Bootstrap Testing for Long Range Dependence

Argjir Butka[#1], Llukan Puka[*2], Ilir Palla[#3]

[#]Department of Mathematics, "Fan S. Noli" University,
Shëtitorja "Rilindasit", Korçë, Albania

[*]Department of Mathematics, University of Tirana,
Blv. "Zog 1", Tirana, Albania

Abstract—**The presence of long-memory or long-range dependence (LRD) in a stochastic process has important consequences in statistical inferences. Lo develops a robust test for detecting the existence of LRD and derives its asymptotic distribution. Teverovsky et al. uncover some drawbacks to using Lo's method in practice. In particular, they find that Lo's method has a strong preference for accepting the null hypothesis of no LRD. The bootstrap provides a practical method to correct the size distortion of the asymptotic tests. In this paper, we introduce a semi-parametric bootstrap testing procedure for detecting LRD. We investigate the size and power of Lo's and its bootstrap tests by means of a computer simulation study. The results suggest that the bootstrap tests correct for size distortions of asymptotic tests for small sample sizes.**

Keywords— **Long-range dependence, fractional differencing, modified R/S statistic, block bootstrap, cycling blocks.**

I. Introduction

Long-memory, or long-range dependence (LRD), denotes the property of a time series to exhibit a significant dependence between very distant observations. Since the early work of Hurst [1], it has been recognised that many time series may exhibit the phenomenon of LRD. Time series with LRD has recently received considerable attention in economics (see [2]-[3] for example), but also in other research areas.

There are several possible definitions of the property of LRD. According to [4], a stationary time series possesses LRD if its autocorrelations are absolutely non-summable. Equivalently, the spectral density function of a LRD process will be unbounded at low frequencies.

The autoregressive fractionally integrated moving average, *ARFIMA*(p, d, q), process has widely been used in different fields, such as hydrology, computer science and finance, to represent a time series with LRD property. It is important to distinguish LRD processes from more common short range dependence processes such that the *ARMA* processes. A number of tests have been proposed to detect the presence of LRD in a given data set. Some of them are described in [5]. Historically, one of the most effective techniques was the rescaled range statistical analysis, first introduced by [1] to describe the long-term dependence of water levels in rivers and reservoirs. This analysis provides a sensitive method for revealing long-run correlations in random processes. What especially makes the rescaled range analysis appealing is that all the information about a complex signal are contained in one parameter only called as *the Hurst exponent* or *the Hurst parameter*, or even *the H parameter*. Lo modifies the rescaled range approach and proposes a test procedure for the null hypothesis of no LRD, in [6]. He derives the limiting distribution of the modified statistic under both short-range and long-range dependence.

In the present study, the Lo's modified test and the corresponding bootstrap test will be used to test for the null hypothesis of no presence of LRD in a stationary time series. We apply a bootstrap procedure that considers resampling blocks composed of cycles, as introduced in [7].

The paper is organized as follows. Section II describes the LRD notion and the modified Lo's test. In section III, we review the block bootstrap methods and the bootstrap hypothesis testing. Then we describe a semi-parametric bootstrap procedure and give the algorithm to test for LRD. In section IV, we outline the experiment design and discuss the results of our Monte Carlo simulation study. At last, conclusions are pointed out in section V.

II. Long Range Dependence and its Detection

In this section, we give a brief description of the LRD notion and *ARFIMA* models. The classical rescaled range (*R/S*) statistic and the modified *R/S* statistic, to test for LRD, are also described.

A. The LRD Notion

Historically, the importance of LRD processes as stochastic models lies in the fact that they provide an elegant explanation and interpretation of an empirical law that is commonly referred to as Hurst's law or the Hurst effect. Even though the LRD notion recently has become important, there are various definitions of it. When definitions are given, they vary from author to author. There are alternative definitions not necessarily equivalent to one-another, in literature (for a collection of different notions behind the concept of LRD, see [8]). Most of the definitions of LRD appearing in literature are based on the second order properties of a stochastic process. Such properties include asymptotic behaviour of covariance, spectral density and variances of partial sums.

Let $\{X_t, t \in Z\}$ be a stationary time series process with $\gamma(k)$, $\rho(k)$, $f(\omega)$ being autocovariance, autocorelation and spectral density function, respectively. In general, long memory is defined by the fact that the autocorrelations of the process are absolutely non-summable, [4], i.e.

$$\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty \tag{1}$$

However, there are alternative definitions. Suppose that there exists a real number $\alpha \in (0,1)$ and a constant $c_\rho > 0$ not depended on k such that:

$$\lim_{k\to\infty} \frac{\rho(k)}{c_\rho k^{-\alpha}} = 1 \tag{2}$$

Then X_t is called a stationary process with LRD, or a stationary process with slowly decaying or long-range correlations, [5].

LRD can equivalently be defined in the frequency domain. Suppose that there exists a real number $\beta \in (0,1)$ and a constant $c_f > 0$ not depended on ω such that:

$$\lim_{\omega\to 0} \frac{f(\omega)}{c_f |\omega|^{-\beta}} = 1 \tag{3}$$

Then X_t is called a stationary process with long memory or LRD or strong dependence, [5].

