Heavy tail estimation with wavelets and internet traffic

Rolf Riedi

with
Paulo Goncalves, INRIA Rhone-Alpes

Hailuoto, June 2005

Diverging Moments

Diverging moments: $\mathbb{E}|X|^q = \infty$

bear on…

- Estimation of tails: $P[|X| > x] \sim x^{-\alpha}$
- Estimators per se: $(X_1^2 + ...X_n^2)/n$
- Asymptotic normality

Moments and Tails

- Let λ>0.

$$\boxed{\mathbb{E}[|X|^r] < \infty \quad \text{for all } r < \lambda}$$

$$\Leftrightarrow$$

$$\boxed{P[|X| > 1/u] \overset{u \to 0}{=} O(|u|^r) \quad \text{for all } r < \lambda}$$

- … not so much an issue of ``convergence'' to Pareto

Moments and characteristic fct

- Characteristic function:

$$\phi(u) = \mathbb{E}[\exp(iuX)]$$

- Moments 101: If E|X|^n is exists then

$$\phi^{(n)}(0) = i^n \mathbb{E}[X^n]$$

- Vice versa: If ϕ has 2p derivates then E|X|^{2p} exists
- Moments 102: (Tauberian Thm) For $0<\lambda<2$

$$\boxed{\mathbb{E}[|X|^r] < \infty \quad \text{for all } r < \lambda} \qquad \Longleftrightarrow$$

$$\boxed{\mathcal{R}e\ \phi(u) - 1 \overset{u \to 0}{=} O(|u|^r) \quad \text{for all } r < \lambda}$$

Extension to orders > 2

- Kawata ('72) / Lukacs ('83) / Ramachandran ('69):
  - Let $2p<\lambda\leq 2p+2$ with integer p.
  - If $\mathbb{E}|X|^\lambda < \infty$ then $\mathcal{R}e\ \phi(u) - \sum_{k=1}^{p} \mathbb{E}[X^{2k}]u^{2k} = O(|u|^\lambda)$
  - Vice versa: If $\mathcal{R}e\ \phi(u) - \sum_{k=1}^{p} a_{2k}u^{2k} = O(|u|^\lambda)$ then $\mathbb{E}|X|^r < \infty$ for all $r < \lambda$...
  - (upon inspection of proof): provided the $a_{2k} = \mathbb{E}[X^{2k}]$ exist.
- $\boxed{\mathbb{E}[|X|^r] < \infty \quad \text{for all } r < \lambda}$ $\Longleftrightarrow$

$$\boxed{\mathcal{R}e\ \phi(u) - \sum_{k=1}^{p} \mathbb{E}[X^{2k}]u^{2k} \overset{u\to 0}{=} O(|u|^r) \quad \text{for all } r < \lambda}$$

Estimating the Regularity of ϕ

- Motivation:
  - exact regularity of ϕ at zero provides the cutoff value for finite moments
- Microscope for regularity: Wavelet transform T

$$T(a,t) =< \mathcal{R}e\ \phi, \psi_{a,b} >= \int \mathcal{R}e\ \phi(t) \cdot \frac{1}{a}\psi\left(\frac{t-b}{a}\right) dt$$

- Simplified regularity theorem: Assume
  - Wavelet regularity N>λ: $\int t^k \psi(t) dt = 0 \quad \text{for } 0 \le k < N$
  - Hoelder polynomial Pϕ of degree ≤N
  - *Transform T(a,t) is maximal at 0*
  - Then

$$\boxed{\begin{array}{lll} \mathcal{R}e\ \phi(u) - P_\phi(u) \overset{u\to 0}{=} O(|u|^r) & & \text{for all } r < \lambda \\ \Leftrightarrow \qquad T(a,0) \overset{a\to 0}{=} O(|a|^r) & & \text{for all } r < \lambda \end{array}}$$

Proof of simplified regularity theorem:

- If $Y(u) - P_Y(u) \overset{u \to 0}{=} O(|u|^r)$
- then $T(a,0) \overset{a \to 0}{=} O(|a|^r)$

$$
\begin{aligned}
|T(a,0)| = \left| < Y, \psi_{a,0} > \right| &= \left| \frac{1}{a} \int_0^a Y(s)\, \psi(s/a)\ ds \right| \\
&= \left| \frac{1}{a} \int_0^a (Y(s) - P_Y(s))\, \psi(s/a)\ ds \right| \\
&\leq C \cdot \frac{1}{a} \int_0^a |s|^r \ |\psi(s/a)|\, ds \\
&\leq C \cdot a^r \frac{1}{a} \int_0^a |\psi(s/a)|\, ds \\
&\leq C \cdot a^r \cdot \int_{\mathbb{R}} |\psi(s)|\, ds
\end{aligned}
$$

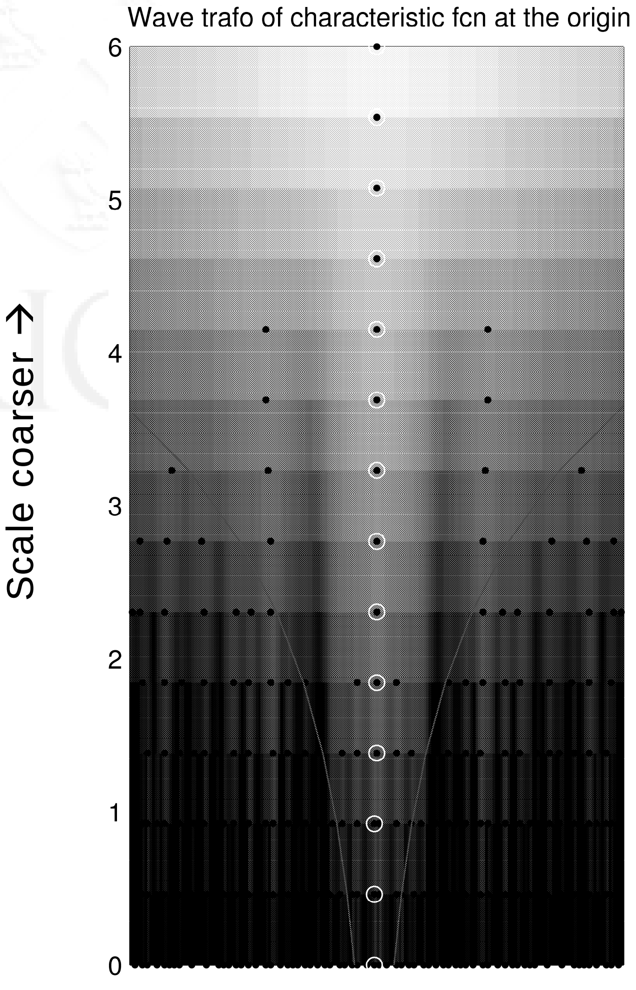
$\int t^k \psi(t) dt = 0$ for $0 \leq k < N$

(1)

