TIME-SCALE BLOCK BOOTSTRAP TESTS FOR NON GAUSSIAN FINITE VARIANCE SELF-SIMILAR PROCESSES WITH STATIONARY INCREMENTS

Herwig Wendt, Patrice Abry

Physics Lab., CNRS UMR 5672, Ecole Normale Supérieure de Lyon, France.
herwig.wendt@ens-lyon.fr, patrice.abry@ens-lyon.fr

SCALING (OR MULTIFRACTAL) ANALYSIS

Scaling

Scale Invariance

$X(t),\ t \in [0, n)$ - Process under analysis

$T_X(a,k)$ - Multiresolution quantities of X jointly depend on analysis scale a and time position t

$$\boxed{\frac{1}{n_a}\sum_{k=1}^{n_a} |T_X(a,k)|^q \simeq c_q |a|^{\zeta(q)}}$$

for statistical orders $q \in [q_*^-, q_*^+]$

for scales $a = 2^j \in [a_m, a_M]$, $a_M/a_m >> 1$

$\zeta(q)$ - Scaling exponents

$D(h)$ - Multifractal spectrum (can be obtained as the Legendre transform of $\zeta(q)$)

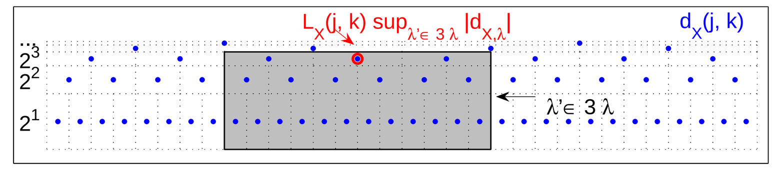
Multiresolution Quantities

Discrete Wavelet Coefficients (vanishing moments N_ψ)

$d_X(j,k) = \langle \psi_{j,k} | X \rangle$

Wavelet Leaders

$L_X(j,k) = \sup_{\lambda' \in 3\lambda_{j,k}} |d_{\lambda'}|$



Scaling Estimation

Structure Functions

$S(2^j,q) = \frac{1}{n_j}\sum_{k=1}^{n_j} |T_X(2^j,k)|^q \simeq c_q 2^{j\zeta(q)}$

Log-Cumulants

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

- $\forall p \geq 1:$ $C(2^j,p) = c_p^0 + c_p \ln 2^j$
  $C(2^j,p)$: p-th cumulant of $\ln |T_X(2^j,\cdot)|$

Estimation Procedures

a) calculate n_j coefficients $T_X(2^j,k)$

b) $S(2^j,q) = \frac{1}{n_j}\sum_{k=1}^{n_j} |T_X(2^j,k)|^q$
  $\hat C(2^j,p)$ - standard estimate of p-th cumulant of $\ln |T_X(2^j,\cdot)|$

c) Linear regressions:
  $\hat\zeta(q) = \sum_{j=j_1}^{j_2} w_j \log_2 S(2^j,q)$
  $\hat c_p = (\log_2 e) \cdot \sum_{j=j_1}^{j_2} w_j \hat C(2^j,p)$

Scaling Modeling

H-sssi processes (e.g. fractional Brownian motion (FBM))

1. deeply tied to additive random walks
2. one single self-similarity parameter H
3. well-defined fully parametric class of processes

Multifractal (MF) processes (e.g. Mandelbrot's cascades)

1. mostly recursive multiplicative constructions
2. whole collection of parameters
3. broad and not well defined class of processes

Consequences
1. Significantly different physical or biological mechanisms underlying data
2. Scaling parameters $\longleftrightarrow$ *physical* parameters
3. MF: extra parameters = better fits, BUT derivation of estimation / test procedures / statistical performance delicate

$\Rightarrow$ Need for practical procedures deciding whether H-sssi or MF are better model for data

