Central European Journal of Economic Modelling and Econometrics

Testing for Long-Range Dependence in Financial Time Series

Manveer Kaur Mangat* and Erhard Reschenhofer†

Submitted: 16.11.2018, Accepted: 3.04.2019

Abstract

Various trading strategies have been proposed that use estimates of the Hurst coefficient, which is an indicator of long-range dependence, for the calculation of buy and sell signals. This paper introduces frequency-domain tests for long-range dependence which do, in contrast to conventional procedures, not assume that the number of used periodogram ordinates grow with the length of the time series. These tests are applied to series of gold price returns and stock index returns in a rolling analysis. The results suggest that there is no long-range dependence, indicating that trading strategies based on fractal dynamics have no sound statistical basis.

Keywords: long-range dependence, fractionally integrated process, frequency domain test, Kolmogorov-Smirnov goodness-of-fit-test

JEL Classification: C12, C14, C15, C22, C58

*University of Vienna; e-mail: manveer.mangat@univie.ac.at; ORCID: 0000-0001-5888-3940
†University of Vienna; e-mail: erhard.reschenhofer@univie.ac.at; ORCID: 0000-0002-4866-2324

1 Introduction

In order to describe the long-range dependence of a financial time series, one often uses the Hurst coefficient H, which is related to the fractional differencing parameter d via $H = d + 0.5$: $H = 0.5$ is indicative of no long-range dependence, whereas $H > 0.5$ corresponds to trend-reinforcing behavior and $H < 0.5$ to a mean-reverting behavior. Building on earlier work by Carbone *et al.* (2004), De Souza and Gokcan (2004) and Batten *et al.* (2013), Auer (2016a) suggested a trading strategy that takes a long position when $H > 0.7 \ \wedge B > 0.5$ or $H < 0.3 \ \wedge B < 0.5$, and a short position when $H < 0.3 \ \wedge B > 0.5$ or $H > 0.7 \wedge B < 0.5$, where B expresses the ratio of positive returns to the total number of returns and H and B are computed using a rolling window over $w = 240$ trading days. When $0.3 \leq H \leq 0.7$, three-month treasury bills are purchased. Applying this trading strategy to daily time series of gold, silver and gold-silver spread returns (from January 1979 to March 2015) and including transaction costs, his strategy outperformed the buy-and-hold strategy regardless of the method used to estimate H. For the periodogram regression method for example, developed by Geweke and Porter-Hudak (1983), he only used the first $K = \lfloor w^{0.35} \rfloor = \lfloor 240^{0.35} \rfloor = 6$ Fourier frequencies. However, Reschenhofer at al. (2018) argue that "while choosing a very small value of K makes perfect sense in the case of macroeconomic time series with large business cycles, the situation is quite different in the case of series of log prices of stocks or commodities. In the latter case, short-run effects are much less severe (if at all), hence the number K of included Fourier frequencies should be much larger".

Another application of the Hurst coefficient can be found in stock markets. Due to a considerable number of publications claiming an increase of efficiency in emerging markets over time, Cajueiro and Tabak (2004) and Auer (2016*b*) computed the Hurst coefficient to verify this claim and found a downward trend in the Hurst coefficient. However, this downward trend does not necessarily have to be facilitated by a change of long-range dependence over time. Indeed, Reschenhofer at al. (2018) observed that the autocorrelation of stock returns changed from positive to negative in the last decades, which implies in the case of a simple AR(1) model that the maximum of the spectral density at frequency zero becomes a minimum and the slope of the spectral density, on which the estimation of d is based, changes accordingly.

More reliable results can be obtained when a formal statistical test is employed. Conventional tests are based on the asymptotic normality of semiparametric estimators of d which only use the periodogram ordinates at the lowest frequencies in order to avoid interference from short-range dependencies. Thereby it is assumed that both the length n of the time series and the number K of used periodogram ordinates are large and that K is small relative to n (Geweke and Porter-Hudak, 1983; Robinson, 1995). Unfortunately, this assumption is often implausible in practice, particularly when d is estimated over a rolling window. A parametric approach based on fractionally integrated ARMA models (see Granger and Joyeux, 1980; Hosking, 1981), which can be estimated by maximum likelihood (Sowell, 1992) or approximate maximum likelihood (Fox and Taqqu, 1986), is not an alternative unless we are sure that these models are correctly specified (see, e.g., Robinson, 1995; Reschenhofer, 2013). An interesting alternative to the specification of a fixed model (which we do not intend to pursue in this paper) would be to formally take into account the model uncertainty, e.g., in a Bayesian approach. For example, Koop *et al.* (1997) argued that "functions of the model parameters, like impulse responses, are not model-specific quantities and it is formally possible to average them over models" (for posterior properties of long-run impulse responses see Koop *et al.*, 1994). In the next section, we therefore propose new frequency domain tests for the fractional differencing parameter d, which are based on a fixed number of periodogram ordinates in the neighborhood of frequency zero. The results of a Monte-Carlo power study are presented in Section 3. Section 4 applies the tests to financial time series before Section 5 concludes.