LRD processes can be represented as integrated $I(d)$ processes with fractional $d \in (0, 0.5)$. This can be done by generalising Box/Jenkins, [9], *ARMA* models to *ARFIMA* models. *ARFIMA* models were introduced by [10] and independently by [11]. Allowing also for short memory terms we have the *ARFIMA*(p, d, q) representation as following

$$\Phi(B)(1-B)^d X_t = \Theta(\varepsilon_t) \tag{4}$$

where B denotes the backshift operator, ε_t is a mean zero finite variance white noise process and $\Phi(\bullet)$, $\Theta(\bullet)$ denote, respectively, the p-order autoregressive and q-order moving average polynomials (see [5]).

B. Detecting LRD

It is important to distinguish LRD processes from more common short range dependence processes. For example, LRD processes violate the central limit theorem (CLT) in that the variance of the sample mean goes to zero more slowly than the usual order of $1/N$ (N denotes the sample size). On the other hand, the CLT is still valid for short memory processes. A number of tests have been proposed to detect the presence of LRD in an underlined data set. Some of them are described in [5].

The most popular test for long memory is the rescaled range (*R/S*) statistic, proposed by [1] and discussed in details in [5] and [12]. Classical *R/S* analysis aims at inferring from an empirical record the value of the Hurst parameter for the LRD process that presumably generated the record at hand.

For a time series $\{X_t, t \in N\}$, let $Y_t = \sum_{j=1}^{t} X_j$ be the aggregated time series. Then, the adjusted range statistic or $R(t,k)$ statistic is defined by

$$R(t,k) = \max_{0 \le i \le k}\left[\left(Y_{t+i} - Y_t\right) - \frac{i}{k}\left(Y_{t+k} - Y_t\right)\right] - \min_{0 \le i \le k}\left[\left(Y_{t+i} - Y_t\right) - \frac{i}{k}\left(Y_{t+k} - Y_t\right)\right] \qquad (5)$$

This range can be interpreted as the ideal capacity of a water reservoir with inflows at time i denoted by X_i for the time span between t and $t+k$ (see [5]). For the adjusted range $R(t,k)$ be independent of a scale, it is standardized by standard deviation $S(t,k) = \sqrt{\frac{1}{k}\sum_{i=t+1}^{t+k}\left(X_i - \bar{X}_{t,k}\right)^2}$, where $\bar{X}_{t,k} = \frac{1}{k}\sum_{i=t+1}^{t+k} X_i$. Then, for all possible values of *k* and *t*, the rescaled adjusted range statistic is given by

$$\left(R/S\right)_k = \frac{R(t,k)}{S(t,k)} \qquad (6)$$

To estimate the LRD parameter *H*, we plot the logarithm *R/S* against logarithm of *k* for all possible values of *k*. Hurst, [1], observed that many empirical records may be well represented by the relation $\log(R/S)_k \approx c + H\log(k)$ as *k* becomes large, with typically values of the Hurst parameter in the interval $]0.5,1[$ and *c* a finite positive constant that does not depend on *k*. Then *H* can be estimated as the slope ordinary least squares (OLS) estimator.

Lo, [6], proposed a modified *R/S* statistic that is obtained by replacing the denominator of the classical *R/S*, given by (6), by a consistent estimator of the square root of the variance of the partial sum. Lo, [6], derives the limiting distribution of his modified *R/S* statistic under both short-range and long-range dependence. Following [6], we define

$$S_q(N) = \left(\frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2 + \frac{2}{N}\sum_{i=1}^{q}\omega_i(q)\sum_{k=i+1}^{N}\left(X_k - \bar{X}\right)\left(X_{k-i} - \bar{X}\right)\right)^{1/2} \qquad (7)$$

where $\bar{X} = \sum_{i=1}^{N} X_i$ and $\omega_k(q) = 1 - \frac{k}{q+1}$ for a fixed integer $q < N$. Then the Lo's modified *R/S* statistic, usually denoted by $V_q(N)$, is defined by

$$V_q(N) = N^{-1/2}\frac{R(N)}{S_q(N)} \qquad (8)$$

As *N* increases without bound, the truncation lag $q = q(N)$ also increases such that $q \square o\left(N^{1/4}\right)$.

Recently, it is established that Lo's modified *R/S* method does not provide a safe ultimate test for LRD detection. Teverovsky et al., [13], identify a number of serious drawbacks to using Lo's method in practice. They also show that the value of Lo's modified *R/S* statistic is sensitive to the choice of the truncation lag *q*, used to estimate the standard deviation of the process. As the truncation lag increases, the test statistic has a strong bias towards accepting the null of no long run dependence, even when the DGP is a basic fractionally integrated (*FI*) process.

