

Bootstrap for Log Wavelet Leaders Cumulant based Multifractal Analysis

Herwig Wendt, Stephane G. Roux, Patrice Abry

Physics Lab., CNRS UMR 5672
Ecole Normale Supérieure de Lyon, France

EUSIPCO 2006, Firenze, 07.09.2006



Motivation

- **Multifractal Analysis (MFA)**
 - Scaling in data: numerous applications of very different nature
 - Usually based on moments of wavelet coefficients
- **Wavelet Leaders:**
 - significant theoretical/practical qualities
- **Log-cumulants based MFA:**
 - emphasizes difference mono- & multi-fractal processes

Goal:

Practical procedure for obtaining accurate log-cumulant estimates and confidence intervals from one single realization of data

- Do **Wavelet Leaders** improve current estimation procedures?
- Does **Bootstrap** provide reliable confidence intervals?

Motivation

- **Multifractal Analysis (MFA)**
 - Scaling in data: numerous applications of very different nature
 - Usually based on moments of wavelet coefficients
- **Wavelet Leaders:**
significant theoretical/practical qualities
- **Log-cumulants** based MFA:
emphasizes difference mono- & multi-fractal processes

Goal:

Practical procedure for obtaining accurate log-cumulant estimates and confidence intervals from one single realization of data

- Do **Wavelet Leaders** improve current estimation procedures?
- Does **Bootstrap** provide reliable confidence intervals?

Motivation

- **Multifractal Analysis (MFA)**
 - Scaling in data: numerous applications of very different nature
 - Usually based on moments of wavelet coefficients
- **Wavelet Leaders:**
significant theoretical/practical qualities
- **Log-cumulants** based MFA:
emphasizes difference mono- & multi-fractal processes

Goal:

Practical procedure for obtaining accurate log-cumulant estimates and confidence intervals from one single realization of data

- Do **Wavelet Leaders** improve current estimation procedures?
- Does **Bootstrap** provide reliable confidence intervals?

Motivation

- **Multifractal Analysis (MFA)**
 - Scaling in data: numerous applications of very different nature
 - Usually based on moments of wavelet coefficients
- **Wavelet Leaders:**
significant theoretical/practical qualities
- **Log-cumulants** based MFA:
emphasizes difference mono- & multi-fractal processes

Goal:

Practical procedure for obtaining accurate log-cumulant estimates and confidence intervals from one single realization of data

- Do **Wavelet Leaders** improve current estimation procedures?
- Does **Bootstrap** provide reliable confidence intervals?

Motivation

- **Multifractal Analysis (MFA)**
 - Scaling in data: numerous applications of very different nature
 - Usually based on moments of wavelet coefficients
- **Wavelet Leaders:**
significant theoretical/practical qualities
- **Log-cumulants** based MFA:
emphasizes difference mono- & multi-fractal processes

Goal:

Practical procedure for obtaining accurate log-cumulant estimates and confidence intervals from one single realization of data

- Do **Wavelet Leaders** improve current estimation procedures?
- Does **Bootstrap** provide reliable confidence intervals?

Wavelet Coefficients and Wavelet Leaders

- Discrete Wavelet Coefficients

$$\begin{aligned}\psi_{j,k}(t) &= 2^{-j}\psi_0(2^{-j}t - k) && \text{dyadic grid} \\ d_X(j, k) &= \langle \psi_{j,k} | X \rangle\end{aligned}$$

- Wavelet Leaders $L_X(j, k) = \sup_{\lambda' \subset 3\lambda_{j,k}} |d_{\lambda'}|$

$$\lambda_{j,k} = [k2^j, (k+1)2^j), \quad 3\lambda_{j,k} = \lambda_{j,k-1} \cup \lambda_{j,k} \cup \lambda_{j,k+1}$$

*Supremum:
taken on $d_X(j, k)$,
in time neighborhood $3\lambda_{j,k}$,
over all finer scales $2^{j'} < 2^j$*

