University of Zagreb
Faculty of Science
Department of Mathematics

Danijel Grahovac

SCALING PROPERTIES OF STOCHASTIC PROCESSES WITH APPLICATIONS TO PARAMETER ESTIMATION AND SAMPLE PATH PROPERTIES

DOCTORAL THESIS

Zagreb, 2015

Danijel Grahovac 2015 DOCTORAL THESIS

University of Zagreb
Faculty of Science
Department of Mathematics

Danijel Grahovac

SCALING PROPERTIES OF STOCHASTIC PROCESSES WITH APPLICATIONS TO PARAMETER ESTIMATION AND SAMPLE PATH PROPERTIES

DOCTORAL THESIS

Supervisors:
Prof. Nikolai N. Leonenko
Izv. prof. dr. sc. Mirta Benšić

Zagreb, 2015

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Matematički odsjek

Danijel Grahovac

SVOJSTVA SKALIRANJA SLUČAJNIH PROCESA S PRIMJENAMA U PROCJENI PARAMETARA I SVOJSTVIMA TRAJEKTORIJA

DOKTORSKI RAD

Mentori:
Prof. Nikolai N. Leonenko
Izv. prof. dr. sc. Mirta Benšić

Zagreb, 2015.

Acknowledgements

First of all, I would like to thank my supervisor Prof. Nikolai Leonenko for his guidance, constant support and many interesting and memorable discussions. His admiring enthusiasm and commitment will always be a source of inspiration to me. I thank to my second supervisor, Prof. Mirta Benšić for her support and persistence in providing everything to make my research possible. I thank to Nenad Šuvak, my dear colleague and friend for his numerous advices and encouragements. I also thank to my collaborators, who shaped many of the parts of this thesis.

I am grateful to my parents and my brother for always being there for me and believing in me. A special thanks goes to my wife Maja for her love and understanding. Her constant support and encouragement was in the end what made this work possible.

Summary

Scaling properties of stochastic processes refer to the behavior of the process at different time scales and distributional properties of its increments with respect to aggregation. In the first part of the thesis, scaling properties are studied in different settings by analyzing the limiting behavior of two statistics: partition function and the empirical scaling function.

In Chapter 2 we study asymptotic scaling properties of weakly dependent heavy-tailed sequences. These results are applied on the problem of estimation of the unknown tail index. The proposed methods are tested against some existing estimators, such as Hill and the moment estimator.

In Chapter 3 the same problem is analyzed for the linear fractional stable noise, which is an example of a strongly dependent heavy-tailed sequence. Estimators will be developed for the Hurst parameter and stable index, the main parameters of the linear fractional stable motion.

Chapter 4 contains an overview of the theory of multifractal processes, which can be characterized in several different ways. A practical problem of detecting multifractal properties of time series is discussed from the point of view of the results of the preceding chapters.

The last Chapter 5 deals with the fine scale properties of the sample paths described with the so-called spectrum of singularities. The new results are given relating scaling properties with path properties and applied to different classes of stochastic processes.

Keywords: partition function, scaling function, heavy-tailed distributions, tail index, linear fractional stable motion, Hurst parameter, multifractality, Hölder continuity, spectrum of singularities

Sažetak

Važnost svojstava skaliranja slučajnih procesa prvi je put istaknuta u radovima Benoita Mandelbrota. Najpoznatije svojstvo skaliranja u teoriji slučajnih procesa je sebi-sličnost. Pojam multifraktalnosti pojavio se kasnije kako bi opisao modele s bogatijom strukturom skaliranja. Jedan od načina kako se skaliranje može izučavati jest korištenjem momenata procesa i tzv. funkcije skaliranja. Multifraktalni procesi mogu se karakterizirati kao procesi s nelinearnom funkcijom skaliranja. Ovaj pristup prirodno nameće jednostavnu metodu detekcije multifraktalnih svojstava procjenjivanjem funkcije skaliranja korištenjem tzv. particijske funkcije.

Prvi dio ovog rada bavi se statističkim svojstvima takvih procjenitelja s obzirom na različite pretpostavke. Najprije će se analizirati asimptotsko ponašanje empirijske funkcije skaliranja za slučaj slabo zavisnih nizova s teškim repovima. Preciznije, promatrat će se stacionarni nizovi sa svojstvom eksponencijalno brzog jakog miješanja koji imaju marginalne distribucije u klasi distribucija s teškim repovima. Dobiveni rezultati bit će iskorišteni za definiranje metoda procjene repnog indeksa te će biti napravljena usporedba s postojećim procjeniteljima kao što su Hillov i momentni procjenitelj. Osim toga, predložit ćemo i grafičku metodu temeljenu na obliku procijenjene funkcije skaliranja koja može detektirati teške repove u uzorcima.

U sljedećem koraku analizirat će se asimptotska svojstva funkcije skaliranja na jako zavisnim stacionarnim nizovima. Za primjer takvog niza koristit ćemo linearni frakcionalni stabilni šum čija svojstva su određena s dva parametra, indeksom stabilnosti i Hurstovim parametrom. Pokazat ćemo da u ovom slučaju funkcija skaliranja ovisi o vrijednostima ta dva parametra. Na osnovu tih rezultata, definirat će se metode za istodobnu procjenu oba parametra koje predstavljaju alternativu standardnim procjeniteljima.

U drugom dijelu rada prethodno uspostavljeni rezultati će biti analizirani s aspekta multifraktalnih slučajnih procesa. U prvom redu, dobiveni rezultati pokazuju da nelinearnosti procijenjene funkcije skaliranja mogu biti posljedica teških repova distribucije uzorka. Takav zaključak dovodi u pitanje metodologiju temeljenu na particijskoj funkciji.

Svojstva skaliranja često se isprepliću sa svojstvima putova procesa. Osim u terminima globalnih karakteristika kao što su momenti, multifraktalni slučajni procesi često se definiraju i u terminima lokalnih nepravilnosti svojih trajektorija. Nepravilnosti u trajektorijama mogu se mjeriti formiranjem skupova vremenskih točaka u kojima put procesa

ima isti Hölderov eksponent u točki. Hausdorffova dimenzija takvih skupova u ovisnosti o Hölderovom eksponentu naziva se spektar singulariteta ili multifraktalni spektar. Multifraktalni slučajni procesi mogu se karakterizirati kao procesi koji imaju netrivijalan spektar, u smislu da je spektar konačan u više od jedne točke. Dvije definicije mogu se povezati tzv. multifraktalnim formalizmom koji predstavlja tvrdnju da su funkcija skaliranja i spektar singulariteta Legendreova transformacija jedno drugoga. Brojna istraživanja usmjerena su na uvjete pod kojima multifraktalni formalizam vrijedi. Spektar singulariteta dosad je izveden za mnoge primjere slučajnih procesa, kao što su frakcionalno Brownovo gibanje, Lévyjevi procesi i multiplikativne kaskade.

Rezultati o asimptotskom obliku funkcije skaliranja pokazat će da u nekim slučajevima procjena beskonačnih momenata može dati točan spektar korištenjem multifraktalnog formalizma. Ova činjenica motivira dublje istraživanje odnosa između momenata i svojstava trajektorija kojim se bavimo u posljednjem dijelu rada. Hölder neprekidnost i skaliranje momenata povezani su poznatim Kolmogorovljevim teoremom neprekidnosti. S druge strane, dokazat ćemo svojevrsni komplement Kolmogorovljevog teorema koji povezuje momente negativnog reda s izostankom Hölder neprekidnosti trajektorije u svakoj točki. Ova tvrdnja bit će dodatno pojačana formulacijom u terminima momenata negativnog reda maksimuma nekog fiksnog broja prirasta procesa. Iz ovih rezultata, između ostalog, slijedit će da sebi-slični procesi s konačnim momentima imaju trivijalan spektar (npr. frakcionalno Brownovo gibanje). Obratno, svaki sebi-sličan proces s netrivijalnim spektrom mora imati teške repove (npr. stabilni Lévyjevi procesi). Dobiveni rezultati sugeriraju prirodnu modifikaciju particijske funkcije koja će biti testirana na nizu primjera.

Ključne riječi: particijska funkcija, funkcija skaliranja, distribucije s teškim repovima, repni indeks, linearno frakcionalno stabilno gibanje, Hurstov parametar, multifraktalnost, Hölder neprekidnost, spektar singulariteta

Contents

Chapter 1

Introduction

The importance of scaling relations was first stressed in the work of Benoit Mandelbrot. The early references are the seminal papers Mandelbrot (1963) and Mandelbrot (1967); see also Mandelbrot (1997). Scaling properties of stochastic processes refer to the behavior of the process at different time scales. This usually accounts to changes in finite dimensional distributions of the process when the time parameter is scaled by some factor. The best known scaling relation in the theory of stochastic processes is self-similarity. The scaling of time of the self-similar processes by some constant $a > 0$ results in scaling the state space by a factor $b > 0$, in the sense of finite dimensional distributions. More precisely, a stochastic process $\{X(t), t \geq 0\}$ is said to be self-similar if for any $a > 0$ there exists $b > 0$ such that

$$\{X(at)\} \stackrel{d}{=} \{bX(t)\},$$

where equality is in finite dimensional distributions. Suppose $\{X(t)\}$ is self-similar, non-trivial (meaning it is not a.s. constant for every t) and stochastically continuous at 0, that is for every $\varepsilon > 0$, $P(|X(t) - X(0)| > \varepsilon) \to 0$ as $t \to 0$. Then b must be of the form a^H for some $H \geq 0$, i.e.

$$\{X(at)\} \stackrel{d}{=} \{a^H X(t)\}.$$

Constant H is called the Hurst parameter or the self-similarity index. The importance of self-similar processes may be illustrated by the Lamperti's theorem, which states that the only possible limit (in the sense of finite dimensional distributions) of a normalized partial sum process of stationary sequences are self-similar processes with stationary increments (see Embrechts & Maejima (2002) for more details).

As a generalization of self-similarity, models allowing a richer form of scaling were introduced by Yaglom as measures to model turbulence (Yaglom (1966)). Later these models were called multifractal in the work of Frisch and Parisi (Frisch & Parisi (1985)). The concept can be easily generalized to stochastic processes, thus extending the notion of self-similar processes by allowing the factor a^H to be random. Of course, in many examples there is no such exact scaling of finite dimensional distributions as in the case of self-similar or multifractal processes.

If we have a sequence of random variables $(Y_i, i \in \mathbb{N})$, then we can also speak about scaling properties of the partial sum process $\{\sum_{i=1}^{n} Y_i, n \in \mathbb{N}\}$. For example, if $(Y_i, i \in \mathbb{N})$ is an independent identically distributed (i.i.d.) sequence with strictly α-stable distribution, $\alpha \in (0, 2]$, then we know that

$$\sum_{i=1}^{n} Y_i \overset{d}{=} n^{1/\alpha} Y_1, \quad \forall n \in \mathbb{N}. \tag{1.1}$$

The continuous time analog of this case corresponds to Brownian motion ($\alpha = 2$) and strictly α-stable Lévy processes, which are both self-similar with Hurst parameter $1/\alpha$. This parameter appears by taking logarithms in (1.1)

$$\frac{\ln |\sum_{i=1}^{n} Y_i|}{\ln n} \overset{d}{=} 1/\alpha + \frac{\ln |Y_1|}{\ln n}, \quad \forall n \in \mathbb{N},$$

and represents the rate of growth of the partial sum process measured as a power of n. The central limit theorem indicates that for all zero mean (if mean exists) i.i.d. sequences the relation (1.1) holds approximately for large n. Thus, in the general case, scaling can be studied as the behavior of the sequence with respect to aggregation and measured as the rate of growth of the partial sums.

We adopt this point of view and in the next section we define the so-called partition function (sometimes called empirical structure function). Partition function will be used for defining the so-called empirical scaling function. The names of the two come from the theory of multifractal processes, which is a topic we deal with in Chapter 4.

1.1 Partition function

Partition function is a special kind of the sample moment statistic based on the blocks of data. Given a sequence of random variables $Y_1, Y_2, \dots$ we define the partition function to be

$$S_q(n,t) = \frac{1}{\lfloor n/t \rfloor} \sum_{i=1}^{\lfloor n/t \rfloor} \left| \sum_{j=1}^{\lfloor t \rfloor} Y_{(i-1)\lfloor t \rfloor + j} \right|^q, \tag{1.2}$$

where $q \in \mathbb{R}$ and $1 \leq t \leq n$. In words, we partition the data into consecutive blocks of length $\lfloor t \rfloor$, we sum each block and take the power q of the absolute value of the sum. Finally, we average over all $\lfloor n/t \rfloor$ blocks. Notice that for $t = 1$ one gets the usual empirical q-th absolute moment.

The partition function can also be viewed as an estimator of the q-th absolute moment of the process with stationary increments. Indeed, suppose $\{X(t)\}$ is a process with stationary increments and one tries to estimate $E|X(t)|^q$ for fixed $t > 0$ based on a discretely observed sample $X_i = X(i)$, $i = 1, \dots, n$. The natural estimator is given by

$$\frac{1}{\lfloor n/t \rfloor} \sum_{i=1}^{\lfloor n/t \rfloor} \left| X_{i\lfloor t \rfloor} - X_{(i-1)\lfloor t \rfloor} \right|^q .$$

If we denote the one step increments as $Y_i = X(i) - X(i-1)$, then this is equal to (1.2).

In Chapters 2 and 3 we will study asymptotic properties of the partition function in two settings. Instead of keeping t fixed, we take it to be of the form $t = n^s$ for some $s \in (0, 1)$, which allows the blocks to grow as the sample size increases. The partition function will then have the following form

$$S_q(n, n^s) = \frac{1}{\lfloor n^{1-s} \rfloor} \sum_{i=1}^{\lfloor n^{1-s} \rfloor} \left| \sum_{j=1}^{\lfloor n^s \rfloor} Y_{\lfloor n^s \rfloor (i-1)+j} \right|^q . \tag{1.3}$$

Since $s > 0$, $S_q(n, n^s)$ will generally diverge as $n \to \infty$. We are interested in the rate of divergence of this statistic measured as a power of n. This can be obtained by considering the limiting behavior of

$$\frac{\ln S_q(n, n^s)}{\ln n}$$

as $n \to \infty$. One can think of this limiting value as the value of the smallest power of n needed to normalize the partition function in such a way that it will converge to some random variable not identically equal to zero.

1.2 Empirical scaling function

If $\{X(t)\}$ is a H-self-similar process with stationary increments, then $E|X(t)|^q = t^{Hq} E|X(1)|^q$ for $q \in \mathbb{R}$ such that $E|X(t)|^q < \infty$. Taking logarithms we have that

$$\ln E|X(t)|^q = Hq \ln t + \ln E|X(1)|^q.$$

Having in mind that $S_q(n, t)$ can be considered as the estimator of $E|X(t)|^q$, we can expect that $\ln S_q(n, t)$ will be linear in $\ln t$. This motivates considering the slope in the simple linear regression of $\ln S_q(n, t)$ on $\ln t$ based on some points $1 \le t_i \le n$, $i = 1, \dots, N$. These slopes for varying q will be called the empirical scaling function, although linear relation may not always be justified.

Given points $1 \le t_i \le n$, $i = 1, \dots, N$ and using the well known formula for the slope of the linear regression line, we define the empirical scaling function at the point q as

$$\hat{\tau}_{N,n}(q) = \frac{\sum_{i=1}^N \ln t_i \ln S_q(n, t_i) - \frac{1}{N} \sum_{i=1}^N \ln t_i \sum_{j=1}^N \ln S_q(n, t_j)}{\sum_{i=1}^N (\ln t_i)^2 - \frac{1}{N} \left(\sum_{i=1}^N \ln t_i \right)^2}. \tag{1.4}$$

If we write t_i in the form n^{s_i}, $s_i \in (0,1)$, $i = 1, \dots, N$, then the empirical scaling function is given by

$$\hat{\tau}_{N,n}(q) = \frac{\sum_{i=1}^{N} s_i \frac{\ln S_q(n,n^{s_i})}{\ln n} - \frac{1}{N} \sum_{i=1}^{N} s_i \sum_{j=1}^{N} \frac{\ln S_q(n,n^{s_j})}{\ln n}}{\sum_{i=1}^{N} (s_i)^2 - \frac{1}{N} \left(\sum_{i=1}^{N} s_i \right)^2}. \tag{1.5}$$

1.3 Overview

The partition function and the empirical scaling function will be used to study asymptotic scaling properties of different types of stationary sequences. In the next chapter we establish asymptotic behavior for weakly dependent heavy-tailed sequences and in Chapter 3 we do the same analysis for the linear fractional stable noise, which is an example of a heavy-tailed and strongly dependent sequence. Both results will have applications in the parameter estimation problem. In the first setting, we will propose an exploratory method and several estimation methods for the unknown tail index that will be compared with the existing estimators. In the second setting we establish methods for estimating Hurst exponent and stable index of the linear fractional stable motion.

In Chapter 4 we provide an overview of the theory of multifractal processes and consider the implications of the results of Chapters 2 and 3. The analysis will lead to a conclusion that, empirically, it is hard to distinguish multifractal and heavy-tailed processes. In Chapter 5 we study in more details the relation between fine path properties and moments of both positive and negative order. Such analysis will lead to a new definition of the partition function.

Chapter 2

Asymptotic scaling of weakly dependent heavy-tailed stationary sequences

In this chapter we establish limiting behavior of the partition function and the empirical scaling function introduced in (1.3) and (1.5). The results are applied in the tail index estimation problem, which is discussed in Section 2.2.

2.1 Asymptotic scaling

In order to establish our results, we first summarize the assumptions on the sequences considered in this chapter.

2.1.1 Assumptions

Through the chapter we assume that $(Y_i, i \in \mathbb{N})$ is a strictly stationary sequence of random variables. Each Y_i is assumed to have a heavy-tailed distribution with tail index α. This means that it has a regularly varying tail with index $-\alpha$ so that

$$P(|Y_i| > x) = \frac{L(x)}{x^\alpha},$$

where $L(t)$, $t > 0$ is a slowly varying function, that is, for every $t > 0$, $L(tx)/L(x) \to 1$ as $x \to \infty$. In particular, this implies that $E|X|^q < \infty$ for $0 < q < \alpha$ and $E|X|^q = \infty$ for $q > \alpha$, which is sometimes also used to define heavy tails. The parameter α is called the tail index and measures the "thickness" of the tails. Examples of heavy-tailed distributions include Pareto, stable and Student's t-distribution, which will be precisely defined in Subsections 2.2.3 and 2.2.5. For more details on heavy-tailed distributions, regular variation and related topics see Embrechts et al. (1997) and Resnick (2007).

We also impose some assumptions on the dependence structure of the sequence, which go beyond the independent case. First, for two sub-σ-algebras, $\mathcal{A} \subset \mathcal{F}$ and $\mathcal{B} \subset \mathcal{F}$ on the same complete probability space $(\Omega, \mathcal{F}, P)$ we define

$$a(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(A \cap B) - P(A)P(B)|.$$

Now for a process $\{Y_t, t \in \mathbb{N}\}$ or $\{Y_t, t \in [0, \infty)\}$, consider $\mathcal{F}_t = \sigma\{Y_s, s \leq t\}$ and $\mathcal{F}^{t+\tau} = \sigma\{Y_s, s \geq t + \tau\}$. We say that $\{Y_t\}$ has a strong mixing property if $a(\tau) = \sup_{t \geq 0} a(\mathcal{F}_t, \mathcal{F}^{t+\tau}) \to 0$ as $\tau \to \infty$. Strong mixing is sometimes also called α-mixing. If $a(\tau) = O(e^{-b\tau})$ for some $b > 0$ we say that the strong mixing property has an exponentially decaying rate. We will refer to $a(\tau)$ as the strong mixing coefficient function. Through this chapter $(Y_i, i \in \mathbb{N})$ is assumed to have the strong mixing property with an exponentially decaying rate.

In some arguments the proof of the main result of this chapter relies on the limit theory for partial maxima of absolute values of the sequence $(Y_i, i \in \mathbb{N})$. It is well known that for the i.i.d. sequence $(Z_i, i \in \mathbb{N})$ having regularly varying tail with index $-\alpha$ there exists a sequence of the form $n^{1/\alpha} L_1(n)$ with L_1 slowly varying, such that $\max_{i=1,\dots,n} |Z_i|/(n^{1/\alpha} L_1(n))$ converges in distribution to a Fréchet random variable whose distribution is one of the three types of distributions that can occur as a limit law for maxima (see Embrechts et al. (1997) for more details). Following Leadbetter et al. (1982), this can be extended to weakly dependent stationary sequence $(Y_i, i \in \mathbb{N})$ under additional assumptions. We say that $(Y_i, i \in \mathbb{N})$ has extremal index θ if for each $\tau > 0$ there exists a sequence $u_n(\tau)$ such that $nP(|Y_1| > u_n(\tau)) \to \tau$ and $P(\max_{i=1,\dots,n} |Y_i| \leq u_n(\tau)) \to e^{-\theta\tau}$ as $n \to \infty$. If (Y_i) is strong mixing and Y_i heavy-tailed, then it is enough for this to hold for a single $\tau > 0$ in order for θ to be the extremal index. It always holds that $\theta \in [0, 1]$. The i.i.d. sequence $(Z_i, i \in \mathbb{N})$ such that $Z_i =^d Y_i$ for each i is called the associated independent sequence. If $\theta > 0$, then the limiting distribution of $\max_{i=1,\dots,n} |Y_i|$ is of the same type as the limit of the maximum of the associated independent sequence with the same norming constants. In particular, if $\theta > 0$, under our assumptions $\max_{i=1,\dots,n} |Y_i|/(n^{1/\alpha} L_1(n))$ converges in distribution to a Fréchet random variable, possibly with different scale parameter. For our consideration in this chapter, we assume $(Y_i, i \in \mathbb{N})$ has positive extremal index. The case when $\theta = 0$ or does not exist is considered as degenerate and only a few examples are known where this happens under some type of mixing condition assumed (see (Leadbetter et al. 1982, Chapter 3) and references therein). In particular, $\theta > 0$ holds for any example considered later in this chapter.

2.1.2 Main theorems

Asymptotic properties of the partition function $S_q(n,t)$ have been considered before in the context of multifractality detection (Heyde (2009), Sly (2005); see also Heyde & Sly (2008)). Notice that if we keep t fixed, behavior of $S_q(n,t)$ as $n \to \infty$ accounts to the standard limit theory for partial sums of the sequence

$$\left| \sum_{j=1}^{\lfloor t \rfloor} Y_{(i-1)\lfloor t \rfloor + j} \right|^q, \quad i = 1, 2, \dots. \tag{2.1}$$

If $(Y_i, i \in \mathbb{N})$ is i.i.d. and $q < \alpha$, the weak law of large numbers implies that $S_q(n,t)$ converges in probability to the expectation of (2.1) as $n \to \infty$. To get more interesting limit results and analyze the effect of the block size, we take $t = n^s$. It is clear that $S_q(n, n^s)$ will diverge as $n \to \infty$ and we will measure the rate of divergence of this statistic as a power of n. To obtain the limiting value, we analyze $\ln S_q(n,n^s)/\ln n$ representing the rate of growth.

The next theorem summarizes the main results on the rate of growth. We additionally assume that the sequence has a zero expectation in case $\alpha > 1$. For practical purposes, this is not a restriction as one can always subtract the mean from the starting sequence. For the case $\alpha \leq 1$ this is not necessary. The proof of the theorem is given in Subsection 2.1.3. A special case of this theorem has been proved in Sly (2005) and cited in Heyde (2009).

Theorem 1. *Let $(Y_i, i \in \mathbb{N})$ be a strictly stationary sequence that has a strong mixing property with an exponentially decaying rate, positive extremal index and suppose that $Y_i, i \in \mathbb{N}$ has a heavy-tailed distribution with tail index $\alpha > 0$. Suppose also that $EY_i = 0$ if $\alpha > 1$. Then for every $q > 0$ and every $s \in (0,1)$*

$$\frac{\ln S_q(n,n^s)}{\ln n} \xrightarrow{P} R_\alpha(q,s) := \begin{cases} \frac{sq}{\alpha}, & \text{if } q \leq \alpha \text{ and } \alpha \leq 2, \\ s + \frac{q}{\alpha} - 1, & \text{if } q > \alpha \text{ and } \alpha \leq 2, \\ \frac{sq}{2}, & \text{if } q \leq \alpha \text{ and } \alpha > 2, \\ \max\left\{ s + \frac{q}{\alpha} - 1, \frac{sq}{2} \right\}, & \text{if } q > \alpha \text{ and } \alpha > 2, \end{cases} \tag{2.2}$$

as $n \to \infty$, where $\xrightarrow{P}$ stands for convergence in probability.

In order to illustrate the effects of the theorem, consider the simple case in which $(Y_i, i \in \mathbb{N})$ is a zero mean (if $\alpha > 1$) i.i.d. sequence that is in the domain of normal attraction of some α-stable random variable, $0 < \alpha < 2$. This means that $\sum_{i=1}^n Y_i / n^{1/\alpha}$ converges in distribution to some random variable Z with α-stable distribution. A sufficient condition for this to hold is the regular variation of the tail (2.1.1) with L constant at infinity and the balance of the tails (see Gnedenko & Kolmogorov (1968) for more details). Suppose first that $q < \alpha$ and notice that

$$\frac{S_q(n, n^s)}{n^{\frac{sq}{\alpha}}} = \frac{1}{\lfloor n^{1-s} \rfloor} \sum_{i=1}^{\lfloor n^{1-s} \rfloor} \left| \frac{\sum_{j=1}^{\lfloor n^s \rfloor} Y_{\lfloor n^s \rfloor (i-1)+j}}{n^{\frac{s}{\alpha}}} \right|^q .$$

When $n \to \infty$, each of the internal sums converges in distribution to an independent copy of Z. Since $q < \alpha$, $E|Z|^q$ is finite, so the weak law of large numbers applies and shows that the average tends to some nonzero and finite limit. For the case $q > \alpha$, the weak law cannot be applied and the rate of growth will be higher:

$$\frac{S_q(n, n^s)}{n^{s+\frac{q}{\alpha}-1}} = \frac{\sum_{i=1}^{\lfloor n^{1-s} \rfloor} \left| \frac{\sum_{j=1}^{\lfloor n^s \rfloor} Y_{\lfloor n^s \rfloor (i-1)+j}}{n^{\frac{s}{\alpha}}} \right|^q}{n^{(1-s)\frac{q}{\alpha}}} .$$

Internal sums again converge to independent copies of Z. Since $|Z|^q$ has $(-\alpha/q)$-regularly varying tail, it will be in the domain of attraction of (α/q)-stable distribution. Centering is not necessary since $\alpha/q < 1$ and the limit (modulo possibly some slowly varying function) will be some positive random variable.

For the case $\alpha > 2$, the variance is finite and so the central limit theorem holds. When $q < \alpha$ the rate of growth has an intuitive explanation by arguments similar to those just given above. When $q > \alpha$, interesting things happen. Note that the asymptotics of the partition function is influenced by two factors: averaging and the weak law on the one side and distributional limit arguments on the other side. It will depend on s which of the two influences prevails. For larger s, $s + q/\alpha - 1 < sq/2$ and the rate will be as in the case $q < \alpha$, i.e.,

$$\frac{S_q(n, n^s)}{n^{\frac{sq}{2}}} = \frac{1}{\lfloor n^{1-s} \rfloor} \sum_{i=1}^{\lfloor n^{1-s} \rfloor} \left| \frac{\sum_{j=1}^{\lfloor n^s \rfloor} Y_{\lfloor n^s \rfloor (i-1)+j}}{n^{\frac{s}{2}}} \right|^q .$$

Internal sums converge in distribution to normal, which has every moment finite and the weak law applies. But for small s, the rate will be the same as that for the case $\alpha < 2$. What happens is that in this case internal sums have a small number of terms, so convergence to normal is slow, much slower than the effect of averaging. This is the reason why the rate is greater than $sq/2$.

Remark 1. Note that in general, the normalizing sequence for partial sums can be of the form $n^{1/\alpha} L(n)$ for some slowly varying function L. This does not affect the rate of growth. Indeed, if $Z_n / n^a L(n) \to^d Z$ for some non-negative sequence Z_n, then for every $\varepsilon > 0$,

$$P\left(\frac{\ln Z_n}{\ln n} < a - \varepsilon \right) = P\left(Z_n < n^{a-\varepsilon} \right) = \quad P\left(\frac{Z_n}{n^a L(n)} < \frac{1}{L(n) n^{\varepsilon}} \right)$$
$$\leq P\left(\frac{Z_n}{n^a L(n)} < \frac{1}{2n^{\varepsilon}} \right) \to 0,$$

since for n large enough $n^{-\varepsilon} < L(n) < n^{\varepsilon}$, i.e. $\ln L(n)/\ln n \to 0$. Similar argument applies for the upper bound. On the other hand, if $\ln Z_n/\ln n \to^P a$, then Z_n grows at the rate a in the sense that for every $\varepsilon > 0$ there exist constants $c_1, c_2 > 0$ such that $P(c_1 < Z_n/n^a < c_2) \geq 1 - \varepsilon$ for n large enough. This is sometimes denoted as $Z_n = \Theta_P(n^a)$.

Remark 2. A natural question arises from the previous discussion whether it is possible to identify a normalizing sequence and a distributional limit of $S_q(n, n^s)$. In some special cases the limit can be easily deduced. Suppose $(Y_i, i \in \mathbb{N})$ is an i.i.d. sequence with strictly α-stable distribution. When $q < \alpha$, the rate of growth will be sq/α. Dividing the partition function with $n^{sq/\alpha}$ and using the scaling property of stable distributions yields

$$\frac{S_q(n, n^s)}{n^{\frac{sq}{\alpha}}} = \frac{1}{\lfloor n^{1-s} \rfloor} \sum_{i=1}^{\lfloor n^{1-s} \rfloor} \left| \frac{\sum_{j=1}^{\lfloor n^s \rfloor} Y_{\lfloor n^s \rfloor (i-1)+j}}{n^{\frac{s}{\alpha}}} \right|^q \overset{d}{=} \frac{1}{\lfloor n^{1-s} \rfloor} \sum_{i=1}^{\lfloor n^{1-s} \rfloor} |Y_i|^q .$$

Since $q < \alpha$, $E|Y_i|^q < \infty$ and the weak law of large numbers implies

$$\frac{S_q(n, n^s)}{n^{\frac{sq}{\alpha}}} \overset{P}{\to} E|Y_1|^q, \quad n \to \infty.$$

On the other hand, when $q \geq \alpha$ the weak law cannot be applied and the rate of growth is $s + q/\alpha - 1$. Normalizing the partition function gives

$$\frac{S_q(n, n^s)}{n^{s+\frac{q}{\alpha}-1}} = \frac{\sum_{i=1}^{\lfloor n^{1-s} \rfloor} \left| \frac{\sum_{j=1}^{\lfloor n^s \rfloor} Y_{\lfloor n^s \rfloor (i-1)+j}}{n^{\frac{s}{\alpha}}} \right|^q}{n^{(1-s)\frac{q}{\alpha}}} \overset{d}{=} \frac{\sum_{i=1}^{\lfloor n^{1-s} \rfloor} |Y_i|^q}{n^{(1-s)\frac{q}{\alpha}}}.$$

Each $|Y_i|^q$ has $(-\alpha/q)$ regularly varying tail, so it will be in the domain of normal attraction of (α/q)-stable distribution. Since $\alpha/q < 1$, the centering is not needed and by the generalized central limit theorem it follows that

$$\frac{S_q(n, n^s)}{n^{s+\frac{q}{\alpha}-1}} \overset{d}{\to} Z, \quad n \to \infty,$$

with Z having (α/q)-stable distribution.

Using Theorem 1 we can establish asymptotic properties of the empirical scaling function defined by (1.5). First, we show how Theorem 1 can motivate the definition of the scaling function.

Using the notation of Theorem 1, we denote

$$\varepsilon_n := \frac{S_q(n, n^s)}{n^{R_\alpha(q,s)}}.$$

Taking logarithms and rewriting yields

$$\frac{\ln S_q(n, n^s)}{\ln n} = R_\alpha(q, s) + \frac{\ln \varepsilon_n}{\ln n}. \tag{2.3}$$

As follows from Remark 1, ε_n is bounded in probability from above and from bellow, thus, it makes sense to view (2.3) as a regression model of $\ln S_q(n, n^s)/\ln n$ on q and s with the model function $R_\alpha(q, s)$, where $\ln \varepsilon_n/\ln n$ are the errors. One should count on the intercept in the model due to the possible nonzero mean of an error. Notice that, when $\alpha \leq 2$, $R_\alpha(q, s)$ is linear in s, i.e. it can be written in the form $R_\alpha(q, s) = a(q)s + b(q)$ for some functions $a(q)$ and $b(q)$. This also holds if $\alpha > 2$ and $q \leq \alpha$. We can then rewrite (2.3) as

$$\frac{\ln S_q(n, n^s)}{\ln n} = a(q)s + b(q) + \frac{\ln \varepsilon_n}{\ln n}.$$

Fixing q gives the simple linear regression model of $\ln S_q(n, n^s)/\ln n$ on s, thus it makes sense to consider the slope of this regression. This is exactly the empirical scaling function (1.5). If $\alpha > 2$ and $q > \alpha$, $R_\alpha(q, s)$ is not linear in s due to the maximum term in (2.2). It is actually a broken line with the breakpoint depending on the values of q and α. However, this does not prevent us from considering statistic (1.5) anyway. This will be reflected as the peculiar nonlinear shape of the asymptotic scaling function.

Theorem 2. *Suppose that the assumptions of Theorem 1 hold and $\hat{\tau}_{N,n}$ is the empirical scaling function based on the points $s_1, \dots, s_N \in (0, 1)$. Let*

$$\tau_\alpha^\infty(q) = \begin{cases} \frac{q}{\alpha}, & \text{if } q \leq \alpha \text{ and } \alpha \leq 2, \\ 1, & \text{if } q > \alpha \text{ and } \alpha \leq 2, \\ \frac{q}{2}, & \text{if } 0 < q \leq \alpha \text{ and } \alpha > 2, \\ \frac{q}{2} + \frac{2(\alpha-q)^2(2\alpha+4q-3\alpha q)}{\alpha^3(2-q)^2}, & \text{if } q > \alpha \text{ and } \alpha > 2. \end{cases} \tag{2.4}$$

(i) If $\alpha \leq 2$ or $\alpha > 2$ and $q \leq \alpha$ then

$$\operatorname*{plim}_{n\to\infty} \hat{\tau}_{N,n}(q) = \tau_\alpha^\infty(q),$$

where plim *stands for the limit in probability.*

(ii) If $\alpha > 2$ and $q > \alpha$, suppose $s_i = i/N$, $i = 1\dots, N$. Then

$$\lim_{N\to\infty} \operatorname*{plim}_{n\to\infty} \hat{\tau}_{N,n}(q) = \tau_\alpha^\infty(q).$$

Theorem 2 shows that, asymptotically, the shape of the empirical scaling function in the setting considered significantly depends on the value of the tail index α. The limit from case (i) of Theorem 2 does not depend on the choice of points s_i in the computation of the empirical scaling function. In case (ii), we need additional assumptions as in this case we are estimating the slope while the underlying relation is actually nonlinear. Plots of the asymptotic scaling function τ_α^∞ for different values of α are shown in Figure 2.1. When $\alpha \leq 2$, the scaling function has the shape of a broken line (we will refer to this shape as *bilinear*). In this case the first part of the plot is a line with slope $1/\alpha > 1/2$ and the second part is a horizontal line with value 1. A break occurs exactly at the point α. In case $\alpha > 2$, τ_α^∞ is approximately bilinear, the slope of the first part is 1/2 and again the breakpoint is at the α. When α is large, i.e., $\alpha \to \infty$, it follows from (2.4) that $\tau_\alpha^\infty(q) \equiv q/2$. This case corresponds to a sequence coming from a distribution with all moments finite, e.g., an independent normally distributed sample. This line will be referred to as the baseline. In Figure 2.1 the baseline is shown by a dashed line. The cases $\alpha \leq 2$ ($\alpha = 0.5, 1, 1.5$) and $\alpha > 2$ ($\alpha = 2.5, 3, 3.5, 4$) are shown by dot-dashed and solid lines, respectively.

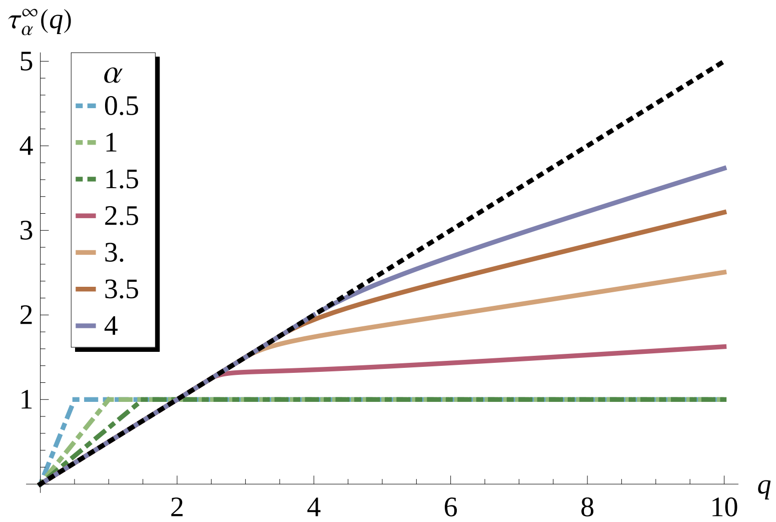


Figure 2.1: Plots of τ_α^∞ for different values of α

2.1.3 Proofs of Theorems 1 and 2

One of the main ingredients in the proof of Theorem 1 is the following version of Rosenthal's inequality for strong mixing sequences, precisely Theorem 2 in Section 1.4.1 of Doukhan (1994):

Lemma 1. *Fix $q > 0$ and suppose $(Z_k, k \in \mathbb{N})$ is a sequence of random variables and let $a_Z(m)$ be the corresponding strong mixing coefficient function. Suppose that there exists $\zeta > 0$ and $c \geq q, c \in \mathbb{N}$ such that*

$$\sum_{m=1}^{\infty} (m+1)^{2c-2} \left(a_Z(m)\right)^{\frac{\zeta}{2c+\zeta}} < \infty, \tag{2.5}$$

and suppose $E|Z_k|^{q+\zeta} < \infty$ and if $q > 1$, $EZ_k = 0$ for all k. Then there exists some *constant K depending only on q and $a_Z(m)$ such that*

$$E\left|\sum_{k=1}^{l} Z_k\right|^q \leq KD(q,\zeta,l),$$

where

$$D(q,\zeta,l) = \begin{cases} L(q,0,l), & \text{if } 0<q\leq 1,\\ L(q,\zeta,l), & \text{if } 1<q\leq 2,\\ \max\left\{L(q,\zeta,l),(L(2,\zeta,l))^{\frac{q}{2}}\right\}, & \text{if } q>2, \end{cases}$$

$$L(q,\zeta,l) = \sum_{k=1}^{l}\left(E\left|Z_k\right|^{q+\zeta}\right)^{\frac{q}{q+\zeta}}.$$

Remark 3. The inequality from Lemma 1 for $q \leq 1$ is a simple consequence of the fact that for $0 < q \leq 1$, $(a+b)^q \leq a^q + b^q$ for all $a, b \geq 0$. Therefore, in this case no assumption on the mixing is needed and more importantly Z_k are not required to be centered.

Proof of Theorem 1. We split the proof into three parts depending whether $q > \alpha$, $q < \alpha$ or $q = \alpha$.

(a) Let $q > \alpha$. First we show an upper bound for the limit in probability.
Let $\epsilon > 0$. Notice that

$$n^{\frac{\ln S_q(n,n^s)}{\ln n}} = S_q(n,n^s) = \frac{1}{\lfloor n^{1-s}\rfloor}\sum_{i=1}^{\lfloor n^{1-s}\rfloor}\left|\sum_{j=1}^{\lfloor n^s\rfloor} Y_{\lfloor n^s\rfloor(i-1)+j}\right|^q.$$

Let $\delta > 0$ and define

$$\begin{aligned} \bar{Y}_{j,n} &= Y_j\,\mathbf{1}\left(|Y_j| \leq n^{\frac{1}{\alpha}+\delta}\right), \quad j=1,\dots,n,\ n\in\mathbb{N},\\ Z_{j,n} &= \bar{Y}_{j,n} - E\bar{Y}_{j,n},\\ \xi_i &= \left|\sum_{j=1}^{\lfloor n^s\rfloor} Z_{n^s(i-1)+j,n}\right|^q, \quad i=1,\dots,\lfloor n^{1-s}\rfloor. \end{aligned}$$

By Remark 3, centering is not needed in Lemma 1 when $\alpha < q \leq 1$ and so we consider $Z_{j,n} = \bar{Y}_{j,n}$ in this case. Before splitting the cases based on different α values, we derive some facts that will be used later. Due to stationarity, (ξ_i) are identically distributed for fixed n, so that $E\left[(1/k)\sum_{i=1}^{k}\xi_i\right] = E\xi_1$. Moments of all orders of $\bar{Y}_{j,n}$ are finite and by using Karamata's theorem (Resnick 2007, Theorem 2.1), for arbitrary $r > \alpha$ it follows that for n large enough

$$\begin{aligned} E|\bar{Y}_{j,n}|^r &= \int_0^\infty P(|\bar{Y}_{j,n}|^r > x)dx = \int_0^{n^{r(\frac{1}{\alpha}+\delta)}} P(|Y_j|^r > x)dx \\ &= \int_0^{n^{r(\frac{1}{\alpha}+\delta)}} L(x^{\frac{1}{r}})x^{-\frac{\alpha}{r}}dx \le C_1 L(n^{\frac{1}{\alpha}+\delta})n^{r(\frac{1}{\alpha}+\delta)(-\frac{\alpha}{r}+1)} \le C_1 n^{\frac{r}{\alpha}-1+\delta(r-\alpha)+\eta}, \end{aligned} \tag{2.6}$$

since for any $\eta > 0$ we can take n large enough to make $L(n^{1/\alpha+\delta}) \le n^\eta$. It follows then that if $r > 1$ by Jensen's inequality

$$\begin{aligned} E|Z_{j,n}|^r &= E|\bar{Y}_{j,n} - E\bar{Y}_{j,n}|^r \le 2^{r-1}\left(E|\bar{Y}_{j,n}|^r + (E|\bar{Y}_{j,n}|)^r\right) \le 2^r E|\bar{Y}_{j,n}|^r \\ &\le C_2 n^{\frac{r}{\alpha}-1+\delta(r-\alpha)+\eta}. \end{aligned} \tag{2.7}$$

On the other hand, if $r \le 1$, the same bound holds as $\alpha < r \le 1$ so there is no centering, i.e. $E|Z_{j,n}|^r = E|\bar{Y}_{j,n}|^r$.

