HELSINGIN KAUPPAKORKEAKOULU

TESTING FOR FRACTIONAL NOISE IN FINANCIAL TIME SERIES

Taloustieteiden kvantitatiivisten menetelmien
pro gradu -tutkielma
Sampsa Kiiski
Kevät 2000

Kansantaloustieteen laitoksen

laitosneuvoston kokouksessa 19 / 5 2000 hyväksytty

arvosanalla Erinomainen (90 p.)

professori Antti Kanto lehtori Tapani Lehtonen

HELSINGIN KAUPPAKORKEAKOULU
Taloustieteiden kvantitatiiviset menetelmät
Pro gradu –tutkielma
Sampsa Kiiski

TIIVISTELMÄ

6.4. 2000

FRAKTIONAALISEN KOHINAN TESTAUS RAHOITUSAIKASARJOISSA

Fraktionaalista kohinaa on tutkittu huomattavasti viime aikoina, koska sitä hyödyntämällä voidaan kasvattaa eräiden yleisesti käytettyjen stokastisten differenssimallien tarkkuutta tiheissä ja pitkäkestoisissa aikasarjoissa. Tässä tutkielmassa testaamme useilla estimointimetodeilla fraktionaalisen kohinan esiintymistä Helsinki HEX 50, Shanghai SHX All Share ja Standard & Poor's 500 osakeindeksien päivätuotoissa, sekä U.S. Federal Funds Rate -koron kuukausittaisissa muutoksissa. Tulokset osoittavat tilastollisesti merkitsevästi näiden aikasarjojen muodostuvan fraktionaalisesta kohinasta, todentaen myös tätä vastaavan pitkän aikavälin riippuvuden.

Klassinen yksinkertainen differointi ei näin ollen ole riittävä kyseessä olevien rahoitusaikasarjojen stationarisoimiseksi. Stationarisointi on sen sijaan suoritettava käyttäen fraktionaalista differenssiä, joka vastaa kyseessä olevan sarjan fraktionaalista integroituvuusastetta. Vain fraktionaalista differointia käyttäen on mahdollista estimoida harhattomasti myös lyhyen aikavälin riippuvuus. Estimoimme yhtäaikaisesti sekä pitkän että lyhyen aikavälin riippuvuutta käyttäen stokastista autoregressiivistä fraktionaalisesti integroitunutta liukuvan keskiarvon ARFIMA(p,d,q) –mallikategoriaa, jolla on saavutettu parempi ennustustarkkuus kuin vastaavilla yksinkertaisesti differoiduilla malleilla.

Avainsanat: tilastotiede, aikasarja-analyysi, stokastisuus, stationaarisuus, integroituvuus, fraktionaalinen, differenssi, estimointi, osakeindeksi, tuotto

HELSINKI SCHOOL OF ECONOMICS AND BUSINESS ADMINISTRATION
Quantitative Methods In Economics And Management Science
Master's Thesis
Sampsa Kiiski

ABSTRACT

April 6, 2000

TESTING FOR FRACTIONAL NOISE IN FINANCIAL TIME SERIES

Fractional noise has recently been an active research subject, since it provides improved accuracy for certain stochastic difference models of long and dense time series. Utilizing several estimation methods, we test for fractional noise in the daily Helsinki HEX 50, Shanghai SHX All Share, and Standard & Poor's 500 stock index returns, and the monthly U.S. Federal Funds Rate changes. The results show statistically significantly that these time series are composed of fractional noise, and verify the corresponding long term dependence.

The classical simple difference is therefore not adequate for stationarization of these financial time series. Instead, stationarization needs to be performed with the fractional difference equal to the fractional order of integration of the series under consideration. Only when the series are fractionally differenced, the short term dependency can be unbiasedly estimated. We estimate simultaneously long and short term dependency using the stochastic Autoregressive Fractionally Integrated Moving Average ARFIMA(p,d,q) class of models, which has been shown to attain higher forecast accuracy compared to the corresponding simply differenced models.

Keywords: statistics, time series analysis, stochastic, stationarity, integration, fractional, difference, estimation, stock index, return

TABLE OF CONTENTS

1. Introduction

A significant amount of time series research has recently concentrated on long memory along with fat tailed distributions, chaotic behaviour, and other forms of nonlinearities. These types of time series models are inclined for extraction of additional useful information, since the more traditional techniques do not account for such 'anomalies' in a straightforward manner.

In this master's thesis, we apply fractional differencing methods to test for fractional noises, or long memory, in both time and frequency domain for three stock market indexes and a short term interest rate series. The final model includes parameters for both short and long range dependence. The returns from daily closing prices of Standard & Poors 500, Helsinki Stock Exchange HEX 50, and Shanghai Stock Exchange SHX indexes, and the monthly average of US. Federal Funds Rate are considered for periods from five to forty-one years. A brief description of the model follows in the next three paragraphs. In the rest of the introductory chapter, we characterize the research problem more throughly, and review some recent research papers. Chapter two considers statistical characteristics of standard types of stationary time series processes, and some special characteristics of economic and financial time series, and shortly reviews spectral analysis. The connection between noise stationary series and long memory is illustrated. The third chapter describes the frequency (spectral) and time domain estimators and tests. Estimation results showing significant long memory effects follow in chapter four.

We look at the univariate model $x = f(t,u;\theta)$ where x is the time series realization, f is the specified functional form including both the possible mean and trend functions and the function for their residual, t is time observed at regular intervals, u is the remainder error term, and θ is the parameter vector to be estimated. For f and θ, we are interested in the Autoregressive Fractionally Integrated Moving Average ARFIMA(p,d,q) class and its difference parameter d. Long memory, the significant long-run autocorrelations, is accounted for by the fractional d. d is estimated using a binomial series expansion, or spectral regression of power on frequency on a frequency band which includes only long wavelengths. The decay of the theoretical autocorrelations can be hyperbolic in the ARFIMA class, it is not restricted to the more quickly fading exponential decay of the widely used ARIMA class with integer order of integration. For fractional d, the order of integration is non-integer and allows for variance that does not necessarily scale with the square root of time. u are assumed to form a white noise sequence, expected to be more pure than the sequences resulting from the ARIMA class due to extraction of the additional information content. t is equally spaced time, including only business days.

We consider fractional Brownian motion, henceforth fBm, as the fundamental fractional noise process. fBm is a direct generalization of standard Brownian motion, which occurs when $d = 0$. The process is persistent with positive long-run correlation for $d > 0$, and antipersistent with negative long-run correlation for $d < 0$. fBm needs to be differenced d times to attain stationarity. There are many other fractional noise processes, e.g. a generalized fBm where the difference is defined as $(1-2uB+B^2)^d$, where also u is a parameter, to account for regular sine fluctuation superimposed on the hyperbolic autocorrelation of the basic fBm. Recently, fBm has become a basis for an increasing number of derived processes.

fBm has peculiar dimensional characteristics. Where a regular Brownian motion meandering in a plane would eventually fill the plane densely, a persistent or antipersistent fBm would leave considerable 'holes' in the coverage. This results in non-integer, or fractal dimensions for fBm processes. The determination of d and the approximation of the fractal dimension D can be viewed as equal problems in this light. While not essential for determination of d, the dimensional concept is outlined in chapter two to provide further insight to the properties of fBm and the test statistics derived from dimensional considerations.

1.1 Level, Difference, Or Noise Stationarity

It is mandatory to reach a 'high degree' of stationarity in the course of statistical modelling of economic or financial time series. The stochastic mechanism which creates the sequence of values is unchanging in a stationary series. When the observed time series does not statistically depart from its proposed governing rules, its statistical properties can be inferred with the appropriate confidence. In this section, we describe different forms of stationarity corresponding to certain typical time series characteristics.

Following Mills (1996), we consider an empirically observed time series, a particular realization of a stochastic process. The process is assumed to be ergodic, i.e. the sample moments of long enough but finite stretches of the particular realization approach their population values, when the length of the realization approaches infinity. It is possible to deduce the unknown parameters of a probability distribution from a single realization only if the process is ergodic, but it is not possible to test the validity of the ergodicity assumption from a single realization. It is usually sufficient to define the index set of the stochastic process to be the length of the realization, (1,*T*), so that the probability space consists of a *T*-dimensional distribution of random variables X_t, $t=1,\ldots T$, from which the realization x_t is drawn.

It must noted that this *T*-dimensional distribution tends to reduce to a drastically lower dimension due to the statistical dependencies in the series. Indeed, each fragment of a continuous series has to start from the coordinates of the previous one. With the stationarity assumptions assigning similar characteristics to the *T* dimensions, the dimensionality of the observable path in practice is finally reduced to more than one but less than a few.

The ideal strict stationarity assumes that the rules governing the stochastic process are such that all sets of joint probability distributions are the same regardless of which time point is considered the starting point of the realization. It applies to all moments, which are assumed to exist to as high an order as necessary. As a complete characterization of all these moments would require $T + T(T-1)/2$ parameters, further simplification is necessary. Hence, the rules are assumed to produce constant mean and variance, and autocovariances and autocorrelations dependent only on the time difference of the bivariate distributions in question. If the T-dimensional distribution is jointly normal, or if the current value is generated by a linear combination of previous values of the process itself and possibly the current and past values of other related processes, the expectations of the first and second moments characterize the process completely. Then, the stationarity concept is called covariance stationarity, weak stationarity, or second-order stationarity.

While many regular phenomena especially in the natural sciences can be adequately represented by covariance stationary linear or linearized processes, there are some typical features frequent in economic and financial time series requiring other forms of representations; various forms of apparent nonlinear trends, near cyclical behaviour, and sudden periods of large fluctuations both in mean and variance among others. Over a short time span, a series may be stationary around a linear approximation. Often, long time periods have to be considered, and the nonstationarities become evident.

Some commonly used methods of reducing such nonstationary characteristics are mostly supported by particular knowledge of the series under inspection, and judgement over which characteristics to include or exclude is necessary. For example, application of the logarithmic transformation, a specific treatment of statistical outliers, and estimation and removal of a polynomial and/or trigonometric time trend or a combination of these procedures can render the residual covariance stationary. We call this, informally, *level stationarity*. If nonstationarities remain, or as an alternative altogether, it is possible to test if the series is integrated of first or rarely of second degree, denoted I(d) with $d = 1$ or $d = 2$, by considering the appropriately differenced series instead. The logarithmic transformation and statistical outlier treatment are applied when necessary. The process is then *difference stationary*.

However, assuming that a series is difference stationary, i.e. has a unit root, may not be the most realistic possibility. Nearly always second differencing and sometimes even first differencing increases the estimated overall variance, or causes the spectral density to approach zero near the zero frequency, or both. These are signs of overdifferencing. In addition, the memory of a series with a unit root is not infinite with distinct characteristics differentiating it from other series. Instead, the series is the running sum of the past shocks without any damping factors. Thus, the effect of a shock at a time point is retained as is, or independently, to all following periods and consequently all autocorrelations approach one as T approaches infinity.

The process behind the unit root is Brownian motion. Brownian motion fills the plane densely, and its sample paths are continuous. However, the prices in markets are discontinuous and may jump over some values especially when expectations change, so a process which does not fill the plane as uniformly as Brownian motion could have advantages. If a subset of the plane filled by the process is regular enough, it is possible to determine the fractal dimension and also the fractional difference precisely related to that subset. Other drawbacks for Brownian motion include the behaviour of the sample variance of observed economic or financial series. When different subsamples are considered, they may have variances of different orders of magnitude, often due to a few outliers. Often the variance does not stabilize in the long run. The power spectrum suggests that the variance is extremely large. If a flat spectrum can be attained by taking the difference, the unit root assumption may hold. Often the differenced series still shows somewhat higher or lower power in the low frequencies compared to the high ones, and in this case it is clear that some other process would better represent the series.

Unit root tests such as Dickey-Fuller consider the hypothesis of difference stationarity versus level or trend stationarity, preferably extended to allow for short-run serial correlation which considerably relaxes the strict independency of total randomness. Still, these tests are not completely satisfactory as they may behave suspiciously with regard to the time span of the data. It has been observed that more frequent data does not increase the power of the test much, but a longer time span does (Mills, 1996). This ought not to occur in a level or difference stationary process. Such time dependency could be well explained by a long memory process, not by an infinite or nonexistent memory process.

Another related issue is the thick tails and a sharp peak at the mean, frequently observed in distributions of many financial and in some economic series. The basic normally distributed white noise process and its integrated random walk trail do not have such characteristics. It is natural that the cause for the departures from normality may depend not only on another form of distribution such as Student's t or a stable distribution applying at each time point regardless of the specified functional form of the levels or differences, but also on the endogenous dynamics of the process. Fractional Brownian motion can produce also non-normal trails even though the increments can be conveniently interpreted as white noise, which could be normally distributed. Other distributions of the increments are not considered in this paper, although the R/S statistic and the spectral estimators in section four can accommodate various forms of distributions of the observed time series, as they do not require particular distributional assumptions for the observed series in order to estimate the parameters; the maximum likelihood (ML) estimator is also able to estimate the parameters with slightly more involved theoretical difficulties. Slight biases can be shown to occur as a side effect of these properties.

In the frequency domain, different time distances are transformed to different wavelengths, and periodic time dependencies are visible in an uneven distribution of power among the wavelengths in the frequency spectrum. White noise has even power in all wavelengths, whereas pink noise has more power in the short wavelengths and black noise in the long wavelengths. Various methods have been devised to 'denoise' a process into a white noise from pink or black noise. The process is then *noise stationary*. When the removal of a seasonal peak over some wavelengths, or the corresponding more common time-domain procedure produce stationarity, the process is seasonally stationary, which is in a way a subset of other forms of stationarity. An overview of time-evolving spectrums, another alternative for a noise stationary process, is given in Priestley (1989).