The right choice of the truncation point *q* is essential so long as the Lo's results are asymptotic whereas, in practice, the sample size is always finite. Lo, [6], suggests a data-driven formula for choosing *q* as

$$q_L = \left[\left(\frac{3N}{2}\right)^{1/3}\left(\frac{2\hat{\rho}}{1-\hat{\rho}^2}\right)^{2/3}\right] \qquad (9)$$

where $\hat{\rho}$ is the estimated first-order autocorrelation coefficient, N is the sample length and the symbol $[\bullet]$ denotes the greatest integer function. Lo, [6], shows that this choice of q is appropriate if the data generation process is an $AR(1)$ process. Some simulation studies ([6], [13], [14]) have shown that, in general, the larger the q, the less likely the null hypothesis is to be rejected. After a trial-error procedure, Wang et al., [14], suggest the following modified formula to choose the truncation lag q.

$$q_W = \left[\left(\frac{N}{10} \right)^{1/4} \left(\frac{2\hat{\rho}}{1-\hat{\rho}^2} \right)^{2/3} \right] \tag{10}$$

Murphy and Izzeldin, [15], use a truncation lag that does not depend on the data, given by

$$q_M = \left[8 \left(\frac{N}{100} \right)^{1/4} \right] \tag{11}$$

We use all these three truncation lags in our Monte Carlo study.

III. Block Bootstrap Methods and Hypothesis Testing

A. The Block Bootstrap Methods

Bootstrap method, originally introduced by Efron, [16], is a resample technique that provides better estimations in small sample sizes than the classical methods. However, Efron's bootstrap classical application in the context of dependent data, such as a time series, fails to work. In this case, the bootstrap technique must be carried out in such a way that the dependence structure of the original time series to be preserved in the bootstrap time series. Block bootstrap methods for dependent data were introduced by [17]-[19], among others. In these methods, the data is divided in blocks which are approximately independent and the joint distribution of the variables in different blocks is about to the same due to stationarity. The block bootstrap for time series consists of randomly resampling blocks of consecutive values of the given data and aligning these blocks into a bootstrap sample. Different methods differ in the way as blocks are constructed. Kunsch, [19], and Liu and Singh, [20], have independently formulated a resampling procedure, called the moving-block bootstrap (MBB), for general weakly-dependent observations. For a theoretical comparison of some block bootstrap methods see [21].

Lahiri, [22], showed that in general, the MBB procedure fails to provide a valid approximation to the distribution of normalized sample mean under LRD. One of the reasons behind this is that joining independent bootstrap blocks to form the bootstrapped statistic destroys the strong dependence of the underlying observations. Carlstein et al., [23], sampled blocks dependently, attempting to follow each block with one that might realistically follow it in the underlying process. Hesterberg, [24], used the matched-block bootstrap for LRD processes, investigating block matching rules, based on linear combinations of observations in the block. Ekonomi and Butka, [7], introduced the bootstrap with cycling blocks (BCB) to estimate the fractional parameter in ARFIMA models. The validity of BCB for stationary time series is shown in [25]. In this paper, we consider BCB with circular moving blocks similar to [26]. We briefly describe the BCB procedure in following.

Let $X_t, t = 1, 2, ..., N$ be an observed time series. A cycle is defined as a pair of alternating high and low runs of the data that are created when the terms of the time series cross the sample mean. Then we define a block composed of a fixed number of consecutive cycles. Let $C_1, C_2, ..., C_k$ denote the created cycles and let $n_j, j = 1, 2, ..., k$ be the number of terms of the cycle C_j. Then, the created cycles would be written as $C_1 = \{X_1, X_2, ..., X_{n_1}\}$, $C_2 = \{X_{n_1+1}, X_{n_1+2}, ..., X_{n_1+n_2}\}, \ldots,$ $C_k = \{X_{n_1+n_2+...+n_{k-1}+1}, ..., X_N\}$. The cycle lengths and the number of cycles are random variables so long as they are determined automatically by the data. The number of cycles of a block is a tuning parameter and is analogous with the block length in MBB. We consider the circular moving-block bootstrap approach, which amounts to wrapping the data around in a

circle before the blocks are created, as in [26]. If the number of cycles per block is set to be *s*, then the blocks would be $B_j = \{C_j, C_{j+1}, ..., C_{j+s-1}\}$ for $j = 1, 2, ..., k$ (if $i > k$, we assign $C_i \equiv C_{i(\text{mod } k)}$). We treat a cycle as an inseparable observation. Since the cycles, consequently the blocks, are created automatically by the series' crossings of the sample mean we pretend that the dependence structure of the bootstrap sample mimics the dependence structure of the original data.