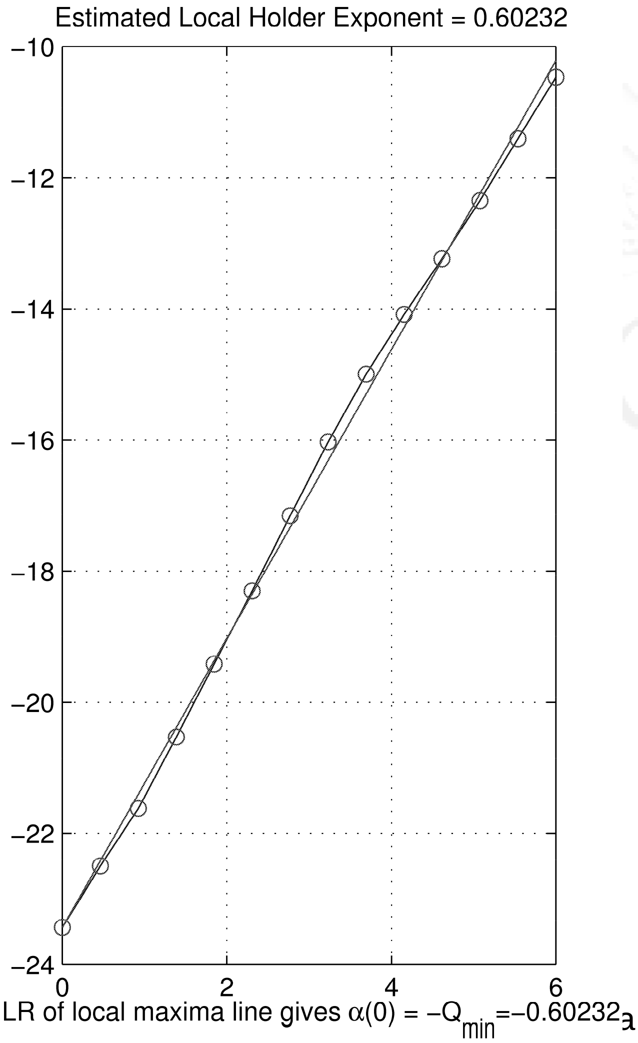
Numerical demonstration

Wavelet Transform

Estimate of scaling exponent



circles correspond to local maxima coeff



LR of local maxima line gives $\alpha(0) = -Q_{min} = -0.60232$

Wavelet Transform of ϕ

- Fourier transform:

$$\Psi_{a,t}(x) = \frac{1}{a}\int \psi(\frac{s-t}{a})e^{isx}ds = \int \psi(u)e^{i(au+t)x}du = e^{ixt}\Psi(ax)$$

- Parseval:

$$T(a,t) =< \mathcal{R}e\ \phi, \psi_{a,t} >= \mathcal{R}e < F, \Psi_{a,t} >= \mathcal{R}e\,\mathbb{E}[\Psi_{a,t}(X)]$$

- Assume: Fourier Transform Ψ of ψ is real positive.
  - then:

$$|T(a,t)| \leq \mathbb{E}[|\Psi_{a,t}(X)|] = \mathbb{E}[|\Psi(aX)|] = |T(a,0)|$$

  - in other words: *T(a,0)* *maximal*

- Ex:

$$\Psi(u) = u^{2n}\exp(-u^2) \geq 0$$

$$\psi(x) = (-1)^n(\tfrac{d}{dx})^{2n}\exp(-x^2)$$

Wavelet Transform of ϕ

- Parseval:

$$T(a,t) = \mathbb{E}[\Psi_{a,t}(X)] = \mathbb{E}[e^{ixt}\Psi(ax)]$$

- Assume: Ψ is real positive then T(a,0) maximal
- Recall equivalent conditions for 0<λ<2:

$$(1) \quad \mathcal{R}e\ \phi(u) - 1 \overset{u\to 0}{=} O(|u|^r) \quad \text{for all } r < \lambda$$

$$(2) \quad T(a,0) \overset{a\to 0}{=} O(|a|^r) \quad \text{for all } r < \lambda$$

- → estimate regularity of Re(ϕ) by the powerlaw

$$\boxed{|T(a,0)| = \mathbb{E}[|\Psi(aX)|] \sim a^{\lambda}}$$

Extension to orders > 2: Differentiability

- Kawata'72 / Lukacs'83 / Ramachandran'69:
  - Let 2p<λ≤2p+2 with integer p.
  - If $\mathbb{E}|X|^\lambda < \infty$ then $\mathcal{R}e\ \phi(u) - \sum_{k=1}^{p} \mathbb{E}[X^{2k}]u^{2k} = O(|u|^\lambda)$
  - and vice versa: If moments up to $\mathbb{E}[X^{2p}]$ exist
    and $\mathcal{R}e\ \phi(u) - \sum_{k=1}^{p} \mathbb{E}[X^{2k}]u^{2k} = O(|u|^\lambda)$
    then $\mathbb{E}|X|^r < \infty$ for all $r < \lambda$.

- Wavelets are blind to *any* polynomials, provide no estimate of differentiability: Ex a function Y(t) with
  - $Y(t) = 1 + t + t^2 + t^{3.5}\sin(1/t)$
  - Taylor polynomial 1+t: once differentiable at t=0
  - Hoelder polynomial $1+t+t^2$:best polynomial approximation
  - Regularity 3.5

$$Y'(t) = 1 + 2t + 3.5t^{2.5}\sin(1/t) - t^{1.5}\cos(1/t)$$
$$Y''(t) = 2 + 3.5 \cdot 2.5t^{1.5}\sin(1/t) + ... + t^{-.5}\sin(1/t)$$

Direct link via fractional wavelets

- Consider fractional Wavelets defined in frequency:

$$\Psi_\nu(u) = c|u|^\nu \exp(-u^2) \geq 0$$

- Lemma: If either side of the following exists then

$$\mathrm{Sup}_a\, T_\nu(a,0)\, a^{-\nu} = c\, E[\, |X|^\nu\,]$$

Proof:

$$\begin{aligned} \underline{T_\nu(a,0)a^{-\nu}} &= a^{-\nu}\, \frac{1}{a}\int \phi_X(u)\psi_\nu(u/a)du \\ &= a^{-\nu}\int \Psi_\nu(ax)dF_X(x) \\ &= c\int |x|^\nu \exp(-(ax)^2)dF_X(x) \overset{a\to 0}{\to} \underline{c\int |x|^\nu dF_X(x)} \end{aligned}$$

Parseval

Monotone convergence

- Fill `gap’ of Lukacs/Ramachandran

$$\mathcal{R}e\ \phi(u) - \sum_{k=1}^{p} a_{2k}u^{2k} = O(|u|^\lambda) \Rightarrow \mathbb{E}|X|^{2p} < \infty$$

Extension to orders > 2: Differentiability

- Kawata'73 / Lukacs'82 / Ramachandran'69:
  - Easy direct fix via monotone convergence
  - Let 2p<λ≤2p+2 with integer p.

$$\boxed{\mathbb{E}[|X|^r] < \infty \quad \text{for all } r < \lambda}$$

$$\Leftrightarrow \boxed{\mathcal{R}e\ \phi(u) - \sum_{k=1}^{p} a_{2k} u^{2k} \overset{u \to 0}{=} O(|u|^r) \quad \text{for all } r < \lambda}$$

$$\Leftrightarrow \boxed{T(a,0) \overset{a \to 0}{=} O(|a|^r) \quad \text{for all } r < \lambda}$$

Numerical Implementation

$$\boxed{|T(a,0)| = \mathbb{E}[|\Psi(aX)|] \sim a^{\lambda}}$$

The estimator of T(a,0) of ϕ is

- …simple:

$$\hat{T}(a,0) = \hat{\mathbb{E}}[\Psi(aX)] = 1/N \sum_{k=1}^{N} \Psi(aX_k)$$

- …unbiased
- …non-parametric!

