L M J L
UMR 6629 - Nantes

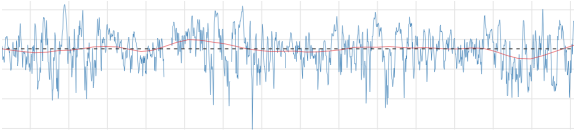
UNIVERSITÉ DE NANTES

Quelques résultats sur les processus à longue mémoire.

Anne Philippe
Université de Nantes

10 ième Biennale Française des Mathématiques Appliquées et Industrielles

anne.philippe@univ-nantes.fr



Long memory processes

Random Sampling & continuous time processes

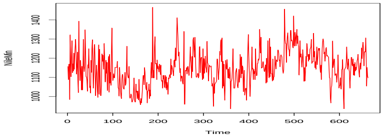
Testing long memory processes

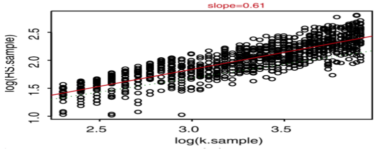
Hydrology : average discharge of the Nile

- Hurst (1951) studied the hydrological properties of the Nile basin

- Observations : $X_1, \ldots, X_n$: the annual flows of river over n years.
- the cumulative flows over time

$$D_k = \sum_{i=1}^{k} X_i - \frac{k}{n} \sum_{i=1}^{n} X_i$$





Statistic of interest :

Range

$$R(n) = \max_{k=1,\ldots,n} D_k - \min_{k=1,\ldots,n} D_k$$

or its rescaled version $R(n)/S(n)$ where $S(n)^2$ is the empirical variance.

LIMIT THERORE M

$(X_j)_{j=1,2,\ldots}$ are iid random variables with finite mean μ and finite variance σ^2

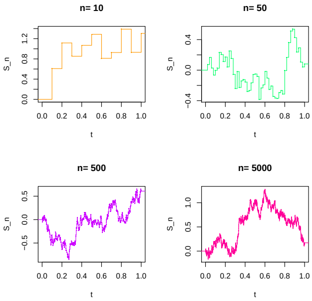
$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j-\mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2)$$

Weak convergence of the partial sums Donsker (1952)

we have

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}(X_j-\mathbb{E}X_1) \xrightarrow{\mathcal{L}}_{(D[0,1],\sup)} \sigma^2 B(t), \qquad t\in[0,1].$$

- B is the standard Brownian motion with zero mean and covariance $\operatorname{cov}(B(t),B(s)) = s\wedge t$



Asymptotic behavior of R/S

$$\frac{\ln(R(n)/S(n))}{\ln n} \xrightarrow{P} 1/2$$

- We consider $\mathcal{X} = \{X_t, t \in \mathbb{N}\}$ a stochastic process.
- We suppose that $\mathcal{X}$ is a stationary process in the L^2 sense i.e.
  1. $\mathbb{E}(X_t) = \mu$ does not depend on t
  2. $\text{cov}(X_s, X_t) = \mathbb{E}(X_t - \mu)(X_s - \mu) = \sigma_X(|t-s|)$
- We quantify the memory of the process from the asymptotic behavior of σ_X.

AR(1) PROCESS

Consider $\{X(t),\ t \in \mathbb{Z}\}$ the solution of

$$X(t) = aX(t-1) + \zeta(t), \quad t \in \mathbb{Z},$$

where

- $a \in (-1, 1)$
- $\{\zeta(t)\}$ a sequence of iid random variables $\mathbb{E}\zeta = 0$, $\mathbb{E}\zeta^2 = 1$,

Properties

- The unique strictly stationary solution is given by

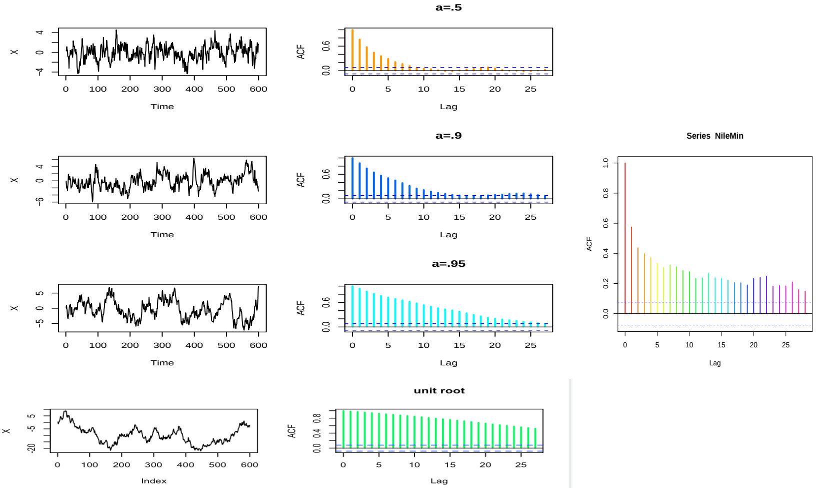
$$X(t) = \sum_{s \le t} a^{t-s} \zeta(s)$$