It is notable that the stochastic trend can be much 'more' stochastic compared to integer differencing. Parameters for deterministic trends may still be necessary, if the stochastic trend estimated by the fractional differencing parameter is not sufficient to make the series stationary, given this type of model is needed for some purpose. In addition, more detailed statistical information around the deterministic trend can be obtained. Seasonal differencing in addition to fractional differencing may be necessary if the scaling of the long memory does not coincide with seasonals.

The noise stationarity associated with a fractionally differenced series is necessarily of a narrower character than the stationarity associated with a level or difference stationary series, especially when considering comparison of more than one series. Not only an integer 0, 1, or 2, the order of integration of the series and their cointegration may be non-integer. The simultaneous use of such economic or financial series demands special attention to avoid misjudgment. Fractionally cointegrated processes can be conveniently analyzed in the frequency domain, but they are beyond the scope of this paper.

1.2 Long Memory And Fractional Noise

We consider a long memory process to have hyperbolically decaying autocorrelations or impulse response weights, or equally, hyperbolically shaped logarithmic power spectrum in which case the concept is called fractional noise. The effect of significant hyperbolic autocorrelations over long time distances allows for a broader range of behaviour by compressing or boosting the effects of the conventional exponentially decaying short-run autocorrelations. The hyperbolic decay of fBm and the fractional difference is well in line with the typical spectral shape of an economic variable observed e.g. in Granger (1966). After trends in mean and seasonal components are removed, the logarithmic power spectrum appears like a hyperbola. In many cases, the overall hyperbolic shape is quite clear even with trends and seasonalities included, and the definition of long memory can be extended to cover such processes. Furthermore, long memory captures mean reverting type behaviour, although it is not necessary to assume a stable mean for long memory. Over long time horizons, mean reversion generally occurs in stock returns and many other financial and economic time series, whereas over short time horizons they are positively correlated. Short memory processes are generally modelled with lags over the positively correlated part only.

For a discrete time series process x_t with autocorrelation function ρ_j, a general definition of long memory is

$$\lim_{n\to\infty}\sum_{j=-n}^{n}\left|\rho_j\right|$$

be divergent. In terms of the spectral density, this corresponds to unbounded low frequencies. The more specific definition is that all processes are long memory if their autocovariance function γ_k is of the following type for k large:

$$\gamma_k \approx \Xi(k)k^{2H-2}$$

where $\Xi(k)$ is some slowly varying function at infinity and H is the Hurst exponent in the range $0 \le H \le 1$. A slowly varying function becomes asymptotically a constant; a common one is the logarithm. Hyperbolic distribution has $\Pr(U > 0) = \infty$, but it often suffices to assume that the variance is at worst very large instead of infinite.

In contrast, short memory processes have finite variance $\sigma^2 = \lim_{T\to\infty} E\left[T^{-1}\sum_{t=1}^{T} x_t^2\right]$, and

$$B_T(r) = \frac{1}{\sigma\sqrt{T}}\sum_{t=1}^{[rT]} x_t \xrightarrow{d} B(r)$$

the normalized and scaled partial sum $B_T(r)$ of x_t, up to the integer part of rT, asymptotically converges in distribution to standard Brownian motion $B(r)$ for all $r \in [0,1]$. This also illustrates the property that for all time periods at all starting points, scaling a short memory, white noise series by the square root of time retains constant variance. For long memory processes, the scaling factor will be different from the square root in order to retain similarity for all r.

Additionally, Mandelbrot (1982) claims that the hyperbolic distribution has advantages for data quality considerations. In many cases, data is composed through various collection and preparation methods, and often several steps of estimations or are performed. It is possible that such data may have the underlying true distribution and an unspecified filter on top. Asymptotically hyperbolic distributions are very robust with respect to a variety of filters. Conversely, scaling invariance demands an asymptotically hyperbolic distribution for many common transformations.

Another possibility for long memory is a deterministically chaotic process, where a strange attractor is generating the data. This type of behaviour is often observed in nature, and it may not be uncharacteristic of events at a trading floor. As all points of a deterministic process are correlated, long memory will be observable in the realization, although the series may appear essentially unpredictable. In this case, predictions would be have an asymptotic sense only, since already in the short run the process itself loses predictability quickly. Different locations on the attractor may imply very different short run realizations, and even very small measurement errors are critical. Furthermore, empirical measurements are ineffective if the length of the attractor cannot be clarified. In practice, it is difficult to be certain whether the attractor is strictly strange, but the approximate case remains valid due to the shadowing lemma, which states that within a small neighbourhood ε from the observed series, there will be an exact orbit with the same statistical properties. While the shadowing lemma provides justification for inference from a single realization, it allows the determination of certain dimensional characteristics such as the Lyapunov exponents and even the rescaled range statistic, which aid in assessing the validity of long memory assumptions and give information on the appropriate lengths of the time intervals used for parameter estimation. The largest Lyapunov exponent provides an estimate of the average information loss in forecasts, but it is currently not a readily estimable statistic.

1.3 A Fractal Market Hypothesis

Peters (1996) hypothesized that the stock market has several intrinsic time scales in the form of different investment horizons, corresponding to the scaling properties behind fractional Brownian motion or a multifractal process. When investors have different time horizons, they keep the market liquid by definition: as a three-hour investor observes a high-risk event at the scale of, say, 6σ, the same event is minuscule to a three-year investor who determines the event scale with respect to the average three-year market and/or stock volatility which is of different order than the three-hour volatility. If the three-hour investor is forced to liquidate, the three-year investor will buy the shares if the drop has been large enough to expect gains during the longer time scale. Only if all investors have short-run time horizons as a result of lost confidence in the long-run information, the market loses liquidity and may experience large erratic movements. A complete characterization of the different horizons may require a time-evolving multifractal structure, but the average scaling in time detected by fractional noise may be a reasonable approximation.

The same principle could be generalized to other securities and markets as well. Furthermore, autocorrelation is frequent in market prices and in economic fundamentals, and can be expected to be observable in individual agent's behaviour as well. Several authors have showed that independent autoregressive processes produce a fractionally integrated aggregate process, for example Granger (1980). Linden (1999) constructs a ARFIMA-related aggregated process from AR(1)-processes with uniform distribution. Thus, fractional processes may be expected to be a rule rather than the exception due to the heterogeneity of the background forces driving the observed processes.

1.4 Recent Research On Fractional Differencing

A selection of the growing literature on fractional differencing is shortly reviewed here without explicit clarifications of the terminology; these follow in the next chapters. Brock and de Lima (1996) give an overview of theoretical and practical findings on nonlinear modelling including long memory. Baillie (1996) throughly reviews fractional differencing.

There exist several alternative estimators for the fractional difference. Li and McLeod (1986) consider aspects of ML estimation for fractional differencing. Fox and Taqqu (1986) develop a Whittle-type estimator for approximate MLE to avoid the increase in computation time from second derivative checks in the exact MLE. This estimator and its variants approximate the ML reasonably well (Taqqu and Teverovsky, 1998). Baillie and Chung (1993) present a conditional sum of squares estimator which approximates the MLE well in moderate and large samples. Chen et al (1994) estimate the fractional differencing parameter by various standard lag windows, and note a tendency for smaller mean square error but a larger bias compared to frequency band of the periodogram used by Geweke and Porter-Hudak (1983). Cheung and Diebold (1994) compare approximate frequency domain ML to Sowell's exact time-domain ML and suggest that in practical application with the presence of unknown mean, the efficiency of the frequency domain ML is good especially in medium or large sample sizes, while the efficiency of time domain ML with estimated mean is still somewhat higher.

Taqqu and Teverovsky (1997) examine cases of misleading estimates from rescaled range analysis due to a jump in the mean or a slow trend; they find that a method with differenced variance is able to correct several situations. In addition, they find long memory in the number of bytes and packets transmitted during 10 ms intervals over an Ethernet monitoring system. Taqqu and Teverovsky (1998) examine the robustness of several estimators with simulations. Some estimators are similar to the R/S analysis but concerned mainly with the variance instead of range and variance. Also, various stable distributions with infinite variance are implemented. They find variance-type estimators and Whittle frequency domain estimators to be relatively robust with respect to deviations from Gaussian series. In addition, non-zero AR and MA components are found to have strong effect on all estimators, requiring correct specification. Chong and Lui (1999) study the asymptotic bias of an estimator based on the partial autocorrelation function finding that the estimator is not attenuated in usual cases of measurement error. Koop et al (1997) discuss Bayesian analysis with ARFIMA models.

Robinson (1995) considers multivariate frequency domain regression. Robinson and Hidalgo (1997) derive central limit theorems, asymptotic normality and $n^{1/2}$-consistency for multivariate frequency domain GLS, and consider FGLS and NLS as well. A limit theory for a non-Gaussian quasi-maximum likelihood is derived in Hosoya (1997). Martin and Wilkins (1999) develop a framework for indirect estimation of ARFIMA and vector ARFIMA (VARFIMA) models.

Gil-Alana and Robinson (1997) apply Robinson's tests allowing fractional integration to an extended macroeconomic data set, the predecessor of which was used in the Nelson and Plosser (1982) paper suggesting the I(1) hypothesis for macroeconomic series instead of the deterministic trend hypothesis prevailing at the time. Gil-Alana and Robinson find that the I(1) hypothesis is neither adequate, but different I(d) alternatives depending on the series and distributional assumptions give better results. They conclude that consumer prices and money stock are the most nonstationary series, followed by GDP deflator and wages, and unemployment rate followed by industrial production are closest to stationarity. Diebold and Rudebusch (1989) found long memory in real US per capita GNP. Delgado and Robinson (1994) apply various estimators to 1939-1991 Spanish monthly general price index and find it to be integrated of order 1.3-1.4. Chambers (1998) finds long memory in UK macroeconomic series, and also notes some differences in the performance of fully parametrized and semi-parametrized periodogram estimators. Using various ARFIMA-GARCH models, Mazaheri (1999) finds the implied convenience yield for petroleum and petroleum products to be driven by a mean reverting long memory process.

Blasco and Santamaria (1996) find some evidence of long memory in the Spanish stock market. Barkoulas et al (2000) find long memory in the Greek stock market in weekly returns index, and also note that for large stock markets there has been little or no empirical evidence of long memory. In other mimeos, Baum and Barkoulas find that weekly returns do not show long memory, whereas daily returns of the same period may show long memory. Slight long memory is found in the Canadian stock market by Beveridge and Oickle (1997).

Baillie and Bollerslev (1994) find the forward exchange premium of five major currencies to be well described by a long memory process. Hauser et al (1994) find little evidence of long memory in seven major exchange rate series. Significant GARCH(1,1) effects with parametres summing to one are found in the variance. Hauser et al conduct also Monte Carlo experiments to investigate the determination of long memory with simultaneous conditional heteroskedasticity. They find that heteroscedastic series may be erroneusly interpreted as long-term dependent, and vice-versa to a smaller extent. Some tables for critical values for the distribution of d are given, showing higher departures from the asymptotic distribution of a white noise series as the GARCH parameter values increase. Breidt et al (1998) find long memory in stock market volatility using a long memory stochastic volatility (LMSV) model based on incorporating ARFIMA in a standard stochastic volatility model. A further recent development to this direction is the combined ARFIMA-FIGARCH model.

Seasonal fractional integration is considered in Porter-Hudak (1990) and Silvapulle (1995). Some evidence for improvement compared to the usual seasonal ARMA is obtained.

Soofi (1998) tests for fractional cointegration in purchasing power parity among 9 OPEC countries, using the GPH estimator to measure fractional integration of an Engle-Granger type cointegrating vector. For four countries, monthly exchange rates and CPI are found fractionally cointegrated. Tse (1998) tests the effect of GARCH on fractional cointegration with simulations.

Gray et al (1989) (with a 1994 correction) extend the fractionally differenced process $(1-B)^d$ to a Gegenbauer process $(1-2uB+B^2)^\lambda$ allowing explicit presentation of periodic or quasi-periodic long-term dependencies through various forms of sinusoidal fluctuations in the autocorrelation function. In the spectrum, there is a peak at the Gegenbauer frequency. Chung (1996) conducts Monte Carlo experiments on Gegenbauer ARMA (GARMA) and ARFIMA series, and detects long memory and persistent 9-year cycles in the US Wholesale price index with a GARMA model.

Sutcliffe (1994) obtains better forecasting results with ARFIMA compared to ARIMA using e.g. the airline data from Box and Jenkins (1976).

2. Fractional Brownian Motion Processes

2.1 Deviations From White Noise And Martingales

The Wold decomposition, a fundamental theorem in time series analysis, states that every weakly stationary stochastic process can be expressed as a stationary noise process through a linear combination, or a linear filter, of a sequence of uncorrelated random variables. The fundamental noise process is the white noise, a sequence of uncorrelated random variables a_t with distribution properties

$$
\begin{aligned}
&\mathrm{E}(a_t) = 0,\\
&\mathrm{Var}(a_t) = \mathrm{E}(a_t^2) = \sigma^2 < \infty,\\
&\mathrm{Cov}(a_t, a_{t-k}) = \mathrm{E}(a_t a_{t-k}) = 0 \ \forall \ k \neq 0.
\end{aligned}
$$

The linear filter representation is $x_t = a_t + \phi_1 a_{t-1} + \phi_2 a_{t-2} + \ldots$, $\sum_{j=0}^{\infty} \phi_j^2 < \infty$. The sum of the squared ϕ-weights is assumed to be bounded, which is in this case equal to assuming that x_t is second-order stationary, the variance is finite, and all moments exist and are independent of time origin. Sometimes the absolute ϕ-weights are required to converge, which is a stronger assumption e.g. for minimizing the absolute mean deviations.