- Estimation of critical order λ = sup{q: E[|X|^q]<∞ }

$$\boxed{1/N \sum_{k=1}^{N} \Psi(aX_k) \simeq a^{\lambda} \quad \text{as } a \to 0}$$

Practical Considerations

$$\boxed{1/N \sum_{k=1}^{N} \Psi(aX_k) \simeq a^{\lambda} \quad \text{as } a \to 0}$$

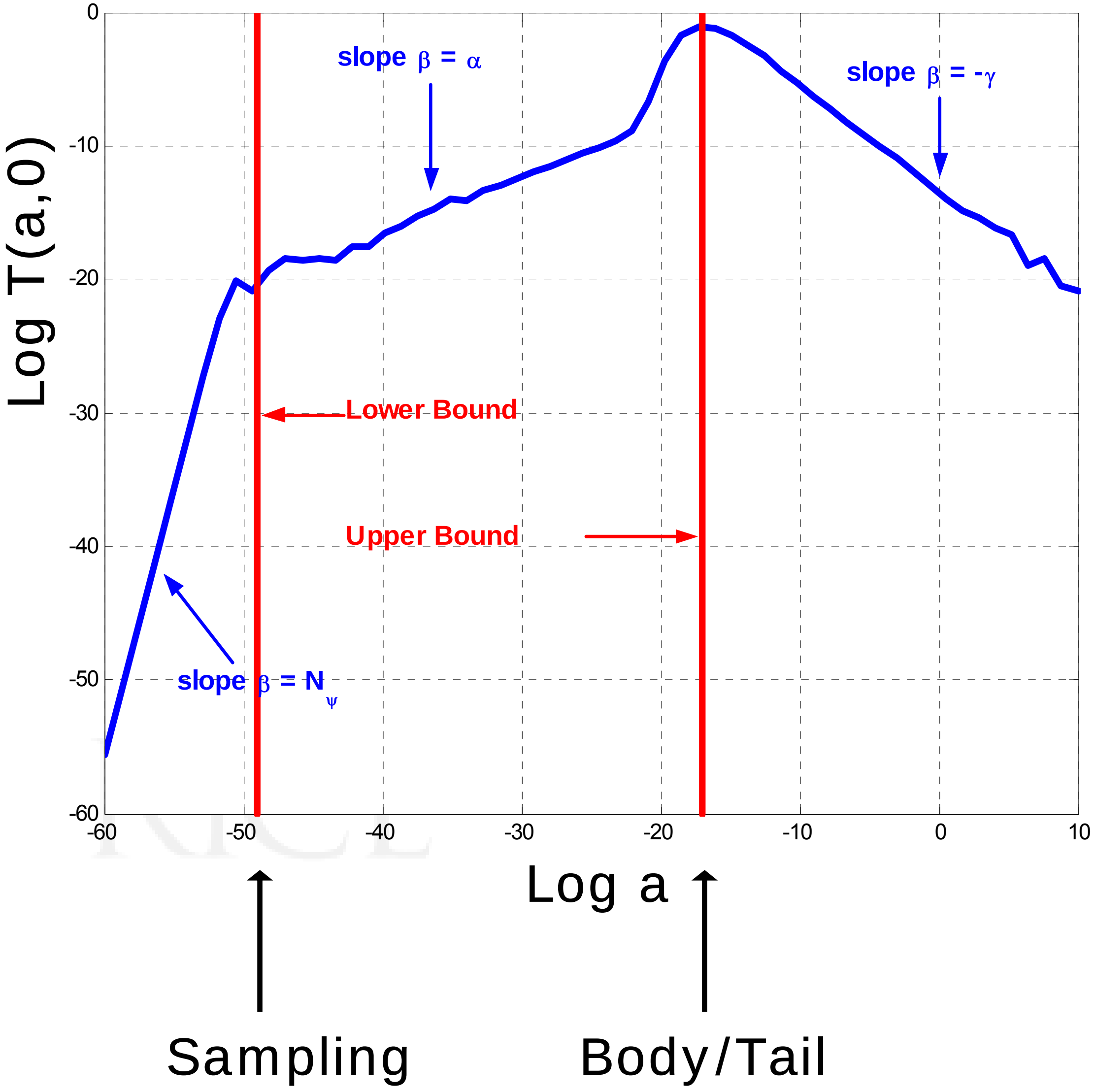
- Choose a wavelet
  - With high enough regularity (N>λ)
  - With real positive Fourier transform (ex: even derivatives of Gaussian kernel)
- Cutoff scales J0 < j < J1
  - Shannon argument on max $\{x_i\}$: lower bound J0
  - Body / Tail frontier : upper bound J1
- Interpretation of estimator:
  - Weight-average of samples with weight $\Psi(aX)$
  - Shift weights out to large samples by scaling a→0

Cutoff scales

Ex: Hybrid distribution

(Gamma body and stable tails)

- (for $x >= \delta$)
  - $\mathbf{x} \sim \alpha$-stable (β=1),
  - E $|x|^r = \infty$, $r >= \alpha$
- (for $x < \delta$)
  - $\mathbf{x} \sim \Gamma(\gamma)$
  - E $|x|^r = \infty$, $r <= -\gamma$



Competing for stable parameter

Alpha-stable Laws:

- compare with Koutrouvelis’80 and McCullogh’86 are parametric (stable distribution)
- non-parametric wavelet based estimator is
  - competitive
  - especially for intermediate to small a

α	0.2	0.6	1	1.4	1.8
Wavelet based	0.196 ± 0.007	0.58 ± 0.018	1.0 ± 0.035	1.46 ± 0.066	1.74 ± 0.02
$\hat{\alpha}$ (Koutrouvelis)	ND	0.60 ± 0.007	1.0 ± 0.009	1.403 ± 0.013	1.80 ± 0.012
$\hat{\alpha}$ (McCullogh)	0.59 ± 0.0018	0.605 ± 0.009	1.0 ± 0.009	1.40 ± 0.016	1.80 ± 0.022

Competing for Pareto parameter

1/Gamma Laws:

- Pareto
- Koutrouvelis'80 and McCullogh'86 are parametric (stable distribution)
- non-parametric wavelet based estimator is
  - superior

