The test for stationarity versus trends and unit roots for a wide class of dependent errors *

Liudas Giraitis†
University of York

Remigijus Leipus
Vilnius University,
Institute of Mathematics and Informatics

Anne Philippe
Université de Lille
Laboratoire de Mathématiques Appliquées CNRS FRE 2222

December 4, 2002

Abstract

We suggest a rescaled variance type test for stationarity (null hypothesis) against deterministic trends and unit roots. The asymptotic distribution of the test is derived and critical values tabulated for a wide class of stationary errors with short, long or negative dependence structure. The proposed test detects a deterministic trend, that can be presented as a general function in time, for example non-parametric, linear or polynomial regression, abrupt changes in the mean, plus unobserved stationary error process which has an unspecified dependence structure (short, long or negative memory). The test is also applicable for unit root alternatives with/without deterministic trend. Simulation study shows that the power of the test significantly improves by increasing the number of observations. In cases when the weak dependence of the error process can be justified, the suggested testing procedure possesses exceptionally good power properties. The case of small samples is also considered. The empirical performance of the test is illustrated for the Nile river and S&P 500 data.

Key words: Test for stationarity, trend, unit root, rescaled variance test, long memory

Running head: The test for stationarity

Corresponding Author: Liudas Giraitis, Department of Mathematics and Economics, Heslington York YO10 5DD, United Kingdom; lg8@york.ac.uk

*The last two authors supported by a cooperation agreement CNRS/LITHUANIA (4714)
†Supported by the ESRC grants R000238212 and R000239538.

1 Introduction

The enormous number of studies in econometrics deals with the question whether economic and financial data are best characterized by the deterministic trend or stochastic trend (unit root) models (see Maddala and Kim (1998) and references therein). The test proposed Dickey and Fuller (1979) became standard for testing unit root against the alternative of stationarity. Since the paper of Nelson and Plosser (1982) evidence of the unit root in many economic time series was established and Nelson–Plosser data set has been used as example data set. On the other hand, it was found that in not large samples the unit root tests have a low power against the relevant alternatives such as long memory fractionally integrated errors (see Diebold and Rudebusch (1991)) or stable autoregressive model with roots near unity (DeJong *et al.* (1992)). The low power of these tests may lead to the acceptance of the unit root hypothesis of the Nelson–Plosser series, when other approaches, developed by Perron (1989) and DeJong and Whiteman (1991) found very few of the Nelson–Plosser series to have unit roots.

These limitations suggest that it would be useful to perform the test for stationarity *as a null hypothesis*. One of the oldest techniques to test the stationarity when the alternative corresponds to some kind of non-stationarity or long-range dependent stationarity is based on the rescaled range, or R/S, statistic of Hurst (1951), later modified by Lo (1991) (to take into account the short-range dependent errors). Bhattacharya *et al.* (1983) applied the R/S test for detecting the presence of deterministic trend in the data. Kwiatkowski *et al.* (1992) proposed a widely used Lagrange multiplier (KPSS) test for the null hypothesis that an observable series is stationary (or, more generally, can contain a linear trend) whereas the alternative hypothesis corresponds to the unit root (or unit root plus linear trend) model. KPSS test indicates that for many series of the Nelson–Plosser data the hypothesis of trend stationarity cannot be rejected at 5% critical level. Lee and Schmidt (1996) have shown that the KPSS test is also consistent against stationary long memory alternatives, such as fractionally integrated $I(d)$ model with $0 < |d| < 1/2$, which means that stationary long memory models are excluded from the null hypothesis of stationarity. Recently Xiao (2001) proposed another stationarity test, similar to KPSS, based on the fluctuation measure in time series.

Concerning more recent unit root studies, Cavaliere (2001) showed that the modified R/S statistic can be used to test for the unit roots. Shin and Schmidt (1992), Lee and Amsler (1997) found that, as it could be expected, standard unit roots tests (e.g., Dickey–Fuller test) can give better results than the KPSS test. Dolado *et al.* (2002) extended the Dickey–Fuller test, originally designed for the $I(1)$ against $I(0)$ case, to the more general, parametric Fractional Dickey–Fuller procedure.

To formulate the testing problem considered in our paper, suppose that under null hypothesis $X_1, \ldots, X_n$ is a sample from the stationary sequence with the alternative that $X_1, \ldots, X_n$ are generated by the non-stationary model

$$X_k = \mu + g_n(k) + \xi_k, \tag{1.1}$$

where $g_n(k)$ is a deterministic trend and $\{\xi_k\}$ is a second order stationary sequence (noise), or by the unit root model

$$X_k - X_{k-1} = \mu + g_n(k) + \xi_k, \tag{1.2}$$

which will serve as the main data generating schemes of non-stationarity in our paper. The models (1.1), (1.2) describe much wider class of non-stationary alternatives including the stochastic trends (unit roots) and deterministic trends such as change point, non-parametric regression and monotonic trends, not investigated in the literature on R/S and KPSS tests, which mostly focuses on the unit root testing procedures.

The stationarity test we consider in this paper is based on the *rescaled variance*, or V/S, statistic introduced by Giraitis et al. (2001) which gives somewhat better balance of size and power than the KPSS test and the modified R/S test of Lo (1991). Both V/S and KPSS statistics are integral-type statistics, so that criterion of the weak convergence in space L_p of Cremers and Kadelka (1986) can be applied to derive the corresponding limiting distributions.

It should be noted that the classical stationarity testing literature deals mostly with the stationary error process $\{\xi_k\}$ possessing specific dependence structure, such as iid or weak dependence, or I(d) with the memory parameter d known a priori. In practice, however, the dependence structure of the error process is not known, the variables might have short, long or negative memory. As it was mentioned, the tests designed for weakly dependent errors are incapable of distinguishing between strong dependence and deterministic trends (structural changes) or unit roots, consequently they can misspecify long memory as a "spurious trend" and vice versa (see Lobato and Savin (1998)). It might be very difficult to distinguish small trends (change points in the mean) and the long memory graphically since the paths of stationary time series under strong dependence contain patterns which resemble "spurious" local trends.

Therefore, it is important to have a test for stationarity which allows to detect trends/unit roots under all types of dependence of stationary noise sequence including long memory. In our paper, we construct such test based on the V/S statistic. The model can be extended to the non-stationary errors, i.e. some heterogeneity is allowed. It is shown that the proposed test is consistent against wide range of alternatives corresponding to nonlinear trends and unit roots, and has good power in large samples. In cases when the assumption of weak dependence of the error process can be justified, the corresponding testing procedure possess exceptionally good power properties.

The plan of the paper is the following. The models under consideration and the test statistic are introduced in Section 2. The asymptotic results are presented in Section 3. Monte-Carlo simulation results and empirical examples are presented in sections 4 and 5. Section 6 discuss some results of Cremers and Kadelka (1986) on the convergence in space $L_2[0,1]$ and their applications. Finally, Appendix contains the proofs of results.

2 Assumptions of the model and testing procedure

In the models considered in this paper we assume that the stationary noise sequence $\{\xi_k\}$ may have *short memory*, characterized by absolutely summable covariance $\gamma_k = \mathrm{Cov}(\xi_k, \xi_0)$,

$$\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty, \tag{2.1}$$

(this holds as a rule for iid variables and mixing process); the ξ_k may exhibit *long memory* which is usually characterized by slowly decaying covariance

$$\gamma_k \sim ck^{2d-1} \tag{2.2}$$

as $k \to \infty$, for some $c \neq 0$ and $0 < d < 1/2$ or even have *negative memory* described by the property

$$\gamma_k \sim ck^{2d-1}, \quad \sum_{k=-\infty}^{\infty} \gamma_k = 0 \tag{2.3}$$

for some $c \neq 0$, $-1/2 < d < 0$.

We shall write $(\gamma_k) \in G(d)$, $d \in (-1/2, 1/2)$ to denote that γ_k satisfies asymptotics (2.1)–(2.3) where $d = 0$ corresponds to the short memory case (2.1).

Now we describe rigorously the hypotheses considered in our paper.

Hypothesis H_0 (stationarity). We say that the random variables X_k satisfy the null hypothesis H_0 if

$$X_k = \mu + \xi_k, \tag{2.4}$$

where μ is a real number and $\{\xi_k\}$ is a stationary sequence with zero mean and covariance function $(\gamma_k) \in G(d)$. We assume that $d \in [-a_1, a_2]$ where $0 \leq a_1, a_2 < 1/2$.

Theorem 3.1 below, under additional conditions, establishes asymptotics of the test statistic under H_0.

The departure from stationarity is modeled by the alternative H_1 which includes the commonly used deterministic trend plus noise and the unit root hypotheses.

Alternative H_T (deterministic trend plus noise). We say that the X_k satisfy alternative hypothesis H_T if

$$X_k = \mu + g_n(k) + \xi_k, \tag{2.5}$$

where $\mu \in \mathbf{R}$, $\{\xi_k\}$ is a stationary sequence with zero mean and covariance function $(\gamma_k) \in G(d)$, $d \in [-a_1, a_2]$ with $0 \leq a_1, a_2 < 1/2$. Deterministic trend function $g_n(k)$ which leads to the asymptotically robust procedure is described in Proposition 3.1, whereas trends corresponding to the asymptotically consistent procedure are given in Theorem 3.2.

Alternative H_U (unit root). We say that the X_k satisfy alternative hypothesis H_U if

$$X_k - X_{k-1} = \mu + g_n(k) + \xi_k, \tag{2.6}$$

where μ and the ξ_k are the same as under hypothesis H_T. Theorem 3.3 establishes additional conditions on the ξ_k and $g_n(k)$ under which the proposed testing procedure is consistent.

The following assumption on the functions $g_n(k)$, characterizing the class of trends, plays an important role in the proof of consistency results.

Assumption T(γ). There exists constant $\gamma > -1/2$ such that

(g1) for some function $g^* \in L_2[0,1]$ uniformly in $t \in [0,1]$ and $n \geq 1$,

$$n^{-\gamma}|g_n([nt])| \leq g^*(t); \tag{2.7}$$

(g2) for almost all $t \in [0,1]$, there exists the limit, as $n \to \infty$,

$$n^{-\gamma} g_n([nt]) \to g_*(t) \neq \text{const}. \tag{2.8}$$

Dominated convergence theorem implies that $g_* \in L_2[0,1]$; we assume that $g_*(t) \neq$ const in L_2.

Remark 2.1 Assumption T(γ) is satisfied for the main classes of trends of practical interest. For example,

(a) *polynomial trend*

$$g_n(k) = ck^\beta, \quad c \neq 0,\ \beta > -1/2$$

satisfies Assumption T(γ) with $\gamma = \beta$;

(b) *non-parametric regression*

$$g_n(k) = g(k/n), \quad k = 1, \dots, n,$$

where $g(t), t \in [0,1]$ is a bounded function, satisfies T(γ) with $\gamma = 0$. This class of trends covers also structural breaks in the mean (single and multiple change points) of type

$$g_n([nt]) = \begin{cases} \mu_1 & \text{for } 0 \le t \le \tau, \\ \mu_2 & \text{for } \tau < t \le 1 \end{cases}$$

with $0 < \tau < 1$;
(c) a large variety of *deterministic regression models.*

In a sequel, the crucial assumption for our testing procedure will be the linear structure of errors ξ_k (although main theorems hold for more general models, see Section 3):

$$\xi_k = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{k-j}, \tag{2.9}$$

where the a_j are real weights, $\sum_j a_j^2 < \infty$, and the ε_j are iid random variables with zero mean, unit variance and finite fourth moment $E\varepsilon_0^4 < \infty$. We assume in addition that either

$$\sum_j |a_j| < \infty, \tag{2.10}$$

or with some $c \neq 0$, $0 < d < 1/2$

$$a_j \sim c j^{d-1} \quad (j \to \infty), \tag{2.11}$$

or with some $c \neq 0$ and negative $-1/2 < d < 0$

$$\sum_{j=-\infty}^{\infty} a_j = 0, \qquad a_j \sim c j^{d-1} \quad (j \to \infty). \tag{2.12}$$

As covariance function of the ξ_k equals $\gamma_k = \sum_{j=-\infty}^{\infty} a_{j+k} a_j$, it is easy to check that, under assumption (2.10), $\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty$ and $(\gamma_k) \in G(0)$ holds; under assumption (2.11) γ_k has property $\gamma_k \sim C_d k^{2d-1}$ with some $C_d \neq 0$ and $(\gamma_k) \in G(d)$ holds; under (2.12) γ_k has property $\sum_{k=-\infty}^{\infty} \gamma_k = 0$, $\gamma_k \sim C_d k^{2d-1}$ $(C_d \neq 0)$ and $(\gamma_k) \in G(d)$ holds with negative $d \in (-1/2, 0)$.

2.1 Testing procedure

Our test is based on the rescaled variance (or V/S) statistic, see Giraitis *et al.* (2003), Giraitis *et al.* (2001). It is defined as the ratio $V_n/\hat{s}_{n,q}^2$, where

$$V_n = \frac{1}{n^2} \left[\sum_{k=1}^{n} (S_k^*)^2 - \frac{1}{n} \left(\sum_{k=1}^{n} S_k^* \right)^2 \right], \tag{2.13}$$

where $S_k^* = \sum_{j=1}^{k} (X_j - \bar{X}_n)$ and

$$\hat{s}_{n,q}^2 = q^{-1} \sum_{i,j=1}^{q} \hat{\gamma}_{i-j} \tag{2.14}$$

with the $\hat{\gamma}_j$ being the sample covariances, $\hat{\gamma}_j = n^{-1}\sum_{i=1}^{n-j}(X_i - \bar{X}_n)(X_{i+j} - \bar{X}_n), \quad 0 \le j < n$, where the bandwidth $q = q_n$ satisfies $q \to \infty$, $q/n \to 0$. This statistic is similar to the R/S statistic introduced by Hurst (1951) and modified by Lo (1991) to take into account short range dependence. Compared with the R/S statistic, the numerator (2.13) is based on the sample variance of the sums S_k^*, rather than their range in Lo's statistic.

Define the test function

$$T_n(\hat{d}) = \left(\frac{q}{n}\right)^{2\hat{d}} \frac{V_n}{\hat{s}_{n,q}^2}, \tag{2.15}$$

where $\hat{d}$ is any estimate of d such that under H_0

$$\hat{d} - d = o_P(1/\log n). \tag{2.16}$$

Recall that we assume under H_0 that $d \in [-a_1, a_2]$ with $0 \le a_1, a_2 < 1/2$. To estimate d, we use the local Whittle estimation procedure carried out over the interval $[-a_1, a_2] = [-0,4; 0,4]$.

As follows from the results in Section 3, under H_0 statistic (2.15) weakly converges to the random variable Z_d given in (2.21) which depends on the *unknown* memory parameter $d \in [-a_1, a_2]$. Denote $c_\alpha(d)$ an $(1-\alpha)$-quantile of Z_d:

$$c_\alpha(d) = \inf\{x : F_{Z_d}(x) \ge \alpha\}, \tag{2.17}$$

where $F_{Z_d}(x)$ is a distribution function of random variable Z_d. The properties of quantile $c_\alpha(d)$ are presented in Subsection 2.2. The testing procedure is based on the quantities $T_n(\hat{d})$ and $c_\alpha(\hat{d})$ in which estimate $\hat{d}$ replaces the unknown parameter d.

