The increment ratio test for long memory

Marijus Vaičiulis

Institute of Mathematics and Informatics, Vilnius, Lithuania

The increment ratio statistic

D.Surgailis, G.Teyssière and M.Vaičiulis

Journal of Multivariate Analysis, forthcoming

Definitions of a stationary process with long memory

Definition 1.(in the time domain terms) Let $\{X_t,\ t \in Z\}$ be a stationary process. If

$$\sum_{t=0}^{\infty} |\mathrm{Cov}(X_0, X_t)| = \infty, \tag{1}$$

then X_t is called a stationary process with long memory, or long range dependence, or strong dependence.

Definition 2.(in the spectral domain terms) Let $\{X_t,\ t \in Z\}$ be a stationary process. If its spectrum $f(\lambda)$ is unbounded in a close positive neighborhood of the zero frequency, then then X_t is called a stationary process with long memory.

Example of a stationary process with long memory

The Fractionally Integrated process(denoted as either FI(d) ir I(d)),defined as

$$(1-L)^d X_t = \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2) \tag{2}$$

where L stands for the lag operator, $LX_t = X_{t-1}$, the fractional difference operator $(1-L)^d$ associated with the degree of fractional integration $d \in (-1/2, 1/2)$ is defined as

$$(1-L)^d = \sum_{j=0}^{\infty} b_j L^j, \tag{3}$$

where

$$b_0 = 1, \ b_j \prod_{k=1}^{j} \left(1 - \frac{1+d}{k}\right). \tag{4}$$

The covariance:

$$Cov(X_0, X_t) \sim \frac{\Gamma(1-d)}{\Gamma(d)} |t|^{2d-1}, \quad |t| \to \infty. \tag{5}$$

The spectrum:

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \left| 2 \sin\left(\frac{\lambda}{2}\right) \right|^{-2d}. \tag{6}$$

Example of a stationary process with short memory

The process $\{X_t,\ t \in Z\}$ is said to be an $AR(1)$ process if X_t is stationary and if for every t,

$$X_t - aX_{t-1} = \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2), \quad |a| < 1. \tag{7}$$

The covariance:

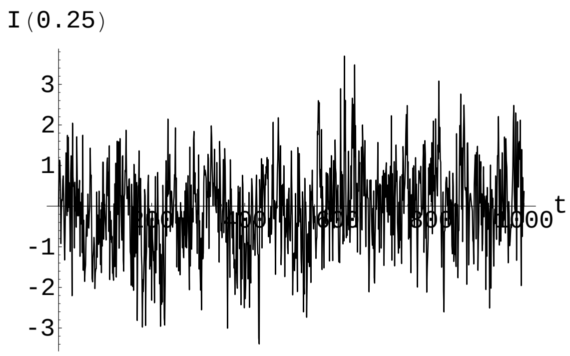
$$Cov(X_0, X_t) = \sigma_\varepsilon^2 \frac{a^{|t|}}{1 - a^2}, \quad t = 0, \pm 1, \pm 2, \ldots \tag{8}$$

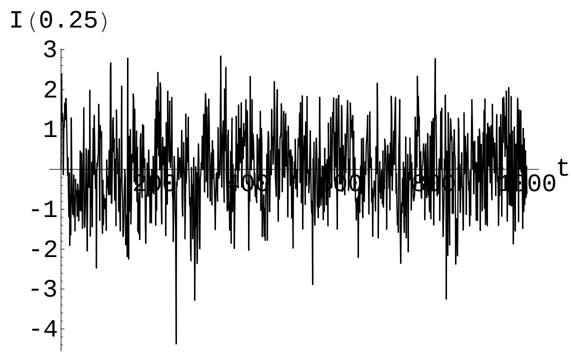
The spectrum:

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1 - a^2}{1 - 2a\cos(\lambda) + a^2}. \tag{9}$$

