Diverging moments and parameter estimation

Paulo Gonçalvès

INRIA Rhône-Alpes
on leave (2003-2004) at Univ. of Tech. Lisbon
IST-ISR, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal

Fax: +351 21 841 8291
E-mail: `paulo.goncalves@inrialpes.fr`

Rudolf Riedi *

Department of Statistics
Rice University, MS 138
Houston, TX 77005–1892, USA

Fax: +1 713/348–5476
E-mail: `riedi@rice.edu`

Abstract

Heavy tailed distributions enjoy increased popularity and become more readily applicable as the arsenal of analytical and numerical tools grows. They play key roles in modeling approaches in networking, finance, hydrology to name but a few. The tail parameter α is of central importance as it governs both the existence of moments of positive order and the thickness of the tails of the distribution. Some of the best known tail estimators such as Koutrouvelis and Hill are either parametric or show lack in robustness or accuracy. This paper develops a shift and scale invariant, non-parametric estimator for both, upper and lower bounds for orders with finite moments. The estimator builds on the equivalence between tail behavior and the regularity of the characteristic function at the origin and achieves its goal by deriving a simplified wavelet analysis which is particularly suited to characteristic functions.

Keywords: Diverging moments, heavy tailed distributions, characteristic functions, wavelet transform.

---

*Financial support of R. Riedi comes in part from NSF ANI-00099148, DARPA/AFRL F30602–00–2–0557, NSF ANI-0338856 and Texas ATP 003604-0036-2003.

1 Introduction

Heavy tailed distributions enjoy increased popularity and become more readily applicable as the arsenal of analytical and numerical tools grows. They play key roles in modeling approaches in networking, finance, hydrology to name but a few applications. Examples of interest include the stable, the Pareto and certain extreme value distributions. The tail parameter α is of importance in its own right as the central parameter for several of the mentioned distributions.

In addition, it sets the upper bound for the orders r beyond which moments $\mathbb{E}[|X|^r]$ do not exist. Indeed, recall that a random variable X is called *heavy tailed* with tail parameter α if

$$P[|X| > x] = x^{-\alpha} L(x) \tag{1}$$

where L is a slowly varying function, i.e., $L(tx)/L(x) \to 1$ as $x \to \infty$ for any $t > 0$. It is called Pareto, if $L(x)$ is a constant and (1) holds for all $|x| > \delta$. For the Pareto distribution, it is clear that moments are finite exactly up to order α, a fact that can be generalized using standard facts.

The issue of finiteness of moments is particularly pressing in view of the abundance and usefulness of moment estimators. They are not only important for parameter estimation, when the underlying distribution law is known, but also for data fitting and model selection, i.e., identifying unknown distributions from sample data. To recall but two instances, the Kurtosis statistic hypothesis test resolves Gaussianity versus non-Gaussianity, whereas for a Poisson random variable mean and variance should be equal. In addition, many applications integrate moment estimates as a crucial ingredient. That is the case in multifractal analysis, where the $q-$th order absolute moments of the increments (or the wavelet coefficients) of a process hold valuable information on the local behavior of its paths.

Pathologies emerge when moments are infinite or not defined, such as for the Cauchy distribution which has infinite second moment and undefined mean. As infinite moments may degrade the performance of estimators (possibly introducing some systematic errors) or reduce the speed of convergence to limiting laws, special attention must be dedicated to their theoretical existence. We refer once more to multifractal analysis where infinite moments may indicate phase transitions that are highly informative about the process regularity.

All this motivates the development of statistical methods to determine the finiteness of moments of a distribution given finite sample data [10], more precisely, to determine the (positive and negative) *critical order* λ^-, λ^+ of a distribution, by which we mean here

$$\begin{aligned} \lambda^+ &\triangleq \sup\{r > 0 \,:\, \mathbb{E}[|X|^r] < \infty\} \\ \lambda^- &\triangleq \inf\{r < 0 \,:\, \mathbb{E}[|X|^r] < \infty\}. \end{aligned} \tag{2}$$

Related estimators will not only provide useful for the tail parameter, but also for the analogous parameter governing the distribution around zero, or any center of choice μ after a translation $X \mapsto X_\mu$. To this end, we propose an approach that combines two facts.

First, the characteristic function $\phi(u) \triangleq \mathbb{E}[\exp(iuX)]$, being the Fourier transform of the distribution of X, has as many continuous derivatives at zero as the distribution has finite moments of positive orders. In particular, for even k we have $\phi^{(k)}(0) = i^k \mathbb{E}[X^k]$ whenever one of the two is defined [21]. A crucial ingredient to our methodology is a more general relation of this sort. To this end, we resort to the concept of the *characteristic exponent* ρ^+ of a distribution by which we mean the (generalized) degree of Lipschitz continuity of the real part of the characteristic function at the origin. Provided in lies in $(0, 2)$ the characteristic exponent can be written simply as

$$\rho^+ = \sup\{r > 0 \,:\, 1 - \operatorname{Re}\phi(u) = O(u^r) \text{ as } u \to 0^+\} \tag{3}$$

It follows from basic known facts that

$$\lambda^+ = \rho^+ \tag{4}$$

as long as these values lie between 0 and 2. Estimation of the critical exponent can then be achieved via the regularity of ϕ. Replacing the random variable X by $1/X$ we find ρ^- and an estimate for λ^-.

Part of the paper will address the extension of this approach for orders larger than 2 from the estimation point of view; the mathematical foundation of this extensions is developed in a companion paper [24] and can also be found in [9]. While effective for model selection, the characteristic exponent provides, however, only bounds for the critical order in that case: $\rho^+ \leq \lambda^+ \leq \lfloor \rho^+ \rfloor + 1$ (see corollary 6).

Second, having reduced the problem at hand to an estimation of local regularity, it proves effective to leverage the power of the wavelet transform. In a nutshell, the decay of the wavelet coefficients of a function $W[g](u,s)$ for u close to 0 provides quite precise information on the local regularity of the function g at 0. As will be established, this wavelet analysis becomes particularly simple for a characteristic function ϕ and requires only the knowledge of $W[\phi](0,s)$.

In summary, the recipe of our estimator for ρ^+ is dramatically simple:

- From the sample data set $\{X_n,\ n=1,\ldots,N\}$ compute solely the wavelet coefficients at zero of the empirical characteristic function $\widehat{\Phi}(u) = N^{-1}\sum_n \exp\{iuX_n\}$, i.e., $W[\widehat{\Phi}](0,s)$; as it appears, this amounts to computing the non-parametric un-biased kernel estimator $\widehat{W}(s)$:

$$\widehat{W}(s) = \frac{1}{N}\sum_{k=1}^{N} \Psi(s{\cdot}X_k) \tag{5}$$

  where the kernel Ψ is a the Fourier transform of a semi-definite wavelet (see text).

- The estimators of the two characteristic exponents, i.e., $\widehat{\rho^+}$ and $\widehat{\rho^-}$ are obtained from simple linear regressions of $\log \widehat{W}(s)$ against $\log s$ within some predefined scale intervals. These estimators are scale-invariant, can be made shift-invariant, and are asymptotically un-biased.

- Since wavelets can not capture regularities higher than their own regularity N_ψ, the procedure should be repeated with wavelets of increasing regularity (reasonably up to $N_\psi = 4$).

We will demonstrate the effectiveness of this estimator looking at symmetrical stable distributions in comparison with well established estimators such as Koutrouvelis'. Recall that stable distributions appear as limiting distributions of properly renormalized sums of iid random variables with (possibly) infinite variance. The symmetrical stable laws are defined by

$$\phi_X(u) = \mathbb{E}[\exp(iuX)] = \exp(-\sigma^\alpha |u|^\alpha + i\mu u) \tag{6}$$

and their heavy tail parameter is known to be equal to α [25]. Combining this with the fact that their densities, though not explicitly known, are symmetrical and uni-modal, they possess finite absolute moments of order r exactly for $r \in (-1, \alpha)$. On the other hand, the Taylor expansion of $\exp(\cdot)$ implies readily that $\operatorname{Re}\phi(u) = 1 - \sigma^\alpha |u|^\alpha + O(u^2)$, which verifies (4).

2 Background

In this section we collect well known facts on the existence of moments as well as the wavelet analysis of irregular signals.

2.1 Tail Estimators

Most well-known tests for the existence of moments emerge as by-products of tail estimators and appear in parameterized settings. For instance, Nolan in [20] proposed a maximum likelihood estimator for general alpha-stable laws (including Gaussian and Cauchy) based on a large sample data set. As no closed form exists for these distribution functions (aside from some particular rational values of the parameters), he proposes an efficient numerical resolution of the maximum likelihood equation.

Previously, Koutrouvelis [16] and Mc Culloch [18, 19], among many others, have proposed two different estimators of the parameters of $\alpha-$stable laws, based either on Pareto approximation for $\alpha-$stables tails, or on the analytic form of the characteristic function.

More recently, Bianchi and Meerschaert [3] proposed a quadratic estimator of tail index α, based on the asymptotic of the sample variance. This robust estimator has the advantage over Hill estimator [11], to be shift and scale invariant, and also to perform well in situation where the Hill estimator is inefficient, namely for stable distributions with $1.5 < \alpha < 2$.