- $\{X(t),\ t \in \mathbb{Z}\}$ has zero mean and the auto covaraince function is

$$\mathbb{E}[X(0)X(t)] = \sigma_X(t) = \frac{a^{|t|}}{1-a^2}$$

- Auto covariance function : $\sigma_X \in \ell^1$ short memory

EMPIRICAL AUTOCOVARIANCE FUNCTION AR(1)



WEAK DEPENDENCE (SHORT MEMORY)

Rosenblatt (1956), Billingsley (1961), Ibragimov (1962), etc ...

Theorem

Let (X_t) to be a stationary process L^2
We assume that

$$(\mathrm{cov}(X_1, X_k))_{k\in\mathbb{N}} \in \ell^1 \text{ + conditions}$$

Then

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}(X_j - \mathbb{E}X_1) \xrightarrow{\mathcal{L}} \sigma^2 B(t), \text{ with } \sigma^2 = \sum_{k=-\infty}^{\infty} \mathrm{cov}(X_1, X_k)$$

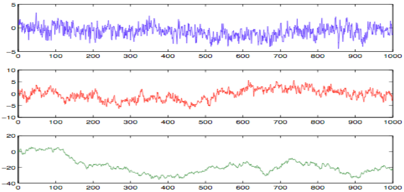
Conclusion

Hurst's findings are incompatible with the assumption of weak dependence

Linear model

Consider a linear process
$\mathcal{X} = \{X_n, n \in \mathbb{N}\}$ defined by

$$X_n = \sum_{j=0}^{\infty} \psi_j \epsilon_{n-j}$$



where

- (ϵ_n) are iid with zero mean and finite variance
- $(\psi_j)_j$ is a deterministic sequence in ℓ^2

Theorem (Davydov 70)

If

$$\mathrm{var}\left(\sum_{i=1}^{n} X_i\right) = n^{2H} L(n)$$

where $H \in (0, 1)$ and where L slowly varying function at ∞
Then

$$\frac{1}{n^H L(n)^{1/2}} \sum_{j=1}^{[nt]} (X_j - \mathbb{E}X_1) \xrightarrow{D[0,1]} W_H(t)$$

where W_H is the fractional Brownian motion (fBm) with parameter $0 < H < 1$.

APPLICATION

If the coefficients $(\psi_j)_{j\in\mathbb{N}}$ behave as $\psi_j \sim j^{d-1}L(j)$ as $j \to \infty$ with

- $d \in (0, \frac{1}{2})$
- L is a slowly varying function at infinity

Then

- $\sigma_x(k) = k^{2d-1}L(k) \notin \ell^1$
- $$\mathrm{var}(\sum_{i=1}^{n} X_i) = n^{2H}L(n)$$
  avec $H = 1/2 + d$
- Davydov's theorem holds
- Asymptotic behavior of R/S

$$\frac{\ln(R(n)/S(n))}{\ln n} \xrightarrow{P} 1/2 + d$$

LONG-RANGE DEPENDENCE

Definition (General)

Let $\mathcal{X} = \{X_n, n \in \mathbb{N}\}$ be a stationary process with auto-covariance function σ_X

$$\sigma_X(k) = \text{cov}(X_0, X_k).$$

$\mathcal{X}$ is said to have long memory if

$$\sum_{k\in\mathbb{Z}} |\sigma_X(k)| = \infty$$

Asymptotic theory and statistical inference is possible if we specify the rate of convergence of σ_X

Definition (Hyperbolically decay)

$\mathcal{X}$ is said to have LM if there exist $d \in]0, 1/2[$ and L slowly varying function at ∞ such that

$$\sigma_X(k) = k^{2d-1}L(k)$$

d is the long memory parameter.