1000 observations of the series

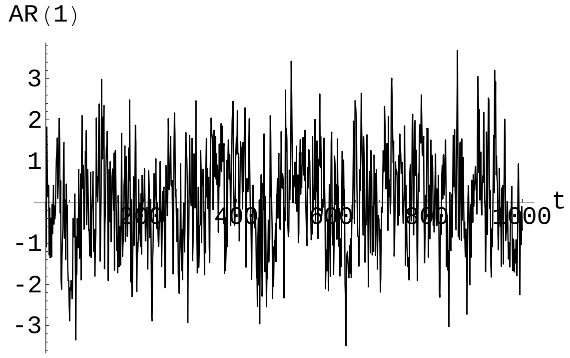
$$(1-L)^{1/4}X_t = \varepsilon_t, \quad \varepsilon_t \sim N(0,1)$$

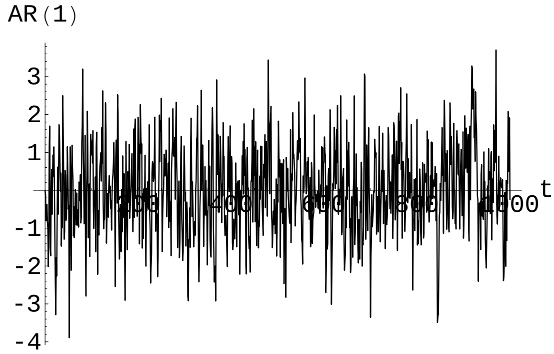




1000 observations of the series

$$X_t - 0.5X_{t-1} = \varepsilon_t, \quad \varepsilon_t \sim N(0,1)$$





The problem

Suppose that $X_1 X_2, \ldots, X_N$ are observations of some stationary process. How to test the hypothesis about unknown parameter d, e.g.

$$H_0 : d = d_0?$$

1. Score test (Robinson(1994));
2. Lagrange multiplier test (Lobato and Robinson(1998));
3. The modified R/S statistic (Lo(1991));
4. The KPSS statistic (Kwiatkowski et. al. (1992));
5. The V/S statistic (Giraitis et. al. (2003)).

The increment ratio statistic

$$IR = \frac{1}{N-3m} \sum_{k=0}^{N-3m-1} \frac{|V_m(k) + V_m(k+m)|}{|V_m(k)| + |V_m(k+m)|},$$

- $X_1, \ldots, X_N$ is a given sample of length N;
- $V_m(k) = \sum_{t=k+1}^{k+m} X_{t+m} - X_t$;
- $m = 1, 2, \ldots$ – bandwidth parameter;
- IR is location and scale free (i.e., it does not change if the X_ts are replaced by $aX_t + b$).

IR - sum of ratios of 2nd order increments of partial sums process $S_n(\tau) = \sum_{t=1}^{[n\tau]} X_t$:

$$IR = \frac{1}{(N/m)-3} \int_0^{(N/m)-3} \frac{\left| \Delta^2 S_{[m\tau]} + \Delta^2 S_{[m(\tau+1)]} \right|}{\left| \Delta^2 S_{[m\tau]} \right| + \left| \Delta^2 S_{[m(\tau+1)]} \right|} \, \mathrm{d}\tau,$$

where

$$\begin{aligned} \Delta f(\tau) &:= f(\tau+1) - f(\tau), \\ \Delta^2 f(\tau) &:= \Delta(\Delta f(\tau)), \end{aligned}$$

$[a]$ - integer part of a real number a

Asymptotic results

The limit function is defined as

$$\Lambda(d) := E\left[\frac{|Z_d(0)+Z_d(1)|}{|Z_d(0)|+|Z_d(1)|}\right],$$

where $(Z_d(0), Z_d(1))$ have a jointly Gaussian distribution, with zero mean, unit variances and the covariance

$$\begin{aligned}\rho(d) &:= \operatorname{cov}(Z_d(0), Z_d(1)) \\ &= \frac{-9^{d+.5}+4^{d+1.5}-7}{2(4-4^{d+.5})}.\end{aligned}$$

