Tests of long memory, a bootstrap approach

Pilar Grau-Carles*

Preliminary draft

Abstract

Many time series in diverse fields have been found to exhibit long memory. This paper analyzes the behaviour of some of the most used tests of long memory: the R/S analysis, the modified R/S, the GPH (Geweke and Porter-Hudak) test and the DFA (Detrended Fluctuation Analysis). Some of these tests exhibit size distortions in small samples. It is well known that the bootstrap procedure may correct this fact. In this paper, size and power for those tests, for finite samples and different distributions, such as normal, uniform and lognormal are investigated. In the case of short memory processes such as AR, MA and ARCH and long memory such ARFIMA, p-values are calculated using the post-blackening moving block bootstrap. The Monte Carlo study suggests that the bootstrap critical values perform better. The results are applied to financial return time series.

1 Introduction

Time series with long memory appear in many contexts, for example in financial economics, networks traffic, hydrology, cardiac dynamics, meteorology, etc. Long memory or long-range dependence is characterized by hyperbollically decaying autocovariance function, by spectral density that tends to infinity as the frequencies tend to zero and by the self-similarity of aggregated summands.

The intensity of this phenomena can be measured either by a parameter d, used as differencing parameter or by the parameter H, that is a scaling parameter. Both parameters are related, in the case of finite variance processes by $H = d + \frac{1}{2}$, and in the case of infinite variance processes by $H = d + \frac{1}{\alpha}$ (Taqqu *et al.*, 1998).

Several test of long-range dependence are available in the literature and some of them are described in detail in Beran (1994). In particular the R/S analysis, the modified R/S, the GPH (Geweke and Porter-Hudak) and the DFA (Detrended Fluctuation Analysis) tests are investigated in this paper. As it was

*Departamento de Economía Aplicada, Universidad Rey Juan Carlos, Paseo de los Artilleros s/n, 28032 Madrid, *e-mail*:grau@fcjs.urjc.es

found in Cheung (1993), the modified R/S and the GPH suffer for size distortion. Andersson and Gredenhoff (1997) implement a bootstrap method, in order to size adjust fractional integration tests. Bootstrap is a technique that allows the simulation of the probability distribution of any statistic. The key idea is to resample form the original data to create replicate data sets, from which the empirical probability distribution of interest can be found. In the case of dependent data, the resampling should preserve the temporal statistical structure of the original series. Two are the most popular bootstrapping dependent data approaches, the Model Based Resampling and the Moving Block Bootstrap. We will use a strategy intermediate between them, called Post-blackening Moving Block Bootstrap that is explained in Davison and Hinkley (1997).

2 Tests of long memory

The oldest and best-known method for detecting long memory is the R/S analysis. This method, proposed by Mandelbrot and Wallis (1968) and based on previous hydrological analysis of Hurst (1951), allows the calculation of the self-similarity parameter H. This parameter measures the intensity of long-range dependence in a time series. For a time series $\{x_t\}$, $t = 1, ..., T$ we define the range R_n

$$R_n = \left\{ \max_{1 \le i \le n} \sum_{t=1}^{i} (x_t - \overline{x}) - \min_{1 \le i \le n} \sum_{t=1}^{i} (x_t - \overline{x}) \right\} \tag{1}$$

where $\overline{x}$ is the sample mean of the time series. If the range is rescaled with the sample standard deviation S, the R/S statistics asymptotically follows the relation

$$(R/S)_t \propto Ct^H \tag{2}$$

The value of H is generally obtained from the linear regression over a sample of growing temporal horizons ($s = t_1, t_2, ..., T$)

$$\ln(R/S)_s = \ln(C) + H \ln(s) \tag{3}$$

An estimated value of $H = 1/2$ means that the process has no memory, but $H \neq 1/2$ would mean that the process has long memory.