Next, notice that, for fixed n, $Z_{j,n}$, $j = 1,\dots,n$ is a stationary sequence. By definition $EZ_{j,n} = 0$ and also $E|Z_{j,n}|^{q+\zeta} < \infty$ for every $\zeta > 0$. Since $Z_{j,n}$ is no more than a measurable transformation of Y_j, the mixing properties of $Z_{j,n}$ are inherited from those of sequence (Y_j). This means that there exists a constant $b > 0$ such that the mixing coefficients sequence satisfies $a_Z(m) = O(e^{-bm})$ as $m \to \infty$. It follows that

$$\sum_{m=1}^{\infty}(m+1)^{2c-2}\left(a_Z(m)\right)^{\frac{\zeta}{2c+\zeta}} \le \sum_{m=1}^{\infty}(m+1)^{2c-2}K_1 e^{-bm\frac{\zeta}{2c+\zeta}} < \infty$$

for every choice of $c \in \mathbb{N}$ and $\zeta > 0$. Hence we can apply Lemma 1 for n fixed to get

$$E\xi_1 = E\left|\sum_{j=1}^{\lfloor n^s \rfloor} Z_{j,n}\right|^q \le \begin{cases} KL(q,0,\lfloor n^s \rfloor), & \text{if } 0 < q \le 1, \\ KL(q,\zeta,\lfloor n^s \rfloor), & \text{if } 1 < q \le 2, \\ K\max\left\{L(q,\zeta,\lfloor n^s \rfloor), (L(2,\zeta,\lfloor n^s \rfloor))^{\frac{q}{2}}\right\}, & \text{if } q > 2. \end{cases} \tag{2.8}$$

Notice that none of the previous arguments uses assumptions on α. Now we split the cases:

$\boxed{\bullet\alpha > 2}$ Because for $q > \alpha$, utilizing Equation (2.7), we can choose ζ small enough so that $\zeta < q\delta\alpha$ (in order to achieve $n^{-\frac{q}{q+\zeta}(1+\delta\alpha)} < n^{-1}$) to obtain

$$\begin{aligned} L(q,\zeta,\lfloor n^s \rfloor) &= \sum_{j=1}^{\lfloor n^s \rfloor}\left(E\left|Z_{j,n}\right|^{q+\zeta}\right)^{\frac{q}{q+\zeta}} \le C_3 n^s n^{\left(\frac{q+\zeta}{\alpha}-1+\delta(q+\zeta-\alpha)+\eta\right)\left(\frac{q}{q+\zeta}\right)} \\ &\le C_3 n^{s+\frac{q}{\alpha}-\frac{q}{q+\zeta}(1+\delta\alpha)+\delta q+\eta} \le C_3 n^{s+\frac{q}{\alpha}-1+\delta q+\eta}, \end{aligned} \tag{2.9}$$

and

$$(L(2,\zeta,\lfloor n^s \rfloor))^{\frac{q}{2}} = \left(\sum_{j=1}^{\lfloor n^s \rfloor} \left(E\,|Z_{j,n}|^{2+\zeta} \right)^{\frac{2}{2+\zeta}} \right)^{\frac{q}{2}} \leq n^{\frac{sq}{2}} \left(E|Y_1|^{2+\zeta} \right)^{\frac{q}{2+\zeta}} \leq C_4 n^{\frac{sq}{2}}.$$

Hence $E\xi_1 \leq C_5 n^{\max\left\{s+\frac{q}{\alpha}-1+\delta q+\eta, \frac{sq}{2}\right\}}$.

$\boxed{\bullet 1 < \alpha \leq 2}$ Bound for $L(q,\zeta,\lfloor n^s \rfloor)$ is the same as in (2.9), so if $\alpha < q \leq 2$ we have $E\xi_1 \leq KL(q,\zeta,\lfloor n^s \rfloor) \leq C_5 n^{s+q/\alpha-1+\delta q+\eta}$. If $q > 2$, using Equation (2.7) and choosing $\zeta < 2\delta\alpha$ yields

$$\begin{aligned}(L(2,\zeta,\lfloor n^s \rfloor))^{\frac{q}{2}} &= \left(\sum_{j=1}^{\lfloor n^s \rfloor} \left(E\,|Z_{j,n}|^{2+\zeta} \right)^{\frac{2}{2+\zeta}} \right)^{\frac{q}{2}} \leq n^{\frac{sq}{2}} \left(C_2 n^{\frac{2+\zeta}{\alpha}-1+\delta(2+\zeta-\alpha)+\eta} \right)^{\frac{q}{2+\zeta}} \\ &\leq C_6 n^{\frac{sq}{2}+\frac{q}{\alpha}-\frac{q}{2+\zeta}(1+\delta\alpha)+\delta q+\eta} \leq C_6 n^{\frac{sq}{2}+\frac{q}{\alpha}-\frac{q}{2}+\delta q+\eta}.\end{aligned}$$

But for $q > 2$

$$s + \frac{q}{\alpha} - 1 - \frac{sq}{2} - \frac{q}{\alpha} + \frac{q}{2} = \left(1 - \frac{q}{2}\right)(s-1) > 0,$$

so that $s + \frac{q}{\alpha} - 1 > \frac{sq}{2} + \frac{q}{\alpha} - \frac{q}{2}$ and

$$K \max\left\{ L(q,\zeta,\lfloor n^s \rfloor), (L(2,\zeta,\lfloor n^s \rfloor))^{\frac{q}{2}} \right\} \leq C_7 n^{s+\frac{q}{\alpha}-1+\delta q+\eta}.$$

We conclude that for every $q > \alpha$, $E\xi_1 \leq C_8 n^{s+\frac{q}{\alpha}-1+\delta q+\eta}$.

$\boxed{\bullet 0 < \alpha \leq 1}$ If $q > 1$ we can repeat the arguments from the previous case. If $\alpha < q \leq 1$, again by (2.7)

$$L(q,0,\lfloor n^s \rfloor) = \sum_{j=1}^{\lfloor n^s \rfloor} E|Z_{j,n}|^q \leq C_2 n^{s+\frac{q}{\alpha}-1+\delta(q-\alpha)+\eta} \leq C_2 n^{s+\frac{q}{\alpha}-1+\delta q+\eta},$$

so for every $q > \alpha$

$$E\xi_1 \leq C_9 n^{s+\frac{q}{\alpha}-1+\delta q+\eta}. \tag{2.10}$$

Next, notice that

$$P\left(\max_{i=1,\dots,n} |Y_i| > n^{\frac{1}{\alpha}+\delta} \right) \leq \sum_{i=1}^{n} P\left(|Y_i| > n^{\frac{1}{\alpha}+\delta} \right) \leq n \frac{L(n^{\frac{1}{\alpha}+\delta})}{(n^{\frac{1}{\alpha}+\delta})^{\alpha}} \leq C_{10} \frac{L(n^{\frac{1}{\alpha}+\delta})}{n^{\alpha\delta}}.$$

If $\alpha > 1$, since $EY_i = 0$ we have from Karamata's theorem

$$
\begin{aligned}
|E\bar{Y}_{j,n}| &= \left|E\left(Y_j - \bar{Y}_{j,n}\right)\right| \leq E\left|Y_j - \bar{Y}_{j,n}\right| = E\left|Y_j \mathbf{1}\left(|Y_j| > n^{\frac{1}{\alpha}+\delta}\right)\right| \\
&= \int_{n^{\frac{1}{\alpha}+\delta}}^{\infty} P\left(|Y_j| \mathbf{1}\left(|Y_j| > n^{\frac{1}{\alpha}+\delta}\right) > x\right) dx + \int_0^{n^{\frac{1}{\alpha}+\delta}} P\left(|Y_j| \mathbf{1}\left(|Y_j| > n^{\frac{1}{\alpha}+\delta}\right) > x\right) dx \\
&= \int_{n^{\frac{1}{\alpha}+\delta}}^{\infty} P\left(|Y_j| > x\right) dx + n^{\frac{1}{\alpha}+\delta} P\left(|Y_j| > n^{\frac{1}{\alpha}+\delta}\right) \\
&= \int_{n^{\frac{1}{\alpha}+\delta}}^{\infty} L(x) x^{-\alpha} dx + n^{\frac{1}{\alpha}+\delta} L(n^{\frac{1}{\alpha}+\delta}) n^{-1-\alpha\delta} \\
&\leq C_{11} L(n^{\frac{1}{\alpha}+\delta}) n^{(\frac{1}{\alpha}+\delta)(-\alpha+1)} + n^{\frac{1}{\alpha}+\delta} L(n^{\frac{1}{\alpha}+\delta}) n^{-1-\alpha\delta} \leq C_{12} n^{\frac{1}{\alpha}-1+\delta(1-\alpha)+\eta}
\end{aligned}
$$

and thus

$$
\begin{aligned}
E\left|\sum_{j=1}^{\lfloor n^s \rfloor} \bar{Y}_{j,n}\right|^q &= E\left|\sum_{j=1}^{\lfloor n^s \rfloor} \left(Z_{j,n} + E\bar{Y}_{j,n}\right)\right|^q \leq 2^{q-1} E\left|\sum_{j=1}^{\lfloor n^s \rfloor} Z_{j,n}\right|^q + 2^{q-1}\left|\sum_{j=1}^{\lfloor n^s \rfloor} E\bar{Y}_{j,n}\right|^q \\
&= 2^{q-1} E\xi_1 + 2^{q-1} n^{sq} \left|E\bar{Y}_{1,n}\right|^q \leq 2^{q-1} E\xi_1 + 2^{q-1} C_{13} n^{sq+\frac{q}{\alpha}-q+q\delta(1-\alpha)+q\eta}.
\end{aligned}
$$

If $\alpha \leq 1$ and $q > 1$ we can use (2.6) to get

$$
\begin{aligned}
E\left|\sum_{j=1}^{\lfloor n^s \rfloor} \bar{Y}_{j,n}\right|^q &\leq 2^{q-1} E\left|\sum_{j=1}^{\lfloor n^s \rfloor} Z_{j,n}\right|^q + 2^{q-1}\left|\sum_{j=1}^{\lfloor n^s \rfloor} E\bar{Y}_{j,n}\right|^q \\
&= 2^{q-1} E\xi_1 + 2^{q-1} n^{sq} \left(E|\bar{Y}_{1,n}|\right)^q \\
&\leq 2^{q-1} E\xi_1 + 2^{q-1} C_1 n^{sq+\frac{q}{\alpha}-q+\delta q(1-\alpha)+q\eta}.
\end{aligned}
\tag{2.11}
$$

By partitioning on the event $\{Y_i = \bar{Y}_i, i = 1, \dots, n\} = \{\max_{i=1,\dots,n} |Y_i| \leq n^{\frac{1}{\alpha}+\delta}\}$ and its complement, using Markov's inequality and preceding results we conclude for the case $\alpha > 2$:

$$
\begin{aligned}
& P\left(\frac{\ln S_q(n, n^s)}{\ln n} > \max\left\{s + \frac{q}{\alpha} - 1, \frac{sq}{2}\right\} + \delta q + \epsilon\right) \\
&= P\left(S_q(n, n^s) > n^{\max\left\{s+\frac{q}{\alpha}-1, \frac{sq}{2}\right\}+\delta q+\epsilon}\right) \\
&\leq P\left(\frac{1}{\lfloor n^{1-s} \rfloor} \sum_{i=1}^{\lfloor n^{1-s} \rfloor} \left|\sum_{j=1}^{\lfloor n^s \rfloor} \bar{Y}_{\lfloor n^s \rfloor (i-1)+j,n}\right|^q > n^{\max\left\{s+\frac{q}{\alpha}-1, \frac{sq}{2}\right\}+\delta q+\epsilon}\right) + P\left(\max_{i=1,\dots,n} |Y_i| > n^{\frac{1}{\alpha}+\delta}\right) \\
&\leq \frac{E\left|\sum_{j=1}^{\lfloor n^s \rfloor} \bar{Y}_{j,n}\right|^q}{n^{\max\left\{s+\frac{q}{\alpha}-1, \frac{sq}{2}\right\}+\delta q+\epsilon}} + C_{10} \frac{L(n^{\frac{1}{\alpha}+\delta})}{n^{\alpha\delta}}
\end{aligned}
$$

$$\leq \frac{2^{q-1}E\xi_1 + 2^{q-1}C_{13}n^{sq+\frac{q}{\alpha}-q+q\delta(1-\alpha)+q\eta}}{n^{\max\left\{s+\frac{q}{\alpha}-1,\frac{sq}{2}\right\}+\delta q+\epsilon}} + C_{10}\frac{L(n^{\frac{1}{\alpha}+\delta})}{n^{\alpha\delta}}$$
$$\leq \frac{2^{q-1}C_5 n^{\max\left\{s+\frac{q}{\alpha}-1+\delta q+\eta,\frac{sq}{2}\right\}} + 2^{q-1}C_{13}n^{sq+\frac{q}{\alpha}-q+q\delta(1-\alpha)+q\eta}}{n^{\max\left\{s+\frac{q}{\alpha}-1,\frac{sq}{2}\right\}+\delta q+\epsilon}} + C_{10}\frac{L(n^{\frac{1}{\alpha}+\delta})}{n^{\alpha\delta}} \to 0,$$

as $n \to \infty$, since $sq + q/\alpha - q + q\delta(1-\alpha) + q\eta < s + q/\alpha - 1 + \delta q + \epsilon$ if we take $\eta < \epsilon/q$. As ϵ and δ are arbitrary, it follows that

$$\operatorname*{plim}_{n\to\infty} \frac{\ln S_q(n, n^s)}{\ln n} \leq \max\left\{s + \frac{q}{\alpha} - 1, \frac{sq}{2}\right\}.$$

In case $1 < \alpha \leq 2$ we can repeat the previous with $n^{s+q/\alpha-1+\delta q}$ instead of $n^{\max\{s+q/\alpha-1+\delta q, sq/2\}}$ and get

$$\operatorname*{plim}_{n\to\infty} \frac{\ln S_q(n, n^s)}{\ln n} \leq s + \frac{q}{\alpha} - 1.$$

If $\alpha \leq 1$ and $q > 1$ we use (2.11) and similarly get

$$P\left(\frac{\ln S_q(n, n^s)}{\ln n} > s + \frac{q}{\alpha} - 1 + \delta q + \epsilon\right)$$
$$\leq \frac{2^{q-1}C_9 n^{s+\frac{q}{\alpha}-1+\delta q+\eta} + 2^{q-1}C_1 n^{sq+\frac{q}{\alpha}-q+\delta q(1-\alpha)+q\eta}}{n^{s+\frac{q}{\alpha}-1+\delta q+\epsilon}} + C_{10}\frac{L(n^{\frac{1}{\alpha}+\delta})}{n^{\alpha\delta}} \to 0,$$

as $n \to \infty$. Finally, if $\alpha < q \leq 1$ there is no centering and (2.10) gives

$$P\left(\frac{\ln S_q(n, n^s)}{\ln n} > s + \frac{q}{\alpha} - 1 + \delta q + \epsilon\right) \leq \frac{E\xi_1}{n^{s+\frac{q}{\alpha}-1+\delta q+\epsilon}} + C_{10}\frac{L(n^{\frac{1}{\alpha}+\delta})}{n^{\alpha\delta}} \to 0,$$

as $n \to \infty$.

We next show the lower bound in two parts.
We first consider the case $\alpha > 2$ and assume that $s + q/\alpha - 1 \leq sq/2$. Let

$$\sigma^2 = \lim_{n\to\infty} \frac{E\left(\sum_{j=1}^{n} Y_j\right)^2}{n},$$

$$\rho_n = P\left(\left|\sum_{j=1}^{\lfloor n^s \rfloor} Y_{n^s(i-1)+j}\right| > n^{\frac{s}{2}}\sigma\right).$$

Since the sequence (Y_j) is stationary and strong mixing with an exponential decaying rate and since $E|Y_j|^{2+\zeta} < \infty$ for $\zeta > 0$ sufficiently small, the central limit theorem holds (see (Hall & Heyde 1980, Corollary 5.1.)) and σ^2 exists. Since $P(|\mathcal{N}(0,1)| > 1) > 1/4$, it follows that for n large enough $\rho_n > 1/4$. Recall that if $\mathcal{MB}(n,p)$ is the sum of n stationary mixing indicator variables with expectation p, then ergodic theorem implies $\mathcal{MB}(n,p)/n \to p$, a.s.

Now we have

$$\begin{aligned}
P\left(\frac{\ln S_q(n,n^s)}{\ln n} < \frac{sq}{2} - \epsilon\right) &= P\left(S_q(n,n^s) < n^{\frac{sq}{2}-\epsilon}\right) \\
&\leq P\left(\sum_{i=1}^{\lfloor n^{1-s}\rfloor} \left|\sum_{j=1}^{\lfloor n^s\rfloor} Y_{n^s(i-1)+j}\right|^q < n^{\frac{sq}{2}-\epsilon+1-s}\right) \\
&\leq P\left(\sum_{i=1}^{\lfloor n^{1-s}\rfloor} \mathbf{1}\left(\left|\sum_{j=1}^{\lfloor n^s\rfloor} Y_{n^s(i-1)+j}\right| > n^{\frac{s}{2}}\sigma\right) < \frac{n^{\frac{sq}{2}-\epsilon+1-s}}{n^{\frac{sq}{2}}\sigma^q}\right) \\
&= P\left(\sum_{i=1}^{\lfloor n^{1-s}\rfloor} \mathbf{1}\left(\left|\sum_{j=1}^{\lfloor n^s\rfloor} Y_{n^s(i-1)+j}\right| > n^{\frac{s}{2}}\sigma\right) < \frac{n^{1-s-\epsilon}}{\sigma^q}\right) \\
&\leq P\left(\mathcal{MB}(\lfloor n^{1-s}\rfloor, 1/4) < \frac{n^{1-s-\epsilon}}{\sigma^q}\right) \to 0,
\end{aligned}$$

hence

$$\operatorname*{plim}_{n\to\infty} \frac{\ln S_q(n,n^s)}{\ln n} \geq \frac{sq}{2}.$$

For the second part, assume that $s + q/\alpha - 1 > sq/2$. Notice that in this case it must hold $1/\alpha - s/2 > 0$. We can assume that $\epsilon < 1/\alpha - s/2$. Indeed, otherwise we can choose $0 < \tilde{\epsilon} < 1/\alpha - s/2$ and continue the proof with it in place of ϵ by observing that

$$P\left(\frac{\ln S_q(n,n^s)}{\ln n} < s + \frac{q}{\alpha} - 1 - \epsilon\right) \leq P\left(\frac{\ln S_q(n,n^s)}{\ln n} < s + \frac{q}{\alpha} - 1 - \tilde{\epsilon}\right).$$

The main fact behind the following part of the proof is that $\sum |Y_i|^q \approx \max |Y_i|^q$ and that s is small, which makes the blocks to grow slowly. As discussed in Subsection 2.1.1, the assumption that the extremal index is positive ensures that $\max_{j=1,\dots,n} |Y_j|/(n^{1/\alpha}L_1(n))$ with some L_1 slowly varying converges in distribution to some positive random variable, so that

$$P\left(\max_{j=1,\dots,n} |Y_j| < 2n^{\frac{1}{\alpha}-\epsilon}\right) \to 0.$$

Let $l \in \mathbb{N}$ be such that $|Y_l| = \max_{j=1,\dots,n} |Y_j|$. Then, for some $k \in \{1, 2, \dots, \lfloor n^{1-s}\rfloor\}$ we have $l \in \mathcal{J} := \{\lfloor n^s\rfloor(k-1)+1, \dots, \lfloor n^s\rfloor k\}$. Assumption $\alpha > 2$ ensures that $E|Y_1|^{2+\zeta} < \infty$ for some $\zeta > 0$. Applying Markov's inequality and then Lemma 1 yields

$$\begin{aligned}
P\left(\left|\sum_{j\in\mathcal{J},j\neq l} Y_j\right| > n^{\frac{1}{\alpha}-\epsilon}\right) &\leq \frac{E\left(\sum_{j\in\mathcal{J},j\neq l} Y_j\right)^2}{n^{\frac{2}{\alpha}-2\epsilon}} \leq \frac{K_1 \sum_{j\in\mathcal{J},j\neq l}\left(E|Y_j|^{2+\zeta}\right)^{\frac{2}{2+\zeta}}}{n^{\frac{2}{\alpha}-2\epsilon}} \\
&\leq \frac{K_2 n^s}{n^{\frac{2}{\alpha}-2\epsilon}} = K_2 n^{s-\frac{2}{\alpha}+2\epsilon} \to 0, \quad \text{as } n \to \infty,
\end{aligned}$$

since $s - 2/\alpha + 2\epsilon < 0$ by the assumption in the proof.

Combining this it follows that

$$\begin{aligned} P\left(\frac{\ln S_q(n,n^s)}{\ln n} < s+\frac{q}{\alpha}-1-q\epsilon\right) &= P\left(S_q(n,n^s) < n^{s+\frac{q}{\alpha}-1-q\epsilon}\right) \\ &\leq P\left(\sum_{i=1}^{\lfloor n^{1-s}\rfloor}\left|\sum_{j=1}^{\lfloor n^s\rfloor} Y_{n^s(i-1)+j}\right|^q < n^{\frac{q}{\alpha}-q\epsilon}\right) \\ &\leq P\left(\left|\sum_{j\in\mathcal{J}} Y_j\right|^q < n^{\frac{q}{\alpha}-q\epsilon}\right) = P\left(\left|\sum_{j\in\mathcal{J}} Y_j\right| < n^{\frac{1}{\alpha}-\epsilon}\right) \\ &\leq P\left(|Y_l| < 2n^{\frac{1}{\alpha}-\epsilon}\right) + P\left(\left|\sum_{j\in\mathcal{J},j\neq l} Y_j\right| > n^{\frac{1}{\alpha}-\epsilon}\right) \to 0, \end{aligned}$$

as $n\to\infty$. Hence,

$$\operatorname*{plim}_{n\to\infty}\frac{\ln S_q(n,n^s)}{\ln n} \geq \max\left\{s+\frac{q}{\alpha}-1, \frac{sq}{2}\right\}.$$

For the case $0<\alpha\leq 2$ we just need a different estimate for the sum containing maximum. Choose γ such that $0<\gamma<\alpha$. Again we use Markov's inequality

$$P\left(\left|\sum_{j\in\mathcal{J},j\neq l} Y_j\right| > n^{\frac{1}{\alpha}-\epsilon}\right) \leq \frac{E\left|\sum_{j\in\mathcal{J},j\neq l} Y_j\right|^{\alpha-\gamma}}{n^{1-\alpha\epsilon-\frac{\gamma}{\alpha}+\epsilon\gamma}}.$$

From Lemma 1 one can easily bound this expectation by $K_3 n^s$ for some constant K_3. Choosing ϵ and γ small enough to make $s-1+\alpha\epsilon+\gamma/\alpha-\epsilon\gamma<0$, we get

$$P\left(\left|\sum_{j\in\mathcal{J},j\neq l} Y_j\right| > n^{\frac{1}{\alpha}-\epsilon}\right) \leq \frac{K_3 n^s}{n^{1-\alpha\epsilon-\frac{\gamma}{\alpha}+\epsilon\gamma}} \to 0, \quad \text{as } n\to\infty,$$

and this completes the (a) part of the proof.

(b) Now let $q<\alpha$. We first show the upper bound on the limit, i.e. we analyze

$$\begin{aligned} P\left(\frac{\ln S_q(n,n^s)}{\ln n} > \frac{sq}{\beta(\alpha)}+\epsilon\right) &= P\left(S_q(n,n^s) > n^{\frac{sq}{\beta(\alpha)}+\epsilon}\right) \\ &\leq P\left(\frac{1}{\lfloor n^{1-s}\rfloor}\sum_{i=1}^{\lfloor n^{1-s}\rfloor}\left|\sum_{j=1}^{\lfloor n^s\rfloor} Y_{n^s(i-1)+j,n}\right|^q > n^{\frac{sq}{\beta(\alpha)}+\epsilon}\right) \leq \frac{E\left|\sum_{j=1}^{\lfloor n^s\rfloor} Y_j\right|^q}{n^{\frac{sq}{\beta(\alpha)}+\epsilon}}, \end{aligned}$$

where $\beta(\alpha)=\alpha$ or 2 according to $\alpha\leq 2$ or $\alpha>2$. To show that this tends to zero, we first consider the case $\alpha>2$. If $q>2$, using Lemma 1 with ζ small enough it follows that

$$E\left|\sum_{j=1}^{\lfloor n^s\rfloor} Y_j\right|^q \leq C_1\max\{n^s, n^{\frac{sq}{2}}\}.$$

For the case $q \leq 2$ we combine Jensen's inequality with Lemma 1:

$$E\left|\sum_{j=1}^{\lfloor n^s \rfloor} Y_j\right|^q \leq \left(E\left|\sum_{j=1}^{\lfloor n^s \rfloor} Y_j\right|^2\right)^{\frac{q}{2}} \leq C_2 n^{\frac{sq}{2}}.$$

In the case $\alpha \leq 2$ we choose γ small enough to make $q < \alpha - \gamma < \alpha$ and get

$$E\left|\sum_{j=1}^{\lfloor n^s \rfloor} Y_j\right|^q \leq \left(E\left|\sum_{j=1}^{\lfloor n^s \rfloor} Y_j\right|^{\alpha-\gamma}\right)^{\frac{q}{\alpha-\gamma}} \leq C_3 n^{\frac{sq}{\alpha-\gamma}}.$$

We next prove the lower bound. For the case $\alpha > 2$ the proof is the same as the proof of (a). Assume $\alpha \leq 2$. The arguments go along the same line, but we avoid using limit theorems for partial sums of stationary sequences. Instead we use before mentioned asymptotic behavior of the partial maximum, that is, we use the fact that $\max_{j=1,\dots,\lfloor n^s \rfloor} |Y_j|/(n^{s/\alpha} L_1(n^s))$ converges in distribution to some positive random variable, for some slowly varying L_1. This means we can choose some constant $m > 0$ such that for large enough n

$$P\left(\frac{\max_{j=1,\dots,\lfloor n^s \rfloor} |Y_j|}{n^{\frac{s}{\alpha}}} > 2m\right) > \frac{1}{4}.$$

Let $|Y_l| = \max_{j=1,\dots,\lfloor n^s \rfloor} |Y_j|$. Then it follows that

$$P\left(\left|\sum_{j=1}^{\lfloor n^s \rfloor} Y_j\right| > mn^{\frac{s}{\alpha}}\right) \geq P\left(|Y_l| > 2mn^{\frac{s}{\alpha}}\right) + P\left(\left|\sum_{j=1, j\neq l}^{\lfloor n^s \rfloor} Y_j\right| < mn^{\frac{s}{\alpha}}\right) > \frac{1}{4}.$$

Now we conclude as before, denoting by $\mathcal{MB}(n,p)$ the sum of n stationary mixing indicator variables with mean p and noting that the ergodic theorem implies $\mathcal{MB}(n,p)/n \to p > 0$, a.s.:

$$\begin{aligned}
P\left(\frac{\ln S_q(n,n^s)}{\ln n} < \frac{sq}{\alpha} - \epsilon\right) &=\leq P\left(\sum_{i=1}^{\lfloor n^{1-s} \rfloor} \left|\sum_{j=1}^{\lfloor n^s \rfloor} Y_{n^s(i-1)+j}\right|^q < n^{\frac{sq}{\alpha}-\epsilon+1-s}\right) \\
&\leq P\left(\sum_{i=1}^{\lfloor n^{1-s} \rfloor} \mathbf{1}\left(\left|\sum_{j=1}^{\lfloor n^s \rfloor} Y_{n^s(i-1)+j}\right| > n^{\frac{s}{\alpha}} m\right) < \frac{n^{\frac{sq}{\alpha}-\epsilon+1-s}}{n^{\frac{sq}{\alpha}} m^q}\right) \\
&\leq P\left(\sum_{i=1}^{\lfloor n^{1-s} \rfloor} \mathbf{1}\left(\left|\sum_{j=1}^{\lfloor n^s \rfloor} Y_{n^s(i-1)+j}\right| > n^{\frac{s}{\alpha}} m\right) < \frac{n^{1-s-\epsilon}}{m^q}\right) \\
&\leq P\left(\mathcal{MB}(\lfloor n^{1-s} \rfloor, 1/4) < \frac{n^{1-s-\epsilon}}{m^q}\right) \to 0,
\end{aligned}$$

as $n \to \infty$. This proves the lower bound.

(c) It remains to consider the case $q = \alpha$. For every $\delta > 0$, we have for n large enough

$$\frac{\ln S_{q-\delta}(n, n^s)}{\ln n} \leq \frac{\ln S_q(n, n^s)}{\ln n} \leq \frac{\ln S_{q+\delta}(n, n^s)}{\ln n}.$$

Thus, the limit must be monotone in q and the claim follows from the previous cases. □

Proof of Theorem 2. Fix $q > 0$. First we show that

$$\operatorname*{plim}_{n\to\infty} \hat{\tau}_{N,n}(q) = \frac{\sum_{i=1}^N s_i R_\alpha(q, s_i) - \frac{1}{N}\sum_{i=1}^N s_i \sum_{j=1}^N R_\alpha(q, s_j)}{\sum_{i=1}^N (s_i)^2 - \frac{1}{N}\left(\sum_{i=1}^N s_i\right)^2}. \tag{2.12}$$

Let $\varepsilon > 0$ and $\delta > 0$ and denote

$$C = \sum_{i=1}^N (s_i)^2 - \frac{1}{N}\left(\sum_{i=1}^N s_i\right)^2 > 0.$$

By Theorem 1, for each $i = 1, \dots, N$ there exists $n_i^{(1)}$ such that

$$P\left(\left|\frac{\ln S_q(n, n^{s_i})}{\ln n} - R_\alpha(q, s_i)\right| > \frac{\varepsilon C}{2 s_i N}\right) < \frac{\delta}{2N}, \quad n \geq n_i^{(1)}.$$

It follows then that for $n \geq n_{max}^{(1)} := \max\{n_1^{(1)}, \dots, n_N^{(1)}\}$

$$\begin{aligned} P&\left(\left|\sum_{i=1}^N s_i \frac{\ln S_q(n, n^{s_i})}{\ln n} - \sum_{i=1}^N s_i R_\alpha(q, s_i)\right| > \frac{\varepsilon C}{2}\right) \\ &\leq P\left(\sum_{i=1}^N s_i \left|\frac{\ln S_q(n, n^{s_i})}{\ln n} - R_\alpha(q, s_i)\right| > \frac{\varepsilon C}{2}\right) \\ &\leq \sum_{i=1}^N P\left(\left|\frac{\ln S_q(n, n^{s_i})}{\ln n} - R_\alpha(q, s_i)\right| > \frac{\varepsilon C}{2 s_i N}\right) < \frac{\delta}{2}. \end{aligned}$$

Similarly, for each $i = 1, \dots, N$ there exist $n_i^{(2)}$ such that

$$P\left(\left|\frac{\ln S_q(n, n^{s_i})}{\ln n} - R_\alpha(q, s_i)\right| > \frac{\varepsilon C}{2\left(\sum_{i=1}^N s_i\right)}\right) < \frac{\delta}{2N}, \quad n \geq n_i^{(2)},$$

and for $n \geq n_{max}^{(2)} := \max\{n_1^{(2)}, \dots, n_N^{(2)}\}$

$$P\left(\left|\frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} \frac{\ln S_q(n, n^{s_j})}{\ln n} - \frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} R_\alpha(q, s_j)\right| > \frac{\varepsilon C}{2}\right)$$
$$\leq P\left(\sum_{j=1}^{N}\left|\frac{\ln S_q(n, n^{s_j})}{\ln n} - R_\alpha(q, s_j)\right| > \frac{N\varepsilon C}{2\left(\sum_{i=1}^{N} s_i\right)}\right)$$
$$\leq \sum_{j=1}^{N} P\left(\left|\frac{\ln S_q(n, n^{s_j})}{\ln n} - R_\alpha(q, s_j)\right| > \frac{\varepsilon C}{2\left(\sum_{i=1}^{N} s_i\right)}\right) < \frac{\delta}{2}.$$

Finally then, for $n \geq \max\{n_{max}^{(1)}, n_{max}^{(2)}\}$ it follows that

$$P\left(\left|\hat{\tau}_{N,n}(q) - \frac{\sum_{i=1}^{N} s_i R_\alpha(q, s_i) - \frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} R_\alpha(q, s_j)}{\sum_{i=1}^{N}(s_i)^2 - \frac{1}{N}\left(\sum_{i=1}^{N} s_i\right)^2}\right| > \varepsilon\right)$$
$$\leq P\left(\left|\sum_{i=1}^{N} s_i \frac{\ln S_q(n, n^{s_i})}{\ln n} - \sum_{i=1}^{N} s_i R_\alpha(q, s_i)\right|\right.$$
$$\left. + \left|\frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} \frac{\ln S_q(n, n^{s_j})}{\ln n} - \frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} R_\alpha(q, s_j)\right| > \varepsilon C\right)$$
$$\leq P\left(\left|\sum_{i=1}^{N} s_i \frac{\ln S_q(n, n^{s_i})}{\ln n} - \sum_{i=1}^{N} s_i R_\alpha(q, s_i)\right| > \frac{\varepsilon C}{2}\right)$$
$$+ P\left(\left|\frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} \frac{\ln S_q(n, n^{s_j})}{\ln n} - \frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} R_\alpha(q, s_j)\right| > \frac{\varepsilon C}{2}\right) < \delta,$$

and this proves (2.12). To show (i), notice that in this case $R_\alpha(q, s)$ from (2.2) can be written in the form $R_\alpha(q, s) = \tau_\alpha^\infty(q)s + b(q)$. Now the right hand side in (2.12) is

$$\frac{\tau_\alpha^\infty(q)\sum_{i=1}^{N} s_i^2 + b(q)\sum_{i=1}^{N} s_i - \frac{1}{N}\sum_{i=1}^{N} s_i\left(\tau_\alpha^\infty(q)\sum_{j=1}^{N} s_j + Nb(q)\right)}{\sum_{i=1}^{N}(s_i)^2 - \frac{1}{N}\left(\sum_{i=1}^{N} s_i\right)^2} = \tau_\alpha^\infty(q).$$

For (ii), dividing denominator and numerator of the fraction in limit (2.12) by N yields

$$\operatorname*{plim}_{n\to\infty} \hat{\tau}_{N,n}(q) = \frac{\frac{1}{N}\sum_{i=1}^{N} \frac{i}{N} R_\alpha\left(q, \frac{i}{N}\right) - \left(\frac{1}{N}\sum_{i=1}^{N} \frac{i}{N}\right)\left(\frac{1}{N}\sum_{j=1}^{N} R_\alpha\left(q, \frac{i}{N}\right)\right)}{\frac{1}{N}\sum_{i=1}^{N}(s_i)^2 - \left(\frac{1}{N}\sum_{i=1}^{N} s_i\right)^2}.$$

One can see all the sums involved as Riemann sums based on the equidistant partition. Functions involved, $s \mapsto sR_\alpha(q, s)$, $s \mapsto R_\alpha(q, s)$, $s \mapsto s$ and $s \mapsto s^2$, are all bounded continuous on $[0, 1]$, so all sums converge to integrals when partition is refined, i.e. when

$N \to \infty$. Thus

$$\lim_{N\to\infty} \operatorname*{plim}_{n\to\infty} \hat{\tau}_{N,n}(q) = \frac{\int_0^1 sR_\alpha(q,s)ds - \int_0^1 sds \int_0^1 R_\alpha(q,s)ds}{\int_0^1 s^2 ds - \left(\int_0^1 sds\right)^2}.$$

By solving the integrals using the expression for $R_\alpha(q,s)$, one gets τ_α^∞ as in (2.4). Indeed, let $\overline{s} = (1-q/\alpha)/(1-q/2)$. For the numerator we have

$$\begin{aligned}
&\int_0^1 sR_\alpha(q,s)ds - \int_0^1 sds \int_0^1 R_\alpha(q,s)ds \\
&= \int_0^{\overline{s}} s^2 ds + \left(\frac{q}{\alpha}-1\right)\int_0^{\overline{s}} sds + \frac{q}{2}\int_{\overline{s}}^1 s^2 ds - \frac{1}{2}\int_0^{\overline{s}} sds - \frac{1}{2}\left(\frac{q}{\alpha}-1\right)\int_0^{\overline{s}} ds - \frac{1}{2}\frac{q}{2}\int_{\overline{s}}^1 sds \\
&= \frac{\overline{s}^3}{3} + \left(\frac{q}{\alpha}-1\right)\frac{\overline{s}^2}{2} + \frac{q}{2}\left(\frac{1}{3}-\frac{\overline{s}^3}{3}\right) - \frac{1}{2}\frac{\overline{s}^2}{2} - \frac{1}{2}\left(\frac{q}{\alpha}-1\right)\overline{s} - \frac{q}{4}\left(\frac{1}{2}-\frac{\overline{s}^2}{2}\right) \\
&= \frac{\overline{s}^3}{3}\left(1-\frac{q}{2}\right) - \frac{\overline{s}^2}{2}\left(1-\frac{q}{\alpha}\right) - \frac{\overline{s}^2}{4}\left(1-\frac{q}{2}\right) + \frac{\overline{s}}{2}\left(1-\frac{q}{\alpha}\right) + \frac{q}{6} - \frac{q}{8} \\
&= \frac{q}{24} + \frac{1}{3}\frac{\left(1-\frac{q}{\alpha}\right)^3}{\left(1-\frac{q}{2}\right)^2} - \frac{1}{2}\frac{\left(1-\frac{q}{\alpha}\right)^3}{\left(1-\frac{q}{2}\right)^2} - \frac{1}{4}\frac{\left(1-\frac{q}{\alpha}\right)^2}{\left(1-\frac{q}{2}\right)} + \frac{1}{2}\frac{\left(1-\frac{q}{\alpha}\right)^2}{\left(1-\frac{q}{2}\right)} \\
&= \frac{q}{24} + \frac{\left(1-\frac{q}{\alpha}\right)^2}{\left(1-\frac{q}{2}\right)^2}\left(-\frac{1}{6}\left(1-\frac{q}{\alpha}\right) + \frac{1}{4}\left(1-\frac{q}{2}\right)\right) \\
&= \frac{q}{24} + \frac{4(\alpha-q)^2}{\alpha^2(2-q)^2}\left(\frac{1}{12} + \frac{1}{6}\frac{q}{\alpha} - \frac{1}{8}q\right) = \frac{q}{24} + \frac{1}{12}\frac{2(\alpha-q)^2}{\alpha^3(2-q)^2}(2\alpha + 4q - 3\alpha q).
\end{aligned}$$

Since

$$\int_0^1 s^2 ds - \left(\int_0^1 sds\right)^2 = \frac{1}{12},$$

we arrive at the form given in (2.4). □

2.2 Applications in the tail index estimation

This section deals with the applications of the partition function and the empirical scaling function in the analysis of the tail index of heavy-tailed data. Heavy-tailed distributions are of considerable importance in modeling a wide range of phenomena in finance, geology, hydrology, physics, queuing theory and telecommunications. Pioneering work was done in Mandelbrot (1963), where stable distributions with index less than 2 have been advocated for describing fluctuations of cotton prices. In the field of finance, distributions of logarithmic asset returns can often be fitted extremely well by Student's t-distribution (see Heyde & Leonenko (2005) and references therein).

Two important practical problems arise in this context. First, if we have data sampled from some stationary sequence $(Y_i, i \in \mathbb{N})$, the question is whether this data comes from some heavy-tailed distribution or not. Usually, methods for this purpose are graphical. The second problem is the estimation of the unknown tail index for samples coming from some heavy-tailed distribution.

Before we apply our results on these problems, we provide a brief overview of the existing methods.

2.2.1 Overview of the existing methods

The problem of estimation of the tail index is widely known and there have been numerous approaches to it. Probably the best known estimator of the tail index is the Hill estimator (Hill (1975)). For what follows, $Y_{(1)} \geq Y_{(2)} \geq \cdots \geq Y_{(n)}$ will denote the order statistics of the sample $Y_1, Y_2, \ldots, Y_n$. For $1 \leq k \leq n$, the Hill estimator based on k upper order statistics is

$$\hat{\alpha}_k^{Hill} = \left(\frac{1}{k} \sum_{i=1}^{k} \log \frac{Y_{(i)}}{Y_{(k+1)}} \right)^{-1} . \tag{2.13}$$

The Hill estimator possesses many desirable asymptotic properties, for example weak consistency provided the sample is i.i.d. and $k = k(n)$ is a sequence satisfying $\lim_{n\to\infty} k(n) = \infty$ and $\lim_{n\to\infty} (k(n)/n) = 0$. Under additional assumptions on the sequence $k(n)$ and second order regular variation properties of the underlying distribution, even asymptotic normality holds. Properties of the Hill estimator have been extensively studied in settings different than i.i.d. (for example, see Hsing (1991) for mixing sequences).

Another estimator of the tail index is the so-called moment estimator proposed by Dekkers et al. (1989). Define for $r = 1, 2$

$$H_k^{(r)} = \frac{1}{k} \sum_{i=1}^{k} \left(\log \frac{Y_{(i)}}{Y_{(k+1)}} \right)^r .$$

A moment estimator based on k order statistics is given by

$$\hat{\alpha}_k^M = \left(1 + H_k^{(1)} + \frac{1}{2} \left(\frac{(H_k^{(1)})^2}{H_k^{(2)}} - 1 \right)^{-1} \right)^{-1} . \tag{2.14}$$

Originally, moment estimator is defined as $1/\hat{\alpha}_k^M$ and is an estimator of the extreme value index ξ, which coincides with the reciprocal of the tail index when $\xi > 0$. For more details on both estimators as well as the definitions of others, like e.g. the Pickands estimator, see Embrechts et al. (1997) and De Haan & Ferreira (2007).

