A BOOTSTRAP APPROACH FOR DETECTION OF LONG RANGE DEPENDENCE

Argjir Butka[1], Gjergji Capollari[2]

[1]"Fan S. Noli" University, Shetitorja "Rilindasit", Korce, Albania,
argjirbutka@yahoo.com

[2]"Fan S. Noli" University, Shetitorja "Rilindasit", Korce, Albania,
gjergjicapollari@yahoo.com

Abstract

Time series with long range dependence or long memory appear in many contexts, for example in hydrology, financial economics, astrology, psychology, cardiology, etc. The Autoregressive Fractionally Integrated Moving Average, ARFIMA(p,d,q), model has widely been used to represent a time series with long memory. The intensity of the memory can be measured by the fractional difference parameter, d. Since the classical methods cease to be valid in presence of long range dependence, the detection of long memory has important implications. Some classical tests on the long memory exhibit size distortions in small samples. It is well known that the bootstrap techniques are a good remedy in these situations. A bootstrap method, widely used for dependent data, is the block bootstrap method. In this paper we consider a bootstrap technique, which uses blocks compound by cycles, to perform hypothesis testing on the fractional difference parameter in the ARFIMA model. The approach is applied to two semi-parametric estimators of fractional difference parameter, well known in the literature, derived by the log-periodogram and the smoothed log-periodogram regression methods. The R software is used to conduct a simulation study. Both short memory and long memory processes are generated in order to investigate the behaviour of the proposed bootstrap tests. By means of Monte Carlo experiments the performance of the tests is obtained calculating the empirical significance level and empirical power of the tests. Different block sizes are considered.

Keywords: *Long range dependence, ARFIMA model, fractional difference parameter, block bootstrap, hypothesis test.*

1. INTRODUCTION

Long memory or long range dependence is an important aspect of many types of data that has gained much attention during the last decades. Long memory is a characteristic of a stationary process in which the underlying time series realizations display significant temporal dependence at very distant observations. The autocorrelation function of a process with long memory takes far longer to decay than that of the common processes modeled by the vast class of ARMA (Auto Regression Moving Average) models usually called as short memory processes. The presence of long memory firstly evidenced in hydrology is now evidenced in many fields, such as economics, finance, network traffic, psychology etc. It is important to distinguish long memory processes from short memory ones. The classical methods applied in short memory processes are not still valid in the presence of long memory. Also, the prediction and the effects of shocks in a time series data are very different for long and short memory processes (for theory and methods of long memory see Palma (2007)). As an example, note that the decay of the variance of the sample mean depends on the memory of the process having different rates of convergence for short or long memory. This distinction between rates of convergence is certainly relevant when computing tests and confidence intervals for the mean of the process. If the distinction is ignored standard statistical methods may yield unrealistic results.

There are several tests in literature in order to test the presence of the long memory (Beran 1994, chapter 4-6). The classical tests are based on the asymptotic distribution of an estimator of a parameter, which by definition provides a measure of the long memory in the data. It is well known that these tests suffer from finite sample distortion.

In this paper we present a bootstrap technique of hypothesis testing in order to test for long memory. A simulation Monte Carlo study is conducted to empirically estimate the performance of these bootstrap tests.

The outline of the paper is as follows. In Section 2 we specify some definitions of long memory and ARFIMA process. Then we describe two semi-parametric tests for long memory. In section 3 we consider a review of the bootstrap resampling techniques for time series and their application on hypothesis testing. Section 4 shows some simulation results.

2. LONG MEMORY AND ARFIMA MODELLING

2.1. Long memory

In contrast to the short memory of weakly dependet processes such as all Gaussian finite-order ARMA models (see Box & Jenkins, 1976, chapter 4), natyral phenomena often display long-term memory in the form of non-periodic cycles in the sense that they exhibit dependence even over very long time spans.