If the aim in modelling a univariate time series is to produce optimum linear forecasts, this objective has been reached when a linear transformation is found that reduces the series to white noise. Such transform is unique for a linear process. When a_t and a_{t-k} are independent for $k \neq 0$, the process is a pure white noise. It cannot be forecast from its own past whereas the non-independent white noise cannot be forecast linearly. If a nonlinear process is reduced to a white noise process through a linear transformation, the transform is not necessarily unique. (Granger and Watson 1984)

It seems likely that the more strictly pure the empirical white noise sequence is, the smaller the number of further transforms capable of achieving an even more pure white-noise sequence. The strictness is not well measurable in our empirical setting especially with the nonlinear functional forms, but less unexplained variance, and less serial correlation are the necessary ingredients as usual. Independency is proxied by correlation in the time domain, and equivalently by power in the frequency domain.

Cramer's extension of the Wold decomposition theorem represents time series through a combination of a measurable trend with the noise process. In the stochastic differential equation

$$dx_t = \mu(x_k; k \leq t)dt + \sigma(x_k; k \leq t)dM_t$$

x_t is a random process endogenously determined firstly by the time-dependent process $\mu(x_k; k \leq t)dt$, the expected 'instantaneous' change in x, if there is need for one. The second term consists of the predictable process $\sigma(x_k; k \leq t)$, the 'instantaneous' standard deviation, which is driven by a martingale *M* analoguously to a forcing function of a dynamical system

$$(\sigma(x) * M)_t = \int_{k=0}^{t} \sigma(x_k) dM_k$$

so that

$$dE(\sigma(x_k) * M)_t = \sigma(x)_t^2 dE(M)_t .$$

Martingale processes allow the sequence of independent random variables to be nonstationary, relaxing the assumption required by a white noise process. The combined process $\sigma(x_k)dM_t$ which produces the unexpected change, is also martingale. (Karr, 1990)

When M is standard Brownian motion, the differential equation represents the Itô process. Fractional Brownian motion can be fitted into the differential equation to more accurately describe the time series. Then, the motion generating noise process is either pink or black if there are short-run or long-run dependencies, respectively. It is more convenient to define the motion generating noise process white as usual, and transfer the statistical deviation from white noise to $\sigma_{fBm}(x_k)$ instead of replacing M with it. The increments still converge to a stationary sequence. The observable trail is nonstationary but not martingale, since the sequence resulting from the changed process $\sigma_{fBm}(x_k)dM_t$ is more persistent or antipersistent than pure white noise.

Mandelbrot (1971) notes that under long-run dependence in mean, perfect arbitraging may not be possible as it would not be difficult to identify promising assets with rescaled range analysis or some other method and deteriorate the effect quickly. In order for long-run dependence to remain, the process of identifying and arbitraging should then create further long memory characteristics which remain similar throughout the time series path along with the fundaments such as business cycles.

2.2 Spectral Representation of a Stochastic Process

In spectral, or harmonic analysis, a periodic function is analyzed with its Fourier series, and a nonperiodic function with the analoguous Fourier integral. The Fourier series of a function $f(x)$ is defined as $\frac{1}{2}a_o + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x) + \ldots$. $f(x)$ is required to be continuous except for a finite number of discontinuities and to have a finite number of maxima and minima. The real and complex Fourier representations can be obtained and manipulated via the identities $\sin(a+b) = \sin a \cos b + \cos a \sin b$ and $\sin c = (e^{ic} - e^{-ic})/2i$.

The argument of a time series, assumed equally spaced, is transferred to the circular range $[0,2\pi]$. For convenience, we assume the series to have odd length. The spectral decomposition of the series can then be obtained by composition of a number of harmonic functions $A_\lambda \sin(\lambda t + \varphi_\lambda)$, where amplitude A_λ is the maximum displacement at a frequency, phase φ is the displacement of the zero point of the wave along the time scale, and λ is some angular frequency defined in radians as $\lambda = 2\pi f_\lambda$. The frequency f_λ measures cycles per unit time, $f_\lambda = 1/\tau_\lambda$. The period, τ_λ, is the number of time units required for a complete cycle in radians. (Koopmans 1974)

Henceforth, we refer to λ as frequency and drop the subscripts when they are not essential. The time scale can be expressed in radians as $\vartheta = \lambda t + \varphi$, so for $t = 1,...,T$, a complete period τ can occur fT times. Taking $\varphi = 0$, each point t of the time series, τ units apart, is traversed by the harmonic $\sin(2\pi t/\tau)$. The frequencies $2\pi t/\tau$ are called, accordingly, the harmonic frequencies. It can be seen that the points traversed by shorter-wavelength harmonics $2\lambda, 3\lambda, 4\lambda, \ldots$ also belong to a longer 'base' wavelength λ.

In the spectral representation of a continuous stochastic process, there is an infinite number of harmonic frequencies, which have normally distributed, uncorrelated amplitude functions with constant common variance, also continuous. As the spectrum is symmetric with respect to the zero frequency in the same manner as autocorrelations, it suffices to restrict attention to $[0, \pi]$. The spectral representation (Cramer's representation) is (Harvey, 1982)

$$x_t = \int_0^\pi u(\lambda)\cos\lambda t\, d\lambda + \int_0^\pi v(\lambda)\sin\lambda t\, d\lambda .$$

Cycles longer than π are aliased to frequencies with a shorter period, possibly creating some inaccuracies in the harmonic functions. In time domain estimation, the possible inaccuracy from significant autocorrelations longer than π units apart becomes decomposed in a different manner. Equally, the series cannot be sampled more accurately than the unit time interval, and any shorter cycles, should they exist, are also aliased.

In practice, the spectral density is obtained via the discrete Fourier transform. The series is represented as a linear function of a finite number of harmonic frequencies so that all points in the series are traversed. Fast Fourier transform algorithms are commonly used to reduce computing times. The discrete amplitude parametres are the Fourier coefficients. The energy, or 'covariance', at each frequency can be conveniently shown with the squared Fourier coefficients as $p(\lambda) = u(\lambda)^2 + v(\lambda)^2$. (Harvey 1982)

The theoretical autocovariance function $\gamma(\tau)$ can be decomposed into the spectral density $f(\lambda)$ as

$$f(\lambda) = \frac{1}{2\pi}\left(\gamma_0 + 2\sum_{\tau=1}^{\infty}\gamma(\tau)\cos\lambda\tau\right)$$

assuming γ_0 is finite. If x_t is white noise, $f(\lambda) = \sigma^2 / 2\pi$. The convenient complex spectral density is $f(\lambda) = (2\pi)^{-1}\sum_{\tau=-\infty}^{\infty}\gamma(\tau)e^{i\lambda\tau}$; for each λ, $e^{i\lambda\tau} = \cos(\lambda\tau) + i\sin(\lambda\tau)$ represents a harmonic oscillation in τ. $\gamma(\tau)$ can be obtained from $f(\lambda)$ as $\gamma(\tau) = \int_{-\pi}^{\pi} e^{-i\lambda\tau} f(\lambda)d\lambda$, or $\gamma(\tau) = 2\int_{0}^{\pi} f(\lambda)\cos\lambda\tau\, d\lambda$. (Harvey 1982)

In addition to sampling, the estimation of the harmonics is delicate. The periodogram obtained from the squared Fourier coefficients is the most straightforward estimator, but different windows, frequency band shapes, or an averaged band are generally considered necessary for better estimators (Koopmans 1974). Some leakage from one frequency to another is unavoidable due to side lobes and other such features resulting from operating with superimposed trigonometric functions. "Care has to be taken in estimating the spectrum since it seems that it is impossible to have an unbiased and consistent estimate" (Granger, 1990), which seems a natural property of wave decomposition.

As the periodogram is a linear combination of the autocovariances, the sum over all frequencies is equal to the variance of x_t when it is defined. This is more clear in the notation

$$I(\lambda_j) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n}\hat{\gamma}(t)e^{i\lambda_j t}\right|^2 = \frac{1}{2\pi n}\left|\sum_{t=1}^{n}(x_t - \bar{x})e^{i\lambda_j t}\right|^2$$

(Kanto, 1983). The periodogram is an unbiased estimator of the spectral density, since their ratio is $\sim \chi^2$. The variance of the periodogram equals $f^2(\lambda_j)$ for $j = 1,\ldots,n$, $\lambda \neq \pi$, and $2f^2(\pi)$, $\lambda = \pi$, not dependent on *T*, so the periodogram is not consistent. The properties are asymptotic for a series which is not white noise. The frequency domain estimators aim to achieve better consistency by utilizing an appropriately chosen bandwith which slowly decays to zero as *T* approaches infinity.

Between time and frequency domain, there is a one-to-one correspondence in terms of the time distances and wavelengths. Usually, both approaches yield similar results but the other may be better suited for certain applications due to its computational or theoretical properties.

2.3 Dimensional Properties of Fractional Noise

This section discusses dimensional interpretations of some estimators and tests, providing some additional insights to the characteristics of fBm. As fBm adjusts itself for the scaling properties of the series under consideration, it introduces a modification in the coordinates, which causes fBm to have a fractal dimension distinct from the fractal dimension of the standard Brownian motion.

When attempting a statistical measurement, it is essential that the frame of reference, roughly the dimension where the object is to be measured, is orthogonal enough. For example, the angles of a triangle on a plane always sum to 180°. Moving one integer dimension up to the surface of a sphere, the angles of a triangle sum to more than 180°

and up to 270° depending on the relative scale of the triangle and the sphere. Then it becomes necessary to ask, which frame of reference is meaningful for measurement without prior knowledge, if orthogonality depends on the scale of the frame? If the triangle itself is a statistical object and its orthogonal properties are to be measured, how is it possible to differentiate between the object and its frame of reference at all? One answer is the scaling dependence, which has been empirically shown to occur in many natural phenomena. When the orthogonality of statistical measurements on the object at different scales within the frame is retained, the highest dimensional one of such frames is the correct one. Viewed this way, it is not possible to separate the object from its frame of reference, since the frame is inherent in the stable scales of the object.

There are several possible dimensional definitions and associated transformations which concentrate on different aspects of the process to be measured. Whichever is the most suitable for the problem at hand, could be in practice designated as the effective dimension of an object. If the dimension is non-integer, it is called a fractal dimension. The concept of a fractal and the fractal dimension have not been exactly defined (e.g. Mandelbrot 1997). The elementary ingredients of a fractal object are one or more repeating patterns, which appear similar when magnification is increased or decreased by a suitable amount. Often the surface of a fractal object is not smooth.

For example, a wiffle ball is not a complete three-dimensional ball since it is full of holes. It does not occupy the whole of the volume of its circumference, preventing other objects from interference within its circumference. On the other hand, it cannot be completely represented in two dimensions either since it is a ball. Thus, its effective dimension is a fractal dimension between two and three. The embedding dimension is three; it is the space which is spanned by the circumference, the next higher integer dimension from the effective dimension.

In this paper, we hold a working definition which roughly classifies a time series a fractal if it is not level stationary, and the relative strengths of the autocorrelations are invariant with respect to a contraction or expansion of the series by a certain time scale. This definition assumes the series is a unifractal where scaling properties remain similar throughout all resolutions. It encompasses both fBm and standard Brownian motion with short memory such as an AR(1)-model. Cases where the time series realization is a multifractal, having several fractal realizations or differently repeating patterns superimposed, or different stationary processes superimposed, are not considered.

A traditional mathematical approach to evaluating the area of a planar shape is to cover the shape with small squares arranged along planar coordinates, and then calculating the sum of the sides of the squares raised to the power $D = 2$. Following Cantor and Minkowski, Carathéodory extended this approach and the ideas of 'length' or 'area' in 1914. The reliance on the planar coordinate axis should be avoided, since it already implies that the planar shape is two-dimensional. The shape can instead be imbedded in a three-dimensional space, and covered by balls. Then the approximate contents or 'volume' of the object can be computed by adding the circular shapes together. If the shape is planar, the area would be $\sum \pi\rho^2$ since the balls reduce to discs. In the general case with a d-dimensional shape, the hypervolumes of d-dimensional balls of radius ρ are computed as $\sum \pi(d)\rho^d$, where $\pi(d) = \frac{[\Gamma(\frac{1}{2})]^d}{\Gamma(1+d/2)}$ is the volume of a ball with unit radius. (Mandelbrot 1982)

The 1919 Hausdorff dimension forms a theoretical basis for dimensional calculations in non-integer dimensions. There is a number of practical applications of the Hausdorff dimension. Unless otherwise noted, the following descriptions draw on Peitgen et al (1992), which includes theoretical discussions as well.

The pointwise dimension, or mass dimension, is approximated by Hölder exponent α_H. α_H is applied in R/S analysis as the Hurst exponent (Mandelbrot 1997). Disks $B_r(x, y)$ of various radiuses r are centred on a point (x, y) on the object to be measured. The probability of visits of the attractor (the trail of the process) to the disk, $\mu(B_r(x, y))$, is taken to be proportional to a power of the radius, $\mu(B_r(x, y)) \propto c \cdot r^{\alpha_H}$, so that the Hölder exponent can be computed as

$$\alpha_H = \lim_{r \to 0} \frac{\ln \mu(B_r(x,y))}{\ln r}.$$

If the same exponent holds for all points, the object is likely to be unifractal. If the exponent varies for different points, it is multifractal.

