Bootstrap Testing for Fractional Integration*

Michael K. Andersson† and Mikael P. Gredenhoff†

Working Paper Series in Economics and Finance No. 188

August 1997

Abstract

Asymptotic tests for fractional integration, such as the Geweke-Porter-Hudak test, the modi ed rescaled range test and Lagrange multiplier type tests, exhibit size-distortions in small-samples. This paper investigates a parametric bootstrap testing procedure, for size-correction, by means of a computer simulation study. The bootstrap provides a practical method to eliminate size-distortions in the case of an asymptotic pivotal statistic while the power, in general, is close to the corresponding size-adjusted asymptotic test. The results are very encouraging and suggest that a bootstrap testing procedure does correct for size-distortions.

Key-words: Long-memory; ARFIMA; Parametric resampling; Small-sample; Monte Carlo simulation; Size-correction

JEL-classi cation: C15; C22; C52

---

*Valuable discussions with and comments by Tor Jacobson, Sune Karlsson and Rolf Tschernig are gratefully acknowledged. We also thank participants at the International Symposium on Forecasting (ISF 97), Bridgetown meeting. The usual disclaimer applies.

†Department of Economic Statistics, Stockholm School of Economics, E-mail: *stma@hhs.se, stmg@hhs.se.*

1. Introduction

The fractionally integrated autoregressive moving average ($ARFIMA$) model has recently received considerable attention in economics, but also in other research areas. ARFIMA processes generalize linear $ARIMA$ models by allowing for non-integer differencing powers and do thereby provide a more exible framework for analyzing time series data. This exibility enables fractional processes to model stronger data dependence than what is allowed in stationary ARMA models without resorting to non-stationary unit-root processes. However, estimators of the fractional model exhibit larger bias and are computationally more demanding. It is, therefore, bene cial to discriminate fractionally integrated processes from ARIMA speci cations in a rst modelling step, that is to test the null-hypothesis of an integer differencing power against a fractional alternative. For this purpose the literature frequently utilizes the Geweke and Porter-Hudak (1983) test, the modi ed rescaled range test of Lo (1991) and Lagrange multiplier tests, see e.g. Agiakloglou and Newbold (1994). The size and power of these asymptotic tests are investigated by Cheung (1993) and Agiakloglou and Newbold. One nding in their studies is the existence of non-negligible small-sample size-distortions.

To improve inference, classical statistical theory employs expansions to provide analytical corrections. By numerical means, similar corrections can be given by bootstrap methods. While analytical corrections modi es the test statistic to approach the asymptotic distribution more rapidly, the bootstrap adjusts the critical values so that the true size of the test converges to its nominal value. This paper applies a parametric bootstrap testing procedure. The size-distortion of such a procedure, based on parameter estimates under the null, will according to Davidson and MacKinnon (1996a) be at least one full order, $O(T^{-1})$, smaller the distortion of the asymptotic test. Thus, the bootstrap is said to provide a trustworthy technique to perform inference in small samples and yields, under regularity conditions, exact or close to exact tests. The purpose of this paper is to examine this claim.

The paper is organized as follows. Section 2 brie y describes the tests and section 3 introduces the bootstrap testing procedure. Section 4 contains the Monte Carlo simulation study, where size and power of the tests are compared with their bootstrap analogues. Section 5 concludes.

2. Testing for Fractional Integration

A time series $\{x_t\}$ follows an ARFIMA(p, d, q)[1] process if

[1] The properties of the fractionally integrated ARMA model are presented by Granger and Joyeux (1980) and Hosking (1981).

$$\phi(B)(1-B)^d x_t = \theta(B) a_t, \tag{2.1}$$

where $\{a_t\}$ is a series of independently and identically distributed disturbances with mean zero and variance $\sigma_a^2 < \infty$, and $\phi(B)$ and $\theta(B)$ are the autoregressive and moving average polynomials in the backshift operator B. If the roots of $\phi(B)$ and $\theta(B)$ are outside the unit circle and $d < |0.5|$, x_t is both stationary and invertible. When $d > 0$, x_t is persistent in the sense that the autocorrelations are not absolutely summable. Thus, there exists a region $(0 < d < 0.5)$ where the ARFIMA model is capable of generating stationary series which are persistent.[2] If $d \neq 0$ the process displays long-memory characteristics, such as a hyperbolic autocorrelation decay , while the stationary ARMA model exhibits a faster geometrical decay (given the existence of AR parameters).