2 Testing for Long-Range Dependence

Suppose that the mean-corrected observations y_t can be described by a fractionally integrated ARMA (ARFIMA) process

$$y_t = (1 - \phi_1 L - \cdots - \phi_p L^p)^{-1} (1 - L)^{-d} (1 + \theta_1 L + \cdots + \theta_q L^q) u_t \tag{1}$$

(Granger and Joyeux, 1980; Hosking, 1981) with spectral density

$$\begin{aligned} f(\omega) &= |1 - e^{-i\omega}|^{-2d} \underbrace{\frac{\sigma^2}{2\pi} \left(\left| 1 + \sum_{j=1}^{q} \theta_j e^{-i\omega j} \right|^2 \left| 1 - \sum_{j=1}^{p} \phi_j e^{-i\omega j} \right|^{-2} \right)}_{f_0(\omega)} = \\ &= (2\sin(\omega/2))^{-2d} f_0(\omega), \end{aligned} \tag{2}$$

where the fractional differencing parameter $-0.5 < d < 0.5$ takes care of any long-range dependence, while the autoregressive (AR) parameters $\phi_1, \ldots, \phi_p$ and the moving average (MA) parameters $\theta_1, \ldots, \theta_q$ describe the short-range dependence. Since the ARMA component $f_0(\omega)$ of the spectral density is approximately constant near frequency zero, the spectral density can be approximated by

$$f(\omega) \sim c\omega^{-2d}, \quad c > 0, \tag{3}$$

in a neighborhood of frequency zero. Subsection 2.1 proposes frequency-domain tests for testing hypotheses about the parameter d and Subsection 2.2 checks the plausibility of certain assumptions on which these tests are based.

2.1 Test Statistics

In this subsection, we develop test statistics for testing hypotheses of the form

$$H_0 : d \geq d_0 > 0 \text{ vs } H_A : d < d_0 \tag{4}$$

and

$$H_0 : d \leq d_0 < 0 \text{ vs } H_A : d > d_0. \tag{5}$$

However, being skeptical about the presence of long-range dependence in return series, we use a more specific alternative for the construction of the test statistics, namely $H_A : d = 0$. Considering the case

$$H_0 : d = d_0 \neq 0 \text{ and } H_A : d = 0, \tag{6}$$

where both hypotheses are simple, we obtain the most powerful test for comparing these hypotheses by using the likelihood ratio as test statistic (Neyman-Pearson Lemma). We use the frequency-domain likelihood (Whittle likelihood), which is based on the assumption that the periodogram ordinates

$$I_k = I(\omega_k) = \frac{1}{2\pi n}\Big|\sum_{t=1}^{n} y_t e^{-i\omega_k t}\Big|^2 \tag{7}$$

at the Fourier frequencies $\omega_k := 2\pi k/n$, $k = 1, \ldots, m = \lfloor (n-1)/2 \rfloor$ are approximately independent and exponentially distributed with means $f_k := f(\omega_k)$ (see Subsection 2.2 for a discussion of the plausibility of this assumption), because it allows us to achieve robustness against short-range autocorrelation simply by using only frequencies $\omega_k, k = 1, \ldots, K << m$, in a neighborhood of frequency zero. The only term of interest in the log ratio

$$\begin{aligned} &\log\left(\frac{\prod_{k=1}^{K}(1/f_k^A)\exp(-I_k/f_k^A)}{\prod_{k=1}^{K}(1/f_k^0)\exp(-I_k/f_k^0)}\right) = \\ &= \sum_{k=1}^{K}\log(f_k^0) - \sum_{k=1}^{K}\log(f_k^A) + \sum_{k=1}^{K}\frac{I_k}{f_k^0} - \sum_{k=1}^{K}\frac{I_k}{f_k^A} \sim \\ &\sim \sum_{k=1}^{K}\log(c_0\omega_k^{-2d_0}) - K\log(c_A) + \frac{1}{c_0}\sum_{k=1}^{K} I_k/\omega_k^{-2d_0} - \frac{1}{c_A}\sum_{k=1}^{K} I_k \end{aligned} \tag{8}$$

is the third one, which has, after multiplication by 2, approximately a χ^2-distribution with $2K$ degrees of freedom. The problem that the scale parameter c_0 is unknown can be addressed by rescaling. Unfortunately, the sum of the rescaled periodogram ordinates

$$I_j^0 = \frac{I_j/f_j^0}{\sum_{k=1}^{K} I_k/f_k^0} \sim \frac{I_j/\omega_j^{-2d_0}}{\sum_{k=1}^{K} I_k/\omega_k^{-2d_0}}, \quad j = 1, \ldots, K, \tag{9}$$

is equal to one and is therefore useless as test statistic. However, the random variables

$$J_k^0 := \sum_{j=1}^{k} I_j^0, \quad k = 1, \ldots, K-1, \tag{10}$$

are under the null hypothesis approximately distributed as the order statistics of a random sample of size $K-1$ from a uniform distribution on $[0,1]$. If the null hypothesis is false and the alternative hypothesis is true, i.e. $d=0$, then the spacings (distances between successive order statistics) I_j^0 will have a downward trend if $d_0<0$. In this case, the cumulative distribution function (CDF) will first be flat and then become steeper and steeper, hence the true distribution will be stochastically greater than a uniform distribution.