The correlation dimension measures the correlation between points on the attractor. For a single observed series of length T, an m-history with a sufficiently large number of lags $(m > 2T+1)$ is used to represent the series in an embedding dimension which includes all dimensions of the attractor. The (Euclidean) distance of two points X_i^m and X_j^m is compared to a predefined distance ε in order to see if the two points are 'close' or 'correlated'. If the distance between the two points is smaller than ε, the function $H(u)$ is set to 1, otherwise to 0 in the correlation integral

$$C_{m,T}(\varepsilon) = \frac{1}{T^2 - T} \sum_{i=1}^{T} \sum_{j=1}^{T} H\left(\varepsilon - \left\| X_i^m - X_j^m \right\|\right), \qquad i \neq j$$
$$C_m(\varepsilon) = \lim_{T \to \infty} C_{m,T}(\varepsilon)$$

(Brock et al 1996). In practice, after computing the results for a range of m and ε, it is possible to obtain the scalar correlation dimension by first determining the slope of the regression of $\ln C_m(\varepsilon)$ on $\ln \varepsilon$ for each sufficiently large m. As m increases, the slope will converge to a constant (in case of most fractals) which is equal to the correlation dimension D_C. (Creedy and Martin 1994)

Lyapunov exponents are measures of average attraction and average repulsion to all possible directions over the realization of a process. They can be used to distinguish a deterministically chaotic process from a stochastic process; a positive first Lyapunov exponent indicates that nearby orbits are moving apart which is a characteristic of a chaotic attractor. A negative first exponent would imply that nearby orbits converge and the process is stabile.

There exist as many Lyapunov exponents as there are dimensions in the object. The Lyapunov exponents are ordered so that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots$. The sum of all Lyapunov exponents is negative, characterizing how fast area, volume, or hypervolume shrinks towards the attractor. As $x_{n+1} = f(x_n)$ is iterated with two nearby starting points x_0 and u_0 with an initial error $E_0 = x_0 - u_0$, after n iterations the error will be $E_n = u_n - x_n = c^n(x_0 + E_0) - c^n x_0 = c^n E_0$, and the respective error ratio is defined as

$$c^n = \left|\frac{E_n}{E_0}\right|.$$

The error ratio can be written as a product $\frac{E_n}{E_0} = \frac{E_n}{E_{n-1}} \cdot \frac{E_{n-1}}{E_{n-2}} \cdot \ldots \cdot \frac{E_1}{E_0}$. Taking logarithms, dividing by *n*, and noting that c^n is a derivative of the process,

$$\ln c = \lim_{E_0 \to 0} \frac{1}{n} \sum_{k=1}^{n} \ln\left|\frac{E_k}{E_{k-1}}\right| = \frac{1}{n} \sum_{k=1}^{n} \ln|f'(x_{k-1})|.$$

Letting $n \to \infty$, the Lyapunov exponent is defined as

$$\lambda(x_0) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \ln|f'(x_{k-1})|.$$

A small error in an initial point x_0 will be scaled by the factor e^{λ} on the average in each iteration. The largest Lyapunov exponent measures the average rate of loss of predictive power; if e.g. $\lambda(x_0) = 1$, and the initial condition is measured to the eighth significant decimal, the chaoticity causes all information to be lost in eight iterations (Creedy and Martin 1994). This requires representation of the exponent with base 2 logarithms.

The Lyapunov dimension is defined as

$$D_L = m + \frac{1}{|\lambda_{m+1}|}\sum_{k=1}^{m}\lambda_k, \ \sum_{k=1}^{m}\lambda_k \geq 0 \ ,$$

which is the point where the cumulative sum of the Lyapunov exponents linearly interpolated to form a continuous line is equal to 0. Lyapunov exponents can theoretically be computed for a high-dimensional, e.g. 10-dimensional object without much change in their rather statistical accuracy, although they are in general not as precise as the other dimensions mentioned here.

2.4 Fractional Brownian Motion And 1/λ Noises

The differences between the scaling properties in trails of standard and various fractional Brownian motions have been summarized already by Leibniz (quote from Mandelbrot 1982):

> "I have diverse definitions for the straight line. The straight line is a curve, any part of which is similar to the whole, and it alone has this property, not only among curves but among sets."

Fractional Brownian motion processes were introduced by Mandelbrot and others in the 1960's, following Hurst's findings on the dependence of the river Nile in the 1950's.

The properties of standard and fractional Brownian motion are asymptotic. Standard Brownian motion *B*(*t*) maps from a linear time scale through a white noise process to the observable Brown trail. A realization of *B*(*t*) produces a sequence *t* of independent normally distributed increments so that

$$\Pr\Big((B(t+\Delta t) - B(t))/|\Delta t|^H < x\Big) \sim N(0,1)$$

where H is fixed to ½, which is 'straight line' with a specific similarity property: at least the first and second moments of its distribution are independent of time when rescaled by the square root of time as $t^{-1/2}B(\Delta t)$, for all time distances Δt (Mandelbrot, 1982). It could be then expected that after Δt, the expected value of the function has changed by $o\left(\sqrt{\Delta t}\right)$. This scaling property is often assumed to hold without checking if the correct rescaling factor is $t^{-1/2}B(\Delta t)$ or some other $t^{-u}B(\Delta t)$.

The process observable in the trail of discrete Brownian motion is also known as the random walk, $x_t = x_{t-1} + \varepsilon_t$. The mean of the random walk is $E(x_t) = E(x_0)$, which satisfies the first requirement of weak stationarity since it is constant over time when x_0 is fixed as usual. The process is non-stationary since the variance is not constant at a finite limit; $Var(x_t) = t\sigma^2$ and $Cov(x_t, x_{t-\tau}) = |t-\tau|\sigma^2$. It is I(1), since $\Delta x_t = x_t - x_{t-1} = \varepsilon_t$, which is the stationary white noise process. Notably, Brownian motion is nowhere differentiable (with a probability of one). The properties hold almost surely almost everywhere.

The spectral representation of a Brownian motion could be called the circle-to-line Brownian motion, or a random Fourier-Brown-Wiener series. White noise has a flat power spectrum, which is not dependent on the frequency λ under consideration. This could be expressed as being proportional to λ^0. The spectrum of the trail of Brownian motion is proportional to λ^{-2}, which is again a special case related to the non-stationarity.

Fractional Brownian motion $B_H(t)$ agrees better with the special characteristics in financial and economic series than the standard Brownian motion, and it can well account for the typical spectral shape. It is defined to have variable inherent scaling character as outlined in the dimensional properties section. The properties given here are from Mandelbrot (1982). Fractional Brownian motion has $0 < H < 1$, and cases with $H \neq ½$ are properly fractional. Now the sample correlations satisfy

$$E\left((B_H(t+\Delta t) - B_H(t))^2\right) = |\Delta t|^{2H}$$

so that the variance scales at faster or slower instead of the same pace as the square root of time; in the discrete case $\sum_{t=1}^{T} x_t$ is $o\left(T^H\right)$ for any $0 < H < 1$. The spectral density is proportional to λ^{-2H-1}. The process is everywhere continuous and nondifferentiable like standard Brownian motion.

The line-to-line fBm can be obtained from the line-to-line standard Brownian motion by application of the Riemann-Liouville fractional integral or differential of order $H - \frac{1}{2}$ as

$$B_H(t) = \frac{1}{\Gamma(H + \frac{1}{2})} \int_{-\infty}^{t} (t-s)^{H-\frac{1}{2}} dB(s)$$

The integral is divergent, but increments such as $B_H(t) - B_H(0)$ converge. The increments are fractional noise when the process is properly fractional. When the order of integration is $H - \frac{1}{2} > 0$, the transform is a fractional form of integration, because it increases the smoothness of the function. When the order is $H - \frac{1}{2} < 0$, irregularity increases which is a characteristic of differentiation. The definition is strongly asymmetric in *t*, and when the process ought to be reversible for stationarity considerations, Mandelbrot (1982) proposes a symmetric definition

$$B_H(t) = \frac{1}{\Gamma(H + \frac{1}{2})} \left(\int_{-\infty}^{t} (t-s)^{H-\frac{1}{2}} dB(s) - \int_{t}^{\infty} |t-s|^{H-\frac{1}{2}} dB(s) \right)$$

In the frequency domain, ordinary integration of a nonperiodic function corresponds to multiplying the Fourier transform by $1/\lambda$. The Fourier transform has to be defined for the nonperiodic function in question, but the representation can safely be assumed to hold for the series we are concerned with. Fractional integro-differentiation multiplies the Fourier coefficients by $(1/\lambda)^{H+\frac{1}{2}}$ so that the spectrum of the fractional Fourier-Brown-Wiener series becomes proportional to $(1/\lambda)^{2(H+\frac{1}{2})} = \lambda^{-2H-1}$. The spectral exponent λ^{-B} is thus $B = 2H - 1, -1 \leq B \leq 1$. In the estimators, the task is to find the proper amount to subtract from or add to the moduli of the Fourier coefficients at low frequencies.

Mandelbrot (1982) suggests a 'most straightforward procedure' to compute the fractional Brown line-to-line function as first to compute the circle-to-line function, and then to discard it except for a small portion corresponding to a small subinterval $0 < t < t^*$ (t^* has to tend to 0 as $H \to 1$), and finally to add a separately computed low frequency component. The specification of the included frequencies is crucial, since changing the arguments of the hyperbolic shape near zero results in very large changes in ordinates, and the area of interest, the next higher frequencies, could be easily outweighed in estimation.

The long-run correlation of fractional Brownian motion does not vanish, contrary to standard Brownian motion. This enables the estimation of a long memory parameter. By setting $B_H(0) = 0$, and defining the past increment as $B_H^-(t)$ and the future increment as $B_H^+(t)$,

$$
\begin{aligned}
&\frac{E\left(B_H^-(t)B_H^+(t)\right)}{E\left((B_H(t))^2\right)} \\
&= \frac{\frac{1}{2}\left(E\left((B_H^+(t) - B_H^-(t))^2\right)\right) - 2E\left((B_H(t))^2\right)}{E\left((B_H(t))^2\right)} \\
&= \frac{(2t)^{2H} - t^{2H}}{2t^{2H}} \\
&= 2^{2H-1} - 1
\end{aligned}
$$

For $H > \frac{1}{2}$ the correlation is positive and the series is persistent (in the case of an I(1) series the correlation equals 1). For $H < \frac{1}{2}$, the series is antipersistent with negative long-run correlation. (Mandelbrot 1997)

The fractal dimension of fractional Brownian motion is $D_{fBm} = 2 - H$ (in terms of the fractional difference parameter, $d = H - ½$, $D_{fBm} = 1½ - d$). In addition to the ability to distinguish processes of different orders of integration, the fractional Brownian motion process is an important simplification since it is invariant with respect to fractal dependencies in the time scale and thus seems not necessarily to introduce error when applied to a process driven by a fractal time scale such as trading time. Standard Brownian motion is not invariant to these considerations. (Mandelbrot 1997)

The circle-to-line fBm converges to a continuous sum for all $H > 0$. Below $H = 0$, or below $D = 1$, both the circle-to-line and line-to-line functions are not defined. Above $H = 1$, or $D = 2$, the sum is differentiable. There is no upper limit for integration on the frequency domain, but the time domain function is only defined up to $H = 1$. It is possible to extend the definition by integer differencing the series under consideration before applying the fractional difference.

2.5 Fractional Difference

The continuous time fractional noise process has to be discretized in order to fit it into observed discrete data. Two modelling approaches exist, the 'actual' fractional noise from fBm and the fractional difference. Fractional difference can be shown to approximately result from fractional Brownian motion processes, and they represent equal processes for $d = H - ½$ (Geweke and Porter-Hudak, 1983).

At first, Mandelbrot (later with Van Ness) defined discrete time fractional noise so that its correlation function is the same as the correlation function of the process of unit increments of fBm, $\Delta B_H(t) = B_H(t) - B_H(t-1)$,

$$\gamma_k = C(½)\gamma_0\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right)$$

for some $C > 0$, $\frac{1}{2} < H < 1$. As $k \to \infty$, $\gamma_k = CH(2H-1)|k|^{2(H-1)}$. The spectral density of this process is due to Jonas (Geweke and Porter-Hudak, 1983)

$$f(\lambda) = \frac{\sigma^2}{(2\pi)^{2H+2}} \Gamma(2H+1)\sin(\pi H) 4\sin^2\frac{\lambda}{2}\sum_{n=-\infty}^{\infty}\left|n+\frac{\lambda}{2\pi}\right|^{-2H-1}$$

Hosking (1981) defines the discrete analog starting from the definition of fractional noise as the $(\frac{1}{2}-H)$th fractional derivative and discretizes the derivative, which results in a simpler model valid for a larger range of *H*. The fractional difference operator Δ^d is defined for any real $d > -1$ by binomial series as

$$\Delta^d = \sum_{k=0}^{\infty}\binom{d}{k}(-B)^k = 1 - dB - \frac{d(1-d)}{2!}B^2 - \frac{d(1-d)(2-d)}{3!}B^3 - \dots$$

The fractionally differenced white noise process is weakly stationary for $d < \frac{1}{2}$ (or $H < 1$, since $d = H - \frac{1}{2}$) and invertible for $d > -\frac{1}{2}$ (or $H > 0$). The infinite order autoregressive representation is

$$x_t = \sum_{k=0}^{\infty}\phi_k x_{t-k} + \varepsilon_t \ ,$$
$$\phi_k = \prod\nolimits_{0\le j\le k}\frac{j-1-d}{j} = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}$$

The infinite moving average representation, or the Wold decomposition, is

$$x_t = \sum_{k=0}^{\infty}\theta_k \varepsilon_k \ ,$$
$$\theta_k = \prod\nolimits_{0\le j\le k}\frac{j-1+d}{j} = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}$$

The cumulative impulse response is $\sum_{j=0}^{\infty}\theta_j$ as usual. Expressions for autocorrelations are

$$\rho_k = \prod_{0 \le j \le k} \frac{j-1+d}{j-d} = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)}$$

The impulse response weights, the infinite autoregressive coefficients, and autocorrelation coefficients decay hyperbolically and monotonically to zero for large k. This can be shown using Stirling's approximation $\Gamma(k+a)/\Gamma(k+b) \approx k^{a-b}$, giving

$$\begin{aligned}
\phi_k &\approx c_1 k^{d-1} \\
\theta_k &\approx c_2 k^{-d-1} \\
\rho_k &\approx c_3 k^{2d-1}
\end{aligned}$$

for some constants c_i. The power spectrum is

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left(2\sin\frac{\lambda}{2}\right)^{-2d} = \frac{\sigma^2}{2\pi}\left(1-e^{-i\lambda}\right)^{-2d}$$

2.6 The Autoregressive Fractionally Integrated Moving Average Process

With the fractional difference at hand, the ARFIMA(p,d,q) process can be defined as

$$\phi(L)(1-L)^d(x_t - \mu) = \theta(L)\varepsilon_t$$

incorporating both short and long range dependence in the same equation. The roots of $\phi(L)$ and $\theta(L)$ lie outside the unit circle and ε_t is white noise as usual. The process is covariance stationary for $-0.5 < d < 0.5$ and mean reverting for $d < 1$. For $d > 0.5$ the process has infinite variance.