Both equations, (2.13) and (2.14), actually yield a sequence of estimated values for different values of k. The choice of optimal k is considered to be the main disadvantage of these estimators as their performance can vary significantly with k. If additional assumptions are imposed on the second order regular variation of the distribution, a sequence $k(n)$ giving an optimal asymptotic mean square error can be obtained (see De Haan & Ferreira (2007)). Although there is no much practical significance of such results, estimators of the sequence $k(n)$ can be derived. This leads to the adaptive selection methods for k, an example of which can be found in Beirlant et al. (2006) (see also De Haan & Ferreira (2007) and references therein). A more common approach is to plot estimated values against k. A heuristic rule is to look for the place where the graph stabilizes and report this as the estimated value. For the Hill estimator this is usually called the Hill plot. We use this approach later in the examples.

Tail index estimators are usually based on upper order statistics and their asymptotic properties. Alternatively, in Meerschaert & Scheffler (1998), an estimator based on the asymptotics of the sample variance has been proposed. More precisely, the authors define

$$\hat{\alpha} = \frac{2 \ln n}{\ln n + \ln \hat{\sigma}^2},$$

where $\hat{\sigma}^2$ is the usual sample variance. The estimator is consistent for i.i.d. samples in the domain of attraction of a stable law with index $\alpha < 2$. This approach is, however, appropriate mostly for the case $\alpha < 2$, otherwise one would need to transform the data, e.g. to square it when $2 < \alpha < 4$. In a certain way, the underlying idea of our method is also based on the asymptotic properties of the sum. Our approach is, however, more general and independent of the results in Meerschaert & Scheffler (1998). As we will see, the block structure of the partition function enables extracting more information about the tail index. Moreover, we go beyond the i.i.d. case and consider weakly dependent samples.

Before applying any of the tail index estimators, one should make sure that the heavy-tailed model is appropriate. Usually, various graphical techniques are used for this purpose. It is important to stress that Hill plots cannot be used as a graphical tool for establishing heavy tail property of the data as they can be misleading in cases when the tails are light. On the other hand, extreme value index estimators, like moment estimator, can be used for this purpose. Plotting the values $1/\hat{\alpha}_k^M$ for varying k can indicate that the tails are light if the values are around zero (see Resnick (2007)). There are other exploratory tools for inspecting whether the tails are heavy or not. One of the most frequently used tool is a variation of the QQ plot. The basic idea comes from the fact that if $P(Y > x) \sim x^{-\alpha}$, then $P(\ln Y > x) = P(Y > e^x) \sim e^{-\alpha x}$, i.e. the log-transformed Pareto random variable has an exponential distribution. By choosing $1 \leq k \leq n$ one can plot the points

$$\left(-\ln \left(\frac{i}{k+1} \right), \ln Y_{(i)} \right), \quad i = 1, \dots, k.$$

If the data is heavy-tailed with index α, the plot should be roughly linear with slope $1/\alpha$. This is no more than the standard QQ plot of log-transformed data on exponential quantiles. This graphical method can be used to define estimator (Resnick (2007)), however, we will use it only as an exploratory tool. For $k = n$ this plot is sometimes called Zipf's plot. We will refer to it simply as the QQ plot.

2.2.2 Graphical method

As a first step, we propose a graphical method useful for exploratory analysis of the tails of the underlying distribution. Since the scaling function shape is strongly influenced by the tail index value, this motivates the use of a plot of the empirical scaling function to detect the tail index of a distribution. In particular, the asymptotic results suggest that there should be sharp differences between the plots for distributions with infinite variance ($\alpha < 2$) and the others ($\alpha > 2$).

Based on a finite sample and chosen points $s_i \in (0, 1)$, $i = 1, \dots, N$, one can estimate the scaling function by equation (1.5) for a fixed value of q. Repeating this for a range of q values makes it possible to give a plot of the empirical scaling function $\hat{\tau}_{N,n}$.

By examining the plot and comparing it with the baseline, it is possible to say something about the nature of the tails of the underlying distribution. If $\hat{\tau}_{N,n}(q)$ is above the baseline for $q < 2$ and nearly horizontal afterward, then true α is probably less than 2. By examining the point where the graph breaks, one can roughly estimate the interval containing α. If $\hat{\tau}_{N,n}(q)$ coincides with the baseline for $q < 2$ and diverges from it somewhere after $q > 2$, then the true α is probably greater than 2. The point at which deviation starts can be an estimate for α. This also establishes a graphical method for distinguishing two cases, whether $\alpha \leq 2$ or $\alpha > 2$.

If the graph coincides with the baseline, then we can suspect that the data does not exhibit heavy tails and that the moments are finite for the considered range of q. This way one can distinguish between heavy tails or not.

In the next subsection we show how the estimated scaling functions look like on several sets of simulated data. Subsection 2.2.6 contains examples of how conclusions can be made from the shape of the scaling functions.

2.2.3 Plots of the empirical scaling functions

The shape of the empirical scaling function is not always ideal as its asymptotic form. However, most plots are very close to their theoretical form. To illustrate this, we simulate 10 independent samples of size 1000 in six different settings. The first three cases studied are i.i.d. samples and others are stationary and weakly dependent, in accordance with the assumptions of Theorem 1. Figure 2.2 summarizes the plots of the empirical scaling functions (dotted) together with the corresponding asymptotic form (solid) and the baseline (dot-dashed). Here, s_i, $i = 1, \dots, N$ in (1.5) are chosen equidistantly in the interval $[0.1, 0.9]$ with $N = 23$. The scaling function is estimated at the points q_j chosen in the interval $[0, 10]$ with step 0.1.

The first group of samples is generated from a stable distribution with stable index equal to 1. A random variable Y has an α-stable distribution with index of stability $\alpha \in (0, 2)$, scale parameter $\sigma \in (0, \infty)$, skewness parameter $\beta \in [-1, 1]$ and shift parameter $\mu \in \mathbb{R}$, denoted by $Y \sim S_\alpha(\sigma, \beta, \mu)$ if its characteristic function has the following form

$$E\left[e^{i\zeta Y}\right] = \begin{cases} \exp\left\{-\sigma^\alpha |\zeta|^\alpha \left(1 - i\beta \operatorname{sign}(\zeta) \tan \frac{\pi\alpha}{2}\right) + i\zeta\mu\right\}, & \text{if } \alpha \neq 1, \\ \exp\left\{-\sigma|\zeta| \left(1 + i\beta \frac{2}{\pi} \operatorname{sign}(\zeta) \ln|\zeta|\right) + i\zeta\mu\right\}, & \text{if } \alpha = 1, \end{cases} \quad \zeta \in \mathbb{R}. \tag{2.15}$$

If $\mu = 0$ and $\alpha \neq 1$, or if $\beta = 0$ and $\alpha = 1$, then Y is said to have strictly stable distribution. The second group of samples is generated from the Student t-distribution with 3 degrees of freedom, a parameter that corresponds to the tail index. Recall that the probability density function of the Student t-distribution $T(\nu, \sigma, \mu)$ is

$$f_{T(\nu,\sigma,\mu)}(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu}\sigma\sqrt{\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}, \tag{2.16}$$

where $\sigma > 0$ is the scale parameter, ν the tail parameter (usually called degrees of freedom) and $\mu \in \mathbb{R}$ the location parameter. Figures 2.2a and 2.2b show that for both stable and Student case the empirical scaling functions are close to their theoretical form. Both plots are approximately bilinear and by identifying the breakpoint, one can roughly guess the tail index value. Also, it is clear from the shape of the empirical scaling functions that the variance is infinite in the first case and finite in the second. The third sample is generated from a standard normal distribution. From Figure 2.2c one can surely doubt the existence of heavy-tails in these samples since the empirical scaling functions almost coincide with the baseline $q/2$. This shows that the estimated scaling functions have the potential of providing a self-contained characterization of the tail.

Examples shown in Figures 2.2d-2.2f are based on dependent data. Dependent samples are generated as sample paths of two types of stochastic processes: Ornstein-Uhlenbeck (OU) type processes and diffusions. Recall that a stochastic process $X = \{X(t), t \geq 0\}$ is said to be of OU type if it satisfies a stochastic differential equation (SDE) of the form

$$dX(t) = -\lambda X(t)dt + dL(\lambda t), \quad t \geq 0, \tag{2.17}$$

where $L = \{L(t), t \geq 0\}$ is the background driving Lévy process (BDLP) and $\lambda > 0$. We consider strictly stationary solutions of SDE (2.17). The α-stable OU type process with parameter $\lambda > 0$ and $0 < \alpha < 2$ is the solution of the SDE (2.17) with L being the α-stable Lévy process. Since the distribution of increments for the BDLP L is known in this case, we use Euler's scheme of simulation by replacing differentials in Equation (2.17) with differences. Student OU type process has been introduced in Heyde & Leonenko (2005). It can be shown that for arbitrary $\lambda > 0$ there exists a strictly stationary stochastic process $X = \{X(t), t \geq 0\}$, which has a marginal distribution $T(\nu, \sigma, \mu)$ with density function (2.16) and BDLP L such that (2.17) holds. This stationary process X is referred to as the Student OU type process. Moreover, the cumulant transform of the BDLP L can be expressed as

$$\kappa_{L_1}(\zeta) = \log E\left[e^{i\zeta L_1}\right] = i\zeta\mu - \sqrt{\nu}\sigma|\zeta| \frac{K_{\nu/2-1}(\sqrt{\nu}\sigma|\zeta|)}{K_{\nu/2}(\sqrt{\nu}\sigma|\zeta|)}, \quad \zeta \in \mathbb{R},\ \zeta \neq 0,$$

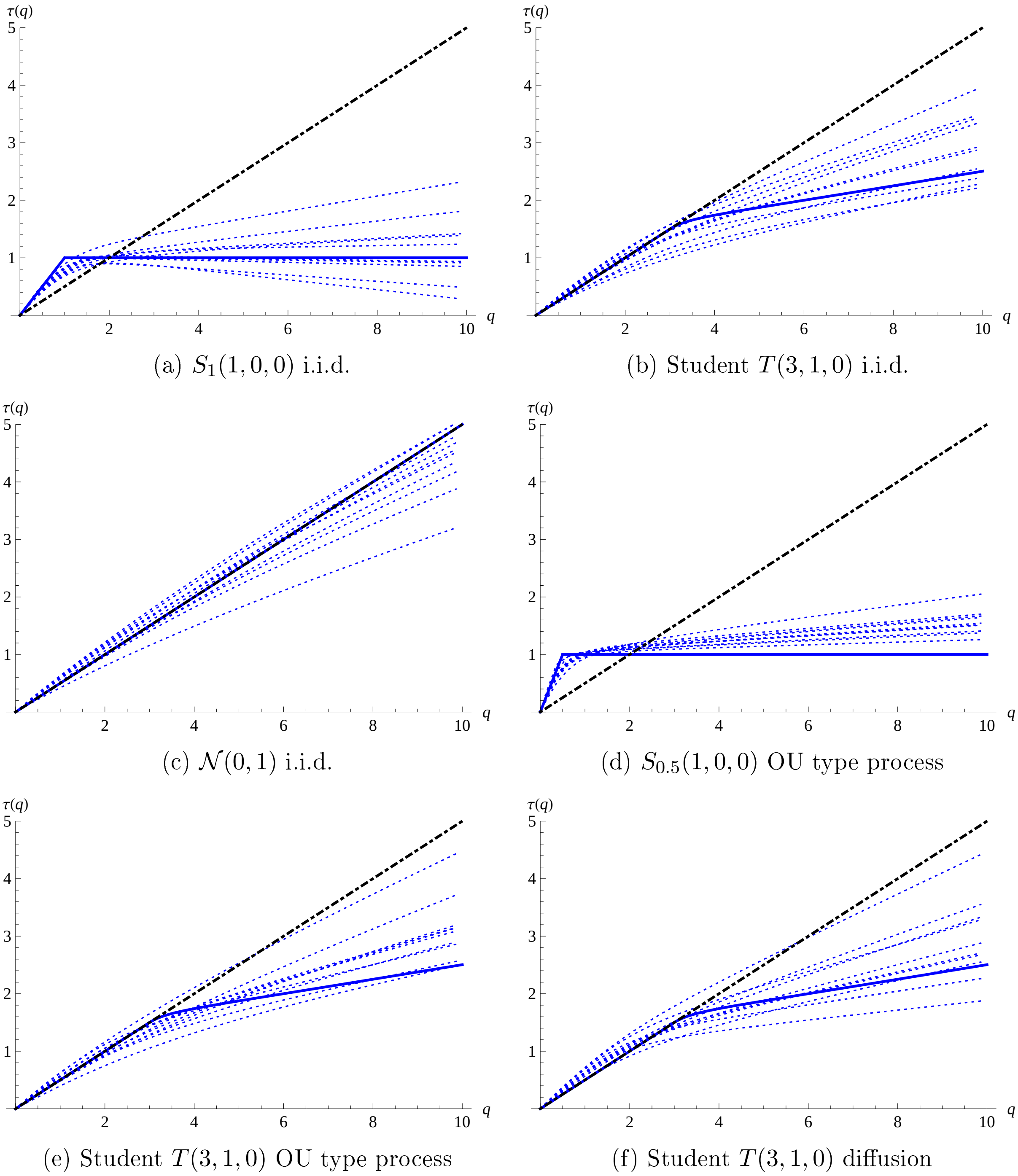
where K is the modified Bessel function of the third kind and $\kappa_{L_1}(0) = 0$ (Heyde & Leonenko (2005)). Since for the Student OU process the exact law of the increments of the BDLP is unknown, we use the approach introduced in Taufer & Leonenko (2009) to simulate Student OU process. This approach circumvents the problem of simulating the jumps of the BDLP and is easily applicable when an explicit expression of the cumulant transform is available. Both OU processes considered can be shown to posses strong mixing property with an exponentially decaying rate (see Masuda (2004)).

The last process considered is a stationary Student diffusion. In order to define the Student diffusion, we introduce the SDE:

$$dX(t) = -\theta\left(X(t) - \mu\right) dt + \sqrt{\frac{2\theta\sigma^2}{\nu - 1}\left(\nu + \left(\frac{X(t) - \mu}{\sigma}\right)^2\right)} dB(t), \quad t \geq 0, \tag{2.18}$$

where $\nu > 2$, $\sigma > 0$, $\mu \in \mathbb{R}$, $\theta > 0$, and $B = \{B(t), t \geq 0\}$ is a standard Brownian motion (BM) (see Bibby et al. (2005) and Heyde & Leonenko (2005)). The SDE (2.18) admits a unique ergodic Markovian weak solution $X = \{X(t), t \geq 0\}$, which is a diffusion process with the Student invariant distribution given by probability density function (2.16). The diffusion process which solves the SDE (2.18) is called the Student diffusion. If $X(0) =^d T(\nu, \sigma, \mu)$, the Student diffusion is strictly stationary. According to Leonenko & Šuvak (2010), the Student diffusion is a strong mixing process with an exponentially decaying rate. For the simulation of paths of the Student diffusion process with known values of parameters, we have used the Milstein scheme (for details see Iacus (2008)). Both OU processes were generated with autoregression parameter $\lambda = 1$ and diffusion was generated with $\theta = 2$.

From the examples on dependent data we can conclude that the shape of the empirical scaling function is not affected with this weak form of dependence present. Again, empirical scaling functions are very near their asymptotic form.



(a) $S_1(1,0,0)$ i.i.d.

(b) Student $T(3,1,0)$ i.i.d.

(c) $\mathcal{N}(0,1)$ i.i.d.

(d) $S_{0.5}(1,0,0)$ OU type process

(e) Student $T(3,1,0)$ OU type process

(f) Student $T(3,1,0)$ diffusion

Figure 2.2: Plots of the empirical scaling functions

2.2.4 Estimation methods

Besides the graphical method, a simple estimation methods for the unknown tail index can also be established based on the asymptotic behavior of the partition function and the empirical scaling function. As follows from the assumptions of Theorem 1, the estimators defined here should work well for stationary strong mixing samples, thus extending the problem from the simplest i.i.d. case. We propose here three methods and test their performance in the next subsection by means of simulation.

The basic idea of a **method M1** is to estimate α by fitting the empirical scaling function to the asymptotic form τ_α^∞. This is done by the ordinary least squares method. First we fix some points $s_i \in (0,1)$, $i = 1\dots,N$ in the definition of the empirical scaling function. For example, in simulations below we take equidistant points in the interval $[0.1, 0.9]$ with $N = 23$. Now, for points $q_i \in (0, q_{max})$, $i = 1,\dots,M$, we can calculate $\hat{\tau}_i = \hat{\tau}_{N,n}(q_i)$ using Equation (1.5). The estimator is defined as

$$\hat{\alpha}_1 = \arg\min_{\alpha\in(0,\infty)} \sum_{i=1}^{M} (\hat{\tau}_i - \tau_\alpha^\infty(q_i))^2. \tag{2.19}$$

For practical reasons, due to complexity of the expression for τ_α^∞, the method is divided into two cases: $\alpha \leq 2$ and $\alpha > 2$; i.e., the corresponding part of τ_α^∞ is used as a model function in (2.19), depending where the true value of α is. Therefore, it is necessary to first detect whether we are in the case of infinite variance or not. This can be accomplished by using the graphical method described earlier. In the inconclusive case, it is advisable to compute both estimates and compare the quality of the fit. For simulations, points q_i are chosen equidistantly in the interval $[0, 8]$ with step 0.1.

If $\alpha \leq 2$, the information on α in the asymptotic form of the empirical scaling function is hidden in the slope of the first part and the breakpoint. When $\alpha > 2$, the information on α appears in the breakpoint and in the complicated nonlinear expression of the second part. Method M1 tries to use all three parts in estimating α. Alternatively, we can base the method only on a breakpoint. Since the shape of τ_α^∞ is bilinear or approximately bilinear, we define **method M2** by fitting the following general continuous bilinear function to the empirical scaling function

$$\varsigma(q) = \begin{cases} aq, & \text{if } 0 < q \leq b, \\ cq + b(a-c), & \text{if } q > b. \end{cases} \tag{2.20}$$

The parameter of interest is b which corresponds to a breakpoint α and the estimator by method M2, $\hat{\alpha}_2$, is defined as

$$(\hat{a}, \hat{\alpha}_2, \hat{c}) = \arg\min_{(a,b,c)\in(0,\infty)\times(0,\infty)\times\mathbb{R}} \sum_{i=1}^{M} (\hat{\tau}_i - \varsigma(q_i))^2. \tag{2.21}$$

This method has the advantage of not depending on whether $\alpha \leq 2$ or $\alpha > 2$. Moreover, the second part depends on the rate of divergence of infinite moments and may not precisely follow the shape of τ_α^∞ on finite samples. This part is, however, usually approximately linear and fitting (2.20) makes the method more robust on the discrepancies from τ_α^∞.

For the third method, we go one step back to the asymptotic behavior of the partition function. As already discussed, results of Theorem 1 motivate viewing (2.3) as the regression model. It makes sense then to estimate α from a bivariate nonlinear regression of $\ln S_q(n, n^s)/\ln n$ on q and s with model function $R_\alpha(q,s)$. However, this complicated regression model may not always give good results and that is why we approach it in two steps. When defining the empirical scaling function we fixed q and first considered s as the variable in regression, while for the third method we go the other way around.

For the moment let us fix $s \in (0,1)$. The limit $R_\alpha(q,s)$ in Theorem 1 has the following form for $\alpha \leq 2$:

$$R_\alpha(q,s) = \begin{cases} \frac{s}{\alpha} q, & \text{if } q \leq \alpha, \\ \frac{1}{\alpha} q + s - 1, & \text{if } q > \alpha, \end{cases} \tag{2.22}$$

and if $\alpha > 2$ the limit is

$$R_\alpha(q,s) = \begin{cases} \frac{s}{2} q, & \text{if } q \leq \alpha, \\ \max\left\{\frac{1}{\alpha} q + s - 1, \frac{s}{2} q\right\}, & \text{if } q > \alpha. \end{cases}$$

Notice that in this case bilinear function $q \mapsto \max\left\{\frac{1}{\alpha} q + (s-1), \frac{s}{2} q\right\}$ has a breakpoint at

$$\overline{q} = \frac{s-1}{\frac{s}{2} - \frac{1}{\alpha}}.$$

If $s \in (2/\alpha, 1)$, $\overline{q} < 0$ and there is no breakpoint in the range of positive q values. If $s \in (0, 2/\alpha)$, we can write for the case $\alpha > 2$:

$$R_\alpha(q,s) = \begin{cases} \frac{s}{2} q, & \text{if } q \leq \overline{q}, \\ \frac{1}{\alpha} q + (s-1), & \text{if } q > \overline{q}. \end{cases} \tag{2.23}$$

So, if $s \in (0, 2/\alpha)$, $q \mapsto R_\alpha(q,s)$ is bilinear continuous and, motivated by the regression model (2.3), we can fit function (2.20) to points

$$\left\{ \left(q_i, \frac{\ln S_{q_i}(n, n^s)}{\ln n} \right) : i = 1, \dots, M \right\}. \tag{2.24}$$

This way we find the estimated parameters of (2.20)

$$(\hat{a}_s, \hat{b}_s, \hat{c}_s) = \underset{(a,b,c) \in (0,\infty) \times (0,\infty) \times \mathbb{R}}{\arg\min} \sum_{i=1}^{M} \left(\frac{\ln S_{q_i}(n, n^s)}{\ln n} - \varsigma(q_i) \right)^2.$$

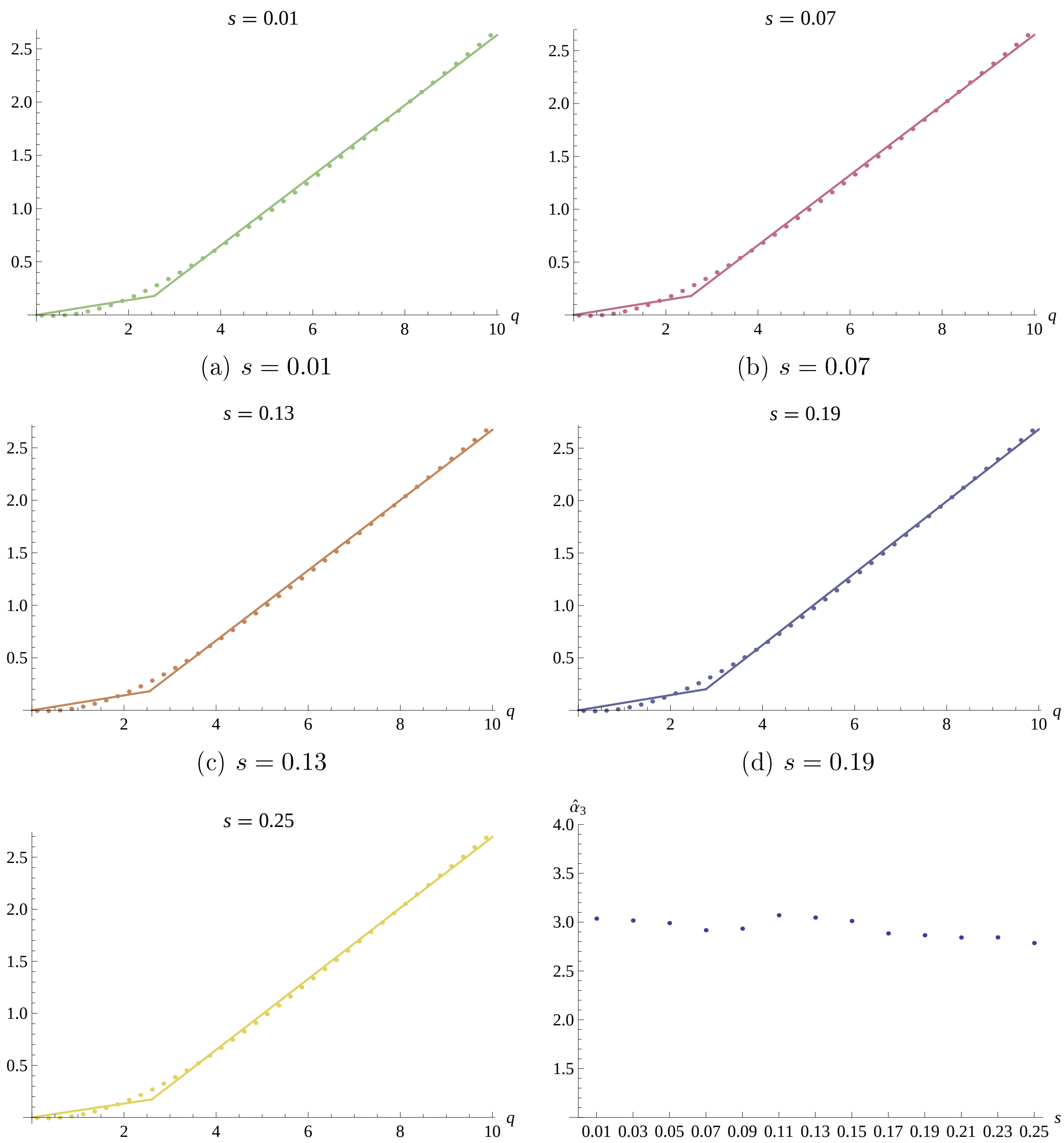
In order to define a method that does not depend on $\alpha \leq 2$ or not, we notice that, if $s \in (0, 2/\alpha)$, the common part of (2.22) and (2.23) giving information on α is the slope of the second part. Therefore, $1/\hat{c}_s$ is an estimate for α for each $s \in (0, 2/\alpha)$. Since we do not consider problems with tail index greater than, say 8, s can be chosen in the interval $(0, 0.25)$. We define a **method M3** estimator by averaging over a set of values $s_j \in (0, 0.25)$, $j = 1, \dots, N$:

$$\hat{\alpha}_3 = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{c}_{s_i}}. \tag{2.25}$$

In simulations and examples below we take s_j equidistantly in the interval $[0.01, 0.25]$ with step 0.02.

To make the estimation process by this method more clear, we illustrate it on a simple example with data consisting of 1000 points generated from the Student t distribution $T(3, 1, 0)$ (Figure 2.3). Figures 2.3a-2.3e show sets of points (2.24) for different s values, together with the fitted bilinear function (2.20). The reciprocal of the slope of the second part corresponds to the tail index α. These values for a range of s are shown in Figure 2.3f. By averaging, we obtain the final estimate by method M3 to be 2.942.

Although method M3 may seem promising, it has a serious drawback of not being scale invariant. Indeed, scaling the data by some factor c would scale the partition function by a factor $|c|^q$. As the samples are finite, this produces an additional term $\ln |c|^q / \ln n$ in the ordinate of the set of points (2.24) that affects the estimation process. This makes the practical use of the method very limited. Nonetheless, we will include it in the simulation study below in which the generated data will be chosen from distributions with scale parameter equal to 1. These problems do not appear for the methods based on the empirical scaling functions. Scaling functions are robust to scale change as they are based only on the slope obtained for a fixed value of q.



(a) $s = 0.01$

(b) $s = 0.07$

(c) $s = 0.13$

(d) $s = 0.19$

(e) $s = 0.25$

(f) Estimates for a range of s values

Figure 2.3: Estimation process by method M3 on Student $T(3, 1, 0)$ data

2.2.5 Simulation study

In this subsection we provide a simulation study in order to investigate finite sample properties of the estimators defined in the previous subsection. We choose to generate i.i.d. random samples from the following distributions: stable distribution $S_\alpha(1, 0, 0)$ with $\alpha = 0.5$ and $\alpha = 1.5$, Student t-distribution $T(\alpha, 1, 0)$ with $\alpha = 0.5, 1.5, 2.5, 3, 4$ and Pareto distribution with tail index $\alpha = 0.5, 1, 1.5$ and scale parameter 1. Recall that the random variable X has Pareto distribution if its tail distribution is given by

$$P(X > x) = \begin{cases} \left(\frac{x_m}{x}\right)^\alpha, & x \geq x_m, \\ 1, & x < x_m, \end{cases}$$

where $x_m > 0$ is the minimal possible value (scale parameter) and $\alpha > 0$ is the tail index. For each distribution we have generated 250 samples of length 1000 and computed the estimators (2.19), (2.21) and (2.25). In addition to i.i.d. samples, we have generated samples from stable and Student OU type process and Student diffusion, in the same way as it was done in Subsection 2.2.3.

In order to give a picture of the performance of the estimators, we compare them with the Hill estimator. Since the Hill estimator depends on the number of order statistics k, we do this in the following manner. For each sample we compute the value of the Hill estimator for each k in the range $\{1, \dots, 250\}$. After this is done for all 250 samples, we choose k such that the mean square error (MSE) is minimal. So this is the smallest MSE that can be achieved with the Hill estimator on the generated samples if k is fixed for each distribution. It is clear that this comparison is unfair to the estimators proposed in Subsection 2.2.4, as in practice one can hardly choose an optimal k for the Hill estimator. We note that besides the Hill estimator, we computed in the same way the moment estimator and the Pickands estimator. However, as neither of these is significantly better than the Hill estimator, we do not report their results in the following.

Table 2.1: Bias $(\hat{\alpha} - \alpha)$ of the estimators based on 250 samples of length 1000

distribution	α	M1	M2	M3	Hill	optimal k
stable	0.5	0.1252	0.1085	**0.0126**	-0.0218	98
stable	1.5	0.1533	0.3352	0.2914	**-0.0275**	211
Student	0.5	0.1182	0.1194	0.0276	**-0.0065**	238
Student	1.5	**0.0716**	0.1607	0.0834	-0.0938	94
Student	2.5	0.5669	-0.1396	**0.0025**	-0.2375	44
Student	3	0.3432	-0.4268	**-0.1053**	-0.4241	39
Student	4	**-0.1861**	-1.1929	-0.4638	-0.6033	20
Pareto	0.5	0.0654	0.1204	-0.0048	**0.0011**	250
Pareto	1	-0.1391	0.2349	**0.0003**	0.0023	250
Pareto	1.5	0.1736	0.1449	0.0381	**0.0034**	250
stable OU	0.5	0.1805	0.1965	0.0876	**-0.0289**	159
stable OU	1.5	0.0846	0.3940	0.3113	**-0.0161**	209
Student OU	3	1.1333	-0.4035	0.9498	**-0.3224**	32
Student OU	4	**0.4769**	-1.2158	1.3796	-0.6695	23
Student diff	3	**0.0073**	-0.9425	-0.4411	-0.8703	31
Student diff	4	**-0.5066**	-1.6429	-0.5128	-1.0137	14

Table 2.1 reports the estimated bias as well as the optimal k for the Hill estimator, while Table 2.2 shows the root of the mean square error (RMSE). Bold values show the

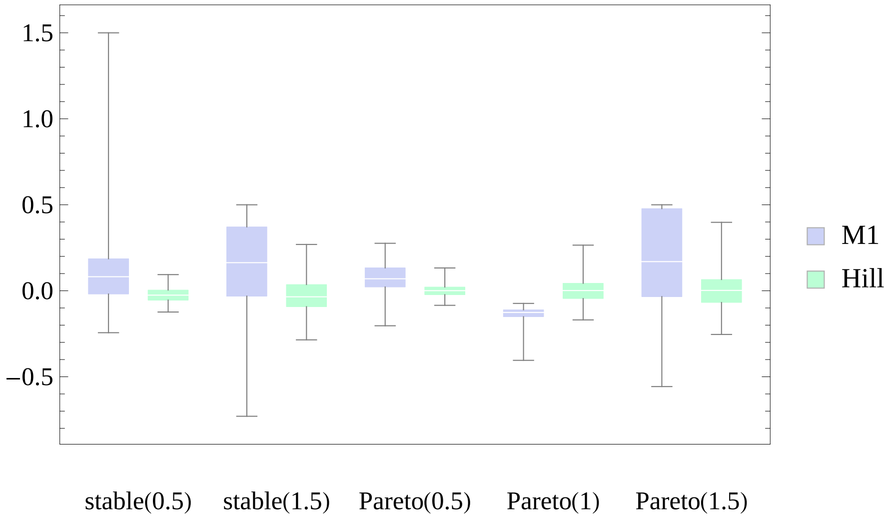
Table 2.2: RMSE of the estimators based on 250 samples of length 1000

distribution	α	M1	M2	M3	Hill
stable	0.5	0.2474	0.1602	0.0861	**0.0504**
stable	1.5	0.3027	0.4654	0.4549	**0.1001**
Student	0.5	0.2039	0.1606	0.0869	**0.0307**
Student	1.5	0.2708	0.2897	0.2783	**0.1652**
Student	2.5	0.9884	**0.2894**	0.4030	0.3950
Student	3	1.0232	0.4913	**0.4556**	0.5704
Student	4	1.3163	1.2171	**0.6687**	0.9418
Pareto	0.5	0.1098	0.1673	0.0754	**0.0316**
Pareto	1	0.1470	0.3387	0.1565	**0.0632**
Pareto	1.5	0.3352	0.2906	0.2628	**0.0948**
stable OU	0.5	0.2778	0.2386	0.1389	**0.0668**
stable OU	1.5	0.2563	0.5112	0.4752	**0.1203**
Student OU	3	1.9149	**0.4612**	1.1514	0.5337
Student OU	4	1.7082	1.2410	1.7274	**0.9081**
Student diff	3	**0.5985**	0.9810	0.6821	0.9757
Student diff	4	1.1394	1.6655	**0.9150**	1.3743

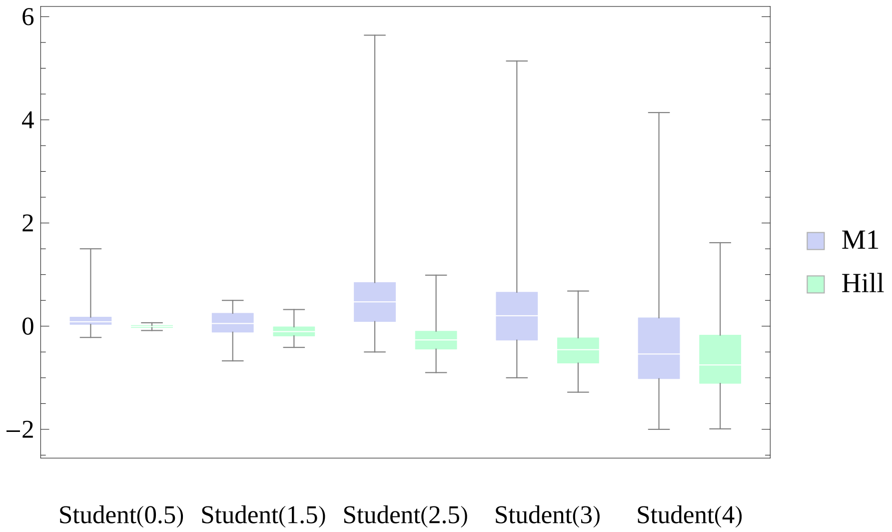
best value for each case considered.

First, if we compare the three proposed methods, one can notice that method M2 is outperformed by M1 and M3 by bias, while there is no much difference when it comes to RMSE. The methods M1 and M3 have similar performance with M3 slightly better when RMSE is considered. The disadvantage of method M1 is that it depends on the knowledge of whether $\alpha \leq 2$ or $\alpha > 2$ and in some cases the estimated value may depend on the choice of the maximal q value taken into consideration. On the other hand, method M3 is not scale invariant which makes it less suitable for the practical examples. To summarize these arguments, in a general situation a method M1 is recommended to use.

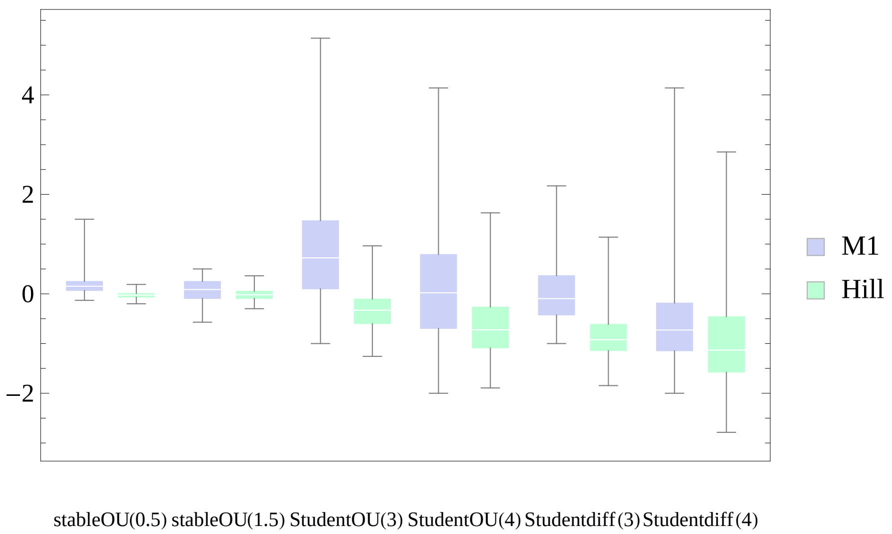
When compared by the RMSE, the Hill estimator shows smaller variability than the other three estimators, especially for smaller values of α, while for larger α the proposed methods give better results. However, one should notice that the differences are not substantial. Moreover, the Hill estimator's performance is reported for the optimal choice of k, which may not be achieved in practice. On the other hand, computation of the proposed estimators is straightforward and does not require choosing extra parameters. If we take this in mind, the new methods can be seen as a good alternative to Hill estimator. The comparison of method M1 and the Hill estimator is also shown by boxplots in Figure 2.4. We illustrate more advantages of the estimator M1 on examples in the next subsection.



(a) Stable and Pareto distribution



(b) Student distribution



(c) Dependent samples

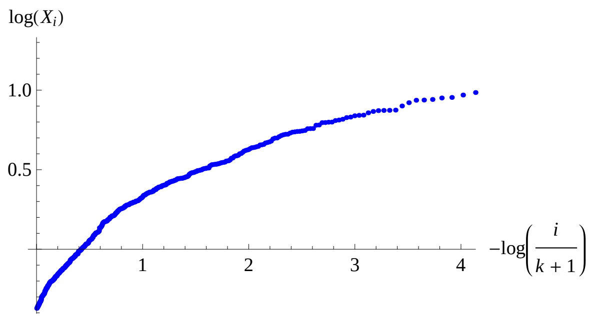
Figure 2.4: Boxplots of $\hat{\alpha} - \alpha$ for $\hat{\alpha}_1$ and $\hat{\alpha}^{Hill}$

2.2.6 Examples and comparison

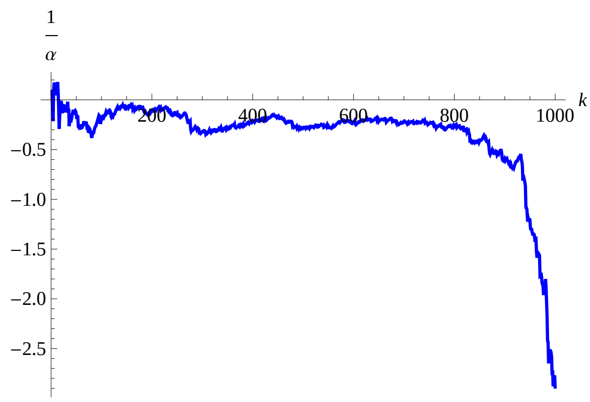
In this subsection we provide several examples to illustrate how the proposed methods work and compare them with the existing methods.

Example 1 - non heavy-tailed data

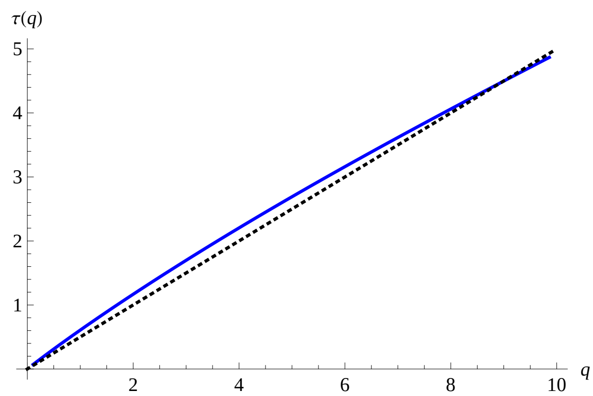
With this example we try to illustrate the potential of the empirical scaling functions as a graphical method that can distinguish between heavy-tailed and light-tailed scenario. Different methods are tested on a random sample from standard normal distribution of size 2000. The results are shown in Figure 2.5. QQ plot for 500 largest data points exhibits nonlinearity, thus indicating that Pareto type tail is not a good fit for the data (Figure 2.5a). For the purpose of tail analysis we plot the extreme value index ($1/\alpha$) values estimated by the moment estimator. From Figure 2.5b one can see that the plot stabilizes at a negative value near zero. This is an indication of light tails. The scaling function shown in Figure 2.5c is completely in accordance with this analysis. Indeed, it almost coincides with the baseline that corresponds to a non heavy-tailed data.



(a) QQ plot ($k = 500$)



(b) Moment estimator plot

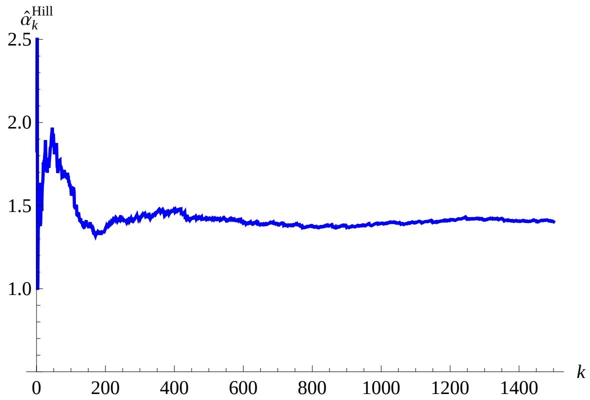


(c) Empirical scaling function

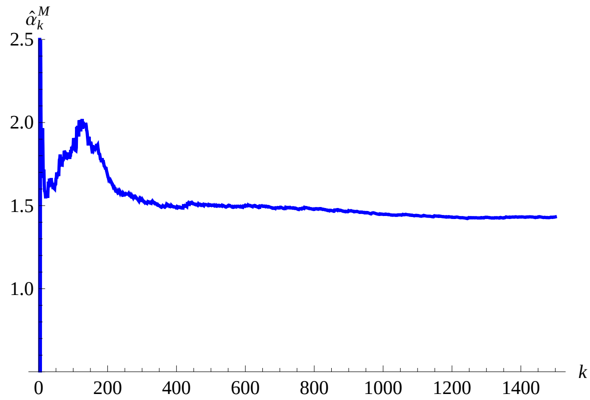
Figure 2.5: Example 1 - non heavy-tailed data

Example 2 - Danish fire insurance claims

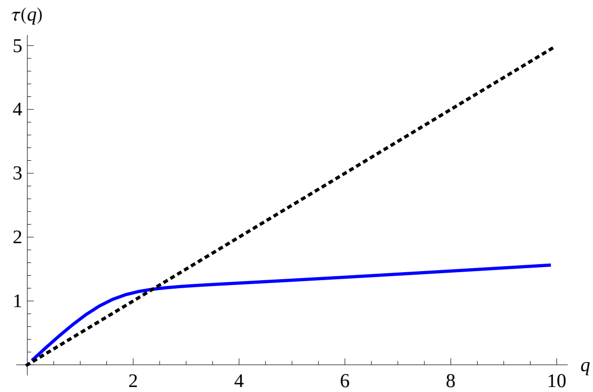
The second example is a practical one. The data we analyze are the Danish fire insurance claims in the period from 1980 to 1990. There are 2167 observations and the amounts are in millions of Danish Kroner.[1] The same example has been considered in (Embrechts et al. 1997, Example 6.2.9) and in (Resnick 2007, Fig. 1.6., 4.5., 4.7.). The analysis made in Resnick (2007) suggests that the data exhibits heavy-tails and the tail index estimate is around 1.4 (see (Resnick 2007, Fig. 4.5.)). Hill and moment estimator plots (Figures 2.6a and 2.6b) confirm the index value is around 1.5. The empirical scaling function, together with the baseline, is shown in Figure 2.6c. Mean has been subtracted from the data to adjust to the assumptions of Theorem 1. The scaling function is approximately bilinear: the first part of the plot has a slope greater than the baseline and the second part is nearly horizontal. This points out that the variance is infinite. The breakpoint occurs at around 1.5, which indicates a possible value of the tail index. Estimation by method M1 gives $\hat{\alpha}_1 = 1.419$, which is consistent with the previous analysis done in Resnick (2007).