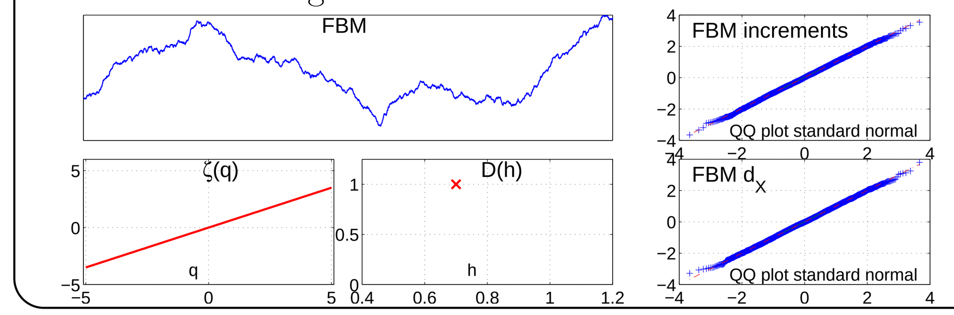
Finite Variance Self Similar Processes

Self similar with stationary increments (H-sssi):

- $\{X(t)\}_{t\in\mathbb{R}} \overset{fdd}{=} \{a^H X(t/a)\}_{t\in\mathbb{R}}, \forall a > 0$
- $\{X(t+a) - X(t)\}_{t\in\mathbb{R}} \overset{fdd}{=} \{X(a)\}_{a\in\mathbb{R}}$
- self-similarity parameter H, $0 < H < 1$

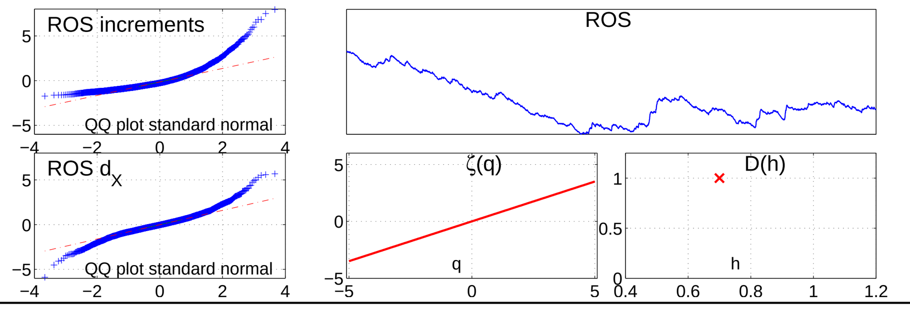
Fractional Brownian motion (FBM):

- only Gaussian H-sssi process
- Scaling properties: $\boxed{\zeta(q) = qH,\ c_1 = H,\ \forall p \geq 2:\ c_p \equiv 0}$
  - Long range dependence for $H \in (1/2, 1)$
  - Gaussian marginal distributions



Rosenblatt process (ROS):

- Most well known finite variance non Gaussian H-sssi process
- $Z_H(t) = k_H \int'_{\mathbb{R}^2} \left\{ \int_0^t (s-u)_+^{\frac{H}{2}-1} (s-v)_+^{\frac{H}{2}-1} ds \right\} dB(u)dB(v)$
  $\int'_{\mathbb{R}^2}$ - double Wiener-Itô integral
  $H \in (1/2, 1)$, $x_+ = \max\{x, 0\}$ for $x \in \mathbb{R}$
  $B(t),\ t \in \mathbb{R}$ - standard Brownian motion
- Scaling properties: $\boxed{\zeta(q) = qH,\ c_1 = H,\ \forall p \geq 2:\ c_p \equiv 0}$
  - Same covariance as FBM, but higher order dependence
  - non Gaussian highly skewed marginal distributions



Summary Recently, we showed that bootstrap procedures for scaling analysis efficiently discriminate Gaussian self similar processes from multifractal processes. In the present contribution, we propose a new joint time-scale block bootstrap and investigate the relevance of estimation and bootstrap procedures to discriminate non Gaussian finite variance self similar processes with stationary increments and multifractal processes.