For high lags the decay of the autocorrelations is hyperbolic, as the ARMA part of the process is negligible. The impulse response weights can be obtained by first differencing x_t so that

$$(1-L)x_t = (1-L)^{1-d}\frac{\theta(L)}{\phi(L)}\varepsilon_t = A(L)\varepsilon_t$$

The impact of a unit innovation after *k* units of time for x_{t+k} is then $1+\sum_{j=1}^{k} A_j$.

The spectral density function is

$$\begin{aligned} f(\omega) &= \frac{\sigma^2}{2\pi}\left(\frac{\left|\theta(e^{-i\lambda})\right|}{\left|\phi(e^{-i\lambda})\right|}\right)^2 \left|1-e^{-i\lambda}\right|^{-2d} \\ &= \frac{\sigma^2}{2\pi}\left(\frac{\left|\theta(e^{-i\lambda})\right|}{\left|\phi(e^{-i\lambda})\right|}\right)^2 \left(2\left|1-\cos(\lambda)\right|\right)^{-2d} \end{aligned}$$

For low frequencies, $f(\lambda) \approx \frac{\sigma^2}{2\pi}\left(\frac{\theta(1)}{\phi(1)}\right)^2 \lambda^{-2d}$. The spectral density is infinitely differentiable at all frequencies except zero (Baillie 1996). This is a characteristic of a fractal process.

3. Estimators And Tests

We apply two tests, the KPSS and the BDS tests in order to detect nonlinear departures from linearity. The KPSS test was originally devised for unit root testing, but was found to be able to assess more general situations also. A number of other tests for similar purposes exist, but these are widely known and often used in the course of testing for long memory. The R/S statistic is also considered as a test as it is not suitable for very accurate parameter estimation.

Estimation of the largest Lyapunov exponent to test for chaos was attempted with an application of Peters (1996) of the Wolf et al algorithm. The process was slow, and initial results from two series using several parameter values showed no convergence of the highest Lyapunov exponent. If valid for a number of parameter ranges, this result would imply that the series were not chaotic – however, we cannot make such conclusion based on a small number of runs with short series. As Gwilym et al (1999) among others point out, the algorithm is very sensitive to noise in the data, and this drawback is especially troubling with noisy economic and financial data having a much smaller than desirable number of data points. Gwilym et al (1999) apply also a more recent method by Dechert and Gencay, which has better properties with noisy data. The method utilizes a three-layer neural network with user selected number of inputs that correspond to the embedding dimension, and sigmoid activation functions. Gwilym et al extend the method to contain several hidden layers. Jacobian matrixes for each individual node are utilized in computing all Lyapuov exponents. One run of computations with 1-10 layers with a single input took 24 hours of CPU time on a Sun SparcCentre 2000 with eight processors. Armed with a Pentium 233 MMX, we decided not to pursue the Lyapunov computations further.

The estimation of an ARFIMA process can proceed along two lines. Analoguously to regular differencing, the fractional difference can be first estimated with a semiparametric frequency domain estimate, which preferably does not include a full parametrization of the logarithmic periodogram or other global assumptions of the spectral density of the process. The estimated fractional difference is taken by multiplying the Fourier transform of the series x_t by $\left(1-e^{-i\lambda}\right)^d$, the inverse Fourier transform is computed from the resulting series, and the ARMA parameters are estimated from the resulting series. The steps are iterated until convergence. A drawback of this two-stage method is that no published results exist for distributions of the second-stage parameters. Some Monte Carlo studies show that the distributions may be rather different from the conventional ones. With respect to the original series, the R/S statistic is distribution-free, and so are the three frequency domain estimators; in $x = f(t, u; \theta)$, semiparametric methods estimate θ consistently when the distribution of u is unknown but f is specified (Cosslett, 1990). Most of the complicated derivations for expressions of variances for the semiparametric estimators have assumed normality for simplicity, although it is not necessary for all results (e.g. Lobato and Robinson 1996). Semiparametric estimators are not as efficient as a correctly parametrized parametric estimator.

The other possibility is parametric estimation in the time domain, in general with the maximum likelihood estimator or its approximations. However, the accuracy of ML depends on the correct estimation of the mean and the correct order of the AR and MA polynomials, and ML also relies on the normality of the error u. The estimates are still consistent for non-normal distributions.

Smith et al (1997) compare the bias and misspecification in estimating the ARFIMA model with simulations using three estimators used also here; Geweke – Porter-Hudak semiparametric estimator, Robinson's averaged periodogram estimator, and Sowell's maximum likelihood estimator. They find that the ML outperforms the other two estimators in terms of bias and mean square error both in the short-term and long-term parameter estimates. The biases may easily cause misspecification with formal model selection criteria. Even when the misspecification biases of the ML are taken into account, the worst-case scenario for the ML is found to have the smallest bias. In addition, Taqqu and Teverovsky (1998) stress the importance of correct specification with all types of estimators they examine. Thus, we favour the ML estimator with several specifications applied to each series under inspection, and use the other estimators mostly with a purpose to obtain supporting evidence for the ML.

3.1 R/S Statistic

The R/S statistic was originally proposed by Hurst in an 1951 paper when he was studying the fluctuations of the river Nile. The sample sequential range $R(t,s)$ is defined as (Brock and de Lima, 1996)

$$R(t,s) = \max_{0\le k\le s}\left(X^*_{t+k} - \left(X^*_t + \frac{k}{s}\left(X^*_{t+s} - X^*_t\right)\right)\right) - \min_{0\le k\le s}\left(X^*_{t+k} - \left(X^*_t + \frac{k}{s}\left(X^*_{t+s} - X^*_t\right)\right)\right)$$

for a time series X_t and any arbitrary time interval $[t, t+s]$. $X^*_t = \sum_{u=1}^{t} X_u$ with $X^*_0 = 0$.

Each sample range is divided by its standard deviation,

$$S(t,s) = \sqrt{\frac{1}{s}\sum_{k=1}^{s} X^2_{t+k} - \frac{1}{s^2}\left(\sum_{k=1}^{s} X_{t+k}\right)^2}$$

and the rescaled range statistic is

$$R/S = \frac{R(t,s)}{S(t,s)}$$

which is convenient to estimate by regressing $\ln\left(R(t,s)/S(t,s)\right)$ on $\ln(s)$ and a constant. The variance of the series is assumed constant within the time intervals of different lengths, but may vary between them. The slope of the regression is an estimate of *H*. Mandelbrot has shown that R/S converges asymptotically almost surely for series with infinite variance. However, the estimate may be very much biased if short memory is present in the series. In general, R/S is biased towards 0.7 (Taqqu and Teverovsky, 1998). Davies and Harte (1987) conclude that it is hard to see how one could be expected to differentiate between fractional Gaussian noise from short memory such as AR(1) with R/S without a very long series. Computing the R/S from the residual of an autoregressive short memory filter may help to reduce the bias. Peters (1994) proposes that the variance of H is $1/T$ for Gaussian IID series, and provides some simulation results.

The V statistic for the R/S statistic was also proposed by Hurst.

$$V_n = \frac{(R/S)_n}{\sqrt{n}}$$

For the random walk, plotting the V statistic against $\ln(n)$ would produce a horizontal line. When the series is persistent, the square root standardization is not sufficient and the line has positive slope. In the antipersistent case, the slope is negative. The statistic provides a clearer graphical interpretation than the R/S, for which such fluctuations are less pronounced when plotted against $\ln(n)$. In a properly fractal series, the scaling continues indefinitely, and both R/S and V have constant slope. However, if there are proper cycles, periodic or nonperiodic, which are not caused by intermittent fractal trends, the constant upward slope falls when correlation over the cycle is exhausted. In this manner, the longest absolute or average cycle length can be approximated even in a chaotic process. (Peters, 1994)

It is still possible to qualitatively distinguish between random changes and fractional Brownian motion by scrambling the filtered series to break the correlation structure, which should result in a random walk with $H = ½$ (e.g. Peters, 1994). In general, the R/S statistic is to be considered only an indicator of the possibility of long memory, and further examination is necessary.

The modified R/S statistic is presented in Lo (1991). Instead of the usual sample variance, Lo uses the Newey-West heteroscedasticity and autocorrelation consistent variance estimate

$$\widetilde{S}(1,n)^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2 + \frac{2}{n}\sum_{j=1}^{q}\omega_j(q)\left(\sum_{k=j+1}^{n}(X_k - \overline{X})(X_{k-j} - \overline{X})\right)$$

where $\omega_j(q)$ are weights of a Bartlett window, $\omega_j(q) = 1 - \frac{j}{q+1}$, $q < n$. q should be large enough to include all 'large' autocorrelations in the estimator. Asymptotic results show that for a larger q the test size decreases as power increases. The sample mean $\overline{X}$ is taken over the whole series as is the modified *R/S* statistic also. In the notation of the original R/S, the modified statistic is expressed as

$$Q(n) = \frac{R(1,n)}{\widetilde{S}(1,n)}.$$

This statistic shows usually less evidence of long memory compared to the original R/S statistic. This may be intuitively attributed in part to the statistic taken at once over the whole series, although the biasedness of the original R/S is likely to be more important. The result is still strongly dependent on the choice of q and additional conditions and small sample properties remain sensitive to short memory effects. This estimator is not utilized in the estimation, as it has been shown to have similar properties to the original R/S statistic on one hand, and to the level stationary KPSS test on the other.

3.2 KPSS Test

Kwiatkowski, Phillips, Schmidt, and Shin (1992) test the null hypothesis of short memory deviations from a deterministic linear time trend as opposed to the alternative of trend or I(0) stationarity using a Lagrange multiplier test. The series y_t is assumed to be generated by the process

$$\begin{aligned} y_t &= \psi + \xi t + z_t, \qquad t = 1,2,\ldots,T \\ z_t &= r_t + \varepsilon_t \end{aligned}$$

(in the more intuitive notation of Lee and Schmidt (1996)). The deviations from the trend, z_t, are assumed to consist of a random walk $r_t = r_{t-1} - v_t$ with $r_0 = 0$ where v_t are iid with mean 0 and variance σ_v^2. ε_t is a stationary short memory process allowing z_t to asymptotically converge to the standard Brownian motion. The null hypothesis is then $\mathrm{H}_0 : \sigma_v^2 = 0$, under which $z_t = \varepsilon_t$ and the series y_t is trend stationary short memory. A special case of $\xi = 0$ which assumes the series y_t is stationary around a level ψ is considered as well.

The statistic under the null is computed as follows. A regression is run on $y_t = \psi + \xi(t) + \varepsilon$ and the residuals ε_t are assumed to have long run variance σ_ε^2 estimated by the Newey-West estimator as

$$s^2(l) = T^{-1}\sum_{t=1}^{T} e_t^2 + 2T^{-1}\sum_{s=1}^{l} W(s,l)\sum_{t=s+1}^{T} e_t e_{t-s}$$

$$W(s,l) = 1 - \frac{s}{1+l}$$

For the estimator to be consistent under the null, the lag truncation parameter $l \to \infty$ as $T \to \infty$; it is customary to use $l = o\left(T^{1/2}\right)$. Then, the KPSS statistic is

$$\hat{\eta}_\tau = T^{-2}\sum_{t=1}^{T}\frac{S_t^2}{s^2(l)}, \qquad S_t^2 = \sum_{j=1}^{t} e_j,\ t = 1,\ldots,T$$

The level stationary KPSS statistic $\hat{\eta}_\mu$ uses the residuals $e_t = y_t - \bar{y}$ (a regression on an intercept only). The test is conducted as an upper tail test. Under the null of $z_t = \varepsilon_t$ is short memory, the KPSS test statistic η_τ is distributed as a so-called second order Brownian bridge. Kwiatkowski et al (1992) provide tabulations of the upper tail critical values. η_τ is $o(1)$ under the null and $o(T/l)$ under the unit root alternative (Δz_t is short memory). Similar results apply for the statistic η_μ. Lee and Schmidt (1996) show that the KPSS test is consistent against stationary long memory processes where the order of integration $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, $d \neq 0$.