Test. We reject H_0 if

$$T_n(\hat{d}) > c_\alpha(\hat{d}). \tag{2.18}$$

As it follows from Theorem 3.1, the proposed test is correctly specified, i.e. the first-type error asymptotically equals significance level α:

$$P\{T_n(\hat{d}) > c_\alpha(\hat{d}) | H_0\} \to \alpha. \tag{2.19}$$

By Theorems 3.2 and 3.3, the test is consistent under H_T or H_U, and H_0 will be rejected with probability tending to 1:

$$P\{T_n(\hat{d}) > c_\alpha(\hat{d}) | H_1\} \to 1. \tag{2.20}$$

Remark 2.2

a) The results of this paper can be easily formulated in terms of the KPSS statistic.

b) The use of smaller a_2 in H_0 assumption $d \in [-a_1, a_2]$ will increase the power of the test. Of course, in such case the stationary errors ξ_k with the memory parameter $d \in (a_2, 1/2)$ are excluded from the null hypotheses (they can be included into alternative, however the test will detect such alternatives with a low power when d is close to a_2). In applications, a_1 could be chosen between 0 and 0.4, without significant reduction of power; meanwhile estimation with smaller $a_2 \in [0, 1/2)$ will increase the power of the test.

c) In a special case of the null hypothesis, when the errors ξ_k have short memory with parameter $d = 0$, which is common in practice, instead of testing procedure (2.18) the critical region

$$T_n(0) > c_\alpha(0)$$

should be used. Such procedure leads to a significant increase of the power, see Subsection 4.2.

d) It can be shown that the test is consistent in the case of integrated errors $I(d)$, $d > 3/2$.

e) The iid noise ε_k in (2.9) can be replaced by the martingale difference-type sequence $\{\varepsilon_k\}$ having properties $E(\varepsilon_k|\mathcal{F}_{k-1}) = 0$, $E(\varepsilon_k^i|\mathcal{F}_{k-1}) = \mu_i$, $i = 2, 3, 4$, where $\mathcal{F}_k$ is the $\sigma-$field generated by $\varepsilon_s, s \le k$.

2.2 Quantiles $c_\alpha(d)$

Suppose that $W_H(t), 0 \le t \le 1$ denotes the fractional Brownian motion with parameter H, i.e. a Gaussian process with zero mean and covariance $EW_H(t_1)W_H(t_2) = 1/2\ (t_1^{2H} + t_2^{2H} - |t_1 - t_2|^{2H})$. In case $H = 1/2$, $W_{1/2}(t)$ is the standard Wiener process (Brownian motion).

Set

$$Z_d = \int_0^1 (W^0_{1/2+d}(t))^2 dt - \left(\int_0^1 W^0_{1/2+d}(t) dt\right)^2, \quad |d| < 1/2 \tag{2.21}$$

where $W_H^0(t), 0 \le t \le 1$ is the fractional Brownian bridge $W_H^0(t) = W_H(t) - tW_H(1)$. Let $c_\alpha(d)$, (2.17), be a $(1-\alpha)$-quantile of Z_d. It can be easily shown that Z_d has continuous strictly increasing distribution function for all $-1/2 < d < 1/2$, hence its quantile has the desirable continuity property (see Lemma 2.1).

In Table 1 we provide the values of the quantiles $c_\alpha(d)$ calculated using two approaches: first consists in 10 000 Monte-Carlo simulations of the statistic $(q/n)^{2d}(V_n/\hat{s}^2_{n,q})$ for the sequence FARIMA(0,d,0), second consists in the discretization of the integrals in expression of the random variable Z_d in (2.21). Both approaches give almost identical results. To obtain the values of the quantile $c_\alpha(d)$ for all $d \in (-1/2, 1/2)$ we use a polynomial approximation. Numerical experiments show that this approximation fits very well. For $\alpha = 5\%$ the standard MATLAB procedure gives the following formula: $c_{5\%}(d) = -1.98d^5 + 0.73d^4 - 0.05d^3 + 0.63d^2 - 0.66d + 0.19$; for $\alpha = 10\%$, $c_{10\%}(d) = -0.27d^5 + 0.48d^4 - 0.55d^3 + 0.66d^2 - 0.53d + 0.14$.

Table 1: Quantiles $c_\alpha(d)$ for the statistic Z_d with $\alpha = 5\%$ and 10%.

d	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4
$c_{5\%}(d)$	0.597	0.457	0.349	0.262	0.190	0.130	0.083	0.049	0.022
$c_{10\%}(d)$	0.524	0.393	0.292	0.213	0.153	0.105	0.067	0.039	0.016

Figure 1 shows the values of quantile $c_\alpha(d)$ for $d \in [-0.45, 0.45]$ and several values of α.

In case $d = 0$ the analytical properties of the random variable Z_0 are known; the distribution function is given by formula

$$F_{Z_0}(x) = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2\pi^2 x} \tag{2.22}$$

and was established by Watson (1961) in the context of goodness-of-fit tests on a circle. Note that $F_{Z_0}(x) = F_K(\pi\sqrt{x})$, where F_K is the asymptotic distribution function of the standard

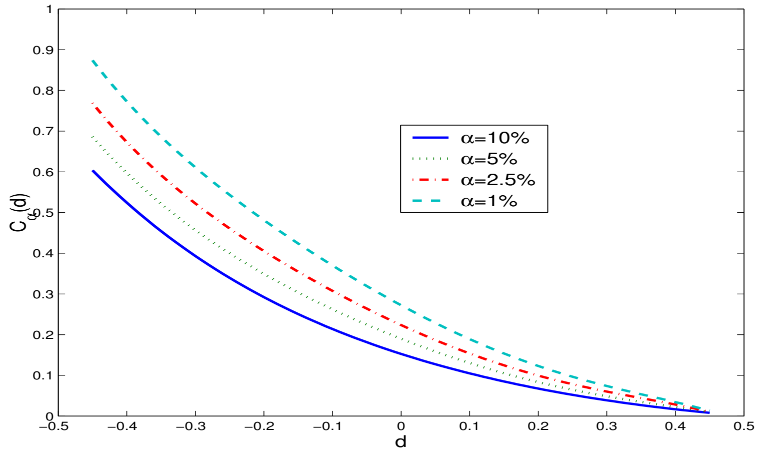

Figure 1: Quantiles $c_\alpha(d)$.

Kolmogorov statistic. Moreover, Z_0 admits the representation

$$Z_0 = \frac{1}{4\pi^2} \sum_{j=1}^{\infty} \frac{Y_{2j-1}^2 + Y_{2j}^2}{j^2},$$

where the Y_j are independent standard normal variables, and $EZ_0 = 1/12$, $\mathrm{Var}\, Z_0 = 1/360$ (see for more details Giraitis *et al.* (2003)).

LEMMA 2.1 *Suppose that* $\hat{d}, d \in [-a_1, a_2]$ $(0 \le a_1, a_2 < 1/2)$ *and* $\hat{d} \xrightarrow{P} d$ *as* $n \to \infty$. *Then*

$$c_\alpha(\hat{d}) \xrightarrow{P} c_\alpha(d). \tag{2.23}$$

Proof is given in Appendix A.

2.3 Local Whittle estimate of d

Our testing procedure requires a consistent estimator of the memory parameter d. For this purpose we use the local Whittle estimator $\hat{d}$ (see Künsch (1987), Robinson (1995)) defined as

$$\hat{d} = \mathrm{argmin}_{[-a_1, a_2]} U_n(d), \tag{2.24}$$

where $0 \le a_1, a_2 < 1/2$ and $U_n(d)$ is a local contrast function

$$\begin{aligned} U_n(d) &= \log\left(\frac{1}{m}\sum_{j=1}^{m} \lambda_j^{2d} I_n(\lambda_j)\right) - \frac{2d}{m}\sum_{j=1}^{m} \log \lambda_j \\ &= \log\left(\frac{1}{m}\sum_{j=1}^{m} j^{2d} I_n(\lambda_j)\right) - \frac{2d}{m}\sum_{j=1}^{m} \log j. \end{aligned}$$

Here $\lambda_k = \frac{2\pi k}{n}$, $k = 1, \ldots, [(n-1)/2]$ are the Fourier frequencies and $I_n(\lambda) = (2\pi n)^{-1} \left|\sum_{j=1}^{n} e^{ij\lambda} X_j\right|^2$ is the periodogram.

We assume that ξ_k is a linear process (2.9) with spectral density satisfying following assumptions:

$$f(\lambda) = c\lambda^{-2d} + o(\lambda^{\beta}) \quad \text{as } \lambda \to 0+, \tag{2.25}$$

where $c > 0, \beta > 0$ and $-1/2 < d < 1/2$; $f(\lambda)$ is differentiable in a neighborhood $(0, \delta)$ of the origin and

$$\frac{d}{d\lambda} \log f(\lambda) = O(\lambda^{-1}) \quad \text{as } \lambda \to 0+; \tag{2.26}$$

the bandwidth $m \equiv m_n$ satisfies $n^{\epsilon} \leq m \leq n^{1-\epsilon}$ with $\epsilon > 0$.

Note that properties (2.25), (2.26) are satisfied under some additional regularity assumptions on a_j's. The proof of Theorem 1 of Robinson (1995) implies the validity of (2.16):

Proposition 2.1 *Suppose X_k is a linear process* (2.9) *satisfying assumptions* (2.25), (2.26) *and $\hat{d}$ is a local Whittle estimate* (2.24). *Then*

$$\hat{d} - d = o_P(1/\log n).$$

In case of Gaussian sequence X_k, under assumptions above, it holds $\hat{d} - d = O_P((m/n)^{\beta} + m^{-1/2})$ (see Lemma 5.7 of Giraitis and Robinson (2001)).

The rate of convergence of the local Whittle estimator in the non-parametric regression model with dependent errors was discussed in Robinson (1997), Hurvich *et al.* (2002) among others. Note that the parametric testing procedure suggested in Dolado *et al.* (2002) requires much stronger assumption of $n^{1/2}$-consistency of estimator $\hat{d}$.

3 Asymptotic theory

In the previous section we dealt with the case where the errors ξ_k have a moving average representation (2.9). Consider now *general* errors ξ_k.

Let $(\gamma_k) \in G(d)$, $|d| < 1/2$, and define

$$s_0^2 = \sum_{j=-\infty}^{\infty} \gamma_j \ \text{ if } d = 0, \quad s_d^2 = C_d/(d(2d+1)) \ \text{ if } d \neq 0, \tag{3.1}$$

where C_d is given in (2.2) or (2.3). Denote by $\xrightarrow{D}$ the convergence in distribution.

3.1 Asymptotic null distribution and robustness

Theorem 3.1 *Suppose that $X_1, \ldots, X_n$ is a sample obtained from*

$$X_k = \mu + \xi_k,$$

where $\mu \in \mathbf{R}$ and $\{\xi_k\}$ is a zero mean covariance stationary sequence with $(\gamma_k) \in G(d)$, $d \in [-a_1, a_2]$, $0 \leq a_1, a_2 < 1/2$. Assume that

(a1) *finite dimensional distributions of the process*

$$X_n(t) = n^{-1/2-d} \sum_{j=1}^{[nt]+1} \xi_j, \quad 0 \leq t \leq 1 \tag{3.2}$$

converge to those of the process $s_d W_{1/2+d}(t)$*;*

(a2) $\hat{d} \in [-a_1, a_2]$ *is an estimate of* d *such that*

$$\hat{d} - d = o_P(1/\log n); \tag{3.3}$$

(a3)

$$\frac{\sum_{i,j=1}^{q} \hat{\gamma}_{i-j}}{\sum_{i,j=1}^{q} \gamma_{i-j}} \xrightarrow{P} 1. \tag{3.4}$$

Then

$$T_n(\hat{d}) \xrightarrow{D} Z_d \tag{3.5}$$

and

$$P\{T_n(\hat{d}) > c_\alpha(\hat{d})\} \to \alpha, \tag{3.6}$$

where $c_\alpha(d)$ *is an* $(1-\alpha)$*-quantile of* Z_d*.*

The following proposition describes the trends which do not change the asymptotic distribution of the test statistic, i.e. such trends asymptotically can not be detected by our testing procedure.

PROPOSITION 3.1 (Negligible trend). *Suppose that*

$$X_k = \mu + g_n(k) + \xi_k,$$

where $\mu \in \mathbf{R}$*, the* ξ_k *satisfy assumptions of Theorem 3.1 and* $g_n(k)$ *is a deterministic trend such that*

$$\sum_{k=1}^{n} (g_n(k) - \bar{g}_n)^2 = o(n^{2d}). \tag{3.7}$$

Then, for $\hat{d}$ *such as in* (3.3)*, relations* (3.5), (3.6) *remain valid.*

REMARK 3.1 Assumptions (a1)–(a3) of Theorem 3.1 hold for a linear process ξ_k (2.9) with covariances $(\gamma_k) \in G(d)$ and weights a_j satisfying either (2.10), or (2.11), or (2.12). Indeed, in the case of linear processes with $E\varepsilon_k^2 < \infty$, property (a1), i.e. the convergence of finite dimensional distributions is well-known, see, e.g., Ibragimov and Linnik (1971, Theorem 18.6.5), Davydov (1970), and the local Whittle estimator (2.24) has property (a2), see Section 2.3. It is very likely that under assumption $(\gamma_k) \in G(d)$, the local Whittle estimate $\hat{d}$ (2.24) is a consistent estimate of the memory parameter d and property (a2) is valid for a much wider class of errors ξ_k than linear processes.

Assumption (a3) is also not restrictive. It means that the variance $\mathrm{Var}(\sum_{j=1}^{q} X_j) = \sum_{i,j=1}^{q} \gamma_{i-j}$ is well approximated by the empirical variance $\sum_{i,j=1}^{q} \hat{\gamma}_{i-j}$. In Lemma 8.4 in Appendix B it is shown that (3.4) is satisfied for a wide class of errors ξ_k such that

$$\sup_h \sum_{r,s=-\infty}^{\infty} |\kappa(h,r,s)| \le C n^{2d} \quad \text{if} \quad d \ge 0, \tag{3.8}$$

$$\sum_{h,r,s=-\infty}^{\infty} |\kappa(h,r,s)| \le C \quad \text{if} \quad d < 0, \tag{3.9}$$

where $\kappa(h, r, s)$ is a fourth-order cumulant defined by

$$\kappa(h, r, s) = E[\xi_k \xi_{k+h} \xi_{k+r} \xi_{k+s}] - (\gamma_h \gamma_{r-s} + \gamma_r \gamma_{h-s} + \gamma_s \gamma_{h-r}). \tag{3.10}$$

The validity of assumption (3.8) for a linear process ξ_k in case $d \geq 0$ was shown in Giraitis *et al.* (2003); in case $d < 0$, (3.9) follows from $\kappa(h, r, s) = (E\varepsilon_0^4 - 3) \sum_{k=-\infty}^{\infty} a_k a_{k+h} a_{k+r} a_{k+s}$ and $\sum_{k=-\infty}^{\infty} |a_k| < \infty$.