(a) Hill plot



(b) Moment estimator plot



(c) Empirical scaling function

Figure 2.6: Example 2 - Danish fire insurance claims

[1] The data can be obtained from: `http://www.ma.hw.ac.uk/~mcneil/data.html`

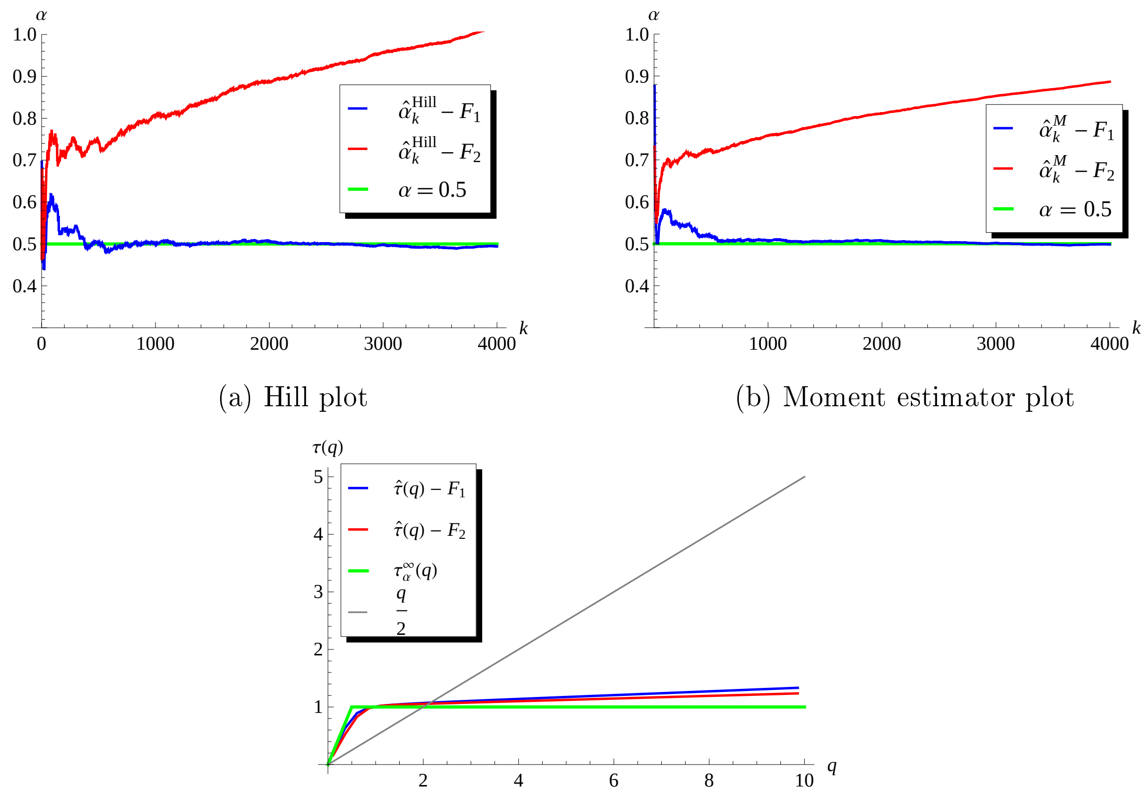
Example 3 - departure from Pareto tail

The Hill estimator, as well as many others, is known to behave poorly if the slowly varying function in the tail is far away from a constant. We compare this behavior with the performance of the empirical scaling functions on the same samples. Consider two distribution F_1, F_2 defined by their tail distribution functions

$$\overline{F}_1(x) = 1 - F_1(x) = \frac{1}{x^{\frac{1}{2}}}, \quad x \geq 1, \tag{2.26}$$

$$\overline{F}_2(x) = 1 - F_2(x) = \frac{e^{\frac{1}{2}}}{x^{\frac{1}{2}} \ln x}, \quad x \geq e. \tag{2.27}$$

Both distributions are heavy-tailed with tail index equal to 1/2. We generate samples from these two distributions with 5000 observations. The corresponding Hill plots are shown in Figure 2.7a. While for the Pareto distribution F_1 the Hill plot provides very good results, for F_2 it is impossible to draw any conclusion about the value of the tail index. The plot fails to stabilize at some value and produces a departure from the true index value. This is sometimes called the Hill horror plot (see Embrechts et al. (1997)). The result is similar with the moment estimator: a non-constant slowly varying function in the tail produces a significant bias, as shown in Figure 2.7b.



(a) Hill plot

(b) Moment estimator plot

(c) Scaling functions

Figure 2.7: Example 3 - departure from Pareto tail

Figure 2.7c shows the empirical scaling functions for the same samples, together with the theoretical one and the baseline. One can see that the empirical scaling functions are very close to their asymptotic shape, especially in the first part of the plot, before the breakpoint. It seems that nonconstant slowly varying function affects the estimation but the effect is not so dramatic as for the standard estimators. Calculating estimates using method M1 yields $\hat{\alpha}_1 = 0.645$ for F_1 and $\hat{\alpha}_1 = 0.72$ for F_2.

Chapter 3

Asymptotic scaling of the linear fractional stable noise

In the previous chapter, we have analyzed asymptotic properties of the partition function for weakly dependent heavy-tailed sequences. In this chapter we do this for the linear fractional stable noise, which is an example of a self-similar stationary sequence exhibiting both strong dependence and heavy-tails.

We start with the definition and basic properties of the linear fractional stable motion and then establish asymptotic behavior of the partition function and the empirical scaling function. These results are used to define estimation methods for the parameters of the process. The methods are tested by simulations and on some real data examples.

3.1 Linear fractional stable motion

Empirical time series which appear in many applications display both the "Joseph" and "Noah" effects, as coined by Mandelbrot after the biblical figures of Joseph and Noah (see e.g. Mandelbrot (1997)). While the Joseph effect refers to long-range dependence of the increments, the Noah effect refers to their high variability, as expressed by the power law tails of the marginal distributions. Fractional Brownian motion (FBM) is an example of a process exhibiting only the Joseph effect: its increments are long-range dependent but with normal marginal distribution. On the other hand, the α-stable Lévy process with $0 < \alpha < 2$ exhibits only the Noah effect: it has independent but heavy-tailed increments with tail index equal to α.

An example of a stochastic process which exhibits both effects is the linear fractional stable motion (LFSM). LFSM can be defined through the stochastic integral

$$X(t) = \frac{1}{C_{H,\alpha}} \int_{\mathbb{R}} \left((t-u)_+^{H-1/\alpha} - (-u)_+^{H-1/\alpha} \right) M(du), \quad t \in \mathbb{R}, \tag{3.1}$$

where $\alpha \in (0,2)$, $0 < H < 1$, $(x)_+ = \max(x,0)$ and where M is a random noise. More specifically, M is an α-stable random measure on $\mathbb{R}$ with Lebesgue control measure λ and skewness β. This means, first, that M is a σ-additive mapping from $\mathcal{E} = \{A \in \mathcal{B}(\mathbb{R}) : \lambda(A) < \infty\}$ to the space of random variables and that it is independently scattered: if $A_1, A_2 \in \mathcal{E}$ are disjoint sets, then $M(A_1)$ and $M(A_2)$ are independent random variables. Secondly, for all sets $A \in \mathcal{E}$, $M(A)$ has an α-stable distribution with scale parameter $\lambda(A)^{1/\alpha}$ and skewness parameter β, i.e. $M(A) \sim S_\alpha(\lambda(A)^{1/\alpha}, \beta, 0)$. If $\alpha = 1$ we assume $\beta = 0$, but for other values of α, M is allowed to be skewed. Recall that stable distributions $S_\alpha(\sigma, \beta, \mu)$ were defined in (2.15).

If the constant $C_{H,\alpha}$ in the representation (3.1) is chosen such that the scaling parameter of $X(1)$ equals 1, i.e.

$$C_{H,\alpha} = \left(\int_{\mathbb{R}} \left| (1-u)_+^{H-1/\alpha} - (-u)_+^{H-1/\alpha} \right|^\alpha du \right)^{1/\alpha},$$

then the process is called standard LFSM. The stationary sequence $Y_i = X(i) - X(i-1)$, $i \in \mathbb{N}$ is referred to as the linear fractional stable noise (LFSN). The LFSM process $\{X(t)\}$ is H-self-similar with stationary increments (see (Samorodnitsky & Taqqu 1994, Proposition 7.4.2)). For each t, $X(t)$ has a strictly stable distribution with stable index α (Samorodnitsky & Taqqu 1994, Proposition 7.4.3). The parameter α governs the tail behavior of the marginal distributions in the sense that for each t, $X(t)$ is heavy-tailed with tail index α, i.e. $P(|X(t)| > x) = L(x)x^{-\alpha}$ where L is a slowly varying function, more precisely, L is constant at infinity. In particular, we have that $E|X(t)|^q = \infty$ for $q \geq \alpha$. More details on the LFSM and its properties can be found in the monograph Samorodnitsky & Taqqu (1994).

Setting $\alpha = 2$ in (3.1) reduces the LFSM to the FBM. Dependence structure of the FBM $\{B_H(t), t \geq 0\}$ can be described using autocovariance function $\gamma(h) = EY_1Y_{h+1}$, where $Y_i = B_H(i) - B_H(i-1)$, $i \in \mathbb{N}$. It can be shown that $\gamma(h) \sim H(2H-1)h^{2H-2}$ as $h \to \infty$. This implies that $\sum_{h=1}^{\infty} \gamma(h) = \infty$ for $1/2 < H < 1$, which is a property known as the long-range dependence. The case $0 < H < 1/2$ is referred to as the negative dependence, since the coefficient $H(2H-1)$ is negative in this case. As the second moment of the LFSM is infinite, dependence cannot be characterized with the covariance function. For stable processes this can be done using codifferences (Samorodnitsky & Taqqu (1994)), but there are also many other competing approaches (see e.g. Magdziarz (2009)). Nevertheless, by analogy to the FBM, the case of LFSM with $H > 1/\alpha$ is referred to as a long-range dependence and the case $H < 1/\alpha$ as negative dependence.

3.2 Asymptotic scaling

Suppose $\{X(t), t \geq 0\}$ is LFSM that is sampled in a regularly spaced time instants, $X(\delta), X(2\delta), \dots, X(n\delta)$. For simplicity of notation, we assume $\delta = 1$, so we have a sample $X_1, \dots, X_n$. We denote by $Y_i = X_i - X_{i-1}$, $i \in \mathbb{N}$ the corresponding LFSN. Partition function can now be written as

$$S_q(n,t) = \frac{1}{\lfloor n/t \rfloor} \sum_{i=1}^{\lfloor n/t \rfloor} \left| \sum_{j=1}^{\lfloor t \rfloor} Y_{(i-1)\lfloor t \rfloor + j} \right|^q = \frac{1}{\lfloor n/t \rfloor} \sum_{i=1}^{\lfloor n/t \rfloor} \left| X_{i\lfloor t \rfloor} - X_{(i-1)\lfloor t \rfloor} \right|^q,$$

where $q \in \mathbb{R}$ and $1 \leq t \leq n$. Asymptotic properties of $S_q(n,t)$ for LFSM have been considered in the context of multifractality detection in Heyde & Sly (2008). We go over the methodology of the previous chapter to establish the asymptotic properties. Although dependence restricts many of the arguments used in the previous chapter, self-similarity simplifies the proofs at many points.

In our analysis we will also include a range of negative q values. Although this may seem unusual, finite negative order moments provide additional information on the value of the Hurst parameter H. In particular, for $q \in (-1, 0)$, stable-distributed random variables have finite q-th absolute moment, since their probability density function is bounded (see e.g. Zolotarev (1986)).

The main argument in establishing asymptotic properties of the partition function is based on the following lemma. A similar result has been proved in Heyde & Sly (2008), yet we prove it here by much simpler arguments.

Lemma 2. *Suppose $(Y_i, i \in \mathbb{N})$ is a LFSN. Then for $q > \alpha$,*

$$\frac{\ln\left(\sum_{i=1}^n |Y_i|^q\right)}{\ln n} \xrightarrow{P} \frac{q}{\alpha},$$

as $n \to \infty$.

Proof. Let $\varepsilon > 0$. Suppose $\delta < \varepsilon/(q-\alpha)$ and define

$$Z_{i,n} = Y_i \mathbf{1}\left(|Y_i| \leq n^{\frac{1}{\alpha}+\delta}\right), \quad i = 1, \dots, n, \; n \in \mathbb{N}.$$

It follows from Karamata's theorem (Resnick (2007)) that for arbitrary $r > \alpha$

$$\begin{aligned} E|Z_{i,n}|^r &= \int_0^\infty P(|Z_{i,n}|^r > x)dx = \int_0^{n^{r(\frac{1}{\alpha}+\delta)}} P(|Y_i|^r > x)dx \\ &= \int_0^{n^{r(\frac{1}{\alpha}+\delta)}} L(x^{\frac{1}{r}})x^{-\frac{\alpha}{r}}dx \leq C_1 L(n^{\frac{1}{\alpha}+\delta})n^{r(\frac{1}{\alpha}+\delta)(-\frac{\alpha}{r}+1)} = C_1 L(n^{\frac{1}{\alpha}+\delta})n^{\frac{r}{\alpha}-1+\delta(r-\alpha)} \end{aligned} \tag{3.2}$$

Next, notice that

$$P\left(\max_{i=1,\dots,n}|Y_i|>n^{\frac{1}{\alpha}+\delta}\right)\leq\sum_{i=1}^{n}P\left(|Y_i|>n^{\frac{1}{\alpha}+\delta}\right)\leq C_2 n\frac{L(n^{\frac{1}{\alpha}+\delta})}{(n^{\frac{1}{\alpha}+\delta})^{\alpha}}\leq C_2\frac{L(n^{\frac{1}{\alpha}+\delta})}{n^{\alpha\delta}}.$$

Now by partitioning on the event

$$\{Y_i=Z_{i,n},\ i=1,\dots,n\}=\left\{Y_i\leq n^{\frac{1}{\alpha}+\delta},\ i=1,\dots,n\right\}=\left\{\max_{i=1,\dots,n}|Y_i|\leq n^{\frac{1}{\alpha}+\delta}\right\},$$

using Markov's inequality and (3.2) we have

$$\begin{aligned}
&P\left(\frac{\ln\left(\sum_{i=1}^{n}|Y_i|^q\right)}{\ln n}>\frac{q}{\alpha}+\varepsilon\right)=P\left(\sum_{i=1}^{n}|Y_i|^q>n^{\frac{q}{\alpha}+\varepsilon}\right)\\
&\leq P\left(\sum_{i=1}^{n}|Y_i|^q>n^{\frac{q}{\alpha}+\varepsilon},\max_{i=1,\dots,n}|Y_i|\leq n^{\frac{1}{\alpha}+\delta}\right)+P\left(\max_{i=1,\dots,n}|Y_i|>n^{\frac{1}{\alpha}+\delta}\right)\\
&\leq P\left(\sum_{i=1}^{n}|Z_{i,n}|^q>n^{\frac{q}{\alpha}+\varepsilon}\right)+P\left(\max_{i=1,\dots,n}|Y_i|>n^{\frac{1}{\alpha}+\delta}\right)\leq\frac{nE\,|Z_{i,n}|^q}{n^{\frac{q}{\alpha}+\varepsilon}}+C_2\frac{L(n^{\frac{1}{\alpha}+\delta})}{n^{\alpha\delta}}\\
&\leq\frac{n}{n^{\frac{q}{\alpha}+\varepsilon}}C_1L(n^{\frac{1}{\alpha}+\delta})n^{\frac{q}{\alpha}-1+\delta(q-\alpha))}+C_2\frac{L(n^{\frac{1}{\alpha}+\delta})}{n^{\alpha\delta}}\leq C_1L(n^{\frac{1}{\alpha}+\delta})n^{\delta(q-\alpha)-\varepsilon}+C_2\frac{L(n^{\frac{1}{\alpha}+\delta})}{n^{\alpha\delta}}\to 0,
\end{aligned}$$

as $n\to\infty$, since $\delta(q-\alpha)-\varepsilon<0$ and $L(x)$ is slowly varying, thus bounded by any positive power of x, as $x\to\infty$.

For the reverse inequality notice that since Y_i is a stationary strictly α-stable sequence corresponding to a dissipative flow it follows by (Samorodnitsky 2004, Theorem 4.8) that $\max_{i=1,\dots,n}|Y_i|/n^{1/\alpha}$ converges in distribution to some positive random variable. So for any $\delta>0$

$$P\left(\max_{i=1,\dots,n}|Y_i|<n^{\frac{1}{\alpha}-\delta}\right)\to 0,\qquad\text{as } n\to\infty.$$

Now it follows that

$$P\left(\frac{\ln\left(\sum_{i=1}^{n}|Y_i|^q\right)}{\ln n}<\frac{q}{\alpha}-\varepsilon\right)=P\left(\sum_{i=1}^{n}|Y_i|^q<n^{\frac{q}{\alpha}-\varepsilon}\right)\leq P\left(\max_{i=1,\dots,n}|Y_i|<n^{\frac{1}{\alpha}-\frac{\varepsilon}{q}}\right)\to 0,$$

as $n\to\infty$ which proves the statement.

□

Theorem 3. *Suppose $(Y_i, i\in\mathbb{N})$ is a LFSN. Then for every $q>-1$ and every $s\in(0,1)$*

$$\frac{\ln S_q(n,n^s)}{\ln n}\xrightarrow{P}R_{H,\alpha}(q,s):=\begin{cases}sqH, & \text{if } q\leq\alpha,\\ s\left(1+qH-\frac{q}{\alpha}\right)+\frac{q}{\alpha}-1, & \text{if } q>\alpha,\end{cases}\tag{3.3}$$

as $n\to\infty$.

Proof. By H-self-similarity of the LFSM it follows that

$$S_q(n,n^s) \stackrel{d}{=} \frac{n^{sqH}}{n^{1-s}} \sum_{i=1}^{n^{1-s}} |X(i)-X(i-1)|^q = \frac{n^{sqH}}{n^{1-s}} \sum_{i=1}^{n^{1-s}} |Y_i|^q . \tag{3.4}$$

The LFSN sequence $(Y_i, i \in \mathbb{N})$ is a stable mixed moving average, which is known to be ergodic (see Cambanis et al. (1987), Surgailis et al. (1993) or Pipiras & Taqqu (2002)). For $q \in (-1,\alpha)$, $E|Y_i|^q < \infty$, so it follows by the ergodic theorem that

$$\frac{1}{n^{1-s}} \sum_{i=1}^{n^{1-s}} |Y_i|^q \to E|X(1)|^q \text{ a.s.}$$

and in particular

$$\ln\left(\frac{n^{sqH}}{n^{1-s}} \sum_{i=1}^{n^{1-s}} |Y_i|^q\right) - \ln n^{sqH} \stackrel{P}{\to} \ln E|X(1)|^q.$$

Since

$$\ln\left(\frac{n^{sqH}}{n^{1-s}} \sum_{i=1}^{n^{1-s}} |Y_i|^q\right) - \ln n^{sqH} \stackrel{d}{=} \ln n \left(\frac{\ln S_q(n,n^s)}{\ln n} - sqH\right),$$

this implies the statement of the theorem for $q<\alpha$.

Now we consider the case $q > \alpha$. We have by (3.4) that

$$\begin{aligned}
&\frac{\ln S_q(n,n^s)}{\ln n} - s\left(1+qH-\frac{q}{\alpha}\right) - \left(\frac{q}{\alpha}-1\right) \\
&\stackrel{d}{=} \frac{(sqH-1+s)\ln n + \ln\left(\sum_{i=1}^{n^{1-s}} |Y_i|^q\right)}{\ln n} - s\left(1+qH-\frac{q}{\alpha}\right) - \left(\frac{q}{\alpha}-1\right) \\
&= \frac{\ln\left(\sum_{i=1}^{n^{1-s}} |Y_i|^q\right)}{\ln n} - \frac{q}{\alpha}(1-s) \\
&= \frac{\ln\left(\sum_{i=1}^{n^{1-s}} |Y_i|^q\right)}{\ln n^{1-s}}(1-s) - \frac{q}{\alpha}(1-s).
\end{aligned}$$

Since by Lemma 2

$$\frac{\ln\left(\sum_{i=1}^{n^{1-s}} |Y_i|^q\right)}{\ln n^{1-s}} \stackrel{P}{\to} \frac{q}{\alpha}, \qquad \text{as } n\to\infty,$$

it follows that

$$\begin{aligned}
&P\left(\left|\frac{\ln S_q(n,n^s)}{\ln n} - s\left(1+qH-\frac{q}{\alpha}\right) - \left(\frac{q}{\alpha}-1\right)\right| > \varepsilon\right) \\
&\qquad = P\left(\left|\frac{\ln\left(\sum_{i=1}^{n^{1-s}} |Y_i|^q\right)}{\ln n^{1-s}} - \frac{q}{\alpha}\right|(1-s) > \varepsilon\right) \to 0,
\end{aligned}$$

as $n\to\infty$ and this proves the second case. The case $q=\alpha$ follows by the same argument as (c) part of the proof of Theorem 1. □

The previous theorem can be seen as an analogue of Theorem 1. As opposed to the setting of the previous chapter, the limit now additionally depends on the value of the parameter H. It is clear from (3.3) that $\ln S_q(n, n^s)/\ln n$ should behave approximately linearly in s. It thus makes sense to focus on the slope of the simple linear regression of $\ln S_q(n, n^s)/\ln n$ on s for a fixed value of q. We now establish an analog of Theorem 2 on the empirical scaling function.

Theorem 4. *Suppose that the assumptions of Theorem 3 hold and fix $s_1, \dots, s_N$ in the definition of the empirical scaling function (1.5). Then for every $q > -1$*

$$\hat{\tau}_{N,n}(q) \xrightarrow{P} \tau^{\infty}_{H,\alpha}(q) := \begin{cases} Hq, & \text{if } q \leq \alpha, \\ \left(H - \frac{1}{\alpha}\right) q + 1, & \text{if } q > \alpha, \end{cases} \tag{3.5}$$

as $n \to \infty$.

Proof. Fix $q > -1$ and let $\varepsilon > 0$, $\delta > 0$ and

$$C = \sum_{i=1}^{N} (s_i)^2 - \frac{1}{N}\left(\sum_{i=1}^{N} s_i\right)^2 > 0.$$

By Theorem 3, for each $i = 1, \dots, N$ there exist $n_i^{(1)}$ such that

$$P\left(\left|\frac{\ln S_q(n, n^{s_i})}{\ln n} - R_{H,\alpha}(q, s_i)\right| > \frac{\varepsilon C}{2 s_i N}\right) < \frac{\delta}{2N}, \quad n \geq n_i^{(1)}.$$

It follows then that for $n \geq n_{max}^{(1)} := \max\{n_1^{(1)}, \dots, n_N^{(1)}\}$

$$\begin{aligned} P&\left(\left|\sum_{i=1}^{N} s_i \frac{\ln S_q(n, n^{s_i})}{\ln n} - \sum_{i=1}^{N} s_i R_{H,\alpha}(q, s_i)\right| > \frac{\varepsilon C}{2}\right) \\ &\leq P\left(\sum_{i=1}^{N} s_i \left|\frac{\ln S_q(n, n^{s_i})}{\ln n} - R_{H,\alpha}(q, s_i)\right| > \frac{\varepsilon C}{2}\right) \\ &\leq \sum_{i=1}^{N} P\left(\left|\frac{\ln S_q(n, n^{s_i})}{\ln n} - R_{H,\alpha}(q, s_i)\right| > \frac{\varepsilon C}{2 s_i N}\right) < \frac{\delta}{2}. \end{aligned}$$

Similarly, for each $i = 1, \dots, N$ there exist $n_i^{(2)}$ such that

$$P\left(\left|\frac{\ln S_q(n, n^{s_i})}{\ln n} - R_{H,\alpha}(q, s_i)\right| > \frac{\varepsilon C}{2\left(\sum_{i=1}^{N} s_i\right)}\right) < \frac{\delta}{2N}, \quad n \geq n_i^{(2)},$$

and for $n \geq n^{(2)}_{max} := \max\{n^{(2)}_1, \dots, n^{(2)}_N\}$

$$\begin{aligned} &P\left(\left|\frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} \frac{\ln S_q(n, n^{s_j})}{\ln n} - \frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} R_{H,\alpha}(q, s_j)\right| > \frac{\varepsilon C}{2}\right) \\ &\qquad \leq P\left(\sum_{j=1}^{N} \left|\frac{\ln S_q(n, n^{s_j})}{\ln n} - R_{H,\alpha}(q, s_j)\right| > \frac{N\varepsilon C}{2\left(\sum_{i=1}^{N} s_i\right)}\right) \\ &\qquad \leq \sum_{j=1}^{N} P\left(\left|\frac{\ln S_q(n, n^{s_j})}{\ln n} - R_{H,\alpha}(q, s_j)\right| > \frac{\varepsilon C}{2\left(\sum_{i=1}^{N} s_i\right)}\right) < \frac{\delta}{2}. \end{aligned}$$

Finally then, for $n \geq \max\{n^{(1)}_{max}, n^{(2)}_{max}\}$ it follows that

$$\begin{aligned} &P\left(\left|\hat{\tau}_{N,n}(q) - \frac{\sum_{i=1}^{N} s_i R_{H,\alpha}(q, s_i) - \frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} R_{H,\alpha}(q, s_j)}{\sum_{i=1}^{N} (s_i)^2 - \frac{1}{N}\left(\sum_{i=1}^{N} s_i\right)^2}\right| > \varepsilon\right) \\ &\leq P\Bigg(\left|\sum_{i=1}^{N} s_i \frac{\ln S_q(n, n^{s_i})}{\ln n} - \sum_{i=1}^{N} s_i R_{H,\alpha}(q, s_i)\right| \\ &\qquad + \left|\frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} \frac{\ln S_q(n, n^{s_j})}{\ln n} - \frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} R_{H,\alpha}(q, s_j)\right| > \varepsilon C\Bigg) \\ &\leq P\left(\left|\sum_{i=1}^{N} s_i \frac{\ln S_q(n, n^{s_i})}{\ln n} - \sum_{i=1}^{N} s_i R_{H,\alpha}(q, s_i)\right| > \frac{\varepsilon C}{2}\right) \\ &\qquad + P\left(\left|\frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} \frac{\ln S_q(n, n^{s_j})}{\ln n} - \frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} R_{H,\alpha}(q, s_j)\right| > \frac{\varepsilon C}{2}\right) < \delta, \end{aligned}$$

and thus

$$\hat{\tau}_{N,n}(q) \xrightarrow{P} \frac{\sum_{i=1}^{N} s_i R_{H,\alpha}(q, s_i) - \frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} R_{H,\alpha}(q, s_j)}{\sum_{i=1}^{N} (s_i)^2 - \frac{1}{N}\left(\sum_{i=1}^{N} s_i\right)^2}, \quad \text{as } n \to \infty.$$

It remains to show that the right hand side is exactly $\tau^{\infty}_{H,\alpha}(q)$ from (3.5). Indeed, when $q \leq \alpha$ we have

$$\frac{\sum_{i=1}^{N} qHs_i^2 - \frac{1}{N}\sum_{i=1}^{N} s_i \sum_{j=1}^{N} qHs_j}{\sum_{i=1}^{N} (s_i)^2 - \frac{1}{N}\left(\sum_{i=1}^{N} s_i\right)^2} = Hq$$

and if $q > \alpha$

$$\frac{\sum_{i=1}^N s_i \left(\left(1+qH-\frac{q}{\alpha}\right)s_i + \frac{q}{\alpha} - 1\right) - \frac{1}{N}\sum_{i=1}^N s_i \sum_{j=1}^N \left(\left(1+qH-\frac{q}{\alpha}\right)s_j + \frac{q}{\alpha} - 1\right)}{\sum_{i=1}^N (s_i)^2 - \frac{1}{N}\left(\sum_{i=1}^N s_i\right)^2}$$

$$= \frac{\left(1+qH-\frac{q}{\alpha}\right)\sum_{i=1}^N s_i^2 - \left(1+qH-\frac{q}{\alpha}\right)\frac{1}{N}\left(\sum_{i=1}^N s_i\right)^2}{\sum_{i=1}^N (s_i)^2 - \frac{1}{N}\left(\sum_{i=1}^N s_i\right)^2}$$

$$+ \frac{\left(\frac{q}{\alpha}-1\right)\sum_{i=1}^N s_i - \frac{1}{N}\sum_{i=1}^N s_i \sum_{j=1}^N \left(\frac{q}{\alpha}-1\right)}{\sum_{i=1}^N (s_i)^2 - \frac{1}{N}\left(\sum_{i=1}^N s_i\right)^2}$$

$$= \left(H - \frac{1}{\alpha}\right)q + 1 + \frac{\left(\frac{q}{\alpha}-1\right)\sum_{i=1}^N s_i - \left(\frac{q}{\alpha}-1\right)\sum_{i=1}^N s_i}{\sum_{i=1}^N (s_i)^2 - \frac{1}{N}\left(\sum_{i=1}^N s_i\right)^2} = \left(H - \frac{1}{\alpha}\right)q + 1.$$

□

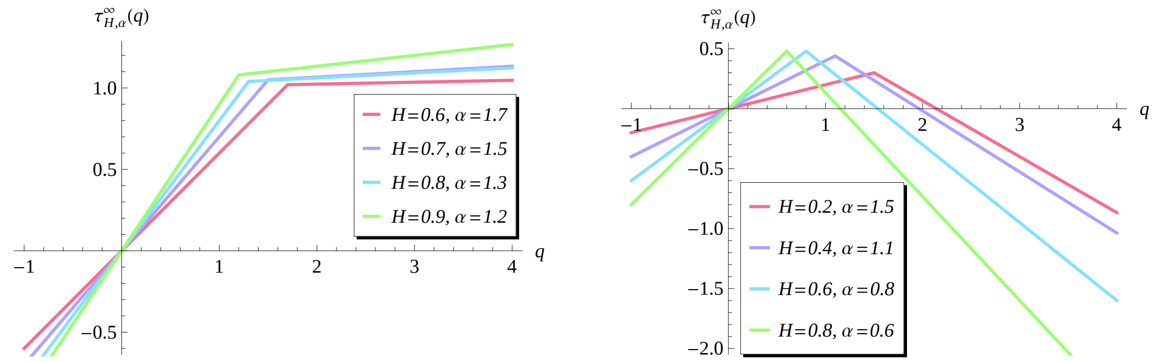
3.3 Applications in parameter estimation

Since LFSM combines both heavy-tails and long-range dependence it provides a rich modeling potential (see e.g. Willinger et al. (1998)). It is therefore important to have methods for estimating the main parameters α and H. Standard estimators of the Hurst exponent H usually assume that the underlying process has finite variance and this makes them inappropriate for the case of LFSM. Also, estimators of the tail index are known to behave well mostly on independent or weakly dependent samples (see e.g. Embrechts et al. (1997)). It is therefore necessary to construct estimators of both parameters that take into account the special structure of the LFSM.

A wavelet based estimator of the parameter H for the LFSM has been proposed in Stoev & Taqqu (2003) (see also Stoev & Taqqu (2005) and Pipiras et al. (2007)). The authors define two estimators, both of which are shown to be strongly consistent and asymptotically normal under some conditions. These estimators do not require the knowledge of α. In Ayache & Hamonier (2012) a wavelet-based estimator of α has been defined. However, this method requires one to know the H value first. For the more general model of linear multifractional stable motion, estimators of α and the Hurst functional parameter $H(\cdot)$ have been developed in Ayache & Hamonier (2013).

In this section we develop methods that are able to simultaneously estimate both parameters α and H. The methods are based on the asymptotic behavior of partition function and the empirical scaling function established in Theorem 4. The asymptotic shape of the empirical scaling function $\tau^{\infty}_{H,\alpha}$ defined in (3.5) is shown in Figure 3.1 for a range of values of H and α. Figure 3.1a shows the long-range dependent case. As indicated in (3.5), the scaling function is bilinear in q with the first part having the slope

H. A break occurs at α and the plot is linear again but now with the slope $H - 1/\alpha$. In the negative dependence case $H < 1/\alpha$ (Figure 3.1b), the second part has a negative slope.



(a) Case $H > \frac{1}{\alpha}$

(b) Case $H < \frac{1}{\alpha}$

Figure 3.1: Asymptotic scaling function $\tau^{\infty}_{H,\alpha}$

Figure 3.2 shows the empirical scaling functions (dashed) for some of the H and α values presented in Figure 3.1. Their computation is based on one sample path realization of length 15784 (explained later). Here, and in every other example of this chapter, s_i, $i = 1, \dots, N$ in (1.4) are chosen equidistantly in the interval $[0.1, 0.9]$ with $N = 23$. The scaling function is estimated at points q_j chosen equidistantly in the interval $[-1, 4]$ with step 0.1. On each plot in Figure 3.2 the corresponding true scaling function is shown by a solid line. Although the break is not sharp, one can notice the bilinear shape.

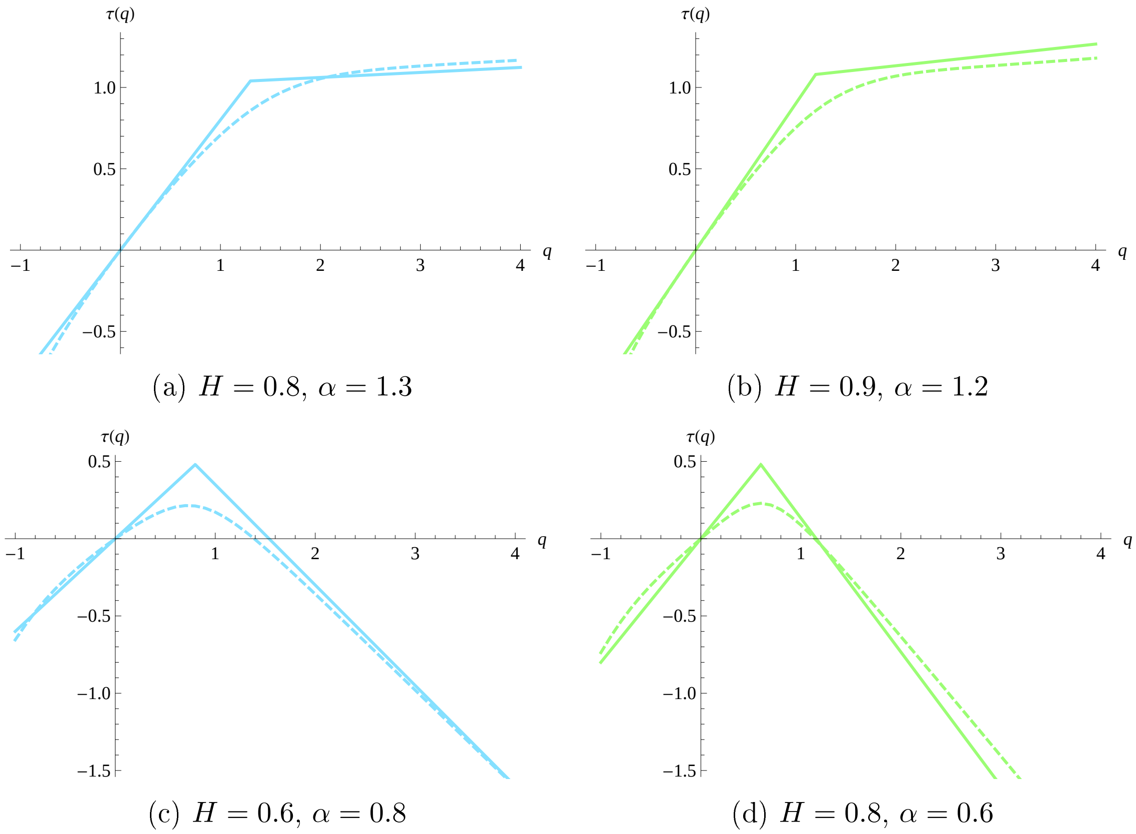


Figure 3.2: Empirical scaling functions (dashed) with the corresponding $\tau^{\infty}_{H,\alpha}$

3.3.1 Estimation methods

We now specify the estimation methods for the parameters H and α, similarly as it was done in Subsection 2.2.4.

Let us first mention that one method can be based on the results of Theorem 3. By choosing points $0 \leq s_1 < \cdots < s_N \leq 1$ and $q_j \in (-1, q_{max})$, $j = 1, \dots, M$, based on the sample of length n, we can calculate

$$\left\{ \frac{\ln S_{q_j}(n, n^{s_i})}{\ln n} \; : \; i = 1, \dots, N, \; j = 1, \dots, M \right\}. \tag{3.6}$$

As $\ln S_{q_j}(n, n^{s_i}) / \ln n$ is expected to behave as $R_{H,\alpha}(q_j, s_i)$ defined in (3.3), we can define the **method MI** estimator for (H, α) by minimizing the difference between the two in the sense of the ordinary least squares, i.e.

$$(\hat{H}_1, \hat{\alpha}_1) = \underset{(H,\alpha) \in (0,1) \times (0,2)}{\arg\min} \sum_{i=1}^{N} \sum_{j=1}^{M} \left(\frac{\ln S_{q_j}(n, n^{s_i})}{\ln n} - R_{H,\alpha}(q_j, s_i) \right)^2 . \tag{3.7}$$

Although this method follows naturally from Theorem 3, it has a disadvantage of being sensitive to the scale parameter of the data similarly as method M3 of the previous chapter. Again, scaling the data by factor c would produce an additional term $\ln |c|^q / \ln n$

in Equation (3.7). If the underling process is standard LFSM, then the scale parameter equals 1 and this would not cause the problem. However, for real data this is usually not the case. For this reason, we would not consider this method in more details, although it will be tested by simulations in the next subsection. To avoid this issue, we specify two other methods which are based only on the slope of $R_{H,\alpha}$ and are not affected by an underlying scale parameter.

Method MII is based on the empirical scaling function (1.4) and Theorem 4. In contrast to MI, we proceed here in two steps. First, based on the data set (3.6) for each q_j, $j = 1, \dots, M$ we compute the empirical scaling function $\hat{\tau}_{N,n}(q_j)$ as defined in (1.4). Since for large samples this converges to $\tau^{\infty}_{H,\alpha}(q_j)$ for each j, we define estimators based on the empirical scaling function as

$$(\hat{H}_2, \hat{\alpha}_2) = \underset{(H,\alpha)\in(0,1)\times(0,2)}{\arg\min} \sum_{j=1}^{M} \left(\hat{\tau}_{N,n}(q_j) - \tau^{\infty}_{H,\alpha}(q_j)\right)^2. \tag{3.8}$$

In simulations and examples below we choose q_j, $j = 1, \dots, M$ equidistantly in the interval $[-1, 4]$ with step 0.2, in order to cover the range of $\alpha \in (0, 2)$ values.

The slope of the first part ($q < \alpha$) of the empirical scaling function corresponds to H, the breakpoint corresponds to α and the slope of the second part ($q > \alpha$) contains information about both parameters H and α. In some examples, as well as in those in Figure 3.2, the slope of the second part does not give the value $H - 1/\alpha$ very precisely, although the first part and the breakpoint behave as expected from (3.5). The second part corresponds to the rate of growth under infinite moments, which makes it a sensitive quantity to measure. Moreover, it depends on both parameters H and α. This can affect the estimation even when there is an obvious bilinear shape. For this reason we provide an alternative estimation method which uses only the information from the first part of the scaling function and the breakpoint, analogues to method M2 of Subsection 2.2.4.

Method MIII fits the following general continuous bilinear function to the empirical scaling function

$$\varsigma(q) = \begin{cases} aq, & \text{if } q \leq b, \\ cq + b(a - c), & \text{if } q > b. \end{cases} \tag{3.9}$$

Here we are interested in two parameters: a which corresponds to H and b which corresponds to α. The estimators by method MIII are now defined as

$$(\hat{H}_3, \hat{\alpha}_3, \hat{c}) = \underset{(a,b,c)\in(0,1)\times(0,2)\times\mathbb{R}}{\arg\min} \sum_{j=1}^{M} \left(\hat{\tau}_{N,n}(q_j) - \varsigma(q_j)\right)^2. \tag{3.10}$$

3.3.2 Simulation study

We use simulation to test the bias and variability of the estimators (3.7), (3.8) and (3.10). We also compare the methods to see which one provides the best results.

In order to simulate paths of LFSM we have used fast Fourier transform (FFT) based algorithm, described in Stoev & Taqqu (2004). All generated sample paths are of length 15784 and additional parameters of the generator are chosen to be $m = 128$ and $M = 600$. This makes $m(M + 15784)$ to be a power of 2 and the algorithm uses FFT (see Stoev & Taqqu (2004) for more details). In all cases we use symmetric α-stable LFSM and the scale parameter of $X(1)$ is set to 1.

