Counting the number of different scaling exponents in multivariate scale-free dynamics: Clustering by bootstrap in the wavelet domain

C.-G. Lucas[1], P. Abry[1], H. Wendt[2], G. Didier[3]

[1] ENSL, CNRS, Laboratoire de physique, F-69342 Lyon, France, firstname.lastname@ens-lyon.fr
[2] IRIT, Univ. Toulouse, CNRS, Toulouse, France, herwig.wendt@irit.fr
[3] Math. Dept., Tulane University, New Orleans, USA, gdidier@tulane.edu

Goals

- Multivariate self-similarity: model for multivarate data with scale-free dynamics
- Eigen-wavelet estimation for the vector of self-similarity exponents: $\underline{H} = (H_1, \dots, H_M)$
- Count the number of H_m actually different
- Count the number of components of $\underline{H}$ with same H_m

Methods

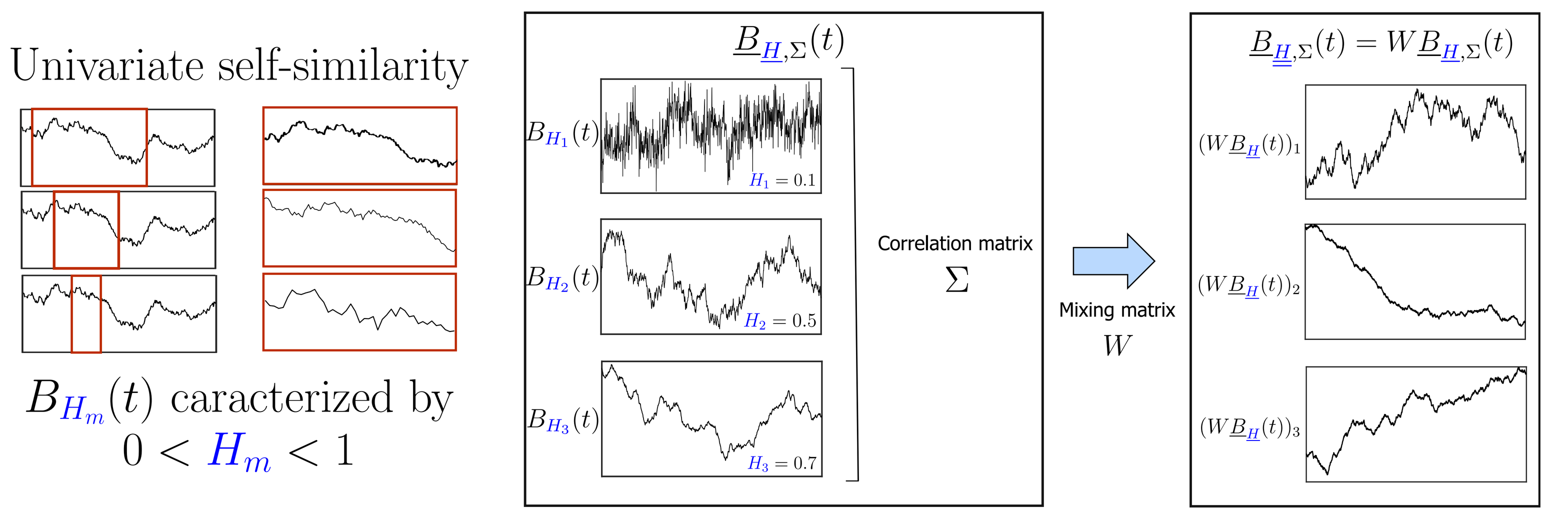
- Pairwise tests $H_m = H_{m+1}$
- Half-normal test statistics under the null hypothesis
- Multivariate wavelet block-bootstrap for test statistics estimation
- Multiple hypothesis corrections and clustering

Conclusions and perspectives

- Bootstrap reproduces the null Hypothesis
- Decent clustering performance
- Non ranked pairwise tests $H_m = H_{m'}$?
- Large dimension ?

