Robust Testing for Fractional Integration using the Bootstrap*

Michael K. Andersson† Mikael P. Gredenhoff†

Working Paper Series in Economics and Finance No. 218

January 1998

Abstract

Asymptotic tests for fractional integration are usually badly sized in small samples, even for normally distributed processes. Furthermore, tests that are well-sized under normality may be severely distorted by non-normalities and ARCH errors. This paper demonstrates how the bootstrap can be implemented to correct for such size distortions. It is shown that a well-designed bootstrap test based on the MRR and GPH tests is exact, and a procedure based on the REG test is nearly exact.

Key words: Long-memory; Resampling; Skewness and kurtosis; ARCH; Monte Carlo; Size correction.

JEL-classi cation: C12; C15; C22; C52.

---

*We are greatly indebted to Tor Jacobson and Sune Karlsson for useful discussions and comments. Financial support from the Tore Browaldh Foundation is gratefully acknowledged. The usual disclaimer applies.

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

1. Introduction

Many financial time series display characteristic features such as observations that are non-normally distributed (i.e. with excess skewness and kurtosis), conditionally heteroskedastic and ruled by long-memory. For instance Ding, Granger and Engle (1993) report evidence of autocorrelations between distant lags for long lags in the absolute returns of the Standard and Poor 500, S&P500, composite stock index. Furthermore, Granger and Ding (1995) show that the absolute value of the rate of return for a variety of stock prices, commodity prices and exchange rates exhibit excess skewness and kurtosis.

Long-memory is usually described by fractionally integrated specifications, hence testing for long-memory may be performed via a test for a fractional differencing power. For this purpose, several tests have been proposed and some of the most popular are thoroughly investigated by Cheung (1993), who also studies the influence of ARCH disturbances. Andersson and Gredenhoff (1997) implement a bootstrap method, in order to size-adjust fractional integration tests. The bootstrap provides a trustworthy technique for estimation of the small-sample distribution of a statistic. When using a bootstrap test the null-distribution is retrieved by bootstrap methods and hence the critical values are adjusted to give exact tests.

This paper investigates some fractional integration tests when the data are non-normal or the residuals are heteroskedastic. Again, the bootstrap is used to correct for size distortions. Another extension is a comparison of parametric, non-parametric and heteroskedasticity invariant resampling algorithms. The aim of the paper is to find tests that are robust to non-normalities and ARCH effects in data, and thus are well-suited when testing for long-memory in financial and economic time series.

The results suggest that the performance of a bootstrap testing procedure depends to some extent on the chosen resampling algorithm. However, all (but one) bootstrap tests are superior to the original version of the tests, in the sense that the bootstrap tests have better size properties.

The paper is organized as follows. Section 2 describes the bootstrap testing procedure and Section 3 contains a Monte Carlo simulation study where the sizes of the tests are presented for normal and non-normal data and processes with ARCH errors. Section 4 concludes the paper.

2. The Bootstrap Testing Procedure

The bootstrap, see for instance Efron and Tibshirani (1993), provides a feasible method for estimation of the small-sample distribution of a statistic. The basic principle is to approximate this distribution by a bootstrap distribution, which can be retrieved by simulation. In short, this is done by generating a large number of resamples, based on the original sample, and by computing the statistics of interest in each resample. The collection of bootstrap statistics, suitably ordered, then constitutes the bootstrap distribution.

2.1. The Bootstrap Test

The objective of a general (two-sided) test is to compute the p-value function

$$p(\hat{\tau}) = p(|\tau| \geq |\hat{\tau}| \,|\Psi_0, T) \tag{2.1}$$

where Ψ_0 is the data generating process (DGP) under the null hypothesis, and $\hat{\tau}$ is the realized value of a test statistic τ based on a sample of length T. Since Ψ_0 is unknown this p-value function has to be approximated, which is regularly done using asymptotic theory. For asymptotic theory to be valid it is required that $p(\hat{\tau})$ should not depend on Ψ_0 and T, which is usually not true in small samples. An alternative to an asymptotic solution is to estimate the nite-sample DGP by the bootstrap DGP $\hat{\Psi}_0$, that is to use a bootstrap test. According to Davidson and MacKinnon (1996a), a bootstrap test is understood as a test for which the signi cance level is calculated using a bootstrap procedure.