The process $Z_d(\tau)$ (2nd increment of FBM) is defined by

$$Z_d(\tau) := \frac{1}{\sqrt{|4-4^{d+.5}|}}\Delta^2 B_{d+.5}(\tau), \ -.5 < d < .5,$$

and, if $.5 < d < 1.5$,

$$Z_d(\tau) := \frac{\sqrt{2d(2d+1)}}{\sqrt{|4-4^{d+.5}|}}\int_0^1 \Delta B_{d-.5}(\tau+s)\mathrm{d}s.$$

$$\Lambda(d) := E\left[\frac{|Z_d(0)+Z_d(1)|}{|Z_d(0)|+|Z_d(1)|}\right],$$

In particular,

$$\begin{aligned}\Lambda(0) &= \frac{2}{\pi}\arctan\sqrt{\frac{1}{3}}+\frac{1}{\pi}\sqrt{\frac{1}{3}}\log 4 = .588\ldots, \\ \Lambda(1) &= \frac{2}{\pi}\arctan\sqrt{\frac{5}{3}}+\frac{1}{\pi}\sqrt{\frac{5}{3}}\log\left(\frac{8}{5}\right) = .774\ldots.\end{aligned}$$

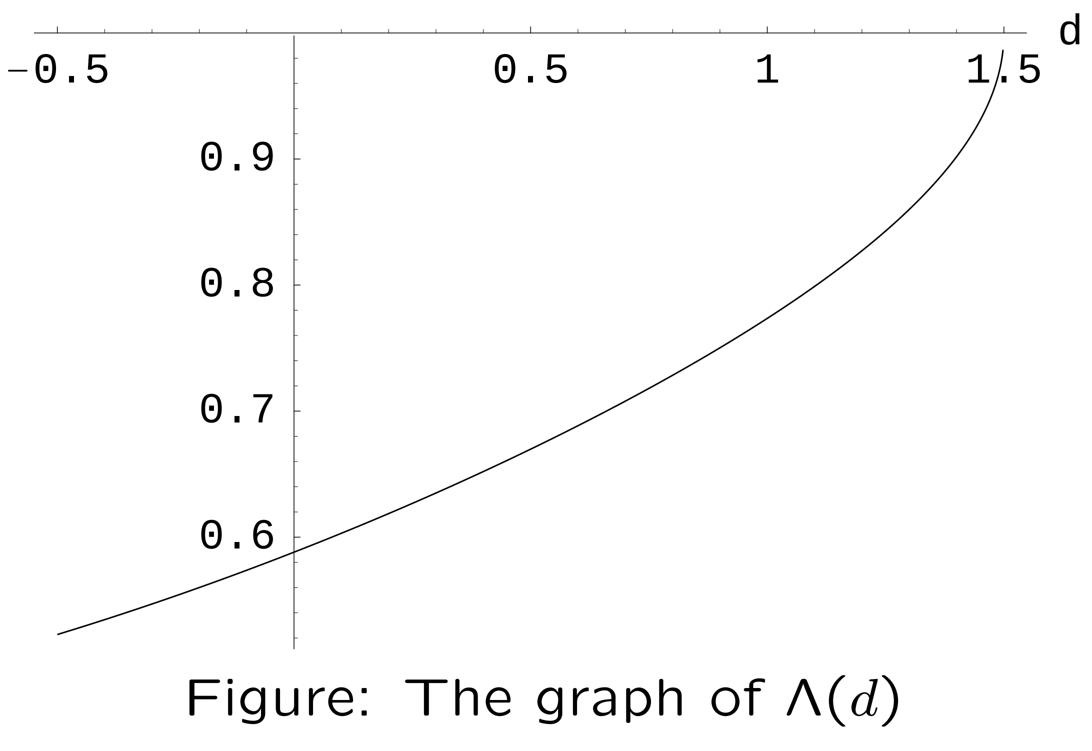


Figure: The graph of $\Lambda(d)$

Theorem 1

Assume either

(A1) $-1/2 < d < 1/2$ and "distant partial sums of X_t tend to independent FBM's with Hurst index $H = d + (1/2)$;

or

(A2) $-1/2 < d < 1/2$ and that "distant increments of X_t tend to independent FBM's with Hurst index $H = d - (1/2)$.