Multivariate self-similarity

Model [Didier et al., 2011]

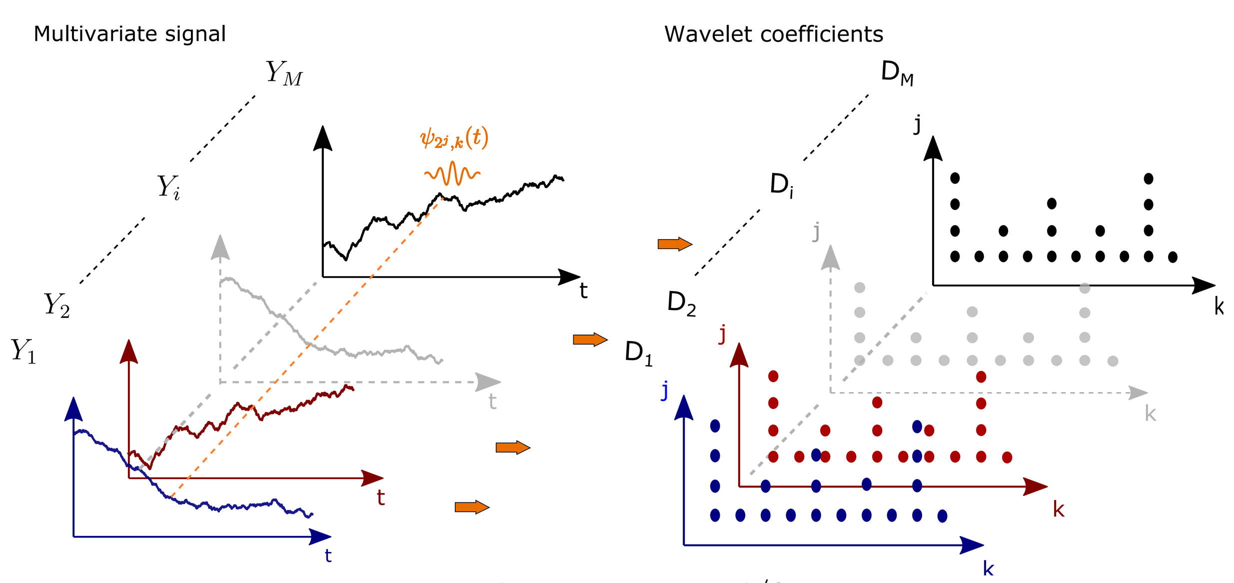


Multivariate self-similarity exponent: $\underline{H} = (H_1, \dots, H_M)$,

$$0 < H_1 \leq \dots \leq H_M < 1$$

Estimation [Lucas et al., 2021]

1. Multivariate wavelet transform:



- Univariate wavelets $D_m(2^j, k) = \langle 2^{-j/2}\psi_{j,k}(t) | Y_m(t) \rangle$
- Multivariate wavelets $D(2^j, k) = (D_1(2^j, k), \dots, D_M(2^j, k))$

2. Wavelet spectra computed from n_{j_2} wavelet coefficients:

$$S^{(w)}(2^j) \triangleq \frac{1}{n_{j_2}} \sum_{k=1+(w-1)n_{j_2}}^{wn_{j_2}} D(2^j,k)D(2^j,k)^*, \ w = 1, \dots, 2^{j-j_2}$$

3. Eigenvalues of $S^{(w)}(2^j)$: $\{\lambda_1^{(w)}(2^j), \dots, \lambda_M^{(w)}(2^j)\}$
   → similar repulsion between eigenvalues at all scales $j \in \{j_1, \dots, j_2\}$
   → asymptotical power law: $\lambda_m^{(w)}(2^j) \propto 2^{j(2H_m+1)}$

4. Averaged log-eigenvalues: $\bar{\lambda}_m(2^j) \triangleq 2^{j_2-j} \sum_{w=1}^{2^{j-j_2}} \log_2(\lambda_m^{(w)}(2^j))$