Thus, the null hypothesis can be tested by applying a one-sided Kolmogorov-Smirnov goodness-of-fit test which is based on the supremum

$$D_{K-1}^{-}=\sup_x(F_0(x)-F_{K-1}(x)) \tag{11}$$

of the differences between the hypothesized CDF and the empirical distribution function

$$F_{K-1}(x)=\frac{1}{K-1}\sum_{k=1}^{K-1}I_{(-\infty,x]}J_k^0. \tag{12}$$

Analogously, the test statistic

$$D_{K-1}^{+}=\sup_x(F_{K-1}(x)-F_0(x)) \tag{13}$$

can be used if $d_0>0$.

Naturally, the statistics D_{K-1}^{-} and D_{K-1}^{+} can also be used for testing the composite null hypotheses (4) and (5). The fact that the rescaled periodogram ordinates I_j^0 exhibit a downward trend or an upward trend under the null hypothesis $d_0<0$ and $d_0>0$, respectively, which implies a convex or a concave CDF for the cumulative variables J_k^0, is crucial for the performance of these tests because of the well-known inefficiency of the Kolmogorov-Smirnov test in case of more complex (e.g., multimodal) alternatives (see Reschenhofer and Bomze, 1992; Reschenhofer, 1997).

Finally, in case we suspect there is indeed some long-range dependence and want to find supporting evidence, we may simply apply a one-sided or two-sided Kolmogorov-Smirnov test to the cumulative normalized periodogram

$$J_k=\sum_{j=1}^{k}I_j/\sum_{j=1}^{K}I_j,\quad k=1,\dots,K-1. \tag{14}$$

For $K=m$, this test reduces to a standard frequency-domain test for white noise.

2.2 Plausibility Checks

In order to ensure that for $d\neq 0$ the normalized periodogram ordinates $I(\omega_J)/f(\omega_J),\dots,I(\omega_K)/f(\omega_K)$ are asymptotically i.i.d. standard exponential, we have to let J grow with n (i.e., $J/\sqrt{n}\to\infty$; see Künsch, 1986). Clearly, the omission of the very first (and possibly most informative) periodogram ordinates $I_1, \ldots, I_{J-1}$ comes with a price. For example, Reisen *et al.* (2001) compared several methods for the estimation of the parameter d and found that already the omission of only the first Fourier frequency leads to an increase in the mean squared error. We therefore decided against omitting any frequencies. Further justification for this decision is given below.

For illustration, we assume that the spectral density $f(\omega)$ of y_t is given by (3) in the neighborhood of frequency zero and, in addition, the autocovariance function $\gamma(j)$ of y_t is given by

$$\gamma(j) \sim C(d) j^{2d-1} \tag{15}$$

for large j, which holds true for ARFIMA processes (e.g., Brockwell and Davis, 1987, 469).

In the case of a fractionally integrated white noise

$$(1-L)^d y_t = u_t, \tag{16}$$

where u_t is white noise with mean zero and variance σ^2, we have $c = \sigma^2/2\pi$ and

$$C(d) = \frac{\sigma^2 \Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} = \sigma^2 \Gamma(1-2d) \frac{\sin(\pi d)}{\pi}, \tag{17}$$

(see e.g., Brockwell and Davis (1987), 466-467), hence Robsinson's (1995) approximation (1.5) of $\mathrm{E}[I_k/f_k]$ becomes

$$2\Gamma(1-2d)\sin(\pi d) \frac{k^{2d}}{(2\pi)^{2d-1}} \Big\{ \frac{1}{d(2d+1)} - 4 \int_0^1 (1-\lambda)\lambda^{2d-1} sin^2(\pi k \lambda) d\lambda \Big\}. \tag{18}$$

