# Modeling and estimation of stationary and nonstationary long memory

Donatas Surgailis

Vilnius University

Plan of the lectures:

1. The notion of long memory (LM). Partial sums and self-similarity

2. Modeling of LM processes

   2.1 Linear models and fractional integration

   2.2 Time-varying fractionally integrated processes

   2.3 Nonlinear functions of linear LM processes

   2.4 ARCH and stochastic volatility LM processes

   2.5 ON/OFF and duration based LM models

3. Estimation of LM

   3.1 Local Whittle estimator

   3.2 Increment Ratio estimator

   3.3 Testing for a change of memory parameter

# 1. The notion of long memory (LM)

## 1.1 Covariance LM

A 2nd order discrete time process $X = \{X_t, t \in \mathbb{Z} := \{\cdots, -1, 0, 1, \cdots\}\}$ is said *covariance stationary* if $EX_t = \mu = const$ and the covariance

$$\text{cov}(X_t, X_s) = \text{cov}(X_0, X_{t-s}) =: \gamma(t-s)$$

depends on $t-s$ only.

The decay rate of $\gamma(j)$ as $j \to \infty$ characterizes the dependence in $X$ [the dependence between lagged variables $X_t$ and $X_s$ as $|t-s| \to \infty$].

Equivalently, the dependence in stationary process $X$ can be characterized by the behavior of spectral density $f(x)$ as $x \to 0$

$$\gamma(k) = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}kx} f(x) \mathrm{d}x, \qquad k \in \mathbb{Z}.$$

<u>Definition 1.1</u> A covariance stationary process $\{X_j\}$ with is said:

*covariance short memory (SM)* if

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| < \infty \quad \text{and} \quad \sum_{k\in\mathbb{Z}} \gamma(k) > 0;$$

*covariance long memory (LM)* if there exist $0 < d < 1/2$, $c_\gamma > 0$ s.t.

$$\gamma(k) \ \sim \ c_\gamma |k|^{2d-1}, \qquad k \to \infty;$$

*covariance negative memory (NM)*, or *antipersistence*, if there exist $-1/2 < d < 0$, $c_\gamma < 0$ s.t.

$$\gamma(k) \ \sim \ c_\gamma |k|^{2d-1}, \quad k \to \infty, \quad \text{and} \quad \sum_{k\in\mathbb{Z}} \gamma(k) = 0.$$

- covariance LM $\Rightarrow \sum_{k\in\mathbb{Z}} |\gamma(k)| = \infty$ ($d \in (0, 1/2)$)
- covariance NM $\Rightarrow \sum_{k\in\mathbb{Z}} |\gamma(k)| < \infty$ ($d \in (-1/2, 0)$)
- $d \in (-1/2, 1/2)$ is called the *memory parameter*
- covariance SM: $d := 0$
- "Almost equivalent" (spectral domain) definition of SM, LM, and NM:

$$f(x) \sim c_f |x|^{-2d}, \;\; x \to 0, \text{ for some } -\tfrac{1}{2} < d < \tfrac{1}{2}, \, c_f > 0$$

$d > 0$: spectral density grows to $\infty$ at $x = 0$: LM

$d < 0$: spectral density decays to 0 at $x = 0$: NM

$d = 0$: spectral density is continuous at $x = 0$ with $f(0) > 0$: SM

- for each spectral density $f$, $\exists$ a corresponding Gaussian process $X$.

$\Rightarrow$ *Gaussian LM, SM and NM processes exist*

- sample mean $\bar{X} := n^{-1} \sum_{t=1}^{n} X_i$

$$\text{var}(\bar{X}) \sim s_d^2 n^{2d-1} \qquad (-1/2 < d < 1/2), \qquad s_d^2 := \frac{c_\gamma}{d(1+d)} \tag{1}$$

- convergence rate $n^{2d-1}$ under LM is much worse than the "classical" rate $n^{-1}$ corresponding to $d = 0$
- 2nd moment analysis is important but not sufficient for making inferences on LM processes

*Proof of (1)*: use

$$\mathrm{var}(\bar{X}) = n^{-1}\sum_{j=-n}^{n}(1-\frac{|j|}{n})\gamma(j)$$

which follows from

$$\begin{aligned}\mathrm{var}(\bar{X}) &= n^{-2}E(\sum_{i=1}^{n}(X_j-\mu))^2\\ &= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{cov}(X_i,X_j) \;=\; n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma(i-j)\end{aligned}$$

In the SM case,

$$n\mathrm{var}(\bar{X}) = \sum_{j=-n}^{n}(1-\frac{|j|}{n})\gamma(j) \;\to\; \sum_{j=-\infty}^{\infty}\gamma(j) = s_0^2.$$

In the LM case,

$$\begin{aligned}n^{1-2d}\mathrm{var}(\bar{X}) &= n^{-2d}\sum_{j=-n}^{n}(1-\frac{|j|}{n})\gamma(j) \;\sim\; c_\gamma n^{-1}\sum_{j=-n}^{n}(1-\frac{|j|}{n})|\frac{j}{n}|^{-(1-2d)}\\ &\sim 2c_\gamma\int_0^1(1-x)x^{-(1-2d)}\mathrm{d}x = \frac{c_\gamma}{2d(1+2d)} = s_d^2.\end{aligned}$$

In the NM case,

$$\begin{aligned}n^{1-2d}\mathrm{var}(\bar{X}) &= n^{-2d}\sum_{j=-n}^{n}\gamma(j) - n^{-2d}\sum_{j=-n}^{n}\frac{|j|}{n}\gamma(j)\\ &= -n^{-2d}\sum_{|j|>n}\gamma(j) - n^{-2d}\sum_{j=-n}^{n}\frac{|j|}{n}\gamma(j)\\ &\sim c_\gamma n^{-1}\sum_{|j|>n}|\frac{j}{n}|^{-(1-2d)} - c_\gamma n^{-1}\sum_{j=-n}^{n}\frac{|j|}{n}|\frac{j}{n}|^{-(1-2d)}\\ &\sim 2c_\gamma\int_1^\infty x^{-(1-2d)}\mathrm{d}x - 2c_\gamma\int_0^1 x^{2d}\mathrm{d}x = \frac{c_\gamma}{2d(1+2d)} = s_d^2.\end{aligned}$$

## 1.2 Distributional LM

<u>Definition 1.2</u> Let $\{X_j\}$ be a stationary process. Assume there exist normalizing constants $A_n \to \infty$ and $B_n$ s.t. the normalized partial sums process converges:

$$A_n^{-1} \sum_{j=1}^{[nt]} (X_j - B_n) \to_{FDD} Z(t), \qquad t \geq 0,$$

where $Z(t) \not\equiv 0$. We say that $\{X_j\}$ has:

*distributional long memory (LM)* if $\{Z(t)\}$ has dependent increments

*distributional short memory (SM)* if $\{Z(t)\}$ has independent increments

- Def. 1.2 is due to Cox (1984): a clear boundary between SM and LM

- Under weak additional conditions, the normalizing constants $A_n$ are regularly varying with index $H > 0$ (in particular, $A_n = n^H$), and the limit process $\{Z(t)\}$ is $H-$*self-similar with stationary increments* ($H-$sssi) (Lamperti theorem), i.e.

$$c^{-H} Z(ct) =_{FDD} Z(t), \qquad \forall c > 0.$$

- $H-$sssi + independent increments $\Rightarrow \{Z(t)\}$ is Lévy stable + finite variance $\Rightarrow \{Z(t)\}$ is Brownian motion

• $H-$sssi + dependent increments: a rich class + Gaussianity $\Rightarrow$ $\{Z(t)\}$ is a *fractional Brownian motion (fBM) with Hurst parameter* $0 < H < 1, H \neq 1/2$:

$$\mathrm{E}[Z(t)Z(s)] = \frac{\sigma^2}{2}(t^{2H} + s^{2H} - |t-s|^{2H}), \quad t, s \geq 0.$$

• $H = 1/2$ corresponds to usual BM with independent increments

• Covariance LM and distributional LM are related but different notions

• In "good" cases, $\{X_j\}$ may have both covariance LM with parameter $d \in (0, 1/2)$ and distributional LM with $A_n = n^H, H = 0.5 + d$, $\{Z(t)\}=$ fBM

• However, there are simple examples of covariance LM processes whose partial sums tend to a $\alpha-$stable Lévy process and therefore such process has distributional SM

## 1.3 Nonstationary LM

Arises in several contexts:

• Stationary increments. A discrete time series $\{X_j\}$ is said $d-$integrated with parameter $1/2 < d < 3/2$ ( $\{X_j\} \in I(d)$) if the differenced process

$$\xi_j := X_j - X_{j-1}$$

is covariance stationary and has LM with memory parameter $d_\xi := d - 1 \in (-1/2, 1/2)$.

$\{X_j\}$ behaves as partial sums process of $\{\xi_j\}$ and var$(X_j) \to \infty$ as $j \to \infty$

• Stationary + trend model: $X_j = \xi_j + g_n(j)$, $\{\xi_j\}$: stationary LM/SM, $g_n(j)$: deterministic trend, e.g.

$$g_n(j) := \begin{cases} \mu_1, & 1 \le j \le n/2, \\ \mu_2, & n/2 < j \le n \end{cases}$$

A change of the mean can be confused with LM ("spurious LM")

• changing memory parameter (time-varying FARIMA)

• continuous time/spatial data

## 1.4 Examples of data sets with LM

• Hydrology: "Hurst effect" of Nile river annual minima (Hurst, 1951): yearly minimal water levels of the Nile river measured near Cairo for the years 622-1281

• Finance: "Phenomenon of LM in asset returns": while daily returns $r_t$ (on SP500 and other market indices, stocks, foreign exchange rates etc) are almost uncorrelated, absolute $|r_t|$ and squared $r_t^2$ returns have autocorrelations which decay very slowly (nonnegligible for $\approx$ 500 lags)

• Geophysics: Hasslet and Raftery (1989): wind speed measurements on Ireland coast

- Global warming: Smith (1991)

- Measurements of 1 kg standard weight by the US National Bureau of Standards: correlations of these measurements were found to decay as $|k|^{-0.8}$

Other examples and discussion: Beran (1992, 1994), Baillie (1996)

## 1.5 Statistical application: testing for long memory and breaks in the mean

In applications it is important to differentiate the given data between SM and LM, or between LM and breaks in the mean (which can be confused with LM)

Hypothesis $H_0$ (stationarity) $X_1, \cdots, X_n$ is a sample from a stationary process $\{X_j = \mu + \xi_j\}$ having distributional SM/LM, in the sense that for some $d \in (-1/2, 1/2)$

$$\frac{1}{n^{d+.5}} S_{[nt]} \to_{FDD} s_d \, B_{d+.5}(t), \quad t \in [0,1], \quad S_n := \sum_{j=1}^{n} \xi_j, \tag{2}$$

where $s_d > 0$ is some constant and $B_{d+.5}$ is fBM with Hurst parameter $H = d + .5$.

The V/S statistic $V/S(q)$ [V = Variance, S = Scale] is defined as a ratio

$$V/S(q) := \frac{V_n}{\hat{s}_q^2},$$

where

$$V_n \quad := \quad n^{-2}\sum_{k=1}^{n}(\mathcal{S}_k-\bar{\mathcal{S}})^2, \qquad \bar{\mathcal{S}} := n^{-1}\sum_{k=1}^{n}\mathcal{S}_k$$

is the empirical variance of centered partial sums

$$\mathcal{S}_k := \sum_{j=1}^{k}(X_j-\bar{X}), \quad \bar{X} := n^{-1}\sum_{j=1}^{n}X_j$$

and $\widehat{s}_q^2$ is the Newey-West estimator of the long-run variance $s_d^2 = \lim_{n\to\infty}\mathrm{var}(S_n)/n^{1+2d}$:

$$\widehat{s}_q^2 := q^{-1}\sum_{|i-j|\le q}\widehat{\gamma}(i-j), \quad \widehat{\gamma}(j) := n^{-1}\sum_{i=1}^{n-j}(X_i-\bar{X})(X_{i+j}-\bar{X})$$

- $q = 0, 1, \cdots$ is the bandwidth parameter, $q = q_n \to \infty$, $q = o(n)$

- Assumption (2) + weak additional assumptions imply

$$(q/n)^{2d}V/S(q) \ \to_{\mathrm{law}} \ Z_d := \int_0^1 (B^0_{d+.5}(t))^2\mathrm{d}t - \Big(\int_0^1 B^0_{d+.5}(t)\mathrm{d}t\Big)^2$$

where $B^0_{d+.5}(t) := B_{d+.5}(t) - tB_{d+.5}(1)$ is fractional Brownian bridge (fBB).

- The distribution of r.v. $Z_d$ depends on a single parameter $d$. Let $c_d(\alpha)$ be the upper $\alpha-$quantile of $Z_d$:

$$\mathrm{P}(Z_d > c_d(\alpha)) = \alpha.$$

- The decision rule of $\alpha-$level V/S test for testing the stationarity hypothesis $H_0$: Reject $H_0$ if

$$V/S(q) > (n/q)^{2\hat{d}}c_{\hat{d}}(\alpha), \tag{3}$$

where $\hat{d}$ is a $o(1/\log(n))-$consistent estimator of unknown memory parameter $d$

- The test (3) can be used to test the null (stationarity) hypothesis $H_0$ against several types of alternatives:

- Alternative $H_T$ (deterministic trend): $X_j = \mu + g_n(j) + \xi_j$, where $\{\xi_j\}$ is the same as in $H_0$ and $g_n(j)$ is a deterministic trend (e.g., $g_n(j) = g(j/n)$ or $g_n(j) = cj^\beta + o(j^\beta)$);

- Alternative $H_U$ (unit root): $X_j - X_{j-1} = \mu + \xi_j$

- The V/S test was developed in Giraitis et al. (2003, 2006). It is a version of the R/S test of Hurst (1951) modified by Lo (1991), and KPSS test of Kwiatkowski et al. (1992)

## 1.6 Some conclusions

- LM inference requires new methods and approaches

- Nonstandard limit distributions and worse convergence rates

- There is no general estimation theory for all LM models

- New probabilistic models of LM processes are of interest

- Studying partial sums limits for concrete LM models is important (and sometimes difficult)

- Tests and inferential procedures often require an approximate knowledge of the memory parameter $d$ which must be estimated

## 1.7 References

Articles:

Beran, J. (1992) Statistical methods for data with long-range dependence (with discussion). *Statistical Sci.* **7**, 404–416.

Cox, D.R. (1984) Long-range dependence: A review. In: H. A. David and H. T. David (eds.) *Statistics: An Appraisal*, pp. 55–74. Iowa State Univ. Press, Iowa.

Diebold, F.X. and Inoue, A. (2001) Long memory and regime switching. *J. Econometrics* **105**, 131–159.

Giraitis, L., Kokoszka, P., Leipus, R. and Teyssière, G. (2003) Rescaled variance and related tests for long memory in volatility and levels. *J. Econometrics* **112**, 265–294.

Giraitis, L., Leipus, R. and Philippe, A. (2006) A test for stationarity versus trends and unit roots for a wide class of dependent errors. *Econometric Theory*, **22**, 989–1029.

Granger, C.W.J. and Hyung, N. (2004) Occasional structural breaks and long memory with an application to the S&P500 absolute stock returns. *J. Empirical Finance* **11**, 399–421.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *J. Econometrics* **54**, 159–178.

Lo, A. (1991) Long-term memory in stock market prices. *Econometrica* **59**, 1279–1313.

Mikosch, T. (2003) Modelling dependence and tails of financial time series. In *Extreme Values in Finance, Telecommunications and the Environment*, eds B. Finkenstädt and H. Rootzén, Chapman and Hall, New York, pp. 185–286.

Books:

Beran, J. (1994) *Statistics for Long Memory processes.* Chapman and Hall, New York.

Doukhan, P., Oppenheim, G. and Taqqu, M.S. (eds.) (2003) *Long Range Dependence: Theory and Applications.* Birkhäuser, Boston 2003.

Robinson, P.M. (ed.) (2003) *Long Memory Time Series.* Advanced Texts in Econometrics, Oxford Univ. Press, Oxford.

Palma, W. (2007) *Long Memory Time Series.* Wiley, New Jersey.

Giraitis, L., Koul, H.L. and Surgailis, D. (forthcoming) *Large Sample Inference for Long Memory Processes*, Imperial College Press.

# Large Sample Inference for Long Memory Processes

**Liudas Giraitis** University of London
**Hira L. Koul** Michigan State University
**Donatas Surgailis** Vilnius University

# Large sample inference *for* Long Memory Processes

Box and Jenkins (1970) made the idea of obtaining a stationary time series by differencing the given, possibly nonstationary, time series popular. Numerous time series in economics are found to have this property. Subsequently, Granger and Joyeux (1980) and Hosking (1981) found examples of time series whose fractional difference becomes a short memory process, in particular, a white noise, while the initial series has unbounded spectral density at the origin, i.e. exhibits long memory.

Further examples of data following long memory were found in hydrology and in network traffic data while in finance the phenomenon of strong dependence was established by dramatic empirical success of long memory processes in modeling the volatility of the asset prices and power transforms of stock market returns.

At present there is a need for a text from where an interested reader can methodically learn about some basic asymptotic theory and techniques found useful in the analysis of statistical inference procedures for long memory processes. This text makes an attempt in this direction. the authors provide in a concise style a text at the graduate level summarizing theoretical developments both for short and long memory processes and their applications to statistics. the book also contains some real data applications and mentions some unsolved inference problems for interested researchers in the field.

Imperial College Press
www.icpress.co.uk

P591 hc
ISBN-13 978-1-84816-278-5
ISBN-10 1-84816-278-2
9 781848 162785

Large sample inference *for* Long Memory Processes

ICP

# Large Sample Inference *for* Long Memory Processes

Liudas Giraitis • Hira L Koul • Donatas Surgailis

Imperial College Press

# Contents

## 2. Modeling of long memory (LM) processes

### 2.1 Linear process

A linear, or moving average (MA($\infty$)) process is given by

$$X_j = \sum_{i=0}^{\infty} a_i \zeta_{j-i}, \quad j \in \mathbb{Z}, \tag{1}$$

where $\{\zeta_j\} \sim IID(0, \sigma^2)$ (i.i.d. with zero mean and variance $\sigma^2$) and $a_i, i = 0, 1, \cdots$ are deterministic MA coefficients. The series (1) converges in $L^2$ and a.s. if and only if $\sum_{i=0}^{\infty} a_i^2 < \infty$ and defines a strictly stationary and covariance stationary process with zero mean $\mathsf{E}X_j = 0$, covariance

$$\gamma(k) = \mathsf{E}[X_0 X_k] = \sigma^2 \sum_{i=0}^{\infty} a_i a_{i+k}$$

and spectral density

$$f(x) = \frac{\sigma^2}{2\pi} | \sum_{j=0}^{\infty} a_j \mathrm{e}^{-\mathrm{i}jx} |^2, \quad x \in \Pi := [-\pi, \pi].$$

- Any stationary Gaussian process with $\int_{\Pi} \log f(x) \mathrm{d}x > -\infty$ admits the MA representation (1) (Wold Decomposition)

- The linear model (1) is mathematically tractable and has simple structure

- LM/SM is essentially due to slow/fast decay of MA coefficients $a_i$ as $i \to \infty$

**Proposition 2.1** (i) (Case $d = 0$) Let

$$\sum_{k=0}^{\infty} |a_k| < \infty, \qquad \sum_{k=0}^{\infty} a_k \neq 0.$$

Then $\sum_{k=0}^{\infty} |\gamma(k)| < \infty, \sum_{k=0}^{\infty} \gamma(k) > 0$, i.e. $\{X_j\}$ of (1) has covariance short memory (SM).

(ii) (Case $0 < d < 1/2$) Let

$$a_k \sim c_a k^{d-1}, \quad k \to \infty$$

for some $0 < d < 1/2$, $c_a \neq 0$. Then $\gamma(k) \sim c_\gamma |k|^{2d-1}$, $k \to \infty$, $c_\gamma := \sigma^2 c_a^2 B(d, 1-2d)$, i.e. $\{X_j\}$ of (1) has covariance long memory LM.

(iii) (Case $-1/2 < d < 0$) Let

$$a_k = c_a j^{d-1}(1 + O(k^{-1})), \quad k \to \infty, \quad \sum_{k=0}^{\infty} a_k = 0$$

for some $-1/2 < d < 0$, $c_a \neq 0$. Then $\gamma(k) \sim c_\gamma |k|^{2d-1}$, $k \to \infty$, $c_\gamma := \sigma^2 c_a^2 B(d, 1-2d)$, $\sum_{k\in\mathbb{Z}} \gamma(k) = 0$, i.e. $\{X_j\}$ of (1) has covariance negative memory (NM).