5. Linear regression: $\hat{H}_m = \frac{1}{2}\sum_{j=j_1}^{j_2} \omega_j \bar{\lambda}_m(2^j) + \frac{1}{2}$

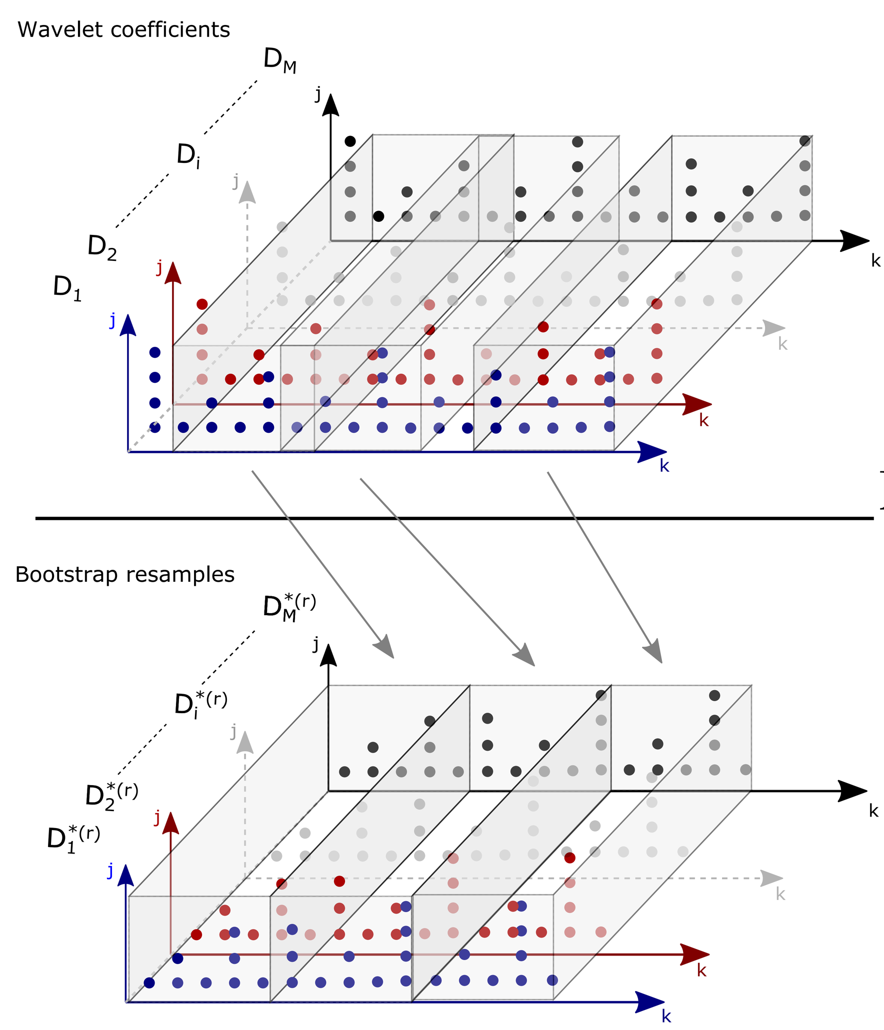
Clustering strategy

Test procedure from a single observation

1. $M-1$ null hypotheses: $H_m = H_{m+1}, \ m = 1, \dots, M-1$
2. Sorted estimates: $\hat{\underline{H}}_{\tau(\cdot)} = \mathrm{sort}(\hat{\underline{H}})$
3. Test statistics: $\tilde{\delta}_m = \hat{H}_{\tau(m+1)} - \hat{H}_{\tau(m)}$
4. Under null hypothesis, $\tilde{\delta}_m \simeq$ half-normal (σ_m)
5. Test decisions: rejects $H_m = H_{m+1}$ if $\tilde{\delta}_m > \gamma_m(\sigma_m)$

σ_m unknown ⇒ bootstrap estimation

Bootstrap resampling



⇒ R resamples
$D^{*(r)} = (D_1^{*(r)}, \dots, D_M^{*(r)})$
⇓
R bootstrap estimates
$\hat{\underline{H}}^{*(r)} = (\hat{H}_1^{*(r)}, \dots, \hat{H}_M^{*(r)})$
⇓
Null hypothesis reproduction:
$\bar{H}_m^{*(r)} = \hat{H}_m^{*(r)} - \langle \hat{H}_m^* \rangle$
⇓
Sorted bootstrap estimates:
$\bar{H}^{*(r)}_{\tau^*(r,1)} < \dots < \bar{H}^{*(r)}_{\tau^*(r,M)}$
⇓
Bootstrap test statistics
$\tilde{\delta}_m^{*(r)} = \bar{H}^{*(r)}_{\tau^*(r,m+1)} - \bar{H}^{*(r)}_{\tau^*(r,m)}$