Simulations were conducted as follows. We chose for α values $0.3, 0.7, 1, 1.3, 1.7$ and for H values $0.2, 0.4, 0.6, 0.8$, which makes a total of 20 cases. For each case, 100 sample paths of length 15784 have been simulated. For each sample we compute the estimates $(\hat{H}_1, \hat{\alpha}_1)$, $(\hat{H}_2, \hat{\alpha}_2)$ and $(\hat{H}_3, \hat{\alpha}_3)$ corresponding to each of the methods. The mean bias and RMSE of each estimator have been computed for each case. The results are shown in Table 3.1 and Table 3.2. We compare the methods by indicating the better values in bold for each parameter separately.

Table 3.1: Bias of the estimators based on 100 sample paths

H	α	$\hat{H}_1 - H$	$\hat{H}_2 - H$	$\hat{H}_3 - H$	$\hat{\alpha}_1 - \alpha$	$\hat{\alpha}_2 - \alpha$	$\hat{\alpha}_3 - \alpha$
0.2	0.3	**0.1076**	0.1898	0.1888	0.0199	0.0124	**-0.0089**
	0.7	**0.0349**	0.0472	0.0869	0.0325	**0.0317**	-0.0798
	1	0.0250	**-0.0162**	0.0304	**0.0396**	0.0958	-0.1137
	1.3	0.0227	-0.0377	**0.0112**	**0.0590**	0.1854	-0.1465
	1.7	0.0266	0.0048	**0.0014**	0.1640	0.2253	**-0.0950**
0.4	0.3	**-0.0129**	0.0877	0.0876	0.0243	0.0167	**-0.0062**
	0.7	-0.0208	-0.0208	**0.0138**	**0.0456**	0.0486	-0.0632
	1	**-0.0071**	-0.0590	-0.0085	**0.0543**	0.1222	-0.1005
	1.3	0.0071	-0.0533	**-0.0053**	**0.0707**	0.1951	-0.1265
	1.7	0.0176	0.0040	**0.0024**	0.1572	0.2045	**-0.0870**
0.6	0.3	-0.1210	-0.0240	**-0.0227**	0.0295	0.0242	**-0.0013**
	0.7	-0.0650	-0.0910	**-0.0515**	0.0627	0.0793	**-0.0442**
	1	**-0.0234**	-0.0855	-0.0277	**0.0663**	0.1423	-0.0828
	1.3	0.0064	-0.0432	**-0.0033**	**0.0579**	0.1546	-0.1174
	1.7	0.0139	-0.0022	**-0.0016**	0.1395	0.1899	**-0.0921**
0.8	0.3	-0.2230	-0.1159	**-0.1156**	0.0360	0.0305	**0.0030**
	0.7	**-0.1109**	-0.1507	-0.1195	0.0824	0.1097	**-0.0098**
	1	**-0.0309**	-0.0883	-0.0513	0.0622	0.1334	**-0.0360**
	1.3	**0.0072**	-0.0379	-0.0079	0.0276	0.1126	**-0.1016**
	1.7	**-0.0176**	-0.0356	-0.0466	0.1626	0.2261	**-0.0382**

The comparison based on the RMSE from Table 3.2 shows that method MIII provides the best results for both parameters. Table 3.1 indicates that MI and MIII provide smaller bias than MII, although the differences between the estimators are not substantial. Having

Table 3.2: RMSE of the estimators based on 100 sample paths

H	α	$\hat{H}_1$	$\hat{H}_2$	$\hat{H}_3$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
0.2	0.3	0.2733	0.2537	**0.2516**	0.0416	0.0423	**0.0318**
	0.7	0.1301	**0.1155**	0.1333	**0.0866**	0.1124	0.1104
	1	0.0902	0.0899	**0.0838**	**0.1133**	0.2014	0.1513
	1.3	0.0652	0.0833	**0.0644**	**0.1461**	0.2902	0.1903
	1.7	0.0444	**0.0432**	0.0483	0.2238	0.2621	**0.2030**
0.4	0.3	0.2597	0.2000	**0.1975**	0.0454	0.0431	**0.0307**
	0.7	0.1438	0.1148	**0.1073**	0.0952	0.1128	**0.0940**
	1	0.0996	0.1167	**0.0811**	**0.1257**	0.2128	0.1385
	1.3	**0.0691**	0.0970	0.0731	**0.1607**	0.2940	0.1820
	1.7	**0.0418**	0.0432	0.0500	0.2266	0.2537	**0.1815**
0.6	0.3	0.2846	0.1965	**0.1935**	0.0503	0.0478	**0.0320**
	0.7	0.1654	0.1465	**0.1176**	0.1095	0.1258	**0.0828**
	1	0.1091	0.1292	**0.0963**	0.1390	0.2105	**0.1268**
	1.3	0.0731	0.0958	**0.0695**	0.1592	0.2577	**0.1517**
	1.7	**0.0413**	0.0469	0.0531	0.2206	0.2512	**0.1775**
0.8	0.3	0.3357	0.2264	**0.2261**	0.0566	0.0495	**0.0313**
	0.7	0.1881	0.1859	**0.1523**	0.1288	0.1446	**0.0612**
	1	0.1094	0.1350	**0.1003**	0.1415	0.1984	**0.0916**
	1.3	0.0746	0.0939	**0.0699**	0.1461	0.2255	**0.1363**
	1.7	**0.0446**	0.0563	0.0681	0.2305	0.2601	**0.1541**

in mind that MI is not scale invariant, we can definitely recommend the method MIII as the best one. The performance of MIII is also shown in Figure 3.3. Mean estimates based on 100 samples are shown as points in the (H, α) plane and the gridlines show the true value of the parameters. For each value of H the corresponding points and gridline are shown with different colors. We also plot the function $1/H$ to distinguish between long-range dependence ($\alpha > 1/H$) and negative dependence ($\alpha < 1/H$) case.

It can be seen from Figure 3.3 that the value of H does not seem to have a significant influence on the estimation of the tail index α. However, the quality of the estimates of H worsens as α takes smaller values. This can be explained from (3.3) since for small α, the value of H has only a small impact on the shape of $R_{H,\alpha}$. One can see this also from the shape of the scaling function (3.5). The scaling function contains information on H in the slopes of the two parts of the broken line. When α is small, the first linear part is short and the value $1/\alpha$ dominates in the slope of the second part. This makes it hard to estimate H in this case.

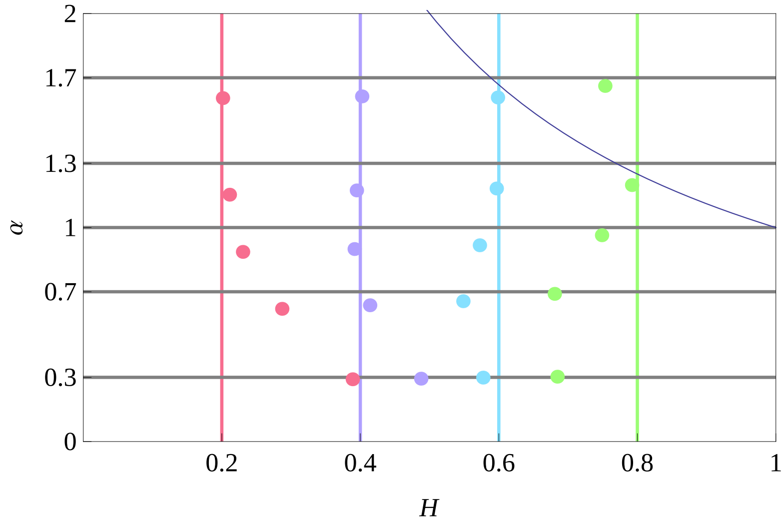
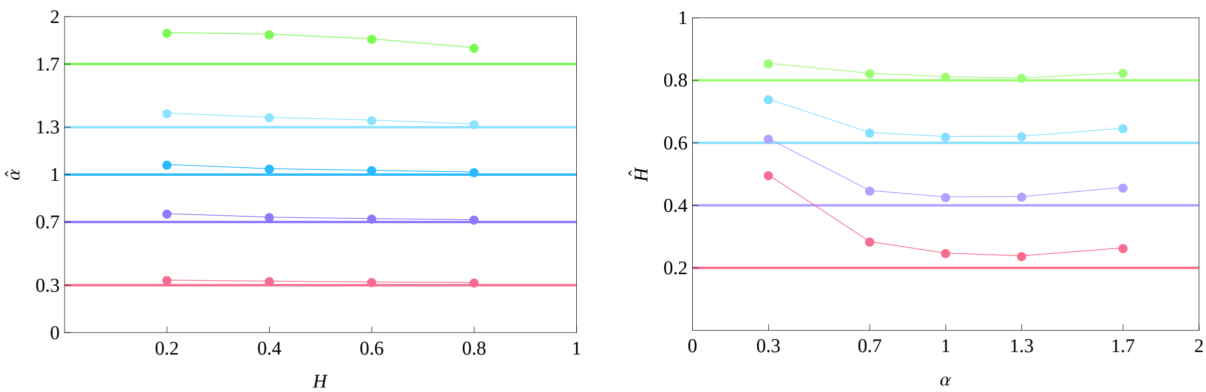


Figure 3.3: Mean estimates of $(\hat{H}_3, \hat{\alpha}_3)$

All three methods can estimate both parameters H and α simultaneously. In view of the results of Chapter 2, we find it interesting to also study how an estimation of the tail index would behave if the dependence structure is stronger and the parameter H measuring dependence is known. So we included in the simulation the behavior of the estimators MII and MIII when H is known. We also did this for the estimators of H assuming α is known. In these cases (3.8) and (3.10) reduce to the minimization of an univariate function.

When one of the parameters is known, both methods MII and MIII behave equally well. Here we present only the mean estimates by method MII (Figure 3.4). The mean estimates of α when H is known are shown in Figure 3.4a. Figure 3.4b shows a similar plot for the estimated value of H assuming α is known. In the case of the estimation of α (Figure 3.4a), one sees that estimators based on the scaling function, like the one proposed in the previous chapter, can perform well even under complicated dependence structure.



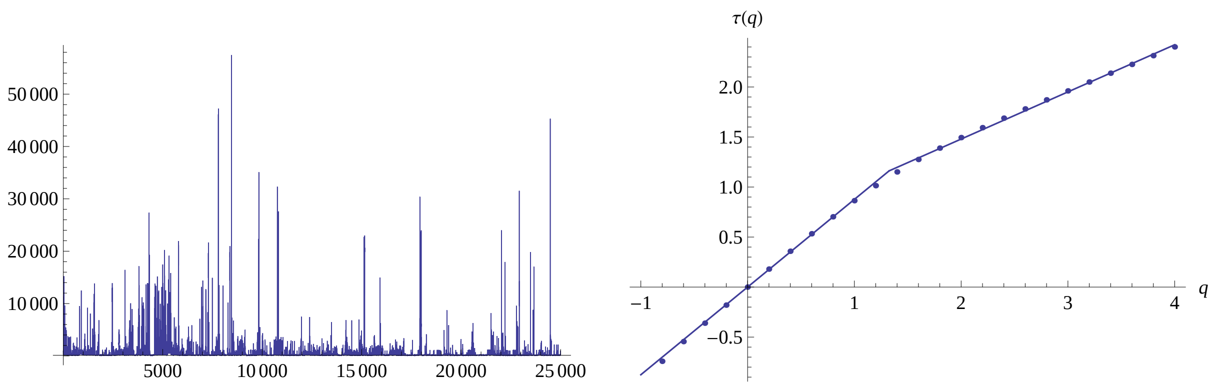
(a) Estimates $\hat{\alpha}_2$, H known (b) Estimates $\hat{H}_2$, α known

Figure 3.4: Mean estimates assuming one parameter is known

3.3.3 Real data applications

Empirical studies show that the network traffic data can exhibit both self-similarity and heavy tails (see e.g. Willinger et al. (1998), Leland et al. (1994)). Many models have been built explaining this behavior. In Karasaridis & Hatzinakos (2001), the authors propose to model network traffic as a linear transformation of the totally skewed linear fractional stable noise. Here we take one network traffic data set and assuming the data is a realization of this model, we estimate the self-similarity and tail parameters.

The data we analyze is the Ethernet trace recorded at the Bellcore Morristown Research and Engineering facility (BC-Oct89Ext) (see Leland & Wilson (1991) and Leland et al. (1994) for more details). It contains packet arrival times (in seconds) and the number of packets (in bytes). The original data has been modified by counting the packets in the blocks of 1 second. We express the time series as the number of packets per time unit and take only the first 25000 values, which is around 20% of all data (Figure 3.5a). The sample mean has been subtracted according to the model and the empirical scaling function (dotted) is shown in Figure 3.5b with the fitted bilinear function (3.9) (solid). The shape indeed resembles the one characteristic for the LFSM and estimation with MIII yields values $\hat{H} = 0.88$ and $\hat{\alpha} = 1.33$. The same data set has been analyzed in Karasaridis & Hatzinakos (2001). The authors report the estimated value 0.8 for the Hurst parameter and 1.63 for the tail index, which is in accordance with our analysis.



(a) Data

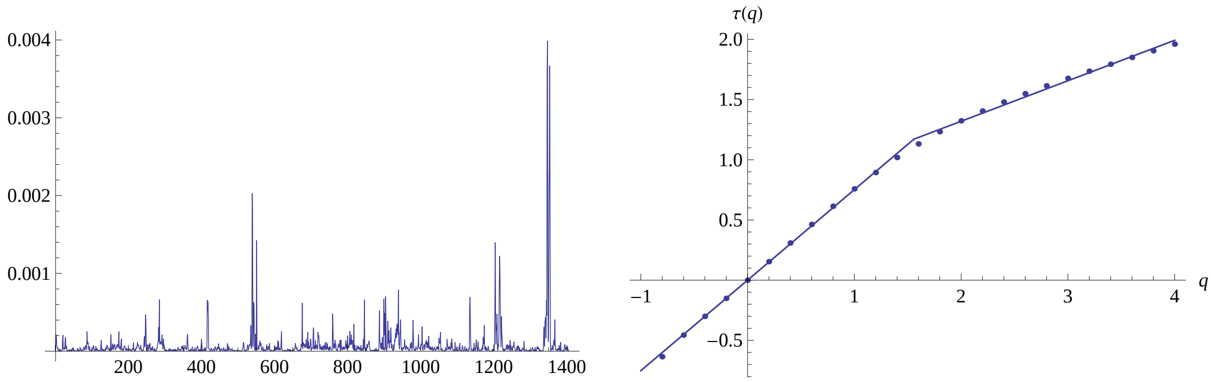
(b) Empirical and fitted scaling function

Figure 3.5: BC-Oct89Ext trace

For the second illustration we analyze the solar flare X-ray data observed by GOES satellite[1]. This type of data is considered to exhibit both self-similarity and heavy tails and claimed to be modeled well with the LFSM (see Weron et al. (2005) and Stanislavsky et al. (2009)). Assuming the data is indeed a realization of the mean shifted linear fractional stable noise, we estimate the parameters H and α. The data contains the information about the time of appearance and energy of the solar flares. We take the data in the

[1] Data is publicly available at http://www.ngdc.noaa.gov/stp/solar/solarflares.html

period from August, 1999 to December, 2003, aggregate the maximum flux values on a daily basis and set the mean to 0, which provides 1405 data points. Figure 3.6a shows the plot of the data and Figure 3.6b the empirical scaling function with the fitted bilinear function (3.9). The estimated values of the parameters are $\hat{H} = 0.75$ and $\hat{\alpha} = 1.56$.



(a) Data (b) Empirical and fitted scaling function

Figure 3.6: Solar flare X-ray data

Chapter 4

Detecting multifractality of time series

In this chapter we introduce the notion of a multifractal stochastic process. There is no unique definition, so we provide an overview of different properties usually referred to as multifractality. The importance of such processes is still a subject of debate, mainly because there is no reliable statistical method that would confirm the occurrence of multifractal properties in empirical time series. We make a contribution to this problem in the second part of this chapter, where we analyze the implications of the results of Chapters 2 and 3 on one of the detection methods.

4.1 Multifractal stochastic processes

The starting point of the multifractal theory can be traced back to the work of Mandelbrot in the context of turbulence modeling, continuing the earlier work of Kolmogorov, Yaglom and Obukhov. In his seminal papers Mandelbrot (1972) and Mandelbrot (1974), Mandelbrot introduced multiplicative cascades in the setting of measures, but also allowing the randomness in the construction. The term "multifractal" was coined to refer to cascades and is attributed to Frisch and Parisi (Frisch & Parisi (1985)). Cascades have many interesting properties which became a model for characterizing more general objects. Geometrical properties of the sets of irregularities of cascades have led to extending the notion of multifractality to functions and motivated studying fine scale properties of functions (see e.g. Jaffard (1996)). In this setting, multifractal analysis deals with the local irregularities of functions characterized by the Hausdorff dimension of sets of points having the same pointwise Hölder exponent. For multifractal functions these sets are interwoven and have noninteger Hausdorff dimension. This notion can be generalized to stochastic processes by simply applying the definition for a function on the sample paths.

On the other hand, scaling properties of cascades can be seen as a generalization of the self-similarity. This led to a consideration of a new class of stochastic processes, called multifractal, characterized by global scaling properties that extend self-similarity in different manners. However, this can lead to discrepancy. For example, strictly α-stable

Lévy processes with $0 < \alpha < 2$ are known to be self-similar with Hurst parameter $1/\alpha$. On the other hand, the sample paths of these processes exhibit multifractal features (see Jaffard (1999)).

The example of cascade suggests that local irregularities are closely related with global scaling properties. This is precisely described with the so-called multifractal formalism. There has been an extensive research questioning when this formalism holds in different settings (see Jaffard (1997*a*,*b*, 2000), Riedi (1995, 2003)). Numerically, fine irregularities are unreachable without such property. For this reason, many different definitions have been introduced in order to achieve the validity of the formalism.

In this section we provide an overview of different scaling relations for stochastic processes that are usually referred to as multifractality. Examples are provided illustrating each of these properties.

4.1.1 Definitions of multifractality

The best known scaling relation in the theory of stochastic processes is self-similarity. A stochastic process $\{X(t), t \geq 0\}$ is said to be self-similar if for any $a > 0$, there exists $b > 0$ such that

$$\{X(at)\} \stackrel{d}{=} \{bX(t)\}, \tag{4.1}$$

where $\{\cdot\} \stackrel{d}{=} \{\cdot\}$ stands for the equality of finite dimensional distributions. A process $\{X(t), t \geq 0\}$ is said to be stochastically continuous at 0 if for every $\varepsilon > 0$, $P(|X(h) - X(0)| > \varepsilon) \to 0$ as $h \to 0$. If $\{X(t), t \geq 0\}$ is self-similar, nontrivial and stochastically continuous at 0, then b in (4.1) must be of the form a^H for some $H \geq 0$, i.e.

$$\{X(at)\} \stackrel{d}{=} \{a^H X(t)\}. \tag{4.2}$$

The proof of this fact can be found in Embrechts & Maejima (2002). These weak assumptions are assumed to hold for every self-similar process considered latter on. The exponent H is called the Hurst parameter and we say $\{X(t), t \geq 0\}$ is H-ss or H-sssi if it also has stationary increments.

Following Mandelbrot et al. (1997), the definition of a multifractal that we present first is motivated by generalizing the scaling rule of self-similar processes in the following manner:

Definition 1. A stochastic process $\{X(t)\}$ is said to be multifractal if

$$\{X(ct)\} \stackrel{d}{=} \{M(c)X(t)\}, \tag{4.3}$$

where for every $c > 0$, $M(c)$ is a random variable independent of $\{X(t)\}$ whose distribution does not depend on t.

When $M(c)$ is nonrandom for every $c > 0$, the process is self-similar and $M(c) = c^H$ if the process is nontrivial and stochastically continuous at 0. The scaling factor $M(c)$ is assumed to satisfy the following property:

$$M(ab) \stackrel{d}{=} M_1(a)M_2(b), \tag{4.4}$$

for every choice of a and b, where M_1 and M_2 are independent copies of M. This generalizes the property of the nonrandom factor for H-ss processes $(ab)^H = a^H b^H$. A motivation for this property can be found in Mandelbrot et al. (1997).

However, instead of Definition 1, scaling is usually specified in terms of moments. The idea of extracting the scaling properties from average type quantities, like L^p norm, dates back to the work of Frisch and Parisi (Frisch & Parisi (1985)).

Definition 2. A stochastic process $\{X(t)\}$ is said to be multifractal if there exist functions $c(q)$ and $\tau(q)$ such that

$$E|X(t) - X(s)|^q = c(q)|t - s|^{\tau(q)}, \quad \forall t, s \in \mathcal{T},\ \forall q \in \mathcal{Q}, \tag{4.5}$$

where $\mathcal{T}$ and $\mathcal{Q}$ are intervals on the real line with positive length and $0 \in \mathcal{T}$.

The function $\tau(q)$ is called the scaling function. Set $\mathcal{Q}$ can also include negative reals. The definition can also be based on the moments of the process instead of the moments of increments, i.e. $E|X(t)| = c(q)t^{\tau(q)}$. If the increments are stationary, these definitions coincide. It is clear that if $\{X(t)\}$ is H-sssi, then $\tau(q) = Hq$ where it is defined.

The following argument shows that τ must be concave. Let $q_1, q_2 \in \mathcal{Q}$, $w_1, w_2 > 0$, $w_1 + w_2 = 1$ and $q = w_1 q_1 + w_2 q_2$. From Hölder's inequality it follows

$$E|X(t) - X(s)|^q \leq \left(E|X(t) - X(s)|^{q_1}\right)^{w_1} \left(E|X(t) - X(s)|^{q_2}\right)^{w_2}.$$

Taking logarithms gives

$$\ln c(q) + \tau(q) \ln|t - s| \leq (w_1\tau(q_1) + w_2\tau(q_2)) \ln|t - s| + (w_1 \ln c(q_1) + w_2 \ln c(q_2)).$$

Dividing by $\ln|t-s| < 0$ and letting $t \to s$ gives $\tau(q) \geq w_1\tau(q_1) + w_2\tau(q_2)$, so τ is concave.

If $\mathcal{T} = [0, \infty)$, then by letting $t \to \infty$ we would get the opposite inequality $\tau(q) \leq w_1\tau(q_1) + w_2\tau(q_2)$, showing τ is linear. Therefore, strict concavity can hold only over a finite time horizon, otherwise $\tau(q)$ would be linear. This is not considered to be a problem for practical purposes (see Mandelbrot et al. (1997) for details). Since the scaling function is linear for self-similar processes, every departure from linearity can be attributed to multifractality. However, for this reasoning to make sense, one must assume moment scaling to hold as otherwise self-similarity and multifractality are not complementary notions.

The drawback of involving moments in the definition is that they can be infinite. This narrows the applicability of the definition and as we show later, can produce practical problems.

It is easy to see that under stationary increments the defining property (4.3), along with the property (4.4), implies multifractality Definition 2. Indeed, (4.4) implies that $E|M(c)|^q$ must be of the form $c^{\tau(q)}$ and the claim follows from $X(t) =^d M(t)X(1)$. One has to assume finiteness of the moments involved in order for the statements like (4.5) to have sense. Also notice that both definitions imply $X(0) = 0$ a.s., which will be used later.

There exist many variations of Definition 2. Some processes obey the definition only for a small range of values t or for asymptotically small t. The stationarity of increments can also be imposed. When referring to multifractality we will make clear which definition we mean. However, we exclude self-similar processes from the preceding definitions.

The prominent examples of multifractals are the so-called multiplicative cascades. They were first introduced as measures to model turbulence and the basic idea goes back to Richardson in 1920s (see Lovejoy & Schertzer (2013) for the historical account). A motivating example is the binomial measure, which can be constructed on the interval $[0, 1]$ as follows. Suppose $p \in (0, 1)$ and $q = 1-p$. Take μ_0 to be the Lebesgue measure on $[0, 1]$. Measure μ_1 is defined by assigning mass p to interval $[0, 1/2]$ and mass q to interval $[1/2, 1]$. Repeating this process, at the step n we arrive at measure μ_n such that for the dyadic interval $I_{a_1 a_2 \cdots a_n} := \left[\sum_{i=1}^n \frac{a_i}{2^i}, \sum_{i=1}^n \frac{a_i}{2^i} + \frac{1}{2^n}\right)$, $a_i \in \{0, 1\}$, $i = 1, \dots, n$, it holds

$$\mu_n(I_{a_1 a_2 \cdots a_n}) = p^{n - \sum_{i=1}^n a_i} q^{\sum_{i=1}^n a_i}.$$

One can then show that the sequence $(\mu_n, n \in \mathbb{N})$ converges weakly to a finite measure μ defined on the Borel σ-algebra on $[0, 1]$ (Kahane & Peyriere (1976)). Several extensions of this construction can be made. First, instead of splitting the intervals in half, one can take some $b > 2$ and spread masses $p_1, \dots, p_b$ that add up to one. Secondly, the mass allocated to each interval can be made random and this leads to a discrete multiplicative cascade. More precisely, suppose $\{W_{a_1 a_2 \cdots a_n} : a_i \in \{0, 1, \dots, b-1\}, n \in \mathbb{N}\}$ is a family of independent copies of the non-negative random variable W such that $EW = 1$. The previous construction gives at the step n a random measure μ_n such that for the b-adic interval $I_{a_1 a_2 \cdots a_n} := \left[\sum_{i=1}^n \frac{a_i}{b^i}, \sum_{i=1}^n \frac{a_i}{b^i} + \frac{1}{b^n}\right)$, $a_i \in \{0, \dots, b-1\}$, $i = 1, \dots, n$, it holds

$$\mu_n(I_{a_1 a_2 \cdots a_n}) = b^{-n} W_{a_1} W_{a_1 a_2} \cdots W_{a_1 a_2 \cdots a_n}.$$

The factor b^{-n} ensures conserving the mass on average, $E\left[\mu_n([0, 1])\right] = 1$. We can write the preceding equation in the form

$$\mu_{n+1}(I_{a_1a_2\cdots a_{n+1}}) = b^{-1}W_{a_1a_2\cdots a_{n+1}}\mu_n(I_{a_1a_2\cdots a_n}),$$

which can be seen as a discrete form of the multifractal property (4.3) with $t = b^{-n}$ and $c = b^{-1}$. Moreover, we have that, assuming the moments involved are finite,

$$E\left[(\mu_n(I_{a_1a_2\cdots a_n}))^q\right] = b^{-nq}\left(EW^q\right)^n = \left(b^{-n}\right)^{-\log_b EW^q + q},$$

which is a version of (4.5) for $[s,t] = I_{a_1a_2\cdots a_n}$ with $\tau(q) = -\log_b EW^q + q$. It can be shown that, a.s., the sequence $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to a random measure μ. Conditions for the nondegeneracy of μ and existence of moments can be found in Kahane & Peyriere (1976). We can now define a stochastic process $X(t) = \mu([0,t])$, $t \in [0,1]$, which we will refer to as the discrete multiplicative cascade. The resulting measure has another interesting property: if we denote $Y_n = \mu_n([0,1])$, $n \in \mathbb{N}$, then

$$\begin{aligned} Y_{n+1} &= b^{-(n+1)} \sum_{a_1a_2\cdots a_{n+1}\,:\,a_i\in\{0,\dots,b-1\}} W_{a_1}W_{a_1a_2}\cdots W_{a_1a_2\cdots a_{n+1}} \\ &= b^{-1}\sum_{j=0}^{b-1} W_j \left(b^{-n} \sum_{a_2\cdots a_{n+1}\,:\,a_i\in\{0,\dots,b-1\}} W_{ja_2}\cdots W_{ja_2\cdots a_{n+1}} \right). \end{aligned}$$

This means that Y_{n+1} satisfies

$$Y_{n+1} \overset{d}{=} b^{-1}\sum_{j=0}^{b-1} W_j Y_n(j),$$

where $Y_n(j)$, $j = 0,\dots,b-1$ are independent, distributed as Y_n and independent of $W_0,\dots,W_{b-1}$. Taking the limit as $n \to \infty$ we get that for $Y_\infty = \mu([0,1])$

$$Y_\infty \overset{d}{=} b^{-1}\sum_{j=0}^{b-1} W_j Y_\infty(j),$$

where $Y_\infty(j)$, $j = 0,\dots,b$ are independent copies of Y_∞ independent of i.i.d. W_j, $j = 0,\dots,b$. Thus, Y_∞ is a solution of the following problem: given nonnegative i.i.d. $A_0,\dots,A_{b-1}$, a non-negative random variable Z is the *fixed point of the smoothing transform* if

$$Z \overset{d}{=} \sum_{j=0}^{b-1} A_j Z_j,$$

where $Z_0,\dots,Z_{b-1}$ are independent copies of Z, independent of $A_0,\dots,A_{b-1}$. Describing the solutions of such equations has attracted a lot of attention (see Alsmeyer et al. (2012)

and references therein). If A_j are nonrandom, then it is well known that the only solutions are strictly stable distributed random variables.

Although cascades are considered as a model example of multifractals, they satisfy multifractal properties only on a discrete grid of time points. Moreover, the discrete cascade process $\{X(t)\}$ does not have stationary increments. Several constructions have been proposed to obtain continuous scaling properties and stationary increments starting with Barral & Mandelbrot (2002) and followed by Muzy & Bacry (2002), Bacry & Muzy (2003), Chainais et al. (2005) and more recently, Barral & Jin (2014). Of all these, mostly equivalent constructions, we will use only the log-normal cascade process which is derived from a class of log-infinitely divisible multifractal random measures proposed in Bacry & Muzy (2003). These measures can be constructed as follows. Let $\phi(q)$, $q \in \mathbb{R}$ be the logarithm of the characteristic function of some infinitely divisible distribution. Suppose $\mathcal{P}$ is an independently scattered infinitely divisible random measure on $S^+ = \{(t,l) : t \in \mathbb{R}, l \geq 0\}$ with intensity measure $\mu(dt,dl) = l^{-2}dtdl$. This means that for every sequence of disjoint Borel sets $(A_n) \subset S^+$, $(\mathcal{P}(A_n))$ are independent random variables, $\mathcal{P}\left(\bigcup_n A_n\right) = \sum_n \mathcal{P}(A_n)$ a.s. and for every Borel set $A \subset S^+$

$$Ee^{iq\mathcal{P}(A)} = e^{\phi(q)\mu(A)}.$$

Given $T > 0$, define

$$f(k) = \begin{cases} k, & k \leq T, \\ T, & k > T. \end{cases}$$

For $t \in \mathbb{R}$ and $l > 0$ define sets (cones)

$$A_l(t) = \{(s,k) : k \geq l, \ -f(k)/2 < s - t < f(k)/2\},$$

and consider process $w_l(t) = \mathcal{P}(A_l(t))$. Now for $l > 0$ and Lebesgue measurable set I, we define measure $M_l(I) = \int_I e^{w_l(t)}dt$. Under certain conditions (see Bacry & Muzy (2003)), a.s., M_l converges weakly to a random measure M, as $l \to 0$. This limiting measure is called log-infinitely divisible cascade measure. If the starting infinitely divisible distribution is normal, then we arrive at the log-normal cascade measure that in the discrete construction corresponds to a cascade constructed with multipliers W such that $\ln W$ has normal distribution. In this case, we will refer to a process $\{\theta(t), t \in [0,T]\}$, $\theta(t) = M([0,t])$ as the log-normal cascade (LNC). LNC with *intermittency parameter* $\lambda^2 \in (0, 1/2)$ has stationary increments and the property that for every $0 < c < 1$

$$\{\theta(ct)\} \stackrel{d}{=} \{ce^{2\Gamma_c}\theta(t)\}, \tag{4.6}$$

where Γ_c is normally distributed random variable such that $E\Gamma_c = -Var(\Gamma_c) = \lambda^2 \ln c$, independent of $\{\theta(t)\}$. For every $t \in [0,T]$ and for the range of finite moments, moment scaling of the form (4.5) holds with

$$\tau_{LNC}(q) = q(1+2\lambda^2) - 2\lambda^2 q^2. \tag{4.7}$$

Cascades are heavy-tailed, and for the log-normal cascade the range of finite moments is $(-\infty, 1/(2\lambda^2))$ by applying (Bacry & Muzy 2003, Lemma 3) and by (Bacry et al. 2013, Proposition 5).

Modeling abilities of cascades are restricted as they are nondecreasing and can take only positive values. Several models have been proposed to address the need for a more general multifractal processes. Two approaches are the most common. The first is based on the time change of some self-similar process, i.e. $X(t) = Y(\theta(t))$, where $\{\theta(t)\}$ is some type of continuous cascade process independent of $\{Y(t)\}$. A typical choice for $\{Y(t)\}$ is FBM and resulting process is called (fractional) Brownian motion in multifractal time or multifractal (fractional) random walk (MRW). The second construction is based on the stochastic integration of some continuous cascade process $\{\theta(t)\}$ with respect to some self-similar process, i.e. $X(t) = \int_0^t \theta(s)dY(s)$. Generally, two approaches lead to different processes (see Bacry et al. (2001), Muzy & Bacry (2002), Ludeña (2008) and Abry et al. (2009) for more details). However, if $\{Y(t)\}$ is BM ($H = 1/2$) and $\{\theta(t)\}$ log-infinitely divisible cascade process, then the process $\{Y(\theta(t))\}$ has the same finite dimensional distributions as the process

$$X(t) = \lim_{l \to 0} \int_0^t e^{w_l(s)/2} dY(s),$$

where $\{w_l(s)\}$ is defined in the construction of the cascade measure. If $\{\theta(t)\}$ is the log-normal cascade process, we will refer to a process $\{X(t) = Y(\theta(t)), t \in [0,T]\}$ as the log-normal multifractal random walk (LNMRW). Multifractal properties of $\{X(t)\}$ are inherited from those of $\{\theta(t)\}$ and for every $0 < c < 1$

$$\{X(ct)\} \stackrel{d}{=} \{c^{\frac{1}{2}} e^{\Gamma_c} X(t)\}, \tag{4.8}$$

with Γ_c as in (4.6). Since $E|X(t)|^q = E|\theta(t)|^{q/2} E|Y(1)|^q$, the scaling function is given by

$$\tau_{LNMRW}(q) = q\left(\frac{1}{2} + \lambda^2\right) - \frac{\lambda^2}{2} q^2, \tag{4.9}$$

when q-th moment is finite, i.e. when $q \in (-1, 1/\lambda^2)$. Moments of negative order $q \leq -1$ are infinite as this is a property of the normally distributed $Y(1)$.

Continuous cascade processes and multifractal random walks are the main examples of multifractals with properties (4.3) and (4.5). In what follows when we say multifractal process we mean a process of this kind. More advanced models can also be built by generalizing the previous construction, see e.g. Anh et al. (2008, 2009*a*,*b*, 2010).

4.1.2 Spectrum of singularities

Previous definitions involve "global" properties of the process. Alternatively, one can base the definition on the "local" scaling properties, such as roughness of the process sample paths measured by the pointwise Hölder exponents. There are different approaches on how to develop the notion of a multifractal function. First, we say that a function $f : [0, \infty) \to \mathbb{R}$ is $C^{\gamma}(t_0)$ if there exists a constant $C > 0$ such that for all t in some neighborhood of t_0

$$|f(t) - f(t_0)| \leq C|t - t_0|^{\gamma}. \tag{4.10}$$

Another common definition in the literature is to define that f is Hölder continuous of order γ at point t_0 if $|f(t) - P_{t_0}(t)| \leq C|t - t_0|^{\gamma}$ for some polynomial P_{t_0} of degree at most $\lfloor \gamma \rfloor$. If the Taylor polynomial of this degree exists, then P_{t_0} is that Taylor polynomial. Thus, if P_{t_0} is constant, then $P_{t_0} \equiv f(t_0)$ and two definitions coincide. This happens, in particular when $\gamma < 1$. For other conditions of equivalence and more details see Riedi (2003). In what follows we will use the first definition as in many cases we consider only processes whose sample paths are $C^{\gamma}(t_0)$ with $\gamma < 1$ at any point t_0.

It is clear that if $f \in C^{\gamma}(t_0)$, then $f \in C^{\gamma'}(t_0)$ for each $\gamma' < \gamma$. A pointwise Hölder exponent of the function f at t_0 is

$$H(t_0) = \sup \left\{ \gamma : f \in C^{\gamma}(t_0) \right\}. \tag{4.11}$$

Consider sets $S_h = \{t : H(t) = h\}$ consisting of the points in the domain where f has the Hölder exponent of value h. If we consider the Lebesgue measure of the sets S_h for varying h, then usually only one of them would have a full Lebesgue measure while all the others would have Lebesgue measure zero. To properly measure the size of the sets S_h, we introduce the Hausdorff dimension, which is based on the concept of Hausdorff measure. Denote by $|U| = \sup\{\|x - y\| : x, y \in U\}$ the diameter of the set $U \subset \mathbb{R}^n$, where $\| \cdot \|$ is the Euclidean norm. For $F \subset \mathbb{R}^n$, $F \neq \emptyset$ and $s \geq 0$, let for every $\delta > 0$

$$H^s_\delta(F) = \inf \left\{ \sum_{i=1}^{\infty} |U_i|^s : F \subset \bigcup_{i=1}^{\infty} U_i, |U_i| \leq \delta \right\},$$

which increases as $\delta \to 0$. The s-dimensional Hausdorff measure of $F \subset \mathbb{R}^n$ is defined as

$$H^s(F) = \lim_{\delta \to 0} H^s_\delta(F).$$

One can show that this is indeed a measure on the Borel sets on $\mathbb{R}^n$ and for $s = n$ it is a multiple of the Lebesgue measure. Moreover, it has the scaling property $H^s(\lambda F) =$

$\lambda^s H^s(F)$. Next, notice that for $|U_i| \leq \delta$ and $t > s$, $\sum_i |U_i|^t \leq \sum_i |U_i|^{t-s}|U_i|^s \leq \delta^{t-s} \sum_i |U_i|^s$ so that $H^t_\delta(F) \leq \delta^{t-s} H^s_\delta(F)$. Letting $\delta \to 0$ we see that if $H^s(F) < \infty$, then $H^t(F) = 0$ for $t > s$ and there must be a critical value of s where $H^s(F)$ changes from ∞ to 0. This is the Hausdorff dimension of $F \subset \mathbb{R}^n$, denoted by $\dim_H F$. More precisely,

$$\dim_H F = \inf \{s \geq 0 : H^s(F) = 0\} = \sup \{s \geq 0 : H^s(F) = \infty\}$$

with the convention that $\sup \emptyset = 0$. If $F = \emptyset$, then we define $\dim_H F = -\infty$. For more details on the Hausdorff dimension see Falconer (2003).

In the interesting examples, sets S_h are fractal in the sense that they have noninteger Hausdorff dimension. This means that the local regularity of a function, measured by the pointwise Hölder exponents, changes erratically from point to point. The mapping $h \mapsto d(h) = \dim_H S_h$ is called the spectrum of singularities (also multifractal or Hausdorff spectrum) or simply the spectrum. We will refer to a set of h such that $d(h) \neq -\infty$ as the support of the spectrum. A function f is said to be multifractal if the support of its spectrum is nontrivial, in the sense that it is not a one point set. This is naturally extended to stochastic processes:

Definition 3. A stochastic process $\{X(t)\}$ on some probability space $(\Omega, \mathcal{F}, P)$ is said to have multifractal paths if for (almost) every $\omega \in \Omega$, $t \mapsto X(t, \omega)$ is a multifractal function.

When considered for a stochastic process, Hölder exponents are random variables and S_h random sets. However, in many cases the spectrum is deterministic, that is, for almost every sample path spectrum is equally defined. Moreover, spectrum is usually homogeneous, in the sense it is the same when considered over any nonempty subset $A \subset [0, \infty)$. All the examples considered in the following will have these two properties. An example of a process with random, nonhomogeneous spectrum can be found in Barral et al. (2010).

To perform a multifractal analysis usually means to determine the spectrum of a process or a function. This has been done so far for many examples. For instance, some famous classical functions are shown to have a nontrivial spectrum of singularities (Jaffard (1996)). FBM with Hurst parameter H is known to have a trivial spectrum consisting of only one point, that is a.s.

$$d_{FBM}(h) = \begin{cases} 1, & \text{if } \ h = H \\ -\infty, & \text{otherwise.} \end{cases} \tag{4.12}$$

This property is sometimes referred to as monofractality. In the early days, this example led many to think that all self-similar processes have a trivial spectrum and that a more general scaling property is needed to obtain a nontrivial spectrum. Strictly stable Lévy processes give a counterexample to this thought, as they are $1/\alpha$-sssi and their spectrum is a.s. (Jaffard (1999))

$$d_{SLP}(h) = \begin{cases} \alpha h, & \text{if } \; h \in [0, 1/\alpha], \\ -\infty, & \text{if } \; h > 1/\alpha. \end{cases} \tag{4.13}$$

More generally, many other Lévy processes have multifractal paths. This was established in Jaffard (1999) and extended in Balança (2014) under weaker assumptions. Denote by β the Blumenthal-Getoor (BG) index of a Lévy process

$$\beta = \inf \left\{ \gamma \geq 0 : \int_{|x| \leq 1} |x|^{\gamma} \pi(dx) < \infty \right\},$$

where π is the corresponding Lévy measure. We assume the drift is zero if $\beta < 1$. If σ is the Brownian component of the characteristic triplet, define

$$\beta' = \begin{cases} \beta, & \text{if } \; \sigma = 0, \\ 2, & \text{if } \; \sigma \neq 0. \end{cases}$$

Spectrum of singularities of almost every path of the Lévy process is given by

$$d_{LP}(h) = \begin{cases} \beta h, & \text{if } \; h \in [0, 1/\beta'), \\ 1, & \text{if } \; h = 1/\beta', \\ -\infty, & \text{if } \; h > 1/\beta'. \end{cases} \tag{4.14}$$

Another example is provided by the LFSM studied in Chapter 3. Local regularity of the LFSM has been established recently in Balança (2014). If $\alpha \in [1, 2)$, $H \in (0, 1)$ and $H > 1/\alpha$, then a.s.

$$d_{LFSM}(h) = \begin{cases} \alpha(h - H) + 1, & \text{if } \; h \in [H - \frac{1}{\alpha}, H], \\ -\infty, & \text{otherwise.} \end{cases} \tag{4.15}$$

It is known that in the case $H < 1/\alpha$ sample paths are nowhere bounded, which explains the assumptions.

