big)^2 \wedge \Big(n^{-1/2} \sum_{s=1}^n 1 \wedge \|\delta_s(z)\|_{\mathbb{R}^m}\Big)^3 \mu(\mathrm{d}z) \\ &\leq \sum_{i,j=1}^m \int_{\mathbb{R}^2} \big(A_n^i(z)^2 \wedge A_n^j(z)^3\big) \mu(\mathrm{d}z). \end{aligned}$$

From this point on, we can literally follow the proof of Lemma 4.6 in [2]. In fact, this shows that for any $p \in [0,2]$, $q > 2$ and $i, j \in \{1, \dots, m\}$, one has that

$$\int_{\mathbb{R}^2} \big(A_n^i(z)^p \wedge A_n^j(z)^q\big) \mu(\mathrm{d}z) \leq C \begin{cases} n^{1-q/2}, & \text{if } \underline{\alpha}\beta > q, \\ n^{1-q/2}\log(n), & \text{if } \underline{\alpha}\beta = q, \\ n^{(2-\underline{\alpha}\beta)/2}, & \text{if } 2 < \underline{\alpha}\beta < 3. \end{cases}$$

Indeed, these bounds rely solely on the tail behaviour (in terms of the α_i's) of the kernels g_i, where $\underline{\alpha}$ reflects the *weakest* behaviour, and the power behaviour (in terms of the κ_i's) around 0, all of which satisfies the condition $\kappa_i > -1/\beta$. This completes the argument. □

Proof of Theorem 2.3. According to Theorem 2.1 we have that for any $n \in \mathbb{N}$,

$$d_3(V_n(X; f), Y) \leq \sum_{i,k=1}^d (\gamma_1(F_i, F_k) + \gamma_2(F_i, F_k)) + \gamma_3.$$

Using now Lemmas 5.3, 5.4 and 5.5 we see that

$$\begin{aligned} d_3(V_n(X; f), Y) &\leq C(n^{-1/2} + n^{-1/2}) + C \begin{cases} n^{-1/2}, & \text{if } \underline{\alpha}\beta > 3, \\ n^{-1/2}\log(n), & \text{if } \underline{\alpha}\beta = 3, \\ n^{(2-\underline{\alpha}\beta)/2}, & \text{if } 2 < \underline{\alpha}\beta < 3, \end{cases} \\ &\leq C \begin{cases} n^{-1/2}, & \text{if } \underline{\alpha}\beta > 3, \\ n^{-1/2}\log(n), & \text{if } \underline{\alpha}\beta = 3, \\ n^{(2-\underline{\alpha}\beta)/2}, & \text{if } 2 < \underline{\alpha}\beta < 3. \end{cases} \end{aligned}$$

This completes the argument. □

Acknowledgement

We would like to thank Mark Podolskij for initiating this collaboration. MML acknowledges financial support from the project 'Ambit fields: probabilistic properties and statistical inference' financed by the Villum Fonden.

References

[1] Andreas Basse-O'Connor, Claudio Heinrich and Mark Podolskij (2019). »*On limit theory for functionals of stationary increments Lévy driven moving averages*«. Electronic Journal of Probability 24(79), 42 pp. doi: 10.1214/19-EJP336.

[2] Andreas Basse-O'Connor, Mark Podolskij and Christoph Thäle (2019). »*A Berry–Esseén theorem for partial sums of functionals of heavy-tailed moving averages*«. arXiv: 1904.06065.

[3] Norio Kôno and Makoto Maejima (1991). »*Self-similar stable processes with stationary increments*«. In: »*Stable processes and related topics: A selection of papers from the Mathematical Sciences Institute Workshop, January 9–13, 1990*«. Edited by Stamatis Cambanus, Gennady Samorodnitsky and Murad S. Taqqu. Birkhäuser Boston, 275–295. isbn: 978-1-4684-6778-9. doi: 10.1007/978-1-46 84-6778-9_13.

[4] Günter Last, Giovanni Peccati and Matthias Schulte (2016). »*Normal approximation on Poisson spaces: Mehler's formula, second order Poincaré inequalities and stabilization*«. Probability Theory and Related Fields 165(3), 667–723. doi: 10.1007/s00440-015-0643-7.

[5] Günter Last and Mathew D. Penrose (2011). »*Poisson process Fock space representation, chaos expansion and covariance inequalities*«. Probability Theory and Related Fields 150(3–4), 663–690. doi: 10.1007/s00440-010-0288-5.

[6] Günter Last and Mathew D. Penrose (2017). »*Lectures on the Poisson process*«. Institute of Mathematical Statistics Textbooks 7. Cambridge University Press. isbn: 978-1-316-10447-7. doi: 10.1017/9781316104477.

[7] Günter Last, Mathew D. Penrose, Matthias Schulte and Christoph Thäle (2014). »*Moments and central limit theorems for some multivariate Poisson functionals*«. Advances in Applied Probability 46(2), 348–364. doi: 10.1239/aap/14013696 98.

[8] Mathias Mørck Ljungdahl and Mark Podolskij (2018). »*A limit theorem for a class of stationary increments Lévy moving averages processes with multiple singularities*«. Modern Stochastics: Theory and Applications 5(3), 297–316. doi: 10.15559/18-VMSTA111.

[9] Mathias Mørck Ljungdahl and Mark Podolskij (2020). »*A minimal contrast estimator for the linear fractional motion*«. Statistical Inference for Stochastic Processes 23, 381–413. doi: 10.1007/s11203-020-09216-2.

[10] Ivan Nourdin and Giovanni Peccati (2012). »*Normal approximation with Malliavin calculus: From Stein's method to universality*«. Cambridge Tracts in Mathematics 192. Cambridge University Press. ISBN: 978-1-139-08465-9. DOI: 10.1017/CBO9781139084659.

[11] David Nualart and Eulalia Nualart (2018). »*Introduction to Malliavin calculus*«. Institute of Mathematical Statistics Textbooks 9. Cambridge University Press. ISBN: 978-1-139-85648-5. DOI: 10.1017/9781139856485.

[12] Giovanni Peccati and Matthias Reitzner, editors (2016). »*Stochastic analysis for Poisson point processes: Malliavin calculus, Wiener–Itô chaos expansions and stochastic geometry*«. Bocconi & Springer Series 7. Springer International Publishing. ISBN: 978-3-319-05232-8. DOI: 10.1007/978-3-319-05233-5.

[13] Giovanni Peccati and Cengbo Zheng (2010). »*Multi-dimensional Gaussian fluctuations on the Poisson space*«. ELECTRONIC JOURNAL OF PROBABILITY 15(48), 1487–1527. DOI: 10.1214/EJP.v15-813.

[14] Vladas Pipiras and Murad S. Taqqu (2003). »*Central limit theorems for partial sums of bounded functionals of infinite-variance moving averages*«. BERNOULLI 9(5), 833–855. DOI: 10.3150/bj/1066418880.

[15] Balram S. Rajput and Jan Rosiński (1989). »*Spectral representations of infinitely divisible processes*«. PROBABILITY THEORY AND RELATED FIELDS 82(3), 451–487. DOI: 10.1007/BF00339998.

[16] Jan Rosiński (2018). »*Representations and isomorphism identities for infinitely divisible processes*«. THE ANNALS OF PROBABILITY 46(6), 3229–3274. DOI: 10.1214/17-AOP1246.

[17] Ken-Iti Sato (1999). »*Lévy processes and infinitely divisible distributions*«. Cambridge Studies in Advanced Mathematics 68. Cambridge University Press. ISBN: 978-0-521-55302-5.

[18] Matthias Schulte and Joseph E. Yukich (2019). »*Multivariate second order Poincaré inequalities for Poisson functionals*«. ELECTRONIC JOURNAL OF PROBABILITY 14(130), 1–42. DOI: 10.1214/19-EJP386.

[19] T. Watanabe (2007). »*Asymptotic estimates of multi-dimensional stable densities and their applications*«. TRANSACTIONS OF THE AMERICAN MATHEMATICAL SOCIETY 358(6), 2851–2879.

E Paper

Multi-Dimensional Parameter Estimation of Heavy-Tailed Moving Averages

Mathias Mørck Ljungdahl

Department of Mathematics, Aarhus University

Mark Podolskij

Department of Mathematics, University of Luxembourg

Abstract. In this paper we present a parametric estimation method for certain multi-parameter heavy-tailed Lévy-driven moving averages. The theory relies on recent multivariate central limit theorems obtained in [3] via Malliavin calculus on Poisson spaces. Our minimal contrast approach is related to the papers [15, 14], which propose to use the marginal empirical characteristic function to estimate the one-dimensional parameter of the kernel function and the stability index of the driving Lévy motion. We extend their work to allow for a multi-parametric framework that in particular includes the important examples of the linear fractional stable motion, the stable Ornstein–Uhlenbeck process, certain CARMA$(2,1)$ models and Ornstein–Uhlenbeck processes with a periodic component among other models. We present both the consistency and the associated central limit theorem of the minimal contrast estimator. Furthermore, we demonstrate numerical analysis to uncover the finite sample performance of our method.

Key words: Heavy tails, low frequency, Lévy processes, parametric estimation, limit theorems

AMS 2010 subject classifications: 62F12, 60F05, 60G51, 60G52

1 Introduction

Steadily through the last decades estimation procedures for various classes of continuous time moving averages and related processes have been proposed, see, e.g. [2, 11, 16] for estimation of the parameters in the linear fractional stable motion model and [9, 10] for the more general class of self-similar processes among many others. The bedrock of these techniques are of course the underlying limit theory for various functionals of the processes at hand. One such seminal paper is [18], which gives conditions for bounded functionals of a large class of moving averages and was later extended in [19] to certain unbounded functions. In a similar framework [5] gives an almost complete picture of the 'law of large numbers' for the classical case of the power variation functional. The article [4] extends the functionals from power variation to a large class of statistically interesting functionals and for a class of symmetric β-stable moving averages. This paper also provides an almost complete picture of the corresponding weak limit theorems, at least in the setting of Appell rank > 1 (such as is the case for power variation and the (real part) of the characteristic function).

Previous estimation methods suggested in [15, 14, 16] relied on functionals of the one-dimensional marginal law of the process and specific properties of the process at hand. Since the marginal distribution of the considered models have been symmetric β-stable, only the scale and the stability parameters can be estimated via such statistics. In particular, they are typically not sufficient to estimate kernel functions that depend on a multi-dimensional parameter. Indeed, this discrepancy is observed in [15], where the characteristic function of the one-dimensional law is not sufficient and instead the authors have to rely on a combination with other statistics to ensure estimation of all parameters. The aim of this paper is to construct estimators of the kernel function and the stability index in the general setting of a multi-dimensional parameter space. Instead of relying on existing theory [4, 5, 18], which only accounts for the marginal law of the underlying model, we shall use the framework from the recent paper [3], which is tailor-made for the study of Gaussian fluctuations of functionals of multiple heavy-tailed moving averages, to estimate the multi-dimensional parameter.

Let us now define the class of moving average processes for which the underlying limit theory applies. Let $L = (L_t)_{t\in\mathbb{R}}$ be a standard symmetric β-stable Lévy process and consider the model

$$X_t = \int_{-\infty}^{t} g(t-s)\,\mathrm{d}L_s, \qquad t \in \mathbb{R}, \tag{1.1}$$

for some measurable $g : \mathbb{R} \to \mathbb{R}$. Necessary and sufficient conditions for the integral to exist are given in [20] and we mention that in our setting a sufficient condition is $\int_{\mathbb{R}} |g(s)|^\beta\,\mathrm{d}s < \infty$. The kernel function g is assumed to have a power behaviour around 0 and at infinity. More specifically, we shall assume the existence of a constant $K > 0$ together with powers $\alpha > 0$ and $\kappa \in \mathbb{R}$ for which it holds

$$|g(x)| \le K\Big(x^\kappa \mathbb{1}_{[0,1)}(x) + x^{-\alpha}\mathbb{1}_{[1,\infty)}(x)\Big) \qquad \text{for all } x \in \mathbb{R}. \tag{1.2}$$

We are interested in (scaled) partial sums of multivariate functionals of the vectors $((X_{s+1},\dots,X_{s+m}))_{s\ge 0}$:

$$V_n(X;f) = \frac{1}{\sqrt{n}} \sum_{s=0}^{n-m} \Big(f(X_{s+1},\dots,X_{s+m}) - \mathbb{E}[f(X_1,\dots,X_m)] \Big), \tag{1.3}$$

where $f : \mathbb{R}^m \to \mathbb{R}^d$ is a suitable Borel function. Adhering to [3, Remark 2.4(iii)] the following result holds. Below $\mathcal{C}_b^2(\mathbb{R}^m, \mathbb{R}^d)$ denotes the space of twice differentiable functions $f : \mathbb{R}^m \to \mathbb{R}^d$ such that f and all of its first and second order derivatives are bounded and continuous.

