# Estimation of mixed fractional stable processes using high-frequency data

Mies, Fabian; Podolskij, Mark

**DOI**
10.1214/23-AOS2312

**Publication date**
2023

**Document Version**
Final published version

**Published in**
Annals of Statistics

*The Annals of Statistics*
2023, Vol. 51, No. 5, 1946–1964
https://doi.org/10.1214/23-AOS2312


# ESTIMATION OF MIXED FRACTIONAL STABLE PROCESSES USING HIGH-FREQUENCY DATA

BY FABIAN MIES[1,a] AND MARK PODOLSKIJ[2,b]

[1]*Department of Applied Mathematics, Delft University of Technology,* [a]*f.mies@tudelft.nl*
[2]*Department of Mathematics, University of Luxembourg,* [b]*mark.podolskij@uni.lu*

The linear fractional stable motion generalizes two prominent classes of stochastic processes, namely stable Lévy processes, and fractional Brownian motion. For this reason, it may be regarded as a basic building block for continuous time models. We study a stylized model consisting of a superposition of independent linear fractional stable motions and our focus is on parameter estimation of the model. Applying an estimating equations approach, we construct estimators for the whole set of parameters and derive their asymptotic normality in a high-frequency regime. The conditions for consistency turn out to be sharp for two prominent special cases: (i) for Lévy processes, that is, for the estimation of the successive Blumenthal–Getoor indices and (ii) for the mixed fractional Brownian motion introduced by Cheridito. In the remaining cases, our results reveal a delicate interplay between the Hurst parameters and the indices of stability. Our asymptotic theory is based on new limit theorems for multiscale moving average processes.

## 1. Introduction.

1.1. *Overview.* The linear fractional stable motion (lfsm) is a self-similar stochastic process defined as

$$Y_t^{H,\beta} = \int_{-\infty}^t (t-s)_+^{H-\frac{1}{\beta}} - (-s)_+^{H-\frac{1}{\beta}} \, dZ_s^\beta, \quad t \in \mathbb{R},$$

for a Hurst parameter $H \in (0,1)$ and a standard symmetric $\beta$-stable Lévy motion $Z^\beta$ with $\beta \in (0,2]$, that is,

$$\mathbb{E}\exp(i\lambda Z_t^\beta) = \exp(-t|\lambda|^\beta).$$

Here, we use the notation $(x)_+ = x\mathbb{1}_{x>0}$. Notable special cases of the lfsm are fractional Brownian motion ($\beta = 2$), and the $\beta$-stable Lévy motion itself ($H = 1/\beta$). For general parameters $H$ and $\beta$, the increments of the lfsm exhibit long memory and heavy tails. Another prominent feature of lfsm is its self-similarity, namely $(Y_{\gamma t}^{H,\beta})_{t\ge 0} \overset{d}{=} (\gamma^H Y_t^{H,\beta})_{t\ge 0}$ for any $\gamma > 0$. The lfsm is in some sense prototypical, since it arises as a scaling limit of various moving average processes in discrete time; see, for example, Astrauskas (1983). Hence, parameter estimation for the lfsm may be regarded as an idealized test bed for more general, nonsemimartingale processes.

In this paper, we study the mixed fractional stable process of the form

$$X_t = \sum_{j=1}^q b_j Y_t^{H_j,\beta_j}, \tag{1}$$

Received August 2022; revised June 2023.
*MSC2020 subject classifications.* Primary 60G22, 62F12; secondary 62E20, 62M09, 60F05.
*Key words and phrases.* High frequency data, linear fractional stable motion, Lévy processes, parametric estimation, self-similar processes.

where $Y^{H_j,\beta_j}$, $j = 1, \dots, q$ are independent fractional stable motions and $b_j > 0$. Unless all $H_j$ are identical, the mixed process is no longer self-similar. Throughout this work, we assume that $H_1 < \cdots < H_q$. In this case, we have $\gamma^{-H_1} X_{\gamma t} \Rightarrow b_1 Y_t^{H_1,\beta_1}$ as $\gamma \downarrow 0$, and $\gamma^{-H_q} X_{\gamma t} \Rightarrow b_q Y_t^{H_q,\beta_q}$ as $\gamma \uparrow \infty$. Loosely speaking, the process looks different when zooming in ($\gamma \downarrow 0$) or zooming out ($\gamma \uparrow \infty$). In the sequel, we are mostly interested in the scaling as $\gamma \downarrow 0$. In particular, we want to estimate the parameters $(b_j, H_j, \beta_j)$ based on high-frequency, discrete observations of $X$ in the interval $[0, 1]$. Based on the scaling limit, it is not surprising that $(b_1, H_1, \beta_1)$ may be estimated consistently as $n \to \infty$. Can we also estimate the remaining parameters, and at which rate?

Estimation of the lfsm has been recently studied by Mazur, Otryakhin and Podolskij (2020) for high-frequency observations, and by Ljungdahl and Podolskij (2021) for low-frequency observations. To see why estimation of the mixed lfsm is more complicated, we briefly review the methodology of Mazur, Otryakhin and Podolskij (2020). Crucially, they exploit the self-similarity of the lfsm, $h^{-H}(Y_{t+h}^{H,\beta} - Y_t^{H,\beta}) \overset{d}{=} (Y_{t+1}^{H,\beta} - Y_t^{H,\beta})$ to transfer the high-frequency setting to the low-frequency setting. In particular, they suggest to first estimate $H$ by a log-ratio statistic, and then estimate $(b, \beta)$ based on the empirical characteristic function of the rescaled increments. For the mixed lfsm, on the other hand, the log-ratio estimator will only estimate the dominant, that is, the smallest Hurst exponent $H_1$. It is thus not clear how to extend their procedure to the mixed case. Notably, the mixed lfsm is no longer self-similar, and we may not switch between the low-frequency and the high-frequency regime.

In this paper, we propose the first estimators for the mixed lfsm, based on $n$ discrete observations $X_{\Delta_n}, \dots, X_{n\Delta_n}$ at frequency $\Delta_n$. Of particular interest is the high-frequency case $\Delta_n = 1/n$, but we also derive asymptotic results for the regime $\Delta_n \to 0$, $n\Delta_n \to \infty$. Our estimation method are based upon a system of nonlinear equation of the form $\mathbb{E} f_n(X_{i\Delta_n}, \dots, X_{(i+k)\Delta_n})$ for a suitable choice of functions $f_n$. These moments may be estimated by their sample means, and a parameter estimator may be obtained via the generalized method of moments. In contrast to the classical literature, the function $f_n$ as well as the process $X_{i\Delta_n}$ now depend explicitly on $n$. To be precise, we will choose $f_n(x_0, \dots, x_k) = f(u_n \sum_{r=0}^{k} \binom{k}{r}(-1)^r x_r)$, that is, we take the $k$th order increments of the process, scaled by a factor $u_n \to \infty$ and the function $f$ will be bounded. In the first step, we will derive the asymptotic theory for such class of functionals, which will provide the basis for statistical inference. This limit theory is new and has interest in its own right. In the second step, we will define an estimator for the unknown parameters of the model via a system of estimating equations. We will study two related approaches, where the second one (smooth threshold) explicitly accounts for the presence of a Gaussian component ($\beta = 2$). We prove the asymptotic normality of our estimator and discuss the identifiability issues.

1.2. *Related work.* While the estimation problem under consideration is new, there exists a large body of work in more restrictive settings. This builds a natural comparison basis for our new methodology.

(i) *Linear fractional stable motion.* Estimation of a single lfsm, that is, $q = 1$, has been studied in several papers; early references include a nonrigorous treatment by Abry, Pesquet-Popescu and Taqqu (1999) and a study by Abry, Delbeke and Flandrin (1999) who obtain a suboptimal rate of convergence for the parameter $H$. An asymptotically normal estimator for $H$ is proposed by Stoev, Pipiras and Taqqu (2002), Stoev and Taqqu (2005), using a central limit theorem published in Pipiras, Taqqu and Abry (2007). A consistent estimator for the stability index $\beta$, provided that $H$ is known, is given by Ayache and Hamonier (2012), and consistent estimators for $(H, \beta)$ are suggested by Basse-O'Connor, Lachièze-Rey and Podolskij (2017), Dang and Istas (2017), Grahovac, Leonenko and Taqqu (2015). Estimation of the full parameter $(b, H, \beta)$ has been investigated by Mazur, Otryakhin and Podolskij (2020). Their approach is based on power variations with negative exponents, and they derive the asymptotic normality of their estimator in the low- and high-frequency regime. For the low-frequency regime, these results have been further refined in Ljungdahl and Podolskij (2020) and Ljungdahl and Podolskij (2021). We remark that there is currently no statistical lower bound for estimation of the general lfsm.

(ii) *Mixed fractional Brownian motion.* The mixed fractional Brownian motion corresponds to the model (1) with $\beta_i = 2$ for all $i$. It has been originally studied in Cheridito (2001). It has been shown in van Zanten (2007) that the measures induced by $C_t^1 = b_1 Y_t^{H_1,2}$ and $C_t^2 = b_1 Y_t^{H_1,2} + b_2 Y_t^{H_2,2}$, $t \in [0, 1]$ are equivalent if $H_2 > H_1 + \frac{1}{4}$, and singular otherwise. Hence, in the Gaussian case, the parameter $(b_j, H_j)$ is identifiable if and only if $H_i < H_1 + \frac{1}{4}$. Estimation of the parameters of a mixed fractional Brownian motion has been studied by Xiao, Zhang and Zhang (2011), although in a very restricted setting where the $H_j$ are assumed to be known. An estimator for a much more general nonstationary model with $d = 2$ and $H_2 < H_1 = \frac{1}{2}$ has been recently suggested by Chong, Delerue and Li (2021). To the best of our knowledge, the latter paper is the first investigation of inference for the general mixed fractional Brownian motion. An extension has been studied in Chong, Delerue and Mies (2022), where the two processes are allowed to be correlated, and optimal rates are derived for the special case of nonzero correlation.

