TESTS FOR FRACTIONAL INTEGRATION: A MONTE CARLO INVESTIGATION

By Yin-Wong Cheung

University of California at Santa Cruz

First version received June 1991

Abstract. The performance of the Geweke–Porter-Hudak (GPH) test, the modified rescaled range (MRR) test and two Lagrange multiplier (LM) type tests for fractional integration in small samples is examined using Monte Carlo methods. Both the GPH and MRR tests are found to be robust to moderate autoregressive moving-average components, autoregressive conditional heteroskedasticity effects and shifts in the variance. However, these two tests are sensitive to large autoregressive moving-average components and shifts in the mean. It is also found that the LM tests are sensitive to deviations from the null hypothesis. As an illustration, the GPH test is applied to two economic data series.

Keywords. Tests for fractional integration; Monte Carlo experiment; ARMA; ARCH; shifts in parameters.

1. INTRODUCTION

Recently, autoregressive fractionally integrated moving-average (ARFIMA) processes have received considerable attention in the literature. Granger and Joyeux (1980) and Hosking (1981) are credited as the two seminal studies on ARFIMA processes. Recent theoretical and empirical studies on this topic include those of Diebold and Rudebusch (1989, 1991a), Fox and Taqqu (1986), Geweke and Porter-Hudak (1983), Li and McLeod (1986), Lo (1991), Porter-Hudak (1990), Robinson (1991), Shea (1991), Sowell (1990a, b), and Yajima (1988, 1989).

ARFIMA processes generalize standard linear ARIMA(p, d, q) models by allowing the degree of integration d to assume non-integer values. This generalization provides a more flexible framework to study time series data. In particular, the class of fractional processes can be used to model data dependence that is stronger than allowed in stationary ARMA processes and weaker than implied by unit root processes. This ability to describe strong dependence without resorting to non-stationary unit root processes attracts, for example, researchers who are interested in studying persistence in economic and financial time series.

One common test for fractional integration is the frequency-domain regression-based procedure introduced by Geweke and Porter-Hudak (1983). Lo (1991) proposes another test which is modified from the classical rescaled range statistic (Hurst, 1951; Mandelbrot, 1972). In addition, Robinson (1991) develops a Lagrange multiplier (LM) type test for fractional integration. Most applied work using these techniques relies on asymptotic results to make small-sample inferences, and there is limited Monte Carlo evidence regarding the robustness of these tests to different data-generating mechanisms.

In this paper we examine the finite-sample properties of these three tests by analyzing their sensitivity to different data-generating mechanisms via Monte Carlo methods. Critical values used in the Monte Carlo experiment are based on the asymptotic distributions of these tests. Section 2 briefly describes the tests for fractional integration. Monte Carlo results based on simulated ARMA processes, autoregressive conditional heteroskedastic (ARCH) processes, fractional processes and processes with shifts in mean or variance are presented in Section 3. Two examples drawn from economic data are provided in Section 4. Concluding remarks are offered in Section 5.

2. TESTS FOR FRACTIONAL INTEGRATION

$\{X_t\}$ is said to be generated from an ARFIMA(p, d, q) process if

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t, \tag{1}$$

where $\varepsilon_t \sim \mathrm{IID}(0, \sigma^2)$, B is the backward-shift operator, $\Phi(B)$ is the AR polynomial, $\Theta(B)$ is the MA polynomial, and the fractional integrating operator $(1-B)^d$ is defined by $(1-B)^d = \sum_{k=0}^{\infty}\Gamma(k-d)B^k/\Gamma(k+1)\Gamma(-d)$ with $\Gamma(\cdot)$ being the gamma function.

X_t is both stationary and invertible if the roots of $\Phi(B)$ and $\Theta(B)$ are outside the unit circle and $d < |0.5|$. When $d=0$, an ARFIMA process reduces to an ARMA process. In contrast with the spectral density of a unit root process which is approximated by $S(\omega) \propto \omega^{-2}$ as $\omega \to 0$, the spectral density of an ARFIMA process behaves as $S(\omega) \propto \omega^{-2d}$ as $\omega \to 0$ (Hosking, 1981). A wide range of low frequency behavior can thus be modeled when d is not restricted to the integer domain. Further, the existence of fractional integration can be determined by the significance of the sample d parameter.

Three tests for fractional integration are considered. The first test is based on a modified rescaled range (MRR) statistic derived by Lo (1991). The MRR statistic Q_T is defined by

$$Q_T = R/\sigma_T(q) \tag{2}$$

where

$$R = \max_{0<i\leq T} \sum_{t=1}^{i} (X_t - \bar{X}) - \min_{0<i\leq T} \sum_{t=1}^{i} (X_t - \bar{X}) \tag{3}$$

$$\sigma_T^2(q) = \sigma^2 + 2\sum_{j=1}^{q} \sum_{i=j+1}^{T} \left(1 - \frac{j}{q}\right)(X_i - \bar{X})(X_{i-j} - \bar{X}) \tag{4}$$

$$\sigma^2 = \sum_{t=1}^{T} \frac{(X_t - \bar{X})^2}{T}, \tag{5}$$

with q being set equal to the integer part of $(3T/2)^{1/3}\{2\hat{\rho}/(1-\hat{\rho}^2)\}^{2/3}$; $\hat{\rho}$ is the sample first-order autocorrelation coefficient and $\bar{X}$ is the sample mean. R measures the range of cumulative departures from $\bar{X}$. $\sigma_T^2(q)$ is a heteroskedasticity- and autocorrelation-consistent variance estimator. Extreme values of Q_T are regarded as signs of fractional integration. Critical values of the MRR test are given in Lo (1991).