If d is integer-valued, the ARFIMA process reduces to an ARIMA process. The tests are applicable on stationary and invertible series and the series are, subsequently, differenced or summed until this is satis ed. $d = 0$ is thus a natural null-hypothesis when testing for fractional integration.

2.1. The Periodogram Regression Test of Geweke and Porter-Hudak

Geweke and Porter-Hudak (1983), henceforth GPH, proposed the following non-parametric periodogram regression,

$$\ln\{I_x(\omega_j)\} = \alpha - d\ln\left\{4\sin^2(\omega_j/2)\right\} + v_j, \tag{2.2}$$

for the estimation of the fractional difference parameter. $I_x(\omega_j)$ is the periodogram at the harmonic frequencies $\omega_j = 2\pi j/T$, where $j = 1, ..., g(T)$. Under a proper choice of $g(T)$, the ordinary least squares (OLS) estimator of d is consistent and the distribution of $\left(\hat{d}_{OLS} - d\right)/SE\left(\hat{d}_{OLS}\right)$ is asymptotically normal. The known variance of v, $\pi^2/6$, is used to increase the efficiency of the test and $g(T)$ is commonly selected as $T^{1/2}$.

2.2. The Modi ed Rescaled Range Test

The rescaled range statistic was proposed by Hurst (1951) and has been re ned by Mandelbrot (1972) and MacLeod and Hipel (1978). A version of the statistic, which is robust to short-range dependence in data was suggested by Lo (1991). This modi ed rescaled range (MRR) statistic is de ned by the ratio

[2]Persistence is commonly found in Economic time series, i.e. real exchange rates and unemployment.

$$\tilde{Q}_T = \frac{R_T}{\hat{\sigma}_T(k)}, \tag{2.3}$$

where the range and standard error are calculated by

$$R_T = \max_{0<i\leq T} \sum_{t=1}^{i} (x_t - \overline{x}) - \min_{0<i\leq T} \sum_{t=1}^{i} (x_t - \overline{x}) \tag{2.4}$$

$$\hat{\sigma}_T^2(k) = \hat{\sigma}^2 + 2\sum_{j=1}^{k}\sum_{i=j+1}^{T}\left(1 - \frac{j}{k+1}\right)(x_i - \overline{x})(x_{i-j} - \overline{x}). \tag{2.5}$$

The truncation lag k depends on the short-term correlation structure of the series and is set, according to Andrews (1991) data dependent formula, to the integer part of $(3T/2)^{\frac{1}{3}}\left\{2\hat{\rho}/\left(1-\hat{\rho}^2\right)\right\}^{\frac{2}{3}}$, where $\hat{\rho}$ denotes the sample rst-order autocorrelation coefficient and $\hat{\sigma}^2$ the maximum likelihood variance estimate. Asymptotic critical values of the MRR test is given by Lo (1991).

2.3. A Lagrange Multiplier Test

The LM test, denoted REG, of Agiakloglou and Newbold (1994) is carried out through the likelihood based auxiliary regression

$$\hat{a}_t = \sum_{i=1}^{p} \beta_i W_{t-i} + \sum_{j=1}^{q} \gamma_j Z_{t-j} + \delta K_m + u_t, \tag{2.6}$$

where

$$K_m = \sum_{j=1}^{m} j^{-1}\hat{a}_{t-j},\ \hat{\theta}(B) W_t = x_t,\ \hat{\theta}(B) Z_t = \hat{a}_t \text{ and } u_t \text{ is } iid \text{ normal.}$$

$\hat{a}_t$ and $\hat{\theta}(B)$ are the estimated residual and MA polynomial from the ARFIMA speci cation (2.1) under the null-hypothesis.

