Modelling and Estimating long memory in non linear time series

Philippe Soulier

Université Paris X

www.tsi.enst.fr/∼soulier

1. Introduction
2. Definition of long memory
3. Linear long memory processes
4. Non linear long memory processes
5. Estimation of long memory

I. Introduction

Long memory processes (linear or non linear) are considered mainly in two fields of applications: teletraffic and financial time series. The topics covered in these lectures are related to modelling financial time series.

For fundamental results, references and applications, cf.
Theory and applications of long-range dependence.

Edited by Paul Doukhan, George Oppenheim and Murad Taqqu (2003).

$\{P_t\}$ financial time series. Returns:

$$R_t = (P_t - P_{t-1})/P_{t-1}.$$

Volatility :

$$\sigma_t{}^2 = \mathrm{E}\left[R_t{}^2 \mid R_{t-1}, R_{t-2}, \dots\right].$$

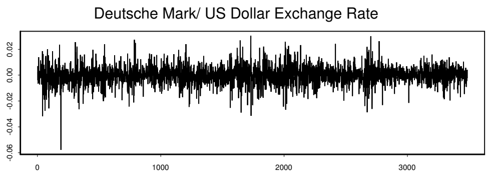
"Stylised facts" Lecture notes of T. Mikosch (2004)

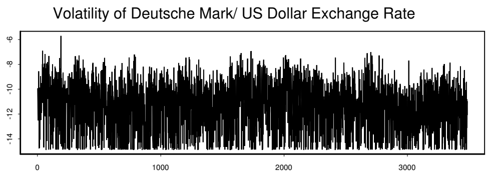
- Returns are uncorrelated, martingale difference sequence.

- Non linear transformations of the returns are (strongly) correlated. *E.g.* log-squared returns.

- Issue: the returns form a stationary process.

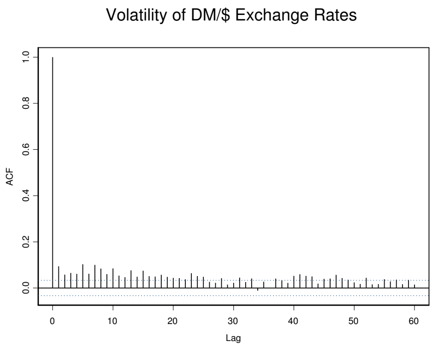
Not to be discussed here.





Exchange rate US \$/ DM; 1985/09/26 to 1998/05/12

stationary process !



Sample autocorrelation of log-squared returns.

Figures taken from Hurvich and Ray (2003).

Two issues:

- What processes can be used to model long memory in the volatility of financial time series ?
- What are the properties of these processes, what statistical procedures can be used?

II. Definition of long memory

A stationary process $\{X_t\}$ is said to have long memory if the asymptotic behaviour of some usual statistics is *very* different from that of a weakly dependent (or short memory) process, such as i.i.d. or strongly mixing sequences.

Examples:

- Weak convergence of partial sums with an unusual normalisation and/or limiting distribution.
- Empirical autocorrelation;
- Discrete Fourier transforms;
- Many others, not to be discussed here.

Long memory is an asymptotic property,
with no precise definition

Working definition of long memory

A weakly stationary process $\{X_t\}$ exhibits second order long memory if its autocovariance function is regularly varying:

$$\mathrm{cov}(X_0, X_t) = L(t)\,|t|^{2H-2}.$$

$H \in (1/2, 1)$ is the Hurst index;

L is slowly varying at infinity, *i.e.*

$$\forall\, x > 0, \quad \lim_{t\to\infty} L(tx)/L(t) = 1.$$

Consequence:

$$\text{var}\left(\sum_{k=1}^{n} X_k\right) = \hat{L}(t)\, n^{2H},$$

where $\hat{L}$ is slowly varying at infinity and equivalent to L, up to a multiplicative constant, depending only on H.

Sometimes taken as a definition.

Equivalent formulation : the spectral density function:

$$f(x) = \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} \text{cov}(X_0, X_t) e^{ikx}.$$

$\{X_t\}$ has long memory if f is regularly varying at zero:

$$f(x) = \tilde{L}(x) x^{1-2H}.$$

$\tilde{L}$ is slowly varyig at zero, $\lim_{t\to\infty} L(t)/\tilde{L}(1/t) = c_H$.

Not exactly equivalent. See Tauberian theorems for regularly varying functions in Bingham, Goldie and Teugels (1987).

III. Gaussian and Linear Long Memory Processes

Gaussian process

A Gaussian process has long memory if and only if its autocovariance function is regularly varying with index $2-2H$, $1/2 < H < 1$.

Linear process

$$X_t = \sum_{j\in\mathbb{Z}} a_j \epsilon_{t-j}, \ \sum a_j^2 < \infty,$$

$\{\epsilon_t\}$ i.i.d. sequence with zero mean and finite variance.
If $\{\epsilon_t\}$ is Gaussian White Noise, then $\{X_t\}$ is Gaussian.