Geweke and Porter-Hudak (hereafter GPH), in their 1983 article, proposed a semi-nonparametric procedure to test for fractional integration (also see Yajima, 1989). The procedure is motivated by the log spectral density of the ARFIMA process, and amounts to estimating the least squares regression

$$\ln\{I(\omega_j)\} = c - d\ln\{4\sin^2(\omega_j/2)\} + \eta_j \qquad (j = 1, \ldots, n) \tag{6}$$

where $I(\omega_j)$ is the periodogram of $\{X_t\}$ at frequency ω_j, $\omega_j = 2\pi j/T$ $(j = 1, \ldots, T-1)$ and $n = g(T) \ll T$.

There is evidence of fraction integration if $\hat{d}$, the least squares estimate of d, is significantly different from zero. With a proper choice of n, the asymptotic distribution of $\hat{d}$ depends on neither the order of the ARMA part nor the distribution of the error term. It is suggested to set $n = T^{0.5}$ and use the known variance of η_j, $\pi^2/6$, to compute the sample variance of $\hat{d}$.

The third test for fractional integration is an LM type test developed by Robinson (1991). Specifically, we consider the null hypothesis of a white noise process against the fractional white noise, ARFIMA$(0, d, 0)$, alternative. Under these specifications two variants of the LM test are derived:

$$\lambda_1 = T^{1/2}\left(\sum_{j=1}^{T-1} j^{-2}\right)^{-1/2} \sum_{j=1}^{T-1} C_j(C_0 j)^{-1} \tag{7}$$

and

$$\lambda_2 = T^{1/2}\left(\sum_{j=1}^{T-1} K_j j^{-2}\right)^{-1/2} \sum_{j=1}^{T-1} C_j j^{-1} \tag{8}$$

where C_j and K_j are defined by

$$C_j = T^{-1} \sum_{t=1}^{T-j} (X_t - \bar{X})(X_{t+j} - \bar{X}) \tag{9}$$

and

$$K_j = T^{-1} \sum_{t=1}^{T-j} (X_t - \bar{X})^2(X_{t+j} - \bar{X})^2. \tag{10}$$

Under the null hypothesis, both λ_1 and λ_2 have an asymptotic standard normal distribution. In addition λ_2 is robust to conditional heteroskedasticity.

3. MONTE CARLO RESULTS

In the Monte Carlo experiment, we consider sample sizes T of 100, 300 and 500. In each case $T + 50$ observations are generated and the last T observations are used to reduce the effect of initial values. Each data series is tested for fractional integration using the tests described in Section 2. The simulation results are based on 1000 replications. For the GPH test, n is set equal to $T^{0.5}$. Observed rejection percentages at the two-sided nominal 5% significance level are reported. Note that under the null hypothesis of no fractional integration the 95% confidence interval of the rejection percentage is equal to 5% $\pm$ 1.4%.

3.1. *ARMA and fractional processes*

The Monte Carlo experiment examining the effect of different ARMA and ARCH specifications is based on the following simulated processes:

$$X_t = \phi X_{t-1} + \varepsilon_t \tag{11}$$

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1} \tag{12}$$

and

$$X_t = u_t \qquad u_{t|t-1} \sim \mathrm{N}(0, h_t)$$
$$h_t = \alpha_0 + \alpha u_{t-1}^2, \tag{13}$$

where $\varepsilon_t \sim \mathrm{IIDN}(0, 1)$. The AR parameter ϕ and the MA parameter θ are set equal to ± 0.1, ± 0.5, ± 0.7 and ± 0.9. The ARCH parameter α is set equal to 0.1, 0.3, 0.5, 0.7 and 0.9. α_0 is set equal to 1.

The sensitivity to AR components is reported in Table I. The simulation result suggests that the MRR test is conservative in the sense that it tends to reject the null less frequently than the nominal 5% significance level. The rejection frequency is quite robust to changes in the AR parameter. On the other hand, the GPH test is sensitive to large AR parameters. When ϕ is in the range of 0.7–0.9, the GPH test rejects the null hypothesis of no fractional integration much too often (recall that the 95% confidence interval is 5% $\pm$ 1.4%). The over-rejection rate decreases rather slowly with sample size. With $\phi = 0.9$ and $T = 500$, the GPH test still rejects the null hypothesis 622 times out of 1000 trials. Moreover, when we examine the breakdown of rejection frequencies (not shown), we find that large AR parameters bias the GPH test in favor of $d > 0$ alternatives. For the λ_1 and λ_2 tests the rejection rate is significantly larger than the nominal 5% level even in the presence of weak serial correlation. It is noted that a positive (negative) AR parameter leads to large (small) λ_1 and λ_2 estimates, and hence the rejection of the null.