In all these examples the spectrum is linear. The cascade processes provide a more general shape. For example, LNC process has a parabolic spectrum a.s. given by (Bacry et al. (2008))

$$d_{LNC}(h) = \begin{cases} 1 - \frac{(h-1-2\lambda^2)^2}{8\lambda^2}, & \text{if } \; h \in \left[1 + 2\lambda^2 - 2\sqrt{2\lambda^2}, 1 + 2\lambda^2 + 2\sqrt{2\lambda^2}\right], \\ -\infty, & \text{otherwise.} \end{cases} \tag{4.16}$$

In the case of LNMRW we have a.s.

$$d_{LNMRW}(h) = \begin{cases} 1 - \frac{(h-1/2-\lambda^2)^2}{2\lambda^2}, & \text{if } \; h \in \left[1/2 + \lambda^2 - \sqrt{2\lambda^2}, 1/2 + \lambda^2 + \sqrt{2\lambda^2}\right], \\ -\infty, & \text{otherwise.} \end{cases} \tag{4.17}$$

See Barral & Seuret (2007) and Bacry et al. (2008) for more details.

4.1.3 Multifractal formalism

Multifractal formalism relates local and global scaling properties by connecting singularity spectrum with the scaling function via the Legendre transform:

$$d(h) = \inf_q \left(hq - \tau(q) + 1\right). \tag{4.18}$$

Since the Legendre transform is concave, the spectrum is always a concave function, provided the multifractal formalism holds. If the multifractal formalism holds, then $\inf_q (hq - \tau(q) + 1) = -\infty$ implies that $S_h = \emptyset$ so that h is not the Hölder exponent at any point. In addition, formalism gives the possibility of estimating the spectrum as the Legendre transform of the estimated scaling function.

The following loose heuristics may motivate the formalism for functions. In this case, scaling function may be defined by

$$\int |f(s+t) - f(s)|^q ds \sim |t|^{\tau(q)}.$$

If the Hölder exponent in s is h, then in a small interval of length $|t|$ around s we have $|f(s+t) - f(s)| \sim |t|^{qh}$. Since the dimension of such singularities if $d(h)$, there is around $|t|^{-d(h)}$ such intervals of length $|t|$, so that they contribute to the integral with $|t|^{qh-d(h)+1}$. The largest contribution as $t \to 0$ is given by the smallest exponent, thus

$$\int |f(s+t) - f(s)|^q ds \sim |t|^{\inf_h (qh - d(h) + 1)}.$$

The preceding arguments actually illustrate that τ is the Legendre transform of d, but for concave functions double Legendre transform returns the original function. This and other facts about the Legendre transform can be found in (Riedi et al. 1999, Appendix A).

Substantial work has been done to investigate when this formalism holds, in the beginning for measures and later for functions in the work of Jaffard. A somewhat surprising result was given in Jaffard (2000), showing that the formalism holds for a dense G_δ subset (countable intersection of open sets) of a certain naturally chosen function space. Multifractal formalism ensures that the spectrum can be estimated from computable global quantities and is therefore a desirable property of the object considered. For this reason many authors seek for different definitions of global and local scaling properties that would always be related by a certain type of multifractal formalism. To this end, local scaling properties can also be based on replacing the Hausdorff dimension with the box-counting dimension or using the so-called coarse Hölder exponents for defining local regularity. Coarse Hölder exponents can be defined as

$$H_k^{(n)} = \frac{\log_2 \sup \left\{ |X(t) - X(s)| : \frac{k}{2^n} \leq s \leq t \leq \frac{k+1}{2^n} \right\}}{\log_2 2^{-n}}.$$

In this setting, we fix one path of the process $\{X(t)\}$ and measure the spectrum discretely. Let

$$N^{(n)}(h, \varepsilon) = \sum_{k=0}^{2^n-1} 1_{\{|H_k^{(n)} - h| < \varepsilon\}}$$

denote the number of coarse Hölder exponents in the interval $[h-\varepsilon, h+\varepsilon]$. In this sense, we can think of $N^{(n)}(h,\varepsilon)/2^n$ as the probability to select k from the set $\{0, 1, \dots, 2^n - 1\}$ such that $H_k^{(n)} \in [h-\varepsilon, h+\varepsilon]$. This roughly estimates the probability that h is a Hölder exponent. Typically, by the argument based on the law of large numbers, this probability will be more and more concentrated at some point as $n \to \infty$. This is the most probable Hölder exponent, while the others are derived by the large deviations principle, which measures a probability of observing deviant spectrum values. The Legendre transform comes here into play as a part of the large deviations principle. Notice that the path is fixed and the randomness is considered through the choice of k. Such type of statements are sometimes referred to as the weak multifractal formalism. More details on such definitions and related formalism can be found in Riedi (2003). Another path to the multifractal formalism is investigated by changing the global property involved. For this purpose, scaling function can be specified using wavelets, wavelet leaders or oscillations (see Jaffard et al. (2014) and references therein).

Multifractal formalism and other approaches will be discussed in more details in Chapter 5. Even though Legendre transform of the scaling function may not represent the spectrum of singularities, in many applications it is estimated based on the estimated scaling function and used for the classification and model selection. We will use this principle until the end of this chapter and estimate the spectrum without any reference to path properties.

Another important issue is which range of q should one use for taking infimum in (4.18). Since τ may not be defined for all q it is reasonable to take only the range of finite moments. As the example of the LNC shows, it is important to consider also the moments of negative order to get the full spectrum. Indeed, it is easy to check using (4.7) and (4.16) that

$$d_{LNC}(h) = \inf_{q\in(-\infty,1/(2\lambda^2))} \left(hq - \tau_{LNC}(q) + 1\right),$$

thus the multifractal formalism holds for the LNC. The same is not true for LNMRW, as the moments of order $q \leq -1$ are infinite. However, if the infimum is taken only over positive q values, we can get the left (increasing) part of the parabolically shaped spectrum (4.17):

$$\inf_{q\in(0,1/\lambda^2)} \left(hq - \tau_{LNMRW}(q) + 1\right) = \begin{cases} 1 - \frac{(h-1/2-\lambda^2)^2}{2\lambda^2}, & \text{if } \ h \in \left[1/2+\lambda^2-\sqrt{2\lambda^2}, 1/2+\lambda^2\right], \\ 1, & \text{if } \ h > 1/2+\lambda^2, \\ -\infty, & \text{otherwise.} \end{cases}$$

We defer any further discussion on this topic until Chapter 5 where we will study the influence of infinite moments on the path properties. Through this chapter spectrum is used only as a graphical method and all the computation is performed by taking infimum over $q \in (0, \infty)$.

4.2 Statistical analysis of multifractal processes

The first goal of the statistical analysis is to detect multifractal properties of a given time series data. Definition 2, which is a direct consequence of Definition 1 if (4.4) is assumed, provides a simple criterion for detecting multifractal stochastic processes. Before using Definition 2, one must determine that the moment scaling of the form (4.5) holds. If this is true, then the method can be based on exploiting the fact that the scaling function is linear for self-similar processes where it is defined. Every departure from linearity can therefore be accredited to multifractality. So, the main problem is to check if the moment scaling holds and then estimate the scaling function from the data and inspect its shape. This provides a connection with the work done in Chapters 2 and 3. One can already notice from the earlier results that the effects may be blurred with infinite moments. If the data is suspected to be a realization of some multifractal process, then some of the multifractal models described in Subsection 4.1.1 can be fitted based on the shape of the estimated scaling function or spectrum. We now present the methodology provided in Fisher et al. (1997) for detecting multifractality of a time series data.

Suppose $\{X(t), t \in [0,T]\}$ is a stationary increments stochastic process with $X(0) = 0$. If $\{X(t)\}$ is multifractal in the sense of Definition 2, then we have from (4.5) that

$$\ln E|X(t)|^q = \tau(q)\ln t + \ln c(q). \tag{4.19}$$

Thus, to check if the moment scaling holds, one should check if, for each q, $\ln E|X(t)|^q$ is linear in $\ln t$ for varying t. As we already remarked in Chapter 1, a natural estimator for $E|X(t)|^q$ if $\{X(t)\}$ is sampled on the interval $[0,T]$, is the partition function

$$S_q(T,t) = \frac{1}{\lfloor T/t \rfloor} \sum_{i=1}^{\lfloor T/t \rfloor} |X(it) - X((i-1)t)|^q \,.$$

If we denote Y_i, $i = 1,\dots,T$ to be the one step increments $Y_i = X(i) - X(i-1)$, we get the usual form (1.2). Since $ES_q(T,t) = E|X(t)|^q$, for multifractal processes we have

$$\ln ES_q(T,t) = \tau(q)\ln t + \ln c(q). \tag{4.20}$$

It thus makes sense to consider the slope in the simple linear regression of $\ln S_q(T,t)$ on $\ln t$. This is exactly the empirical scaling function $\hat{\tau}_{N,T}(q)$ as defined in (1.4). Alternatively, taking $\log_T$ instead of ln, we would get

$$\log_T ES_q(T,t) = \tau(q)\log_T t + \log_T c(q), \tag{4.21}$$

and consider the slope of the linear regression of $\log_T S_q(T,t)$ on $\log_T t$. This is the empirical scaling function in the form (1.5), which is equivalent to (1.4) by changing n to T and s to $\log_T t$.

Multifractal behavior is inspected through the use of Equation (4.19). Based on the data sample, X_i, $i = 1,\dots T$, the following methodology is presented in Fisher et al. (1997):

(1) For fixed value of q, one computes the logarithm of the partition function for a range of values t and plots it against $\ln t$. If the scaling exists, the plot should be approximately a linear line.

(2) Following Equation (4.20) the slope of the line can be estimated by a linear regression of $\ln ES_q(T,t)$ on $\ln t$. The value obtained provides an estimate $\hat{\tau}_{N,T}(q)$ for the scaling function $\tau(q)$ at point q.

(3) Repeating this for a range of q values, one is able to plot the empirical scaling function. If the plot is nonlinear then one can suspect the existence of the multifractal scaling.

(4) After an estimate $\hat{\tau}_{N,T}$ of the scaling function is obtained, it is possible to calculate the spectrum using Equation (4.18) with τ replaced by $\hat{\tau}_{N,T}$.

In Fisher et al. (1997) the method is applied to the DEM/USD exchange rate as well as to other financial data. The examples suggest a linear relation of the form (4.19) holds and the scaling function exhibits nonlinear behavior. We next explain that these effects can also be contributed to the presence of heavy tails.

Remark 4. Although the definition (1.4) follows naturally from the moment scaling relation (4.5), it is not the only one appearing in the literature. A common approach is to estimate the scaling function by using only the smallest time scale available. For example, for the cascade process on the interval $[0, T]$ the smallest interval is usually of the length $2^{-j}T$ for some j. One can then estimate the scaling function at point q as

$$\frac{\log_2 S_q(T, 2^{-j}T)}{-j}. \tag{4.22}$$

The estimator (1.4) estimates the scaling function across different time scales and is therefore more general than (4.22).

Fixed domain asymptotic properties of the estimator (4.22) for the discrete multiplicative cascade have been established in Ossiander & Waymire (2000), where it was shown under some assumptions that when $j \to \infty$ estimator (4.22) tends a.s. to

$$\tau_C^\infty(q) = \begin{cases} h_0^- q, & \text{if } q \leq q_0^-, \\ \tau(q), & \text{if } q_0^- < q < q_0^+ \\ h_0^+ q, & \text{if } q \geq q_0^+, \end{cases}$$

where

$$\begin{aligned} q_0^+ &= \inf\{q \geq 1 : q\tau'(q) - \tau(q) + 1 \leq 0\}, \\ q_0^- &= \sup\{q \leq 0 : q\tau'(q) - \tau(q) + 1 \leq 0\}, \end{aligned} \tag{4.23}$$

and $h_0^+ = \tau'(q_0^+)$, $h_0^- = \tau'(q_0^-)$. For the LNC we have

$$q_0^+ = 1/\sqrt{2\lambda^2}, \tag{4.24}$$

$$q_0^- = -1/\sqrt{2\lambda^2}. \tag{4.25}$$

So the estimator (4.22) is consistent for a certain range of q, while outside this interval the so-called linearization effect happens, which is still not entirely understood. See Bacry et al. (2010) for a discussion, where similar results have been established in a mixed asymptotic framework.

4.3 Detecting multifractality under heavy tails

In this section we apply the results of Chapters 2 and 3 on the problem of detecting multifractal properties of time series. It is clear that these results show that the scaling function can be estimated as nonlinear due to heavy tails and that this effect can be mistakenly regarded as multifractality. We make this point more clear in the rest of the chapter. Although multifractal models have attracted a lot of attention among practitioners, the evidence of multifractal properties has been questioned in many references. In Barunik et al. (2012), it has been reported by simulations that these properties may have been confused with heavy-tails. The first proper treatment in this direction was given in Heyde & Sly (2008) and Heyde (2009). Our goal here is to provide a full analysis of the problem and all the repercussions it has.

First, we define a class of stochastic processes which fall into our consideration. We will call these processes to be of type $\mathfrak{L}$.

Definition 4. A stochastic process $\{X(t), t \geq 0\}$ is said to be of type $\mathfrak{L}$, if $Y_i = X(i) - X(i-1)$, $i \in \mathbb{N}$ is a strictly stationary sequence having heavy-tailed marginal distribution with index α, positive extremal index, satisfying strong mixing property with an exponentially decaying rate and such that $EY_i = 0$ when $\alpha > 1$.

This class includes many examples like all Lévy processes with $X(1)$ heavy-tailed, for example, strictly α-stable Lévy processes with $0 < \alpha < 2$. A richer modeling ability is provided by the Student Lévy process, which allows for arbitrary tail index parameter. Since the Student t-distribution defined in (2.16) is infinitely divisible, a Lévy process such that $X(1) \stackrel{d}{=} T(\nu, \sigma, \mu)$ surely exists (see Heyde & Leonenko (2005) for details). We assume $\mu = 0$ if $\nu > 1$.

The class $\mathfrak{L}$ also includes cumulative sums of stationary processes like Ornstein-Uhlenbeck (OU) type processes or diffusions with heavy-tailed marginal distributions. Recall from Subsection 2.2.3 that OU type or diffusion process $\{Y(t), t \geq 0\}$ is a strictly stationary process having strong mixing property with an exponentially decaying rate. Heavy-tailed examples are Student OU type process and Student diffusion. For more examples of heavy-tailed diffusions see Avram et al. (2013). If $\{Y(t), t \geq 0\}$ is a strictly stationary OU type process or diffusion with heavy-tailed marginals having the strong mixing property with an exponentially decaying rate, then the process

$$X(t) = \sum_{i=0}^{\lfloor t \rfloor} Y(i), \quad t \geq 0$$

will be of type $\mathfrak{L}$. This provides a variety of examples with dependent increments.

For what follows, we will assume that $X_1, \dots, X_T$ is a sample observed at discrete equally spaced time instants from a stochastic process $\{X(t), t \geq 0\}$. Moreover, we assume that the process is sampled at time instants $1, 2, \dots T$. Notice that it is not a restriction, as if $\{X(t)\}$ is multifractal by Definition 2 and is sampled at regularly spaced time instants $\delta, 2\delta, \dots, n\delta$ and $\delta \neq 1$, then we know the values of the process $\{\tilde{X}_t\} = \{X_{\delta t}\}$ at times $1, 2, \dots n$. But then $\tau_{\tilde{X}}(q) = \tau_X(q)$, since

$$E|\tilde{X}(t)|^q = E|X(\delta t)|^q = \left(c(q)\delta^{\tau(q)}\right) t^{\tau(q)}.$$

We can therefore assume that the process is sampled at time instants $1, 2, \dots T$.

We first discuss the implications of Theorem 1. If $\{X(t)\}$ is of type $\mathfrak{L}$, this theorem establishes the limit of $\ln S_q(T, T^s)/\ln T$ when $T \to \infty$. Repeating the discussion preceding Theorem 2, we denote

$$\varepsilon_T = \frac{S_q(T,t)}{T^{R_\alpha(q, \log_T t)}},$$

taking the logarithm and rewriting yields

$$\log_T S_q(T,t) = R_\alpha(q, \log_T t) + \log_T \varepsilon_T. \tag{4.26}$$

In the step (1) of the methodology presented in the previous section, we should check that $\ln S_q(T,t)$ is linear in $\ln t$, or equivalently that $\log_T S_q(T,t)$ is linear in $\log_T t$. By the same argument as in Subsection 2.1.2, we can view (4.26) as the regression model with errors $\log_T \varepsilon_T$. Moreover, when $\alpha \leq 2$ or, $\alpha > 2$ and $q \leq \alpha$, $R_\alpha(q, \log_T t)$ is linear in $\log_T t$, i.e. we can rewrite (4.26) as

$$\log_T S_q(T,t) = a(q) \log_T t + b(q) + \log_T \varepsilon_T, \tag{4.27}$$

for some functions $a(q)$ and $b(q)$. It follows that the relation of type (4.21) holds up to some random variable and the data sampled from processes of type $\mathfrak{L}$ will behave as if they obey the moment scaling relation (4.19). Thus, for processes of type $\mathfrak{L}$ step (1) of the methodology from Section 4.2 will always be satisfied. If $\alpha > 2$ and $q > \alpha$, $R_\alpha(q, \log_T t)$ is not linear in $\log_T t$. It is actually bilinear with the breakpoint depending on the values of q and α. However, if q is not much greater than α, $s \mapsto R_\alpha(q,s)$ is very close to a linear function. The relation of type (4.27) would again hold approximately.

As follows from the preceding discussion, for processes of type $\mathfrak{L}$, it makes sense to consider the slope of the linear regression of $\ln S_q(T,t)$ on $\ln t$. This is step (2) of the methodology given in Section 4.2 that leads to the empirical scaling function defined by (1.4). Theorem 2 derives the asymptotic form of the scaling function. The additional assumption in the case $q > \alpha > 2$ forces t_i to be of the form $T^{\frac{i}{N}}$. This ensures that, as the sample size grows ($T \to \infty$), the number of points included in the regression based on (4.19) grows.

To conclude, if the underlying process has stationary, heavy-tailed, zero mean, weakly dependent increments, then the scaling function estimated from the data will behave approximately as τ_α^∞ defined in (2.4). The shape of the scaling function is determined by the value of the tail index α, as shown in Figure 2.1. Thus, under the assumptions considered, the difference between linear and nonlinear estimated scaling function can be described by the presence of heavy tails.

This means that heavy-tailed processes exhibit features that can be confused with multifractality. It is therefore dangerous to conclude multifractality by examining the estimated scaling function under heavy tails. Moreover this effect is so regular that one can make inference about the unknown tail index of the underlying distribution by analyzing the scaling function behavior, as it was done in Chapter 2. We illustrate these points by examples given in Section 4.4, but before that we analyze how the spectrum would be estimated for processes of type $\mathfrak{L}$.

4.3.1 Estimation of the spectrum

As mentioned before, using Equation (4.18) the spectrum can be estimated as the Legendre transform of the estimated scaling function, i.e.

$$\hat{d}_{N,T}(h) = \inf_{q} \left(hq - \hat{\tau}_{N,T}(q) + 1 \right). \tag{4.28}$$

At this point, spectrum is not used with any reference to path properties, but only as an exploratory tool for the model choice. In this chapter we will consider (4.28) by taking infimum only over $q > 0$. This means considering only moments of positive order. More detailed discussion on the range of moments used will be provided in Chapter 5.

One way to assess the spectrum numerically is to interpolate $\hat{\tau}_{N,T}$ based on some estimated points and then proceed with numerical minimization. Since $\hat{\tau}_{N,T}(q)$ can be estimated at any point q, the interpolation can be made arbitrary precise. In the graphical presentation we choose not to plot the values of the minimum achieved for $q = 0$ as these trivially give the value 1.

The alternative approach is based on the following geometrical meaning of the Legendre transform. Consider $d(h) = \inf_q (hq - \tau(q))$ and suppose τ is concave. Given q_0 we can find the tangent at q_0 on τ, call it $s(q) = aq + b$, such that $\tau(q) \leq s(q)$ with equality at q_0. If τ is differentiable, this tangent will be unique. Then $aq - \tau(q) \geq aq - s(q) = -b$ with equality at q_0 and so

$$d(a) = aq_0 - \tau(q_0) = -b.$$

If we suppose τ is differentiable at q_0, then s is unique and

$$d(\tau'(q_0)) = q_0\tau'(q_0) - \tau(q_0).$$

One can show that d is concave (see e.g. (Riedi et al. 1999, Appendix A)). Thus, the line $l_{q_0}(h) = q_0 h - \tau(q_0)$ is a tangent of d at point a. This gives an idea of how to estimate the spectrum graphically, as simply plotting l_{q_0} for a range of q_0 values will yield an envelope for d. We will use this method in one of the examples in the next section.

Having in mind the asymptotic behavior of the empirical scaling function, the estimated spectrum for processes of type $\mathcal{L}$ is expected to behave as

$$d_\infty(h) = \inf_{q>0} \left(hq - \tau_\alpha^\infty(q) + 1 \right). \tag{4.29}$$

We consider the cases $\alpha \leq 2$ ($\underline{d}_\infty$) and $\alpha > 2$ ($\overline{d}_\infty$) separately.

When $\alpha \leq 2$, we can explicitly calculate

$$\underline{d}_\infty(h) = \min\left\{ \inf_{0<q\leq\alpha} \left(hq - \frac{q}{\alpha} + 1\right),\ \inf_{q\geq\alpha} hq \right\} = \begin{cases} \alpha h, & \text{if } 0 \leq h \leq \frac{1}{\alpha} \\ 1, & \text{if } h > \frac{1}{\alpha}. \end{cases}$$

If the infimum is taken over all q, the part of the spectrum for $h > 1/\alpha$ would depend on $\tau_\alpha^\infty(q)$ for negative q. This part corresponds to the right (decreasing) part of the spectrum and requires negative order moments to be estimated.

When $\alpha > 2$ we have

$$\overline{d}_\infty(h) = \min\left\{ \inf_{0<q\leq\alpha} \left(hq - \frac{q}{2} + 1\right),\ \inf_{q\geq\alpha} \left(hq - \frac{q}{2} - \frac{2(\alpha-q)^2(2\alpha+4q-3\alpha q)}{\alpha^3(2-q)^2}\right) \right\}.$$

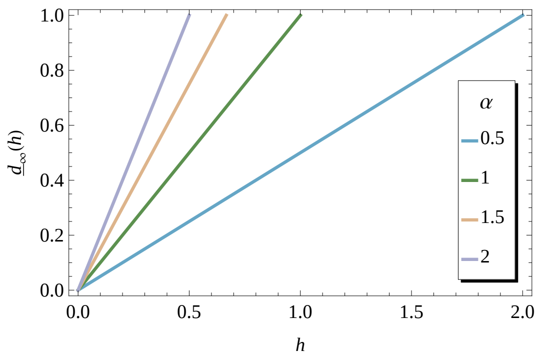
Values $h > 1/2$ yield the right part of the spectrum and $\overline{d}_\infty(h) = 1$. On the other hand if

$$h < \lim_{q\to\infty} \frac{\tau_\alpha^\infty(q)}{q} = \frac{(\alpha-2)^2(\alpha+4)}{2\alpha^3},$$

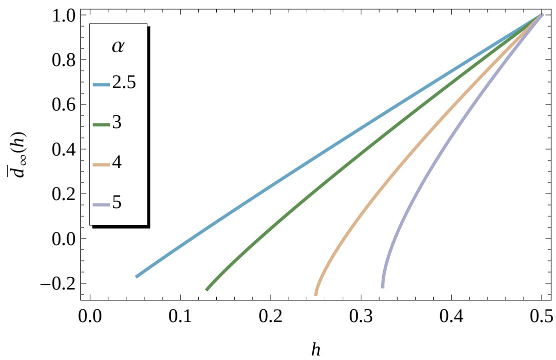
then $\overline{d}_\infty(h) = -\infty$ is attained when $q \to \infty$. Thus, the left part of the spectrum is finite for

$$h \in \left[\frac{(\alpha-2)^2(\alpha+4)}{2\alpha^3}, \frac{1}{2}\right].$$

On this interval the spectrum is nonlinear and approximately parabolic but the explicit formula is complicated. Figure 4.1 shows the shape of the spectrum one would expect when estimation is done using scaling function. We conclude that processes of type $\mathfrak{L}$ can yield a nontrivial estimated spectrum in the presence of heavy-tails.



(a) Case $\alpha \leq 2$



(b) Case $\alpha > 2$

Figure 4.1: Spectrum estimated from τ_α^∞

4.4 Simulations and examples

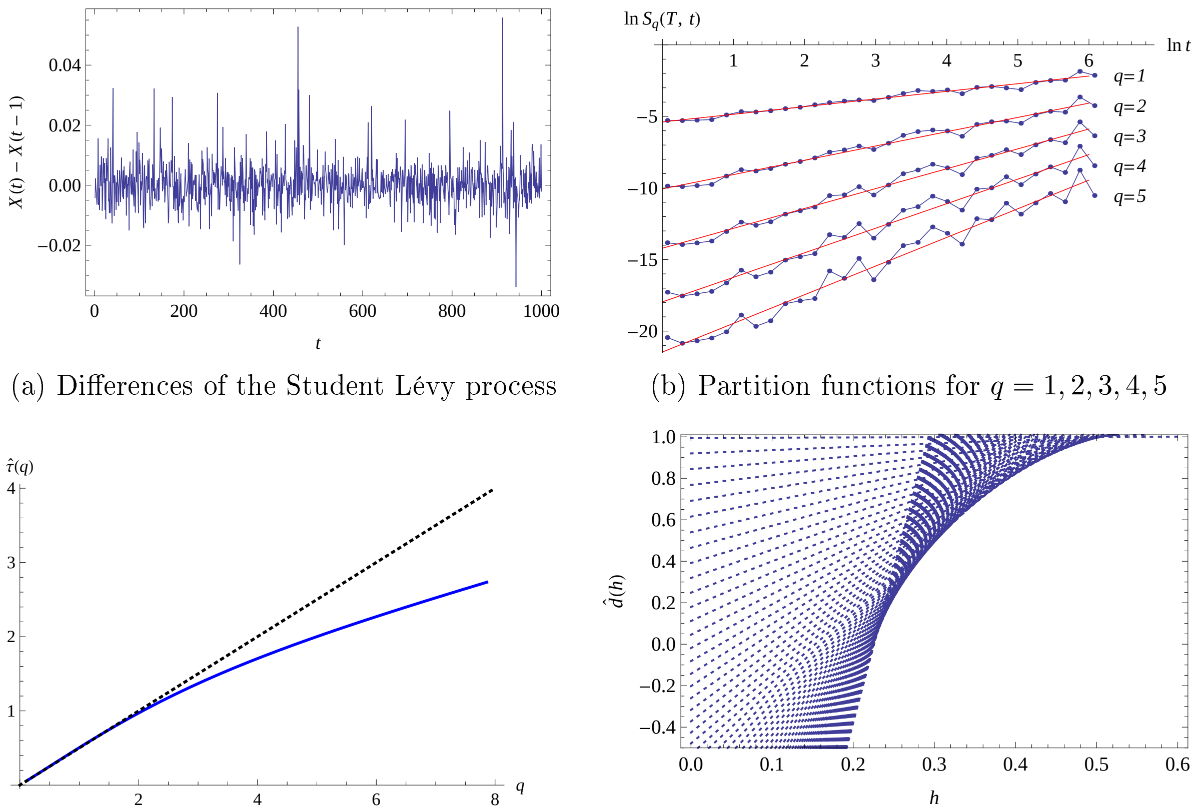
In this section we provide examples showing that nonlinearity of the estimated scaling functions can be reconstructed just by using a process with heavy-tailed increments. For this purpose we set $\{X(t)\}$ to be a Student Lévy process, i.e. a stochastic process with stationary independent increments such that $X(0) = 0$ and $X(1)$ has Student's t-distribution.

It is important to stress that we do not advocate using Student Lévy process as a model in any of the examples. Besides, independence of increments is an unrealistic property for financial data. Our goal is simply to show that nonlinear scaling functions can be reproduced using heavy-tailed models. However, our results also cover some weakly dependent processes. In Heyde & Leonenko (2005) and Leonenko et al. (2011) the authors provide examples of Student processes with different dependence structures that could be more appropriate for financial data.

4.4.1 Example 1

In Calvet & Fisher (2002) (see also Fisher et al. (1997) and Calvet & Fisher (2008)), the authors provide an example with DM/USD exchange rate data with the plot of the estimated scaling function (Calvet & Fisher 2002, Figure 6). The concavity is ascribed to the multifractal property of the data. Considering the discussion of the preceding section and comparing the plot with Figure 2.1 one can conclude that the data exhibit heavy-tailed characteristics and a rough estimate of the tail index is around 4. This is consistent with other studies suggesting risky asset returns are usually heavy-tailed with tail index between 3 and 5 (see Hurst & Platen (1997) and Heyde & Liu (2001)).

We try to reproduce the same figures as in Calvet & Fisher (2002) by simulating the data taking $\{X(t)\}$ to be the Student Lévy process. Figure 4.2a shows the one step increments of a sample path of length 1000 of a Student Lévy process with $X(1) \overset{d}{=} T(4, 0.005, 0)$. A linear behavior of the partition function in the sense of relation (4.19) is confirmed by Figure 4.2b, which shows the plots of $\ln t$ against $\ln S_q(T,t)$ for five different values of q. Adjusted R^2 values for the linear fit are approximately 0.97 for $q = 1, \dots, 5$, which confirms the linear relation. Similar analysis was done in (Calvet & Fisher 2002, Figure 5) and we want to stress the similarity of the two plots. Figure 4.2c shows the estimated scaling function together with the baseline. Concavity is a consequence of heavy tails and one can notice the resemblance with Figure 6 in Calvet & Fisher (2002). Finally, we estimate the spectrum by plotting tangents forming an envelope of the spectrum. The shape of the spectrum is almost identical to the one presented in Figure 7 in Calvet & Fisher (2002).



(a) Differences of the Student Lévy process

(b) Partition functions for $q = 1, 2, 3, 4, 5$

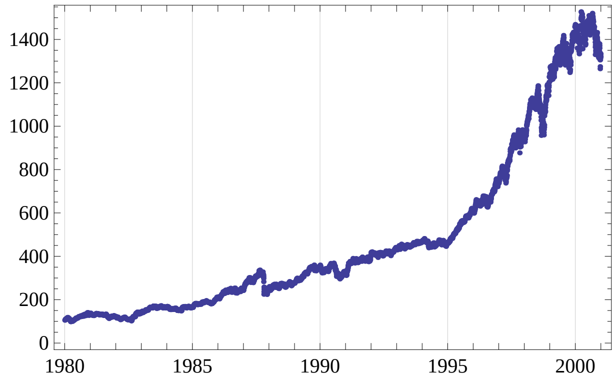
(c) Estimated scaling function

(d) Estimated spectrum

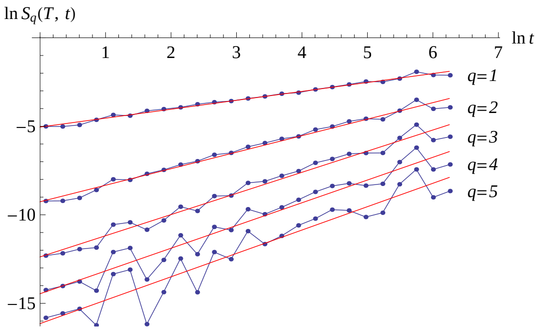
Figure 4.2: Student Lévy process

4.4.2 Example 2

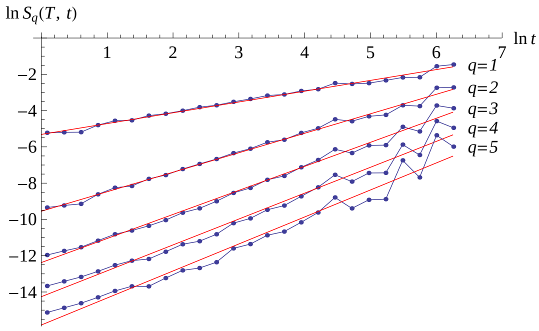
We provide another example based on the financial data. The data consist of 5307 daily closing values of the S&P500 stock market index collected in the period from January 1, 1980 until December 31, 2000 (Figure 4.3a). For the analysis we consider log-differences of this series and subtract the mean. Figure 4.3b confirms linear relation of the partition function and time in the log-log scale. For estimating the scaling function at every point q the time points chosen are $1, 2, 3, 4, 5, 7, 15, 30, 60, 90, 180$. The scaling function estimated from the data is shown in Figure 4.3d, together with the baseline and plot of τ_α^∞ for $\alpha = 2.5$ (dashed). One can see the resemblance which indicates the data may be heavy-tailed with tail index around 2.5. We additionally generate a sample path of the same length for a Student Lévy process with $X(1) \stackrel{d}{=} T(2.5, 0.0046, 0)$, where the second parameter is estimated from the data by the method of moments. Partition functions again scale linearly (Figure 4.3c) and Figure 4.3e shows the estimated scaling function for the generated process. The similarity of the two scaling functions is also naturally reflected in the estimated multifractal spectrum shown in Figures 4.3f and 4.3g.



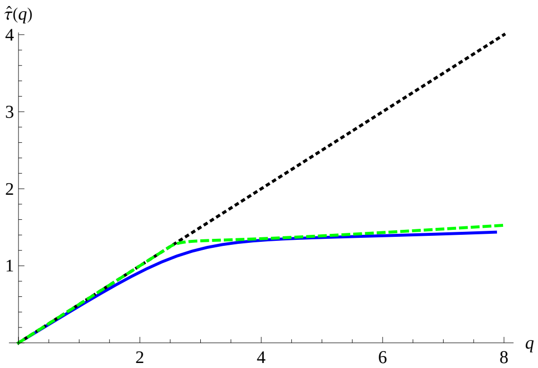
(a) S&P 500 index time series



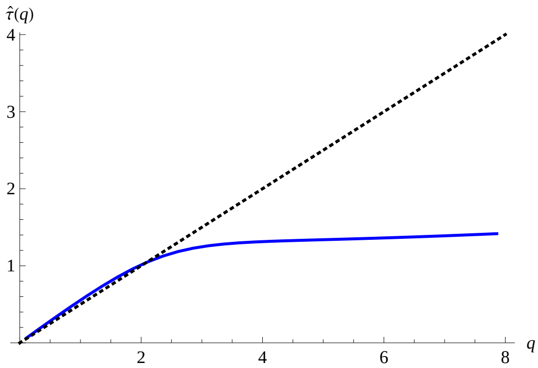
(b) Partition functions for the S&P 500 index



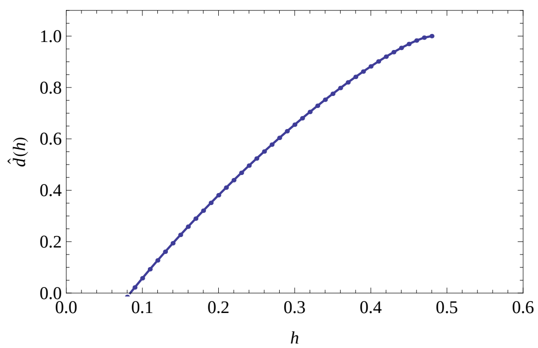
(c) Partition functions for the generated Student Lévy process



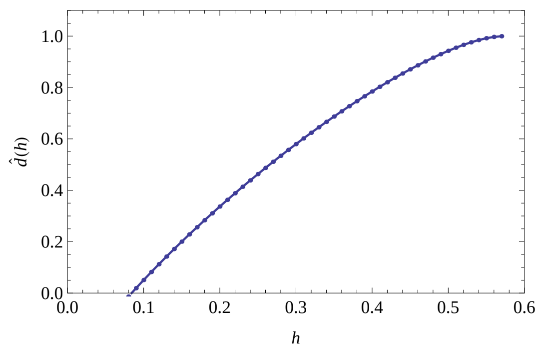
(d) Estimated scaling function of the S&P 500 index with τ_α^∞ for $\alpha = 2.5$



(e) Estimated scaling function of the generated Student Lévy process



(f) Estimated spectrum of the S&P 500 index



(g) Estimated spectrum of the generated Student Lévy process

Figure 4.3: S&P 500 index

4.4.3 Conclusion

To briefly summarize the main findings of this section, we showed that the data coming from heavy-tailed processes exhibit multifractal features. The first conclusion from this analysis is that one must be very careful when advocating multifractal model (like cascade process or MRW) based on the shape of the estimated scaling function. A large class of processes will behave as the moment scaling holds and this is hardly verifiable in practice. Even if the moment scaling does hold, heavy tails will produce concavity of the empirical scaling function.

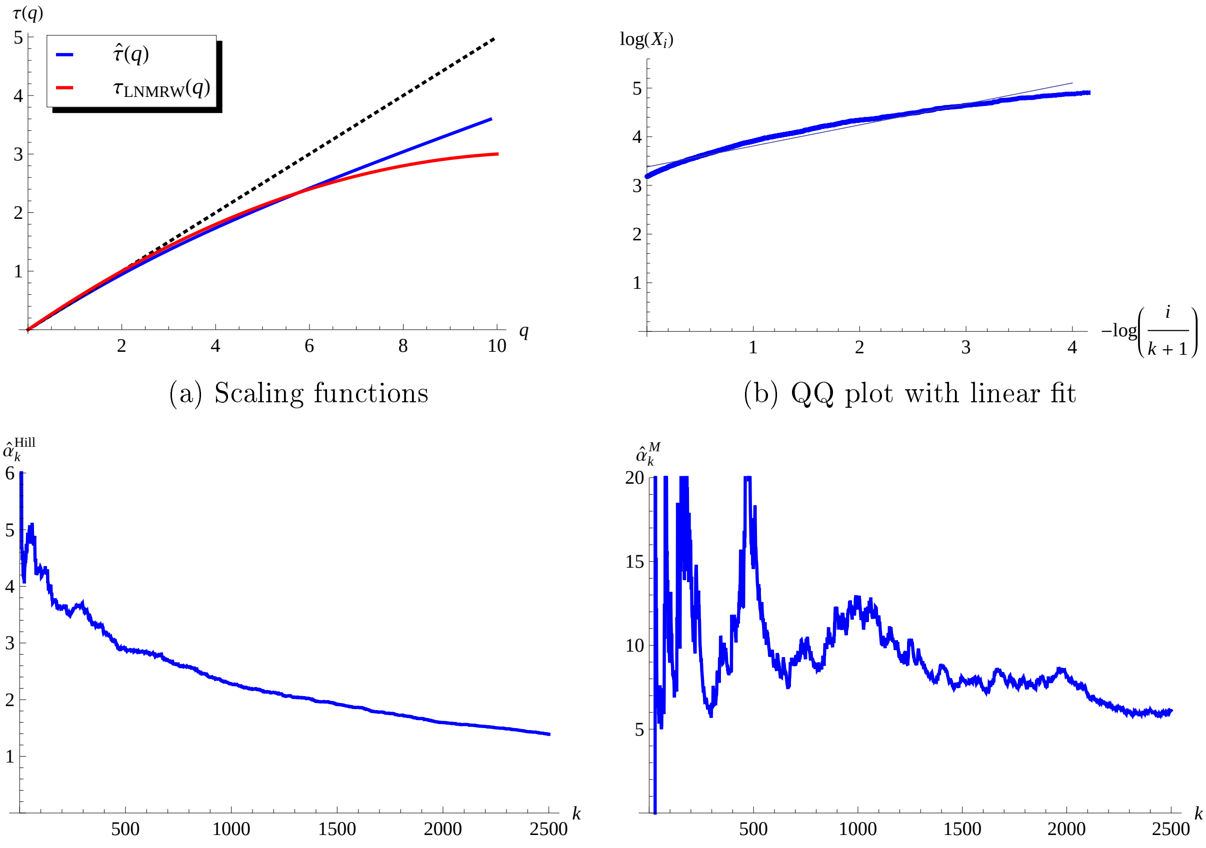
The nonlinearity of the estimated scaling function comes from estimating moments that are actually infinite. So it is obvious what we are doing wrong. The reasonable solution one may suggest is to first estimate the range of finite moments and then estimate the scaling function only for this range, as otherwise moment scaling (4.5) does not make sense. Therefore the first step should be to estimate the tail index and suppose we want to do this with the empirical scaling function as proposed in Chapter 2. For truly multifractal processes, e.g. cascades, the estimated scaling function would be nonlinear as there is nonlinear scaling and from the point of view of Chapter 2 this would lead us to classify the data as heavy-tailed. Thus, we arrive at the converse problem: multifractal processes with nonlinear scaling function may exhibit heavy tails. To illustrate this point and to show that this problem is not restricted only to methods of Chapter 2, but appears for other tail index estimators too, we present a simulation based example.

Example

Suppose $\{X(t)\}$ is LNMRW with intermittency parameter λ^2. We will analyze the distribution tail of the increments of one sample path of such process. In financial applications, based on fitting the estimated scaling function, parameter λ^2 is usually taken around 0.05. LNMRW is itself heavy-tailed (see Subsection 4.1.1) with the tail index equal to $1/\lambda^2$ and for $\lambda^2 = 0.05$ this gives the tail index value 20. Such large value is hardly observable in practice and this has been a major critique of multifractal models in modeling asset returns, as it is widely accepted that this data is heavy-tailed with the tail index between 3 and 5.

It has been observed first in Calvet & Fisher (2004) that the data from multifractal process may exhibit heavy-tailed properties with tail index much lower than its true value. In Muzy et al. (2006), the authors provide a discussion based on the analogy with a certain kind of physical system. They conclude, among other things, that estimating tail index of the log-normal cascade (which is $1/(2\lambda^2)$) with the classical estimators such as Hill and Pickands, will yield values around $\sqrt{(1-\nu)/(2\lambda^2)}$. Here ν represents the log-ratio of upper order statistics used in computing the estimator, i.e. $\ln k/\ln n$ for the sample of size n. This means, for example, that if for the sample of size 10000 of the LNC with

$\lambda^2 = 0.05$ we compute the Hill estimator using $k = 1000$, instead of the true value 10, we will get a value around 1.58.



