Journal of Applied Statistics
Vol. 36, No. 9, September 2009, 945–972

Taylor & Francis
Taylor & Francis Group

Comparison of non-parametric and semi-parametric tests in detecting long memory

Mohamed Boutahar*

Luminy Faculty of Sciences, Departement of Mathematics, University of Méditerranée, Marseille, France

(*Received 10 December 2007; final version received 14 October 2008*)

The first two stages in modelling times series are hypothesis testing and estimation. For long memory time series, the second stage was studied in the paper published in [M. Boutahar *et al.*, *Estimation methods of the long memory parameter: monte Carlo analysis and application*, J. Appl. Statist. 34(3), pp. 261–301.] in which we have presented some estimation methods of the long memory parameter. The present paper is intended for the first stage, and hence completes the former, by exploring some tests for detecting long memory in time series. We consider two kinds of tests: the non-parametric class and the semi-parametric one. We precise the limiting distribution of the non-parametric tests under the null of short memory and we show that they are consistent against the alternative of long memory. We perform also some Monte Carlo simulations to analyse the size distortion and the power of all proposed tests. We conclude that for large sample size, the two classes are equivalent but for small sample size the non-parametric class is better than the semi-parametric one.

Keywords: hypothesis testing; long memory; power; short memory; size

1. Introduction

The last two decades of statistical research has resulted in a vast array of important contributions in the area of long memory modelling, from both a theoretical and an empirical perspective. Interested readers for this concept are referred to the books of Beran [3], Deniau *et al.* [12], Doukhan *et al.* [14] and Robinson [39]. This statistical tool has been applied in many areas: In hydrology [25,27], finance [33,34,52], macroeconomic [22], geophysics [4,20], telecommunication [44], demography [17] and in psychology [47].

*Email: boutahar@univmed.fr

ISSN 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/02664760802562464
http://www.informaworld.com

The first step in statistical modelling is hypothesis testing. One can naturally ask the following questions:

- Whether the studied series has really a long memory?
- Which statistical test one can use to decide the presence of such persistence?

The second step in statistical modelling is the estimation. In this step, the same problem arises: Which estimation method can be used ? This problem was studied in many papers in which the authors compare several estimation methods (see [6,43] among others).

In this paper, we propose the use of some new test statistics and try, by a Monte Carlo study, to shed light on the tests which are always used in empirical works. For instance, we will point out that the popular tests based on the Geweke and Porter-Hudak's [16] and Robinson's [38] statistics suffer from a size distortion for small sample size.

This paper is organized as follows. In Section 2, we recall some definitions of the long memory concept. In Section 3, we present some tests and propose the use of some new non-parametric tests. We establish the limiting distribution of the non-parametric tests under the null of short memory and show that they are consistent against the alternative of long memory. In Section 4, we perform a Monte Carlo study to compare all the tests introduced in Section 3. Section 5 concludes.

2. Definitions of long memory

Let (y_t) be a covariance stationary process with covariance function $\gamma(k)$ and spectral density $f(\lambda)$. The following are five common definitions of long memory:

- There exist $d \in (0, 1/2)$ and a constant $c_1 > 0$ such that: $\gamma(k) \sim c_1 k^{2d-1}$ as $k \to \infty$, (1)

where $a_k \sim b_k$ means that $a_k/b_k \to 1$ as $k \to \infty$.

- There exist $d \in (0, 1/2)$ and a constant $c_2 > 0$ such that: $f(\lambda) \sim c_2 |\lambda|^{-2d}$ as $\lambda \to 0$, (2)

(see [3]).

- The covariance of (y_t) is not absolutely summable, i.e.: $\sum_{k\in\mathbb{Z}} |\gamma(k)| = +\infty$, (3)

(see [42]).

- The spectral density of (y_t) can be written as: $f(\lambda) = |\lambda|^{-2d} L(|\lambda|^{-1}), 0 < d < 1/2$, (4)

where L is a slowly varying function (i.e. for all $a > 0$, $L(at)/L(t) \to 1$ if $t \to +\infty$), bounded on every finite interval, (see [9]).

- The process (y_t) has the linear and causal representation: $y_t = \sum_{j=0}^{\infty} \psi_j u_{t-j}$, (5)

where $\psi_j \sim \delta j^{d-1}$ as $j \to \infty$, $\delta > 0$, $0 < d < 1/2$, $(u_j) \sim$ i.i.d.$(0, \sigma^2)$ such that $E(u_j^4) = \sigma^4(3 + \kappa) < \infty$, for some constant $\kappa \geq 0$, (see [26]).

The process (y_t) is called a short memory process if

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| < +\infty.$$

There is no evident link between the five definitions of long memory. Under some assumptions, (1) and (2) are equivalent with $c_2 = (c_1/\pi)\Gamma(2d)\sin(1/2 - d)\pi$, where Γ is the gamma function (see [42]). The definition (4) is a relaxed version of (2). The definition (3) is the weaker one, since (3) can be obtained from (1) or (2) or (4). Two long memory processes are frequently evoked:

- The fractional Gaussian noise (FGN), i.e. the stationary Gaussian process with mean 0 and covariance

$$\gamma(k) = \frac{\sigma^2}{2}\{|k+1|^{2d+1} - 2|k|^{2d+1} + |k-1|^{2d+1}\}. \tag{6}$$

The FGN is self-similar with the parameter $H = d + 1/2$, further properties of the FGN are discussed in Mandelbrot and Taqqu [35], it has been used in various domains: geophysical data [4,20], communication [32,34], see also the references therein. It satisfies (1) with $c_1 = \sigma^2 d(2d+1)$ (see [3, p. 52]).

- The Autoregressive Fractionality Integrated Moving Average (ARFIMA), (p, d, q) process

$$\phi(L)(1-L)^d y_t = \theta(L)u_t,$$

where $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$, $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$, $d \in \mathbb{R}$ is the memory parameter, L the backward shift operator $Ly_t = y_{t-1}$, u_t a white noise with mean 0 and variance σ^2, $(1-L)^d$ the fractional difference operator defined by the binomial series

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} L^j.$$

The spectral density of the ARFIMA process is given by

$$f(\lambda) = \frac{\sigma^2}{2\pi}|\theta e^{i\lambda})|^2 |\phi(e^{i\lambda})|^{-2} |1 - e^{i\lambda}|^{-2d}$$

and hence satisfies (2) with $c_2 = (\sigma^2/2\pi)|\theta(1)|^2|\phi(1)|^{-2}$. It can also be shown that the ARFIMA process satisfies (5) with $\psi_j \sim (|\theta(1)||\phi(1)|^{-1}/\Gamma(d)) j^{d-1}$ as $j \to \infty$ (see [26, p. 272]).

This class of processes was introduced by Granger and Joyeux [21] and Hosking [24] to extend the Autoregressive Integrated Moving Average (ARIMA) modelling of Box and Jenkins [8]. It has been used in macroeconomic [24] and [22], demography [17], psychology [47] and many other applications.

3. Short or long memory?

To test if a given time series has short memory there are several tests, which can be gathered into two classes: The non-parametric class of tests and the semi-parametric one. Before presenting such tests, we will precise the null and the alternative hypotheses of interest. For the null hypothesis, we focus on the following class of short memory processes:

H_0: The process (y_t) is given by

$$y_t = \sum_{j\in\mathbb{Z}} b_j u_{t-j}, \sum_{j\in\mathbb{Z}} |b_j| < \infty, \tag{7}$$

where $(u_j) \sim$ i.i.d.$(0, \sigma^2)$ such that $E(u_j^4) = \sigma^4(3+\kappa) < \infty$, for some constant $\kappa \geq 0$.

The alternative hypothesis is the following class of long memory processes:

H_1: The process (y_t) satisfies (5).

Our results remain true if we assume that the process (y_t) satisfies under the null the following Functional Central Limit Theorem (FCLT):

$$\frac{1}{\hat{\sigma}_y\sqrt{n}} \sum_{k=1}^{[nt]} y_k \Longrightarrow B(t), \tag{8}$$

where $\hat{\sigma}_y^2$ is any consistent estimator of the long run variance σ_y^2 which is defined as

$$\sigma_y^2 = \lim_{n\to\infty} \text{var}\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} y_k\right) \tag{9}$$

$[x]$ is the integer part of x, $X_n \Longrightarrow X$ denotes the weak convergence of a sequence of random elements X_n in D to a random element X in D, and $D = D[0, 1]$ is the space of random functions that are right-continuous and have left limits, endowed with the Skorohod topology, $B(t)$ is a Brownian motion.

The assumption (8) is weaker than (7) since there are many processes, not necessary linear, that satisfy the FCLT (8), with σ_y^2 instead of $\hat{\sigma}_y^2$ (see [18,23] among others). The main difficulty is to construct $\hat{\sigma}_y^2$ for such nonlinear processes. However, for linear processes (7), the long run variance (9) is equal to $\sum_{k\in\mathbb{Z}} \gamma(k) = 2\pi f(0)$, where γ is the covariance function and f the spectral density of y_t, and the convergence (8) holds with $\hat{\sigma}_y^2 = 2\pi \hat{f}(0)$ and $\hat{f}(0)$ is any consistent estimator of $f(0)$ (see [11]). Moreover, it is easy to build a consistent estimator of the spectral density for the class (7) of linear processes, in opposite of nonlinear processes for which the spectral theory is not developed enough.

If the null hypothesis of short memory is rejected then there are many sources to explain this rejection such as the nonlinearity or/and the non-stationarity. If the time series is linear,[1] and the non-stationary tests[2] fail to reject the stationarity, then one can consider the alternative hypothesis of long memory. Such alternative can be corroborated by looking at the graph of the autocorrelation function (slow decay) and plotting the periodogram (high peak at the zero frequency) of the time series under study.

We do not assume the normality of the time series neither under H_0, nor under H_1. Such assumption is strong and usually supposed in this context (see [31,33]).

3.1 *Non-parametric tests*

Consider the time series y_t, $t = 1, \ldots, n$, to be tested. To define the non-parametric test statistics let $\hat{\sigma}_y^2$ be a consistent estimator of the long run variance $\sigma_y^2 = \lim_{n\to\infty} \text{var}((1/\sqrt{n}) \sum_{k=1}^{n} y_k)$ and

$$S_k = \sum_{j=1}^{k} (y_j - \bar{y}_n), \quad \bar{y}_n = \frac{1}{n} \sum_{j=1}^{n} y_j. \tag{10}$$

We consider the following six test statistics:

- *The modified Hurst statistic* (Hurst [27])

$$mR/S = \frac{R(n)}{\hat{\sigma}_y}, \quad \text{where } R(n) = \max_{0<k<n} S_k - \min_{0<k<n} S_k. \tag{11}$$

- *The Kolmogorov statistic* (Kulperger and Lockhart [29])

$$K_n = \frac{1}{\hat{\sigma}_y} \max_{1<k<n} |S_k|. \tag{12}$$

- *The Anderson–Darling statistic* (Anderson and Darling [2])

$$\mathrm{AD}_n(g) = \frac{1}{n^2 \hat{\sigma}_y^2} \sum_{k=1}^{n} g^2 \left(\frac{k}{n+1} \right) S_k^2, \tag{13}$$

  where g is a weight function continuous on $(0, 1)$.
- *The Cramèr-von Mises statistic* (Anderson and Darling [2]). If $g(x) = 1$ for all x in (13), then the Anderson–Darling statistic is reduced to the one of Cramèr-von Mises:

$$\mathrm{CVM}_n = \frac{1}{n^2 \hat{\sigma}_y^2} \sum_{k=1}^{n} S_k^2. \tag{14}$$

- *The Linear statistic* (Kulperger and Lockhart [29])

$$L_n(h) = \frac{1}{n \hat{\sigma}_y} \sum_{k=1}^{n} h \left(\frac{k}{n+1} \right) S_k, \tag{15}$$

  where h is a weight function continuous on $(0, 1)$.
- *The Quadratic statistic*

$$Q_n(g, h) = \frac{1}{n^2 \hat{\sigma}_y^2} \left\{ \sum_{k=1}^{n} g^2 \left(\frac{k}{n+1} \right) S_k^2 - \frac{1}{n} \left(\sum_{k=1}^{n} h(\frac{k}{n+1}) S_k \right)^2 \right\}, \tag{16}$$

  where g and h are two weight functions continuous on $(0, 1)$, such that $h \le g$, i.e. $h(x) \le g(x)$ for any $x \in (0, 1)$.