Table II presents the effect of MA components on the estimated size of the tests. Rejection frequencies of both the MRR and GPH tests are significantly larger than the nominal significance level when the MA parameter is near

TABLE I

REJECTION PERCENTAGE OF THE NOMINAL 5% FRACTIONAL INTEGRATION TEST WHEN THE DATA FOLLOW AN AR(1) PROCESS

	MRR	GPH	λ_1	λ_2
$T = 100$				
$\phi = -0.9$	3.6	5.0	100.0	100.0
$\phi = -0.7$	0.2	3.8	100.0	100.0
$\phi = -0.5$	0.9	5.0	96.8	95.6
$\phi = -0.1$	7.4	4.4	7.1	7.0
$\phi = 0.1$	5.1	4.9	14.2	14.8
$\phi = 0.5$	2.0	6.4	98.9	98.8
$\phi = 0.7$	1.6	16.6	100.0	100.0
$\phi = 0.9$	0.9	71.7	100.0	100.0
$T = 300$				
$\phi = -0.9$	0.0	4.0	100.0	100.0
$\phi = -0.7$	1.0	5.9	100.0	100.0
$\phi = -0.5$	2.4	5.1	100.0	100.0
$\phi = -0.1$	5.8	5.9	22.1	22.3
$\phi = 0.1$	6.1	5.2	29.5	29.7
$\phi = 0.5$	3.2	4.9	100.0	100.0
$\phi = 0.7$	2.8	8.4	100.0	100.0
$\phi = 0.9$	1.6	69.5	100.0	100.0
$T = 500$				
$\phi = -0.9$	0.0	5.9	100.0	100.0
$\phi = -0.7$	0.9	5.1	100.0	100.0
$\phi = -0.5$	2.8	4.3	100.0	100.0
$\phi = -0.1$	6.8	4.1	39.4	39.5
$\phi = 0.1$	5.9	5.2	43.0	42.8
$\phi = 0.5$	4.1	6.4	100.0	100.0
$\phi = 0.7$	3.7	6.3	100.0	100.0
$\phi = 0.9$	1.4	62.2	100.0	100.0

MRR, modified rescaled range test; GPH, Geweke–Porter-Hudak test; λ_1, the LM test which is not robust to heteroskedasticity; λ_2, the LM test that is robust to heteroskedasticity. The number reported in the table is the rejection percentage of the two-sided 5% test. Under the null of no fractional integration, the 95% confidence interval of the rejection percentage is 5 ± 1.4. The data were constructed to follow $X_t = \phi X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, 1)$.

−0.9. It is interesting to note that, with $\theta = -0.9$, the rejection percentage of the MRR test increases as the sample size rises. In contrast with the case of AR components, the breakdown of rejection frequencies indicates that negative MA parameters bias both the GPH and MRR tests towards $d < 0$ alternatives. Once again, we observe that both the λ_1 and λ_2 tests are very sensitive to serial correlation in the data. For example, more than 5% of the λ_1 and λ_2 estimates are significantly larger (smaller) than zero when θ equals 0.1 (−0.1).

The effect of conditional heteroskedasticity on the tests is shown in Table III. Both the MRR and GPH tests are robust to ARCH effects. The only significant deviation from the nominal 5% size is the case of $\alpha_1 = 0.1$ and

TABLE II

REJECTION PERCENTAGE OF THE NOMINAL 5% FRACTIONAL INTEGRATION TEST WHEN THE DATA FOLLOW AN MA(1) PROCESS

	MRR	GPH	λ_1	λ_2
T = 100				
$\theta = -0.9$	13.9	50.3	100.0	100.0
$\theta = -0.7$	7.5	15.6	100.0	100.0
$\theta = -0.5$	6.4	5.1	99.5	99.7
$\theta = -0.1$	7.6	4.6	7.3	7.1
$\theta = 0.1$	5.2	5.0	13.0	13.7
$\theta = 0.5$	3.3	4.9	80.6	78.3
$\theta = 0.7$	2.7	5.7	90.3	86.7
$\theta = 0.9$	2.6	5.8	93.1	90.3
T = 300				
$\theta = -0.9$	55.7	49.9	100.0	100.0
$\theta = -0.7$	11.5	9.3	100.0	100.0
$\theta = -0.5$	5.8	5.2	100.0	100.0
$\theta = -0.1$	6.1	5.9	23.9	24.0
$\theta = 0.1$	6.5	5.2	26.2	26.2
$\theta = 0.5$	3.6	4.7	99.8	99.8
$\theta = 0.7$	4.3	5.5	100.0	100.0
$\theta = 0.9$	2.5	4.8	100.0	100.0
T = 500				
$\theta = -0.9$	80.4	51.2	100.0	100.0
$\theta = -0.7$	14.0	7.4	100.0	100.0
$\theta = -0.5$	5.2	3.7	100.0	100.0
$\theta = -0.1$	7.1	4.1	42.2	42.3
$\theta = 0.1$	5.9	6.2	38.4	39.0
$\theta = 0.5$	5.1	5.4	100.0	100.0
$\theta = 0.7$	3.8	5.3	100.0	100.0
$\theta = 0.9$	4.2	4.9	100.0	100.0

See the note to Table I. The data were constructed to follow $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$, $\varepsilon_t \sim N(0, 1)$.