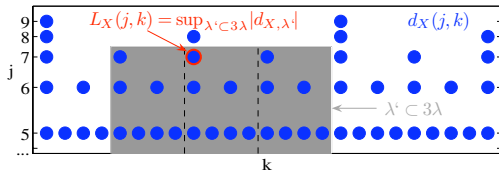
Wavelet Coefficients and Wavelet Leaders

- Discrete Wavelet Coefficients

$$\begin{aligned}\psi_{j,k}(t) &= 2^{-j}\psi_0(2^{-j}t - k) && \text{dyadic grid} \\ d_X(j,k) &= \langle \psi_{j,k} | X \rangle\end{aligned}$$

- Wavelet Leaders $L_X(j,k) = \sup_{\lambda' \subset 3\lambda_{j,k}} |d_{\lambda'}|$

$$\lambda_{j,k} = [k2^j, (k+1)2^j), \quad 3\lambda_{j,k} = \lambda_{j,k-1} \cup \lambda_{j,k} \cup \lambda_{j,k+1}$$



Supremum:
taken on $d_X(j,k)$,
in time neighborhood $3\lambda_{j,k}$,
over all finer scales $2^{j'} < 2^j$

Scaling and Multifractal

- **Scale Invariance:**

$$\frac{1}{n_j} \sum_{k=1}^{n_j} L_X(j, k)^q = F_q |a|^{\zeta(q)} \quad (1)$$

for $a \in [a_m, a_M]$, $a_M/a_m \gg 1$ ($a = 2^j$)

- **Multifractal Analysis:**

$\zeta(q) \longrightarrow$ singularity spectrum $D(h)$

- **Scaling exponent $\zeta(q)$:**

$\zeta(q) = qH$ linear $\rightarrow X$ monofractal

$\zeta(q) \neq qH$ non linear $\rightarrow X$ multifractal

Scaling and Multifractal

- **Scale Invariance:**

$$\frac{1}{n_j} \sum_{k=1}^{n_j} L_X(j, k)^q = F_q |a|^{\zeta(q)} \quad (1)$$

for $a \in [a_m, a_M]$, $a_M/a_m \gg 1$ ($a = 2^j$)

- **Multifractal Analysis:**

$\zeta(q) \longrightarrow$ singularity spectrum $D(h)$

- **Scaling exponent $\zeta(q)$:**

$\zeta(q) = qH$ linear $\rightarrow X$ monofractal

$\zeta(q) \neq qH$ non linear $\rightarrow X$ multifractal

Scaling and Multifractal

- **Scale Invariance:**

$$\frac{1}{n_j} \sum_{k=1}^{n_j} L_X(j, k)^q = F_q |a|^{\zeta(q)} \quad (1)$$

for $a \in [a_m, a_M]$, $a_M/a_m \gg 1$ ($a = 2^j$)

- **Multifractal Analysis:**

$\zeta(q) \longrightarrow$ singularity spectrum $D(h)$

- **Scaling exponent $\zeta(q)$:**

$\zeta(q) = qH$ linear $\rightarrow X$ monofractal

$\zeta(q) \neq qH$ non linear $\rightarrow X$ multifractal

Log-Cumulants

Some classes of processes:

- Eq. (1) $\longrightarrow \mathbb{E}L_X(j, \cdot)^q = F_q |2^j|^{\zeta(q)}$ (2)

- $\ln \mathbb{E}e^{q \ln L_X(j, \cdot)} = \sum_{p=1}^{\infty} C_p^j \frac{q^p}{p!}$

C_p^j - cumulants of random variable $\ln L_X(j, \cdot)$

$$C_p^j = c_p^0 + c_p \ln 2^j, \quad \forall p \geq 1 \quad (3)$$

- Eqs. (2) + (3):

$$\Rightarrow \zeta(q) = \sum_{p=1}^{\infty} c_p \frac{q^p}{p!}$$

Log-Cumulants

Some classes of processes:

- Eq. (1) $\longrightarrow \mathbb{E}L_X(j, \cdot)^q = F_q |2^j|^{\zeta(q)}$ (2)
- $\ln \mathbb{E}e^{q \ln L_X(j, \cdot)} = \sum_{p=1}^{\infty} C_p^j \frac{q^p}{p!}$

C_p^j - **cumulants** of random variable $\ln L_X(j, \cdot)$

$$C_p^j = c_p^0 + c_p \ln 2^j, \quad \forall p \geq 1 \quad (3)$$

- Eqs. (2) + (3):

$$\Rightarrow \zeta(q) = \sum_{p=1}^{\infty} c_p \frac{q^p}{p!}$$

Log-Cumulants

Some classes of processes:

- Eq. (1) $\longrightarrow \mathbb{E}L_X(j, \cdot)^q = F_q |2^j|^{\zeta(q)}$ (2)
- $\ln \mathbb{E}e^{q \ln L_X(j, \cdot)} = \sum_{p=1}^{\infty} C_p^j \frac{q^p}{p!}$

C_p^j - **cumulants** of random variable $\ln L_X(j, \cdot)$

$$C_p^j = c_p^0 + c_p \ln 2^j, \quad \forall p \geq 1 \quad (3)$$

- Eqs. (2) + (3):

$$\Rightarrow \zeta(q) = \sum_{p=1}^{\infty} c_p \frac{q^p}{p!}$$

Log-Cumulants

- Measurement of $\zeta(q)$ replaced by those of **log-cumulants** c_p :

$$\zeta(q) = c_1 q + c_2 \frac{q^2}{2} + c_3 \frac{q^3}{6} + \dots$$

- X **monofractal**: $\zeta(q)$ linear
 $\Rightarrow \forall p > 1 : c_p \equiv 0$
- X **multifractal**: $\zeta(q)$ non linear
 $\Rightarrow \exists p > 1 : c_p \neq 0$
- Estimation of c_p ?

Log-Cumulants

- Measurement of $\zeta(q)$ replaced by those of **log-cumulants** c_p :

$$\zeta(q) = c_1 q + c_2 \frac{q^2}{2} + c_3 \frac{q^3}{6} + \dots$$

- X **monofractal**: $\zeta(q)$ linear
 $\Rightarrow \forall p > 1 : c_p \equiv 0$
- X **multifractal**: $\zeta(q)$ non linear
 $\Rightarrow \exists p > 1 : c_p \neq 0$
- Estimation of c_p ?

Estimating the log-cumulants c_p

- At each scale j :
 - compute n_j **Leaders** $L_X(j, k)$
 - estimate **cumulants** \hat{C}_p^j of $\ln L_X(j, \cdot)$
- \hat{c}_p : $C_p^j = c_p^0 + c_p \ln 2^j \rightarrow$ **linear regressions** \hat{C}_p^j vs. $\ln 2^j$

$$\hat{c}_p = \log_2 e \sum_{j=j_1}^{j_2} w_j \hat{C}_p^j$$

Weights w_j : reflect confidence granted to each \hat{C}_p^j
here $w_j = 1/n_j$

Equivalent procedures for coefficients: $L_X(j, k) \rightarrow |d_X(j, k)|$

Estimating the log-cumulants c_p

- At each scale j :
 - compute n_j **Leaders** $L_X(j, k)$
 - estimate **cumulants** \hat{C}_p^j of $\ln L_X(j, \cdot)$
- \hat{c}_p : $C_p^j = c_p^0 + c_p \ln 2^j \rightarrow$ **linear regressions** \hat{C}_p^j vs. $\ln 2^j$

$$\hat{c}_p = \log_2 e \sum_{j=j_1}^{j_2} w_j \hat{C}_p^j$$

Weights w_j : reflect confidence granted to each \hat{C}_p^j
here $w_j = 1/n_j$

Equivalent procedures for coefficients: $L_X(j, k) \rightarrow |d_X(j, k)|$

Estimating the log-cumulants c_p

- At each scale j :
 - compute n_j **Leaders** $L_X(j, k)$
 - estimate **cumulants** \hat{C}_p^j of $\ln L_X(j, \cdot)$
- \hat{c}_p : $C_p^j = c_p^0 + c_p \ln 2^j \rightarrow$ **linear regressions** \hat{C}_p^j vs. $\ln 2^j$