3.2 Consistency of the test

In this subsection we specify general conditions for the consistency of the test.

THEOREM 3.2 *Let $X_1, \ldots, X_n$ be a sample obtained from the model*

$$X_k = \mu + g_n(k) + \xi_k, \tag{3.11}$$

where $\mu \in \mathbf{R}$ and $g_n(k)$ satisfies Assumption T(γ). Assume that ξ_k are zero mean random variables with uniformly bounded variance:

$$E\xi_k^2 \leq C. \tag{3.12}$$

Then, for any $\hat{d} \in [-a_1, a_2]$, convergence

$$T_n(\hat{d}) \xrightarrow{D} \infty \tag{3.13}$$

holds (hence H_0 is rejected) in the following cases:

(h1) $\gamma > 0$ *with no additional restrictions on the ξ_k;*

(h2) $\gamma = 0$ *and*

$$E(\bar{\xi}_n)^2 \to 0, \tag{3.14}$$

where $\bar{\xi}_n := n^{-1} \sum_{j=1}^{n} \xi_j$;

(h3) $\gamma < 0$, *$\{\xi_k\}$ is a covariance stationary sequence with $(\gamma_k) \in G(d)$, $-1/2 < d < 1/2$, and d, γ satisfy restrictions*

$$d - 1/2 < \gamma < 0 \tag{3.15}$$

and

$$n^{-2\gamma} \leq C q^{1-2d}. \tag{3.16}$$

REMARK 3.2

a) The case (h1) includes polynomial trends ck^γ with $\gamma > 0$; Theorem 3.2 shows that such trends can be detected both for stationary and non-stationary errors ξ_k satisfying (3.12).

The case (h2) covers non-parametric regression model (including change points in the mean) with

$$g_n(k) = G_n(k/n),$$

where $G_n([nt]/n) \to g_*(t)$ for any $t \in [0, 1]$, and uniformly in t, n, $|G_n([nt]/n)| \leq g^*(t)$ for some $g^* \in L_2[0, 1]$.

Assumption (3.14) is obviously satisfied for covariance stationary errors ξ_k if $\mathrm{Cov}(\xi_k, \xi_0) \to 0$.

The case (h3) deals with small trends of type ck^γ with $\gamma < 0$; according to (3.15), when $0 < d < 1/2$, the test detects only trends with small $|\gamma| < 1/2 - d$.

b) Note that under alternatives H_T, H_U the estimate $\hat{d} \in [-a_1, a_2]$ does not have to satisfy condition (3.3) (as a rule it does not hold).

Consider now the general unit root model.

THEOREM 3.3 *Let $X_1, \dots, X_n$ be a sample obtained from the model*

$$X_k - X_{k-1} = \mu + g_n(k) + \xi_k, \tag{3.17}$$

where $\mu \in \mathbf{R}$ and $g_n(k)$ is a deterministic trend. Assume that the ξ_k are zero mean random variables with uniformly bounded variance:

$$E\xi_k^2 \le C. \tag{3.18}$$

Then, for any $\hat{d} \in [-a_1, a_2]$, convergence

$$T_n(\hat{d}) \xrightarrow{D} \infty \tag{3.19}$$

holds (and H_0 is rejected) in the following cases:

(h1*) *trend $g_n(k)$ satisfies Assumption T(γ) with $\gamma > 0$;*

(h2*) *$g_n(k)$ and ξ_k satisfy Assumption T(γ) with $\gamma = 0$ and* (3.14) *correspondingly;*

(h3*) *$\mu \ne 0$, ξ_k satisfy* (3.14) *and $g_n(k)$ are such that*

$$\sum_{j=1}^{n} (g_n(j) - \bar{g}_n)^2 = o(n^{1/2}); \tag{3.20}$$

(h4*) *$\mu = 0$, $\{\xi_k\}$ is a covariance stationary sequence with $(\gamma_k) \in G(d)$, $-1/2 < d < 1/2$; finite dimensional distributions of the process $X_n(t) = n^{-1/2-d} \sum_{j=1}^{[nt]+1} \xi_j$ converge to those of the process $X_0(\cdot) \ne 0$; and trend $g_n(j)$ satisfies*

$$\sum_{j=1}^{n} (g_n(j) - \bar{g}_n)^2 = o(n^{2d}). \tag{3.21}$$

The proof of the latter statement follows from the general theorem below which might be useful to show that the test can detect also higher order integrated alternatives $I(d)$, $d > 3/2$.

THEOREM 3.4 *Let $X_k, k \in \mathbf{Z}$ are random variables such that for some $\theta > -1/2$ process*

$$\widetilde{X}_n(t) = n^{-1/2-\theta}(X_{[nt]+1} - X_0), \quad 0 \le t \le 1$$

satisfies the following conditions:
(a) finite dimensional distributions of $\widetilde{X}_n(\cdot)$ converge to those of the process $X_0(\cdot) \ne 0$;
(b) uniformly in t and n,

$$\operatorname{Var} \widetilde{X}_n(t) \le C; \tag{3.22}$$

(c) for any $t \in [0,1]$, as $n \to \infty$, there exists the limit

$$\operatorname{Var} \widetilde{X}_n(t) \to h(t). \tag{3.23}$$

Then (3.19) *holds.*

Note that the limiting process $X_0(t)$ in theorems 3.3 and 3.4 can be non-Gaussian.

4 Simulation study

4.1 Large samples

In this subsection we present some results of a simulation study examining the finite sample performance of the proposed test. In all examples we assume that $\hat{d}$ is the local Whittle estimate, the minimization is carried out over $d \in [-0.4, 0.4]$ and $m = n^{0.9}$. The bandwidth q is equal to $n^{1/2}$.

We first study the empirical size of the test assuming that the data generating process under null hypothesis is FARIMA$(0, d, 0)$ where $d \in (-1/2, 1/2)$. The FARIMA$(0, d, 0)$ processes are simulated using the circulant matrix embedding method (see Bardet *et al.* (2002) for a review on the simulation of such processes). Figure 2 shows that the size is close to the nominal level α when the long memory parameter d is not too large, i.e. $d \leq 0.35$.

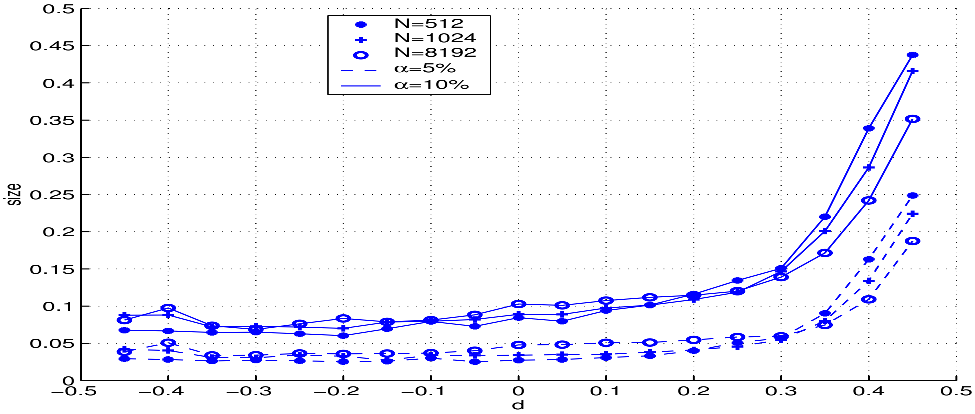

Figure 2: Empirical test size under FARIMA$(0, d, 0)$ model with iid standard normal innovations. The significance level $\alpha = 5\%$ and 10%, the sample size $n = 512,\ 1024,\ 8192$. Estimations are based on 5 000 independent replications.

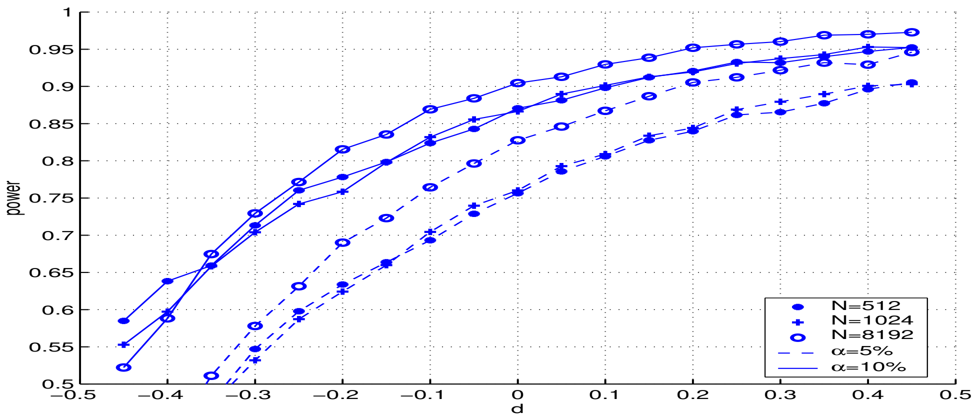

Figure 3: Empirical power of the test under unit root. The ξ_k are FARIMA(0,d,0) with iid standard normal innovations. The significance level $\alpha = 5\%$ and 10%, the sample size $n = 512,\ 1024,\ 8192$. Estimations are based on 5 000 independent replications.

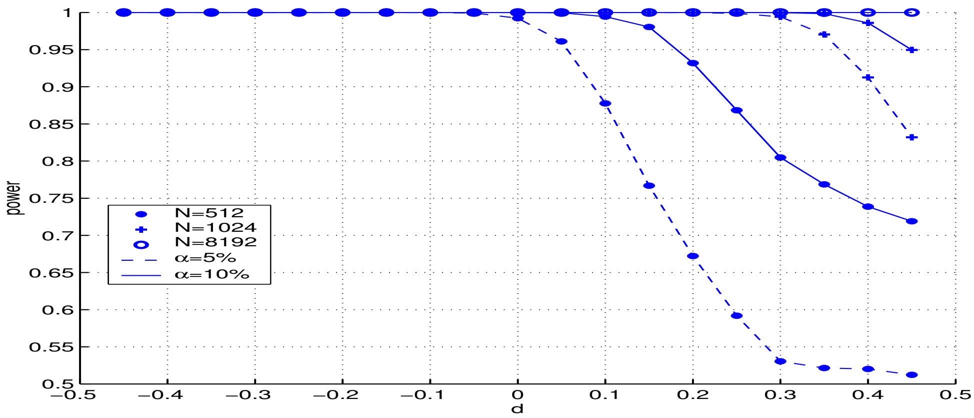

Figure 4: Power of the test under the linear trend model with $g(k) = ck$ and $c = 0.005$. The ξ_k are FARIMA(0,d,0). The significance level $\alpha = 5\%$ and 10%, the sample size $n = 512$, 1024, 8192. Estimations are based on 5 000 independent replications.

To illustrate the empirical power of the test we study several non-stationary models corresponding to alternative hypothesis. First example is a standard unit root model

$$X_k = X_{k-1} + \xi_k,$$

where the noise ξ_k is a FARIMA$(0, d, 0)$ process. Clearly, it is more difficult to detect the non-stationarity when the parameter d is negative (see Figure 3). However, we obtain rather efficient procedure and the power is greater than 0.75 at significance level 10% (respectively, 5%) for $d > -0.25$ (respectively, $d \geq 0$).

Figure 4 demonstrates the power in the case of linear trend model

$$X_k = ck + \xi_k,$$

where the ξ_k are FARIMA$(0, d, 0)$. As it is seen, we get very efficient procedure, with the power close to 1, for any $d \in [-0.4, 0.4]$ and even for very small c.

To illustrate the properties of the test in the case of non-parametric regression

$$X_k = G_n(k/n) + \xi_k$$

with FARIMA(0,d,0) errors ξ_k, we consider two examples of function $G_n(k/n)$:

(i) the change point model

$$G_n(k/n) = \begin{cases} 0 & \text{for } 1 \leq k \leq [n\tau], \\ \Delta & \text{for } [n\tau] < k \leq n; \end{cases}$$

(ii) the non-linear model $G_n(k/n) = c(k/n)^{\beta}$.

Figure 5 gives the empirical power in both cases. We see again that the procedure is less efficient when the memory parameter is close to 1/2. However, the results are satisfactory when $-0.4 \leq d \leq 0.35$. Moreover, for large sample size (here $n = 8192$) the power becomes greater than 0.75 for any $d \in [-0.4, 0.4]$.

Remark 4.1 Under the alternatives H_T or H_U, the rejection rate is approximately $(n/q)^{1-2a_2}$ (see the proof of Lemma 8.1), so that the test power significantly increases when we minimize over the region $[-a_1, a_2]$ with smaller a_2.