(iv) In all three cases (i) - (iii),

$$n^{-d-1/2} \sum_{j=1}^{[nt]} X_j \to_{FDD} s_d B_{d+.5}(t), \tag{2}$$

where $B_{d+.5}$ is fBM with $H = d + .5$.

(v) In addition to (i) - (iii) assume

$$\mathrm{E}|\zeta_0|^p < \infty \quad \text{for some} \quad p > \tfrac{1}{H} = \tfrac{1}{.5+d}.$$

Then the convergence in (2) holds in the Skorohod space $D[0,1]$ with the sup-topology.

- Stationary ARMA$(p,q)$ processes are SM and have MA$(\infty)$ representation with exponentially decaying coefficients $a_j$

- (iii)-(iv) go back to Ibragimov (1959) and Davydov (1970). The first paper showed that for linear process, $\sigma_n^2 := \mathrm{var}(S_n) \to$

$\infty$, $S_n = \sum_{j=1}^n X_j$ implies $\sigma_n^{-1} S_n \to_{\text{law}} \mathcal{N}(0,1)$. The second paper showed that (2) follows from regular variation of $\sigma_n^2$

- [GKS]: various extensions of the above CLT for weighted sums $\sum_{j=1}^n w_{nj} X_j$ of linear process

- tightness in $D[0,1]$ is easy from the Kolmogorov criterion

## 2.2 Fractional integration

$Lx_t := x_{t-1}$ shift

$(1-L)x_t = \Delta x_t = x_t - x_{t-1}$ discrete derivative

$(1-L)^{-1} x_t = \Delta^{-1} x_t = \sum_{j=-\infty}^t x_j$ discrete integral (sum)

$\Delta^d$, $-1 \le d \le 1$: a family of interpolating operators between $\Delta$ and $\Delta^{-1}$

$$\Delta^d = \begin{cases} \text{discrete fractional derivative}, & 0 < d < 1, \\ \text{usual discrete derivative}, & d = 1, \\ \text{identity}, & d = 0, \\ \text{discrete fractional integral}, & -1 < d < 0, \\ \text{discrete usual integral}, & d = -1 \end{cases}$$

group property: $\Delta^{d_1}\Delta^{d_2} = \Delta^{d_1+d_2}$

Definition of $\Delta^d$:

$$\begin{aligned}
\Delta^d x_t &:= (1-L)^d x_t \\
&= \sum_{j=0}^{\infty} \psi_j(d) L^j x_t \\
&= \sum_{j=0}^{\infty} \psi_j(d) x_{t-j},
\end{aligned}$$

where for $|z| < 1$

$$\begin{aligned}
(1-z)^d &= \sum_{j=0}^{\infty} \psi_j(d) z^j \qquad \text{Taylor expansion,} \\
\psi_j(-d) &= \left(\frac{d}{1}\right)\left(\frac{d+1}{2}\right)\cdots\left(\frac{d+j-1}{j}\right) = \frac{\Gamma(d+j)}{j!\Gamma(d)}
\end{aligned}$$

From Stirling formula

$$\psi_j(-d) = \frac{1}{\Gamma(d)} j^{d-1}(1+O(j^{-1})) \sim \frac{1}{\Gamma(d)} j^{d-1}, \quad j \to \infty.$$

We have

$$\sum\nolimits_{j=0}^{\infty} |\psi_j(-d)| = \infty, \qquad 0 < d < 1,$$

$$\sum_{j=0}^{\infty} \psi_j^2(-d) \;< \infty, \qquad -1 < d < 1/2,$$

$$\sum_{j=0}^{\infty} |\psi_j(-d)| \;< \infty, \qquad -1 < d < 0,$$

$$\sum_{j=0}^{\infty} \psi_j(-d) \;= 0, \qquad -1 < d < 0,$$

where the last equality follows from

$$0 = (1-1)^d = \sum_{j=0}^{\infty} \psi_j(d), \qquad 0 < d < 1.$$

ARFIMA$(0, d, 0)$ [AutoRegressive Fractional Integrated Moving Average] is defined as stationary solution of

$$\Delta^d X_t = (1-L)^d X_t = \zeta_t, \qquad \{\zeta_t\} \sim IID(0, \sigma^2)$$

The solution is given by

$$X_t = \Delta^{-d}\zeta_t = \sum_{j=0}^{\infty} \psi_j(-d)\zeta_{t-j} \tag{3}$$

• (3) is well-defined for $d < 1/2$

• (3) is invertible for $d > -1/2$

• for any $-1/2 < d < 1/2$ ARFIMA$(0, d, 0)$ of (3) satisfies the assumptions (i) - (iii) of Proposition 2.1

• the last conclusion applies also to ARFIMA$(p, d, q)$ process which is defined as a stationary solution of the difference equation

$$\phi(L)(1-L)^d X_t \;=\; \theta(L)\zeta_t,$$

where

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_q z^q, \qquad \theta(z) = 1 + \theta_1 z + \cdots + \theta_p z^p$$

are polynomials of degrees $p \geq 0, q \geq 0$ which have no common zeros and which have no zeros on the complex unit disc $\{|z| \leq 1\}$

• spectral density of ARFIMA$(p, d, q)$:

$$\begin{aligned} f(x) &= \frac{\sigma^2}{2\pi}|1 - \mathrm{e}^{-\mathrm{i}x}|^{-2d}|\frac{\theta(\mathrm{e}^{-\mathrm{i}x})}{\phi(\mathrm{e}^{-\mathrm{i}x})}|^2 \\ &\sim \frac{\sigma^2}{2\pi}|\frac{\theta(1)}{\phi(1)}|^2|x|^{-2d}, \qquad x \to 0. \end{aligned}$$

## 2.3 Nonhomogeneous fractional integration

Aim: to define "ARFIMA with changing memory parameter $d = d_t$

- nonstationary LM
- mathematically interesting
- different approaches possible
- Want to understand how can $d_t$ control the memory
- straightforward approach of plugging $d \to d_t$ in ARFIMA$(0, d, 0)$

MA representation is not good

$\boldsymbol{D} = \{\boldsymbol{d} := (d_t, t \in \mathbb{Z}) : |d_t| < C (\exists C < \infty), d_t \notin \{-1, -2, \cdots\}\} =$ a class of "infinite dimensional memory parameters" $\boldsymbol{d}$

$-\boldsymbol{d} := (-d_t, t \in \mathbb{Z})$

For any $\boldsymbol{d} \in \boldsymbol{D}$ define two operators (time-varying filters) $A(\boldsymbol{d}), B(\boldsymbol{d})$

$$A(\boldsymbol{d})x_t \;:=\; \sum_{j=0}^{\infty} a_j(t)x_{t-j}, \qquad B(\boldsymbol{d})x_t \;:=\; \sum_{j=0}^{\infty} b_j(t)x_{t-j},$$

such that

$$B(-\boldsymbol{d})A(\boldsymbol{d}) \;=\; A(-\boldsymbol{d})B(\boldsymbol{d}) \;=\; I \quad \text{(identity)}, \tag{4}$$

and

$$A(\boldsymbol{d}) \;=\; B(\boldsymbol{d}) \;=\; (1-L)^{-d} \qquad \text{if } d_t \equiv d = \text{constant} \tag{5}$$

<u>Definition of $A(\boldsymbol{d})$, $B(\boldsymbol{d})$.</u> Put: $a_0(t) = b_0(t) := 1$,

$$a_j(t) \;:=\; \frac{d_{t-1}}{1}\cdot\frac{d_{t-2}+1}{2}\cdot\frac{d_{t-3}+2}{3}\cdots\frac{d_{t-j}+j-1}{j},$$

$$b_j(t) \;:=\; \frac{d_{t-1}}{1}\cdot\frac{d_{t-j}+1}{2}\cdot\frac{d_{t-j-1}+2}{3}\cdots\frac{d_{t-2}+j-1}{j}.$$

- $a_j(t)$ and $b_j(t)$ are obtained from each other by permutation ("reflection"):

$$\begin{array}{cccccccl}
t-j & t-j+1 & \ldots & t-3 & t-2 & | & t-1 & \\
d_{t-j} & d_{t-j+1} & \ldots & d_{t-3} & d_{t-2} & | & d_{t-1} & : a_j(t) \\
d_{t-2} & d_{t-3} & \ldots & d_{t-j+1} & d_{t-j} & | & d_{t-1} & : b_j(t)
\end{array}$$

- Property (5) is immediate from the definitions

- Property (4) reduces to infinite system of equations:

$$\sum_{j=0}^{n} b_j^-(t)a_{n-j}(t-j) = 0, \tag{6}$$
$$\sum_{j=0}^{n} a_j^-(t)b_{n-j}(t-j) = 0, \quad \forall n, t,$$

where $a_j^-(t;\boldsymbol{d}) = a_j(t;-\boldsymbol{d})$, $b_j^-(t;\boldsymbol{d}) = b_j(t;-\boldsymbol{d})$

- In turn, (6) reduces to the following polynomial identity (by putting $x_k = d_{t-k+1} - n + k$): for any $1 \le k < n$

$$\begin{aligned} P_{n,k}(x_1, \cdots, x_k) &:= \sum_{j=0}^{n} (-1)^{n-j} \binom{n}{j} (j+x_1)\cdots(j+x_k) \\ &\equiv 0 \end{aligned} \tag{7}$$

- (7) is not true for $k \ge n$: $\frac{1}{n!}P_{n,k}(0, \cdots, 0) = \left\{ \begin{matrix} k \\ n \end{matrix} \right\}$ is the Stirling number [= the number of ways $k$ different object can be placed into $n$ boxes]

- In view of (4),

$$B(-\boldsymbol{d})^{-1} := A(\boldsymbol{d}), \qquad A(-\boldsymbol{d})^{-1} =: B(\boldsymbol{d})$$

- We are interested in (nonstationary "tv-ARFIMA") processes defined as $L^2-$bounded solutions of equations

$$A(-\boldsymbol{d})X_t = \zeta_t, \qquad B(-\boldsymbol{d})Y_t = \zeta_t, \qquad t \in \mathbb{Z},$$

where $\{\zeta_t\} \sim IID(0, \sigma^2)$ and $\boldsymbol{d} \in \boldsymbol{D}$ is a given infinite dimensional memory parameter.

- The above processes are defined as

$$X_t = A(-\boldsymbol{d})^{-1}\zeta_t = B(\boldsymbol{d})\zeta_t = \sum_{j=0}^{\infty} b_j(t)\zeta_{t-j}, \qquad (8)$$

$$Y_t = B(-\boldsymbol{d})^{-1}\zeta_t = A(\boldsymbol{d})\zeta_t = \sum_{j=0}^{\infty} a_j(t)\zeta_{t-j} \qquad (9)$$

- Natural questions: When $\{X_t\}, \{Y_t\}$ in (8), (9) are well-defined? What is the covariance decay of $\{X_t\}, \{Y_t\}$? Partial sums limits?

- The answers to above questions clearly depend on the memory sequence $\boldsymbol{d}$

- For illustration consider the particular case when $\boldsymbol{d}$ is *periodic* with period $T = 100$, say, and such that the maximal value

$$d_{\max} := \max_{t \in \mathbb{Z}} d_t \;>\; \bar{d} := \frac{1}{100} \sum_{t=1}^{100} d_t$$

is attained at points $t = 0, 100, 200, 300$ etc

- Consider also the nonstationary ARFIMA$(0, d_t, 0)$ process with the same $\boldsymbol{d}$, defined by

$$Z_t \quad := \quad \sum_{j=0}^{\infty} \psi_j(-d_t) \zeta_{t-j}, \tag{10}$$

where $\psi_j(d)$ are ARFIMA$(0, d, 0)$ coefficients.

Let $\bar{d} \in (0, 1/2)$. Then

$$n^{-\bar{d}-.5} \sum_{j=1}^{[nt]} X_t \quad \to_{FDD} \quad c_X B_{\bar{d}+.5}(t),$$

$$n^{-\bar{d}-.5}\sum_{j=1}^{[nt]} Y_t \quad \to_{FDD} \quad c_Y B_{\bar{d}+.5}(t).$$

Let $d_{\max} \in (0,1/2)$. Then

$$\begin{aligned} n^{-d_{\max}-.5}\sum_{j=1}^{[nt]} Z_t \quad &= \quad n^{-d_{\max}-.5}\sum_{k=1}^{[nt/100]} Z_{100k} + o_p(1) \\ &\to_{FDD} \quad c_Z B_{d_{\max}+.5}(t). \end{aligned}$$

• $\{Z_t\}$: memory changes abruptly with $t$. The parameter $d_t$ necessarily satisfies $d_t < 1/2$. Only moments $t$ with $d_t = d_{\max}$ contribute to partial sums limit

• $\{X_t\}, \{Y_t\}$: memory changes smoothly with $t$. The parameter $d_t$ may take arbitrary large values provided the Cesaro mean $\bar{d} < 1/2$. All moments $t = 1, \cdots, n$ contribute to partial sums limit

*Partial sums limits for general memory sequences* $\boldsymbol{d}$

**Definition 2.1** A sequence $\boldsymbol{d} = (d_t, t \in \mathbb{Z})$ is said *almost periodic at* $+\infty$ (denoted by $\boldsymbol{d} \in AP(+\infty)$) if for any $\epsilon > 0$ there exist $T_\epsilon > 0$ and a periodic sequence $\boldsymbol{d}_\epsilon = (d_{\epsilon,t}, t \in \mathbb{Z})$ such that $\sup_{t>T_\epsilon} |d_t - d_{\epsilon,t}| < \epsilon$.

A sequence $\boldsymbol{d} = (d_t, t \in \mathbb{Z})$ is said *almost periodic at* $-\infty$ (denoted by $\boldsymbol{d} \in AP(-\infty)$) if the sequence $(d_{-t}, t \in \mathbb{Z}) \in AP(+\infty)$.

- For any $\boldsymbol{d} \in AP(+\infty)$ there exists its mean value $\bar{d}_+$ at $t = +\infty$, viz.

$$\bar{d}_+ := \lim_{n\to\infty} \frac{1}{n} \sum_{t=s}^{s+n} d_t \qquad \text{uniformly in} \quad s \geq 0.$$

- If $\lim_{t\to\infty} d_t = d_+$ exists then $\boldsymbol{d} \in AP(+\infty)$ and $\bar{d}_+ = d_+$

- If $\boldsymbol{d}$ is almost periodic then $\boldsymbol{d} \in AP(+\infty)$

**Theorem 2.1** *Let* $\boldsymbol{d} \in AP(+\infty) \cap AP(-\infty)$ *and* $\bar{d}_\pm \in (0, 1/2)$. *Then*

$$n^{-\bar{d}_+ - .5} \sum_{j=1}^{[nt]} Y_j \quad \to_{D[0,1]} \quad c_Y (J_{\bar{d}_+}(t) + U_{\bar{d}_+, \bar{d}_-}(t)),$$

*where $c_Y > 0$ is some constant and $J_{\bar{d}_+}(t), U_{\bar{d}_+,\bar{d}_-}(t)$ are Gaussian processes defined as stochastic integrals w.r.t. Gaussian white noise $W(\mathrm{d}x)$ with variance $\mathrm{d}x$:*

$$
\begin{aligned}
J_{\bar{d}_+}(t) &:= \int_0^t W(\mathrm{d}x) \int_x^t (y-x)^{\bar{d}_+ - 1} \mathrm{d}y, \\
U_{\bar{d}_+,\bar{d}_-}(t) &:= \int_{-\infty}^0 W(\mathrm{d}x) \int_0^t (y-x)^{\bar{d}_- - 1} y^{\bar{d}_+ - \bar{d}_-} \mathrm{d}y.
\end{aligned}
$$

- If $\bar{d}_+ = \bar{d}_- =: d$ the process $J_d(t) + U_{d,d}(t)$ is a fBM with $H = d + .5$

- The process $J_d(t), t \geq 0$ is called type II fBM (Robinson and Marinucci, 1999) (the rough part of fBM)

- The process $U_{\bar{d}_+,\bar{d}_-}(t)$ has a.s. infinitely differentiable trajectories on $(0, \infty)$

- All three processes $\{J_{\bar{d}_+}(t)\}, \{U_{\bar{d}_+,\bar{d}_-}(t)\}$ and $\{J_{\bar{d}_+}(t) + U_{\bar{d}_+,\bar{d}_-}(t)\}$ are $H-$self-similar with $H = \bar{d}_+ + 0.5$

- $\{U_{\bar{d}_+,\bar{d}_-}(t)\}$ has asymptotically vanishing increments:

$$U_{\bar{d}_+,\bar{d}_-}(t+T) - U_{\bar{d}_+,\bar{d}_-}(T) \to_{FDD} 0, \qquad T \to \infty.$$

- $\{J_{\bar{d}_+}(t)\}$ has asymptotically stationary increments tending to those of fBM

- Similar results hold for $X_t = B(\boldsymbol{d})\zeta_t$; however there are some differences between the limit processes for filters $A(\boldsymbol{d})$ and $B(\boldsymbol{d})$

- The mean value $\bar{d}_+ = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} d_t$ can be interpreted as the memory intensity of filters $A(\mathbf{d})$ ir $B(\mathbf{d})$ at "distant future" $t = +\infty$, while $\bar{d}_- = \lim_{n\to\infty} \frac{1}{n}\sum_{t=-1}^{-n} d_t$ - as the corresponding memory intensity at "distant past" $t = -\infty$. The limit distribution of partial sums of $Y_t = A(\boldsymbol{d})\zeta_t$ depends on both $\bar{d}_+$ and $\bar{d}_-$ but the normalization depends on $\bar{d}_+$ only. For $X_t = B(\boldsymbol{d})\zeta_t$ the normalization of partial sums is by $n^{\max(\bar{d}_+,\bar{d}_-)+0.5}$

• The case when $-1/2 < \bar{d}_+ < 0$ or $-1/2 < \bar{d}_- < 0$ (negative memory) is open

• An interesting case is when the sequence $\boldsymbol{d} = (d_t, t \in \mathbb{Z})$ is *random* (i.i.d.) In this case partial sums of $X_t = B(\boldsymbol{d})\zeta_t$ and $Y_t = A(\boldsymbol{d})\zeta_t$ tend to fBM with $H = \mathrm{E}d_0 + 0.5$.