Long memory and long-range dependence are synonymous notions. Even though the long memory notion recently has become important, there are various definitions of it. When definitions are given, they vary from author to author. Most of the definitions of long range dependence 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. In fact, long range dependent processes are characterized by slowly decaying autocorrelations or by a spectral density function exhibiting a pole at the origin. Let $X_t, t = 1, 2, ...$ be a stationary process with finite variance. We can write $E(X_t) =$ « and

$Cov(X_t, X_{t+k}) = \gamma(k)$ for any moment *t*. Here $\gamma(k)$ for $k \in Z$ denotes the autocovariance function of the process. The autocorrelation function is then defined by equation

(1) $$\rho(k) = \frac{\gamma(k)}{\gamma(0)}$$

Also for a stationary process we can define its spectral density function by equation

(2) $$f(\lambda) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma(k)e^{ik\lambda} = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma(k)\cos(k\lambda)$$

In general long memory is defined by the fact that the autocorrelations of the process are absolutely non-summable, i.e. $\sum_{k=-\infty}^{\infty}\left|\rho(k)\right| = \lim_{n\to\infty}\sum_{k=-n}^{n}\left|\rho(k)\right| = \infty$. However, there are alternative definitions. According to Beran (1994) let give two of more common definitions of long memory (Beran, 1994, p. 42).

Definition in time domain:

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

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

Then X_t is called a stationary process with long memory or long-range dependence or strong dependence, or a stationary process with slowly decaying or long-range correlations.

Definition in spectral domain:

Suppose that there exists a real number $\in (0,1)$ and a constant $c_f > 0$ not depended on λ such that

(4) $$\lim_{\lambda\to 0}\frac{f(\lambda)}{c_f\left|\lambda\right|^{-}} = 1$$

Then X_t is called a stationary process with long memory or long-range dependence or strong dependence.

2.2. The ARFIMA(p,d,q) model

The ARFIMA(p,d,q) model proposed by Granger and Joyeux (1980) and Hosking (1981) is a generalization of an ARIMA (p,d,q) model obtained by permitting the differential parameter, *d* to take any real value instead being restricted to integer values. It is defined as

(5) $$\Phi(B)(1-B)^d(X_t - \mu) = \Theta(B)\varepsilon_t$$

where B is the back-shift operator and $\{\varepsilon_t\}$ is a white noise process with zero mean and finite variance σ_ε^2. All the roots of the polynomials $\Phi(B)$ and $\Theta(B)$ lie outside the unit circle, while $-0.5 < d < 0.5$ is the fractional differencing parameter. The process $\{X_t\}_{t\in Z}$ is both stationary and invertible. For $0 < d < 0.5$ the process exhibits long memory in the sense of equation (3) if we set $\alpha = 1 - 2d$. If $-0.5 < d < 0$ we say that the process has intermediate memory and when $d = 0$ we have the ARMA (p,q) process (short memory). In the simple case when $p = q = 0$ equation (5) defines a ARFIMA(0,d,0) process also called a FI(d) (Fractionally Integrated) process with parameter d. The spectral density function of the FI(d) process is $f(µ) = \frac{±_£^2}{2®}\left(2\sin\left(\frac{µ}{2}\right)\right)^{-2d}$, $µ > 0$. In this case equation (4) holds if we set $c_f = ±_£^2$ and $\;= 2d$.

As we see the intensity of the memory can be measured by the difference parameter, d. Since several estimators with known asymptotic distribution are available in the literature we are able to perform a hypothesis testing for the null hypothesis $H_0 : d = 0$. Supposing that the time series data at the hand may be sufficiently approximated by an ARFIMA model the null hypothesis would be that the data series has no long memory.

2.3. Regression estimators of the fractional parameter

There have been proposed several estimators for the fractional parameter in the literature. In our simulation study, we use two semi-parametric estimators based on the regression equation constructed from the logarithm of the spectral density function.