Starting from a closed form for the characteristic function (recall (6)) or in some cases a numerical approximation of the density function all these methods aim at finding the minimum of the log-likelihood function, given the data. As a result, it is well known that these approaches are optimal in the sense of minimum variance and achieve the Cramér Rao bound [6, 16, 20]. However, being parametric, these estimators may perform poorly whenever the true underlying distributions do not match the model.

In this paper, we propose a *non-parametric* estimation procedure with convincing robustness properties for the characteristic exponents ρ^+ and ρ^- which do not rely on any assumption on the density model. In particular, not even the semi-parametric assumption of heavy tails (1) is made and can be tested via this approach. The resulting estimates can be used for estimating the tail parameter and the body parameter.

Notably, both exponents are estimated in the same procedure. Indeed, the problem of existing negative moments could be reformulated with a simple change of variables $x \mapsto x^{-1}$, as a positive moment existence problem. Then, we could apply our estimator to X^{-1} instead of X directly, allowing thus for determining a lower negative bound for the existence of $\int_{-\infty}^{\infty} |x|^{-r} \mathrm{d}F(x)$, $r > 0$. However, as we will demonstrate both, the positive and the negative characteristic exponent can be evaluated at once, using the same procedure applied to the same data set of i.i.d. samples $\{X_i\}_{i=1,\ldots N}$.

2.2 Characteristic Function and Moments

Let us recall a well known relation between high order moments of a distribution function $F(x)$ of a random variable X and its so-called *characteristic function* which is defined as:

$$\phi(u) = \mathbb{E}e^{iu\mathbf{x}} = \int e^{iux} \mathrm{d}F(x). \tag{7}$$

Using simple duality argument between time and frequency (via the Fourier transform in (7)), the behavior of the characteristic function at the origin relates to the tail behavior of the distribution F for large $|x|$. In particular, whenever the integer p-th order moment of F exists, the p-th derivative of ϕ at the origin exits as well and they simply relate as follows

$$\phi^{(p)}(0) = \frac{d^p}{du^p}\phi(u)\bigg|_{u=0} = i^p \mathbb{E}[X^p] = i^p \int x^p \mathrm{d}F(x). \tag{8}$$

This justifies ϕ to be also referred to as a *moment generating function*.

Conversely, when p is *even*, existence of $\phi^{(p)}(0)$ implies existence of $\mathbb{E}[X^p]$. Notably, pathologies can occur when p is odd. As the example of $\phi(u) = C^{-1}\sum_{j=2}^{\infty} \cos ju/(j^2 \log j)$ demonstrates (cpre. [15, p. 411]), the existence of $\phi^{(1)}(0)$ does not necessarily guarantee the existence of $\mathbb{E}[X]$.

As we strive towards a generalization of a relation between moments and characteristic function to non-integer orders $r > 0$, let us first introduce the absolute moments of order $r \in \mathbb{R}$:

$$\mathcal{M}_r \triangleq \mathbb{E}[|X|^r] = \int |x|^r \mathrm{d}F(x) \tag{9}$$

where we allow the value ∞. Let us emphasize that $\mathbb{E}[X^p]$ exists if and only if $\mathcal{M}_p$ is finite, in other words, if and only if both $\mathbb{E}[\max(X,0)^p]$ and $\mathbb{E}[\max(-X,0)^p]$ are finite.

We first recall the definition of λ^+ in (2) and note a simple fact:

Lemma 1 *For any distribution F we have*

$$\begin{aligned} \lambda^+ &= \sup\{r>0 \,:\, \mathcal{M}_r < \infty\} \\ &= \sup\{r>0 \,:\, 1-F(x)+F(-x) = O(x^{-r}) \textit{ as } x\to\infty\} \end{aligned} \tag{10}$$

Note that a priori there is no information on the behavior in (10) for r exactly equal to λ^+.

Proof

To obtain one half of the lemma recall the Markov inequality which states that

$$\mathrm{P}\left[|X|>a\right] \le a^{-r}\mathbb{E}|X|^r, \qquad \forall r>0,\, \forall a>0. \tag{11}$$

Consequently, $1-F(a)+F(-a)$ is $O(a^{-r})$ for all $r>0$ with finite $\mathcal{M}_r$. The other half of the lemma follows from theorem [15, Th. 11.3.1] which states that $1-F(a)+F(-a)=O(a^{-r})$ implies that $\mathcal{M}_{r'}$ is finite for all $r'<r$. $\diamondsuit$

Next, we apply a theorem[1] due to Binmore and Stratton [4, 15] which relates the Lipschitz regularity of ϕ at the origin to the tail decay of F for orders less than 2. Recalling the definition of the Lipschitz exponent of ϕ given in (3) we find:

Corollary 2 *If either λ^+ or ρ^+ is known to be strictly less than 2 then:*

$$\lambda^+ = \rho^+. \tag{12}$$

With this in mind, we present wavelet theory in the next section with particular emphasis on their natural abilities to detect and estimate the local regularity of a function.

2.3 Wavelets and Local Regularity

A wavelet analysis consists in a linear decomposition of a signal g onto a set of analyzing functions[2]

$$\{\psi_{t,s}(u) \triangleq |s|^{-1}\psi((u-t)/s),\, (t,s)\in\Re\times\Re_+^*\} \tag{13}$$

[1]Let $0<r<2$. Then, $1-\mathrm{Re}\,\phi(u)=O(u^r)$ if and only if $P[|X|>x]=O(x^{-r})$.

[2]We restrict ourselves to the case of real continuous wavelet transforms, even though all theoretic results we present here transpose directly to the discrete framework of real orthogonal wavelets.

through the inner product

$$W[g](t,s) \triangleq \int g(u)\, \psi_{t,s}(u)\, du. \tag{14}$$

Conceptually, this transform can be viewed as a partitioning of the time-frequency space, where $W[g](t,s)$ measures the correlation between g and each elementary atom $\psi_{t,s}$. All of these time-frequency cells $\psi_{t,s}$ are time-shifted and scale changed versions of a unique prototype function ψ. Therefore, for the time-frequency tiling to be consistent, the mother wavelet must be localized in the time and in the frequency domain simultaneously. Formally, these constraints transpose to the following: We call a wavelet ψ *admissible of regularity* N_ψ, if it has the following three properties:

- $|\psi^{(k)}(t)| \le C_1(1+|t|)^{-N_\psi-1}$ for $k=0,\dots N_\psi$,
- $\int t^k \psi(t)\,\mathrm{d}t = 0$ for $k=0,\dots N_\psi - 1$, and
- $\int_0^\infty |\Psi(\nu)|^2/\nu\,\mathrm{d}\nu = \int_0^\infty |\Psi(-\nu)|^2/\nu\,\mathrm{d}\nu = 1$.

Now, because equation (14) conveys information on the local oscillatory behavior of the analyzed function g, it is possible to assess the local Lipschitz exponent of g from the dynamic of wavelet coefficients across scales. A simple fact reads as follows (see [12, 14], also [23]):

Theorem 3 *Consider an admissible wavelet ψ of regularity $N_\psi \ge r$. Assume that $g(u) - g(0) = O(u^r)$ as $u \to 0$. Then, there is a constant C such that*

$$|W[g](0,s)| \ \le\ Cs^r \qquad \textit{as } s \to 0^+. \tag{15}$$

Reciprocally, somewhat more precise knowledge[3] as $u \to 0$ of the decay of $W[g](u,s)$ for all u allows to determine the local continuity of the function g [14, 13]. As we will elaborate, a certain type of wavelet analysis of the Lipschitz continuity of characteristic functions simplifies dramatically due to the fact that the wavelet transform is in this particular case maximal at the origin.

[3]For a simplified version consider $0 < r < 1$. The following condition implies that $g(u) - g(0) = O(u^r)$ [12, 14, 13]: there exist numbers C and $q < r$ such that

$$|W(u,s)| \ \le\ C\left(s^r + \frac{|u|^q}{\log|u|}\right), \qquad \text{for } s \to 0^+. \tag{16}$$

3 Wavelet analysis of Characteristic Functions

We start by demonstrating how the wavelet analysis of characteristic functions can be simplified tremendously.

3.1 Semi-definite Wavelets

As it turns out it is particularly simple to estimate the wavelet coefficients of a characteristic function provided the wavelet ψ is *semi-definite* by which we mean that its Fourier transform $\Psi(\nu) = \int \psi(t) \exp(-it\nu)\mathrm{d}t$ is real and does not change sign. In other words, ψ is either positive semi-definite, i.e., $\Psi(\cdot) \geq 0$, or it is negative semi-definite, i.e., $\Psi(\cdot) \leq 0$. Examples of such wavelets are the derivatives of even order of the Gaussian kernel: set

$$\psi_p(t) \triangleq c_p \frac{\mathrm{d}^{2p}}{\mathrm{d}t^{2p}} \exp(-\sigma^2 t^2) \tag{17}$$

where c_p is a normalization constant and p is a positive integer. One finds the semi-definite Fourier transform

$$\Psi_p(\nu) = C_p(-1)^p \nu^{2p} \exp\left(\frac{-\nu^2}{4\sigma^2}\right). \tag{18}$$

Lemma 4 *If the Fourier transform Ψ of the wavelet ψ is real, square integrable and semi-definite then*

$$|W[\mathrm{Re}\,\phi](t,s)| \leq |W[\mathrm{Re}\,\phi](0,s)| \tag{19}$$

In other words, for fixed scale s the modulus of the wavelet transform of the real part of a characteristic function is maximal at $t = 0$ for semi-definite Ψ.