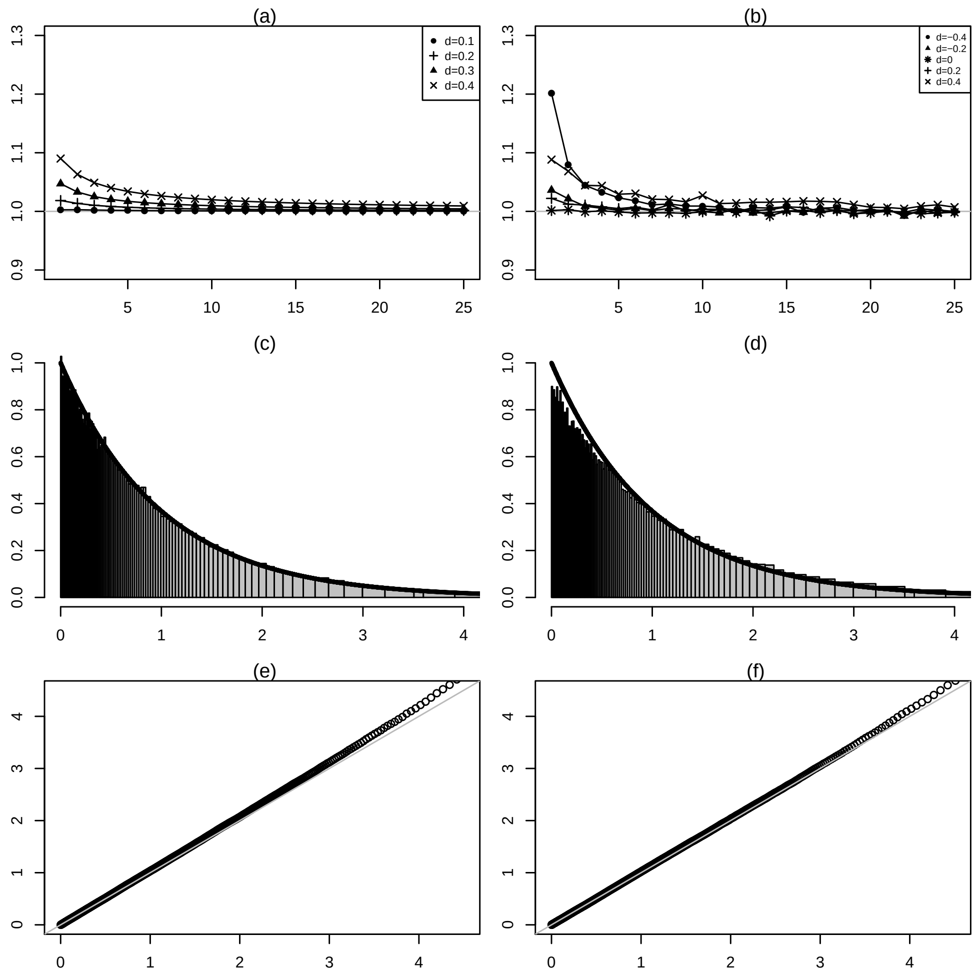
Figure 1.a plots the approximation (18) for the values $k = 1, \ldots, 25$, $d = 0.1, 0.2, 0.3, 0.4$. Apparently, the deviations from 1 are relatively small. This finding is corroborated by Figure 1.b, which displays the means of the normalized periodogram ordinates I_k/f_k obtained from 100,000 realizations $y_1, \ldots, y_n$ of an ARFIMA(1,d,0) process

$$(1-\phi L)(1-L)^d y_t = u_t \tag{19}$$

for $n = 250$, $\phi = 0.1$, $k = 1, 2, \ldots, 25$, $d = -0.4, -0.2, 0, 0.2, 0.4$. Moreover, not only are the means of the normalized periodogram ordinates close to 1, but their distributions look similar to the standard exponential distribution (see Figure 1.c-f). All computations are carried out with the free statistical software R (R Core Team, 2017).

Finally, the simulation results presented in the next section show that the tests proposed in this paper attain the advertised levels of significance, which also speaks against a serious violation of the assumptions underlying our tests.

Figure 1: Distributions of the normalized periodogram ordinates I_k/f_k



(a) Large-sample approximations of the means for $k = 1, \ldots, 25$, $d = 0.1, 0.2, 0.3, 0.4$

(b) Sample means obtained from 100,000 realizations $y_1, \ldots, y_n$ of a fractionally integrated ARMA(1,0) process for $n = 250$, $\phi = 0.1$, $k = 1, \ldots, 25$, $d = -0.4, -0.2, 0, 0.2, 0.4$

(c), (d) Histogram of I_1/f_1 from $100,000$ realizations for $d = 0.25$ (c), 0.49 (d) versus probability density of the standard exponential distribution (bold black line)

(e), (f) $Q - Q$ plots of sample quantiles of I_1/f_1 (e) and I_5/f_5 (f) from $100,000$ realizations for $d = 0.25$ versus quantiles of the standard exponential distribution (the gray line is the 45-degree reference line)

3 Monte Carlo Power Study

For a closer examination of the performance of the new tests, a Monte Carlo power study was conducted. Two standard tests were also included in this study to provide evidence of the need for a new test. Both tests are based on the asymptotic normality of the estimator of d (see Hurvich *et al.*, 1998) obtained from the log periodogram regression

$$\log I(\omega_k) = c + d(-\log|1-e^{-i\omega_k}|^2) + v_k, \quad k = 1, \dots, K, \tag{20}$$

(see Geweke and Porter-Hudak, 1983). The first test, T_a, uses the asymptotic variance and the second, T_{LS}, the variance formula of the LS estimator of the slope in a simple linear regression, which depends on the sample size. The difference between the two tests may be substantial for small values of K but vanishes as K increases. For $d_A = -0.4, -0.3, \dots, 0.3, 0.4$ and $\phi = -0.1, 0, 0.1$, 5,000 pseudorandom