## 2.4 Partial sums under slowly changing memory

• Theorem 2.1 refers to the case of *fast changing memory parameter* $\boldsymbol{d} = (d_t, t \in \mathbb{Z})$

• It is of interest to obtain partial sums limit of the nonstationary LM processes $\{X_t\}, \{Y_t\}, \{Z_t\}$ of (8) - (10) when the memory parameter $\boldsymbol{d} = (d_t, t \in \mathbb{Z})$ *changes slowly* in the sense that

$$d_t = d(t/n), \qquad t = 1, \cdots, n$$

where $d(x), x \in [0, 1]$ is piece wise monotone function on the interval $[0, 1]$ (we call $d(x)$ the *memory function*)

• In fact of importance is the behavior of $d(x)$ near the maximum point $x_{\max} = \operatorname{argmax}(d(x), x \in [0,1])$ only

**Assumption 2.1** The function $d(x), x \in [0,1]$ is a measurable function taking values in the interval $(0, 1/2)$ and having a *unique* supremum $d(x_{\max}) =: d_{\max} \in (0, 1/2)$ at some point $x_{\max} \in (0,1)$. Moreover, for some $\gamma > 0$, the exist the limits

$$\lim_{u \downarrow 0} u^{-\gamma}(d(x_{\max}) - d(x_{\max} \pm u)) \; =: \; \Delta_{\pm} > 0.$$

• Assumption 2.1 means that $d(x)$ behaves like a power function with exponent $\gamma$ near the maximum point: $d(x_{\max} \pm u) = d(x_{\max}) - u^{\gamma}(\Delta_{\pm} + o(1)), \; u \downarrow 0$

For $\{X_t\}, \{Y_t\}, \{Z_t\}$ of (8) - (10), denote

$$S_n^X(t) := \sum_{j=1}^{[nt]} X_j, \quad S_n^Y(t) := \sum_{j=1}^{[nt]} Y_j, \quad S_n^Z(t) := \sum_{j=1}^{[nt]} Z_j, \quad t \in [0,1]$$

**Theorem 2.2** *Let $d(\cdot)$ satisfy Assumption 2.1. Then*

$$\Big(\frac{\log^{1/\gamma} n}{n}\Big)^{d_{\max}+1/2} S_n^Y\Big(x_{\max} + \frac{u}{\log^{1/\gamma} n}\Big) \quad \to_{D(\mathbb{R})} \quad V(u),$$

*where the limit process*

$$V(u) \quad := \quad \frac{1}{\Gamma(d_{\max})} \int_{-\infty}^{u} \eta(\mathrm{d}w) \int_{w}^{u} (v-w)^{d_{\max}-1} \mathrm{e}^{-\Delta_{\mathrm{sgn}(v)}|v|^{\gamma}} \mathrm{d}v, \qquad u \in \mathbb{R}$$

*is well-defined as a stochastic integral with respect to a Gaussian white noise* $\eta(\mathrm{d}w)$ *on the real line, with zero mean and variance* $\mathrm{E}(\eta(\mathrm{d}w))^2 = \mathrm{d}w$*.*

- Time is rescaled by factor $\log^{1/\gamma} n$ in the vicinity of $x_{\max}$
- $\{V(u)\}$ has a.s. continuous trajectories and finite limits $\lim_{u\to\pm\infty} V(u) =: V(\pm\infty)$, $V(-\infty) = 0$
- For $\Delta_+ = \Delta_- = 0$, the process $\{V(u) - V(0), u \in \mathbb{R}\}$ is a fBM with $H = d_{\max} + 0.5$
- Theorem 2.1 holds for $S_n^Z$ but not for $S_n^X$ [the limit apparently is different]

- Theorem 2.1 implies

$$\Big(\frac{\log^{1/\beta} n}{n}\Big)^{d_{\max}+1/2} S_n^Y(t) \ \to_{FDD} \ \begin{cases} 0, & t < x_{\max}, \\ V(0), & t = x_{\max}, \\ V(+\infty), & t > x_{\max}, \end{cases}$$

## 2.5 Continuous time nonhomogeneous fractional integration and multifractional processes

- from discrete time to continuous time

- motivation: "statistics of nanoscience"

- modeling of fractal processes with time-varying Hurst parameter

- Tool: nonhomogeneous fractional integration with continuous time $t \in \mathbb{R}$

Classical (homogeneous) Liouville fractional operators $D^\alpha, I^\alpha$ ($0 < \alpha < 1$):

$D^\alpha$ interpolates between $D^0$ and $D^1 = D$

$I^\alpha$ interpolates between $I^0$ and $I^1 = I$

Here:

$D =$ differentiation: $(Df)(t) := \frac{\mathrm{d}f(t)}{\mathrm{d}t}$

$I =$ integration: $(If)(t) := \int_{-\infty}^{t} f(s)\mathrm{d}s$

$I^0 = D^0 =$ identity: $(I^0 f)(t) = (D^0 f)(t) := f(t)$

<u>Definition:</u>

$$\begin{aligned} I^\alpha f(t) &:= \frac{1}{\Gamma(\alpha)} \int_{-\infty}^{t} f(s)(t-s)^{\alpha-1}\mathrm{d}s, \\ D^\alpha f(t) &:= \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\Gamma(1-\alpha)} \int_{-\infty}^{t} f(s)(t-s)^{-\alpha}\mathrm{d}s \end{aligned}$$

Basic property:

$$D^\alpha I^\alpha f = f \qquad (\forall\, f \in L^1 \cap L^\infty). \tag{11}$$

*Proof*: By definition, the l.h.s. of (11) equals

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\Gamma(1-\alpha)} \int_{-\infty}^{t} (t-s)^{-\alpha}\mathrm{d}s \frac{1}{\Gamma(\alpha)} \int_{-\infty}^{s} f(u)(s-u)^{\alpha-1}\mathrm{d}u \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \int_{-\infty}^{t} f(u)\mathrm{d}u \underbrace{\frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} \int_{u}^{t} (t-s)^{-\alpha}(s-u)^{\alpha-1}\mathrm{d}s}_{\equiv 1} \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \int_{-\infty}^{t} f(u)\mathrm{d}u \\ &= f(t). \end{aligned}$$

- fBM = integral of fractionally integrated (differentiated) white noise $\dot{B}(t) := \frac{\mathrm{d}B(t)}{\mathrm{d}t}$
- $B(t) = \int_0^t \dot{B}(s)\mathrm{d}s$
- Let $0 < \alpha < 1/2$. Then

$$\begin{aligned} X(t) &:= \int_0^t (I^\alpha \dot{B})(s)\mathrm{d}s \\ &= \int_0^t \Big\{ \frac{1}{\Gamma(\alpha)} \int_{-\infty}^s \dot{B}(u)(s-u)^{\alpha-1}\mathrm{d}u \Big\} \mathrm{d}s \\ &= \int_{-\infty}^t \dot{B}(u) \Big\{ \frac{1}{\Gamma(\alpha)} \int_0^t (s-u)_+^{\alpha-1}\mathrm{d}s \Big\} \mathrm{d}u \\ &= \frac{1}{\Gamma(1+\alpha)} \int_{-\infty}^t \Big( (t-u)^\alpha - (-u)_+^\alpha \Big) \mathrm{d}B(u) \\ &= \text{fBM} \qquad \text{with} \qquad H = \alpha + 0.5. \end{aligned}$$

- Similarly, let $-1/2 < \alpha < 0$. Then

$$\begin{aligned} X(t) &:= \int_0^t (D^{-\alpha} \dot{B})(s)\mathrm{d}s \\ &= \cdots \\ &= \frac{1}{\Gamma(1+\alpha)} \int_{-\infty}^t \Big( (t-u)^\alpha - (-u)_+^\alpha \Big) \mathrm{d}B(u) \\ &= \text{fBM} \qquad \text{with} \qquad H = \alpha + 0.5. \end{aligned}$$

## Nonhomogeneous fractional integration in cont. time:

Want to define linear operators

$$(D^{\alpha(\cdot)}f)(t) \quad \text{and} \quad (I^{\alpha(\cdot)}f)(t)$$

depending on "functional parameter"

$$\alpha(\cdot) = \{\alpha(t), t \in \mathbb{R}\}, \qquad \alpha(t) \in (0,1)$$

and such that

$$D^{\alpha(\cdot)}I^{\alpha(\cdot)}f(t) \;=\; f(t), \qquad \text{for } f \in L^1 \cap L^\infty, \quad \text{say,} \tag{12}$$

$$D^{\alpha(\cdot)} = D^\alpha, \quad I^{\alpha(\cdot)} = I^\alpha \qquad \text{if } \alpha(\cdot) \equiv \alpha = \text{constant} \tag{13}$$

- unusual: differentiation order $\alpha(t)$ changes with point $t$

Definition:

$$(I^{\alpha(\cdot)}f)(t) \;:=\; \frac{1}{\Gamma(\alpha(t))}\int_{-\infty}^{t} f(s)(t-s)^{\alpha(t)-1}\mathrm{e}^{A_-(s,t)}\mathrm{d}s,$$

$$(D^{\alpha(\cdot)}f)(t) \;:=\; \frac{\mathrm{d}}{\mathrm{d}t}\int_{-\infty}^{t} \frac{1}{\Gamma(1-\alpha(s))} f(s)(t-s)^{-\alpha(s)}\mathrm{e}^{A_+(s,t)}\mathrm{d}s$$

where for $s < t$

$$A_-(s,t) := \int_s^t \frac{\alpha(u)-\alpha(t)}{t-u}\mathrm{d}u, \quad A_+(s,t) := \int_s^t \frac{\alpha(s)-\alpha(v)}{v-s}\mathrm{d}v$$

- For $\alpha(\cdot) \equiv \alpha =$ const., $A_\pm(s,t) \equiv 0 \;\Rightarrow$ (13) immediate

- Proof of (12): plug $g(t) := I^{\alpha(\cdot)}f(t)$ into $D^{\alpha(\cdot)}g(t)$ and use Fubini as in the case of constant $\alpha$. Then (12) follows from the integral identity:

$$\int_s^t (x-s)^{\alpha(x)-1}(t-x)^{-\alpha(x)} \exp\Big\{\int_s^t \frac{\alpha(v)-\alpha(x)}{x-v}\mathrm{d}v\Big\} \times \frac{\sin(\pi\alpha(x))}{\pi}\mathrm{d}x \;\equiv\; 1 \quad (14)$$

- No elementary proof of (14) available

- The proof of (14) as well as the form of $I^{\alpha(\cdot)}$ and $D^{\alpha(\cdot)}$: via discrete time approximations and orthogonality relation $A(\boldsymbol{d})B(-\boldsymbol{d}) = I^0$ for discrete time nonhomogeneous fractional operators with slowly changing memory parameter $d_t = \alpha(t/n)$

- Sufficient condition for (14): $0 < \inf_{t\in\mathbb{R}} \alpha(t) \le \sup_{t\in\mathbb{R}} \alpha(t) < 1$. Moreover, $\alpha(\cdot)$ may have a finite number of discontinuity points on each finite interval. Between those points, $\alpha(\cdot)$ is a $\delta-$Hölder function with some $\delta > 0$.

## Multifractional Gaussian processes

Recall $\dot{B}(t) := \frac{\mathrm{d}B(t)}{\mathrm{d}t}$: Gaussian WN (= derivative of BM)

For given function $H(t) = \alpha(t) + 1/2$, introduce Gaussian processes

$$X_t := \int_0^t (I^{\alpha(\cdot)}\dot{B})(s)\mathrm{d}s, \qquad Y_t := \int_0^t (D^{-\alpha(\cdot)}\dot{B})(s)\mathrm{d}s$$

Less formally:

$$\begin{aligned} X_t &= \int_{-\infty}^t \Big\{ \int_0^t \frac{1}{\Gamma(\alpha(s))}(s-u)^{\alpha(s)-1}\mathrm{e}^{A_-(u,s)}\mathrm{d}s\Big\}\mathrm{d}B(u), \\ Y_t &= \int_{-\infty}^t \frac{1}{\Gamma(1+\alpha(u))}\Big((t-u)^{\alpha(u)}\mathrm{e}^{A_+(u,t)} - (-u)_+^{\alpha(u)}\mathrm{e}^{A_+(u,0)}\Big)\mathrm{d}B(u) \end{aligned}$$

Also introduce

$$Z_t := \int_{-\infty}^t \frac{1}{\Gamma(1+\alpha(t))}\Big((t-u)^{\alpha(t)} - (-u)_+^{\alpha(t)}\Big)\mathrm{d}B(u)$$

- For $\alpha(t) \equiv \alpha =$ const., $\{X_t\}, \{Y_t\}$ and $\{Z_t\}$ coincide with fBM with $H = \alpha + 0.5$

- $\{Z_t\}$ is called *multifractional Brownian motion (mBM)* (Peltier and Lévy Véhél (1995), Benassi, Jaffard and Roux (1997)). mBM is one of the main statistical models of multifractional processes

- Under some boundedness and regularity conditions on $\alpha(\cdot)$, $\{X_t\}, \{Y_t\}$, $\{Z_t\}$ are *locally self-similar* at each point $t \in \mathbb{R}$ and admit a fBM as a *tangent process*:

$$\lambda^{-\alpha(t)-.5} \left\{ \begin{array}{c} X_{t+h\lambda} - X_t \\ Y_{t+h\lambda} - Y_t \\ Z_{t+h\lambda} - Z_t \end{array} \right\} \quad \to_{FDD} \quad B_{\alpha(t)+.5}(h), \qquad \lambda \to 0$$

- The main advantage of $\{X_t\}$ and $\{Y_t\}$ vs. mBM $\{Z_t\}$ is that $\alpha(\cdot)$ in the former processes can be very rough (logarithmic modulus of continuity); for $\{Z_t\}$ the natural condition is $H(t) = \alpha(t) + 1/2 < \min(1, \delta)$, for $\alpha(\cdot) \in C^\delta$

- We expect that increments $\{X_t\}$ and $\{Y_t\}$ have better decorrelation rates than increments of $\{Z_t\}$

## 2.6 References


Bardet, J.-M. and Surgailis, D. (2011) Measuring the roughness of random paths by increment ratios. *Bernoulli* **17** (2011), 749–780.

Bardet, J.-M. and Surgailis, D. (2010) Nonparametric estimation of the local Hurst function of multifractional processes. Preprint.

Benassi, A., Jaffard, S. and Roux, D. (1997) Gaussian processes and pseudodifferential elliptic operators. *Rev. Mat. Iberoamericana* **13**, 19–89.

Bružaitė, K., Surgailis, D. and Vaičiulis, M. (2007) Time-varying fractionally integrated processes with finite or infinite variance and nonstationary long memory. *Acta Appl. Math.* **96**, 99–118.

Davydov, Y. (1970) The invariance principle for stationary process. *Theory Probab. Appl.* **15**, 145–180.

Falconer, K. (2003) The local structure of random processes.

*J. London Math. Soc.* **67**, 657–672.

Giraitis, L., Koul, H.L. and Surgailis, D. *Large Sample Inference for Long Memory Processes* (forthcoming)

Ibragimov, I.A. (1959) Some limit theorems for strictly stationary stochastic processes. *Doklady Akad. Nauk SSSR* **125**, 711–714.

Leipus, R. and Surgailis, D. (2011) Asymptotics of partial sums of linear processes with changing memory parameter. Preprint.

Marinucci, D. and Robinson, P.M. (1999). Alternative forms of fractional Brownian motion. *J. Statist. Plann. Inference* **80**, 111-122.

Peltier, R. and Lévy-Véhél, J. (1994) A new method for estimating the parameter of fractional Brownian motion. Preprint INRIA. Preprint available on `http://hal.inria.fr/docs/ 00/07/42/79/PDF/RR-2396.pdf`.

Philippe, A., Surgailis, D. and Viano, M.-C. (2006) Invariance principle for a class of non stationary processes with long memory. *C. R. Acad. Sci. Paris Ser. 1* **342**, 269–274.

Philippe, A., Surgailis, D. and Viano, M.-C. (2008) Time-varying fractionally integrated processes with nonstationary long memory. *Th. Probab. Appl.* **52**, 651-673.

Surgailis, D. (2008) Nonhomogeneous fractional integration and multifractional processes. *Stochastic Process. Appl.* **118**, 171–198.

# 3. Modeling of long memory (LM) processes: nonlinear models

## 3.1 Nonlinear functions of linear LM processes

Let

$$
\begin{aligned}
X_j &= \sum_{i=0}^{\infty} a_i \zeta_{j-i}, \quad j \in \mathbb{Z}, && (1) \\
a_i &\sim c_a i^{d_X - 1}, \quad i \to \infty, && \exists\ 0 < d_X < 1/2,\ c_a \neq 0
\end{aligned}
$$

be a linear LM process as in Lecture 2, $\{\zeta_t\} \sim IID(0,1)$.

Let $g(x), x \in \mathbb{R}$ be a nonlinear function s.t. $\mathrm{E}g^2(X_0) < \infty$. Then

$$
Y_j := g(X_j), \qquad j \in \mathbb{Z}
$$

is a stationary and covariance stationary process. The natural question are:

- Does $\{Y_j\}$ has SM or LM?
- What is the memory parameter of $\{Y_j\}$?

• What is partial sums limit of $\{Y_j\}$?

• The above questions are of interest for modeling of LM processes, but even more important in statistical inference of LM processes

• Linear Gaussian process $\{X_j\}$ is often used as the first approximation to real data, while many statistics are expressed as a sum $\sum_{j=1}^n g(X_j)$ of some nonlinear function $g$ of observations

• Particular examples of $g$: $g(x) = \mathbf{1}(x \le y)$ (empirical d.f.), $g(x) = x^2$ (empirical variance), $g(x) = \mathrm{e}^x$ (stochastic volatility model)

### *3.1.1 Gaussian $\{X_j\}$: Hermite expansion*

Let $\Phi$ be the standard normal d.f. of r.v. $Z \sim N(0,1)$

Any function $g \in L_2(\mathbb{R}, \Phi)$ can be expanded in Hermite polynomials $H_k(x) := (-1)^k \mathrm{e}^{\mathrm{x}^2/2} \mathrm{d}^{\mathrm{k}} \mathrm{e}^{-\mathrm{x}^2/2}/\mathrm{dx}^{\mathrm{k}}$:

$$g(x) = \sum_{k=0}^{\infty} \frac{J_k}{k!} H_k(x), \qquad J_k = \mathrm{E} g(Z) H_k(Z)$$

The *Hermite rank* of $g \in L_2(\mathbb{R}, \Phi)$ is the index of the first nonzero Hermite coefficient of $g - J_0$:

$$\text{H-rank}(g) := \min\{k \geq 1 : J_k \neq 0\}$$

The Hermite polynomials have the following remarkable orthogonality property. Let $(X, Y)$ be a Gaussian vector with zero means $\mathsf{E}X = \mathsf{E}Y = 0$, unit variances $\mathsf{E}X^2 = \mathsf{E}Y^2 = 1$ and correlation coefficient $\rho = \mathsf{E}XY \in [-1, 1]$. Then

$$\begin{aligned} \mathsf{E}X_k(X)H_j(Y) &= 0, \qquad k \neq j, \\ &= \rho^k k!, \qquad k = j \end{aligned}$$

For any two $g_1, g_2 \in L_2(\mathbb{R}, \Phi)$

$$\mathsf{E}g_1(X)g_2(Y) = \sum_{k,j=0}^{\infty} \frac{J_{1k}J_{2j}}{k!j!} \mathsf{E}H_k(X)H_j(Y) = \sum_{k=0}^{\infty} \frac{J_{1k}J_{2k}}{k!} \rho^k$$

Let $\kappa := \max(\text{H-rank}(g_1), \text{H-rank}(g_2))$. Then

$$|\text{cov}(g_1(X), g_2(Y))| \quad \leq \quad |\rho|^\kappa \sum_{k=\kappa}^{\infty} \frac{|J_{1k}J_{2k}|}{k!}$$

$$\begin{aligned} &\le \quad |\rho|^\kappa \Big(\sum_{k=\kappa}^\infty \frac{J_{1k}^2}{k!}\Big)^{1/2} \Big(\sum_{k=\kappa}^\infty \frac{J_{2k}^2}{k!}\Big)^{1/2} \\ &= \quad |\rho|^\kappa \sqrt{\mathrm{var}(g_1(X))}\sqrt{\mathrm{var}(g_2(Y))} \end{aligned} \qquad (2)$$

Let $\{X_j\}$ be a LM stationary Gaussian process with zero mean, unit variance and covariance

$$\gamma_X(j) \sim c_X j^{2d_X-1}, \quad j \to \infty, \quad \exists\, c_X > 0,\ 0 < d_X < 1/2.$$

Let $g \in L_2(\mathbb{R}, \Phi)$, $\; k^* := \text{H-rank}(g) \ge 1$ and

$$Y_j := g(X_j), \qquad \gamma_Y(j) := \mathrm{cov}(Y_0, Y_j).$$

We have the decomposition of the Gaussian subordinated process $\{Y_j\}$:

$$Y_j - \mathrm{E}Y_j \; = \; \sum_{k=k^*}^\infty \frac{J_k}{k!} H_k(X_j) \; = \; Y_j^0 + R_j, \qquad (3)$$

$$Y_j^0 := \frac{J_{k^*}}{k^*!} H_{k^*}(X_j), \quad R_j := \sum_{k>k^*} \frac{J_k}{k!} H_k(X_j)$$

In (3), $\{Y_j^0\}$ is the main term and $\{R_j\}$ the remainder term which are orthogonal: $\mathrm{E}Y_j^0 R_k = 0$ $(\forall\, k, j \in \mathbb{Z})$ and

$$\gamma_Y(j) \ \sim \ \gamma_{Y^0}(j) = \frac{J_{k^*}^2}{k*!}\gamma_X^{k^*}(j) \ \sim \ c_Y j^{-k^*(1-2d_X)},$$
$$\gamma_R(j) \ = \ O(j^{-(k^*+1)(1-2d_X)}), \qquad j \to \infty$$

- If $k^*(1-2d_X) < 1$ then $\{Y_j = g(X_j)\}$ has covariance LM with memory parameter $1-2d_Y = k^*(1-2d_X)$, or

$$\begin{aligned} d_Y \ &:= \ \frac{1}{2} - k^*(\frac{1}{2} - d_X) \in (0, 1/2) \\ &= \ d_X, \qquad \text{if} \quad k^* = 1, \\ &< \ d_X, \qquad \text{if} \quad k^* > 1 \end{aligned}$$

- Particularly, if $0 < d_X < 1/2$, $h \in L_2(\mathbb{R}, \Phi)$ and $k^* :=$ H-rank$(g) = 1$ then

$$n^{-d_X-.5}\sum_{j=1}^{[nt]}(g(X_j) - \mathsf{E}g(X_j)) \quad \to_{FDD} \quad J_1 s_{d,X}\ B_{d_X+.5}(t)$$

- In the general case $0 < d_X < 1/2$, $g \in L_2(\mathbb{R}, \Phi)$, $k^* :=$ H-rank$(g) \geq 1$ and $k^*(1-2d_X) < 1$,

$$n^{-d_Y-.5}\sum_{j=1}^{[nt]}(Y_j - \mathsf{E}Y_j) \quad \to_{FDD} \quad \frac{J_{k^*}c_a^{k^*}}{k^*!}\ \mathcal{H}_{k^*}(t) \qquad (4)$$

where

$$\mathcal{H}_k(t) \quad := \quad \int_{\mathbb{R}^k} \Big\{ \int_0^t \prod_{i=1}^k (v-u_i)_+^{d_X-1} \mathrm{d}v \Big\} W(\mathrm{d}u_1)\cdots W(\mathrm{d}u_k) \tag{5}$$

is $k-$th order *Hermite process* defined as multiple Wiener-Itô integral w.r.t. Gaussian white noise $W(\mathrm{d}x)$ with zero mean and variance $\mathrm{d}x$

- $\mathcal{H}_k(t)$ is well-defined for $1 \le k < 1/(1-2d_X)$, is $H-$sssi with $H = 1 - k(\frac{1}{2} - d_X)$. $\mathcal{H}_1(t)$ is a fBM and $\mathcal{H}_k(t)$, $k \ge 2$ are non-Gaussian

- Proof of *non-CLT* in (4): write the l.h.s. as "discrete multiple integrals" with respect to "discrete noise" $W_n(A) := n^{-1/2}\sum_{(i/n)\in A}\zeta_i$ and then prove the $L^2(\mathbb{R}^k)-$convergence of the corresponding integrands towards the integrand $\int_0^t \prod_{i=1}^k (v-u_i)_+^{d_X-1}\mathrm{d}v$ of the limiting multiple Wiener-Itô integral

- Dobrushin and Major (1979), Taqqu (1979)

*The case $k^*(1-2d_X) > 1$: CLT*

• If $k^*(1-2d_X) > 1$, the Gaussian subordinated process $\{Y_j = g(X_j)\}$ *has covariance SM*. Indeed, by (2) inequality,

$$
\begin{aligned}
|\gamma_Y(j)| \ = \ |\mathrm{cov}(g(X_0), g(X_j))| \ &\le \ |\gamma_X(j)|^{k^*} \mathsf{E} g^2(X_0) \\
&\le \ const. |j|^{-k^*(1-2d_X)}
\end{aligned}
$$

hence $\sum_{j\in\mathbb{Z}} |\gamma_Y(j)| < \infty$.