One of the most used estimators is proposed by Geweke and Porter-Hudak (1983) and is usually noted as GPH estimator. This estimator is based on the log-periodogram regression and is obtained applying the least square method to the linear regression

(6) $u_j = a + bv_j + e_j,\ j = 1, 2, ..., g$

in which $u_j = \ln\left(I(µ_j)\right)$, $a = \ln\left(f_{ARMA}(0)\right)$, $b = -d$, $v_j = \ln\left(4\sin^2\left(µ_j/2\right)\right)$ and $e_j = \ln\left(\frac{I(µ_j)}{f(µ_j)}\right)$ are the random errors. Here $f_{ARMA}(µ)$ denotes the spectral density function of the corresponding ARMA(p,q) process of (5), $I(µ) = \frac{1}{2®}\left[R(0) + 2\sum_{k=1}^{n-1} R(k)\cos(kµ)\right]$ is the periodogram of the observed data $x_1, x_2, ..., x_n$, and $µ_j = \frac{2® j}{n}$ for $j = 1, 2, ..., g < n$ are the Fourier frequencies. Then *d* can be estimated by

(7) $$\hat{d}_{GPH} = -\hat{b} = -\frac{\sum_{j=1}^{g}\left(v_j - \bar{v}\right)u_j}{\sum_{j=1}^{g}\left(v_j - \bar{v}\right)^2}$$

where $\bar{v} = \frac{1}{g}\sum_{j=1}^{g} v_j$. An important practical problem in the implementation of the GPH estimator is the choice of the truncation parameter $g = g(n)$, which has an important influence on the bias and variance of the estimator. Hurvich, Deo and Brodsky (1998) showed that $g(n) = O\left(n^{4/5}\right)$ is an optimal value which can minimize the MSE (Mean Squared Error) of $\hat{d}_{GPH}$. Reisen (1994) suggested using a smoothed periodogram instead of the sample periodogram. The smoothed periodogram is defined as

(8) $$I_{SP}(µ) = \frac{1}{2®}\left[ª_0 R(0) + 2\sum_{k=1}^{m} ª_k R(k)\cos(kµ)\right]$$

where $\{a_k\}$ is a set of weights called lag window. Reisen (1984) used the Parzen window for the lag window $\{a_k\}$ .Then, the smoothed periodogram (hereinafter denoted as SP) estimator of *d* is

$$(9) \qquad \hat{d}_{SP} = -\frac{\sum_{j=1}^{g}\left(v_j - \overline{v}\right)u_j^*}{\sum_{j=1}^{g}\left(v_j - \overline{v}\right)^2}$$

where $u_j^* = \ln\left(I_{SP}(\mu_j)\right)$ while v_j are as before. For this estimator we have to choose two tuning parameters, the truncation parameter of the regress $g = g(n)$ and the truncation parameter of the lag window, $m = m(n)$ in the smoothed periodogram.

Both $\hat{d}_{GPH}$ and $\hat{d}_{SP}$ are asymptotically unbiased and normally distributed but they suffer from finite-sample bias.

3. BOOTSTRAP HYPOTHESIS TESTING.

The bootstrap methodology, proposed originally by Efron (1979) is a computer-intensive technique that presents solutions when the parametric methods and statistical theory do not work. The idea behind the bootstrap is treating the original data of a sample as to be the population of interest and resampling with replacement new data from the original data. When the data at the hand are not independent and identically distributed (i.i.d.), such as a time series, the resample 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. In the last years many bootstrap methods for time series have been developed. The most common are the block bootstrap methods. They consider a set of consecutive observations to define blocks. Then the bootstrap time series is constructed by resampling (with replacement) blocks and concatenating them still yielding a time series usually with equal length to original series. It is pretended that the data within each block have the same dependence structure as they have had in the original time series and the data in different blocks are asymptotically independent. Different methods differ in the way as blocks are constructed. For a theoretical comparison of some methods see Lahiri (1999).

In this paper we consider blocks that are composed by one or more consecutive cycles. A cycle is defined as a pair of alternating positive and negative terms that are made when the time series crosses its mean (supposed to be zero). This bootstrap method is proposed by Park and Willemain (1999) to estimate the sample mean of a stationary time series with short memory. This method resamples a set of random-length cycles. The number of cycles is also random. We consider both the non-overlapping block and the moving block bootstrap. In the moving block case a cycle is considered as a single observation. The number of cycles that should be included in a block is a tuning parameter that should be determined appropriately since it is close related with the block length.