Sufficient condition for long memory:

$$a_j = \ell(j) j^{-d-1/2}, \quad d \in (0, 1/2), \quad \ell \text{ slowly varying}.$$

Then $\{X_t\}$ has long memory with Hurst index $H = d + 1/2$.

Examples

- Fractional Gaussian Noise:

Gaussian process with autocovariance function:

$$\text{cov}(X_0, X_t) = \frac{\sigma^2}{2}\{|t+1|^{2H} - 2|t|^{2H} + |t-1|^{2H}\}.$$

Spectral density:

$$f(x) = c_H\, \sigma^2 \sin^2(x/2) \sum_{k\in\mathbb{Z}} |2k\pi + x|^{-2H-1}.$$

- ARFIMA(p, d, q) process

$$X_t = (I - B)^{-d} U_t = \sum_{j=0}^{\infty} \frac{\Gamma(d+j)}{j!\,\Gamma(d)} U_{t-j}.$$

U_t : ARMA(p, q) stationary and invertible.
$H = d + 1/2$.

Spectral density

$$f(x) = |1 - \mathrm{e}^{\mathrm{i}x}|^{-2d} f_U(x).$$

f_U is smooth and positive around zero.

Asymptotic properties

Let X_t be a Gaussian or linear long memory process with Hurst index H.

- Convergence of partial sums

$$L^{-1}(n) n^{-H} \sum_{k=1}^{[nt]} X_k \xrightarrow{\mathcal{D}} \sigma_H B_H(t)$$

B_H: fractional Brownian motion:

$$\mathrm{cov}(B_H(s), B_H(t)) = \frac{1}{2}\{|t|^{2H} - |t-s|^{2H} + |s|^{H}\}$$

The normalisation is the standard deviation of the partial sums.

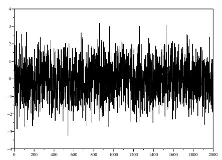
- Convergence of empirical variance

$$\ell(n)\, n^{-\alpha} \sum_{t=1}^{n} (X_t{}^2 - 1) \xrightarrow{w} Z$$

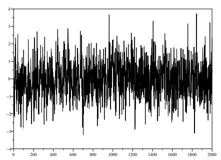
  - $\alpha = 1/2$ and Z Gaussian if $H < 3/4$,
  - $\alpha = 2H - 1$ and Z non gaussian if $H \geq 3/4$.

- Estimation of the Hurst index by spectral methods

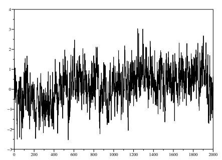
next talk



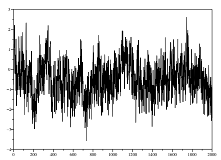
Gaussian white noise
$H = 1/2$



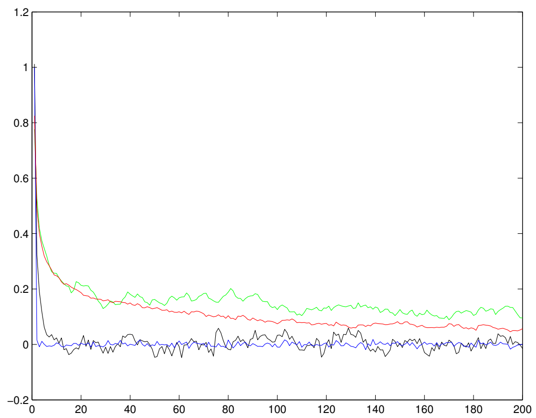
Fractional Gaussian noise
$H = 0.7$



Fractional Gaussian noise
$H = 0.9$

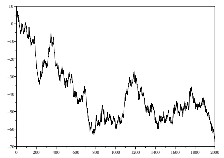


ARFIMA(0, 0.4, 0)
$H = 0.9$

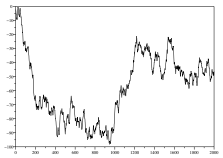


Empirical autocorrelations :

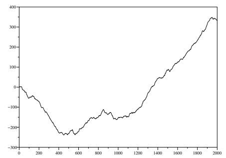
Gaussian white noise, FBM 0.7, FBM 0.9, ARFIMA(0,0.4,0)