BOOTSTRAP PROCEDURES

Time Block and Time-Scale Block Bootstrap

Bootstrap Procedure

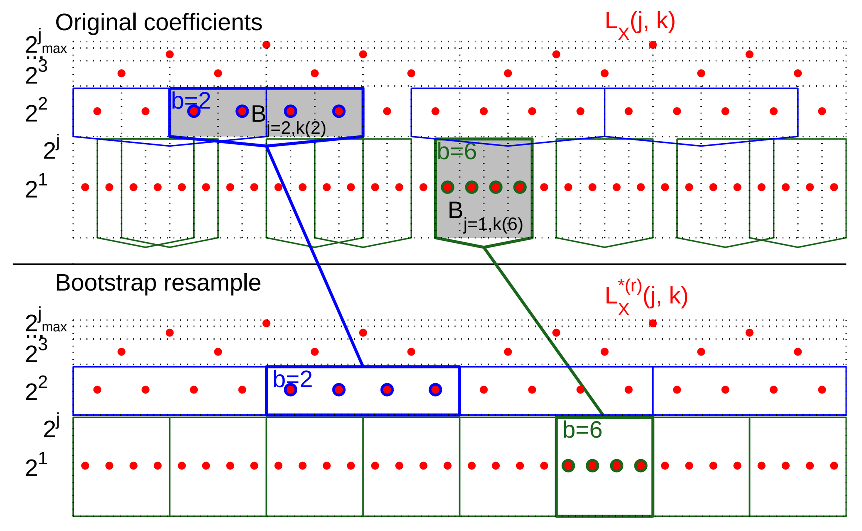
For $r = 1, \cdots, R$:

a) Draw randomly, with replacement, blocks of coefficients from original sample L_X $\longrightarrow$ resample $L_X^{*(r)}$

b) Calculate $S^{*(r)}(2^j,q)$ and $\hat C^{*(r)}(2^j,p)$ on $L_X^{*(r)}$

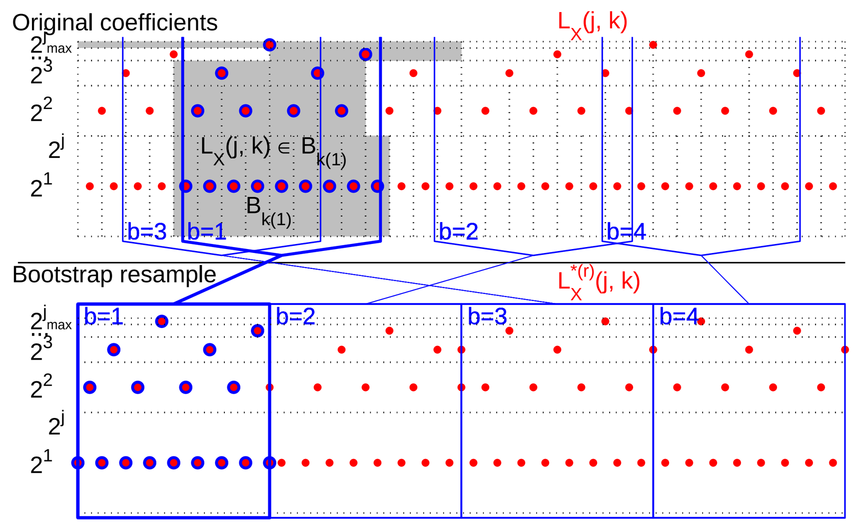
c) Calculate $\hat\zeta^{*(r)}(q)$ and $\hat c_p^{*(r)}$ by linear regressions on $S^{*(r)}(2^j,q)$ and $\hat C^{*(r)}(2^j,p)$

Time Block



Blocks $B_{j,k(b)}$ of $2N_\psi$ coefficients drawn at each scale independently
$\longrightarrow$ capture time dependence of coefficients