Theorem 1.1 ([3, Theorem 2.3]).
Let $(X_t)_{t\in\mathbb{R}}$ *be a moving average as in* (1.1) *with kernel function* g *satisfying* (1.2)*. Assume that* $\alpha\beta > 2$ *and* $\kappa > -1/\beta$*. Let* $f = (f_1,\dots,f_d) \in \mathcal{C}_b^2(\mathbb{R}^m,\mathbb{R}^d)$ *and consider the statistic* $V_n(X;f)$ *introduced at* (1.3)*. Then as* $n\to\infty$

$$\Sigma_n^{i,j} := \mathrm{Cov}(V_n(X;f)) \to \Sigma^{i,j} := \sum_{s\in\mathbb{Z}} \mathrm{Cov}(f_i(X_{s+1},\dots,X_{s+m}), f_j(X_1,\dots,X_m)) \tag{1.4}$$

for any $1\le i,j\le d$*. Moreover,* $V_n(X;f) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0,\Sigma)$ *as* $n\to\infty$*.*

The paper [3] additionally provides Berry–Esseen type bounds for an appropriate distance between probability laws on $\mathbb{R}^d$, but Theorem 1.1 is sufficient for our statistical analysis. We remark that the limit theory for bounded f in the case of $m = 1$ and general $d\in\mathbb{N}$ is handled in [19], but it is actually the reverse situation, i.e. $m\in\mathbb{N}$ and $d = 1$, which we shall need. Specifically, f will be the empirical characteristic function of the joint distribution $(X_{s+1},\dots,X_{s+m})$, which then grants us the ability to estimate parameters which are not determined by the one-dimensional distribution of X_1, see Examples 2.4–2.8 below.

The paper is organized as follows. In Section 2 we introduce the parametric model, numerous assumptions and the main theoretical results of the paper, which show the strong consistency and the asymptotic normality of the minimal contrast estimator. Section 3 is devoted to a numerical analysis of the finite sample performance of our estimator. Finally, all proofs are collected in Section 4.

2 The Setting and Main Results

2.1 The Model and Assumptions

In the following we will consider a Lévy-driven moving average $X = (X_t)_{t\in\mathbb{R}}$ given by

$$X_t = \int_{\mathbb{R}} g_{\beta,\theta}(t-s)\,\mathrm{d}L_s, \qquad t\in\mathbb{R}, \tag{2.1}$$

where L is a symmetric β-stable Lévy process with scale parameter 1 and $\beta\in\Upsilon$ for some open subset $\Upsilon\subseteq(0,2)$, and $\{g_{\beta,\theta} \mid \beta\in\Upsilon,\theta\in\Theta\}$ is a measurable family of functions parametrized by an open subset $\Upsilon\times\Theta\subseteq(0,2)\times\mathbb{R}^d$ for some $d\ge 1$. For ease of notation we shall often denote the joint parameter with $\xi = (\beta,\theta)$ and the open subset by $\Xi = \Upsilon\times\Theta$.

The main goal of this section is to extend the theory of [14] from a one-dimensional parameter space, i.e. $d = 1$, to a general multi-dimensional theory. Such multi-dimensional parameter spaces include important examples of the linear fractional stable motion, the stable Ornstein–Uhlenbeck process, certain CARMA(2,1) models, and Ornstein–Uhlenbeck processes with a periodic component among others. One of the main difficulties in extending from $d = 1$ to $d\in\mathbb{N}$ is that, quite naturally, the parameters (β,θ) should be identifiable from the (theoretical) statistic, which in the case of [14] is the one-dimensional characteristic function:

$$\phi_{\beta,\theta}(u) = \mathbb{E}[\mathrm{e}^{\mathrm{i}uX_1}] = \exp(-\|ug_{\beta,\theta}\|_\beta^\beta).$$

This identification can very well be an unreasonable assumption if $d > 1$, see Example 2.4. But if we instead consider the characteristic function of the joint distribution $(X_1,\dots,X_m)$

$$\varphi^m_{\beta,\theta}(u_1,\dots,u_m) = \mathbb{E}\big[e^{\mathrm{i}\sum_{k=1}^m u_k X_k}\big] = \exp\Big(-\Big\|\sum_{k=1}^m u_k g_{\beta,\theta}(\,\cdot\,+k)\Big\|_\beta^\beta\Big), \tag{2.2}$$

such an identification may be possible. Let us discuss this in more details. The underlying stability index β is always identifiable from (2.2) since the stability index of a stable random variable is unique. The problem is then reduced to whether the parametrisation of the kernel $\theta \mapsto g_{\beta,\theta}$ specifies the distribution of X uniquely. The question now becomes a matter of uniqueness for the spectral representation of moving averages, which has been studied in, e.g. [21]. Translating the question to the characteristic functions of the finite dimensional distributions, $(X_1,\dots,X_m)$, $m \in \mathbb{N}$, we ask whether the β-norm of linear combinations of translations of the kernel specifies $g_{\beta,\theta}$ uniquely. This is known as *Kanter's theorem* in the literature and first appeared in [12], but for exposition sake let us repeat it here. Suppose $\beta \in (0,\infty)$ is not an even integer and let $g,h \in \mathcal{L}^\beta(\mathbb{R})$. Then Kanter's theorem states that if for all $n \in \mathbb{N}$ and $u_1,t_1,\dots,u_n,t_n \in \mathbb{R}$ it holds that

$$\Big\|\sum_{i=1}^n u_i g(\,\cdot\,+t_i)\Big\|_\beta^\beta = \Big\|\sum_{i=1}^n u_i h(\,\cdot\,+t_i)\Big\|_\beta^\beta,$$

then there exists an $\epsilon \in \{\pm 1\}$ and a $\tau \in \mathbb{R}$ such that $g = \epsilon h(\,\cdot\,+\tau)$ almost everywhere. Kanter's theorem then implies that the distribution of X is the same under θ and θ' if and only if there exists $\epsilon \in \{\pm 1\}$ and $\tau \in \mathbb{R}$ such that

$$\epsilon g_{\beta,\theta}(\,\cdot\,+\tau) = g_{\beta,\theta'} \qquad \text{almost everywhere.}$$

For many concrete examples of the kernel family $\{g_\xi \mid \xi \in \Xi\}$ it is often straightforward to check that such an identity only occurs if $\epsilon = 1$, $\tau = 0$ and $\theta = \theta'$.

Due to the preceding discussion it is reasonable to make the following assumptions on the family of kernels and we note that similar identification requirements are often explicitly or implicitly required in the literature. An important remark is that our theory allows for a general $m \in \mathbb{N}$ instead of only $m \in \{1,2\}$, where the statistics in the case $m = 2$ are often autocorrelations. We denote by $\partial_{z_1,z_2} f_\xi$ the partial derivative of f with respect to the parameters z_1, z_2 evaluated at $\xi \in \Xi$.

Assumption (A). There exists an $m \in \mathbb{N}$ such that:

(1) $0 < \|g_{\beta,\theta}\|_\beta < \infty$ for all $(\beta,\theta) \in \Upsilon \times \Theta$.

(2) The map $\theta \mapsto \varphi^m_{\beta,\theta}$ given in (2.2) is injective.

(3) The function $(\beta,\theta) \mapsto \|\sum_{i=1}^m u_i g_{\beta,\theta}(\,\cdot\,+i)\|_\beta^\beta$ is $C^2(\Upsilon \times \Theta)$ for each $u_1,\dots,u_m \in \mathbb{R}$.

(4) $u \mapsto \partial_\beta \varphi^m_\xi(u), \partial_{\theta_1}\varphi^m_\xi(u),\dots,\partial_{\theta_d}\varphi^m_\xi(u)$ are linearly independent continuous functions.

Let us give some remarks about the imposed conditions.

Remark 2.2.

(A) The assumption (A)(1) is a necessary and sufficient condition for X to be well-defined and non-degenerate. Moreover, (A)(1) makes it apparent why an explicit dependence on β of the kernel $g_{\beta,\theta}$ could be useful. This case of dependence is also necessary for some processes such as increments of the linear fractional stable motion, see Example 2.5 below.

(B) Condition (A)(2) is necessary to ensure that the model (2.1) is parametrized properly. Note that the non-existence of an $m \in \mathbb{N}$ such that (A)(2) holds would imply that the parameters could never be inferred from any *finite* data sample making the inference of θ impossible in practice. The identification of the parameters in a continuous time model from samples at equidistant time points is known in the literature as the *aliasing problem*.

(C) Condition (A)(3) is a minimal requirement for our method of proof (see also [14, Assumption (A)]). In particular, it ensures existence of the derivatives in (A)(4). $\diamond$

In order to use Theorem 1.1 we need to make additional assumptions on our kernel and for this we need to introduce some more notation. Consider a strictly positive weight function $w \in \mathcal{L}^1(\mathbb{R}_+^m)$ and define the weighted inner product and norms

$$\langle g,h\rangle_w = \int_{\mathbb{R}_+^m} g(x)h(x)w(x)\,\mathrm{d}x \qquad \text{and} \qquad \|h\|_{w,p}^p = \int_{\mathbb{R}_+^m} |h(x)|^p w(x)\,\mathrm{d}x, \qquad p \in \{1,2\}.$$

Let $\mathcal{L}_w^p(\mathbb{R}_+^m)$ denote the corresponding Banach $\mathcal{L}^p$-space of Borel functions.

Assumption (B).

(1) Assume that for all $(\beta,\theta) \in \Upsilon \times \Theta$ there exist $\kappa \in \mathbb{R}$ and $\alpha > 0$ such that $\kappa > -1/\beta$, $\alpha\beta > 2$ and (1.2) holds for $g_{\beta,\theta}$.

(2) The functions $u \mapsto |\partial_{\xi_i,\xi_k}\varphi_\xi(u)|, |\partial_{\xi_i}\varphi_\xi(u)|, i,k \in \{1,\dots,d+1\}$, are locally dominated in $\mathcal{L}_w^2(\mathbb{R}_+^m)$. That is, there exists for all $\xi \in \Xi$ a neighbourhood $\Xi_0 \ni \xi$ such that the supremum of these functions over $\xi \in \Xi_0$ are dominated by a function in $\mathcal{L}_w^2(\mathbb{R}_+^m)$.

Assumption (B)(1) is imposed to ensure that we may employ Theorem 1.1. While (B)(2) seems strict it is always satisfied in the one-dimensional case $m = 1$ and we shall need it to ensure validity of the implicit function theorem in our setup.

We now demonstrate some examples, which satisfy Assumption (A) for $m \geq 2$ but not for $m = 1$.

Example 2.4 (Stable Ornstein–Uhlenbeck process).
Let $(X_t)_{t\in\mathbb{R}}$ denote the β-stable Ornstein–Uhlenbeck process with parameter $\lambda > 0$ and scale parameter $\sigma > 0$. That is, $(X_t)_{t\in\mathbb{R}}$ is a stationary solution of the stochastic differential equation

$$\mathrm{d}X_t = -\lambda X_t\,\mathrm{d}t + \sigma\,\mathrm{d}L_t.$$

It has the representation (2.1) with kernel function $g_\theta(u) = \sigma\exp(-\lambda u)\mathbb{1}_{(0,\infty)}(u)$ and $\theta = (\sigma,\lambda) \in (0,\infty)^2$. It is clear that the one-dimensional characteristic function does not characterize the parameter θ, hence Assumption (A)(2) is not satisfied for $m = 1$. Consider therefore the case $m = 2$. Here the characteristic function is uniquely determined by θ if the β-norms are. Indeed, using the binomial series one may deduce the following formula:

$$\|u_1 g_\theta + u_2 g_\theta(\,\cdot\,+1)\|_\beta^\beta = \frac{\sigma^\beta}{\beta\lambda}\big[u_2^\beta(1-\mathrm{e}^{-\beta\lambda}) + (u_1 + u_2\mathrm{e}^{-\lambda})^\beta\big], \qquad u_1 > u_2 \geq 0.$$

It is then straightforward to check that these equations in $u_1 > u_2 \geq 0$ determine $\theta \in (0,\infty)^2$ uniquely. Additionally, (A)(4) can be checked in a manner similar to Example 2.6 below and we refer to Section 4.4 for the derivation of these statements.

