Semi-parametric estimation Large-Time Scaling (LRD).

Fourier vs Wavelets

E. Moulines (ENST)
C. Hurvich (NYU)
P. Soulier (U. Paris X)
F. Roueff (ENST)
M. Taqqu (BU)

The need for sound statistical methods

The richness of traffic is such that one is always in need of more powerful data gathering and processing infrastructures on the one hand, and statistical analysis methods on the other. For existing estimation techniques, the most urgent requirement is increasing their robustness to nonstationarities of various types, which will always be present, despite the luxury of huge data sets which allow apparently stationary subsets to be selected. Closely related to this is the need for formal hypothesis tests to more rigorously select between competing conclusions, and closely related in turn is the need for reliable confidence intervals to be computable, computed, and used intelligently

"Self-Similar Traffic and Network Dynamics", by Erramili, Roughan, Veitch, Willinger (Proc. IEEE 2002)

OUTLINE OF THE TALK

- Today, I will survey methods to detect and assess large time scaling (definition to come)...
- I will not spend time to explain why large time scaling is important and what are the plausible models explaining large time scaling properties (read the excellent survey paper mentioned in the first slide !)
- As you might know, the statisticians already comme too few and too late... (like the US cavalry) ! I hope the issue is still of some importance to some of you !!!

- Introduction
- Fourier Methods
- Wavelet Methods
- Pros and Cons and conclusion

FRACTIONAL MODELS

- A covariance stationary process $\{X_t\}$ is said to be fractional if its spectral density is given

$$f(x) = |1 - \mathrm{e}^{\mathrm{i}x}|^{-2d} f^*(x), \qquad d < 1/2$$

where f^* is continuous at zero frequency.

- Allowing d to take non integer values produces a fundamental change in the correlation structure of the process as compared to the correlation structure of a standard times series... The covariance coefficients decay hyperbolically

$$\rho(\tau) := \mathrm{Cov}(X_\tau, X_0) = O(\tau^{-1+2d}) \qquad \text{as} \qquad \tau \to \infty.$$

SEMI-PARAMETRIC ESTIMATION

- In the semi-parametric setting (SPS), a full parametric model is not specified for the ”smooth part” of the spectral density f^*: f^* is considered as an **infinite dimensional nuisance parameter**.

- Two distinct approaches:

  1. Local-to-zero methods : estimators that estimate d and $f^*(0)$ and which are consistent without any restrictions on f^* away from zero, apart from integrability on $[-\pi, +\pi]$.
  2. global methods : estimators that jointly estimate d and f^* over the whole frequency range, and which are consistent over classes of functions implying ”global” regularity conditions.

OUTLINE OF THE TALK

- The Semi-parametric setting
- Fourier Methods
- Wavelet Methods
- Pros and Cons and conclusion

PERIODOGRAM AT FOURIER FREQUENCIES

- The oldest and most natural tool for spectral estimation is the periodogram (100 years Before Internet !)
- Given an observation $X_1, \cdots, X_n$, the ordinary discrete Fourier transform (DFT) and the periodogram are respectively defined as

$$d_n^X(x) = (2\pi n)^{-1/2} \sum_{t=1}^{n} X_t \mathrm{e}^{\mathrm{i}tx},$$

$$I_n^X(x) = |d_n^X(x)|^2 = (2\pi n)^{-1} \left| \sum_{t=1}^{n} X_t \mathrm{e}^{\mathrm{i}tx} \right|^2 .$$

PERIODOGRAM AT FOURIER FREQUENCIES

Under miscellaneous weak dependence conditions,

- the periodogram is an asymptotically unbiased estimate of the spectral density, *i.e.*

$$\mathbb{E}[I_n^X(x_k)] = f(x_k) + O(n^{-1}), \qquad 1 \leq k \leq \tilde{n},$$

  where the $O(n^{-1})$ term is uniform in k,

- the periodogram ordinates are asymptotically uncorrelated,

$$\mathrm{var}(I_n^X(x_k)) = f(x_k)^2 + O(n^{-1})$$
$$\mathrm{cov}(I_n^X(x_k), I_n^X(x_l)) = O(n^{-1}), \quad k \neq l$$

  where the $O(n^{-1})$ term is uniform w.r.t k, l,

THE PERIODOGRAM OF A FRACTIONAL PROCESS: BAD NEWS

- For LRD processes ($0 < d < 1/2$), none of the above mentioned properties remains valid (Künsch (1986), Hurvich and Beltrao (1993)) !
- The bias is not vanishingly small: for any given $k \in \{1, \cdots, [(n-1)/2]$,

$$\lim_{n \to \infty} \mathbb{E}[I_n^X(x_k)]/f(x_k) \neq 1 ,$$

- The periodogram coordinates are not asymptotically uncorrelated, for any given k, j, $1 \leq k < j \leq [(n-1)/2]$

$$\lim_{n \to \infty} |\mathrm{cov}\left(I_n^X(x_k)/f(x_k) , I_n^X(x_j)/f(x_j)\right)| \neq 0.$$

THE PERIODOGRAM OF A FRACTIONAL PROCESS: GOOD NEWS

- Nevertheless, the bias is vanishingly small for frequencies sufficiently far away from zero.

$$|\mathbb{E}[I_n(x_k)/f(x_k)] - 1| \leq Ck^{-1}$$

- The normalized periodogram ordinates are asymptotically uncorrelated

$$|\mathrm{cov}\left(I_n(x_k)/f(x_k)\,,\, I_n(x_j)/f(x_j)\right)| \leq Ck^{-2d}(j-k)^{2d-1}, \quad k < j$$

THE GEWEKE PORTER-HUDAK (GPH) ESTIMATOR

- In the neighborhood of the zero frequency, $f(x) \approx |1 - \mathrm{e}^{\mathrm{i}x}|^{-2d} f^*(0)$. Therefore,