DEFINITION RCAR(1)

Consider $\{X(t),\ t \in \mathbb{Z}\}$ the solution of

$$X(t) = aX(t-1) + \zeta(t), \quad t \in \mathbb{Z},$$

where

- $a \in (-1, 1)$ the random coefficient.
- $\{\zeta(t)\}$ a sequence of iid random variables $\mathbb{E}\zeta = 0$, $\mathbb{E}\zeta^2 = 1$,

Properties

Under the assumptions

- $\mathbb{E}(1-a^2)^{-1} < \infty$,
- a is independent of $\{\zeta(t)\}$,

1. The unique strictly stationary solution is given by $X(t) = \sum_{s \le t} a^{t-s}\zeta(s)$
2. $\{X(t),\ t \in \mathbb{Z}\}$ has zero mean and the auto covaraince function is

$$\mathbb{E}[X(0)X(t)] = \mathbb{E}\left(\frac{a^{|t|}}{1-a^2}\right) < \infty.$$

3. The process is not ergodic

LONG MEMORY RCAR(1) PROCESS

Proposition

If the density of $a \in (0,1)$ is of the form

$$g(x) \sim g_1(1-x)^{\beta-1},\ x \to 1$$

for some $\beta > 1$ and $g_1 > 0$, then

$$\sigma_X(k) \sim Ck^{1-\beta} (\text{as } t \to \infty)$$

Moreover

$$\sum_{k=-\infty}^{\infty} |\sigma_X(k)| = \infty \quad \text{if and only if} \quad \beta \in (1,2)$$

$\rightsquigarrow$ long memory processes with parameter $d = 1 - \frac{\beta}{2}$

AGGREGATION

Under the assumptions

- $\{X_i(t)\}$, $i = 1, 2, \ldots$: independent copies of RCAR(1).
- (a_i) are iid and $\mathbb{E}(1 - a_1^2)^{-1} < \infty$,
- for each i : a_i is independent of $\{\zeta_i(t)\}$,

we have

$$N^{-1/2} \sum_{i=1}^{N} X_i(t) \to_{\text{fdd}} \mathcal{X}(t), \quad N \to \infty,$$

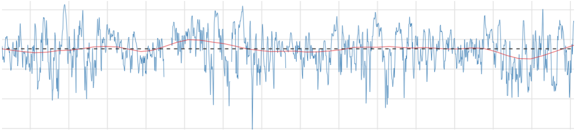
where

- $\mathcal{X}$ is Gaussian with zero mean
- its autocovariance function is the same as RCAR(1) process

$$\mathbb{E}[\mathcal{X}(0)\mathcal{X}(t)] = \mathbb{E}\left(\frac{a^{|t|}}{1 - a^2}\right) < \infty.$$

Reference

Granger 80, Zaffaroni 04, Puplinskaitė, Surgailis 11, Philippe, Puplinskaitė, Surgailis 14



Long memory processes

Random Sampling & continuous time processes

Testing long memory processes

Irregularly spaced time series

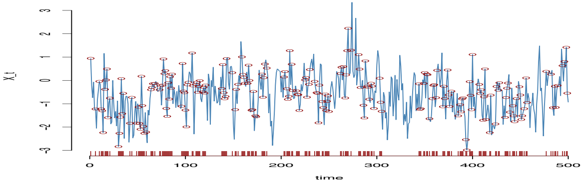
- Irregularly observed time series occur in many fields such as
  - astronomy,
  - finance,
  - environmental,
  - biomedical sciences.
- We do not control the way data are observed, as they are recorded at irregular time points.
- The data are then interpreted as a realization of a continuous temporal process observed at random times

Reference

1. Philippe, Anne, Caroline Robet, and Marie-Claude Viano. 2021. Random Discretization of Stationary Continuous Time Processes. Metrika.https://doi.org/10.1007/s00184-020-00783-1, 84(3), 375 –400 .
2. Ould Haye, Mohamedou, Anne Philippe, and Caroline Robet. 2019. Inference for continuous-time long memory randomly sampled processes.https://hal.archives-ouvertes.fr/hal-02266684

MODEL

- We observe the discrete-time process $\mathbf{Y}$



- We start with a continuous time process.

$$\mathbf{X} = (X_t)_{t \in \mathbb{R}+}$$

- We assume that it is observed at random times $(T_n)_{n \geq 0}$
- We suppose that $\mathbf{Y}$ is of the form

$$Y_n = X_{T_n}, \quad n \in \mathbb{N}.$$

PROPERTIES OF SAMPLED PROCESS Y

Assumptions $\mathcal{H}$:

$\mathcal{H}1$: $\mathbf{X} = (X_t)_{t\in\mathbb{R}+}$ is a second-order stationary continuous time process with zero mean and autocovariance function σ_X.