Proof

Recall that $|s|^{-1}\psi((u-t)/s)$ and $\Psi(sx)\exp(-itx)$ form Fourier pairs, as well as ϕ and F. Applying Parseval's identity yields

$$\begin{aligned} W[\mathrm{Re}\,\phi](t,s) &= \mathrm{Re} \int |s|^{-1}\psi((u-t)/s)\phi(u)\mathrm{d}u \\ &= \mathrm{Re} \int \Psi(sx)\exp(-itx)\mathrm{d}F(x) \end{aligned} \tag{20}$$

Using the simple estimate $|\mathrm{Re}\,x| \leq |x|$ as well as the fact that Ψ is semi-definite and does not change its sign we obtain

$$\Big|W[\mathrm{Re}\,\phi](t,s)\Big| \leq \int |\Psi(sx)\exp(-itx)|\,\mathrm{d}F(x) = \int |\Psi(sx)|\,\mathrm{d}F(x)$$

$$= \quad \left|\int \Psi(sx)\mathrm{d}F(x)\right| = \left|W[\mathrm{Re}\,\phi](0,s)\right| \tag{21}$$

as desired. $\diamondsuit$

As a corollary from (20) we note

$$W[\mathrm{Re}\,\phi](0,s) \quad = \quad \int \Psi(sx)\mathrm{d}F(x) = \mathbb{E}[\Psi(sX)] \tag{22}$$

3.2 Critical orders smaller than 2

We are now in a position to combine the above results into the anticipated tight connection between a wavelet analysis and the critical order λ^+. For orders larger than 2 the connection is less tight, yet still useful (see Section 3.3).

Theorem 5 *Consider an admissible, semi-definite wavelet ψ of regularity $N_\psi \geq 2$. Then,*

$$\lambda^+ = \rho^+ = \sup\{r > 0 \,:\, |W[\mathrm{Re}\,\phi](0,s)| \;\leq\; Cs^r \text{ for } s \to 0^+\}, \tag{23}$$

provided that either term is known a priori to be strictly less than 2.

¿From a wavelet point of view we can not stress enough that the above result owes its simplicity to the fact that the wavelet coefficients of $\mathrm{Re}\,\phi$ are maximal at 0. Also, $\lambda^+ = \rho^+$ was noted earlier.

Proof

Due to lemma 4, the condition (16) of footnote 3 follows trivially from $W(0,s) \leq Cs^r$. The extension to $0 < r < 2$ exploits the symmetry of $\mathrm{Re}\,\phi$ to conclude that the best polynomial approximation of $\mathrm{Re}\,\phi$ of degree 1 is still constant (for a full argument see the companion paper [24] or [9]). $\diamondsuit$

3.3 Critical orders higher than 2

Attempting to extend the appealing three-fold connection of theorem 5 to orders higher than 2 we face two hurdles, one surmountable due to special properties of the characteristic function, the other more profound.

For a better understanding, we need to extend the concept of Lipschitz continuity to higher orders. To this end, we define the Taylor rest-term of order $2p$ at zero as:

$$Q_{2p}(t) = \operatorname{Re}\phi(t) - 1 - \sum_{k=1}^{p} \frac{t^{2k}}{(2k)!}\phi^{(2k)}(0) \tag{24}$$

whenever it exists. Thus, the general definition of ρ^+ reads then as

$$\rho^+ = \sup\{r > 0 \,:\, Q_{2p}(u) = O(u^r) \text{ as } u \to 0^+, \text{ for } 2p \leq r < 2(p+1)\} \tag{25}$$

Also, we require a more general version of corollary 2. The higher order extension of Binmore and Stratton [4] is found in Kawata [15] and relates the finiteness of moments, i.e., the value of λ^+ to a smoothness condition of Q_{2p}.

The first hurdle concerns the fact the wavelet analysis is a powerful tool for assessing the local degree of regularity, but does in general not allow to make conclusions on differentiability of the analyzed function. To make the point, functions which behave at zero as $|\cdot|^{2.5}$ (modulo a polynomial) but have only one derivative are easily constructed. In other words, the corrective polynomial does not have to be the Taylor polynomial as in (24). This difficulty is overcome by proving existence of moments directly via monotone convergence from the decay of appropriate wavelet coefficients (see the companion paper [24] or [9]). Finite moments imply then that ϕ was indeed differentiable and that wavelet analysis indeed reflects the regularity ρ^+ of Q_{2p}.

The second hurdle stems from the fact that Kawata's smoothness condition[4] (which allows to compute λ^+) is in terms of an integral and weaker than the Lipschitz condition (25) (which is the one resulting from wavelet analysis). However, we are able to obtain bounds. We state only the final result and leave mathematical details to a companion paper [24] (see also [9]).

Corollary 6 *In general, the Lipschitz regularity of a characteristic function (25) is related to the critical order of moments (2) via*

$$\rho^+ \leq \lambda^+ \leq \lfloor \rho^+ \rfloor + 1. \tag{27}$$

Note that $\lfloor \rho^+ \rfloor + 1$ is the smallest integer strictly larger than ρ^+.

[4]Assume that $\phi^{(2p)}(0)$ exists. Then $\mathcal{M}_r$ exists if and only if [15]

$$\int_0^\infty \frac{1}{t^{r+1}} |Q_{2p}(t)| \,\mathrm{d}t < \infty. \tag{26}$$

3.4 Negative Critical Orders

We are now interested in estimating the negative critical order λ^- defined in (2), of a random variable X with density $\mathrm{d}F_X(x)$. Let us define a new random variable Y using the one to one mapping from $\mathbb{R}$ to $\mathbb{R}$: $Y = g(X) = X^{-1}$. Fixing $Y = y$, equation $y = g(x)$ has only one root $x = y^{-1}$, and $|g'(x)| = y^2$, from which we deduce the distribution of Y, $\mathrm{d}F_Y(y) = y^{-2}\mathrm{d}F_X(y^{-1})$. The negative qth order of X simply corresponds to the positive $-q$th order of random variable Y:

$$\mathbb{E}[|X|^q] = \mathbb{E}[|1/Y|^q] = \mathbb{E}[|Y|^{-q}]. \tag{28}$$

Therefore, to estimate $\lambda^-(X)$ of X, we can directly apply general results obtained in Section 3.3 for positive higher orders, to get

$$\lambda^-(X) = -\lambda^+(1/X) \tag{29}$$

4 Estimation procedure

In this section, we elaborate on the implementation of our estimator for λ^+, in particular the choice of wavelet and scales to consider, its bias, robustness and optimality properties.

4.1 Implementation

Given X_i, $(i = 1, \dots, N)$ a set of N observed i.i.d. samples of the distribution $\mathrm{d}F(x)$, we use the empirical estimator for the characteristic function

$$\widehat{\phi}(u) \triangleq \phi_N(u) \triangleq \frac{1}{N}\sum_{k=1}^{N} \exp\{iuX_k\} \tag{30}$$

For our purpose, we need to evaluate this function on a properly sampled interval $u_j = j \cdot \delta u$, $j = 0, \dots, K-1$, that we will make more precise later.

We now recall some convergence properties of this empirical characteristic function (see [7, 8] for details), justifying its use in the rest of our method. First, $\phi_N(u)$ converges almost surely when N goes to infinity towards $\phi(u)$ in the L^∞ sense, over some finite interval T

$$\sup_{|u|\le T} |\phi_N(u) - \phi(u)| \to 0. \tag{31}$$

Second, consider the random process $Y_N(u) = \sqrt{N}(\phi_N(u) - \phi(u))$ and let $Y(u) = \overline{Y}(-u)$ be a zero mean complex Gaussian process with covariance structure $\mathbb{E}Y(u)Y(v) = \phi(u+v) - \phi(u)\phi(v)$. Then, the sequence $Y_N(u_1), Y_N(u_2), \ldots, Y_N(u_m)$ converges in distribution to $Y(u_1), Y(u_2), \ldots, Y(u_m)$.

It is also shown in [8] that $Y_N(u)$ converges weakly towards $Y(u)$ in any finite interval, provided that $\mathbb{E}|X|^{1+\delta} < \infty$.