The autoregressive and moving average orders p and q are estimated by the Bayesian information criterion (BIC) of Schwartz (1978). According to Agiakloglou and Newbold a small value of the truncation lag m is preferable, therefore m is set equal to ve. The equation (2.6) is tted by non-linear least squares (the $IMSL$ routine DNSLSE) over the time period $t = m+1, ..., T$. The usual t-test of the hypothesis $\delta = 0$ together with asymptotically normal critical values constitutes the LM test.

2.4. The Size and Power of the GPH, MRR and REG tests

Cheung (1993) presents size and power for the MRR and GPH tests. This is done for a variety of AR(1), MA(1) and ARFIMA(0,d,0) processes with positive and negative parameter values. The MRR test is conservative for autoregressions, that is the empirical size is smaller than the nominal, for almost every parameter value and serial length. For large positive AR parameters, the GPH test is severely over-sized whereas it is well-sized for the remaining parameter values. Rejection frequencies of both the MRR and GPH are notably larger than the nominal signi cance level when the MA parameter is close to -0.9.

The empirical size of the REG test is similar to the asymptotic size according to Agiakloglou and Newbold (1994). In contrast to the thorough investigation of the MRR and GPH test, the size of the REG test is only computed for $\phi = 0.5$ and 0.9, and $T = 100$. Under the, unrealistic, assumption of a known AR order the REG test exhibits high rejection frequencies when the true process is fractionally integrated. A lower power is expected when the lag-order is unknown. The MRR test has difficulties to detect positive fractional integration, especially in moderate sample sizes. Independently of the serial length, the GPH test displays a low rejection frequency for weakly persistent processes. Our study con rms these conclusions and extend them for the REG test.

3. The Bootstrap Test

The nite-sample distribution of a test statistic may not always coincide with its asymptotic distribution. One feasible way to estimate the small-sample distribution is through a bootstrap procedure, see for instance Horowitz (1995) for an introduction and overview. The size-distortion of a bootstrap test is of order $T^{-1/2}$ smaller than that of the corresponding asymptotic test. A further re nement of order $T^{-1/2}$ can be obtained in the case of an asymptotically pivotal statistic, i.e. a statistic whose limiting distribution is independent of unknown nuisance parameters. This is achieved without the complex derivations of analytical higher order expansions. If the signi cance level of a test is calculated using a bootstrap procedure, an exact or close to exact test is often the result, which enables more reliable inference in nite-samples. Following Davidson and MacKinnon (1996a), such a procedure will be referred to as a bootstrap test.

The objective of the test is to compute the p-value function,

$$p(\hat{\tau}) = p(\tau \geq \hat{\tau} \,|\Psi_0, T), \tag{3.1}$$

where Ψ_0 is the true data generating process (DGP) under the null hypothesis, T is the sample size and $\hat{\tau}$ is a realized value of the test statistic τ based on the original sample $\mathbf{x} = [x_1, ..., x_T]'$. The DGP Ψ_0 is characterized by an unknown ARMA(p, q) speci cation. Since the null model, and hence Ψ_0, is unknown the estimated (bootstrap) DGP $\hat{\Psi}_0$ is employed to create the bootstrap samples. The basic idea is to create a large number of such samples which all obey the null-hypothesis and, as far as possible, resembles the original sample.