(a) Scaling functions

(b) QQ plot with linear fit

(c) Hill plot

(d) Moment estimator plot

Figure 4.4: LNMRW generated with $\lambda^2 = 0.05$

Let us illustrate this phenomenon with the sample path of LNMRW of length 10000 generated with $\lambda^2 = 0.05$. Such process has finite moments up to order 20. Figure 4.4a shows the estimated scaling function with the corresponding τ_{LNMRW}. From the point of view of the results of Chapter 2, one may say that the scaling function exhibits nonlinearity due to heavy-tails. However, the moments considered are finite and it is completely legitimate to estimate them. One can observe the linearization effect as the discrepancy between the two plots starting at around $q = 6$. Other methods also indicate heavy-tails. QQ plot with $k = 2000$ (Figure 4.4b) shows approximate linearity and the estimated slope is 1/2.31. The Hill plot shows very low values, although it is not very informative (Figure 4.4c). The moment estimator plot stabilizes around $5 - 6$ (Figure 4.4d). Following Muzy et al. (2006), tail index estimators should correspond to $\sqrt{2/\lambda^2} \approx 6.3$ or lower, which is confirmed in this case.

If we would adopt an estimate 3 or 4, the scaling function is almost linear for this range and there would be no reason to consider multifractal models. To conclude, by using empirical scaling functions it is impossible to distinguish whether the data are heavy-tailed or multifractal.

One of the goals of the statistical analysis of scaling properties is to distinguish whether the data come from a process that is self-similar (4.2) or there is a more general multifractal scaling (4.3). Both properties can be reduced to moment scaling (4.5) and described by the scaling function, which is linear for self-similar processes and nonlinear for multifractals defined by (4.3). This holds for a range of finite moments, otherwise, the definition of the scaling function does not make sense. In theory, this is a completely clear distinction. In practice, one estimates the scaling function using estimated moments. For this to work, one must first examine if the moments under consideration are finite. This is a serious practical issue. First of all, this is a very challenging problem from the statistical point of view. Secondly, the previous example shows that for the main examples of multifractals the tail index is seriously underestimated. In the example considered, it is completely legitimate to estimate the scaling function as the moments are finite, but the data itself will behave differently. So, determining a range of finite moments before estimation would not be a good approach.

One may wonder if some other method may give better results. Wavelets are surely the most widely used tool for the multifractal analysis. However, the same problem seems to affect wavelets, as it is reported in Gonçalves & Riedi (2005). In this paper, the authors propose a tail index estimator to determine the range of finite moments.

Before jumping to conclusion that the estimated scaling functions are not very useful, one should remember that if the multifractal formalism holds, they can be used to estimate the spectrum of singularities. It is to expect that the nonlinearities that are artefact of the estimation method have nothing to do with the path properties. Although this is generally true, some examples may suggest differently. We leave this discussion for the next chapter where we investigate the relation between moments and small scale path properties.

Chapter 5

Moments and the spectrum of singularities

As we have seen in the previous chapter, empirical scaling functions are not capable to detect multifractal data as infinite positive order moments are empirically indistinguishable from multifractality. In this chapter we investigate the role of the scaling functions in the multifractal formalism. For stochastic processes, multifractal formalism would mean that it is possible to determine fine path properties using an object (scaling function) derived from the process distribution. Moreover, this object, like moments, should be equally defined for many, if not all processes. It is hard to believe that such general property may hold. In this chapter, we make contribution to the problem by identifying properties that make the spectrum nontrivial. We start with some motivating examples.

5.1 Motivation

Estimating infinite positive order moments is what makes the empirical scaling function to look misleading. However, the following examples suggest that if we are interested in the spectrum of singularities, empirical scaling functions can be the right approach.

Suppose $\{X(t)\}$ is a strictly α-stable Lévy process, $0 < \alpha < 2$. This process is $1/\alpha$-sssi and has multifractal paths with spectrum given by (4.13). Scaling function τ_{SLP} is defined on $(-1, \alpha)$ and $\tau_{SLP}(q) = q/\alpha$. The empirical scaling function asymptotically behaves as τ_α^∞ defined in Theorem 2, i.e.

$$\tau_\alpha^\infty(q) = \begin{cases} \frac{q}{\alpha}, & \text{if } 0 < q \leq \alpha, \\ 1, & \text{if } q > \alpha. \end{cases}$$

If we consider Legendre transforms of these two functions and take infimum over all positive q where they are defined, then one can easily check that

$$\inf_{0<q<\alpha} (hq - \tau_{SLP}(q) + 1) = \inf_{0<q<\infty} (hq - \tau_{\alpha}^{\infty}(q) + 1) = \begin{cases} \alpha h, & \text{if } h \in [0, 1/\alpha], \\ 1, & \text{if } h > 1/\alpha. \end{cases}$$

This actually coincides with the true spectrum (4.13), except for the part $h > 1/\alpha$, which is the infimum obtained when $q \to 0$. To correctly estimate this part one needs negative order moments, which will be discussed later. Thus, although we are estimating infinite moments we arrive at the correct spectrum.

The second example is more compelling. Suppose $\{X(t)\}$ is a LFSM with $\alpha \in [1, 2)$, $0 < H < 1$ and $H > 1/\alpha$. Since $\{X(t)\}$ is H-sssi, the scaling function is $\tau_{LFSM}(q) = Hq$ for $q \in (-1, \alpha)$. Theorem 3 justifies estimating the scaling function in the same manner as in Section 4.2 and Theorem 4 gives the asymptotic behavior:

$$\tau_{H,\alpha}^{\infty}(q) = \begin{cases} Hq, & \text{if } 0 < q \leq \alpha, \\ \left(H - \frac{1}{\alpha}\right) q + 1, & \text{if } q > \alpha. \end{cases}$$

Considering Legendre transform of τ_{LFSM} over $(0, \alpha)$ gives

$$\inf_{0<q<\alpha} (hq - \tau_{LFSM}(q) + 1) = \begin{cases} \alpha(h - H) + 1, & \text{if } h \in [0, H], \\ 1, & \text{if } h > H. \end{cases}$$

Although the expression is similar to the true spectrum d_{LFSM} defined in (4.15), the support is different. On the other hand, it is easy to check that

$$\inf_{0<q<\infty} \left(hq - \tau_{H,\alpha}^{\infty}(q) + 1\right) = \begin{cases} -\infty, & \text{if } h < H - 1/\alpha, \\ \alpha(h - H) + 1, & \text{if } h \in [H - 1/\alpha, H], \\ 1, & \text{if } h > H. \end{cases}$$

Thus, the empirical scaling function will lead to a correct left part of the spectrum using formalism. This reveals that the validity of the formalism may be narrow if τ is specified as in (4.5). Secondly, it shows the potential of the empirical scaling function and indicates that infinite positive order moments may be related with path properties. To further support this conjecture, consider FBM. FBM is H-sssi with all positive order moments finite and has a trivial spectrum consisting of only one point (4.12). This raises the question if infinite moments are what makes the spectrum of stable Lévy process and LFSM nontrivial. We deal with this in the next section. The following example illustrates that generally two properties are not related, however, the difference with the previous examples is that there is no exact scaling of finite dimensional distributions, like (4.1) or (4.3).

Suppose $\{X(t), t \geq 0\}$ is a Lévy process. The Lévy processes in general do not satisfy moment scaling of the form (4.5). The only such examples are the BM and strictly α-stable Lévy process. If $X(1)$ is heavy-tailed and with zero mean, the empirical scaling function will asymptotically behave as τ_α^∞ defined in (2.4). On the other hand, spectrum (4.14) depends on the BG index β. The estimated scaling function and the spectrum are not related as they depend on the different parts of the Lévy measure. The shape of the estimated scaling function is governed by the tail index, which depends on the behavior of the Lévy measure π at infinity since for $q > 0$, $E|X(1)|^q < \infty$ is equivalent to $\int_{|x|>1} |x|^q \pi(dx) < \infty$. On the other hand, the spectrum is determined by the behavior of π around origin, i.e. by the BG index. Stable Lévy processes are the special case since then $\alpha = \beta$.

5.1.1 Negative order moments

An example of the LNC suggests that to get the spectrum from the multifractal formalism, negative order moments must also be included. Negative order moments depend on the behavior of the distribution around 0. If taken for the increments of stochastic process, they may describe small variations of paths. The problem with negative order moments is that they are usually infinite. For example, if random variable X has continuous density f such that $f(0) > 0$, then $E|X|^{-1} = \infty$ and consequently, $E|X|^q = \infty$ for every $q \leq -1$. However, if f is continuous and bounded near zero, then $E|X|^q < \infty$ for every $-1 < q < 0$.

One can guess that infinite negative order moments may also produce nonlinear behavior of the estimated scaling function. In the spirit of the previous discussion, one may ask if these nonlinearities are related to the spectrum. We will show that this is not so in the next section.

One of the first references reporting this problem is Muzy et al. (1993), where the authors introduced wavelet methods in the multifractal analysis. They proposed the method called wavelet transform modulus maxima, which is based on a continuous wavelet transform of a process where the partition function is defined by taking maxima over the same scale wavelet transforms. Later the method based on wavelet leaders was introduced, where the partition function is based on the local suprema of wavelet coefficients (see e.g. Jaffard et al. (2007) and references therein). This resembles the method we propose in Section 5.5, although our motivation comes entirely from the results of Section 5.2. Two main advantages of the wavelet methods over the increments based partition function are emphasized in the literature. First, wavelet leaders are more stable for divergent negative order moments. In Section 5.5 we modify the partition function to have the same property and this idea relies on the precise results about the spectrum. The second advantage is the insensitivity of wavelets to polynomial trends, which makes it possible to study more generally defined Hölder exponents (see Subsection 4.1.2). We conform to the definition (4.10) in order to simplify the analysis and relate infinite moments with path properties.

5.2 Bounds on the support of the spectrum

Our goal in this section is to identify the property of the process that makes the spectrum nontrivial. We do this by deriving the bounds on the support of the spectrum. The lower bound is a consequence of the well-known Kolmogorov's continuity theorem. For the upper bound we prove a sort of complement of this theorem.

Before we proceed, we fix the following notation for a process $\{X(t), t \in \mathcal{T}\}$ where $\mathcal{T} = [0, T]$ or $\mathcal{T} = [0, +\infty)$. We denote the range of finite moments as $\mathcal{Q} = (\underline{q}, \overline{q})$, i.e.

$$\begin{aligned} \overline{q} &= \sup\{q > 0 : E|X(t)|^q < \infty, \ \forall t\}, \\ \underline{q} &= \inf\{q < 0 : E|X(t)|^q < \infty, \ \forall t\}. \end{aligned} \tag{5.1}$$

If $\{X(t)\}$ is multifractal in the sense of Definition 2 with the scaling function τ define

$$\begin{aligned} H^- &= \sup\left\{\frac{\tau(q)}{q} - \frac{1}{q} : q \in (0, \overline{q}) \ \& \ \tau(q) > 1\right\}, \\ \widetilde{H^+} &= \inf\left\{\frac{\tau(q)}{q} - \frac{1}{q} : q \in (\underline{q}, 0) \ \& \ \tau(q) < 1\right\}, \end{aligned} \tag{5.2}$$

with the convention that $\sup \emptyset = 0$ and $\inf \emptyset = +\infty$. In this context, we always assume that (4.5) holds on the whole $\mathcal{T}$ and $\mathcal{Q}$. All the processes $\{X(t), t \in \mathcal{T}\}$ considered here are defined on some probability space $(\Omega, \mathcal{F}, P)$ and measurable, meaning that $(t, \omega) \mapsto X(t, \omega)$ is $\mathcal{B}(\mathcal{T}) \times \mathcal{F}$-measurable. Furthermore, we assume that $\{X(t), t \in \mathcal{T}\}$ is separable with respect to any dense countable set $\mathbb{T} \subset \mathcal{T}$, in the sense that for all $t \in \mathcal{T}$ there exists a sequence (t_n) in $\mathbb{T}$, $t_n \to t$ such that a.s. $X(t_n) \to X(t)$. We say that the two processes $\{X(t), t \in \mathcal{T}\}$ and $\{\tilde{X}(t), t \in \mathcal{T}\}$ defined on the same probability space are modifications of each other if for every $t \in \mathcal{T}$, $P(X(t) = \tilde{X}(t)) = 1$. If $P(X(t) = \tilde{X}(t), \forall t \in \mathcal{T}) = 1$, we say that the two processes are indistinguishable. Every stochastic process $\{X(t), t \in \mathcal{T}\}$ has a separable modification (see e.g. Doob (1953)).

5.2.1 The lower bound

Using the well-known Kolmogorov's criterion it is easy to derive the lower bound on the support of the spectrum. Before stating the theorem, we define $f : \mathcal{T} \to \mathbb{R}$ to be locally Hölder continuous of order γ if for every compact $K \subset \mathcal{T}$ there exists a constant $C(K)$ such that

$$|f(t) - f(s)| \leq C(K)|t - s|^\gamma, \quad \forall t, s \in K.$$

It is clear that the local Hölder continuity at some domain implies pointwise Hölder continuity of the same order at any point. The proof of the following theorem can be found in (Karatzas & Shreve 1991, Theorem 2.8) or (Kallenberg 2002, Theorem 3.23).

Theorem 5 (Kolmogorov-Chentsov). *Suppose that a process $\{X(t), t \in \mathcal{T}\}$ satisfies*

$$E|X(t) - X(s)|^{\alpha} \leq C|t - s|^{1+\beta}, \quad \forall t, s \in \mathcal{T}, \tag{5.3}$$

for some constants $\alpha > 0$, $\beta > 0$ and $C > 0$. Then there exists a modification $\{\tilde{X}(t), t \in \mathcal{T}\}$ of $\{X(t), t \in \mathcal{T}\}$ having continuous sample paths. Furthermore, a.s. $\{\tilde{X}(t)\}$ is locally Hölder continuous of order γ for every $\gamma \in (0, \beta/\alpha)$.

Proposition 1. *Suppose $\{X(t), t \in \mathcal{T}\}$ is multifractal in the sense of Definition 2. If $\tau(q) > 1$ for some $q \in (0, \overline{q})$, then there exists a modification of $\{X(t)\}$ which is a.s. locally Hölder continuous of order γ for every*

$$\gamma \in \left(0, \frac{\tau(q)}{q} - \frac{1}{q}\right).$$

In particular, there exists a modification such that a.s.

$$H^{-} \leq H(t), \quad \forall t \in \mathcal{T},$$

where $H(t)$ is defined by (4.11) and H^{-} by (5.2).

Proof. This is a simple consequence of Theorem 5 since Definition 2 implies

$$E|X(t) - X(s)|^{q} = c(q)|t - s|^{1+(\tau(q)-1)}.$$

For the second part, if $H^{-} = 0$ there is nothing to prove. Otherwise, by (5.2), for each $\gamma < H^{-}$ there is $q \in (0, \overline{q})$ such that $\tau(q) > 1$ and $\gamma < (\tau(q) - 1)/q$, and thus, by the first part there is modification which is a.s. locally Hölder continuous of order γ. Since all continuous modifications are indistinguishable (see e.g. (Karatzas & Shreve 1991, Problem 1.5), we have the desired modification. This implies that a.s. the pointwise Hölder exponent is everywhere greater than H^{-}. □

In the sequel we always suppose to work with the modification from Proposition 1. If $H^{-} > 0$, we conclude that the spectrum $d(h) = -\infty$ for $h \in (0, H^{-})$. This way we can establish an estimate for the left endpoint of the support of the spectrum. It also follows that if the process is H-sssi and has finite moments of every positive order, then $H^{-} = H \leq H(t)$. Thus, when the moment scaling holds, path irregularities are closely related with infinite moments of positive order. We make this point stronger later.

Theorem 5 is valid for general stochastic processes. Although moment condition (5.3) is appealing, the condition needed for the proof of Theorem 5 can be stated in a different form.

Corollary 1. *For a process $\{X(t), t \in \mathcal{T}\}$ there exists a modification which is a.s. locally Hölder continuous of order $\gamma > 0$ if for some $\eta > 1$ it holds that for every $K > 0$ there* *exists* $C > 0$ *such that*

$$\limsup_{t\to 0} \frac{P\left(|X(s+t)-X(s)| \geq Kt^{\gamma}\right)}{t^{\eta}} \leq C, \quad \forall s \in \mathcal{T}.$$

Proof. This is obvious from the proof of Theorem 5; see (Karatzas & Shreve 1991, Theorem 2.8). □

5.2.2 The upper bound

To establish the bound on the right endpoint of the support of the spectrum, one needs to show that a.s. the sample paths are nowhere Hölder continuous of some order γ, i.e. that a.s. $t \mapsto X(t) \notin C^{\gamma}(t_0)$ for each $t_0 \in \mathcal{T}$. To show this we first use a criterion based on the negative order moments, similar to (5.3). The resulting theorem can be seen as a sort of a complement of the Kolmogorov-Chentsov theorem. We then apply this to moment scaling multifractals to get an estimate for the support of the spectrum.

In proving the statements involving negative order moments we use the following two simple facts at several places. The first is a Markov's inequality for negative order moments. If X is a random variable, $\varepsilon > 0$ and $q < 0$, then

$$P(|X| \leq \varepsilon) = P(|X|^q \geq \varepsilon^q) \leq \frac{E|X|^q}{\varepsilon^q}.$$

The second fact is the expression for the q-th order moment, $q < 0$,

$$E|X|^q = -\int_0^{\infty} q y^{-q-1} P(1/|X| \geq y) dy = -\int_0^{\infty} q y^{q-1} P(|X| \leq y) dy.$$

Theorem 6. *Suppose that a process* $\{X(t), t \in \mathcal{T}\}$ *satisfies*

$$E|X(t)-X(s)|^{\alpha} \leq C|t-s|^{1+\beta}, \quad \forall t,s \in \mathcal{T}, \tag{5.4}$$

for some constants $\alpha < 0$, $\beta < 0$ *and* $C > 0$. *Then a.s.* $\{X(t)\}$ *is nowhere Hölder continuous of order* γ *for every* $\gamma > \beta/\alpha$.

Proof. For typographical convenience we sometimes write X_t for $X(t)$. First, it suffices to prove the statement by fixing arbitrary $\gamma > \beta/\alpha$. Indeed, this would give events Ω_γ, $P(\Omega_\gamma) = 0$ such that for $\omega \in \Omega \backslash \Omega_\gamma$, $t \mapsto X_t(\omega)$ is nowhere Hölder continuous of order γ. If Ω_0 is the union of Ω_γ over all $\gamma \in (\beta/\alpha, \infty) \cap \mathbb{Q}$, then $\Omega_0 \in \mathcal{F}$, $P(\Omega_0) = 0$ and $\Omega \backslash \Omega_0$ would fit the statement of the theorem.

Secondly, it is enough to consider only restrictions to the interval $[0,1)$, as, if needed, for $n \in \mathbb{N}$ we get from this the proof for the interval $[n, n+1)$ by using the process $X'(t) = X(n+t) - X(n)$, $t \in [0,1)$. Removing null sets for all $n \in \mathbb{N}$ would imply the general statement.

For $j, k \in \mathbb{N}$ define the set

$$M_{jk} := \bigcup_{t\in[0,1)} \bigcap_{h\in[0,1/k]} \{\omega \in \Omega : |X_{t+h}(\omega) - X_t(\omega)| \leq jh^\gamma\}.$$

It is clear that if $\omega \notin M_{jk}$ for every $j, k \in \mathbb{N}$, then $t \mapsto X_t(\omega)$ is nowhere Hölder continuous of order γ. As there is countably many M_{jk}, it is enough to fix arbitrary $j, k \in \mathbb{N}$ and show that $M_{jk} \subset A$ for some $A \in \mathcal{F}$ such that $P(A) = 0$.

Suppose $n > 2k$ and $\omega \in M_{jk}$. Then there is some $t \in [0, 1)$ such that

$$|X_{t+h}(\omega) - X_t(\omega)| \leq jh^\gamma, \quad \forall h \in [0, 1/k]. \tag{5.5}$$

Take $i \in \{1, \dots, n\}$ such that

$$\frac{i-1}{n} \leq t < \frac{i}{n}. \tag{5.6}$$

Since $n > 2k$ we have

$$0 \leq \frac{i}{n} - t < \frac{i+1}{n} - t \leq \frac{i+1}{n} - \frac{i-1}{n} = \frac{2}{n} < \frac{1}{k},$$

and from (5.5) it follows that

$$|X_{\frac{i+1}{n}}(\omega) - X_{\frac{i}{n}}(\omega)| \leq |X_{\frac{i+1}{n}}(\omega) - X_t(\omega)| + |X_t(\omega) - X_{\frac{i}{n}}(\omega)| \leq 2^{\gamma+1}jn^{-\gamma}.$$

Put $A_i^{(n)} = \left\{|X(\frac{i+1}{n}) - X(\frac{i}{n})| \leq 2^{\gamma+1}jn^{-\gamma}\right\}$. Since ω was arbitrary it follows that

$$M_{jk} \subset \bigcup_{i=1}^{n} A_i^{(n)}.$$

Using Markov's inequality for $\alpha < 0$ and the assumption of the theorem we get

$$\begin{aligned} P(A_i^{(n)}) &\leq \frac{E|X(\frac{i+1}{n}) - X(\frac{i}{n})|^\alpha}{(2^{\gamma+1}j)^\alpha n^{-\gamma\alpha}} \leq C(2^{\gamma+1}j)^{-\alpha}n^{\gamma\alpha-1-\beta}, \\ P\left(\bigcup_{i=1}^{n} A_i^{(n)}\right) &\leq \sum_{i=1}^{n} P(A_i^{(n)}) \leq C(2^{\gamma+1}j)^{-\alpha}n^{-(\beta-\gamma\alpha)}. \end{aligned} \tag{5.7}$$

If we set

$$A = \bigcap_{n>2k} \bigcup_{i=1}^{n} A_i^{(n)},$$

then $A \in \mathcal{F}$ and $M_{jk} \subset A$. Since $\gamma > \beta/\alpha$, it follows that $\beta - \gamma\alpha > 0$ and hence $P(A) = 0$. This proves the theorem. □

Proposition 2. *Suppose $\{X(t), t \in \mathcal{T}\}$ is multifractal in the sense of Definition 2. If $\tau(q) < 1$ for some $q \in (\underline{q}, 0)$, then a.s. $\{X(t)\}$ is nowhere Hölder continuous of order γ*

for every

$$\gamma \in \left(\frac{\tau(q)}{q} - \frac{1}{q}, \ +\infty \right).$$

In particular, a.s.

$$H(t) \leq \widetilde{H^+}, \quad \forall t \in \mathcal{T}.$$

Proof. Definition 2 implies

$$E|X(t) - X(s)|^q = c(q)|t - s|^{1+(\tau(q)-1)}.$$

Since $q < 0$, $\tau(q) < 0$ and the statement follows from Theorem 6. □

This proposition shows that $d(h) = -\infty$ for $h \in (\widetilde{H^+}, \infty)$. Recall that $\widetilde{H^+}$ is defined in (5.2).

Remark 5. Statements like the ones in Proposition 1 and 2 are stronger than saying, for example, that for every $t \in \mathcal{T}$, $H(t) \leq U$ a.s. Indeed, an application of the Fubini's theorem would yield that for almost every path, $H(t) \leq U$ for almost every t. If we put $h = U + \delta$, then the Lebesgue measure of the set $S_h = \{t : H(t) = h\}$ is zero a.s. This, however, does not imply that $d(h) = -\infty$ and hence, it is impossible to say something about the spectrum of almost every sample path. On the other hand, it is clear that this type of statements are implied by Propositions 1 and 2.

For the example of this weaker type of the bound, consider $\{X(t), \in \mathcal{T}\}$ multifractal in the sense of Definition 2. If there is $q \in (\underline{q}, 0)$, then for every $t \in \mathcal{T}$

$$H(t) \leq \frac{\tau(q)}{q} \text{ a.s.}$$

Indeed, let $\delta > 0$ and suppose $C > 0$. Since $q < 0$, by Markov's inequality

$$P\left(|X(t+\varepsilon) - X(t)| \leq C|\varepsilon|^{\frac{\tau(q)}{q}+\delta} \right) \leq \frac{E\left|X(t+\varepsilon) - X(t)\right|^q}{C^q|\varepsilon|^{\tau(q)+\delta q}} = \frac{c(q)}{C^q|\varepsilon|^{\delta q}} \to 0,$$

as $\varepsilon \to 0$. We can choose a sequence (ε_n) that converges to zero such that

$$P\left(|X(t+\varepsilon_n) - X(t)| \leq C|\varepsilon_n|^{\frac{\tau(q)}{q}+\delta} \right) \leq \frac{1}{2^n}.$$

Now, by the Borel-Cantelli lemma

$$\frac{|X(t+\varepsilon_n) - X(t)|}{|\varepsilon_n|^{\frac{\tau(q)}{q}+\delta}} \to \infty \quad \text{a.s., as } n \to \infty.$$

Thus, for arbitrary $\delta > 0$ it holds that for every t, $H(t) \leq \frac{\tau(q)}{q} + \delta$ a.s. However, this result does not allow us to say anything about the spectrum.

Consider for the moment the FBM. The range of finite moments is $(-1,\infty)$ and $\tau(q)=Hq$ for $q\in(-1,\infty)$, so we have $\widetilde{H^+}=H+1$. Thus, the best we can say from Proposition 2, is that $d(h)=-\infty$ for $h>H+1$. However, we know that $d(h)=-\infty$ for $h>H$. If the infimum in the definition of $\widetilde{H^+}$ could be considered over all negative q, we would get exactly the right endpoint of the support of the spectrum.

The fact that the bound derived in Proposition 2 is not sharp enough for some examples points that negative order moments may not be the right paradigm to explain the spectrum. We therefore provide more general conditions that do not depend on the finiteness of moments. First of them is obvious from the proof of Theorem 6, Equation (5.7).

Corollary 2. *A process $\{X(t), t\in\mathcal{T}\}$ is a.s. nowhere Hölder continuous of order $\gamma>0$ if for some $\eta>1$ it holds that for every $K>0$ there exists $C>0$ such that*

$$\limsup_{t\to 0}\frac{P\left(|X(s+t)-X(s)|\leq Kt^{\gamma}\right)}{t^{\eta}}\leq C,\quad \forall s\in\mathcal{T}.$$

Theorem 7. *A process $\{X(t), t\in\mathcal{T}\}$ is a.s. nowhere Hölder continuous of order $\gamma>0$ if for some $\eta>1$ and $m\in\mathbb{N}$ it holds that for every $K>0$ there exists $C>0$ such that*

$$\limsup_{t\to 0}\frac{P\left(\max\limits_{l=1,\dots,m}|X(s+lt)-X(s+(l-1)t)|\leq Kt^{\gamma}\right)}{t^{\eta}}\leq C,\quad \forall s\in\mathcal{T}. \tag{5.8}$$

Proof. The first part of the proof goes exactly as in the proof of Theorem 6. Fix $j,k\in\mathbb{N}$ and take $n\in\mathbb{N}$ such that

$$n>(m+1)k.$$

If $\omega\in M_{jk}$, then there is some $t\in[0,1)$ and $i\in\{1,\dots,n\}$ such that (5.5) and (5.6) hold. Choice of n ensures that for $l\in\{1,\dots,m\}$

$$0<\frac{i+l-1}{n}-t<\frac{i+l}{n}-t<\frac{i+l}{n}-\frac{i-1}{n}=\frac{l+1}{n}\leq\frac{1}{k}.$$

It follows from (5.5) that for each $l\in\{1,\dots,m\}$

$$|X_{\frac{i+l}{n}}(\omega)-X_{\frac{i+l-1}{n}}(\omega)|\leq j\left(\frac{l+1}{n}\right)^{\gamma}+j\left(\frac{l}{n}\right)^{\gamma}\leq 2j\left(\frac{m+1}{n}\right)^{\gamma}.$$

Let

$$\begin{aligned}A_{i,l}^{(n)}&=\left\{|X(\tfrac{i+l}{n})-X(\tfrac{i+l-1}{n})|\leq 2j\left(\tfrac{m+1}{n}\right)^{\gamma}\right\},\\ A_i^{(n)}&=\bigcap_{l=1}^{m}A_{i,l}^{(n)}.\end{aligned}$$

It then follows that

$$M_{jk} \subset \bigcup_{i=1}^{n} A_i^{(n)}.$$

From the assumption, there exists $C > 0$ such that

$$P(A_i^{(n)}) = P\left(\max_{l=1,\dots,m} |X(\tfrac{i+l}{n}) - X(\tfrac{i+l-1}{n})| \le 2j(m+1)^{\gamma}\left(\tfrac{1}{n}\right)^{\gamma}\right) \le Cn^{-\eta},$$

$$P\left(\bigcup_{i=1}^{n} A_i^{(n)}\right) \le \sum_{i=1}^{n} P(A_i^{(n)}) \le Cn^{-(\eta-1)}.$$

Now setting

$$A = \bigcap_{n>(m+1)k} \bigcup_{i=1}^{n} A_i^{(n)} \in \mathcal{F},$$

it follows that $P(A) = 0$, since $\eta > 1$. □

Theorem 7 enables one to avoid using moments in deriving the bound. As an example, we consider how Theorem 7 can be applied in the simple case when $\{X(t)\}$ is BM. Since $\{X(t)\}$ is 1/2-sssi we have

$$P\left(\max_{l=1,\dots,m} |X(lt) - X((l-1)t)| \le Kt^{\gamma}\right) = P\left(\max_{l=1,\dots,m} |X(l) - X(l-1)| \le Kt^{\gamma-1/2}\right).$$

Due to independent increments:

$$P\left(\max_{l=1,\dots,m} |X(l) - X(l-1)| \le Kt^{\gamma-1/2}\right) = \left(P\left(|X(1)| \le Kt^{\gamma-1/2}\right)\right)^m \le Ct^{m(\gamma-1/2)}.$$

This holds for every $\gamma > 1/2$ and $m \in \mathbb{N}$ and by taking $m > 1/(\gamma - 1/2)$ we conclude that $d(h) = -\infty$ for $h > 1/2$.

Before we proceed on applying these results, we state the following simple corollary that expresses the criterion (5.8) in terms of the negative order moments, but now moments of the maximum of increments. This is a generalization of Theorem 6, which enables bypassing infinite negative order moments under very general conditions. From this criterion, in the next section we derive strong statements about H-sssi processes.

Corollary 3. *Suppose that a process $\{X(t), t \in \mathcal{T}\}$ satisfies*

$$E\left[\max_{l=1,\dots,m} |X(s+lt) - X(s+(l-1)t)|\right]^{\alpha} \le Ct^{1+\beta}, \quad \forall t, s \in \mathcal{T}, \tag{5.9}$$

for some $\alpha < 0$, $\beta < 0$, $m \in \mathbb{N}$ and $C > 0$. Then a.s. $\{X(t)\}$ is nowhere Hölder continuous of order γ for every $\gamma > \beta/\alpha$.

Proof. This follows directly from the Markov's inequality for negative order moments and

Theorem 7 since

$$P\left(\max_{l=1,\dots,m}|X(s+lt)-X(s+(l-1)t)|\leq Kt^{\gamma}\right)$$
$$\leq K^{-\alpha}t^{-\gamma\alpha}E\left[\max_{l=1,\dots,m}|X(s+lt)-X(s+(l-1)t)|\right]^{\alpha}\leq K^{-\alpha}Ct^{-\alpha\gamma+1+\beta},$$

and $1+\beta-\alpha\gamma>1$. □

5.3 Applications

In this section we consider in more details processes with scaling properties and apply the results of the previous section to derive some very general statements about these processes.

5.3.1 The case of self-similar stationary increments processes

In this subsection we refine our results for the case of H-sssi processes by using Corollary 3. These results can also be viewed in the light of the classical papers Vervaat (1985) and Takashima (1989). To be able to apply Corollary 3, we need to make sure that the moment in (5.9) can be made finite by choosing m large enough. We state this condition explicitly for reference.

Condition 1. *Suppose $\{X(t),t\in\mathcal{T}\}$ is a stationary increments process. For every $\alpha<0$ there is $m_0\in\mathbb{N}$ such that*

$$E\left[\max_{l=1,\dots,m_0}|X(l)-X(l-1)|\right]^{\alpha}<\infty.$$

One way of assessing the Condition 1 is given in the following lemma, which is strong enough to cover all the examples considered later. Recall the definition of the range of finite moments $\underline{q}$ and $\overline{q}$ given in (5.1).

Lemma 3. *Suppose $\{X(t),t\geq 0\}$ is a stationary increments process which is ergodic in the sense that if $E|f(X_1)|<\infty$ for some measurable f, then*

$$\frac{\sum_{l=1}^{m}f(X_l-X_{l-1})}{m}\overset{a.s.}{\to}Ef(X_1),\quad as\ m\to\infty.$$

Suppose also that $\underline{q}<0$. Then Condition 1 holds.

Proof. Let $r < 0$ be such that $E|X(1) - X(0)|^r < \infty$. Then

$$\begin{aligned}\inf_{l\in\mathbb{N}} |X(l) - X(l-1)|^r &= \lim_{m\to\infty} \min_{l=1,\dots,m} |X(l) - X(l-1)|^r \\ &\leq \lim_{m\to\infty} \frac{\sum_{l=1}^m |X(l) - X(l-1)|^r}{m} = E|X(1) - X(0)|^r =: M \quad \text{a.s.}\end{aligned}$$

For $\alpha < 0$ it follows

$$\inf_{l\in\mathbb{N}} |X(l) - X(l-1)|^\alpha = \left(\inf_{l\in\mathbb{N}} |X(l) - X(l-1)|^r\right)^{\frac{\alpha}{r}} \leq M^{\frac{\alpha}{r}} \quad \text{a.s.}$$

and $\inf_{l\in\mathbb{N}} |X(l) - X(l-1)|^\alpha$ is bounded and thus has finite expectation. Given $\alpha < 0$ we can choose m_0 such that

$$\begin{aligned}&\left[\max_{l=1,\dots,m_0} |X(l) - X(l-1)|\right]^\alpha = \left[\frac{1}{\max_{l=1,\dots,m_0} |X(l) - X(l-1)|}\right]^{-\alpha} \\ &= \left[\min_{l=1,\dots,m_0} \frac{1}{|X(l) - X(l-1)|}\right]^{-\alpha} = \min_{l=1,\dots,m_0} |X(l) - X(l-1)|^\alpha \leq M^{\frac{\alpha}{r}} \quad \text{a.s.,}\end{aligned}$$

which implies the statement. □

Remark 6. Two examples may provide insight of how far the assumptions of Lemma 3 are from Condition 1. If $X(t) = tX$ for some random variable X, then

$$\max_{l=1,\dots,m} |X(l) - X(l-1)| = X$$

and thus, Condition 1 depends on the range of finite moments of X. For the second example, suppose $X(l) - X(l-1)$ is an i.i.d. sequence such that $P(|X(1) - X(0)| \leq x) = -\ln 2/\ln x$ for $x \in (0, 1/2)$. This implies, in particular, that $E|X(1) - X(0)|^r = \infty$ for any $r < 0$. Moreover,

$$\begin{aligned}E\left[\max_{l=1,\dots,m} |X(l) - X(l-1)|\right]^\alpha &= -\int_0^\infty \alpha y^{\alpha-1} P\left(\max_{l=1,\dots,m} |X(l) - X(l-1)| \leq y\right) dy \\ &= -\int_0^\infty \alpha y^{\alpha-1} \frac{1}{(\ln y)^m} dy = \infty,\end{aligned}$$

for every $\alpha < 0$ and $m \in \mathbb{N}$, thus Condition 1 does not hold.

We are now ready to prove a general theorem about H-sssi processes.

Theorem 8. *Suppose $\{X(t), t \geq 0\}$ is H-sssi stochastic process such that Condition 1 holds and $H - 1/\overline{q} \geq 0$. Then a.s.*

$$H - \frac{1}{\overline{q}} \leq H(t) \leq H, \quad \forall t \geq 0.$$

Proof. By the same argument as in the beginning of the proof of Theorem 6 it is enough to take arbitrary $\gamma > H$. Given γ we take $\alpha < 1/(H-\gamma) < 0$ which implies $\gamma > H - 1/\alpha$. Due to Condition 1, we can choose $m_0 \in \mathbb{N}$ such that $E\left[\max_{l=1,\dots,m_0} |X(lt) - X((l-1)t)|\right]^\alpha < \infty$. Self-similarity then implies that

$$E\left[\max_{l=1,\dots,m_0} |X(lt) - X((l-1)t)|\right]^\alpha = t^{H\alpha} E\left[\max_{l=1,\dots,m_0} |X(l) - X(l-1)|\right]^\alpha = Ct^{1+(H\alpha-1)}.$$

The claim now follows immediately from Corollary 3 with $\beta = H\alpha - 1$ since $\gamma > \beta/\alpha$. □

A simple consequence of the preceding is the following statement.

Corollary 4. *Suppose that Condition 1 holds. A H-sssi process with all positive order moments finite has a trivial spectrum, i.e. $d(h) = -\infty$ for $h \neq H$.*

This applies to FBM, but also to all Hermite processes, like e.g. Rosenblatt process (see Section 5.4). Thus, under very general conditions a self-similar stationary increments process with a nontrivial spectrum must be heavy-tailed. This shows clearly how infinite moments can affect path properties when the scaling holds. The following simple result shows this more precisely.

Proposition 3. *Suppose $\{X(t), t \geq 0\}$ is H-sssi. If $\gamma < H$ and $d(\gamma) \neq -\infty$, then $E|X(1)|^q = \infty$ for $q > 1/(H-\gamma)$.*

Proof. Suppose $E|X(t)|^q < \infty$ for $q > 1/(H-\gamma)$. Then for $\varepsilon > 0$ we can apply Markov's inequality to get

$$P\left(|X(t)| \geq Kt^\gamma\right) = P\left(|X(1)| \geq Kt^{\gamma-H}\right) \leq \frac{E\left|X(1)\right|^{\frac{1}{H-\gamma}+\varepsilon}}{K^{\frac{1}{H-\gamma}+\varepsilon} t^{-1-\varepsilon(H-\gamma)}} \leq Ct^{1+\varepsilon(H-\gamma)}.$$

By Corollary 1 this implies $d(\gamma) = -\infty$, which is a contradiction. □

5.3.2 The case of multifractal processes

Our next goal is to show that in the definition of $\widetilde{H^+}$ one can essentially take the infimum over all $q < 0$. At the moment this makes no sense as τ from Definition 2 may not be defined in this range. It is therefore necessary to redefine the meaning of the scaling function and thus we work with the more general Definition 1.

In the next section we will see on the example of the LNC process that when the multifractal process has all negative order moments finite, the bound derived in Proposition 2 is sharp. In general, this would not be the case for any multifractal in the sense of Definition 1. Take for example a LNMRW, which is a compound process $X(t) = B(\theta(t))$ where B is BM and θ is an independent LNC. By the multifractality of the cascade for $t < 1$, $\theta(t) =^d M(t)\theta(1)$ and multifractality of LNMRW implies $X(t) =^d (M(t)\theta(1))^{1/2} B(1)$.

Now by the independence of B and θ, if $E|B(1)|^q = \infty$, then $E|X(t)|^q = \infty$. Since $B(1)$ is Gaussian, the moments will be infinite for $q \leq -1$.

We thus provide a more general bound which only has a restriction on the moments of the random factor from Definition 1. Therefore, if the process satisfies Definition 1 and if the random factor M is multifractal by Definition 2 with scaling function τ, we define

$$H^+ = \inf\left\{\frac{\tau(q)}{q} - \frac{1}{q} : q < 0 \ \& \ E|M(t)|^q < \infty\right\}.$$

Corollary 5. *Suppose $\{X(t), t \in \mathcal{T}\}$ has stationary increments and Condition 1 holds. Suppose also it is multifractal by Definition 1 and the random factor M satisfies Definition 2 with scaling function τ. If $E|M(t)|^q < \infty$, $\forall t \in \mathcal{T}$ for some $q < 0$, then a.s. $\{X(t)\}$ is nowhere Hölder continuous of order γ for every*

$$\gamma \in \left(\frac{\tau(q)}{q} - \frac{1}{q}, \ +\infty\right).$$

In particular, a.s.

$$H(t) \leq H^+, \quad \forall t \in \mathcal{T}.$$

Proof. By Condition 1 for m large enough it follows from the multifractal property (4.3) that

$$E\left[\max_{l=1,\dots,m} |X(lt) - X((l-1)t)|\right]^q = E|M(t)|^q E\left[\max_{l=1,\dots,m} |X(l) - X(l-1)|\right]^q = Ct^{1+\tau(q)-1}.$$

The claim now follows from Corollary 3 with $\alpha = q$ and $\beta = \tau(q) - 1$ and by the argument at the beginning of the proof of Theorem 6. □

In summary, we provide bounds on the support of the multifractal spectrum. We show that the lower bound can be derived using positive order moments and link infinite moments with path properties for the case of H-sssi process. In general, negative order moments are not appropriate for explaining the right part of the spectrum. To derive an upper bound on the support of the spectrum, we use negative order moments of the maximum of increments. This avoids the nonexistence of the negative order moments, which is a property of the distribution itself.

5.4 Examples

In this section we list several examples of stochastic processes and investigate how the results of Sections 5.2 and 5.3 apply in these cases. We also discuss how the multifractal formalism could be achieved.

5.4.1 Self-similar processes

It follows from Theorem 8 and Corollary 4 that if H-sssi process is ergodic with finite positive order moments, then the spectrum is simply

$$d(h) = \begin{cases} 1, & \text{if } \ h = H \\ -\infty, & \text{otherwise.} \end{cases}$$

This applies to all Hermite processes, e.g. BM, FBM and Rosenblatt process. Hermite process $\{Z_H^{(k)}(t), t \geq 0\}$ with $H \in (1/2, 1)$ and $k \in \mathbb{N}$ can be defined as

$$Z_H^{(k)}(t) = C(H,k) \int_{\mathbb{R}^k}' \int_0^t \left(\prod_{j=1}^{k} (s - y_j)_+^{-\left(\frac{1}{2} + \frac{1-H}{k}\right)} \right) ds dB(y_1) \cdots dB(y_k), \quad t \geq 0,$$

where $\{B(t)\}$ is the standard BM and the integral is taken over $\mathbb{R}^k$ except the hyperplanes $y_i = y_j$, $i \neq j$. The constant $C(H,k)$ is chosen such that $E[Z_H^{(k)}(1)]^2 = 1$ and $(x)_+ = \max(x, 0)$. Hermite processes are H-sssi. For $k = 1$ one gets the FBM and for $k = 2$ the Rosenblatt process. See Embrechts & Maejima (2002) for more details.

Hermite processes have all positive order moments finite and the increments are ergodic (see e.g. (Samorodnitsky 2007, Section 7)), so they have a trivial spectrum. The spectrum of Hermite processes has been studied so far only numerically (Wendt et al. (2012)). We now discuss heavy tailed examples of H-sssi processes.