Next to consider is the wavelet decomposition of $\phi_N(u)$ which simplifies to

$$\begin{aligned} W[\phi_N](t,s) &= \int \psi_{t,s}(v)\,\phi_N(v)\,dv \\ &= \frac{1}{N}\sum_k \int \psi_{t,s}(v)\exp\{iX_k v\}\,dv \\ &= \frac{1}{N}\sum_k \exp\{iX_k t\}\int \psi(u)\exp\{iX_k s u\} \\ &= \frac{1}{N}\sum_k \Psi(s\cdot X_k)\exp\{iX_k t\}. \end{aligned}$$

Two-point Estimation Procedure

(1) Assuming that Ψ is real, semi-definite we finally arrive at the surprisingly simple estimator for the maximal wavelet coefficient of $\mathrm{Re}\,\phi$ of scale s, which is the main ingredient in our method:

$$\widehat{W}(s) \triangleq W[\mathrm{Re}\,\phi_N](0,s) = (1/N)\sum_{k=1}^{N}\Psi(s\cdot X_k). \tag{32}$$

(2) Finally, according to theorem 5 the characteristic exponent ρ^+ is estimated from the powerlaw exponent which steers the decay of $\widehat{W}(s)$. An estimator of the critical moment order results from either corollary 2 or corollary 6. Taking the logarithmic of this powerlaw model yields the linear trend

$$\log\widehat{W}(s) \approx \widehat{\rho^+}\log s \,+\, \log C, \tag{33}$$

where $\widehat{\rho^+}$ is simply obtained via a standard (weighted) linear regression procedure of $\log\widehat{W}(s)$ against s restricted to some scaling interval $(s_{\min}, s_{\max})$ to be specified.

Robustness Since we assume nothing on the distribution we obtain thus a non-parametric estimator. We also note immediately, that the estimation can be *made shift invariant* by subtracting the sample average from the data X and that it is *scale invariant*.

Indeed, consider the data $X'_i = aX_i$. Then $\widehat{W}[X'](s) = \widehat{W}[X](as)$. Rewriting $\log(s)$ as $\log(as) - \log(a)$ we find that the regression data of X' and X differ merely by a shift, leading to the same estimated least square slope.

4.2 Statistics of the estimator

Let us study the bias of both, the simple estimator of the wavelet coefficient (32) itself, as well as the derived estimation of the scaling exponent (33).

Since all observations are drawn from the same distribution, we may write

$$\mathbb{E}[\widehat{W}(s)] = \frac{1}{N}\sum_{k=1}^{N}\mathbb{E}[\Psi(sX_k)] = \mathbb{E}[\Psi(s \cdot X)] = \int \Psi(sx)\mathrm{d}F(x). \qquad (34)$$

This shows that as an estimator of the wavelet coefficient $W(0,s)$ itself, $\widehat{W}(s)$ is unbiased. However, as we will show, a bias is introduced as we estimate the powerlaw decay of $W(0,s)$ through the powerlaw decay of $\widehat{W}(s)$. This result is similar to the one obtained in [1] where it is shown that using log-periodograms (Welch estimator) to analyze processes with spectra of the type $\Gamma_X(f) \sim \sigma^2|f|^{-\alpha}$ leads to a systematic bias on the estimate of α. On the other hand, using a wavelet-based spectral analysis (the frequency marginal of a wavelet decomposition) yields an asymptotically unbiased estimator for exponent α. This is due to the constant relative bandwidth of wavelets that performs a logarithmic tiling of the time axis. The resulting *time-band* analysis has joint time and frequency resolutions that match naturally powerlaw decays as in $\Gamma_X(f)$, or in our case, in $\phi(u)$ around the origin.

Estimating the critical order: A showcase To explore the properties of an estimator of the characteristic exponent ρ^+ through the wavelet coefficients we first treat a simple case where we assume that

- the distribution is Pareto, i.e., $F'(x) = p_X(x) = c_0 x^{-\alpha-1}$ for $|x| > \delta$ and vanishes elsewhere, with $c_0 = \alpha\delta^\alpha$;
- the wavelet is bandlimited, actually require that $\Psi(\nu) = 0$ for $|\nu| \le \underline{\nu}$, where $\underline{\nu} > 0$ is some constant.

Such wavelets are known to exist. For instance, by construction, the auto-correlation function of any admissible and band-limited wavelet is itself a symmetric in time, band limited and positive definite admissible wavelet.

Inserting the particular form of $p_X(x/s)$ into the bias formula (34) we can extract the scale s through a substitution. Provided s is *small enough*, i.e., $s < \underline{\nu}/\delta$, the remaining integral is independent of the scale due to the band limitation of Ψ. This reads as:

$$\begin{aligned}\mathbb{E}[\widehat{W}(s)] &= \int_{\delta}^{\infty} \Psi(xs) c_0 x^{-\alpha-1} \mathrm{d}x = s^{\alpha} \cdot \int_{\delta s}^{\infty} \Psi(y) c_0 y^{-\alpha-1} \mathrm{d}y \\ &= s^{\alpha} \cdot \int_{\underline{\nu}}^{\infty} \Psi(y) c_0 y^{-\alpha-1} \mathrm{d}y = C_{\Psi}(\alpha) \cdot s^{\alpha}. \end{aligned} \tag{35}$$

Thus, the exact powerlaw of the density translates into one of $\widehat{W}(s)$ thanks to the band limitation of the wavelet. Apart from this show-case, approximatively the same decay of $\widehat{W}(s)$ can be observed under much less restrictive assumptions, as we are about to show.

Estimating the critical order of fat tail distributions We relax the above assumptions to the following scenario:

- We consider a simple, heavy tailed probability density function which is symmetrical, constant around the origin and which follows an exact powerlaw in the tails:
$$F'(x) = p_X(x) = \begin{cases} c_1 & \text{if } |x| < \delta, \\ c_2 |x|^{-\alpha-1} & \text{if } |x| \geq \delta, \end{cases} \tag{36}$$
  where
$$c_1 = \frac{1}{2\delta} \cdot \frac{\alpha}{\alpha + v} \qquad \text{and} \qquad c_2 = \frac{\delta^{\alpha}}{2} \cdot \frac{\alpha}{1 + \alpha/v}.$$
- The wavelet is sufficiently regular:
$$\Psi(\nu) \leq d_{\psi} |\nu|^{N_{\psi}}. \tag{37}$$

Let us comment on this choice. Despite its special form, this distribution will be sufficient to explore general fat tail distributions. Clearly, it has finite moments of the orders between $\lambda^- = -1$ and $\lambda^+ = \alpha$. Also, $v = p_X(\delta)/p_X(0)$ is the ratio of the tail amplitude to the constant value around zero. Clearly, the bound (37) is restrictive only at small ν, as Ψ is integrable and must decay at infinity.

To show that $\mathbb{E}[\widehat{W}(s)]$ scales as s^α, we need to generalize (35) and split the integral of (34) into two parts, $|x| < \delta$ and $|x| \geq \delta$. We claim that the first part is of the order s^{N_ψ} while the second term behaves as s^α plus a term of the order s^{N_ψ}. In summary, the wavelet estimator decays indeed as s^α with an error term in the order s^{N_ψ}, which may introduce a bias in the estimation of $\lambda^+ = \alpha$.

Applying (37) we find

$$\int_{-\delta}^{\delta} \Psi(sx) p_X(x) \mathrm{d}x \leq c_1 \cdot \int_{-\delta}^{\delta} \Psi(x) \mathrm{d}x \leq c_1 \cdot 2\delta \cdot d_\psi \delta^{N_\psi} \cdot s^{N_\psi} \tag{38}$$

as claimed. Next, similarly to (35) we obtain

$$\int_{\delta}^{\infty} \Psi(sx) p_X(x) \mathrm{d}x = \frac{1}{s} \int_{s\delta}^{\infty} \Psi(y) p_X(y/s) \mathrm{d}y = s^\alpha \cdot c_2 \int_{s\delta}^{\infty} \Psi(y) y^{-\alpha-1} \mathrm{d}y. \tag{39}$$

To the contrary of (35) this integral depends on s. Thus, we write it as $\int_0^\infty - \int_0^{s\delta}$. The first term is now a constant, leading to the announced behavior as s^α. To estimate the second term, we estimate Ψ in a way similar to (38):

$$\int_0^{s\delta} \Psi(y) y^{-\alpha-1} \mathrm{d}y \leq d_\psi \int_0^{s\delta} y^{N_\psi} y^{-\alpha-1} \mathrm{d}y = d_\psi \frac{(s\delta)^{N_\psi - \alpha}}{N_\psi - \alpha}. \tag{40}$$

Collecting (38) - (40) we find that $\mathbb{E}[\widehat{W}(s)] = As^\alpha + O(s^{N_\psi})$. Bounding the error relied on the regularity (37) of the wavelet, while the exact scaling derives directly from the exact powerlaw (36) of the distribution. We generalize this result as follows:

Proposition 7 *Assume that Ψ is positive semi-definite. Assume, the distribution has a density p_X which can be bounded as follows:*

$$p_X(x) \begin{cases} \leq a|x|^{-\alpha-1} & \textit{for } |x| \geq \delta, \\ \geq b|x|^{-\alpha-1} & \textit{for } |x| \geq \delta, \\ \leq c & \textit{for } |x| \leq \delta. \end{cases} \tag{41}$$

Assume that the regularity of the wavelet ψ is larger than the critical order, i.e., $N_\psi \geq \alpha$. Then,

$$\tilde{a} \cdot s^\alpha + O(s^{N_\psi}) \geq \mathbb{E}[\widehat{W}(s)] \geq \tilde{b} \cdot s^\alpha + O(s^{N_\psi}), \tag{42}$$

with $\tilde{a}/\tilde{b} = a/b$.