Time-Scale Block



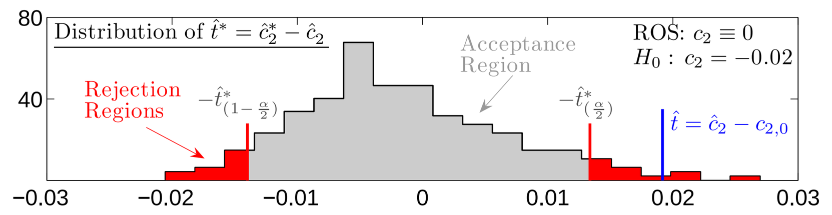
Blocks $B_{k(b)}$ are stripes of time length $2l$ in the time-scale plane
$\longrightarrow$ capture joint time-scale dependence of coefficients

Bootstrap Confidence Intervals and Tests

Bootstrap hypothesis tests

Test d_α:

- test $H_0:\ c_p = c_{p,0}$ against double sided alternative $c_p \neq c_{p,0}$
  $\hat t = \hat c_p - c_{p,0}$ - test statistic
  $\left\{\hat t^{*(r)} = \hat c_p^{*(r)} - \hat c_p\right\}_{r=1}^{R}$ - null distribution estimate
  (α quantile $\hat t^*_\alpha$)
  $\hat{\mathcal{T}}_{(1-\alpha)} = [-\hat t^*_{(1-\frac{\alpha}{2})}, -\hat t^*_{\frac{\alpha}{2}}]$ - acceptance region
- $d_\alpha = 1$ if $\hat t \notin \hat{\mathcal{T}}_{(1-\alpha)}$, and $d_\alpha = 0$ otherwise.



Variance and confidence interval estimation

- $\hat\sigma^{*2}_{\hat\Theta} = \widehat{\mathrm{Var}}^*\hat\Theta^*$ - sample variance of bootstrap resamples ($\hat\Theta \in \{\hat\zeta(q), \hat c_p\}$)
- $\widehat{\mathrm{CI}}_\Theta = \left[\hat\Theta^*_{\frac{\alpha}{2}}, \hat\Theta^*_{(1-\frac{\alpha}{2})}\right]$ - equi-tailed percentile confidence limit

References

S. Jaffard, "Wavelet techniques in multifractal analysis," in *Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot, M. Lapidus et M. van Frankenhuijsen Eds., Proceedings of Symposia in Pure Mathematics.* 2004, vol. 72(2), pp. 91–152, AMS.

R. H. Riedi, "Multifractal processes," *in "Theory and applications of long range dependence", eds. Doukhan, Oppenheim and Taqqu, Birkhäuser,* 2003.

H. Wendt, P. Abry, S. Jaffard, "Bootstrap for empirical multifractal analysis with application to hydrodynamic turbulence," *IEEE Signal Processing Mag.*, vol. 24, no. 4, pp. 38–48, 2007.

H. Wendt, P. Abry, "Multifractality tests using bootstrapped wavelet leaders," *IEEE Trans. on Signal Processing*, to appear, 2007.

G. Samorodnitsky, M. Taqqu, *Stable non-Gaussian random processes*, Chapman and Hall, New York, 1994.

M. Rosenblatt, "Independence and dependence," in *Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability*, University of Carolina Press, Berkeley, CA, 1961, vol. 2, pp. 431–443.

P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. on Info. Theory*, vol. IT-38, no. 2, pp. 910–917, 1992.

P. Abry, V. Pipiras, "Wavelet-based synthesis of the rosenblatt process," *Signal Processing*, vol. 86, no. 9, pp. 2326–2339, 2006.

B. Efron, *The Jackknife, the Bootstrap, and Other Resampling Plans*, SIAM, 1982.

S.N. Lahiri, *Resampling Methods for Dependent Data*, Springer, New York, 2003.