Stable Lévy processes

As already discussed in Section 5.1, the spectrum of a strictly α-stable Lévy process is nontrivial and supported on $[0, 1/\alpha]$. These are exactly the bounds given in Theorem 8 as in this case $H = 1/\alpha$ and $\overline{q} = \alpha$.

Linear fractional stable motion

The spectrum of the LFSM for $\alpha \in [1, 2)$, $H \in (0, 1)$ and in the long-range dependence case $H > 1/\alpha$ is supported on $[H - \frac{1}{\alpha}, H]$. Also, increments of the LFSM are ergodic (see e.g. Cambanis et al. (1987)). Since $\overline{q} = \alpha$, sharp bounds on the support of the spectrum follow from Theorem 8.

Inverse stable subordinator

Lévy process $\{Y(t), t \geq 0\}$ such that $Y(1) \sim S_\alpha(\sigma, 1, 0)$, $0 < \alpha < 1$ is called the stable subordinator. It is nondecreasing and $1/\alpha$-sssi. The inverse stable subordinator $\{X(t), t \geq 0\}$ is defined as

$$X(t) = \inf \{s > 0 : Y(s) > t\}.$$

It is α-ss with dependent, nonstationary increments, nondecreasing and corresponds to the first passage time of the stable subordinator strictly above level t. For more details see Meerschaert & Straka (2013) and references therein.

The application of the results of the previous section for the inverse stable subordinator is not straightforward as it has nonstationary increments, yet we can prove that it has a trivial spectrum such that $d(\alpha) = 1$.

To derive the lower bound we use Theorem 5. First recall that $a^\alpha + b^\alpha \leq (a+b)^\alpha$ for $a, b \geq 0$ and $\alpha \in (0,1)$. Taking $a = t - s$, $b = s$ when $t \geq s$ and $a = t$, $b = s - t$ when $t < s$ gives that $|t^\alpha - s^\alpha| \leq |t - s|^\alpha$. Since $\{X(t)\}$ has finite moments of every positive order we have for arbitrary $q > 0$ and $t, s > 0$

$$E|X(t) - X(s)|^q = |t^\alpha - s^\alpha|^q E|X(1)|^q \leq E|X(1)|^q |t-s|^{1+\alpha q - 1}.$$

By Theorem 5 there exists modification which is a.s. locally Hölder continuous of order $\gamma < \alpha - 1/q$. Since q can be taken arbitrarily large, we can get the modification such that a.s. $H(t) \geq \alpha$ for every $t \geq 0$.

For the upper bound we use Theorem 7. Given $\gamma > \alpha$ we choose $m \in \mathbb{N}$ such that $m > 1/(\gamma - \alpha)$. If $\{Y(t)\}$ is the corresponding stable subordinator, from the property $\{X(t) \leq a\} = \{Y(a) \geq t\}$ we have for every $t_1 < t_2$ and $a > 0$

$$\{X(t_2) - X(t_1) \leq a\} = \{Y_{X(t_1)+a} \geq t_2\} = \{Y_{X(t_1)+a} - t_1 \geq t_2 - t_1\}.$$

By (Bertoin 1998, Theorem 4, p. 77), for every $t_1 > 0$, $P(Y_{X(t_1)} > t_1) = 1$, thus, on this event

$$\{Y_{X(t_1)+a} - t_1 \geq t_2 - t_1\} \subset \{Y_{X(t_1)+a} - Y_{X(t_1)} \geq t_2 - t_1\}.$$

Now by the strong Markov property choosing t small enough and stationarity of increments of $\{Y(t)\}$ we have

$$\begin{aligned}
&P\left(\max_{l=1,\dots,m} |X(s+lt) - X(s+(l-1)t)| \leq Kt^\gamma\right) \\
&= P\left(X(s+t) - X(s) \leq Kt^\gamma, \dots, X(s+mt) - X(s+(m-1)t) \leq Kt^\gamma\right) \\
&\leq P\left(Y_{X(s)+Kt^\gamma} - Y_{X(s)} \geq t, \dots, Y_{X(s+(m-1)t)+Kt^\gamma} - Y_{X(s+(m-1)t)} \geq t\right) \\
&\leq \left(P\left(Y(Kt^\gamma) \geq t\right)\right)^m = \left(P\left(Y(1) \geq K^{-\frac{1}{\alpha}} t^{1-\frac{\gamma}{\alpha}}\right)\right)^m \leq \left(Ct^{\gamma-\alpha}\right)^m,
\end{aligned}$$

by the regular variation of the tail for t sufficiently small. Due to the choice of m, $m(\gamma - \alpha) > 1$. This property of the first-passage process has been noted in (Bertoin 1998, p. 96).

5.4.2 Lévy processes

Suppose $\{X(t), t \geq 0\}$ is a Lévy process. The Lévy processes in general do not satisfy the moment scaling of the form (4.5). As there is no exact moment scaling, Propositions 1 and 2 cannot be applied. Thus, in order to establish bounds on the support of the spectrum we use other criteria from Section 5.2. We present two analytically tractable examples to illustrate the use of these criteria.

Inverse Gaussian Lévy process

The inverse Gaussian Lévy process is a subordinator such that $X(1)$ has an inverse Gaussian distribution $IG(\delta, \lambda)$, $\delta > 0, \lambda \geq 0$, given by the density

$$f(x) = \frac{\delta}{\sqrt{2\pi}} e^{\delta\lambda} x^{-3/2} \exp\left\{-\frac{1}{2}\left(\frac{\delta^2}{x} + \lambda^2 x\right)\right\}, \quad x > 0.$$

The expression for the cumulant reveals that for each t, $X(t)$ has $IG(t\delta, \lambda)$ distribution. Lévy measure is absolutely continuous with the density given by

$$g(x) = \frac{\delta}{\sqrt{2\pi}} x^{-3/2} \exp\left\{-\frac{\lambda^2 x}{2}\right\}, \quad x > 0,$$

thus, the BG index is $\beta = 1/2$. See Eberlein & v. Hammerstein (2004) for more details. Inverse Gaussian distribution has moments of every order finite and for every $q \in \mathbb{R}$ we can express them as

$$\begin{aligned} E|X(1)|^q &= \int_0^\infty x^q f(x) dx = \frac{\delta}{\sqrt{2\pi}} e^{\delta\lambda} \left(\frac{2}{\lambda^2}\right)^{q-1/2} \int_0^\infty x^{q-3/2} \exp\left\{-x - \frac{\delta^2\lambda^2}{4x}\right\} dx \\ &= \frac{\delta}{\sqrt{2\pi}} e^{\delta\lambda} \left(\frac{2}{\lambda^2}\right)^{q-1/2} K_{-q+\frac{1}{2}}(\delta\lambda) 2 \left(\frac{\delta\lambda}{2}\right)^{q-\frac{1}{2}} = \sqrt{\frac{2}{\pi}} e^{\delta\lambda} \delta^{q+\frac{1}{2}} \lambda^{-q+\frac{1}{2}} K_{-q+\frac{1}{2}}(\delta\lambda), \end{aligned}$$

where we have used (Olver et al. 2010, Equation 10.32.10) and K_ν denotes the modified Bessel function of the second kind. This implies that

$$E|X(t)|^q = \sqrt{\frac{2}{\pi}} e^{t\delta\lambda} t^{q+\frac{1}{2}} \delta^{q+\frac{1}{2}} \lambda^{-q+\frac{1}{2}} K_{-q+\frac{1}{2}}(t\delta\lambda) \sim C t^{q+\frac{1}{2}} t^{-|-q+\frac{1}{2}|}, \quad \text{as } t \to 0,$$

since $K_\nu(z) \sim \frac{1}{2}\Gamma(\nu)(\frac{1}{2}z)^{-\nu}$ for $z > 0$ and $K_{-\nu}(z) = K_\nu(z)$. For any choice of $\gamma > 0$ condition of Corollary 1 cannot be fulfilled, so the best we can say is that the lower bound is 0, in accordance with (4.14). Since negative order moments are finite, Corollary 2 yields the sharp upper bound on the spectrum. Indeed, given $\gamma > 1/\beta = 2$ we have for $q < 1/(2-\gamma) < 0$

$$P\left(|X(t)| \leq K t^\gamma\right) \leq \frac{E|X(t)|^q}{K^q t^{\gamma q}} \leq C t^{-q(\gamma-2)},$$

for t sufficiently small. It follows that the upper bound is 2 which is exactly the reciprocal of the BG index.

Tempered stable subordinator

The positive tempered stable distribution is obtained by exponentially tilting the Lévy density of the α-stable distribution, $0 < \alpha < 1$. The tempered stable subordinator is a Lévy process $\{X(t)\}$ such that $X(1)$ has a positive tempered stable distribution given by the cumulant function

$$\Phi(\theta) = \log E\left[e^{-\theta X(1)}\right] = \delta\lambda - \delta\left(\lambda^{1/\alpha} + 2\theta\right)^{\alpha}, \quad \theta \geq 0,$$

where δ is the scale parameter of the stable distribution and λ is the tilt parameter. In this case BG index is equal to α (see Schoutens (2003) for more details). We use Corollary 2 for $\gamma > \alpha$ to get

$$P\left(|X(t)| \leq Kt^{\gamma}\right) \leq eE\left[e^{-\frac{X(t)}{Kt^{\gamma}}}\right] = e^{1+t\Phi(K^{-1}t^{-\gamma})} = O(e^{-t^{1-\gamma/\alpha}}), \quad \text{as } t \to 0.$$

As this decays faster than any power of t as $t \to 0$, the upper bound follows.

5.4.3 Multifractal processes

In this subsection, results are applied to examples of processes that are multifractal by Definition 1 and 2.

Log-normal cascade

We first analyze the log-normal cascade process. The spectrum of the LNC is supported on the interval $\left[1 + 2\lambda^2 - 2\sqrt{2\lambda^2}, 1 + 2\lambda^2 + 2\sqrt{2\lambda^2}\right]$, as follows from (4.16).

The condition $\tau(q) > 1$ of Proposition 1 yields $q \in (1, 1/(2\lambda^2))$. We then get that

$$H^{-} = 1 + 2\lambda^2 - 2\sqrt{2\lambda^2}.$$

This is exactly the left endpoint of the interval where the spectrum of the cascade is finite, in accordance with Proposition 1. This maximal lower bound is achieved for $q = 1/\sqrt{2\lambda^2}$, which by (4.24) is exactly the point q_0^{+} at which the linearization effect occurs. If q^{-} is the point at which the maximal lower bound H^{-} is achieved, then

$$\left(\frac{\tau(q)}{q} - \frac{1}{q}\right)' = \frac{1}{q^2}\left(q\tau'(q) - \tau(q) + 1\right)$$

must be equal to 0 at q^{-}. This is exactly defined in (4.23). Although the range of finite moments is not relevant for computing H^{-} in this case, in general it can depend on $\overline{q}$.

For the upper bound, since all negative order moments are finite we get that

$$\widetilde{H^+} = H^+ = 1 + 2\lambda^2 + 2\sqrt{2\lambda^2}$$

achieved for $q = -1/\sqrt{2\lambda^2}$, which is equal to q_0^- from (4.25). Thus, again, the bound from Proposition 2 is sharp, giving the right endpoint of the interval where the spectrum is supported.

Log-normal multifractal random walk

The example of LNMRW illustrates that we may have $\widetilde{H^+} \neq H^+$ and that the definition of the scaling function needs to be adjusted to avoid infinite moments of negative order. The scaling function is defined for the range of finite moments, which is $(-1, 1/\lambda^2)$ as explained earlier. However, if we consider τ as the scaling function of the random factor $M(c) = c^{1/2}e^{\Gamma_c}$ defined in (4.8), then the definition of τ makes sense for all $q \in (-\infty, 1/\lambda^2)$. The spectrum is supported on the interval $\left[1/2 + \lambda^2 - \sqrt{2\lambda^2}, 1/2 + \lambda^2 + \sqrt{2\lambda^2}\right]$ and given by (4.17).

The random factor $M(c)$ is the source of multifractality, has the same scaling function (4.9), but all negative order moments are finite. Thus we get

$$\begin{aligned} H^- &= 1/2 + \lambda^2 - \sqrt{2\lambda^2}, \\ \widetilde{H^+} &= \frac{3}{2} + \frac{3\lambda^2}{2}, \\ H^+ &= 1/2 + \lambda^2 + \sqrt{2\lambda^2}. \end{aligned}$$

H^- and H^+ give the sharp bounds, while $\widetilde{H^+}$ is affected by the divergence of negative order moments. This shows that when the multifractal process has infinite negative order moments, one should specify scaling in terms of the random factor. The optimal bounds H^- and H^+ are attained for $q^- = \sqrt{2/\lambda^2}$ and $q^+ = -\sqrt{2/\lambda^2}$, respectively.

5.5 Robust version of the partition function

In Section 5.2 using Corollary 3 we managed to avoid the problematic infinite moments of negative order and prove results like Theorem 8 and Corollary 5. When the scaling function (1.4) is estimated from the data, spurious concavity may appear for negative values of q, due to the effect of diverging negative order moments. We use the idea of Corollary 3 to develop a more robust version of the partition function.

Instead of using increments in the partition function (1.2), we can use the maximum of some fixed number m of the same length increments. This will make the negative order moments finite for some reasonable range and prevent divergencies. The underlying idea resembles the wavelet leaders method, where leaders are formed as the maxima of the wavelet coefficients over some time scale (see Jaffard et al. (2007)). Since m is fixed, this does not affect the true scaling. The same idea can be used for $q > 0$ by an argument following from Corollary 1. It is important to stress that the estimation of the scaling function makes sense only if the underlying process is known to possess scaling property of the type (4.3).

Suppose $\{X(t)\}$ has stationary increments and $X(0) = 0$. Divide the interval $[0, T]$ into $\lfloor T/(mt) \rfloor$ blocks each consisting of m increments of length t and define the modified partition function:

$$\widetilde{S}_q(T,t) = \frac{1}{\lfloor T/(mt) \rfloor} \sum_{i=1}^{\lfloor T/(mt) \rfloor} \max_{l=1,\dots,m} |X(imt + lt) - X(imt + (l-1)t)|^q . \tag{5.10}$$

One can see $\widetilde{S}_q(T,t)$ as a natural estimator of the moment in (5.9). Analogously we define the modified scaling function as in (1.4) by using $\widetilde{S}_q(n, t_i)$:

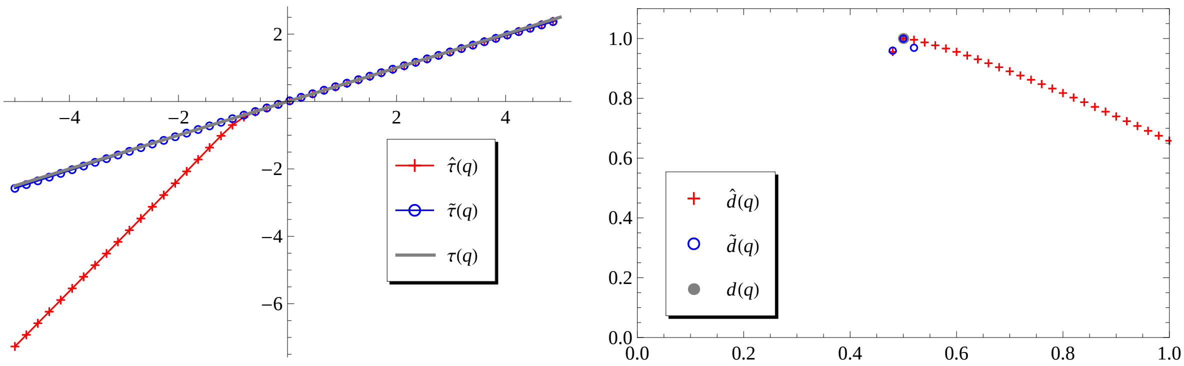
$$\widetilde{\tau}_{N,T}(q) = \frac{\sum_{i=1}^{N} \ln t_i \ln \widetilde{S}_q(n,t_i) - \frac{1}{N} \sum_{i=1}^{N} \ln t_i \sum_{j=1}^{N} \ln \widetilde{S}_q(n,t_i)}{\sum_{i=1}^{N} (\ln t_i)^2 - \frac{1}{N} \left(\sum_{i=1}^{N} \ln t_i \right)^2}. \tag{5.11}$$

The definition can be altered only for $q < 0$ although there is no much difference between two forms when $q > 0$.

To illustrate how this modification makes the scaling function more robust we present several examples comparing (1.4) and (5.11). We generate sample paths of several processes and estimate the scaling function by both methods. We also estimate the spectrum numerically using (4.18). The results are shown in Figures 5.1-5.4. Each figure shows the estimated scaling functions and the estimated spectrum by using standard definition (1.4) and by using (5.11). We also added the plots of the scaling function that would yield the correct spectrum via multifractal formalism and the true spectrum of the process.

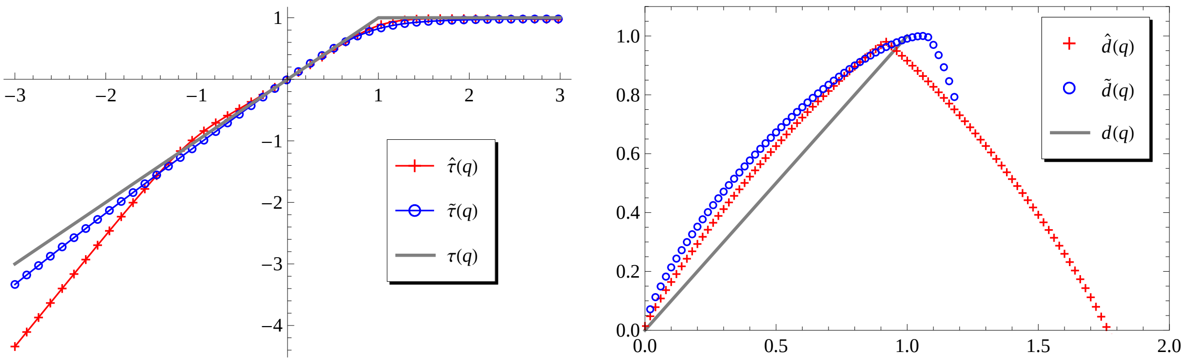
For the BM (Figure 5.1) and the α-stable Lévy process (Figure 5.2) we generated sample paths of length 10000 and used $\alpha = 1$ for the latter. LFSM (Figure 5.3) was generated using $H = 0.9$ and $\alpha = 1.2$ with path length 15784. Finally, the sample path of the LNMRW of length 10000 was generated with $\lambda^2 = 0.025$ (Figure 5.4). For each case we take $m = 20$ in defining the modified partition function (5.10).

In all the examples considered, the modified scaling function is capable of yielding the correct spectrum of the process with the multifractal formalism. As opposed to the standard definition, it is unaffected by diverging negative order moments. Moreover, it captures the divergence of positive order moments which determines the shape of the spectrum.



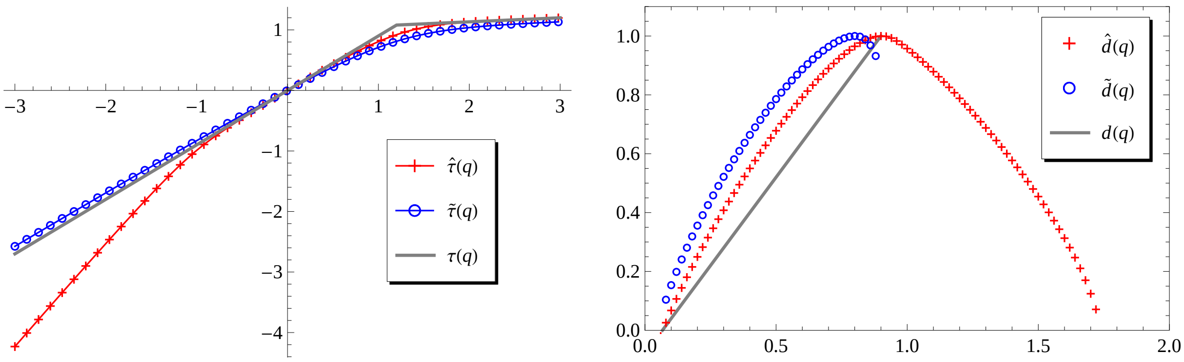
(a) Scaling functions (b) Spectrum

Figure 5.1: Brownian motion



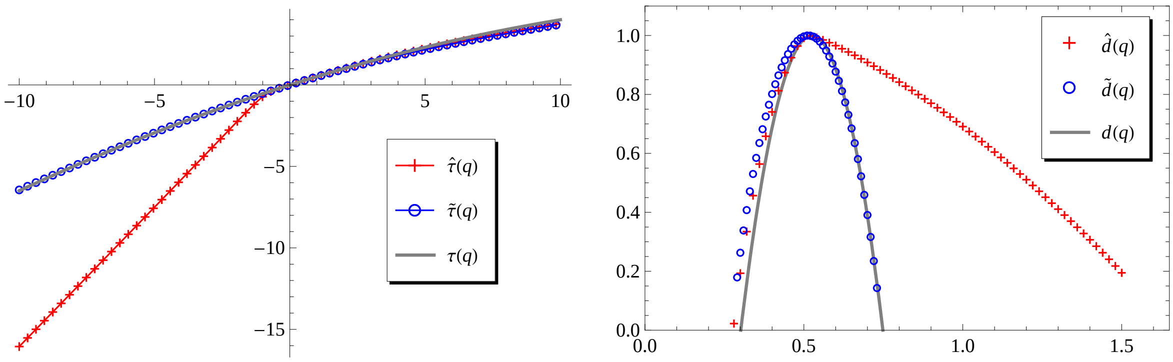
(a) Scaling functions (b) Spectrum

Figure 5.2: Stable Lévy process $\alpha = 1$



(a) Scaling functions (b) Spectrum

Figure 5.3: LFSM $H = 0.9$, $\alpha = 1.2$



(a) Scaling functions

(b) Spectrum

Figure 5.4: LNMRW $\lambda^2 = 0.025$

Bibliography

Abry, P., Chainais, P., Coutin, L. & Pipiras, V. (2009), 'Multifractal random walks as fractional Wiener integrals', *Information Theory, IEEE Transactions on* **55**(8), 3825–3846.

Alsmeyer, G., Biggins, J. & Meiners, M. (2012), 'The functional equation of the smoothing transform', *The Annals of Probability* **40**(5), 2069–2105.

Anh, V. V., Leonenko, N. N. & Shieh, N.-R. (2008), 'Multifractality of products of geometric Ornstein-Uhlenbeck-type processes', *Advances in Applied Probability* **40**(4), 1129–1156.

Anh, V. V., Leonenko, N. N. & Shieh, N.-R. (2009*a*), 'Multifractal products of stationary diffusion processes', *Stochastic Analysis and Applications* **27**(3), 475–499.

Anh, V. V., Leonenko, N. N. & Shieh, N.-R. (2009*b*), 'Multifractal scaling of products of birth-death processes', *Bernoulli* **15**(2), 508–531.

Anh, V. V., Leonenko, N. N., Shieh, N.-R. & Taufer, E. (2010), 'Simulation of multifractal products of Ornstein-Uhlenbeck type processes', *Nonlinearity* **23**(4), 823–843.

Avram, F., Leonenko, N. & Šuvak, N. (2013), 'On spectral analysis of heavy-tailed Kolmogorov-Pearson diffusions', *Markov Processes and Related Fields* **19**(2), 249–298.

Ayache, A. & Hamonier, J. (2012), 'Linear fractional stable motion: A wavelet estimator of the α parameter', *Statistics & Probability Letters* **82**(8), 1569–1575.

Ayache, A. & Hamonier, J. (2013), 'Linear multifractional stable motion: wavelet estimation of $H(\cdot)$ and α parameters', *arXiv preprint arXiv:1304.2995* .

Bacry, E., Delour, J. & Muzy, J.-F. (2001), 'Multifractal random walk', *Physical Review E* **64**(2), 026103.

Bacry, E., Gloter, A., Hoffmann, M. & Muzy, J. F. (2010), 'Multifractal analysis in a mixed asymptotic framework', *The Annals of Applied Probability* **20**(5), 1729–1760.

Bacry, E., Kozhemyak, A. & Muzy, J. (2013), 'Log-normal continuous cascade model of asset returns: aggregation properties and estimation', *Quantitative Finance* **13**(5), 795–818.

Bacry, E., Kozhemyak, A. & Muzy, J.-F. (2008), 'Continuous cascade models for asset returns', *Journal of Economic Dynamics and Control* **32**(1), 156–199.

Bacry, E. & Muzy, J. F. (2003), 'Log-infinitely divisible multifractal processes', *Communications in Mathematical Physics* **236**(3), 449–475.

Balança, P. (2014), 'Fine regularity of Lévy processes and linear (multi)fractional stable motion', *Electronic Journal of Probability* **19**(101), 1–37.

Barral, J., Fournier, N., Jaffard, S. & Seuret, S. (2010), 'A pure jump Markov process with a random singularity spectrum', *The Annals of Probability* **38**(5), 1924–1946.

Barral, J. & Jin, X. (2014), 'On exact scaling log-infinitely divisible cascades', *Probability Theory and Related Fields* **160**(3-4), 521–565.

Barral, J. & Mandelbrot, B. B. (2002), 'Multifractal products of cylindrical pulses', *Probability Theory and Related Fields* **124**(3), 409–430.

Barral, J. & Seuret, S. (2007), 'The singularity spectrum of Lévy processes in multifractal time', *Advances in Mathematics* **214**(1), 437–468.

Barunik, J., Aste, T., Di Matteo, T. & Liu, R. (2012), 'Understanding the source of multifractality in financial markets', *Physica A: Statistical Mechanics and its Applications* **391**(17), 4234–4251.

Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. (2006), *Statistics of extremes: theory and applications*, Wiley, New York.

Bertoin, J. (1998), *Lévy processes*, Cambridge University Press, Cambridge.

Bibby, B. M., Skovgaard, I. M. & Sørensen, M. (2005), 'Diffusion-type models with given marginal distribution and autocorrelation function', *Bernoulli* **19**(2), 191–220.

Calvet, L. E. & Fisher, A. J. (2004), 'How to forecast long-run volatility: Regime switching and the estimation of multifractal processes', *Journal of Financial Econometrics* **2**(1), 49–83.

Calvet, L. E. & Fisher, A. J. (2008), *Multifractal volatility: theory, forecasting, and pricing*, Academic Press, New York.

Calvet, L. & Fisher, A. (2002), 'Multifractality in asset returns: theory and evidence', *Review of Economics and Statistics* **84**(3), 381–406.

Cambanis, S., Hardin Jr, C. D. & Weron, A. (1987), 'Ergodic properties of stationary stable processes', *Stochastic Processes and their Applications* **24**(1), 1–18.

Chainais, P., Riedi, R. & Abry, P. (2005), 'On non-scale-invariant infinitely divisible cascades', *Information Theory, IEEE Transactions on* **51**(3), 1063–1083.

De Haan, L. & Ferreira, A. (2007), *Extreme value theory: an introduction*, Springer, New York.

Dekkers, A., Einmahl, J. & De Haan, L. (1989), 'A moment estimator for the index of an extreme-value distribution', *The Annals of Statistics* **17**(4), 1833–1855.

Doob, J. L. (1953), *Stochastic processes*, Wiley, New York.

Doukhan, P. (1994), *Mixing: properties and examples*, Springer, New York.

Eberlein, E. & v. Hammerstein, E. (2004), Generalized hyperbolic and inverse Gaussian distributions: Limiting cases and approximation of processes, *in* R. Dalang, M. Dozzi & F. Russo, eds, 'Seminar on stochastic analysis, random fields and applications IV', Vol. 58 of *Progress in Probability*, Birkhäuser, Basel, pp. 221–264.

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997), *Modelling extremal events: for insurance and finance*, Vol. 33, Springer, Berlin.

Embrechts, P. & Maejima, M. (2002), *Selfsimilar processes*, Princeton University Press, Princeton, NJ.

Falconer, K. (2003), *Fractal geometry: mathematical foundations and applications*, 2nd edn, Wiley, Chichester.

Fisher, A., Calvet, L. & Mandelbrot, B. B. (1997), Multifractality of Deutschemark/US Dollar exchange rates, Cowles Foundation Discussion Papers 1165, Cowles Foundation for Research in Economics, Yale University.

Frisch, U. & Parisi, G. (1985), On the singularity structure of fully developed turbulence, *in* M. Gil, R. Benzi & G. Parisi, eds, 'Turbulence and predictability in geophysical fluid dynamics and climate dynamics', Elsevier, Amsterdam, North-Holland, pp. 84–88.

Gnedenko, B. & Kolmogorov, A. (1968), *Limit distributions for sums of independent random variables*, Addison-Wesley, Reading, MA.

Gonçalves, P. & Riedi, R. (2005), 'Diverging moments and parameter estimation', *Journal of the American Statistical Association* **100**(472), 1382–1393.

Hall, P. & Heyde, C. C. (1980), *Martingale limit theory and its application*, Academic Press, New York.

Heyde, C. C. (2009), 'Scaling issues for risky asset modelling', *Mathematical Methods of Operations Research* **69**(3), 593–603.

Heyde, C. C. & Leonenko, N. (2005), 'Student processes', *Advances in Applied Probability* **37**(2), 342–365.

Heyde, C. C. & Liu, S. (2001), 'Empirical realities for a minimal description risky asset model. The need for fractal features', *Journal of the Korean Mathematical Society* **38**, 1047–1059.

Heyde, C. & Sly, A. (2008), 'A cautionary note on modeling with fractional Lévy flights', *Physica A: Statistical Mechanics and its Applications* **387**(21), 5024–5032.

Hill, B. M. (1975), 'A simple general approach to inference about the tail of a distribution', *The Annals of Statistics* **3**(5), 1163–1174.

Hsing, T. (1991), 'On tail index estimation using dependent data', *The Annals of Statistics* **19**(3), 1547–1569.

Hurst, S. R. & Platen, E. (1997), The marginal distributions of returns and volatility, *in* Y. Dodge, ed., 'L_1-statistical procedures and related topics', Vol. 31 of *Lecture Notes–Monograph Series*, Institute of Mathematical Statistics, Hayward, CA, pp. 301–314.

Iacus, S. M. (2008), *Simulation and inference for stochastic differential equations: with R examples*, Springer, New York.

Jaffard, S. (1996), 'Old friends revisited: the multifractal nature of some classical functions', *Journal of Fourier Analysis and Applications* **3**(1), 1–22.

Jaffard, S. (1997*a*), 'Multifractal formalism for functions part I: results valid for all functions', *SIAM Journal on Mathematical Analysis* **28**(4), 944–970.

Jaffard, S. (1997*b*), 'Multifractal formalism for functions part II: self-similar functions', *SIAM Journal on Mathematical Analysis* **28**(4), 971–998.

Jaffard, S. (1999), 'The multifractal nature of Lévy processes', *Probability Theory and Related Fields* **114**(2), 207–227.

Jaffard, S. (2000), 'On the Frisch–Parisi conjecture', *Journal de Mathématiques Pures et Appliquées* **79**(6), 525–552.

Jaffard, S., Abry, P. & Wendt, H. (2014), Irregularities and scaling in signal and image processing: Multifractal analysis, *in* M. Frame & N. Cohen, eds, 'Benoit Mandelbrot: A Life in Many Dimensions', World Scientific Publishing, Singapore.

Jaffard, S., Lashermes, B. & Abry, P. (2007), Wavelet leaders in multifractal analysis, *in* T. Qian, M. Vai & Y. Xu, eds, 'Wavelet analysis and applications', Applied and Numerical Harmonic Analysis, Birkhäuser Basel, pp. 201–246.

Kahane, J.-P. & Peyriere, J. (1976), 'Sur certaines martingales de Benoit Mandelbrot', *Advances in mathematics* **22**(2), 131–145.

Kallenberg, O. (2002), *Foundations of modern probability*, Springer, New York.

Karasaridis, A. & Hatzinakos, D. (2001), 'Network heavy traffic modeling using α-stable self-similar processes', *Communications, IEEE Transactions on* **49**(7), 1203–1214.

Karatzas, I. & Shreve, S. E. (1991), *Brownian motion and stochastic calculus*, Springer, New York.

Leadbetter, M. R., Lindgren, G. & Rootzén, H. (1982), *Extremes and related properties of random sequences and processes*, Springer, Berlin.

Leland, W. E., Taqqu, M. S., Willinger, W. & Wilson, D. V. (1994), 'On the self-similar nature of ethernet traffic (extended version)', *Networking, IEEE/ACM Transactions on* **2**(1), 1–15.

Leland, W. E. & Wilson, D. V. (1991), High time-resolution measurement and analysis of LAN traffic: Implications for LAN interconnection, *in* 'Tenth annual joint conference of the IEEE computer and communications societies', IEEE, Bal Harbour, FL, pp. 1360–1366.

Leonenko, N. N., Petherick, S. & Sikorskii, A. (2011), 'The Student subordinator model with dependence for risky asset returns', *Communications in Statistics - Theory and Methods* **40**(19-20), 3509–3523.

Leonenko, N. N. & Šuvak, N. (2010), 'Statistical inference for Student diffusion process', *Stochastic Analysis and Applications* **28**(6), 972–1002.

Lovejoy, S. & Schertzer, D. (2013), *The weather and climate: emergent laws and multifractal cascades*, Cambridge University Press, Cambridge.

Ludeña, C. (2008), 'L_p-variations for multifractal fractional random walks', *The Annals of Applied Probability* **18**(3), 1138–1163.

Magdziarz, M. (2009), 'Correlation cascades, ergodic properties and long memory of infinitely divisible processes', *Stochastic Processes and their Applications* **119**(10), 3416–3434.

Mandelbrot, B. (1963), 'The variation of certain speculative prices', *The Journal of Business* **36**(4), 394–419.

Mandelbrot, B. (1967), 'The variation of some other speculative prices', *The Journal of Business* **40**(4), 393–413.

Mandelbrot, B. B. (1972), Possible refinement of the lognormal hypothesis concerning the distribution of energy dissipation in intermittent turbulence, *in* M. Rosenblatt & C. Van Atta, eds, 'Statistical models and turbulence', Vol. 12 of *Lecture Notes in Physics*, Springer, Berlin, pp. 333–351.

Mandelbrot, B. B. (1974), 'Intermittent turbulence in self-similar cascades - divergence of high moments and dimension of the carrier', *Journal of Fluid Mechanics* **62**(2), 331–358.

Mandelbrot, B. B. (1997), *Fractals and scaling in finance: discontinuity, concentration, risk*, Springer, New York.

Mandelbrot, B., Fisher, A. & Calvet, L. (1997), A multifractal model of asset returns, Cowles Foundation Discussion Papers 1164, Cowles Foundation for Research in Economics, Yale University.

Masuda, H. (2004), 'On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process', *Bernoulli* **10**(1), 97–120.

Meerschaert, M. M. & Scheffler, H. P. (1998), 'A simple robust estimation method for the thickness of heavy tails', *Journal of Statistical Planning and Inference* **71**(1), 19–34.

Meerschaert, M. M. & Straka, P. (2013), 'Inverse stable subordinators', *Mathematical Modelling of Natural Phenomena* **8**(02), 1–16.

Muzy, J.-F. & Bacry, E. (2002), 'Multifractal stationary random measures and multifractal random walks with log-infinitely divisible scaling laws', *Physical Review E* **66**(5), 056121.

Muzy, J.-F., Bacry, E. & Arneodo, A. (1993), 'Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method', *Physical Review E* **47**(2), 875.

Muzy, J.-F., Bacry, E. & Kozhemyak, A. (2006), 'Extreme values and fat tails of multifractal fluctuations', *Physical Review E* **73**(6), 066114.

Olver, F. W. J., Lozier, D. W., Boisvert, R. F. & Clark, C. W., eds (2010), *NIST Handbook of Mathematical Functions*, Cambridge University Press, New York.

Ossiander, M. & Waymire, E. C. (2000), 'Statistical estimation for multiplicative cascades', *The Annals of Statistics* **28**(6), 1533–1560.

Pipiras, V. & Taqqu, M. S. (2002), 'The structure of self-similar stable mixed moving averages', *Annals of Probability* **30**(2), 898–932.

Pipiras, V., Taqqu, M. S. & 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**(4), 1091–1123.

Resnick, S. I. (2007), *Heavy-tail phenomena: probabilistic and statistical modeling*, Springer, New York.

Riedi, R. H. (1995), 'An improved multifractal formalism and self-similar measures', *Journal of Mathematical Analysis and Applications* **189**(2), 462–490.

Riedi, R. H. (2003), Multifractal processes, *in* P. Doukhan, G. Oppenheim & M. S. Taqqu, eds, 'Theory and applications of long-range dependence', Birkhäuser Basel, pp. 625–716.

Riedi, R. H., Crouse, M. S., Ribeiro, V. J. & Baraniuk, R. G. (1999), 'A multifractal wavelet model with application to network traffic', *Information Theory, IEEE Transactions on* **45**(3), 992–1018.

Samorodnitsky, G. (2004), 'Extreme value theory, ergodic theory and the boundary between short memory and long memory for stationary stable processes', *The Annals of Probability* **32**(2), 1438–1468.

Samorodnitsky, G. (2007), 'Long range dependence', *Foundations and Trends in Stochastic Systems* **1**(3), 163–257.

Samorodnitsky, G. & Taqqu, M. S. (1994), *Stable non-Gaussian random processes: stochastic models with infinite variance*, Chapman & Hall, New York.

Schoutens, W. (2003), *Lévy processes in finance: pricing financial derivatives*, Wiley, New York.

Sly, A. (2005), 'Self-similarity, multifractionality and multifractality', *MPhil Thesis, Australian National University* .

Stanislavsky, A., Burnecki, K., Magdziarz, M., Weron, A. & Weron, K. (2009), 'FARIMA modeling of solar flare activity from empirical time series of soft X-ray solar emission', *The Astrophysical Journal* **693**(2), 1877.

Stoev, S. & Taqqu, M. (2003), Wavelet estimation for the Hurst parameter in stable processes, *in* G. Rangarajan & M. Ding, eds, 'Processes with long-range correlations', Vol. 621 of *Lecture Notes in Physics*, Springer Berlin Heidelberg, pp. 61–87.

Stoev, S. & Taqqu, M. S. (2004), 'Simulation methods for linear fractional stable motion and FARIMA using the Fast Fourier Transform', *Fractals* **12**(01), 95–121.

Stoev, S. & Taqqu, M. S. (2005), 'Asymptotic self-similarity and wavelet estimation for long-range dependent fractional autoregressive integrated moving average time series with stable innovations', *Journal of Time Series Analysis* **26**(2), 211–249.

Surgailis, D., Rosinski, J., Mandrekar, V. & Cambanis, S. (1993), 'Stable mixed moving averages', *Probability Theory and Related Fields* **97**(4), 543–558.

Takashima, K. (1989), 'Sample path properties of ergodic self-similar processes', *Osaka Journal of Mathematics* **26**(1), 159–189.

Taufer, E. & Leonenko, N. N. (2009), 'Simulation of Lévy-driven Ornstein-Uhlenbeck processes with given marginal distribution', *Computational Statistics and Data Analysis* **53**(6), 2427–2437.

Vervaat, W. (1985), 'Sample path properties of self-similar processes with stationary increments', *The Annals of Probability* **13**(1), 1–27.

Wendt, H., Jaffard, S. & Abry, P. (2012), Multifractal analysis of self-similar processes, *in* 'Statistical Signal Processing Workshop (SSP)', IEEE, Ann Arbor, MI, pp. 69–72.

Weron, A., Burnecki, K., Mercik, S. & Weron, K. (2005), 'Complete description of all self-similar models driven by Lévy stable noise', *Physical Review E* **71**(1), 016113.

Willinger, W., Paxson, V. & Taqqu, M. S. (1998), Self-similarity and heavy tails: structural modeling of network traffic, *in* R. Adler, R. Feldman & M. S. Taqqu, eds, 'A Practical Guide to Heavy Tails', Birkhäuser, Boston, pp. 27–53.

Yaglom, A. (1966), 'Effect of fluctuations in energy dissipation rate on the form of turbulence characteristics in the inertial subrange', *Doklady Akademii Nauk SSSR* **166**, 49–52.

Zolotarev, V. M. (1986), *One-dimensional stable distributions*, American Mathematical Society, Providence, RI.

Curriculum vitae

Danijel Grahovac was born on November 15, 1986 in Vukovar, Croatia. After finishing high school in Vukovar, he enroled in the undergraduate program in mathematics at the Department of Mathematics, J. J. Strossmayer University of Osijek. In September 2010, he obtained his master's degree in mathematics, Financial and Business Mathematics at the Department of Mathematics, University of Osijek. In November 2010, he enroled in the Croatian doctoral program in Mathematics at the Department of Mathematics, Faculty of Science of the University of Zagreb.

From November 2010 he has been employed as a research assistant at the Department of Mathematics, University of Osijek on the project "Statistical aspects of estimation problem in nonlinear parametric models" under the leadership of izv. prof. dr. sc. Mirta Benšić, financed by the Ministry of Science, Education and Sports of the Republic of Croatia.

During his career he visited the School of Mathematics at the University of Cardiff, UK as a visiting researcher at several occasions and also participated in other programmes of professional development. He attended four international conferences so far, giving a talk at two of them. He received a scientific award from United Kingdom Association of Alumni and Friends of Croatian Universities. In 2013 he was granted Erasmus programme scholarship for staff mobility.

List of publications:

Journal Publications:

1. Grahovac, D., Leonenko, N. N. (2014), Detecting multifractal stochastic processes under heavy-tailed effects, *Chaos, Solitons & Fractals* 66, 78-89.

2. Grahovac, D., Leonenko, N. N., Taqqu, M. S. (2014), 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.

3. Grahovac, D., Jia, M., Leonenko, N. N., Taufer, E. (2014), Asymptotic properties of the partition function and applications in tail index inference of heavy-tailed data, *Statistics*, accepted for publication

4. Grahovac, D., Leonenko, N. N. (2014), Bounds on the support of the multifractal spectrum of stochastic processes, *submitted*

Refereed Proceedings:

1. Grahovac, D., Jia, M., Leonenko, N. N., Taufer, E. (2014), On the tail index inference based on the scaling function method, *Proceedings of the 18th European Young Statisticians Meeting*, Šuvak, N. (ed.), J. J. Strossmayer University of Osijek, Department of Mathematics, 39-45.

Others:

1. Grahovac, D. (2010), Dvodimenzionalni interpolacijski spline, *Osječki matematički list* 10(1), 59-69.