In this paper we use a parametric bootstrap algorithm[3], for which the DGP $\hat{\Psi}_0$ is based on parameter estimates under the null, that is retrieving $\hat{\Psi}_0$ from the estimated ARMA$(\hat{p}, \hat{q})$ model,

$$\left(1 - \hat{\phi}_1 B - ... - \hat{\phi}_{\hat{p}} B^{\hat{p}}\right) x_t = \left(1 + \hat{\theta}_1 B + ... + \hat{\theta}_{\hat{q}} B^{\hat{q}}\right) \hat{a}_t, \tag{3.2}$$

where $\hat{a}_t$ is the residual at time t. Alternatively, the re-sampling model may be estimated by

$$\left(1 - \tilde{\phi} B - ... - \tilde{\phi}_{\tilde{p}} B^{\tilde{p}}\right) x_t = \tilde{a}_t, \tag{3.3}$$

which can be regarded as the estimated AR representation of the bootstrap DGP. The models (3.2) and (3.3) are estimated, conditional on stationarity and invertibility conditions, by the BIC and non-linear least squares and OLS (the *IMSL* routines DNSLSE and DRLSE) respectively. The orders p and q are allowed to a maximum lag of ve for the ARMA model, whereas a maximum lag p of 30 is allowed for the AR speci cation.[4]

The bootstrap samples, each denoted $\mathbf{x}_r^*, r = 1, ..., R$, are created recursively using

$$x_{r,t}^* = \phi^* (B)^{-1} \theta^* (B) a_t^*, \tag{3.4}$$

where $\phi^* (B)$ and $\theta^* (B)$ are the estimated polynomials of $\hat{\Psi}_0$. The a_t^*:s are independent draws from a normal distribution with mean zero and variance $s_{\hat{a}}^2$ or $s_{\tilde{a}}^2$.

If R bootstrap re-samples, each of size T, and their respective test statistics τ_r^* are generated, the estimated bootstrap p-value function, for a two-sided test, is de ned by

$$p^* (\hat{\tau}) = R^{-1} \sum_{r=1}^{R} I \left(|\tau_r^*| \geq |\hat{\tau}|\right), \tag{3.5}$$

where $I(\cdot)$ equals one if the inequality is satis ed and zero otherwise, and the number of bootstrap replicates R is chosen as 1000. The null hypothesis is rejected when the selected signi cance level exceeds $p^* (\hat{\tau})$.

[3]The use of a parametric bootstrap is motivated by the assumed normality of the data. Further resampling procedures are evaluated by Andersson and Gredenhoff (1997).

[4]Preliminary results suggest that no signi cant AR parameters enter the estimated polynomial after the $30th$ lag when the true process is an MA(1) with $\theta = 0.9$.

Davidson and MacKinnon (1996b) show that the power of a bootstrap test, based on a pivotal statistic, is generally close to the size-adjusted asymptotic test. Even if the statistic is only close to pivotal this is generally true.

4. The Monte Carlo Study

The experiment covers first order autoregressions and moving averages, and fractional noise series of lengths $T = 50$, 100, 300 and 500. We generate $T + 100$ normally distributed pseudo random numbers, using the $IMSL$ routine DRNNOA, and discard the first 100 observations to reduce the effect of initial values. The AR and MA series are then constructed recursively and the fractional noise series are generated using the algorithm of Diebold and Rudebusch (1991).

The Monte Carlo study involves 1000 replicates (series), where each series is tested for fractional integration using the tests described in Section 2 and 3.

The bootstrap resamples are created by the ARMA (3.2) and AR (3.3) specifications. Reported results are based on the AR resampling model, due to its better performance. The AR specification works better than a pure MA resampling model even when the true process is a moving average, regardless of parameter values.

Estimated size and power of the different processes in the study are computed as the rejection frequencies of the non-fractional null hypothesis.

4.1. AR and MA Processes

The empirical size of the tests are examined for the specifications

$$x_t = \phi x_{t-1} + a_t \tag{4.1}$$

and

$$x_t = a_t + \theta a_{t-1}, \tag{4.2}$$

where the members of $\{a_t\}$ are iid $N(0, 1)$. The AR and MA parameters ϕ and θ are set equal to ± 0.1, ± 0.5 and ± 0.9. Table 4.1[5] presents the sensitivity, at a nominal 5% level of significance, of the empirical size with respect to AR and MA parameters.