$T = 100$. In contrast, the rejection percentage of the λ_1 test increases with both the ARCH parameter and the sample size. The rejection frequency is statistically larger than the nominal size when the sample size $T \geqslant 300$ or when there is a moderate ARCH effect, say $\alpha_1 \geqslant 0.3$. On the other hand, the estimated size of the λ_2 test, which is expected to be robust to ARCH effects, is very close to the theoretical 5% level.

3.2. *Fractional processes*

The data $\{X_t\}$ used to examine the estimated power of the tests against ARFIMA(0, d, 0) alternatives are constructed to follow

$$(1 - B)^d X_t = \varepsilon_t \tag{14}$$

and the fractional differencing parameter d is set equal to ± 0.05, ± 0.25 and ± 0.45. The data are generated according to Hosking (1984).

TABLE III

REJECTION PERCENTAGE OF THE NOMINAL 5% FRACTIONAL INTEGRATION TEST WHEN THE DATA FOLLOW AN ARCH PROCESS

	MRR	GPH	λ_1	λ_2
$T = 100$				
$\alpha = 0.1$	8.0	6.7	5.5	4.5
$\alpha = 0.3$	5.4	5.0	8.8	4.6
$\alpha = 0.5$	5.8	5.2	10.7	3.9
$\alpha = 0.7$	4.2	5.1	17.3	4.9
$\alpha = 0.9$	2.6	5.9	21.6	4.1
$T = 300$				
$\alpha = 0.1$	7.3	5.3	6.8	5.6
$\alpha = 0.3$	7.2	5.7	8.8	4.2
$\alpha = 0.5$	5.0	4.5	15.8	4.9
$\alpha = 0.7$	4.6	4.9	23.4	3.8
$\alpha = 0.9$	4.4	4.8	36.2	6.5
$T = 500$				
$\alpha = 0.1$	5.1	4.1	6.1	5.0
$\alpha = 0.3$	5.3	5.4	9.8	4.6
$\alpha = 0.5$	5.3	5.7	17.3	4.9
$\alpha = 0.7$	4.4	5.4	27.4	5.0
$\alpha = 0.9$	3.9	5.0	41.0	5.1

See the note to Table I. The data were constructed to follow $X_t = u_t$, $u_t \sim N(0, h_t)$, $h_t = \alpha_0 + \alpha u_{t-1}^2$.

Results in Table IV show that the MRR test detects negative differencing parameters better than positive differencing parameters. In general, the estimated power of the MRR test increases with either the absolute value of d or the sample size. An exception is that the power recorded for an ARFIMA(0, 0.45, 0) process is lower than that for an ARFIMA(0, 0.35, 0) process. A possible explanation is as follows. The truncation lag q in the MRR test controls the amount of serial correlation to be discounted in computing the statistic. As q increases with d, which determines the strength of dependence in the data, we may encounter two problems. First, a large q may mean too much dependence due to fractional integration being discounted and thus lowering the power of the test. Second, Lo and MacKinlay (1989) show that unpredictable behavior of Q_T can occur when q is large relative to the sample size.

When $d < 0.25$, the estimated power of the GPH test is lower than that of the MRR test. With the sample sizes examined, the GPH test has practically no power to detect fractional processes with $d = -0.05$ or 0.05. However, the estimated power rises quickly when $|d| \geq 0.25$.

The last two columns of Table IV report the estimated power of the λ_1 and λ_2 tests. These two variants of the LM test have good power to detect fractional integration.

TABLE IV

REJECTION PERCENTAGE OF THE NOMINAL 5% FRACTIONAL INTEGRATION TEST WHEN THE DATA FOLLOW AN ARFIMA(0, 1, 0) PROCESS

	MRR	GPH	λ_1	λ_2
T = 100				
$d = -0.45$	37.4	25.7	98.5	98.7
$d = -0.25$	29.7	13.4	60.2	62.0
$d = -0.05$	11.4	6.8	4.7	4.9
$d = 0.05$	4.5	5.4	5.0	5.4
$d = 0.25$	5.5	15.6	71.1	71.4
$d = 0.45$	1.1	31.5	98.7	98.9
T = 300				
$d = -0.45$	78.7	50.0	100.0	100.0
$d = -0.25$	49.0	21.3	99.9	99.9
$d = -0.05$	10.3	6.3	15.9	16.9
$d = 0.05$	7.2	6.3	13.8	13.7
$d = 0.25$	27.2	23.7	99.7	99.7
$d = 0.45$	22.7	66.8	100.0	100.0
T = 500				
$d = -0.45$	93.1	66.1	100.0	100.0
$d = -0.25$	58.2	28.1	100.0	100.0
$d = -0.05$	12.6	6.6	24.6	24.8
$d = 0.05$	9.4	5.2	26.8	27.4
$d = 0.25$	36.5	31.3	100.0	100.0
$d = 0.45$	41.1	77.8	100.0	100.0

See the note to Table I. The data were constructed to follow $(1 - B)^d X_t = \varepsilon_t$, $\varepsilon_t \sim N(0, 1)$.