Table 1: Rejection rates at the 1% level of tests based on the test statistics D^+_{K-1} ($d_A \leq d_0$) and D^-_{K-1} ($d_A > d_0$), respectively, for $K = 6$ (top value), $K = 24$ (bottom value) and $n = 250$

$d_0\backslash d_A$	−0.4	−0.3	−0.2	−0.1	0	0.1	0.2	0.3	0.4
−0.4	**0.006**	0.026	0.042	0.078	0.119	0.175	0.259	0.359	0.445
	0.006	0.073	0.25	0.481	0.652	0.921	0.975	0.992	0.999
−0.3	0.013	**0.009**	0.022	0.047	0.081	0.118	0.188	0.268	0.341
	0.021	**0.01**	0.063	0.32	0.475	0.781	0.952	0.968	0.991
−0.2	0.022	0.014	**0.008**	0.022	0.04	0.065	0.122	0.188	0.257
	0.091	0.034	**0.008**	0.061	0.213	0.486	0.74	0.91	0.981
−0.1	0.038	0.024	0.011	**0.008**	0.02	0.036	0.079	0.12	0.18
	0.197	0.092	0.035	**0.01**	0.065	0.221	0.501	0.751	0.925
0	0.039	0.025	0.012	0.01	**0.009**	0.011	0.024	0.045	0.078
	0.258	0.139	0.056	0.022	**0.01**	0.035	0.163	0.418	0.751
0.1	0.075	0.058	0.054	0.027	0.026	**0.01**	0.018	0.041	0.078
	0.53	0.41	0.378	0.102	0.104	**0.01**	0.067	0.235	0.52
0.2	0.096	0.076	0.054	0.04	0.026	0.019	**0.009**	0.019	0.042
	0.689	0.545	0.38	0.225	0.104	0.037	**0.006**	0.065	0.25
0.3	0.124	0.102	0.081	0.055	0.04	0.032	0.016	**0.009**	0.025
	0.812	0.715	0.548	0.375	0.217	0.1	0.0317	**0.009**	0.074
0.4	0.175	0.138	0.107	0.08	0.062	0.046	0.028	0.015	**0.009**
	0.904	0.825	0.7	0.551	0.374	0.214	0.093	0.03	**0.009**

samples of size $n = 250$ from the ARFIMA(1,d,0) process (19) were obtained with the help of the R package *fracdiff*. For each sample, the hypothesis (4) with $d_0 = 0.1, 0.2, 0.3, 0.4$ is tested using the test statistic D^+_{K-1}, the hypothesis (5) with $d_0 = -0.1, -0.2, -0.3, -0.4$ is tested using the test statistic D^-_{K-1}, and the hypothesis

$$H_0: d = d_0 = 0 \text{ vs } H_A: d \neq 0 \tag{21}$$

is tested using the test statistic D^+_{K-1} when $d_A < 0$ or D^-_{K-1} when $d_A > 0$.
For $\phi = 0$, Tables 1 and 2 show the rejection rates at the 1% and 5% level, respectively. We omitted the tables for $\phi = -0.1, 0.1$ as they do not substantially differ from these tables. At the 5% level and $K = 24$, the power is 0.301 (0.435) if $d_A = 0$ and $d_0 = 0.2$ (-0.2). The corresponding values for $K = 6$ are 0.114 and 0.153. These cases are particularly relevant when we suspect that there is no long-range dependence at all and we want to reject the null hypothesis that $d_0 \geq 0.2$ ($d_0 \leq -0.2$). The power is relatively low. However, in a rolling analysis, we obtain a large number of test results, hence it may be possible for us to draw our conclusions from the overall picture. Tables 3 an 4 are analogous to Tables 1 and 2. They show the rejection rates at the 0.01 and 0.05 level of significance, respectively, for the conventional test T_{LS}. For the 5% level and $K = 24$, the power is 0.114 (0.101) if $d_A = 0$ and $d_0 = 0.2$ (-0.2). The corresponding values for $K = 6$ are 0.006 and 0.005. Obviously, the power is much lower than that of the new test. The tables for the test T_a are omitted because this test is completely useless in case of small sample sizes. It incorrectly rejects the null hypothesis at the 1% level with a probability of about 0.05 and at the 5% level with a probability of about 0.1 if $K = 6$. The corresponding probabilities are about 0.02 and 0.06, respectively, if $K = 24$.