γ	0.2	0.4	0.6	0.8
Wavelet based	0.204 ± 0.007	0.395 ± 0.008	0.589 ± 0.015	0.793 ± 0.03
$\hat{\alpha}$ (Koutrouvelis)	ND	0.433 ± 0.006	0.56 ± 0.007	0.67 ± 0.009
$\hat{\alpha}$ (McCullogh)	0.513 ± 0.000	0.514 ± 0.000	0.583 ± 0.009	0.72 ± 0.013

Interlude

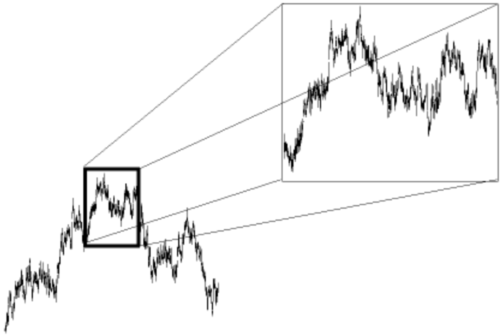
Statistical scaling

Statistical Self-similarity

- H-self-similar:

  $B(at) =^{fdd} a^H B(t)$

  stationary increments



- H-ss examples
  - Gaussian: unique, fractional Brownian motion
  - Stable: not unique, Levy motion

How do self-similar processes occur?

- X_k: stationary time series

If

$$\frac{X_1 + \ldots + X_{\lfloor tn \rfloor}}{f(n)} \overset{f.d.d.}{\longrightarrow} Z(t)$$

then necessarily

$$H = \lim_{n\to\infty} \frac{\log f(n)}{\log(n)}$$

and $Z(t)$ is H-self-similar.

- → ADDITIVE SCHEME

- X_k iid, finite variance: H=1/2, Z is Brownian motion
- X_k LRD: H>1/2, Z is fractional Brownian motion

Scaling of moments

Self-similarity: $\mathbb{E}[|B(t+\delta)-B(t)|^q] \simeq \delta^{qH}$

Multifractal scaling: $\mathbb{E}[|X(t+\delta)-X(t)|^q] \simeq \delta^{\tau(q)}$

- Multifractal:
  - In distribution, log |X(t)-X(t+δ)| looks like a convolution,
  - Thus, |X(t)-X(t+δ)| looks in distribution like a product

$$\mathbb{E}[e^{q \log |X(\delta)|}] \simeq \exp(\tau(q)\log(\delta)) = \underbrace{(\exp(\tau(q)))^{\log(\delta)}}_{\log\delta\text{-fold convolution}}$$

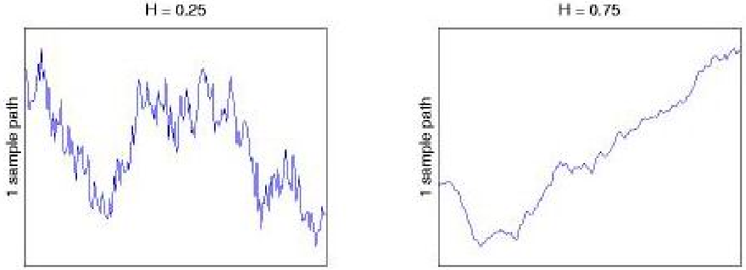
- τ(q)=Hq:
  - Multifractal regresses to self-similarity (mono-fractal).
  - X looks statistically like a constant
- Model identification:
  - Additive versus multiplicative
  - Linear versus strictly convex tau(q)

Monofractal versus Multifractal

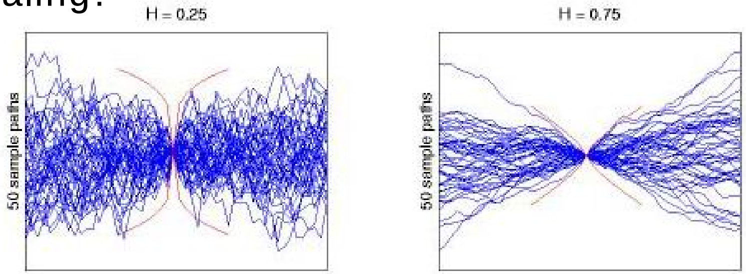
- Fractional Brownian motion
- Levy modulus of continuity:

  $| B_H(t+d)-B_H(t) | \sim |d|^H$ for all t

Dependence (Memory):

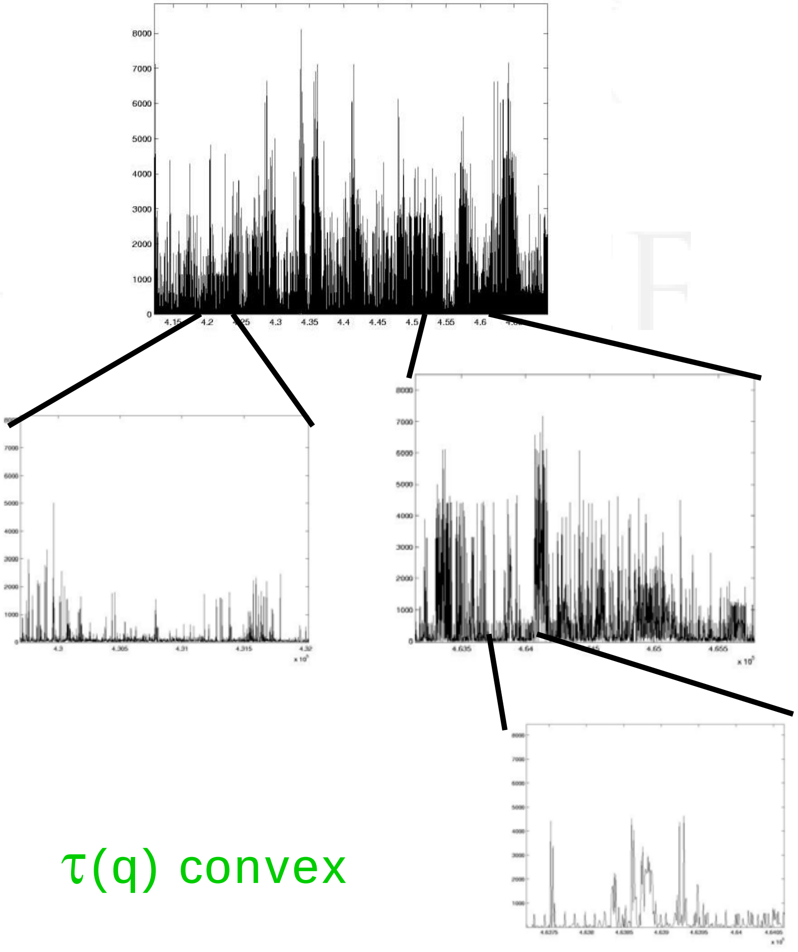


Scaling:



$\tau(q)$ linear

- Measured data traffic:
  - Smooth and bursty regions are ubiquitous



$\tau(q)$ convex

Rudolf Riedi Rice University

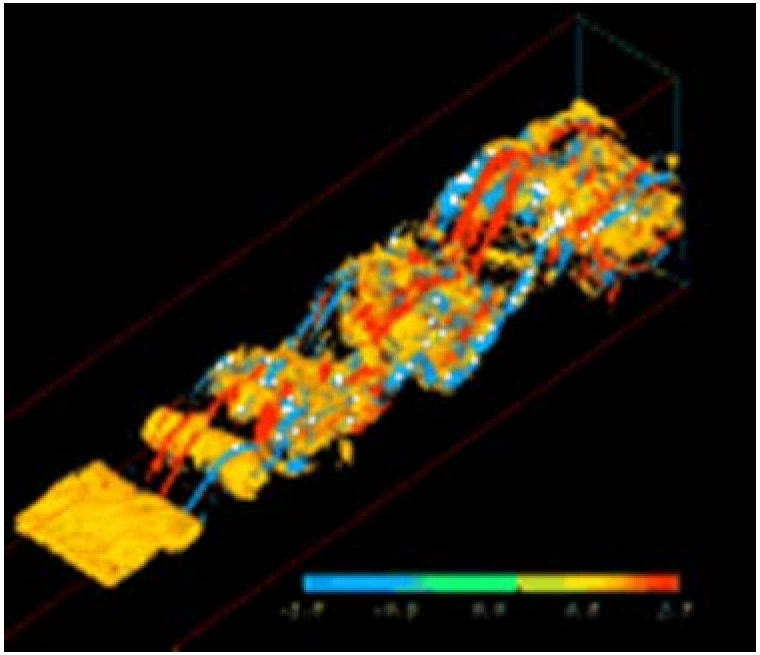
Model identification

…through scaling of moments

Why Moments and Scaling

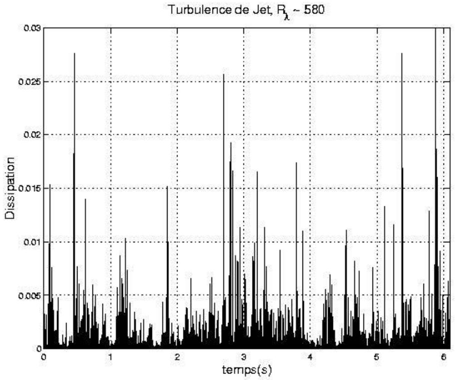
Turbulence: models wanted

- Velocity field v(x)



- Kolmogorov 1941:
  - $\mathbb{E}|v(t+\delta)-v(t)|^q \simeq \delta^{q/3}$
  - Linear model, fBm
- Kolmogorov 1962:
  - $\mathbb{E}|v(t+\delta)-v(t)|^q \simeq \delta^{\tau(q)}$
  - Multiplicative model, Cascade
- More recent:
  - non-powerlaw scaling
  - Infinitely divisible cascades

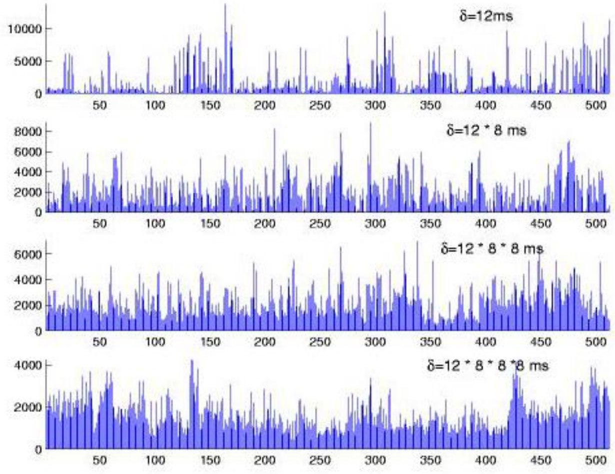
Courtesy P. Chainais

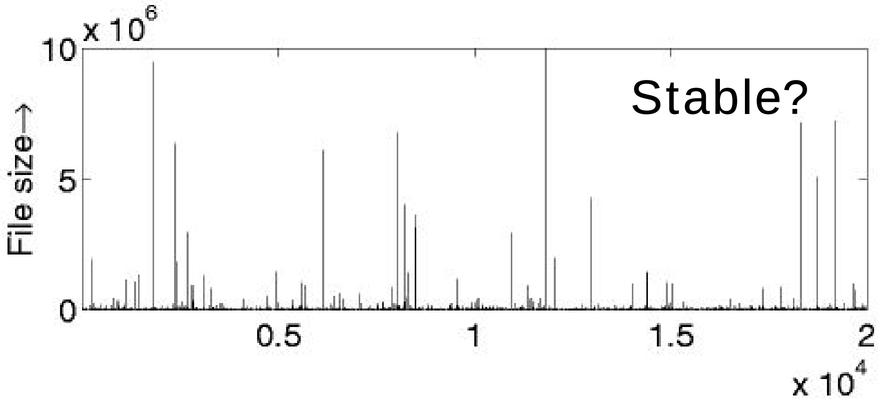


Scaling and statistical aspects

- Networks
  - Non-Gaussianity / Long-memory
  - Model identification (cascade?)
- WWW
  - File size distribution
- Stock Markets
  - Long-memory

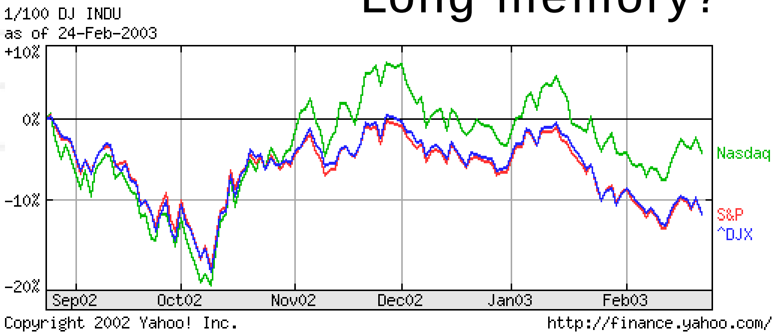
Log-normal?





Stable?

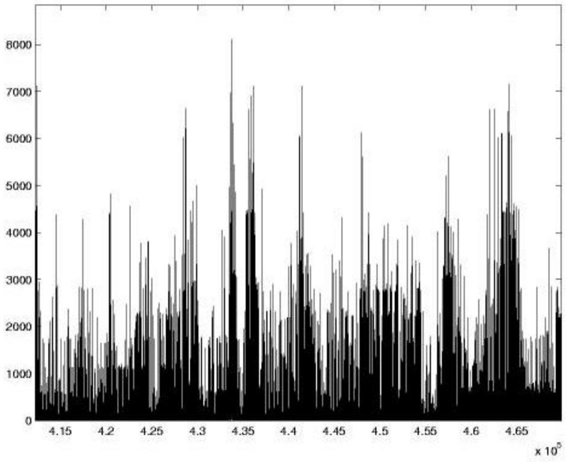
Long memory?