3.3. *Mean or variance shifting processes*

A caveat for interpreting the result of fractional integration tests is that structural changes may be observationally equivalent to fractional integration. Specifically, tests for fractional integration may not distinguish fractional integration from shifts in the mean of a series. Klemes (1974) illustrates that the classical rescaled range statistic (i.e. Q_T with $q = 0$) is biased toward fractional integration alternatives when there are shifts in the mean.

The construction of the MRR statistic provides a clue to the possible effect of shifts in the mean on the MRR test. Given T observations $\{X_1, \ldots, X_T\}$, the range R of cumulative deviations from the sample mean attains the maximum when the observations above (or below) the mean are clustered in one subsample. This pattern of data clustering is likely to occur when there is one shift in the mean. This suggests that a shift in the mean can bias the test in favor of fractional integration alternatives.

The effect of shifts in the mean on the GPH, λ_1 and λ_2 tests is less obvious. However, given the possible impact on shifts in the mean on the Q_T statistic discussed above, we speculate that these tests will be affected in a similar way and will be biased in favor of fractional integration alternatives.

By construction the MRR test is robust to heteroskedasticity. Therefore, we expect that shifts in variance do not affect the empirical size of the MRR test. Once again, the impact of time-varying variance on the GPH, λ_1 and λ_2 tests is not certain.

The Monte Carlo experiment examining the sensitivity of the tests to shifts in parameters is based on the following data-generating processes:

$$X_t = \varepsilon_t$$
$$E\varepsilon_t = 0.5(-1)^k, \qquad (k-1)N < t \leq kN, \tag{15}$$

and

$$X_t = \varepsilon_t$$
$$V(\varepsilon_t) = 1.5 + 0.5(-1)^k, \qquad (k-1)N < t \leq kN, \tag{16}$$

where $N = \text{int}\{T/(g+1)\}$, g is the number of shifts in the mean or variance, and $k = 1, 2, \ldots, g+1$. Essentially, (15) is a normal white noise process with the mean alternating between -0.5 and 0.5 after each N observations. Similarly, (16) is a random process with the variance alternating between 1 and 2 after each N observation. Table V presents the simulation results for $g = 1, 2, 4$ and 9. For notational convenience, we denote (15) with $g = 1$ as M1, with $g = 2$ as M2, with $g = 4$ as M4 and with $g = 9$ as M9, and (16) with $g = 1$ as V1, with $g = 2$ as V2, with $g = 4$ as V4 and with $g = 9$ as V9.

The results in Table V indicate that all the tests are sensitive to shifts in the mean. Consistent with the intuition discussed above, the over-rejection rate is positively related to the sample size and inversely related to the number of regime changes. The λ_1 and λ_2 statistics are relatively more sensitive to shifts in the mean. The breakdown of rejection frequencies indicates that, in the presence of shifts in the mean, these tests tend to yield spurious evidence for $d > 0$ alternatives. On the other hand, the empirical size of these tests appears to be robust to the patterns of variance shifts examined.

4. EXAMPLES

As an illustration this section applies the GPH test, which is commonly used in the literature, to two economic time series. First, we consider the annual real per capita US gross domestic product (GDP) data from 1870 to 1989. The data through 1979 are taken from Maddison (1982) and the rest are taken from various issues of the OECD *Main Economic Indicators*. Figure 1 shows the first log differences of the per capita data. No shifts in the mean are observed although turbulence in the 1930s and 1940s is prominent.

When the GPH test is applied to the first differenced series, the asymptotic t statistic of $\hat{d}$ (the point estimate is -0.67) is equal to -2.28, indicating that there is less than unit root persistence in the per capita GDP data. When the test is applied to the levels of the data, an asymptotic t statistic of 3.35 is

TABLE V

REJECTION PERCENTAGE OF THE NOMINAL 5% FRACTIONAL INTEGRATION TEST WHEN THE DATA HAVE A NON-CONSTANT MEAN OR VARIANCE

	MRR	GPH	λ_1	λ_2
T = 100				
M1	79.1	32.0	89.7	90.3
M2	63.2	33.6	73.3	75.6
M4	5.4	21.9	58.1	61.5
M9	6.3	4.5	20.0	21.0
V1	7.7	3.7	3.4	2.9
V2	6.3	4.2	2.9	2.7
V4	7.3	5.7	1.9	2.0
V9	6.6	5.1	4.2	4.3
T = 300				
M1	100.0	67.4	100.0	100.0
M2	100.0	83.0	100.0	100.0
M4	80.0	76.8	100.0	100.0
M9	0.4	4.8	99.3	99.3
V1	5.6	5.4	5.9	4.9
V2	6.5	5.3	4.6	4.0
V4	6.2	5.3	5.1	4.0
V9	5.8	5.6	5.2	4.7
T = 500				
M1	100.0	82.5	100.0	100.0
M2	100.0	94.6	100.0	100.0
M4	99.3	93.9	100.0	100.0
M9	0.8	5.6	100.0	100.0
V1	6.3	5.7	4.7	3.8
V2	6.0	5.4	5.4	4.3
V4	5.7	5.5	4.1	3.6
V9	5.7	5.5	4.8	4.0