The estimated size of the MRR test for both AR and MA processes differs, in general significantly at the 5% level, from the nominal size. Significant differences, based on a 95% confidence interval, are obtained when the rejection frequencies lie outside $(3.6, 6.4)$. The MRR test is, in general, for AR as well as MA processes

[5] All results are approximately valid for the 1% and 10% nominal significance level and for T=300 and 500.

Table 4.1: Rejection percentage of the nominal 5 percent fractional integration test when the data follows an AR(1) or MA(1) process of length T.

	MRR		GPH		REG	
ϕ/θ	Orig.	Boot.	Orig.	Boot.	Orig.	Boot.
T=50			*AR(1) Processes*			
-0.9	**18.9**	**2.2**	4.7	5.4	**6.9**	4.7
-0.5	**0.9**	**3.1**	4.7	5.3	6.2	4.1
-0.1	**6.7**	4.5	5.3	5.1	5.8	3.9
0.1	5.0	3.8	5.5	4.8	5.7	4.1
0.5	**2.5**	**3.1**	**8.2**	5.1	**6.8**	4.7
0.9	**1.2**	4.1	**63.6**	3.6	**7.8**	4.8
T=100						
-0.9	6.4	**2.1**	6.4	5.6	5.6	4.8
-0.5	**0.8**	5.0	5.7	5.0	5.1	4.6
-0.1	**6.6**	4.8	4.9	4.5	6.2	3.6
0.1	**6.6**	5.9	4.9	4.5	6.2	4.6
0.5	**2.5**	4.8	**8.3**	4.4	**6.5**	4.4
0.9	**0.9**	3.6	**71.8**	**2.7**	5.2	5.2
T=50			*MA(1) Processes*			
-0.9	4.8	5.4	**41.2**	**25.8**	**9.3**	**9.5**
-0.5	5.5	5.0	**7.7**	5.0	**3.3**	**1.6**
-0.1	**7.3**	4.5	5.0	4.5	**6.6**	5.4
0.1	**6.7**	4.9	4.3	4.0	**7.4**	5.3
0.5	3.7	4.3	4.4	4.8	4.0	4.4
0.9	**2.0**	**3.2**	5.2	3.7	**10.0**	6.3
T=100						
-0.9	**9.9**	**10.5**	**50.1**	**36.3**	**8.0**	**7.2**
-0.5	4.2	4.8	**7.9**	5.6	**2.8**	**2.8**
-0.1	**6.6**	5.1	4.9	5.5	5.6	4.3
0.1	5.4	4.7	5.0	5.4	5.2	4.1
0.5	**3.2**	5.7	5.4	5.7	**2.7**	4.0
0.9	**2.3**	4.8	6.0	4.4	**6.8**	4.4

The number reported in the table is the rejection percentage of the two-sided 5% test. Numbers in bold face denote signi cant deviations from the nominal size. Under the null hypothesis of no fractional integration, the 95% con dence interval of the rejection percentage equals (3.6, 6.4). In the table head, Orig. denotes the original test and Boot. the corresponding bootstrap test.

over-sized for large negative and conservative for large positive parameters. Exactly as in Cheung (1993), AR series with parameter $\phi = -0.5$ leads to very low rejection frequencies. The MRR test is always conservative for autoregressions near the unit circle for larger sample sizes ($T = 300$ and 500), whereas the rejection frequencies increase with T for moving averages with $\theta = -0.9$. The GPH test is well-sized, except for highly short-term AR/MA dependent series with positive roots[6]. Agiakloglou *et al.* (1993) show that large positive AR and MA roots biases the periodogram (2.2), resulting in biased estimates of d and hence large test statistics. These results are close to those of Cheung.

Extending the results of Agiakloglou and Newbold (1994), we nd the REG test well-sized, compared with the other tests, for the entire AR parameter space when $T = 100$. However, the test is over-sized for $T = 50$ and $|\phi|$ close to unity. This over-sizing tendency, close to the unit circle, is enhanced for moving average processes. This is most pronounced for series of length $T = 50$, where large empirical sizes also occur for small parameters. The performance of the REG test improves with the serial length (considering also $T = 300$ and 500).