If R bootstrap samples, each of size T, are generated in accordance with $\hat{\Psi}_0$ and their respective test statistics τ_r^* are calculated using the same test statistic τ as above, the estimated bootstrap p-value function is de ned by the quantity

$$p^*(\hat{\tau}) = R^{-1} \sum_{r=1}^{R} I(|\tau_r^*| \geq |\hat{\tau}|), \tag{2.2}$$

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

The bootstrap testing procedure is a general tool and can be applied to all tests that allow for the implementation of the null-hypothesis in the bootstrap. Davidson and MacKinnon conclude that the size distortion of a bootstrap test is of the order $T^{-1/2}$ smaller than that of the corresponding asymptotic test. A further re nement of the 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 paper employs the bootstrap technique on fractional integration tests. In order to handle non-normal or conditionally heteroskedastic data, we re ne the bootstrap testing procedure of Andersson and Gredenhoff (1997) to include these cases. The bootstrap tests are based on the periodogram regression test of Geweke and Porter-Hudak, GPH, (1983), the modi ed rescaled range, MRR, test (Lo, 1991) and the Lagrange multiplier REG test of Agiakloglou and Newbold (1993).[1]

A fractionally integrated autoregressive moving average (ARFIMA) time series process is described by

$$\phi(B)(1-B)^d x_t = \theta(B) a_t, \qquad t = 1, ..., T \tag{2.3}$$

where the roots of $\phi(B)$ and $\theta(B)$ have all roots outside the unit circle and a_t is *iid* with mean zero and variance $\sigma_a^2 < \infty$. The differencing parameter d is allowed to take any real number, but if d is restricted to the set of integers the speci cation (2.3) reduces to an ARIMA process. The sample autocorrelation function of a long-memory process may be approximated by a fractionally integrated model, hence testing for long-memory can be done by a test on d. Such tests are applied to stationary and invertible series and $d = 0$ is thus a natural null-hypothesis.[2]

When testing for fractional integration, the DGP Ψ_0 is characterized by an unknown ARMA(p,q) speci cation. Since the null model, and consequently Ψ_0, is unknown, the estimated (bootstrap) DGP $\hat{\Psi}_0$ is used to create the bootstrap samples.

2.2. Construction of the Bootstrap Samples

The original non-parametric bootstrap of Efron (1979), designed for *iid* observations, usually fails for dependent observations, e.g. time series, since the order of the observations is affected. Dependencies in data can be maintained in the bootstrap resample by using a model-based bootstrap, which is the natural way to proceed in our case since a well-de ned model forms the null-hypothesis. A model free procedure, such as a moving blocks bootstrap or a spectral resampling

[1] These tests are brie y described in *Appendix* A.

[2] Stationarity and invertibility require that $d < |1/2|$. The ARFIMA model is presented in greater detail by Granger and Joyeux (1980) and Hosking (1981).

scheme, may also preserve dependencies. However, model free approaches deviate from the bootstrap testing idea of Davidson and MacKinnon (1996a, b), in the sense that the resemblance between the bootstrap samples and the original sample is sacrificed. This is due to the implementation of the null-hypothesis, which in this situation is done by filtering the series through the long-memory filter $(1-B)^{\hat{d}}$, where $\hat{d}$ is an estimate of the differencing parameter. A further drawback is that the bootstrap test would then in general be sensitive to the estimate of d.

For the bootstrap fractional integration tests we use the resampling model,

$$\left(1-\hat{\phi}_0-\hat{\phi}_1 B-\ldots-\hat{\phi}_{\hat{p}} B^{\hat{p}}\right) x_t=\hat{a}_t, \tag{2.4}$$

which clearly obeys the null-hypothesis and can be regarded as the estimated AR representation of the process. The autoregressive order $\hat{p}$ is selected from the values $(0, 1\ldots, 5)$ for the size evaluation and up to 25 for the power, by the Bayesian information criterion (BIC) of Schwartz (1978), and the parameters are estimated by ordinary least squares (OLS). The use of the BIC is motivated by comparisons, not reported in the paper, with the AIC of Akaike (1974). Furthermore Andersson and Gredenhoff (1997) use the AR approximation as well as an ARMA resampling model, and find that the former performs better.

The bootstrap samples $\mathbf{x}_r^*$, $r=1,...,R$, are created recursively by the equation

$$x_{r,t}^*=\hat{\phi}(B)^{-1} a_t^*, \tag{2.5}$$

where $\hat{\phi}(B)$ is the polynomial of (2.4) and a_t^* are the bootstrap residuals. In this study the number of bootstrap replicates is $R=1,000$.