$$\hat{c}_p = \log_2 e \sum_{j=j_1}^{j_2} w_j \hat{C}_p^j$$

Weights w_j : reflect confidence granted to each \hat{C}_p^j
here $w_j = 1/n_j$

Equivalent procedures for coefficients: $L_X(j, k) \rightarrow |d_X(j, k)|$

Recall Goals

- 1 Accurate log-cumulant estimates: Coefficients or Leaders ?
- 2 Confidence intervals from single realization ?
→ Non parametric bootstrap

Recall Goals

- 1 Accurate log-cumulant estimates: Coefficients or Leaders ?
- 2 Confidence intervals from single realization ?
→ **Non parametric bootstrap**

Resampling and Bootstrap Estimates

Non parametric **Bootstrap**:

Statistical properties of estimate from single realization by **repeated resampling from available data**

At each scale j : $\mathcal{L}_j = \{L_X(j, 1), \dots, L_X(j, n_j)\}$

- $\mathcal{L}_j \rightarrow$ estimates \hat{C}_p^j, \hat{c}_p
- $\mathcal{L}_j \rightarrow R$ **bootstrap resamples** $\mathcal{L}_j^{*(1)}, \dots, \mathcal{L}_j^{*(R)}$:
$$\mathcal{L}_j^* = \{L_X^*(j, 1), \dots, L_X^*(j, n_j)\}$$

drawn **blockwise**, with **replacement** from \mathcal{L}_j .
- $\mathcal{L}_j^* \rightarrow R$ **bootstrap estimates** $\hat{C}_p^{j*}, \hat{c}_p^*$: empirical distributions

Resampling and Bootstrap Estimates

Non parametric **Bootstrap**:

Statistical properties of estimate from single realization by
repeated resampling from available data

At each scale j : $\mathcal{L}_j = \{L_X(j, 1), \dots, L_X(j, n_j)\}$

- $\mathcal{L}_j \rightarrow$ estimates \hat{C}_p^j, \hat{c}_p
- $\mathcal{L}_j \rightarrow R$ **bootstrap resamples** $\mathcal{L}_j^{*(1)}, \dots, \mathcal{L}_j^{*(R)}$:
$$\mathcal{L}_j^* = \{L_X^*(j, 1), \dots, L_X^*(j, n_j)\}$$

drawn **blockwise**, with **replacement** from \mathcal{L}_j .
- $\mathcal{L}_j^* \rightarrow R$ **bootstrap estimates** $\hat{C}_p^{j*}, \hat{c}_p^*$: empirical distributions

Resampling and Bootstrap Estimates

Non parametric **Bootstrap**:

Statistical properties of estimate from single realization by **repeated resampling from available data**

At each scale j : $\mathcal{L}_j = \{L_X(j, 1), \dots, L_X(j, n_j)\}$

- $\mathcal{L}_j \rightarrow$ estimates \hat{C}_p^j, \hat{c}_p
- $\mathcal{L}_j \rightarrow R$ **bootstrap resamples** $\mathcal{L}_j^{*(1)}, \dots, \mathcal{L}_j^{*(R)}$:
$$\mathcal{L}_j^* = \{L_X^*(j, 1), \dots, L_X^*(j, n_j)\}$$

drawn **blockwise**, with **replacement** from \mathcal{L}_j .
- $\mathcal{L}_j^* \rightarrow R$ **bootstrap estimates** $\hat{C}_p^{j*}, \hat{c}_p^*$: empirical distributions

Resampling and Bootstrap Estimates

Non parametric **Bootstrap**:

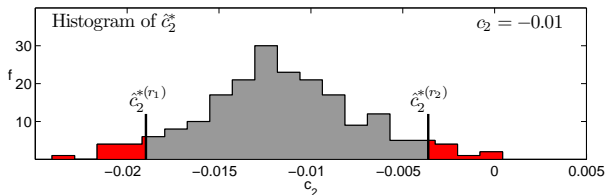
Statistical properties of estimate from single realization by **repeated resampling from available data**

At each scale j : $\mathcal{L}_j = \{L_X(j, 1), \dots, L_X(j, n_j)\}$

- $\mathcal{L}_j \rightarrow$ estimates \hat{C}_p^j, \hat{c}_p
- $\mathcal{L}_j \rightarrow R$ **bootstrap resamples** $\mathcal{L}_j^{*(1)}, \dots, \mathcal{L}_j^{*(R)}$:
$$\mathcal{L}_j^* = \{L_X^*(j, 1), \dots, L_X^*(j, n_j)\}$$

drawn **blockwise**, with **replacement** from \mathcal{L}_j .
- $\mathcal{L}_j^* \rightarrow R$ **bootstrap estimates** $\hat{C}_p^{j*}, \hat{c}_p^*$: empirical distributions

Bootstrap Confidence Intervals



- $100(1 - \alpha)\%$ confidence intervals for c_p s:

$$CI_p = \left(Q_p \left(\frac{\alpha}{2} \right), Q_p \left(1 - \frac{\alpha}{2} \right) \right) = \left(\hat{c}_p^{*(r_1)}, \hat{c}_p^{*(r_2)} \right)$$

- $Q_p(\alpha)$ - α -th quantile of empirical distribution \hat{c}_p^* :
 $r_1 = \lfloor \frac{R\alpha}{2} \rfloor$ and $r_2 = R - r_1 + 1$

Monte Carlo Simulation

- We have now: $\hat{c}_p, CI \rightarrow$ Performance ?
- Apply procedures to **large number** N_{MC} of realizations of synthetic multifractal processes with known scaling properties.
- **Scaling Processes** with stationary increments:
 - Fractional Brownian Motion (**FBM**):
only Gaussian exactly selfsimilar process
 $\zeta(q) = qH, c_1 = H, p \geq 1 : c_p \equiv 0 \leftarrow$ monofractal
 - Multifractal Random Walk (**MRW**):
simple multifractal (hence non Gaussian) process
 $c_1 \neq 0, c_2 \neq 0, p \geq 2 : c_p \equiv 0$

Monte Carlo Simulation

- We have now: $\hat{c}_p, CI \rightarrow$ Performance ?
- Apply procedures to **large number** N_{MC} of realizations of synthetic multifractal processes with known scaling properties.
- **Scaling Processes** with stationary increments:
 - Fractional Brownian Motion (**FBM**):
only Gaussian exactly selfsimilar process
 $\zeta(q) = qH, c_1 = H, p \geq 1 : c_p \equiv 0 \leftarrow$ monofractal
 - Multifractal Random Walk (**MRW**):
simple multifractal (hence non Gaussian) process
 $c_1 \neq 0, c_2 \neq 0, p \geq 2 : c_p \equiv 0$

Monte Carlo Simulation

- We have now: $\hat{c}_p, CI \rightarrow$ Performance ?
- Apply procedures to **large number** N_{MC} of realizations of synthetic multifractal processes with known scaling properties.
- **Scaling Processes** with stationary increments:
 - Fractional Brownian Motion (**FBM**):
only Gaussian exactly selfsimilar process
 $\zeta(q) = qH, c_1 = H, p \geq 1 : c_p \equiv 0 \leftarrow$ monofractal
 - Multifractal Random Walk (**MRW**):
simple multifractal (hence non Gaussian) process
 $c_1 \neq 0, c_2 \neq 0, p \geq 2 : c_p \equiv 0$