Note that all the test statistics (11)–(16) are functional of $\hat{\sigma}_y$ and S_k. Consequently, we need to define a consistent estimator $\hat{\sigma}_y^2$ of σ_y^2. For linear processes we have $\sigma_y^2 = 2\pi f(0)$ and hence it is sufficient to estimate $f(0)$. Many consistent estimators are available in the literature. For example, one can consider the smoothed periodogram (see [7]). In this paper, we will use the following spectral estimator:

$$\hat{f}(0) = \frac{1}{2\pi} \sum_{|k| \le q} w \left(\frac{k}{q+1} \right) \hat{\gamma}(k), \quad q < n, \tag{17}$$

where $w(.)$ is a spectral window, i.e. a continuous function on $[0, 1]$ with $w(0) = 1$ and $|w(x)| \le 1$ for all x, $w(x) = 0$ for all $x \notin [0, 1]$, q is the truncation parameter, and

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} (y_{t+k} - \bar{y}_n)(y_t - \bar{y}_n).$$

Choice of the truncation parameter q: For the class of processes (7), the consistency of $\hat{f}(.)$ holds under the assumption: $(1/q) + (q/n) \to 0$ as $n \to \infty$. However, to obtain the same order for the bias and variance, we should make the following choice:

$$q = \left[Cn^{1/(2\delta+1)}\right], \tag{18}$$

where δ is the largest exponent for which (see [1, p. 533])

$$\lim_{x\to 0} \frac{1 - w(x)}{|x|^{\delta}} = K < \infty. \tag{19}$$

For example $\delta = 1$ for the Bartlett window $w(x) = 1 - |x|$, $\delta = 2$ for the Parzen window $w(x) = 1 - x^2$. The mean square error (MSE) $E(\hat{f}(0) - f(0))^2$ is of order $n^{-2\delta/(2\delta+1)}$. Hence the Parzen window is to be preferred than the Bartlett window. The choice of the constant C in (18) can be achieved by performing a data driven window (see Section 4).

If $\hat{\sigma}_y^2 = 2\pi \hat{f}(0)$, where $\hat{f}(0)$ is the spectral estimator (17), then the following two classical statistics can be obtained as a particular case of the mR/S statistic:

- If $q = 0$ then mR/S is reduced to the well known R/S statistic introduced by Hurst [27].
- If $w(x) = 1 - |x|$, the Bartlett window, then mR/S is reduced to the statistic proposed by Lo [33].

We shall denote it by $m1R/S$. Lo [33] proved the consistency of the test based on the statistic $m1R/S$ against the alternative of Gaussian long memory processes. We will show that this result remains true without assuming the normality assumption.

The statistics K_n, $\mathrm{AD}_n(g)$ and $L_n(h)$, with $\hat{\sigma}_y^2 = \hat{\gamma}(0)/2\pi$, have been used by Kulperger and Lockhart [29] to test independence in time series. In this paper we suggest the use of them to test the presence of long memory by using a more general estimator $\hat{\sigma}_y^2$ of the long run variance σ_y^2. We do this because $\hat{\gamma}(0)/2\pi$ is a non-consistent estimator of σ_y^2 unless the underlying process is a white noise. Taking $\hat{\sigma}_y^2 = \hat{f}(0)/2\pi$, where $\hat{f}(0)$ is given by (17) with $q \neq 0$, we obtain versions of K_n, $\mathrm{AD}_n(g)$ and $L_n(h)$ which are robust to weak dependence.

The Kwiatkowski, Phillips, Schmidt and Shin (KPSS) statistic, proposed by Kwiatkowski *et al.* [30] to test stationarity of time series against the alternative of unit root, can be obtained as a particular case of the Cramèr-von Mises statistic CVM_n if we choose the Bartlett window in (17). Lee and Schmidt [31] have proposed the use of the KPSS statistic to test the presence of long memory and shown its consistency against the alternative of Gaussian ARFIMA$(0, d, 0)$ process. We will show that the KPSS test is still consistent against the large class of long memory processes (5).

If $g(x) = h(x) = 1$ for all x in (16) and we consider the spectral estimator (17) by using the Bartlett window, then the quadratic statistic $Q_n(g, h)$ is reduced to the one proposed by Giraitis *et al.* [19], and denoted by V/S.

3.1.1 *Limiting distribution of non-parametric tests under the null of short memory*

PROPOSITION 1 *Under the null H_0, let $\hat{f}(0)$ be any arbitrary consistent estimator of $f(0)$ and consider the statistics* (11)–(16) *with $\hat{\sigma}_y^2 = 2\pi \hat{f}(0)$, then the following convergences hold*

$$\frac{mR/S}{\sqrt{n}} \xrightarrow{\mathcal{L}} V = \sup_{0\le t\le 1} B_0(t) - \inf_{0\le t\le 1} B_0(t), \tag{20}$$

$$\frac{K_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} K_0 = \sup_{0\le t\le 1} |B_0(t)|, \tag{21}$$

$$\mathrm{AD}_n(g) \xrightarrow{\mathcal{L}} \mathrm{AD}(g) = \int_0^1 (g(t)B_0(t))^2\, \mathrm{d}t, \tag{22}$$

$$L_n(h) \xrightarrow{\mathcal{L}} L(h) = \int_0^1 h(t)B_0(t)\,\mathrm{d}t, \tag{23}$$

$$Q_n(g,h) \xrightarrow{\mathcal{L}} Q(g,h) = \int_0^1 (g(t)B_0(t))^2\,\mathrm{d}t - \left(\int_0^1 h(t)B_0(t)\,\mathrm{d}t\right)^2, \tag{24}$$

where $\xrightarrow{\mathcal{L}}$ *denotes the convergence in distribution and* $B_0(t) = B(t) - tB(1)$ *is a Brownian bridge.*

Proof See Appendix. ■

The random variable $L(h)$ is Gaussian with zero mean and variance $\sigma^2 = \int_0^1 \int_0^1 h(t)h(s)(\min(t,s) - ts)\,\mathrm{d}t\,\mathrm{d}s$. The random variable AD$(g)$ has the same probability density as $\sum_{j=1}^{\infty} \lambda_j Z_j^2$, where $(Z_j) \sim$ i.i.d.$N(0,1)$ and λ_j are the eigenvalues of the following integral equation:

$$\lambda \int_0^1 g(s)g(t)(\min(t,s) - ts)f(s)\,\mathrm{d}s = f(t);$$

there are two choices frequently used for the function g:

- $g(t) = 1$, in this case, $\lambda_j = 1/\pi^2 j^2$ and the limiting distribution is known as the Cramèr-von Mises distribution.
- $g(t) = \{t(t-1)\}^{-1/2}$, in this case $\lambda_j = 1/(j(j+1))$ and the limiting distribution is the Anderson–Darling distribution.

The cumulative probability distributions of AD(g), with $g(t) = 1$ and $g(t) = \{t(t-1)\}^{-1/2}$, and those of V, K_0, $Q(1,1)$ have a known analytic expressions (see Appendix).

3.1.2 *Consistency of non-parametric tests against the alternative of long memory*

To establish the consistency for the non-parametric tests, we limit ourselves to the spectral estimator (17) with the Bartlett window as a particular choice for $w(.)$.

PROPOSITION 2 *Under the hypothesis* H_1, *consider the non-parametric tests* (11)–(16) *where* $\hat{\sigma}_y^2 = \hat{\sigma}_n^2(q) = \hat{\gamma}(0) + 2\sum_{j=1}^{q}(1 - j/(q+1))\hat{\gamma}(j)$, *and assume that* $q \to \infty$ *as* $n \to \infty$ *and*

$$\begin{cases} \dfrac{q}{n} \longrightarrow 0 & \text{if } \dfrac{1}{4} < d < \dfrac{1}{2} \\ \dfrac{q\log n}{n} \longrightarrow 0 & \text{if } d = \dfrac{1}{4} \\ \dfrac{q^{1-2d}}{n^{1/2}} \longrightarrow 0 & \text{if } 0 < d < \dfrac{1}{4}, \end{cases} \tag{25}$$

then we have

$$\frac{\hat{\sigma}_n^2(q)}{(q+1)^{2d}} \xrightarrow{P} C^2(d,\delta), \tag{26}$$

$$\left(\frac{q+1}{n}\right)^d \frac{m1R/S}{\sqrt{n}} \xrightarrow{\mathcal{L}} V_{d+1/2} = \sup_{0\le t\le 1} W_{d+1/2}(t) - \inf_{0\le t\le 1} W_{d+1/2}(t), \tag{27}$$

$$\left(\frac{q+1}{n}\right)^d \frac{K_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} K_d = \sup_{0\le t\le 1} |W_{d+1/2}(t)|, \tag{28}$$

$$\left(\frac{q+1}{n}\right)^{2d} \mathrm{AD}_n(g) \xrightarrow{\mathcal{L}} \mathrm{AD}(g,d) = \int_0^1 (g(t)W_{d+1/2}(t))^2\,\mathrm{d}t, \tag{29}$$

$$\left(\frac{q+1}{n}\right)^{d} L_n(h) \xrightarrow{\mathcal{L}} L(h,d) = \int_0^1 h(t)W_{d+1/2}(t)\,\mathrm{d}t, \tag{30}$$

$$\left(\frac{q+1}{n}\right)^{2d} Q_n(g,h) \xrightarrow{\mathcal{L}} Q(g,h,d) = \int_0^1 (g(t)W_{d+1/2}(t))^2\,\mathrm{d}t - \left(\int_0^1 h(t)W_{d+1/2}(t)\,\mathrm{d}t\right)^2, \tag{31}$$

where

$$C^2(d,\delta) = \frac{\sigma^2\delta^2\Gamma^2(d)}{(2d+1)\Gamma(2d+1)\sin((d-1/2)\pi)},$$

$W_{d+1/2}(t) = B_{d+1/2}(t) - tB_{d+1/2}(1)$ *is the fractional Brownian bridge.*

Proof see Appendix. ■

Remark It is not easy to verify the condition (25), since in practice the degree of integration d is unknown. However, there are many choices for q independent of d and satisfying (25), for example $q \sim C_1 n^\gamma$ with $0 < \gamma \le 1/2$ and for some strictly positive constant C_1.

Recall that the Lo [33] statistic is given by $m1R/S = R(n)/\hat{\sigma}_n(q)$. By assuming that the process (y_t) is Gaussian satisfying

$$\gamma(k) \sim k^{2d-1}L(k) \quad \text{as } k \to \infty \tag{32}$$

and by using the fact that in this case the following convergence holds

$$\frac{1}{n^{d+1/2}L(n)^{1/2}} \sum_{k=1}^{[nt]} y_k \Longrightarrow B_{d+1/2}(t), \tag{33}$$

Lo [33] has shown that

$$\frac{m1R/S}{\sqrt{n}} \xrightarrow{P} +\infty$$

which implies that the test based on the statistic $m1R/S/\sqrt{n}$ is consistent.

By assuming that (y_t) is an ARFIMA$(0,d,0)$ process, i.e. $(1-L)^d\ y_t = u_t$ where u_t is a Gaussian white noise, Lee and Schmidt [31] have shown that

$$\left(\frac{q}{n}\right)^{2d} \mathrm{KPSS} \xrightarrow{\mathcal{L}} \int_0^1 (W_{d+1/2}(t))^2\,\mathrm{d}t, \tag{34}$$

where the KPSS is given by (14) with $\hat{\sigma}_y^2 = 2\pi \hat{f}(0)$ and $\hat{f}(0)$ is computed from (17) with the Bartlett window ($w(x) = 1 - |x|$). This implies that the test based on the KPSS statistic is consistent against the alternative of fractionally integrated processes. However, to obtain the convergence

(34) the normality assumption is not needed (see the proof of proposition 2 in the Appendix). Indeed if we assume that (y_t) belongs to the class (5) then (34) follows from the Functional Non-Central Limit Theorem (FNCLT)

$$\frac{1}{n^{d+1/2}} \sum_{k=1}^{[nt]} y_k \Longrightarrow C(d, \delta) B_{d+1/2}(t). \tag{35}$$

The convergence (31) was established by Giraitis *et al.* [19] for the particular case $g(x) = h(x) = 1$ for all x, by assuming that the process is stationary satisfying (1) and (35), and imposing a condition on the cumulants of order 4 (see condition (3.7), p. 272). Unfortunately, such condition is very difficult to check and is not necessary to obtain the convergence (31), (see the proof of Proposition 2 in the Appendix).