Four resampling algorithms are utilized to generate the bootstrap residuals a_t^*. The *first* algorithm, b_1, makes use of a normality assumption for the disturbances a_t in (2.3), and is denoted the simple parametric bootstrap. In this resampling the residuals a_t^* are independent draws from a normal distribution with mean zero and variance $s_{\hat{a}}^2$.

A *second* similar but non-parametric resampling scheme (denoted b_2) does not impose distributional assumptions but is directly based on the estimated residuals $\hat{a}_t$. The bootstrap residuals are drawn, with replacement, from the recentered and degrees of freedom corrected residual vector. One typical bootstrap residual is constructed as

$$a_t^*=\sqrt{\frac{T}{T-\hat{p}-1}}\times\hat{a}_s,$$

where s is $U(\hat{p}+1, T)$ distributed.

The *third* and *fourth* resampling algorithms are constructed to preserve ARCH(1) dependence in the residuals. ARCH is introduced to the autoregression, $\phi(B) x_t = a_t$, by the equation $a_t = \sqrt{\omega_t}\varepsilon_t$, where the conditional variance is given by $\omega_t = \beta_0 + \beta_1 a_{t-1}^2$. The assumed normality of ε_t allows joint estimation of the parameters through maximization of the log-likelihood function

$$l(\phi_0, \ldots, \phi_p, \beta_0, \beta_1 | \mathbf{x}) = -\frac{1}{2T}\sum_{t=1}^{T}\left(\log \omega_t + \frac{a_t^2}{\omega_t}\right).$$

For the optimization, we use the numerical method of Davidon, Fletcher and Powell, see for instance Press *et al.* (1992). The resamplings are based on a parametric or a non-parametric algorithm, similar to those above. In the parametric case (denoted b_3), a residual series $\tilde{\varepsilon}_t$ is created by independent draws from a $N(0, s_{\hat{\varepsilon}}^2)$ distribution. For the non-parametric (b_4) scheme the members of $\{\tilde{\varepsilon}_t\}$ are drawn from the degrees of freedom adjusted elements of $\{\hat{\varepsilon}_t\}$. The bootstrap residuals are then built by imposing the estimated conditional dependency, according to the equations

$$\tilde{\omega}_t = \hat{\beta}_0 + \hat{\beta}_1 a_{t-1}^{*2}$$

and

$$a_t^* = \tilde{\varepsilon}_t\sqrt{\tilde{\omega}_t}.$$

This implies that a_t^* has an unconditional variance of $\hat{\beta}_0 / \left(1 - \hat{\beta}_1\right)$.

3. The Monte Carlo Study

The Monte Carlo study involves 1000 replicates (series), where each series is tested for fractional integration using the original tests and the different bootstrap tests described in Section 2. The rejection frequencies of the non-fractional null-hypothesis, i.e. the empirical sizes, are evaluated and compared. The power of a bootstrap test is in general close to that of the size adjusted asymptotic test (see Davidson and MacKinnon 1996*b*). In particular, for the asymptotic tests in this study and normal processes, Andersson and Gredenhoff (1997) demonstrate this for a bootstrap test with the simple parametric resampling scheme.

To evaluate the size of the tests rst order autoregressions,

$$(1 - \phi B) x_t = a_t, \tag{3.1}$$

of length $T = 100$ are generated and the parameter ϕ is set equal to the values $\{0, 0.1, 0.5, 0.7, 0.9\}$. To reduce the initial-value effect, an additional 100 observations are generated. We construct the data in order to display three different characteristics: normality, non-normality (skewness and excess kurtosis) and ARCH errors. The characteristics are introduced via the disturbances a_t.

3.1. Normal Processes

The experiment examining the empirical size of the tests under normality is based on the process (3.1) where the disturbances $\{a_t\}$ are *iid* normally distributed with mean zero and variance equal to unity. Table 3.1 presents the sensitivity of the empirical size with respect to the investigated AR parameters for a nominal 5% level of signi cance. Signi cant differences from the nominal size are obtained when the rejection frequencies lie outside the 95% acceptance interval $(3.6, 6.4)$.

The estimated size of the original MRR test always differs signi cantly from the 5% nominal level. In particular, the MRR test is strongly conservative for large positive parameters. The GPH test is severely over-sized for highly short-term dependent series, which is explained by a biased periodogram regression estimate due to large positive AR roots, see Agiakloglou *et al.* (1993). Compared with the other original tests the REG test is well-sized; only one signi cant size-distortion can be found. A more detailed presentation of the tests is given in Andersson and Gredenhoff (1997).