In briefly, the implementation of the bootstrap for hypothesis testing can be summarized as follows (for details see MacKinnon (2007)). Suppose that $\hat{2}$ is the observed value of a test statistic 2 with cumulative distribution function F_2 under the null hypothesis. Let we wish to perform a test at level $\sim$ that rejects the null hypothesis when $\hat{2}$ is in the upper tail. Than the p-value of $\hat{2}$ is $p(\hat{2}) = 1 - F_2(\hat{2})$. If we knew F_2 , we would simply calculate $p(\hat{2})$ and reject the null whenever $p(\hat{2}) < \sim$. But in practice F_2 is unknown or the asymptotic approximation of it may perform poorly for finite sample sizes. An increasingly popular alternative is to perform a bootstrap test. In the spirit of the bootstrap we first generate B bootstrap samples and compute the bootstrap test statistic 2_b^* for each one, most commonly by the same procedure used to calculate $\hat{2}$ from the real sample. Then, if we wish to reject when $\hat{2}$ is in the upper tail, the bootstrap p-value is

$$(10) \qquad \hat{p}^*(\hat{2}) = \frac{1}{B}\sum_{b=1}^{B} I(2_b^* > \hat{2})$$

where $I(\bullet)$ denotes the indicator function. When we wish to perform a two-tailed test, and we are willing to assume that 2 is symmetrically distributed around zero, we can use the symmetric bootstrap p-value

$$(11) \qquad \hat{p}_s^*(\hat{2}) = \frac{1}{B}\sum_{b=1}^{B} I\left(\left|2_b^*\right| > \left|\hat{2}\right|\right)$$

If we are not willing to make this assumption, we can use the equal-tail bootstrap p-value

$$(12) \qquad \hat{p}_{et}^*\left(\hat{2}\right) = 2\min\left(\frac{1}{B}\sum_{b=1}^{B} I\left(2_b^* \le \hat{2}\right), \frac{1}{B}\sum_{b=1}^{B} I\left(2_b^* > \hat{2}\right)\right)$$

If the mean of 2_b^* is far from zero, $\hat{p}_s^*$ and $\hat{p}_{et}^*$ may be very different, and tests based on them may have very different properties under both null and alternative hypothesis (MacKinnon, 2007, p. 4).

In our Monte Carlo study we use an iterative bootstrap procedure, called pretesting bootstrap, proposed in Davidson and MacKinnon (2000). The idea is to start with a relatively small value of B and then increase it, if necessary, until we are confident, at some prechosen significance level, that $\hat{p}^*(\hat{2})$ is either greater or less than $\sim$. If the procedure stops with a small value of B, $\hat{p}^*(\hat{2})$ may differ substantially from $p^*(\hat{2})$, the asymptotic bootstrap p-value with infinite replications, but only when $p^*(\hat{2})$ is not close to $\sim$. Thus we ensure low probability that the feasible and ideal bootstrap tests yield different outcomes (Davidson and MacKinnon, 2000, chapter 3).

4. MONTE CARLO RESULTS

4.1 The experiment design

To test $H_0 : d = d_0$ against $H_1 : d \neq d_0$ we use the test statistic $\hat{2} = \dfrac{\hat{d} - d_0}{sd(\hat{d})}$, where $\hat{d}$ is the GPH or SP estimator and $sd(\hat{d})$ is the corresponding estimator of the standard deviation. In our case we set $d_0 = 0$. Since $\hat{2}$ is asymptotically pivotal it is an appropriate statistic to be used in a bootstrap hypothesis testing procedure.

In a Monte Carlo experiment we considered a wide class of models from the ARFIMA(p,d,q) processes. In order to obtain empirically size and power estimates of the test we considered ARFIMA(1,d,0) models with auto regression coefficient $¤ = 0.1, 0.5$, ARFIMA(0,d,1) with moving average coefficient $¯ = 0.1, 0.5$ and ARFIMA(0,d,0). In all models we considered the values $d \in [-0.45, 0.45]$ in increments of 1.5. This set of values includes both the short memory (for $d = 0$) and the long memory (for $d \neq 0$) cases. Using the ‘fracdiff” package in