B. Bootstrap Hypothesis Testing

In briefly, the implementation of the bootstrap for hypothesis testing can be summarized as follows (for details see [27]). Suppose that $\hat{\tau}$ is the observed value of a test statistic τ with cumulative distribution function F_τ under the null hypothesis. Let we wish to perform a test at level α that rejects the null hypothesis when $\hat{\tau}$ is in the upper tail. Than the p-value of $\hat{\tau}$ is $p(\hat{\tau}) = 1 - F_\tau(\hat{\tau})$. If we knew F_τ, we would simply calculate $p(\hat{\tau})$ and reject the null whenever $p(\hat{\tau}) < \alpha$. But in practice F_τ is unknown or the asymptotic approximation of it may perform poorly for finite sample sizes. A popular alternative is to perform a bootstrap test. In the spirit of the bootstrap we first generate *Q* bootstrap samples and compute the bootstrap test statistic τ_j^* for each one, most commonly by the same procedure used to calculate $\hat{\tau}$ from the real sample. The basic idea is to create a large number of such samples which all obey the null hypothesis and, as far as possible, resamples the original sample. Then, if we wish to reject when $\hat{\tau}$ is in the upper tail, the bootstrap *p*-value is $\hat{p}^*(\hat{\tau}) = \frac{1}{Q}\sum_{j=1}^{Q} I(\tau_j > \hat{\tau})$ where $I(\bullet)$ denotes the indicator function. When we wish to perform a two-tailed test, and we are willing to assume that τ is symmetrically distributed around zero, we can use the symmetric bootstrap *p*-value given by

$$\hat{p}_s^*(\hat{\tau}) = \frac{1}{Q}\sum_{j=1}^{Q} I\left(\left|\tau_j^*\right| > \left|\hat{\tau}\right|\right) \tag{12}$$

If we are not willing to make the assumption that τ is symmetrically distributed around zero, we can use the equal-tail bootstrap *p*-value given by

$$\hat{p}_{et}^*(\hat{\tau}) = 2\min\left(\frac{1}{Q}\sum_{j=1}^{Q} I\left(\tau_j^* \le \hat{\tau}\right), \frac{1}{Q}\sum_{j=1}^{Q} I\left(\tau_j^* > \hat{\tau}\right)\right) \tag{13}$$

In this paper we consider both $\hat{p}_s^*$ and $\hat{p}_{et}^*$ bootstrap *p*-values.

C. The Algorithm

Suppose that we wish to test the null hypothesis H_0: "*There is no LRD in the process under consideration*" against the alternative hypothesis, H_1: "*There is LRD*". Let $\{X_1, X_2, ..., X_N\}$ be the observed time series. The algorithm for testing for LRD, used in this paper, consists as follows.

Step 1. Calculate the test statistic $\tau = V_q(N)$ defined by equation (8) with truncation lag (9), (10) or (11).

Step 2. Fit an $AR(p_0)$ model, choosing the parameter p_0 such that $AR(p_0)$ is the best fitting data model in all $\{AR(p), p=1,2,...,30\}$ models. We use AIC or BIC information criteria to select the best fitting model.

Step 3. Create a new time series $\{X_1', X_2', ..., X_N'\}$ that is generated by p_0 *AR* terms and a random sample of normal innovations. Let B_i, $i = 1, 2, ..., k$ be the blocks of this time series defined by BCB procedure.

Step 4. Sample with replacement from the set B_i, $i = 1, 2, ..., k$ and lay the resampled blocks end to end so producing the bootstrap sample $\{X_1^*, X_2^*, ..., X_N^*\}$ (also called as pseudoseries).

Step 5. From the pseudoseries calculate the bootstrap test statistic $\tau^* = V_q^*(N)$ following equation (8) with truncation lag (9), (10) or (11), but using X_j^* instead of X_j.

Step 6. Repeat steps 4 to 5 Q times and the bootstrap p-value, p^*, is calculated by equation (12) or (13).

Step 7. For a nominal significance level α, the null hypothesis H_0 is rejected only if $p^* < \alpha$.

IV. Simulation Study

A. Experiment Design

In a Monte Carlo experiment, we consider a wide class of models from the *ARFIMA*(p, d, q) processes, given by equation (4), to test for the LRD. The null hypothesis is that the data series has no long memory.

The Monte Carlo experiment covers first order autoregressions and moving averages as well as autoregressive fractionally integrated series of lengths N = 100, 200, 300, 500, and 1000. The empirical size of the tests are examined for the *AR*(1) and *MA*(1) processes, whilst the empirical power of the test are examined against *ARFIMA*(0, d, 0) and *ARFIMA*(1, d, 0) processes. We consider eight *AR*(1), eight *MA*(1), five *ARFIMA*(0, d, 0) and four *ARFIMA*(1, d, 0). The *AR*(1) model is defined by equation $X_t = \varphi X_{t-1} + \varepsilon_t$, the *MA*(1) model by $X_t = \theta\varepsilon_{t-1} + \varepsilon_t$, the *ARFIMA*(0, d, 0) by $(1-B)^d X_t = \varepsilon_t$ and the *ARFIMA*(1, d, 0) model by $(1-\varphi B)(1-B)^d X_t = \varepsilon_t$, where B is the back-shift operator, $(1-B)^d = \sum_{k=0}^{\infty}\binom{d}{k}B^k$ is the generalized binomial expansion, and $\{\varepsilon_t\}$ is a white noise process with zero mean and finite variance σ_ε^2. The considered coefficients are ± 1, ± 4, ± 5 and ± 9 for *AR*(1) and *MA*(1) models. For the fractionally integrated noise, we consider the values $d = 0.05$, 0.25, 0.30, $1/3$, and 0.45 in order to investigate different intensities of the LRD. For the *ARFIMA*(1, d, 0) models, all combinations of $\varphi = 0.5$, 0.9 and $d = 0.3$, 0.45 are considered. Using the "fracdiff" package in R, we generate an artificial time series from each model under consideration with 300 first terms removed in order to reduce the effect of initial values. The white noise process $\{\varepsilon_t\}$ is generated from the standard normal distribution. For each combination of the parameters we apply both the symmetric and the equal-tail bootstrap p-values. The autoregressive order p_0 of the fitted *AR*(p_0) model is derived using AIC and BIC information criteria. Then, in each simulated case we calculate the empirical size or power of the original and bootstrap tests, based on 1000 Monte Carlo replications. The nominal size (type I error) of the test is chosen to be $\alpha = 0.01$, 0.05, 0.10. We use Q = 1000 bootstrap replications. We set the average block length approximately at order $O(N^{0.5})$ using different numbers of cycles per block.