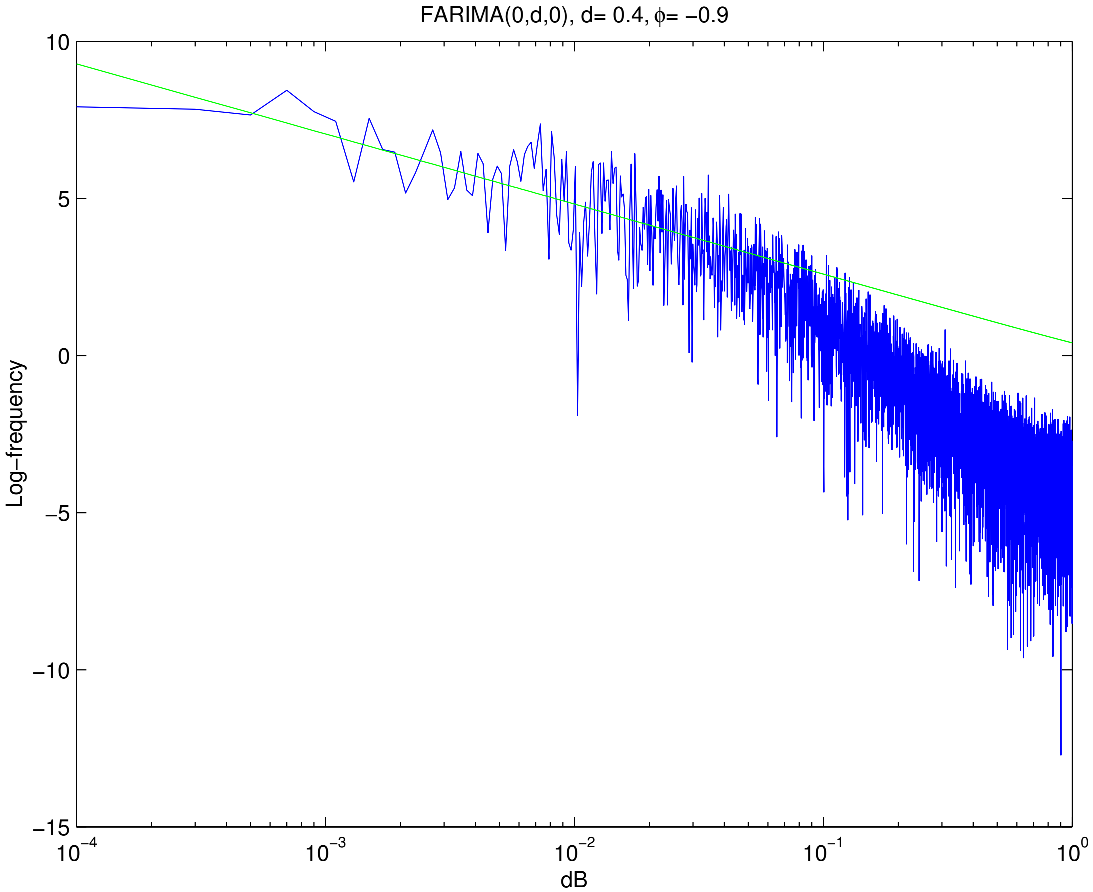
$$\log f(x) \approx dg(x) + \log f^*(0) \quad g(x) = -2 \log |1 - \mathrm{e}^{\mathrm{i}x}|$$

- Writing $\log I_n^X(x_k) = \log f(x_k) + \log I_n^X(x_k)/f(x_k)$ and plugging the expression above,

$$\log I_n^X(x_k) = dg(x_k) + c + \left(\log I_n^X(x_k)/f(x_k) - \gamma\right)$$

- This suggests to estimate d as the regression coef. associated to g !

$$(\hat{d}^{\mathrm{GPH}}(M), \hat{c}^{\mathrm{GPH}}(M)) = \arg\min_{\bar{d},\bar{c}} \sum_{k=1}^{M} (\log(I_n^X(x_k)) - \bar{d}g(x_k) - \bar{c})^2,$$



GPH estimator for a FARIMA(1,d,0) process, $(I - B)^d(1 - \phi B)X = Z$, $\phi = 0.9$. Blue line: log-periodogram. Green Line: least square fit of the intercept.

THE LOCAL WHITTLE ESTIMATOR (LWE)

- Assume that $(d_{n,1}, \ldots, d_{n,M})$ are independent zero-mean complex gaussian random variables satisfying

$$\mathbb{E}|d_{n,k}|^2 = s_{n,k} \ , \quad \mathbb{E}d_{n,k}^2 = 0 \ .$$

- The negated log-likelihood of $(d_{n,1}, \ldots, d_{n,M})$ is

$$\sum_{i=1}^{M} \log(s_{n,k}) + \frac{|d_{n,k}|^2}{s_{n,k}}$$

- Idea: approximate the log-likelihood of $(d_n^X(x_1), \ldots, d_n^X(x_M))$ by

$$-\sum_{i=1}^{M} \log(f(x_k)) + \frac{I_n^X(x_k)}{f(x_k)}$$

Of course, this is not quite true (see the comments above) but we may nevertheless expect that this approximation yields to sensible estimates.

THE LOCAL WHITTLE ESTIMATOR

The Local Whittle Estimator (LWE) is defined as the minimum of

$$(\hat{d}_M^{\mathrm{GSE}}, \hat{C}_M) = \mathrm{argmin}_{\bar{d},\bar{C}} M^{-1} \sum_{k=1}^{M} \left\{ \log(\bar{C}|1 - \mathrm{e}^{\mathrm{i}x_k}|^{-2\bar{d}}) + \frac{I_n^X(x_k)}{\bar{C}|1 - \mathrm{e}^{\mathrm{i}x_j}|^{-2\bar{d}}} \right\}$$

where M is a bandwidth parameter.

- Contrary to GPH, there is no closed form solution...
- however, the problem can be solved for C for any given d, yielding a profile quasi-likelihood which depends only on a single parameter d... this is not a tough optimization problem !!

FEXP ESTIMATOR

- Principle Estimate d and the coefficients of a truncated expansion of $\log f^*$ on the cosine basis.
- Define $h_0 = 1/\sqrt{2\pi}$ and $h_j(x) = \cos(jx)/\sqrt{\pi}$, $j \geq 1$. The log-periodogram regression estimator of d is defined by

$$(\hat{d}^{\mathrm{FEXP}}(q), \hat{\theta}_0, \cdots, \hat{\theta}_q) =$$

$$\arg \min_{\bar{d}, \bar{\theta}_0, \cdots, \bar{\theta}_q} \sum_{k=1}^{K} \left(\log(I_n^X(x_k)) - \bar{d} g(x_k) - \sum_{j=0}^{q} \bar{\theta}_j h_j(x_k) \right)^2,$$

- the choice of the bandwidth parameter M is replaced here by the choice of the truncation index q.