(iii) *Mixed Lévy processes.* Some statistical results have been obtained in the setting of mixed Lévy processes, which corresponds to model (1) with $H_j = 1/\beta_j$ (in other words, $Y^{H,\beta} = Z^{\beta}$). If $\beta_1 < 2$ denotes the largest stability index, Aït-Sahalia and Jacod (2012) showed that the measures induced by $D_t^1 = b_1 Z_t^{\beta_1}$ and $D_t^2 = b_1 Z_t^{\beta_1} + b_2 Z_t^{\beta_2}$ are equivalent if $\beta_2 < \beta_1/2$, and singular otherwise. Thus, the parameter $(b_j, \beta_j)$ is identifiable if and only if $\beta_i > \beta_1/2$. See also Aït-Sahalia and Jacod (2008) for an earlier treatment of the special case $q = 2$. Another known case is given by mixed Lévy processes with the additional restriction that $\beta_1 = 2$, and $\beta_2 < 2$ being the largest of the remaining indices. In this setting, the identifiability condition becomes $\beta_j > \beta_2/2$. In other words, the identifiability of the stable components is unaffected by the presence of a Gaussian component. This can be explained by the observation that given a full trajectory of the process $X$, the continuous and the discontinuous components could be perfectly separated. In discrete samples, however, the presence of the Gaussian component has an adverse effect on the rate of convergence of any estimator of $\beta_j$, $j \geq 2$; see Aït-Sahalia and Jacod (2012). Note that this case is of particular interest for financial econometrics, where many models for high-frequency asset prices contain both, continuous and discontinuous components. Here, the sup of a Brownian motion and a stable Lévy motion may be regarded as the prototype of a semimartingale with infinite jump activity. The corresponding statistical problem has been first studied in the econometric literature in Aït-Sahalia and Jacod (2009), and more recently by Bull (2016) and Mies (2020).

Besides these three special cases, the mixed lfsm has not been studied in the literature. The currently unexplored regimes include, for example, the sum of fBm and a stable Lévy process; the sum of Brownian motion and lfsm; the sum of non-Gaussian lfsms; the sum of fBm and non-Gaussian lfsm. Our unified theory covers all these regimes, while matching the results in the cases, which have already been investigated.

1.3. *Outline of the paper.* The paper is structured as follows. In Section 2, we present a new limit theory for functionals of a mixed fractional stable motion. These results not only provide a necessary basis for statistical estimation, but also have an interest in their own right. Section 3 is devoted to statistical inference for mixed fractional stable motions. Here, we employ the idea of estimating equations that may specifically account for the presence of a Gaussian component (Section 3.2). In Section 3.3, we discuss the rates of convergence of our estimators, and we present a result on the singularity of measures induced by the mixed lfsm. In the Appendix, we derive a general result on consistency and asymptotic normality of solutions of estimating equations, which generalizes existing results in the literature (Appendix A). The proofs of the main results are presented in the Supplementary Material (Mies and Podolskij (2023), Appendix B).

1.4. *Notation.* Throughout the paper, we use the following notation. We denote by $C^p(\mathbb{R}^{d_1};\mathbb{R}^{d_2})$ the space of $p$ times continuously differentiable functions $f:\mathbb{R}^{d_1}\to\mathbb{R}^{d_2}$, and denote the first derivative matrix as $Df(x)_{i,j}=\frac{d}{dx_j}f_i(x)$. For a function $f:\mathbb{R}\to\mathbb{R}$, we denote its $k$th derivative as $f^{(k)}$. The space $L_p(\mathbb{R}^d)$ is the collection of all functions $f$ satisfying $\int_{\mathbb{R}^d}\|f(x)\|^p\,dx<\infty$. We write $a_n\sim b_n$ when there exist constants $c_1,c_2>0$ such that $c_1a_n\le b_n\le c_2a_n$ for all $n\ge 1$. The notation $a_n\gg b_n$ means that $b_n/a_n\to 0$ as $n\to\infty$. Throughout this paper, all positive constants are denoted by $C$, or by $C_p$ if we want to stress their dependence on some external parameter $p$, although they may change from line to line.

**2. Limit theorems for multiscale moving average processes.** In this section, we present a novel limit theorem for functionals of multiscale moving average processes, which will provide the theoretical basis for the statistical procedures investigated in the next section. Our main motivation comes from statistics of higher-order increments of $X$, which are defined as follows: For a frequency $\Delta_n\to 0$, $\gamma\in\mathbb{N}$ and $k\in\mathbb{N}$, the $k$th order increments of $X$ at the frequency $\gamma\Delta_n$ are given as

$$X_{l,n,\gamma}=\sum_{v=0}^{k}(-1)^v\binom{k}{v}X_{(l-v\gamma)\Delta_n},\quad l=1,\dots,n. \tag{2}$$

The factor $\gamma$ allows us to identify parameters of the lfsm's via their different temporal scaling. In particular, the self-similarity of the lfsm yields

$$\begin{aligned} X_{l,n,\gamma} &\overset{d}{=}\sum_{j=1}^{q}b_j\Delta_n^{H_j}\int_{-\infty}^{l}g_j(s-l)\,dZ_s^{\beta_j}\quad\text{where}\\ g_j(s)&=\sum_{v=0}^{k}\binom{k}{v}(-1)^v(v\gamma-s)_+^{H_j-\frac{1}{\beta_j}}. \end{aligned} \tag{3}$$

This higher-order differencing is common when working with fractional processes, as higher orders $k$ improve the decay of the autocovariances of $f(X_{l,n,\gamma})$ for suitably bounded functions $f$. In particular, the integral kernels decay as $g_j(s)\sim s^{H_j-k-1/\beta_j}$, making this effect evident.

Motivated by this example, we consider a more general class of discrete models, namely a multiscale array of moving averages $(X_{t,n})_{t\in\mathbb{N}}$ defined as

$$X_{t,n}=\sum_{i=1}^{q_1}a_{n,i}\int_{-\infty}^{t}h_i(s-t)\,dB_s^i+\sum_{j=1}^{q_2}b_{n,j}\int_{-\infty}^{t}g_j(s-t)\,dZ_s^j, \tag{4}$$

for nonnegative sequences $a_{n,i}$ and $b_{n,j}$. Here, for $i=1,\dots,q_1$, the $B^i$ are independent standard Brownian motions, and for $j=1,\dots,q_2$, the $Z^j$ are independent symmetric $\beta_j$-stable motions, standardized such that $\mathbb{E}\exp(i\lambda Z_1^j)=\exp(-|\lambda|^{\beta_j})$, $\beta_j\in(0,2)$. We allow the kernels $g$, $h$ to be $d$-dimensional, that is, $g_i,h_i:\mathbb{R}\to\mathbb{R}^d$. This allows us to account for different $\gamma$'s in (2) by deriving a multivariate limit theory. Throughout this section, the kernels are assumed to satisfy the following conditions:

(i) $h_i(x) = g_j(x) = 0$ for $x \geq 0$, (causality)
(ii) $h_i \in L_2(\mathbb{R})$ with $\|h_i(x)\| \leq C|x|^{\delta_0}$, and $\|h_i^l\|_{L_2} \leq \sigma < \infty$ for $h_i = (h_i^l)_{l=1}^d$,
(iii) $g_j \in L_{\beta_j}(\mathbb{R})$ and $\|g_j(x)\| \leq C|x|^{\delta_j}$ for some $\delta_j < -1/\beta_j$.

Definition (4) explicitly distinguishes between the Gaussian and the non-Gaussian components. As demonstrated in the sequel, this distinction is motivated by a qualitatively different effect of the Gaussian component on the limiting behavior.

Our focus is on statistics of the form

$$S_n(f) = \frac{1}{n} \sum_{t=1}^{n} [f(X_{t,n}) - \mathbb{E} f(X_{t,n})], \tag{5}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is a nonlinear function. We will assume that it belongs to the following class of functions.

DEFINITION 2.1. For $\eta \geq 0$, define the class $\mathfrak{F}_\eta$ of all functions $f : \mathbb{R}^d \to \mathbb{R}$ such that:

(F1) $f \in C^5(\mathbb{R}^d; \mathbb{R})$ and $\|D^j f\|_\infty \leq 1$ for $j = 0, \dots, 5$,
(F2) $f$ is even, and $f(0) = 0$,
(F3) $D^2 f(x) = D^2 f(0)$ for $\|x\| < \eta$.

Moreover, define the class $\mathfrak{F}_\eta^0 \subset \mathfrak{F}_\eta$, which additionally satisfies:

(F4) $f(x) = 0$ for $\|x\| < \eta$.

The class $\mathfrak{F}_\eta$ contains functions, which are quadratic on the interval $(-\eta, \eta)$, whereas functions in $\mathfrak{F}_\eta^0$ are smooth thresholds. Note that the class $\mathfrak{F}_0 = \mathfrak{F}_0^0$ imposes the least regularity.

The array $X_{t,n}$ is a superposition of multiple processes with long memory and heavy tails of different severity. Naively, one would expect that the limiting behavior of $S_n(f)$ is determined by the component with the largest scaling factor $a_{n,i}$, respectively, $b_{n,j}$. However, we observe two interesting phenomena: First, for the stable components, it is not the scaling factor $b_{n,j}$ itself which distinguishes the dominant component, but rather the power $b_{n,j}^{\beta_j}$, revealing an interesting interplay between the scale and the tail decay. Second, if $f$ vanishes near zero, then the effect of the Gaussian component is asymptotically negligible compared to the stable components, even if $\max_i a_{n,i} \gg \max_j b_{n,j}$. This is made precise by the following theorem.

THEOREM 2.2 (Variance bound). *Suppose that the exponent of the tail decay satisfies* $\delta^\star = \max(2\delta_0, \delta_1\beta_1, \dots, \delta_{q_2}\beta_{q_2}) < -2$, *and that* $a_{n,i}$ *and* $b_{n,j}$ *are bounded. Then for all* $n$, *and for all* $f \in \mathfrak{F}_0$:

$$\operatorname{Var}(S_n(f)) \leq \frac{C}{n} \left[ \sum_{i=1}^{q_1} a_{n,i}^4 + \sum_{j=1}^{q_2} b_{n,j}^{\beta_j} \right].$$

*If* $\eta > 0$, *then for any* $\lambda > 0$, *there exists a constant* $C \in (0, \infty)$ *such that for all* $n$, *and for all* $f \in \mathfrak{F}_\eta^0$,

$$\operatorname{Var}(S_n(f)) \leq \frac{C}{n} \left[ \left( \sum_{i=1}^{q_1} a_{n,i}^2 \right) \exp\left( -\frac{\eta^2 \lambda^2}{2d\sigma^2 \sum_{i=1}^{q_1} a_{n,i}^2} \right) + \sum_{j=1}^{q_2} b_{n,j}^{\beta_j} \right].$$

We remark that the exponential term in the variance bound for $\eta > 0$ vanishes rapidly if $a_{n,i} \to 0$ at a polynomial speed, such that the contribution of the Gaussian component is negligible in this case. That is, the smooth threshold $f \in \mathfrak{F}_\eta^0$ effectively filters out the Gaussian component.