Monte Carlo Simulation

- We have now: $\hat{c}_p, CI \rightarrow$ Performance ?
- Apply procedures to **large number** N_{MC} of realizations of synthetic multifractal processes with known scaling properties.
- **Scaling Processes** with stationary increments:
 - Fractional Brownian Motion (**FBM**):
only Gaussian exactly selfsimilar process
 $\zeta(q) = qH, c_1 = H, p \geq 1 : c_p \equiv 0 \leftarrow$ monofractal
 - Multifractal Random Walk (**MRW**):
simple multifractal (hence non Gaussian) process
 $c_1 \neq 0, c_2 \neq 0, p \geq 2 : c_p \equiv 0$

Statistical Performance: MSE

$$\text{MSE} = \sqrt{\text{Var}_{\text{MC}}\{\hat{c}_p\} + (c_p - \mathbb{E}_{\text{MC}}\{\hat{c}_p\})^2}$$

MSE $\times 10^3$	FBM				
	c_1	c_2	c_3	c_4	c_5
Coefficients	15.5	37.3	187.7	1251	9803
Leaders	10.8	4.1	1.8	1.0	0.7

MSE $\times 10^3$	MRW				
	c_1	c_2	c_3	c_4	c_5
Coefficients	35.3	42.8	200.7	1366	11068
Leaders	32.8	17.5	18.4	30.0	50.1

→ Leaders significantly outperform coefficients

Statistical Performance: MSE

$$\text{MSE} = \sqrt{\text{Var}_{\text{MC}}\{\hat{c}_p\} + (c_p - \mathbb{E}_{\text{MC}}\{\hat{c}_p\})^2}$$

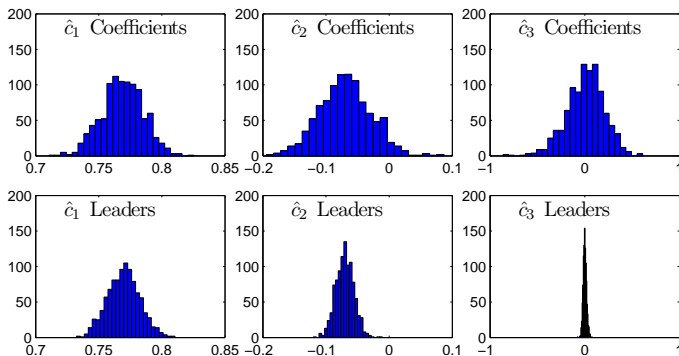
MSE $\times 10^3$	FBM				
	c_1	c_2	c_3	c_4	c_5
Coefficients	15.5	37.3	187.7	1251	9803
Leaders	10.8	4.1	1.8	1.0	0.7

MSE $\times 10^3$	MRW				
	c_1	c_2	c_3	c_4	c_5
Coefficients	35.3	42.8	200.7	1366	11068
Leaders	32.8	17.5	18.4	30.0	50.1

→ Leaders significantly outperform coefficients

Statistical Performance continued

Monte Carlo empirical distributions of estimates:



Confidence Intervals by Bootstrap

Nominal Coverage: 95%

$$\text{Empirical Coverage} = \frac{\#(c_p \in CI_p)}{N_{MC}} \quad (\text{bias corrected})$$

Empirical Coverage (in %)	FBM				
	c_1	c_2	c_3	c_4	c_5
Coefficients	92.1	92.3	91.3	87.8	87.7
Leaders	83.4	90.3	94.7	97.2	98.0

Empirical Coverage (in %)	MRW				
	c_1	c_2	c_3	c_4	c_5
Coefficients	98.6	95.5	92.6	90.3	90.8
Leaders	98.8	97.0	96.4	95.4	96.6

→ Bootstrap CI: nominal coverage reproduced satisfactorily

Confidence Intervals by Bootstrap

Nominal Coverage: 95%

$$\text{Empirical Coverage} = \frac{\#(c_p \in CI_p)}{N_{MC}} \quad (\text{bias corrected})$$