Then, as $N \to \infty$, $m \to \infty$ and $m/N \to \infty$,

$$E\mathrm{IR} \to \Lambda(d) \quad \text{and} \quad E(\mathrm{IR} - \Lambda(d)^2 \to 0.$$

Theorem 2("Bias")

Assume:

(B1) $-1/2 < d < 1/2$ and X_t is stationary Gaussian process having spectral density $f(\lambda)$ such that there exist constants $c_f > 0$, $0 < \beta < 2d + 1$ such that

$$f(\lambda) = |\lambda|^{-2d}\left(c_f + O(|x|^\beta\right),\ x \to 0;$$

(B2) $1/2 < d < 3/2$ and $U_t = X_t - X_{t-1}$ is a stationary Gaussian process with zero mean and spectral density $f(\lambda)$ such that there exist constants $c_f > 0$, $0 < \beta < 2d - 1$ such that

$$f(\lambda) = |\lambda|^{2-2d}\left(c_f + O(|x|^\beta\right),\ x \to 0;$$

Then $EIR - \Lambda(d) = O(m^{-\beta})$.

Theorem 3(" CLT")

Assume (B1)/(B2), $-1/2 < d < 5/4$, and

$$|f'(x)| \leq \begin{cases} |x|^{-1-2d}, & -1/2 < d < 1/2; \\ |x|^{1-2d}, & 1/2 < d < 5/4. \end{cases}$$

Then, as, $, m, N/m \to \infty$,

$$\frac{N}{m}\text{Var(IR)} \to \sigma^2(d),$$

and

$$\sqrt{\frac{N}{m}}(IR - E(IR)) \to N\left(0, \sigma^2(d)\right),$$

where

$$\begin{aligned} \sigma^2(d) &= 2\int_0^\infty \text{Cov}\,(\eta_0, \eta_\tau)\,\text{d}\tau, \\ \eta_\tau &= \frac{|Z_d(\tau) + Z_d(\tau+1)|}{|Z_d(\tau)| + |Z_d(\tau+1)|}. \end{aligned}$$

$$\begin{aligned}\sigma^2(d) &= 2\int_0^\infty \mathrm{Cov}\,(\eta_0, \eta_\tau)\, \mathrm{d}\tau, \\ \eta_\tau &= \frac{|Z_d(\tau) + Z_d(\tau+1)|}{|Z_d(\tau)| + |Z_d(\tau+1)|}.\end{aligned}$$