Copyright 2002 Yahoo! Inc. http://finance.yahoo.com/

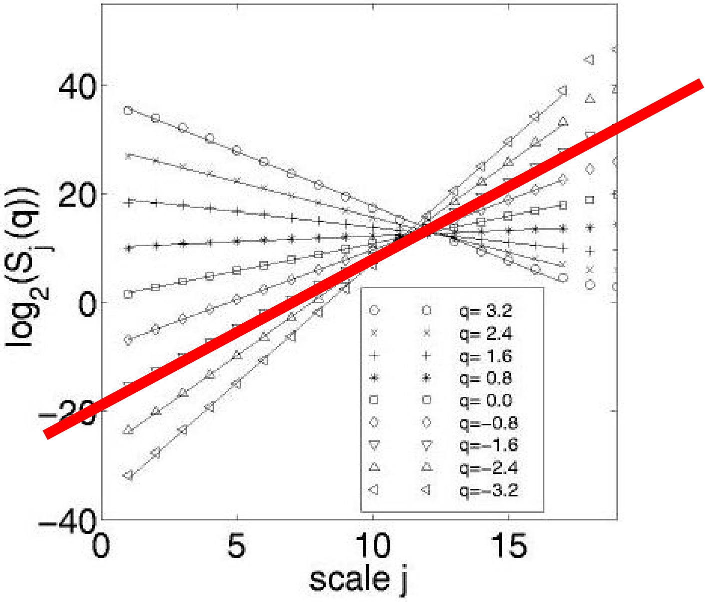
Praxis of estimating $\tau(q)$

- Data: $Y_k = A(k+1)-A(k)$ traffic load per time unit
- $S(j,q) = \Sigma_k |Y_k|^q$
- $\tau(q)$ = slope of $j \rightarrow \log S(j,q)$



Data = Bellcore 1993
traffic arrival per time bin

$j \rightarrow \log S(j,q)$ for several q



Slope = $\tau(q)$

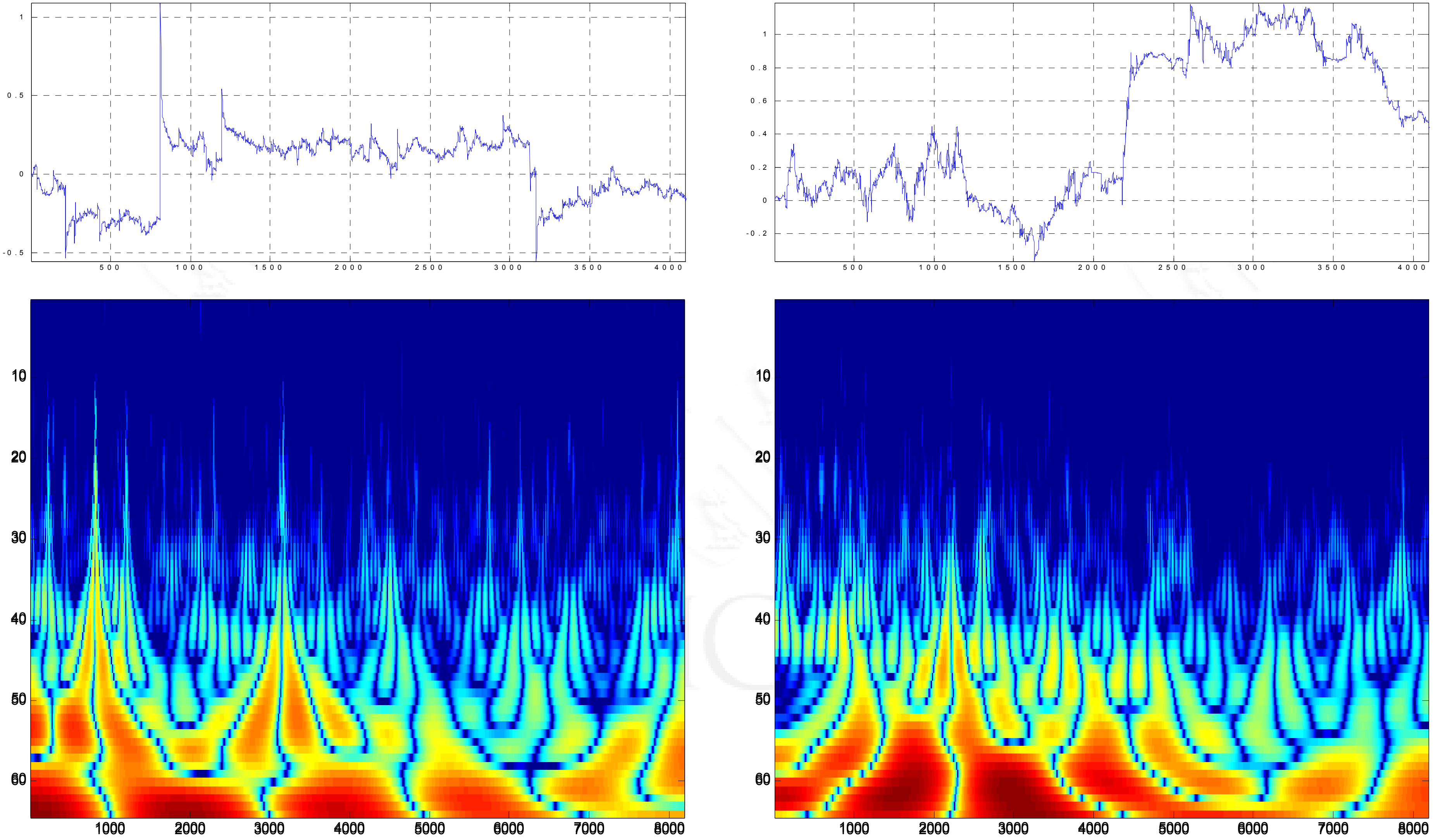
Model identification

back

Identify the Multifractal

- One of these signals is a stable Levy flight,
- …the other is a multiplicative cascade.
- Which is which?

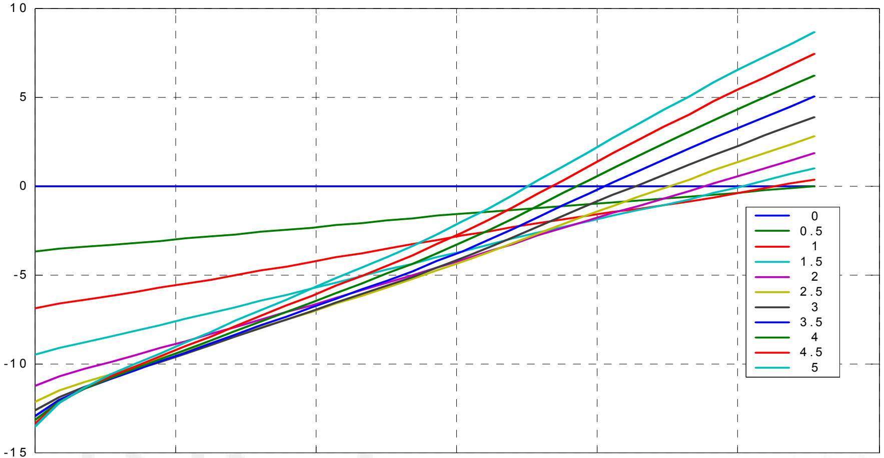
Wavelet transform: $a^{-½} \int x(t)\, \psi((t-b)/a)\, dt$