ASYMPTOTIC NORMALITY : GPH / GSE estimator

- (Loc1) There exist a real $d < 1/2$, a square summable sequence $\{\psi_j\}$ and a zero-mean unit variance white noise $\{Z_t\}_{t\in\mathbb{Z}}$ such that

$$X_t = (I - B)^{-d} Y_t, \quad \text{and} \quad Y_t = \sum_{k=-\infty}^{\infty} \psi_k Z_{t-k} \; .$$

We denote $\hat{\psi}(x)$ the Fourier transform of $\{\psi_k\}$ and $f^* = |\hat{\psi}|^2$.

- (Loc2) $1/L \leq f^*(0) \leq L$ and $|f^*(x) - f^*(0)| \leq L|x|^\beta$, for $x \in [0, \Delta]$ and some $L > 0$.

- (Loc3) The bandwidth $M = M_n$ is a non-decreasing function of n (the sample size) which verifies

$$\lim_{n\to\infty} (M_n^{-1} + M_n n^{-\frac{2\beta}{1+2\beta}}) = 0.$$

ASYMPTOTIC NORMALITY : GPH / GSE estimator

Assume (Loc1-3)

- If Z is a martingale increment sequence + conditions then

$$\sqrt{M_n}(\hat{d}^{\mathrm{GSE}}(M_n) - d) \to_d \mathcal{N}(0, 1/4).$$

- If Z is Gaussian or Z is i.i.d. and satisfies moments + Cramer's condition (non-lattice)

$$\sqrt{M_n}(\hat{d}^{\mathrm{GPH}}(M_n) - d) \to_d \mathcal{N}(0, \pi^2/24)$$

ASYMPTOTIC NORMALITY : GPH / GSE estimator

- There is a loss is asymptotic efficiency due to the use of the log-periodogram: this loss can be partially corrected by pooling the periodogram ordinates.
- The maximal rate of convergence is $n^{2/5}$ ($\beta = 2$) no matter how smooth the spectral density is in the neighborhood of zero frequency: can be corrected by using a local polynomial regression (Phillips and co-authors)
- Because of the ”log”, stronger assumptions on the noise are required for the GPH !
- Bad news ... nothing is known for non-linear processes, such as those currently used in traffic analysis... for linear processes, establishing such results proved to be extremely involved, and it would presumably require a tremendous effort to carry this analysis for ”complex” non linear models.

- Glob1 X is a Gaussian process with spectral density $f(x) = |1 - e^{ix}|^{-2d} f^*(x)$, where $-1/2 < d < 1/2$ and the function $x \to l^*(x)$ is continuous. The Fourier coefficients

$$\theta_j(l^*) := (2\pi)^{-1} \int_{-\pi}^{\pi} l^*(x) \cos(jx) dx$$

are absolutely summable.

- Glob2 $|xf^*(x) - yf^*(y)| \leq L|x - y|$,
- Glob3 $q := q_n$ is a non-decreasing sequence of integers such that

$$\lim_{n\to\infty} (q_n^{-1} + q_n \log^5(n) n^{-1}) = 0, \quad \text{and} \quad \lim_{n\to\infty} \sqrt{n/q_n} \sum_{k=q_n}^{\infty} |\theta_j(l^*)| = 0.$$

$$\sqrt{n/q_n}(\hat{d}^{\mathrm{FEXP}}(m, q_n) - d) \to_d \mathcal{N}(0, m\psi'(m))$$

Quasi-parametric rate of convergence can be achieved for analytic function $|\theta_j| \leq Ce^{-\beta j}$, for some $\beta > 0$. In such case one may set $q_n = \log(n)/(2\beta)$ and the rate of convergence is $\sqrt{n/\log(n)}$.

Conclusion - Global vs Local methods

- The FEXP estimator performs ”uniformly well” over a class that contain ”all” the FARIMA processes with roots inside a disk $\{|z| \geq e^{\gamma}\}$,
- The bound explodes as as $\gamma \to 0$... means that some poles of the short-memory part get close to the unit circle !
- Results extend to general function class, including Sobolev classes, quasi-analytic class, class of entire functions...
- For smooth functions global methods should be preferred to local methods... but local methods do not require much