3.3 BDS Test

Brock, Dechert, Scheinkman, later with LeBaron (see Brock et al 1996), devise a test for detecting a chaotic data generation process using the correlation dimension. The BDS statistic has been found to be consistent against several different types of departures from the IID (independent and identically distributed) null hypothesis. The altenative hypothesis is not specified, and the statistic has been in wide use in testing different nonlinear specifications. The BDS statistic

$$W_{m,n}(\varepsilon) = \sqrt{n}\,\frac{T_{m,n}(\varepsilon)}{V_{m,n}(\varepsilon)}$$

is distribution free and converges in distribution to $N(0,1)$. $T_{m,n} = C_{m,n}(\varepsilon) - C_{1,n}(\varepsilon)^m$, where C represents the correlation integral. and $V_{m,n}$ is computed from

$$\begin{aligned}\tfrac{1}{4}V_m^2 &= m(m-2)C^{2m-2}(K - C^2) + K^m - C^{2m} \\ &\quad + 2\sum_{j=1}^{m-1}\left(C^{2j}\left(K^{m-j} - C^{2m-2j}\right) - mC^{2m-2}\left(K - C^2\right)\right)\end{aligned}$$

where K and C are also dependent on the distance ε . See Brock et al (1996) for a more complete derivation which is lengthy to be included here; the essential idea relies on the convergence of the correlation integral as noted in the previous chapter.

3.4 Geweke – Porter-Hudak Semiparametric Estimator

Geweke and Porter-Hudak (1983) estimate the fractional differencing parameter from the log spectrum

$$\ln(f(\lambda)) = \ln\frac{\sigma^2 f_u(0)}{2\pi} - d\ln\left(4\sin^2\frac{\lambda}{2}\right) + \ln\frac{f_u(\lambda)}{f_u(0)}$$

where $f(\lambda) = (\sigma^2/2\pi)(4\sin^2\lambda)^{-d} f_u(\lambda)$ and $f_u(\lambda)$ is the spectral density of the stationary linear process u_t in $(1-B)^d x_t = u_t$. The logarithmic periodogram can be written as

$$\ln(I(\lambda_j)) = \ln\frac{\sigma^2 f_u(0)}{2\pi} - d\ln\left(4\sin^2\frac{\lambda_j}{2}\right)$$
$$+\ln\frac{f_u(\lambda_j)}{f_u(0)} + \ln\frac{I(\lambda_j)}{f(\lambda_j)}$$

The second term on the RHS is considered in the OLS regression as the explanatory variable with the slope coefficient $-d$, and the third term is negligible when only the low frequencies having power similar to the zero frequency are considered. The regression intercept consists of the first term along with the mean of the last term. The last term is the regression disturbance, which is assumed asymptotically IID with a distribution of Gumbel type. Its asymptotic mean is the negative of Euler's constant, -.57721... and the variance is $\pi^2/6$.

The regression is run over the range of $j = 1,\ldots,g(T)$. $g(T)$ is the highest frequency used, and it should be chosen so that $\lim_{T\to\infty} g(T) = \infty$, $\lim_{T\to\infty}\frac{g(T)}{T} = 0$, and $\lim_{T\to\infty}\frac{\ln(T)^2}{g(T)} = 0$ in order to achieve asymptotic normality of the distribution and consistency of the slope estimator also in the presence of autocorrelation. In practice, little is known of the optimal choice of $g(T)$. The usual $\sqrt{T}$ has been suggested, and also experimenting with the regression so that the residual variance is approximately $\pi^2/6$. An a priori view of the shortest cycle length in actual time to be considered long memory could be used when possible.

Robinson has shown that the estimator is asymptotically consistent for $d < 0.50$, but only if some very first ordinates are truncated also. However, the small sample properties may still be seriously affected by short memory (Brock and de Lima, 1996). $\operatorname{var}(\hat{d}) = O(g(T)^{-1})$, which may cause the second-stage short memory parameters to be seriously biased. Clifford and Ray (1995) also note the bias of the estimator especially for non-stationary series.

3.5 Robinson's Gaussian Semiparametric Estimator

Robinson (1995) develops the statistical properties of an approximate frequency domain Gaussian likelihood estimator, proposed by Künsch. The process $x_t, t = 1,\ldots,n$ is assumed covariance stationary. The long memory assumption is $f(\lambda) \sim G\lambda^{1-2H}$, $\lambda \to 0+$, where $G \in (0,\infty)$, $H \in (0,1)$. The Fourier transform and periodogram are defined without correction for an unknown mean, as the correction only affects λ_0 and the computations involve only $\lambda_j = 2\pi j/n$, $j = 1,\ldots,m$ (with $m < \frac{1}{2}n$).

The objective function

$$Q(G,H) = \frac{1}{m}\sum_{j=1}^{m}\left(\ln G\lambda_j^{1-2H} + \frac{\lambda_j^{2H-1}}{G} I(\lambda_j)\right)$$

is minimized with the likelihood

$$\hat{H} = \arg\min_{H\in\Theta} \ln\left(\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2H-1} I(\lambda_j)\right) - (2H-1)\frac{1}{m}\sum_{j=1}^{m}\ln\lambda_j$$

where $\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2H-1} I(\lambda_j)$ is the estimator $\hat{G}(H)$. G (distinct from $\hat{G}(H)$) and H denote any admissible parameter values, as the estimate is not defined in closed form. The closed interval of $\hat{H}$ is defined as $\Theta = [\Delta_1, \Delta_2]$, $0 < \Delta_1 < \Delta_2 < 1$. The interval can be arbitrarily close to 0 and 1, or it can be chosen if prior knowledge exists. For example, if the spectral density does not approach zero at the zero frequency, the lower limit can be set to $\Delta_1 = \frac{1}{2}$.

Robinson (1995) includes lengthy proofs of weak consistency and asymptotic normality. Some related Monte Carlo experiments are conducted in Cheung and Diebold (1994). The assumptions for weak consistency are as follows: a regularity condition

$$\frac{d}{d\lambda} \ln f(\lambda) = O(\lambda^{-1}), \quad \lambda \to 0+$$

is required. In the Wold representation, the innovations are assumed to form a square-integrable martingale difference sequence that satisfies a homogeneity restriction but is not required to be strictly stationary. The last assumption is

$$\frac{1}{m} + \frac{m}{n} \to 0, \quad n \to \infty$$

since m must tend to infinity for consistency but at a slower rate than n in order to create a neighbourhood of zero frequency that slowly degenerates to zero as the sample size tends to infinity. This causes the estimate to lose some efficiency with respect to a theoretical parametric model; on the other hand, x_t need not be Gaussian and the whole range of $0 < H < 1$ is applicable. Selection of m is left to user. The conditions for asymptotic normality are stronger and require x_t to be fourth-order stationary.

3.6 Robinson's Averaged Periodogram Estimator

Robinson (1994) and Lobato and Robinson (1996) discuss estimation of $H \in (\frac{1}{2}, 1)$ with averages over bands of equally spaced discrete periodogram frequencies in a neighbourhood of zero frequency, instead of the logarithmic frequencies used e.g. in the GPH estimator. When logarithms are used also on the frequency axis, they move most of the frequencies towards the right end of the axis, which can be shown to have a considerable, possibly problematic effect on the regression estimates.

The estimator is based on

$$\hat{F}(\lambda) = \frac{2\pi}{n}\sum_{j=1}^{m} I\left(\lambda_j\right), \; m = \left[\frac{\lambda n}{2\pi}\right]$$

where [.] denotes integer part and n is the number of observations. x_t is assumed to be a linear process consisting of martingale differences that satisfy certain regularity conditions and the band decays to zero in the same way as above when sample size increases.

The theoretical band $F(\lambda_m)$, assumed now continuous, is able to detect long memory as

$$F(\lambda) = \int_0^{\lambda} f(\theta)d\theta \sim \frac{C\lambda^{2-2H}}{2-2H}, \qquad \lambda \to 0+$$

The estimator is

$$\hat{H}_q = 1 - \frac{1}{2\ln q}\ln\frac{\hat{F}(q\lambda_m)}{\hat{F}(\lambda_m)}$$

for a user-chosen $q \in (0,1)$.

Lobato and Robinson (1996) derive the limiting distributions, where x_t is assumed Gaussian. For $H \in (\frac{1}{2}, \frac{3}{4})$, the distribution is normal and for $H \in (\frac{3}{4}, 1)$ nonnormal and the estimator is biased also asymptotically. Therefore, the estimator is stated to appeal principally in the normal range, for which optimal q is tabulated with respect to H. There exists also an optimal choice of m. Monte Carlo simulations are employed to study the estimator's sensitivity to m and q with a Gaussian process having fractional noise autocovariance. Smaller $m = 32$ shows increasing negative skewness and leptokurtosis compared to $m = 64$. Negative skewness and leptokurtosis are stronger for higher H in all cases. $q = \frac{1}{2}$ or smaller is found reasonable for the whole range of H, and a stronger deterioration for $q = 0.8$ in case of small H.

3.7 Sowell's Maximum Likelihood Estimator

Sowell (1992) presents the MLE of the ARFIMA(p,d,q) class. The autocorrelation function is

$$\gamma_k = \sigma^2 \sum_{j=1}^{p} \xi_j \sum_{n=0}^{q} \sum_{m=0}^{q} \theta_n \theta_m C(d, d, p+n-m-k, \lambda_j)$$

where λ_j is the jth root (assumed distinct) of the autoregressive polynomial $\phi(L)$, and

$$\xi_j = \left(\lambda_j \prod_{i=1}^{p} (1 - \rho_i \rho_j) \prod_{\substack{k=1 \\ k \neq i}}^{p} (\rho_i - \rho_k) \right)^{-1}$$

$$C(w, v, k, p) = G(w, v, k)\left(\rho^{2p} F(v+k, 1; 1-w+k; \rho) + F(w-k, 1; 1-v-k; \rho\right)$$

$$G(w, v, k) = \frac{\Gamma(1-w-v)\Gamma(v+k)}{\Gamma(1-w+k)\Gamma(1-v)\Gamma(v)}$$

Doornik and Ooms (1996, 1999) give an alternative representation which they utilize in their application:

$$\gamma_k = \sigma_\varepsilon^2 \sum_{k=-q}^{q} \sum_{j=1}^{p} \psi_k \xi_j C(d, p+k-i, \rho_j),$$

$$\phi_k = \sum_{s=|k|}^{q} \theta_s \theta_{s-|k|},$$

$$\xi_j^{-1} = \left[\rho_j \prod_{i=1}^{p} \left(1 - \rho_i \rho_j\right) \prod_{m \neq j} \left(\rho_j - \rho_m\right) \right]$$

where

$$C(d,h,p)=\frac{\Gamma(1-2d)}{[\Gamma(1-d)]^2}\frac{(d)_h}{(1-d)_h}\cdot$$
$$\left[\rho^{2p}F(d+h,1;1-d+h;\rho)+F(d-h,1;1-d-h;\rho)-1\right]$$

$$F(a,1;b;z)=\sum_{i=0}^{\infty}\frac{(a)_i}{(b)_i}z^i,$$
$$(u)_i=u(u+1)(u+2)\cdots(u+i-1),\ (u)_0=1$$

The Levinson algorithm is used in the computations to reduce the computer memory storage requirement from j^2 to j during computations of the concentrated likelihood function. The common residual variance is concentrated out by writing $\Omega=\mathbf{R}\sigma_\varepsilon^2$. Differentiating and maximizing with respect to σ_ε^2 leads to $\sigma_\varepsilon^2=T^{-1}\mathbf{x}'\mathbf{R}^{-1}\mathbf{x}$, and the concentrated log-likelihood is

$$\ell_c(d,\phi,\theta)=-\frac{T}{2}\ln(2\pi)-\frac{T}{2}-\frac{1}{2}\ln|\mathbf{R}|-\frac{T}{2}\ln\left[T^{-1}\mathbf{x}'\mathbf{R}^{-1}\mathbf{x}\right]$$

For the maximization, the constants are removed to obtain

$$\max\left(-\frac{1}{2}\left\{\frac{1}{T}\ln|\mathbf{R}|+\ln\left[T^{-1}\mathbf{x}'\mathbf{R}^{-1}\mathbf{x}\right]\right\}\right)$$

Maximization is carried out using BFGS with numerical derivatives. Stationarity is imposed for each iteration step by rejecting parameter value combinations where $d\le -5$ or $d>0.49999$, and/or some autoregressive root $|\rho_i|\ge 0.9999$.

4. Results And Conclusions

4.1 Estimation Results

The stock market series are logarithmic and first differenced unless otherwise required. The interest rate series is only differenced. The conventional manner is to use the logarithmic differenced series, but we felt the unlogged case is worth studying as well, since there is no clear theoretical justification for logarithms when concerned with the economic interpretation of the effects of changes in interest rates. All series look nonstationary in the logarithmic graphs, and all of them show the typical spectral shape. In addition, HEX, S&P 500, and Federal Funds Rate seem to have a declining first differenced spectrum. The periodograms in the figures are limited from top and bottom. Only SHX seems stationary after taking the first difference.

The KPSS test, frequency domain and R/S estimations were carried out with RATS packages contributed and updated by various authors. BDS test was undertaken with Dechert's software. Estimation of the ARFIMA models was conducted with Ox ARFIMA 1.0 package.

The tests conducted before estimation (Table 1) clearly confirm the nonstationarity. The conventional Dickey-Fuller unit root test detects a unit root – or an approximate one – in all cases, and the AR(1) estimates for the Jarque-Bera test are very close to one. The J-B rejects normality, as all series have a sharp peak at the mean and noticeable leptokurtosis. KPSS also suggests that a linear trend is not enough for modelling these series.