The results suggest that the MRR and GPH bootstrap tests, regardless of resampling, give exact tests in the sense that the estimated sizes of the tests coincide with the nominal. The bootstrap REG test based on the simple parametric resampling is almost exact, whereas the non-parametric resampling (which does not incorporate the normality) produces notably large sizes for strongly dependent AR processes. The resamplings that account for the (non-existing) ARCH effects have resonable estimated sizes, however conservative for ϕ equal to 0.7 and 0.9.

3.2. Non-Normal Processes

In the non-normal case, the disturbances a_t are distributed with mean, variance, skewness and kurtosis equal to $0, 1, \gamma_s$ and γ_k respectively. The members of $\{a_t\}$ are generated by the transformation,

$$a_t = c_0 + c_1\alpha_t + c_2\alpha_t^2 + c_3\alpha_t^3 \qquad \alpha_t \sim N(0,1), \tag{3.2}$$

Table 3.1: Rejection percentage of the nominal 5 percent fractional integration test when the data follow an AR(1), of length 100, with normal errors.

Test		ϕ 0.0	0.1	0.5	0.7	0.9
MRR	o	**7.6**	**6.8**	**2.3**	**1.3**	**0.8**
	b_1	5.3	5.9	4.8	4.7	3.9
	b_2	4.5	4.2	4.9	5.0	4.4
	b_3	6.2	6.4	6.0	5.2	3.9
	b_4	6.0	6.4	5.5	5.3	4.3
GPH	o	4.9	4.9	**8.3**	**17.9**	**71.8**
	b_1	5.0	5.2	4.7	4.3	3.7
	b_2	5.0	5.1	4.6	4.2	4.0
	b_3	5.3	5.8	5.0	4.4	3.6
	b_4	5.4	5.5	5.0	4.4	3.6
REG	o	5.9	**6.9**	6.4	5.1	5.1
	b_1	6.0	5.1	4.7	4.2	4.7
	b_2	4.5	3.7	5.6	**7.9**	**19.2**
	b_3	5.0	4.6	3.7	**3.3**	**3.4**
	b_4	4.8	4.8	3.6	**3.5**	**3.5**

The number reported is the rejection percentage of the two-sided 5% test. Bold face denotes a signi cant deviation from the nominal size. Under the null-hypothesis of no fractional integration, the 95% acceptance interval of the rejection percentage equals (3.6, 6.4). o denotes the original test and $b_1 - b_4$ the bootstrap testing procedures described in Section 2.

proposed by Fleichmann (1978). In this situation, the sequence $\{a_t\}$ will have a distribution dependent upon the constants c_i, which can be solved for using a non-linear equation system speci ed as a function of selected skewness and kurtosis. γ_s and γ_k are chosen to generate series x_t with a skewness and kurtosis of 2 and 9 respectively.[3] The empirical size of the tests under non-normality is reported in Table 3.2.

The original MRR and GPH tests are robust to excess skewness and kurtosis, in the sense that the results are similar to the normal case. This does not imply that the tests are well-sized, since the distortions of the MRR test are slightly more

[3]The expressions for the determination of $c_i, i = 0, ..., 3$ and how the residual skewness and kurtosis depend on those of the time series process are given in *Appendix* B.

Table 3.2: Rejection percentage of the nominal 5 percent fractional integration test when the data follow an AR(1), of length 100, with non-normal errors.

		ϕ				
Test		0.0	0.1	0.5	0.7	0.9
MRR	o	**7.7**	6.0	**2.2**	**2.0**	**0.5**
	b_1	5.0	5.5	5.3	5.4	4.8
	b_2	4.7	4.8	4.4	4.8	4.6
	b_3	4.9	6.0	5.5	4.2	4.2
	b_4	4.7	6.0	5.2	4.5	3.8
GPH	o	4.9	5.6	**8.0**	**16.3**	**72.2**
	b_1	4.2	4.2	3.7	4.5	3.9
	b_2	4.4	4.3	5.0	4.7	3.6
	b_3	4.2	5.4	5.3	5.0	3.7
	b_4	4.4	5.4	5.1	5.2	4.1
REG	o	5.5	**7.3**	**7.7**	**6.6**	**8.7**
	b_1	4.9	5.0	3.8	4.7	**2.4**
	b_2	5.1	3.9	4.5	**6.9**	**13.8**
	b_3	3.9	5.8	4.5	**3.1**	**1.5**
	b_4	3.8	4.7	4.2	**1.8**	**1.1**

See note to Table 3.1. The skewness and kurtosis of the disturbances are selected in order to give a skewness of 2.0 and a kurtosis of 9.0, for all processes.

articulated and the GPH test is still severely over-sized for large parameters. In contrast to the other tests, the original REG test is sensitive when the data do not ful ll the normality assumption. The difference between the empirical and nominal size is in general signi cant.