See the note to Table I. The shift-in-mean models considered are M1, M2, M4 and M9 where the number after M denotes the number of shifts in the mean. The shift-in-variance models considered are V1, V2, V4 and V9 where the number after V denotes the number of shifts in the variance. See the text for a more detailed description on the construction of these data series.

obtained for $\hat{d}$. These results suggest that the order of integration of the annual real per capita GDP series is between 0 and 1.

We estimate an ARIMA specification for the data to examine the possibility that the fractional integration result is driven by the ARMA component in the GDP data. ARMA(p, q) models with $p, q \leqslant 3$ are considered. Both the Akaike information criterion and the Bayesian information criterion choose the ARMA(1, 1) model. The maximum likelihood estimates of this model are

$$\underset{}{y_t} = \underset{(-2.11)}{-4.3698} + \underset{(138.1)}{0.9979 y_{t-1}} + \varepsilon_t + \underset{(2.61)}{0.2354\varepsilon_{t-1}}. \tag{17}$$

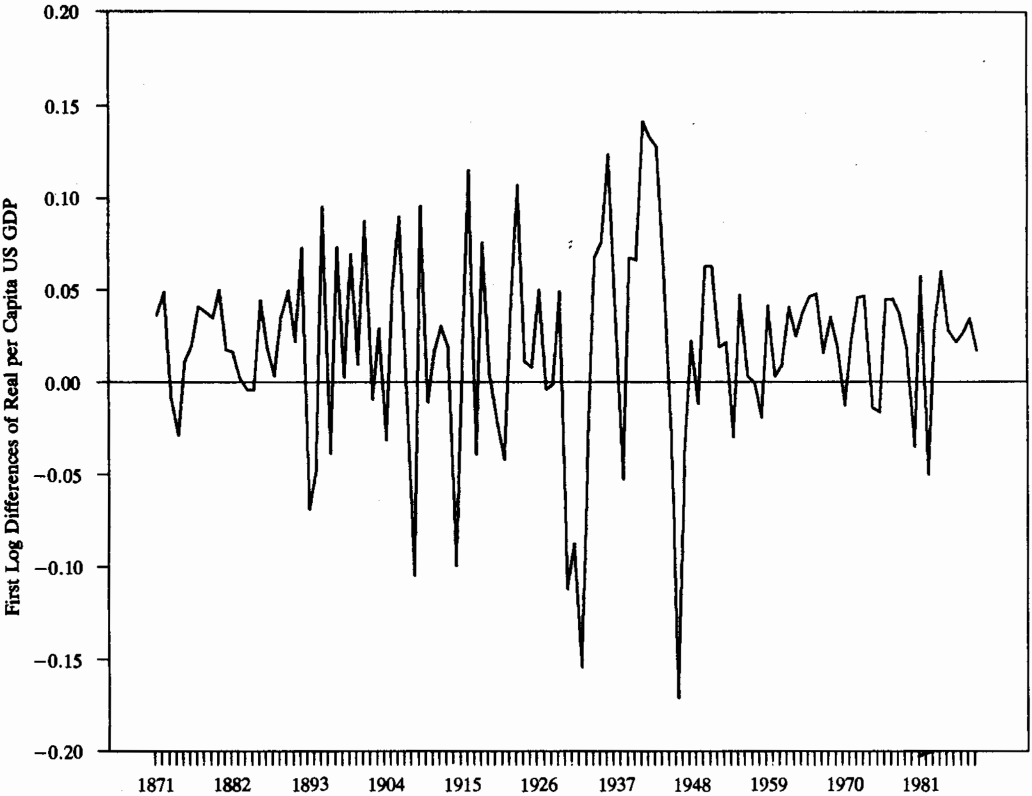

FIGURE 1. Real per capita US gross domestic product.

Asymptotic t-statistics are given in parentheses. The Ljung–Box $Q(12)$ and $Q(24)$ statistics are 8.39 and 26.01 respectively, indicating no serial correlation in the residuals. However, the augmented Dickey–Fuller test indicates that there is a unit root in the data. When we re-estimate an ARIMA(0, 1, 1) model we have

$$\underset{}{(1-B)y_t} = \underset{(3.40)}{0.0190} + \varepsilon_t + \underset{(1.95)}{0.1771}\varepsilon_{t-1}. \tag{18}$$

Again, a $Q(12)$ statistic of 14.67 and a $Q(24)$ statistic of 21.85 indicate no serial correlation in the residuals.