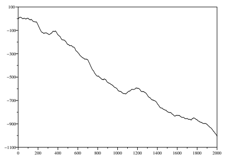
Brownian motion
$H = 1/2$



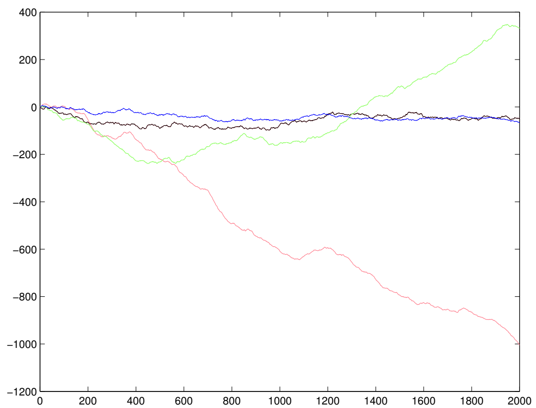
Fractional Brownian Motion
$H = 0.7$



Fractional Brownian Motion
$H = 0.9$



cumulative sum
ARFIMA(0, 0.4, 0)



cumulative sums: ARFIMA(0,0.4,0);
Brownian motion, FBM 0.7 and FGN 0.9

We have recalled the main properties of Gaussian and linear long memory processes. They are quantitatively different from those of short memory processes.

- Different normalizations, but as in the short memory case, linked to standard deviation.
- Asymptotic distributions are Gaussian or related to Gaussian (weighted chi-squared).
- Invariance principle to Fractional Brownian motion.

IV Non linear long memory Processes

We will next describe two classes of non linear long memory models, present some of their asymptotic properties, and see how they differ from those of Gaussian and linear long memory processes.

Statistical issues will be addressed in the second lecture.

There are many other models, that we will not have time to discuss. *E.g.* Random coefficients AR or ARCH processes; Markov switching processes; Stochastic unit root, processes linked to queues; etc.

Bilinear type models

Stochastic volatility models

$$R_t = \sigma_t \epsilon_t,$$

$\{\epsilon_t\}$ i.i.d., $\mathrm{E}[\epsilon] = 0$, $\mathrm{E}[\epsilon^2]=1$,

${\sigma_t}^2 = \mathrm{E}[{R_t}^2 \mid R_s,\ s \le t-1]$ is the conditional variance, *i.e.* the volatility.

$\{R_t\}$ is a martingale difference (hence uncorrelated) sequence.

Can $\{{\sigma_t}^2\}$, $\{|R_t|\}$, $\{{R_t}^2\}$, $\{\log({R_t}^2)\}$ have long memory?

- ARCH processes

$$ {\sigma_t}^2 = a + \sum_{j=1}^{\infty} a_j {R_{t-j}}^2, \quad a_j \ge 0. $$

- If $\sum_{j=1}^{\infty} a_j < 1$, $\{\sigma_t\}$ is strictly stationary and has *short memory*.

- If $\sum_{j=1}^{\infty} a_j = 1$, no weakly stationary solution: $\mathrm{E}[{\sigma_t}^2] = \infty$. Strictly stationary solution if $|a_j| \le c\rho^j$, $\rho < 1$.

Kazakevičius et Leipus (2003)

- The "FIGARCH problem": does there exist a strictly stationary solution if

$$ a_j = -\Gamma(-d+j)/(\Gamma(-d)\, j!) \quad d \in (0, 1/2) \ ? $$

or more generally, $a_j \sim c j^{-d-1/2}$, $\sum_{j=1}^{\infty} a_j = 1$.

- Exponential models

To avoid the difficulties of ARCH type processes, model directly the log-squared volatility:

$$ R_t = \sigma_t \epsilon_t, \quad \sigma_t = \mathrm{e}^{Y_t/2}, \quad Y_t = \sum_{j=1}^{\infty} \psi_j \zeta_{t-j}. $$

$\{\epsilon_t\}$, $\{\zeta_t\}$ i.i.d. sequences;
for each t, ζ_t is independent of $\{\epsilon_s,\ s \ne t\}$,
ϵ_t and ζ_t can be either independent or dependent.

If $\psi_j \sim c j^{d-1}$, $d \in (0, 1/2)$, then $\{\log({\sigma_t}^2)\}$ has long memory with Hurst index $H = d + 1/2$.

• FI-EGARCH processes (Nelson 1991, Bollerslev and Mikkelsen 1996).

$$\zeta_t = g(\epsilon_t),$$

E.g. Nelson (1991) suggested $g(z) = \theta z + \gamma(|z| - E[|\epsilon_0|])$, modelizes the so-called leverage effect.

• LMSV process (Harvey 1998, Breidt *et al.* 1998)

$\{\zeta_t\}$ is independent of $\{\epsilon_t\}$.

In both cases, if $\psi_j \sim cj^{d-1}$, $d \in (0, 1/2)$, $\{\sigma_t\}$, $\{R_t\}$, $\{|R_t|^u\}$ have long memory with same Hurst index $H = d + 1/2$ and

$$n^{-H} \sum_{k=1}^{[nt]} U_k \overset{\mathcal{D}}{\Longrightarrow} \sigma_U B_H(t)$$

with $\{U_t\}$ any of the above processes, under moment conditions, satisfied if $\{\zeta_t\}$ is Gaussian.