RESULTS

Monte Carlo Simulation

Experimental set-up

Apply procedures to N_{MC} realizations of length n of FBM and ROS

- $N_{MC} = 500$, $n = 2^{15}$
- $H = 0.7$
- $N_\psi = 3$, $[j_1, j_2] = [3, 8]$
- $R = 199$, $l = 2^{10}$

Estimation performance evaluation

- bias
  $\gamma_{\hat\Theta} = \hat{\mathbb{E}}_{N_{MC}}\hat\Theta - \Theta$
- standard deviation
  $\sigma_{\hat\Theta} = \sqrt{\widehat{\mathrm{Var}}_{N_{MC}}\hat\Theta}$
- (root) mean squared error
  $\mathrm{MSE}_{\hat\Theta} = \sqrt{\sigma_{\hat\Theta}^2 + \gamma_{\hat\Theta}^2}$

($\widehat{\mathbb{E}}_{N_{MC}}$, $\widehat{\mathrm{Var}}_{N_{MC}}$ - sample mean/variance N_{MC} realizations)

Bootstrap performance evaluation

- bootstrap mean standard deviation
  $\sigma_{\hat\Theta*} = \sqrt{\widehat{\mathbb{E}}_{N_{MC}}\hat\sigma^{*2}_{\hat\Theta}}$
- bias-corrected empirical coverage of CI
  $\mathcal{C}_{\Theta,bc} = \widehat{\mathbb{E}}_{N_{MC}}\mathbb{I}\left\{\Theta + \gamma^{\hat\Theta} \in \widehat{\mathrm{CI}}_\Theta\right\}$
- rejection rates of tests $H_0:\ c_p = c_{p,0}$
  $\hat\beta(\alpha, c_{p,0}, c_p) = \widehat{\mathbb{E}}_{N_{MC}}\left\{\hat d_\alpha | c_{p,0}, c_p\right\}$

($\mathbb{I}\{\cdot\}$ - indicator function of event $\{\cdot\}$)

Performance of Estimation

FBM		d_X			L_X		
Θ		bias	std	mse	bias	std	mse
$\zeta(-2)$	−1.400	−0.995	1.171	1.537	0.033	0.018	0.038
$\zeta(-1)$	−0.700	−0.107	0.323	0.340	0.014	0.009	0.017
$\zeta(1)$	0.700	−0.002	0.012	0.012	−0.009	0.009	0.013
$\zeta(2)$	1.400	−0.004	0.023	0.023	−0.014	0.020	0.024
c_1	0.700	−0.002	0.017	0.017	−0.011	0.009	0.014
c_2	0.000	−0.001	0.042	0.042	0.005	0.003	0.006
c_3	0.000	0.008	0.217	0.217	−0.001	0.001	0.002

ROS		d_X			L_X		
Θ		bias	std	mse	bias	std	mse
$\zeta(-2)$	−1.400	−1.002	1.101	1.489	0.026	0.035	0.043
$\zeta(-1)$	−0.700	−0.114	0.266	0.289	0.013	0.018	0.022
$\zeta(1)$	0.700	0.001	0.023	0.023	−0.011	0.022	0.024
$\zeta(2)$	1.400	−0.004	0.052	0.052	−0.021	0.050	0.054
c_1	0.700	0.004	0.024	0.024	−0.012	0.019	0.023
c_2	0.000	−0.006	0.043	0.044	0.002	0.009	0.009
c_3	0.000	0.006	0.211	0.211	0.001	0.006	0.006

Conclusions

$\zeta(q<0)$ - d_X meaningless, L_X good

$\zeta(q>0)$ - d_X and L_X good and
and c_1 similar performance

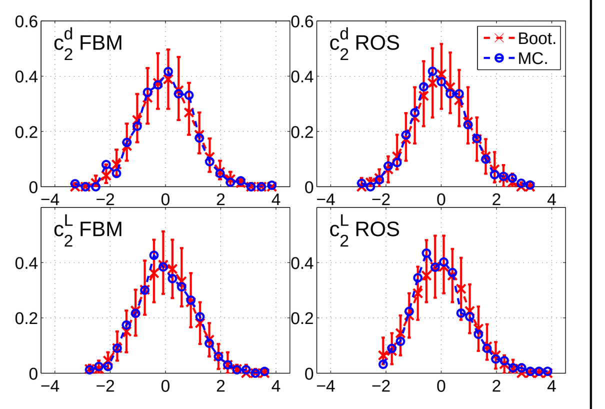
c_2, c_3 - L_X excellent and much better than d_X