Here we obtain a more general results than those of Lo [33] and Lee and Schmidt [31], since Proposition 2 implies that the tests based on the statistics $m1R/S/\sqrt{n}$, $K_n/\sqrt{n}$, $\mathrm{AD}_n(g)$, $L_n(h)$ and $Q_n(g, h)$ are consistent against the alternative H_1 of long memory.

3.2 *Semi-parametric tests*

The semi-parametric tests are based on semi-parametric estimators of the long memory parameter. When such estimators are used, the spectral density of the underlying process is always assumed to have the following form:

$$f(\lambda) = |1 - \mathrm{e}^{\mathrm{i}\lambda}|^{-2d} f^*(\lambda), \tag{36}$$

where f^* is a bounded spectral density. Then a test of long memory can be formulated as follows:

$$H_0\colon d = 0$$

against

$$H_1\colon d \neq 0.$$

To perform such a semi-parametric test, let $\hat{d}$ be an asymptotically normal estimator of d:

$$\frac{\hat{d} - d}{\sigma_n(d)} \xrightarrow{\mathcal{L}} N(0, 1), \tag{37}$$

for a given sequence $\sigma_n^2(d) > 0$.

The null hypothesis H_0: $d = 0$ is rejected at the significance level α if $|\hat{d}|/\sigma_n(d) > u_{1-\alpha/2}$, where $u_{1-\alpha/2}$ is the $1 - \alpha/2$ percentile of $N(0, 1)$.

In this paper we will be interested to the following three semi-parametric estimators.

The Geweke and Porter-Hudak estimator: It is given by

$$\hat{d}_{\mathrm{GPH}} = \frac{\sum_{j=1}^{m_n} (R_j - \overline{R}) \log[I_n(\lambda_j)]}{\sum_{j=1}^{m_n} (R_j - \overline{R})^2}, \quad \overline{R} = \frac{1}{m_n} \sum_{j=1}^{m_n} R_j, \tag{38}$$

where

$$I_n(\lambda) = \frac{1}{n} \left| \sum_{t=1}^{n} y_t \mathrm{e}^{\mathrm{i}t\lambda} \right|^2,$$

is the periodogram of (y_t), λ_j are the Fourier frequencies and $R_j = -\log\{4\sin^2(\lambda_j/2)\}$. Geweke and Porter-Hudak [16] showed that $\hat{d}_{\mathrm{GPH}}$ is asymptotically normal for $d < 0$ (i.e. the memory of the process is antipersistent) and the truncation parameter m_n is such that $1/m_n + m_n/n \to 0$ as $n \to \infty$, for example $m_n = n^\gamma$, $0 < \gamma < 1$. The sequence in Equation (37) is $\sigma_n^2(d) = \pi^2/[6\sum_{j=1}^{m_n}(R_j - \overline{R})^2]$.

Hurvich *et al.* [28] extended the Geweke and Porter-Hudak's result to the general case $|d| < 1/2$. Moreover, based on the behaviour of MSE, they recommended $m_n = \mathrm{O}(n^{4/5})$ as an optimal choice for the truncation parameter.

The modified Geweke and Porter-Hudak estimator: To reduce the bias of the estimator $\hat{d}_{\mathrm{GPH}}$, Robinson [37] proposed to discard the first l low frequencies

$$\hat{d}_{Rm} = \frac{\sum_{j=l+1}^{m_n}(R_j - \overline{R})\log\{I_n(\lambda_j)\}}{\sum_{j=l+1}^{m_n}(R_j - \overline{R})^2}, \quad \overline{R} = \frac{1}{m_n - l}\sum_{j=l+1}^{m_n} R_j, \quad 0 \le l < m_n. \tag{39}$$

Under the normality of (y_t) and some regularity conditions, he showed the asymptotic normality of $\hat{d}_{Rm}$ with $\sigma_n^2(d) = \pi^2/24m_n$. Velasco [49] has extended Robinson's [37] results to linear but non-Gaussian processes.

The Robinson estimator: The spectral density of (y_t) is assumed to have the form (36) with $-1/2 < d < 1/2$, then the Robinson estimator of d is given by

$$\hat{d}_{\mathrm{GSP}} = \arg\min_{d\in\Theta} R(d), \tag{40}$$

where $\Theta = [\Delta_1, \Delta_2]$, Δ_1 and Δ_2 are chosen arbitrary between $-1/2$ and $1/2$ such that $-1/2 < \Delta_1 < \Delta_2 < 1/2$, and

$$R(d) = \log \hat{G}(d) - 2d\frac{1}{m_n}\sum_{j=1}^{m_n}\log\lambda_j, \quad \hat{G}(d) = \frac{1}{m_n}\sum_{j=1}^{m_n}\lambda_j^{2d} I_n(\lambda_j).$$

Robinson [38] proved the asymptotic normality of $\hat{d}_{\mathrm{GSP}}$ by assuming that (y_t) is Gaussian, and the following condition for the truncation parameter m_n:

$$\frac{1}{m_n} + \frac{m_n^{2\beta+1}}{n^{2\beta}}(\log m_n)^2 \longrightarrow 0 \quad \text{if } n \longrightarrow \infty, \tag{41}$$

where β depends on the regularity of the spectral density of (y_t) through the relation:

$$f(\lambda) \sim G\lambda^{-2d}(1 + \mathrm{O}(\lambda^\beta)) \quad \text{if } \lambda \longrightarrow 0+, 0 < d < 1/2, 0 < \beta \le 2.$$

The convergence (37) holds with $\sigma_n^2(d) = 1/4m_n$.

4. Monte Carlo simulations

We use the *S-PLUS 6.0* software to carry out the Monte Carlo experiments. We study the size and the power of the eleven following tests:

The non-parametric tests based on the following statistics:

- R/S: the statistic (11) with $\hat{\sigma}_y^2 = \hat{\gamma}(0) = 1/n\sum_{t=1}^{n}(y_t - \bar{y}_n)^2$.
- $m1R/S$: the statistic (11) with $\hat{\sigma}_y^2 = 2\pi\hat{f}(0)$ and $\hat{f}(0)$ is computed from (17) with the Bartlett window ($w(x) = 1 - |x|$), and the truncation parameter q is data-dependent given by Equation (43) below.
- $m2R/S$: the statistic (11) with $\hat{\sigma}_y^2 = 2\pi\hat{f}(0)$ and $\hat{f}(0)$ is computed from (17) with the Parzen window ($w(x) = 1 - x^2$) and the truncation parameter q is data-dependent given by Equation (44) below.

For the following other non-parametric tests we use $\hat{\sigma}_y^2 = 2\pi \hat{f}(0)$, and $\hat{f}(0)$ is computed from (17) with the Parzen window ($w(x) = 1 - x^2$) and the truncation parameter q is data-dependent given by Equation (44) below.

- K_n: the Kolmogorov statistic (12).
- $\mathrm{CVM} = \mathrm{AD}_n(1)$: the Cramèr-von Mises statistic i.e. the statistic (13) with $g(t) = 1$.
- $\mathrm{AD}_n = \mathrm{AD}_n(g)$: the Anderson–Darling statistic (13) with $g(t) = \{t(t-1)\}^{-1/2}$.
- $L_n = L_n(1)$: the linear statistic (15) with $h(t) = 1$.
- $Q_n = Q_n(1, 1)$: the quadratic statistic with $g(t) = h(t) = 1$.

The semi-parametric tests based on the following statistics:

- $\mathrm{GPH} = \hat{d}_{\mathrm{GPH}}/\sigma_n(d)$: the Geweke–Porter-Hudak statistic, $\hat{d}_{\mathrm{GPH}}$ is given by (38) and

$$\sigma_n^2(d) = \left[\frac{\pi^2}{6\sum_{j=1}^{m_n}(R_j - \overline{R})^2}\right].$$

- $\mathrm{GPHT} = \hat{d}_{Rm}/\sigma_n(d)$: the truncated Geweke–Porter-Hudak statistic, $\hat{d}_{Rm}$ is given by (39) and $\sigma_n^2(d) = \pi^2/24m_n$.
- $R = \hat{d}_{\mathrm{GSP}}/\sigma_n(d)$: the Robinson statistic, $\hat{d}_{\mathrm{GSP}}$ is given by (40) and $\sigma_n^2(d) = 1/4m_n$.

4.1 *Choice of the truncation parameters*

All the non-parametric (resp.semi-parametric) tests, except R/S, depend on the truncation parameter q (resp. m_n). The optimal choice for q remains an open problem. For m_n, we use the adaptive procedure proposed by Hurvich and Deo [28] when we compute the statistic GPH. However, for the statistic R, no optimal choice is known in the literature and this problem is not yet resolved.

Non-parametric tests: A good estimator for $f(0)$ will give a test with a correct size. However, if we use the spectral estimator (17), this one depends on the truncation parameter q. If we limit ourselves to the class (7) of linear processes, a choice which makes a trade-off between the bias and the variance is given by (18), the constant C can be computed by minimizing the mean square error, MSE, of the estimator $\hat{f}(0)$. Indeed, from Anderson [1]:

$$\mathrm{MSE} = E(\hat{f}(0) - f(0))^2 \sim n^{-2\delta/(2\delta+1)}\left\{AC + \frac{1}{C^{2\delta}}B\right\},$$

$$A = 2f^2(0)\int_{-1}^{1} w^2(x)\,\mathrm{d}x, \quad B = K^2\left(\frac{1}{2\pi}\sum_{r\in\mathbb{Z}}|r|^{\delta}\gamma(r)\right)^2,$$

the constant K is given by (19). By minimizing the MSE with respect to C we get $C_{\mathrm{opt}} = (2\delta A/B)^{1/(2\delta+1)}$, and hence

$$q = \left[\left(\frac{2\delta A}{B}n\right)^{1/(2\delta+1)}\right],$$

A and B depend on the spectral density of the process (y_t). For example, if (y_t) is an AR(1):

$$y_t = ay_{t-1} + u_t, \tag{42}$$

then an optimal choice for q should be

$$q = \left[\left(\frac{7.5\hat{a}^2}{(1-\hat{a})^4} n \right)^{1/3} \right] \quad \text{(for the Bartlett window)}, \tag{43}$$

$$q = \left[\left(\frac{6\hat{a}^2}{(1-\hat{a}^2)^2} n \right)^{1/5} \right] \quad \text{(for the Parzen window)}, \tag{44}$$

where $\hat{a}$ is the least squares estimator of the parameter a in the model (42). However, such choices are not optimal since the dynamic of the time series is unknown and usually different from Equation (42). Nevertheless, for all non-parametric tests, except the $m1R/S$ test, we use the Parzen window, i.e. $\hat{\sigma}_y^2 = \tilde{\sigma}_n^2(q) = \hat{\gamma}(0) + 2\sum_{j=1}^{q}(1-(j/(q+1)^2))\hat{\gamma}(j)$, where q is data-dependent via the formula (44).

Semi-parametric tests: For the GPH statistic we use the data-driven truncation parameter suggested by Hurvich and Deo [28]

$$m_n = \left[\left(\frac{27}{128\pi^2} \hat{K}^{-2} n^4 \right)^{1/5} \right],$$

where $\hat{K}$ is the least squares estimator of the third parameter in regressing $\log\{I_n(\lambda_j)\}$ on $(1, \log|2\sin(\lambda_j/2)|, \lambda_j^2/2)$.

To compute the Robinson statistic R we take $m_n = [n^{1/2}]$, with this choice the condition (41) holds if the process (y_t) is assumed to be an ARFIMA.