B. Results

Table I presents the empirical size of the tests, at a nominal $\alpha = 0.05$ level of significance, for *AR*(1) models. BIC criteria is used (AIC and BIC criteria yielded similar results). Numbers in bold face denote significant evidence for the nominal size of $\alpha = 0.05$. Under the null hypothesis of no LRD, the 95% confidence interval of the rejection percentage is]3.6, 6.4[, for the $\alpha = 0.05$ nominal significance level of the test. Results suggest that, in general, the bootstrap testing procedure is able to improve the tests. Moreover, each one of bootstrap tests has, in most of cases, better size properties than the corresponding original tests. It is more evident for *AR*(1) models with negative coefficients or with moderate positive coefficients. It is true also for the *MA*(1) models (not reported in this paper) when the truncation lag (11) is used. Exception is the case of *MA*(1) model with $\theta = -0.9$. We note that for *MA*(1) model with $\theta = -0.9$ all original tests overestimate the nominal size, whereas, the bootstrap tests underestimate the nominal size

when the symmetric p-value (12) is used and overestimate the nominal size when the equal-tail p-value (13) is used.

TABLE I
EMPIRICAL SIZE OF THE LO'S TESTS FOR $AR(1)$ MODELS ($\alpha = 0.05$)

φ	N	q_L lags (9)			q_W lags (10)			q_M lags (11)		
		Orig.	**Symm.**	**Equal.**	**Orig.**	**Symm.**	**Equal.**	**Orig.**	**Symm.**	**Equal.**
-.9	100	**5.1**	22.8	16.7	0.7	13.5	8.2	3.5	3.4	13.3
	200	1.2	12.3	7.7	1.1	8.2	**4.5**	0.5	6.6	**6.3**
	300	0.3	8.6	**5.0**	1.0	6.5	**4.2**	9.8	1.9	8.1
	500	0.1	**6.3**	**3.8**	2.6	**5.9**	**4.6**	2.9	**4.2**	**4.5**
	1000	0.3	**4.6**	**3.6**	**3.7**	**6.2**	**4.9**	13.6	2.7	6.8
-.5	100	0.7	**6.3**	**4.8**	10.5	**6.1**	**6.0**	0.2	**6.3**	6.7
	200	1.2	**4.4**	**3.7**	6.9	**4.4**	**4.6**	0.3	**4.0**	3.4
	300	1.9	**5.5**	**4.2**	**6.2**	**5.5**	**4.6**	1.7	**5.3**	**4.4**
	500	3.4	6.7	**6.4**	7.4	**5.9**	**6.2**	3.1	6.5	**6.2**
	1000	3.2	**5.6**	**4.5**	6.7	**5.7**	**5.2**	3.1	**5.4**	**4.7**
-.4	100	2.2	**4.8**	**5.2**	10.8	**4.8**	**5.5**	0.3	**4.7**	**4.4**
	200	2.3	**5.3**	**4.9**	7.9	**5.4**	**5.8**	1.1	**4.4**	**5.0**
	300	2.0	**4.5**	**4.3**	7.7	**5.6**	**4.7**	1.1	**4.4**	**3.7**
	500	**4.2**	**6.3**	**5.7**	8.9	7.3	6.5	3.5	6.5	**5.2**
	1000	**3.9**	**4.3**	**4.0**	**6.3**	**5.2**	**5.1**	3.4	**4.3**	**4.8**
-.1	100	7.4	7.4	8.9	10.9	7.8	8.8	0.5	**5.9**	**5.0**
	200	**5.6**	**5.5**	**5.5**	8.8	**5.2**	**5.9**	0.9	**4.9**	**5.0**
	300	**5.5**	6.5	6.9	8.0	**6.2**	7.5	1.5	**5.9**	**5.3**
	500	**6.1**	**4.2**	**5.5**	8.0	**4.0**	**6.2**	**3.8**	**4.3**	**5.3**
	1000	**5.3**	**4.2**	**4.9**	6.8	**4.1**	**5.5**	**3.6**	**4.4**	**4.8**
.1	100	**5.8**	7.8	7.4	**5.3**	11.0	8.3	0.9	**3.9**	**4.8**
	200	**5.4**	9.2	7.5	**6.1**	12.0	9.1	1.4	7.1	**6.3**
	300	**5.4**	8.3	6.9	**5.7**	10.2	7.9	2.6	**6.4**	**5.7**
	500	**5.3**	7.5	**6.4**	**5.3**	9.0	7.2	2.6	**6.4**	**5.0**
	1000	**6.1**	6.9	**6.2**	**6.2**	7.3	6.9	**4.0**	**5.6**	**5.8**
.4	100	3.4	7.7	6.9	**5.7**	9.5	8.4	2.0	**5.0**	**5.8**
	200	**4.9**	7.8	6.6	6.8	8.7	8.6	2.5	6.9	6.7
	300	3.4	**6.0**	**5.4**	**5.8**	7.0	**5.5**	2.6	**5.4**	**4.9**
	500	**3.9**	**6.0**	**4.9**	**5.6**	**6.2**	**5.9**	2.6	**5.4**	**4.3**
	1000	**5.3**	**5.9**	**5.5**	6.6	**6.2**	**5.6**	**4.6**	**5.3**	**5.3**
.5	100	2.7	6.7	**5.8**	**4.1**	9.1	8.4	1.9	**4.2**	**4.5**
	200	2.9	7.8	**6.0**	6.8	9.0	7.5	2.3	7.6	**5.5**
	300	2.6	**4.9**	**4.5**	**5.9**	7.5	**5.5**	2.7	**4.5**	**4.9**
	500	3.5	6.6	**5.6**	7.1	6.7	6.5	3.3	**6.3**	**5.8**
	1000	**3.7**	**4.2**	**4.5**	**5.1**	**4.7**	**5.3**	**3.8**	**4.2**	**4.3**
.9	100	1.0	3.0	2.9	0.5	7.8	**5.6**	0.5	10.7	7.2
	200	1.4	**5.8**	**4.1**	1.7	11.3	8.9	**4.5**	15.3	10.4
	300	2.0	**5.1**	**4.0**	3.4	8.0	**5.8**	9.8	10.2	8.2
	500	2.1	**4.5**	**4.8**	**6.2**	7.9	7.0	15.8	10.3	7.9
	1000	2.3	**5.0**	**4.3**	6.5	6.9	**5.7**	12.6	8.2	**6.0**