ROS - estimation more difficult
c_p by L_X remain excellent

Performance of Bootstrap Estimation

FBM	time-scale block						time block	
	d_X			L_X			L_X	
Θ	$\sigma_{\hat\Theta}$	$\sigma_{\hat\Theta*}$	$\mathcal{C}_{\Theta,bc}$	$\sigma_{\hat\Theta}$	$\sigma_{\hat\Theta*}$	$\mathcal{C}_{\Theta,bc}$	$\sigma_{\hat\Theta*}$	$\mathcal{C}_{\Theta,bc}$
$\zeta(-2)$	1.171	0.718	82.0	0.018	0.017	91.6	0.014	87.0
$\zeta(-1)$	0.323	0.200	88.6	0.009	0.008	91.4	0.007	87.1
$\zeta(1)$	0.012	0.012	93.4	0.009	0.009	94.0	0.007	87.4
$\zeta(2)$	0.023	0.023	94.4	0.020	0.019	94.6	0.016	89.0
c_1	0.017	0.016	94.2	0.009	0.009	92.2	0.007	87.7
c_2	0.042	0.040	91.8	0.003	0.003	93.6	0.003	92.8
c_3	0.217	0.200	90.8	0.001	0.001	94.0	0.001	94.7

ROS	time-scale block						time block	
	d_X			L_X			L_X	
Θ	$\sigma_{\hat\Theta}$	$\sigma_{\hat\Theta*}$	$\mathcal{C}_{\Theta,bc}$	$\sigma_{\hat\Theta}$	$\sigma_{\hat\Theta*}$	$\mathcal{C}_{\Theta,bc}$	$\sigma_{\hat\Theta*}$	$\mathcal{C}_{\Theta,bc}$
$\zeta(-2)$	1.101	0.703	81.8	0.035	0.026	88.0	0.020	78.5
$\zeta(-1)$	0.266	0.181	89.4	0.018	0.013	86.8	0.010	76.9
$\zeta(1)$	0.023	0.017	84.4	0.022	0.015	85.2	0.012	74.3
$\zeta(2)$	0.052	0.037	84.0	0.050	0.036	83.8	0.030	75.7
c_1	0.024	0.019	88.8	0.019	0.014	85.8	0.011	76.0
c_2	0.043	0.041	93.2	0.009	0.008	88.8	0.006	85.0
c_3	0.211	0.196	91.0	0.006	0.006	87.6	0.005	85.0



Variance estimation and confidence intervals

FBM $\zeta(q)$, c_p - estimation of $\sigma_{\hat\Theta}$ and CI excellent
under-estimation < 5%

ROS $\zeta(q)$, c_1 - estimation of $\sigma_{\hat\Theta}$ and CI more difficult
under-estimation < 25% ($\sigma_{\hat\Theta}$), < 10% (CI)

c_2, c_3 - estimation of $\sigma_{\hat\Theta}$ and CI excellent
under-estimation < 5% ($\sigma_{\hat\Theta}$), < 7% (CI)

Time block vs. time-scale block bootstrap

d_X - no difference in performance (results not reported here)

L_X - time-scale block bootstrap better than time block bootstrap
small difference for FBM
significant difference for ROS
$\longrightarrow$ non-linear construction of L_X introduces inter-scale dependence

Performance of Bootstrap Hypothesis Tests

$H_0:\ c_p = c_{p,0}$ when c_p: $\{c_1, c_2, c_3\} = \{0.7, 0, 0\}$

Significance

Rejection rate $c_{p,0} \equiv c_p$

- $100 - \mathcal{C}_{\Theta,(bc)}$ - same conclusions as for CI

Power

Rejection rate $c_{p,0} \neq c_p$

- for $p = 2{,}3$: tests rejecting MF
- tests much more powerful for L_X
- rejecting MF more difficult for ROS