4.2 *Size*

The R/S, $m1R/S$, $m2R/S$, K_n, CVM, AD_n and Q_n are one side tests, i.e. the null is rejected if the p-value is lesser than α at a significance level α. The other tests are two sides, i.e. the null is rejected if the p-value is lesser than $\alpha/2$ at a significance level α. To study the size of the tests, we consider two data generating processes (DGP):

- The AR(1):

$$y_t = \phi y_{t-1} + u_t, \quad (u_t) \sim \text{i.i.d. } N(0,1).$$

- The ARMA(1,1):

$$y_t - \phi y_{t-1} = u_t - \theta u_{t-1}, \quad (u_t) \sim \text{i.i.d.} N(0,1) \quad \text{with } \phi = 0.5, \theta = -3.$$

For the AR(1), the parameter ϕ takes two values, $\phi = 0.5$ (the process is weakly dependent), and $\phi = 0.9$ (the process is in the neighbourhood of the unit root, and hence nearly non-stationary). We carry out an experiment of 1000 replications, and we use three different sample sizes, $n = 100$, $n = 500$ and $n = 1000$.

The Analysis of the Simulation Results for an AR(1)

From Table 1, we observe that the size distortions are reasonable for the non-parametric tests except the R/S test. For the R/S test the distortion is severe, this was also pointed out by Lo [33] and Giraitis *et al.* [19]. For example, if $n = 500$ and $\alpha = 5\%$, then we obtain an empirical size equal to 66.7**%**, Lo (1991) reported the value 55.9**%** and Giraitis *et al.* [19] reported the value 66.58%. The size distortions of $m2R/S$ (the modified R/S test with Parzen window) are a bit greater than $m1R/S$ (the modified R/S with Bartlett window). For example if $n = 500$ and $\alpha = 5\%$, then the empirical size of $m1R/S$ is 2.8% whereas the empirical size of $m2R/S$ is 2%.

Table 1. Empirical test sizes (in %).

	Sample size n	$n = 100$			$n = 500$			$n = 1000$		
Test	Level α	1%	5%	10%	1%	5%	10%	1%	5%	10%
R/S		26.3	46.8	59.7	43.7	66.7	59.7	49.9	70.9	79.9
$m1R/S$		0	0.3	1.4	0.4	2.8	7.4	0.7	4.3	8.2
$m2R/S$		0	0.2	1	0.3	2	5.6	0.7	3.8	7.3
K_n		0.1	0.7	2.3	0.4	1.9	3.7	0.2	2	5.2
CVM		0	6.2	12.9	0.7	5	10.1	1.1	4.8	10.7
AD_n		0.8	3.9	10.8	0.3	3.6	9.5	0.7	3.3	10.5
L_n		0	6.8	13.5	0.8	5.2	11.3	1.1	5.3	10.8
Q_n		0.6	2.8	9.3	0.7	3.8	9.8	1.4	5.1	9.8
GPH		18.4	33.2	43.1	24.8	35.9	43.8	24	35.7	45.8
GPHT		60.4	71.3	75.6	42	54.3	62.4	38.6	51.4	59.4
R		11.7	29.1	36.8	6.4	14.5	23.7	4.9	12.1	17.8

Note: The table contains rejection frequencies of the null hypothesis of short memory using the 11 tests. Rejection frequencies are based on 1000 replications generated from the DGP: $y_t = 0.5y_{t-1} + u_t$, $u_t \sim$ i.i.d.$N(0, 1)$ where the nominal significance levels are 1% , 5% and 10%, the sample sizes are $n = 100$, $n = 500$ and $n = 1000$.

The K_n test is more conservative, i.e. the empirical size is always lesser that the nominal one. The empirical sizes of CVM, AD_n, L_n and Q_n are much closer to their nominal values. The size distortions are large for all semi-parametric tests. The size distortions of GPHT are greater than GPH, for example if $n = 1000$ and $\alpha = 10\%$, then the empirical size of GPHT is 59.4% whereas the empirical size of GPH is 45.8%. The R test has also a size distortion which decreases as the sample size n increases.

Remark For $m1R/S$ (the modified R/S with Bartlett window), the averages of q over the 1000 replications take values around 3,4 and 5 if the sample size n is equal to 100,500 and 1000, respectively. For the other test we use the Parzen window and the averages of q over the 1000 replications take values around 14,25 and 31 if the sample size n is equal to 100,500 and 1000, respectively.

To save space in the rest of the paper we present only the results of the following six tests[3]:

- The two non-parametric R/S and $m2R/S$ tests. The Lo [33] test $m1R/S$ has similar properties as $m2R/S$, and then is discarded.
- The two non-parametric L_n and Q_n tests. The first is a linear functional of the partial sums S_k given by Equation (10), the second is quadratic. The results of the K_n test are worse and hence are removed. The CVM, $AD_n(g)$ and Q_n statistics are functional of S_k^2, hence we suppress the results of CVM and $AD_n(g)$ which are similar to those obtained for the Q_n test.
- The two semi-parametric GPH and R tests. The GPHT test gives a worse results than GPH, and hence is removed.

From Table 2, we observe that if the parameter $\phi = 0.9$ (the process is near a unit root process which is non-stationary), then all the tests suffer from a large size distortion. For instance, among the 1000 replications, no one is retained as a short memory by the GPH test! The size distortions of the non-parametric tests, except the R/S test, are clearly lesser than the semi-parametric ones. For example, if $n = 500$ and $\alpha = 5\%$, then the empirical sizes of $m2R/S$, L_n and Q_n are, respectively, 16%, 17% and 27.6% whereas the sizes of GPH and R are 100% and 84.5%. By increasing the sample size the tests are not improved. These results are not surprising since the

Table 2. Empirical test sizes (in %).

Test	Sample size n	$n = 100$			$n = 500$			$n = 1000$		
	Level α	1%	5%	10%	1%	5%	10%	1%	5%	10%
R/S		95.4	98.3	99.1	99.8	100	100	100	100	100
$m2R/S$		0	0	0.9	3.8	16	27.6	7.2	20.7	31.2
L_n		0.5	23.7	35.1	6.3	17	24	5.1	14.9	25
Q_n		0.3	15	35	11.7	27.6	41.4	11.2	26.6	38.5
GPH		100	100	100	100	100	100	100	100	100
R		85.8	91.3	94	73.7	84.5	89.1	55	72	79.9

Note: The table contains rejection frequencies of the null hypothesis of short memory using the six tests. Rejection frequencies are based on 1000 replications generated from the DGP: $y_t = 0.9y_{t-1} + u_t$, $u_t \sim$ i.i.d.$N(0, 1)$ where the nominal significance levels are 1% , 5% and 10%, the sample sizes are $n = 100$, $n = 500$ and $n = 1000$.

limiting distributions established in the proposition 1 under the null of short memory no longer hold if $\phi = 1$.

The analysis of the simulation results for an ARMA(1,1).

By comparing Tables 1 and 3 (i.e. by adding a moving average component to the AR(1) process) we observe the following: the size distortion of the R/S increases, for example if $n = 500$ and $\alpha = 5\%$, then the empirical size in the AR(1) is 66.7% and increases to 81.7% in the ARMA(1,1). The size distortion of $m2R/S$, L_n and Q_n are almost the same for the two DGPs. The size distortion of GPH increases, for example if $n = 500$ and $\alpha = 5\%$, then the empirical size of GPH in the AR(1) is 35.9% and increases to 100% in the ARMA(1,1). The size distortions of R are almost the same for the two DGPs, for example if $n = 500$ and $\alpha = 5\%$, then the empirical size of R in the AR(1) is 14.5% and decreases to 13.9% in the ARMA(1,1).

We can point out that in the class of the semi-parametric tests, the GPH test is less robust to short memory than the R test.

4.3 *Power*

To study the power of all tests, we generate two long memory processes. The first one is an ARFIMA generated by using the function arima.fracdiff.sim in SPLUS 6.0. The second is a FGN by using Beran's [4] code.

The analysis of the simulation results for an ARFIMA(0,d,0).

Table 3. Empirical test sizes (in %).

Test	Sample size n	$n = 100$			$n = 500$			$n = 1000$		
	Level α	1%	5%	10%	1%	5%	10%	1%	5%	10%
R/S		38.8	59.9	71.8	61.2	81.7	88.7	65.5	83.9	91.8
$m2R/S$		0	0.6	1.7	0.6	4.1	7.9	0.3	4	8.5
L_n		1.2	6.3	12.7	1.4	6.1	12.3	1	5.4	10.2
Q_n		0.3	5.6	12.3	1.5	6.8	13.3	0.9	5.8	11.6
GPH		94.6	98.5	99.2	100	100	100	100	100	100
R		15.4	28.5	37	6.5	13.9	20.7	4.3	10.4	17.6

Note: The table contains rejection frequencies of the null hypothesis of short memory using the six tests. Rejection frequencies are based on 1000 replications generated from the DGP: $y_t = 0.5y_{t-1} + u_t + 3u_{t-1}$, $u_t \sim$ i.i.d. $N(0, 1)$ where the nominal significance levels are 1%, 5% and 10%, the sample sizes are $n = 100$, $n = 500$ and $n = 1000$.

The DGP is the ARFIMA(0,d,0):

$$(1 - L)^d y_t = u_t, \quad \text{where}(u_t) \sim \text{i.i.d. } N(0, 1), \tag{45}$$

we consider three values for d, $d = 0.1$, $d = 0.3$ and $d = 0.4$.

From Tables 4–6, we observe that for a given sample size n, the power of the tests are improved considerably as we increase the degree of integration d. For example, when $n = 500$ and $\alpha = 5\%$, the empirical power of $m2R/S$ for d equals to 0.1, 0.3 and 0.4 are 24.2%,52.1% and 72.4%, respectively.

We observe also that for d fixed, the rejection frequencies of the null increase with the sample size n. For example, if $d = 0.3$ (Table 5) and $\alpha = 10\%$, then the empirical power of $m2R/S$ is 10.6% if $n = 100$. Increasing n to 500 and to 1000 observations, the empirical power becomes respectively 63.5% and 79.9%. The empirical power of R is 55.1% if $n = 100$, increases to 81.1% if $n = 500$ and then increases to 90.6% if $n = 1000$.

The analysis of the simulation results for FGN

The DGP is the FGN given by (6) with $d = 0.4$ and $\sigma^2 = 1$.

From Tables 5 and 7, we note that the rejection frequencies of H_0 for FGN with $d = 0.4$ are greater than those obtained for the ARFIMA (0, 0.3, 0). For example if $n = 500$ and $\alpha = 5\%$, then the empirical power of $m2R/S$ is 52.1% if the DGP is an ARFIMA (0, 0.3, 0) and increases to 67.4% if the DGP is a FGN. This can be explained by the behaviour of the spectral densities of the two processes in the neighbourhood of 0. The spectral density of a FGN is equivalent to $c_1|\lambda|^{-2d}$, the one of ARFIMA behaves also like $c_2|\lambda|^{-2d}$. Consequently the memory of a FGN with $d = 0.4$

Table 4. Empirical test powers (in %).

	Sample size n	$n = 100$			$n = 500$			$n = 1000$		
Test	Level α	1%	5%	10%	1%	5%	10%	1%	5%	10%
R/S		3.4	11.1	18.2	21	38.4	50.8	25.9	44.5	56.7
$m2R/S$		0.6	3.7	9.5	9.1	24.2	35.4	14.2	28.4	39
L_n		2.3	10	15.9	5.7	17.9	26.6	7.6	17.3	24.7
Q_n		1.6	10.6	18.6	10.7	27	39.9	12.7	26.5	38.8
GPH		2.8	12.3	20.4	20.2	41.6	54.7	41.5	64.3	73.1
R		15.7	28.2	36.4	10.2	22.6	33	10.7	22.9	32.5

Note: The table contains rejection frequencies of the null hypothesis of short memory using the six tests. Rejection frequencies are based on 1000 replications generated from the DGP: $(1 - L)^{0.1} y_t = u_t$, $u_t \sim$ i.i.d. $N(0, 1)$ where the nominal significance levels are 1%, 5% and 10%, the sample sizes are $n = 100$, $n = 500$ and $n = 1000$.

Table 5. Empirical test powers (in %).

