Computational Economics **22:** 187–212, 2003.

Bilateral Bootstrap Tests for Long Memory: An Application to the Silver Market

CHRISTIAN DE PERETTI
GREQAM, Centre de la Vieille Charité 2, rue de la Charité, F-13002 Marseille, France;
E-mail: peretti@ehess.cnrs-mrs.fr

(Accepted 29 January 2003)

Abstract. Many time series in diverse fields of application may exhibit long-memory. The class of fractionally integrated (FI) processes can be used to try to model this strong data dependence. Asymptotic tests for FI include the re-scaled range statistic test and its modified form, the frequency-domain regression-based procedure, the modified Higuchi's test and Jensen's test. De Peretti and Marimoutou (2002) finds that proper finite-sample inferences are not possible using these techniques without correcting for size distortions. Some attempt this correction through 'bootstrapping', but this method is not perfect and needs more study and improvements. In this paper, I examine and compare the finite-sample properties of parametric and nonparametric bootstrap tests by using graphical techniques of Davidson and MacKinnon (1998a) for showing whether they properly correct the distortions while retaining their power relative to the corresponding asymptotic tests. One of the tests uses a double bootstrap that provide better true power and size properties. I use a bilateral P value that permits the true power of the tests to grow when the size distortions are asymmetric. We then apply these procedures to a real time series to investigate its long memory properties.

Key words: parametric and nonparametric bootstrap, long memory, tests, P value plots, corrected size-power curves

1. Introduction

Many time series in diverse fields of application may exhibit long-memory or long-range dependence. This occurs when the autocorrelation function (a.c.f.) at large lags decays to zero at a slower rate than data following an ARMA(p, q) model and at a quicker rate than that of a unit root processes (Mandelbrot et al., 1969, 1971). The class of fractionally integrated (FI) processes is characterised by hyperbolic decay rate, and it is often used to try to capture this strong data dependence. Granger and Joyeux (1980) and Hosking (1981) are two seminal articles on ARFIMA(p, d, q) processes.

Many methods for estimating and testing the long memory parameter d are described in Beran (1994). A classical test for FI is the re-scaled range statistic test (Hurst, 1951; Mandelbrot, 1972). Lo (1991) proposes another which is a modified form of the Hurst statistic. Other tests for FI are the frequency-domain regression-based procedure introduced by Robinson (1995), and the test of Jensen (1994). I also construct a test based on the estimator of Higuchi (1988).

Most applications of these techniques rely on asymptotics to make small-sample inferences. De Peretti and Marimoutou (2002) examines and compares the finite-sample properties of these five tests using the graphical techniques developed by Davidson and MacKinnon (1998a). That study examines the size correction needed to show the true power of the tests rather than their nominal power, it uncovers two severe problems.

First, under the null hypothesis H_0, there are very large size distortions for all tests, all sample sizes, and all parameter values. It is impossible to make correct inferences without correcting for the distortions. While correction can be made using bootstrap techniques (Andersson and Gredenhoff, 1998), this method is not perfect and needs further study. In this paper, I examine and compare the finite-sample properties of bootstrap tests using the graphical techniques of Davidson and MacKinnon (1998a) to show the true power of these tests, and to see if there is a loss of power of the bootstrap test relative to that of the corresponding asymptotic test on a proper size-corrected basis (Davidson and MacKinnon, 1996).

The second problem is that, under the alternative hypothesis H_1, for all tests, H_1 can be accepted more often under H_0 than under H_1 when there is persistent[1] long-memory against persistent short-memory. One of the tests employed uses a double bootstrap that provides better true power and size properties. I also use a bilateral P value that permits the true power of the tests to grow when the size distortions are asymmetric.

This approach is of interest when we wish to detect long-memory behavior on real data. The purpose of this work is to explain in detail this method and to show how it can be used on real data with reasonable precautions.

Section 2 details bootstrap tests for FI. Monte Carlo results based on simulated ARMA processes and ARFIMA processes are presented in Section 3. In Section 4, the described methods are applied to real data to determine the presence of long-memory behavior. Concluding remarks are offered in Section 5.

2. Asymptotic Tests for Long-Memory

2.1. THE MODEL

I restrict attention to univariate, linear FI models of the ARFIMA[2] form:

$$\phi(L)(1-L)^d x_t = \theta(L)\varepsilon_t \quad t \in \{1, \ldots, T\}, \tag{1}$$

$$\{\varepsilon_t\} \sim i.i.d.\, D(0, \sigma_\varepsilon^2), \tag{2}$$

where

- ϕ and θ are polynomials that have all roots outside the unit circle,
- $\sigma_\varepsilon^2 < \infty$,

- L is the lag operator,
- d is the differencing parameter and takes a real value,
- $D(0, \sigma_\varepsilon^2)$ is a distribution with zero mean and σ_ε^2 finite variance.

In some circumstances, a long-memory process may be approximated by a FI model; hence testing for long-memory can be done by a test on d. Such tests are applied to stationary and invertible series (require that $|d| < 1/2$), and $H_0 : d = 0$ is thus a natural null hypothesis.

2.2. FIVE TESTS FOR LONG-MEMORY

I test the null hypothesis $H_0 : d = 0$ against $H_1 : d \in (-0.5, 0) \cup (0, 0.5)$, in the five following ways:

2.2.1. *The R/S Method*

The R/S or re-scaled adjusted range statistic is introduced in Hurst (1951), where the question of how to regularise the flow of the Nile river is investigated. Mandelbrot (1975) proves that under regularity conditions with $h_0 : d = 0$, this statistic converges in distribution to a non-degenerate random variable. In Appendix A.1, I use the Mandelbrot proof to construct the asymptotic test.

2.2.2. *The Modified R/S Method*

A weakness of the standard R/S analysis is its sensitivity to the short-range dependence. To avoid short memory perturbations, Lo (1991) suggests modifying the R/S statistic to use a consistent estimator of the variance of the partial sums correction (Newey and West (1987) method) that takes into account the short-term dependence. See Appendix A.2 for more details.

2.2.3. *The Log-Periodogram Method*

This method is introduced in Geweke and Porter-Hudak (1983) for the stationary Gaussian case, where, in a semi-parametric framework, the fractional parameter d of long-range dependence behaviour is estimated. The only information required is the behaviour of the spectral density near the origin. Robinson (1995) develops this method by maximisation of an objective function in the frequency domain, and I use this extension for the statistic and the asymptotic test.

The choice of the break point m that determines the point that truncates the log-periodogram regression (hence in the objective function), is very important for robustness against short-memory effects, see Appendix A.3. m must be an integer less than or equal to $T/2$ so that $m = o(T)$ to have consistency. I optimise m for each sample size T with $d = 0.3$. Specifically, I choose that m for which the true power curve is the greatest. I find that the rule $m = O(\sqrt{T})$ is good, and in fact choose for m following the rule $[m = \sqrt{2 \times T}]$, see Table I. While m is small, it

Table I. Choice of m.

T	≤ 128	256	512	1024	2048
m	Inapplicable	23	32	45	64

Table II. Choice of *kmax*.

T	≤ 128	256	512	1024	2048
kmax	Inapplicable	128	256	256	512

is necessarily so for the true power of the tests. The optimal choice for the value of m depends on many parameters, but there are no known rules to obtain it.

2.2.4. *The Modified Higuchi Method*

The Higuchi's estimator is introduced in Higuchi (1988) for measuring the fractal dimension D of a non periodic and irregular time series, such as, for example, $(d + 0.5)$-self-similar processes (we have $d = 1.5 - D$). If the time series has no long-range dependence, then we should have $D = 1.5$. To perform the test procedures, I use the statistic

$$\frac{\hat{d} - 0.5}{\hat{\sigma}(\hat{d})}, \tag{3}$$

where $\hat{\sigma}(\hat{d})$ is the consistent estimator of $\sigma(\hat{d})$.

If I use the OLS estimator of $\sigma(\hat{d})$, the performance of the asymptotic test is very poor, and therefore I prefer to use the bootstrap estimator. I then apply the asymptotic test procedure. For computational raisons, in the context of the Monte Carlo experiments, I do only 39 bootstrap replications. Even, with this very small number, this test is the best, so I restrict my presentation to this case. In the context of real data analyse, I suggest taking at least several hundred bootstrap replications.

As in the log-periodogram method, I find that the choice of the break point *kmax* that defines the maximal time interval for the aggregated series used to construct the fractal curves is very important, see Appendix A.4. For small scales, the short memory component dominates the behaviour of the series, and therefore they must de truncated. Table II shows the values I use for *kmax*. There are no known rules to determine the optimal value of *kmax*. All the remarks on the choice of truncation point made for the log-periodogram method hold here.