Proof

Since ψ has N_ψ vanishing moments we know that (37) holds. Proceeding as before, we write

$$\mathbb{E}[\widehat{W}(s)] \quad = \quad \int_{-\delta}^{\delta} \Psi(sx) p_X(x) \mathrm{d}x + s^{-1} \cdot \int_{|y|>s\delta} \Psi(y) p_X(y/s) \mathrm{d}y \tag{43}$$

The first term is bounded from above as $O(s^{N_\psi})$ as in (38). The second term maybe framed using the tail bounds on p_X as

$$s^\alpha \cdot aI \geq s^{-1} \cdot \int_{|y|>s\delta} \Psi(y) p_X(y/s) \mathrm{d}y \geq s^\alpha \cdot bI, \tag{44}$$

where

$$I = \int_{|y|>s\delta} \Psi(y)|y|^{-\alpha-1} \mathrm{d}y = \int_{-\infty}^{\infty} \Psi(y)|y|^{-\alpha-1} \mathrm{d}y - \int_{-s\delta}^{s\delta} \Psi(y)|y|^{-\alpha-1} \mathrm{d}y.$$

Here, the last term can be bounded as $O(s^{N_\psi - \alpha})$ as in (40). It combines with the factor s^α of (44) to a $O(s^{N_\psi})$. So, only one term behaves as s^α and we find

$$\mathbb{E}[\widehat{W}(s)] \quad = \quad As^\alpha + Bs^{N_\psi} \tag{45}$$

where

$$a \cdot \int_{-\infty}^{\infty} \Psi(y)|y|^{-\alpha-1} \mathrm{d}y \geq A \geq b \cdot \int_{-\infty}^{\infty} \Psi(y)|y|^{-\alpha-1} \mathrm{d}y. \tag{46}$$

A more careful computation reveals that

$$B \leq 2\delta^{N_\psi+1} \cdot c \cdot d_\psi + 2ad_\psi \frac{\delta^{N_\psi-\alpha}}{(N_\psi - \alpha)}. \tag{47}$$

$\diamondsuit$

4.3 Numerical robustness

Provided that the observations X_k ($k = 1, \ldots, N$) are un-correlated one finds easily

$$\mathrm{var}\, \widehat{W}(s) = (1/N) \mathrm{var}\left(\Psi(sX)\right). \tag{48}$$

Moreover, under the assumptions of proposition 7 we conclude that $\mathbb{E}\Psi(sX) \sim s^\alpha$ and, considering Ψ^2 as a wavelet, $\mathbb{E}\Psi^2(sX) \sim s^\alpha$. Thus,

$$\begin{aligned} \mathrm{var}\left(\Psi(sX)\right) \quad &= \quad \mathbb{E}\Psi^2(sX) - \left(\mathbb{E}\Psi(sX)\right)^2 \\ &\sim \quad s^\alpha \mathbb{E}\Psi^2(X) - s^{2\alpha}\left(\mathbb{E}\Psi(X)\right)^2. \end{aligned}$$

To provide a more rigorous error estimate let us assume that we consider the elementary, yet admissible, wavelet

$$\Psi(x) = \begin{cases} A_\psi, & x \in I_\psi = [\nu_\psi - \frac{1}{2}N_\psi^{-1/2}, \nu_\psi + \frac{1}{2}N_\psi^{-1/2}] \\ 0, & \text{otherwise.} \end{cases} \tag{49}$$

This somewhat crude boxcar approximation for the wavelet becomes reasonably accurate for the derivatives of the Gaussian kernel ψ_p (17) as we set $\nu_\psi = \sqrt{p}\sigma/\pi$. Indeed, $|\Psi_p|$ reaches its maximal value $c_p(p\sigma^2/\pi^2)^p \exp(-p)$ at this ν_ψ. Clearly, the approximation becomes more accurate as the regularity increases, i.e., $N_\psi \to \infty$.

For the box-car wavelet we get

$$\text{var}\widehat{W}(s) = (1/N)\text{var}\left(\Psi(sX)\right) = \frac{A_\psi^2}{N}\left(p_X\left[sX \in I_\psi\right] - p_X^2\left[sX \in I_\psi\right]\right).$$

Assuming an exact powerlaw for the tail as in (36) we may write, provided the scale is sufficiently small, i.e., $s < (\nu_\psi - \sqrt{N_\psi}/2)/\delta$:

$$p_X\left[sX \in I_\psi\right] = \int_{(\nu_\psi - \frac{1}{2}N_\psi^{-1/2})/s}^{(\nu_\psi + \frac{1}{2}N_\psi^{-1/2})/s} c_2 x^{-\alpha-1}\text{d}x = s^\alpha \cdot c_2 \int_{I_\psi} y^{-\alpha-1}\text{d}y.$$

Using the mean value theorem we may rewrite the integral by $y_\psi^{-\alpha-1} \cdot N_\psi^{-1/2}$ where y_ψ is some number in I_ψ, thus, $y_\psi \sim \nu_\psi$. Finally, given ψ has unit energy, i.e. $A_\psi = N_\psi^{1/4}$

$$\text{var}\widehat{W}(s) \sim \frac{c_2}{N\nu_\psi^{\alpha+1}}\, s^\alpha \left(1 - \frac{s^\alpha c_2}{\nu_\psi^{\alpha+1}\sqrt{N_\psi}}\right) \tag{50}$$

For small scales ($s \to 0$), the variance behaves like $\text{var}\widehat{W}(s) \sim O\left(s^\alpha\right)$. Figures 4.3(a)–(c) show empirical variance $\text{var}\widehat{W}(s)$ varying with parameters N, s and N_ψ, attesting the good agreement between experimental and theoretical results.

Let us now consider the new variable $\log\widehat{W}(s)$. With a central limit theorem argument, we can say that $\widehat{W}(s)$ is asymptotically normal with mean $\delta_s \approx As^\alpha$ and variance $\sigma_s^2 \approx Cs^\alpha$. Then, in first approximation, using a result on functions of asymptotically Gaussian variables [21, 26], we conclude that $\log\widehat{W}(s)$ is asymptotically Gaussian and

$$\begin{cases} \mathbb{E}\log\widehat{W}(s) \approx \log\mathbb{E}\widehat{W}(s) \approx \log A + \alpha \cdot \log(s) \\ \\ \text{var}\log\widehat{W}(s) \approx |\mathbb{E}\widehat{W}(s)|^{-2}\text{var}\widehat{W}(s) \approx B/A \cdot s^{-\alpha} \end{cases} \tag{51}$$

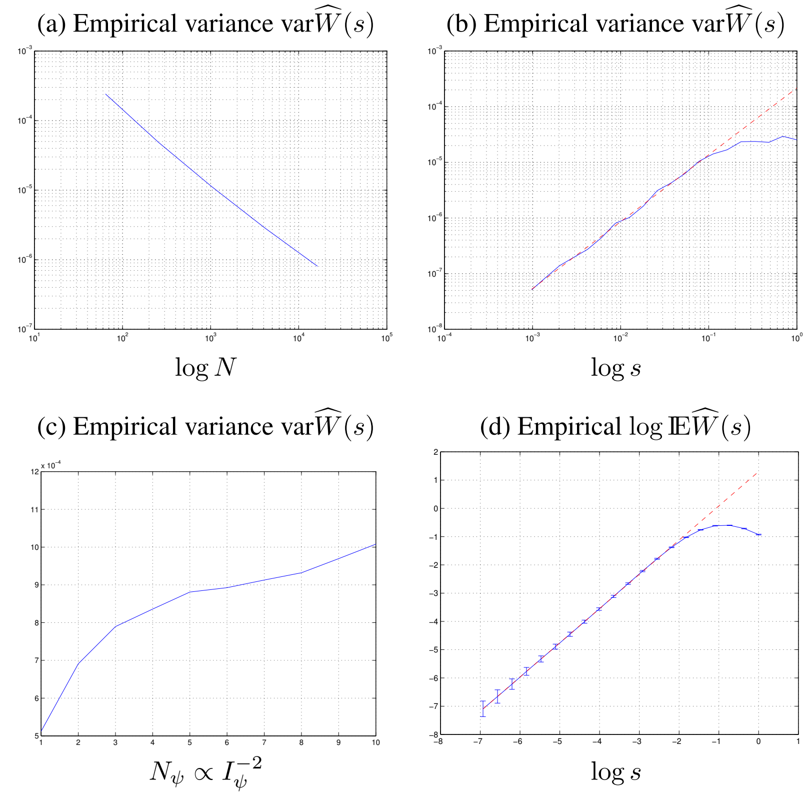

Figure 1: Experimental verification of expressions (50) and (51). (a)–(c): Empirical estimates of $\mathrm{var}\widehat{W}(s)$ estimated over a set of 100 independent realizations of α-stable processes of length N. Evolution of $\mathrm{var}\widehat{W}(s)$ is plotted as a function of: (a) $\log N$ ($\alpha = 1.2$, $N_\psi = 4$, $s = 0.0087$); (b) $\log s$ ($\alpha = 1.2$, $N_\psi = 4$, $N = 2^{14}$); (c) N_ψ ($\alpha = 1.2$, $s = 0.0087$, $N = 2^{14}$). (d) Empirical estimation of $\log \mathbb{E}\widehat{W}(s)$ versus $\log s$. The error bars correspond to the standard deviation of $\log \widehat{W}(s)$. The dashed line materializes the theoretical law $\log \mathbb{E}\widehat{W}(s) = \alpha \cdot \log(s) + C'$ ($\alpha = 1.2$, $N_\psi = 4$, $N = 2^{14}$).