As a consequence of Theorem 2.2, we obtain a law of large numbers with rate of convergence. In particular,

$$\frac{1}{n}\sum_{t=1}^{n} f(X_{t,n}) = \mathbb{E} f(X_{t,n}) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{q_1} a_{n,i}^4 + \sum_{j=1}^{q_2} b_{n,j}^{\beta_j}}\right). \tag{6}$$

We highlight that the centering term $\mathbb{E} f(X_{t,n})$ is not explicit as a function of $\theta$, and also depends on the index $n$.

The next theorem is the main result of this section. It demonstrates a central limit theorem for the statistic $S_n(f)$ in various settings.

THEOREM 2.3 (Central limit theorem). *Let $i^\star \in \{1, \ldots, q_1\}$ such that $a_{n,i^\star}^2 \gg \max_{i \neq i^\star} a_{n,i}^2$, and let $j^\star \in \{1, \ldots, q_2\}$ such that $b_{n,j^\star}^{\beta_{j^\star}} \gg \max_{j \neq j^\star} b_{n,j}^{\beta_j}$. Assume that the kernel decay satisfies* $\max(2\delta_0, \beta_1\delta_1, \ldots, \beta_{q_2}\delta_{q_2}) < -2$, *and suppose furthermore that*

$$a_{n,i^\star} \to a \in [0,\infty), \qquad b_{n,j^\star} \to b \in [0,\infty).$$

*Fix some $f \in \mathfrak{F}_0$, and let one of the following conditions hold*:

(i) $a > 0, b = 0$;
(ii) $a = 0, b > 0$;
(iii) $a = b = 0$, *and* $a_{n,i^\star}^4 \ll b_{n,j^\star}^{\beta_{j^\star}}$, $n b_{n,j^\star}^{\beta_{j^\star}} \to \infty$;
(iv) $a = b = 0$, *and* $a_{n,i^\star}^4 \gg b_{n,j^\star}^{\beta_j^\star}$, $n a_{n,i^\star}^4 \to \infty$ *and* $D^2 f(0) \neq 0$.

*Then*

$$\frac{\sqrt{n}}{\sqrt{\max(a_{n,i^\star}^4, b_{n,j^\star}^{\beta_{j^\star}})}} S_n(f) \quad \Rightarrow \quad \mathcal{N}(0, \xi^2),$$

*where the asymptotic variance $\xi^2$ is given in* (B.20) *in the Appendix.*

*Moreover, if* (*case* (v)):

- $a = b = 0$, $n a_{n,i^\star}^4 \to \infty$, $n b_{n,j^\star}^{\beta_{j^\star}} \to \infty$,
- $f_1 \in \mathfrak{F}_0$ *such that* $D^2 f(0) \neq 0$, *and* $f_2 \in \mathfrak{F}_\eta^0$ *for some* $\eta > 0$,
- *for some* $\lambda \in (0,1)$,

$$\exp\left(-\frac{\eta^2\lambda^2}{2d\sigma^2\sum_{i=1}^{q_1} a_{n,i}^2}\right) \ll b_{n,j^\star}^{\beta_{j^\star}} \ll a_{n,i^\star}^4,$$

*then*

$$\sqrt{n}\begin{pmatrix}\sqrt{a_{n,i^\star}^4} & 0 \\ 0 & \sqrt{b_{n,j^\star}^{\beta_{j^\star}}}\end{pmatrix}^{-1}\begin{pmatrix} S_n(f_1) \\ S_n(f_2)\end{pmatrix} \quad \Rightarrow \quad \mathcal{N}\left(0, \begin{pmatrix}\gamma_{f_1,0}^2 & 0 \\ 0 & \zeta_{f_2,0}^2\end{pmatrix}\right),$$

*and $\gamma_{f_1,0}^2$ and $\zeta_{f_2,0}^2$ are defined in* (B.20) *in the Appendix.*

We remark that limit theorems for the sum of Gaussian and non-Gaussian fractional processes are rather rare in the literature as the mathematical tools are different in the two cases. Indeed, Theorem 2.3 seems to be the first result in this setting.

There exist numerous central and noncentral limit theorems for statistics of lfsm's or more general Lévy moving average processes, all of them focusing on the case $q = 1$. Pipiras and Taqqu (2003) considered a bounded function $f$ and investigated some extensions to nonbounded functions in Pipiras, Taqqu and Abry (2007). The asymptotic theory for power variation statistics of lfsm has been studied in Basse-O'Connor, Lachièze-Rey and Podolskij (2017), Basse-O'Connor and Podolskij (2017) and later extended to more general functionals in Basse-O'Connor, Heinrich and Podolskij (2019). Further results on high frequency statistics of lfsm's and related models can be found in Mazur, Otryakhin and Podolskij (2020), Basse-O'Connor, Heinrich and Podolskij (2018) and Azmoodeh, Ljungdahl and Thäle (2022).

Notably, all available results for the high-frequency regime scale the increments by $\Delta_n^{-H}$, where $H$ is the Hurst exponent of the lfsm (cf. Theorem 2.3(ii)). This would correspond to the scaling $\Delta_n^{-H_1}$ in our mixed setting. In contrast, we also allow for different scalings, such that the argument of the nonlinearity vanishes (cases (iii) and (iv)), and we also allow for a dominant Gaussian component (case (i)). The special case of smooth thresholds in Theorem 2.3(v), has been studied for Lévy processes in Mies (2020), but not for processes with dependent increments.

The proof of the central limit theorem is performed by approximating $X_{t,n}$ via an $m$-dependent sequence. This is similar in spirit to the methodology of Pipiras and Taqqu (2003) and Basse-O'Connor, Heinrich and Podolskij (2019). However, in the setting of Theorem 2.3, we obtain more refined bounds on the autocovariances, which explicitly account for the scaling terms $a_{n,i}$ and $b_{n,j}$; see Lemma B.4 in the Appendix.

REMARK 1. One of the key conditions for the validity of the central limit theorem is the assumption $\max(2\delta_0, \beta_1\delta_1, \dots, \beta_{q_2}\delta_{q_2}) < -2$. Here, we would like to compare this condition with classical conditions in central limit theorems for statistics of Gaussian and non-Gaussian fractional processes.

We consider the non-Gaussian case first, that is, $q_1 = 0$, $q_2 = 1$ and let the kernel $g$ be given as in (3). Then the decay rate of the kernel is given as $\delta = H - k - 1/\beta$ and the condition of Theorem 2.3 becomes $H\beta - k\beta - 1 < -2$. Since the function $f$ is assumed to be even, its *Appell rank* is 2 or larger, and the latter condition coincides with the one from Basse-O'Connor, Heinrich and Podolskij (2019).

On the other hand, when considering the Gaussian case $q_1 = 1$ and $q_2 = 0$, our condition seems to be suboptimal (although sufficient for statistical applications). Indeed, when $k = 1$ the condition translates to $H < 1/2$. However, when dealing with functions of *Hermite rank* 2, the optimal condition is known to be $H < 3/4$.

## 3. Estimation of the mixed fractional stable motion.

Observing the representation (3), the characteristic function of the $k$th order increments takes the form

$$\mathbb{E}\cos(\lambda X_{l,n,\gamma}) = \exp\big(-\psi_n(\lambda,\gamma)\big), \quad \psi_n(\lambda,\gamma) = \sum_{j=1}^{q} \widetilde{b}_j \lambda^{\beta_j} \gamma^{\beta_j H_j} \Delta_n^{\beta_j H_j},$$

$$\widetilde{b}_j = b_j^{\beta_j} \int_{\mathbb{R}} \left| \sum_{v=0}^{k} \binom{k}{v} (-1)^v (v-s)_+^{H_j - \frac{1}{\beta_j}} \right|^{\beta_j} ds. \tag{7}$$

Since there is a one-to-one correspondence between $(b_j, H_j, \beta_j)$ and $(\widetilde{b}_j, H_j, \beta_j)$, we focus on the estimation of the latter. We summarize the unknown parameters as

$$\theta = (\widetilde{b}_1, H_1, \beta_1, \dots, \widetilde{b}_q, H_q, \beta_q) \in \Theta,$$

$$\Theta = \big\{\theta \in ((0,\infty) \times (0,1) \times (0,2])^q : H_1(\theta) < \cdots < H_q(\theta)\big\} \subset \mathbb{R}^{3q}.$$

In the sequel, $\theta_0$ will denote the true but unknown parameter vector.

Formula (7) for the characteristic exponents suggests that one can identify the stability indices $\beta_j$ by varying the scaling factor $\lambda$. This property can also be exploited for parameter estimation for Lévy processes and related semimartingales; see Aït-Sahalia and Jacod (2012), Bull (2016), Mies (2020), Reiß (2013). Moreover, formula (7) reveals that one can identify the Hurst exponents $H_j$ by varying the lag $\gamma$. This leads to the intuitive idea of using empirical characteristic functions to estimate the parameters of the model, which has been first proposed in Mazur, Otryakhin and Podolskij (2020) in the setting $q = 1$ (see also Ljungdahl and Podolskij (2021)). However, their method has several drawbacks, one of which is the need of preestimation of the smallest Hurst exponent $H_1$. It leads to singular limit distributions (cf. Mazur, Otryakhin and Podolskij (2020), Theorem 3.1) and its performance is far from obvious in the mixed case $q > 1$.

Instead we follow a different strategy, which relies on a system of estimating equations, and does not require prior information about the parameters. Section 3.1 presents the asymptotic theory for the new approach. Section 3.2 demonstrates a further refinement of the estimation method, the smooth threshold, which specifically accounts for the presence of the Gaussian component in the model.