More recently Lo (1991) discusses the lack of robustness of the R/S statistic in the presence of short memory or heteroskedasticity. Lo suggests the *modified rescaled range statistics*, henceforth MRR, which replaces the denominator S, the standard deviation, by a consistent estimator of the square root of the variance of the partial sum of x.

$$Q_T = \left\{ \max_{1 \le i \le T} \sum_{t=1}^{i} (x_t - \overline{x}) - \min_{1 \le i \le T} \sum_{t=1}^{i} (x_t - \overline{x}) \right\} / s_T(q) \tag{4}$$

where

$$s_T(q) = \left\{ \frac{1}{T} \sum_{i=1}^{T} (x_i - \overline{x})^2 + \frac{2}{T} \sum_{j=1}^{q} \tau_j(q) \left(\sum_{i=j+1}^{T} (x_i - \overline{x})(x_{i-j} - \overline{x}) \right) \right\}^{1/2} \quad (5)$$

and also $\tau_j(q) = 1 - \frac{j}{q+1}$ with $q < n$

Lo derives the limiting distribution of $T^{-1/2}Q_T$ under the null of no memory, and shows that the *modified rescaled range statistics* is robust to short-range dependence. Critical values of the distribution are tabulated in Lo (1991) table II.

In a recent paper Teverlosky, Taqqu and Willinger (1999) show that MRR test tends to reject the null hypothesis of no long-range dependence when the series is in fact long dependent and that the choice of the truncation lag q is crucial.

The other method is a semi-parametric procedure to obtain an estimate of the fractional differencing parameter d. This technique, proposed by Geweke and Porter-Hudak (1983), henceforth GPH, is based on the slope of the spectral density function around the angular frequency $w = 0$. The spectral regression is defined by

$$\ln \{I (w_\lambda)\} = a + b \ln \left\{ 4 \sin^2 \left(\frac{w_\lambda}{2} \right) \right\} + n_\lambda \qquad \lambda = I, ..., v \quad (6)$$

where $I(w_\lambda)$ is the periodogram of the time series at the frequencies $w_\lambda = \frac{2\pi\lambda}{T}$ with $\lambda = 1, ..., (T-1)/2$, T is the number of observations and v is the number of Fourier frequencies included in the spectral regression. The least square estimate of the slope coefficient provides an estimation of d. The theoretical error variance is $\pi^2/6$ and allows the construction of the t-statistics for the fractional differencing parameter d. A mayor issue on the application of this method is the choice of v.

The last method we have test that measures the long-range dependence is the Detrended Fluctuation Analysis (DFA) proposed by Peng *et al.* (1994) and improved in Viswanathan *et al.* (1997). The advantage of DFA over Hurst analysis is that avoids spurious detection of apparent long-range correlation that is an artefact of non-stationarities. The method can be summarized as follows. First, the integrate time series $y(t')$ is obtained, $y(t') = \sum_{T=1}^{t'} x(t)$. Next the integrate series is divided into non-overlapping intervals, containing each interval m data. In each interval a least squared line is fitted to the data. The y coordinate of the straight line segments is denoted by $y_m(t')$. Next the root of the mean square fluctuation of the integrated and detrended time series is calculated

$$F(m) = \sqrt{\frac{1}{T} \sum_{t'=1}^{T} [y(t') - y_m(t')]^2} \quad (7)$$

This calculation is repeated over all intervals. A linear relationship on a double log graph of $F(m)$ and the interval size m indicates the presence of a power law scaling. If there is no correlation or only short correlation $F(m) \propto m^{1/2}$, but if there is long-range power-law correlations then $F(m) \propto m^{\alpha}$ with $\alpha \neq 1/2$.

3 Bootstrap testing

The bootstrap technique can be used for estimation of the small-sample distribution of a statistics. Introduced by Efron (1979), the bootstrap procedure enables correction of size distortions. For details of bootstrap tests see Davidson and MacKinnon (1996).

The original test, designed for *iid* observations, fails for dependent observations. The moving block bootstrap and model based resampling methods perform better for short range dependence. Short and long memory processes will be examined using the post-blackened moving block bootstrap method discussed in Davison and Hinkley (1997) and studied by Srinivas and Srinivasan (2000), which is an intermediate approach between both and appears to capture the dependence structure of the data, even using a small number of bootstrap replications. The idea that underlies the block resampling is that if block are long enough, the original dependence will be preserved in the resampled series.