There are a number of alternative estimation methods for a stable Ornstein–Uhlenbeck model. When the stability parameter β is known, λ can be estimated with convergence rate $(n/\log n)^{1/\beta}$ as it has been shown in [24]. In the discrete-time setting of the AR(1) model with heavy-tailed i.i.d. noise, it is known that a Gaussian limit can be obtained, cf. [13], but this method again lacks joint estimation with the parameter β. In a similar framework the paper [1] investigates the asymptotic behaviour of the maximum likelihood estimator. In particular, their results imply that the parameters σ and β can be estimated with a $\sqrt{n}$-precision, while the drift parameter λ has a faster convergence rate of $n^{1/\beta}$. ○

Example 2.5 (Linear fractional stable motion).
Let $(Y_t)_{t\in\mathbb{R}}$ be the linear fractional stable motion with self-similarity $H \in (0,1)$, stability index $\beta \in (0,2)$ and scale parameter $\sigma > 0$. That is,

$$Y_t = \int_{\mathbb{R}} \sigma\big[(t-s)_+^{H-1/\beta} - (-s)_+^{H-1/\beta}\big]\,\mathrm{d}L_s.$$

Consider the low frequency kth order increment at rate r ($k, r \in \mathbb{N}$) defined as

$$X_i := \Delta_{i,k}^r Y = \sum_{j=0}^{k} (-1)^j \binom{k}{j} Y_{i-rj}, \qquad i \geq rk.$$

If $r = 1$ or $k = 1$ we remove the corresponding index. In the case of $k = r = 1$ then $X_i = \Delta_i Y = Y_i - Y_{i-1}$ is simply the increments of Y and for $k = 2$ we have that

$$\Delta_{i,2}^r Y = Y_i - 2Y_{i-r} + Y_{i-2r}, \qquad i \geq 2r.$$

The corresponding kernel of X becomes

$$g_{\beta,H,\sigma}(u) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} (u - rj)_+^{H-1/\beta},$$

where $x_+ = x \vee 0$ is the positive part and $x_+^a := 0$ for all $x \leq 0$. We note the asymptotic behaviour

$$\frac{g_{\beta,H,\sigma}(u)}{K u^{H-1/\beta-k}} \longrightarrow 1 \qquad \text{as } u \to \infty$$

for some constant $K > 0$ depending on α, H and k. Hence, (1.2) holds with $\kappa = H - 1/\beta > -1/\beta$ and $\alpha = k + 1/\beta - H > 0$. In this case Assumption (B) can simply be translated into an assumption on the parameter space $\Upsilon \times \Theta$, e.g.

$$\Upsilon \times \Theta = \big\{(\beta, H, \sigma) \mid 0 < H < k - 1/\beta, 1/C < \sigma < C\big\},$$

for some arbitrary but finite constant $C > 0$. It is well-known that X has a version with continuous paths if and only if $H - 1/\beta > 0$, so if we want to do inference in the continuous case we have the two parameter inequalities:

$$0 < H - 1/\beta \qquad \text{and} \qquad H < k - 1/\beta. \tag{2.3}$$

Note that these inequalities never hold for $k = 1$. But they are always satisfied for $k \geq 2$, which shows the usefulness of higher order increments. Moreover, the H-self-similarity of X implies that

$$\frac{\mathbb{E}[|\Delta^2_{2k,k} Y|^p]}{\mathbb{E}[|\Delta_{k,k} Y|^p]} = 2^{pH} \qquad \text{for } p \in (-1, 0).$$

For $k = 2$ the term $\Delta^2_{4,2} Y$ is a linear combination of $\Delta_{2,2} Y$, $\Delta_{3,2} Y$, $\Delta_{4,2} Y$. Hence, H is identifiable from the characteristic function of the three-dimensional distribution (X_1, X_2, X_3), in other words, $m = 3$ in the case $k = 2$. ○

Example 2.6 (OU-type model with a periodic component).
The next example we consider is a periodic extension of the stable Ornstein–Uhlenbeck process from Example 2.4. Let $\theta = (\theta_1, \theta_2) \in (0, \infty)^2$ and consider the kernel function:

$$g_\theta(u) = \exp(-\theta_1 u - \theta_2 f(u)) \mathbb{1}_{(0,\infty)}(u), \qquad u \in \mathbb{R},$$

where $f : \mathbb{R} \to \mathbb{R}$ is a bounded measurable function which is either non-negative or non-positive and has period 1, i.e. $f(x+1) = f(x)$ for all x. If f does not vanish except on Lebesgue null set, then $\theta \mapsto \varphi^m_{\beta,\theta}$ for $m = 2$ is injective. If, in addition, f is negative then Assumption (B)(2) is satisfied except possibly at $\beta = 1$. We refer to Section 4.4 for the proof of these statements. ○

Example 2.7 (Modulated OU).
Consider the process X defined at (2.1) with kernel given by

$$g_\theta(s) = \theta_1 s \exp(-\theta_2 s) \mathbb{1}_{(0,\infty)}(s), \qquad s \in \mathbb{R}. \tag{2.4}$$

Under the assumptions on the parameters $\theta \in (0, \infty)^2$ and $\beta \in (1, 2)$ it is possible to prove that θ is not identifiable from $m = 1$ while it is in the case $m = 2$. We refer to Section 4.5 for the full exposition of these claims. ○

Example 2.8 (CARMA processes).
Consider integers $p > q$. The CARMA(p, q) process $(Y_t)_{t \in \mathbb{R}}$ with parameters $a_1, \ldots, a_p$, $b_0, \ldots, b_{q-1} \in \mathbb{R}$ driven by L is the solution to the stochastic differential equation

$$X_t = b^\top Y_t \qquad \text{with } \mathrm{d}Y_t - A Y_t \, \mathrm{d}t = e \, \mathrm{d}L_t, \tag{2.5}$$

where e and b are the p-dimensional column vectors given by

$$e = (0, \ldots, 0, 1)^\top \qquad \text{and} \qquad b = (b_0, \ldots, b_{p-1})^\top,$$

where $b_q = 1$ and $b_i = 0$ for all $q < i < p$ and A is the $p \times p$ matrix given by

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ -a_p & -a_{p-1} & a_{p-2} & \cdots & -a_1 \end{pmatrix}.$$

CARMA(p,q) processes fits within the framework of (2.1) since if the eigenvalues of A have strictly negative real part, then a unique stationary solution of (2.5) exists and is given by

$$X_t = \int_{\mathbb{R}} b^\top \mathrm{e}^{A(t-s)} e \mathbb{1}_{[0,\infty)}(t-s)\,\mathrm{d}L_s, \qquad t \in \mathbb{R},$$

see [7, Proposition 1]. In this example we discuss a specific three-dimensional subclass of CARMA$(2,1)$ processes, which corresponds to the choice $\lambda := -\sqrt{a_2}$ and $a_1 = 2\sqrt{a_2} = -2\lambda$. The parameter of interest becomes $\xi = (\beta, b_0, \lambda)$ and we further assume that $\beta \in (1,2)$ and $\theta := b_0 + \lambda > 0$. In this setting the matrix A is given by

$$A = \begin{pmatrix} 0 & 1 \\ -\lambda^2 & 2\lambda \end{pmatrix}$$

and $\lambda < 0$ is the only eigenvalue of A. We thus obtain the Jordan normal form

$$A = S\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} S^{-1}, \qquad S = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}, \qquad S^{-1} = \begin{pmatrix} 1 & 0 \\ -\lambda & 1 \end{pmatrix}.$$

Using this representation elementary matrix algebra yields the identity

$$g(s) = b^\top \exp(sA) e \mathbb{1}_{[0,\infty)}(s) = (1+\theta s)\exp(\lambda s)\mathbb{1}_{[0,\infty)}(s).$$

In Section 4.6 we show that the parameters of the model are identifiable in the case $m = 2$. ○

2.2 Parametric Estimation via Minimal Contrast Approach

We note first that the discrete time process $(X_t)_{t\in\mathbb{Z}}$ is ergodic according to [8], and so is the sequence

$$Y_i = f(X_{i+1}, \ldots, X_{i+m}), \qquad i \in \mathbb{Z},$$

for any measurable function f. Hence, we obtain by Birkhoff's ergodic theorem the strong consistency (of the real part) of the joint empirical characteristic function:

$$\begin{aligned} \varphi_n(u_1,\ldots,u_m) &= \frac{1}{n}\sum_{i=0}^{n-m} \cos\Big(\sum_{k=1}^{m} u_k X_{i+k}\Big) \\ &\xrightarrow{\text{a.s.}} \mathbb{E}\Big[\cos\Big(\sum_{k=0}^{m-1} u_k X_{1+k}\Big)\Big] = \varphi_\xi^m(u_1,\ldots,u_m), \end{aligned} \tag{2.6}$$

where $\xi = (\beta, \theta) \in \Xi$ denotes the unknown parameter of the model. To reduce cumbersome notation we drop the dependence on m in the characteristic function and simply write φ_ξ from now on. For a weight function w introduced in the previous section, we denote by $F : \mathcal{L}_w^2(\mathbb{R}_+^m) \times \Xi \to \mathbb{R}$ the map

$$F(\psi, \xi) = \|\psi - \varphi_\xi\|_{w,2}^2.$$

The minimal contrast estimator ξ_n of ξ is then defined as

$$\xi_n \in \operatorname*{argmin}_{\xi\in\Xi} F(\varphi_n, \xi) = \operatorname*{argmin}_{\xi\in\Xi} \int_{\mathbb{R}_+^m} (\varphi_n(u) - \varphi_\xi(u))^2 w(u)\,\mathrm{d}u, \tag{2.7}$$

and we remark that ξ_n can be chosen universally measurable by [23, Theorem 2.17(d)]. To obtain the asymptotic normality of the minimal contrast estimator ξ_n we will show a central limit theorem for the statistic $\sqrt{n}(\varphi_n(u_1,\dots,u_m) - \varphi_\xi(u_1,\dots,u_m))$ using Theorem 1.1 and then apply a functional version of the implicit function theorem. For this purpose we introduce a centred Gaussian field $(G_u)_{u\in\mathbb{R}_+^m}$ whose covariance kernel is defined as

$$\mathrm{Cov}(G_u, G_v) = \sum_{l\in\mathbb{Z}} \mathrm{Cov}\Big(\cos(\langle u, Z_0\rangle_{\mathbb{R}^m}), \cos(\langle v, Z_l\rangle_{\mathbb{R}^m})\Big), \tag{2.8}$$

where $Z_k = (X_{1+k},\dots,X_{m+k})$. The main theoretical result of the paper is the strong consistency and asymptotic normality of the minimal contrast estimator ξ_n.

Theorem 2.9. *Let* (ξ_n) *be the minimal contrast estimator at* (2.1) *associated with the true parameter* $\xi_0 = (\beta_0, \theta_0)$. *Suppose that Assumptions (A) and (B) hold for the underlying family of kernels* $(g_\xi)_{\xi\in\Xi}$. *Assume that the weight function* w *is continuous and* $\int_{\mathbb{R}_+^m} \|u\|_{\mathbb{R}^m}^2 w(u)\,\mathrm{d}u < \infty$.

(i) $\xi_n \to \xi_0$ *almost surely as* $n \to \infty$.

(ii) *The convergence as* $n \to \infty$

$$\sqrt{n}(\xi_n - \xi_0) \xrightarrow{\mathcal{L}} \big(\nabla_\xi^2 F(\varphi_{\xi_0}, \xi_0)\big)^{-1} \big(\langle \partial_{\xi_i} \varphi_{\xi_0}, G\rangle_w\big)_{i=1,\dots,d+1}$$

holds, where $G = (G_u)_{u\in\mathbb{R}_+^m}$ *is a continuous zero-mean Gaussian random field with covariance kernel defined by* (2.8). *In particular, the above limit is a normally distributed* $(d+1)$*-dimensional random vector.*

We note that due to Assumption (A)(4) the matrix $\nabla_\xi^2 F(\varphi_{\xi_0}, \xi_0)$ is invertible. In principle, the normal limit in Theorem 2.9 is explicit up to the knowledge of the parameter ξ_0, but due to the complex covariance kernel of the process G it is hard to apply the central limit theorem to obtain confidence regions. Instead one may use a parametric bootstrap approach as it has been suggested in [15, Section 4.2].

We remark that the convergence rate is $\sqrt{n}$ for all parameters. Due to the non-Markovian structure of the general model (2.1) it is a non-trivial task to assess the optimality of this rate. As we have discussed in Example 2.4 the rate $\sqrt{n}$ can be suboptimal in the particular case of the drift parameter in an Ornstein–Uhlenbeck model.

Remark 2.10 (Extension to general Lévy drivers).
If we drop the requirement for estimation of β we can consider a larger class of Lévy drivers. Indeed, according to [3] the statement of Theorem 1.1 still holds for a symmetric Lévy process L, which admits a Lévy density ν such that

$$\nu(x) \le C|x|^{-1-\beta} \qquad \text{for all } x \neq 0.$$

In this case the characteristic function takes on a more complicated form. Indeed, by [20, Theorem 2.7] it holds that

$$\mathbb{E}\big[e^{\mathrm{i}\langle u,(X_1,\dots,X_m)\rangle_{\mathbb{R}^m}}\big] = \exp\Big(\int_{\mathbb{R}}\int_{\mathbb{R}} [\cos(\langle u, x(g_\xi(z+i))_{i=0,\dots,m-1}\rangle_{\mathbb{R}^m}) - 1]\,\nu(\mathrm{d}x)\,\mathrm{d}z\Big).$$

In principle, the asymptotic theory of Theorem 2.9 can be extended to this more general setting. However, the proof of the asymptotic normality relies on the existence of a continuous modification of the random field $(G_u)_{u\in\mathbb{R}_+^m}$ and the behaviour of $\mathbb{E}[G_u^2]$ in $u \in \mathbb{R}_+^m$ (cf. Section 4.1), which requires a different treatment compared to the β-stable case. ◇

3 A Simulation Study

In this section we will demonstrate the finite sample performance of our estimator for three examples, which are supposed to highlight different aspects of the minimal contrast approach. First, we will consider the linear fractional stable motion (cf. Example 2.5) and use $m = 3$ to estimate the three-dimensional parameter of the model. The second example is the *generalized modulated OU-process*, which has not been shown to satisfy the main assumptions of the paper. We will use $m = 2$ to estimate the three-dimensional parametric model and test how our method works in this framework. The third model is the Ornstein–Uhlenbeck process considered in Example 2.4 with a fixed and known scale parameter σ. In this setting both $m = 1$ and $m = 2$ can be used to estimate the drift λ and the stability index β, and the aim of the numerical simulation is to test how the choice of higher index m affects the performance of the estimator.