The simulation results suggest, in general, that the bootstrap testing procedure is able to improve the tests. Moreover, every bootstrap test has better size properties than any of the original test. In more detail, the bootstrap MRR test is found conservative when $\phi = -0.9$, whereas $\theta = -0.9$ leads to higher rejection frequencies than the nominal signi cance level. Over the parameter space, the dispersion of the sizes for each test is smaller than that of the original tests. The bootstrap procedure improves the MRR test, that is only two out of twelve (AR and MA) empirical sizes differ signi cantly from the nominal size at sample size 100, compared to nine for the original test.

The size problems encountered by the GPH test for autoregressions are adjusted by the bootstrap procedure. In particular, the bootstrap correction is remarkable for $\phi = 0.9$ processes. The bootstrap is also able to correct for size-distortions due to intermediate positive MA roots and partly adjust the size for large positive roots. The empirical size for $\theta = -0.9$ is unfortunately still very large for the bootstrap GPH test. One might think that this is due to the AR resampling, but the size-adjustment is even smaller when using a pure MA resampling. Furthermore, the bootstrap procedure does not impose distortions where the original GPH test is well-sized. In general, the bootstrap GPH test works considerably better than the original test.

The bootstrap procedure corrects the size distortions of the REG test for autoregressive processes, that is the rejection frequencies of the bootstrap REG test are always within the con dence bounds. The bootstrap also corrects when the

[6]Due to the de nition of the AR and MA polynomials, positive $\phi's$ imply positive roots and positive $\theta's$ negative roots.

Table 4.2: Rejection percentage of the nominal 5 percent fractional integration test when the data follows an ARFIMA(0,d,0) process of length T=100.

	MRR		GPH		REG	
d	Orig.	Boot.	Orig.	Boot.	Orig.	Boot.
-0.45	14.6	16.5	21.8	21.0	31.0	31.3
-0.25	11.3	13.6	8.3	11.1	14.1	18.7
-0.05	3.7	6.1	3.9	5.3	4.6	5.6
0	5.0	5.3	5.0	4.4	5.0	3.5
0.05	5.8	6.9	4.4	5.3	10.1	6.6
0.25	14.7	14.3	17.0	16.5	26.1	27.1
0.45	4.0	18.6	40.9	11.8	17.2	21.5

See note to Table 4.1. The original tests are size-adjusted.

process is an MA with $\theta > -0.5$, whereas $\theta \leq -0.5$ processes leads to signi cant size distortions for all serial lengths. Except for these cases, the bootstrap REG test is correctly (on the 95% level) sized and more robust than its original version, in particular for MA series.

4.2. Fractional Processes

The power of the tests against ARFIMA$(0, d, 0)$ are studied using data constructed by

$$(1 - B)^d x_t = a_t, \tag{4.3}$$

where the fractional differencing parameter, d, is set equal to ± 0.05, ± 0.25 and ± 0.45. Table 4.2 presents the power, as a function of d, of the tests.The simulation results verify that the power of the bootstrap tests are close to the power of the size-adjusted asymptotic tests. We nd the dispersion of the different power functions least pronounced for the REG test, which relates to the small size-improvements of the bootstrap. The REG tests and the original MRR test reduce in power when the true d is close to 0.5 compared to a slightly lower d value, which is not the case of the bootstrap MRR.

When specifying the auxiliary regression (2.6), for the REG test, a large true fractional differencing power is interpreted as a large autoregressive order, yielding decreased rejection frequencies for the test. For the MRR test, the truncation lag k in (2.3) increases with d, i.e. too many autocorrelations are included in the variance correction term (2.5), resulting in a negatively biased estimate of the d parameter which lower the power of the original test.