2.2.5. *The Wavelet Method*

Jensen (1994), following Wornell and Oppenheim (1992), suggests estimating d using wavelet analysis. A wavelet set is the set of dilatations and translations of a mother wavelet function, and they form an orthonormal basis in $L^2(\mathbb{R})$, that allows one to decompose $L^2(\mathbb{R})$ in a growing sequence of subspaces V_j that approximate $L^2(\mathbb{R})$. V_j can be viewed as containing the functions with which one cannot examine details smaller than 2^{-j}. We must truncate the subspace set by eliminating those corresponding to high frequencies. The aim is to reduce the disturbances caused by the short memory. See Appendix A.5 for more details.

A problem arises because one cannot optimise the truncation point with respect to sample size since the frequencies do not depend upon the sample size. Rather, they depend on d, the parameter we wish to estimate. So, I must, therefore, examine all possible values for the truncation point. But, another problem appears: the performance is not satisfactory for sample sizes smaller than 2049. This arises because the covariances between the wavelet coefficients are not taken into account (Barnet et al., 2000). In Jensen (1994), this problem does not arise because processes under the null are IID that is not realistic for real data. We do not present the results concerning Jensen's test in this paper.

3. Bootstrap Tests for Long-Memory

While the approaches described above are asymptotically valid, the tests for the statistics based on the asymptotic distributions are not exact in finite samples, and so, it is natural to 'bootstrap' them. For further information on the bootstrap, see Efron (1979), Davidson and MacKinnon (1993, 1996), and (1998b).

3.1. THE BOOTSTRAP PROCEDURE

The procedure is as follows:

1. Compute the test statistic (Hurst, Lo, Robinson, Higuchi, or Jensen), which will be denoted $\hat{\tau}$.
2. Estimate the model (1)–(2) by maximum likelihood under the $H_0 : d = 0$, (where the model is reduced to an ARMA(p', q')), to obtain $(\hat{\phi}, \hat{\theta}, \hat{\sigma}_\varepsilon^2)$ and $\hat{\varepsilon}$.
3. Draw B sets of bootstrap error terms, ε^b, and use them to generate B bootstrap samples x^b. There are numerous ways to drawn the error terms, four of which are described below. The elements of x^b are generated recursively from the equation

$$x_t^b = [1 - \hat{\phi}(L)]x_t^b + \hat{\theta}(L)\varepsilon_t^b\,, \tag{4}$$

where the elements of x_t^b are equal to the observed values of x_t if they correspond to values of x_t prior to period $\hat{p} + 1$, and equal to the appropriate lagged values of x_t^b otherwise.

4. For each bootstrap sample, compute the statistic (Hurst, Lo, Robinson, Higuchi, or Jensen), denoted τ^b, with x^b instead of x.
5. Then compute the estimated bootstrap P value (see (6) or (7)–(8)).

I examine four ways for generating the ε_t^b:

1. The parametric bootstrap, called b_0: the ε_t^b are independent draws from the $N(0, \hat{\sigma}_\varepsilon^2)$ distribution.
2. The simplest nonparametric bootstrap, called b_1: the ε_t^b are obtained by re-sampling with replacement from the vector of $\{\hat{\varepsilon}_t\}_{t=\hat{p}+1}^T$.
3. A slightly more complicated from of nonparametric bootstrap called b_2: the ε^b are generated by re-sampling with replacement from the vector

$$\left\{\sqrt{\frac{T}{T-2\hat{p}-1}}\left(\hat{\varepsilon}-\frac{1}{T-\hat{p}}\sum_{i=\hat{p}+1}^{T}\hat{\varepsilon}_i\right)\right\}_{t=\hat{p}+1}^{T}. \tag{5}$$

4. The most complicated nonparametric bootstrap, called b_3: the ε^b are generated by re-sampling from the vector with typical element et constructed as follows:
   - let d_t be the tth diagonal element of $P_{[1-\phi(L)]}$, the matrix projecting onto the space spanned by $1 - \phi(L)$;
   - divide each element of $\hat{\varepsilon}$ by $\sqrt{1-d_t}$;
   - re-centre the resulting vector;
   - re-scale it so that it has variance $\hat{\sigma}_\varepsilon^2$.

   This type of procedure is advocated in Weber (1984).

3.2. THE CHOICE OF THE BOOTSTRAP P VALUE

By drawing large numbers of bootstrap statistics τ^b, a bootstrap P value can be computed as

$$\hat{p}_{uni}(\hat{\tau}^2) = \frac{1}{B}\sum_{b=1}^{B} I((\tau^b)^2 > \hat{\tau}^2), \tag{6}$$

(Davidson and MacKinnon, 1993). This formula corresponds to a unilateral test, but similar formulae are often associated with symmetric bilateral tests. However, the size distortion is not necessarily symmetric. Thus, for bilateral (asymmetric) tests, I prefer to use the formula

$$\hat{p}_{bil}(\hat{\tau}) = 2\ \min\{\hat{p}(\hat{\tau}), 1 - \hat{p}(\hat{\tau})\}, \tag{7}$$

where

$$\hat{p}(\hat{\tau}) = \frac{1}{B}\sum_{b=1}^{B} I(\tau^b > \hat{\tau}). \tag{8}$$

This sort of P value can be found in Chapter 5 of Davidson and MacKinnon, 1993, in the context of confidence regions.

3.3. THE ESTIMATION UNDER THE NULL

Best way to estimate the model (1)–(2) under the null hypothesis is to consider the ARMA(p', q') model. I select $(\hat{p}', \hat{q}')$, the estimates of (p', q'), by the BIC (Schwarz, 1978). In the context of Monte-Carlo experiments, I use an AR(p'') for reasons of computation time. However, I recommend to those who only want to use the bootstrap tests on true data to consider the full ARMA model to estimate H_0, because $\hat{p}''$ in the AR model can be large under H_1. It is of interest to note that this issue does not much affect the Monte Carlo results. An AR(p'') process with a large p'' has essentially the same appearance of long memory as its corresponding ARMA(p', q').

3.4. THE NUMBER OF BOOTSTRAP REPLICATIONS

To compute the estimated bootstrap P value, one must draw B sets of bootstrap error terms. I use a small value for B, $B = 99$, in the Monte Carlo experiments for reasons of computing time. Again, I recommend to those who want only use bootstrap tests to consider larger values, which allows some (but not much) gain true power and better properties for the size distortion. The gain is small, however (see Figure 4, where the true power curves of the bootstrap tests are not far from those of the corresponding asymptotic tests). The bootstrap size correction is not quasi-perfect in this context because the P-value functions are strongly sloped (Figures 1 and 2). This means that the distributions of the statistics depend greatly on the parameter values, and thus, the bootstrap error in estimating the null hypothesis become more important.

4. Monte-Carlo Experiments

The theoretical results of Davidson and MacKinnon suggest that all the bootstrap tests should work well with finite samples. But this need not be so, depending on circumstances. It is possible then for the to be removed from the true distribution or even be unstable and be inferior to the corresponding asymptotic test. I now provide evidence, based on Monte-Carlo experiments, that the bootstrap is not unstable, but it can encounter difficulties.

All the experiments deal with tests for short memory in the context of a model without a constant term. Under H_0, the test statistics depend only on the parameters $(\phi, \theta, \sigma_\varepsilon^2)$ and T. σ_ε^2 is a scale parameter and its value does not affect the statistics. Thus I focus on the coefficients (ϕ, θ), setting σ_ε^2 to unity. I use $T = 2^n$, where n is an integer. For $n \leq 7$, all the asymptotic statistic tests display problems. Either they have negligible true power or they do not work (de Peretti and Marimoutou, 2002).