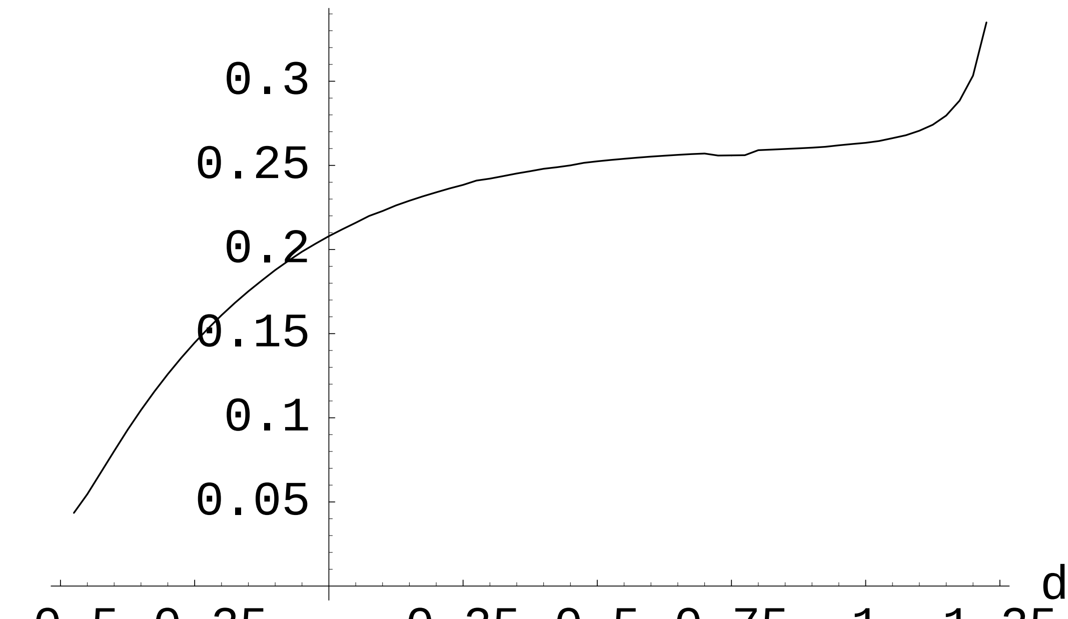


Figure: The graph of $\sigma(d)$

Testing the null hypothesis $d = 0$

The IR test: reject of short memory ($d = 0$) in favor of long memory ($d > 0$) whenever

$$IR - \Lambda(0) > z_\alpha \sigma(0) \sqrt{\frac{m}{N - 3m}}, \tag{10}$$

where $\Lambda(0) = 0.588$, $\sigma(d) = 0.208$, z_α - $N(0,1)$ quantile.

The V/S test: reject of short memory ($d = 0$) in favor of long memory ($d > 0$) whenever

$$V_N/\widehat{s^2_{N,q}} > c_\alpha N, \tag{11}$$

where

$$\begin{aligned} V_N &= \frac{1}{N}\Big[\sum_{k=1}^{N}\Big(\sum_{j=1}^{k}(X_j - \bar{X}_N)^2 \\ &- \frac{1}{N}\Big(\sum_{k=1}^{N}\sum_{j=1}^{k}(X_j - \bar{X}_N)\Big)^2\Big] \end{aligned}$$

is sample variance of the sums $S_k = \sum_{j=1}^{k}(X_j - \bar{X}_N)$,

$$\widehat{s^2_{N,q}} = \frac{1}{q_N + 1}\sum_{i,j=1}^{q_N+1}\widehat{\gamma}_{i-j}, \tag{12}$$

where the $\widehat{\gamma}_j$ being the sample covariances, $\bar{X}_N$ is the sample mean, q_N is a bandwidth parameter, c_α is a quantile of r.v. W, where

$$P(W \le x) = 1 + 2\sum_{k=1}^{\infty}(-1)^k \mathrm{e}^{-2k^2\pi^2 x}. \tag{13}$$

Stochastic and deterministic trends

The empirical sizes (probabilities of Type I error) of the tests IR test and V/S test are studied for short memory observations X_t of the form

$$
\begin{aligned}
X_t &= Y_t + f_{t,N}, \quad t = 1, \ldots, N, && (14) \\
Y_t &= aY_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid } \mathcal{N}(0,1), && (15) \\
f_{t,N} &= \sum_{i=1}^{t} b_{i,N} c_i, \quad c_i \sim \text{iid } \mathcal{N}(0, b^2) && (16)
\end{aligned}
$$

where $b_{i,N}$ are iid Bernoulli r.v., $P(b_{i,N} = 1) = \pi_N = 1 - P(b_{i,N} = 0)$.

Frequency of rejection of the null hypothesis of short memory for sequences of AR(1) + mixture trend processes, having on average 5 $\mathcal{N}(0, b^2)$-distributed jumps in a sample, ($\pi_N = 5/1000$). Test size 5%. $N = 1000$ (based on 10000 replications)