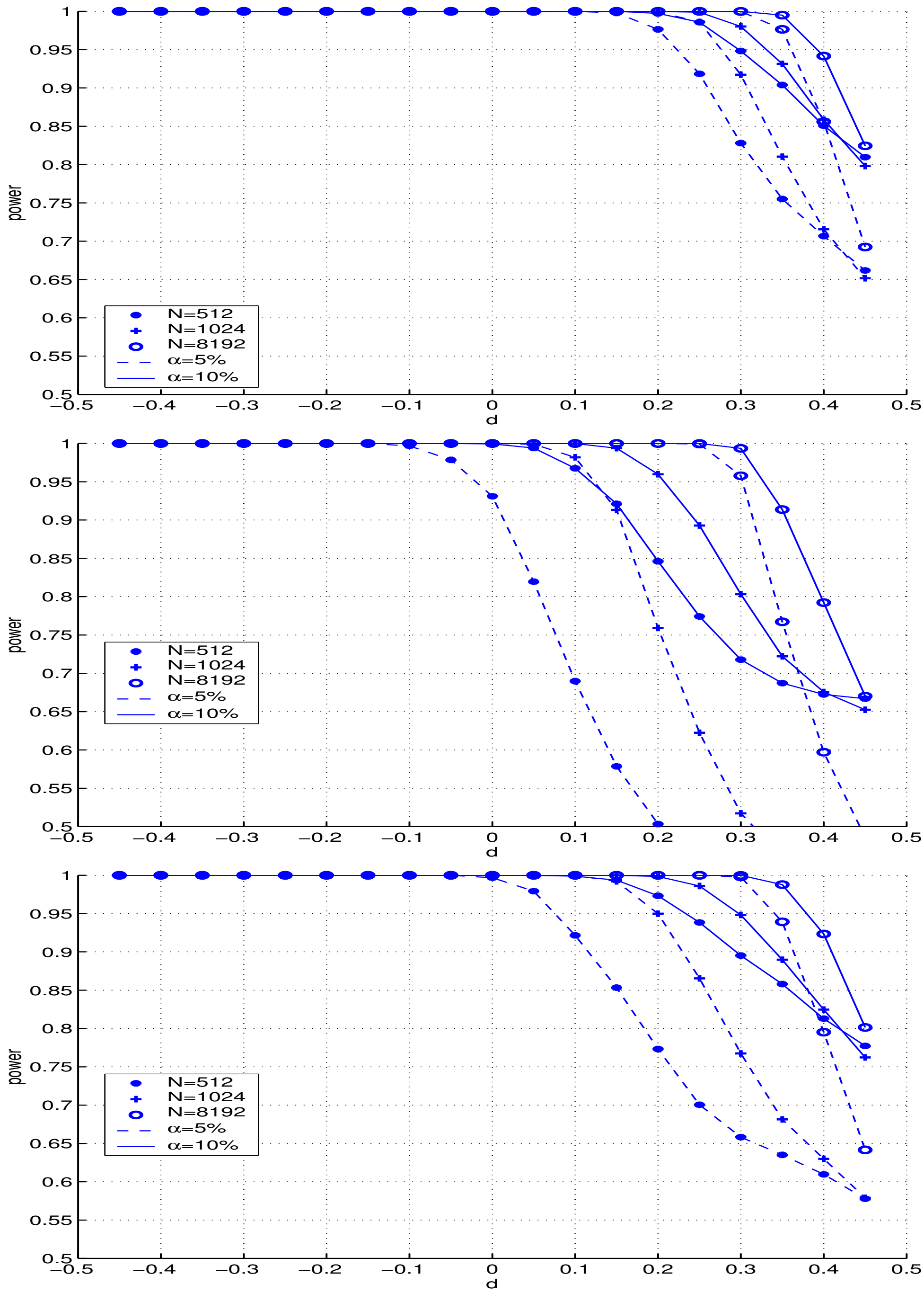

Figure 5: Power of the test under models of the form: $X_k = G_n(k/n)+\xi_k$. The top picture corresponds to the change point model (i) with $\Delta = 1.5$; the middle picture corresponds to the model (ii) with $c = 3$ and $\beta = 1/2$; the bottom picture corresponds to the (ii) with $c = 3$ and $\beta = 1$. The ξ_k are FARIMA(0,d,0). The significance level $\alpha = 5\%$ and 10%, the sample size $n = 512,\ 1024,\ 8192$. Estimations are based on 5 000 independent replications.

4.2 The short memory case $d = 0$

The most common assumption in the econometrical literature is the short memory (weak dependence) of the errors ξ_k:

$$\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty, \qquad \sum_{k=-\infty}^{\infty} \gamma_k = 2\pi f(0) \neq 0,$$

which corresponds to $d = 0$. We denote such null hypotheses of stationarity by H_0^*. We reject H_0^* if

$$T_n(0) > c_\alpha(0),$$

here $c_{5\%}(0) = 0.190, c_{10\%}(0) = 0.153$ (see Table 1).

It is natural to expect that in the case $d = 0$ (or more generally when d is known) the properties of the test are better than for unknown d. To illustrate this, consider the stationary short memory AR(1) model $X_k = aX_{k-1} + \varepsilon_k$. The empirical size of the test for different values of a is shown in Figure 6. We observe the distortion of the size when the parameter $a < 1$ approaches the non-stationarity region $a = 1$. The test is moderately conservative (undersized) for $300 \leq n \leq 500$ and has good size for large $n > 500$.

The power of the test (Figure 7) is illustrated on three types of alternatives:
(a) unit root model $X_k = X_{k-1} + \xi_k$ with AR(1) errors $\xi_k = a\xi_{k-1} + \varepsilon_k$;
(b) linear trend model $X_k = ck + \varepsilon_k$;
(c) the change point model $X_k = \Delta \mathbf{1}_{\{k > [n\tau]\}} + \varepsilon_k$.

Figure 7 demonstrates behavior of the power when the parameters a, c and Δ vary. We see that in all cases the test based on $T_n(0)$ has very good power properties. The first picture shows that in case of unit root the test is very powerful even for moderate samples sizes $n \geq 300$; the second picture shows that the presence of linear trend ck is effectively detected even for very small $c \geq 0.005$ in samples $n \geq 300$ and the power tends to 1 when c increases; the third picture shows that the small changes in the mean can be detected very efficiently when n increases (for instance, when $n \approx 500$ we can detect jumps of magnitude $\Delta \geq 0.35$ with probability 0.75).

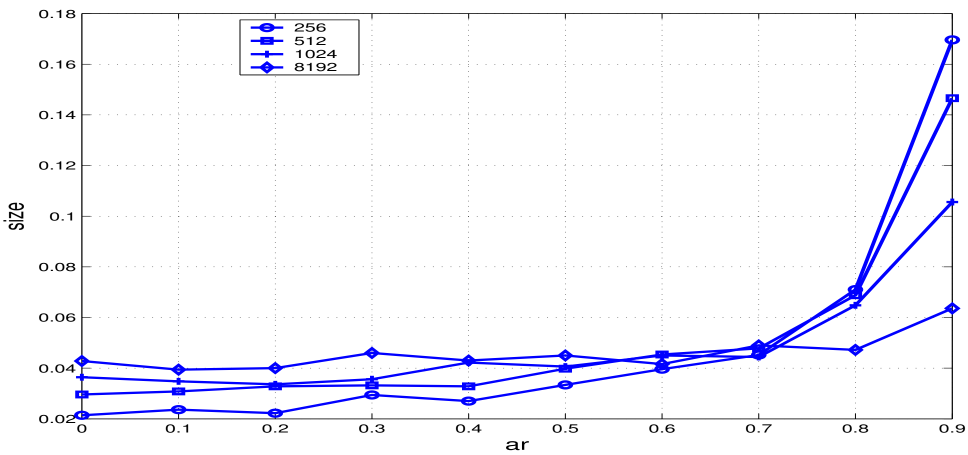

Figure 6: Empirical size of the test under AR(1) model $X_k = aX_{k-1} + \varepsilon_k$, where ε_k are iid $N(0, 1)$. The significance level $\alpha = 5\%$. Estimations are based on 5 000 independent replications.

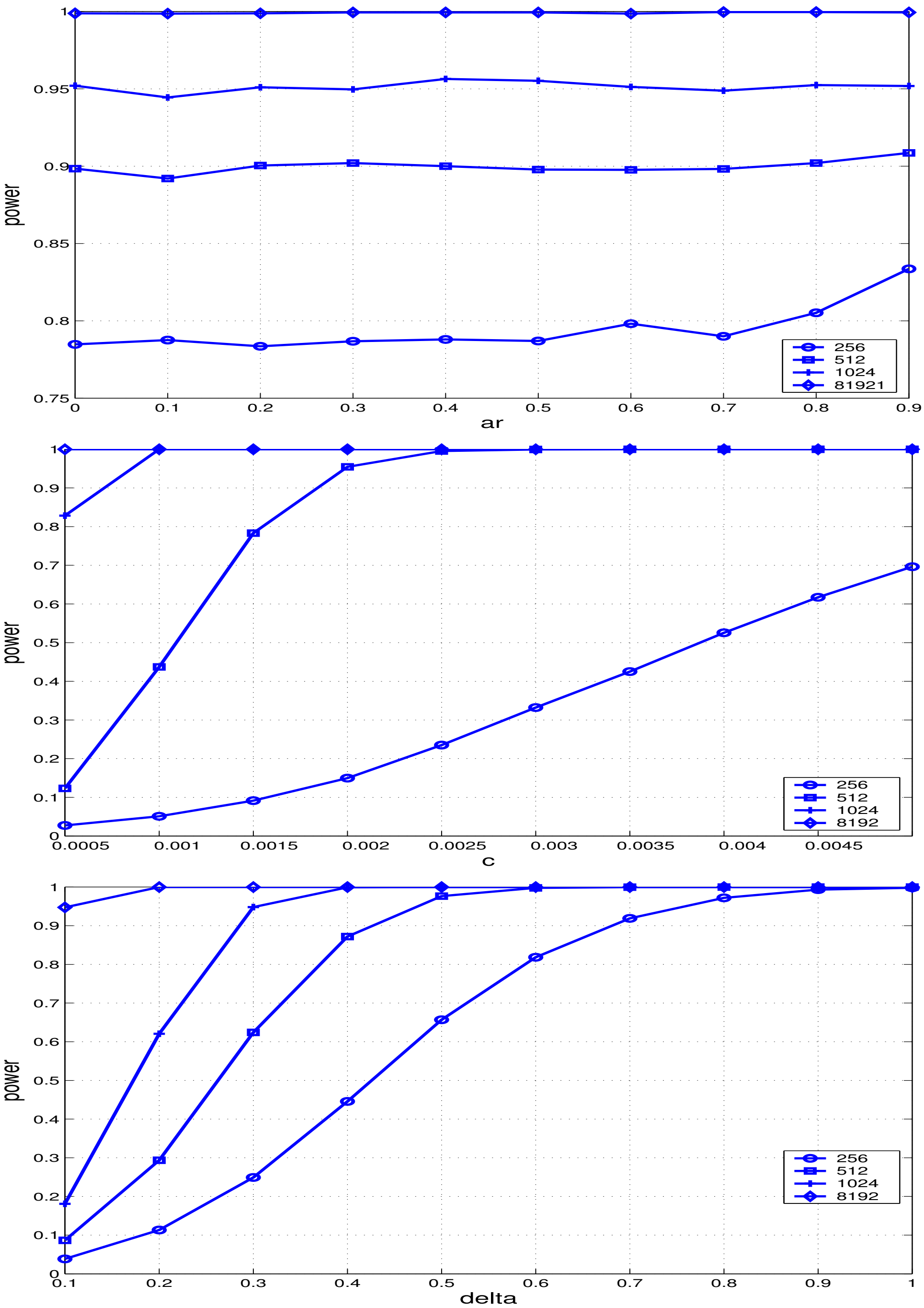

Figure 7: Empirical power of the test under unit root model $X_k = X_{k-1} + \xi_k$, where $\xi_k = a\xi_{k-1} + \varepsilon_k$ with a varying from 0 to 1 (top); linear trend model $X_k = ck + \varepsilon_k$ with c varying from 0.0005 to 0.005 (middle); change point model $X_k = \Delta \mathbf{1}_{\{k>[n/2]\}} + \varepsilon_k$ with Δ varying from 0.1 to 1 (bottom). The innovations ε_k are iid $N(0,1)$. Significance level $\alpha = 5\%$. Estimations are based on 5 000 independent replications.

Remark 4.2 Under hypothesis H_0^* ($d=0$), the interval $[-a_1, a_2]$ of the hypothetical memory parameter shrinks to a single point $a_1 = a_2 = 0$, so that stationary sequences $\{X_k\}$ with positive memory parameter $d > 0$ or negative $d < 0$ are not covered by H_0^*. To describe the behavior of the test in such cases, note that for $0 < d < 1/2$, $T_n(0) \geq O_P(1)(n/q)^d \to \infty$ (see the proof of Theorem 3.1); and thus, if $d > 0$, H_0^* is rejected. Formally, the stationary long memory case $d > 0$ could be considered as an alternative, but the power of the test for such alternative will be low if d approaches 0 (see Figure 8). For negative $-1/2 < d < 0$ it holds $T_n(0) = O_P(1)(n/q)^d \to 0$ and in such case the test accepts H_0^*.

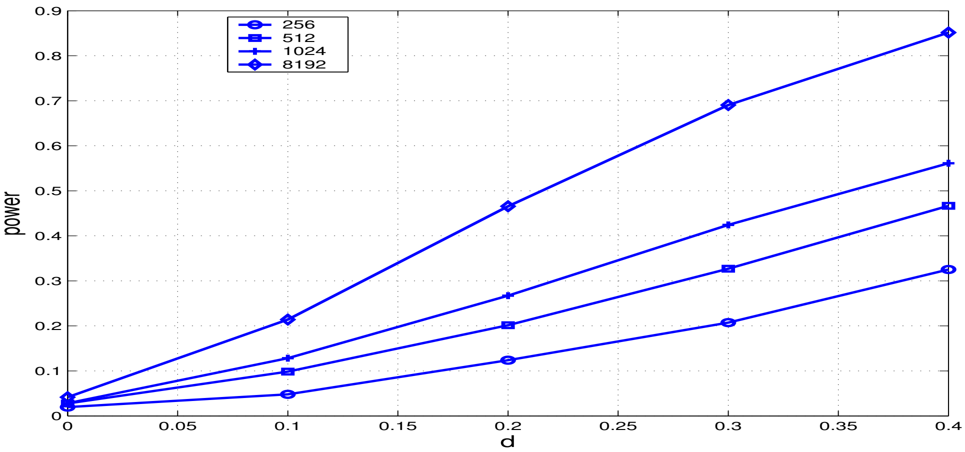

Figure 8: Empirical power of the test under FARIMA(0,d,0), d varying from 0 to 0.4. Innovations ε_k are iid $N(0,1)$. Significance level $\alpha = 5\%$. Estimations are based on 5 000 independent replications.

4.3 Testing in small samples

The semiparametric type testing procedure of Subsection 2.1 is applicable for a wide class of semiparametrically specified errors. It depends on the choice of bandwidth parameters m, q and includes the estimation of memory parameter d. We have seen in Subsection 4.1 that the test has a good size and power when the sample size n is large. Now we shall focus ourselves on the case of small samples. Since the test based on asymptotic quantiles $c_\alpha(d)$ is strongly undersized in small samples, see Figure 9, to improve the size properties we shall use the empirical quantiles $c_\alpha^{(n)}(d)$ calculated from M independent replications of the test statistic $T_n(d)$. We restrict the class of stationary processes under H_0 to the FARIMA(0,d,0) with iid N(0,1) innovations. Figure 9 presents the graphs of the size for $n = 50, \ldots, 500$, derived using quantiles $c_\alpha^{(n)}(d)$. As before, we set $m = n^{0.9}$, $q = \sqrt{n}$. The use of empirical quantiles gives improvement of the size in the interval $d \in [-0.2, 0.2]$. Figure 9 also shows that using asymptotic quantiles $c_{5\%}(d)$ the test is moderately conservative (undersized) in small samples ($300 \leq n \leq 500$) and therefore could be used for hypothesis testing in case when the errors might not have the FARIMA(0,d,0) structure. In such small samples restriction $d \in [-0.2, 0.2]$ under H_0 would be needed to avoid the distortion of the size.