$\mathcal{H}2$: The random walk $(T_n)_{n\geq 0}$ is independent of $\mathbf{X}$.

$\mathcal{H}3$: $T_0 = 0$. The increments $\Delta_j = T_{j+1} - T_j$ ($j \in \mathbb{N}$) are iid from continuous distribution supported by $\mathbb{R}^+$.

Proposition

Under Assumption $\mathcal{H}$, the discrete-time process

1. $\mathbf{Y}$ is also second-order stationary
2. $\mathbf{Y}$ has with zero mean
3. its autocovariance sequence is

$$\begin{cases} \sigma_Y(0) = \mathrm{var}(Y_1) = \sigma_X(0), \\ \sigma_Y(h) = cov(Y_1, Y_{h+1}) = \mathbb{E}\left[\sigma_X(T_h)\right], \quad h \geq 1. \end{cases}$$

Preservation of the memory when $\mathbb{E}[T_1] < \infty$

We assume that the autocovariance σ_X is regularly varying function at infinity of the form

$$\sigma_X(t) = t^{-1+2d} L(t), \qquad \forall t \geq 1 \tag{1}$$

where

- $0 < d < 1/2$
- L is ultimately non-increasing and slowly varying at infinity, in the sense that L is positive on $[t_0, \infty)$ for some $t_0 > 0$ and

$$\lim_{x \to +\infty} \frac{L(ax)}{L(x)} = 1, \qquad \forall a > 0.$$

Theorem

Under Assumption $\mathcal{H}$ the discrete time process $\mathbf{Y}$ has a long memory and its covariance function behaves as

$$\sigma_Y(h) \sim c h^{-1+2d} L(h), \qquad h \to \infty.$$

with $c = \mathbb{E}[T_1]^{-1+2d}$

Reduction of the memory

Assume that the covariance of $\mathbf{X}$ satisfies

$$|\sigma_X(t)| \le c \min(1, t^{-1+2d}) \qquad \forall t \in \mathbb{R}^+,$$

where $0 < d < 1/2$.
Assumption on the distribution of T_1 : there exists $\beta \in (0,1)$ such that

$$\liminf_{x\to\infty} \left(x^\beta P(T_1 > x)\right) > 0$$

Under these hypotheses there exists $C > 0$ such that

$$|\sigma_Y(h)| \le C h^{\frac{-1+2d}{\beta}}.$$

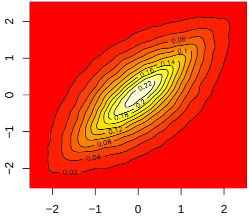
comments

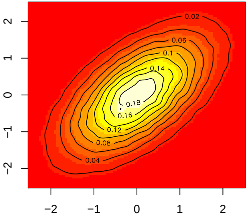
- Condition on T_1 implies that $\mathbb{E}[T_1^\beta] = \infty$.
- the long memory parameter d of the initial process $\mathbf{X}$ is not identifiable using the sampled process. Information on probability distribution of Δ_1 is required.

GAUSSIAN PROCESS

If $\mathbf{X}$ is a Gaussian process satisfying Assumption $\mathcal{H}$ then

1. the marginals of the sampled process $\mathbf{Y}$ are Gaussian.
2. if σ_X is not almost everywhere constant on the set $\{x : s(x) > 0\}$, then $\mathbf{Y}$ is not a Gaussian process.