Surgailis and Viano (2002)

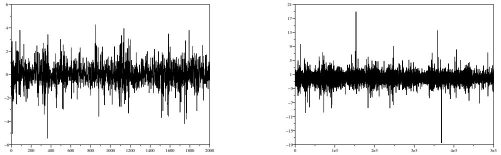
If $\mathrm{P}(\epsilon_t = 0) = 0$,

$$\log(R_t{}^2) = Y_t + \log(\epsilon_t{}^2)$$

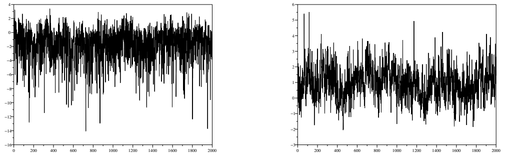
has long memory with Hurst index $H = d + 1/2$.

Hence it is sufficient to study the memory of the log-squared returns.

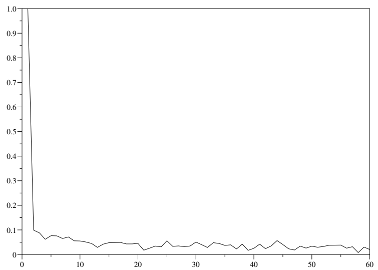
Simulation of LMSV processes : $R_t = \epsilon_t e^{Y_t/2}$, $\{Y_t\}$ ARFIMA(0,r,0), $d = 0.4$.



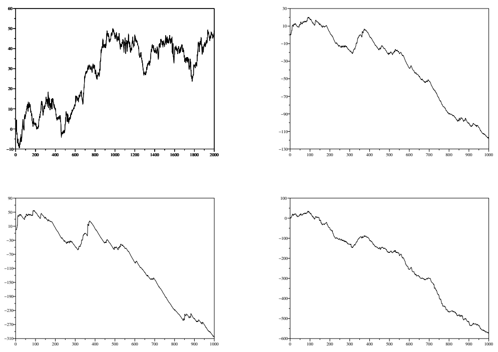
$\{\epsilon_t\}$ Gaussian $\{\epsilon_t\}$ Pareto 4



log squared returns



Sample autocorrelation of log-squared returns of LMSV process.



Cumulative sums of LMSV process with Gaussian shocks:
returns, absolute, squared and log-squared returns

Duration driven long memory

Spurrious long memory:
It has been known for a long time that in finite samples, slow decay of the autocovariance function can be an artefact:

- of polynomial trends (Bhattacharya *et al.* 1983)
- of structural break (Diebold and Inoue 2000, Mikosch and Starica 2004)

But

- polynomial trends can be eliminated by differencing;
- stationary models with structural change have been introduced, in particular as a model for volatility.

Stationary models with structural change

- Renewal-Reward (Taqqu and Levy 1986);
- Long memory shot noise (Giraitis, Molchanov and Surgailis 1993, Klüppelberg and Kühn 2004);
- ON-OFF process (Taqqu, Willinger and Sherman 1997, Heath, Resnick and Samorodnitsky 1998);
- Infinite source Poisson (Mikosch, Resnick, Rootzen, Stegeman 2002), with random transmission rate (Maulik, Resnick and Rootzén 2002);
- Error duration (Parke 1999, Hsieh, Hurvich and Soulier 2003).
- Other examples in Kulik and Szekli (2002).

General shot noise process

$\{S_n\}_{n\in\mathbb{Z}}$: points of a stationary renewal process: birth dates;
$\{U_n\}$: i.i.d. independant of $\{S_n\}$: rewards or pulses or shocks;
$\{\eta_n\}$: i.i.d. non-negative: durations;
g function.

$$X_t = \sum_{n\in\mathbb{Z}} U_n \, g(\{t - S_n\}/\eta_n).$$

When is it defined ?

- Simplification: $g(t) = \mathbf{1}_{[0,1)}(t)$.

Rules out the shot noise processes of Giraitis *et al.* (1993), Klüppelberg and Kühn (2004).

Model now

$$X_t = \sum_{n\in\mathbb{Z}} U_n \, \mathbf{1}_{[S_n, S_n+\eta_n)}(t).$$

The shock U_k created at time S_k survives for a duration η_k.

The process is well defined iff only a finite number of rewards survive at time t. Sufficient condition:

$$\mathrm{E}[\eta] < \infty.$$

The process is then strictly stationary and weakly stationary if

$$\mathrm{E}[U^2] < \infty.$$

Stationary renewal process

$\{\tau_k\}_{k\geq 1}$ i.i.d. sequence of nonnegative r.v. with distribution function F and finite mean $\lambda^{-1} = \int_0^\infty (1 - F(s))\mathrm{d}s$.
τ_0 independent of $\{\tau_k\}_{k\geq 1}$ with distribution function

$$F_0(t) = \mathrm{P}(\tau_0 \leq t) = \lambda \int_0^t (1 - F(s))\mathrm{d}s$$

Let τ_{-1} be independent of $\{\tau_j\}_{j\geq 0}$ distributed as $-\tau_0$;
$\{\tau_j\}_{j<-1}$ be independent of $\{\tau_j\}_{j\geq -1}$, distributed as $\{-\tau_j\}_{j\geq 1}$.