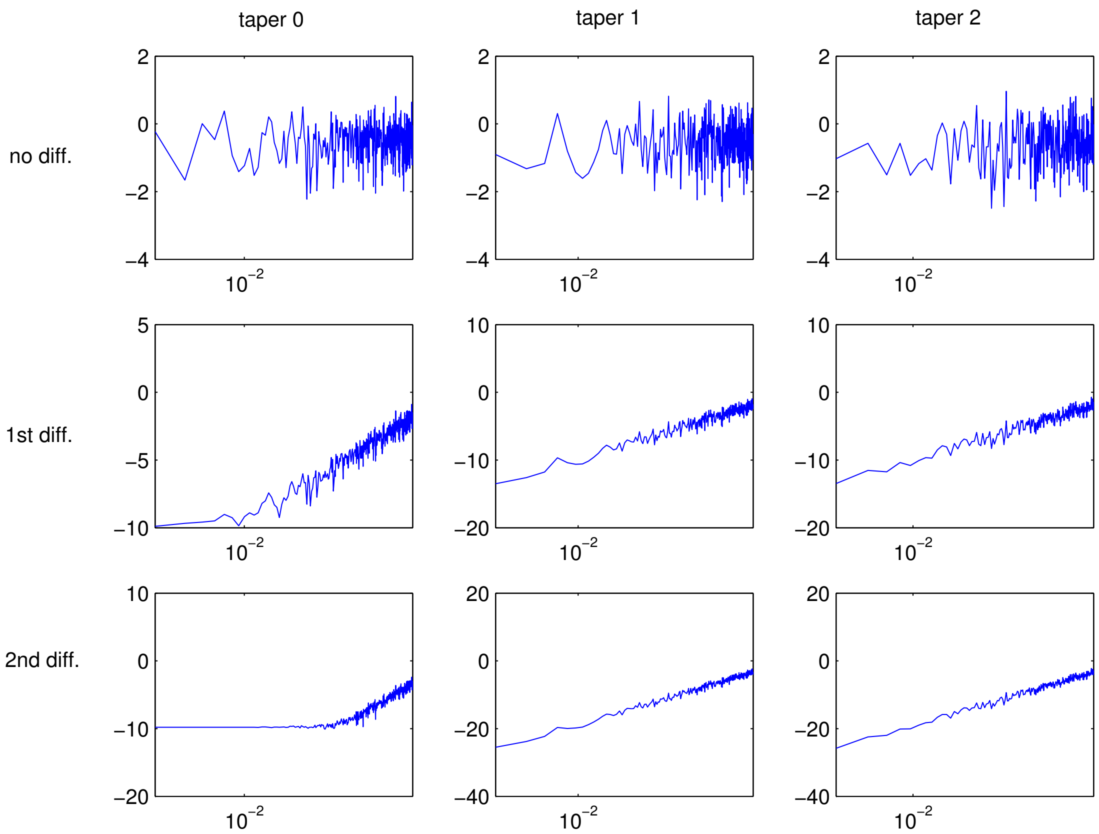
TRENDS, NONSTATIONARITY

- Instead of working directly on the data, it might be worthwhile to work on the M-th difference of the process. Denoting $\{Y_t\}$ the observations, define the M-th difference process as

$$X_t = \Delta^M(B)Y_k, \quad \text{where} \quad \Delta(B) = I - B \; .$$

- Because we differentiate, working with $\{X_t\}$ presents some distinctive advantages:
  - (i) we are insensitive to polynomial trends (of order M) in the observations (and approximately insensitive to ”smooth” trends in the mean)
  - (ii) we can deal with process which are genuinely non-stationary but whose increments are stationary (of importance in econometrics and quantitative finance, where unit-root processes abound).
- When we work on the increments, some care should be taken and the plain Fourier analysis will not work, due to frequency leakage !

FREQUENCY-DOMAIN LEAKAGE



TAPERING

- A solution to control leakage is to taper the observed data prior to computing the DFT.
- The tapered DFT and periodogram are defined as

$$d^X_{h,n}(x) := (2\pi \sum_{t=1}^{n} |h_{t,n}|^2)^{-1/2} \sum_{t=1}^{n} h_{t,n} X_t \mathrm{e}^{\mathrm{i}tx} \text{ and } I^X_{h,n}(x) := |d^X_{h,n}(x)|^2.$$

  where $\{h_{t,n}\}$ is a (real or complex) taper function.

- There a zillions of papers on tapers... I usually used the Hurvich * and Chen (2000) tapers, which generalizes Hanning windows...

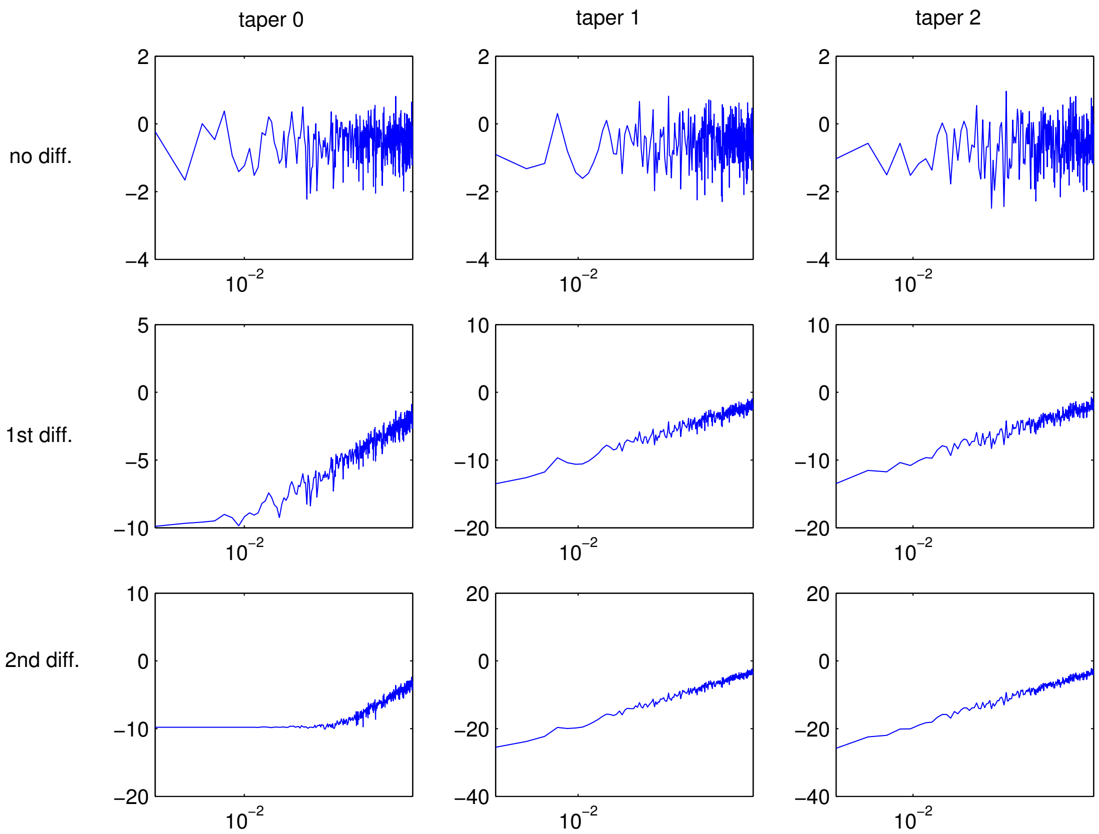
$$h_{t,n} = \left(1 - \mathrm{e}^{2\mathrm{i}\pi t/n}\right)^p$$

  because p gives an explicit control on the rate of decay of the taper in the tails.