A substantially lower power is found for the bootstrap GPH compared its original version when $d = 0.45$. A large differencing power results in a rich parameter structure of the ARMA resampling model in the bootstrap procedure. The rich parametrization implies that the resample periodograms resemble the periodogram of the original, highly persistent process. Thus, the bootstrap GPH test will have difficulties to distinguish fractional processes from ARMA speci cation, which can be seen in Table 4.2.

The power properties suggest that the REG test is superior when testing for fractional integration in small samples. In larger samples the MRR test is more powerful when $d < 0$ and the GPH test when $d > 0$.

5. Conclusions

The bootstrap testing procedure has better size properties than the original tests, that is the bootstrap corrects existing size-distortions without introducing new ones. All bootstrap tests are close to exact on the 95% con dence level, with an exception for MA(1) processes with a large positive root.

In general, the power of the bootstrap tests are close to the power of the corresponding size-adjusted asymptotic tests. The REG test is the most powerful test in small samples and by using the bootstrap version we get a test which is robust to ARMA components and has power properties similar to the original test. In larger samples the bootstrap MRR and GPH have higher power, when the alternative hypothesis is one-sided.

We conclude that a bootstrap testing procedure provides a practical and effective method to improve existing tests for fractional integration.

6. References

Agiakloglou, C., P. Newbold and M. Wohar (1993) Bias in an Estimator of the Fractional Difference Parameter. *Journal of Time Series Analysis* 14 (3), 235-246.

Agiakloglou, C. and P. Newbold (1994) Lagrange Multiplier Tests for Fractional Difference. *Journal of Time Series Analysis* 15 (3), 253-262.

Andersson, M. K. and M. P. Gredenhoff (1997) Robust Testing for Fractional Integration using the Bootstrap, Work in Progress, *Dept. of Economic Statistics, Stockholm School of Economics, Sweden.*

Andrews, D. (1991) Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* 59, 817-858.

CHEUNG, Y.-W. (1993) Tests for Fractional Integration: A Monte Carlo Investigation. *Journal of Time Series Analysis* 14 (4), 331-345.

DAVIDSON, R. and J. G. MACKINNON (1996a) The Size Distortion of Bootstrap Tests. *GREQAM Document de Travail No.* 96A15 and *Queens Institute for Economic Research Discussion Paper No.* 936.

DAVIDSON, R. and J. G. MACKINNON (1996b) The Power of Bootstrap Tests. *Queens Institute for Economic Research Discussion Paper No.* 937.

DIEBOLD, F. X. and G. D. RUDEBUSCH (1991) On the Power of Dickey-Fuller Tests against Fractional Alternatives. *Economics Letters* 35, 155-160.

GEWEKE, J. and S. PORTER-HUDAK (1983) The Estimation and Application of Long Memory Time Series Models. *Journal of Time Series Analysis* 4 (4), 221-238.

GRANGER, C. W. J. and R. JOYEUX (1980) An Introduction to Long-Memory Time Series Models and Fractional integration. *Journal of Time Series Analysis* 1 (1), 15-29.

HOROWITZ, J. L. (1995) Bootstrap Methods in Econometrics: Theory and Numerical Performance. Paper presented at the 7th World Congress of the Econometric Society, Tokyo.

HOSKING, J. R. M. (1981) Fractional Differencing. *Biometrica* 68, 165-176.

HURST, H. E. (1951) Long Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers* 116, 770-799.

LO, A. W. (1991) Long Term Memory in Stock Market Prices. *Econometrica* 59 (5), 1279-1313.

MACLEOD, A. I. and K. W. HIPEL (1978) Preservation of the Rescaled Adjusted Range, Part I, A Reassessment of the Hurst Phenomenon. *Water Resources Research* 14 (3), 491-518.

MANDELBROT, B. B. (1972) Statistical Methodology for Non-Periodic Cycles: From the Covariance to R/S Analysis. *Annals of Economic and Social Measurements* 1, 259-290

SCHWARTZ, G. (1978) Estimating the order of a model. *The Annals of Statistics* 2, 461-464