See figure 4.3(d).

To summarize the above, we propose to estimate the characteristic exponent ρ^+ via the estimator of the scaling exponent α of the wavelet coefficients in 51. For (asymptotically) Gaussian random variables such as 51, the maximum likelihood estimator of α is simply obtained from a linear regression of $log\widehat{W}(s)$ against $logs$, as already suggested in 33. Asymptotically, the resulting estimate converges to ρ^+. In practice though, the finite size data set limits the regression range to some interval $s \in (s_{\min} > 0; s_{\max} < \infty)$. The important issue of properly choosing this scaling region is treated in the next section.

4.4 Choice of the scale range

We have defined an estimator for ρ^+ via a log-log linear fit. While in theory the wavelet coefficients should decay as a powerlaw of the scale, we are in practice faced with the fact that the scaling deviates significantly from the theoretical ideal case for both large and small scales. Here we discuss the reasons for this deviation and explain how to choose the scaling region.

4.4.1 Lower bound of scaling region

We have two different approaches to determine a lower bound for the scale range of the linear regression $\log \widehat{W}(s)$ versus $\log s$ in (33).

The first one is based on a Shannon-like theorem. Our estimator estimates the singularity of the characteristic function at the origin. In practice, we use the empirical estimator for the characteristic function, i.e., $\hat{\phi}(u_k) = N^{-1} \sum_j \exp(iu_k X_i)$. The maximum variation of $\hat{\phi}$ is controlled by the maximum value of X_j. Therefore sampling $\hat{\phi}$ at a higher rate than approximately the Nyquist rate $(\overline{X})^{-1}$ with $\overline{X} = \max\{X_j,\, j = 1, \cdots N\}$, does not bring any finer information on the regularity of $\phi(u)$ at $u = 0$. On the contrary, when the analyzing scale goes below the minimum bound $s_{\min} = (\overline{X})^{-1}$, the measured regularity is overestimated, as the function under analysis reduces to the sole C^∞ component $\exp(iu_k\overline{X})$, oversampled at the vicinity of the origin. Thus, concordantly with theorem 5, when $\widehat{\alpha}$ is estimated from data below this minimum scale bound it reflects the regularity N_ψ of the analyzing wavelet rather than the targeted regularity of the characteristic function.

The second approach starts from the expression (32). In order to be consistent, we need to ensure that at least one sample X_j falls inside the equivalent support of $\Psi(sx)$. For small s, only the largest values of X_j are retained to enter the sum (32). As a result, if $\overline{X}$ is the maximum sample of the series X_j, ν_ψ/s is the *central* frequency of the wavelet at scale s. We then want $\overline{X} \approx \nu_\psi/s$, which leads to $s_{\min} \approx \nu_\psi/\overline{X}$.

In summary, both arguments above lead to the same conclusion that the lower cut-off scale should be chosen proportionally to $1/\overline{X}$. For the numerical analysis in this paper we adopted:

$$s_{\min} = 1/\overline{X}. \tag{52}$$

Using a stable law with index of stability (or shape parameter) α, we present in figure 4.4.2 the theoretical lower scale bound $s_{\min} = (\overline{X})^{-1}$. A linear regression of $\log \widehat{W}(s)$ versus $\log s$ for $s > s_{\min}$ yields an accurate estimate of the characteristic exponent $\rho^+ = \alpha$. Moreover, on this same plot, we verify that for $s < s_{\min}$, the wavelet estimator $\widehat{W}(s)$ behaves like s^{N_ψ}, in accordance with the aforementioned argument that the function under analysis is now the C^∞ exponential $\exp(iu\overline{X})$.

4.4.2 Upper bound and negative moments

As we saw, existence of moments is dictated by the tail decay of the distribution $F(x)$ for $x \to \infty$. For instance, it is shown in [25], that the asymptotic tail behavior of a stable law is Pareto when $0 < \alpha < 2$. Defining when exactly this asymptotic behavior starts seems to be a tough problem (see [20]), as it depends heavily on the parameterization that is used to model the distribution (in the parametric context). We just pretend here, that the upper cutoff scale $s_{\max}$ below which $\widehat{W}(s)$ behaves like s^α is also determined by this cutoff value of X separating the tail behavior (as a Pareto law for instance) from the *body* of the distribution. We illustrate this with a compound distribution, made out of a uniform distribution for $|X| \leq \delta$ and $\alpha-$stable distribution for $|X| > \delta$. We show with this simple example (see Figure 4.4.2), that the upper cutoff scale is of order:

$$s_{\max} = \delta^{-1} \tag{53}$$

where δ marks the transition from body to tail behavior in the distribution. In practical applications one might choose δ from prior knowledge (rendering the estimator semi-parametric) or estimate δ itself from the scaling plots (see Figure 4.4.2).

Beyond this upper limit, the wavelet estimator $\log \widehat{W}(s)$ decays with slope -1. This particular value of the slope depends only upon the distribution we have chosen for the body of our compound distribution. In our example, the uniform distribution has negative moments only for $p > \lambda^-(X) = -1$. That is precisely this bound that is estimated in 4.4.2, when $s > s_{\max}$. To support our claim, we simply follow the same lines as for the tail estimator (35) : Given $P_X(x) \sim |x|^{\gamma-1},\ x \to 0$, then

$$\mathbb{E}W(s) = \int_0^{\delta} \Psi(sx) P_X(x) \mathrm{d}x = \int_0^{\delta.s} \Psi(x) |s^{-1}x|^{\gamma-1} s^{-1} \mathrm{d}x,$$

and recalling that Ψ is band-limited, we get :

$$\begin{aligned} \mathbb{E}W(s) &= s^{-\gamma} \int_0^{\overline{\nu}} \Psi(x) |x|^{\gamma-1} \mathrm{d}x,\ \forall s \text{ s.t. } \overline{\nu} \leq \delta s \\ &\sim B_{\psi} s^{\gamma}. \end{aligned}$$

The same value for $\lambda^-(X)$ would have been estimated, if instead of X directly we had analyzed the new random variable $Y = X^{-1}$ as discussed in Section 3.4, and estimated $\lambda^+(Y) = -\lambda^-(X)$ from the tail decay of the transformed distribution.

This observation bears a convenient consequence as far as negative moments are concerned: We can fully exploit the behavior of $\widehat{W}(s)$ for $s > s_{\max}$, leading us to a simple estimator of λ^- in (2). To illustrate this, we now choose a compound process similar to before but replace the uniform distribution for $|X| \leq \delta$ with a Gamma distribution of parameter $0 < \gamma < 1$. The density of the Gamma distribution behaves as $|\cdot|^{\gamma-1}$ around the origin, whence negative moments exist exactly for negative orders $p > \lambda^- = -\gamma$. Therefore, we estimate the slope of $\log \widehat{W}(s)$ for $s > s_{\max} = \delta^{-1}$ and compare this estimate against the theoretical value $\lambda^- = -\gamma$ (see table 2).

To summarize, given K i.i.d. random variables $\{X_j,\ j = 1, \cdots K\}$, the wavelet estimator (32) behaves like:

- $\widehat{W}(s) \sim s^{N_\psi}$, for $s < s_{\min} = (\max_j \{X_j\})^{-1}$,
- $\widehat{W}(s) \sim s^{\rho^+}$, for $s_{\min} < s < s_{\max}$, where $s_{\max}$ corresponds to the inverse of the cut-off value separating the tail from the body of the underlying distribution,
- $\widehat{W}(s) \sim s^{-\rho^-}$, for $s > s_{\max}$.