Point process on the line: $..S_{-2} < S_{-1} < 0 \leq S_0 < S_1 < ...,$

$$S_k = \sum_{j=0}^{k} \tau_j, \ k \geq 0, \qquad S_k = \sum_{j=-1}^{-k} \tau_j, \ k < 0,$$

$$N(s,t] = \sum_k \mathbf{1}_{(s,t]}(S_k).$$

N is translation invariant. $\quad \mathrm{E}N(s,t] = \lambda(t-s)$.

Exemples of stationary renewal processes.

- Deterministic renewal process: $S_n = \mu n$.

- Poisson process: $F(t) = F_0(t) = 1 - \mathrm{e}^{-\lambda t}$.

- Renewal process with heavy tailed inter-arrival distribution:

$$F(t) = L(t)(1+t)^{-\alpha}, \quad F_0(t) = \tilde{L}(t)(1+t)^{1-\alpha}, \quad 1 < \alpha < 2.$$

Duration driven long memory processes

Renewal-reward: $\quad \tau_{n+1} = \eta_n$.
Exactly one random shock at a time.

ON-OFF: $\quad U_n \equiv 1, \quad \tau_{n+1} = \eta_n + \zeta_n$.
Succession of ones and zeros, at most one shock at a time.

Infinite source Poisson: $\quad \{S_n\}$ are the points of a homogeneous Poisson process. U_n is the transmission rate: constant or random.

Error duration: $\quad S_n = n$.

Autocovariance function

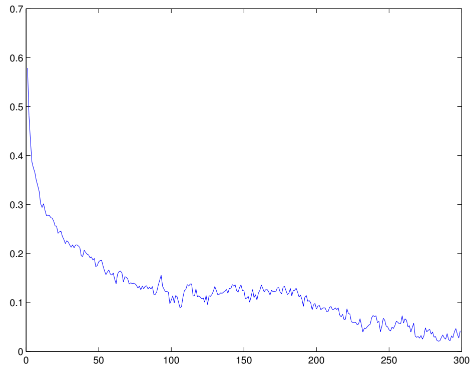
- Centered shocks.

$$\mathrm{cov}(X_0, X_t) = \lambda \ \mathrm{E}[U] \ \mathrm{E}[(\eta - t)_+].$$

If η is regularly varying: $\boxed{\mathrm{P}(\eta > t) = \ell(t) t^{-\beta},}$
then:

$$\boxed{\mathrm{cov}(X_0, X_t) = \tilde{\ell}(t) t^{1-\beta}.}$$

Long memory with Hurst index: $H = (3 - \beta)/2$.



Sample autocorelation
of Error Duration process, $H = 0.9$.

Asymptotic properties of duration driven long memory processes

- Centered shocks.

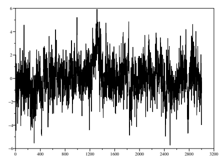
$$\hat{\ell}(n) n^{-1/\beta} \sum_{k=1}^{[nt]} X_k \ \xrightarrow{fi.\,di.} \ \Lambda_\beta(t).$$

β-stable Lévy process (with independent increments).

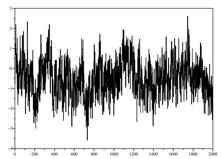
But: $\quad n^{-H} \sum_{k=1}^{n} X_k \ \xrightarrow{\mathrm{P}} 0,$

because $\quad \mathrm{var}\left(\sum_{k=1}^{n} X_k\right) = L(t) \ n^{3-\beta} = L(t) n^{2H}$

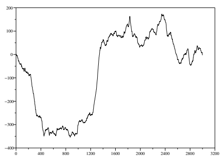
and $\quad H = (3 - \beta)/2 > 1/\beta.$



Error duration, $\beta = 1.2$
$H = 0.9$



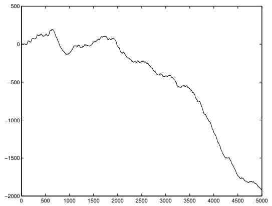
ARFIMA(0,0.4,0)
$H = 0.9$



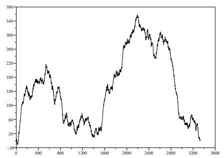
cumulative sum

cumulative sum

1.2-stable Levy process

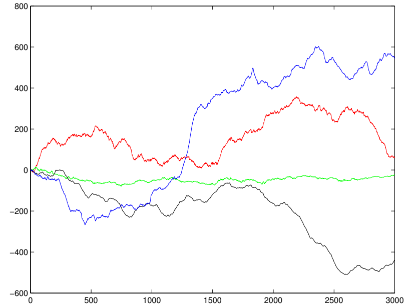


FBM $H = 0.9$



cumulative sum of log-squared returns US\$-DM
Levy or Gauss ?