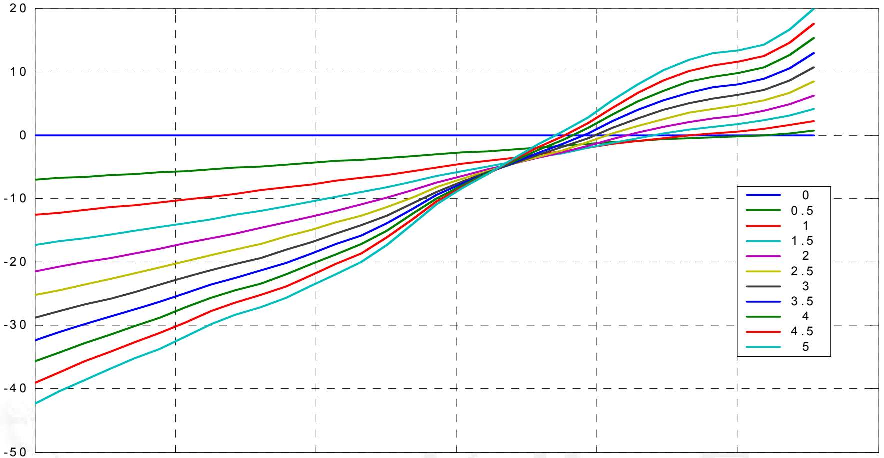
Challenge: which wavelet to use.