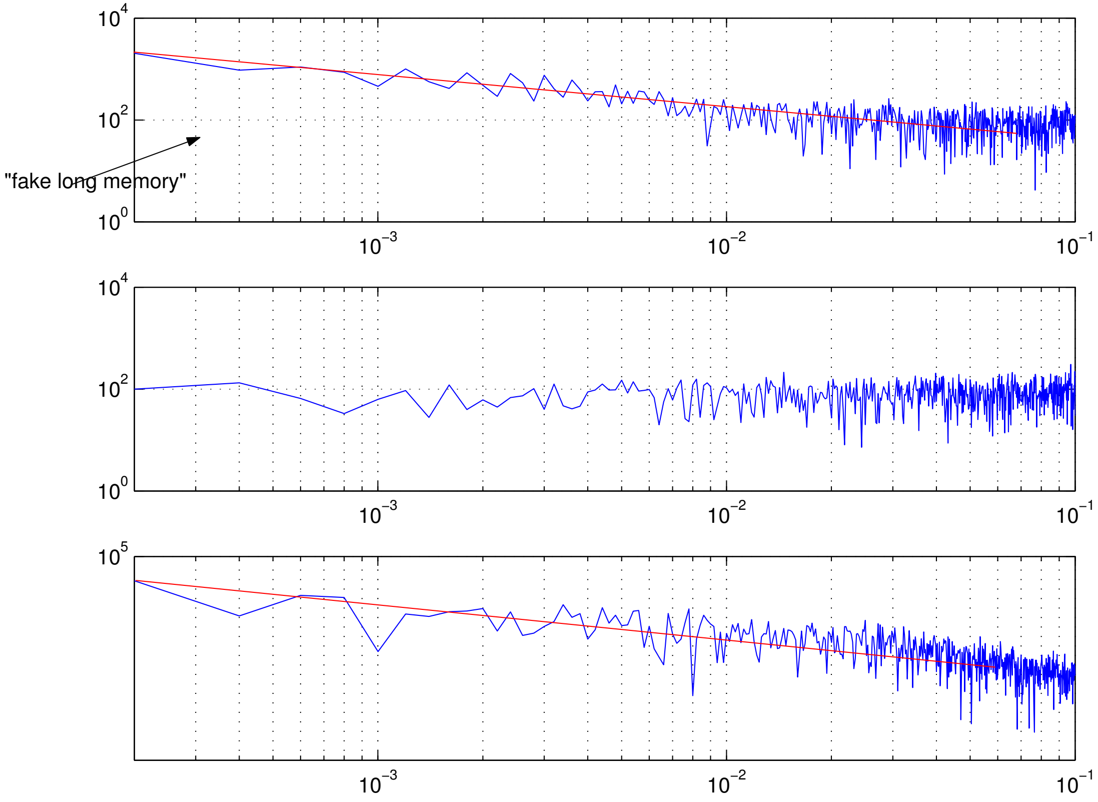
- Using $p = M$ (the taper order = the differentiation order) it is possible to retrieve all the results obtained above (Hurvich, Moulines and Soulier,2002).