3.1. *Estimation via adaptive moment equations.* We consider an even Schwartz function $f : \mathbb{R} \to \mathbb{R}$, that is, $f \in C^\infty(\mathbb{R})$ such that

$$\sup_{x\in\mathbb{R}} |x|^p \big| f^{(j)}(x) \big| < \infty$$

for all $p > 0$ and all $j \geq 0$, for example, $f(x) = \exp(-x^2/2)$. For $\lambda_r \in \mathbb{R}$ and $\gamma_r \in \mathbb{N}$, we define the statistic

$$S_n\big(f; u_n(\theta); (\lambda_r, \gamma_r)_{r=1}^{3q}\big) = \left[\frac{1}{n}\sum_{l=1}^{n} f\big(u_n(\theta)\lambda_r X_{l,n,\gamma_r}\big)\right]_{r=1}^{3q}, \qquad u_n(\theta) = \Delta_n^{-H_1(\theta)}.$$

We introduce the random vector

$$\mathcal{G}_n(\theta) = S_n\big(f; u_n(\theta); (\lambda_r, \gamma_r)_{r=1}^{3q}\big) - \mathbb{E}_\theta S_n\big(f; u_n(\theta); (\lambda_r, \gamma_r)_{r=1}^{3q}\big), \tag{8}$$

for some $\lambda_r, \gamma_r, r = 1, \ldots, 3q$. Note that the expectations in (8) can be determined numerically for any $\theta \in \Theta$, and hence the quantity $\mathcal{G}_n(\theta)$ can be computed from data. We now define an estimator $\widehat{\theta}_n$ of the unknown parameter $\theta_0$ as a solution of the equation

$$\mathcal{G}_n(\widehat{\theta}_n) = 0. \tag{9}$$

REMARK 2. To determine statistical properties of $\widehat{\theta}_n$, it is convenient to solve the equation $\mathcal{G}_n(\widehat{\theta}_n) = 0$ over an open set rather than over $\Theta$. This can be achieved by extending the parameter set $(0, 2]$ for $\beta$ to an arbitrary open set containing $(0, 2]$. While a $\beta$-stable distribution does not exist for $\beta > 2$, the expectation $\mathbb{E}_\theta f(\lambda X_{l,n,\gamma})$ can be extended to all values $\beta \in (0, \infty)$. Indeed, we obtain via Fourier transform:

$$\begin{aligned} \mathbb{E}_\theta f(\lambda X_{l,n,\gamma}) &= \int \widehat{f}(v) \exp\big(-\psi_n(v\lambda, \gamma, \theta)\big)\, dv, \\ \widehat{f}(v) &= \frac{1}{2\pi} \int \cos(vx) f(x)\, dx. \end{aligned} \tag{10}$$

Since $f$ is a Schwartz function, its Fourier transform $\widehat{f}(v)$ is a Schwartz function as well, and in particular decays rapidly as $v \to \infty$. On the other hand, definition (7) of the function $\psi_n$ is also sensible for $\beta_j > 2$, such that the integral (10) is finite. In the sequel, we use (10) as definition of the expectation for the case $\beta_j > 2$. In practice, we extend the original parameter set $\Theta$ to an arbitrary open set that contains $\Theta$.

To formulate our result on the asymptotic behavior of the estimator $\widehat{\theta}_n$, we introduce the following rate matrix:

$$\overline{C}_n^j(\theta) = \frac{\Delta_n^{\beta_j(\theta)(H_1(\theta)-H_j(\theta))}}{\sqrt{n}} \begin{pmatrix} 1 & -\widetilde{b}_j\beta_j \log(\Delta_n) & \widetilde{b}_j H_j \log(\Delta_n) \\ 0 & 1 & -H_j/\beta_j \\ 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 3},$$

$$\overline{C}_n(\theta) = \operatorname{diag}(\overline{C}_n^1, \dots, \overline{C}_n^q) \in \mathbb{R}^{3q\times 3q}.$$

We also define the matrix $\overline{W}(\theta) \in \mathbb{R}^{3q\times 3q}$ by

$$\overline{W}(\theta)_{i,r} = \int \widehat{f}(v) \exp(-\widetilde{b}_1 |\lambda_r v|^{\beta_1} \gamma_r^{\beta_1 H_1}) \partial_{\theta_i} \left( \sum_{j=1}^q \widetilde{b}_j \gamma_r^{\beta_j H_j} |\lambda_r v|^{\beta_j} \right) dv.$$

The main result of this section is the following central limit theorem.

THEOREM 3.1. *Suppose that the order of differencing $k$ is large enough, such that*

$$k > H_j + \frac{1}{\beta_j}, \quad j = 1, \dots, q.$$

*Choose $\lambda_r \in \mathbb{R}$, $\gamma_r \in \mathbb{N}$, $r = 1, \dots, 3q$, such that the matrix $\overline{W}(\theta_0)$ is regular. Assume the identifiability condition*

$$\frac{\Delta_n^{\beta_j(\theta_0)[H_1(\theta_0)-H_j(\theta_0)]}}{\sqrt{n}} \ll \frac{1}{|\log \Delta_n|^2}, \quad j = 1, \dots, q. \tag{11}$$

*Then there exists a sequence of random vectors $\widehat{\theta}_n$ such that $P(\mathcal{G}_n(\widehat{\theta}_n) = 0) \to 1$, which satisfies*

$$\overline{C}_n^{-1}(\widehat{\theta}_n - \theta_0) \quad \Rightarrow \quad \mathcal{N}(0, \Sigma),$$

*with asymptotic covariance matrix $\Sigma = \overline{W}(\theta_0)^{-1}\widetilde{\Sigma}(\overline{W}(\theta_0)^{-1})^T \in \mathbb{R}^{3q\times 3q}$, and $\widetilde{\Sigma} = \widetilde{\Sigma}(\theta_0) \in \mathbb{R}^{3q\times 3q}$ is given by formula (B.35) in the Appendix.*

The proof of Theorem 3.1 is based upon the asymptotic theory for the random vector $\mathcal{G}_n(\theta_0)$, which is obtained via an application of Theorem 2.3 from the previous section, and a general theory for solutions of estimating equations; see Theorem A.2. The solution of the estimating equation $\mathcal{G}_n(\hat{\theta}_n) = 0$ is asymptotically unique in the sense that for $n$ large enough, there exists at most one solution within a $r_n$-neighborhood of the true $\theta_0$ for $r_n = \|\widehat{\theta}_n - \theta_0\|^{1/2}/|\log \Delta_n|$; see Remark 5. Establishing global uniqueness of the system of nonlinear estimation equations requires different techniques and is beyond the scope of our investigation. In practice, the equations will be solved numerically, and the numerical solution will be chosen as estimator.

From the rate matrix $\overline{C}_n(\theta_0)$, we may derive constraints on the parameter $\theta_0$ to ensure that all components of the process may be estimated consistently. If $\Delta_n = n^{-\rho}$ for some $\rho \geq 0$, then $\|\widehat{\theta}_n - \theta_0\| \to 0$ if and only if

$$H_j < H_1 + \frac{1}{2\rho\beta_j}. \tag{12}$$

For the classical high-frequency regime $\rho = 1$, that is, $\Delta_n = 1/n$, this identifiability condition is sharp for two prominent special cases:

(i) If $\beta_j = 2$ for all $j$, then each $Y_t^{H_j,2}$ is a fractional Brownian motion, and (12) recovers the identifiability condition of van Zanten (2007).

(ii) If $H_j = 1/\beta_j$ and for all $j$, then each $Y_t^{1/\beta_j,\beta_j}$ is a $\beta_j$-stable Lévy process. If additionally $\beta_j < 2$ for all $j$, then (12) recovers the identifiability condition of Aït-Sahalia and Jacod (2012).

However, condition (12) is not sharp when the process contains both, Gaussian and non-Gaussian components. If we consider the special case $q = 2$ and $\beta_1 = 2$, $H_1 = 1/2$, $\beta_2 < 2$, $H_2 = 1/\beta_2$, it is well known that all parameters are identifiable as $\Delta_n \to 0$; see, for example, Aït-Sahalia and Jacod (2009). However, the moment estimator based on $\mathcal{G}_n$ requires the restriction $\beta_2 > 1$. We will address this issue in the next subsection. Another disadvantage of the estimator based on $\mathcal{G}_n$ is that regularity of $\overline{W}(\theta_0)$ is nontrivial to verify because $\overline{W}(\theta_0)$ depends nonlinearly on all parameters. This issue will also be addressed by the alternative estimator presented in the next subsection.

3.2. *Estimation via smooth thresholds*. In this section, we improve upon the estimator of Section 3.1 by exploiting the Gaussianity of some components in order to obtain weaker conditions for identifiability and faster rates of convergence. Here, to demonstrate the main ideas, we restrict our attention to a simplified model consisting of two Gaussian components and two non-Gaussian components, and we assume that the Gaussian component is dominant. To be specific, we study the model

$$\begin{aligned} &X_t = a_1 Y_t^{H_1,2} + a_2 Y_t^{H_2,2} + b_1 Y_t^{\overline{H}_1,\beta_1} + b_2 Y_t^{\overline{H}_2,\beta_2}, \\ &H_1 < \min(H_2, \overline{H}_1, \overline{H}_2), \qquad (\overline{H}_1 - H_1)\beta_1 < (\overline{H}_2 - H_1)\beta_2. \end{aligned} \tag{13}$$

Note that the fractional Brownian motion $Y^{H_1,2}$ is the dominant component on small scales, in the sense that $\gamma^{-H_1} X_{\gamma t} \Rightarrow a_1 Y_t^{H_1,2}$ as $\gamma \downarrow 0$. The model has ten parameters and we denote the corresponding parameter set with the constraints as above by $\overline{\Theta} \subset \mathbb{R}^{10}$. As before (see the transform in (7)), we switch to an equivalent representation of scale parameters and obtain

$$\theta = (\widetilde{a}_1, H_1, \widetilde{a}_2, H_2, \widetilde{b}_1, \overline{H}_1, \beta_1, \widetilde{b}_2, \overline{H}_2, \beta_2) \in \mathbb{R}^{10}.$$

As the Gaussian part $Y^{H_1,2}$ is dominant, it is particularly hard to estimate the non-Gaussian components of the model. To improve the results from the previous section, we employ the idea of a smooth threshold, which aims at filtering out the continuous part at small scales. To this end, we consider two Schwartz functions $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ such that, for some $\eta > 0$,

$$f_1''(0) \neq 0, \qquad f_2(x) = 0 \quad \text{for } |x| \le \eta.$$

Moreover, let $\lambda_{r,n}(\theta) = \lambda_r u_n(\theta)$ for $\lambda_r \in \mathbb{R}$, and $\gamma_r \in \mathbb{N}$, $r = 1, \ldots, 10$. We suggest to estimate $\theta \in \mathbb{R}^{10}$ as a solution $\widehat{\theta}_n$ of the estimating equation

$$\begin{aligned} &\mathcal{H}_n(\widehat{\theta}_n) = 0 \quad \text{where} \\ &\mathcal{H}_n(\theta) = \begin{pmatrix} S_n(f_1; u_n(\theta); (\lambda_r, \gamma_r)_{r=1}^4) - \mathbb{E}_\theta S_n(f_1; u_n(\theta); (\lambda_r, \gamma_r)_{r=1}^4) \\ S_n(f_2; u_n(\theta); (\lambda_r, \gamma_r)_{r=5}^{10}) - \mathbb{E}_\theta S_n(f_2; u_n(\theta); (\lambda_r, \gamma_r)_{r=5}^{10}) \end{pmatrix} \in \mathbb{R}^{10}. \end{aligned} \tag{14}$$