The estimated density of the centered couple (Y_1, Y_2) where $\Delta_j \sim E(1)$ and $\sigma_X(t) = (1 + t^{0.9})^{-1}$. Gaussian vector (W_1, W_2) with the same covariance matrix Σ_{Y_1,Y_2}

LIMIT THEOREM

We consider the process of partial sums

$$S_n(\tau) = \sum_{j=1}^{[n\tau]} Y_j, \quad 0 \leq \tau \leq 1. \tag{2}$$

Let $\mathbf{X}$ be a Gaussian process with regularly varying covariance function

$$\sigma_X(t) = L(t)t^{-1+2d}$$

, where

- $0 < d < 1/2$
- L slowly varying at infinity and ultimately non-increasing

If Assumption $\mathcal{H}$ holds and $\mathbb{E}[T_1] < \infty$,
then

$$L(n)^{-1/2} n^{-1/2-d} S_n(\tau) \Rightarrow c_d B_{\frac{1}{2}+d}(\tau), \quad \text{in } \mathcal{D}[0,1]$$

where $B_{\frac{1}{2}+d}$ is the fractional Brownian motion with parameter $\frac{1}{2}+d$,

INFERENCE OF d

We have

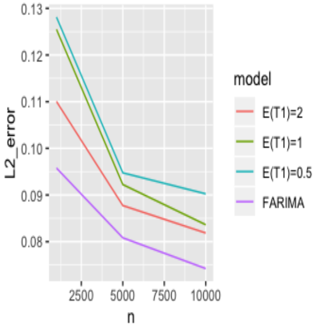
$$\frac{1}{L(n)^{1/2}n^{1/2+d}}\frac{R_n}{S_n}\xrightarrow[n\to\infty]{d}\mathcal{R}:=\sqrt{\frac{\gamma_d}{\sigma_X(0)}}\left(\max_{0\le t\le 1}B^0_{\frac{1}{2}+d}(t)-\min_{0\le t\le 1}B^0_{\frac{1}{2}+d}(t)\right)$$

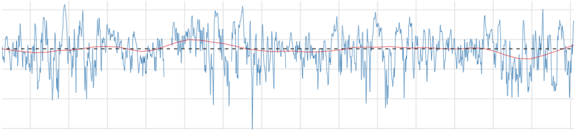
where $B^0_{\frac{1}{2}+d}(t)=B_{\frac{1}{2}+d}(t)-tB_{\frac{1}{2}+d}(1)$ is a fractional Brownian bridge

Particular case : $\sigma_X(t)\sim ct^{2d-1}$ as $t\to\infty$

$$\log\left(\frac{R_n}{S_n}\right)=(1/2+d)\log(n)+\log(\sqrt{c}\mathcal{R})+\varepsilon_n,$$

where $\varepsilon_n\xrightarrow{P}0$





Long memory processes

Random Sampling & continuous time processes

Testing long memory processes

Short memory vs Long memory

Consider linear process

$$X_t = \sum \psi_j \epsilon_{t-j}$$

Short memory
$\psi_k \in \ell^1$
$H = 1/2$
the limit process has independent increments

Long memory
$\psi_k \overset{k\to\infty}{\sim} ck^{d-1}L(k),\ d \in (\frac{-1}{2}, \frac{1}{2}) - \{0\}$
$H = 1/2 + d$
the limit process has dependent increments

The behavior of S_n is used to test

short memory i.e. $H = 1/2$ *(or* $d = 0$ *)*
vs
long memory i.e. $H \neq 1/2$ *(or* $d \neq 0$ *)*

$$H = d + 1/2$$

The most standard statistics

- R/S (Lo, 1991) : based on the range of S_n,
- KPSS (Kwiatkowski *et al.*, 1992) : based on $E(S_n^2)$,
- V/S (Giraitis, Leipus & Philippe, 2003) : based on $Var(S_n)$.

V/S STATISTIC FOR THE SERIES X GIRAITIS, LEIPUS & PHILIPPE, 2003

- V_n is the empirical variance of the partial sums of X

$$V_n = n^{-2}\sum_{k=1}^{n}\left(\sum_{t=1}^{k}(X(t)-\overline{X}_n)\right)^2 - n^{-3}\left(\sum_{k=1}^{n}\sum_{t=1}^{k}(X(t)-\overline{X}_n)\right)^2$$

$\overline{X}_n$ denotes the sample mean of X

- S_q estimates the variance of the limiting law of the partial sums

$$S_q = \sum_{h=-q}^{q}\left(1-\frac{|h|}{q+1}\right)\hat{\gamma}(h)$$

$\hat{\gamma}(h)$ the empirical covariance function of X.

Limit theorem

For linear models (+ assumptions on the white noise)

$$\left(\frac{n}{q}\right)^{-2d} V_n/S_q \Longrightarrow U(d) \qquad \text{when } n, q, n/q \to \infty$$

where

$$U(d) = \int_0^1 (W^0_{1/2+d}(t))^2 dt - \left(\int_0^1 W^0_{1/2+d}(t)dt\right)^2,$$

where $W^0_{1/2+d}(t)) = W_{1/2+d}(t) - tW_{1/2+d}(t)$ is the Fractional Brownian bridge.