Note: **Orig.** denotes the original test, **Symm.** denotes the corresponding bootstrap test with symmetric bootstrap p-value and **Equal.** denotes that one with equal-tail p-value.
The fitted model is $AR(p)$ with p selected using BIC criterion.

Over all the parameter values, the dispersion of the sizes for each bootstrap test is smaller than that of the original test. For example, there are only ten out of forty bootstrap empirical sizes that differ significantly from the nominal size for the Lo statistic with truncation lag (9) and with p-value (13), compared to twenty four significant deviations for the corresponding original test (see third and fifth columns to table I). On the other hand the size underestimates encountered by the original tests for $AR(1)$ models with $|\varphi| = 0.5$ and $|\varphi| = 0.9$ are adjusted by the bootstrap procedure.

TABLE II
EMPIRICAL POWER OF THE LO'S TESTS FOR *ARFIMA*(0, *d*, 0) MODELS (α = 0.05)

d	N	q_L lags (9)			q_W lags (10)			q_M lags (11)		
		Orig.	**Symm.**	**Equal.**	**Orig.**	**Symm.**	**Equal.**	**Orig.**	**Symm.**	**Equal.**
.05	100	4.0	12.3	9.9	6.2	**14.9**	10.5	0.3	6.2	5.9
	200	5.8	14.6	11.4	7.0	**16.0**	12.4	1.1	7.6	5.6
	300	6.0	14.9	10.6	7.8	**17.1**	12.6	1.6	9.1	6.1
	500	7.5	17.3	11.7	9.1	**19.5**	14.2	3.4	10.0	6.3
	1000	10.4	18.6	12.6	11.7	**21.6**	13.5	5.3	12.5	8.5
.25	100	3.8	25.8	17.2	19.5	**35.3**	26.5	0.8	9.5	6.3
	200	17.1	36.0	26.6	40.3	**44.8**	35.0	3.2	24.2	16.0
	300	24.6	39.3	29.4	48.4	**48.6**	39.0	8.4	28.7	20.5
	500	38.0	48.4	37.9	**64.4**	59.8	48.2	22.7	39.8	30.3
	1000	54.9	60.6	51.8	**79.9**	68.9	60.2	39.0	52.3	41.8
.3	100	3.5	25.0	17.6	23.9	**36.8**	28.2	0.8	10.1	6.7
	200	13.3	31.4	23.4	43.3	**44.2**	33.6	3.5	24.0	16.9
	300	27.6	44.9	33.4	**61.0**	53.4	45.0	14.3	35.4	25.4
	500	42.1	51.2	41.4	**71.3**	60.3	51.1	30.6	46.8	35.4
	1000	58.5	63.6	53.4	**84.3**	72.1	63.7	47.7	58.4	48.4
1/3	100	3.3	25.8	17.3	30.0	**39.5**	28.7	0.5	13.1	7.1
	200	13.5	35.7	26.7	**48.6**	48.1	38.2	4.5	29.6	19.7
	300	26.2	42.9	32.9	**62.8**	54.2	44.5	16.3	36.1	27.0
	500	41.5	51.6	40.6	**73.2**	61.9	52.9	32.4	46.5	36.5
	1000	61.5	66.5	55.9	**87.8**	75.1	66.9	55.4	62.6	53.7
.45	100	1.0	18.8	11.9	32.3	**37.5**	28.2	0.7	16.1	10.9
	200	6.4	33.3	23.6	**56.3**	49.9	38.0	9.3	35.7	25.3
	300	18.6	40.1	30.8	**68.3**	55.1	45.5	28.8	43.8	34.9
	500	39.4	53.3	41.3	**81.8**	66.2	57.2	52.6	57.6	48.6
	1000	65.1	69.3	61.5	**92.9**	79.6	72.6	75.9	75.1	65.2