	Sample size n	$n = 100$			$n = 500$			$n = 1000$		
Test	Level α	1%	5%	10%	1%	5%	10%	1%	5%	10%
R/S		42.2	61.5	70.3	91.9	97.1	98.4	98.5	99.5	99.7
$m2R/S$		0.4	3.9	10.6	27.2	52.1	63.5	53.2	72	79.9
L_n		6.8	22.8	32.5	24.2	36.9	46.7	29.6	44.3	52.4
Q_n		3.4	18.8	34.5	35	58.6	69.9	55.7	71.4	80.2
GPH		25.1	48.1	58.9	80.4	92.7	95.7	94.8	98.7	99.4
R		33.1	47.9	55.1	55.4	74.2	81.1	72.8	86.6	90.6

Note: The table contains rejection frequencies of the null hypothesis of short memory using the six tests. Rejection frequencies are based on 1000 replications generated from the DGP: $(1 - L)^{0.3} y_t = u_t$, $u_t \sim$ i.i.d. $N(0, 1)$ where the nominal significance levels are 1% , 5% and 10%, the sample sizes are $n = 100$, $n = 500$ and $n = 1000$.

Table 6. Empirical test powers (in %).

	Sample size n	$n=100$			$n=500$			$n=1000$		
Test	Level α	1%	5%	10%	1%	5%	10%	1%	5%	10%
R/S		64.4	80.8	86.6	98.9	99.8	99.8	100	100	100
$m2R/S$		0.6	10.2	22.5	55.9	72.4	82.1	73.4	87	91.5
L_n		13	30.1	40.1	36.4	51.2	59.6	44.3	55.9	62.9
Q_n		10.3	33.6	49.9	60.7	78.8	86.8	74.6	87.2	91.5
GPH		74.2	88.2	92.2	100	100	100	100	100	100
R		47.8	63.5	70.9	81.9	91.4	93.3	91.4	97.3	98.1

Note: The table contains rejection frequencies of the null hypothesis of short memory using the six tests. Rejection frequencies are based on 1000 replications generated from the DGP: $(1-L)^{0.4}y_t = u_t$, $u_t \sim$ i.i.d. $N(0,1)$ where the nominal significance levels are 1% , 5% and 10%, the sample sizes are $n=100$, $n=500$ and $n=1000$.

Table 7. Empirical test powers (in %).

	Sample size n	$n=100$			$n=500$			$n=1000$		
Test	Level α	1%	5%	10%	1%	5%	10%	1%	5%	10%
R/S		71.1	85	89.1	98.6	99.9	99.9	99.7	99.9	100
$m2R/S$		0.1	7	17.4	46.5	67.4	76.6	73.1	87.3	91.8
L_n		14.6	30.4	40.3	31.5	46.1	54.1	42.9	57.3	62
Q_n		6.7	31.3	46.2	54	72.7	81.2	73.8	86.6	91.8
GPH		88.7	95.1	97.6	100	100	100	100	100	100
R		49.6	64.2	70.8	79.1	88.5	91.8	92.4	96.3	97.8

Note: The table contains rejection frequencies of the null hypothesis of short memory using the six tests. Rejection frequencies are based on 1000 replications generated from the DGP which is the FGN given by (6) with $d=0.4$ and $\sigma^2=1$ where the nominal significance levels are 1%, 5% and 10%, the sample sizes are $n=100$, $n=500$ and $n=1000$.

is greater than the one of an ARFIMA(0, 0.3, 0) [in the sense that the spectral density of the FGN goes, as $\lambda \to 0$, to infinity faster than the one of the ARFIMA(0, 0.3, 0)]. The ARFIMA(0, 0.4, 0) and the FGN with $d=0.4$ have the same memory, and the rejection frequencies of H_0 for the two processes are remarkably similar: Compare Tables 6 and 7.

4.4 *Nominal size-empirical size curves*

The size distortion of tests can easily be identified by plotting the curve $(x_i, \hat{F}_0(x_i))_{1\le i\le m}$, where

$$\hat{F}_0(x_i) = \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{\{p_j \le x_i\}}, \tag{46}$$

where p_j is the p-value of the test corresponding to the significance level x_i obtained in the jth simulation, and the DGP satisfies the null hypothesis H_0.

For the nominal sizes we choose the grid

$$x_i = 0.001, 0.002, \ldots, 0.01; 0.015, \ldots, 0.3 \quad (m=68). \tag{47}$$

We generate an AR(1) with $\phi = 0.6$, and an ARMA(1,1) with $\phi=0.4$ and $\theta=-3$. For each DGP we perform $N=5000$ replications. We use two sample sizes $n=100$ and $n=500$.

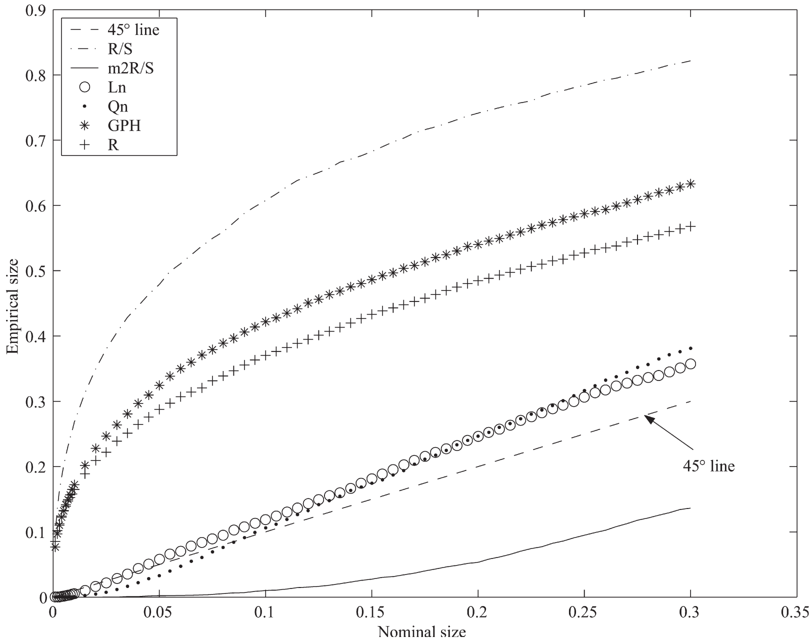

Figure 1. The size distortions.
Note: The DGP is the AR(1): $y_t = 0.6y_{t-1} + u_t$, the sample size is $n = 100$.

Figure 1 shows that, if the process is weakly dependent, then the R/S test has always a positive bias, whereas the one of the $m2R/S$ test is always negative. The tests L_n and Q_n have a negative bias for small nominal sizes, which becomes positive when the nominal size increases. The semi-parametric GPH and R tests have a positive bias. As the sample size increases (Figure 2),

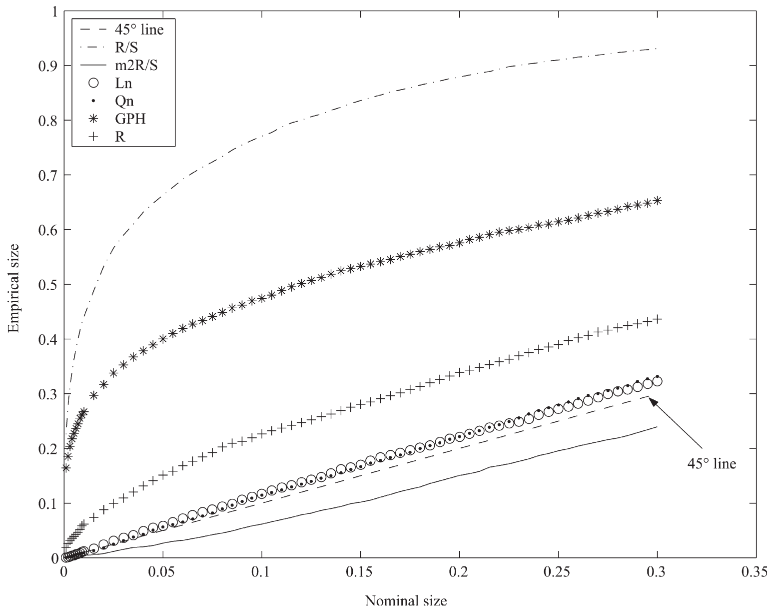

Figure 2. The size distortions.
Note: The DGP is the AR(1): $y_t = 0.6y_{t-1} + u_t$, the sample size is $n = 500$.

the size distortion of R/S increases, the bias of $m2R/S$ remains negative but decreases, the bias of L_n and Q_n decrease and become positive for almost significance level x_i, the size distortion of GPH increases, finally the bias of R is still positive but decreases.

Figures 3 and 4, compared with Figures 1 and 2, show that the behaviour of the non-parametric R/S, $m2R/S$ and Q_n tests for ARMA(1,1) process is fairly similar to the one obtained for AR(1) process. The size distortion of L_n increases. The curves of GPH and R are reversed.

4.5 *Size-power curves*

The comparison of tests can easily be performed by plotting the size-power curves, i.e. the empirical size is plotted in the horizontal axis and the empirical power is plotted in the vertical axis. Such method was suggested by Davidson and MacKinnon [10] to adjust the power to the correct size. The most useful test is the one having the nearest curve to the shape Γ, joining the points (0,0),(0,1) and (1,1). For each test we plot the curve $(\hat{F}_0(x_i), \hat{F}_1(x_i))$, where $\hat{F}_0(x_i)$ is given by Equation (46),

$$x_i = 0.001, 0.002, \ldots, 0.01; 0.015, \ldots, 0.985; 0.990, 0, 0991, \ldots, 0.999 \quad (m = 215), \quad (48)$$

and

$$\hat{F}_1(x_i) = \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{\{p_j \le x_i\}},$$

where p_j is the p-value of the test corresponding to the significance level x_i obtained in the jth simulation, and the DGP satisfies the alternative hypothesis H_1.

We carry out four experiments.

Experiment 1: To compute the empirical size and the empirical power, we generate an AR(1) with $\phi = 0.6$, and an ARFIMA(0,0.3,0), respectively.

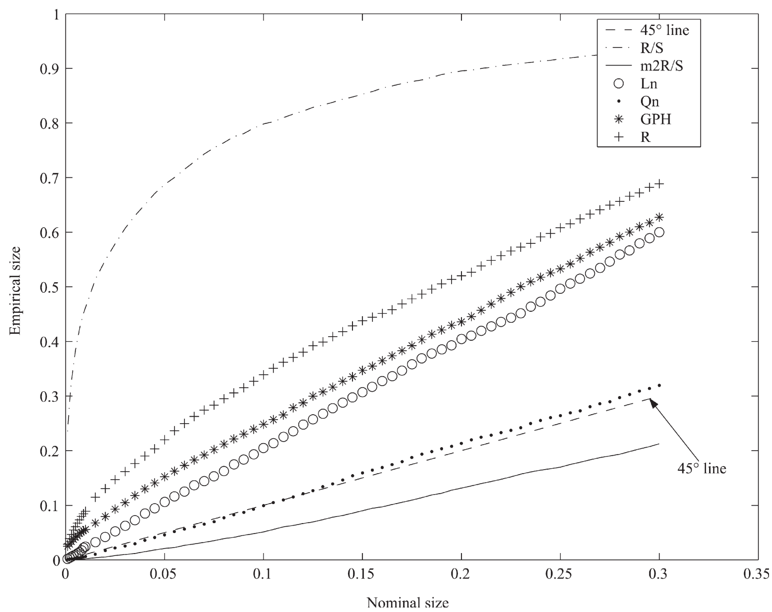

Figure 3. The size distortions.
Note: The DGP is the ARMA(1,1): $y_t = 0.4y_{t-1} + 3u_{t-1} + u_t$, the sample size is $n = 100$.