Modelling long memory in volatility:
fractional differencing or duration driven long memory?



cumulative sums: returns and log-squared returns US\$/DM
FBM $H = 0.9$, Levy 1.2 stable process.

Other problems

- Non centered shocks: tails of interarrival distribution vs. tails of durations.

- Convergence in $\mathcal{D}$?
Proved for the infinite source Poisson (Resnick and Van den Berg, 2000)

- Convergence of empirical process ?
Finite dimensional convergence obvious for the ON-OFF process.

- Sample autocorrelations: $\widehat{\gamma}_n(p) = n^{-1} \sum_{k=1}^{n-|p|} X_k X_{k+|p|}$.
Centered shocks: $n^{-1/\beta}\{\widehat{\gamma}_n(p) - \gamma(p)\}$ converges weakly to a β-stable limit.

V. Estimation of long memory

Working definition : $\{X_t\}$ is a long memory process if it is weakly stationary with spectral density f regularly varying at 0:

$$f(x) = x^{1-2H} L(x),$$

L is a slowly varying function.

For statistical purpose, it is equivalent and more convenient to write:

$$f(x) = |2\sin(x/2)|^{1-2H} L(x),$$

Given this definition, estimating long memory means estimating the Hurst index H.

Two types of methods : time domain and frequency domain.

Time domain methods: R/S, Agregated variance, DFA...

Drawbacks: asymptotic properties not always established, rates of convergence are often bad and asymptotic variance depends on H.

Frequency domain methods

The discrete Fourier transform and the periodogram

$$d_k = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n X_t \mathrm{e}^{\mathrm{i} t x_k}, \qquad I_k = |d_k|^2,$$

$x_k = 2k\pi/n, \quad 1 \le k < n/2.$

The periodogram is an asymptotically unbiased but inconsistent estimator of the spectral density f of a weakly stationary process.

$$I_k = \frac{1}{2\pi n} \sum_{p=1-n}^{n-1} \widehat{\gamma}_n(p) \mathrm{e}^{\mathrm{i} t x_k}$$

$\widehat{\gamma}_n(p) = n^{-1} \sum_{k=1}^{n-|p|} X_k X_{k+|p|}$: empirical autocovariance,

estimator of the p-th Fourier coefficient of f.

Heuristic approximation: the random variables

$$\xi_k := \frac{I_k}{f(x_k)}, \quad 1 \le k < n/2.$$

are i.i.d. exponentially distributed.

Always wrong (except for Gaussian white noise), even asymptotically, but useful:

the conclusions drawn from this wrong approximation can be justified theoretically in some cases.*

* est-ce clair ? à chaque fois que je le dis, on me pose des questions qui montrent que ça ne l'est pas.

More precisely, in the case <u>Gaussian</u> or <u>linear</u> processes, for any fixed integer u and sequences of integers $k_1, \dots, k_u$, the vector

$$\left(\frac{d_{k_1}}{f(x_{k_1})}, \dots, \frac{d_{k_1}}{f(x_{k_1})}\right)$$

converges weakly to

- a vector of u independent standard complex Gaussian random variables in the <u>weakly dependent</u> case;

- a vector of u dependent complex Gaussian random variables with dependent complex and imaginary parts in the <u>long memory</u> case.

But this dependency is concentrated on <u>low frequencies</u>, and vanishes for <u>increasing frequencies</u>.

The GPH estimator Geweke Porter-Hudak (1983)

Assume that the spectral density can be expanded as

$$f(x) = c|2\sin(x/2)|^{1-2H}\{1 + o(1)\}.$$

true for ARFIMA(p, d, q) and FGN.

Write then:

$$\begin{aligned}\log(I_k) &= \log(f(x_k)) + \log(\xi_k) \\ &= \log(c) + (1 - 2H)\log(2|\sin(x_k/2)|) + \log(\xi_k) + \text{ bias}.\end{aligned}$$

The GPH estimator $\widehat{H}_{GPH}$ is the OLS estimator based on m frequencies:

$$\begin{aligned}\widehat{H}_{GPH} &= \sum_{k=1}^{m} \nu_k \log(I_k) \\ &= H + \sum_{k=1}^{m} \nu_k \log(\xi_k) + \text{bias} \approx H + \sqrt{\frac{\pi^2}{24m}}\mathcal{N}(0,1) + \text{bias}.\end{aligned}$$

$$\nu_k = \frac{\{\log(2|\sin(x_k/2)|) - m^{-1}\sum_{j=1}^{m}\log(2|\sin(x_j/2)|)\}}{\sum_{\ell=1}^{m}\{\log(2|\sin(x_\ell/2)|) - m^{-1}\sum_{j=1}^{m}\log(2|\sin(x_j/2)|)\}^2}.$$

Main problem : choice of m.