Table 2: Rejection rates at the 5% level of tests based on the test statistics D^+_{K-1} ($d_A \leq d_0$) and D^-_{K-1} ($d_A > d_0$), respectively, for $K = 6$ (top value), $K = 24$ (bottom value) and $n = 250$

$d_0\backslash d_A$	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4
-0.4	**0.041**	0.099	0.145	0.214	0.31	0.375	0.498	0.5988	0.67
	0.041	0.22	0.462	0.709	0.9	0.97	0.995	0.998	0.999
-0.3	0.072	**0.044**	0.095	0.149	0.209	0.289	0.403	0.508	0.58
	0.121	**0.048**	0.189	0.438	0.701	0.892	0.972	0.992	0.999
-0.2	0.097	0.074	**0.044**	0.12	0.153	0.208	0.299	0.414	0.492
	0.282	0.135	**0.044**	0.182	0.435	0.716	0.903	0.965	0.991
-0.1	0.109	0.079	0.069	**0.051**	0.086	0.138	0.218	0.307	0.391
	0.138	0.292	0.1316	**0.052**	0.185	0.435	0.703	0.881	0.952
0	0.121	0.082	0.07	0.054	**0.049**	0.059	0.095	0.131	0.226
	0.531	0.352	0.189	0.095	**0.053**	0.118	0.405	0.62	0.85
0.1	0.238	0.251	0.192	0.112	0.101	**0.051**	0.09	0.103	0.222
	0.829	0.825	0.67	0.298	0.298	**0.046**	0.161	0.431	0.73
0.2	0.293	0.261	0.231	0.15	0.114	0.089	**0.042**	0.085	0.152
	0.918	0.837	0.695	0.494	0.301	0.131	**0.042**	0.203	0.481
0.3	0.352	0.3	0.243	0.198	0.143	0.104	0.081	**0.046**	0.09
	0.967	0.92	0.858	0.679	0.507	0.315	0.13	**0.047**	0.208
0.4	0.405	0.371	0.3	0.234	0.192	0.146	0.101	0.07	**0.042**
	0.991	0.962	0.917	0.825	0.641	0.385	0.2898	0.136	**0.046**

Table 3: Rejection rates at the 1% level of the test T_{LS} (right-tailed if $d_A > d_0$ and left-tailed if $d_A \leq d_0$) for $K = 6$ (top value), $K = 24$ (bottom value) and $n = 250$

$d_0 \backslash d_A$	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4
-0.4	**0.000**	0.001	0.000	0.000	0.001	0.000	0.000	0.002	0.007
	0.002	0.003	0.021	0.085	0.252	0.492	0.723	0.887	0.963
-0.3	0.001	**0.001**	0.000	0.000	0.001	0.000	0.000	0.002	0.002
	0.010	**0.008**	0.003	0.023	0.076	0.245	0.498	0.726	0.882
-0.2	0.001	0.001	**0.002**	0.000	0.000	0.000	0.000	0.002	0.000
	0.034	0.010	**0.004**	0.002	0.012	0.080	0.262	0.498	0.731
-0.1	0.001	0.001	0.002	**0.000**	0.000	0.000	0.000	0.000	0.000
	0.083	0.027	0.017	**0.003**	0.000	0.022	0.083	0.247	0.497
0	0.003	0.002	0.002	0.000	**0.001**	0.000	0.000	0.000	0.000
	0.199	0.077	0.038	0.009	**0.003**	0.008	0.016	0.102	0.253
0.1	0.004	0.002	0.003	0.000	0.002	**0.001**	0.000	0.000	0.000
	0.427	0.202	0.103	0.031	0.009	**0.001**	0.004	0.023	0.090
0.2	0.006	0.002	0.004	0.000	0.002	0.002	**0.001**	0.000	0.000
	0.672	0.421	0.244	0.096	0.029	0.010	**0.003**	0.005	0.026
0.3	0.011	0.004	0.005	0.002	0.002	0.002	0.001	**0.003**	0.000
	0.872	0.658	0.469	0.235	0.096	0.032	0.010	**0.004**	0.004
0.4	0.018	0.006	0.005	0.002	0.002	0.002	0.001	0.003	**0.001**
	0.960	0.868	0.680	0.456	0.224	0.096	0.034	0.011	**0.000**

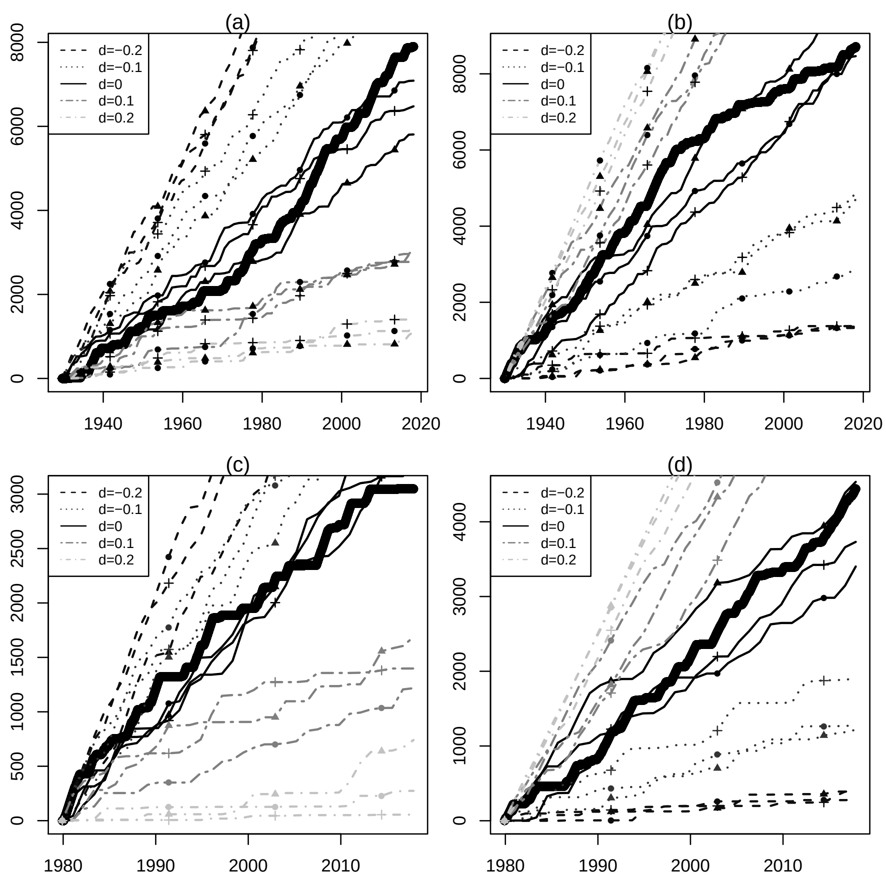
Table 4: Rejection rates at the 5% level of the test T_{LS} (right-tailed if $d_A > d_0$ and left-tailed if $d_A \leq d_0$) for $K = 6$ (top value), $K = 24$ (bottom value) and $n = 250$