		V/S		IR	
a	b	$q = 10$	$q = 30$	$m = 10$	$m = 30$
0.0	0.0	0.0444	0.0363	0.0515	0.0465
0.0	0.2	0.6709	0.6062	0.0566	0.0914
0.0	1.0	0.9581	0.9103	0.1833	0.4669
0.2	0.0	0.0531	0.0387	0.0845	0.0560
0.2	0.2	0.5947	0.5286	0.0874	0.0848
0.2	1.0	0.9484	0.8969	0.2026	0.4120
0.4	0.0	0.0648	0.0417	0.1351	0.0679
0.4	0.2	0.4885	0.4125	0.1438	0.0834
0.4	1.0	0.9292	0.8724	0.2410	0.3394
0.6	0.0	0.0867	0.0472	0.2802	0.0885
0.6	0.2	0.3636	0.2679	0.2854	0.0946
0.6	1.0	0.8893	0.8146	0.3634	0.2689
0.8	0.0	0.1836	0.0680	0.8023	0.1483
0.8	0.2	0.2918	0.1422	0.7960	0.1492
0.8	1.0	0.7929	0.6545	0.8163	0.2231

The case of deterministic trends, with a possible break at time $t = [\delta N]$

$$X_t \;=\; c_0 t + c_1 I_{\{t>[\delta N]\}}(t - [\delta N]) + \varepsilon_t, \quad (17)$$

where $\delta \in (0,1)$, $\varepsilon_t \sim$ iid $\mathcal{N}(0,1)$. We set $\delta = 0.5$, i.e., the break in the trend occurs in the middle of the sample.

Frequency of rejection of the null hypothesis of short memory for sequences of iid process with a deterministic linear trend and a possible break. Test size 5%. $N = 1000$ (based on 10000 replications)

		V/S		IR	
c_0	c_1	$q = 10$	$q = 30$	$m = 10$	$m = 30$
10^{-3}	0.0	0.044	0.036	0.055	0.065
10^{-3}	0.002	1.000	1.000	0.058	0.204

Robustness to memory breaks

Consider the so-called "FARIMA $(0, d, 0)$ with memory breaks" model, defined by

$$X_t = \varepsilon_t + \sum_{j=1}^{\infty} \varepsilon_{t-j}\psi(j) \prod_{i=1}^{j} (1 - b_{t-i,N}), \quad (18)$$

where $\varepsilon_t \sim$ iid $\mathcal{N}(0, 1)$, $\psi(j)$ are the FARIMA$(0, d, 0)$ coefficients, and $b_{t,N}$ are iid Bernoulli.

Frequency of rejection of the null hypothesis of short memory for sequences of FARIMA$(0, d, 0)$ with memory breaks processes, with the average distance 333.3 between breaks ($\pi_N = 15/5000$). Test size 5%. $N = 5000$ (based on 10000 replications)

d	$q = 10$	$q = 30$	$m = 10$	$m = 30$
0.40	0.9775	0.8329	1.0000	0.9753
0.30	0.8946	0.6692	0.9973	0.8602
0.20	0.6564	0.4363	0.9177	0.5826
0.10	0.3017	0.2069	0.4678	0.2473

Application to financial times series

We consider three series of daily squared returns $X_{1,t}^2$, $X_{2,t}^2$, $X_{3,t}^2$, where

$$X_{i,t} = 100 \times \log(P_{i,t}/P_{i,t-1}),$$

where $P_{i,t}$ are shares on Bank of America (BoA), Oracle, and SAP, observed between April 1999 and April 2002, $N = 752$.

	V/S		IR	
Series	q	P–values	m	P–values
BoA	10	0.0002	10	0.0121
	20	0.0019	20	0.1229
	30	0.0063	30	0.3441
	$[N^{1/2}]$	0.0048	$[N^{1/2}]$	0.3612
	$[N^{1/3}]$	0.0001	$[N^{1/3}]$	0.0027
Oracle	10	0.0061	10	0.0652
	20	0.0202	20	0.0033
	30	0.0402	30	0.1895
	$[N^{1/2}]$	0.0335	$[N^{1/2}]$	0.1527
	$[N^{1/3}]$	0.0053	$[N^{1/3}]$	0.2639
SAP	10	0.0073	10	0.3774
	20	0.0212	20	0.0350
	30	0.0380	30	0.5163
	$[N^{1/2}]$	0.0344	$[N^{1/2}]$	0.5142
	$[N^{1/3}]$	0.0062	$[N^{1/3}]$	0.0660