* a student of J. Tukey

FREQUENCY-DOMAIN LEAKAGE

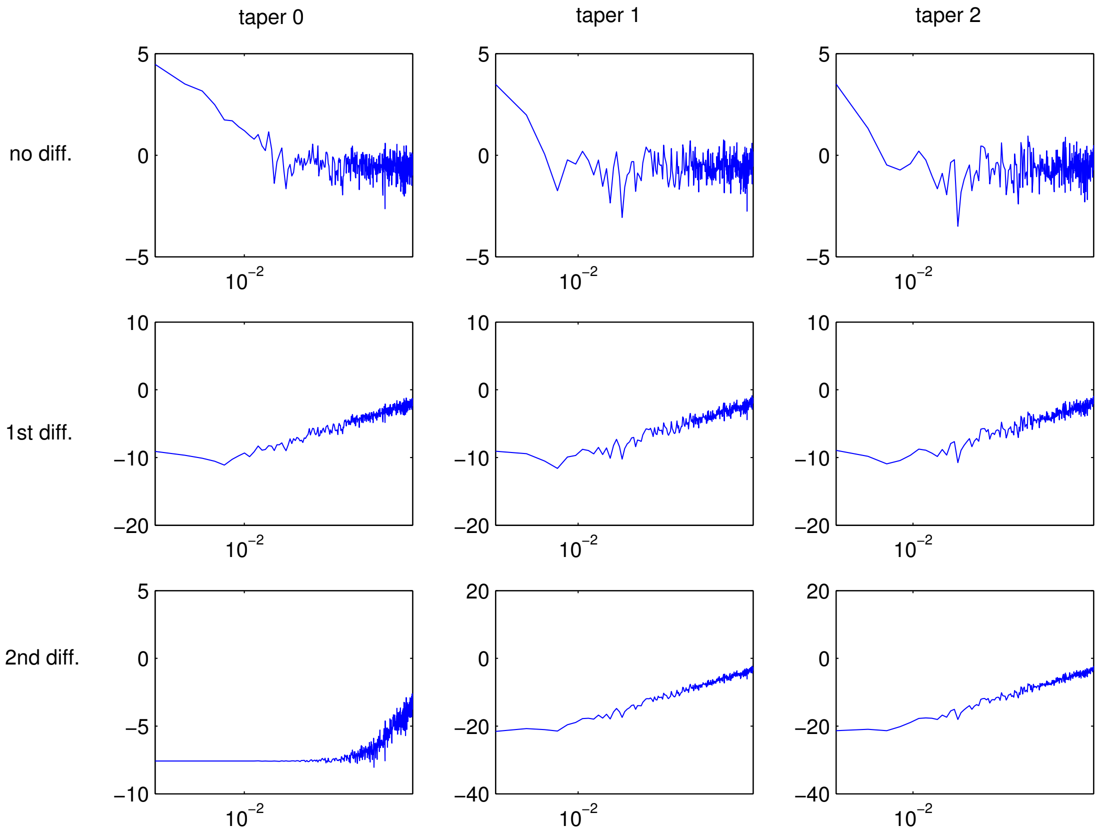


TRENDS, NONSTATIONARITY



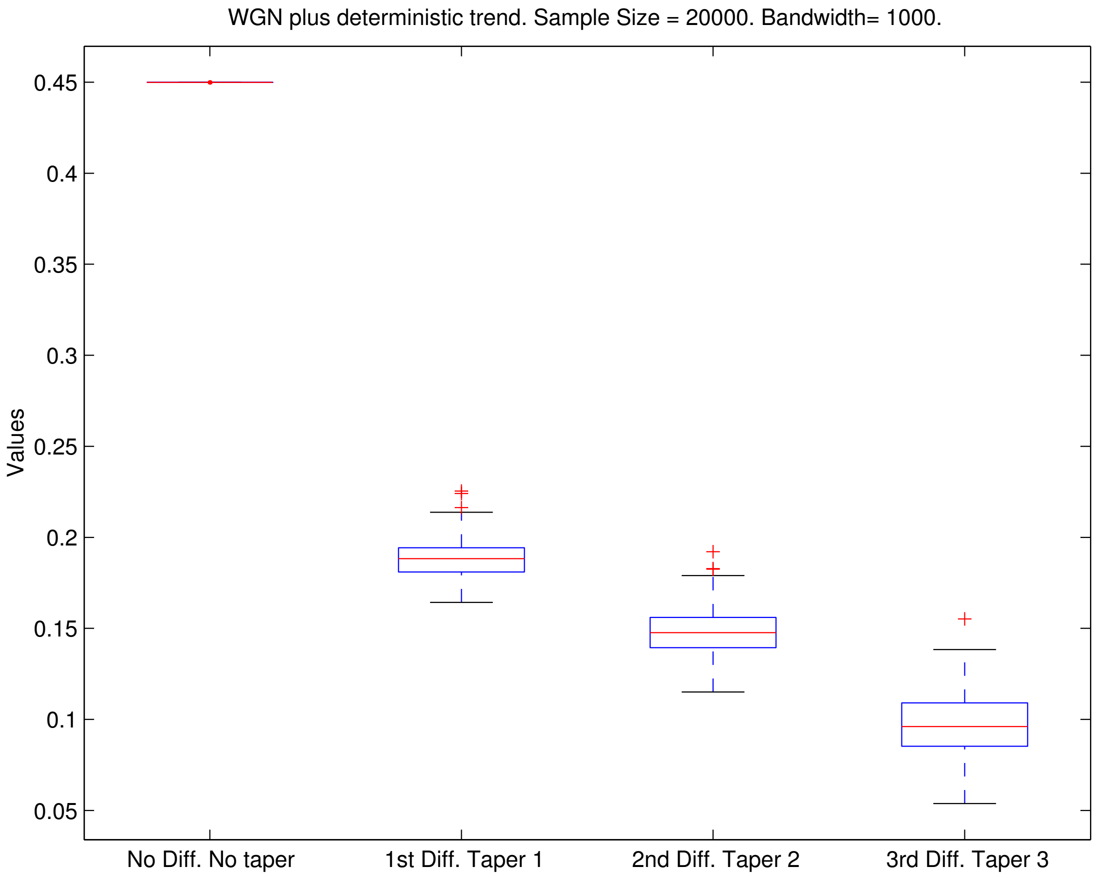
Top plot: WGN + additive trend. Middle plot: WGN. Bottom Plot: FARIMA(1,d,0).

DIFFERENTIATION AND TAPERING



White Noise + additive trend

GSE estimator: WGN + additive trend



”WAVELET” ANALYSIS

- A ”wavelet” analysis is based on a pair of functions (ϕ, ψ), the so called father and mother wavelets.
- Our key assumptions are
  (i) The ”father” wavelet ψ has M vanishing moments, *i.e.* $\int t^l \psi(t)\, dt = 0$ for all $l = 0, \dots, M-1$, or equivalently, $\hat{\psi}(\xi) = O(|\xi|^M)$ in the neighborhood of the zero frequency
  (ii) The ”mother” wavelet ϕ is such that $t \to \sum_{k \in \mathbb{Z}} k^l \phi(\cdot - k)$ is a polynomial of degree l for all $l = 0, \dots, M-1$.

MULTIRESOLUTION

- These assumptions are automatically fulfilled when the functions (ϕ, ψ) are define a multiresolution analysis. Recall that, in this case, $\psi_{j,k}(t) = 2^{-j/2}\,\psi(2^{-j}t - k)$ is an orthonormal basis of $L^2(\mathbb{R})$.
- In fact, for all what follow, it is not required that (ϕ, ψ) define a multi-resolution analysis: there is thus more flexibility in the choice of these functions !
- Nevertheless, classical compactly supported wavelets (like Daubechies) are usually good candidates.

WAVELET COEFFICIENTS

- The computation of the wavelet coefficients of a sequence $x = \{x_k, \; k \in \mathbb{Z}\}$ is (conceptually) in two steps

  (i) interpolation using the father wavelet

$$\mathbf{x}_n(t) := \sum_{k=1}^{n} x_k \, \phi(t-k) \quad \text{and} \quad \mathbf{x}(t) := \sum_{k \in \mathbb{Z}} x_k \, \phi(t-k).$$

  (ii) definition of the (details) wavelet coefficients

$$d_{j,k} := \int \mathbf{x}(t) \psi_{j,k}(t) \, dt \quad (j,k) \in \Lambda.$$

- In practice, the sequence $d_j \stackrel{\text{def}}{=} \{d_{j,k}\}$ is computed by convolving the sequence $\{x_k\}$ with a FIR filter $F_j = \{F_{j,l}\}$ and downsampling, *i.e.* $d_{j,k} = [F_j \star x] \downarrow 2^j$ where

$$F_{j,l} := 2^{-j/2} \int \phi(t+l) \psi(2^{-j} t) \, dt.$$

- Because the wavelets are compactly supported, there is no need to do ”nasty” tricks to handle end effects.

ASSUMPTIONS

(i) The k-th difference $X_t = \Delta^k Y_t$ of the process $\{Y_t\}$ is covariance stationary with spectral density $f(x) = |1 - \mathrm{e}^{\mathrm{i}x}|^{-2d} f^*(x)$, with $|d| < 1/2$.

(ii) $|f^*(\lambda) - f^*(0)| \leq L\lambda^{-\beta}$ with $\beta \in (0, 2]$

(iii) The number M of vanishing moments of ψ is larger than k.

- Since $\{X_t\}$ is covariance stationary, then at any given scale j, the WC sequence $d_j = \{d_{j,k}\}_{k \in \mathbb{Z}}$ is a covariance stationary process.
- The wavelet coefficients process is also stationary across scales (by properly stacking the wavelet coefficients to account for downsampling).

VARIANCE WAVELET COEFFICIENTS

- The variance of the wavelet coefficient grows $\sigma^2(d, f^*(0)) \times 2^{2jd}$; more precisely, for $j \geq 0$,

$$\left| \mathrm{var} \left(\frac{d_{j,k}}{\sigma^2(d, f^*(0))\, 2^{2jd}} \right) - 1 \right| \leq C\, 2^{-\beta j},$$

where $\sigma^2(d, f^*(0)) = f^*(0)\, K_{\phi,\psi}(d)$, with $d \mapsto K_{\psi,\phi}(d)$ a known function depending only on (ψ, ϕ).