The procedure works as follows for a given time series $\{x_t\}, t = 1, \ldots, T$:

Step 1. Compute the tests statistics (Lo, GPH, Hurst and DFA) and obtain $\hat{\tau}$.

Step 2. "Pre-white" the time series fitting an $AR(p)$ model with a suitable large number of lags of the time series and obtain the estimated residuals e_t and the centered residuals $e_t - \bar{e}$. The order of the autorregresive model will be estimated using the Schwartz's criterion.

Step 3. Resample blocks of the centered residuals from the estimated model using the moving block bootstrap to generate B bootstrap samples of them.

Step 4. "Post-black" the resampled centered residuals using the estimated parameters of the AR model to generate B bootstrap samples of x denoted x^b.

Step 5. For each bootstrapped sample compute the statistics τ^b.

Step 6. Compute the statistics p-value. For a two-side test is defined by

$$p^*(\hat{\tau}) = \frac{1}{B} \sum_{b=1}^{B} I(|\tau^b| \geq |\hat{\tau}|)$$

where $I(\cdot)$ equals one if the inequality is satisfied and zero otherwise. The null is rejected when the selected significance level exceeds $p^*(\hat{\tau})$.

4 Monte Carlo study

For the Monte Carlo study 1000 replicated series of each model will be used. Each series will be tested for long-memory using the tests described in section 2, for the no long-memory null hypothesis. In all the cases $T + 100$ observations are generated and only the T last are used to reduce the effect of initial values. The block size has been set in 5.

4.1 Normal, uniform and lognormal distributions

Results of the tests when the data follows uniform, normal and lognormal distributions are reported in Table 1. The simulation result shows that all the tests perform correctly and do not show significant deviations from the nominal size.

Table 1: Rejection percentage of the nominal 5% fractional integration test when the data follows uniform, normal and lognormal distributions of length T.

	MRR	GPH	R/S	DFA [1]
Uniform				
T=100	4.1	4.3	4.8	3.8
T=200	5.0	5.0	4.7	4.5
Normal				
T=100	5.7	4.6	5.3	5.6
T=200	4.5	4.4	4.1	4.4
Lognormal				
T=100	4.3	5.2	4.2	4.5
T=200	6.0	4.3	4.4	4.5

4.2 AR and MA specification

The following processes will be simulated

$$x_t = \phi x_{t-1} + a_t$$

and

$$x_t = a_t + \theta a_{t-1}$$

where $a_t \sim$ IID $N(0, 1)$. The AR and MA parameters are set equal to ± 1, ± 4 and ± 9. Table 2 reports the sensitivity of the empirical size to AR and MA components for time series of length T=100. The simulation results suggests that MRR and GPH for both AR and MA models perform well, but rejection frequencies are significantly larger than the nominal significance level when $\theta =$

[1] MRR: modified rescaled range test GPH: Geweke-Porter-Hudak test; R/S: Hurst test DFA: detrended fluctuation analysis. Under the null hypothesis of no fractional integration, the 95% confidence interval is $5\% \pm 1.4\%$

-0.9. These results are slightly better than the ones obtained in Andersson and Gedenhoff (1997) using the original no parametric bootstrap. The R/S and DFA tests show similar behavior, they are always very conservative in the sense that they tend to reject the null less frequently than the nominal.

MRR and GPH post-blackened bootstrap give exact tests because the estimated sizes of the tests coincide with the nominal, whereas R/S and DFA are very conservative in the case of moving average and autoregressive models and small samples.

Table 2: Rejection percentage of the nominal 5% fractional integration test when the data follows AR and MA processes of length T=100.

	MRR	GPH	R/S	DFA
ϕ		AR process		
-0.9	4.7	4.2	0.6	0.6
-0.4	5.0	3.8	0.4	0.6
-0.1	5.1	4.8	0.3	0.4
0.1	5.3	4.5	0.9	2.7
0.4	5.3	3.7	0.6	1.4
0.9	5.9	4.8	2.4	1.5
θ		MA process		
-0.9	7.5	17.9	0.1	0.0
-0.4	6.3	3.8	0.1	0.1
-0.1	5.2	4.8	0.3	0.2
0.1	5.0	4.6	0.9	0.4
0.4	5.6	5.1	0.5	0.3
0.9	5.9	5.1	0.3	0.2

4.3 ARCH specification

Now, the effect of heteroskedasticity of ARCH type is considered. The simulated model is

$$x_t = u_t \qquad u_{t/t-1} \sim N(0, h_t)$$
$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2$$

Results in Table 3 show that MRR and GPH are quite robust, but, again R/S and DFA tests are conservative with a very small rejection rate, so the post-blackening bootstrap test allows the exact estimation of the long-range dependence for small samples in the presence of heteroskedasticity.