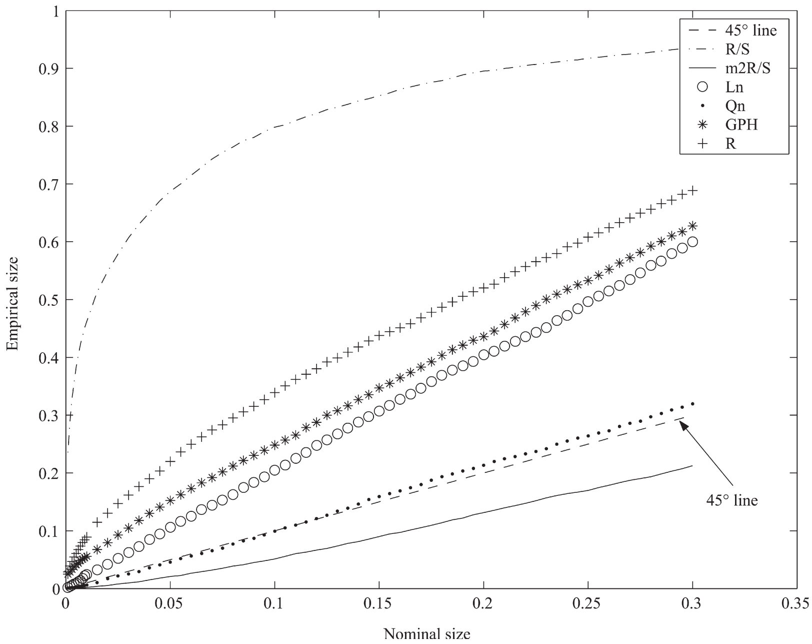

Figure 4. The size distortions.
Note: The DGP is the ARMA(1,1): $y_t = 0.4y_{t-1} + 3u_{t-1} + u_t$, the sample size is $n = 500$.

Recall that the most useful test is the one having the nearest curve to the shape Γ, joining the points (0,0),(0,1) and (1,1). Figure 5 shows that the $m2R/S$ and Q_n tests are more useful than the others. The curve of GPH is over the other curves for large empirical size, this should be

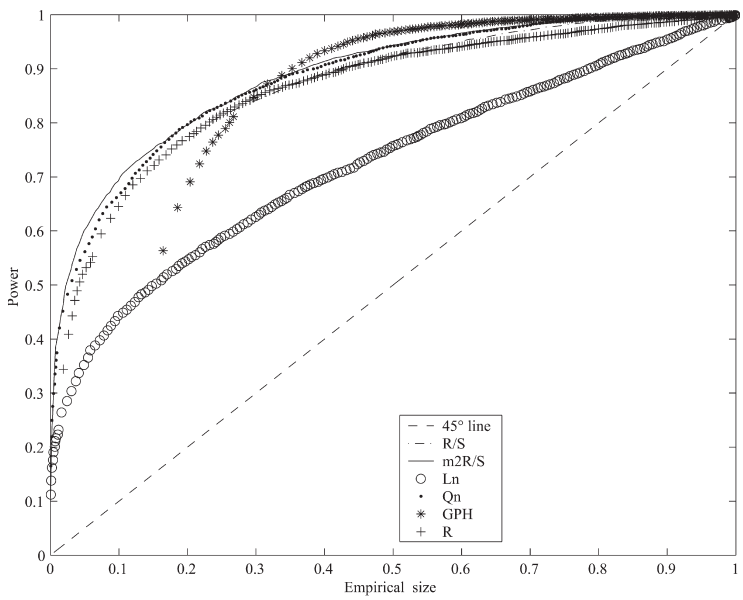

Figure 5. Size-power curves.
Note: The DGP under H_0 is the AR(1): $y_t = 0.6y_{t-1} + u_t$, the DGP under H_1 is the ARFIMA(0,0.3,0): $(1 - L)^{0.3}y_t = u_t$, the sample size is $n = 100$.

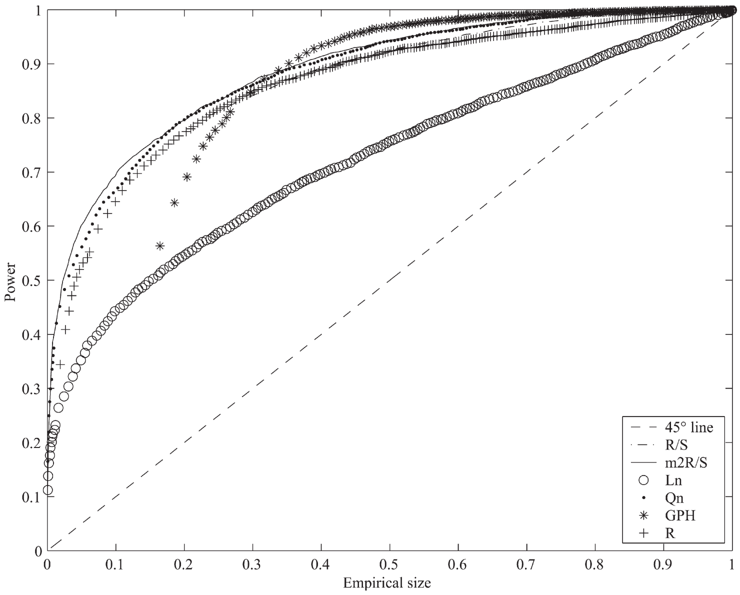

Figure 6. Size-power curves.
Note: The DGP under H_0 is the AR(1): $y_t = 0.6y_{t-1} + u_t$, the DGP under H_1 is the ARFIMA(0,0.3,0): $(1 - L)^{0.3}y_t = u_t$, the sample size is $n = 500$.

interpreted with a care. Indeed, the GPH test does not perform better than the other since its curve is far from the point (0,0). This worse property can be explained as follows: we observe from Figure 1 and 2 that the curve of GPH is always over the 45° line with a high values, which means that the test over-rejects the null. Moreover, it is well known that tests who over-reject the null have always a high power. Note that the empirical size is greater than 16.44% and the empirical power is greater than 56.34% for all significance level $x_i \geq 0.001$. The R/S test has the same problem as GPH.

As the sample size n increases, Figure 6, the $m2R/S$ and Q_n tests remain preferable with a slight superiority of $m2R/S$. The curve of R becomes closer to the ones of $m2R/S$ and Q_n, which means that the performance of the R test becomes equivalent to $m2R/S$ and Q_n. However, the R test suffers from a small size distortion: the observed rejection frequencies exceed 1.88% for every significance level $x_i \geq 0.001$.

Experiment 2: The DGP under H_0 is an ARMA(1,1) with $\phi = 0.4$ and $\theta = -3$, and the DGP under H_1 is an ARFIMA(0,0.3,0).

Figures 7 and 8 provide similar results to those observed in Figures 5 and 6 when AR(1) process is used as a DGP under H_0.

Experiment 3: The DGP under H_0 is an ARMA(1,1) with $\phi = 0.4$ and $\theta = -3$, and the DGP under H_1 is a FGN with $d = 0.4$.

If the DGP under H_1 is a FGN, then the results are different from those obtained when the DGP is an ARFIMA. Figures 9 and 10 show that the R/S, GPH and R tests suffer from a size distortion since their curves are far from the point $(0, 0)$. The performances of $m2R/S$ and Q_n are reversed comparing to the case when the DGP is an ARFIMA (Figures 7 and 8): the Q_n test becomes preferable to $m2R/S$. As the sample size n increases, Figure 10, the spread between $m2R/S$ and Q_n decreases. Only the curves of $m2R/S$, L_n and Q_n approach the shape Γ.

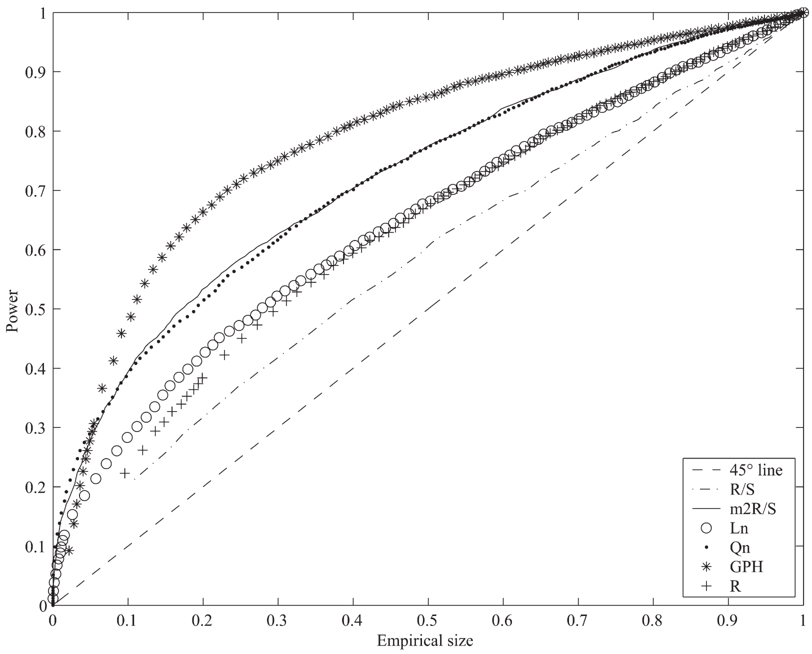

Figure 7. Size-power curves.
Note: The DGP under H_0 is the ARMA(1,1): $y_t = 0.4y_{t-1} + 3u_{t-1} + u_t$, the DGP under H_1 is the ARFIMA(0,0.3,0): $(1 - L)^{0.3}y_t = u_t$, the sample size is $n = 100$.

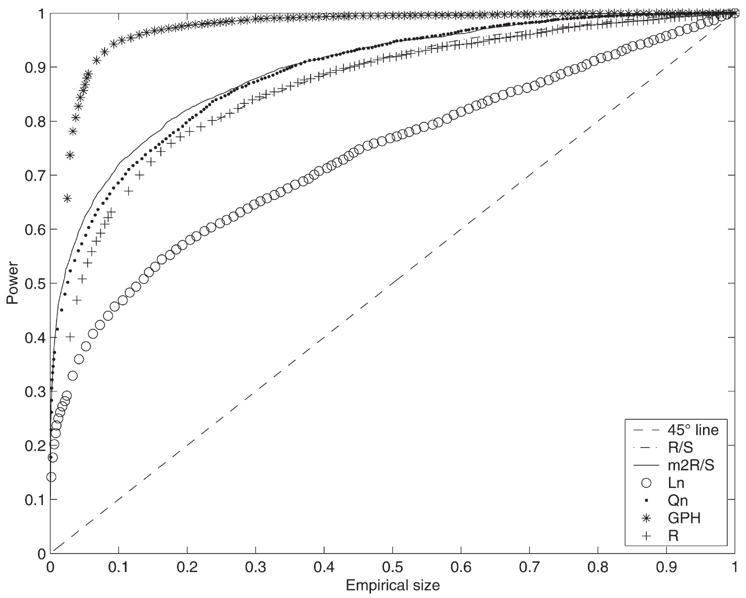

Figure 8. Size-power curves.
Note: The DGP under H_0 is the ARMA(1,1): $y_t = 0.4y_{t-1} + 3u_{t-1} + u_t$, the DGP under H_1 is the ARFIMA(0, 0.3, 0): $(1 - L)^{0.3}y_t = u_t$, the sample size is $n = 500$.

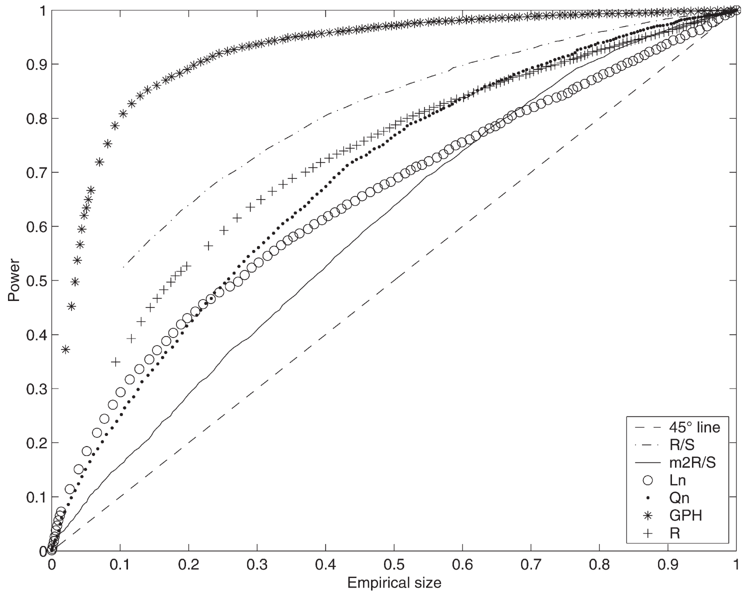

Figure 9. Size-power curves.
Note: The DGP under H_0 is the ARMA(1,1): $y_t = 0.4y_{t-1} + 3u_{t-1} + u_t$, the DGP under H_1 is the FGN with $d = 0.4$, the sample size is $n = 100$.