- If $\{Y_t\}$ is exactly self-similar (FBM) and if the CWT (no sampling) is used, then: $\mathrm{var} \left(\frac{d_{j,k}}{\sigma^2(d, f^*(0))\, 2^{2jd}} \right) = 1$ over all scales ! In particular, there is no bias term !
- It is interesting to note that the bias is exactly controlled by the behavior of the smooth part of the spectral density at zero frequency

COVARIANCE OF WAVELET COEFFICIENTS

- for any scale $j \geq 0$, the spectral density $D(\lambda; f, j)$ of $\{d_{j,k}\}_{k\in\mathbb{Z}}$ is bounded ! The wavelet filters transform LRD into SRD !

- In addition, this spectral density may be approximated, up to the scale factor $f^*(0)$ by a spectral density depending only on d (but otherwise not on f^*).

$$\left| \frac{D(\lambda; f, j)}{f^*(0)\, D_{\phi,\psi}(\lambda; d)\, 2^{2jd}} - 1 \right| \leq C\, 2^{-\beta j},$$

  where

$$D_{\phi,\psi}(\lambda; d) := |\hat{\phi}(0)|^2 \sum_{l\in\mathbb{Z}} |\lambda + 2l\pi|^{-2d}\, |\hat{\psi}(\lambda + 2l\pi)|^2.$$

  Note that $\hat{\psi}$ being null at zero, all the terms in this sum are bounded. Since the objective is to get wavelet coefficients which are approximately a white noise, this function should be (ideally) flat !

- Expressions of this type were known for the CWT of the FBM (which is exactly self-similar)... the striking result is that valid under much weaker assumptions.

VARIANCE WAVELET COEFFICIENTS

- The logged variance $V_j = \mathrm{var}\,(d_{j,k})$ is given by

$$\log_2(V_j) \approx d(2j) + \log_2(\sigma^2(d, f^*(0))) \qquad \text{as } j \to \infty,$$

which suggests a linear regression approach to estimate d. The slope of the regression estimates d and the intercept is related to $\sigma^2(d, f^*(0))$.

- In practice, $\mathrm{var}\,(d_{j,k})$ is not known but estimating this quantity is easy. The empirical variance of the WC at scale j

$$\hat{V}_j = \frac{1}{N_j} \sum_{k=1}^{N_j} d_{j,k}^2$$

is a sensible candidate, because the wavelet filter kills the LRD (the theory is however harder than it might seem ! see (Bardet et al, 1999).

THE ABRY-VEITCH ESTIMATOR

- Because $\mathrm{var}\left(\hat{V}_j\right)$ depends on the scale j it is wise to perform a weighted regression, which yields to the popular Abry-Veitch (1999) estimator which soon becomes a standard in the network community.
- The resulting estimator is very similar to the GPH estimator... the bandwidth selection problem is here replaced by scale selection !
- Not much is known on the limiting behavior of the estimator: consistency and asymptotic normality has been obtained by Bardet (2000) for the FGN. Consistency and rates are given for the CWT in (Bardet, Lang, Moulines, Soulier, 2000)

WAVELET LOCAL WHITTLE ESTIMATOR

- The (negated) log-likelihood of an array $\{c_{j,k}\}$ of centered independent gaussian random variables with variance $\sigma^2_{j,j}$ is

$$\sum_{(j,k)\in\Delta} \left(c^2_{j,k}/\sigma^2_{j,k} + \log(\sigma^2_{j,k})\right).$$

- If we identify $\{c_{j,k},\ (j,k)\in\Delta\}$ with the wavelet coefficients $\{d^{\mathbf{X}}_{j,k},\ (j,k)\in\Delta\}$ and put $\sigma^2_{j,k} = \sigma^2 2^{2jd}$, we get a proxy for the likelihood of the WC, provided that $\{X_k\}$ is a fractional process of index d. Exploits that
  - (i) $\{d^{\mathbf{X}}_{j,k}\}$ is approximately Gaussian (well supported by numerical evidence showing)
  - (ii) $\{d^{\mathbf{X}}_{j,k}$ are approximately uncorrelated (depends on the choice of (ϕ,ψ) but is also achieved with a reasonable accuracy)

- This suggests to estimate (d,σ^2) by minimizing

$$(\sigma^2,d) \to \sum_{j=J}^{[\log_2(n)]} \left(\frac{d^2_{j,k}}{\sigma^2 2^{2jd}} + \log(\sigma^2\, 2^{2jd})\right)$$

WAVELET LOCAL WHITTLE ESTIMATOR

- The k-th difference $X_t = \Delta^k Y_t$ of the process $\{Y_t\}$ is covariance stationary with spectral density $f(x) = |1 - \mathrm{e}^{\mathrm{i}x}|^{-2d} f^*(x)$, with $|d| < 1/2$.
- $|f^*(\lambda) - f^*(0)| \leq L\lambda^{-\beta}$ with $\beta \in (0, 2]$
- The number M of vanishing moments of ψ is larger than k.
- The lower scale J scale indices satisfy:

$$\frac{2^J}{n} + \frac{n^{1/(1+2\beta)}}{2^J} \to 0$$

$$(n2^{-J})^{1/2}(\hat{d}_J - d_0) \to \mathcal{N}(0, \sigma^2(d_0)) \ .$$

We thus obtain (not surprisingly) the same rate of convergence than Fourier methods.

Selected Topics

- Level II asymptotics: minimax Lower bound, minimax rate optimality
- Adaptations, bandwidth selection

MINIMAX LOWER BOUNDS : LOCAL-TO-ZEROS METHODS

Let $\delta, \Delta > 0$, $\alpha \in (0, \pi]$, $\beta > 0$ and $\mu \geq 1$. There exists a constant $c > 0$ such that,

$$\liminf_{n} \inf_{\hat{d}_n} \sup_{-\Delta \leq d \leq \delta} \sup_{f^* \in \mathcal{F}^*(\alpha,\beta,\mu)} \mathbb{P}_{d,f^*}\left(n^{\beta/(2\beta+1)} |\hat{d}_n - d| \geq c\right) > 0,$$

where

- the infimum $\inf_{\hat{d}_n}$ is taken over all possible estimators d based on $\{X_1, \cdots, X_n\}$ of a covariance stationary process $\{X_t\}_{t \in \mathbb{Z}}$ with spectral density $f = e^{dg} f^*$.
- $\mathcal{F}^*(\alpha, \beta, \mu)$ is the set of functions f^* verifying

$$\int_{-\pi}^{\pi} f^*(x)dx \leq \mu, \qquad 1/\mu \leq f^*(0) \leq \mu, \qquad |\phi(x) - \phi(0)| \leq \mu |x|^\beta, \forall |x| \leq \alpha.$$