Table 2 gives the coefficients of the polynomial approximation of quantiles $c_{5\%}^{(n)}(d)$ corrected for the small samples for different values of d. These quantiles should be used for the hypothesis testing in small samples, assuming that the errors ξ_k under null hypothesis follow

Table 2: Coefficients a_i, $i = 0, \ldots, 5$ in the polynomial approximation of the quantiles $c^{(n)}_{5\%}(d) = a_0 + a_1 d + a_2 d^2 + a_3 d^3 + a_4 d^4 + a_5 d^5$. The values are based on 10 000 replications of the FARIMA(0,d,0) with iid standard normal innovations.

n	a_0	a_1	a_2	a_3	a_4	a_5
50	0.145	-0.509	0.827	-0.885	1.086	-1.056
100	0.154	-0.564	0.909	-0.932	1.210	-1.272
150	0.162	-0.595	0.833	-0.599	1.765	-3.016
200	0.163	-0.596	0.926	-1.051	1.353	-0.955
300	0.169	-0.620	0.868	-0.725	1.696	-2.442
400	0.171	-0.624	0.855	-0.812	1.959	-2.504
500	0.173	-0.633	0.789	-0.554	2.271	-3.701

the FARIMA(0,d,0) model. If the value of statistic $T_n(\hat{d})$ is greater than the quantile $c^{(n)}_{5\%}(\hat{d})$ (calculated from the polynomial approximation), the null hypothesis is rejected; it is accepted otherwise. Note that to keep the size close to 5% for small samples ($150 \leq n \leq 500$), H_0 should include the memory parameter from a narrow interval $[-a_1, a_2]$ centered around zero, say, $[-a_1, a_2] = [-0.2, 0.2]$. The empirical quantiles $c^{(n)}_{5\%}(d)$ are adjusted only to the FARIMA(0,d,0) with $d \in [-0.2, 0.2]$ and should not be used for a wider class of stationary FARIMA(p, d, q) models with $p \geq 1$ or $q \geq 1$ since that leads to the distortion of the size.

The first picture of the Figure 10 presents the power of the test in the case of linear trend $X_k = ck + \varepsilon_k$ with $c = 0.05$. In general, the simulations show that when c increases the power of the test tends to 1 for any $d \in [-0.2, 0.2]$ and for any sample size $n \geq 150$, i.e. the test detects linear trend very efficiently. The second picture shows that in the case of change point of the magnitude $\Delta = 1.5$ such non-stationarity can be detected with a power $\geq 75\%$ for any memory parameter $d \in [-0.2, 0.2]$ in rather small samples $n \geq 150$.

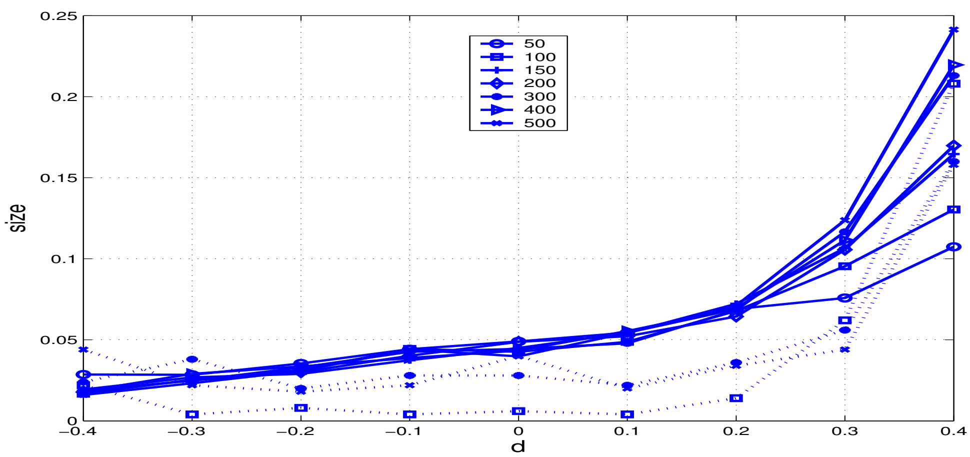

Figure 9: Empirical test size under FARIMA$(0, d, 0)$ model with iid standard normal innovations. The solid (resp., dashed) lines show the size of the test based on empirical quantiles $c^{(n)}_{5\%}(d)$ (asymptotic quantiles $c_{5\%}(d)$). The sample sizes $n = 50, \ldots, 500$. Estimations are based on 5 000 independent replications.

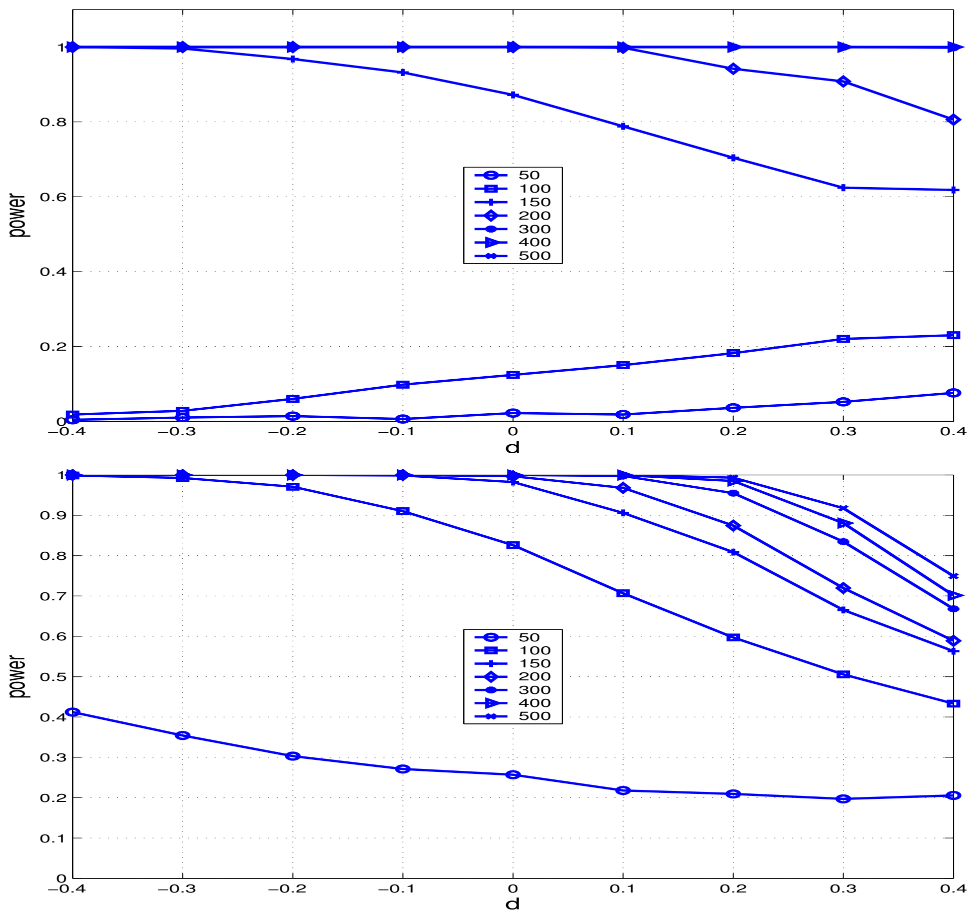

Figure 10: Empirical test power for linear trend model $X_k = ck + \varepsilon_k$ with $c = 0.05$ (top) and change point model $X_k = \Delta \mathbf{1}_{\{k>[n/2]\}} + \varepsilon_k$ with $\Delta = 1.5$ (bottom). ε_k are iid standard normal variables, significance level $\alpha = 5\%$, the sample sizes $n = 50, \ldots, 500$. Estimations are based on 5 000 independent replications.

5 Empirical applications

In this section we apply the theory for two data sets, the Nile river data and S&P 500 daily returns.

Nile river data. The data set consists of 660 subsequent annual minimum level data begining from the year 622; it was examined by Hurst (1951) and led to the famous phenomenon called the "Hurst effect" and invention of the R/S statistic. In a number of subsequent studies (see Mandelbrot and Wallis (1968), Bhattacharya *et al.* (1983), Giraitis *et al.* (2001), among others) different explanations of the Hurst effect were provided, such as stationary long memory, time-varying mean, piece-wise stationarity. Figure 11 presents the graphs of the original time series and the sample autocorrelation function at the first 50 lags which exhibits a slow decay.

Table 3 shows that in the case when the short memory hypothesis of stationarity H_0^* (with $d = 0$) is tested, then $T_n(0) = 0.290$ is greater than $c_\alpha(0) = 0.190$, and the null hypothesis is rejected. However, the local Whittle estimate $\hat{d} = 0.367$ of the memory parameter d indicates that the assumption $d = 0$ is hardly justified and d should be estimated. Testing hypothesis H_0 with unknown memory parameter d the value of the statistic $T_n(\hat{d}) = 0.027$ is smaller than

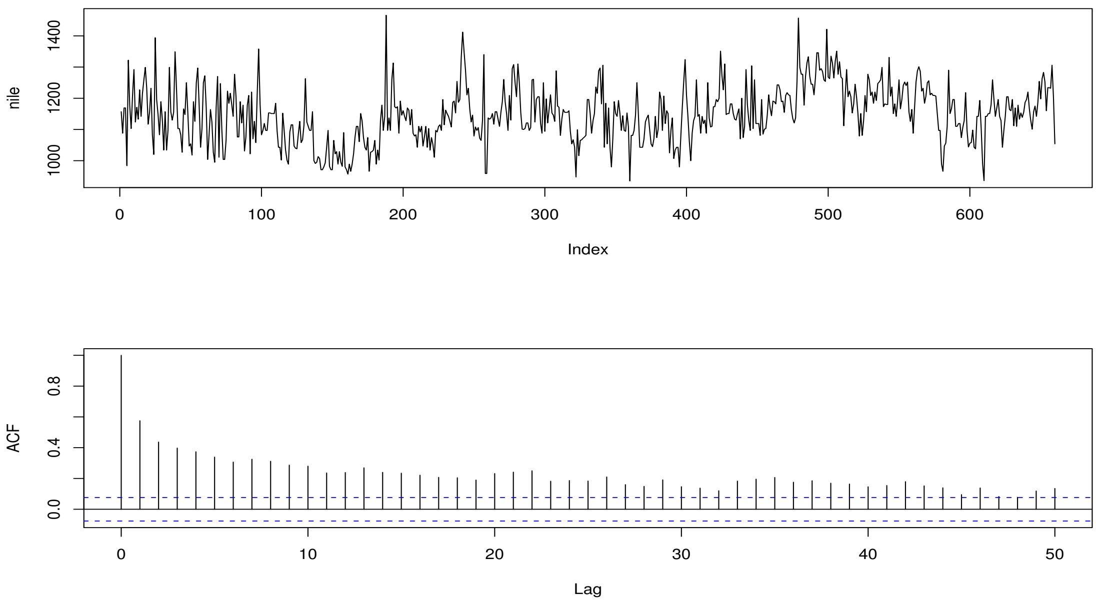

Figure 11: Nile river data: the original 660 annual data and autocorrelation function.

quantile $c_{5\%}(\hat{d}) = 0.030$. (Recall that $q = \sqrt{n}$, $m = n^{0.9}$.) Thus, testing procedure shows that, with 5% significance level, the Nile river data can be described by the stationary long memory model with large enough d.

Table 3: Values of the test statistic and quantiles for the Nile data. H_0^* corresponds to the stationary short memory model $d = 0$ (first row) and H_0 corresponds to the stationary model with unknown memory parameter d (second row). Significance level $\alpha = 5\%$.

	$T_n(d)$	$c_\alpha(d)$	Decision
$d = 0$	0.290	0.190	reject H_0^*
$\hat{d} = 0.367$	0.027	0.030	accept H_0

S&P 500 daily returns. Figure 12 presents the Standard and Poor's 500 Index daily closing values from year 1928 to 1993 and their log returns.

It is well known that, for many financial data, the returns itself are almost uncorrelated whereas their transformations such as squares or absolute returns can be significantly correlated, see, e.g., Ding and Granger (1996). This indicates that the returns do not behave as independent random variables and hence, to investigate the nature of this phenomenon, it is useful to examine the dependence structure of the transformed returns, e.g. $|r_k|, r_k^2$. Since the presence of the long memory is rather likely, the stationarity test for the hypothesis H_0 with unknown d should be applied. The graphs of corresponding autocorrelations are shown in Figure 13 and the statistical results are presented in Table 4.

The results confirm the empirical findings of stationarity for returns r_k, whereas for squares and absolute values the null hypothesis H_0 with unknown d is rejected (since the statistic $T_n(\hat{d})$

Table 4: Values of the test statistic and quantiles for the S&P 500 daily returns, squares, absolute values and 3rd powers of returns, 1/3/28–5/28/93. H_0 corresponds to the stationary model with unknown memory parameter d. Significance level $\alpha = 5\%$.

	$T_n(\hat{d})$	$c_\alpha(\hat{d})$	Decision
	returns r_k		
$\hat{d} = 0.017$	0.053	0.179	accept H_0
	squares r_k^2		
$\hat{d} = 0.198$	0.130	0.084	reject H_0
	absolute values $\lvert r_k \rvert$		
$\hat{d} = 0.216$	0.146	0.077	reject H_0
	3rd powers r_k^3		
$\hat{d} = -0.019$	0.079	0.203	accept H_0

takes greater values than the quantile $c_{5\%}(\hat{d})$). It is interesting to point out that, similarly to r_k, for an (odd) 3rd power r_k^3 the stationarity hypothesis H_0 is not rejected and the small value of $\hat{d} = -0.019$ does not indicate the presence of the long memory.

Our statistical procedure shows that the absolute powers of returns, $|r_k|^\delta, \delta \geq 1$ do not follow stationary model. This finding together with a number of other researches (see, e.g., Lobato and Savin (1998), Mikosch and Stărică (1999), where authors provide arguments of the "spurious long memory" phenomenon, such as structural changes in data, monotonic trends, non-existence of higher order moments, etc.) confirms the phenomenon of "spurious" stationarity (and long memory) in powers $|r_k|^\delta$.

Finally, as it should be expected, our test for stationarity shows that the initial series of prices p_t (Figure 12) and log-prices $\log p_k$ do not possess stationarity property and hypothesis H_0 is rejected in both cases. Moreover, since the stationarity of r_k is not rejected (recall that $r_k = \log p_k - \log p_{k-1}$), $\log p_k$ is a nice empirical example of unit root series. Application of the test shows that, in case of p_k, $T_n(\hat{d}) = 0.033$, $c_{5\%}(\hat{d}) = 0.022$ and H_0 is rejected; similarly in case of $\log p_k$, $T_n(\hat{d}) = 0.049$, $c_{5\%}(\hat{d}) = 0.022$ and H_0 is rejected too.

6 Weak convergence in space $L_2[0, 1]$

Most of the classical weak convergence studies of partial sum processes, empirical processes, quantile processes etc. are conducted in the space $D[0, 1]$ using the Skorokhod topology framework. However, investigation of the asymptotic properties of the KPSS, V/S and many classical integral-type statistics, such as the Cramér–von Mises ω^2 statistic, the Watson statistic, the Anderson–Darling statistic require weaker, only L_2 or L_p, topology on the space of paths (see Billingsley (1968, page 123, Problem 2)). Sufficient conditions of the weak convergence in $L_2[0, 1]$, formulated below, are less restrictive and easier to verify than those in the space $D[0, 1]$. Note that $L_2[0, 1]$ topology is not applicable for the maximum type Kolmogorov–Smirnov, Kuiper, R/S and other statistics where convergence in $D[0, 1]$ or $C[0, 1]$ should be used.