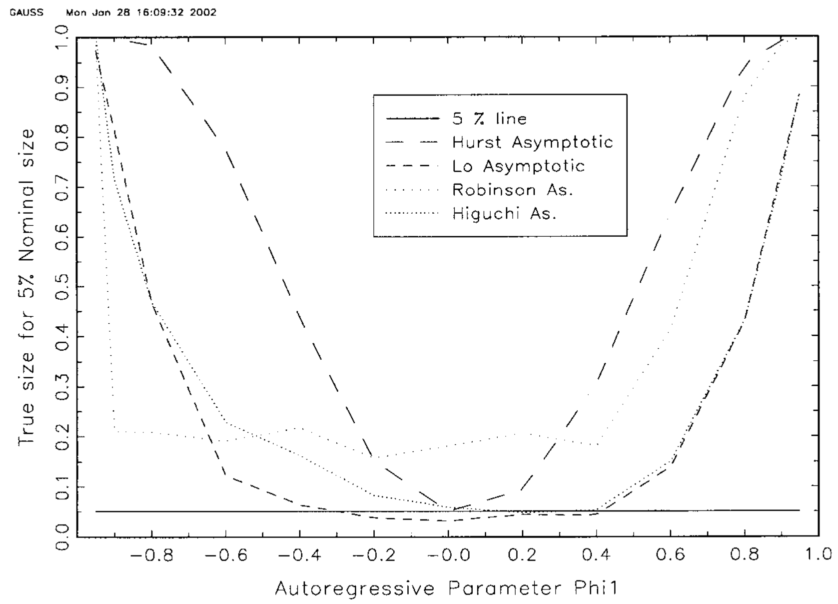

Figure 1. P value functions for long memory tests at 0.05 level, $T = 128$.

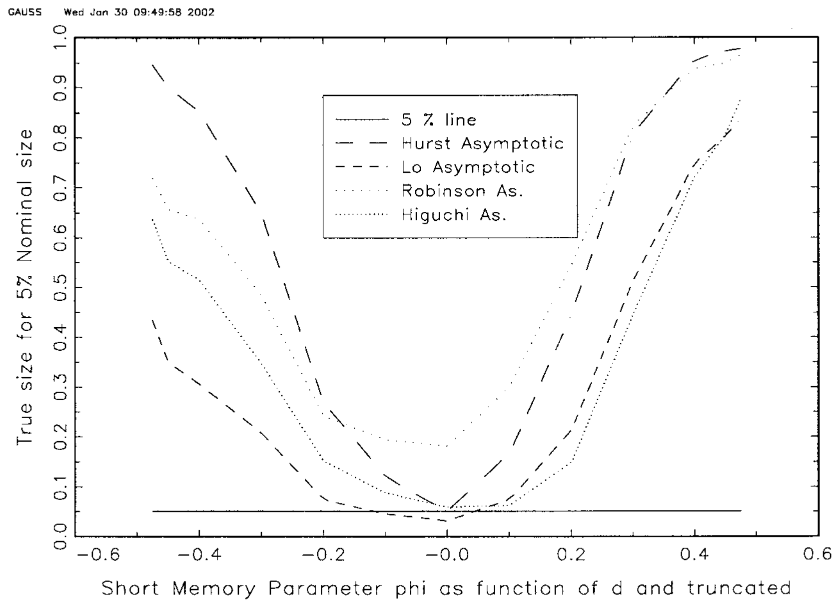

Figure 2. P value functions for long memory tests at 0.05 level, $T = 128$.

The bootstrap cannot solve problems of this sort since it corrects only the size distortion. So, in the following, I discuss performances only for $n \in \{8, 9, 10, 11\}$.

4.1. GRAPHICAL METHODS

Techniques of Davidson and MacKinnon (1998a) for the graphical display of simulation evidence of the size and power of tests of hypotheses are used. Two types of display – called P value plots[3] (Appendix B.1) and size-power curves[4] (Appendix B.2) – are used. For the second type of display, I need to use a DGP that satisfies the null hypothesis. A reasonable choice is the pseudo-true null (Appendix B.2), and to calculate it, I do an asymptotic estimation of the ARFIMA(p, d, q) process using an ARMA(p', q') filter selected by the BIC. Monte Carlo experiments on the properties of Long Memory test are investigated through these figures.

4.2. BOOTSTRAP LONG MEMORY TESTS UNDER THE NULL

P value plots are examined for the parametric bootstrap test b_0 and for the nonparametric bootstrap tests b_1, b_2, b_3 applied to the Hurst, Lo, Robinson and Higuchi's test statistics. They are compared to their corresponding asymptotic test. Four cases of DGP for the null hypothesis are selected by the P value functions (see below). The P value plots are based on an experiment with 800 replications. More precisely, each panel shows the proportion of replications with P values less than size a for each of the five tests, as a function of $\alpha \in [0, 1]$.

4.2.1. *Case of AR(1) Processes*

The Data Generating Process (DGP) under H_0 is

$$x_t = \phi_1 x_{t-1} + u_t\ , \tag{9}$$

$$\phi_1 \in (-1, 1), \tag{10}$$

$$u_t \sim i.i.d.\ N(0, 1), \tag{11}$$

Figure 1 shows P value functions (PVF), for $\phi = (\phi_1, 0, 0, \ldots)$, $T = 128$, and a normal distribution for the error terms, based on 800 replications for each value of ϕ_1. These PVFs are constructed using the asymptotic distributions. One observes severe over rejection in some cases, notably when $|\phi_1| > 0.4$, especially for the Hurst test.

Figure 1 is used to decide what case to investigate. Case 1 is chosen as typical, since it has a plausible value of ϕ, (>0), and no great curvature. Cases 2 and 3 are chosen to be ones where the bootstrap tests might encounter problems, because the PVFs display considerable curvature. A fourth case is considered in which the parameters are the same as in the Case 3, but the error terms have a $t(5)$ distribution instead of $N(0, 1)$.

Table III. Choice of ϕ_1.

Cases	Hurst	Lo	Robinson	Higuchi	Jensen	T
1	0.1	–0.2	–0.2	0.2	–	512
2	0.6	0.9	0.7	0.9	–	1024
3	–0.4	–0.9	–0.9	–0.9	–	256
4	–0.4	–0.9	–0.9	–0.9	–	256

As in de Peretti and Marimoutou (2002), the asymptotic tests have size distortions too large to produce correct inferences. One sees that all the bootstrap tests correct quasi perfectly the size distortions, even for large values of $|\phi_1|$, see Figure 3 for a few examples for Case 3 of the parameters (Table III). So, in this case, the use of the bootstrap is essential.

4.2.2. *Case of AR(p) Processes*

The DGP under H_0 is

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-2} + u_t \,, \tag{12}$$

$$(\phi_1 \cdots \phi_p) \text{ such as } x_t \text{ stationary}, \tag{13}$$

$$u_t \sim i.i.d.\, N(0, 1). \tag{14}$$

The $(\phi_1 \cdots \phi_p)$ are chosen so they are the asymptotic estimations of an ARFIMA$(0, d, 0)$ process selected by the BIC, where $d \in (-0.5, 0.5)$.

Figure 2 shows P value functions (PVF), as function of $(\phi_1 \cdots \phi_p)$ (i.e., as function of d and truncated by the BIC), for $T = 128$, and a normal distribution for the error terms, based on 800 replications for each value of d, using the asymptotic distributions.

As before, Figure 2 is used to decide which cases to investigate, see Table IV (Case 4 is the student case). p is often large, about five or ten.[5]

One observes over-rejection for the Hurst, Lo, and Robinson bootstrap tests. As explained, the true reason for which the bootstrap size correction is not quasi-perfect in this context is that the P value functions are highly sloping. The Higuchi bootstrap tests are quasi-perfect, not because the underlying Higuchi's estimator is better than the others, but because we use a double bootstrap method. In all the cases, the bootstrap test distortions are all lower than the asymptotic test ones. See Figure 4 for a few examples for Case 2 of the parameters (Table IV). Thus, I advise use of the bootstrap tests in these cases too. If possible, it is better to use double bootstrap tests, but the computing time can be large.