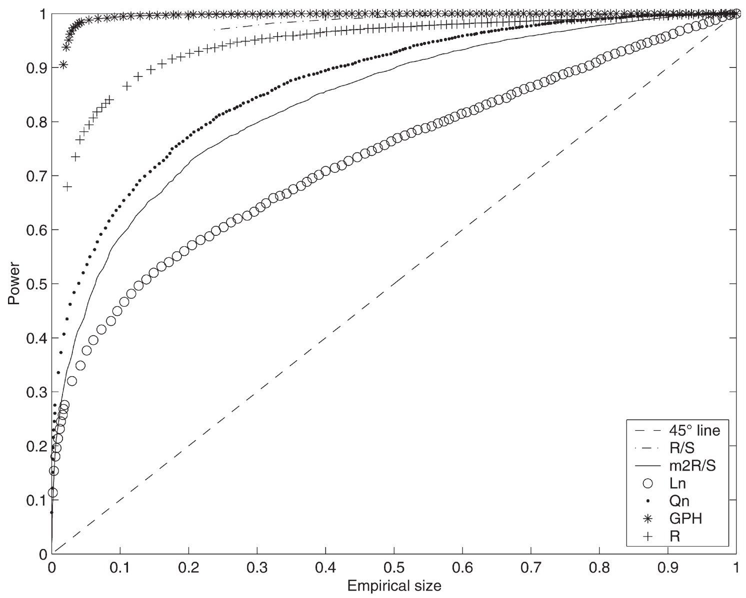

Figure 10. Size-power curves.
Note: The DGP under H_0 is the ARMA(1,1): $y_t = 0.4y_{t-1} + 3u_{t-1} + u_t$, the DGP under H_1 is the FGN with $d = 0.4$, the sample size is $n = 500$.

Experiment 4: The DGP under H_0 is an ARMA(1,1) with $\phi = 0.4$ and $\theta = -3$, and the DGP under H_1 is the square of an ARFIMA(0,0.3,0):

$$y_t = \frac{X_t^2}{E(X_t^2)} - 1, \ (1-L)^{0.3}X_t = u_t, \quad \text{where } (u_t) \sim \text{i.i.d. } N(0,1). \tag{49}$$

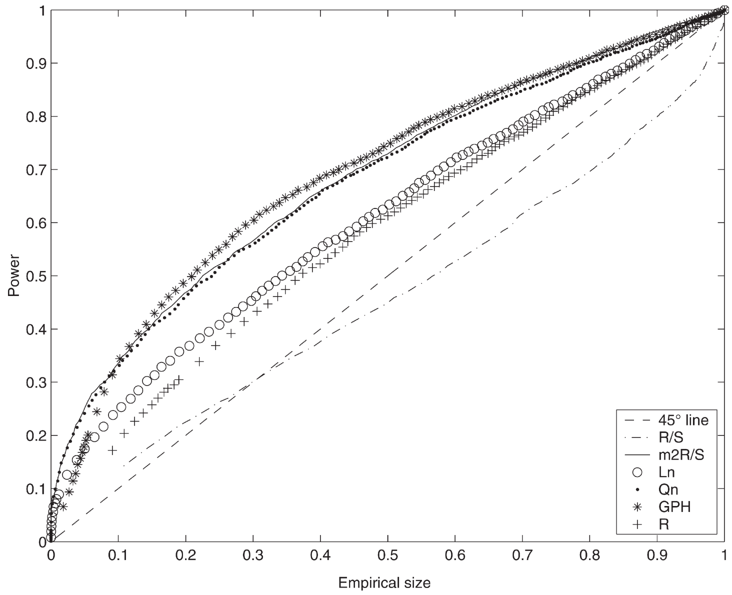

Figure 11. Size-Power curves.
Note: The DGP under H_0 is the ARMA(1,1): $y_t = 0.4y_{t-1} + 3u_{t-1} + u_t$, the DGP under H_1 is the square of the ARFIMA given by Equation (49), the sample size is $n = 100$.

The process (y_t) is nonlinear, and hence does not satisfy the alternative hypothesis H_1. However it has long memory. Indeed, if (X_t) is an ARFIMA(0, d_1, 0) then the autocovariance function of (y_t) is such that $\gamma(k) \sim c_1 k^{2(2d_1-1)}$ as $k \to \infty$, and hence satisfies the definition (1), (see Taqqu [41]).

If the DGP under H_1 is a nonlinear process, the results are also different from those obtained in the two precedent experiments. Figure 11 shows that R/S is a biased test (i.e. the empirical size is greater than the power). The curves of GPH and R are far from (0,0) which implies that the tests suffer from a size distortion. The curve of L_n is near the 45° line and then the test is conservative (i.e. the size and the power are low). The $m2R/S$ and Q_n tests are equivalent and better than the other tests. As the sample size n increases, Figure 12, the R/S test becomes non-biased but has, like the GPH test, a size distortion. The test L_n remains always conservative. Finally, the spreads between $m2R/S$, Q_n and R decrease, and hence the performance of R approaches those of $m2R/S$ and Q_n.

5. Conclusion

We have presented and compared some tests to detect long memory in time series. Based on a limited Monte Carlo experiments, some remarks can be drawn from this study:

- The R/S test has worse properties. This is a well-known result, the test over-rejects the null, see Lo [33].
- The $m2R/S$ test is more useful.
- The K_n and L_n tests, which are functional of the partial sums S_k given by Equation (10), are conservative with a low power. Consequently they are less useful.

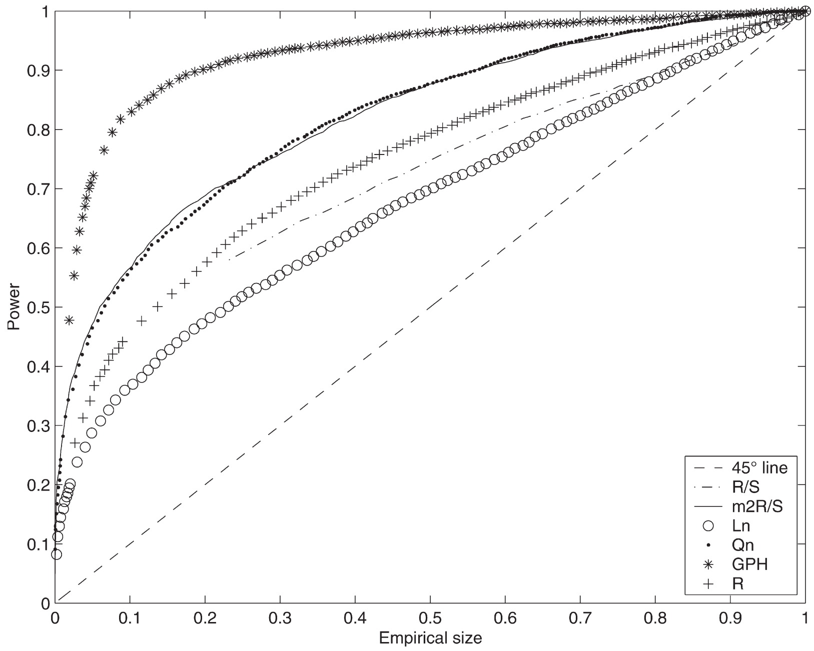

Figure 12. Size-Power curves.
Note: The DGP under H_0 is the ARMA(1,1): $y_t = 0.4y_{t-1} + 3u_{t-1} + u_t$, the DGP under H_1 is the square of the ARFIMA given by Equation (49), the sample size is $n = 500$.

- The $\mathrm{AD}_n(g)$ and $Q_n(g, h)$ tests, which are quadratic functional of the partial sums S_k, have a good performance even for small sample size. Such tests are also more useful in goodness-of-fit testing (see [2]).
- The semi-parametric tests have worse properties for small sample size. The GPH and GPHT tests have a size distortion greater than the one of R. The latter has the same performance as the non-parametric tests $m2R/S$ and $Q_n(g, h)$ for large sample size.

As a conclusion we suggest the use of the non-parametric mR/S, $\mathrm{AD}_n(g)$ and $Q_n(g, h)$ tests if the sample size is small. This is the case for low frequency data (annual, quarterly and monthly). For example macroeconomic samples are often small.

If the sample size is large, as in the case for financial data, then the semi-parametric R test can also be used to detect the presence of long memory. Finally, the R/S,GPH and GPHT tests must be used with a care.

Acknowledgements

I would like to thank the anonymous referees for their helpful comments and valuable suggestions. I also thank the editor for their patient reviews.

Notes

1. Such hypothesis can be tested by using the nonlinearity tests (see [45,46]).
2. Such hypothesis can be tested by using the KPSS test of Kwiatkowski *et al.* [30], the ADF test of Dickey and Fuller [13] and the PP test of Phillips and Perron [36].
3. Results of the other tests are available upon request.

References

[1] T.W. Anderson, *The Statistical Analysis of Time Series*, Wiley, New York, 1971.

[2] T.W. Anderson and D.A. Darling, *Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes*, Ann. Math. Statist. 23 (1952), pp. 193–212.

[3] J. Beran, *Statistics For Long-Memory Processes*, Chapman & Hall, New York, 1994.

[4] J. Beran and N. Terrin, Testing *for a change of the long-memory parameter*, Biometrika 83(3) 1996, pp. 627–638.

[5] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.

[6] M. Boutahar, V. Marimoutou, and L. Nouira, *Estimation methods of the long memory parameter: Monte Carlo analysis and application*, J. Appl. Statist. 34(3) (2007), pp. 261–301.

[7] P.J. Brockwell and R.A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1991.

[8] G.E.P. Box and G.M. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden Day, San Francisco, 1970.

[9] N.H. Chan and N. Terrin, *Inference for unstable long-memory processes with applications to fractional unit root autoregression*, Ann. Statist. 23(5) 1995, pp. 1662–1683.

[10] R. Davidson and J.G. MacKinnon, *Graphical methods for investigating the size and the power of hypothesis tests*, The Manchester School 66 (1998), pp. 1–26.

[11] Yu.A. Davydov, *The invariance principle for stationary processes*, Theory Probab. Appl. 15 (1970), pp. 487–498.

[12] Cl. Deniau, P. Doukhan, G. Oppenheim, and E. Renault, *Théorèmes limites et Longue Mémoire en Statistique*, Stat. Inference Stoch. Process. 3(1–2) (2000), pp. 1–198.

[13] D.A. Dickey and W.A. Fulller, *Likelihood ratio statistics for autoregressive time series with a unit root*, Econometrica 49 (1981), pp. 1057–1072.

[14] P. Doukhan, G. Oppenheim, and M.S. Taqqu, *Theory and Applications of Long-Range Dependence*, Birkhäuser, Boston, MA, 2003.

[15] W. Feller, *The asymptotic distribution of the range of sums of independent random variables*, Ann. Math. Statist. 22 (1951), pp. 427–432.

[16] J. Geweke and S. Porter-Hudak, *The estimation and application of long-memory time series models*, J. Time Ser. Anal. 4 (1983), pp. 221–238.

[17] L.A. Gil-Alana, *A fractional integration of the population in some OECD countries*, J. Appl. Statist. 30(10) (2003), pp. 1–13.

[18] L. Giraitis and D. Surgailis, *CLT and other limit theorems for functionals of Gaussian sequences*, Z. Wahrschein-lickeitstheac Verwandte Geb 70 (1985), pp. 191–212.

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

[20] H. Graf, *Long range correlations and estimation of self-similarity parameter*, Ph.D. thesis, ETH Zürik, 1983.

[21] C.W.J. Granger and R. Joyeux, *An introduction to long-memory time series models and fractional differencing*, J. Time Ser. Anal. 1 (1980), pp. 15–29.

[22] U. Hassler, and J. Wolters, *Long memory in inflation rates: International evidence*, J. Bus. Econom. Statist. 13 (1995), pp. 37–45.