• A remarkable property of instantenuous Gaussian subordinated functionals is that their covariance SM implies distributional LM

**Theorem 3.1** (Breuer and Major, 1983) *Let $\{Y_j = h(X_j)\}$ be Gaussian subordinated process, $g \in L_2(\mathbb{R}, \Phi)$, $k^* :=$ H-rank$(g) \ge 1$. Assume that*

$$
\sum_{j\in\mathbb{Z}} |\gamma_Y(j)| < \infty. \tag{6}
$$

*Then*

$$
n^{-1/2} \sum_{j=1}^{[nt]} (Y_j - \mathsf{E} Y_j) \quad \to_{FDD} \quad \sigma B(t), \quad \sigma^2 := \sum_{k\in\mathbb{Z}} \gamma_Y(k),
$$

*where* $\{B(t)\}$ *is a standard BM. Condition (6) is equivalent to*

$$\sum_{j\in\mathbb{Z}} |\gamma_X(j)|^{k*} < \infty. \qquad (7)$$

• Theorem 3.1 was generalized/extended by several authors: Arcones (1994), Soulier (2001), Nourdin, Peccati and Podolskij (2010), Bardet and Surgailis (2010)

• The "classical" method of the proof: method of moments (cumulants): it suffices to show for $S_n := \sum_{j=1}^n (Y_j - \mathrm{E}Y_j)$, $Z \sim N(0,1)$ that for any $k = 2, 3, \cdots$

$$n^{-k/2}\mathrm{E}S_n^k \quad \to \quad \sigma^k \mathrm{E}Z^k = \begin{cases} \sigma^k(k-1)!!, & k \text{ even}, \\ 0, & k \text{ odd}, \end{cases}$$

or

$$n^{-k/2}\mathrm{cum}_k(S_n) \quad \to \quad \sigma^k \mathrm{cum}_k(Z) = \begin{cases} \sigma^2, & k = 2, \\ 0, & k \geq 3, \end{cases}$$

• Using the orthogonality of Hermite polynomials, the proof of the CLT can be reduced to the case when $g$ is a finite sum of Hermite polynomials, or $S_n = \sum_{k=1}^{K} a_k S_n^{(k)}$, $S_n^{(k)} := \sum_{j=1}^{n} H_k(X_j)$, or

$$\begin{aligned}\mathrm{cum}(S_n^{(k_1)}, \cdots, S_n^{(k_p)}) &= \sum_{j_1,\cdots,j_p=1}^{n} \mathrm{cum}(H_{k_1}(X_{j_1}), \cdots, H_{k_p}(X_{j_p})) \\ &= o(n^{p/2}),\end{aligned}$$

for any $p \geq 3$ and $k_1 \geq k^*, \cdots, k_p \geq k^*$

- The last relation is proved using the *diagram formula* for joint cumulants of Hermite polynomials of Gaussian r.v.'s

- [GKS]: a simpler proof of the CLT for $\sum_{j=1}^{n} H_k(X_j)$ which does not use diagrams but uses approximation by $m-$dependent processes [applies to Gaussian MA processes $\{X_j\}$ of (1)]

### *3.1.2 Nongaussian underlying linear process* $\{X_j\}$

Let

$$X_j = \sum_{i=0}^{\infty} a_i \zeta_{j-i}, \quad j \in \mathbb{Z}, \tag{8}$$
$$a_i \sim c_a i^{d_X - 1}, \quad i \to \infty, \qquad \exists\ 0 < d_X < 1/2,\ c_a \neq 0$$

where $\{\zeta_j\} \sim IID(0,1)$ are *nongaussian* r.v.'s, so that $\{X_j\}$ is also a nongaussian stationary process.

Let $g(x), x \in \mathbb{R}$ be a nonlinear function s.t. $\mathrm{E}g^2(X_0) < \infty$. Again, we are interested in LM properties of the subordinated process

$$Y_j := g(X_j), \qquad j \in \mathbb{Z}$$

- The situation is roughly similar as in the Gaussian case but more complex and technically involved

- The role of Hermite rank now is played by the *Appell rank of* $g$:

$$k_A^* = \text{A-rank}(g) := \min\{k \ge 1 : g_k \neq 0\},$$
$$g_k := \frac{\partial^k \mathrm{E}g(X_0 + y)}{\partial y^k}\Big|_{y=0}$$

• For $X_0 \sim N(0,1)$, A-rank($g$) = H-rank($g$) follows from the properties of Hermite polynomials:

$$H_k'(x) = kH_{k-1}(x), \qquad \mathsf{E}H_k(X_0) = 0 \quad (k \ge 1)$$

• The above properties are shared by *Appell polynomials* $A_k(x) := (-\mathrm{i})^k \frac{\partial^k}{\partial y^k}(\mathrm{e}^{\mathrm{i}yx}/\mathrm{E}\mathrm{e}^{\mathrm{i}yX_0})|_{y=0}$

• If $k_A^*(1-2d_X) < 1$ (+ ... some additional assumptions) then $\{Y_j = g(X_j)\}$ has LM and normalized partial sums of $\{Y_j\}$ tend to a Hermite process of order $k_A^*$ [similarly as in the case of Gaussian $\{X_j\}$]

• If $k_A^*(1-2d_X) < 1$ (+ ... some additional assumptions) then $\{Y_j = g(X_j)\}$ has SM and CLT holds

• ... some additional assumptions include $\mathsf{E}\zeta_0^4 < \infty$. If $\mathsf{E}\zeta_0^2 < \infty$, $\mathsf{E}\zeta_0^4 = \infty$ and the d.f. of $\zeta_0^2$ has $\alpha-$regularly tail with $1 < \alpha < 2$ then partial sums of $\{Y_j = g(X_j)\}$ may converge to a $2\alpha(1-d_X)-$ stable distribution for bounded $g$ (Surgailis, 2004)

- More precise moment condition: $\mathsf{E}|\zeta_0|^r < \infty$ for

$$r > \max\Big(2, \frac{1 + k_A^*(1-2d_X)}{1-d_X}\Big) \qquad \text{[mom]}$$

Condition ([mom]) is satisfied for any $1 \le k_A^* < 1/(1-2d_X)$ if $r = 4$ or $\mathsf{E}\zeta_0^4 < \infty$. Condition ([mom]) is always satisfied if $k_A^* = 1$ and $r > 2$.

- In particular, for bounded $g$ with $k_A^* = \text{A-rank}(g) = 1$ and $\mathsf{E}|\zeta_0|^{2+\delta} < \infty$, the limit distribution of partial sums process is Gaussian:

$$\begin{aligned} n^{-d_X-.5}\sum_{j=1}^{[nt]}(g(X_j) - \mathsf{E}g(X_j)) \quad &= \quad n^{-d_X-.5} g_1 \sum_{j=1}^{[nt]} X_j + o_p(1) \\ &\to_{FDD} \quad g_1 s_{d,X}\ B_{d_X+.5}(t) \end{aligned}$$

where

$$g_1 := -\int_{\mathbb{R}} g(x) f'(x) \mathrm{d}x$$

is the first Appell coefficient and $f(x) = \mathrm{d}F(x)/\mathrm{d}x$ is the marginal p.d. of $X_0$, $F(x) := \mathsf{P}(X_0 \le x)$, which exists and belongs to $C^\infty$ under very mild regularity assumption on the noise distribution

- The above result in important for the study of the centered *empirical process*:

$$\widehat{F}_n(y) - F(y) := n^{-1} \sum_{j=1}^{n} [\mathbf{1}(X_j \le y) - \mathsf{E}1(X_j \le y)],$$

which corresponds to $g(x) = \mathbf{1}(x \le y)$. Then

$$g_1 = - \int_{-\infty}^{y} f'(x) \mathrm{d}x = -f(y)$$

and

$$\begin{aligned} n^{-d_X+.5}(\widehat{F}_n(y) - F(y)) &= -f(y) n^{-d_x+.5} \bar{X} \\ &\to_{D(\bar{R})} s_{d,X} f(y) Z, \end{aligned} \tag{9}$$

where $\bar{X} = n^{-1} \sum_{j=1}^{n} X_j$, $Z \sim N(0,1)$

- (9) is known as the *Uniform Reduction Principle* (Dehling and Taqqu, 1988). The URP is fundamental to many inferences under long memory (estimation of unknown mean and regression coefficients, hypothesis testing about the form of the marginal distribution).

- The URP implies that the empirical process under LM is *asymptotically degenerated*: it behaves like a random constant $s_{d,X}Z$ times a *deterministic process* $f(y)$ (= the marginal p.d.). This fact is in complete contrast to what is known when the observations $\{X_j\}$ are i.i.d. or weakly dependent.

- A complete description of limit laws of partial sums of $\{Y_j = g(X_j)\}$ is open

- Proofs are based on martingale decomposition (Ho and Hsing, 1996)

## 3.2 ARCH and stochastic volatility LM processes

### *3.2.1 Motivation: modeling financial returns*

Stylized facts of financial (daily) returns:

- returns $X_t = \log(p_t/p_{t-1})$ are uncorrelated: $\mathrm{corr}(X_t, X_s) \approx 0$ $(t \neq s)$

- squared and absolute returns have long memory: $\mathrm{corr}(X_t^2, X_s^2) \neq 0$, $\mathrm{corr}(|X_t|, |X_s|) \neq 0$ $(|t-s| = 100 \div 500)$

- heavy tails: $\mathrm{E}X_t^4 = \infty$

- conditional mean $\mu_t = \mathrm{E}\left[X_t|F_{t-1}\right] \approx 0$, conditional variance $\sigma_t^2 = \mathrm{E}\left[X_t^2|F_{t-1}\right]$ "randomly varying" (conditional heteroskedasticity)

- leverage effect: past returns and future volatilities negatively correlated: $\mathrm{corr}\left(X_s, \sigma_t^2\right) < 0$ $(s < t)$

- volatility clustering

### *3.2.2 GARCH, ARCH($\infty$) and Linear ARCH (LARCH)*

GARCH$(p,q)$ :

$$X_t = \sigma_t\zeta_{t,} \qquad \sigma_t^2 = \omega \ + \ \sum_{i=1}^{p}\beta_i\sigma_{t-i}^2 + \sum_{i=1}^{q}\alpha_i X_{t-i}^2,$$

$\omega \geq 0, \alpha_i \geq 0, \beta_i \geq 0, p, q = 0, 1, \cdots,$ $\{\zeta_t\}$ i.i.d., $\mathrm{E}\zeta_t = 0$, $\mathrm{E}\zeta_t^2 = 1$

ARCH($\infty$):

$$X_t = \sigma_t\zeta_{t,} \qquad \sigma_t^2 = \omega \ + \ \sum_{i=1}^{\infty}\alpha_i X_{t-i}^2,$$

- GARCH$(p,q)$ : Engle (1982), Bollerslev (1986), Bougerol and Pickard (1992), ..., Teräsvirta (2010, review)

- ARCH($\infty$): Giraitis, Kokoszka and Leipus (2000), Kazakevičius and Leipus (2002, 2003), ..., Giraitis et al. (2010, review)

- $\exists$ stationary solution of ARCH($\infty$) with $\mathrm{E}X_t^2 < \infty \Leftrightarrow \sum_{i=1}^{\infty} \alpha_i < 1$

- ARCH($\infty$) does not allow for long memory in $(X_t^2)$

- $\{\zeta_t\}$ symmetric $\Rightarrow$ no leverage

- Linear ARCH (LARCH)($\infty$) (Robinson (1991), Giraitis et al. (2000, 2004), Berkes and Horváth (2003), Schützner (2009)):

$$X_t = \sigma_t \zeta_t, \qquad \sigma_t \;=\; \omega \;+\; \sum_{i=1}^{\infty} a_i X_{t-i}, \tag{10}$$

$\sum_{i=1}^{\infty} a_i^2 < 1,\ \omega \neq 0,\ \ a_i \in \mathbb{R},\ \{\zeta_t\} \sim IID(0,1)$

- $a_i \sim c i^{d-1}$ $(i \to \infty,\ \exists\, c \neq 0,\ d \in (0, 1/2)$ (e.g., ARFIMA$(0, d, 0)$)

- LARCH model allows for LM in $\{X_t^2\}$ and the leverage effect

- stationary solution of LARCH($\infty$):

$$\sigma_t \;=\; \omega \;+\; \sum_{i=1}^{\infty} a_i X_{t-i}$$

$$
\begin{aligned}
&= \omega + \sum_{i=1}^{\infty} a_i \zeta_{t-i} \sigma_{t-i} \\
&= \omega + \omega \sum_{i=1}^{\infty} a_i \zeta_{t-i} + \sum_{i_1, i_2 = 1}^{\infty} a_{i_1} a_{i_2} \zeta_{t-i_1} \zeta_{t-i_1-i_2} \sigma_{t-i_1-i_2} \\
&= \omega + \omega \sum_{i=1}^{\infty} a_i \zeta_{t-i} + \omega \sum_{i_1, i_2 = 1}^{\infty} a_{i_1} a_{i_2} \zeta_{t-i_1} \zeta_{t-i_1-i_2} + \cdots \\
&= \omega \Big( 1 + \sum_{k=1}^{\infty} \sum_{s_k < \ldots < s_1 < t} a_{t-s_1} a_{s_1-s_2} \cdots a_{s_{k-1}-s_k} \zeta_{s_1} \cdots \zeta_{s_k} \Big) \\
&= \omega \sum_S a_t^S \zeta^S \qquad (11)
\end{aligned}
$$

where

$$
S := \{s_k, \cdots, s_1\} \subset \mathbb{Z}, \quad a_t^S := a_{t-s_1} a_{s_1-s_2} \cdots a_{s_{k-1}-s_k}, \quad \zeta^S := \zeta_{s_1} \cdots \zeta_{s_k}
$$

Note that

$$
\mathsf{E} \zeta^{S_1} \zeta^{S_2} = \begin{cases} 0, & S_1 \neq S_2, \\ 1, & S_1 = S_2, \end{cases} , \qquad \forall\, S_1, S_2 \subset \mathbb{Z}.
$$

Therefore

$$
\mathsf{E} \sigma_t^2 = \omega^2 \sum_{S \subset \mathbb{Z}} (a_t^S)^2
$$

$$
\begin{aligned}
&= \omega^2\Big(1+\sum_{k=1}^{\infty}\big(\sum_{i=1}^{\infty}a_i^2\big)^k\Big) \\
&= \frac{\omega^2}{1-A^2}, \qquad A^2 := \sum_{i=1}^{\infty} a_i^2 < 1 \qquad (12)
\end{aligned}
$$

- Condition $A < 1$ of (12) is necessary and sufficient for the existence of covariance stationary causal solution of LARCH equations (10). Moreover, with $\mathcal{F}_t := \sigma\{\zeta_s, s \le t\}$ = "historic information set",

$$
\mathsf{E}[X_t|\mathcal{F}_{t-1}] = 0, \qquad \mathsf{E}[X_t^2|\mathcal{F}_{t-1}] = \sigma_t^2 = \mathrm{var}(X_t|\mathcal{F}_{t-1})
$$

- (11) is an *orthogonal* Volterra expansion in $\zeta_s, s < t$ convergent in $L^2$. Whence it easily follows

$$
\begin{aligned}
\mathrm{cov}(\sigma_0, \sigma_t) &= \omega^2 \sum_{\emptyset \ne S \subset \mathbb{Z}} a_0^S a_t^S \\
&= \frac{\omega^2}{1-A^2}\sum_{i=1}^{\infty} a_i a_{t+i} \\
&\sim c_\sigma t^{2d-1}, \quad (t \to \infty), \qquad (13)
\end{aligned}
$$

provided

$$
a_i \sim c_a i^{d-1},\ 0 < d < 0.5, \qquad A^2 = \sum\nolimits_{i=1}^{\infty} a_i^2 < 1, \quad (14)
$$

where $c_\sigma := \frac{c_a^2 \omega^2 B(d,1-2d)}{1-A^2}$.

- (13) means that volatility $\{\sigma_t\}$ has covariance LM.

- Under (14), $\{\sigma_t\}$ has distributional LM and its normalized partial sums converge to a fBM

Indeed, from the LARCH equations (10) it follows $\mathrm{E}[X_t|\mathcal{F}_{t-1}] = \sigma_t \mathrm{E}[\zeta_t|\mathcal{F}_{t-1}] = 0$, $\mathcal{F}_t := \sigma\{\zeta_s, s \le t\}$, i.e. $\{X_t, \mathcal{F}_t\}$ is a square integrable martingale difference sequence. Then $\omega = \mathrm{E}\sigma_t$ and

$$\sigma_t - \omega = \sigma_t - E\sigma_t = \sum_{i=1}^{\infty} a_i X_{t-i}$$

is a linear (moving average) process in martingale differences with regularly decaying weights $a_i$. Hence,

$$n^{-d-.5} \sum_{j=1}^{[nt]} (\sigma_j - \mathrm{E}\sigma_j) \ \to_{FDD} \ s_d B_{d+.5}(t)$$

- However, volatility is *not* directly observable. LM is empirically observed in absolute powers of return series: $\{|X_j|^\delta, j \in \mathbb{Z}\}$, for $\delta > 0$.