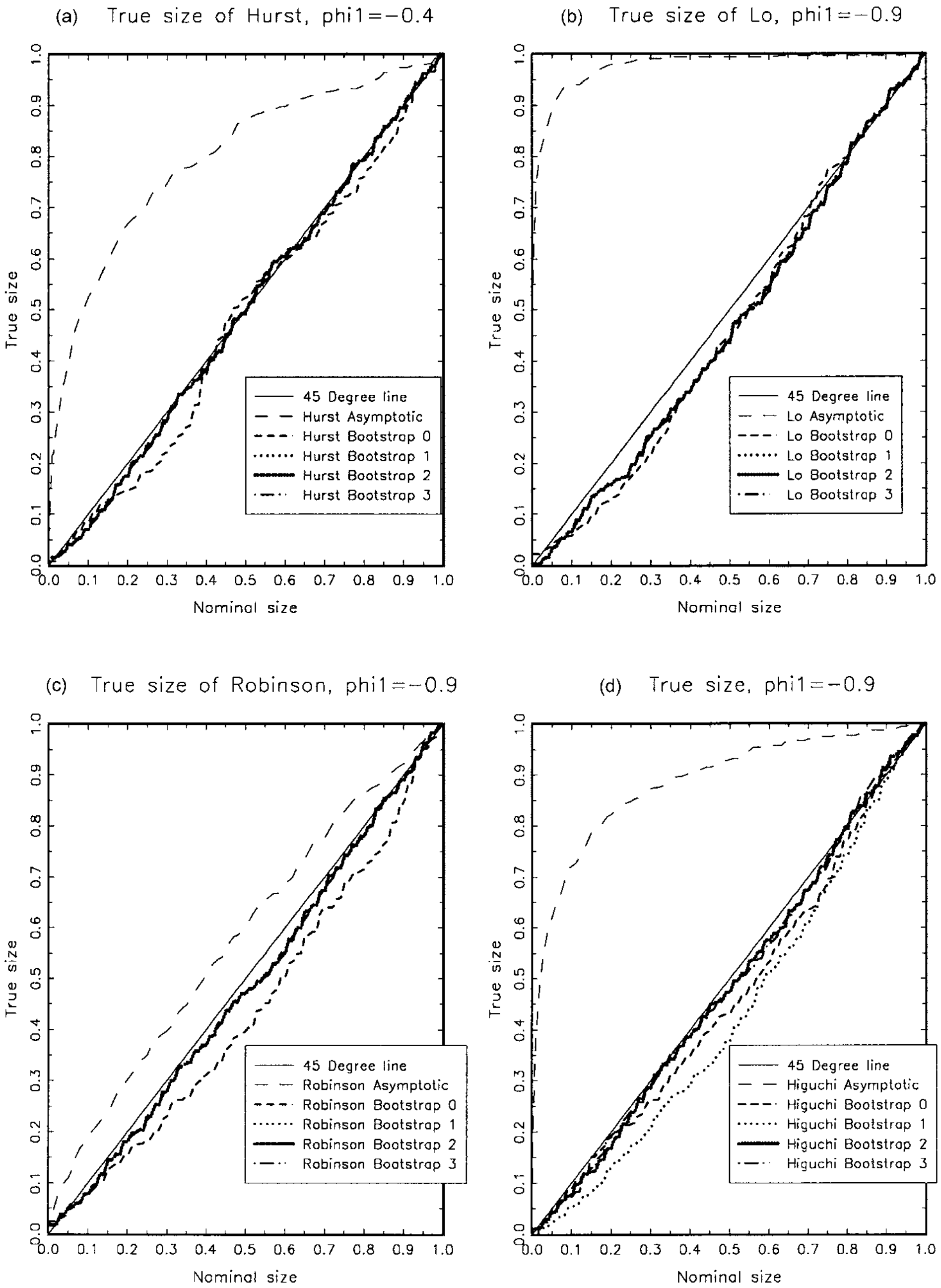

Figure 3. Size distortions for long memory tests. We are in the case of AR(1) processes for the Case 3 of the parameters (Table III).

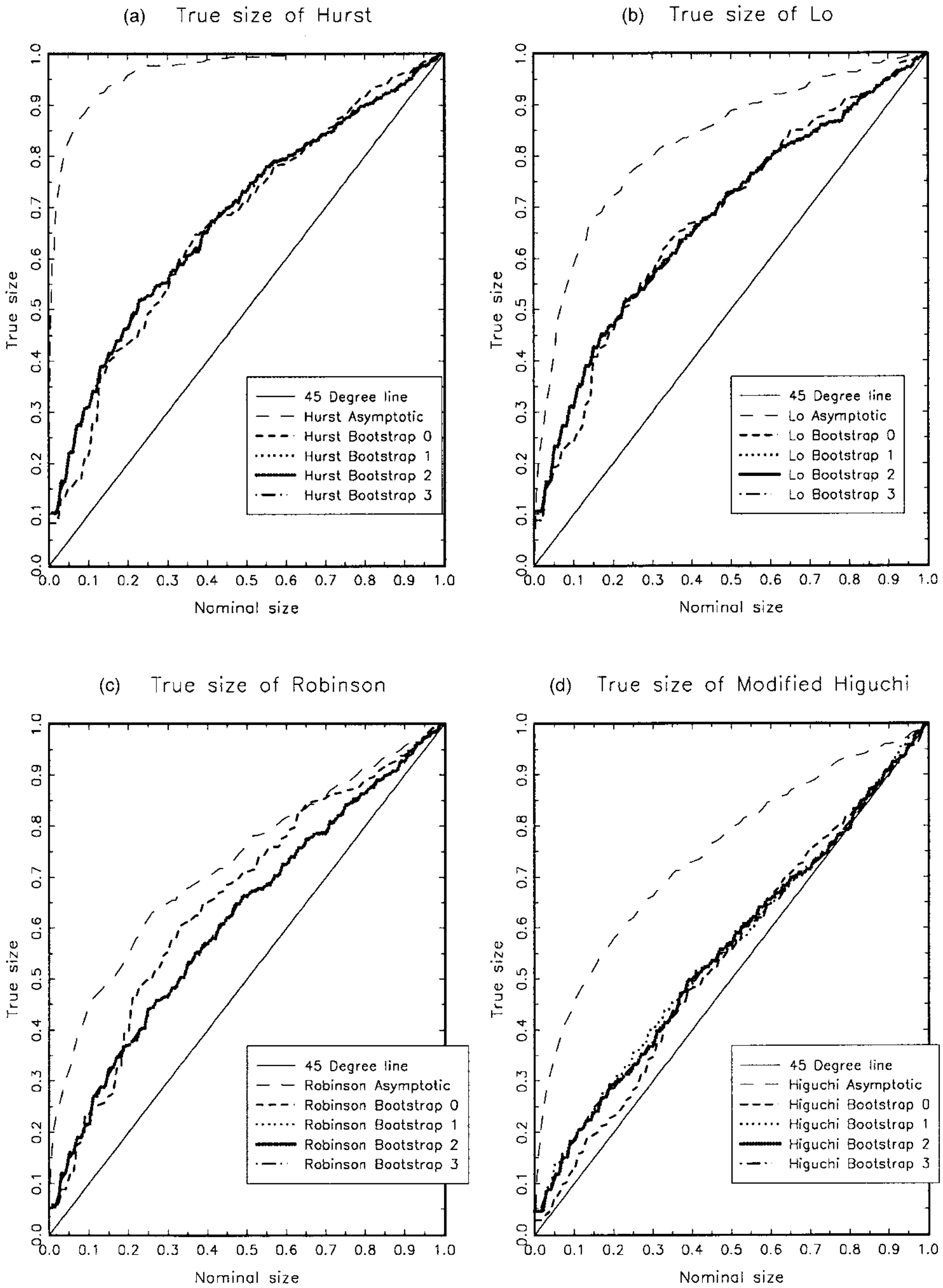

Figure 4. Size distortions for long memory tests. We are in the case of AR(p) processes for the Case 2 of the parameters (Table IV).

Table IV. Choice of d.

Cases	Hurst	Lo	Robinson	Higuchi	Jensen	T
1	–0.45	–0.45	–0.45	–0.45	–	256
2	–0.3	–0.3	–0.3	–0.3	–	512
3	0.3	0.3	0.3	0.3	–	2048
4	–0.3	–0.3	–0.3	–0.3	–	512

4.3. SIZE-POWER CURVES OF BOOTSTRAP LONG MEMORY TESTS

The size-power curves are studied for the parametric bootstrap test b_0 and for the nonparametric bootstrap tests b_1, b_2, b_3, as compared to the corresponding asymptotic tests. There is no unique way to measure size-corrected power (Davidson and MacKinnon, 1996). I choose the null DGP characterised by the 'pseudo-true values', in the sense of White (1982), that corresponds to the fixed DGP which is, asymptotically at least, the closest null to a given fixed DGP. I then pick combinations of d and T in Table IV to investigate, and run experiments with 400 replications under H_1 and under H_0, using the same random numbers (to avoid experimental errors).

4.3.1. *Case of ARFIMA*$(0, d, 0)$ *Processes*

The DGP under H_1 is

$$x_t \sim \text{Gaussian or Student ARFIMA}(0, d, 0) \tag{15}$$

$$d \in (-0.5, 0.5). \tag{16}$$

d is taken from Table IV. The DGP under H_0 is generated as in Section 3.3.

According to theory, the bootstrap test power should be similar to that of the corresponding asymptotic test using the size correction. We verify here that the bootstrap does not result in power loss. In all the cases, the intrinsic power curves of the bootstrap are quite similar to those obtained for the asymptotic distribution, even for parameters chosen so that bootstrap tests could encounter problems. See Figure 5 for a few examples for Case 2 of the parameters (Table IV).