Since the weight function w depends on m implicitly via its domain we need a function, which is reasonably compatible between different dimensions and we consider therefore throughout this study the m-dimensional Gaussian density with zero mean and a scaled unit covariance matrix $\nu^2 I_m$:

$$w_\nu(u) = (2\pi\nu^2)^{-m/2}\exp\Big(-\frac{\|u\|_{\mathbb{R}^m}^2}{2\nu^2}\Big), \qquad u \in \mathbb{R}^m,\ \nu > 0. \tag{3.1}$$

The choice of ν varies between the three example process and it is a subject for future research to automatically determine an optimal weight. For the computation of the weighted integral in (2.7) we use Gauss–Laguerra quadrature which is a weighted sum of function values and the number of weights will also vary depending on the process.

We note additionally that the minimization involved in computing the minimal contrast estimator at (2.7) has to be done numerically and for this we use the method of [17], which requires picking a starting point which naturally will depend on the example kernel at hand. Lastly, we remark that the β-norm of the kernel function is generally not known explicitly, hence the theoretical characteristic function is approximated as well.

All tables in this section are based on at least 200 Monte Carlo repetitions.

3.1 Linear Fractional Stable Motion

Recall from the discussion in Example 2.5 that it is prudent to take higher order increments, and we fix throughout $k = 2$. Moreover, to properly identify the parameters we consider the characteristic function of the three-dimensional joint distribution, hence $m = 3$. Next we consider throughout the weight function at (3.1) with standard deviation $\nu = 10$ and the weighted integral is approximated with $12^3 = 1728$ number of weights. The starting point for the minimization algortihm is $(\beta, H, \sigma) = (1.5, 0.5, 2)$.

The estimator is tested in the continuous case, so only parameter combinations resulting in the equality $H - 1/\beta > 0$ are considered. Table E.1 reports the bias and standard deviation in the case of $n = 1000$ for different parameter combinations, while Table E.2 explores the case $n = 10\,000$. We observe a rather good performance of all estimators with superior results in the setting $n = 10\,000$ as expected from our theoretical statements. We note that the estimator of the scale parameter σ performs the best, which is in line with earlier findings of [16].

Table E.1. Absolute value of bias (|Bias|) and standard deviation (Std) for $n = 1000$ and $\sigma = 0.3$ for the linear fractional stable motion.

		|Bias|			Std		
H	β	$\widehat{\beta}_n$	$\widehat{H}_n$	$\widehat{\sigma}_n$	$\widehat{\beta}_n$	$\widehat{H}_n$	$\widehat{\sigma}_n$
0.6	1.8	0.0176	0.0478	0.0362	0.1950	0.2730	0.0705
0.7	1.6	0.0705	0.1710	0.0834	0.2409	0.3581	0.0947
	1.8	0.0106	0.0041	0.0120	0.1766	0.2094	0.0429
0.8	1.4	0.0862	0.2444	0.0862	0.2348	0.3457	0.1044
	1.6	0.0250	0.0597	0.0270	0.1783	0.2466	0.0541
	1.8	0.0120	0.0060	0.0044	0.1452	0.1578	0.0287

Table E.2. Absolute value of bias (|Bias|) and standard deviation (Std) for $n = 10\,000$ and $\sigma = 0.3$ for the linear fractional stable motion.

		|Bias|			Std		
H	β	$\widehat{\beta}_n$	$\widehat{H}_n$	$\widehat{\sigma}_n$	$\widehat{\beta}_n$	$\widehat{H}_n$	$\widehat{\sigma}_n$
0.6	1.8	0.0133	0.0456	0.0254	0.1272	0.2007	0.0532
0.7	1.6	0.0238	0.0818	0.0331	0.1005	0.2147	0.0685
	1.8	0.0060	0.0147	0.0066	0.0869	0.1153	0.0173
0.8	1.4	0.0347	0.1536	0.0504	0.1095	0.2546	0.0865
	1.6	0.0078	0.0053	0.0008	0.0665	0.0843	0.0085
	1.8	0.0032	0.0020	0.0009	0.0597	0.0732	0.0067

3.2 Generalized Modulated OU Process

The generalized modulated OU process is defined via equation (2.1) with kernel function

$$g_\theta(s) = s^\sigma \exp(-\lambda s)\mathbb{1}_{(0,\infty)}(s), \qquad s \in \mathbb{R},$$

where $\theta = (\sigma, \lambda) \in (0,\infty)^2$. We recall that this class of kernels has not been shown to satisfy the main assumption of the paper, but it easily seen that $m = 1$ is not enough to identify the parameters in θ. We take $m = 2$ and increase the number of weights to 20, hence the weighted integral approximation is based on $20^2 = 400$ nodes. Moreover, the weight function is as in (3.1) with $\nu = 0.1$. Lastly, we pick as starting point for the minimization algorithm $(\beta, \lambda, \sigma) = (1.5, 1, 1)$.

Tables E.3 and E.4 report the finite sample performance of the estimators for $n = 10\,000$, and $\sigma = 0.5$ and $\sigma = 2$, respectively. We observe a good performance of the estimator $\widehat{\beta}_n$ and a very unsatisfactory performance of the estimator $\widehat{\sigma}_n$. We conjecture that the reason for the suboptimal performance lies in the choice of the weight function w, which may have opposite effects on different parameters of the model, as well as in the minimization algorithm, since it has a tendency to get stuck in local minima.

Table E.3. Absolute value of bias (|Bias|) and standard deviation for $n = 10\,000$ and $\sigma = 0.5$ for the generalized modulated OU kernel.

		|Bias|			Std		
β	λ	$\widehat{\beta}_n$	$\widehat{\lambda}_n$	$\widehat{\sigma}_n$	$\widehat{\beta}_n$	$\widehat{\lambda}_n$	$\widehat{\sigma}_n$
1.8	0.5	0.0111	0.1585	0.5982	0.0460	0.0444	0.1353
	0.75	0.0196	0.0925	0.5620	0.0542	0.0494	0.1718
	1.25	0.0147	0.0064	0.0671	0.0813	0.1152	0.0946
	1.5	0.0029	0.0361	0.0969	0.0856	0.1006	0.1728
1.2	0.5	0.0062	0.1881	0.6967	0.0349	0.0732	0.2415
	0.75	0.0044	0.1787	0.8088	0.0440	0.0443	0.0486
	1.25	0.0103	0.0089	0.6124	0.0468	0.0594	0.1307
	1.5	0.0110	0.0886	0.5869	0.0519	0.0951	0.2115

Table E.4. Absolute value of bias (|Bias|) and standard deviation (Std) for $n = 10\,000$ and $\sigma = 2$ for the generalized modulated OU kernel.

		|Bias|			Std		
β	λ	$\widehat{\beta}_n$	$\widehat{\lambda}_n$	$\widehat{\sigma}_n$	$\widehat{\beta}_n$	$\widehat{\lambda}_n$	$\widehat{\sigma}_n$
1.8	0.5	0.0076	0.0314	0.1458	0.1730	0.2052	0.7289
	0.75	0.0028	0.2089	0.6515	0.0309	0.0241	0.0729
	1.25	0.0314	0.2521	1.3273	0.0727	0.0641	0.1244
	1.5	0.0626	0.0066	1.3147	0.0889	0.1085	0.1725
1.2	0.5	0.0165	0.0220	0.1531	0.2724	0.1923	0.6673
	0.75	0.0011	0.2065	0.6793	0.0335	0.0521	0.1611
	1.25	0.0037	0.2068	0.7685	0.0474	0.0454	0.0362
	1.5	0.0019	0.1720	1.0176	0.0635	0.0995	0.1928

3.3 Ornstein–Uhlenbeck

In this subsection we consider the Ornstein–Uhlenbeck kernel from Example 2.4 with $\sigma = 1$ being fixed and known. In this case Assumption (A) is satisfied for both $m = 2$ and $m = 1$, and we will compare the performance for each of these dimensions. Akin to Section 3.2 we pick 20^m, $m = 1, 2$, number of weights in the integral approximation with weight function chosen as in (3.1) with $\nu = 1$. The starting point for the minimization algorithm is throughout $(\beta, \lambda) = (1.5, 0.5)$.

Tables E.5–E.8 demonstrate the simulation results for $m = 1$ and $m = 2$, respectively. We observe a rather convincing performance for both estimators in all settings, but the choice $m = 1$ clearly outperforms the setting $m = 2$. We conjecture that it has a theoretical background, i.e. the asymptotic variances in Theorem 2.9(ii) are smaller for $m = 1$, and a numerical background. Indeed, the minimization algorithm has a worse performance for higher values of m. For this reason it is advisable to use the minimal m, which identifies the parameters of the model.

Table E.5. Absolute value of bias (|Bias|) and standard deviation (Std) for $m = 1$ and $n = 1000$.

$n = 1000$		|Bias|		Std	
β	λ	$\widehat{\beta_n}$	$\widehat{\lambda_n}$	$\widehat{\beta_n}$	$\widehat{\lambda_n}$
1.2	0.25	0.0185	0.0030	0.1072	0.0544
	0.75	0.0144	0.0061	0.0627	0.0683
	1	0.0107	0.0018	0.0573	0.0755
	1.25	0.0084	0.0062	0.0531	0.0856
	1.5	0.0135	0.0058	0.0561	0.0901
	2	0.0044	0.0028	0.0543	0.1282
	2.5	0.0122	0.0150	0.0583	0.1530
1.4	0.25	0.0097	0.0079	0.1214	0.0530
	0.75	0.0047	0.0029	0.0661	0.0669
	1	0.0036	0.0093	0.0593	0.0646
	1.25	0.0042	0.0018	0.0572	0.0757
	1.5	0.0138	0.0023	0.0550	0.0826
	2	0.0091	0.0039	0.0595	0.1072
	2.5	0.0060	0.0072	0.0608	0.1507
1.6	0.25	0.0099	0.0012	0.1143	0.0513
	0.75	0.0071	0.0066	0.0604	0.0604
	1	0.0076	0.0016	0.0590	0.0669
	1.25	0.0116	0.0042	0.0533	0.0759
	1.5	0.0020	0.0039	0.0563	0.0781
	2	0.0101	0.0074	0.0540	0.1021
	2.5	0.0144	0.0061	0.0567	0.1283
1.8	0.25	0.0106	0.0004	0.1013	0.0417
	0.75	0.0111	0.0007	0.0586	0.0597
	1	0.0021	0.0007	0.0529	0.0649
	1.25	0.0088	0.0043	0.0453	0.0764
	1.5	0.0092	0.0136	0.0494	0.0825
	2	0.0084	0.0025	0.0481	0.1015
	2.5	0.0144	0.0045	0.0446	0.1273

Table E.6. Absolute value of bias (|Bias|) and standard deviation (Std) for $m = 1$ and $n = 10\,000$.

$n = 10\,000$		|Bias|		Std	
β	λ	$\widehat{\beta_n}$	$\widehat{\lambda_n}$	$\widehat{\beta_n}$	$\widehat{\lambda_n}$
1.2	0.25	0.0016	0.0004	0.0321	0.0174
	0.75	0.0039	0.0028	0.0187	0.0199
	1	0.0008	0.0000	0.0197	0.0244
	1.25	0.0000	0.0028	0.0189	0.0265
	1.5	0.0004	0.0033	0.0164	0.0323
	2	0.0003	0.0030	0.0161	0.0365
	2.5	0.0016	0.0090	0.0179	0.0466
1.4	0.25	0.0035	0.0014	0.0377	0.0172
	0.75	0.0017	0.0010	0.0202	0.0194
	1	0.0017	0.0022	0.0196	0.0241
	1.25	0.0016	0.0026	0.0174	0.0274
	1.5	0.0000	0.0092	0.0177	0.0281
	2	0.0016	0.0069	0.0166	0.0362
	2.5	0.0014	0.0155	0.0194	0.0404
1.6	0.25	0.0079	0.0019	0.0439	0.0169
	0.75	0.0022	0.0014	0.0184	0.0170
	1	0.0008	0.0020	0.0182	0.0212
	1.25	0.0015	0.0032	0.0183	0.0239
	1.5	0.0003	0.0056	0.0169	0.0271
	2	0.0005	0.0133	0.0166	0.0325
	2.5	0.0020	0.0192	0.0178	0.0401
1.8	0.25	0.0015	0.0011	0.0392	0.0152
	0.75	0.0013	0.0010	0.0187	0.0176
	1	0.0019	0.0051	0.0162	0.0190
	1.25	0.0020	0.0067	0.0159	0.0232
	1.5	0.0012	0.0113	0.0151	0.0263
	2	0.0032	0.0159	0.0146	0.0298
	2.5	0.0000	0.0259	0.0143	0.0411

Table E.7. Absolute value of bias (|Bias|) and standard deviation (Std) for $m = 2$ and $n = 1000$.