This is impressively demonstrated in Figure 4.4.2. In fine, both ρ^+ and ρ^- can be deduced from a linear regression of $\log \widehat{W}(s)$ versus $\log s$, over the corresponding scale

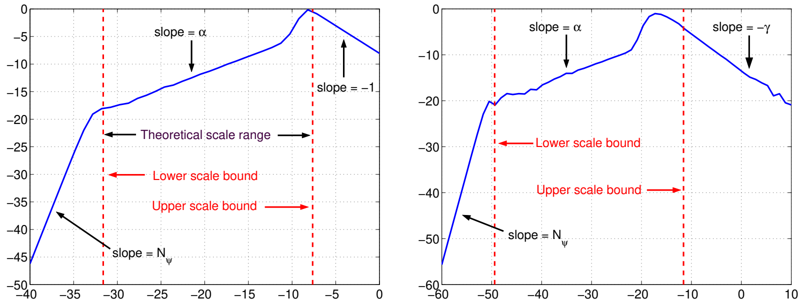

Figure 2: SCALING REGION AND CUTOFF SCALES: Choosing the scale too small, the resolution is fine enough for the wavelet to analyze the individual exponentials that form the estimator $\hat{\phi}$. According to section 4.4.1 the wavelet coefficients decay (at least) with exponent N_ψ. Choosing the scale too large, the estimator samples the body instead of the tail of the distribution; thus, the wavelet coefficients adhere to a powerlaw with exponent λ^-.

intervals. As elaborated in section 3 choosing an appropriate wavelet and according to corollary 6, we have $\lambda^+ = \rho^+$ and $\lambda^- = \rho^-$ whenever these numbers are smaller than 2; in general, $\rho+ \leq \lambda^+ \leq \lfloor \rho^+ \rfloor + 1$ and similar for λ^-.

4.5 Choice of the wavelet

The theoretical results of section 3 form the basis of our estimator. For them to hold the analyzing wavelet ψ is required to have a semi-definite Fourier transform as well as a number of vanishing moments N_ψ larger than $\mathcal{H}_{\mathrm{Re}\,\phi}(0)$.

In practice, we suggest to start with a low regularity wavelet such as the second derivative of the gaussian window $\psi_2(t)$, corresponding to $N_\psi = 2$. If the slope $\widehat{\rho^+}$ obtained from the linear regression of $\log \widehat{W}(s)$ versus $\log s$ is smaller than $N_\psi = 2$, then corollary 2 immediately posits that the positive critical order λ^+ is equal to ρ^+. Now, if the measured slope $\widehat{\rho^+}$ equals $N_\psi = 2$, we need to verify whether the regularity ρ^+ is actually larger than two or not.

To this end, we increase the number of vanishing moments $N_\psi = p$ of $\psi_p(t)$, and repeat the estimation of ρ^+ for increasing integer p as long as the slope $\widehat{\rho^+}$ hits the bound N_ψ. Once we get a $\widehat{\rho^+} < N_\psi$, we should recall corollary 6 which only guarantees that λ^+ can not exceed $\lfloor \rho^+ \rfloor + 1$. This could be of interest in itself for model verification[5].

[5] In [9] and [24], we elaborate on how to use fractional wavelets to get a more accurate estimate for $\widehat{\lambda}^+$.

Despite all this, in all our experiments, we observed that the *basic* estimate ρ^+ obtained with $N_\psi > \mathcal{H}_{\mathrm{Re}\,\phi}(0)$ already accurately estimates λ^+ on its own. In particular, we never encountered the case $\rho^+ < \lambda^+ < \lfloor \rho^+ \rfloor + 1$, that indeed, would necessitate the more refined procedure described in [9, 24], to identify the characteristic exponent precisely.

5 Applications

Application of particular interest in this context are the parameter estimation of stable laws as well as the estimation of the multifractal partition function.

5.1 Estimating Stable and Gamma Parameters

To set notation we recall some classes of distributions well known in the literature, that we will use to illustrate our characteristic exponent estimator.

Pareto. A Pareto density p_X is a simple power law function that take on the form

$$p_X(x) = \begin{cases} \alpha\mu^\alpha x^{-\alpha-1} & \text{if } x > \mu, \\ 0 & \text{else,} \end{cases} \tag{54}$$

with α the shape parameter, and μ the position parameter. A random variable X with Pareto distribution, has positive $r-$th order moments existing only for orders $r < \alpha$, while all negative orders moments exist as $X \geq \mu > 0$ almost surely. The median is $\mu 2^{1/\alpha}$, and if $\alpha > 1$ then the mean exists and equals $\mathbb{E}X = \mu\alpha/(\alpha - 1)$.

Stable. Stable laws form a class of heavy tailed distributions, for which there exists an abundant literature (see e.g. [25] for a detailed introduction). A random variable X follows a stable law that we denote $S\alpha S(\sigma, \beta, \mu)$, if and only if its characteristic function reads :

$$\mathbb{E}[\exp(iuX)] = \exp(-\sigma^\alpha |u|^\alpha (1 - i\beta w_\alpha(t)) + i\mu u), \tag{55}$$

where $w_\alpha(t) = \tan(\pi\alpha \mathrm{sgn}(t)/2)$ for $\alpha \neq 1$ and $w_1(t) = -(2/\pi)\mathrm{sgn}(t) \log |t|$.

Although there exists no closed form for stable distributions except for a handful of special cases, stable laws have a tail behavior that can be approximated as a Pareto distribution (54). Indeed, [25, Property 1.2.15] reads as: If $X \sim S_\alpha(\sigma, \beta, \mu)$ with $0 < \alpha < 2$, then

$$\begin{aligned} \lim_{\lambda \to \infty} \lambda^\alpha P[X > \lambda] &= C_\alpha \frac{1+\beta}{2} \sigma^\alpha \\ \lim_{\lambda \to \infty} \lambda^\alpha P[X < -\lambda] &= C_\alpha \frac{1-\beta}{2} \sigma^\alpha \end{aligned} \tag{56}$$

where $1/C_\alpha = \int_0^\infty x^{-\alpha} \sin(x) \mathrm{d}x$ depends only on α.

The index α is sometimes referred to as the characteristic exponent of the stable law, and for our purpose, it constitutes the most important parameter since absolute moments of order r are finite exactly for $r \in (-1, \alpha)$ ($0 < \alpha < 2$). For $\alpha = 2$ we recover the special case of Gaussian distribution, with existing moments at all orders $r > -1$. The parameter σ indicates scale, since $X \sim S\alpha S(\sigma, \beta, \mu)$, then $aX \sim S\alpha S(a\sigma, \beta, a\mu)$ ($a > 0$). For $\alpha = 2$ we have $\sigma^2 = \text{var}(X)/2$ while for $\alpha < 2$ the second moment $\mathbb{E}[X^2]$ is infinite and the variance is not defined. The parameter μ defines position in the sense that if $X \sim S\alpha S(\sigma, \beta, \mu)$ then $X + c \sim S\alpha S(\sigma, \beta, \mu + c)$. Provided that $\alpha > 1$ we may be even more explicit and identify μ as the expected value: $\mathbb{E}[X] = \mu$. However, in the case $\alpha \le 1$ the mean $\mathbb{E}[X]$ is not even defined; as the most prominent example we mention the Cauchy distribution. Finally, the parameter β provides a measure for the skew, more precisely, X is symmetrical if and only if $\beta = 0$; moreover, if this is the case then (55) reduces to (6).

Gamma. The last case we will comment on is the Gamma distribution. A random variable X has Gamma distribution if

$$p_X(x) = \begin{cases} \lambda x^{\gamma-1} \exp\{-cx\} & \text{if } x \ge 0 \\ 0 & \text{else.} \end{cases} \tag{57}$$

In the above, γ and c are positive numbers, and $\lambda = c^\gamma / \Gamma(\gamma)$, with Γ the generalized factorial function. The special case $\gamma = n/2$, $c = 1/2$ with n an integer, corresponds to the Chi-square density with n degrees of freedom, and for $n = 2$ it reduces to the usual exponential density. As far as moments are concerned, thanks to the dominant exponential decay in (57), all positive order moments exist, and in particular $\mathbb{E}X = \gamma/c$ and $\mathbb{E}X^2 = \gamma(\gamma+1)/c^2$. The negative moments, i.e.,

$$\mathcal{M}_r = \int_0^\infty \lambda x^{r+\gamma-1} \exp\{-cx\} \, \mathrm{d}x, \quad r < 0, \tag{58}$$

Table 1: Estimation of the characteristic exponent α of a stable law, using Koutrouvelis procedure, McCulloch procedure and our wavelet based procedure, using $N = 2^{12}$ i.i.d. samples of a stable variable. Scale, position and skew parameters are fixed ($\sigma = 1$, $\mu = 0$, $\beta = 0$), and α varies in $(0, 2)$. Empirical means and standard deviations (in parenthesis) on the estimates are based upon a 1000 realizations set.

α	0.2	0.6	1	1.4	1.8
$\widehat{\lambda^+}$	0.196 (0.007)	0.58 (0.018)	1.0 (0.035)	1.46 (0.066)	1.74 (0.02)
$\widehat{\alpha}$ (Koutrouvelis)	ND (ND)	0.60 (0.007)	1.0 (0.009)	1.403 (0.013)	1.80 (0.012)
$\widehat{\alpha}$ (McCulloch)	0.59 (0.0018)	0.605 (0.009)	1.0 (0.009)	1.40 (0.016)	1.80 (0.022)

converge only for $r > -\gamma$.