We follow suggestions in Gwilym et al (1999) in computing the BDS test. We assign the value 1 for the ratio ε / σ of the diverging distance to the (estimated) standard deviation of the series. Although other values such as 0.5, 1.5, and 2 are often used, they tend to produce similar results. The embedding dimensions vary from 2 to 6; 6 is very much of an upper limit for any validity of the computations with these sample sizes. The BDS test accepts the IID hypothesis for the differenced SHX and Federal Funds Interest Rate.

This may be somewhat surprising, but the estimations show that at least in the case of SHX, the series simply has a large variance, and possible long memory is intermingled with offsetting short memory. The smaller sample sizes (see Table 2) and the sparsity of the monthly Fed Funds series may have an effect on the accuracy of the test. For HEX and S&P 500, BDS rejects the IID except for the embedding dimension 2. But as the other embedding dimensions are considered, the test rejects IID on average. In the low embedding dimension, a significant part of the process may still be interpreted as linear – assuming that the process fits in the low dimension, of course.

A = Accept, R = Reject	Normality	Unit Root	Stationarity + short memory	IID
	J-B AR(1) residuals	D-F Augm. 10 lags linear trend	KPSS 1-4 lags linear trend y/n	BDS Diff's Emb. D 2-6 $\varepsilon/\sigma = 1$
HEX	R	A	R	2A, 3-6 R
S&P 500	R	A	R	2A, 3-6 R
SHX	R	A	R	A
Fed Funds	R	A	R	A

Table 1. Tests before estimation.

Furthermore, Bartlett's cumulative periodogram test for white noise (see figures 1-4, top right corners) shows that after differencing, there is still a significant amount of power accumulating already in the very beginning of the series, except for SHX. After estimating and taking the fractional difference (in the middle graph of the right-hand side), the periodogram at the low frequencies is reduced to white noise. This still leaves power in the high frequency, transforming the series to pink noise instead. Only after including suitable AR and MA parameters in the complete ARFIMA model, the series are reduced to white noise (the graph in the bottom right corner), which is also confirmed by the test statistic.

The parameter estimates in table 2 show that HEX, S&P 500, and the Federal Funds Rate are all fractional noises with stronger case for HEX and Federal Funds Rate, and a weak but significant case for S&P 500. SHX appears to be I(1), except when the short-run parameters are estimated, long memory becomes visible. However, the Federal

Funds Rate series is problematic, since it has only monthly observations. In table 2, The first number under the estimator is the estimate for $\hat{d}-1$ with the significant ones in boldface. The standard deviation is given under $\hat{d}-1$, and the proposed degree of integration follows. The estimators give similar results in general. The V statistic plots for all differenced series are nearly straight lines with upward slopes, indicating that fractional noise is present, and the scaling continues throughout without a single average cycle length for some approximate, complete cycle measurable within the sample period. Although in Federal Funds Rate series there is a slight drop in the angle of the line at long wavelength, it still continues to clearly slope upwards.

We apply the estimators to the differenced series, as the properties of even those estimators that can handle level series in theory deteriorate quickly as d increases. The results for the spectral estimators are reported for the band parameter 0.5, which uses $\sqrt{T}$ available ordinates for estimation. This is customary; other bands give results that vary according to the shape of the spectrum, but it has been found in simulation studies that the value 0.5 is reliable compared to other ones. For the AVPER (Robinson's averaged periodogram estimator), we use $q = 0.5$. Our results with changes in the bandwidth still classify the series, except SHX, as fractional noises.

HEX	R/S	GPH	GSE	AVPER	ML
Logarithmic	**.647**	.145	.090	.023	**.128**
Differences	.002	.101	.069	.699	.016
	I(1.15)	I(1.15)	I(1.09)	I(1.02)	I(1.13)
870201-971230			$\hat{d}-1$	$=$ **.077**	(.019)
	ML ARFIMA(1,*d*,1): I(1.08)		$\hat{\phi}$	$=$ **−.462**	(.116)
N=2757			$\hat{\theta}$	$=$ **.575**	(.102)
SP500	R/S	GPH	GSE	AVPER	ML
Logarithmic	**.586**	.064	-.010	.002	**.060**
Differences	.001	.072	.051	.720	.009
	I(1.09)	I(1.06)	I(0.99)	I(1.00)	I(1.06)
620702-971213	ML ARFIMA(0,*d*,1): I(0.97)		$\hat{d}-1$	$=$ **−.030**	(.012)
N=8937			$\hat{\theta}$	$=$ **.150**	(.015)
SHX	R/S	GPH	GSE	AVPER	ML
Logarithmic	**.597**	-.004	.020	-.000	-.002
Differences	.003	.123	.082	.722	.020
	I(1.10)	I(1.00)	I(1.02)	I(1.00)	I(1.00)
911507-963112	ML ARFIMA(1,*d*,0): I(1.07)		$\hat{d}-1$	$=$ **.066**	(.033)
N=1381			$\hat{\phi}$	$=$ **−.111**	(.041)
Federal Funds	R/S	GPH	GSE	AVPER	ML
Rate	**.782**	-.048	-.050	.123	**.255**
(monthly)	.007	.170	.107	.604	.049
	I(1.28)	I(0.95)	I(0.95)	I(1.12)	I(1.26)
5701-9812			$\hat{d}-1$	$=$ **−.232**	(.092)
	ML ARFIMA(1,*d*,1): I(0.77)		$\hat{\phi}$	$=$ **.372**	(.136)
N=503			$\hat{\theta}$	$=$ **.278**	(.078)

Table 2. Parameter Estimates.

An interesting feature of the results is that the degree of integration changes notably, when the AR and MA parameters are introduced. Several ARFIMA specifications were tested, and the most parsimonious, significant one was chosen among other possible alternatives. To detect all departures from white noise it is necessary to compromise between the high and low frequencies in accuracy and specification, and it seems that estimation of *d* alone is not sufficient. The actual order of integration is thus not very precise, and perhaps there is even no unique *d*. Even more accurate estimates could be considered – however, at the expense of a relatively convenient and parsimonious representation and at the risk of introducing additional estimation problems.

The Dickey-Fuller and KPSS tests for the residuals (table 3) show that the order of integration could be considered adequately determined by fractional differencing. The series are still not normally distributed, so additional measures need to be explored to reduce the sharp peaks and leptokurtosity of the series. The BDS accepts now the IID hypothesis also for embedding dimension 3 for HEX and S&P 500, indicating that the estimation process removed a large part of the nonlinearities and dependencies, but also notes that dependencies remain in the higher dimensions.

A = Accept, R = Reject	Normality	Unit Root	Stationarity + short memory	IID
	J-B	D-F Augm. 10 lags no trend	KPSS 1-4 lags no trend	BDS Emb. D 2-6 $\varepsilon/\sigma = 1$
HEX	R	R	A	2-3 A, 4-6 R
S&P 500	R	R	A	2-3 A, 4-6 R
SHX	R	R	A	A
Fed Funds	R	R	A	A

Table 3. Tests on ARFIMA residuals.

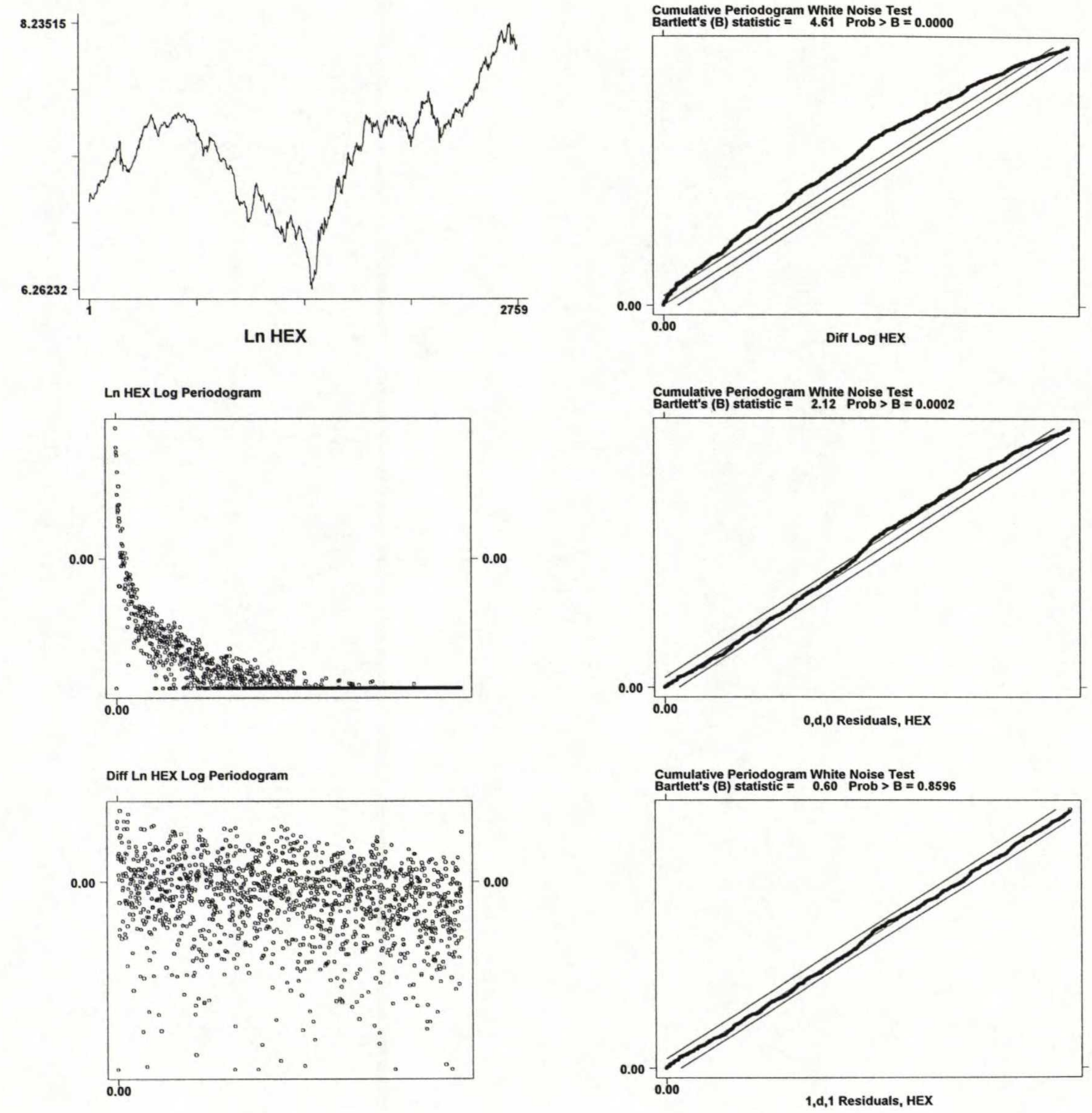

8.23515
6.26232
1
2759
Ln HEX
Cumulative Periodogram White Noise Test
Bartlett's (B) statistic = 4.61 Prob > B = 0.0000
0.00
0.00
Diff Log HEX
Ln HEX Log Periodogram
0.00
0.00
0.00
Cumulative Periodogram White Noise Test
Bartlett's (B) statistic = 2.12 Prob > B = 0.0002
0.00
0.00
0,d,0 Residuals, HEX
Diff Ln HEX Log Periodogram
0.00
0.00
0.00
Cumulative Periodogram White Noise Test
Bartlett's (B) statistic = 0.60 Prob > B = 0.8596
0.00
0.00
1,d,1 Residuals, HEX