Table 3: Rejection percentage of the nominal 5% fractional integration test when the data follows an ARCH (1) model of length T=100.

	MRR	GPH	R/S	DFA
α		ARCH process		
0.3	6.1	4.2	0.5	0.7
0.5	5.9	4.0	0.6	0.2
0.8	7.3	5.0	0.5	0.5

4.4 ARFIMA specification

Fractionally integrated white noise or ARFIMA (0,d,0) is also tested. Data are generated following Hosking (1981):

$$(1-B)^d x_t = a_t$$

where $a_t \sim$ IID $N(0,1)$ and the fractional differencing parameter is d is set equal to ± 0.45, ± 0.25 and ± 0.05. Results in table 4 show that MRR and GPH have good power except for $d = -0.05$ that leads to small rejection frequencies. The R/S and DFA tests have good power for persistent processes, or positive d parameter, but the performance of these tests is really poor for antipersitent series with a very low rejection rates.

Table 4: Rejection percentage of the nominal 5% fractional integration test when the data follows an ARFIMA (0,d,0) processes of length T=100.

	MRR	GPH	R/S	DFA
d		AR process		
-0.45	7.4	9.0	2.6	3.8
-0.25	7.7	7.9	3.4	4.7
-0.05	6.0	4.0	5.3	7.2
0.05	6.8	4.0	7.5	8.3
0.25	14.3	8.5	21.8	12.4
0.45	22.6	14.4	23.0	19.7

5 Application to financial data

Finally all four test are used to test long memory in time series of returns, absolute returns and squared returns of two mayor daily stock indices, the Standard& Poors 500 and the Dow Jones Average. Two samples of each index are analyzed, one year of daily data, from 01/02/2003 to 31/12/2003 yielding 252 observations and a longer series of five years of data, form 01/02/1999 to 31/12/2003, with a total of 1256 observations. The returns are calculated using the logarithmic differenced data

$$r_t = (\ln(x_t) - \ln(x_{t-1})) * 100$$

A common finding in much of the empirical literature is that returns themselves contain little serial correlation, however, the absolute returns and their power transformation present long memory, as discussed in Granger and Ding (1993).

Results of the bootstrap p-values are given in Table 5 and 6.

Table 5: Point estimation and p-value results for test of long memory for the SP500.

	MRR	GPH	R/S	DFA
SP500	T=251			
r_t	1.1968	-0.0615	0.5984	0.3989
p-value	0.4030	0.6870	0.3940	0.7880
$\|r_t\|$	0.3967	0.5746	0.6750	0.4723
p-value	0.0000	0.0010	0.2130	0.6579
r_t^2	0.6814	0.5943	0.6432	0.5481
p-value	0.0010	0.0000	0.3360	0.3032
SP500	T=1255			
r_t	1.5529	0.0850	0.5131	0.4426
p-value	0.1010	0.2980	0.9000	0.8560
$\|r_t\|$	0.4778	0.4699	0.7908	0.7499
p-value	0.0040	0.0000	0.0000	0.0000
r_t^2	0.8106	0.3973	0.7744	0.7523
p-value	0.0340	0.0000	0.0000	0.0000

Table 6: Point estimation and p-value results for test of long memory for the Dow Jones.