If $m \to \infty$ and $m/n \to 0$, $\widehat{H}_{GPH}$ is consistent.

Rate of convergence ?

- The choice of m is linked to the bias term and depends on the (unknown) regularity of $x^{2H-1}f(x)$ around zero.

If f is second order regularly varying:

$$f(x) = c\,|2\sin(x/2)|^{1-2H}\left\{1 + O(x^{\beta})\right\}.$$

Then: $$\text{bias} = O((m/n)^{\beta}).$$

Optimal choice : $m = n^{2\beta/2\beta+1}$,
Optimal rate of convergence : $n^{\beta/2\beta+1}$.

- A rigorous theory is established in Robinson (1995), Giraitis, Robinson et Samarov (1997,2000) for Gaussian processes, Velasco (1999) for linear processes.

Unfortunately, GPH (1983) suggested the "rule of thumb" $m = \sqrt{n}$.
This is not correct but still often used.

Fundamental examples

- ARFIMA(0,d,0).

Spectral density:

$$f(x) = \frac{\sigma^2}{2\pi}|2\sin(x/2)|^{1-2H}.$$

No bias term: parametric rate of convergence!

- ARFIMA(0,d,0)+noise.

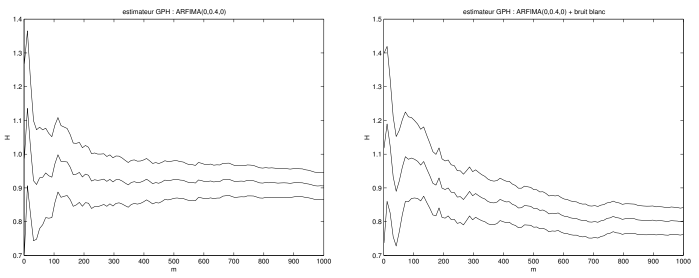
Spectral density:

$$f(x) = \frac{\sigma^2}{2\pi}|2\sin(x/2)|^{2H-1} + \frac{\sigma_N^2}{2\pi}$$
$$= \frac{\sigma^2}{2\pi}|2\sin(x/2)|^{2H-1}\left\{1 + O(x^{1-2H})\right\}.$$

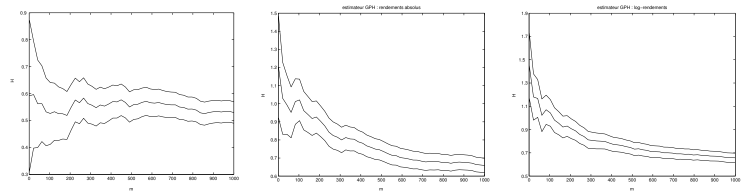
The bias term depends on H, the best possible rate of convergence of the GPH estimator is $n^{(2H-1)/2H}$.

Moreover $\hat{H}_{GPH}$ is negatively biased.

Deo and Hurvich (2001).



GPH estimator:
left ARFIMA(0,0.4,0)
right ARFIMA(0,0.4,0) + white noise
$H = 0.9$



GPH estimator:
left returns US$/DM, $H = 1/2$;
middle absolute returns, $H =$?
right log squarred returns, $H =$?

If the LMSV (or FIEGARCH) model is correct for modelling the volatility, is there a way to estimate H more efficiently ?

Estimation of the Hurst coefficient of the LMSV process

Hurvich, Moulines and Soulier (2001-2004)

$R_t = \mathrm{e}^{Y_t/2}\epsilon_t$,
$\{\epsilon_t\}$ i.i.d. sequence,
independent of the long memory process $\{Y_t\}$
with spectral density $f_Y(x) = c|2\sin(x/2)|^{1-2H}(1 + O(x^\beta))$.

Log-squared returns:

$$X_t = \log(R_t)^2 = Y_t + \eta_t + c,$$

$\eta_t = \log(\epsilon_t{}^2) - c, \quad c = \mathbb{E}[\log(\epsilon_t{}^2)]$.

Signal + noise model.

The spectral density of $\{X_t\}$ is:

$$f_X(x) = f_Y(x) + \frac{\sigma_\eta^2}{2\pi}$$
$$\sim C|2\sin(x/2)|^{1-2H}\left\{1 + \tau|2\sin(x/2)|^{2H-1} + O(x^\beta)\right\}.$$

Assumption : $\beta > 2H - 1$.

The rate of convergence of the GPH estimator is $n^{(2H-1)/4H}$.

Very bad if H is close to 1/2.

Is it possible to improve this rate of convergence ?

Yes: refine the 'heuristic approximation'.