4.4. CASE OF STUDENT ERROR TERMS

To test the nonparametric bootstrap, I simulate leptokurtic data using a Student's t-distribution with five degrees of freedom for the error terms. To capture the excess probability in the tails, I implement more bootstrap replications: $B = 399$. In the case of an AR(1) processes, the correction of the size distortion is quasi-perfect for all the parametric and nonparametric bootstraps, as in the Gaussian cases. In

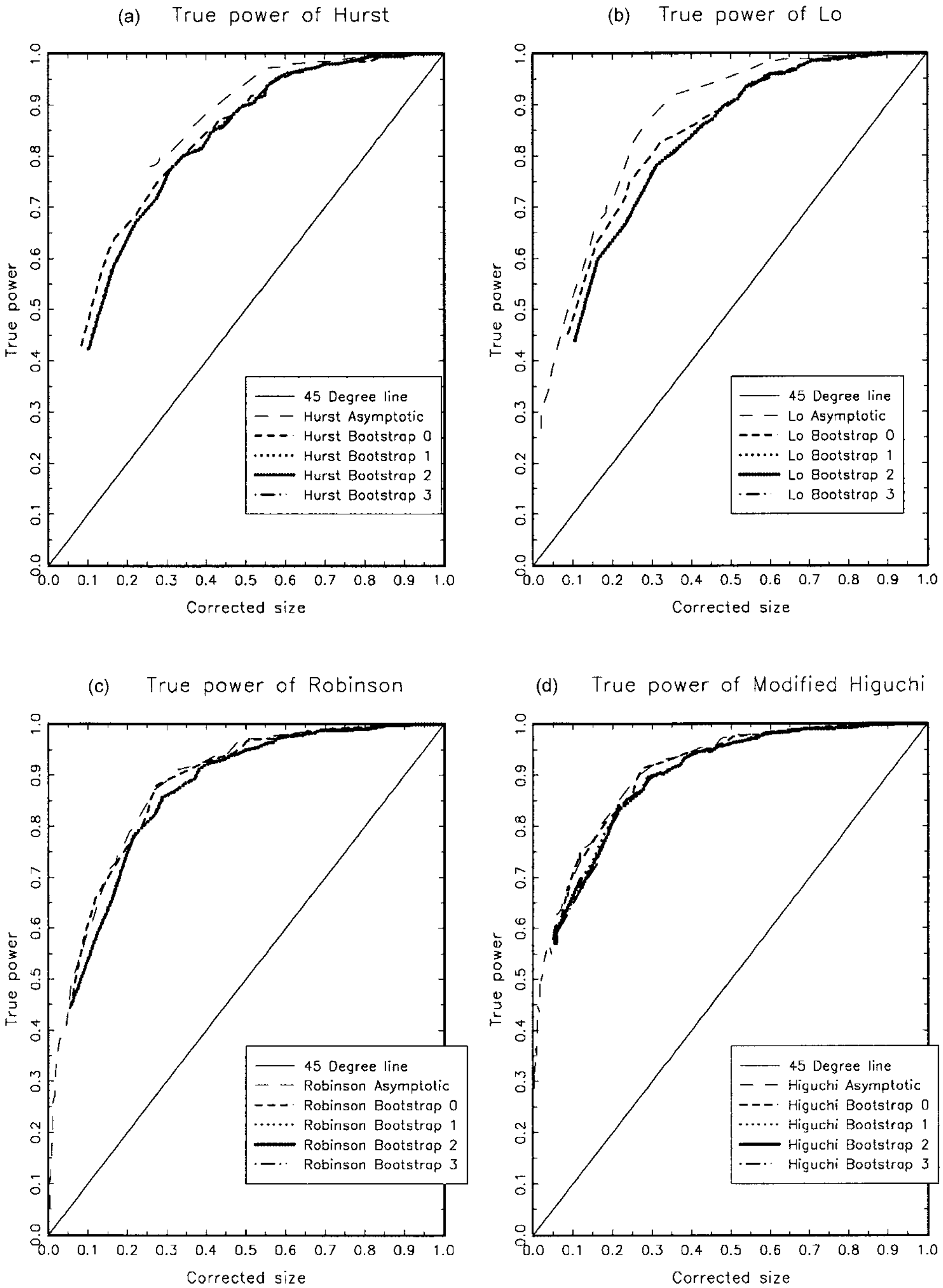

Figure 5. True powers for long memory tests. We are in the case of AR(p) processes in the Case 2 of the parameters (Table IV).

the case of AR(p) processes, the correction of the size distortion is worst than the correction in the Gaussian case, but parametric and nonparametric bootstraps perform quite similarly. In all the Student's t cases, there is no true power loss from the use of bootstrap methods.

4.5. CASE OF UNILATERAL P VALUE

The results for unilateral P value tests are poor compared to bilateral P value tests. The size distortions are closer to the asymptotic distortions than the bilateral bootstrap distortions. Since the power curves are very close, one cannot see the difference between the unilateral tests and the bilateral tests for this criterion. But this is not important, because there is no problem for the true power.

5. Example: Long Memory Analysis for Silver Market Returns

The importance of long-memory in asset markets is considered initially in Mandelbrot (1971). Greene and Fielitz (1977) is perhaps the first study to apply R/S analysis to common stock returns. More recently, the evidence in Fama and French (1988), Lo and MacKinlay (1988), and Poterba and Summers (1995) may be indicative of a long memory component in stock market prices. More recent applications include Booth and Kaen (1979) (gold prices), Booth, Kaen, and Koveos (1982) (foreign exchange rates), and Helms, Kaen, and Rosenman (1984) (futures contracts).

5.1. DATA DESCRIPTION

As an illustration testing for long term memory in stock returns, I apply the previous analysis to a specific data set: daily observations of silver prices available from 01/07/1993 to 30/11/2001 yielding a sample size of 2,140 observations.

From Figure 6a, one sees that silver prices are not stationary. Bootstraped Dickey Fuller tests[6] confirm this. So, I use logarithmic differenced data. Let $\{x_t\}_{t=1}^{T}$ indicate the silver price series at time t, and define

$$dx_t = \ln(x_t) - \ln(x_{t-1}) \tag{17}$$

as the returns series at time $t \in \{1, 2, \ldots, T-1\}$.

I also subject this data set to an analysis developed in Ding et al. (1993), Ding and Granger (1996), and Granger and Ding (1996) dealing with the long memory property of various speculative returns. For this, I consider the absolute returns $|dx_t|$ and the squared returns dx_t^2, which are good measures of the volatility, since $E(dx_t^2)$ estimates the variance and $E(|dx_t|)$ the standard deviation of the series. Figures 6b–d show the plots of dx_t, $|dx_t|$, and dx_t^2.