$d_0 \backslash d_A$	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4
-0.4	**0.003**	0.001	0.003	0.005	0.014	0.015	0.020	0.044	0.070
	0.012	0.035	0.112	0.300	0.583	0.774	0.901	0.966	0.997
-0.3	0.005	**0.002**	0.003	0.003	0.007	0.008	0.012	0.024	0.042
	0.040	**0.011**	0.027	0.106	0.322	0.561	0.769	0.905	0.971
-0.2	0.008	0.003	**0.004**	0.001	0.005	0.005	0.007	0.010	0.024
	0.103	0.035	**0.020**	0.032	0.101	0.297	0.551	0.767	0.895
-0.1	0.013	0.004	0.005	**0.002**	0.004	0.002	0.002	0.007	0.011
	0.251	0.096	0.048	**0.012**	0.024	0.113	0.308	0.563	0.775
0	0.022	0.007	0.006	0.002	**0.002**	0.000	0.000	0.002	0.007
	0.485	0.250	0.126	0.040	**0.013**	0.029	0.124	0.298	0.540
0.1	0.027	0.012	0.009	0.003	0.003	**0.003**	0.000	0.002	0.004
	0.714	0.477	0.280	0.122	0.041	**0.015**	0.029	0.129	0.305
0.2	0.038	0.019	0.014	0.005	0.006	0.004	**0.002**	0.002	0.000
	0.899	0.703	0.509	0.275	0.114	0.044	**0.014**	0.033	0.123
0.3	0.047	0.030	0.021	0.010	0.010	0.005	0.002	**0.005**	0.000
	0.973	0.902	0.739	0.515	0.263	0.124	0.041	**0.015**	0.032
0.4	0.079	0.045	0.041	0.013	0.014	0.005	0.006	0.005	**0.003**
	0.993	0.976	0.911	0.759	0.477	0.274	0.123	0.043	0.014

4 Empirical Results

In this section, we apply our tests both to gold price returns and stock index returns. The gold price returns were obtained from a series of daily gold prices from 1979-01-01 to 2017-11-10 (downloaded from the website www.gold.org of the World Gold Council). The stock index returns were obtained from the daily DJIA from 1928-10-02 to 2018-02-07 (downloaded from Yahoo!Finance). In the latter case, the choice of an extremely long observation period makes sure that periods of varying degrees of efficiency are included. Figure 2 displays the cumulative numbers of rejections

Figure 2: Cumulative numbers of rejections (at the 5% level) by tests based on D^{+}_{K-1} with $d_0 = 0.2$, $K = 24$ (a, c) and D^{-}_{K-1} with $d_0 = -0.2$, $K = 24$ (b, d) for financial time series (bold; a, b: DJIA returns, c, d: gold returns) and synthetic series obtained with various values of d $(d = -0.2, \ldots, 0.2)$ and $\phi = -0.1$ (●), $\phi = 0$ (+), $\phi = 0.1$ (▲)



(at the 5% level of significance) obtained for the gold price returns as well as for the stock index returns in a rolling analysis using the tests based on the statistics D^{+}_{K-1} with $d_0 = 0.2$ and D^{-}_{K-1} with $d_0 = -0.2$ for $K = 24$ and window length $n = 250$. Comparing these lines with the corresponding lines obtained for synthetic time series from (19) with $d = -0.2, -0.1, 0, 0.1, 0.2$ and $\phi = -0.1, 0, 0.1$, we find a high agreement only if $d = 0$. Of course, it is possible to obtain similar rejection rates in case of a balanced mixture of values of d greater than 0.2 and less than -0.2, respectively. However, in the absence of any contrary indication, it is generally preferable to go for the simplest explanation. Furthermore, in order to be of any practical use, the subperiods with extreme values of d must not be too short. Unfortunately, we found no evidence of varying long-range dependence in the cumulative plots. Hence we conclude that there is no indication that $|d|$ could be greater than 0.2 over a longer period of time.