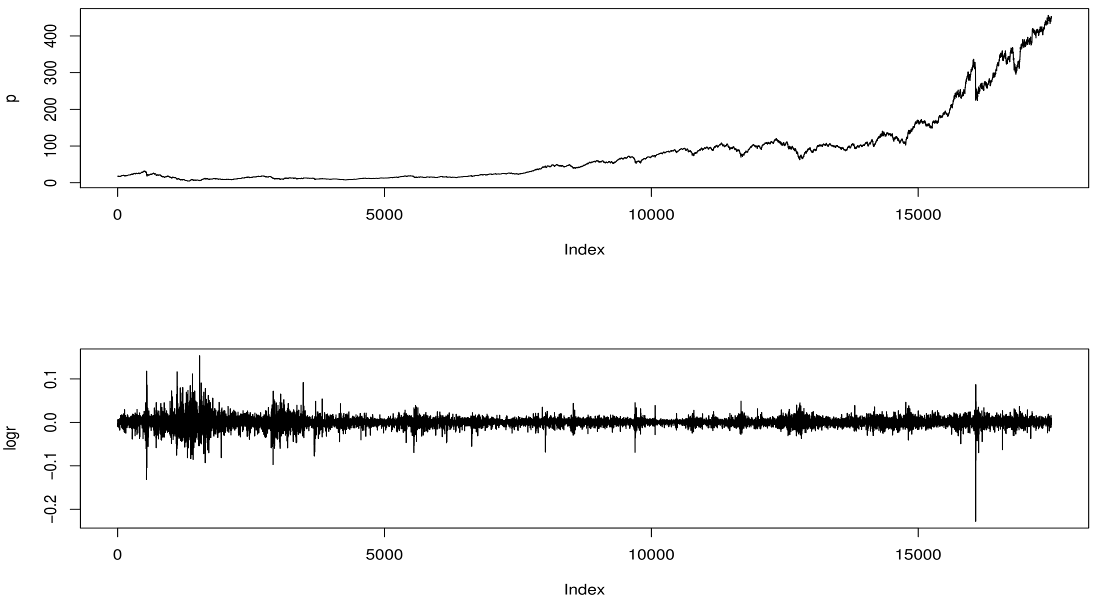

Figure 12: S&P 500 Index closing values and their log returns.

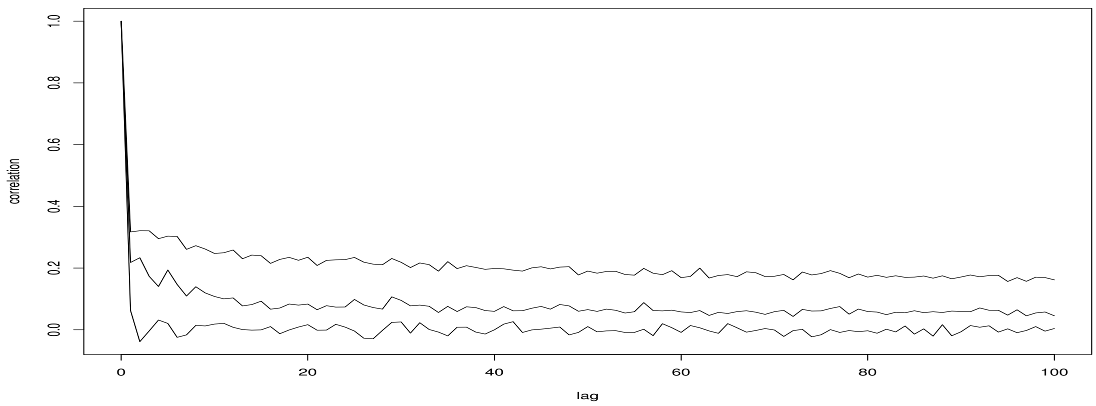

Figure 13: Autocorrelations of S&P 500 daily returns, squares of returns and absolute values of returns (from low to high).

We formulate the criteria of weak convergence in the space $L_2[0,1]$, which presents a useful sufficient condition for the convergence of integrals and sums

$$\int_0^1 X_n^2(t)dt \xrightarrow{D} \int_0^1 X_0^2(t)dt, \quad n^{-1}\sum_{k=1}^n X_n^2\Big(\frac{k}{n}\Big) \xrightarrow{D} \int_0^1 X_0^2(t)dt$$

and other functionals of $X_n(t)$, $0 \le t \le 1$ and follows from the weak convergence criterion in the space $L_p[0,1]$, $p \ge 1$ of Cremers and Kadelka (1986).

THEOREM 6.1 *Assume that $\{X_n(t), t \in [0,1]\}$, $n \geq 0$ is a sequence of real valued measurable processes with the paths from the space $L_2[0,1]$, i.e.*

$$\int_0^1 X_n^2(t)dt < \infty. \tag{6.1}$$

Suppose that the following conditions are satisfied:

(i) finite dimensional distributions of the processes X_n converge to those of the process X_0;

(ii) there exists nonnegative function f satisfying $\int_0^1 f(t)dt < \infty$ such that for any $t \in [0,1]$ and $n \geq 1$

$$EX_n^2(t) \leq f(t);$$

(iii) for any t,

$$EX_n^2(t) \to EX_0^2(t) \quad (n \to \infty).$$

Then X_n weakly converges to X_0 in space $L_2[0,1]$, i.e. for any continuous real valued bounded functional $g : L_2[0,1] \to \mathbf{R}$, it holds, as $n \to \infty$,

$$Eg(X_n) \to Eg(X_0). \tag{6.2}$$

In this case, for any continuous function $F : L_2[0,1] \to L_2[0,1]$, $F(X_n(\cdot))$ weakly converges to $F(X_0(\cdot))$.

Following Cremers and Kadelka (1986), we call such converge *a weak convergence in* $L_2[0,1]$ and write $X_n \stackrel{L_2[0,1]}{\longrightarrow} X_0$. Analogous result can be derived for the spaces $L_p[0,1]$, $1 \leq p < \infty$ (see Cremers and Kadelka (1986)).

REMARK 6.1 Theorem 6.1 implies the convergence

$$F(X_n(\cdot)) \xrightarrow{D} F(X_0(\cdot)), \tag{6.3}$$

for continuous functionals $F : L_2[0,1] \to \mathbf{R}$ of paths $X_n(\cdot)$ such as

$$\int_0^1 X_n^2(t)dt, \quad \int_0^1 X_n(t)dt, \tag{6.4}$$

$$\int_0^1 \Big(X_n(t) - tX_n(1) - \int_0^1 \Big(X_n(u) - uX_n(1)\Big)du\Big)^2 dt, \tag{6.5}$$

$$\int_0^1 \Big(X_n(t) - t\int_0^1 X_n(s)ds - \int_0^1 \Big(X_n(u) - u\int_0^1 X_n(s)ds\Big)du\Big)^2 dt. \tag{6.6}$$

(For more continuous functionals in $L_p[0,1]$ and related discussion see Oliveira and Suquet (1998).)

Convergence of the sums

$$n^{-1}\sum_{k=1}^n X_n\Big(\frac{k}{n}\Big) \xrightarrow{D} \int_0^1 X_0(t)dt, \quad n^{-1}\sum_{k=1}^n X_n^2\Big(\frac{k}{n}\Big) \xrightarrow{D} \int_0^1 X_0^2(t)dt. \tag{6.7}$$

follows under additional unrestrictive regularity assumptions. The criterion of the convergence of general transformations $n^{-1}\sum_{k=1}^n T(n^{-1/2}X_k)$ was studied by Park and Phillips (1999) and Pötcher (2001).

REMARK 6.2 The assumption of measurability of the processes $X_n(t)$ is not restrictive and is needed to assure that integral (6.1) is correctly defined. In fact, to ensure the measurability it is enough to consider the *stochastically continuous* processes. (Recall that random process X is stochastically continuous if for any t and for any $\epsilon > 0$, $P\{|X(t) - X(s)| > \epsilon\} \to 0$ as $s \to t$.)[1] Obviously, the process X with finite second moments is stochastically continuous if for any t, as $s \to t$, $E(X(t) - X(s))^2 \to 0$. Hence, if X is weakly stationary process with variance σ_X^2 and correlation function $\rho_X(\tau)$ then

$$E(X(t) - X(s))^2 = 2\sigma_X^2(1 - \rho_X(t - s)),$$

which tends to zero if and only if $\rho_X(\tau) \to 1$ as $\tau \to 0$, i.e. $\rho_X(\tau)$ is continuous at $\tau = 0$. The stochastic continuity property is satisfied by the partial sum process $\sum_{j=1}^{[nt]} \xi_j$ (as soon as the ξ_j are non-degenerate random variables) and by fractional Brownian motion with parameter $H > 0$.

The numerator V_n, (2.13), appearing in the test statistic (2.15), is the function of the form (6.5). The following proposition provides its limit under rather unrestrictive assumptions.

PROPOSITION 6.1 *Suppose that random variables $X_1, \dots, X_n$ are such that for some $\theta > -1/2$ and $\mu \in \mathbf{R}$, the normalized sums*

$$X_n(t) = n^{-1/2-\theta} \sum_{j=1}^{[nt]+1} (X_j - \mu), \quad (0 \le t \le 1), \quad n \ge 1 \tag{6.8}$$

satisfy assumptions (i), (ii), (iii) of Theorem 6.1. Then, as $n \to \infty$,

$$n^{-2\theta} V_n \xrightarrow{D} V(X_0) \equiv \int_0^1 \Big(X_0(t) - tX_0(1) - \int_0^1 \Big(X_0(u) - uX_0(1) \Big) du \Big)^2 dt. \tag{6.9}$$

Note that the limit process $X_0(u)$ (6.9) which is defined by assumptions (*i*)–(*iii*) can be non-Gaussian.

Acknowledgement. The authors would like to thank Alfredas Račkauskas for drawing our attention to the criterion of Cremers and Kadelka (1986).

7 Appendix A

PROOF OF LEMMA 2.1. (2.23) follows if we show that the mapping $\nu \mapsto c_\alpha(\nu)$, $\nu \in [-a_1, a_2]$ is continuous. Then,

$$E|c_\alpha(\hat{d}) - c_\alpha(d)| = \int_{-a_1}^{a_2} |c_\alpha(\nu) - c_\alpha(d)| F_{\hat{d}}(d\nu) \to 0,$$

where $F_{\hat{d}}$ is a distribution of the estimator $\hat{d}$, to prove (2.23).

[1] Indeed, since $[0, 1]$ is a compact and $\mathbf{R}$ is locally compact, stochastic continuity of processes (see Theorem 1 in Chapter III, §3 of Gikhman and Skorokhod (1980)) implies that there exists stochastically equivalent $\mathcal{F} \otimes \mathcal{B} - \mathcal{B}(\mathbf{R})$ measurable process $\widetilde{X}_n : \Omega \times [0, 1] \to \mathbf{R}$ such that for any n and t it holds $P\{X_n(t, \omega) \neq \widetilde{X}_n(t, \omega)\} = 0$. It remains to observe that the finite dimensional distributions of the two stochastically equivalent processes X_n and $\widetilde{X}_n$ coincide.

Decompose this mapping as $\nu \mapsto \mathcal{L}(Z_\nu) \mapsto c_\alpha(\nu)$ ($\mathcal{L}(X)$ denotes the distribution of random variable X). To show that mapping $\mathcal{L}(Z_\nu) \mapsto c_\alpha(\nu)$ is continuous, it is enough to note that the quantile of a continuous strictly monotonic distribution function is continuous functional in uniform metric, i.e., if c_α^F denotes the quantile of a distribution F, then $c_\alpha^{F_n} \to c_\alpha^F$ whenever $\sup_t |F_n(t) - F(t)| \to 0$. Here we set $F_n = F_{Z_{\nu_n}}$, $F = F_{Z_\nu}$ and notice that, since $F(t)$ is continuous, $Z_{\nu_n} \xrightarrow{D} Z_\nu$ if and only if $\sup_t |F_n(t) - F(t)| \to 0$.

Now we prove that

$$Z_{\nu_n} \xrightarrow{D} Z_\nu \quad \text{as} \quad \nu_n \to \nu, \tag{7.1}$$

i.e. $\nu \mapsto \mathcal{L}(Z_\nu)$ is continuous. Set $Y_\nu(t) := W^0_{1/2+\nu}(t) - \int_0^1 W^0_{1/2+\nu}(s)ds$. Then $Z_\nu = \int_0^1 Y_\nu^2(t)dt$ and

$$E|Z_{\nu_n} - Z_\nu| \le \int_0^1 E|Y^2_{\nu_n}(t) - Y^2_\nu(t)|dt$$

$$\le \int_0^1 [E(Y_{\nu_n}(t) - Y_\nu(t))^2]^{1/2}[E(Y_{\nu_n}(t) + Y_\nu(t))^2]^{1/2}dt \to 0 \quad \text{as} \quad \nu_n \to \nu,$$

because from definition of the fractional Brownian motion, using straightforward calculations it follows that $g(y) = E(Y_y(t) - Y_\nu(t))^2$, $y \in [-a_1, a_2]$ is a continuous bounded function, $g(\nu) = 0$, and

$$E(Y_{\nu_n}(t) + Y_\nu(t))^2 \le 2(EY^2_{\nu_n}(t) + EY^2_\nu(t)) \le C$$

uniformly in ν_n, ν, to prove (7.1). □

Proof of Theorem 3.1. We show that

$$n^{-2d} V_n \xrightarrow{D} s_d^2 Z_d \tag{7.2}$$

and

$$q^{-2d} \hat{s}^2_{n,q} \xrightarrow{P} s_d^2. \tag{7.3}$$

Since under assumption (3.3), $(q/n)^{2\hat{d}-2d} = 1 + o_P(1)$ this implies

$$T_n(\hat{d}) = \left(\frac{q}{n}\right)^{2\hat{d}} \frac{V_n}{\hat{s}^2_{n,q}} \xrightarrow{D} Z_d$$

to prove (3.5). This and Lemma 2.1 imply that $T_n(\hat{d}) - c_\alpha(\hat{d}) \xrightarrow{D} Z_d - c_\alpha(d)$ and hence

$$P\{T_n(\hat{d}) > c_\alpha(\hat{d})\} \to P\{Z_d > c_\alpha(d)\}$$

since Z_d has a continuous distribution function.

To prove (7.2) note that, by assumption (a1) and Lemma 8.2, it follows that the process

$$X_n(t) = n^{-1/2-d} \sum_{j=1}^{[nt]+1} \xi_j$$

satisfies assumptions (i)–(iii) of Theorem 6.1 with $X_0(t) = s_d W_{1/2+d}(t)$ and therefore (7.2) holds by Proposition 6.1.