The first set of moments based on $f_1$ serve to identify the Gaussian components, just as in Section 3.1. The function $f_2$ represents a smooth threshold, as $f_2(u_n X_{l,n,\gamma}) \neq 0$ if and only if $X_{l,n,\gamma} > \eta/u_n$. The scaling factor $u_n$ may thus be seen as a reciprocal threshold value, and the moments based on the smooth threshold $f_2$ identify the stable components. In contrast to Section 3.1, we choose

$$\begin{aligned} &u_n(\theta) = w_n \Delta_n^{-H_1(\theta)} \\ &\text{where } \Delta_n^\epsilon \ll w_n \le \frac{\eta}{a_1 2^{\frac{k+7}{2}} \sqrt{|\log \Delta_n|}} \quad \forall \epsilon > 0. \end{aligned} \tag{15}$$

The crucial observation is that, for this choice of $u_n$, the variance of the smooth thresholds is of a smaller order compared to the empirical characteristic function; see Theorem 2.2. This occurs because the threshold is asymptotically unaffected by the dominant Gaussian component, and the asymptotic distribution is driven by the smaller stable component. The same idea has been employed in Mies (2020) to construct an estimator for Lévy processes, that is, $H_j = 1/\beta_j$.

To formulate our asymptotic result about the estimating equation (14), we define the rate matrix $R_n(\theta)$ as

$$R_n^j(\theta) = \frac{\Delta_n^{2(H_1-H_j)}}{\sqrt{n}} \begin{pmatrix} 1 & -2\widetilde{a}_j \log(\Delta_n) \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}, \quad j = 1, 2,$$

$$\overline{R}_n^j(\theta) = \frac{w_n^{\frac{\beta_1}{2}-\beta_j} \Delta_n^{\frac{\beta_1(\overline{H}_1-H_1)}{2}-\beta_j(\overline{H}_j-H_1)}}{\sqrt{n}} \begin{pmatrix} 1 & 0 & -\widetilde{b}_j \log|w_n| \\ 0 & 1 & -\overline{H}_j/\beta_j \\ 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 3}, \quad j = 1, 2,$$

$$R_n(\theta) = \operatorname{diag}\big(R_n^1, R_n^2, \overline{R}_n^1, \overline{R}_n^2\big) \in \mathbb{R}^{10\times 10}.$$

Furthermore, recalling the notation $\widehat{f}(v) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x) e^{-ivx}\,dx$ of $f$, we define

$$\underline{W}(\theta) = \begin{pmatrix} \underline{W}_1(\theta) & 0 \\ 0 & \underline{W}_2(\theta) \end{pmatrix} \in \mathbb{R}^{10\times 10},$$

$$\begin{aligned} \underline{W}_1(\theta)_{r,i} &= \int \widehat{f}_1(v) \partial_{\theta_i}\left( \sum_{j=1}^2 \widetilde{a}_j \gamma_r^{2H_j} |\lambda_r v|^2 \right) dv \\ &= f_1''(0)|\lambda_r|^2 \partial_{\theta_i}\left( \sum_{j=1}^2 \widetilde{a}_j \gamma_r^{2H_j} \right), \quad r, i = 1, \dots, 4, \end{aligned}$$

$$\underline{W}_2(\theta)_{r-4,i-4} = \int \widehat{f}_2(v) \partial_{\theta_i}\left( \sum_{j=1}^2 \widetilde{b}_j \gamma_r^{\beta_j \overline{H}_j} |\lambda_r v|^{\beta_j} \right) dv, \quad r, i = 5, \dots, 10.$$

The main result of this section is the following theorem.

THEOREM 3.2. *Let $\Delta_n \sim n^{-\rho}$ for some $\rho > 0$. Assume that the order of differencing is large enough, such that*

$$k > \max\left( H_j(\theta_0) + \frac{1}{2}, \overline{H}_j(\theta_0) + \frac{1}{\beta_j(\theta_0)} \right), \quad j = 1, 2.$$

*Assume furthermore that the following identifiability condition holds:*

$$\begin{aligned} & H_2 < H_1 + \frac{1}{4\rho}, \\ (16) \qquad & \overline{H}_1 < H_1 + \frac{1}{\rho\beta_1}, \\ & \overline{H}_2 < H_1 + \frac{1}{2\rho\beta_2} + \frac{\beta_1}{2\beta_2}(\overline{H}_1 - H_1), \end{aligned}$$

*and suppose that $\underline{W}(\theta_0)$ is a regular matrix. Then there exists a sequence of random vectors $\widehat{\theta}_n$ such that $\mathbb{P}(\mathcal{H}_n(\widehat{\theta}_n) = 0) \to 1$, which satisfies*

$$R_n(\theta_0)^{-1}(\widehat{\theta}_n - \theta_0) \quad \Rightarrow \quad \mathcal{N}(0, \Sigma),$$

*with asymptotic covariance matrix* $\Sigma = \underline{W}(\theta_0)^{-1} \operatorname{diag}(\Sigma_1, \Sigma_2)(\underline{W}(\theta_0)^{-1})^T \in \mathbb{R}^{10\times 10}$ *and* $\Sigma_1 \in \mathbb{R}^{4\times 4}$, $\Sigma_2 \in \mathbb{R}^{10\times 10}$ *are given by formula* (*B*.40), *respectively*, (*B*.41) *in the Appendix*.

REMARK 3. The contributions of $f_1$ and $(\lambda_1, \dots, \lambda_4)$ to $\underline{W}(\theta_0)$ and $\Sigma_1$ cancel, and $\Sigma_2$ does not depend on these parameters either. Thus, the choice of $f_1$ and $\lambda_1, \dots, \lambda_4$ does not affect the asymptotic variance.

Regularity of the matrix $\underline{W}(\theta_0)$ implicitly imposes assumptions on both, the parameters $\theta_0$ and the statistical design in terms of $\lambda_r$ and $\gamma_r$. If we choose the latter hyperparameters carefully, the regularity holds for almost all parameters $\theta$.

PROPOSITION 3.3. *Let* $f_2 \geq 0$, $f_2 \not\equiv 0$.

(i) *Choose* $(\gamma_1, \dots, \gamma_4) = (1, 2, 4, 8)$, *and* $\lambda_r \neq 0$ *for* $r = 1, \dots, 4$. *Then* $\underline{W}_1(\theta)$ *is regular for all* $\theta \in \Theta$.

(ii) *Choose* $(\gamma_5, \dots, \gamma_{10}) = (1, 2, 2, 4, 4, 8)$, *and* $(\lambda_5, \dots, \lambda_{10}) = (1, 2, 4, 8, 16, 32)$. *Then* $\underline{W}_2(\theta)$ *is regular if*

$$\beta_1(1 + H_1) \neq \beta_2(1 + H_2) \quad \textit{and} \quad \beta_1(2 + H_1) \neq \beta_2(2 + H_2).$$

(iii) *Choose* $(\gamma_5, \dots, \gamma_{10}) = (1, 2, 2, 4, 8, 8)$, *and* $(\lambda_5, \dots, \lambda_{10}) = (1, 1, 2, 1, 1, 2)$. *Then* $\underline{W}_2(\theta)$ *is regular if*

$$H_1\beta_1 \neq H_2\beta_2.$$

To circumvent the singular edge cases unveiled in Proposition 3.3, a potential solution would be to use both sets of moments, and solve a nonlinear least squares problem instead of the system of estimating equations. However, this extension is beyond the scope of the present paper.

REMARK 4. By virtue of Proposition 3.3, the regularity of $\underline{W}(\theta_0)$ is rather simple to verify. In contrast, the asymptotic analysis of the estimator presented in Section 3.1 requires regularity of the matrix $\overline{W}(\theta_0)$, which is rather unwieldy as it depends nonlinearly on all parameters due to the exponential term. For the matrix $\underline{W}(\theta_0)$, the dependence on all parameters except $\beta_j$ can be handled explicitly. This qualitative difference can be traced back to the different rescaling factors $u_n(\theta)$: in Section 3.1, for $u_n(\theta) = \Delta_n^{-H_1}$, the rescaled increments $u_n(\theta_0)X_{l,n,\gamma}$ converge weakly toward a $\beta_1$-stable random variable; in Section 3.2, we choose a scaling factor $u_n(\theta) \ll \Delta_n^{-H_1}$, hence $u_n(\theta_0)X_{l,n,\gamma} \xrightarrow{\mathbb{P}} 0$.

Following the same strategy as in the proof of Theorem 3.2, it is also possible to derive a central limit theorem for the estimator of Section 3.1 with the scaling factor $u_n(\theta) = w_n\Delta_n^{-H_1}$ for $w_n \to 0$, which will no longer require regularity of $\overline{W}(\theta_0)$, but rather of a matrix similar to $\underline{W}_2$. On the other hand, the rate will be slightly worse by a factor $w_n^\rho$ for some $\rho > 0$. Hence, we do not pursue this direction any further.

To instantiate the estimator $\widehat{\theta}_n$, we need to specify $f_1$, $f_2$, $\gamma$, $\lambda$ and the order of differencing $k$:

- The function $f_1$ should be a Schwartz function with $f''(0) \neq 0$ (locally quadratic); see Proposition 3.3. We suggest to choose $f_1(x) = \exp(-x^2/2)$, as it admits a simple and explicit Fourier transform. Moreover, in view of Remark 3 and Proposition A.1, we may choose $\lambda_1 = \cdots = \lambda_4 = 1$, and $(\gamma_1, \dots, \gamma_4) = (1, 2, 4, 8)$.