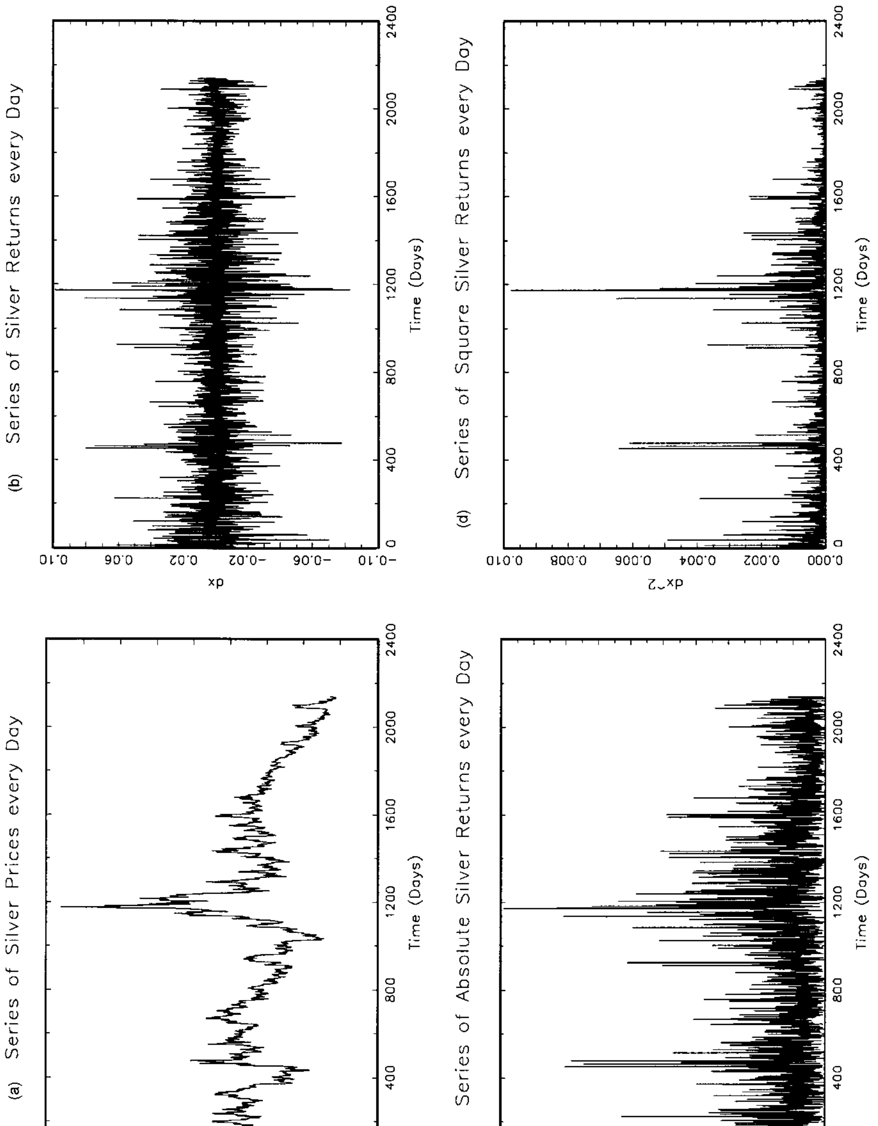

Figure 6. Four transformed series of silver prices every day.

5.2. RESULT DESCRIPTION

The Hurst, Lo, Robinson, and modified Higuchi's asymptotic and bootstrap P values[7] are performed for the returns, the absolute returns, and the square returns for the entire sample period and are given in Table V. $B_{Higuchi} = 999$ bootstrap replications are used to compute the modified Higuchi statistic for the asymptotic test, and $B_{Higuchi} = 39$ for the four bootstrap tests. $B = 399$ bootstrap replications are used to compute the bootstrap P values. There we see that, for the simple returns series, the asymptotic Hurst's and Lo's tests reject the hypothesis of long memory at the 0.01 level. The Robinson's and Modified Higuchi's asymptotic tests, however, accept this hypothesis, which contradicts the notion that financial series do not present evidence for long range dependence in the mean and the conclusions of the Hurst and Lo tests. While this rejection could be due to size distortion in the tests, we should note that all the bootstrap tests, which suffer less from size distortion, reject the hypothesis of long memory at 0.01.

It is also of interest to note that the P values (and the point estimates) indicate persistence of root less than unit root in the silver price first difference data.

Although Table V shows that the Robinson and the modified Higuchi asymptotic P values are statistically significant at the 0.01 level for the daily silver returns, none of the bootstrap P values is. The importance of bootstrapping in calculating these statistics is clear: the ARMA specification of the data drives the fractional integration result. The statistical insignificance of the bootstrap P values indicates that the data respect the short memory null hypothesis.

Consider now the series of absolute returns. All the methods give a P value for long range dependence smaller than 0.05. The results are unambiguous.

The results are opposite for squared returns. However, note that all the bootstrap P-values are much less significant than the asymptotic P-values, so we cannot take into account the asymptotic results. The small P-values for the Robinson and modified Higuchi statistics are possibly due to nonrobustness against greatly skewed and leptokurtic data. The greater P-values for the Hurst and Lo statistics could be due to a break detected in the time series. These cases need further study.

6. Conclusion

I examine the behaviour of four tests for fractional integration: the re-scaled range statistic test (Hurst, 1951), the Lo's test (1991), the frequency-domain regression-based procedure (Robinson, 1995), and the Higuchi's estimator based test (1988). If one uses these asymptotic techniques to make small-sample inferences, two severe problems arise. First, under the null hypothesis H_0, there are very large size distortions. It is impossible to make correct inferences without correcting for the distortions. Second, under the alternative hypothesis H_1 for all the tests, H_1 can be accepted more under H_0 than under H_1 when there is persistent long-memory vs persistent short-memory.

Table V. P value results for long memory for series of transformed returns of silver prices.

Test method	Hurst	Lo	Robinson	Modified Higuchi
P value results for long memory for series of simple returns of silver prices				
Point estimated of d	–0.0880	–0.0880	–0.1973	–0.1249
Asymptotic P value	0.5446	0.6370	0.0013	0.0078
Bootstrap 0 P value	0.3960	0.5163	0.0000	0.1010
Bootstrap 1 P value	0.4712	0.5815	0.0251	0.1212
Bootstrap 2 P value	0.4712	0.5815	0.0251	0.1010
Bootstrap 3 P value	0.4712	0.5815	0.0251	0.0808
P value results for long memory for series of absolute returns of silver prices				
Point estimated of d	–0.0241	–0.0241	0.4473	0.3156
Asymptotic P value	0.0000	0.0000	0.0000	0.0000
Bootstrap 0 P value	0.0201	0.0201	0.0000	0.0202
Bootstrap 1 P value	0.0050	0.0050	0.0050	0.0404
Bootstrap 2 P value	0.0050	0.0050	0.0050	0.0404
Bootstrap 3 P value	0.0100	0.0050	0.0050	0.0202
P value results for long memory for series of square returns of silver prices				
Point estimated of d	0.4720	0.4720	0.8851	0.8289
Asymptotic P value	0.0000	0.0000	0.0000	0.0000
Bootstrap 0 P value	0.1454	0.1253	0.0000	0.0000
Bootstrap 1 P value	0.1604	0.1654	0.0050	0.0150
Bootstrap 2 P value	0.1604	0.1654	0.0050	0.0100
Bootstrap 3 P value	0.1554	0.1504	0.0050	0.0050

Long memory analysis of daily silver stock returns from 01/01/1985 to 30/11/2001 using the classical rescaled range asymptotic and the bootstrap P values, the modified rescaled range asymptotic and bootstrap P values, the Robinson asymptotic and bootstrap P values, and the modified Higuchi asymptotic and bootstrap P values. The bootstrap P values are bilateral.

I find that bootstrapping techniques can be used to provide correction. One of the tests uses a double bootstrap that provides better true power and size properties. Moreover, I use an bilateral bootstrap P value that permits the true power of the tests to grow when the size distortions are asymmetric.

The Monte Carlo results are the following. In the case of AR(1) processes, all the bootstrap tests correct the size distortions quasi perfectly, even for large values of $|\phi_1|$. Size distortions of the asymptotic tests are to large to allow correct inferences. So, in this case, the use of the bootstrap is essential.

In the case of AR(p) processes, there is over rejection by the Hurst, Lo, and Robinson bootstrap tests. The Higuchi (double) bootstrap tests are quasi-perfect.

This method performs better than the others, not because the Higuchi's estimator has better prop erties than the others, but because it uses a better estimation of the variance of d by bootstrapping in the statistic. Of course, we must also apply the bootstrap test procedure to correct the size distortion. The combination of both the bootstraps is called *double bootstrap*. In all the cases, the bootstrap test distortions are lower than the asymptotic test ones. I recommend, then, using bootstrap tests in these cases too. It is better to use double bootstrap tests, but the computing time can be large.

In the case of ARFIMA(0, d, 0) processes, the bootstrap does not result in any power loss. In all the cases studied, the intrinsic power curves of the bootstrap are essentially indistinguishable from the ones obtained for the asymptotic distribution, even for parameter values chosen to give the bootstrap tests possible problems.

With leptokurtic error terms, in an AR(1) processes, the correction of the size distortion is quasi-perfect for all the parametric and nonparametric bootstraps. For an AR(p) processes, the correction of the size distortion is worst than the correction in the Gaussian case, but parametric and nonparametric bootstraps perform similarly. In all the Student's t cases, there is no true power loss from use of bootstrap methods.

The results for unilateral (symmetric) P value tests are poor compared to bilateral (asymmetric) P value tests. The size distortions are closer to the asymptotic distortions than the bilateral bootstrap distortions.

In Section 5, these proposed methods are applied to daily silver data to see if long memory behaviour could be detected. The asymptotic tests can confuse short memory in the series mean with long memory. The bootstrap test performs better in detecting short memory. The long memory of the variance of the series is detected clearly from bootstrapping.

My principal conclusion is that one must use at least a bilateral bootstrap test to detect long-range dependence in time series. One cannot trust the results of non-bootstrapped or unilateral bootstrapped tests because of excessive size distortion. The best way to proceed is to use a double bilateral bootstrap test, even though this can be time consuming.