R we generated 1000 replicated time series of each model under consideration with sample size $n = 100, 300$ and 200 first terms removed in order to reduce the effect of initial values. The white noise process was generated from the standard normal distribution. Then in each simulated case we calculated the empirical size or power of the test. We used $g(n) = Cn^{0.8}$ terms in periodogram regression of $\hat{d}_{GPH}$ estimator with $C = 0.3, 0.45, 1$. In $\hat{d}_{SP}$ estimator case we used $g(n) = n^{0.5}$ regression terms and the value $m = n^{0.9}$ as the truncation parameter of Parzen lag window. We set the block length approximately at order $O(n^{0.5})$ using different numbers of cycles per block. For each combination of the parameters we applied both the moving and non-overlapping block bootstrap methods. We used $B_{\min} = 99$, $B_{\max} = 2999$ and $= 0.007$ in the pretesting bootstrap procedure (Davidson and MacKinnon, 2000, p.10). The nominal size (type I error) of the test was chosen to be $\sim = 0.10, 0.05, 0.01$. We used both the symmetric and equal-tail bootstrap p-values defined by equations (11) and (12).

4.2 Results

Firstly we note that bootstrapping with non-overlapping or overlapping (moving) blocks yielded similar results. Also, the number of cycles per block needs not to be large for our method to perform better. But, depending on the choice of other parameters the results were highly various. There occurred underestimations or overestimations of the test size in most of the cases under considerations.

The use of GPH estimator with $\hat{p}_{et}^{*}$ as a bootstrap p-value yielded always overestimations of the test size, while underestimates of the size occurred when the GPH estimator and $\hat{p}_{s}^{*}$ bootstrap p-value were used. These underestimations were large for C=0.3 and moderate for C=0.45 with an expectation in the case of AR(1) model with coefficient 0.5.

Tests based on the SP estimator underestimated the nominal size for all combinations of parameters and for both $\hat{p}_{s}^{*}$ and $\hat{p}_{et}^{*}$ bootstrap p-values.

The use of regression standard error deviation of $\hat{d}_{GPH}$ or $\hat{d}_{SP}$ instead of the corresponding asymptotic standard deviation yielded size estimations closer to the nominal sizes. Table 1 reports the empirical size estimations implementing the bootstrap procedure presented above using the test statistic based on SP estimator and regression estimator of its standard deviation. Results in table 1 show that the bootstrap test based on $\hat{d}_{SP}$ estimator standardized by the regression standard deviation has good sizes even for ARFIMA(1,0,0) model with coefficient ¤ $= 0.5$.

In table 2 we report the power estimations against the fractional integrated models for the same set of parameters as those reported in table 1. We can see that the bootstrap test has not high power against the fractionally integrated alternatives. It is true in most cases of bootstrap tests using the symmetric bootstrap p-value.

The asymptotic tests generally have biased sizes in the presence of ARMA part as we can see from results of one of the better results cases reported in table 3. But they are often more powerful than the bootstrap tests.

4.3 Conclusions

From the whole Monte Carlo experiment we may conclude that testing the null hypothesis of no memory by mean of the bootstrap procedure presented in this paper must be carefully conducted. Due to a large number of the tuning parameters there can occur misleading results under both the null and alternative hypothesis. In general the bootstrap tests based on both GPH and SP estimators are conservative tests in the sense that they tend to reject the null less frequently than the nominal or to not reject the null of no long memory under the alternative. However, it should be noted that $n = 100$ and $n = 300$ may not be large enough for gaining of power or for some of the methods to perform better in the presence of long memory.

We tentatively suggest using of $z^{*}_{SP} = \dfrac{\hat{d}^{*}_{SP} - \hat{d}_{SP}}{sd_{reg}(\hat{d}^{*}_{SP})}$, where $\hat{d}^{*}_{SP}$ is the bootstrap analogue of (9) and $sd_{reg}\left(\hat{d}^{*}_{SP}\right)$ is the regression estimator of its standard deviation.

Recently many authors suggest the use of a portfolio of tests instead of only one test for detecting the presence of long memory.

Tab. 1 Rejection percentage for short memory data for H_0:d=0 vs. H_1:d≠0.