⇒ Scale parameter estimate: $\hat{\sigma}_m^{*2} = \mathrm{Var}^*(\tilde{\delta}_m^*) / \left(1 - \frac{2}{\pi}\right)$

Multiple hypothesis corrections

1. Bootstrap test p-values: $p_m^* \triangleq 1 - F\left(\frac{\tilde{\delta}_m}{\hat{\sigma}_m^*}\right)$
   F: standardized half-normal cumulative distribution function
2. False discovery rate α
3. Sorted p-values $p^*_{\pi(m)}$
4. Benjamini-Hochberg corrections: $d_\alpha^{(m)} = 1 : p^*_{\pi(m)} < \frac{\alpha}{M-1} m$

Clustering procedure

Rule: $d_\alpha^{(m)} = 1 \Leftrightarrow H_m$ and H_{m+1} in different clusters

Performance evaluation

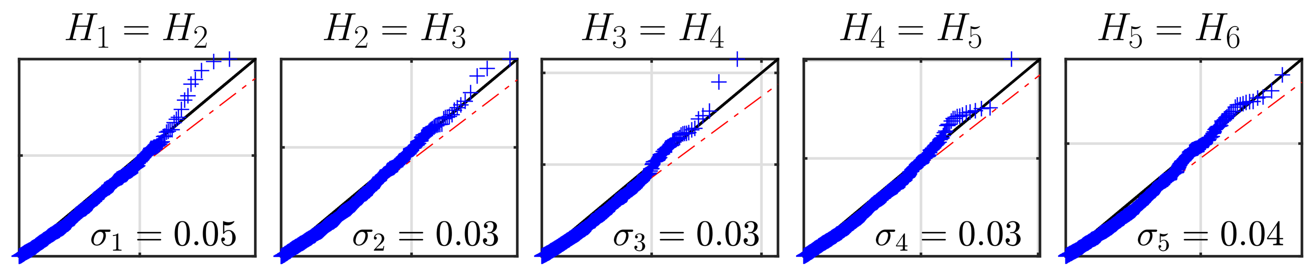
Monte Carlo simulations

$N_{MC} = 1000$ realizations, $M = 6$ components, sample size $N = 2^{16}$

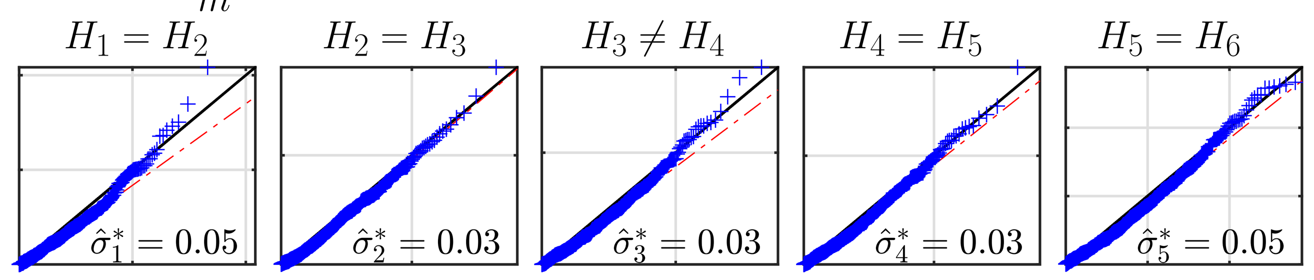
Scenario1 (1 cluster): $\underline{H} = (0.8, 0.8, 0.8, 0.8, 0.8, 0.8)$
Scenario2 (2 clusters): $\underline{H} = (0.6, 0.6, 0.6, 0.8, 0.8, 0.8)$
Scenario3 (3 clusters): $\underline{H} = (0.4, 0.4, 0.6, 0.6, 0.8, 0.8)$
Scenario4 (3 clusters): $\underline{H} = (0.4, 0.6, 0.6, 0.6, 0.8, 0.8)$