**Theorem 3.1** *Let, for integer* $\ell = 2, 3, \cdots,$

$$(4^\ell - 2\ell - 1)\mu_{2\ell}^{1/\ell} A^2 \ < \ 1, \qquad A^2 = \sum\nolimits_{i=1}^\infty a_i^2, \qquad \mu_r := \mathrm{E}\zeta_0^r.$$

*Let* $\{X_t, \sigma_t\}$ *be stationary solution to the LARCH equations (10), with coefficients*

$$a_i \quad \sim \quad c_a i^{d-1}, \qquad 0 < d < 1/2, \quad c_a > 0.$$

*Then* $\mathrm{E}X_j^{2\ell} < \infty$ *and*

$$\mathrm{cov}(X_0^\ell, X_t^\ell) \quad \sim \quad c_\ell^2 t^{2d-1} \quad (t \to \infty)$$

*where*

$$c_\ell \ := \ c_a \omega^{-1} \ell \, \mathrm{E}X_0^\ell \ = \ c_a \frac{\partial \mathrm{E}X_0^\ell}{\partial \omega}.$$

*Moreover,*

$$n^{-d-.5} \sum_{j=1}^{[nt]} X_j^\ell \quad \to_{FDD} \quad c_\ell s_d B_{d+.5}(t) \qquad (= \textit{fBM})$$

- Theorem 3.1 holds trivially for $\ell = 1$ since $c_1 = 0$

- Integer powers $\{X_j^\ell, j \in \mathbb{Z}\},\ l = 2, 3, \cdots$ have both covariance and distributional LM

• Similar results: bilinear models characterized by conditional mean and conditional variance

$$\begin{aligned} \mathrm{E}[X_t|\mathcal{F}_{t-1}] &= \mu + \sum_{j=1}^{\infty} b_j X_{t-j}, & (15) \\ \mathrm{var}(X_t|\mathcal{F}_{t-1}) &= \nu^2 + \Big(\omega + \sum_{j=1}^{\infty} a_j X_{t-j}\Big)^2. & (16) \end{aligned}$$

• (15)-(16) nests MA($\infty$) (case $\omega = a_j \equiv 0$, LARCH (case $\mu = b_j \equiv 0$) but also ARCH($\infty$) and GARCH($p, q$) (Giraitis and Surgailis, 2002)

• (15)-(16) allows for modeling of LM in conditional mean and LM in conditional variance, with distinct memory parameters ("double LM")

• If $\mu = b_j \equiv 0$ then

$$\mathrm{E}[X_t|\mathcal{F}_{t-1}] = 0, \quad \sigma_t^2 = \mathrm{var}(X_t|\mathcal{F}_{t-1}) = \nu^2 + \Big(\omega + \sum_{j=1}^{\infty} a_j X_{t-j}\Big)^2 \qquad (17)$$

(17): important generalization of LARCH: $\nu > 0$, implying strict positivity of conditional variance (volatility): $\sigma_t^2 \geq \nu^2 > 0$ a.s.

- Leverage effect ("past returns and future volatilities are negatively correlated"): $\mathsf{E}[X_0\sigma_t^2] < 0$ $(t > 0)$ reduces to $\omega a_j < 0, j \geq 1$

### 3.2.3 Method of the proof of Thm 3.1

- Case $\ell = 2$ only: $\mathrm{cov}(X_0^2, X_t^2) \sim c_\ell^2 t^{2d-1}$

- Since $X_t^2 = \zeta_t^2 \sigma_t^2$, $\mathsf{E}X_t^2 = \mathsf{E}\zeta_t^2 \mathsf{E}\sigma_t^2 = \mathsf{E}\sigma_t^2$, reduces to

$$\mathrm{cov}(\sigma_0^2, \sigma_t^2) \sim c_\ell^2 t^{2d-1} \qquad (t \to \infty).$$

- Take $\omega = 1$ and recall the Volterra representation

$$\sigma_t = \sum_{S \subset (-\infty, t)} a_t^S \zeta^S$$

$S := \{s_k, \cdots, s_1\} \subset \mathbb{Z}, \quad a_t^S := a_{t-s_1} a_{s_1-s_2} \cdots a_{s_{k-1}-s_k}, \quad \zeta^S := \zeta_{s_1} \cdots \zeta_{s_k}$

Hence

$$\mathrm{cov}(\sigma_t^2, \sigma_0^2) = \sum_{S_1, \cdots, S_4} a_t^{S_1} a_t^{S_2} a_0^{S_3} a_0^{S_4} \, \mathrm{cov}(\zeta^{S_1}\zeta^{S_2}, \zeta^{S_3}\zeta^{S_4}) \quad (18)$$

- $\text{cov}(\zeta^{S_1}\zeta^{S_2}, \zeta^{S_3}\zeta^{S_4}) = 0$ unless sets $S_1, \cdots, S_4$ are "matched":

  each $s \in S_i$ is matched by an element from $\cup_{j \neq i} S_j$, $i = 1, 2, 3, 4$,

  at least one $s \in S_1 \cup S_2$ is matched by an element from $S_3 \cup S_4$.

Fig. 1

- diagram = "pattern of matching (summation)"

Definition 3.1 Let be given a table $I = (I_1, \cdots, I_4)$ consisting of four rows $I_1, \cdots, I_4$ having $k_1 = |I_1|, \cdots, k_4 = |I_4|$ elements, respectively. We write $I = I(k)_4$, $(k)_4 = (k_1, \cdots, k_4)$.

A *diagram* $\gamma = (V_1, \cdots, V_p)$ is a partition of $I$ such that:

1. Each $V_q$ intersects at most 1 element of each row: $|V_q \cap I_j| \le 1$, $q = 1, \cdots, p$, $j = 1, \cdots, 4$

2. $V_1, \cdots, V_q$ are ordered from the left;

3. $2 \le |V_q| \le 4$, $\forall\, q = 1, \cdots, p$

The class of all diagrams over table $I$ is denoted by $\Gamma_I$.

- The sum in (18) can be rewritten by first, choosing a table $I(k)_4$ with some $k_1 = |I_1|, \cdots, k_4 = |I_4|$, then, second, choosing a "matching pattern" $\gamma \in \Gamma_{I(k)_4}$ and, third, summing over all quadruplets $(S)_4 = (S_1, \cdots, S_4) \subset \mathbb{Z}^4, |S_1| = k_1, \cdots, |S_4| = k_4$ which follow this "matching pattern" $\gamma$ (denoted $(S)_4 \sim \gamma$):

$$\mathrm{cov}(\sigma_t^2, \sigma_0^2) = \sum_{(k)_4} \sum_{\gamma \in I(k)_4} \mu_\gamma \sum_{(S)_4 \sim \gamma} a_t^{S_1} a_t^{S_2} a_0^{S_3} a_0^{S_4}, \qquad (19)$$

where

$$\mu_\gamma := \mathrm{cov}(\zeta^{S_1}\zeta^{S_2}, \zeta^{S_3}\zeta^{S_4}) \quad \text{for } (S)_4 \sim \gamma \quad \text{depends only on } \gamma$$

- The problem reduces to the study of the asymptotics of $w_\gamma(t) := \sum_{(S)_4 \sim \gamma} a_t^{S_1} a_t^{S_2} a_0^{S_3} a_0^{S_4}$ as $t \to \infty$.

- Two classes of diagrams: *regular* and *irregular*: $\Gamma_I = \Gamma_I^{\mathrm{reg}} \cup \Gamma_I^{\mathrm{irreg}}$

- The main contribution to the asymptotics of $\mathrm{cov}(\sigma_t^2, \sigma_0^2)$ comes from regular diagrams:

$$w_\gamma(t) \sim c^2(\gamma) t^{2d-1} \quad (\gamma \in \Gamma_I^{\mathrm{reg}}), \qquad w_\gamma(t) = o(t^{2d-1}) \qquad (\gamma \in \Gamma_I^{\mathrm{irreg}})$$

and

$$\sum_{(k)_4} \sum_{\gamma\in I(k)_4}^{\text{reg}} \mu_\gamma w_\gamma(t) \sim c^2 t^{2d-1}, \qquad \sum_{(k)_4} \sum_{\gamma\in I(k)_4}^{\text{irreg}} \mu_\gamma w_\gamma(t) = o(t^{2d-1})$$

- Regular diagrams have a simple graphical structure:

Fig. 2

- The proof is rather technical

### *3.2.4 Stochastic volatility model: EGARCH*

- General SV:

$$X_t = \sigma_t \zeta_t, \qquad \{\zeta_t\} \sim IID(0,1), \qquad 0 < \sigma_t \sim \mathcal{F}_{t-1}$$

where $\mathcal{F}_{t-1}$ is the "past information set" (a sigma-field s.t. $X_s, \zeta_s, s \le t-1$ are $\mathcal{F}_{t-1}-$measurable and $\zeta_s, s \ge t$ are independent of $\mathcal{F}_{t-1}$)

- Note $\mathsf{E}[X_t|\mathcal{F}_{t-1}] = 0$, $\mathsf{var}[X_t|\mathcal{F}_{t-1}) = \sigma_t$

- It is often assumed

$$\sigma_t = f(\eta_t)$$

where $f$ is a (nonlinear) function and $\{\eta_t\}$ is a stationary process (linear or Gaussian)

- LM SV model (Harvey, 1998), (Breidt et al., 1998):

$$\begin{aligned} \sigma_t &= \mathrm{e}^{\eta_t}, \quad \eta_t = a + \sum_{j=1}^{\infty} a_j \xi_{t-j}, \qquad \{(\zeta_t, \xi_t)\} \sim IID, \\ a_j &\sim c_a j^{d-1}, \qquad j \to \infty, \quad c_a > 0, \quad 0 < d < 1/2 \end{aligned}$$

- Case $\xi_j = g(\zeta_j)$ corresponds to the Exponential Generalized ARCH (EGARCH) model of Nelson (1991)

- The natural decomposition:

$$|X_t|^\delta = \underbrace{[|\zeta_t|^\delta - \mathrm{E}|\zeta_t|^\delta]\mathrm{e}^{\delta\eta_t}}_{\text{mart. differ.}} + c_\delta \mathrm{e}^{\delta\eta_t}, \qquad \mathrm{c}_\delta := \mathrm{E}|\zeta_t|^\delta$$

- Reduces studying LM properties of $\{|X_t|^\delta\}$ to those of the exponential process $\{\mathrm{e}^{\delta\eta_t}\}$

- For $\{\zeta_t\} = \{\eta_t\} \sim N(0,1)$ [Gaussian case, Harvey (1998)],

$$\begin{aligned} \mathrm{cov}(|X_0|^\delta, \||X_t|^\delta &= (\mathrm{E}|X_0|^\delta)^2(\mathrm{e}^{\delta^2 \mathrm{cov}(\eta_0,\eta_t)} - 1) \\ &\sim const.\,\mathrm{cov}(\eta_0, \eta_t) \sim const.\, t^{2d-1}, \qquad t \to \infty \end{aligned}$$

- Similar result when $\{\eta_t\}$ is a linear process (Surgailis and Viano, 2002)

- Normalized partial sums of $\{|X_t|^\delta\}$ tend to fBM

## 3.3 ON/OFF and other duration based LM models

- LM can arise in duration based queueing and traffic models when inter-arrival times are heavy tailed

- There is an extensive probabilistic and engineering literature dealing with the phenomenon of self-similarity and long-range dependence in network traffic. See the monograph (Park and Willinger, 2000) and the references herein

- The simplest models: ON/OFF model and the infinite source Poisson model

### 3.3.1 ON/OFF model

*ON/OFF model*: stationary process $\{X_t, t \geq 0\}$ alternates between two states: ON state: $X_t = 1$ and OFF state: $X_t = 0$. The durations

$$U_i^{\mathrm{on}} > 0, \qquad U_i^{\mathrm{off}} > 0, \quad i = 1, 2, \cdots$$

are all independent, have tail distribution functions

$$\bar{F}^{\mathrm{on}}(x) := \mathrm{P}(U^{\mathrm{on}} > x), \qquad \bar{F}^{\mathrm{off}}(x) := \mathrm{P}(U^{\mathrm{off}} > x), \qquad x > 0.$$

and finite means $\mu_{\mathrm{on}} = \mathrm{E}U^{\mathrm{on}}, \mu_{\mathrm{off}} = \mathrm{E}U^{\mathrm{off}}$.

*Stationary version.* Define the initial times $U_0^{\mathrm{on}}, U_0^{\mathrm{off}} \geq 0$ independent of $\{U_i^{\mathrm{on}}, U_i^{\mathrm{off}}, i \geq 1\}$ by

$$\mathrm{P}(U_0^{\mathrm{on}} \geq x) := \mu^{-1} \int_x^\infty \bar{F}^{\mathrm{on}}(y)\mathrm{d}y,$$
$$\mathrm{P}(U_0^{\mathrm{off}} \geq x) := \mu^{-1} \int_x^\infty \bar{F}^{\mathrm{off}}(y)\mathrm{d}y, \qquad \mu := \mu_{\mathrm{on}} + \mu_{\mathrm{off}}$$

and define the alternating point process

$$0 \leq U_0^{\mathrm{on}} < T_0 < T_0 + U_1^{\mathrm{on}} < T_1 + U_2^{\mathrm{on}} < \cdots,$$

where

$$T_0 := U_0^{\text{on}} + U_0^{\text{off}}, \qquad T_n := T_0 + \sum_{i=1}^{n} (U_i^{\text{on}} + U_i^{\text{off}}), \qquad n \geq 1.$$

Then

$$X_t := \mathbf{1}(U_0^{\text{on}} > t) + \sum_{n=0}^{\infty} \mathbf{1}(T_n \leq t < T_n + U_{n+1}^{\text{on}}), \qquad t \geq 0.$$

- Assume that ON durations are heavy tailed:

$$\bar{F}^{\text{on}}(x) \sim \frac{c_{\text{on}}}{x^{\alpha_{\text{on}}}}, \quad x \to \infty, \quad 1 < \alpha_{\text{on}} < 2, \quad c_{\text{on}} > 0,$$

and OFF durations have lighter tails: $\bar{F}^{\text{off}}(x) = o(\bar{F}^{\text{on}}(x))$. Then the ON/OFF process has covariance LM:

$$\text{cov}(X_0, X_t) \quad \sim \quad \Big(\frac{c_{\text{on}}\, \mu_{\text{off}}^2}{(\alpha_{\text{on}} - 1)\mu^3}\Big) \frac{1}{t^{\alpha_{\text{on}}-1}}, \quad t \to \infty. \qquad (20)$$

- Note $\int_0^\infty |\text{cov}(X_0, X_t)|\, \mathrm{d}t = \infty$

- The proof of (20) in Heath et al. (1998) uses renewal methods

• Intuitively: LM in (20) is due to a few very long ON durations. E.g., the initial duration

$$\mathsf{P}(U_0^{\mathsf{on}} \ge t) \sim \frac{c_{\mathsf{on}}}{\mu} \int_t^\infty y^{-\alpha_{\mathsf{on}}} \mathrm{d}y = \frac{c_{\mathsf{on}}}{\mu(\alpha_{\mathsf{on}} - 1)} t^{-(\alpha_{\mathsf{on}} - 1)}$$

decays at the same rate as the r.h.s. of (20).

• On the other hand, "partial sums" (integrated) ON/OFF process tends to an asymmetric $\alpha-$*stable Lévy process* $L_\alpha(t), t \ge 0$:

$$T^{-1/\alpha} \int_0^{tT} (X_u - \mathsf{E}X_u) \mathrm{d}u \to_{FDD} L_\alpha(t) \qquad (21)$$

• According to the terminology of Lecture 1, the above ON/OFF process has distributional short memory

• Intuitively, (21) can be explained that $\int_0^T (X_u - \mathsf{E}X_u)\mathrm{d}u \approx \sum_{i=N_T} (U_i^{\mathsf{on}} - \mathsf{E}U_i^{\mathsf{on}})$, where $N_T$ is a random number of ON intervals in $[0, T]$ and $N_T \sim T/\mu$:

$$N_T/T \to \frac{1}{\mu}$$

- Therefore, $\int_0^T (X_u - \mathsf{E}X_u)\mathsf{d}u$ is approximately a sum of $T/\mu$ i.i.d. r.v.'s in the domain of attraction of $\alpha-$stable law and has $\alpha-$stable limiting distribution

- The above facts have important implications in network traffic modeling: The cumulative traffic from a large number $M$ of independent sources is described as a sum of $M$ independent ON/OFF processes with heavy tailed inter-arrival times.

The integrated aggregated ON/OFF process

$$A_M(t) := \int_0^t \sum_{i=1}^{M} X_u^{(i)} \mathsf{d}u$$

describes the total accumulated work in the system. Depending on the mutual rate of growth of $M$ and $T$, the normalized process $A_M(Tt)$, $t \geq 0$ tends either to a $\alpha-$stable Lévy process, or a fractional Brownian motion (Mikosch et al., 2002)

- Similar results (covariance LM and distributional SM (Lévy process): AR(1) model with randomly switching coefficient:

$$X_t = a_t X_{t-1} + \zeta_t$$

where the processes $\{\zeta_t\} \sim IID(0,1)$ and $\{a_t\}$ are independent and $a_t$ switches from 0 to 1 and back according to an ON/OFF process with heavy-tailed inter-arrival times (Leipus and Surgailis, 2003)

### 3.3.2 *Infinite source Poisson model*

*Infinite source Poisson model (M/G/$\infty$ queue)*: transmissions start at Poisson points $\Gamma_k, k \in \mathbb{Z}$ and have random lengths $U_k^{\text{on}}$. The process $N(t)$ is the number of active transmissions at time $t \in \mathbb{R}$

More precisely:

$$\cdots < \Gamma_{-1} < 0 < \Gamma_0 < \Gamma_1 < \cdots$$

is a rate $\lambda > 0$ homogeneous Poisson process on $\mathbb{R}$, and $\{U_k^{\text{on}} > 0, k \in \mathbb{Z}\}$ is an i.i.d. sequence with common distribution $F^{\text{on}}$, independent of the Poisson process. Then

$$\begin{aligned} N(t) &:= \sum_{k\in\mathbb{Z}} \mathbf{1}(\Gamma_k \le t < \Gamma_k + U_l^{\text{on}}) \\ &= \nu(A_t), \qquad A_t := \{(s,y) \in \mathbb{R} \times \mathbb{R}_+ : s \le t < s + y\} \end{aligned}$$

where

$$\begin{aligned} \nu &= \text{Poisson random measure on } \mathbb{R} \times \mathbb{R}_+ \\ &\quad\ \text{with mean measure } \lambda \text{Leb} \times F^{\text{on}} \end{aligned}$$

We have

$$
\begin{aligned}
\mathrm{cov}(N(0), N(t)) &= \mathrm{cov}(\nu(A_0), \nu(A_t)) \\
&= \mathrm{var}(\nu(A_0 \cap A_t)) \\
&= \mathrm{E}\nu(A_0 \cap A_t) \\
&= \int_{A_0 \cap A_t} \lambda \mathrm{d}s \times F^{\mathrm{on}}(\mathrm{d}x) \\
&= \lambda \int_t^\infty F^{\mathrm{on}}(s)\mathrm{d}s \\
&\sim c_{\mathrm{on}} \int_t^\infty s^{-\alpha_{\mathrm{on}}}\mathrm{d}s \\
&= \Big(\frac{c_{\mathrm{on}}}{\alpha_{\mathrm{on}} - 1}\Big) t^{-(\alpha_{\mathrm{on}}-1)}, \qquad t \to \infty.
\end{aligned}
$$

- Thus, for $1 < \alpha_{\mathrm{on}} < 2$ the Infinite source Poisson model $\{N(t), t \in \mathbb{R}\}$ has covariance LM

- Aggregation of independent M/G/$\infty$ queues is equivalent to letting the intensity $\lambda = \lambda(T) \to \infty$ at certain rate when $T \to \infty$

- the total accumulated input is defined by

$$A_T(t) := \int_0^t N_{\lambda(T)}(u)\mathrm{d}u$$

- Limit results for the Infinite source Poisson model are very similar to those for the ON/OFF process: Depending on the growth of the connection rate of $\lambda(T)$, the normalized process $A_T(t), t \geq 0$ tends either to a $\alpha-$stable Lévy process, or a fractional Brownian motion (Mikosch et al., 2002)

- Similarly as in the ON/OFF model, long memory is due to a small number of very long transmissions

- A similar model: the *error duration model* of Parke (1999). The only difference is that the latter model assumes that transmissions start at *deterministic times* $k \in \mathbb{Z}$ instead of Poisson times.

- Another related and simple model: *renewal reward process*

- Mathematically more difficult: queues with restrictions, e.g. M/G/$k$

- Statistical inference for LM duration models [estimation of LM intensity, empirical d.f.] is rather peculiar and less developed

## 3.4 References

Arcones, M.A. (1994) Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Ann. Probab.* **22**, 2242–2274.

Bardet, J.-M. and Surgailis, D. (2010) A general moment bound for a product of Gaussian vector's functionals and a central limit theorem for subordinated Gaussian triangular arrays. Preprint.