$n = 1000$		|Bias|		Std	
β	λ	$\widehat{\beta_n}$	$\widehat{\lambda_n}$	$\widehat{\beta_n}$	$\widehat{\lambda_n}$
1.2	0.25	0.3988	0.1113	0.1044	0.0623
	0.75	0.0666	0.0363	0.2178	0.1932
	1	0.0150	0.0081	0.0928	0.1339
	1.25	0.0125	0.0104	0.0681	0.1226
	1.5	0.0054	0.0063	0.0626	0.1181
	2	0.0073	0.0090	0.0646	0.1472
	2.5	0.0613	0.1477	0.0636	0.0627
1.4	0.25	0.2028	0.0531	0.1407	0.1061
	0.75	0.0484	0.0204	0.1793	0.1621
	1	0.0063	0.0063	0.0848	0.1165
	1.25	0.0124	0.0067	0.0714	0.1096
	1.5	0.0025	0.0067	0.0721	0.1204
	2	0.0080	0.0203	0.0572	0.1269
	2.5	0.0593	0.1395	0.0734	0.0482
1.6	0.25	0.1120	0.1078	0.3009	0.2139
	0.75	0.0481	0.0210	0.1669	0.1602
	1	0.0165	0.0159	0.0909	0.1164
	1.25	0.0072	0.0017	0.0666	0.1039
	1.5	0.0012	0.0078	0.0667	0.0990
	2	0.0037	0.0133	0.0688	0.1171
	2.5	0.0873	0.1364	0.0850	0.0431
1.8	0.25	0.2478	0.1751	0.3584	0.2232
	0.75	0.0194	0.0015	0.1182	0.1253
	1	0.0112	0.0007	0.0755	0.1010
	1.25	0.0098	0.0083	0.0587	0.0881
	1.5	0.0150	0.0020	0.0540	0.0973
	2	0.0187	0.0121	0.0632	0.1106
	2.5	0.0948	0.1346	0.0802	0.0502

Table E.8. Absolute value of bias (|Bias|) and standard deviation (Std) for $m = 2$ and $n = 10\,000$.

$n = 10\,000$		|Bias|		Std	
β	λ	$\widehat{\beta}_n$	$\widehat{\lambda}_n$	$\widehat{\beta}_n$	$\widehat{\lambda}_n$
1.2	0.25	0.3922	0.1142	0.0464	0.0098
	0.75	0.0005	0.0013	0.0391	0.0480
	1	0.0019	0.0012	0.0260	0.0421
	1.25	0.0007	0.0024	0.0220	0.0377
	1.5	0.0003	0.0030	0.0218	0.0428
	2	0.0005	0.0050	0.0195	0.0454
	2.5	0.0387	0.1186	0.0204	0.0024
1.4	0.25	0.1916	0.0599	0.0731	0.0402
	0.75	0.0024	0.0015	0.0439	0.0500
	1	0.0019	0.0024	0.0257	0.0363
	1.25	0.0009	0.0002	0.0235	0.0361
	1.5	0.0012	0.0004	0.0211	0.0381
	2	0.0027	0.0020	0.0227	0.0397
	2.5	0.0505	0.1184	0.0243	0.0006
1.6	0.25	0.0051	0.0138	0.1794	0.1028
	0.75	0.0084	0.0061	0.0451	0.0479
	1	0.0002	0.0023	0.0253	0.0324
	1.25	0.0003	0.0047	0.0206	0.0305
	1.5	0.0003	0.0040	0.0200	0.0334
	2	0.0015	0.0060	0.0200	0.0389
	2.5	0.0604	0.1185	0.0287	0.0015
1.8	0.25	0.2109	0.1160	0.2539	0.1351
	0.75	0.0016	0.0023	0.0389	0.0395
	1	0.0001	0.0025	0.0243	0.0316
	1.25	0.0004	0.0036	0.0178	0.0266
	1.5	0.0001	0.0042	0.0173	0.0280
	2	0.0012	0.0092	0.0181	0.0371
	2.5	0.0801	0.1184	0.0343	0.0010

4 Proofs

In this section $C > 0$ denotes a generic constant, which may change from line to line. Recall moreover the shorthand $\xi = (\beta, \theta)$ for the joint parameters.

4.1 The Limiting Gaussian Field

To characterize the covariance of the asymptotic Gaussian field $(G_u)_{u \in \mathbb{R}_+^m}$ we define a dependence measure between two m-dimensional stable random vectors $Y = (\int h_1 \, \mathrm{d}L, \ldots, \int h_m \, \mathrm{d}L)$ and $Z = (\int g_1 \, \mathrm{d}L, \ldots, \int g_m \, \mathrm{d}L)$:

$$U_{Y,Z}(u,v) := \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\langle (u,v),(Y,Z)\rangle_{\mathbb{R}^{2m}}}\big] - \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\langle u,Y\rangle_{\mathbb{R}^m}}\big]\mathbb{E}\big[\mathrm{e}^{\mathrm{i}\langle v,Z\rangle_{\mathbb{R}^m}}\big], \qquad u,v \in \mathbb{R}^m.$$

This is a straightforward multivariate extension of the measure defined in [19]. We now apply Theorem 1.1 in conjunction with the smooth and bounded functions

$$f_u(x) = \cos(\langle u, x\rangle_{\mathbb{R}^m}), \qquad u, x \in \mathbb{R}^m,$$

such that we obtain the finite dimensional convergence of the processes:

$$\sqrt{n}\big(\varphi_n(u) - \varphi_\xi(u)\big)_{u \in \mathbb{R}_+^m} \xrightarrow[n\to\infty]{\mathcal{L}\text{-f}} (G_u)_{u \in \mathbb{R}_+^m}.$$

Let $Z_0 = (X_1, \ldots, X_m)$ and $Z_\ell = (X_{1+\ell}, \ldots, X_{m+\ell})$, then the covariance function $R : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ of G is, cf. (1.4), given by

$$R(u,v) = \sum_{\ell \in \mathbb{Z}} r_\ell(u,v),$$

where for $\ell \in \mathbb{Z}$

$$r_\ell(u,v) = \mathrm{Cov}\big(\cos(\langle u, Z_0\rangle), \cos(\langle v, Z_\ell\rangle)\big), \qquad u, v \in \mathbb{R}^m.$$

We will now prove that there exists a version of G, which is locally Hölder continuous up to any order less than $\beta/4$. By Kolmogorov's criteria and Gaussianity it is enough to prove that for any $T > 0$ there exists a constant $C_T \geq 0$ such that

$$\mathbb{E}\big[(G_u - G_v)^2\big] \leq C_T \, \|u - v\|^{\beta/2} \qquad \text{for all } u, v \in [0,T]^m, \tag{4.1}$$

where $\|u - v\| = \sum_{i=1}^m |u_i - v_i|$ denotes the ℓ_1-norm throughtout the rest of this paper. To prove (4.1) note the decomposition

$$\mathbb{E}\big[(G_u - G_v)^2\big] = R(u,u) - R(u,v) + R(v,v) - R(u,v).$$

Hence by symmetry it suffices to consider the term

$$R(u,u) - R(v,v) = \sum_{\ell \in \mathbb{Z}} (r_\ell(u,u) - r_\ell(u,v)).$$

The main difficulty lies in establishing a bound on $r_\ell(u,u) - r(u,v)$ which is both $\frac{\beta}{2}$-Hölder in (u,v) and summable in ℓ. Using the standard identity $\cos(x) = (\mathrm{e}^{\mathrm{i}x} + \mathrm{e}^{-\mathrm{i}x})/2$ and the symmetry of L_1 we deduce the identity

$$2\big(r_\ell(u,u) - r_\ell(u,v)\big) = [U_{Z_0,Z_\ell}(u,-u) - U_{Z_0,Z_\ell}(u,-v)] + [U_{Z_0,Z_\ell}(u,u) - U_{Z_0,Z_\ell}(u,v)].$$

The two terms in the square brackets are treated very similarly so we consider only the first one. Before diving into the tedious calculations we recall the following inequalities for $x, y \in \mathbb{R}$:

$$
\begin{aligned}
|e^{-x} - e^{-y}| &\le |x-y| && \text{if } x, y \ge 0, && (4.2)\\
|x+y|^\beta &\le |x|^\beta + |y|^\beta && \text{for } \beta \in (0,1], && (4.3)\\
\big||x|^\beta - |y|^\beta\big| &\le |x-y|^\beta && \text{for } \beta \in (0,1], && (4.4)\\
\big||x+y|^\beta - |x|^\beta - |y|^\beta\big| &\le |xy|^{\beta/2} && \text{for } \beta \in (0,2). && (4.5)
\end{aligned}
$$

Define additionally the two quantities

$$\rho_i = \int_{\mathbb{R}} |g_\xi(x) g_\xi(x+i)|^{\beta/2}\, dx \qquad \text{and} \qquad \mu_i = \int_{-m}^{\infty} |g_\xi(x+i)|^\beta\, dx, \qquad i \in \mathbb{Z}.$$

We shall need the following lemma.

Lemma 4.1. *Let $i \in \mathbb{N}$. Then it holds:*

(i) $\rho_i \le C i^{-\alpha\beta/2}$.

(ii) *If $i > m$ then $\mu_i \le C(i-m)^{1-\alpha\beta}$.*

Proof. (i) follows as in [6, Lemma 4.1]. For (ii) note if $k > m$ then $x + k > 1$ for any $x > -m$, so according to assumption (1.2)

$$\mu_i \le C \int_{-m}^{\infty} (x+k)^{-\alpha\beta}\, dx = C(k-m)^{1-\alpha\beta},$$

where we used that $\alpha\beta > 2$. □

Using the expression for the characteristic function of a symmetric β-stable random variable we decompose as follows

$$
\begin{aligned}
&U_{Z_0, Z_\ell}(u, -u) - U_{Z_0, Z_\ell}(u, -v)\\
&\quad = \exp\Big(-\Big\|\sum_{i=1}^m u_i(g_\xi(i-\cdot) - g_\xi(i+\ell-\cdot))\Big\|_\beta^\beta\Big) - \exp\Big(-2\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta\Big)\\
&\qquad - \Big[\exp\Big(-\Big\|\sum_{i=1}^m u_i g(i-\cdot) - v_i g(i+\ell-\cdot)\Big\|_\beta^\beta\Big)\\
&\qquad\quad - \exp\Big(-\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta - \Big\|\sum_{i=1}^m v_i g_\xi(i+\ell-\cdot)\Big\|_\beta^\beta\Big)\Big]\\
&\quad = \exp\Big(2\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta\Big)\\
&\qquad \times \Big[\exp\Big(-2\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta\Big)\\
&\qquad\quad - \exp\Big(-\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta - \Big\|\sum_{i=1}^m v_i g_\xi(i-\cdot)\Big\|_\beta^\beta\Big)\Big]
\end{aligned}
$$

$$
\begin{aligned}
&\times\Bigg[\exp\Big(-\Big\|\sum_{i=1}^m u_i(g_\xi(i-\cdot)-g_\xi(i+\ell-\cdot))\Big\|_\beta^\beta\Big)\\
&\quad -\exp\Big(-2\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta\Big)\Bigg]\\
&+\exp\Big(-\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta-\Big\|\sum_{i=1}^m v_i g_\xi(i-\cdot)\Big\|_\beta^\beta\Big)\\
&\times\Bigg[\exp\Big(-\Big\|\sum_{i=1}^m u_i(g_\xi(i-\cdot)-g_\xi(i+\ell-\cdot))\Big\|_\beta^\beta+2\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta\Big)\\
&\quad -\exp\Big(-\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)-v_i g_\xi(i+\ell-\cdot)\Big\|_\beta^\beta\\
&\qquad\quad +\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta+\Big\|\sum_{i=1}^m v_i g_\xi(i-\cdot)\Big\|_\beta^\beta\Big)\Bigg]\\
&=: r_\ell^1(u,v)+r_\ell^2(u,v).
\end{aligned}
$$

For the first term, r_ℓ^1, we notice that the exponential term in front is bounded in $u\in[0,T]^m$ (and of course in $\ell\in\mathbb{Z}$ as well), hence by (4.2)

$$
\begin{aligned}
r_\ell^1(u,v)\le C_T&\Bigg|\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta-\Big\|\sum_{i=1}^m v_i g_\xi(i-\cdot)\Big\|_\beta^\beta\Bigg|\\
&\times\Bigg|\Big\|\sum_{i=1}^m u_i(g_\xi(i-\cdot)-g_\xi(i+\ell-\cdot))\Big\|_\beta^\beta-2\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta\Bigg|.
\end{aligned}
$$

The first absolute value term will give the Hölder continuity of order $\beta/2$ and the second will ensure summability in ℓ. For the first term we may bound as follows in the case $\beta\in(0,1]$ using (4.4) and (4.3)

$$
\begin{aligned}
\Bigg|\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta-\Big\|\sum_{i=1}^m v_i g_\xi(i-\cdot)\Big\|_\beta^\beta\Bigg|&\le\int_{\mathbb{R}}\Big(\sum_{i=1}^m|u_i-v_i||g_\xi(i-x)|\Big)^\beta\,\mathrm{d}x\\
&\le\|u-v\|^\beta\sum_{i=1}^m\int_{\mathbb{R}}|g_\xi(i-x)|^\beta\,\mathrm{d}x\\
&\le C_T\,\|u-v\|^{\beta/2}.
\end{aligned}
$$