Given the previous simulation results, if the GDP series is generated by (18), it is unlikely that its ARMA components explain the GPH test result. In addition, Sowell (1990a) and Diebold and Rudebusch (1991b) show that the Dickey–Fuller test has low power against fractional alternatives. These results suggest favorable evidence of fractional integration in the GDP data. In fact, Cochrane (1988) and Diebold and Rudebusch (1989), using different US GDP data, also show that the persistence in the GDP data is weaker than a unit root process but stronger than a stationary ARMA process. We should point out that the interpretation of the GPH test depends on the ARMA model assigned to the data. For example, if the data are generated according to (17) or other ARMA models that have an AR or MA root close to but not equal to unity, then the GPH test result should be interpreted with caution. This is because, as demonstrated in Section 3, the GPH test can give spurious fractional integration results when the true data-generating mechanism resembles an integrated process.

The second example is the real three-month T-bill rate series from 1960:I to 1990:II. Both the T-bill rates and the inflation rates used to construct the real series were taken from the *International Financial Statistics* data tape. The series is plotted in Figure 2. The graph clearly suggests that there are shifts in the mean of the real T-bill rate series around 1973 and 1980. When the GPH test is applied to this series, the asymptotic t statistic of $\hat{d}$ is equal to 2.65, indicating the presence of fractional integration.

To ensure that the significant $\hat{d}$ is not induced by shifts in the mean of the T-bill data, we split the series into three segments: (i) from 1960:I to 1973:IV, (ii) from 1974:I to 1980:IV and (iii) from 1981:I to 1990:II. The data for deviation from the mean were then constructed by subtracting each real rate from its corresponding subsample mean. The asymptotic t statistic of $\hat{d}$

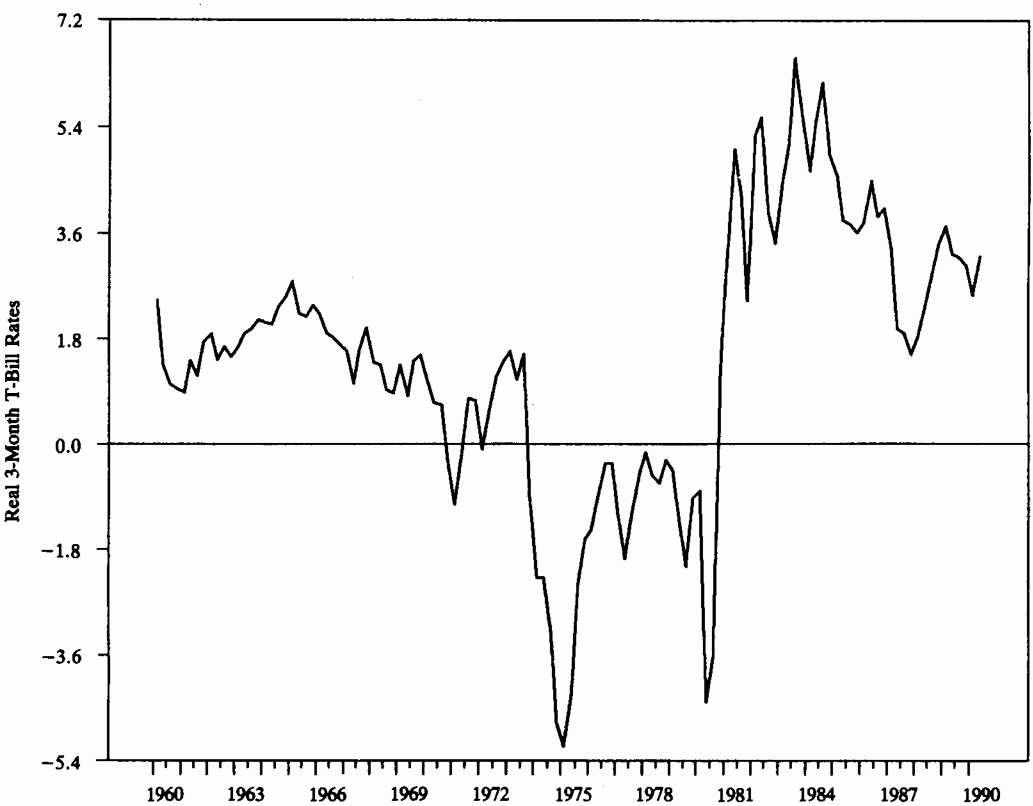

FIGURE 2. Real three-month T-bill rates.

computed from the transformed data is 1.12 and suggests no fractional integration.

An interpretation of the above result is that the rejection of the $d = 0$ hypothesis in the real T-bill series is an artifact of changes in the mean of the series. However, the interpretation is not definitive given no formal analysis of the statistical property of $\hat{d}$ obtained from the de-meaned series. This naturally leads to an interesting further research topic—a test for fractional integration in the presence of structural changes—which is beyond the scope of this paper.

5. CONCLUDING REMARKS

The finite-sample properties of three recently developed tests for fractional integration are investigated using Monte Carlo methods. Compared with simulation results reported in previous studies (e.g. Lo (1991), who reports the performance of the MRR test in the presence of a white noise process, an AR(1) with $\phi = 0.5$ and fractional processes with $d = \pm 1/3$), our experiment considers a much wider class of data-generating mechanisms and a wider range of parameter values. Further, by examining these three representative methods, our exercise provides better information on the general performance of fractional integration tests in finite samples.