	MRR	GPH	R/S	DFA
Dow Jones	T=251			
r_t	1.2619	0.0016	0.6068	0.4318
p-value	0.2540	0.9350	0.4330	0.7090
$\|r_t\|$	0.4219	0.4093	0.7057	0.5150
p-value	0.0000	0.0080	0.1940	0.6910
r_t^2	0.7257	0.4574	0.6761	0.5407
p-value	0.0040	0.0060	0.3090	0.2910
Dow Jones	T=1255			
r_t	1.3249	0.1205	0.5668	0.4858
p-value	0.2940	0.1380	0.5120	0.5510
$\|r_t\|$	0.6348	0.4683	0.8027	0.7726
p-value	0.0000	0.0000	0.0000	0.0000
r_t^2	1.0789	0.3062	0.7710	0.7544

p-value	0.0050	0.0040	0.0020	0.0050

For the shorter time series of both indexes, the MRR and GPH found long memory in absolute returns and squared returns series, but the R/S and DFA do not allow the rejection of the null hypothesis of no long memory.

In the case of the series that covers five years, all of the tests find no memory in returns and long memory in absolute and squared returns.

6 Conclusions

Post-blackening moving blocks bootstrap provides an effective method to improve the test for long memory, especially for MRR and GPH tests and short time series.

R/S and DFA tests can also benefice from the bootstrap approach. They should not be used to test long memory in short time series, but when applied to longer series, p-values that allow hypothesis testing can be obtained.

References

[1] Andersson, M. K. and Gedenhoff, M. P. (1997) Bootstrap testing for fractional integration. *Working papers Series in Economics and Finance* No.188. Dept. of Economic Statistic, Stockhlom School of Economics, Sweden.

[2] Beran, J. (1994) Statistics for Long Memory Processes, Chapman and Hall. London

[3] Cheung, Y.-W. (1993) Tests for fractional integration: A Monte Carlo Investigation, *Journal of time series Analysis* 14 (4), 331-345.

[4] Davidson, R. and MacKinnon, J. (1996) The size distortion of bootstrap tests, *GREQAM Document de Travail No 96A15.*

[5] Davison, A.C. and Hinkley, D.V. (1997)*Bootstrap Methods and their application.* Cambridge: Cambridge Univesity Press.

[6] Efron, B. (1979) Bootstrap methods: Another Look at the Jackknife, *Annals of Statistics* 7, 1-26.

[7] Efron, B. and Tibshirani, R.J. (1993) *An introduction to Bootstrap*, New York, Chapman and Hall.

[8] Geweke, J. and Porter-Hudak, S. (1983) The estimation and Appication of Long Memory Time Series Models, *Journal of Time Series Analysis* 4, 221-238.

[9] Granger, C.W.J. and Ding, Z. (1993) Some Properties fo Absolute Return: An Alternative Measure of Risk, *University of California at San Diego, Economic Working Paper Series* 93-38.

[10] Hosking, J. (1981) Fractional differencing, *Biometrika* 68,165-176.

[11] Hurst, H.E. (1951) Long term storage capacity of reservoirs, *Transactions of the American Society of Civil Engineers* 116, 770-199.

[12] Lo, A. (1991) Long term memory in stock market prices, *Econometrica*, 59 451-474.

[13] Mandelbrot, B.B. and Wallis, J.R. (1968) Noah, Joseph and operational hydrology, *Water Resources Research* 4, 909-918.

[14] Peng, C.-K., Buldyrev, S.V., Havlin, S., Simons, M. and Stanley, H.E. Goldberger, A.L. (1994), Mosaic Organization of DNA sequences *Physical Review E* 49 1684-1989.

[15] Srinivas, V. V. and Srinivasan, K (2000) Post-blackening approach for modeling dependent annual streamflows, *Journal of Hydrology* 230, 86-126.

[16] Taquu, M.S., Teverovsky, V., in: R. Alder, R. Feldman, M.S. Taquu, (Eds.) (1998)*A practical guide to heavy tails: Statistical techniques and applications*, Birkhauser, Boston. Pages 177-217.

[17] Teverlosky,V. Taqqu, M.S. and Willinger,W. (1999) A critical look at Lo modified R/S statistic *Journal of Statistical Planning and Inference* 80, 211-227.

[18] Viswanathan, G.M., Buldyrev, S.V., Havlin, S., Stanley,H. E.(1997) Quantification of DNA Patchiness using correlation measures, *Biophysical Journal* 72, 866-875.