Reproduction of the statistic

Q-Q plot of $\tilde{\delta}_m$ vs. half-normal distribution under Scenario1

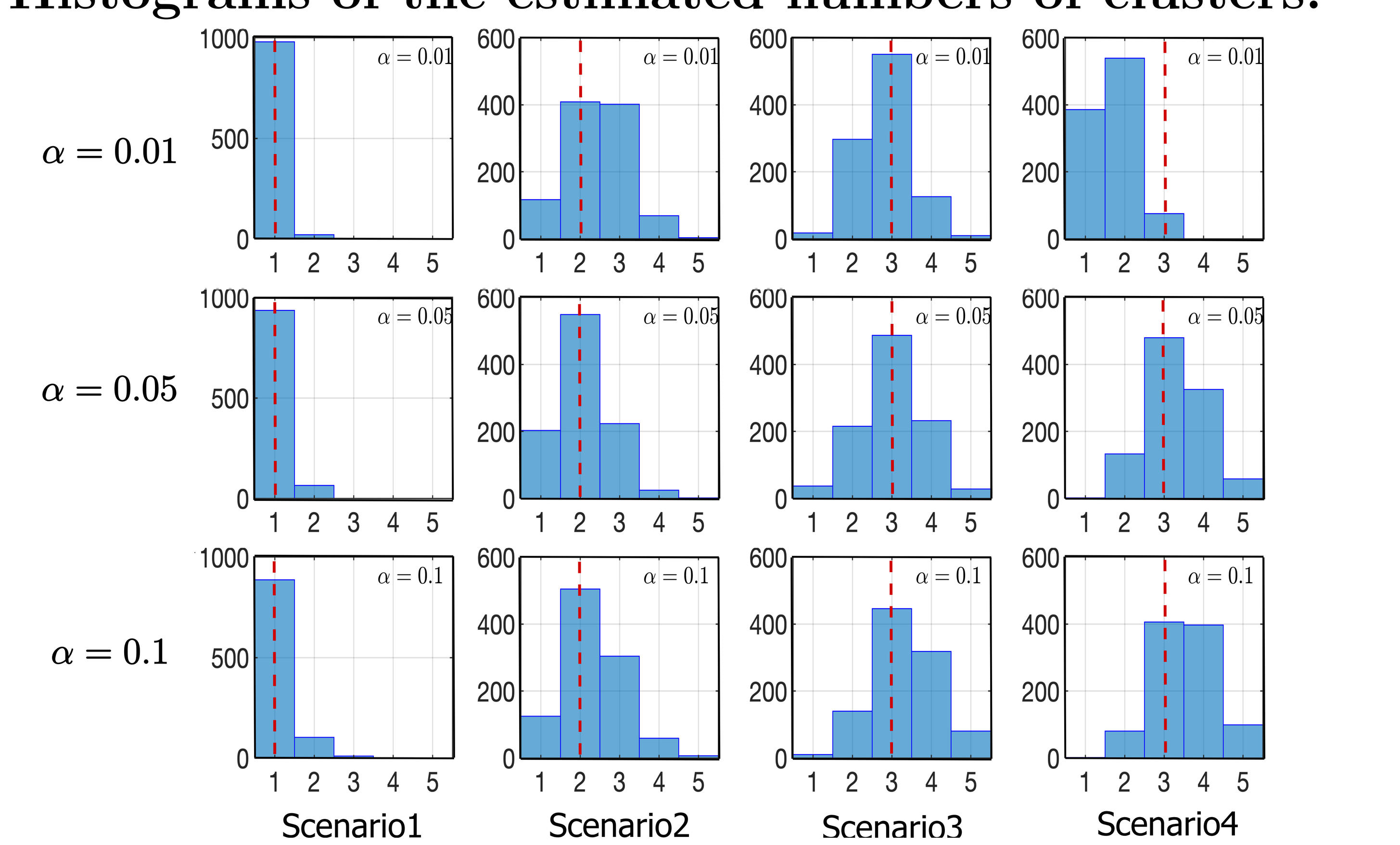


Q-Q plot of $\tilde{\delta}_m^*$ vs. half-normal distribution under Scenario2



Clustering performance

Histograms of the estimated numbers of clusters.



[Didier et al., 2011] G. Didier and V. Pipiras, "Integral representations and proper- ties of operator fractional Brownian motions," Bernoulli, vol. 17, no. 1, pp. 1–33, 2011.
[Lucas et al., 2021] C.-G. Lucas, P. Abry, H. Wendt, and G. Didier, "Bootstrap for testing the equality of selfsimilarity exponents across multivariate time series," in Proc. European Signal Processing Conference (EUSIPCO), Dublin, Ireland, August 2021.