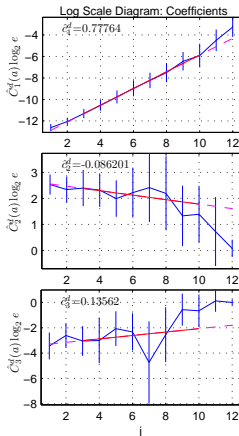
Empirical Coverage (in %)	FBM				
	c_1	c_2	c_3	c_4	c_5
Coefficients	92.1	92.3	91.3	87.8	87.7
Leaders	83.4	90.3	94.7	97.2	98.0

Empirical Coverage (in %)	MRW				
	c_1	c_2	c_3	c_4	c_5
Coefficients	98.6	95.5	92.6	90.3	90.8
Leaders	98.8	97.0	96.4	95.4	96.6

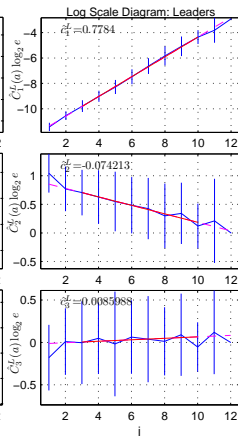
→ Bootstrap CI: nominal coverage reproduced satisfactorily

Practical Procedure

Coefficients



Leaders



$$C_p^j = c_p^0 + c_p \ln 2^j$$

From single realization:

- Estimates \hat{c}_p
- Bootstrap CI for c_p
- Bootstrap CI for C_p^j :
 → regression range

Conclusions and Perspectives

- **Leaders** based estimation procedure **significantly outperforms** Coefficients based one
- Bootstrap provides **highly relevant confidence intervals** for log-cumulants c_p
- Practical procedure that can be applied to a **single, finite sample** of empirical data
- Perspectives:
 - Advanced bootstrap techniques (pivoting, adjusted limits, ...)
 - Bootstrap **hypothesis tests** on c_p : testing mono-vs. multifractal

Conclusions and Perspectives

- **Leaders** based estimation procedure **significantly outperforms** Coefficients based one
- Bootstrap provides **highly relevant confidence intervals** for log-cumulants c_p
- Practical procedure that can be applied to a **single, finite sample** of empirical data
- Perspectives:
 - Advanced bootstrap techniques (pivoting, adjusted limits, ...)
 - Bootstrap **hypothesis tests** on c_p : testing mono-vs. multifractal

Conclusions and Perspectives

- **Leaders** based estimation procedure **significantly outperforms** Coefficients based one
- Bootstrap provides **highly relevant confidence intervals** for log-cumulants c_p
- Practical procedure that can be applied to a **single, finite sample** of empirical data
- Perspectives:
 - Advanced bootstrap techniques (pivoting, adjusted limits, ...)
 - Bootstrap **hypothesis tests** on c_p : testing mono-vs. multifractal

Conclusions and Perspectives

- **Leaders** based estimation procedure **significantly outperforms** Coefficients based one
- Bootstrap provides **highly relevant confidence intervals** for log-cumulants c_p
- Practical procedure that can be applied to a **single, finite sample** of empirical data
- Perspectives:
 - Advanced bootstrap techniques (pivoting, adjusted limits, ...)
 - Bootstrap **hypothesis tests** on c_p : testing mono-vs. multifractal

Multifractal Analysis

Multifractal Analysis:

- Local regularity of X at t : Hölder exponent $h(t)$
 X is $C^\alpha(t_0)$ if $\exists C, \alpha > 0; P_{t_0}(t); \deg(P_{t_0}) < \alpha :$
$$|X(t) - P_{t_0}(t)| < C|t - t_0|^\alpha$$

$$\rightarrow h(t_0) = \sup_\alpha \{\alpha : X \in C^\alpha(t_0)\}$$
- Singularity spectrum $D(h)$:
Hausdorff dimensions of $\{t_i | h(t_i) = h\}$

Empirical Multifractal Analysis:

$D(h)$ obtained as Legendre transform of estimates of $\zeta(q)$