If instead $\beta>1$, then the map is $u\mapsto\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\|_\beta^\beta$ is continuously differentiable, hence by the mean value theorem it is Hölder continuous of any order less than or equal to 1, and since $\beta\in(0,2)$ Hölder continuity of order $\beta/2$ then holds. For the second absolute value term it follows by (4.5) and (4.3)

$$\begin{aligned}
&\Big|\Big\|\sum_{i=1}^m u_i(g_\xi(i-\cdot)-g_\xi(i+\ell-\cdot))\Big\|_\beta^\beta - 2\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta\Big| \\
&\quad= \Big|\Big\|\sum_{i=1}^m u_i(g_\xi(i-\cdot)-g_\xi(i+\ell-\cdot))\Big\|_\beta^\beta - \Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta \\
&\qquad - \Big\|-\sum_{i=1}^m u_i g_\xi(i+\ell-\cdot)\Big\|_\beta^\beta\Big| \\
&\quad\le 2\Big\|\Big(\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big)\Big(\sum_{k=1}^m u_k g_\xi(k+\ell-\cdot)\Big)\Big\|_{\beta/2}^{\beta/2} \\
&\quad\le 2T^\beta \sum_{i,k=1}^m \|g_\xi(i-\cdot)g_\xi(k+\ell-\cdot)\|_{\beta/2}^{\beta/2} \\
&\quad= 2T^\beta \sum_{i,k=1}^m \rho_{\ell+k-i},
\end{aligned}$$

which is summable in ℓ by Lemma 4.1 and the assumption $\alpha\beta > 2$. We now turn our attention to the more complicated second term $r_\ell^2(u,v)$. Utilising (4.2) we have that

$$\begin{aligned}
r_\ell^2(u,v) &\le \Big|\Big\|\sum_{i=1}^m u_i(g_\xi(i-\cdot)-g_\xi(i+\ell-\cdot))\Big\|_\beta^\beta - 2\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta \\
&\quad + \Big\|\sum_{i=1}^m v_i g_\xi(i-\cdot)\Big\|_\beta^\beta + \Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot)\Big\|_\beta^\beta \\
&\quad - \Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot) - v_i g_\xi(i+\ell-\cdot)\Big\|_\beta^\beta\Big| \\
&= \Big|\int_{-m}^\infty \Big[\Big|\sum_{i=1}^m u_i(g_\xi(x+i)-g_\xi(i+\ell+x))\Big|^\beta \\
&\quad - \Big|\sum_{i=1}^m u_i g_\xi(i+x) - v_i g_\xi(i+\ell+x)\Big|^\beta\Big] \\
&\quad + \Big[\Big|\sum_{i=1}^m v_i g_\xi(i+\ell+x)\Big|^\beta - \Big|\sum_{i=1}^m u_i g_\xi(i+\ell+x)\Big|^\beta\Big]\mathrm{d}x\Big| \\
&\le \int_{-m}^\infty \Big|\Big|\sum_{i=1}^m u_i(g_\xi(x+i)-g_\xi(i+\ell+x))\Big|^\beta \\
&\quad - \Big|\sum_{i=1}^m u_i g_\xi(i+x) - v_i g_\xi(i+\ell+x)\Big|^\beta\Big|\mathrm{d}x \\
&\quad + \int_{-m}^\infty \Big|\Big|\sum_{i=1}^m v_i g_\xi(i+\ell+x)\Big|^\beta - \Big|\sum_{i=1}^m u_i g_\xi(i+\ell+x)\Big|^\beta\Big|\mathrm{d}x \\
&=: r_\ell^{2,1}(u,v) + r_\ell^{2,2}(u,v).
\end{aligned}$$

We deal first with the second term $r_\ell^{2,2}$. First, if $\beta \in (0,1]$, then by (4.4) and (4.3)

$$r_\ell^{2,2}(u,v) \le \int_{-m}^\infty \Big|\sum_{i=1}^m (u_i - v_i) g_\xi(i+\ell+x)\Big|^\beta \,\mathrm{d}x \le \|u-v\|^\beta \sum_{i=1}^m \mu_{i+\ell},$$

and by Lemma 4.1(ii) we obtain a bound which is summable in $\ell > m$. If instead $\beta \in (1,2)$ the map

$$h(u) = \int_{-m}^{\infty} \Big| \sum_{i=1}^{m} u_i g_\xi(i+\ell+x) \Big|^{\beta} \,\mathrm{d}x, \qquad u \in \mathbb{R}^m,$$

is continuously differentiable and the absolute value of the derivative is bounded as follows for any $u \in [0,T]^m$ and $\ell > m$:

$$\begin{aligned} \Big| \frac{\partial}{\partial u_k} h(u) \Big| &\leq \int_{-m}^{\infty} \Big| \sum_{i=1}^{m} u_i g_\xi(i+\ell+x) \Big|^{\beta-1} |g_\xi(k+\ell+x)| \,\mathrm{d}x \\ &\leq T^{\beta-1} \sum_{i=1}^{m} \int_{-m}^{\infty} |g_\xi(i+\ell+x)|^{\beta-1} |g_\xi(k+\ell+x)| \,\mathrm{d}x \\ &\leq C T^{\beta-1} m(\ell-m)^{1-\alpha\beta}, \end{aligned}$$

where we have argued as in Lemma 4.1(ii) in the last inequality. Hence, in the case $\beta \in (1,2)$ we obtain by the mean value theorem

$$r_\ell^{2,2}(u,v) \leq \sup_{z \in [0,T]^m} \|\nabla h(z)\| \, \|u-v\| \leq C_T (\ell-m)^{1-\alpha\beta} \|u-v\|^{\beta},$$

and as $\alpha\beta > 2$ we have obtained a bound summable in ℓ.

It remains to consider the term $r_\ell^{2,1}$. Here it follows from the inequality $\big||x|^\beta - |y|^\beta\big| \leq |x^2-y^2|^{\beta/2}$ and the triangle inequality that the integrand is bounded by

$$\begin{aligned} &\bigg| \Big| \sum_{i=1}^{m} u_i (g_\xi(i+x) - g_\xi(i+\ell+x)) \Big|^{\beta} - \Big| \sum_{i=1}^{m} u_i g_\xi(i+x) - v_i g_\xi(i+\ell+x) \Big|^{\beta} \bigg| \\ &\quad \leq \bigg| \sum_{i,k=1}^{m} u_i u_k (g_\xi(i+x) - g_\xi(i+\ell+x))(g_\xi(k+x) - g_\xi(k+\ell+x)) \\ &\qquad\qquad - (u_i g_\xi(i+x) - v_i g_\xi(i+\ell+x))(u_k g_\xi(k+x) - v_k g_\xi(k+\ell+x)) \bigg|^{\beta/2} \\ &\quad = \bigg| \sum_{i,k=1}^{m} \Big[(u_i u_k - v_i v_k) g_\xi(i+\ell+x) g_\xi(k+\ell+x) \\ &\qquad\qquad + u_i (v_k - u_k) g_\xi(i+x) g_\xi(k+\ell+x) \\ &\qquad\qquad + u_k (v_i - u_i) g_\xi(i+\ell+x) g_{\theta,\beta}(k+x) \Big] \bigg|^{\beta/2} \\ &\quad \leq C_T \|u-v\|^{\beta/2} \sum_{i,k=1}^{m} \Big[\big| g_\xi(i+\ell+x) g_\xi(k+\ell+x) \big|^{\beta/2} + \big| g_\xi(i+x) g_\xi(k+\ell+x) \big|^{\beta/2} \Big]. \end{aligned}$$

Hence, we obtain with arguments as in Lemma 4.1(ii) that

$$r_\ell^{2,1}(u,v) \leq C_T \|u-v\|^{\beta/2} \Big((\ell-m)^{1-\alpha\beta} + \sum_{i,k=1}^{m} \rho_{\ell+k-i} \Big),$$

which is summable in ℓ as $\alpha\beta > 2$.

Lastly, we shall prove that $(G_u)_{u \in \mathbb{R}_+^m}$ has paths in $\mathcal{L}_w^1(\mathbb{R}_+^m)$ almost surely, such that $\int_{\mathbb{R}_+^m} G_u w(u) \,\mathrm{d}u$ is well-defined. A sufficient criteria for this is $\int_{\mathbb{R}_+^m} \mathrm{Var}[G_u]^{1/2} w(u) \,\mathrm{d}u < \infty$, since G is centred. For this we need to study $r_\ell(u,u)$ again. Recall that

$$r_\ell(u,u) = U_{Z_0,Z_\ell}(u,-u) + U_{Z_0,Z_\ell}(u,u).$$

As both terms are treated almost identically it suffices to consider the first one. Here it follows from the inequality $|e^x - 1| \le e^{|x|}|x|$, $x \in \mathbb{R}$, and (4.5), that

$$
\begin{aligned}
&|U_{Z_0,Z_\ell}(u,-u)| \\
&\quad = \left|\exp\Big(-\Big\|\sum_{i=1}^m u_i(g_\xi(i-\cdot\,)-g_\xi(i+\ell-\cdot\,))\Big\|_\beta^\beta\Big) - \exp\Big(-2\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot\,)\Big\|_\beta^\beta\Big)\right| \\
&\quad \le \exp\Big(-2\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot\,)\Big\|_\beta^\beta\Big) \\
&\qquad \times \left|\Big\|\sum_{i=1}^m u_i(g_\xi(i-\cdot\,)-g_\xi(i+\ell-x))\Big\|_\beta^\beta - 2\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot\,)\Big\|_\beta^\beta\right| \\
&\qquad \times \exp\left(\left|\Big\|\sum_{i=1}^m u_i(g_\xi(i-\cdot\,)-g_\xi(i+\ell-x))\Big\|_\beta^\beta - 2\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot\,)\Big\|_\beta^\beta\right|\right) \\
&\quad \le \exp\left(-2\Big\|\sum_{i=1}^m u_i g_\xi(i-\cdot\,)\Big\|_\beta^\beta + 2\Big\|\Big(\sum_{i=1}^m u_i g_\xi(i-\cdot\,)\Big)\Big(\sum_{i=1}^m u_i g_\xi(i+\ell-\cdot\,)\Big)\Big\|_{\beta/2}^{\beta/2}\right) \\
&\qquad \times \|u\|^\beta \sum_{i,k=1}^m \rho_{\ell+k-i} \\
&\quad \le \|u\|^\beta \sum_{i,k=1}^m \rho_{\ell+k-i},
\end{aligned}
$$

where we have used the Cauchy–Schwarz inequality in the last line. Summing over ℓ yields an element in $\mathcal{L}^1_w(\mathbb{R}^m_+)$ by the assumption on the weight function w.

4.2 Convergence of Integral Functionals

In Section 4.1 we saw that the empirical characteristic functions suitably scaled and centred converge to a Gaussian process in finite dimensional sense. We wish to extend this convergence to integrals of our processes. For this we need to extend [14, Lemma 1] to a multivariate case. For $x \in \mathbb{R}$ let $\lfloor x \rfloor$ denote the largest integer l such that $l \le x$ and for a vector $u = (u_1, \dots, u_m) \in \mathbb{R}^m$ we set $\lfloor u \rfloor = (\lfloor u_1 \rfloor, \dots, \lfloor u_m \rfloor)$.

Lemma 4.2. *Let $(Y^n_u)_{u \in \mathbb{R}^m_+}$ and $(Y_u)_{u \in \mathbb{R}^m_+}$ be continuous random fields with $Y^n \xrightarrow{\mathcal{L}\text{-f}} Y$. Assume that $\int_{\mathbb{R}^m_+} \mathbb{E}[|Y^n_u|]\,\mathrm{d}u < \infty$ and $\int_{\mathbb{R}^m_+} \mathbb{E}[|Y_u|]\,\mathrm{d}u < \infty$, and set for $k, \ell, n \in \mathbb{N}$*

$$
X_{n,k,\ell} = \int_{[0,\ell]^m} Y^n_{\lfloor uk \rfloor / k}\,\mathrm{d}u \qquad \text{and} \qquad X_{n,\ell} = \int_{[0,\ell]^m} Y^n_u\,\mathrm{d}u.
$$

Suppose that

$$
\begin{gathered}
\lim_{\ell\to\infty} \limsup_{n\to\infty} \int_{\mathbb{R}^{m-1-i}_+} \int_\ell^\infty \int_{\mathbb{R}^i_+} \mathbb{E}[|Y^n_u|]\,\mathrm{d}u = 0, \\
\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(|X_{n,k,\ell} - X_{n,\ell}| > \varepsilon) = 0,
\end{gathered}
$$

where the first convergence holds for all $i \in \{0, \dots, m-1\}$ and the latter for all $\varepsilon, \ell > 0$. Then convergence in distribution holds:

$$
\int_{\mathbb{R}^m_+} Y^n_u\,\mathrm{d}u \xrightarrow{\mathcal{L}} \int_{\mathbb{R}^m_+} Y_u\,\mathrm{d}u \qquad \text{for } n \to \infty.
$$

Proof. Observe for each $\ell > 0$ the decomposition

$$\int_{\mathbb{R}_+^m} Y_u^n \,\mathrm{d}u = X_{n,k,\ell} + (X_{n,\ell} - X_{n,k,\ell}) + \sum_{i=0}^{m-1} \int_{\mathbb{R}_+^{m-1-i}} \int_\ell^\infty \int_{[0,\ell]^i} Y_u^n \,\mathrm{d}u.$$