Testing Short memory versus Long Memory

1. Under H_0 : short memory

$$V_n/S_q \Longrightarrow U(0) \qquad \text{when } n, q, n/q \to \infty$$

2. Under H_1 : long memory

$$V_n/S_q \xrightarrow{P} +\infty$$

We define

$$R = \{V_n/S_q > u_{1-\alpha}(0)\}$$

where

$$P(U(d) < u_\alpha(d)) = \alpha$$

THe region R gives a critical region for testing short memory vs long memory.

- The asymptotic level is α.
- The test is consistent.

STATIONARITY

H_0 X is a stationary process having long or short memory

$$X_t = \sum \psi_j \epsilon_{t-j}$$

Short memory
$\psi_k \in \ell^1$

Long memory
$\psi_k \overset{k\to\infty}{\sim} ck^{d-1}L(k), \; d \in (\frac{-1}{2}, \frac{1}{2}) - \{0\}$

$$X_n = \sum_{j=0}^{\infty} \psi_j \epsilon_{n-j}$$

where

$$\psi_j \sim j^{d-1}L(j)$$

with $d \in]-1/2, 0[\cap]0, 1/2[$ or $(\psi_j)_j \in \ell^1$

H_1 Non stationary process

- Random walk (Stochastic trend) :

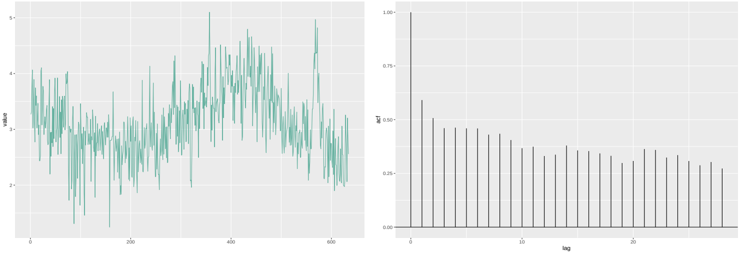
$$X_t = X_{t-1} + \text{stationary process.}$$

- Deterministic trend or structural breaks :

$$X_t = g_n(k) + \text{stationary process,}$$

IS IT A STATIONARY TIME SERIE ?

Illustration on varve data.



DIFFICULTIES

AR(1) model with unit root

- $X_n = aX_n + \epsilon_n$ with a close to 1
- $Y_n = Y_n + \epsilon_n$ Random walk

Strong range dependence

- X_n linear long memory with d close to 1/2
- $Y_n = Y_{n-1} + Z_n$ where Z_n is a stationary linear process with d close to -1/2

TESTING STATIONNARITY *versus* NON STATIONARITY

1. Let $\widetilde{d}$ be a consistent estimate of $d \in]-\frac{1}{2};\frac{1}{2}[$
2. The rejection region
$$R = \{(n/q)^{-2\hat{d}} V_n/S_q > u_{1-\alpha}(\hat{d})\}$$
provides a test with asymptotic level α.
3. The test is consistent against the alternatives H_1 :

In practice :

1. Tests suffer from high empirical sizes, especially near nonstationarity boundary, i.e. when d is close to 1/2.
2. Subject to how good the estimation of d.
3. We have to restrict d to be in a compact $[-a, a]$ with $a < 1/2$.

TIME-DOMAIN VERSUS FREQUENCY-DOMAIN

Assume that $\mathcal{X}$ admits spectral density

$$\sigma_X(h) = \int_{-\pi}^{\pi} e^{i\lambda h} f(\lambda) d\lambda,$$

of the form

$$f(\lambda) = |\lambda|^{-2d} f^*(\lambda), \text{ where } -1/2 < d < 1/2, \text{ and } f^* \text{ is positive and continuous.}$$

The tests are based on the periodogram e.g. Lobato Robinson (1998) ...