Nominal size	ARFIMA (0,0,0)	ARFIMA(1,0,0)		ARFIMA(0,0,1)	
		¤ = 0.1	¤ = 0.5	¯ = 0.1	¯ = 0.5
10%	6.0	6.7	8.2	5.7	8.5
5%	2.3	2.9	3.3	2.9	3.7
1%	0.7	0.6	0.8	0.34	0.8

Notes: The test statistic is based on SP estimator with regression standard deviation. The bootstrap symmetric p-value is used; n=100; the number of cycles of a block: 2

Tab. 2 Rejection percentage for long memory data for H_0:d=0 vs. H_1:d≠0.

Nominal size	d	ARFIMA (0,d,0)	ARFIMA(1,d,0)		ARFIMA(0,d,1)	
			¤ = 0.1	¤ = 0.5	¯ = 0.1	¯ = 0.5
10%	-0.45	23.5	25.4	23	21	22.5
	-0.30	20.4	16.5	15.6	17.3	20.1
	-0.15	12.2	12.7	9.1	11.6	15.9
	0.15	7.2	6.7	18.9	4.1	4.1
	0.30	15.5	18.2	43.4	13.6	4.2
	0.45	38	40.8	65.5	35.2	10.8
5%	-0.45	12.3	13.3	11.3	11.8	10.2
	-0.30	8	9.2	8.7	9.4	8.9
	-0.15	6.6	4.3	4.2	6.4	5.6
	0.15	2	3	10.4	2.3	1.3
	0.30	6.8	9.1	26.1	7.9	1.1
	0.45	24.7	25.9	50.9	19.1	4.7

1%	-0.45	2.2	3.4	1.8	2.4	2.5
	-0.30	1.8	1.9	1.9	1.7	1.5
	-0.15	1.3	1.3	1.1	0.9	1.7
	0.15	0.4	0.4	1.5	0.2	0.3
	0.30	2.1	1.4	7.6	1.1	0.0
	0.45	5.3	7.5	20.9	4.2	1.4

Notes: The test statistic is based on SP estimator with regression standard deviation. The bootstrap symmetric p-value is used; n=100; the number of cycles of a block: 2

Tab. 3 Rejection percentage for short memory data for H_0:d=0 vs. H_1:d≠0.

Nominal size	ARFIMA (0,0,0)	ARFIMA(1,0,0)		ARFIMA(0,0,1)	
		¤ = 0.1	¤ = 0.5	¯ = 0.1	¯ = 0.5
10%	8.7	8.9	29.6	9.3	23.9
5%	4.8	6.4	15.8	5.4	19.3
1%	5.1	1.8	7.2	1.7	6.6

Notes: The test statistic is based on GPH estimator with asymptotic standard deviation. The bootstrap symmetric p-value is used; n=100; C=0.45
The number of cycles of a block: 2

References

Beran, J. (1994). Statistics for long-memory processes. New York: Chapman & Hall.

Box, G.E.P., Jenkins, G.M. (1976). Time series analysis, forecasting and control. Oakland, California: Holden-Day.

Davidson, R., MacKinnon, J.G. (2000). Bootstrap Tests: How Many Bootstraps? *Econometric Reviewers 19*, 55-68.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of statistics 7,* 1-26.

Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series model, *Journal of Time Series Analysis, 4(4),* 221-238.

Granger, C. W. & Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis 1*, 15–39.

Hosking, J. (1981). Fractional differencing. *Biometrika 68(1),* 165-176.

Hurvich, C., Deo, R., Brodsky,J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series, *Journal of Time Series Analysis 19(1),* 19-46.

Lahiri, S.N. (1999). Theoretical comparisons of block bootstrap methods. *The Annals of Statistics 27(1),* 386-404.

MacKinnon, J.G. (2007). Bootstrap Hypothesis Testing. *Working Paper No. 1127,* Department of Economics, Queen's University, Canada.

Palma, W. (2007). Long-memory time series. Hoboken, New Jersey: John Wiley & Sons.

Park, D., Willemain, T.R. (1999). The threshold bootstrap and threshold jackknife. *Computational Statistics & Data Analysis 31*, 187-202.

Reisen, V. A. (1984). Estimation of the fractional difference parameter in the ARIMA(p,d,q) model using the smoothed periodogram. *Journal of Time Series Analysis 15,* 335-350.