Berkes, I. and Horváth, L. (2003) Asymptotic results for long memory LARCH sequences. *Ann. Appl. Probab.* **13**, 641–668.

Breidt, F.J., Crato, N. and de Lima, P. (1998) On the detection and estimation of long memory in stochastic volatility. *J. Econometrics* **83**, 325–348.

Breuer, P. and Major, P. (1983) Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multivariate Anal.* **13**, 425–441.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307–327.

Bougerol, P. and Picard, N. (1992) Stationarity of GARCH processes and some nonnegative time series.*J. Econometrics* **52**, 115–127.

Dehling, H. and Taqqu, M.S. (1989) The empirical process of some long-range dependent sequences with an application to U-statistics. *Ann. Statist.* **17**, 1767–1783.

Dobrushin, R. L. and Major, P. (1979) Non-central limit theorems for nonlinear functionals of Gaussian fields. *Z. Wahrsch. Verw. Gebiete* **50**, 27–52.

Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica* **50**, 987-1008.

Giraitis, L., Robinson, P.M. and Surgailis, D. (2000) A model for long memory conditional heteroskedasticity. *Ann. Appl. Probab* **10**, 1002–1024.

Giraitis, L. and Surgailis, D. (2002) ARCH-type bilinear models with double long memory. *Stoch. Process. Appl.* **100**, 275–300.

Giraitis, L., Leipus, R., Robinson, P.M. and Surgailis, D. (2004) LARCH, leverage and long memory. *J. Financial. Econometrics* **2**, 177–210.

Giraitis, L., Leipus, R. and Surgailis, D. (2009) ARCH($\infty$) models and long memory properties. In: T.G. Andersen, R.A. Davis, J.-P. Kreiss, T. Mikosch (Eds.) *Handbook of Financial Time Series*, pp. 71–84. Springer-Verlag.

Giraitis, L., Leipus, R. and Surgailis, D. (2010) Aggregation of random coefficient GLARCH(1,1) process. *Econometric Theory* **26**, 406–425.

Giraitis, L., Koul, H.L. and Surgailis, D. *Large Sample Inference for Long Memory Processes* (forthcoming)

Giraitis, L., Kokoszka, P.S. and Leipus, R. (2000) Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory* **16**, 3–22.

Harvey, A. (1998) Long memory in stochastic volatility. In: *Forecasting Volatility in the Financial Markets* (eds. J. Knight and S. Satchell), pp. 307–320. Butterworth and Heineman.

Heath, D., Resnick, S. and Samorodnitsky, G. (1998) Heavy tails and long range dependence in ON/OFF processes and associated fluid models. *Math. Oper. Res.* **23**, 145–165.

Ho, H.-C. and Hsing, T. (1996) On the asymptotic expansion of the empirical process of long memory moving averages. *Annals of Statistics* **24**, 992–1024.

Hurvich, C.M. and Soulier, Ph. (2009) Stochastic volatility models with long memory. In: Andersen, T.G., Davis, R.A., J.-P. Kreiss, J.-P. and Mikosch, T., (eds.), *Handbook of Financial Time Series*, 345–354. Springer-Verlag, Berlin Heidelberg.

Kazakevičius, V. and Leipus, R. (2002) On stationarity in the ARCH($\infty$) model. *Econometric Theory* **18**, 1-16.

Leipus, R. and Surgailis, D. (2003) Random coefficient autoregression, regime switching and long memory. *Adv. Appl. Probab.* **35** , 1–18.

Leipus, R., Paulauskas, V. and Surgailis, D. (2005) Renewal regime switching and stable limit laws. *J. Econometrics* **129**, 299–327.

Mikosch, T. (2003) Modelling dependence and tails of financial time series. In: *Extreme Values in Finance, Telecommunications and the Environment* (eds. B. Finkelstädt, H. Rootzén), pp. 185–286. Chapman and Hall.

Mikosch, T., Resnick, S., Rootzén, H. and Stegeman, A. (2002) Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Ann. Appl. Probab.*, **12**, 23–68.

Nelson, D.B. (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometric Th.* **6**, 347–370.

Nourdin, I., Peccati, G. and Podolskij, M. (2010). Quantitative Breuer-Major Theorems.
Preprint available on `http://arxiv.org/abs/1005.3107`

Robinson, P.M. (1991) Testing for strong serial correlation and dynamic heteroskedasticity in multiple regression. *J. Econometrics* **47**, 67–84.

Park, K. and Willinger, W. (2000) Self-similar network traffic: an overview. In: Park, K. and Willinger, W., (eds.), *Self-Similar Network Traffic and Performance Evaluations*, 1-38. Wiley, New York.

Parke, W.R. (1999) What is fractional integration? *Rev. Econom. Statist.* **81**, 632–638.

Schützner, M. (2009) Asymptotic statistical theory for long memory volatility models. Dissertation, Univ. Konstanz.

Soulier, Ph. (2001) Moment bounds and central limit theorem for functions of Gaussian vectors. *Statist. Probab.*

*Lett.* **54**, 193–203.

Surgailis, D. and Viano, M.-C. (2002) Covariance structure and long memory properties of the EGARCH model. *ESAIM Probability and Statistics* **6**, 311–329.

Surgailis, D. (2004) Stable limits of sums of bounded functions of long memory moving averages with finite variance. *Bernoulli* **10**, 327–355.

Surgailis, D. (2008) A quadratic ARCH($\infty$) model with long memory and Levy stable behavior of squares. *Adv. Appl. Probab.* **40**, 1198–1222.

Taqqu, M.S. (1979) Convergence of integrated processes of arbitrary Hermite rank. *Z. Wahrsch. Verw. Gebiete* **50**, 53–83.

Teräsvirta, T. (2009) An introduction to univariate GARCH models. In: T.G. Andersen, R.A. Davis, J.-P. Kreiss, T. Mikosch (Eds.) *Handbook of Financial Time Series*, pp. 17–42. Springer-Verlag.

## 4. Estimation of long memory parameter

I. Parametric estimation:

4.1a General asymptotic theory for linear time-series

4.1b Parametric Whittle estimator (approximate maximum likelihood)

4.1c Example: estimation of ARFIMA$(0, d, 0)$ model

II. Semiparametric estimation:

4.2a Local Whittle semiparameteric estimator

4.2b Log-periodogram semiparametric estimator (Geweke and Porter- Hudak estimator)

4.2c Increment Ratio estimator

## I. Parametric estimation

## 4.1a General asymptotic theory for linear time-series

Example: parametric estimation in short memory ARMA$(p, q)$

$$X_t + a_1 X_{t-1} + \cdots + a_p X_{t-p} = \zeta_t + b_1 \zeta_{t-1} + \cdots + b_q \zeta_{t-q} \tag{1}$$

where:

- polynomials $A(z) = \sum_{j=0}^{p} a_j z^j, \quad B(z) = \sum_{j=0}^{q} b_j z^j$ have no zeros in the unit circle $\{|z| \le 1\}$

- $\{\zeta_j\} \sim \mathrm{WN}(0, \sigma^2)$

Parametrization: ARMA$(p, q)$ model (1) is specified by $p + q + 1$ parameters $\sigma, a_1, \cdots, a_p, \ b_1, \cdots, b_q$; $\sigma$ is called the scale parameter

Denote:

$$\theta = (a_1, \cdots, a_p, b_1, \cdots, b_q)$$

Parameter set: We suppose that the true value $\theta_0$ of parameter $\theta$ is an inner point of a closed parameter set

$$\Theta = I_1 \times \cdots \times I_{p+q} \subset \mathbb{R}^{p+q}$$

where $I_j$ are closed intervals

Spectral density:

$$f(\lambda) \;=\; \frac{\sigma^2}{2\pi}\left|\frac{B(\mathrm{e}^{\mathrm{i}\lambda})}{A(\mathrm{e}^{\mathrm{i}\lambda})}\right|^2$$

is a continuous bounded function on $\Pi := [-\pi, \pi]$, completely specified by parameters $\sigma, \theta$

Problem: to consistently estimate parameters $\sigma, \theta$ from observations $X_1, \cdots, X_n$

Classical time series analysis (LS, MLE, Yule-Walker estimators)

## General linear models (Hannan (1973))

Consider general moving average process:

$$X_t \;=\; \sum_{j=0}^{\infty} b_j(\theta)\zeta_{t-j} \;=\; \zeta_t + \sum_{j=1}^{\infty} b_j(\theta)\zeta_{t-j} \qquad (2)$$

where:

- $b_0(\theta) = 1, \quad \sum_{j=0}^{\infty} b_j^2(\theta) < \infty$

- $\{\zeta_j\} \sim \mathrm{WN}(0, \sigma^2)$

- coefficients $a_j(\theta)$ are uniquely determined by finite dimensional parameter $\theta \in \Theta \subset \mathbb{R}^p$

- (2) includes ARMA$(p, q)$, ARFIMA$(p, d, q)$

- spectral density:

$$f(\lambda) \;=\; \frac{\sigma^2}{2\pi}\left|1 + \sum_{j=1}^{\infty} b_j(\theta)\mathrm{e}^{\mathrm{i}\lambda j}\right|^2 \equiv \sigma^2 k_\theta(\lambda)$$

where

$$k_\theta(\lambda) = \frac{1}{2\pi}\left|1 + \sum_{j=1}^{\infty} b_j(\theta)\mathrm{e}^{\mathrm{i}\lambda j}\right|^2$$

- condition $b_0(\theta) = 1$ *(''standard parametrization'')* allows to estimate $\theta$ and $\sigma$ separately

• "standard parametrization" $a_0(\theta) = 1$ can be also characterized in terms of spectral density: it is equivalent to

$$\int_\Pi \log f(\lambda) \mathrm{d}\lambda \equiv 2\pi \log \sigma^2, \qquad \text{or}$$
$$\int_\Pi \log k_\theta(\lambda) \mathrm{d}\lambda \equiv 0, \quad \forall \theta \in \Theta$$

$\sigma^2$: the variance of the best one step linear prediction

*Whittle's estimates*:

Whittle's objective function:

$$\widetilde{Q}_n(\theta) := \int_\Pi \frac{I_n(\lambda)}{k_\theta(\lambda)} \mathrm{d}\lambda,$$

where

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

is the periodogram (we assume $\mathrm{E}X_t = 0$)

Discretized version of Whittle's objective function:

$$\widehat{Q}_n(\theta) := \frac{2\pi}{n} \sum_{j=1}^n \frac{I_n(\lambda_j)}{k_\theta(\lambda_j)}$$

where $\lambda_j := \frac{2\pi j}{n}, j = 1, 2, \cdots, n$ are called *Fourier frequencies*

$\widetilde{\theta}_n$: minimizes $\widetilde{Q}_n(\theta)$ over $\theta \in \Theta$

$\widehat{\theta}_n$: minimizes $\widehat{Q}_n(\theta)$ over $\theta \in \Theta$

both $\widetilde{\theta}_n$ and $\widehat{\theta}_n$ are called *Whittle's estimates of* $\theta_0$

Theorem 4.1 (Hannan (1973)): If $1/f$ is continuous on $\Pi$ (at $\theta = \theta_0$) then

$$\widetilde{\theta}_n \ \rightarrow \ \theta_0, \qquad \widehat{\theta}_n \ \rightarrow \ \theta_0, \qquad a.s.$$

• The class of time series which can be represented as moving average (2) with uncorrelated noise $\{\zeta_j\}$ is very wide (see Wold decomposition)

• because of this universality, asymptotic properties of Whittle's estimates in the large class might be not very good

• Hannan's consistency result does not give converges rates of Whittle's estimators nor asymptotic normality

• additional restrictions on the model are necessary

*Additional model assumptions:*

$X_1, \cdots, X_n$ is a sample from a stationary *Gaussian* process $\{X_t, t \in \mathbb{Z}\}$ with spectral density

$$f(\lambda) = \sigma^2 k_\theta(\lambda)$$

Parameters $\sigma > 0, \mu = EX_t$ and $\theta \in \Theta$ are *unknown*, $\Theta \subset \mathbb{R}^p$ is a compact set. The function $k_\theta(\cdot)$ is assumed to be *known* (up to the parameter $\theta$) and satisfies the "standard parametrization" condition:

$$\int_\Pi \log k_\theta(\lambda) \mathrm{d}\lambda \equiv 0 \qquad (\forall \theta \in \Theta) \tag{3}$$

• unknown $\theta = (\theta_1, \cdots, \theta_p)$ can include the long memory parameter $d$

## Example 1: ARFIMA$(p, d, q)$

$$f(\lambda) = \frac{\sigma^2 |B(e^{i\lambda})|^2}{2\pi |2\sin(\lambda/2)|^{2d} |A(e^{i\lambda})|^2}$$

where:

• $-1/2 < d < 1/2$ is the memory parameter

•: $\sigma > 0$ is the scale parameter

• $A(z) = \sum_{j=0}^{p} a_j z^j$, $B(z) = \sum_{j=0}^{q} b_j z^j$ are autoregressive polynomials of known orders $p, q$ respectively; $a_0 = b_0 = 1$

Here,

$$\theta = (d, a_1, \cdots, a_p, b_1, \cdots, b_q) \ \in \Theta \subset \mathbb{R}^{p+q+1}$$

and

$$k_\theta(\lambda) \ = \ \frac{|\sum_{j=0}^{q} b_j e^{ij\lambda}|^2}{2\pi |2\sin(\lambda/2)|^{2d} |\sum_{j=0}^{p} a_j e^{ij\lambda}|^2}$$

• "standard normalization" (3) holds because of $a_0 = b_0 = 1$ and $(1-z)^{-d}|_{z=0} = 1$

## Example 2: FEXP$(p, d)$ ("fractional exponential")

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left|1 - e^{i\lambda}\right|^{-2d} \exp\{1 + \sum_{j=2}^{p} \theta_j \cos(j\lambda)\}$$

where:

- $\sigma > 0$ is the scale parameter
- $0 \le d < 1/2$ is the memory parameter
- $\theta = (d, \theta_2, \cdots, \theta_p) \ \in \Theta \subset \mathbb{R}^p$
- $$k_\theta(\lambda) = \frac{1}{2\pi} \left|1 - \mathrm{e}^{\mathrm{i}\lambda}\right|^{-2d} \exp\{1 + \sum_{j=2}^{p} \theta_j \cos(j\lambda)\}$$
- "standard normalization" (3) holds

## 4.1b Parametric Whittle estimate (approximate maximum likelihood)

Recall known facts:

- *Likelihood function*: $L(\theta; x_1, \cdots, x_n)$ = joint probability density of observations $X_1, \cdots, X_n$, as the function of *unknown* parameter $\theta \in \Theta$
- gives the likelihood of the event $X_1 = x_1, \cdots, X_n = x_n$
- if $X_1, \cdots, X_n$ are i.i.d. with common probability density $p(x; \theta)$ then $L(\theta; x_1, \cdots, x_n) = p(x_1; \theta) \cdots p(x_n; \theta)$

• the value $\widehat{\theta}$ which maximizes the likelihood $L(\theta; X_1, \cdots, X_n)$ for a given sample $X_1, \cdots, X_n$ is called the *maximum likelihood estimator* (MLE) of $\theta$

• For i.i.d. or weakly dependent $\{X_t\}$, the MLE is usually $\sqrt{n}-$consistent, asymptotically normal and efficient (has the smallest asymptotic variance)

• for large $n$, the MLE is the "best" (other estimators might be just as good but not better)

## Gaussian likelihood (exact and approximate)

Autocovariance:

$$\gamma(t) = \mathrm{cov}(X_0, X_t) = \sigma^2 r_\theta(t)$$

where

$$r_\theta(t) = \int_\Pi \mathrm{e}^{\mathrm{i}t\lambda} k_\theta(\lambda) \mathrm{d}\lambda.$$

Set

$$R_\theta = (r_\theta(t-s))_{t,s=1,\cdots,n}$$

Then $\sigma^2 R_\theta = (\mathrm{cov}(X_t, X_s))_{t,s=1,\cdots,n}$

*Exact likelihood*:

$$L(\theta; X_1, \cdots, X_n) \; = \; \frac{1}{(2\pi)^{1/2}|\sigma^2 R_\theta|^{1/2}} \exp\{-\frac{1}{2\sigma^2}(X-\mu)' R_\theta^{-1}(X-\mu)\},$$

$(X-\mu)' = (X_1-\mu, \cdots, X_n-\mu)'$

Maximization w.r.t. $\mu$ ($\sigma$ and $\theta$ fixed) gives $\widehat{\mu} = \bar{X} = n^{-1}\sum_{t=1}^{n} X_t$

The MLE of $(\sigma, \theta)$ maximizes the exact log-likelihood:

$$-\frac{1}{2n}\log\sigma^2 - \frac{1}{2n}\log|\det(R_\theta)| - \frac{1}{2n\sigma^2}(X-\bar{X})' R_\theta^{-1}(X-\bar{X})$$

- maximizing exact MLE involves inverting of $R_\theta = (r_\theta(t-s))_{t,s=1,\cdots,n}$
- a complicated task both analytically and computationally ($n$ large!)
- exact Gaussian MLE is sensitive to deviations from Gaussianity

*Approximate maximum likelihood*:

The approximation involves two steps:

Step 1: (under some additional regularity conditions), as $n \to \infty$

$$\frac{1}{n} \log |\det R_\theta| \;\to\; \frac{1}{2\pi} \int\limits_{\Pi} \log k_\theta(\lambda) \mathrm{d}\lambda \equiv 0 \quad (\forall \theta \in \Theta)$$

due to the standard normalization assumption (3)

Step 2: Replacement of the matrix

$$R_\theta^{-1} = \left( \int_\Pi \mathrm{e}^{\mathrm{i}(t-s)\lambda} k_\theta(\lambda) \mathrm{d}\lambda \right)^{-1}_{t,s=\overline{1,n}}$$

by the $n \times n-$matrix

$$S_\theta = \left( \frac{1}{2\pi} \int_\Pi k_\theta^{-1}(\lambda) \mathrm{e}^{\mathrm{i}(t-s)\lambda} \mathrm{d}\lambda \right)_{t,s=\overline{1,n}}$$

The *approximate (Whittle's) likelihood* is

$$-\frac{1}{2} \log \sigma^2 - \frac{1}{2n\sigma^2} (X - \bar{X})' S_\theta (X - \bar{X})$$

Minimization of the approximate likelihood w.r.t. $(\theta, \sigma)$ reduces to the minimization of the objective function:

$$\widetilde{Q}_n(\theta) = \frac{1}{n}(X - \bar{X})' S_\theta (X - \bar{X}) = \int_{-\pi}^{\pi} \frac{I_n^*(\lambda)}{k_\theta(\lambda)} \mathrm{d}\lambda$$

where

$$I_n^*(\lambda) = \frac{1}{2\pi n} \left| \sum\nolimits_{t=1}^{n} \mathrm{e}^{\mathrm{i}t\lambda}(X_t - \bar{X}_n) \right|^2$$

is the periodogram of centered observations $X_1 - \bar{X}, \cdots, X_n - \bar{X}$

The discretized Whittle's objective function is

$$\widehat{Q}_n(\theta) = (2\pi n)^{-1} \sum_{j=1}^{n} \frac{I_n^*(\lambda_j)}{k_\theta(\lambda_j)} = (2\pi n)^{-1} \sum_{j=1}^{n} \frac{I_n(\lambda_j)}{k_\theta(\lambda_j)} \tag{4}$$

- the last equality follows from $I_n^*(\lambda_j) = I_n(\lambda_j)$ for $\lambda_j = 2\pi j/n, j = 1, \ldots, n$
- this follows from $\sum_{j=1}^{n} \mathrm{e}^{\mathrm{i}t\lambda_j} = 0$ for any integer $t$
- in other words: the periodogram is self-centring at Fourier frequencies and the mean-correction is automatically incorporated into the discretized Whittle log-likelihood

**Theorem 4.2** (Fox and Taqqu (1996))

Suppose that $\{X_t\}$ is a stationary Gaussian LM process with spectral density $f(\lambda) = \sigma^2 k_\theta(\lambda)$, where

$$k_\theta(\lambda) = g_\theta(\lambda)|\lambda|^{-2d}$$

where $g_\theta(\lambda)$ is continuous in $(\theta, \lambda)$, $g_\theta(0) > 0$, the parameter $\theta$ includes the long memory parameter $0 < d < 1/2$ and belongs to a compact set $\Theta \subset \mathbb{R}^p$, and the true parameter $\theta_0$ is an inner point of $\Theta$.