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{j=1}^{n} X_j e^{ij\lambda} \right|^2 .$$

Construction of a test statistic

- we split $X_1, \ldots, X_n$ into m blocks each of size ℓ, and we denote $I_{n,i}$ the periodogramof the ith block
- the statitstic is

$$Q_{n,m}(s,d) = m^{-2d} \sum_{j=1}^{s} \frac{I_n(\lambda_j)}{\frac{1}{m} \sum_{i=1}^{m} I_{n,i}(\lambda'_j)}.$$

- Fourier frequencies $\lambda_j = 2\pi j/n$ and $\lambda'_j = 2\pi j/\ell$.

LIMITING DISTRIBUTION OF FREQUENCY STATISTIC

- Under H_0 (stationarity)

$$Q_{n,m}(s,d) \xrightarrow{\mathcal{D}} Q(s,d) = \sum_{i=1}^{2s} \zeta_i(d) Q_i$$

where $Q_1, \ldots, Q_{2s}$ are i.i.d. χ_1^2 random variables and $\zeta_1(d), \ldots \zeta_{2s}(d)$ are eigenvalues of some positive matrix.

- the asymptotic distribution is simply a weighted sum of independent χ^2 random variables.
- The limit distribution is wel defined for $d = 1/2$

Reference Ould Haye,M. and Philippe, A. . 2021. Frequency approach for detecting nonstationarity in dependent data.
`https://hal.archives-ouvertes.fr/hal-02126749`.

Parameter free consistent frequency test. Two alternatives

UnderH_1

1. Stochastic trend (unit root) : $X_t = X_{t-1} + Y_t$ where Y satisfies H_0
2. Deterministic trend (structural breaks) :

$$X_t = g_n(t) + Y_t$$

where

- $g_n(t) = n^{\alpha} g(t/n)$, with $\alpha \geq 0$, and where g is either differentiable with bounded derivative or g is a step function.
- Y_t is a stationary linear process .

Then we have

$$Q_{n,m}(s, 1/2) \xrightarrow{P} \infty.$$

Classification rule :

If $Q_{n,m}(s, 1/2) > Q(s, 1/2)$. then we reject the null hypothesis

- under H_0 : the probability to reject H_0 goes to 0
- under H_1 : the probability to reject H_0 goes to 1

V/S Empirical sizes

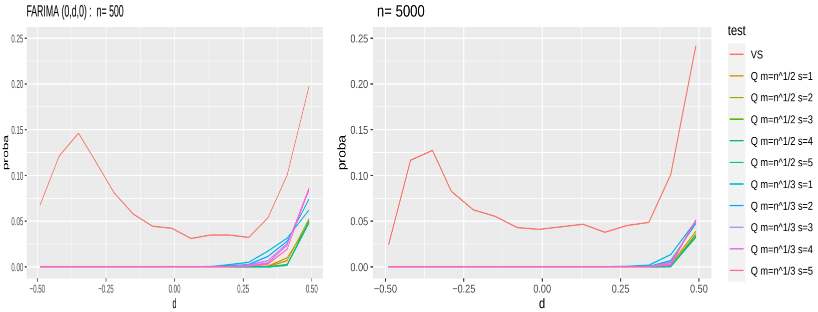


Figure – Empirical level evaluated on FARIMA(0,d,0) process as function of the long memory parameter $d \in (-.5, .5)$. We compare the statistics V/S and $Q_{n,m}(s, 1/2))$ for different values of parameters (m, s) and different sample sizes n

POWER FUNCTION OF THE FREQUENCY TEST

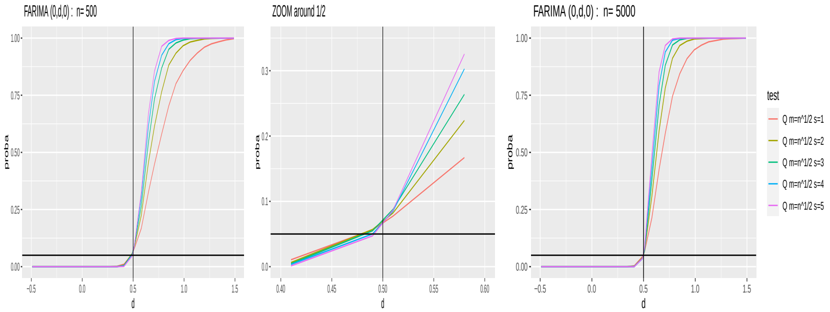


FIGURE – Empirical probability to reject the null hypothesis as function of d the parameter of fractional process with $d \in (-1/2, 3/2)$. We compare $Q_{n,\sqrt{n}}(s, 1/2)$ for different s.