5 Conclusions

In this paper, we question the existence of fractal dynamics in return series and propose new frequency domain tests for testing hypotheses about the fractional differencing parameter d, which we then apply to DJIA returns and gold price returns. Our test procedure is a two step process where we first transform a subsample of periodogram ordinates, so that their cumulative sum is distributed as the order statistics of a random sample of size $K - 1$ from a uniform distribution on $[0, 1]$ under the null hypothesis and either has a concave or a convex CDF under the alternative hypothesis. Then we apply a Kolmogrov-Smirnov goodness-of-fit-test for a uniform distribution on [0,1], which is most powerful against these types of alternatives. In the special case, where the null hypothesis states that there is no long-range dependence (i.e., $d = 0$), our test reduces to a standard frequency-domain test for white noise, though applied only to a part of the periodogram. We also provide evidence that there is no need to omit the very first periodogram ordinates because of their deviating properties in the case of long-range dependence.

In a rolling frequency-domain analysis of a daily financial time series, we can typically only assume that the size of the rolling window is large but not the number of included periodogram ordinates that are not corrupted by short-range dependence. It is therefore a great advantage of our tests that they are based on a fixed number of periodogram ordinates. In contrast, conventional frequency-domain tests critically depend on the assumption that the number of included periodogram ordinates is large. If this assumption is violated they either have extremely low power or do not attain the advertised levels of significance.

The results of our Monte Carlo power study suggest that it may be difficult to distinguish between values of d that are too close to each other, e.g., -0.2 and 0 or 0.2 and 0. However, in our rolling analysis of gold price returns and stock index returns, we obtain a great number of test results which can be compared to the corresponding results obtained from synthetic series of the same length that have been generated with a known value of d. In this comparison, we find a perfect agreement between the returns series and the synthetic series that exhibit no long-range dependence over the whole observation period. Since there are also no indications of significant deviations from short-range dependence in any subperiods, we conclude that trading strategies that are based on fractal dynamics have no sound statistical basis.

Acknowledgments

The authors thank the editor, Jacek Osiewalski, and an anonymous reviewer for helpful suggestions and comments.

References

[1] Auer B. R. (2016a), On the performance of simple trading rules derived from the fractal dynamics of gold and silver price fluctuations, *Finance Research Letters* 16, 255–267.

[2] Auer B. R. (2016b). On time-varying predictability of emerging stock market returns. *Emerging Markets Review* 27, 1–13.

[3] Batten J., Ciner C., Lucey B., Szilagyi P., (2013), The structure of gold and silver spread returns, *Quantitative Finance* 13, 561–570.

[4] Brockwell P.J., Davis R.A., (1987), *Time Series: Theory and Methods*. New York: Springer-Verlag.

[5] Cajueiro D., Tabak B., (2004), The Hurst exponent over time: testing the assertion that emerging markets are becoming more efficient, *Physica A* 336, 521–537.

[6] Carbone A., Castelli G., Stanley H.E., (2004), Time-dependent Hurst exponent in financial time series, *Physica A* 344, 267–271.

[7] De Souza C., Gokcan S., (2004), Hedge fund investing: a quantitative approach to hedge fund manager selection and deselection, *The Journal of Wealth Management* 6, 52–73.

[8] Fox R., Taqqu M.S., (1986), Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *The Annals of Statistics* 14, 517–532.

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

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

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

[12] Hurvich C. M., Deo R., Brodsky J., (1998), The mean square error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series, *Journal of Time Series Analysis* 19, 19–46.

[13] Koop G., Osiewalski J., Steel M., (1994), Posterior properties of long-run impulse responses, *Journal of Business and Economic Statistics* 12, 489–492.

[14] Koop G., Ley E., Osiewalski J., Steel M., (1997), Bayesian analysis of long memory and persistence using ARFIMA models, *Journal of Econometrics* 76, 149–169.

[15] Künsch H.R., (1986), Discrimination between monotonic trends and long-range dependence, *Journal of Applied Probability* 23, 1025–1030.

[16] R Core Team, (2017), R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

[17] Reisen V., Abraham B., Lopes S., (2001), Estimation of parameters in ARFIMA processes: a simulation study, *Communications in Statistics – Simulation and Computation* 30, 787–803.

[18] Reschenhofer E., (1997), Generalization of the Kolmogorov-Smirnov test, *Computational Statistics & Data Analysis* 24, 433–441.

[19] Reschenhofer E., Bomze I.M., (1992), Length tests for goodness-of-fit, *Biometrika* 78, 207–216.

[20] Reschenhofer E., Mangat M.K., Zwatz C., Guzmics S., (2018), Evaluation of current research on stock return predictability, *Working paper*.

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

[22] Sowell F., (1992), Maximum likelihood estimation of stationary univariate fractionally integrated models, *Journal of Econometrics* 53, 165–188.