See note to Table I

Table II and table III report the power of the tests. In bold face is printed the greatest value of all empirical powers of a case study. The simulation results verify that the power values of the bootstrap tests, with truncation lag (9), are greater than those of the asymptotic tests. It is also true for other two truncation lags, (10) and (11), for small sample sizes (always for $N = 100$ and almost always for $N = 200, 300$).

In general, the bootstrap tests with symmetric *p*-value (12) yield better powers than those of the bootstrap tests with equal-tail p-value (13). For the mixed *ARFIMA*(1, *d*, 0) models with $\varphi = 0.9$, the powers are in general low. Original or bootstrap tests with truncation lag (11) perform better in these cases.

We note that the original test statistic (8) using the truncation lag (10), proposed by [14], performs well for *AR*(1) and *MA*(1) models with positive coefficients. It also has good powers for *ARFIMA*(0, *d*, 0) models with moderate or large values of *d*. That is not true for mixed *ARFIMA*(1, *d*, 0) models. However, the bootstrap procedure always improves the performance of the tests for sample size of 100 observations.

V. CONCLUSIONS

In this paper we investigate a bootstrap hypothesis testing procedure. From results obtained by the Monte Carlo study, we conclude that the block bootstrap method, with blocks composed of cycles, provides a practical and effective testing procedure for detecting the LRD, especially for short time series. This bootstrap testing procedure, based in modified *R/S* statistic, has in general better size properties than the corresponding asymptotic test without losing significant power.

TABLE III

EMPIRICAL POWER OF THE LO'S TESTS FOR *ARFIMA*(1, *d*, 0) MODELS (α = 0.05)

d (φ)	N	q_L lags (9)			q_W lags (10)			q_M lags (11)		
		Orig.	Symm.	Equal.	Orig.	Symm.	Equal.	Orig.	Symm.	Equal.
.3 (.5)	100	1.0	7.9	5.7	3.0	**16.9**	11.8	1.0	13.3	9.4
	200	0.7	11.6	6.7	14.1	**23.4**	16.9	4.2	22.1	15.5
	300	1.0	18.2	11.6	26.2	**32.5**	23.2	14.1	29.6	22.3
	500	5.1	26.2	19.5	**39.5**	37.3	30.7	32.0	37.0	29.7
	1000	20.9	41.3	32.0	**56.8**	**56.8**	47.2	50.3	54.8	45.3
.3 (.9)	100	0.6	6.5	8.7	1.0	4.8	5.6	0.3	**17.8**	14.0
	200	0.3	6.7	7.4	0.9	7.6	8.2	14.6	**18.3**	15.4
	300	0.2	5.3	5.1	1.2	8.7	7.8	**32.5**	15.0	11.2
	500	0.4	3.4	3.5	0.9	8.3	6.4	**52.6**	13.7	9.4
	1000	0.6	4.6	4.6	1.5	14.3	9.2	**67.0**	18.4	14.6
.45 (.5)	100	0.6	3.2	2.5	0.3	10.1	6.9	0.5	**14.9**	9.3
	200	0.4	8.0	4.9	5.8	24.8	17.7	10.7	**28.5**	21.6
	300	0.5	10.8	5.7	15.4	30.0	21.3	30.4	**34.2**	26.8
	500	0.6	18.3	10.8	37.7	42.1	33.2	**55.0**	45.9	37.6
	1000	5.2	36.8	26.6	59.9	59.7	50.8	**73.3**	63.9	55.4
.45 (.9)	100	1.5	10.2	17.0	0.8	4.8	7.2	0.2	**22.0**	18.7
	200	1.0	10.6	18.7	1.2	9.3	11.4	**25.2**	21.4	20.6
	300	0.5	9.2	14.6	0.8	7.0	7.0	**49.0**	21.2	18.6
	500	0.1	5.5	8.7	1.2	7.4	6.5	**70.2**	17.0	14.4
	1000	0.0	2.2	5.3	0.7	13.6	10.0	**85.8**	20.3	15.7

See note to Table I

REFERENCES

[1] H. E. Hurst, "Long-term storage capacity of reservoirs", *Transactions of the American Society of Civil Engineers*, vol. 116, pp. 770–808, 1951.