The estimated sizes of all bootstrap MRR and GPH tests never differ signi -cantly from the nominal 5%. The bootstrap REG tests behave as in the normal case. That is, the parametric b_1 works well, the non-parametric b_2 is over-sized for large parameters and b_3 and b_4 are conservative for the same parameters.

3.3. Processes with ARCH Errors

For the nal set of processes the assumption of identically and independently distributed errors is relaxed. Instead we consider the effect of heteroskedasticity of ARCH(1) type, which implies that the disturbances are conditionally distributed as $a_{t|t-1} \sim N(0, \omega_t)$, where $\omega_t = 1 - \beta + \beta a_{t-1}^2$ and $\beta < 1$. The parametrization implies that the unconditional variance of a_t equals unity, and the parameter

β is selected as 0.5 and 0.9. The 0.9 parameter imposes a strong conditional dependence in the disturbances, but the fourth moment of the disturbance process does not exist. As a complement, the weaker ARCH dependence of $\beta = 0.5$ is also investigated.

Results in Table 3.3 show that the MRR test is quite robust also against conditional heteroskedasticity. However, compared with the case of uncorrelated data, cf. $\beta = 0$, the test tends to be more conservative as the ARCH parameter increases. The GPH test is unaffected by ARCH in the disturbances. In short, these tests have the same size problem as with uncorrelated disturbances. On the other hand, the usually well-sized REG test is very sensitive to ARCH effects and exhibits a seriously distorted size for $\beta = 0.5$ and in particular for $\beta = 0.9$.

The robustness of the MRR and GPH tests against ARCH effects can be detected in the bootstrap tests. As a result all bootstrap MRR and GPH tests are exact for all generated combinations of β and ϕ.

The disappointing size of the original REG test is partly inherited by b_1 and b_2. Furthermore, the increasing pattern with the AR parameter for b_2 is still present. However, the size distortions of b_1 and b_2 are smaller for the lower value of β. The REG test, overall, requires that the resampling scheme allows for ARCH effects. This is exactly what bootstraps b_3 and b_4 do, and despite a few conservative values these bootstraps are not only superior to the original test, but also much better than b_1 and b_2.

3.4. Size Comparisons and Power

Table 3.4 supplies an overview of the tests that exhibit the best size properties, judged by the number of signi cant results based on the 95% acceptance region, for the respective processes. All bootstrap MRR and GPH tests work well and have equivalent size properties, for all processes investigated. The simple non-parametric bootstrap REG test is badly sized when the generated process has an AR parameter close to the unit circle, regardless of the characteristics of the disturbance process. Otherwise, all REG tests, even the original, perform well under normality, whereas for non-normality the simple parametric bootstrap is best. When ARCH errors are introduced, the bootstraps that account for the heteroskedasticity clearly adjust the size of the REG test better. Since these resamplings also work for uncorrelated errors, b_3 and b_4 exhibit the best REG performance overall.

A bootstrap MRR or GPH test is shown to be exact, and a well-designed

Table 3.3: Rejection percentage of the nominal 5 percent fractional integration test when the data follow an AR(1), of length 100, with ARCH errors.