Estimating τ(q) from $\mathbb{E}|T(a,b)|^q \simeq a^{\tau(q)}$

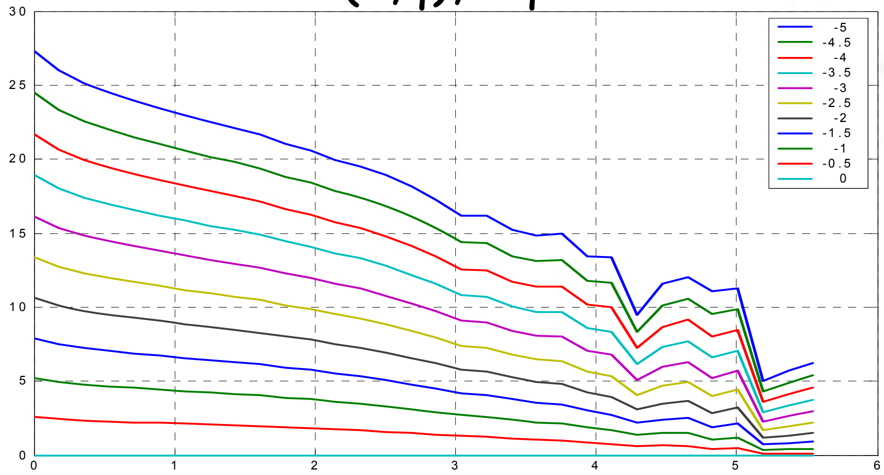
S(a,q), q>0



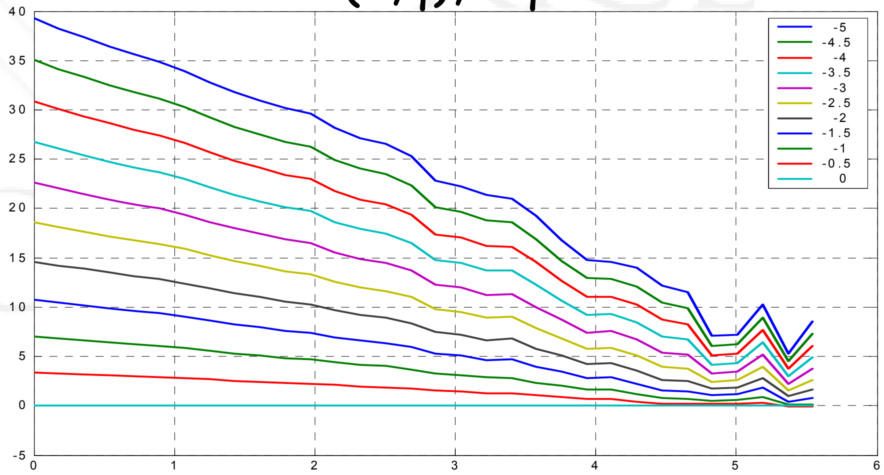
S(a,q), q>0



S(a,q), q<0

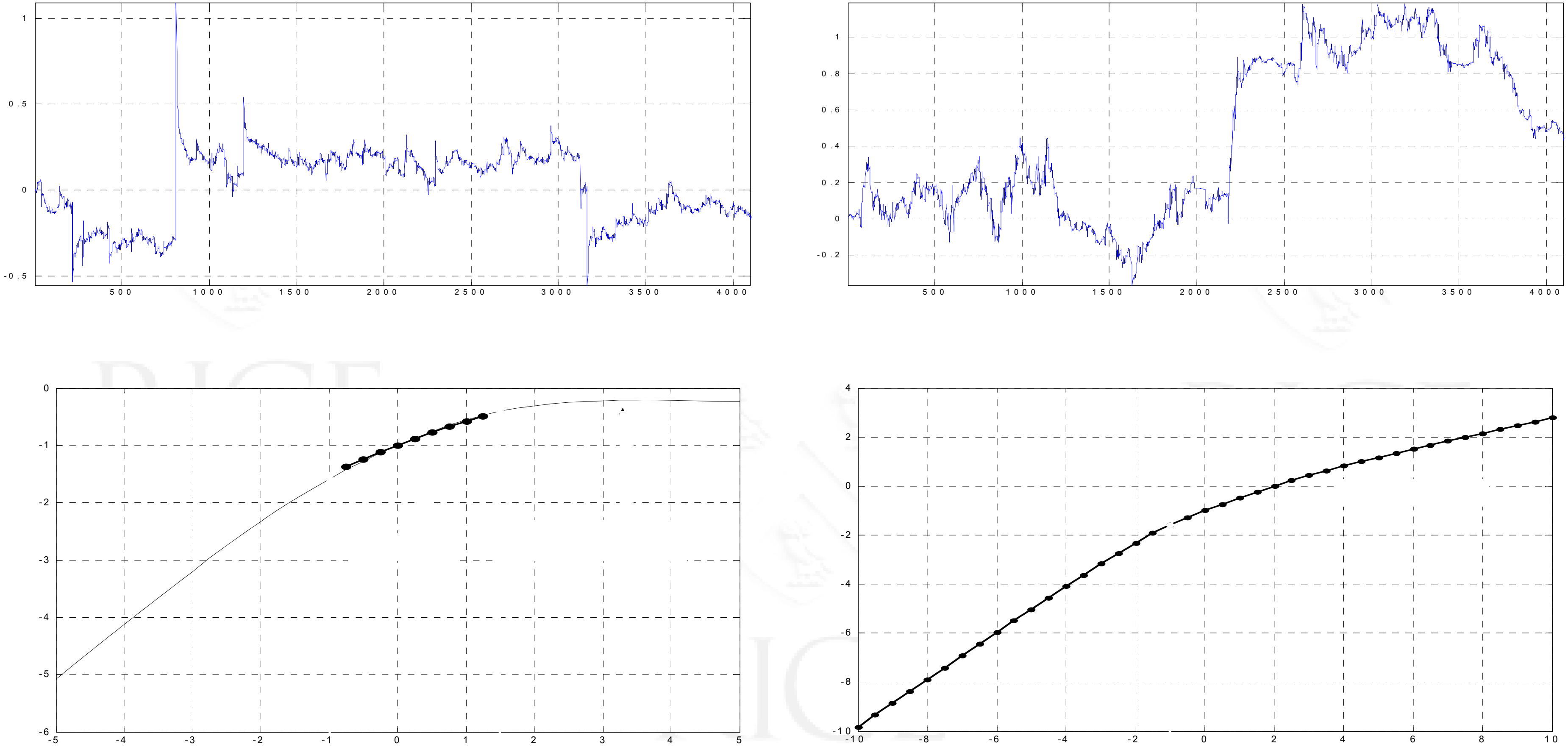


S(a,q), q<0



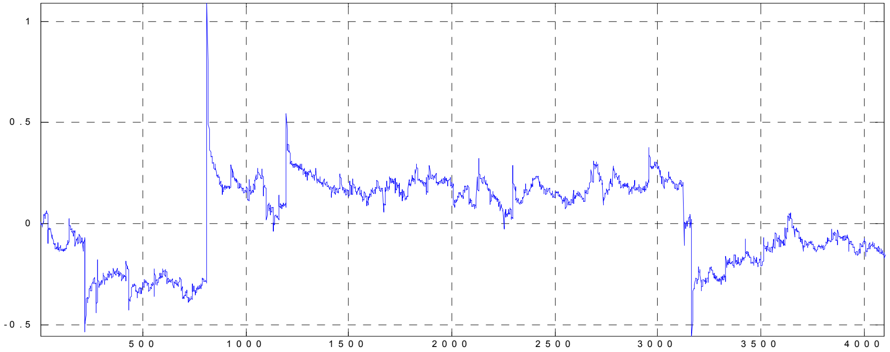
Challenge: which orders to use.

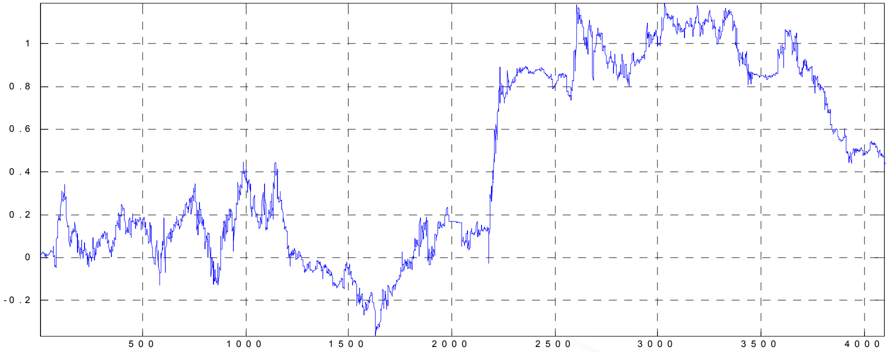
Estimate of τ

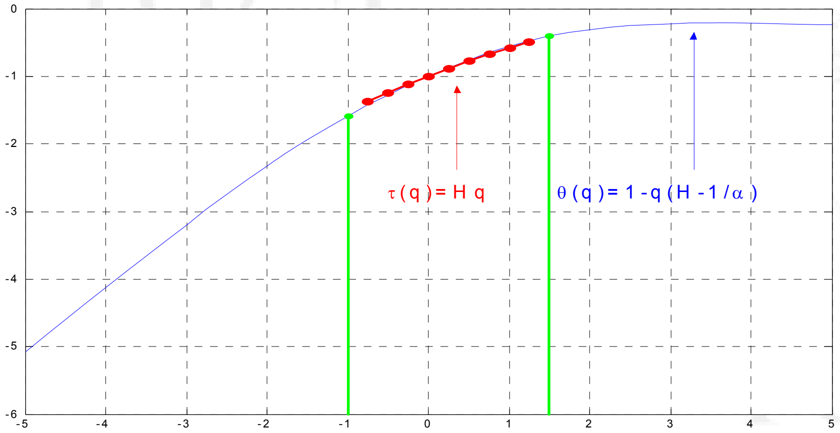


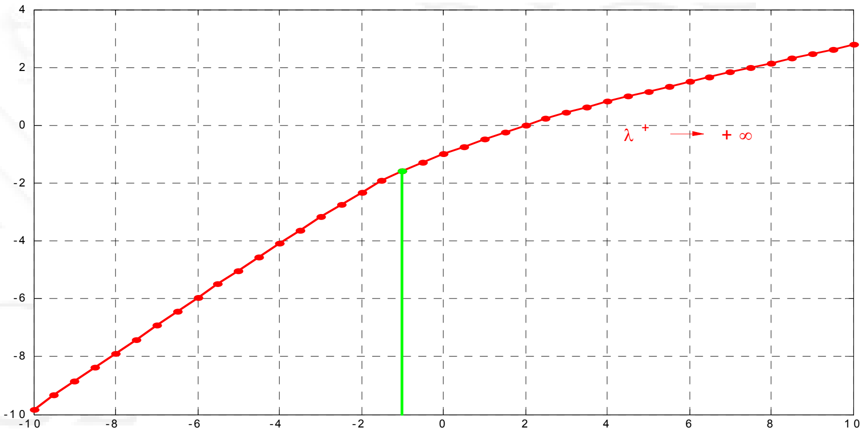
Challenge: Interpretation.

Supervised multifractal estimation









The moments exist only for a few q. The spectrum hints to a monofractal, i.e., Levy flight

The moments exist in a wide range. The spectrum hints to a multifractal, i.e., a cascade.

Summary

- Wavelets useful for non-parametric estimation
- Holder regularity of characteristic function tied to existence of moments beyond order 2
- Estimating critical order of finite moments useful for
  - Tail estimation
  - Model identification

References: Scaling processes

- Beran, J. (1994). *Statistics for Long-Memory Processes*, Chapman & Hall.
- Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*, Chapman and Hall.
- Doukhan, Oppenheim and Taqqu (eds) (2002):

  Long range dependence : theory and applications, Birkhaeuser

- Software:
  - Goncalves: http://www.inrialpes.fr/is2/people/pgoncalv
  - Veitch: http://www.emulab.ee.mu.oz.au/~darryl
  - Riedi: http://www.stat.rice.edu/~riedi