Acknowledgements

I greatly thank Dr. Mohamed Boutahar and Prof. Marimoutou Vêlayoudom for readings and advice. I thank also the 8th International Conference of the Society for Computational Economics in Aix-en-Provence 2002 (France) organised by Christophe Deissenberg (University of Aix Marseille II), the 34e Journées de Statistiques of the Société Française de Statistique in Bruxelles 2002 (Belgium) organised by the Université Catholoque de Louvain, and the 7e colloque des jeunes économètres in Mont Sainte Odile 2002 (France) organised by the Université de Strasbourg for questions and remarks. Finally, I thank the referee of Computational Economics review which corrects the paper.

Appendix A: Long Memory Tests

A.1. THE R/S METHODS

Let $\bar{x}_T = \frac{1}{T}\sum_{i=1}^{T} x_i$ denote the sample mean and $s_T^2 = \frac{1}{T}\sum_{=1}^{T}(x_i - \bar{x}_T)^2$ the sample variance. The R/S statistic is given by

$$\tilde{Q}_T = \frac{R_T}{s_T}, \tag{18}$$

where

$$R_T = \left[\max_{1\le t\le T}\sum_{i=1}^{t}(x_i - \bar{x}_T) - \min_{1\le t\le T}\sum_{j=1}^{t}(x_j - \bar{x}_T)\right], \tag{19}$$

for every $T > 0$. Using Mandelbrot's proof, we have in our case under the null hypothesis:

$$\frac{\tilde{Q}_T}{\sqrt{T}} \to V, \tag{20}$$

where V is the range of a Brownian Bridge on the unit interval.

A.2. THE MODIFIED R/S METHOD

The Newey and West (1987) estimator of the variance is

$$\hat{\sigma}_T^2(q) = \frac{1}{T}\sum_{i=1}^{T}(x_i - \bar{x}_T)^2 + \frac{2}{T}\sum_{j=1}^{q}\omega_j(q)\sum_{k=j+1}^{T}(x_k - \bar{x}_T)(x_{k-j} - \bar{x}_T), \tag{21}$$

where

$$\omega_j(q) = 1 - \frac{j}{q+1}, \quad q > T. \tag{22}$$

There is no criterion to choose q. For our experiments, we set $q = 2$. To test H_0, Lo (1991) proposes the statistic $\frac{Q_T}{\sqrt{T}}$ where $Q_T = \frac{R_T}{\hat{\sigma}_T}$.

A. 3. THE LOG-PERIODOGRAM METHOD

The spectral density of an stationary ARFIMA(p, d, q) process is proportional to λ^{-2d} near the origin, i.e., $f(\lambda) \overset{\lambda\to 0^+}{\sim} G\lambda^{-2d}$. The periodogram $I(\lambda)$ is used to estimate $f(\lambda)$ and thus d using the following method. Robinson (1995) has considered the objective function

$$Q(G, d) = \frac{1}{m}\sum_{i=1}^{m}(\ln(G\lambda_i^{-2d}) + G\lambda_i^{2dm-i} I(\lambda_i)), \tag{23}$$

calculated for the Fourier frequencies $\{\lambda_i = \frac{2\pi i}{T}; i = 1, 2, \ldots, m\}$, where m is an integer lower or equal to $\frac{T}{2}$. m determines the truncation point in the previous regression. Let $\Delta = [\Delta_1, \Delta_2]$ be the set of admissible values of d where $-\frac{1}{2} < \Delta_1 < \Delta_2 < \frac{1}{2}$. One can write

$$\hat{d}_R = \operatorname{argmin}\{R(d); d \in \Delta\}, \tag{24}$$

with

$$R(d) = \ln(\hat{G}_H) - 2d\frac{1}{m}\sum_{i=1}^{m}\ln(\lambda_i), \tag{25}$$

$$\hat{G}_H = \frac{1}{m}\sum_{i=1}^{m}\lambda_i^{2d} I(\lambda_i). \tag{26}$$

Robinson has shown that $\hat{d}_R$ is consistent (under certain conditions) and asymptotically normal

$$\sqrt{m}(\hat{d}_R - d) \overset{n\to\infty}{\longrightarrow} N\left(0, \frac{1}{4}\right). \tag{27}$$

A.4. THE HIGUCHI METHOD

The partial sums of x_t are $y_t = \sum_{i=1}^{t} x_i$. Higuchi (1988) first constructs a new time series, y_k^l, defined as

$$y(l), y(l+k), y(l+2k), \ldots, y\left(l + \left[\frac{T-l}{k}\right]k\right), \quad l \in \{1, 2, \ldots, k\}, \tag{28}$$

where [ ] denotes the integer part and both k and l are integers. He defines the length of the curves, y_k^l, as

$$L_l(k) = \frac{T-1}{[\frac{T-l}{k}]k^2}\sum_{i=1}^{[\frac{T-l}{k}]}|y(l+ik) - y(l+(i-1)k)|. \tag{29}$$

He defines the length of the curve for the time interval k, $L(k)$, as

$$\frac{1}{k}\sum_{i=1}^{k} L_l(k). \tag{30}$$

In the case of long-memory series, we have the following result:

$$E(L(k)) \sim c_L k^{d-1.5}. \tag{31}$$

Then, he plots $\log(L(k))$ against to $\log(k)$.[8] The maximum value of k, *kmax*, is $T/64$ for Higuchi (1988), but we must choose another value for the inference context. The line is fitted according to the least-square procedure.

A.5. THE WAVELET METHOD

A wavelet is defined as

$$\Psi_{j,k}(t) = a^{\frac{1}{2}}\Psi(a^j t - kb), \quad t \in \mathbb{R}, \tag{32}$$

where $(j, k) \in \mathbb{Z} \times \mathbb{Z}$ and $a > 0$, $b > 0$. One often sets $a = 2$ and $b = 1$. The mother wavelet function Ψ must satisfy $\int \Psi(t)dt = 0$. The best known wavelet family is the Daubechies one: C^k-functions with the $K \geq k$ first moments null. We use Daubechies wavelets with $K = 16$. $\{\Psi_{j,k};\ j \in \mathbb{Z}, k \in \mathbb{Z}\}$ forms an orthonormal basis in $L^2(\mathbb{R})$.

Let us decompose $L^2(\mathbb{R})$ in a growing sequence of closed subspaces $\cdots \subset V_{j-1} \subset V_j \subset V_{j+1} \subset \cdots$ that approximate $L^2(\mathbb{R})$ such as V_j and V_{j+1} are similar, i.e., $x_{2t} \in V_{j+1}$. Then $\{\Phi_{j,k};\ \Phi_{j,k}(t) = 2^{j/2}\Phi(2^j t - k), k \in \mathbb{Z}\}$ forms an orthonormal basis for the subspace V_j. The projection of $x \in L^2(\mathbb{R})$ on V_j can be represented as

$$\mathrm{Proj}_{V_j}(x) = \sum_{k\in\mathbb{Z}} c_{j,k}\Phi_{j,k}\,, \tag{33}$$

where $c_{j,k} = \langle x, \Phi_{j,k}\rangle$.

Consider W_j so that $V_{j+1} = V_j \oplus W_j$. $\{\Psi_{j,k};\ \Phi_{j,k}(t) = 2^{j/2}\Psi(2^j t - k), k \in \mathbb{Z}\}$ is an orthonormal basis of W_j. Since $\bigcup_{j\in\mathbb{Z}} V_j = L^2(\mathbb{R})$ and $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$, we see that $L^2(\mathbb{R}) = \oplus_{j\in\mathbb{Z}} W_j$. Since the W_is are mutually orthogonal, $\{\Psi_{j,k};\ j \in \mathbb{Z}, k \in \mathbb{Z}\}$ forms and orthogonal basis of $L^2(\mathbb{R})$. Therefore, all functions $x \in L^2(\mathbb{R})$ can be written as

$$x = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} d_{j,k}\Psi_{j,k}\,, \tag{34}$$

where $d_{j,k} = \langle x, \Psi_{j,k}\rangle$.

A time series x length $T = 2^L$, can be decomposed in a different basis, according to the chosen scale level j_M:

$$x = \sum_{k=0}^{2^{j_M}-1} c_{j_M,k}\Phi_{j_M,k} + \sum_{j=j_M}^{L-1}\sum_{k=0}^{2^j-1} d_{j,k}\Psi_{j,k}\,, \tag{35}$$

with $j_M \in \{0, 1, 2, \ldots, L-1\}$.