Test		ϕ = 0.0	0.1	0.5	0.7	0.9
		$\beta = 0.5$				
MRR	o	5.5	4.0	**2.4**	**1.8**	**0.6**
	b_1	4.4	4.8	5.1	3.9	3.8
	b_2	4.3	4.8	4.7	4.3	4.1
	b_3	5.1	5.1	4.8	5.4	3.9
	b_4	4.7	4.9	4.1	5.2	3.8
GPH	o	4.5	4.3	5.7	**17.6**	**71.6**
	b_1	4.0	4.1	4.0	3.9	3.6
	b_2	3.8	4.2	4.0	3.6	4.1
	b_3	3.8	3.5	4.0	4.5	4.2
	b_4	4.0	3.8	3.9	4.8	4.5
REG	o	**9.8**	**9.7**	**9.0**	**7.4**	**10.4**
	b_1	**7.6**	**8.5**	5.4	4.6	**7.3**
	b_2	**8.7**	4.6	6.6	4.6	**25.2**
	b_3	4.3	3.8	3.9	3.9	3.7
	b_4	4.2	4.0	3.9	3.7	4.0
		$\beta = 0.9$				
MRR	o	**3.4**	**3.1**	**1.4**	**0.8**	**0.7**
	b_1	3.7	3.8	4.0	4.0	5.3
	b_2	4.0	3.6	4.1	3.7	3.7
	b_3	5.2	5.4	4.8	4.3	4.2
	b_4	5.1	5.3	4.8	4.0	4.3
GPH	o	4.7	5.0	**6.5**	**16.8**	**70.9**
	b_1	4.7	5.0	4.1	4.2	4.1
	b_2	5.4	5.3	4.7	3.8	3.8
	b_3	4.0	4.7	4.2	6.2	4.7
	b_4	4.0	4.7	4.4	6.4	4.6
REG	o	**28.6**	**29.3**	**29.5**	**29.2**	**33.4**
	b_1	**7.9**	**8.4**	**7.3**	**7.1**	**7.3**
	b_2	**9.2**	**6.8**	**8.8**	**14.4**	**24.9**
	b_3	**3.1**	3.8	**3.5**	3.8	4.3
	b_4	3.6	3.8	**3.3**	4.1	4.4

See note to Table 3.1. The error processes follow an ARCH process with parameter β.

Table 3.4: Best test based on size properties.

	Normal	*Non − normal*	*ARCH*	*Overall*
MRR	$b_1 - b_4$	$b_1 - b_4$	$b_1 - b_4$	$b_1 - b_4$
GPH	$b_1 - b_4$	$b_1 - b_4$	$b_1 - b_4$	$b_1 - b_4$
REG	$b_1, (o, b_3, b_4)$	$b_1, (b_3, b_4)$	$b_4\,(b_3)$	b_3, b_4

The Table reports the best test based on the smallest number of signi cant size distortions. The test within brackets is regarded as almost equivalent to the ones without.

bootstrap test based on the REG test nearly exact. Furthermore, the stylized DGPs of this study are quite well-behaved, whereas in empirical situations they are not. Consequently, the asymptotic tests are likely to be more distorted and the gain from a bootstrap test to be even larger.

Davidson and MacKinnon (1996*b*) show that the power of a bootstrap test, based on a pivotal statistic, is generally close to the size-adjusted asymptotic test. Table 3.5 presents the power of the tests for fractionally integrated white noise, $(1 - B)^d x_t = a_t$, where the members of $\{a_t\}$ have a normal, non-normal[4] or heteroskedastic distribution. Only the parametric bootstraps are reported, because of the similar power properties of the corresponding non-parametric resamplings. However, combined with the REG test the simple parametric bootstrap b_1 exhibits, at positive differencing parameters, notably better power properties than the simple non-parametric resampling b_2.

The power of the MRR and GPH tests are preserved by all bootstrap procedures, except for processes with $d = 0.45$. In this case too many autocorrelations are included in the variance correction term of the MRR test, resulting in a negatively biased estimate of the fractional differencing parameter which lowers the power of the original test. This phenomenon is not experienced by the bootstrap tests, which have well-behaved power curves. For the GPH test a large differencing parameter results in a rich parameter structure of the resampling model, which implies that the resample periodograms resemble the periodogram of the highly persistent original process. Thus, the bootstrap GPH test has difficulties in distinguishing fractional processes from AR speci cations. The power of the REG test is preserved by the simple parametric bootstrap, whereas the ARCH

[4]The skewness and kurtosis of the residuals as functions of the time series moments are given in *Appendix* B.

Table 3.5: Rejection percentage of the nominal 5 percent fractional integration test when the data follow fractional noise of length 100.