[23] H.C. Ho and T. Hsing, *Limit theorems for functionals of moving average*, Ann. Probab. 25 (1997), pp. 1636–1669.

[24] J.R.M. Hosking, *Fractional differencing*, Biometrika 68 (1981), pp. 165–176.

[25] J.R.M. Hosking, *Modelling persistence in hydrological time series using fractional differencing*, Water Res. Res. 20 (1984), pp. 1898–1908.

[26] J.R.M. Hosking, *Asymptotic distribution of the sample mean, autocovariances, autocorrelations of long-memory time series*, J. Econometrics 73 (1996), pp. 261–284.

[27] H.E. Hurst, *Long-term storage capacity of reservoirs*, Trans. Amer. Soc. Civil Eng. 116 (1951), pp. 770–799.

[28] C.M. Hurvich and R.S. Deo, *Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long-memory time series*, J. Time Ser. Anal. 3 (1998), pp. 331–341.

[29] R.J. Kulperger and R.A. Lockhart, *Tests of independence in time series*, J. Time Ser. Anal. 2(19) (1998), pp. 165–185.

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

[31] D. Lee and P. Schmidt, *On the power of the KPSS unit root test against fractionally integrated alternative*, J. Econometrics 73 (1996), pp. 285–302.

[32] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson, *On the self-similar nature of Ethernet traffic*, IEEE/ACM Trans. Netw. 2(1) (1994), pp. 1–15.

[33] A. Lo, *Long-term memory in stock market prices*. Econometrica 59 (1991), pp. 1279–1313.
[34] B.B. Mandelbrot, *Sur certains prix spéculatifs: faits empiriques et modèle basé sur les processus stables additifs de Paul Lévy*, C. R. Math. Acad. Sci *Paris* 254 (1962), pp. 3968–3970.
[35] B.B. Mandelbrot and M.S. Taqqu, *Robust R/S analysis of long run serial correlation*, Proceeding of the 42nd Session of the International Statistical Institute, Manila, Bulletin of the I.S.I vol. 48(Book 2) (1979), pp. 69–104.
[36] P.C.B. Phillips and P. Perron, *Testing for a unit root in time series regression*, Biometrika 75 (1988), pp. 335–346.
[37] P.M. Robinson, *Log-periodogram regression of time series with long range dependence*, Ann. Statist. 23 (1995a), pp. 1048–1072.
[38] P.M. Robinson, *Gaussian semiparametric estimation of long range dependence*, Ann. Statist. 23 (1995ab), pp. 1630–1661.
[39] P.M. Robinson, *Times Series with Long Memory. Advanced Texts in Econometrics*. Oxford University Press, New York, 2003.
[40] M.S. Taqqu, *Weak convergence to fractional Brownian motion and to the Rosenblatt process*, Z. Wahrscheinlichkeitstheor Verwandte Geb. 31 (1975), pp. 287–302.
[41] M.S. Taqqu, *Fractional Brownian motion and long-range dependence*. in P. Doukhan, G. Oppenheim, and M.S. Taqqu, eds. *Theory and Applications of Long-Range Dependence*, Birkhäuser, Boston, MA, 2003, pp. 5–38.
[42] M.S. Taqqu, V. Teverovsky, and W. Willinger, *Estimators for long-range dependence: an empirical study*, Fractals 3(4) (1995), pp. 785–798.
[43] M.S. Taqqu, W. Willinger, R. Sherman, and D. Wilson, *Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level*, IEEE/ACM Trans. Netw. 5(1) (1997), pp. 71–96.
[44] T. Teräsvirta, *Testing linearity and modelling nonlinear time series*, Kybernetika 30(3) (1994), pp. 319–330
[45] H. Tong, *Non-linear Time Series, A Dynamical System Approach*, Clarendon Press, Oxford, 1990.
[46] K. Torre, D. Delignières, and L. Lemoine, *Detection of long-range dependence and estimation of fractal exponents through ARFIMA modelling*, Br. J. Math. Statist. Psychol. 60(1) (2007), pp. 85–106.
[47] C. Velasco, *Non-Gaussian log-periodogram regression*, Econom. Theory 16 (2000), pp. 44–79.
[48] G.S. Watson, *Goodness-of-fit tests on a circle*, Biometrika 48 (1961), pp. 109–114.
[49] W. Willinger, M.S. Taqqu, and V. Teverovsky, *Stock market prices and long-range dependence*, Finance Stoch. 3 (1999), pp. 1–13.

Appendix

A.1 *Proof of proposition 1*

The convergences (20)–(24) can easily be deduced from the fact that all the statistics can be written as continuous functional of the term in the left-hand side of (8), and by using the continuous mapping theorem: that is for all (X_n), $X \in D[0,1]$, if $X_n \Rightarrow X$ then $g(X_n) \Rightarrow g(X)$ for all continuous function g. For example let us prove the convergence (20). Let

$$Q_n = \frac{R(n)}{\sqrt{2\pi n f(0)}}, \quad X_n(t) = \frac{1}{\sqrt{2\pi n f(0)}} \sum_{k=1}^{[nt]} y_k, \quad Y_n(t) = \frac{\max_{0<k<n} S_{[nt]} - \min_{0<k<n} S_{[nt]}}{\sqrt{2n\pi f(0)}}.$$

We have

$$\frac{mR/S}{\sqrt{n}} = \frac{\sqrt{f(0)}}{\sqrt{\hat{f}(0)}} Q_n, \quad Y_n(t) = g(X_n(t)) \quad \text{and} \quad Q_n = Y_n(1),$$

where $g : D[0,1] \to D[0,1]$ is given by

$$\forall \tau \in (0,1), \quad g(\mathbf{x})(\tau) = \sup_{0\le t\le \tau} (x(t) - tx(1)) - \inf_{0\le t\le \tau} (x(t) - tx(1)),$$

$$\mathbf{x} = (x(t), t \in (0,1)) \in D[0,1].$$

The convergence (8) implies that $Y_n \Rightarrow g(B)$, consequently

$$Q_n = Y_n(1) \xrightarrow{\mathcal{L}} g(B)(1) = \sup_{0\le t\le 1} B_0(t) - \inf_{0\le t\le 1} B_0(t) = V.$$

Finally, since $\sqrt{f(0)}/\sqrt{\hat{f}(0)}$ converges in probability toward 1, the weak convergence of $mR/S/\sqrt{n}$ to V follows immediately.

A.2 *The cumulative distribution functions of the random variables* V, K_0, $AD(g)$ *and* $Q(1,1)$:

The random variable V measures the extent of the Brownian bridge, hence its cumulative distribution function is given by (see [15]):

$$F_V(x) = \sum_{k=-\infty}^{+\infty} (1 - 4x^2k^2)\mathrm{e}^{-2x^2k^2}.$$

The random variable K_0 has a Kolmogorov distribution (see [5, p. 85]),

$$F_{K_0}(x) = 1 + 2\sum_{k=1}^{+\infty}(-1)^k \mathrm{e}^{-2x^2k^2}.$$

The cumulative distribution function of the random variable AD(1) is a somewhat complicated (see [2, p. 202])

$$F_{\mathrm{AD}(1)}(x) = \frac{1}{\pi\sqrt{x\pi}} \sum_{k=0}^{+\infty} \frac{\Gamma(k+1/2)}{k!}\sqrt{4k+1}\mathrm{e}^{-(4k+1)^2/16x} K_{1/4}\left(\frac{(4k+1)^2}{16x}\right),$$

$K_{1/4}$ is the Bessel function:

$$K_{1/4}(y) = \frac{\pi}{\sqrt{2}}(I_{-1/4}(y) - I_{1/4}(y)), \quad I_n(x) = \sum_{k=0}^{+\infty}\left(\frac{x}{2}\right)^{n+2k} \frac{1}{\Gamma(n+k+1)k!}.$$

The cumulative distribution function of the random variable AD(g), with $g(t) = \{t(t-1)\}^{-1/2}$, is given by (see [2, p. 204])

$$F_{\mathrm{AD}(g)}(x) = \frac{1}{x\sqrt{2}} \sum_{k=0}^{+\infty} \frac{(-1)^k\Gamma(k+1/2)}{k!}\sqrt{4k+1}\int_0^1 \mathrm{e}^{-(rx/8)-((4k+1)^2\pi^2/8rx)} \frac{\mathrm{d}r}{r^{3/2}(1-r)^{1/2}}.$$

The random variable $Q(1,1)$ has the following cumulative distribution function (see [50])

$$F_{Q(1,1)}(x) = 1 + 2\sum_{k=1}^{+\infty}(-1)^k \mathrm{e}^{-2x^2\pi^2k^2}.$$

A.3 *Proof of Proposition 2*

The convergence (26) can be deduced from Theorem 3 of Hosking [26]. Let $\|X\|_p = \{E(|X|^p\}^{1/p}$ denotes the Euclidian norm of the random vector $X \in L^p(\Omega)$. From Theorem 3 of Hosking [26], we have for all k

$$\|\hat{\gamma}(k) - \gamma(k)\|_2 \sim \begin{cases} c_1^2 n^{2(2d-1)} & \text{if } \frac{1}{4} < d < \frac{1}{2} \\ c_2^2 \frac{\log n}{n} & \text{if } d = \frac{1}{4} \\ c_3^2 n^{-1} & \text{if } 0 < d < \frac{1}{4}, \end{cases} \tag{A1}$$

for some positive constants c_1, c_2 and c_3. Let $\sigma_n^2(q) = \gamma(0) + 2\sum_{j=1}^{q}(1 - j/(q+1))\gamma(j) = 1/(q+1)\,\mathrm{var}(\sum_{j=1}^{q} y_j)$. Since $\mathrm{var}(\sum_{j=1}^{q} y_j) \sim q^{2d+1}C^2(d,\delta)$, then for n large enough we have $\sigma_n^2(q) \sim q^{2d}C^2(d,\delta)$. Moreover

$$\left\|\frac{\hat{\sigma}_n^2(q)}{(q+1)^{2d}} - C^2(d,\delta)\right\|_1 \le \left\|\frac{\hat{\sigma}_n^2(q)}{(q+1)^{2d}} - \frac{\sigma_n^2(q)}{(q+1)^{2d}}\right\|_1 + \left|\frac{\sigma_n^2(q)}{(q+1)^{2d}} - C^2(d,\delta)\right|. \tag{A2}$$

The second term in the right-hand side of Equation (A2) converges to zero as $q \to \infty$, hence the convergence (26) holds if the first term in the right-hand side of Equation (A2) converges also to zero as $q \to \infty$. We have that

$$
\begin{aligned}
L(n, q, d) :&= \left\| \frac{\hat{\sigma}_n^2(q)}{(q+1)^{2d}} - \frac{\sigma_n^2(q)}{(q+1)^{2d}} \right\|_1 \\
&= \frac{1}{(q+1)^{2d}} \left\| \hat{\gamma}(0) - \gamma(0) + 2\sum_{j=1}^{q}(1 - \frac{j}{q+1})(\hat{\gamma}(k) - \gamma(k)) \right\|_1 \\
&\leq (q+1)^{1-2d} \max_{0 \leq k \leq q} \|(\hat{\gamma}(k) - \gamma(k))\|_1
\end{aligned}
$$

From Equation (A1) it follows that

$$
L(n, d, q) \leq \begin{cases} c_1 \left(\dfrac{q+1}{n}\right)^{1-2d} & \text{if } \dfrac{1}{4} < d < \dfrac{1}{2} \\ c_2 \left(\dfrac{(q+1)\log n}{n}\right)^{1/2} & \text{if } d = \dfrac{1}{4} \\ c_3 \dfrac{(q+1)^{1-2d}}{n^{1/2}} & \text{if } 0 < d < \dfrac{1}{4}. \end{cases}
$$

By using Equation (25), we deduce that $L(n, d, q) \to 0$ as $n \to \infty$, for all $0 < d < 1/2$. The convergences (27)–(31) can be proved by using Equation (26), similar arguments as above and by noting that under $\boldsymbol{H}_1$ the following convergence holds

$$
\frac{1}{n^{d+1/2}} \sum_{k=1}^{[nt]} y_k \quad \Longrightarrow \quad C(d, \delta) B_{d+1/2}(t).
$$