TABLE 1
*Identifiability conditions for the various components, in the sampling regime $\Delta_n = 1/n$, that is, $\rho = 1$*

| | Without threshold ($\mathcal{G}_n$) | With threshold ($\mathcal{H}_n$) |
|---|---|---|
| $H_1$ | $(0, 1)$ | $(0, 1)$ |
| $H_2$ | $< H_1 + \frac{1}{4}$ | $< H_1 + \frac{1}{4}$ |
| $\overline{H}_1$ | $< H_1 + \frac{1}{2\beta_1}$ | $< H_1 + \frac{1}{\beta_1}$ |
| $\overline{H}_2$ | $< H_1 + \frac{1}{2\beta_2}$ | $< H_1 + \frac{1}{2\beta_2} + \frac{\beta_1}{2\beta_2}(\overline{H}_1 - H_1)$ |

- The function $f_2$ should be a Schwartz function, which vanishes in a $\eta$-neighborhood of zero. That is, $f_2$ is a smooth threshold. Note that the specific choice of $\eta$ is somewhat redundant, as it may be compensated by a corresponding choice of $w_n$. Hence, we choose $\eta = 1$, and suggest to employ a function of the form

$$f_2(x) = \exp\left(-\frac{d_1}{(|x| - 1)_+}\right) \cdot \exp\left(-\frac{d_2}{(1 + d_3 - |x|)_+}\right), \quad d_1, d_2, d_3 > 0.$$

  These functions have also been employed in Mies (2020). The choice of $\gamma_5, \dots, \gamma_{10}$ and $\lambda_5, \dots, \lambda_{10}$ should be guided by Proposition 3.3. Note that case (iii) excludes mixtures of Lévy processes, which are a natural candidate model in many situations. Hence, it is preferable to use the specification of (ii). Ideally, we would like to choose $f_2$ and $\gamma_r, \lambda_r$ that minimize the asymptotic variance. Due to the unwieldy form of the asymptotic variance, this could only be achieved numerically.
- For $\beta_j \le 1$, the process is either a Lévy process or explosive. In many cases, we can hence safely assume $\beta_j > 1$, so that $k = 2$ will always satisfy the conditions of Theorem 3.2.

3.3. *Discussion.* For $\rho = 1$, and for estimation of the smoother Gaussian component, that is, $H_2$, the new smooth thresholding estimator presented in Section 3.2 still matches the identifiability restrictions of van Zanten (2007) and Aït-Sahalia and Jacod (2012). In contrast to Section 3.1, we may now also identify the stable Lévy processes if the dominant component is Gaussian, without the restriction $\beta_j > 1$. This is possible because the Gaussian component is effectively filtered by the smooth threshold. A direct comparison with the identifiability condition (12) from Section 3.1 is presented in Table 1. Since $\overline{H}_1 > H_1$, the conditions for the smooth threshold estimator, that is, based on the estimating equations $\mathcal{H}_n$, are strictly weaker.

The rates of convergence of both estimators are presented in Table 2. Again, the thresholding estimator based on $\mathcal{H}_n$ is strictly better than the estimator based on $\mathcal{G}_n$. Furthermore,

TABLE 2
*Rates of convergence in the sampling regime $\Delta_n = 1/n$, that is, $\rho = 1$*

| | Without threshold ($\mathcal{G}_n$) | With threshold ($\mathcal{H}_n$) |
|---|---|---|
| $H_1$ | $n^{-\frac{1}{2}}$ | $n^{-\frac{1}{2}}$ |
| $H_2$ | $n^{2(H_2-H_1)-\frac{1}{2}}$ | $n^{2(H_2-H_1)-\frac{1}{2}}$ |
| $\overline{H}_1$ | $n^{\beta_1(\overline{H}_1-H_1)-\frac{1}{2}}$ | $n^{\frac{\beta_1}{2}(\overline{H}_1-H_1)-\frac{1}{2}}(\log n)^{\frac{\beta_1}{4}}$ |
| $\overline{H}_2$ | $n^{\beta_2(\overline{H}_2-H_1)-\frac{1}{2}}$ | $n^{\beta_2(\overline{H}_2-H_1)-\frac{\beta_1}{2}(\overline{H}_1-H_1)-\frac{1}{2}}(\log n)^{\frac{\beta_2}{2}-\frac{\beta_1}{4}}$ |

we can assess the rates for the special Lévy case $H_1 = 1/2$ and $\overline{H}_1 = 1/\beta_1$, $\overline{H}_2 = 1/\beta_2$, with $\beta_1 > \beta_2$, which has been studied by Mies (2020) and Aït-Sahalia and Jacod (2012). In this case, Theorem 3.2 yields, for $\Delta_n = 1/n$,

$$\widehat{\beta}_1 - \beta_1 = \mathcal{O}_{\mathbb{P}}\big((n/\log n)^{-\frac{\beta_1}{4}}\big), \qquad \widehat{\beta}_2 - \beta_2 = \mathcal{O}_{\mathbb{P}}\big((n/\log n)^{\frac{\beta_1}{4}-\frac{\beta_2}{2}}\big).$$

These are the same rates of convergence, which have been achieved for the Lévy case, and which are conjectured to be optimal for this setting; see the discussion in Mies (2020).[1]

As another benchmark, we may investigate the regime $H_2 = \frac{1}{2}$, $H_1 \in (1/4, 1/2)$, which corresponds to a sum of a classical Brownian motion and a rougher fractional Brownian motion. This setting has been studied by Chong, Delerue and Li (2021), in a generalized model allowing for additional nonstationarity. Both of our estimators, either based on $\mathcal{G}_n$ or on $\mathcal{H}_n$, yield

$$\widehat{H}_1 - H_1 = \mathcal{O}_{\mathbb{P}}\big(n^{-\frac{1}{2}}\big), \qquad \widehat{\widetilde{a}}_1 - \widetilde{a}_1 = \mathcal{O}_{\mathbb{P}}\big(n^{-\frac{1}{2}}\log n\big), \qquad \widehat{\widetilde{a}}_2 - \widetilde{a}_2 = \mathcal{O}_{\mathbb{P}}\big(n^{\frac{1}{2}-2H_1}\log n\big).$$

The rate for $H_1$ is identical to the rate of Chong, Delerue and Li (2021); see Theorem 4.5 therein. On the other hand, our rate for $\widetilde{a}_2$ is slower by a factor $\log n$, which is due to the fact that in our setting, $H_2$ is unknown and needs to be estimated as well.

Except for the special cases discussed above, there are currently no benchmarks for the mixed fractional stable motion (1). Hence, we do not know whether the conditions presented in Table 1 and the rates presented in Table 2 are sharp for all parameter regimes. In fact, we conjecture the results to be not sharp in at least the following two cases.

*Case (i)*: If $\rho \in (0, 1)$, we have $n\Delta_n \to \infty$, such that we effectively observe the process $X_t$ on the increasing interval $[0, T_n]$, $T_n \to \infty$. Since the linear fractional stable motion is ergodic, the same holds for $X_t$, and we should expect that all parameters can be estimated consistently in this regime.

*Case (ii)*: Suppose that $H_j > 1/\beta_j$ for all $j$, and that the process contains no Gaussian component, that is, $\beta_j \in (1, 2)$ for all $j$. In this regime, $X_t$ admits a continuous version. Interestingly, for different parameters $\theta \neq \theta'$ the measures $\mathbb{P}^\theta$ and $\mathbb{P}^{\theta'}$ induced by $(X_t)_{t\in[0,1]}$ on the path space $C[0, 1]$ are singular. This is a consequence of the following identifiability result, which is new and might be of independent interest. The proof is presented in the Appendix.

THEOREM 3.4. *Let $X_t = \sum_{j=1}^q b_j Y_t^{H_j,\beta_j}$ be a mixed fractional stable process, with $\beta_j \in (1, 2)$ for all $j = 1, \dots, q$, and $H_j > 1/\beta_j$. Assume that the parameters $(\beta_j, H_j)$ are pairwise distinct.*[2] *For any two parameters $\theta_1, \theta_2 \in \Theta$, $\theta_1 \neq \theta_2$, satisfying these requirements, with potentially different sizes $q(\theta_1) \neq q(\theta_2)$, the measures $P^{\theta_1}$ and $P^{\theta_2}$ induced on $C[0, 1]$ are mutually singular.*

This result suggests that there should exist a sequence of consistent estimators in the high-frequency regime $\rho = 1$, which would imply that our identifiability condition (12) is too restrictive. However, pairwise singularity of the measures is not sufficient for the existence of a consistent sequence of estimators; see Aït-Sahalia and Jacod (2014), 5.1.1. Hence, Theorem 3.4 does not yield a complete answer about identifiability of the parameters of the mixed lfsm.

---

[1] The formulas (3.2) and (3.3) in Mies (2020) contain an error where the term $n \log n$ should correctly be $n/\log n$. The latter rate is obtained by substituting the value for $u_n \asymp \sqrt{n/\log n}$ therein.

[2] The claim of the theorem is also valid if $H_j = H_{j'}$ for some $j \neq j'$, as long as $\beta_j \neq \beta_{j'}$. In this case, we sort the components of $\theta$ in lexicographical order in $H_j$ and $\beta_j$, such that $H_1 \leq H_2 \dots$, and $\beta_j < \beta_{j+1}$ if $H_j = H_{j+1}$.

Besides these theoretical questions, future research also needs to address various practical aspects regarding the estimation of the mixed stable motion. In particular, for our estimators, we need to choose the functions $f_1$ and $f_2$, the scaling parameter $u_n$, and the values for $\lambda_r$ and $\gamma_r$. All these hyperparameters may affect the asymptotic variance of the estimator in practice. Moreover, the number $q$ of components needs be determined in a data-driven way, raising questions of model selection. Nevertheless, the asymptotic results presented in this paper demonstrate the various intricacies of the mixed lfsm model, and they show that many models studied separately in the literature may in fact be treated by a unified statistical theory.

## APPENDIX A: ESTIMATING EQUATIONS

The estimators proposed in Section 3 fall into the broader framework of estimating equations. In this section, we present some asymptotic results for solutions of general estimating equations. Let $\Theta \subset \mathbb{R}^d$ be an open parameter and let $F_n(\theta)$, $\theta \in \Theta$ be a $d$-variate random vector. Typically, $F_n(\theta)$ is a set of moment equations, and a parameter $\theta$ can be estimated by solving $F_n(\widehat{\theta}_n) = 0$. A survey of the asymptotic theory of estimating equations is given by Jacod and Sørensen (2018). However, for the purpose of this paper, we need to extend their results. It should be noted that the theory presented in this section has already been applied implicitly in the proofs in Mies (2020).