Conclude now as in [14, Lemma 1]. □

4.3 Convergence of the Estimator

First, $\xi_n \xrightarrow{\text{a.s.}} \xi_0$ follows by standard arguments which in particular uses Assumption (A), see, e.g. [14], where one uses

$$\|\varphi_n - \varphi_{\xi_0}\|_w \xrightarrow{\text{a.s.}} 0 \qquad \text{as } n \to \infty, \tag{4.6}$$

which is a direct consequence of (2.2) and Lebesgue's dominated convergence theorem. To derive the central limit theorem for the estimator, we consider instead the requirement

$$\nabla_\xi F(\varphi, \xi) = 0 \qquad \varphi \in \mathcal{L}_w^2(\mathbb{R}_+^m),\ \xi \in \Xi,$$

which is satisfied at (φ_{ξ_0}, ξ_0). The problem may now be viewed from a implicit functional point of view. To this end we recall the implicit function theorem on general Banach spaces. Consider a Fréchet differentiable map $g : U_1 \times U_2 \to B_3$ where U_1 and U_2 are open subsets of the Banach spaces B_1 and B_2, respectively, and B_3 is an additional Banach space. Let $D_{h_i}^i g(p_1, p_2)$, $i \in \{1,2\}$, denote the partial derivatives at the point $(p_1, p_2) \in U_1 \times U_2$ in the direction $h_i \in B_i$. If $(p_1^0, p_2^0) \in U_1 \times U_2$ is a point such that $g(p_1^0, p_2^0) = 0$ and the map $h \mapsto D_h^2 g(p_1^0, p_2^0) : B_2 \to B_3$ is a continuous and invertible function, then there exists open subsets $V_1 \subseteq U_1$ and $V_2 \subseteq U_2$ such that $(p_1^0, p_2^0) \in V_1 \times V_2$ and a Fréchet differentiable and bijective (implicit) function $\Phi : V_1 \to V_2$ such that

$$g(p_1, p_2) = 0 \qquad \Longleftrightarrow \qquad \Phi(p_1) = p_2.$$

In addition, the derivative is given by

$$D_h \Phi(p) = -\Big(D_\cdot^2 g(p, \Phi(p))\Big)^{-1} \Big(D_h^1 g(p, \Phi(p))\Big), \qquad h \in B_1,\ p \in V_1. \tag{4.7}$$

As might be apparent we shall consider the specific setup of $g = \nabla_\xi F$, $B_1 = U_1 = \mathcal{L}_w^2(\mathbb{R}_+^m)$, $U_2 = \Xi \subseteq B_2 = \mathbb{R}^{d+1}$. We note that Assumption (B)(2) ensures the existence and continuity of the first and second order derivatives of F. Moreover, Assumption (A)(4) yields the invertibility of the Hessian $\nabla_\xi^2 F(\varphi_{\xi_0}, \xi_0)$.

In this case

$$\Phi(\varphi_n) = \xi_n \qquad \text{and} \qquad \Phi(\varphi_{\xi_0}) = \xi_0.$$

Hence, by Fréchet differentiability we find that

$$\begin{aligned}\sqrt{n}(\xi_n - \xi_0) &= \sqrt{n}\big(\Phi(\varphi_{\xi_0} + (\varphi_n - \varphi_{\xi_0})) - \Phi(\varphi_{\xi_0})\big) \\ &= D_{\sqrt{n}(\varphi_n - \varphi_{\xi_0})} \Phi(\varphi_{\xi_0}) + \sqrt{n}\|\varphi_n - \varphi_{\xi_0}\|_{w,2} R(\varphi_n - \varphi_{\xi_0}),\end{aligned}$$

where the remainder term satisfies that $R(\varphi_n - \varphi_{\xi_0}) \xrightarrow{\text{a.s.}} 0$ as $\|\varphi_n - \varphi_{\xi_0}\|_{w,2} \xrightarrow{\text{a.s.}} 0$. Recalling the derivative at (4.7) and the representation $F(\varphi, \xi) = \langle \varphi - \varphi_\xi, \varphi - \varphi_\xi \rangle_w$, it suffices to prove that

$$\begin{aligned}\sqrt{n}\|\varphi_n - \varphi_{\xi_0}\|_{w,2} &\xrightarrow{\mathcal{L}} \|G\|_{w,2} \\ (\langle \partial_{\xi_i} \varphi_{\xi_0}, \sqrt{n}(\varphi_n - \varphi_{\xi_0}) \rangle_w)_{i=1,\dots,d+1} &\xrightarrow{\mathcal{L}} (\langle \partial_{\xi_i} \varphi_{\xi_0}, G \rangle_w)_{i=1\dots,d+1}.\end{aligned}$$

Note that since it is the same underlying sequence of processes, $\sqrt{n}(\varphi_n - \varphi_{\xi_0})$, it is for the last convergence enough to consider the case of a fixed $i \in \{1, \dots, d+1\}$; indeed, this requires little modification of Lemma 4.2. We focus on the last convergence as the first is treated similarly, and to use Lemma 4.2 it is sufficient to provide suitable moment estimates for

$$Y_u^n = \partial_{\xi_i}\varphi_{\xi_0}(u)w(u)\sqrt{n}(\varphi_n(u) - \varphi_{\xi_0}(u)) =: h(u)G_u^n, \qquad u \in \mathbb{R}_+^m,\ n \in \mathbb{N}.$$

Using arguments as in [3, Proposition 3.3, page 12] and the variance estimates from Section 4.1 we deduce that

$$\mathbb{E}\big[|Y_u^n|^2\big] \leq (\partial_{\xi_i}\varphi_{\xi_0}(u)w(u))^2 \sum_{\ell \in \mathbb{Z}} |r_\ell(u,u)| \leq C\,\|u\|^\beta (\partial_{\xi_i}\varphi_{\xi_0}(u)w(u))^2.$$

Taking the square root we obtain a bound in $\mathcal{L}^1(\mathbb{R}_+^m)$ of $\mathbb{E}[|Y_u^n|]$ by the Cauchy–Schwarz inequality used together with Assumption (B)(2) and that $u \mapsto \|u\|$ is an element of $\mathcal{L}_w^2(\mathbb{R}_+^m)$. Hence the first condition of Lemma 4.2 is satisfied. The second condition is slightly more involved, but let an $\ell > 0$ be given and consider any $u, v \in [0, \ell]^m$. Then

$$\begin{aligned}\mathbb{E}\big[|Y_u^n - Y_v^n|^2\big]^{1/2} &\leq |h(u) - h(v)|\mathrm{Var}[G_u^n]^{1/2} + |h(v)|\mathrm{Cov}(G_u^n, G_v^n)^{1/2} \\ &\leq C_T(|h(u) - h(v)| + \|u - v\|),\end{aligned}$$

which by Markov's inequality yields the second condition of Lemma 4.2.

4.4 Proof of Statements in Example Example 2.6

Consider the kernel(*) $g_\theta(u) = \exp(-\theta_1 u - \theta_2 f(u))\mathbb{1}_{(0,\infty)}(u)$ for $\theta = (\theta_1, \theta_2) \in (0,\infty)^2$ and where f is a bounded measurable 1-periodic function which does not vanish except on a Lebesgue null set. Assume moreover that f is either non-positive or non-negative. It is straightforward to see that in this case the characteristic function of X_1 does not determine the parameter θ uniquely. Consider instead the joint characteristic function $\varphi_{\beta,\theta}(u_1, u_2)$ of (X_1, X_2) for the moving average X with kernel g_θ, which is given by:

$$\varphi_{\beta,\theta}(u_1, u_2) = \exp\Big(-\|u_1 g_\theta + u_2 g_\theta(\cdot + 1)\|_\beta^\beta\Big), \qquad u_1, u_2 \geq 0.$$

If $\varphi_{\beta,\theta} = \varphi_{\beta,\tilde{\theta}}$ for $\theta, \tilde{\theta} \in (0,\infty)^2$, then the β-norms must be equal. Recalling the generalized binomial theorem

$$(x+y)^\beta = \sum_{k=0}^\infty \binom{\beta}{k} x^{\beta-k} y^k, \qquad x > y \geq 0,$$

we may calculate these norms explicitly for $u_1 > u_2 \geq 0$:

$$\begin{aligned}&\|u_1 g_\theta + u_2 g_\theta(\cdot + 1)\|_\beta^\beta \\ &\quad= u_2^\beta \int_0^1 \exp(-\theta_1 x - \theta_2 f(x))\,\mathrm{d}x \\ &\qquad+ \int_0^\infty \sum_{k=0}^\infty \binom{\beta}{k} u_1^{\beta-k} u_2^k \exp\Big(-(\beta - k)(\theta_1 x + \theta_2 f(x)) \\ &\qquad\qquad - k\big(\theta_1(x+1) + \theta_2 f(x+1)\big)\Big)\mathrm{d}x\end{aligned}$$

(*) Similarly considerations can be done for the Ornstein–Uhlenbeck kernel, albeit easier and more explicit.

$$
\begin{aligned}
&= u_2^\beta \int_0^1 \exp(-\theta_1 x - \theta_2 f(x))\,\mathrm{d}x \\
&\quad + \int_0^\infty \sum_{k=0}^\infty \binom{\beta}{k} u_1^{\beta-k} u_2^k \exp\big(-\beta(\theta_1 x + \theta_2 f(x)) - k\theta_1\big)\,\mathrm{d}x \\
&= u_2^\beta \int_0^1 \exp(-\theta_1 x - \theta_2 f(x))\,\mathrm{d}x \\
&\quad + \big(u_1 + u_2 \exp(-\theta_1)\big)^\beta \int_0^\infty \exp\big(-\beta(\theta_1 x + \theta_2 f(x))\big)\,\mathrm{d}x
\end{aligned}
$$

where the last equality follows from the generalized binomial theorem since $u_1 > u_2 \geq u_2 \exp(-\theta_1)$. Hence if $\varphi_{\beta,\theta} = \varphi_{\beta,\tilde{\theta}}$ then for all $u_1 > u_2 \geq 0$

$$
1 = \frac{u_2^\beta \int_0^1 \mathrm{e}^{-\theta_1 x - \theta_2 f(x)}\,\mathrm{d}x + (u_1 + u_2 \mathrm{e}^{-\theta_1})^\beta \int_0^\infty \mathrm{e}^{-\beta(\theta_1 x + \theta_2 f(x))}\,\mathrm{d}x}{u_2^\beta \int_0^1 \mathrm{e}^{-\tilde{\theta}_1 x - \tilde{\theta}_2 f(x)}\,\mathrm{d}x + (u_1 + u_2 \mathrm{e}^{-\tilde{\theta}_1})^\beta \int_0^\infty \mathrm{e}^{-\beta(\tilde{\theta}_1 x + \tilde{\theta}_2 f(x))}\,\mathrm{d}x}.
$$

Inserting $u_1 = 1 > 0 = u_2$ yields the identity:

$$
K := \int_0^\infty \exp(-\beta(\theta_1 x + \theta_2 f(x)))\,\mathrm{d}x = \int_0^\infty \exp(-\beta(\tilde{\theta}_1 x + \tilde{\theta}_2 f(x)))\,\mathrm{d}x,
$$

hence it suffices to prove that $\theta_1 = \tilde{\theta}_1$. Moreover, inserting the above identity in $\varphi_{\beta,\theta} = \varphi_{\beta,\tilde{\theta}}$ and differentiating with respect to u_1 gives that for all $u_1 > u_2$:

$$
(u_1 + u_2 \exp(-\theta_1))^{\beta-1} K = (u_1 + u_2 \exp(-\tilde{\theta}_1))^{\beta-1} K,
$$

which proves that $\theta_1 = \tilde{\theta}_1$ if $\beta \neq 1$.

Let us additionally show that $u \mapsto \partial_{\theta_1} \varphi_\xi$ and $u \mapsto \partial_{\theta_2} \varphi_\xi$ are linearly independent if the 1-periodic function is negative and bounded and $\beta \neq 1$. Due to their exponential form these derivatives are linearly independent *if* the following functions (note that we only have an explicit formula when $u_1 > u_2 \geq 0$) are linearly independent in $u_1 > u_2 \geq 0$:

$$
\begin{aligned}
\partial_{\theta_1} \|u_1 g_\theta + u_2 g_\theta(\cdot+1)\|_\beta^\beta &= -K_{\theta,1} u_2^\beta - K_{\theta,2}(u_1 + u_2)^{\beta-1} u_2 - K_{\theta,3}(u_1 + u_2 \exp(-\theta_1))^\beta, \\
\partial_{\theta_2} \|u_1 g_\theta + u_2 g_\theta(\cdot+1)\|_\beta^\beta &= K_{\theta,4} u_2^\beta + K_{\theta,5}(u_1 + u_2 \exp(-\theta_1))^\beta,
\end{aligned}
$$

where the constant $K_{\theta,1}, \ldots, K_{\theta,5}$ are strictly positive, indeed the only constants which are not in general positive are:

$$
\begin{aligned}
K_{\theta,5} &= -\int_0^\infty \beta\theta_2 f(x) \exp(-\beta(\theta_1 x + \theta_2 f(x)))\,\mathrm{d}x, \\
K_{\theta,4} &= -\int_0^1 f(x) \exp(-\theta_1 x - \theta_2 f(x))\,\mathrm{d}x,
\end{aligned}
$$

but they are by our assumption $f < 0$. The main observation needed is that these functions are of different order in u_1 when $u_2 \neq 0$ and that their constants are of opposite sign. Indeed, for $a, b \in \mathbb{R}$ we have that

$$
\begin{aligned}
0 &= \Big(a\partial_{\theta_1} \|u_1 g_\theta + u_2 g_\theta(\cdot+1)\|_\beta^\beta + b\partial_{\theta_1} \|u_1 g_\theta + u_2 g_\theta(\cdot+1)\|_\beta^\beta\Big) \Big/ u_1^\beta \\
&\xrightarrow[u_1 \to \infty]{} -aK_{\theta,3} + bK_{\theta,5}.
\end{aligned}
$$

The constants $aK_{\theta,3}$ and $bK_{\theta,5}$ must then be same and we have the following major simplification:

$$
\begin{aligned}
0 &= a\partial_{\theta_1} \|u_1 g_\theta + u_2 g_\theta(\cdot+1)\|_\beta^\beta + b\partial_{\theta_1} \|u_1 g_\theta + u_2 g_\theta(\cdot+1)\|_\beta^\beta \\
&= -(aK_{\theta,1} - bK_{\theta,4})u_2^\beta - aK_{\theta,2}(u_1+u_2)^{\beta-1}u_2.
\end{aligned}
$$

If $\beta > 1$ then this is clearly unbounded in u_1, hence $a = 0$, and therefore $b = 0$ as well since $K_{\theta,4} > 0$. If $\beta < 1$ then differentiating with respect to u_1 yields the simple equation:

$$
0 = aK_{\theta,2}(u_1+u_2)^{\beta-2}u_2 \qquad \text{for all } u_1 > u_2 \geq 0,
$$

which yields $a = 0$ and therefore $b = 0$ since again $K_{\theta,4} > 0$.