While both the MRR and GPH tests are found to be robust to moderate ARMA components, ARCH effects and shifts in the variance, our simulation results suggest several factors that can lead to spurious rejection of the no-fractional-integration hypothesis. For instance, both the MRR and GPH tests are biased towards $d < 0$ alternatives in the presence of large negative MA components. When there are infrequent shifts in the mean, these two tests tend to yield spurious evidence for $d > 0$ alternatives. Further, a large AR parameter also biases the GPH test in favor of $d > 0$ alternatives.

The effect of large ARMA components on the MRR and GPH tests is not unexpected. When ϕ is close to 1, the data series resembles a unit root process. When θ is close to -1, the series resembles an over-differenced process. This explains the results reported in Section 3. The MRR test is robust to large ϕ because its truncation lag parameter q, which controls the amount of autocorrelation to be discounted, is being adjusted according to dependence in the data. For the GPH test, however, the fact that n is fixed at $T^{0.5}$ impedes its ability to adjust for serial correlation in the data. To minimize the effect of ARMA components on the GPH test, the choice of n should be inversely related to the short-run dependence in the data. In fact, Geweke and Porter-Hudak (1983) suggest that n should be kept small if $\hat{d}$ appears sensitive to the choice of n.

It is found that both the λ_1 and λ_2 tests are sensitive to deviations from the null of a white-noise process. The performance of these two tests crucially depends on the specification of the hypotheses under which the statistics are constructed. Our simulation results provide a cautionary note on using these LM tests to detect fractional integration in real-life data.

Overall, our Monte Carlo results and the two economic examples suggest that results of fractional integration tests have to be interpreted with caution, especially when there is good reason to believe that the time series contains a large negative MA component or exhibits shifts in the mean. Preliminary analysis to check for the ARIMA specification and the possibility of structural changes is important for a proper interpretation of the result of fractional integration tests.

ACKNOWLEDGEMENTS

The author is grateful to Francis Diebold, Kevin Hassett, Clive Granger, Kon Lai, Roberto Mariano, Marc Nerlove, Peter Pauly and an anonymous referee for their valuable suggestions and comments.

REFERENCES

Cochrane, J. H. (1988) How big is the random walk in GNP? *J. Polit. Econ.* 96, 893–920.
Diebold, F. X. and Rudebusch, G. D. (1989) Long memory and persistence in aggregate output. *J. Monet. Econ.* 24, 189–209.
—— and —— (1991a) Is consumption too smooth? Long memory and the Deaton paradox. *Rev. Econ. Statist.* 73, 1–17.
—— and —— (1991b) On the power of Dickey–Fuller tests against fractional alternatives. *Econ. Lett.* 35, 155–60.
Fox, R. and Taqqu, M. S. (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* 14, 517–32.
Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long memory time series models. *J. Time Ser. Anal.* 4, 221–38.
Granger, C. W. J. and Joyeux, R. (1980) An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.* 1, 15–29.
Hosking, J. R. M. (1981) Fractional differencing. *Biometrika* 68, 165–76.
—— (1984) Modeling persistence in hydrological time series using fractional differencing. *Water Resour. Res.* 20, 1898–908.
Hurst, H. E. (1951) Long term storage capacity of reservoirs. *Proc. Am. Soc. Civil Eng.* 11, 770–99.
Klemes, K. (1974) The Hurst phenomenon: a puzzle? *Water Resour. Res.* 10, 675–88.
Li, W. K. and McLeod, A. I. (1986) Fractional time series modelling. *Biometrika* 73, 217–21.
Lo, A. W. (1991) Long-term memory in stock market prices. *Econometrica* 59, 1279–314.
—— and MacKinlay, C. (1989) The size and power of the variance ratio test in finite samples: a Monte Carlo investigation. *J. Economet.* 40, 203–38.
Maddison, A. (1982) *Phases of Capitalist Development*, Oxford University Press.
Mandelbrot, B. B. (1972) Statistical methodology for nonperiodic cycles: from the covariance to R/S analysis. *Ann. Econ. Social Meas.* 1, 259–90.
Porter-Hudak, S. (1990) An application of the seasonal fractionally differenced model to the monetary aggregates. *J. Am. Statist. Assoc.* 85, 338–44.
Robinson, P. M. (1991) Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *J. Economet.* 47, 67–84.
Shea, G. S. (1991) Uncertainty and implied variance bounds in long-memory models of the interest rate term structure. *Empirical Econ.* 16, 287–312.

SOWELL, F. (1990a) Fractional unit root distribution. *Econometrica* 50, 495–505.
——(1990b) Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Manuscript, GSIA, Carnegie Mellon University.
YAJIMA Y. (1988) On estimation of a regression model with long-memory stationary errors. *Ann. Statist.* 16, 791–801.
——(1989) A central limit theorem of Fourier transforms of strongly dependent stationary processes. *J. Time Ser. Anal.* 10, 373–83.