In order to derive the asymptotic distribution of $\widehat{\theta}_n$, we impose the following conditions:

(E.1) There exists a sequence of regular matrices $A_n \in \mathbb{R}^{d\times d}$ such that $A_n F_n(\theta_0) \Rightarrow Z$ for some random vector $Z$.

(E.2) The mapping $\theta \mapsto F_n(\theta)$ is $C^1$. There exists a sequence $r_n$ of real numbers, and for each $\theta \in \Theta$ there exist sequences of regular (random) matrices $B_n(\theta)$, $C_n(\theta)$ and a regular matrix $W(\theta)$, such that

$$\sup_{\theta\in \mathbf{B}_{r_n}(\theta_0)} \|B_n(\theta) DF_n(\theta) C_n(\theta) - W(\theta)\| \xrightarrow{\mathbb{P}} 0,$$

$$\sup_{\theta\in \mathbf{B}_{r_n}(\theta_0)} \frac{\|C_n(\theta)\| \|B_n(\theta) A_n^{-1}\|}{r_n} \xrightarrow{\mathbb{P}} 0.$$

(E.3) The mapping $\theta \mapsto (B_n(\theta), C_n(\theta), W(\theta))$ is continuous in the sense that

$$\sup_{\theta\in \mathbf{B}_{r_n}(\theta_0)} \|B_n(\theta)B_n(\theta_0)^{-1} - I\| + \|C_n(\theta)C_n(\theta_0)^{-1} - I\| + \|W(\theta) - W(\theta_0)\| \xrightarrow{\mathbb{P}} 0.$$

If we are only interested in consistency, then (E.1) could be weakened:

(E.1)' There exists a sequence of regular matrices $A_n \in \mathbb{R}^{d\times d}$ such that $A_n F_n(\theta_0) = \mathcal{O}_P(1)$.

Note also that (E.2) and (E.3) imply the following, upon setting $B_n = B_n(\theta_0)$, $C_n = C_n(\theta_0)$ and $W = W(\theta_0)$.

(E.2)' The mapping $\theta \mapsto F_n(\theta)$ is $C^1$. There exists a sequence $r_n$ of real numbers, and for each $\theta \in \Theta$ there exist sequences of regular matrices $B_n$, $C_n$ and a regular matrix $W$, such that

$$\sup_{\theta\in \mathbf{B}_{r_n}(\theta_0)} \|B_n DF_n(\theta) C_n - W\| \xrightarrow{\mathbb{P}} 0,$$

$$\frac{\|C_n\| \|B_n A_n^{-1}\|}{r_n} \xrightarrow{\mathbb{P}} 0.$$

We only need (E.2)' for our theory, but conditions (E.2) and (E.3) might be easier to verify.

PROPOSITION A.1. *Conditions* (*E*.2) *and* (*E*.3) *imply condition* (*E*.2)'.

Jacod and Sørensen (2018) only consider the special case $A_n = B_n = C_n^{-1}$; see Condition 2.10 therein. In contrast, the asymptotic theory presented here allows for additional flexibility. The proof of Section 3.1 uses the case $A_n \neq C_n^{-1}$, and $B_n = I$ as identity matrix, whereas the proof of Section 3.2 also requires a nontrivial $B_n \neq I$.

The following theorem is the main result of this section. Its proof uses central ideas of Lemma 6.2 in Jacod and Sørensen (2018).

THEOREM A.2. *Let conditions* (*E*.1)' *and* (*E*.2)' *hold. Then there exists a sequence of random vectors* $\widehat{\theta}_n \in \Theta$ *such that* $\mathbb{P}(F_n(\widehat{\theta}_n) = 0) \to 1$ *and*

$$A_n B_n^{-1} W C_n^{-1} [\widehat{\theta}_n - \theta_0] = \mathcal{O}_{\mathbb{P}}(1).$$

*The sequence is locally unique in the sense that for any other sequence of random variables* $\widetilde{\theta}_n$ *such that* $\mathbb{P}(F_n(\widetilde{\theta}_n) = 0) \to 1$ *and* $\mathbb{P}(\|\widetilde{\theta}_n - \theta_0\| \leq r_n) \to 1$, *we have* $\mathbb{P}(\widehat{\theta}_n = \widetilde{\theta}_n) \to 1$. *If additionally* $(E.1)$ *holds, then as* $n \to \infty$,

$$A_n B_n^{-1} W C_n^{-1} (\widehat{\theta}_n - \theta_0) \quad \Rightarrow \quad -Z.$$

REMARK 5. Note that Theorem A.2 yields asymptotic uniqueness among all estimators, which converge at least with rate $r_n$. In finite samples, it might still occur that the solution of $F_n(\theta) = 0$ is not unique, and one would need to pick one of those solutions as an estimator. It might also even happen that for large $n$, the estimating equations have two solutions; only one of which yields a consistent estimator. To ensure that these inconvenient scenarios do not occur, one would need additional global properties of the function $F_n$. However, uniqueness of the solution of nonlinear systems of equations is a nontrivial mathematical issue in general. Hence, results about estimating equations typically only yield the existence of a suitable sequence of solutions, as formulated in Theorem A.2 above. Global uniqueness may then be derived on a case-by-case basis. If this is not possible, the solution of $F_n(\theta) = 0$ needs to be determined numerically in practice, and one may use the numerical solution as an estimator.

PROOF OF PROPOSITION A.1. Set $B_n = B_n(\theta_0)$, $C_n = C_n(\theta_0)$, $W = W(\theta_0)$. We have

$$\begin{aligned}\|B_n DF_n(\theta)C_n - W\| \leq {}& \|B_n(\theta)DF_n(\theta)C_n(\theta) - W(\theta)\| + \|W(\theta) - W(\theta_0)\| \\ & + \|B_n(\theta)DF_n(\theta)C_n(\theta) - B_n DF_n(\theta)C_n(\theta)\| \\ & + \|B_n DF_n(\theta)C_n(\theta) - B_n DF_n(\theta)C_n\|.\end{aligned}$$

The first two terms vanish uniformly for $\theta \in \mathbf{B}_{r_n}(\theta_0)$ by (E.2) and (E.3). Regarding the last two terms, we observe that

$$\|B_n DF_n(\theta)C_n(\theta) - B_n DF_n(\theta)C_n\| \leq \|B_n DF_n(\theta)C_n\| \|C_n(\theta)C_n^{-1} - I\| \xrightarrow{\mathbb{P}} 0.$$

The same holds for the remaining term. □

PROOF OF THEOREM A.2. Consistency with rate $r_n$: The equation $F_n(\widehat{\theta}_n) = 0$ holds if and only if $\widehat{\theta}_n$ is a fixed point of the function $\phi(\theta) = \theta - C_n W^{-1} B_n F_n(\theta)$. We show that $\phi$ is a contraction for $n$ sufficiently large. We use the fact that for two matrices $A, B \in \mathbb{R}^{d\times d}$, it holds that $\|AB\|_F = \|BA\|_F$ for the Frobenius norm. In particular, $\|A\|_F = \|C_n^{-1} A C_n\|_F$.

Thus,

$$\begin{aligned}\sup_{\theta\in\mathbf{B}_{r_n}(\theta_0)}\|D\phi(\theta)\|_F &= \sup_{\theta\in\mathbf{B}_{r_n}(\theta_0)}\|I - C_nW^{-1}B_nDF_n(\theta)\|_F\\ &= \sup_{\theta\in\mathbf{B}_{r_n}(\theta_0)}\|I - W^{-1}B_nDF_n(\theta)C_n\|_F\\ &\leq \sup_{\theta\in\mathbf{B}_{r_n}(\theta_0)}\|W^{-1}\|_{\mathrm{op}}\|B_nDF_n(\theta)C_n - W\|_F \xrightarrow{\mathbb{P}} 0.\end{aligned}$$

In the last step, $\|\cdot\|_{\mathrm{op}}$ denotes the Euclidean operator norm. Denote by $\Omega_n$ the event that

$$\Omega_n = \left\{\sup_{\theta\in\mathbf{B}_{r_n}(\theta_0)}\|D\phi(\theta)\|_F \leq \frac{1}{2}, \|C_nW^{-1}B_nF_n(\theta_0)\| \leq \frac{r_n}{3}\right\}.$$

Because $\|C_nW^{-1}B_nF_n(\theta_0)\| = \mathcal{O}_P(\|C_nW^{-1}B_nA_n^{-1}\|_F) \ll r_n$, we have $\mathbb{P}(\Omega_n) \to 1$. Now define the sequence $\theta_k$ recursively by $\theta_k = \phi(\theta_{k-1})$, and $\theta_0$ as above. On the event $\Omega_n$, $\phi$ is a contraction on $\mathbf{B}_{r_n}(\theta_0)$, and we need to show that the sequence $\theta_k$ satisfies $\theta_k \in \mathbf{B}_{r_n}(\theta_0)$. We may show by induction that

$$\begin{aligned}\|\theta_k - \theta_0\| &\leq \sum_{r=1}^{k}\|\theta_r - \theta_{r-1}\| \leq \sum_{r=1}^{k}2^{-r}\|\theta_1 - \theta_0\|\\ &\leq \|\theta_1 - \theta_0\|\\ &= \|C_nW^{-1}B_nF_n(\theta_0)\| \leq \frac{r_n}{3}.\end{aligned}$$

In particular, $\theta_k$ is a Cauchy sequence and converges to a limit value $\theta_\infty \in \mathbf{B}_{r_n}(\theta_0)$, which satisfies $\phi(\theta_\infty) = \theta_\infty$, that is, $F_n(\theta_\infty) = 0$. Moreover, $\theta_\infty$ is measurable since each $\theta_k$ is a measurable random variable.

Define $\widehat{\theta}_n = \theta_\infty$ on the event $\Omega_n$, and $\widehat{\theta}_n = \underline{\theta} \in \Theta$ otherwise, where $\underline{\theta}$ is some arbitrary but fixed parameter value. Then $\mathbb{P}(F_n(\widehat{\theta}_n) = 0) \geq \mathbb{P}(\Omega_n) \to 1$, and $\|\widehat{\theta}_n - \theta_0\| = \mathcal{O}_{\mathbb{P}}(r_n)$. On the event $\Omega_n$, $\phi$ is a contraction on $\mathbf{B}_{r_n}(\theta_0)$, such that $\widehat{\theta}_n$ is also the unique solution on this set. This yields the uniqueness result claimed in the theorem.