The GPH / LWE estimators are rate optimal

MINIMAX LOWER BOUNDS: GLOBAL METHODS

After Iouditski, Moulines, Soulier (2001), Let $\beta > 0$, $\gamma > 0$, $L > 0$, and $\delta < 1/2$. Then

$$\liminf_n \inf_{\hat{d}_n} \sup_{-\delta \le d \le \delta} \sup_{\log(f^*) \in \mathcal{S}(\beta, L)} n^{\frac{2\beta}{2\beta+1}} \mathbb{E}_{d,f^*}[(\hat{d}_n - d)^2] > 0,$$

$$\liminf_n \inf_{\hat{d}_n} \sup_{-\delta \le d \le \delta} \sup_{\log(f^*) \in \mathcal{A}(\beta, L)} n \log^{-1}(n) \mathbb{E}_{d,f^*}[(\hat{d}_n - d)^2] \ge 1/2\beta,$$

where the infimum $\inf_{\hat{d}_n}$ is taken over all possible estimators of d based on $\{X_1, \cdots, X_n\}$ of a covariance stationary process $\{X_t\}_{t \in \mathbb{Z}}$ with spectral density $f = e^{dg} f^*$. Here $\mathcal{S}(\beta, L)$ and $\mathcal{A}(\beta, L)$ are defined as the subsets of $L^2([-\pi, \pi], dx)$ verifying, $\forall q \ge 0$

$$\phi \in \mathcal{S}(\beta, L) \quad \Rightarrow \quad \sum_{j=q}^{\infty} |\hat{\phi}_j| \le L(1+q)^{-\beta},$$

$$\phi \in \mathcal{A}(\beta, L) \quad \Rightarrow \quad \sum_{j=q}^{\infty} |\hat{\phi}_j| \le L e^{-\beta q},$$

where $\hat{\phi}_j := \int \phi(x) \cos(jx) dx$.

THE FEXP ESTIMATOR IS RATE OPTIMAL

Theorem 1 *(Hurvich, Moulines, Soulier, 2001) Let* $\beta > 0$, $\gamma > 0$, $L > 0$, $\delta < 1/2$. *Define* $q_n(\beta, L) = [L^{1/\beta} n^{1/(2\beta+1)}]$ *and* $q_n(\beta) = [\log(n)/2\beta]$.

$$\limsup_n \sup_{|d|\leq\delta} \sup_{\{\log(f^*)\in\mathcal{S}(\beta,L)\}} n^{\frac{2\beta}{1+2\beta}} \mathbb{E}_{d,f^*}[(\hat{d}^{\mathrm{FEXP}}(m, q_n(\beta, L)) - d)^2] \leq L^{\frac{1}{\beta}} m\psi'(m),$$

$$\lim_{n\to\infty} \sup_{|d|\leq\delta} \sup_{\log(f^*)\in\mathcal{A}(\beta,L)} n \log^{-1}(n) \mathbb{E}_{d,f^*}[(\hat{d}^{\mathrm{FEXP}}(m, q_n(\gamma)) - d)^2] = m\psi'(m)/2\beta,$$

where $\mathbb{E}_{d,f^*}$ *denotes the expectation with respect to the distribution of a Gaussian process with spectral density* $e^{dg} f^*$.

- The lower bound is sharp (up to a multiplicative constant between which is less than 2) in the analytic class.
- The restriction to Gaussian processes can be relaxed at the expense of some additional technicalities

Pros and Cons

- Fourier and Wavelets achieve the same rate of convergence (in the situations for which we are able to carry out such analysis !)... yet the asymptotic variance for the wavelet estimators is messier than that of Fourier (depends on d, ϕ, ψ, etc).
- Under rather general assumptions, Fourier estimators are minimax rate optimal and the lower bound is closed for being sharp !...
- Spectral analysis, when appropriately conducted, yields sensible results

Pros and Cons

- Wavelet estimators have ”built-in” robustness to trends in the mean... However, robust Fourier estimators can be obtained by differentiation and tapering.
- Wavelet is directly amenable to adaptive implementations. Fourier estimators can be ”adapted” (e.g. by computing the Fourier transform over blocks and averaging the blocks... whereas such estimators are presumably sensible, theory supporting such estimators is still lacking).
- Wavelet can reveal other features... and particular features that connect large scale with traffic micro-structure - fractals, cascades, etc but here many issues are still open and even the formulation of the problems is not clear cut...
- The theory of wavelet estimators is far less complete than the theory of Fourier estimators... in particular, regressing the logscale is not the only game which can be played (wavelet crossing trees, ”whittle” type estimates)

Selected Contributions

- Moulines, E. and P. Soulier (1999). ”Log-Periodogram Regression of Time Series with Long Range Dependence.” Annals of Statistics 27(1): 1415–1439.
- Bardet, J.-M., G. Lang, et al. (2000). ”Wavelet Estimator of Long-Range Dependent Processes.” Statistical inference for stochastic processes 3(1-2): 85-99.
- Moulines, E. and P. Soulier (2000). ”Data Driven Order Selection for Projection Estimator of the Spectral Density of Time Series with Long Range Dependence.” Journal of Time Series Analysis 21: 193–218.
- Iouditski, A., E. Moulines, et al. (2001). ”Adaptive Estimation of the Fractional Differencing Coefficient.” Bernoulli 7(5): 699-731.
- Hurvich, C., E. Moulines, P. Soulier (2002). ”The FEXP Estimator for Potentially Non-Stationary Linear Time Series.” Stochastic Processes and Their Applications 97: 307–340.
- Moulines, E. and P. Soulier (2002). Long-Range Dependence: Theory and Applications. Theory and applications of long-range dependence. G. O. P. Doukhan, M. Taqqu. Boston, Birkhaüser: 251-301.

Selected Contributions

- Hurvich, C., E. Moulines, P. Soulier (2004). ”Estimation of Long Memory in Stochastic Volatility.” To appear in Econometrica
- Moulines, E., Roueff, F., Taqqu, M. (2004). ”A Wavelet Analog of the Whittle Estimator’ (Manuscript in Preparation)