From (3.4) and Lemma 8.2 it follows that

$$q^{-2d}\hat{s}^2_{n,q} \sim q^{-2d-1} \sum_{i,j=1}^{q} \gamma_{i-j} \sim s_d^2,$$

to prove (7.3). ☐

PROOF OF PROPOSITION 3.1. Similarly as in the proof of Theorem 3.1 it suffices to establish (7.2) and (7.3). (7.3) is shown in Lemma 8.3. To show (7.2) note that condition (3.7) implies

$$\sum_{j=1}^{[nt]+1} |g_n(j) - \bar{g}_n| \le n^{1/2} (\sum_{j=1}^{n} (g_n(j) - \bar{g}_n)^2)^{1/2} = o(n^{1/2+d}).$$

Therefore

$$\begin{aligned} n^{-1/2-d} \sum_{j=1}^{[nt]+1} (X_j - \bar{X}_n) &= n^{-1/2-d} \sum_{j=1}^{[nt]+1} (\xi_j - \bar{\xi}_n) + n^{-1/2-d} \sum_{j=1}^{[nt]+1} (g_n(j) - \bar{g}_n) \\ &= n^{-1/2-d} \sum_{j=1}^{[nt]+1} (\xi_j - \bar{\xi}_n) + o(1) \end{aligned}$$

and (7.2) follows by the same argument as in the proof of Theorem 3.1. ☐

PROOF OF PROPOSITION 6.1. To see (6.9) note that, by definition,

$$V_n = n^{-2} \sum_{k=1}^{n} (S_k^* - n^{-1} \sum_{j=1}^{n} S_k^*)^2 = n^{-1} \int_0^1 (S_{[nt]+1}^* - \int_0^1 S_{[nu]+1}^* du)^2 dt, \tag{7.4}$$

where $S_k^* := \sum_{i=1}^k (X_i - \bar{X}_n) = \sum_{i=1}^k X_i - k\bar{X}_n$. As $n^{-1/2-\theta} S_k^* = X_n((k-1)/n) - ((k-1)/n) X_n(1)$, from (7.4) we have

$$n^{-2\theta} V_n = \int_0^1 \left(X_n\Big(\frac{[nt]}{n}\Big) - \frac{[nt]}{n} X_n(1) - \int_0^1 \Big(X_n\Big(\frac{[nu]}{n}\Big) - \frac{[nu]}{n} X_n(1) \Big) du \right)^2 dt. \tag{7.5}$$

Hence, by Remark 6.1,

$$n^{-2\theta} V_n \xrightarrow{D} \int_0^1 \Big(X_0(t) - t X_0(1) - \int_0^1 (X_0(u) - u X_0(1)) du \Big)^2 dt.$$

☐

8 Appendix B

The proofs of Theorems 3.2–3.4 are based on the following lemma.

LEMMA 8.1 *Let $T_n(\hat{d})$ be given in* (2.15), *where $\hat{d} \in [-a_1, a_2]$, $0 \le a_1, a_2 < 1/2$. If for some $\theta > -1/2$*

$$n^{-1-2\theta} V_n \xrightarrow{D} V_0, \tag{8.1}$$

where $V_0 > 0$ almost surely, and

$$E\hat{s}_{n,q}^2 \le C q n^{2\theta}, \tag{8.2}$$

then

$$T_n(\hat{d}) \xrightarrow{D} \infty. \tag{8.3}$$

Proof of Lemma 8.1. Write

$$T_n(\hat{d}) = \left(\frac{n}{q}\right)^{1-2\hat{d}} \frac{n^{-1-2\theta}V_n}{(n^{2\theta}q)^{-1}\hat{s}_{n,q}^2}.$$

Then, as $\hat{d} < a_2 < 1/2$,

$$\left(\frac{n}{q}\right)^{1-2\hat{d}} \geq \left(\frac{n}{q}\right)^{1-2a_2} =: \alpha_n \to \infty. \tag{8.4}$$

By (8.2),

$$P\left\{(qn^{2\theta})^{-1}\hat{s}_{n,q}^2 > \alpha_n^{1/2}\right\} \leq \alpha_n^{-1/2}(qn^{2\theta})^{-1}E\hat{s}_{n,q}^2 \to 0.$$

Using inequality (8.4) it follows that on the set $(qn^{2\theta})^{-1}\hat{s}_{n,q}^2 \leq \alpha_n^{1/2}$,

$$T_n(\hat{d}) \geq \left(\frac{n}{q}\right)^{1-2\hat{d}} \frac{n^{-1-2\theta}V_n}{\alpha_n^{1/2}} \geq \alpha_n^{1/2}n^{-1-2\theta}V_n.$$

Hence, for any $K > 0$,

$$\begin{aligned} P\{T_n(\hat{d}) > K\} &= P\{T_n(\hat{d}) > K, (qn^{2\theta})^{-1}\hat{s}_{n,q}^2 \leq \alpha_n^{1/2}\} + P\{T_n(\hat{d}) > K, (qn^{2\theta})^{-1}\hat{s}_{n,q}^2 > \alpha_n^{1/2}\} \\ &\geq P\{\alpha_n^{1/2}n^{-1-2\theta}V_n > K\} + o(1) \to 1 \end{aligned}$$

by (8.1), to prove (8.3). □

Proof of Theorem 3.2. (3.13) follows by Lemma 8.1 if we show that

$$n^{-1-2\gamma}V_n \xrightarrow{D} V_0 > 0 \tag{8.5}$$

and

$$E\hat{s}_{n,q}^2 \leq Cqn^{2\gamma}. \tag{8.6}$$

Set

$$X_n(t) = n^{-1-\gamma}\sum_{j=1}^{[nt]+1}(X_j - \mu), \quad G_0(t) = \int_0^t g_*(u)du,$$

where $g_*(u)$ is defined by Assumption T(γ). We shall show that $X_n(t)$ satisfy assumptions (*i*)–(*iii*) of Theorem 6.1 with $X_0(t) \equiv G_0(t)$ which imply

$$n^{-1-2\gamma}V_n \to V(G_0) > 0,$$

where $V(\cdot)$ is given by (6.9). Write

$$X_n(t) = n^{-1-\gamma}\sum_{j=1}^{[nt]} g_n(j) + n^{-1-\gamma}\sum_{j=1}^{[nt]}\xi_j =: X_{n,g}(t) + X_{n,\xi}(t). \tag{8.7}$$

Under Assumption T(γ), the dominated convergence theorem implies

$$X_{n,g}(t) = \int_0^{([nt]+1)/n} n^{-\gamma}g_n([nu])du \to \int_0^t g_*(u)du = G_0(t).$$

Clearly, $X_{n,g}(t)$ satisfies assumptions (i)–(iii) of Theorem 6.1 with $X_0(\cdot) = G_0(\cdot)$. It remains to verify that the second term $X_{n,\xi}(t)$ in (8.7) satisfies these assumptions with the limit $X_0(t) \equiv 0$. To show this we prove that

$$EX^2_{n,\xi}(t) \to 0 \tag{8.8}$$

uniformly in t. In cases (h1), (h2) it holds $\gamma \geq 0$, and therefore

$$EX^2_{n,\xi}(t) = n^{-2-2\gamma} E(\sum_{j=1}^{[nt]} \xi_j)^2 \to 0 \tag{8.9}$$

by assumption (3.12) in case (h1) and by (3.14) in case (h2).

In case (h3), $(\gamma_k) \in G(d)$ and therefore, by (8.16) of Lemma 8.2, $EX^2_{n,\xi}(t) = n^{-2-2\gamma} \sum_{i,j=1}^{[nt]} \gamma_{i-j} \sim C[nt]^{1+2d} n^{-2-2\gamma}$. Hence,

$$EX^2_{n,\xi}(t) \leq Cn^{2d-2\gamma-1} \to 0$$

by assumption (3.15). Thus, in all three cases, (h1)–(h3), (8.8) holds and therefore $X_n(t)$, (8.7), satisfies assumptions (i)–(iii) of Theorem 6.1 with $X_0(t) = G_0(t)$.

To show (8.6) write

$$E\hat{s}^2_{n,q} = \hat{s}^2_{n,q;g} + E\hat{s}^2_{n,q;\xi}, \tag{8.10}$$

where

$$\hat{s}^2_{n,q;g} = q^{-1} \sum_{i,j=1}^{q} \hat{\gamma}_g(i-j), \quad \hat{s}^2_{n,q;\xi} = q^{-1} \sum_{i,j=1}^{q} \hat{\gamma}_\xi(i-j)$$

with

$$\hat{\gamma}_g(j) = n^{-1} \sum_{i=1}^{n-|j|} (g_n(i) - \bar{g}_n)(g_n(i+|j|) - \bar{g}_n), \quad \hat{\gamma}_\xi(j) = n^{-1} \sum_{i=1}^{n-|j|} (\xi_i - \bar{\xi}_n)(\xi_{i+|j|} - \bar{\xi}_n).$$

Note that $\hat{s}^2_{n,q}, \hat{s}^2_{n,q;g}, \hat{s}^2_{n,q;\xi} \geq 0$. Since under Assumption T(γ),

$$|\hat{\gamma}_g(j)| \leq n^{-1} \sum_{i=1}^{n} (g_n(i) - \bar{g}_n)^2 \leq Cn^{2\gamma}$$

it follows that $\hat{s}^2_{n,q;g} \leq Cqn^{2\gamma}$. To prove (8.6) it remains to show that

$$E\hat{s}^2_{n,q;\xi} \leq Cqn^{2\gamma}. \tag{8.11}$$

In cases (h1), (h2), estimating

$$\begin{aligned} |E\hat{\gamma}_\xi(j)| &\leq n^{-1} \sum_{i=1}^{n-|j|} |E(\xi_i - \bar{\xi}_n)(\xi_{i+|j|} - \bar{\xi}_n)| \\ &\leq n^{-1} \sum_{i=1}^{n-|j|} [E(\xi_i - \bar{\xi}_n)^2 E(\xi_{i+|j|} - \bar{\xi}_n)^2]^{1/2} \leq C, \end{aligned}$$

we see that

$$E\hat{s}^2_{n,q;\xi} \leq Cq \leq Cqn^{2\gamma}$$

as $\gamma \geq 0$.

In case (h3), by Lemma 8.2, $E\hat{s}^2_{n,q;\xi} \sim q^{2d} s_d^2$ and thus

$$E\hat{s}^2_{n,q;\xi} \leq Cq^{2d} \leq Cqn^{2\gamma}$$

in view of assumption (3.16). □

PROOF OF THEOREM 3.3. It suffices to show that for some $\theta > -1/2$ process

$$\begin{aligned} \widetilde{X}_n(t) &= n^{-1/2-\theta}(X_{[nt]+1} - X_0) = n^{-1/2-\theta} \sum_{k=1}^{[nt]+1} (X_{k+1} - X_k) \\ &= n^{-1/2-\theta} \sum_{k=1}^{[nt]+1} (\mu + g_n(k) + \xi_k) \end{aligned}$$

satisfies assumptions of Theorem 3.4. In case (h1*)–(h3*) this can be established using similar argument as in the proof of Theorem 3.2; in case (h4*) this is shown in the proof of Proposition 3.1. □

PROOF OF THEOREM 3.4. (3.19) follows by Lemma 8.1 and Proposition 8.1. □

PROPOSITION 8.1 *Suppose that for some $\theta > -1/2$ the differences*

$$\widetilde{X}_n(t) = n^{-1/2-\theta}(X_{[nt]+1} - X_0), \quad 0 \leq t \leq 1 \tag{8.12}$$

satisfy assumptions (a)–(c) of Theorem 3.4. Then

$$n^{-2-2\theta} V_n \xrightarrow{D} \int_0^1 \Big(\int_0^t \widetilde{X}_0(s)ds - \int_0^1 \Big[\int_0^u \widetilde{X}_0(s)ds \Big] du \Big)^2 dt, \tag{8.13}$$

where $\widetilde{X}_0(t) = X_0(t) - \int_0^1 X_0(s)ds$. In addition,

$$E\hat{s}^2_{n,q} \leq Cqn^{1+2\theta}. \tag{8.14}$$

PROOF OF PROPOSITION 8.1. Using (8.12) we can write

$$\begin{aligned} n^{-1/2-\theta}(X_{[ns]+1} - \bar{X}_n) &= n^{-1/2-\theta}\Big(X_{[ns]+1} - X_0 - n^{-1} \sum_{j=1}^n (X_j - X_0) \Big) \\ &= \widetilde{X}_n(s) - \int_0^1 \widetilde{X}_n(u)du. \end{aligned} \tag{8.15}$$

Hence, for $S_k^* = \sum_{i=1}^k (X_i - \bar{X}_n)$ we have

$$n^{-3/2-\theta} S^*_{[nt]+1} = \int_0^{[nt]/n} \Big(\widetilde{X}_n(s) - \int_0^1 \widetilde{X}_n(u)du \Big) ds.$$

This and (7.4) imply

$$\begin{aligned} n^{-2-2\theta} V_n &= \int_0^1 \Big(\int_0^{[nt]/n} \Big(\widetilde{X}_n(s) - \int_0^1 \widetilde{X}_n(v)dv \Big) ds \\ &\quad - \int_0^1 \Big(\int_0^{[nu]/n} (\widetilde{X}_n(s) - \int_0^1 \widetilde{X}_n(v)dv)ds \Big) du \Big)^2 dt \\ &\xrightarrow{D} \int_0^1 \Big(\int_0^t \widetilde{X}_0(s)ds - \int_0^1 [\int_0^u \widetilde{X}_0(s)ds]du \Big)^2 dt \end{aligned}$$

by Remark 6.1.