Asymptotic distribution: On the event $\Omega_n$, we apply the mean value theorem to obtain

$$\begin{aligned}0 = A_nF_n(\widehat{\theta}_n) &= A_nF_n(\theta_0) + A_n\widetilde{F}_n[\widehat{\theta}_n - \theta_0]\\ &= A_nF_n(\theta_0) + A_nB_n^{-1}[B_n\widetilde{F}_nC_n]C_n^{-1}[\widehat{\theta}_n - \theta_0],\end{aligned}$$

where $(\widetilde{F}_n)_{l,r} = \partial_{\theta_r}F_n(\widetilde{\theta}^l)$, for $l, r = 1, \dots, d$ and for some $\widetilde{\theta}^l$ on the line segment between $\theta_0$ and $\widehat{\theta}_n$. Hence, $\widetilde{\theta}^l \in \mathbf{B}_{r_n}(\theta_0)$. By (E.1), we obtain the weak limit

$$A_nB_n^{-1}[B_n\widetilde{F}_nC_n]C_n^{-1}[\widehat{\theta}_n - \theta_0] \quad \Rightarrow \quad -Z.$$

By (E.2)', $B_n\widetilde{F}_nC_n \xrightarrow{\mathbb{P}} W$ as $n \to \infty$. This also yields

$$\begin{aligned}&[A_nB_n^{-1}WC_n^{-1}]^{-1}[A_nB_n^{-1}[B_n\widetilde{F}_nC_n]C_n^{-1}]\\ &\quad = C_n[W^{-1}(B_n\widetilde{F}_nC_n)]C_n^{-1} \xrightarrow{\mathbb{P}} I_{d\times d}.\end{aligned}$$

Here, we use that $\|C_nMC_n^{-1} - I\|_F = \|M - I\|_F$. In particular, Slutsky's theorem yields

$$A_nB_n^{-1}WC_n^{-1}[\widehat{\theta}_n - \theta_0] \quad \Rightarrow \quad -Z.$$

If not (E.1), but only (E.1)' holds, then

$$A_nB_n^{-1}[B_n\widetilde{F}_nC_n]C_n^{-1}[\widehat{\theta}_n - \theta_0] = \mathcal{O}_{\mathbb{P}}(1),$$

and we may proceed analogously as for the central limit theorem. □

**Acknowledgments.** The authors would like to thank Carsten Chong for pointing out an error in a previous version, and two anonymous referees for their very careful reading and constructive comments, which helped to improve the article significantly.

This work was done while Fabian Mies was postdoctoral researcher at the Institute of Statistics, RWTH Aachen University, Germany.

**Funding.** Mark Podolskij gratefully acknowledges financial support of ERC Consolidator Grant 815703 "STAMFORD: Statistical Methods for High-Dimensional Diffusions".

## SUPPLEMENTARY MATERIAL

**Appendix B** (DOI: 10.1214/23-AOS2312SUPP; .pdf). Contains all technical proofs of the results of this paper.

## REFERENCES

ABRY, P., DELBEKE, L. and FLANDRIN, P. (1999). Wavelet based estimator for the self-similarity parameter of $\alpha$-stable processes. In 1999 *IEEE International Conference on Acoustics*, *Speech*, *and Signal Processing*. *Proceedings*. *ICASSP*99 1729–1732.

ABRY, P., PESQUET-POPESCU, B. and TAQQU, M. S. (1999). Estimation Ondelette Des Paramètres de Stabilité et d'autosimilarité Des Processus $\alpha$-Stables Autosimilaires. In $17^e$ *Colloque Sur Le Traitement Du Signal et Des Images*, *FRA*, 1999. *GRETSI*, *Groupe d'Etudes du Traitement du Signal et des Images*.

AÏT-SAHALIA, Y. and JACOD, J. (2008). Fisher's information for discretely sampled Lévy processes. *Econometrica* **76** 727–761. MR2433480 https://doi.org/10.1111/j.1468-0262.2008.00858.x

AÏT-SAHALIA, Y. and JACOD, J. (2009). Estimating the degree of activity of jumps in high frequency data. *Ann. Statist*. **37** 2202–2244. MR2543690 https://doi.org/10.1214/08-AOS640

AÏT-SAHALIA, Y. and JACOD, J. (2012). Identifying the successive Blumenthal–Getoor indices of a discretely observed process. *Ann. Statist*. **40** 1430–1464. MR3015031 https://doi.org/10.1214/12-AOS976

AÏT-SAHALIA, Y. and JACOD, J. (2014). *High-Frequency Financial Econometrics*. Princeton Univ. Press, Princeton, NJ.

ASTRAUSKAS, A. (1983). Limit theorems for sums of linearly generated random variables. *Lith. Math. J.* **23** 127–134. MR0706002

AYACHE, A. and HAMONIER, J. (2012). Linear fractional stable motion: A wavelet estimator of the $\alpha$ parameter. *Statist. Probab. Lett.* **82** 1569–1575. MR2930661 https://doi.org/10.1016/j.spl.2012.04.005

AZMOODEH, E., LJUNGDAHL, M. M. and THÄLE, C. (2022). Multi-dimensional normal approximation of heavy-tailed moving averages. *Stochastic Process. Appl.* **145** 308–334. MR4367889 https://doi.org/10.1016/j.spa.2021.11.011

BASSE-O'CONNOR, A., HEINRICH, C. and PODOLSKIJ, M. (2018). On limit theory for Lévy semi-stationary processes. *Bernoulli* **24** 3117–3146. MR3779712 https://doi.org/10.3150/17-BEJ956

BASSE-O'CONNOR, A., HEINRICH, C. and PODOLSKIJ, M. (2019). On limit theory for functionals of stationary increments Lévy driven moving averages. *Electron. J. Probab.* **24** Paper No. 79, 42. MR4003132 https://doi.org/10.1214/19-ejp336

BASSE-O'CONNOR, A., LACHIÈZE-REY, R. and PODOLSKIJ, M. (2017). Power variation for a class of stationary increments Lévy driven moving averages. *Ann. Probab.* **45** 4477–4528. MR3737916 https://doi.org/10.1214/16-AOP1170

BASSE-O'CONNOR, A. and PODOLSKIJ, M. (2017). On critical cases in limit theory for stationary increments Lévy driven moving averages. *Stochastics* **89** 360–383. MR3574707 https://doi.org/10.1080/17442508.2016.1191493

BULL, A. D. (2016). Near-optimal estimation of jump activity in semimartingales. *Ann. Statist.* **44** 58–86. MR3449762 https://doi.org/10.1214/15-AOS1349

CHERIDITO, P. (2001). Mixed fractional Brownian motion. *Bernoulli* **7** 913–934. MR1873835 https://doi.org/10.2307/3318626

CHONG, C., DELERUE, T. and LI, G. (2021). When frictions are fractional: Rough noise in high-frequency data. Available at arXiv:2106.16149.

CHONG, C., DELERUE, T. and MIES, F. (2022). Rate-optimal estimation of mixed semimartingales. Available at arXiv:2207.10464.

DANG, T. T. N. and ISTAS, J. (2017). Estimation of the Hurst and the stability indices of a $H$-self-similar stable process. *Electron. J. Stat.* **11** 4103–4150. MR3715823 https://doi.org/10.1214/17-EJS1357

GRAHOVAC, D., LEONENKO, N. N. and TAQQU, M. S. (2015). Scaling properties of the empirical structure function of linear fractional stable motion and estimation of its parameters. *J. Stat. Phys.* **158** 105–119. MR3296276 https://doi.org/10.1007/s10955-014-1126-4

JACOD, J. and SØRENSEN, M. (2018). A review of asymptotic theory of estimating functions. *Stat. Inference Stoch. Process.* **21** 415–434. MR3824976 https://doi.org/10.1007/s11203-018-9178-8

LJUNGDAHL, M. M. and PODOLSKIJ, M. (2020). A minimal contrast estimator for the linear fractional stable motion. *Stat. Inference Stoch. Process.* **23** 381–413. MR4123929 https://doi.org/10.1007/s11203-020-09216-2

LJUNGDAHL, M. M. and PODOLSKIJ, M. (2021). Multidimensional parameter estimation of heavy-tailed moving averages. *Scand. J. Stat.* **49** 593–624. MR4428498 https://doi.org/10.1111/sjos.12527

MAZUR, S., OTRYAKHIN, D. and PODOLSKIJ, M. (2020). Estimation of the linear fractional stable motion. *Bernoulli* **26** 226–252. MR4036033 https://doi.org/10.3150/19-BEJ1124

MIES, F. (2020). Rate-optimal estimation of the Blumenthal–Getoor index of a Lévy process. *Electron. J. Stat.* **14** 4165–4206. MR4175392 https://doi.org/10.1214/20-EJS1769

PIPIRAS, V. and TAQQU, M. S. (2003). Central limit theorems for partial sums of bounded functionals of infinite-variance moving averages. *Bernoulli* **9** 833–855. MR2047688 https://doi.org/10.3150/bj/1066418880

PIPIRAS, V., TAQQU, M. S. and ABRY, P. (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** 1091–1123. MR2364228 https://doi.org/10.3150/07-BEJ6143

REISS, M. (2013). Testing the characteristics of a Lévy process. *Stochastic Process. Appl.* **123** 2808–2828. MR3054546 https://doi.org/10.1016/j.spa.2013.03.016

STOEV, S., PIPIRAS, V. and TAQQU, M. S. (2002). Estimation of the self-similarity parameter in linear fractional stable motion. *Signal Process.* **82** 1873–1901.

STOEV, S. and TAQQU, M. S. (2005). Asymptotic self-similarity and wavelet estimation for long-range dependent fractional autoregressive integrated moving average time series with stable innovations. *J. Time Series Anal.* **26** 211–249. MR2122896 https://doi.org/10.1111/j.1467-9892.2005.00399.x

VAN ZANTEN, H. (2007). When is a linear combination of independent fBm's equivalent to a single fBm? *Stochastic Process. Appl.* **117** 57–70. MR2287103 https://doi.org/10.1016/j.spa.2006.05.013

XIAO, W.-L., ZHANG, W.-G. and ZHANG, X.-L. (2011). Maximum-likelihood estimators in the mixed fractional Brownian motion. *Statistics* **45** 73–85. MR2772157 https://doi.org/10.1080/02331888.2010.541254

MIES, F. and PODOLSKIJ, M. (2023). Supplement to "Estimation of mixed fractional stable processes using high-frequency data." https://doi.org/10.1214/23-AOS2312SUPP