[2] G. Teyssiere and A. P. Kirman, Ed., *Long Memory in Economics*, Berlin, Germany: Springer, 2007.

[3] R. T. Baillie, "Long memory processes and fractional integration in econometrics", *Journal of Econometrics*, vol. 73, pp. 5-59, 1996.

[4] A. I. McLeod and K.W. Hipel, "Preservation of the Rescaled Adjusted Range 1. A reassessment of the Hurst Phenomenon", *Water Resources Research*, vol. 14, No. 3, pp. 491-508, 1978.

[5] J. Beran, *Statistics for long-memory processes*, New York: Chapman & Hall, 1994.

[6] A. W. Lo, "Long-Term Memory on Stock Market Prices", *Econometrica*, vol. 59, No. 5, pp. 1279-1313, 1991.

[7] L. Ekonomi and A. Butka, "Jackknife and bootstrap with cycling blocks for the estimation of fractional parameter in ARFIMA model", *Turk. J. Math.*, vol. 35, No. 1, pp. 151-158, 2011.

[8] D. Guegan, "How can we Define the Concept of Long Memory? An Econometric Survey", *Econometric Reviews*, vol. 24, No. 2, pp. 113-149, 2005.

[9] G.E.P. Box and G.M. Jenkins, *Time series analysis, forecasting and control*. Oakland, California: Holden-Day, 1976.

[10] C. W. Granger and R. Joyeux, "An introduction to long memory time series models and fractional differencing", *Journal of Time Series Analysis*, vol. 1, pp. 15–39, 1980.

[11] J. Hosking, "Fractional differencing", *Biometrika*, vol. 68, No. 1, pp. 165-176, 1981.

[12] B. B. Mandelbrot, "Statistical Methodology for Non-Periodic Cycles: From the Covariance to R/S Analysis", Annals of Economic and Social Measurements, vol. 1, No. 3, pp. 256-290, 1972.

[13] V. Teverovsky, M. S. Taqqu, and W. Willinger, "A critical look at Lo's modified R/S statistic", *Journal of Statistical Planning and Inference*, vol. 80, pp. 211-227, 1999.

[14] W. Wang, P. H. A. J. M. Van Gelder, J. K. Vrijling, and X. Chen, "Detecting long-memory: Monte Carlo simulations and application to daily streamflow processes", *Hydrol. Earth Syst. Sci.*, vol. 11, pp. 851-862, 2007.

[15] A. Murphy and M. Izzeldin, "Bootstrapping long memory tests: Some Monte Carlo results", *Computational Statistics and Data Analysis*, vol. 53, pp. 2325-2334, 2009.

[16] B. Efron, "Bootstrap methods: another look at the jackknife", *The Annals of statistics*, vol. 7, No. 1, pp. 1-26, 1979.

[17] P. Hall, "Resampling a coverage pattern", *Stochastic Processes and their Applications*, vol. 20, pp. 231-246, 1985.

[18] E. Carlstein, "The use of subseries values for estimating the variance of a general statistic from a stationary sequence", *The Annals of Statistics*, vol. 14, No. 3, pp. 1171-1179, 1986.

[19] H. Kunsch, "The jackknife and bootstrap for general stationary observations", *The Annals of Statistics*, vol. 17, No. 3, pp. 1217-1241, 1989.

[20] R. Liu and K. Singh, "Moving blocks jackknife and bootstrap capture weak dependence", Exploring the Limits of Bootstrap, Wiley, New York, pp. 225-248, 1992.

[21] S. N. Lahiri, "Theoretical comparisons of block bootstrap methods", *The Annals of Statistics*, vol. 27, No. 1, pp. 386-404, 1999.

[22] S. N. Lahiri, "On the moving block bootstrap under long range dependence", *Statistics & Probability Letters*, vol. 18, pp. 405-413, 1993.

[23] E. Carlstein, K. Do, P. Hall, T. Hesterberg, and H. R. Kunsch, "Matched-block bootstrap for dependent data", *Bernoulli*, vol. 4, No. 3, pp. 305-328, 1998.

[24] T. Hesterberg, "Matched-Block Bootstrap for Long Memory Processes", MathSoft, Inc., Seattle, USA, Research Report No. 66, 1997.

[25] L. Ekonomi and A. Butka, "The Validity of the Bootstrap with Cycling Blocks in Stationary Time Series", *Pensee Journal*, vol. 76, No. 3, pp. 285-303, 2014.

[26] D. Politis and J. A. Romano, "Circular block resampling procedure for stationary data", Exploring the Limits of Bootstrap, Wiley-New York, pp. 263-270, 1992.

[27] J.G. MacKinnon, "Bootstrap Hypothesis Testing", Department of Economics, Queen's University, Canada, Working Paper No. 1127, 2007.