4.5 Proof of Statements in Example 2.7

Recall the moving average kernel from (2.4). First, we show that the one-dimensional characteristic function is not enough to idenitify $\theta = (\theta_1, \theta_2)$. Indeed, we see that for two parameters $\theta, \tilde{\theta} \in (0,\infty)^2$ equality of the one-dimensional characteristic functions gives

$$
\begin{aligned}
\frac{\theta_1^\beta \Gamma(\beta+1)}{(\beta\theta_2)^{\beta+1}} &= \int_0^\infty (\theta_1 s \exp(-\theta_2 s))^\beta \,\mathrm{d}s \\
&= \int_0^\infty (\tilde{\theta}_1 s \exp(-\tilde{\theta}_2 s))^\beta \,\mathrm{d}s = \frac{\tilde{\theta}_1^\beta \Gamma(\beta+1)}{(\beta\tilde{\theta}_2)^{\beta+1}}.
\end{aligned}
\tag{4.8}
$$

We claim that the two-dimensional characteristic function is enough to identity θ. For this we recall the covariation between X_1 and X_0, cf. [22, Section 2.7], which is uniquely determined by the distribution of (X_1, X_0) and hence by its joint characteristic function. If θ denotes the underlying parameter for the moving average X, then the covariation is, cf. [22, Proposition 3.5.2],

$$
\begin{aligned}
[X_1, X_0]_\beta &= \int_{\mathbb{R}} g_\theta(s+1) g_\theta(s)^{\beta-1} \,\mathrm{d}s = \theta_1^\beta \int_0^\infty (s+1)\mathrm{e}^{-\theta_2(s+1)} s^{\beta-1} \mathrm{e}^{-(\beta-1)\theta_2 s} \,\mathrm{d}s \\
&= \theta_1^\beta \mathrm{e}^{-\theta_2} \Big[\int_0^\infty s^\beta \mathrm{e}^{-\beta\theta_2 s} \,\mathrm{d}s + \int_0^\infty s^{\beta-1} \mathrm{e}^{-\beta\theta_2 s} \,\mathrm{d}s \Big] \\
&= \theta_1^\beta \mathrm{e}^{-\theta_2} \Big[\frac{\Gamma(\beta+1)}{(\beta\theta_2)^{\beta+1}} + \frac{\Gamma(\beta)}{(\beta\theta_2)^\beta} \Big] \\
&= \frac{\theta_1^\beta \Gamma(\beta+1)}{(\beta\theta_2)^{\beta+1}} \mathrm{e}^{-\theta_2}(1+\theta_2),
\end{aligned}
\tag{4.9}
$$

where we used the defining property: $\beta\Gamma(\beta) = \Gamma(\beta+1)$. Hence if θ and $\tilde{\theta}$ leads to the same distribution of (X_1, X_0), then combining the identities (4.8) and (4.9) yields

$$
(1+\theta_2)\mathrm{e}^{-\theta_2} = (1+\tilde{\theta}_2)\mathrm{e}^{-\tilde{\theta}_2}.
$$

It is straightforward to check that the function $x \mapsto (1+x)\mathrm{e}^{-x}$ is strictly decreasing on $(0,\infty)$, and therefore injective, which proves that $\theta_2 = \tilde{\theta}_2$ and therefore $\theta_1 = \tilde{\theta}_1$ as well, cf. (4.8).

4.6 Proof of Statements in Example 2.8

We consider a CARMA(2,1) model of the form

$$X_t = \int_{-\infty}^{t} b^\top \exp(A(t-s))e\,\mathrm{d}L_s, \qquad t \in \mathbb{R},$$

where $b = (b_0, 1)^\top$, $e = (0,1)^\top$, L is a symmetric β-stable Lévy process with $\beta \in (1,2)$, and

$$A = \begin{pmatrix} 0 & 1 \\ -\lambda^2 & 2\lambda \end{pmatrix}$$

with $\lambda < 0$. We further assume that $\theta = b_0 + \lambda > 0$. Recall the definition of the incomplete gamma function:

$$\Gamma(\beta; x) = \int_x^\infty y^{\beta-1} \exp(-y)\,\mathrm{d}y, \qquad \beta, x > 0.$$

The following identity is due to partial integration: $\Gamma(\beta+1;x) = \beta\Gamma(\beta;x) + x^\beta \exp(-x)$, or in other words

$$\Gamma(\beta; x) = \beta^{-1}(\Gamma(\beta+1;x) - x^\beta \exp(-x)). \tag{4.10}$$

The one-dimensional characteristic function of X_1 uniquely determines the term

$$\begin{aligned}\int_{\mathbb{R}} |g_\xi(x)|^\beta \,\mathrm{d}x &= \int_0^\infty (1+\theta x)^\beta \exp(\lambda\beta x)\,\mathrm{d}x = \left(\theta \exp(-\lambda\theta^{-1})\right)^\beta \int_{\theta^{-1}}^\infty y^\beta \exp(\lambda\beta y)\,\mathrm{d}y \\ &= -\frac{1}{\lambda\beta}\left(-\frac{\theta\exp(-\lambda\theta^{-1})}{\lambda\beta}\right)^\beta \Gamma(\beta+1; -\lambda\beta\theta^{-1}) =: c.\end{aligned}$$

Now, we compute the covariation $[X_1, X_0]_\beta$:

$$\begin{aligned}[X_1, X_0]_\beta &= \int_{\mathbb{R}} g_\xi(x+1) g_\xi(x)^{\beta-1}\,\mathrm{d}x \\ &= \int_0^\infty (1+\theta(x+1))\exp(\lambda(x+1))(1+\theta x)^{\beta-1}\exp(\lambda(\beta-1)x)\,\mathrm{d}x \\ &= -\frac{1}{\lambda\beta}\left(-\frac{\theta\exp(-\lambda\theta^{-1})}{\lambda\beta}\right)^\beta \exp(\lambda)\left(\Gamma(\beta+1;-\lambda\beta\theta^{-1}) - \lambda\beta\Gamma(\beta;-\lambda\beta\theta^{-1})\right) \\ &= \exp(\lambda)(c(1-\lambda) - \beta^{-1}),\end{aligned}$$

where we used the formula (4.10). Since c is uniquely determined, the quantity $[X_1, X_0]_\beta$ identifies the parameter λ (note that $-c\lambda - \beta^{-1} > 0$, and in particular this term is never equal to 0).

Acknowledgement

The authors acknowledge financial support from the project 'Ambit fields: probabilistic properties and statistical inference' funded by Villum Fonden.

References

[1] Beth Andrews, Matthew Calder and Richard A. Davis (2009). »*Maximum likelihood estimation for α-stable autoregressive processes*«. The Annals of Statistics 37(4), 1946–1982. doi: 10.1214/08-AOS632.

[2] Antoine Ayachea and Julien Hamoniera (2012). »*Linear fractional stable motion: A wavelet estimator of the α parameter*«. Statistics and Probability Letters 82(8), 1569–1575. doi: 10.1016/j.spl.2012.04.005.

[3] Ehsan Azmoodeh, Mathias Mørck Ljungdahl and Christoph Thäle (2020). »*Multi-dimensional normal approximation of heavy-tailed moving averages*«. arXiv: 2002.11335.

[4] Andreas Basse-O'Connor, Claudio Heinrich and Mark Podolskij (2019). »*On limit theory for functionals of stationary increments Lévy driven moving averages*«. Electronic Journal of Probability 24(79), 42 pp. doi: 10.1214/19-EJP336.

[5] Andreas Basse-O'Connor, Raphaël Lachièze-Rey and Mark Podolskij (2017). »*Power variation for a class of stationary increments Lévy driven moving averages*«. The Annals of Probability 45(6B), 4477–4528. doi: 10.1214/16-AOP1170.

[6] Andreas Basse-O'Connor, Mark Podolskij and Christoph Thäle (2019). »*A Berry–Esseén theorem for partial sums of functionals of heavy-tailed moving averages*«. arXiv: 1904.06065.

[7] Peter J. Brockwell, Richard A. Davis and Yu Yang (2011). »*Estimation for non-negative Lévy-driven CARMA processes*«. Journal of Business & Economic Statistics 29(2), 250–259. doi: 10.1198/jbes.2010.08165.

[8] Stamatis Cambanis, Clyde D. Jr. Hardin and Aleksander Weron (1987). »*Ergodic properties of stationary stable processes*«. Stochastic Processes and Their Applications 24(1), 1–18. doi: 10.1016/0304-4149(87)90024-X.

[9] Rainer Dahlhaus (1989). »*Efficient parameter estimation for self-similar processes*«. The Annals of Statistics 17(4), 1749–1766. doi: 10.1214/aos/1176347393.

[10] Thi To Nhu Dang and Jacques Istas (2017). »*Estimation of the Hurst and the stability indices of a H-self-similar stable process*«. Electronic Journal of Statistics 11(2), 4103–4150. doi: 10.1214/17-EJS1357.

[11] Danijel Grahovac, Nikolai N. Leonenko and Murad S. Taqqu (2015). »*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. doi: 10.1007/s10955-014-1126-4.

[12] Marek Kanter (1973). »*The L^p norm of sums of translates of a function*«. Transactions of the American Mathematical Society 179, 35–47. doi: 10.2307/1996488.

[13] Shiqing Ling (2005). »*Self-weighted least absolute deviation estimation for infinite variance autoregressive models*«. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 67(3), 381–393. doi: 10.1111/j.1467-9868.2005.00507.x.

[14] Mathias Mørck Ljungdahl and Mark Podolskij (2019). »*A note on parametric estimation of Lévy moving average processes*«. Springer Proceedings in Mathematics & Statistics 294, 41–56. doi: 10.1007/978-3-030-28665-1_3.

[15] Mathias Mørck Ljungdahl and Mark Podolskij (2020). »*A minimal contrast estimator for the linear fractional motion*«. Statistical Inference for Stochastic Processes 23, 381–413. doi: 10.1007/s11203-020-09216-2.

[16] Stepan Mazur, Dmitry Otryakhin and Mark Podolskij (2020). »*Estimation of the linear fractional stable motion*«. Bernoulli 26(1), 226–252. doi: 10.3150/19-BEJ1124.

[17] J. A. Nelder and R. Mead (1965). »*A simplex method for function minimization*«. The Computer Journal 7(4), 308–313. doi: 10.1093/comjnl/7.4.308.

[18] Vladas Pipiras and Murad S. Taqqu (2003). »*Central limit theorems for partial sums of bounded functionals of infinite-variance moving averages*«. Bernoulli 9(5), 833–855. doi: 10.3150/bj/1066418880.

[19] Vladas Pipiras, Murad S. Taqqu and Patrice Abry (2007). »*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. doi: 10.3150/07-BEJ6143.

[20] Balram S. Rajput and Jan Rosiński (1989). »*Spectral representations of infinitely divisible processes*«. Probability Theory and Related Fields 82(3), 451–487. doi: 10.1007/BF00339998.

[21] Jan Rosiński (1994). »*On uniquness of the spectral representation of stable processes*«. Journal of Theoretical Probability 7(3), 615–634. doi: 10.1007/BF02213572.

[22] Gennady Samorodnitsky and Murad Taqqu (1994). »*Stable non-Gaussian random processes: Stochastic models with infinite variance*«. Stochastic modeling. Chapmann and Hall, New York. isbn: 978-0-412-05171-5.

[23] Maxwell B. Stinchcombe and Halbert White (1992). »*Some measurability results for extrema of random functions over random sets*«. The Review of Economic Studies 59(3), 495–514. doi: 10.2307/2297861.

[24] Shibin Zhang and Xinsheng Zhang (2013). »*A least squares estimator for discretely observed Ornstein–Uhlenbeck processes driven by symmetric α-stable motions*«. Annals of the Institute of Statistical Mathematics 65, 89–103. doi: 10.1007/s10463-012-0362-0.