Denote $\theta = (C, H, \tau)$, $f(\theta, x) = Cx^{1-2H}\{1 + \tau x^{2H-1}\}$ and

$$\xi_k = \frac{I_k}{f(\theta, x)}.$$

Then $\{\xi_k, 1 \le k \le m\}$ can be considered as i.i.d. random variables with standard exponential distribution, if

$$m = o(n^{2\beta/2\beta+1}).$$

Pseudo-maximum likelihood (Whittle) estimator:

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \sum_{k=1}^{m} \left\{\log(f(\theta, x_k)) + \frac{I_k}{f(\theta, x_k)}\right\}.$$

Modification of the local Whittle estimator (Künsch 1986, Robinson 1995).

If $m = o(n^{2\beta/(2\beta+1)})$ and $n^{(4H-1)/4H} = o(m)$,

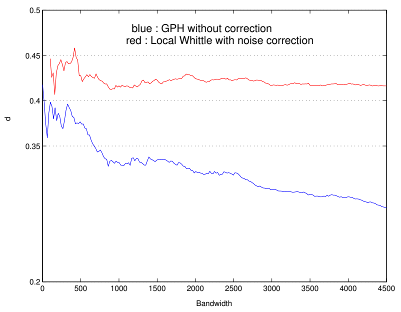
Then $m^{1/2}(\hat{H} - H)$ is asymptotically normal with variance

$$\frac{H^2}{(2H-1)^2}.$$

The rate of convergence is optimal, but the asymptotic variance tends to infinity as H tends to 1/2.

Identifiability problem as H tends to 1/2.

Estimation of $H = d + 1/2 = 0.9$.
ARFIMA(0,0.4,0) + noise.
signal/noise ratio = 1.



	$m = [n^{.05}]$	$m = [n^{.06}]$	$m = [n^{.07}]$	$m = [n^{.08}]$
$\widehat{H}$	0.865	0.878	0.887	1.056
$\widehat{H}_{GPH}$	0.870	0.855	0.774	0.635

GPH and modified Whittle estimators
for US$/DM exchange rate, $n = 3485$.

Quoted from Hurvich and Ray (2003).

Asymptotic properties of the Discrete Fourier transforms of duration driven long memory processes.

$$X_t = \sum_{n\in\mathbb{Z}} U_n \, \mathbf{1}_{[S_n, S_n+\eta_n)}(t).$$

$\{S_n\}_{n\in\mathbb{Z}}$: points of a stationary renewal process;
$\{U_n\}$: i.i.d. zero mean, finite variance, independant of $\{S_n\}$
$\{\eta_n\}$: i.i.d. non-negative durations; with regularly varying tails:

$$\mathrm{P}(\eta > t) = \ell(t) t^{-\beta}.$$

Then $\{X_t\}$ is stationary, has long memory with Hurst index $H = (3-\beta)/2$, its spectral density is regularly varying:

$$f(x) = \tilde{\ell}(x) x^{1-2H}.$$

Discrete Fourier transform:

$$d_k = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t \mathrm{e}^{\mathrm{i}tx_k}$$

What is the behaviour of the renormalized DFT ?

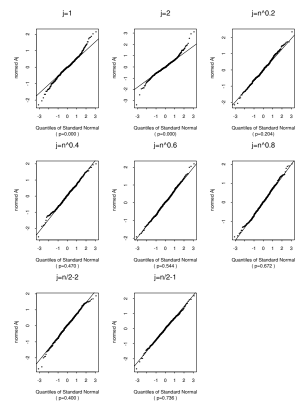
Two regimes:

- Low frequencies: for any fixed integers, u, $k_1, \ldots, k_u$

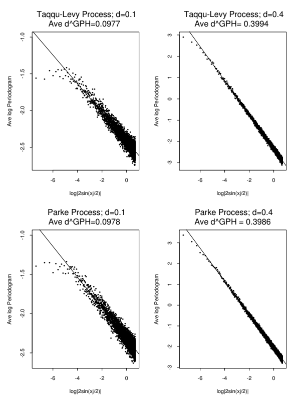
$$n^{1/2-1/\beta}(d_{k_1}, \ldots, d_{k_u})$$

weakly converges to a β-stable vector.

- High frequencies: if k is a sequence of integers such that $k/n \to 0$ and $k/n^{1-1/\beta} \to \infty$, then $d_k/\sqrt{f(x_k)}$ converges to a standard complex Gaussian random variable.



QQ Plots of the Normalized Fourier Cosine Coefficients Re$(d_j)/\sqrt{\frac{1}{2}f(x_j)}$ for the Error duration process; n=10000, $\beta = 1.2$, H=0.4



Scatterplots of Log Periodogram vs. log $|2\sin(x_j/2)|$; j=1,2,...,4999; $H = 0.6$ (left) and $H = 0.9$ (right).

Conclusion

- Non linear long memory processes form a rich class of processes, whose properties are not fully known.

- The statistical procedures to detect long memory seem to be robust, but their theoretical properties are far from well established.

- Practionneers are often not aware of this and apply the GPH estimator blindedly.