Then under some additional regularity conditions on $g_\theta(\lambda)$, the parametric Whittle's estimates $\widetilde{\theta}_n$ and $\widehat{\theta}_n$ (= the discretized version) are $\sqrt{n}$−consistent and asymptotically normal:

$$\left.\begin{array}{l} \sqrt{n}(\widetilde{\theta}_n - \theta_0) \\ \sqrt{n}(\widehat{\theta}_n - \theta_0) \end{array}\right\} \to_{\text{law}} \mathcal{N}\left(0, W_{\theta_0}^{-1}\right)$$

where $\mathcal{N}(0, W_{\theta_0}^{-1})$ denotes the Gaussian vector with zero mean and covariance matrix $W_{\theta_0}^{-1}$

The $p \times p-$matrix $W_{\theta_0}$ has entries

$$w_{\theta_0}(k,j) = \frac{1}{4\pi} \int_{\Pi} \frac{\partial}{\partial \theta_k} \log k_{\theta_0}(\lambda) \frac{\partial}{\partial \theta_j} \log k_{\theta_0}(\lambda) \mathrm{d}\lambda$$

Comments:

• scale parameter $\sigma^2$ can be consistently estimated by either $\widetilde{Q}_n(\widetilde{\theta}_n)$ or $\widehat{Q}_n(\widehat{\theta}_n)$:

$$\widetilde{Q}_n(\widetilde{\theta}_n) \to \sigma^2, \qquad \widehat{Q}_n(\widehat{\theta}_n) \to \sigma^2, \qquad a.s.$$

• Under conditions of Thm 2 (Gaussian set-up), Whittle's estimates $\widetilde{\theta}_n$ and $\widehat{\theta}_n$ are asymptotically efficient (have the smallest asymptotic variance)

• The fact that $\{X_t\}$ has long memory does not affect neither the consistency rate $\sqrt{n}$ nor the limit distribution of Whittle's estimates (the limit covariance matrix $W_{\theta_0}$ is the same as in short memory case)

• to construct confidence intervals, the unknown covariance matrix $W_{\theta_0}$ can be replaced by

$$\widehat{w}(i,k) = \frac{1}{n}\sum_{j=1}^{n} \frac{\partial}{\partial\theta_i} \log k_{\widehat{\theta}_n}(\lambda_j) \frac{\partial}{\partial\theta_k} \log k_{\widehat{\theta}_n}(\lambda_j)$$

• Giraitis and Surgailis (1990) extended the result to LM moving average time series $\{X_t\}$ with $\{\zeta_j\} \sim IID(0,\sigma^2)$. In this case, Whittle's estimator is also $\sqrt{n}$ consistent and asymptotically normal, but the limit matrix $W_{\theta_0}$ contains an additional term

• The proof of the CLT for Whittle's estimator is based on a CLT for Toeplitz quadratic forms

$$Q_n = \sum_{t,s=1}^{n} q_{t-s} X_t X_s$$

where $\{X_t\}$ is a LM process as above and $q_t = (1/2\pi)\int_{\Pi} \mathrm{e}^{\mathrm{i}t\lambda} k_\theta^{-1}(\lambda)\mathrm{d}\lambda$ are coefficients with the property

$$\begin{aligned} \sum_{|t|\le n} q_t &\to (1/2\pi)\int_{\Pi} \Big(\sum_{t\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}t\lambda}\Big) k_\theta^{-1}(\lambda)\mathrm{d}\lambda \\ &= k_\theta^{-1}(0) = 0 \qquad (5) \end{aligned}$$

since $k_\theta^{-1}$ is proportional to $f^{-1}$ and $f(0) = \infty$ under LM

- Condition (5) compensates for LM in $\{X_t\}$ and helps to prove the CLT for $Q_n$ with normalization $n^{1/2}$

- Fox and Taqqu (1995) prove asymptotical normality of $n^{-1/2}Q_n$ by showing that $\text{cum}_k(n^{-1/2}Q_n) \to 0 \, \forall \, k \geq 3$

- For linear (non-gaussian) $\{X_t\}$, a different approach is needed

## 4.1c Example Whittle's estimation of ARFIMA$(0, d, 0)$ parameters

$X_1, \cdots, X_n$ is a sample from a stationary Gaussian ARFIMA$(0, d, 0)$, with fractional differencing parameter $0 < d < 1/2$

Spectral density:

$$f(\lambda) = \frac{\sigma^2}{2\pi |2 \sin(\lambda/2)|^{2d}}$$

ARFIMA$(0, d, 0)$ is completely specified by two parameters: scale parameter $\sigma > 0$ and long memory parameter $0 < d < 1/2$

In particularly,

$$f(\lambda) \sim \frac{\sigma^2}{2\pi|\lambda|^{2d}} \qquad (\lambda \to 0)$$

In this case $\theta$ is the one dimensional parameter $\theta = d$ and the parameter set $\Theta = [0, 1/2]$ contains all possible values of $d$. We assume that the true parameter $d_0$ is an inner point of $\Theta$, i.e.

$$0 < d_0 < 1/2$$

Then

$$f(\lambda) = \sigma^2 k_d(\lambda), \quad k_d(\lambda) = \frac{1}{2\pi|2\sin(\lambda/2)|^{2d}}$$

ARFIMA$(0, d, 0)$ satisfies "standard normalization "condition:

$$\int_{\Pi} \log k_d(\lambda) \mathrm{d}\lambda = 0 \quad \text{for all} \quad d \in \Theta = [0, 1/2]$$

The (discrete) Whittle's estimate $\widehat{d_n}$ of $d$ minimizes the objective function $\widehat{\sigma}_n(d)$ in the interval $d \in [0, 1/2]$:

$$\widehat{d_n} = \mathrm{argmin}_{d \in [0,1/2]} \widehat{\sigma}_n(d),$$

where

$$\widehat{\sigma}_n(d) \;=\; \frac{2\pi}{n}\sum_{j=1}^{n}\frac{I_n(\lambda_j)}{k_d(\lambda_j)} \;=\; \frac{4\pi^2}{n}\sum_{j=1}^{n}|2\sin(\lambda_j/2)|^{2d}I_n(\lambda_j)$$

By Theorem 2,

$$\sqrt{n}(\widehat{d}_n - d_0) \;\to_{\text{law}}\; \mathcal{N}(0, w_{d_0}^{-1}) \qquad (n \to \infty)$$

where $w_{d_0}$ is the scalar ($1 \times 1-$matrix):

$$\begin{aligned} w_{d_0} &= \frac{1}{4\pi}\int_{\Pi}\frac{\partial}{\partial d}\log k_{d_0}(\lambda)\frac{\partial}{\partial d}\log k_{d_0}(\lambda)\mathrm{d}\lambda \\ &= \frac{1}{4\pi}\int_{\Pi}(\log|2\sin(\lambda/2)|)^2\mathrm{d}\lambda \end{aligned}$$

as $\log k_d(\lambda) = -\log(2\pi) - 2d\log|2\sin(\lambda/2)|$ is a linear function in $d$

Hence

$$\sqrt{n}(\widehat{d}_n - d_0) \;\to_{\text{law}}\; \mathcal{N}(0, s^2), \qquad s^2 = \frac{1}{4\pi}\int_{\Pi}(\log|2\sin(\lambda/2)|)^2\mathrm{d}\lambda$$

The scale parameter $\sigma^2$ can be estimated consistently by

$$\widehat{\sigma}_n(\widehat{d}_n) \;=\; \frac{4\pi^2}{n}\sum_{j=1}^{n}|2\sin(\lambda_j/2)|^{2\widehat{d}_n}I_n(\lambda_j)$$

• the asymptotic variance $s^2$ does *not* depend on unknown parameters $d$ and $\sigma$

## II. Semiparametric estimation

In semi-parametric set-up, the full parametric model of spectral density

$$f(\lambda) = |\lambda|^{-2d} g(\lambda)$$

of observations $X_1, \cdots, X_n$, is not specified (the function $g(\lambda)$ is not specified)

It is only assumed that

$$g(\lambda) \;\to\; g(0) \;=\; c > 0 \qquad (\lambda \to 0)$$

so that

$$f(\lambda) \;\sim\; c|\lambda|^{-2d} \qquad (\lambda \to 0) \tag{6}$$

• The parameters of interest are the memory parameter $d \in (-1/2, 1/2)$ and the "scale" parameter $c > 0$

• These parameters specify the behavior of spectral density at $\lambda = 0$ (low frequencies) only

- Many parametric models (ARFIMA$(p, d, q)$, FEXP$(p, d)$, fBN) satisfy (6)

- For these models, the function $g(\lambda)$ is fully specified up to unknown finite-dimensional parameter and efficient parametric Whittle's estimate can be applied to estimate all unknown parameters including the memory parameter $d$

- When $f(\lambda)$ and $g(\lambda)$ are correctly parameterized, the unknown parameter $d$ and other parameters can be precisely estimated, with rate $\sqrt{n}$

- However, if the model is misspecified (e.g. in ARFIMA$(p, d, q)$ the orders $p, q$ are misspecified), parametric Whittle's estimates can be even *inconsistent*

- Semiparametric estimates can consistently estimate unknown parameters $d$ and $c_f$ when $f(\lambda)$ and $g(\lambda)$ are unspecified away from zero frequency

- Semiparametric estimates can be also applied to parametric models to estimate $d$, with some loss of efficiency

## 4.2a The Local Whittle semiparametric estimator (LWE)

• To define LWE, first consider the approximate log-likelihood of a Gaussian process with spectral density $f$:

$$-\frac{1}{2\pi}\int_{\Pi}\Big(\log f(\lambda)+\frac{I_n(\lambda)}{f(\lambda)}\Big)\mathrm{d}\lambda$$

• The above approximation of the quadratic form in the exponent of Gaussian density by $\int_{\Pi}\frac{I_n(\lambda)}{f(\lambda)}\mathrm{d}\lambda$ is similar to Whittle's estimate

• Since the behavior of $f$ near $\lambda = 0$ is only important [we want to estimate the asymptotic parameters $c$ and $d$ only], the above approximate log-likelihood is further simplified by:

- restricting the integral to low frequencies $|\lambda| < \lambda_m := \frac{2\pi m}{n}$, $m \to \infty,\ m/n \to 0$

- replacing $f(\lambda)$ by $c|\lambda|^{-2d}$

- replacing integration by summation over $|\lambda_j| < \lambda_m$

- The resulting approximate log-likelihood is

$$R(c,d) \quad := \quad \frac{1}{m}\sum_{j=1}^{m}\left\{\log(c\lambda_j^{-2d}) + \frac{I_n(\lambda_j)}{c\lambda_j^{-2d}}\right\} \qquad (7)$$

- The Local Whittle estimate minimizes the above *objective function*:

$$(\widehat{c_n}, \widehat{d_n}) := \operatorname{argmin}\Big\{R(c,d) : c \geq 0, d \in [-1/2, 1/2]\Big\}$$

- It is easy (by taking the derivative w.r.t. $c$) to see that for fixed $d$ the minimum of $R(c,d)$ is achieved by

$$c(d) := \frac{1}{m}\sum_{j=1}^{m}|\lambda_j|^{2d} I_n(\lambda_j)$$

- Replacing $c$ in $R(c,d)$ by $c(d)$ and $\lambda_j$ by $2\pi j/m$ yields

$$\begin{aligned} U(d) \quad &:= \quad R(c(d),d) \\ &= \quad \log\left(\frac{1}{m}\sum_{j=1}^{m} j^{2d} I_n(\lambda_j)\right) - \frac{2d}{m}\sum_{j=1}^{m}\log j. \end{aligned}$$

This yields

**Definition 4.1** The Local Whittle estimate of parameters $(c, d)$ is defined by

$$\begin{aligned}\widehat{d}_n &:= \operatorname{argmin}\{U(d) : d \in [-1/2, 1/2]\},\\ \widehat{c}_n &:= c(\widehat{d}_n).\end{aligned}$$

- LWE was proposed by Künsch (1987) and developed by Robinson (1995b)
- Recent developments and simplified exposition: GKS (2011)
- LWE uses only periodogram ordinates belonging to a small neighbourhood of the zero frequency $\lambda = 0$
- To achieve convergence rates, additional assumptions on the bandwidth $m$ are needed, e.g. $m = O(n^{4/5}/\log n)$
- Increasing $m$ reduces standard deviation but increases the bias of LW

• Rate optimal choice of $m$ depends on smoothness assumptions on $g(\lambda)$ in the representation $f(\lambda) = |\lambda|^{-2d} g(\lambda)$

• In practice the choice of $m$ can be difficult

• Because the periodogram $I_n(\lambda_j),\ 1 \le j < n$ is invariant to the shift in the mean at Fourier frequencies, the local Whittle objective function does not depend on unknown mean $\mu = \mathrm{E}X_t$

# Asymptotic results for LWE

• Consistency and rates under weak assumptions: Dalla et al. (2006), including nonlinear models

• CLT for LWE: Robinson (1995a) for linear models (moving averages with i.i.d. innovations)

Consider linear process

$$\begin{aligned} X_t &= (1-L)^{-d} Y_t, \quad |d| < 1/2, \quad Y_t = \sum_{j=0}^{\infty} a_j \zeta_{t-j}, \qquad (8) \\ \{\zeta_t\} &\sim \mathit{IID}(0,1), \quad \sum_{j=0}^{\infty} |a_j| < \infty, \quad \sum_{j=0}^{\infty} a_j \ne 0. \end{aligned}$$

Spectral density of $\{X_t\}$ in (8):

$$f(\lambda) \quad = \quad const. \frac{|A(\lambda)|^2}{|1 - \mathrm{e}^{-\mathrm{i}\lambda}|^{2d}}, \qquad A(\lambda) := \sum_{j=0}^{\infty} a_j \mathrm{e}^{-\mathrm{i}j\lambda}$$

- Satisfies $f(\lambda) = |\lambda|^{-2d} g(\lambda)$ with $g(\lambda) \sim const.|A(\lambda)|^2$ and $c = g(0) > 0$

The following result is due to Robinson (1995b) with some modifications by Dalla et al. (2006).

**Theorem 4.3** Let $\{X_j\}$ be a linear process of (8). Then the LWE $(\widehat{c}_n, \widehat{d}_n)$ are weakly consistent estimates of parameters $(c_0, d_0)$.

Assume in addition $\mathrm{E}|\zeta_0|^4 < \infty$ and that $g(\lambda) := |A(\lambda)|^2$ (= the spectral density of $\{Y_t\}$) satisfies the Hölder condition on with exponent $> 1/2$ and the following relation

$$g(\lambda) = c_0 + b|\lambda|^{\beta} + o(|\lambda|^{\beta}), \qquad \lambda \to 0, \qquad 0 < \beta \leq 2.$$

Then

$$\widehat{d}_n - d_0 \;=\; -\frac{1}{2}Z_m - \frac{1}{2}B_m + o(\frac{1}{m^{1/2}}) + o_p((\frac{m}{n})^{\beta}),$$

where

$$\begin{aligned} \sqrt{m}Z_m &\;\to_{\text{law}}\; \mathcal{N}(0,1), \\ B_m &\;=\; \lambda_m^{\beta}\frac{b}{c_0}\frac{\beta}{(1+\beta)^2} + o((\frac{m}{n})^{\beta}). \end{aligned}$$

- If the bandwidth $m$ increases not too fast: $m = o(n^{2\beta/(1+2\beta)})$ so that the bias $B_m = o(m^{-1/2})$, the asymptotic distribution of the LWE is standard normal:

$$2m^{1/2}\left(\widehat{d}_n - d_0\right) \;\to_{\text{law}}\; \mathcal{N}(0,1) \qquad (9)$$

- The semiparametric LWE of the long memory parameter has a nice limiting distribution: asymptotically normal with a limiting variance that is completely known

- A drawback of LWE (and other semiparametric estimators) is the fact that the *bandwidth* $m$ *must increase more slowly* *than* $n$ and therefore $\widehat{d_n}$ converges to $d_0$ more slowly than $\sqrt{n}$-consistent estimators based on a fully parametric model

- If $n$ is large (as in financial data analysis) then one can choose $m$ large enough to achieve acceptable precision. However, if the time series is of moderate length, $m$ may be too small enough for $\widehat{d_n}$ to achieve a good approximation of $d_0$

- choice of bandwidth $m$: Henry and Robinson (1996)

- Other semiparametric estimators: log-periodogram estimator (Geweke and Porter-Hudak, 1983), (Robinson, 1995a), wavelet estimators (Abry et al., 1998), quadratic variations (Istas and Lang, 1994), Increment Ratio (Surgailis et al., 2008)

## 4.2b. Log-periodogram semiparametric estimator (Geweke and Porter-Hudak)

Similarly to the local Whittle estimate, the log-periodogram estimator is based on the very well-established statistical principle of ‘whitening’ the data and has particularly nice asymptotic statistical properties

These nice properties make the local Whittle and the log-periodogram estimators the most popular estimates of the long memory parameter

We again suppose that $X_1, \cdots, X_n$ is a sample from a stationary Gaussian sequence which satisfies the same assumptions as in Section 4.2a

*The idea:* the asymptotic relation

$$f(\lambda) \ \sim \ c|\lambda|^{-2d} \qquad (\lambda \to 0)$$

implies

$$I_n(\lambda_j) \ = \ f(\lambda_j)\frac{I_n(\lambda_j)}{f(\lambda_j)} \ \sim \ c|\lambda_j|^{-2d} r_j \qquad (\lambda_j \to 0)$$

or

$$\log I_n(\lambda_j) \ \sim \ \log c - 2d \log \lambda_j + \log r_j \qquad (\lambda_j \to 0) \qquad (10)$$

for Fourier frequencies $\lambda_j = 2\pi j/n \to 0$, where

$$r_j = \frac{I_n(\lambda_j)}{f(\lambda_j)}$$

In the case when the Gaussian process $\{X_t\}$ has *short memory* under mild regularity conditions the *whitening principle* says that the ratios $r_j, j = 1, 2, \cdots$ are asymptotically i.i.d. and therefore the $\log r_j$'s can be regarded as "errors" in the "linear regression model" (10)

The asymptotic i.i.d. property of $\log r_j$'s is not true in the long memory case; nevertheless the idea of (10) is very useful

The log-periodogram estimator of Geweke and Porter-Hudak is defined as the least squares estimate of $d$ in the "linear regression model":

$$\log I_n(\lambda_j) = \log c - 2d \log \lambda_j + u_j, \quad j = 1, \cdots, m \tag{11}$$

- "errors" $u_j \approx \log r_j$
- $u_j$ have nonzero asymptotic mean $-\eta = -.5772...$ (Euler's

constant)

• "centered errors":

$$u_j^* = u_j + \eta$$

• (11) can be rewritten as the linear regression with "centered errors":

$$y_j \;=\; \alpha - 2dx_j + u_j^*, \qquad j = 1, \cdots, m \tag{12}$$

with

$$y_j := \log I_n(\lambda_j), \quad \alpha := \log c - \eta, \quad x_j := \log \lambda_j$$

• least squares' in (12):

$$\sum_{j=1}^{m} (y_j - \alpha + 2dx_j)^2 \to \quad \min$$

yielding

$$\begin{aligned} \hat{d} &= -\frac{\sum_{j=1}^{m} (x_j - \bar{x}) y_j}{2 \sum_{j=1}^{m} (x_j - \bar{x})^2} \\ \hat{\alpha} &= \bar{y} + 2\hat{d}\ \bar{x} \end{aligned}$$

• since $\lambda_j = 2\pi j/n$ so

$$x_j - \bar{x} = \log \lambda_j - \frac{1}{m}\sum_{j=1}^{m} \log \lambda_j = \log j - \frac{1}{m}\sum_{j=1}^{m} \log j$$

setting

$$\nu_j := \log j - \frac{1}{m}\sum_{j=1}^{m} \log j$$

we finally obtain

The log-periodogram estimator of $d$ is defined by:

$$\widehat{d} \;:=\; \frac{\sum_{j=1}^{m} \nu_j \log I_n(\lambda_j)}{2\sum_{j=1}^{m} \nu_j^2}$$

• as $\sum_{j=1}^{m} \nu_j^2 \;\sim\; m$ so

$$\widehat{d} \;\sim\; \frac{1}{2m}\sum_{j=1}^{m} \nu_j \log I_n(\lambda_j)$$

*Assumption (f)* (on the spectral density):

$$f(\lambda) \;=\; |\lambda|^{-2d} g(\lambda)$$

where $-1/2 < d < 1/2$ and the function $g(\lambda)$ satisfies

$$g(\lambda) \; = \; c + O(|\lambda|^2) \qquad (\lambda \to 0)$$

Moreover, $g(\lambda)$ is integrable:

$$\int_{-\pi}^{\pi} g(\lambda) \mathrm{d}\lambda < \infty$$

*Assumption* $(m)$ (on the bandwidth):

$$m \to \infty, \qquad m \leq n^{4/5} / \log n$$

• Robinson (1995a) showed (under Assumption $(f)$ and Assumption $(m)$) that the log-periodogram estimator is asymptotically normal:

$$m^{1/2}(\widehat{d} - d_0) \; \Rightarrow \; N\left(0, \frac{\pi^2}{24}\right) \tag{13}$$

• (12) is very important and simple to use, applies in the long memory case $0 < d < 1/2$ and well as in short memory case $d = 0$ and in the negative memory case$-1/2 < d < 0$

- the asymptotic variance $\pi^2/24$ in (12) is independent of $d$

- Velasco (1999) generalized (12) to non-Gaussian moving averages $\{X_t\}$ with iid innovations

## 4.2c Increment Ratio (IR) statistic

For a real-valued function $f = \{f(t), t \in [0,1]\}$, $\Delta_i^{p,n} f$ denotes the $p-$order increment of $f$ at $\frac{i}{n}$, $p = 1, 2, \cdots$, $i = 0, 1, \cdots, n-p$. Let

$$R^{p,n}(f) \quad := \quad \frac{1}{n-p} \sum_{k=0}^{n-p-1} \frac{|\Delta_k^{p,n} f + \Delta_{k+1}^{p,n} f|}{|\Delta_k^{p,n} f| + |\Delta_{k+1}^{p,n} f|}, \tag{14}$$

with the convention $\frac{0}{0} := 1$. In particular,

$$R^{1,n}(f) \;=\; \frac{1}{n-1} \sum_{k=0}^{n-2} \frac{\left| f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right) + f\left(\frac{k+2}{n}\right) - f\left(\frac{k+1}{n}\right) \right|}{\left| f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right) \right| + \left| f\left(\frac{k+2}{n}\right) - f\left(\frac{k+1}{n}\right) \right|}. \tag{15}$$

Note the ratio on the right-hand side of (14) is either 1 or less than 1 depending on whether the consecutive increments $\Delta_k^{p,n} f$ and $\Delta_{k+1}^{p,n} f$ have same signs or different signs; moreover, in the latter case, this ratio generally is small whenever the increments are similar in magnitude ("cancel each other"). Clearly, $0 \le R^{p,n}(f) \le 1$ for any $f, n, p$. Thus *if* $\lim R^{p,n}(f)$ *exists when* $n \to \infty$*, the quantity* $R^{p,n}(f)$ *can be used to estimate* *this limit which represents the "mean roughness of $f$" also called the $p-$th order IR-roughness of $f$* below. We show below that these definitions can be extended to sample paths of very general random processes, e.g. stationary processes, processes with stationary and nonstationary increments, Lévy processes, and even Ł$^q$-processes with $q < 1$.