To estimate $\hat{s}^2_{n,q} = q^{-1}\sum_{i,j=1}^q \hat{\gamma}_{i-j}$, note that $\hat{s}^2_{n,q} \geq 0$. By (8.15) and assumption (b) of Theorem 3.4,

$$En^{-1-2\theta}(X_{[ns]+1} - \bar{X}_n)^2 \leq 2E\widetilde{X}^2_n(s) + 2E\left(\int_0^1 \widetilde{X}_n(u)du\right)^2 \leq C.$$

Hence,

$$E|\hat{\gamma}_j| \leq n^{-1}\sum_{i=1}^{n-|j|}\Big(E(X_i - \bar{X}_n)^2\Big)^{1/2}\Big(E(X_{i+|j|} - \bar{X}_n)^2\Big)^{1/2} \leq Cn^{1+2\theta}$$

and

$$E\hat{s}^2_{n,q} \leq q^{-1}\sum_{i,j=1}^q E|\hat{\gamma}_{i-j}| \leq Cqn^{1+2\theta},$$

to prove (8.14). $\square$

Lemma 8.2 *Let $\{\xi_k\}$ be a covariance stationary zero mean sequence with $(\gamma_k) \in G(d)$, $-1/2 < d < 1/2$. Then*

$$E(n^{-1/2-d}\sum_{i=1}^n \xi_i)^2 = n^{-1-2d}\sum_{i,j=1}^n \gamma_{i-j} \sim s^2_d, \tag{8.16}$$

where s^2_d is defined in (3.1). *Moreover,*

$$E\widehat{s}^2_{n,q;\xi} \sim q^{2d}s^2_d. \tag{8.17}$$

Proof of Lemma 8.2. Suppose that $d > 0$. Then, by (2.2),

$$\begin{aligned} t_n(d) &:= n^{-1-2d}\sum_{i,j=1}^n \gamma_{i-j} \sim C_d n^{-1-2d}\sum_{i,j=1}^n |i-j|^{2d-1} \\ &\sim C_d\int_0^1\int_0^1 |x-y|^{2d-1}dx\,dy \\ &= \frac{C_d}{d(2d+1)} = s^2_d. \end{aligned}$$

In case $d \leq 0$, write

$$t_n(d) = n^{-2d}\sum_{|j|<n}\Big(1 - \frac{|j|}{n}\Big)\gamma_j. \tag{8.18}$$

If $d = 0$ then, by (2.1),

$$t_n(d) \sim \sum_{j=-\infty}^{\infty}\gamma_j = s^2_0.$$

If $d < 0$ then, by (2.3) and (8.18),

$$t_n(d) = -n^{-2d}\sum_{|j|\geq n}\gamma_j - n^{-2d-1}\sum_{|j|\leq n}|j|\gamma_j,$$

where

$$\sum_{|j|\geq n}\gamma_j \sim 2C_d\int_n^\infty x^{2d-1}dx = -\frac{C_d}{d}\,n^{2d},$$

$$\sum_{|j|<n} |j|\gamma_j \sim 2C_d \int_0^n x^{2d} dx = \frac{2C_d}{2d+1}\, n^{2d+1}$$

and therefore

$$t_n(d) \sim \frac{C_d}{d} - \frac{2C_d}{2d+1} = s_d^2.$$

We prove now (8.17). We have $E\hat{s}^2_{n,q;\xi} = q^{-1}\sum_{i,j=1}^{q} E\hat{\gamma}_\xi(i-j)$. Here

$$\begin{aligned} E\hat{\gamma}_\xi(j) &= n^{-1}\sum_{i=1}^{n-|j|} E(\xi_i - \bar{\xi}_n)(\xi_{i+|j|} - \bar{\xi}_n) \\ &= n^{-1}\sum_{i=1}^{n-|j|} E\xi_i\xi_{i+|j|} - n^{-1}\bar{\xi}_n \sum_{i=1}^{n-|j|}(\xi_{i+|j|} + \xi_i) + (\bar{\xi}_n)^2 \\ &= \frac{n-j}{n}\,\gamma_j + O(n^{-1+2d}) \end{aligned}$$

since, by (8.16), $E(\bar{\xi}_n)^2 \le Cn^{-1+2d}$. Hence

$$\begin{aligned} E\hat{s}^2_{n,q;\xi} &= q^{-1}\sum_{i,j=1}^{q}\Big(\frac{n-|i-j|}{n}\,\gamma_{i-j} + O(qn^{-1+2d})\Big) \\ &= q^{-1}\sum_{i,j=1}^{q}\gamma_{i-j} - q^{-1}\sum_{i,j=1}^{q}\frac{|i-j|}{n}\,\gamma_{i-j} + O(qn^{-1+2d}). \end{aligned}$$

By (8.16), the first term satisfies $q^{-1}\sum_{i,j=1}^{q}\gamma_{i-j} \sim q^{2d}s_d^2$. The second term in case $d \ge 0$ can be estimated using the property $(\gamma_k) \in G(d)$, which yields

$$q^{-1}\sum_{i,j=1}^{q}\frac{|i-j|}{n}|\gamma_{i-j}| \le n^{-1}\sum_{i,j=1}^{q}|\gamma_{i-j}| \le Cn^{-1}q^{2d+1} = o(q^{2d}).$$

If $d < 0$ then

$$q^{-1}\sum_{i,j=1}^{q}\frac{|i-j|}{n}|\gamma_{i-j}| \le Cq^{-1}n^{-1}\sum_{i,j=1}^{q}|i-j|^{2d} \le Cn^{-1}q^{1+2d} = o(q^{2d}).$$

Finally, the third term satisfies $qn^{2d-1} = (q/n)^{1-2d}q^{2d} = o(q^{2d})$, to prove (8.17). $\square$

Lemma 8.3 *Let*

$$X_k = \mu + \xi_k, \tag{8.19}$$

where $\{\xi_k\}$ is a fourth-order stationary process with $(\gamma_k) \in G(d), |d| < 1/2$ and, in addition, (3.8), (3.9) hold. Then, if $q \to \infty$ and $q = O(n^{1/2})$, we have, as $n \to \infty$

$$q^{-2d}\hat{s}^2_{n,q}, \xrightarrow{P} s_d^2 \tag{8.20}$$

where s_d^2 is the same as in Lemma 8.2.

(8.20) *remains valid for*

$$X_k = \mu + g_n(k) + \xi_k, \tag{8.21}$$

where $g_n(k)$ satisfies (3.7).

Proof of Lemma 8.3. Consider first the model (8.19). For $d \geq 0$, (8.20) was shown in Giraitis *et al.* (2003, Theorem 3.1). For $d < 0$, the proof follows the same line as in Giraitis *et al.* (2003), i.e., we rewrite

$$\hat{s}^2_{n,q} = \sum_{|j|<q} \Big(1 - \frac{|j|}{q}\Big)\tilde{\gamma}_j + \sum_{|j|<q} \Big(1 - \frac{|j|}{q}\Big)\delta_j =: v_{n,1} + v_{n,2},$$

where

$$\begin{aligned}
\tilde{\gamma}_j &= n^{-1} \sum_{i=1}^{n-|j|} (X_i - \mu)(X_{i+|j|} - \mu), \\
\delta_j &= \Big(1 - \frac{|j|}{n}\Big)(\bar{X}_n - \mu)^2 - n^{-1}(\bar{X}_n - \mu)(Z_{1,n-|j|} + Z_{|j|+1,n})
\end{aligned}$$

and $Z_{k,l} = \sum_{i=k}^{l}(X_i - \mu)$. It suffices to show

$$q^{-2d} v_{n,1} \xrightarrow{P} s_d^2 \quad \text{and} \quad q^{-2d} v_{n,2} \xrightarrow{P} 0.$$

Notice that verification of relations

$$q^{-2d} v_{n,2} \xrightarrow{P} 0 \quad \text{and} \quad q^{-2d} E v_{n,1} \to s_d^2$$

is the same as in Giraitis *et al.* (2003). Hence, it remains to check that

$$E(v_{n,1} - Ev_{n,1})^2 = o(q^{2d}).$$

By assumption $\sum_{h,r,s=-\infty}^{\infty} |\kappa(h,r,s)| < \infty$, similarly as in Giraitis *et al.* (2003) we derive

$$\begin{aligned}
E(v_{n,1} - Ev_{n,1})^2 &\leq n^{-2} \sum_{|j|,|j'|\leq q} \sum_{i=1}^{n}\sum_{i'=1}^{n} \Big(\big|\kappa(|j|, i'-i, i'-i+|j'|)\big| \\
&\quad + |\gamma_{i-i'}|\,|\gamma_{i-i'+|j|-|j'|}| + |\gamma_{i-i'-|j'|}|\,|\gamma_{i-i'+|j|}|\Big) \\
&\leq n^{-2} \sum_{|j|\leq q} \sum_{i=1}^{n} \Big(\sum_{i',j'=-\infty}^{\infty} |\kappa(|j|, i', j')| + (\sum_{k=-\infty}^{\infty} |\gamma_k|)^2 \Big) \\
&\leq C\Big(\frac{1}{n} + \frac{q}{n}\Big) \leq C(q/n) = o(q^{2d})
\end{aligned}$$

since $q/n = q^{2d}(q/\sqrt{n})^{1-2d} n^{-1/2-d} = o(q^{2d})$ and $\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty$.

We now show (8.20) for the model (8.21). Write using notations of (8.10),

$$E\hat{s}^2_{n,q} = \hat{s}^2_{n,q;g} + \hat{s}^2_{n,q;\xi} + r_n, \tag{8.22}$$

where

$$r_n = q^{-1} \sum_{i,j=1}^{q} n^{-1} \sum_{i=1}^{n-|j|} [(g_n(i) - \bar{g}_n)(\xi_{i+|j|} - \bar{\xi}_n) + (\xi_i - \bar{\xi}_n)(g_n(i+|j|) - \bar{g}_n)].$$

Relation (8.20) implies that $q^{2d}\hat{s}^2_{n,q;\xi} \to s_d^2$ and, as in proof of Theorem 3.2, it can be shown that

$$\hat{s}^2_{n,q;g} \leq Cqn^{-1} \sum_{i=1}^{n} (g_n(i) - \bar{g}_n)^2 \leq (q/n)o(n^{2d}) = q^{2d} o((q/n)^{1-2d}) = o(q^{2d})$$

by (3.7). Since $E|(\xi_{i+|j|} - \bar{\xi}_n)| \le C$, then

$$\begin{aligned} E|r_n| &\le Cq^{-1} \sum_{i,j=1}^{q} n^{-1} \sum_{i=1}^{n} |g_n(i) - \bar{g}_n| \\ &\le Cqn^{-1/2} \Big(\sum_{i=1}^{n} (g_n(i) - \bar{g}_n)^2 \Big)^{1/2} = Cqn^{-1/2} o(n^d) = o(q^{2d}), \end{aligned}$$

since $qn^{-1/2}n^d \le Cq^{2d}$ when $q \le Cn^{1/2}$. This and (8.22) prove (8.20). $\square$

Lemma 8.4 *Suppose that assumptions of Lemma 8.3 are satisfied. Then* (3.4) *holds.*

Proof of Lemma 8.4. By Lemma 8.3,

$$q^{-2d} \hat{s}^2_{n,q} \xrightarrow{P} s^2_d,$$

and by (8.16) of Lemma 8.2,

$$q^{-1-2d} \sum_{i,j=1}^{n} \gamma_{i-j} \sim s^2_d,$$

which together with definition (2.14) of $\hat{s}^2_{n,q}$ implies (3.4). $\square$

References

Bardet, J.-M., Lang, G., Oppenheim, G., Philippe, A. and Taqqu, M.S. (2002). Generators of long-range dependent processes: a survey. In *Long-range Dependence: Theory and Applications* (eds P. Doukhan, G. Oppenheim and M. S. Taqqu). Birkhäuser.

Bhattacharya, R. N., Gupta, V. K. and Waymire, E. (1983). The Hurst effect under trends. *Journal of Applied Probability*, **20,** 649–662.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Cavaliere, G. (2001). Testing the unit root hypotheis using generalized range statistics. *Econometrics Journal*, **4,** 70–88.

Cremers, H. and Kadelka, D. (1986). On weak convergence of integral functionals of stochastic processes with applications to processes taking paths in l_p^e. *Stochastic Processes and their Applications*, **21,** 305–317.

Davydov, Y. A. (1970). The invariance principle for stationary processes. *Theory of Probability and its Applications*, **15,** 487–498.

DeJong, D. N., Nankervis, J. C., Savin, N. E. and Whiteman, C. H. (1992). The power problems of unit root tests for time series with autoregressive errors. *Journal of Econometrics*, **53,** 323–343.

DeJong, D. N. and Whiteman, C. H. (1991). Reconsidering 'Trends and random walks in macroeconomic time series'. *Journal of Monetary Economics*, **28,** 221–254.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74,** 427–431.

Diebold, F. X. and Rudebusch, G. D. (1991). On the power of Dickey–Fuller tests against fractional alternatives. *Economics Letters*, **35,** 155–160.

Ding, Z. and Granger, C. W. J. (1996). Modeling volatility persistence of speculative returns: a new approach. *Journal of Econometrics*, **73,** 185–215.

Dolado, J. J., Gonzalo, J. and Mayoral, L. (2002). A fractional Dickey–Fuller test for unit roots. *Econometrica*, **70,** 1963–2006.

Gikhman, I. I. and Skorokhod, A. V. (1980). *The Theory of Stochastic Processes*, volume 1. Springer-Verlag, Berlin.

Giraitis, L., Kokoszka, P. and Leipus, R. (2001). Testing for long memory in the presence of a general trend. *Journal of Applied Probability*, **38,** 1033–1054.

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

Giraitis, L. and Robinson, P. (2001). Edgeworth expansions for semiparametric Whittle estimation of long memory. *Annals of Statistics (forthcoming).*

Hurst, H. (1951). Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, **116,** 770–799.

Hurvich, C. M., Lang, G. and Soulier, P. (2002). Estimation of long memory in the presence of a smooth nonparametric trend. Preprint.

Ibragimov, I. A. and Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters–Noordhoff, Groningen.

Künsch, H. (1987). Statistical aspects of self-similar processes. In *Proceedings of the 1st World Congress of the Bernoulli Society* (eds Yu. A. Prohorov and V. V. Sazonov), volume 1, pp. 67–74. VNU Science Press, Utrecht.

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? *Journal of Econometrics*, **54,** 159–178.

Lee, D. and Schmidt, P. (1996). On the power of the KPSS test of stationarity against fractionally-integrated alternatives. *Journal of Econometrics*, **73,** 285–302.

Lee, H. S. and Amsler, C. (1997). Consistency of the KPSS unit root test against fractionally integrated alternative. *Economics Letters*, **55,** 151–160.

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

Lobato, I. N. and Savin, N. E. (1998). Real and spurious long-memory properties of stock-market data (with comments). *Journal of Business & Economic Statistics*, **16,** 261–283.

Maddala, G. S. and Kim, I.-M. (1998). *Unit Roots, Cointegration and Structural Changes.* Cambridge University Press.

Mandelbrot, B. B. and Wallis, J. M. (1968). Noah, Joseph and operational hydrology. *Water Resources Research*, **4,** 909–918.

Mikosch, T. and Stărică, C. (1999). Change of structure in financial time series, long range dependence and the GARCH model. IWI-preprint 99–5–06.

Nelson, C. R. and Plosser, C. (1982). Trends and random walks in macro-economic time series: some evidence and implications. *Journal of Monetary Economics*, **10,** 139–162.

Oliveira, P. E. and Suquet, Ch. (1998). Weak convergence in $L^p(0,1)$ of the uniform empirical process under dependence. *Statistics and Probability Letters*, **39,** 363–370.

Park, J. and Phillips, P. C. B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory*, **15,** 269–298.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, **57,** 1361–1401.

Pötcher, B. M. (2001). Nonlinear functions and convergence to Brownian motion: beyond the continuous mapping theorem. Preprint.

Robinson, P. (1997). Large-sample inference for nonparametric regression with dependent errors. *Annals of Statistics*, **25,** 2054–2083.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, **23,** 1630–1661.

Shin, Y. and Schmidt, P. (1992). The KPSS stationary test as a unit root test. *Economics Letters*, **38,** 387–505.

Watson, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika*, **48,** 109–114.

Xiao, Z. (2001). Testing the null hypothesis of stationarity against an autoregressive unit root alternative. *Journal of Time Series Analysis*, **22,** 87–105.