If x is fractionally integrated with order $d \in (-\frac{1}{2}, \frac{1}{2})$, Jensen (1994) shows that the coefficients of the wavelet decomposition $(d_{j,k})$ are normally distributed with zero mean and variance $\sigma_j^2 = \sigma^2 s^{2d(j-1)}$, where σ^2 is a positive constant proportional to σ_ε^2. This suggests the estimator of d that maximizes the likelihood function

$$L(\theta) = \prod_{j=j_M}^{L-1} \prod_{k=0}^{2^j-1} (2\pi\sigma_j^2)^{-\frac{1}{2}} \exp\left(-\frac{d_{j,k}^2}{2\sigma_j^2}\right), \tag{36}$$

$$\theta' = (d, \sigma^2). \tag{37}$$

The estimator $\hat{d}$ is asymptotically efficient.

Jensen (1994) shows that the wavelet choice is of little importance for the estimation of d. Likewise, the existence of additional noise (for instance, measure) has only a minor effect.

Appendix B: Graphical Methods

To obtain evidence on the finite-sample properties of hypothesis testing procedures, I use the graphical methods of Davidson and MacKinnon (1998a) that use simulation methods. Consider a Monte-Carlo experiment in which R realisations, denoted by $\{\tau_j;\ j = 1, 2, \ldots, R\}$, of a test statistic t are generated using a DGP that is a special case of the null hypothesis. All of the graphs are based on the empirical distribution function, or EDF, of the P values of the τ_j, named $\{p_j \equiv p(\tau_j);\ j = 1, 2, \ldots, R\}$. The definition of p_j will vary. For the asymptotic P value, if F_τ denotes the c.d.f. of the asymptotic distribution of τ, then

$$p_j = \begin{cases} 1 - F_\tau(\tau_j) & \text{if the test is unilateral (at the right)} \\ 2\min\{1 - F_\tau(\tau_j), F_\tau(\tau_j)\} & \text{if the test is bilateral.} \end{cases} \tag{38}$$

For the bootstrap P value, see Equations (7) and (8).

The estimation of the c.d.f. of $p(\tau)$, at any point α_i in the interval (0, 1), is defined by

$$\hat{F}(\alpha_i) \equiv \frac{1}{R} \sum_{j=1}^{R} I(p_j \leq \alpha_i), \tag{39}$$

where I is an indicator function that takes the value 1 if its argument is true and 0 otherwise.

B.1. P VALUE PLOTS

The first graph that I use is a plot of $\hat{F}(\alpha_i)$ against α_i, referred to a *P value plot* as in Davidson and MacKinnon (1998a). If F is the correct distribution of τ under the null hypothesis, then the p_j are distributed uniformly over (0, 1) for all j and the resulting graph should be close to the 45 degree line. However, because all test statistics behave only approximately, the P value plots show the size distortions of the tests.

B.2. SIZE-POWER CURVES

To compare the power of alternative test statistics, one must plot power against true size. These curves are constructed using two EDFs, one for an experiment in which the null hypothesis is true, and one for an experiment in which it is false.[9] The two approximate EDFs are denoted $\hat{F}_0$ and $\hat{F}_1$, respectively. Plotting the points $(\hat{F}_0(\alpha_i), \hat{F}_1(\alpha_i))$ generates a size-power curve on a correct size-adjusted basis.

There is an infinite number of DGPs that satisfy the null hypothesis. Since the test statistics are not pivotal, the choice of the DGP used to correct the size can matter greatly. Davidson and MacKinnon (1996) argues that a reasonable choice is the pseudo-true null, which is the DGP that satisfies the null hypothesis and is as close as possible, in the sense of the Kullback–Leibler Information Criterion, to the DGP used to generate $\hat{F}_1$; see also Horowitz (1994a, b). I use the pseudo-true null in the experiments.

Notes

[1] Persistent memory as opposed to anti-persistent ($d < 0$).

[2] THe ARFIMA model is presented in greater detail by Granger and Joyeux (1980) and Hosking (1981).

[3] It displays true test size against nominal size.

[4] It displays true power against true (corrected) size.

[5] If we do better estimations with ARMA(p, q) processes, they only better confirm the following results. But the computing times are larger.

[6] I use 1999 bootstrap replications for the three classical Dickey Fuller tests (without constant, with constant, and with trend), and for the three augmented Dickey Fuller tests.

[7] Asymptotic and classical bootstrap P values assigns equal probability to each tail. The bilateral bootstrap does not.

[8] To choose the values of the ks, see Higuchi (1988).

[9] I use the same sequence of random numbers in both experiments to reduce experimental error.

References

Andersson, M.K. and Gredenhoff, M.P. (1998). Robust testing for fractional integration using the bootstrap. *Working Paper Series in Economics and Finance No. 218*.

Bardet, J.M., Lang, G., Moulines, E. and Soulier, P. (2000). Wavelet estimator of long-range dependence processes. *Statistical Inference for Stochastic Processes*, **3**(1–2), 85–99.

Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall. London.

Booth, G. and Kaen, F. (1979). Gold and silver spot prices and market information efficiency. *Financial Review*, **14**, 21–26.

Booth, G., Kaen, F. and Koveos, P. (1982). R/S analysis of foreign exchange rates under two international monetary regimes. *Journal of Monetary Economics*, **10**, 407–415.

Davidson, R. and MacKinnon, J. (1993). *Estimation and Inference in Economics*. Oxford University Press, New York.

Davidson, R. and MacKinnon, J. (1998a). Graphical methods for investigating the size and the power of hypothesis tests. *The Manchester School*, **66**, 1–22.

Davidson, R. and MacKinnon, J. (1998b). The size distortion of bootstrap tests. *Working Paper, GREQAM*.

Davidson, R. and MacKinnon, J.G. (1996). The power of bootstrap tests. *Queen's University Institute for Economic Research, Discussion Paper 937*.

de Peretti, C. and Marimoutou, V. (2002). Are the long memory tests really effective? *GREQAM Working Paper No. 02A14*.

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

Ding, Z., Granger, C.W.J. and Engle, R.F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, **1**, 83–106.

Efron, B. (1979). Bootstrap methods: Another look at the Jacknife. *Annals of Statistics*, **7**, 1–26.

Fama, E. and Frensh, K. (1988). Permanent and temporary component of stock prices. *Journal of Political Economy*, **96**, 246–273.

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

Granger, C.W.J. and Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, **73**, 61–77.

Granger, C.W.J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional integration. *Journal of Time Series Analysis*, **1**(1), 15–29.

Greene, M. and Fielitz, B. (1977). Long-term dependence in common stock returns. *Journal of Financial Economics*, **4**, 339–349.

Helm, B., Kaen, F. and Rosenman, R. (1984). Memory in commodity futures contracts. *Journal of Futures Markets*, **4**, 559–567.

Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica*, **D 31**, 277–283.

Horowitz, J.L. (1994a). Bootstap-based critical values for the information matrix test. *Journal of Econometrics*, **61**, 395–411.

Horowitz, J.L. (1994b). Bootstrap methods in econometrics: Theory and numarical performance. *Paper Presented at the 7th World Congress of the Econometric Society, Tokyo*.

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

Hurst, H.E. (1951). Long-term storage capacity of reservoirs. *Transac. Am. Soc. Civil Eng.*, **116**, 770–808.

Jensen, M.J. (1994). Wavelet analysis of fractionally integrated processes. Dept. of Economics, Washington University, St. Louis, MO 63130.

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

Lo, A.W. and MacKinlay, C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, **1**, 41–66.

Mandelbrot, B. (1971). When can price be arbitraged efficiently? A limit to the validity of the random walk and martingale models. *Review of Economics and Statistics*, **53**, 225–236.

Mandelbrot, B. (1972). Statistical methodology for non-periodic cycles: From the covariance to R/S analysis. *Annals of Economic and Social Measurement*, **1**, 259–290.

Mandelbrot, B. and Wallis, J.R. (1969). Some long-run properties of geophysical records. *Water Resource Research*, **5**, 321–330.

Newey, W.K. and West, K.D. (1987). A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703–708.

Poterba, J. and Summers, L. (1995). Mean reversion in stock returns: Evidence and implications. *Journal of Financial Economics*, **22**, 27–60.

Robinson, P.M. (1995). Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, **23**, 1048–1072.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of statistics*, **6**, 461–464.

Weber, N.C. (1984). On resampling techniques for regression models. *Statistics and Probability Letters*, **2**, 275–278.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–26.

Wornell, G. and Oppenheim, A. (1992). Estimation of fractal signals from noisy measurements using wavelets. *IEEE Transaction on Signal Processing*, **40**, 611–623.