Some properties:

- $f$ monotone $\Longrightarrow R^{1,n}(f) = 1$
- $f$ smooth (absolutely continuous) $\Longrightarrow R^{1,n}(f) \to 1$
- $R^{1,n}(f)$ smaller $\Longrightarrow f$ "more rough"

*Discrete time processes:* S., Teyssière, Vaičiulis (2008), Bružaitė and Vaičiulis (2008), Vaičiulis, M. (2009)

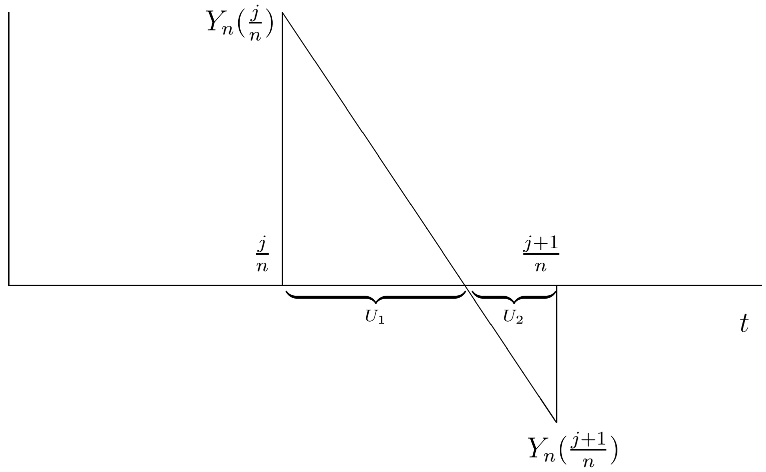
*Relation between the IR stat and zero crossings*

Let $Y_n(t)$, $t \in [0, 1-\frac{1}{n}]$ be the linear interpolation of the "differenced" sequence $\Delta_j^{1,n} X = X(\frac{j+1}{n}) - X(\frac{j}{n})$, $j = 0, 1, \cdots, n-1$:

$$Y_n(t) = n\Big[(\frac{j+1}{n} - t)\Delta_j^{1,n} X + (t - \frac{j}{n})\Delta_{j+1}^{1,n} X\Big], \qquad t \in [\frac{j}{n}, \frac{j+1}{n}),$$

$j = 0, 1, \cdots, n-2$. Then, using Figure **??** as a proof,

$$\begin{aligned} R^{1,n}(X) &= \frac{n}{n-1} \sum_{j=0}^{n-2} \Big| \text{meas}\Big\{ t \in [\frac{j}{n}, \frac{j+1}{n}) : Y_n(t) > 0 \Big\} \\ &\qquad - \text{meas}\Big\{ t \in [\frac{j}{n}, \frac{j+1}{n}) : Y_n(t) < 0 \Big\} \Big| \\ &= \frac{n}{n-1} \sum_{j=0}^{n-2} \Big| \int_{\frac{j}{n}}^{\frac{j+1}{n}} (\mathbf{1}(Y_n(t) > 0) - \mathbf{1}(Y_n(t) < 0)) \mathrm{d}t \Big| \end{aligned} \tag{1.6}$$



The proof of (16): follows by $\frac{|Y_n(\frac{j}{n})+Y_n(\frac{j+1}{n})|}{|Y_n(\frac{j}{n})|+|Y_n(\frac{j+1}{n})|} = n|U_1 - U_2|$.

- Let $\psi(x_1, x_2) := |x_1 + x_2|/(|x_1| + |x_2|)$, $\psi_0(x_1, x_2) := \mathbf{1}(x_1 x_2 \geq 0)$. Clearly, the two quantities $1 - \psi(Y_n(\frac{j}{n}), Y_n(\frac{j+1}{n}))$ and $1 - \psi_0(Y_n(\frac{j}{n}), Y_n(\frac{j+1}{n}))$ both are strictly positive if and only if $Y_n$ crosses the zero level in the interval $[\frac{j}{n}, \frac{j+1}{n})$ but the former quantity measures not only the fact but also the "depth" of the crossing so that $1 - \psi(Y_n(\frac{j}{n}), Y_n(\frac{j+1}{n}))$ attains its maximal value 1 in the case of a "perfect" crossing in the middle of the interval $[\frac{j}{n}, \frac{j+1}{n})$ (see Figure **??**).

## *Asymptotic results for the IR statistic*

*Consistency*

Following Dobrushin (1980), we say that $X = \{X_t,\, t \in \mathbb{R}\}$ has a *small scale limit* $Y^{(t_0)}$ *at point* $t_0 \in \mathbb{R}$ if there exist a normalization $A^{(t_0)}(\delta) \to \infty$ when $\delta \to 0$ and a random process $Y^{(t_0)} = \{Y_\tau^{(t_0)}, \tau \geq 0\}$ such that

$$A^{(t_0)}(\delta)\,(X_{t_0+\tau\delta} - X_{t_0}) \to_{FDD} Y_\tau^{(t_0)}. \tag{17}$$

• Related definitions: Falconer (2002, 2003): $Y^{(t_0)}$ a *tangent process* (at $t_0$); Benassi *et al.* (1997)

• In many cases, $A^{(t_0)}(\delta) = \delta^{H(t_0)}$, $0 < H(t_0) < 1$ and the limit tangent process $Y^{(t_0)}$ is self-similar with index $H(t_0)$ (Falconer, 2003 or Dobrushin, 1980)

• A trivial example: deterministic differentiable $X = f$, with $A^{(t_0)}(\delta) = 1/\delta$ and $Y_\tau^{(t_0)} = f'(t_0)\tau$ = tangent line

- If $X$ satisfies a similar condition to (17), then the statistic $R^{p,n}(X)$ converges to the integral

$$R^{p,n}(X) \to_p \int_0^1 \mathsf{E}\left[\frac{|\Delta_0^p Y^{(t)} + \Delta_1^p Y^{(t)}|}{|\Delta_0^p Y^{(t)}| + |\Delta_1^p Y^{(t)}|}\right] \mathrm{d}t, \qquad (18)$$

where $\Delta_j^p Y^{(t)} = \Delta_j^{p,1} Y^{(t)} = \sum_{i=0}^p (-1)^{p-i}\binom{p}{i} Y_{j+i}^{(t)}$, $j = 0, 1$ is the corresponding increment of the tangent process $Y^{(t)}$ at $t \in [0, 1)$

- In the particular case when $X$ has stationary increments, relation (18) becomes

$$R^{p,n}(X) \to_p \mathsf{E}\left[\frac{|\Delta_0^p Y + \Delta_1^p Y|}{|\Delta_0^p Y| + |\Delta_1^p Y|}\right].$$

*"Fractal" Gaussian processes*

A typical example: FBM $X = B_H$ $(0 < H < 1)$: self-tangent for any $t \in [0,1]$

$$R^{p,n}(X) \quad \to_{a.s.} \quad \Lambda_p(H) \qquad (p = 1, 2), \tag{19}$$

where

$$\begin{aligned} \Lambda_p(H) &:= \lambda(\rho_p(H)), \\ \lambda(r) &:= \frac{1}{\pi}\arccos(-r) + \frac{1}{\pi}\sqrt{\frac{1+r}{1-r}}\log\left(\frac{2}{1+r}\right), \\ \rho_p(H) &:= \mathrm{corr}(\Delta_0^p B_H, \Delta_1^p B_H), \end{aligned}$$

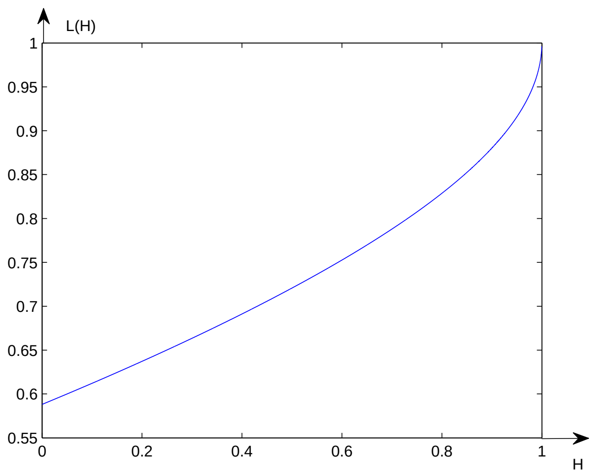
and where $\Delta_j^1 B_H = B_H(j+1) - B_H(j)$, $\Delta_j^2 B_H = B_H(j+2) - 2B_H(j+1) - B_H(j)$ $(j \in \mathbb{Z})$ are respective increments of FBM. Moreover,

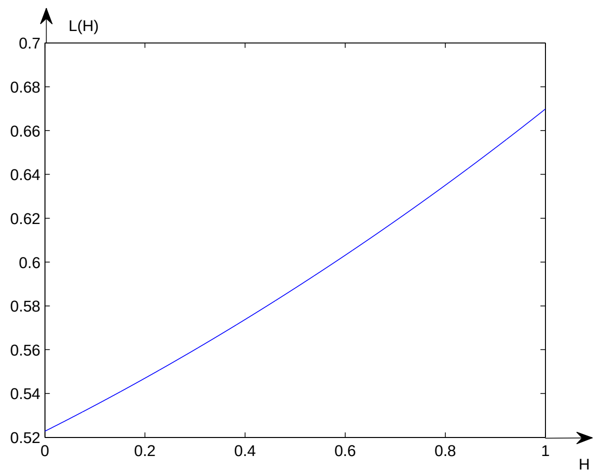
$$\sqrt{n}(R^{p,n}(X) - \Lambda_p(H)) \to_{\mathrm{law}} \mathcal{N}(0, \Sigma_p(H)) \quad \text{if} \quad \begin{cases} p = 1, & 0 < H < 3/4, \\ p = 2, & 0 < H < 1. \end{cases} \tag{20}$$

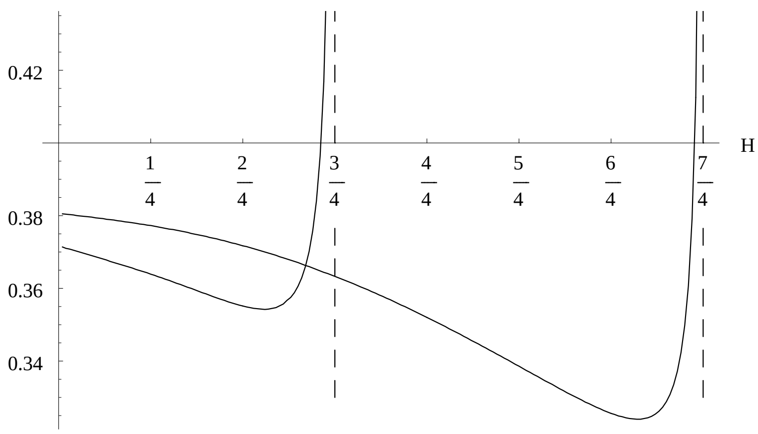
The asymptotic variances $\Sigma_p(H)$ in (20) are given by

$$\Sigma_p(H) := \sum_{j\in\mathbb{Z}} \mathrm{cov}\left(\frac{|\Delta_0^p B_H + \Delta_1^p B_H|}{|\Delta_0^p B_H| + |\Delta_1^p B_H|}, \frac{|\Delta_j^p B_H + \Delta_{j+1}^p B_H|}{|\Delta_j^p B_H| + |\Delta_{j+1}^p B_H|}\right).$$

The graphs of $\Lambda_p(H)$ and $\sqrt{\Sigma_p(H)}$ $(p = 1, 2)$ are given in Figures **??** and **??**.



- The difference in the range of the parameter $H$ for $p = 1$ and $p = 2$ in the CLT (20) are due to the fact that the second order increment process $(\Delta_j^2 B_H, j \in \mathbb{Z})$ is a short memory stationary Gaussian process for any $H \in (0,1)$, in contrast to the first order increment process $(\Delta_j^1 B_H, j \in \mathbb{Z})$ which has long memory for $H > 3/4$.

Generalizations of (19) and (20) to Gaussian processes having *nonstationary increments* are proposed in Section **??**. Roughly speaking, $R^{p,n}(X),\ p = 1, 2$ converge a.s. and satisfy a central limit theorem, provided for any $t \in [0,1]$ the process $X$ admits a FBM with parameter $H(t)$ as a tangent process (more precise assumptions **(A.1)**, **(A.1)**$'$ and **(A.2)**$_p$ are provided in Section **??**). In such frames, the limits in (19) are $\int_0^1 \Lambda_p(H(t))\mathrm{d}t$ instead of $\Lambda_p(H)$ and the asymptotic variances in (20) also change. The case of Gaussian processes with stationary increments is discussed in detail and the results are used to define a $\sqrt{n}-$consistent estimator of $H$, under semiparametric assumptions on the asymptotic behavior of the variogram or the spectral density.

*The main advantages of the IR statistic:*

- Robustness to additive and multiplicative trends

The estimator $R^{p,n}(X)$ essentially depends on local regularity of the process $X$ and not on possible "multiplicative and additive factors" such as diffusion and drift coefficients or smoothly multiplicative and additive trended Gaussian processes. This property is important when dealing with financial data involving heteroscedasticity and volatility clustering. Such a robustness property (also satisfied by the estimators based on generalized quadratic variations of wavelet coefficients) represents a clear advantage versus classical parametric Whittle or semi-parametric log-periodogram estimators.

- Applicability to various "fractional" processes (Lévy, diffusions, . . . )

- Computational simplicity

Does not require any bandwidth or tuning parameters such as the scales for estimators based on quadratic variations or wavelet coefficients.

In the Gaussian case, an estimator of $H$ based on $R^{2,n}(X)$ can be extremely simply computed:

$$\widehat{H}_n^{(2)} \simeq \frac{1}{0.1468}\Big(\frac{1}{n-2}\sum_{k=0}^{n-3} \frac{|X_{\frac{k+2}{n}} - 2X_{\frac{k+1}{n}} + X_{\frac{k}{n}} + X_{\frac{k+3}{n}} - 2X_{\frac{k+2}{n}} + X_{\frac{k+1}{n}}|}{|X_{\frac{k+2}{n}} - 2X_{\frac{k+1}{n}} + X_{\frac{k}{n}}| + |X_{\frac{k+3}{n}} - 2X_{\frac{k+2}{n}} + X_{\frac{k+1}{n}}|}.-0.$$

In the R language, if `X` is the vector $(X_{\frac{1}{n}}, X_{\frac{2}{n}}, \cdots, X_1)$,

$$\widehat{H}_n^{(2)} \simeq (\texttt{mean}(\texttt{abs}(\texttt{diff}(\texttt{diff}(\texttt{X}[-1])) + \texttt{diff}(\texttt{diff}(\texttt{X}[-\texttt{length}(\texttt{X})])))$$

$$/(\texttt{abs}(\texttt{diff}(\texttt{diff}(\texttt{X}[-1]))) + \texttt{abs}(\texttt{diff}(\texttt{diff}(\texttt{X}[-\texttt{length}(\texttt{X})]))))) - 0.517$$

## REFERENCES


J.-M. Bardet, D. Surgailis. Measuring the roughness of random paths by increment ratios. *Bernoulli* **17** (2011), 749–780.

R. Dahlhaus (1989) Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17**, 1749-1766

R. Fox and M.S. Taqqu (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14**, 517-532

J. Geweke and S. Porter-Hudak (1983) The estimation and application of long-memo time series models. *J.Time Series Anal.* **4**, 221-238

L. Giraitis and D. Surgailis (1990) A central limit theorem for quadratic forms in strongly dependent random variables and its application to asymptotical normality of Whittle's estimate. *Probab. Th. Rel. Fields* **86**, 87-104

E.J. Hannan (1973) The asymptotic theory of linear time series models. *J. Appl. Probab.* **10**, 130-145

H.E. Hurst (1951) Long-term store capacity of reservoirs. *trans. Amer. Math. Soc.* **4**, 770-808

H.E. Hurst (1955) Methods of using long-term storage in reservoirs. *Proc. of the Institution of Civil Engineering*, Part I, 519-577

H. Künsch (1987) Statistical aspects of self-similar processes, in: Yu. A. Prohorov and V. V. Sazonov (eds), *Proc. 1st World Congress of the Bernoulli Soc.*, vol.1, Science Press, Utrecht, 67–74

B.B. Mandelbrot and M.S. Taqqu (1979) Robust R/S of long run serial correlation. Proc. 42nd Session of the I. S.I., *Bullet. of the ISI* **48**, 69-104

B.B. Mandelbrot and J.W. van Ness (1968) Fractional Brownianmotions, fractional noise and applications. *SIAM Rev.* **10**, 422-437.

B.B. Mandelbrot and J.R. Wallis (1969a) Some long-run properties of geophysical records. *Water Resources Research* **5**, 321-340

B.B. Mandelbrot and J.R. Wallis (1969b) Robustness of rescaled range $R/S$ in the measurement of noncyclic the long-run statistical dependence. *Water Resources Research* **5**, 967-988

E. Moulines and P. Soulier (2003) Semiparametric spectral estimation for fractional proceeses. In: P. Doukhan, G. Oppenheim and M.S. Taqqu (eds), *Theory and Applications of Long-Range Dependence*, 251-302. Birkhäuser, Boston

P.M. Robinson (1995a) Log-periodogram regression of time series with long range dependence *Ann. Statist.* **23**, 1048-1072

P.M. Robinson (1995b) Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23**, 1630-1661

D. Surgailis, G. Teyssière, M. Vaičiulis (2008) The increment ratio statistic. *J. Multiv. Anal.* **99**, 510–541.

C. Velasco (1999) Non-stationary log-periodogram regression.*J. Econometrics* **91**, 325 - 371