Test		d						
		-0.45	-0.25	-0.05	0.0	0.05	0.25	0.45
		Normal processes						
MRR	o	14.6	11.3	3.7	5.0	5.8	14.7	4.0
	b_1	16.5	13.6	6.1	5.3	6.9	14.3	18.6
	b_3	13.6	12.4	4.4	6.2	6.6	11.9	12.3
GPH	o	21.8	8.3	3.9	5.0	4.4	17.0	40.9
	b_1	21.0	11.1	5.3	5.0	5.3	16.5	11.8
	b_3	17.6	9.3	4.8	5.3	5.3	7.1	12.2
REG	o	43.9	27.8	3.8	5.0	10.8	35.3	22.5
	b_1	41.0	37.6	6.7	6.0	8.0	25.9	19.6
	b_3	19.9	23.9	3.1	5.0	5.6	16.1	14.1
		Non-normal processes						
MRR	o	6.9	8.7	5.0	5.0	6.8	15.8	6.8
	b_1	10.8	8.2	4.7	5.0	5.4	11.8	18.6
	b_3	8.0	8.7	5.3	4.9	6.0	11.6	18.4
GPH	o	25.9	10.8	6.1	5.0	7.0	15.4	40.8
	b_1	20.4	10.5	4.5	4.2	5.3	8.4	10.0
	b_3	19.7	9.2	5.6	4.2	6.6	7.9	8.9
REG	o	42.9	32.6	4.2	5.0	10.8	37.9	22.5
	b_1	42.2	37.7	5.5	4.9	6.1	27.1	15.8
	b_3	23.1	30.2	4.2	3.9	5.3	20.4	16.7
		ARCH processes						
MRR	o	10.8	8.0	4.9	5.0	6.0	19.1	17.1
	b_1	8.3	5.7	3.1	3.7	4.8	11.2	20.3
	b_3	9.4	7.3	4.3	5.2	5.2	13.0	13.7
GPH	o	21.1	10.1	5.9	5.0	6.0	14.8	40.7
	b_1	19.4	11.4	5.6	4.7	4.9	8.9	12.7
	b_3	17.0	9.2	5.5	4.0	5.8	7.7	10.6
REG	o	31.7	16.6	4.5	5.0	7.0	21.2	10.9
	b_1	35.9	25.1	9.5	7.9	9.5	21.6	18.7
	b_3	11.6	9.3	2.8	3.1	2.4	9.3	7.5

The number reported is the rejection percentage of the two-sided 5% test. o denotes the original test, and b_1 and b_3 the bootstrap testing procedures described in Section 2.

resamplings have a lower power throughout.

On the basis of the estimated power, two major situations are detected. If we cannot rule out ARCH effects in the disturbances, the highest power is given by a simple bootstrap MRR or GPH test. However, if there are no ARCH effects (in theory or data) then the simple parametric REG test clearly outperforms all other testing procedures.

4. Conclusions

The concept of bootstrap testing for fractional integration works extraordinarily well. If the signi cance level is calculated by a bootstrap procedure an exact test is almost always the result. However, the choice of resampling algorithm may affect the degree of size adjustment. For instance, if the original test is sensitive to distributional assumptions, in particular ARCH effects, this should be accounted for when specifying the resampling model. However, if the test is robust to ARCH errors, the choice of resampling is not very important for the size properties of that test.

Since economic and nancial data are often heteroskedastic we recommend the use of the parametric ARCH resampling scheme for the REG test. However, if prior information suggests that the investigated series does not have ARCH effects, the simple parametric bootstrap has equivalent size properties and a higher power, and should thus be used.

The MRR and GPH tests, which are robust to deviations from the iid normality of the disturbances, have nice size properties for all bootstrap procedures. Due to the simplicity and the slightly higher power of the simple algorithms, they are preferred when bootstrapping the MRR and GPH tests.

The main conclusions are that the bootstrap tests are remarkably well-sized (whereas the originals are not) and robust to non-normalities and ARCH effects, and that reliable testing for fractional integration in many cases requires a bootstrap test.

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.

AKAIKE, H. (1974) A New Look at the Statistical Model Identi cation. *IEEE Transactions on Automatic Control* 19(6), 716-723

ANDERSSON, M.K. and M.P. GREDENHOFF (1997) Bootstrap Testing for Fractional Integration. *Working Paper Series in Economics and Finance*, No. 188, Dept. of Economic Statistics, Stockholm School of Economics, Sweden.

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 (1996*a*) 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 (1996*b*) The Power of Bootstrap Tests. *Queens Institute for Economic Research Discussion Paper No.* 937.

DING, Z., C. W.J. GRANGER, and R. F. ENGLE (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.

EFRON, B. and R. J. TIBSHIRANI (1993) *An Introduction to the Bootstrap*, New York, Chapman and Hall.

FLEICHMANN, A. I. (1978) A Method for Simulating Non-Normal Distributions. *Psychometrika* 43 (4), 521-532.

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 Z. DING (1995) Stylized Facts on the Temporal and Distributional Properties of Daily Data from Speculative Markets, *Unpublished manuscript,* Department of Economics, University of California, San Diego.

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. *Biometrika* 68, 165-176.

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

PRESS, H., TEUKOLISKY, S., VETTERLING, W. AND B. FLANNERY (1992) *Numerical Recipes in Fortran: The Art of Scienti c Computing,* second edition. Cambridge University Press, New York, USA.