For the above classes of distributions, Pareto, stable and Gamma, there exist efficient procedures aimed at estimating the different sets of parameters. In most cases, these estimators are parametric estimators and they turn out to be optimal (in the sense of maximum likelihood) whenever the specific underlying distribution model and the analyzed data distribution do match. Our estimator (32) is non-parametric, and it should not be expected to outperform a parametric estimator on the distribution it is tailored for. This is for instance very clear on the experiments depicted in Table 1. Considering N i.i.d. samples of a stable variable $X \sim S\alpha S(\sigma, \beta, \mu)$, we compare our estimates (33) of α against two well-known parametric estimators for stable laws : Koutrouvelis [16] and McCulloch [18] procedures.

Superiority of parametric estimators in this appropriate context is not questionable. However, in most real world applications, the true density underlying the data to be analyzed is rarely known, and very likely blind application of parametric estimators will produce aberrant results. A very illustrative example is proposed in Table 2. We consider a Gamma variable X with shape parameter $0 < \gamma < 1$, and form the new variable $Y = X^{-1}$. From (58) we know that $r-$th order moments of Y should only exist for $r < \gamma$. If we now compare the (empirical) densities derived both from Y and from a stable variable with characteristic exponent $\alpha = \gamma$ and skewness parameter $\beta = 1$ (which ensures positivity since $\alpha < 1$) it is quite difficult to dissociate them (Figure 3).

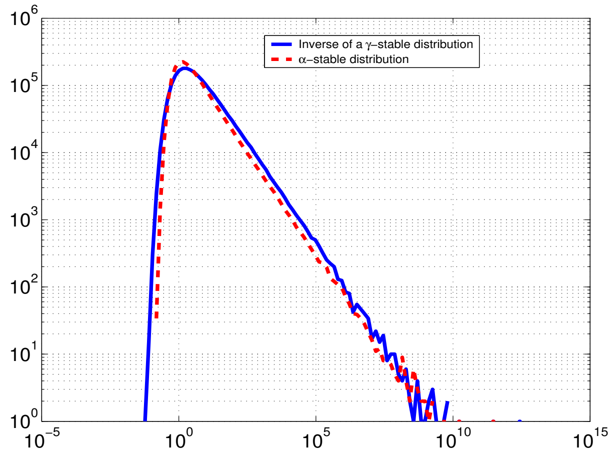

Figure 3: Empirical distributions of the random variables $Y = X^{-1}$ and Z, where X follows a Gamma law with $\gamma = 0.6$, and Z follows a stable law with $\alpha = 0.6$ and $\beta = 1$. For both cases, $\lambda^+ = \alpha = \gamma$. Axis are in logarithmic scale.

Yet, applying crudely stable law designed estimators, like Koutrouvelis or McCulloch, to the raw data Y, yields very bad estimates $\widehat{\alpha} = \widehat{\gamma} = -\widehat{\lambda^-}$. In contrast, determining the characteristic exponent $\lambda^+(Y) = -\lambda^-(X)$ from our wavelet-based regression procedure (as described in Section 3.4), provides us with fairly good estimates of shape parameters γ for Gamma distributions. Hence, because our non-parametric estimator does not assume any a priori distribution for the data, it compares favorably as a general purpose tool to parametric estimators (see for instance the Hill estimator and its various improvements [11, 22, 17, 3], tail estimators [5, 19], and the comparative study conducted in [2]).

Discussion and Conclusions

We itemize the three main results we have derived in this paper.

- We have established a theoretical connection between three exponents namely the critical exponent λ^+ which fixes the highest order of existing moments for a random variable, the tail parameter of its probability distribution and the characteristic exponent ρ^+ which captures the Lipschitz regularity of the characteristic function at origin.

Table 2: Estimation of the shape exponent γ from a Gamma variable X. Koutrouvelis procedure, McCulloch procedure and our wavelet based procedure are applied to the heavy tail transformed variable $Y = X^{-1}$. $N = 2^{12}$ i.i.d. samples of Gamma variable where used. Parameters $c = 1$ is fixed, and γ varies in $(0, 1)$. Empirical means and standard deviations (in parenthesis) on the estimates are based upon a 1000 realizations set.

γ	0.2	0.4	0.6	0.8
$-\widehat{\lambda^-}$	0.204 (0.007)	0.395 (0.008)	0.589 (0.015)	0.793 (0.03)
$\widehat{\alpha}$ (Koutrouvelis)	ND (ND)	0.433 (0.006)	0.56 (0.007)	0.67 (0.009)
$\widehat{\alpha}$ (McCulloch)	0.513 (0.000)	0.514 (0.000)	0.583 (0.009)	0.72 (0.013)

- We proposed a wavelet based estimator of λ^+ and λ^-, that allows for an extraordinarily simple implementation. Moreover, this characteristic exponent estimator is non-parametric and does not assume any a priori knowledge on the underlying distribution, not even Pareto.
- From an application point of view, this estimator shows very useful at characterizing rare events (often responsible for divergence of moments) and measuring power law decays of fat tail distributions. We also mentioned a particularly interesting application of this estimator in the context of model selection in multifractal analysis.

References

[1] P. Abry, P. Gonçalvès, and P. Flandrin. Wavelets, spectrum analysis and $1/f$ processes. In A. Antoniadis and G. Oppenheim, editors, *Lecture Notes in Statistics: Wavelets and Statistics*, volume 103, pages 15–29, 1995.

[2] V. Akgiray and C. G. Lamoureux. Estimation of stable law parameters: a comparative study. *J. Bus. Econ. Stat.*, 7:85–93, 1989.

[3] K. Bianchi and M. Meerschaert. Scale and shift invariant estimators for the heavy tail index alpha. *AMS 2000 Subject Classification: 62F12, 60F05*, 2000.

[4] K. G. Binmore and H. H. Stratton. A note on characteristic functions. *Ann. Math. Statist.*, 40:303–307, 1969.

[5] W. DuMouchel. Estimating the stable index alpha in order to measure tail thickness: A critique. (special invited paper). *Annals of Statistics*, 11:1019–1031, 1983.

[6] A Feuerverger and P. McDunnough. On some fourier methods for inference. *Journal of the American Statistical Association*, 76(374):379–387, 1981.

[7] A Feuerverger and P. McDunnough. On the efficiency of empirical characteristic function procedures. *J. R. Statist. Soc. B*, 43(1):20–27, 1981.

[8] A Feuerverger and R. A. Mureika. The empirical characteristic function and its applications. *Ann. Statist.*, 55:88–97, 1977.

[9] Paulo Goncalves and Rudolf Riedi. Diverging moments and parameter estimation. Research Report 4647, INRIA Rhone-Alpes, Grenoble, France, Nov. 2002.

[10] P. Gonçalvès. Existence test of moments: Application to multifractal analysis. In *Proc. of Int. Conf. on Telecommunications, Acapulco, Mexico*, may 2000.

[11] B. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5):1163–1173, 1975.

[12] S. Jaffard. Exposants de hölder en des points donnés et coefficients d'ondelettes. *C. R. Acad. Sc.*, 308:79–81, 1989.

[13] S. Jaffard. Local behavior of Riemann's function. *Contemporary Mathematics*, 189:287–307, 1995.

[14] S. Jaffard. Méthodes d'ondelettes pour l'analyse multifractale de fonctions. In P. Abry, P. Gonçalvès, and J. Lévy-Véhel, editors, *Lois d'échelle, fractals et ondelettes*, Traité Information–Commande–Communication. Hermès Sc. Pub., Paris, France, 2001.

[15] T. Kawata. *Fourier Analysis in Probability Theory*. Probability and Mathematical Statistics. Academic Press, 1972.

[16] I. A. Koutrouvelis. Regression-type estimation of the parameter of stable law. *J. American Statistical Association*, 75:918–928, 1980.

[17] M. F. Kratz and S. I. Resnick. The qq-estimator and heavy tails. Technical report, Universit Ren Descartes, Paris V and Cornell University, 1995.

[18] J. H. McCulloch. Simple consistent estimator of stable distribution parameter. *Communication on Statistical Simulation*, 15:1109–1136, 1986.

[19] J. H. McCulloch. Measuring tail thickness to estimate the stable index alpha: A critique. *American Statistical Association*, 15:74–81, 1997.

[20] J. P. Nolan. Maximum likelihood estimation and diagnostics for stable distributions. preprint.

[21] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. Mc Graw Hill, 3 edition, 1991.

[22] S. I. Resnick, H. Dress, and L. De Haan. How to make a hill plot. *OR & IE University*, 1998.

[23] R. Riedi. Lois d'échelle multifractales: fondements et approche ondelettes. In P. Abry, P. Gonçalvès, and J. Lévy-Véhel, editors, *Lois d'échelle, fractals et ondelettes*, Traité Information–Commande–Communication. Hermès Sc. Pub., Paris, France, 2001.

[24] R. Riedi and P. Gonçalvès. Diverging moments, characteristic regularity and wavelets. *to be submitted to Applied Probability Trust*, in preparation May 2004.

[25] G. Samorodnitsky and M. Taqqu. *Stable non-Gaussian random processes*. Chapman and Hall, New York ISBN 0-412-05171-0, 1994.

[26] R. Sirfling. *Approximation theorems of Mathematical Statistics*. Wiley, 1980.