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

Appendix

A. Tests for Fractional integration

Consider the regression equation

$$\ln\{I_x(\omega_j)\} = \alpha - d\ln\left\{4\sin^2(\omega_j/2) + v_j\right\}, \tag{A.1}$$

where $I_x(\omega_j)$ is the periodogram at the harmonic frequencies $\omega_j = 2\pi j/T$, and $j = 1, ..., g(T) = T^{1/2}$. The ordinary least squares (OLS) estimator of d is then consistent and the distribution of $\left(\hat{d}_{OLS} - d\right)/SE\left(\hat{d}_{OLS}\right)$ is asymptotically normal. This is the periodogram regression estimation/testing procedure of Geweke and Porter-Hudak (1983).

Lo (1991) proposes a modi ed rescaled range (MRR) statistic when testing for fractional integration. This modi ed rescaled range is de ned by the ratio

$$\tilde{Q}_T = \frac{R_T}{\hat{\sigma}_T(k)}, \tag{A.2}$$

where the range and the standard error are calculated by

$$R_T = \max_{0<i\leq T}\sum_{t=1}^{i}(x_t - \overline{x}) - \min_{0<i\leq T}\sum_{t=1}^{i}(x_t - \overline{x}) \tag{A.3}$$

$$\hat{\sigma}_T^2(k) = \hat{\sigma}^2 + 2\sum_{j=1}^{k}\sum_{i=j+1}^{T}\left(1 - \frac{j}{k+1}\right)(x_i - \overline{x})(x_{i-j} - \overline{x}). \tag{A.4}$$

The truncation lag, k, is set, 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 are given by Lo (1991).

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

$$\hat{a}_t = \sum_{i=1}^{p}\beta_i W_{t-i} + \sum_{j=1}^{q}\gamma_j Z_{t-j} + \delta K_m + u_t, \tag{A.5}$$

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 residuals and MA polynomial under the null and m is a prespeci ed truncation lag. The equation (A.5) is tted by OLS over the time period $t = m+1, ..., T$ and the usual t-statistic for the null hypothesis $\delta = 0$ follows an asymptotic $N(0,1)$ distribution.

B. Generation of Non-normal Data

B.1. The Skewness and Kurtosis Relationship

Under the assumptions in Section 3 the skewness and (raw) kurtosis for the disturbance process are given by

$$\gamma_s = \Gamma_s \frac{(1-\phi^3)}{(1-\phi^2)^{3/2}}$$

and

$$\gamma_k = \frac{\Gamma_k (\phi^2+1) - 6\phi^2}{1-\phi^2},$$

where Γ_1 and Γ_2 are the corresponding moments of the AR(1) process.

In the fractionally integrated case the disturbance skewness and kurtosis are given as

$$\gamma_s = \Gamma_s \frac{\sum_{i=0}^{\infty} \delta_i^3}{var^{3/2}(x)}$$

$$\gamma_k = \left\{ (\Gamma_k - 3) \sum_{i=0}^{\infty} \delta_i^4 + 3 \sum_{j=0}^{\infty} \sum_{k \neq j} \delta_j^2 \delta_k^2 \right\} / var^2(x),$$

where δ_i is the ith weight in the moving average representation,

$$x_t = \sum_{i=0}^{\infty} \delta_i a_{t-i},$$

for the fractionally integrated process. The weights are given by

$$\begin{aligned}
\delta_0 &= 1 \\
\delta_1 &= d \\
\delta_i &= \frac{1}{i}\delta_{i-1}\left(i-1-d\right), \qquad \text{for } i > 1.
\end{aligned}$$

B.2. The Fleichmann Algorithm

The constants in (3.2) are given as the solutions to the following system of equations,

$$\begin{aligned}
c_0 &= -c_2 \\
\gamma_k &= 24\left[c_1c_3 + c_2^2\left(1 + c_1^2 + 28c_1c_3\right) + c_3^2\left(12 + 48c_1c_3 + 141c_2^2 + 225c_3^2\right)\right] \\
c_2 &= \frac{\gamma_s}{2\left(c_1^2 + 24c_1c_3 + 105c_3^2 + 2\right)}
\end{aligned}$$

and

$$2 = 2c_1^2 + 12c_1c_3 + \frac{\gamma_s^2}{\left(c_1^2 + 24c_1c_3 + 105c_3^2 + 2\right)^2} + 30c_3^2.$$