Figure 1. HEX

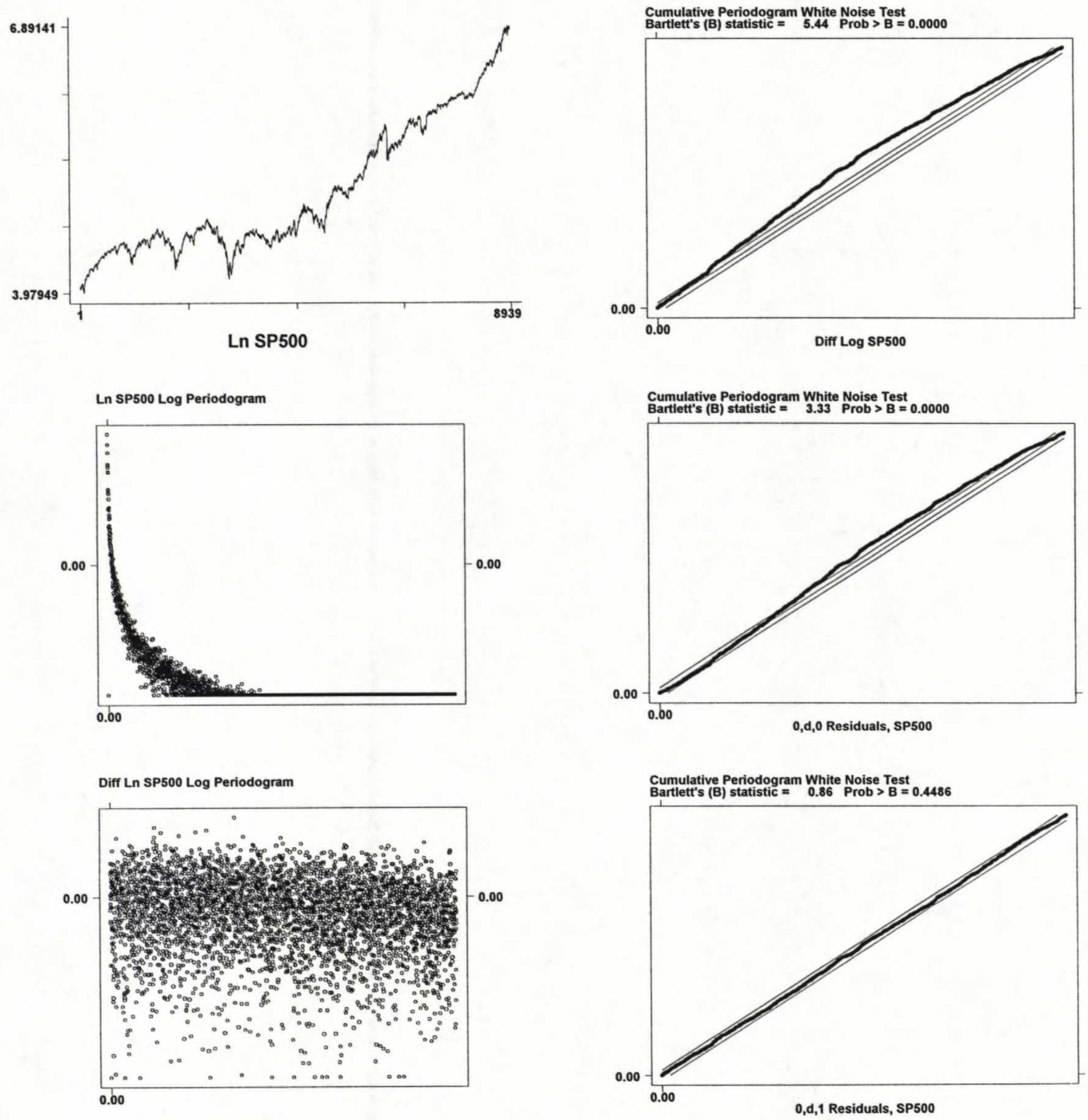
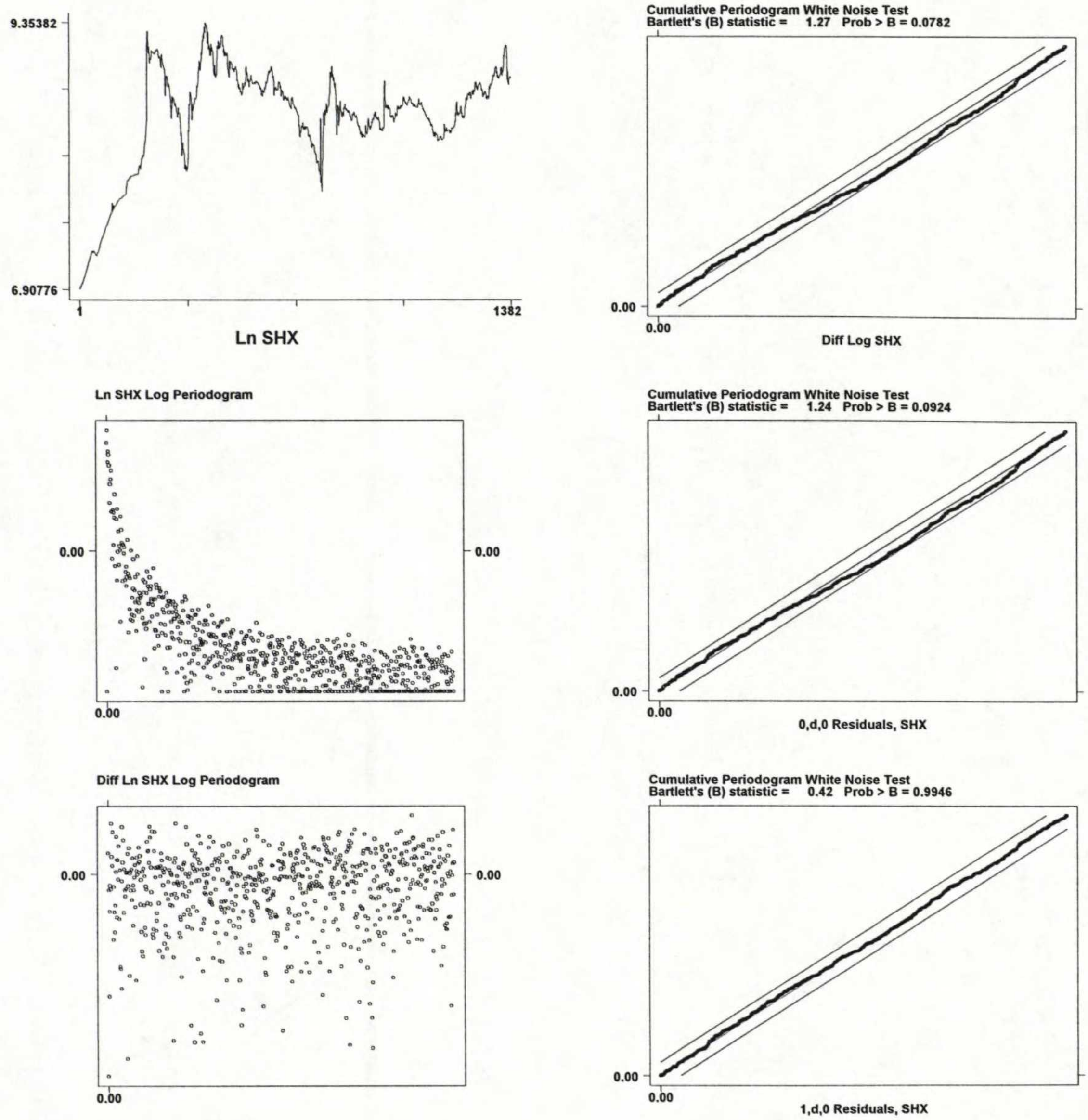


Figure 2. S&P 500



9.35382
6.90776
1
1382
Ln SHX
Cumulative Periodogram White Noise Test
Bartlett's (B) statistic = 1.27 Prob > B = 0.0782
0.00
0.00
Diff Log SHX
Ln SHX Log Periodogram
0.00
0.00
0.00
Cumulative Periodogram White Noise Test
Bartlett's (B) statistic = 1.24 Prob > B = 0.0924
0.00
0.00
0,d,0 Residuals, SHX
Diff Ln SHX Log Periodogram
0.00
0.00
0.00
Cumulative Periodogram White Noise Test
Bartlett's (B) statistic = 0.42 Prob > B = 0.9946
0.00
0.00
1,d,0 Residuals, SHX

Figure 3. SHX

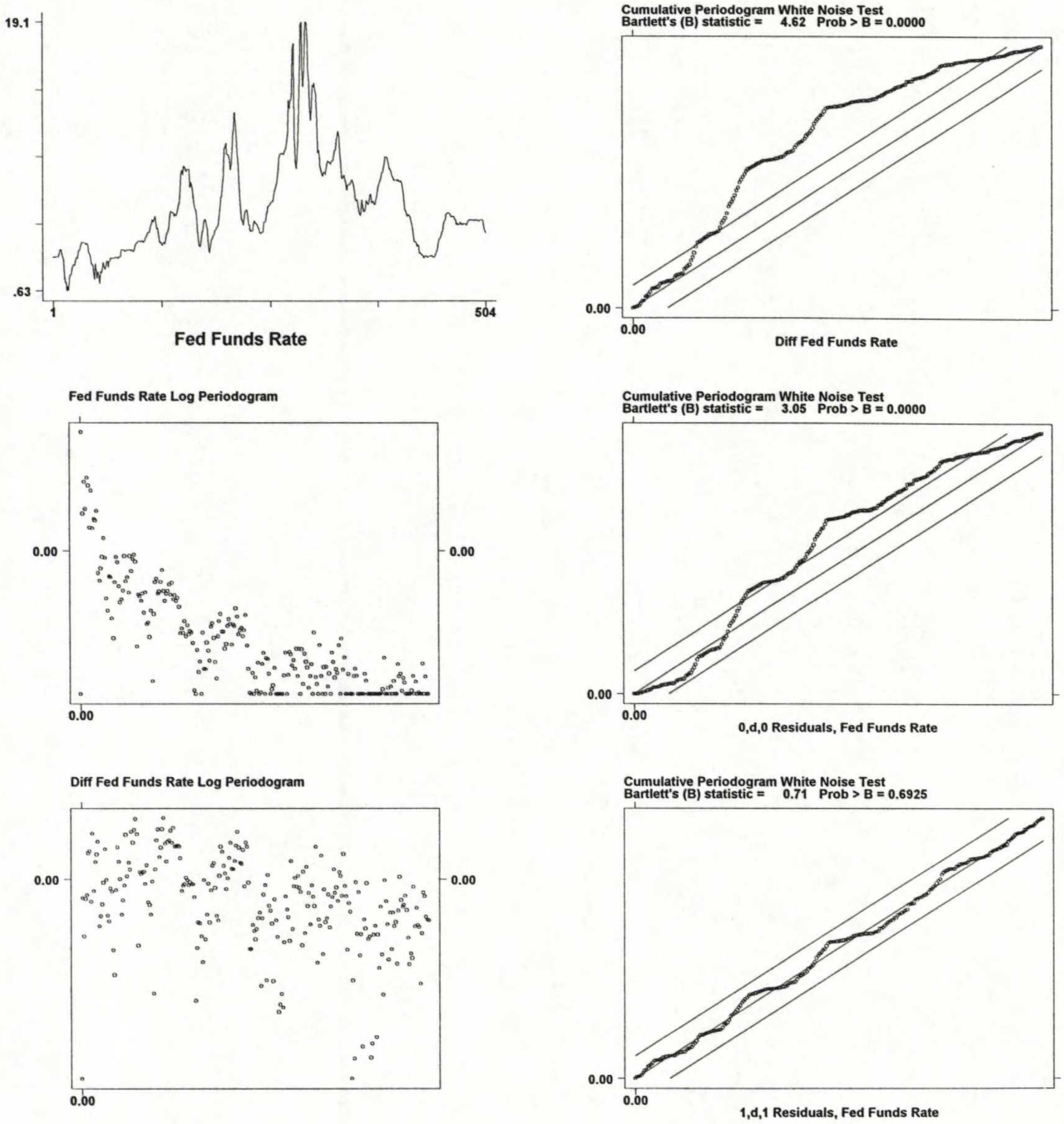


Figure 4. Federal Funds Rate

4.1 Conclusions

We test for fractional noises in four financial time series with various estimators. Fractional noise has a significant amount of low frequency power, and dependencies over long time horizons excert a notable effect on the behaviour of the time series.

The tests confirm that the simple difference does not render the studied series stationary. The nonstationarities are more intricate than what is classically assumed. We attribute the nonstationarities to fractional noise, and estimate the corresponding fractional difference for stationarization. Without fractional differencing, short memory parameters such as the commonly used autoregressive and moving average parameters are erroneously estimated due to the nonstationarities which remain after simple differencing.

We find that the HEX daily closing return series is a strong fractional noise. We find a weaker but significant amount of fractional noise in the S&P 500. According to initial tests and estimates, the Shanghai SHX all share index returns appear first-order integrated. After specifying simultaneous short memory, significant long memory becomes visible. The changes in the U.S. Federal Funds Rate also appear as fractional noise, but the estimates are not very precise possibly due to the monthly sampling frequency.

In estimation and testing, we rely on the maximum likelihood estimator, which is consistent for the model despite of departures from normality of the increments. The frequency domain estimators give similar results, and additionally, show some evidence for further nonlinear characteristics which are also captured by some of the linearity and normality tests employed. The ARFIMA(p,d,q) model specification with simultaneous long and short range dependency is able to reduce all fractional noise series to white noise, and it parsimoniously, although not necessarily uniquely, represents the significant fractional noise characteristics.

References

Adler, R., R. Feldman, and M.S. Taqqu (eds.) (1998), *A Practical Guide To Heavy Tails: Statistical Techniques and Applications*. Birkhauser.

Baillie, R.T. (1996), "Long memory processes and fractional integration in econometrics." *Journal of Econometrics 73, 5-59.*

Baillie, R.T., and T. Bollerslev (1994), "The long memory of the forward premium." *Journal of International Money and Finance 5, 565-571.*

Baillie, R.T., and C-F. Chung (1993), "Small Sample Bias in Conditional Sum-of-Squares Estimators of Fractionally Integrated ARMA Models." *Empirical Economics 18, 791-806.*

Barkoulas, J.T., C.F. Baum, and N. Travlos (2000), "Long Memory In the Greek Stock Market." *Applied Financial Economics 10, 177-184.*

Beveridge, S., and C. Oickle (1997), "Long memory in the Canadian stock market." *Applied Financial Economics 7, 667-672.*

Blasco, N., and R. Santamaria (1996), "Testing memory patterns in the Spanish stock market." *Applied Financial Economics 6, 401-411.*

Box, G.E.P., and G.M. Jenkins (1976), *Time Series Analysis Forecasting And Control, 2nd edition.* Holden-Day.

Breidt, F.J., N. Crato, and P. de Lima (1998), "The detection and estimation of long memory in stochastic volatility." *Journal of Econometrics, 83, 325-348.*

Brock, W.A. and P.J.F. de Lima (1996), "Nonlinear Time Series, Complexity Theory, and Finance." In Maddala, G.S. and C.R. Rao (eds.), *Handbook of Statistics, Vol. 14*. Elsevier Science.

Brock, W.A., W.D. Dechert, J.A. Scheinkman, and B. LeBaron (1996), "A Test For Independence Based On The Correlation Dimension." Mimeo, University of Wisconsin.

Chambers, M.J. (1998), "Long memory and aggregation in macroeconomic time series." *International Economic Review 39, 1053-1072.*

Chen, G., B. Abraham, and S. Peiris (1994), "Lag window estimation of the degree of differencing in fractionally integrated time series models." *Journal of Time Series Analysis 5, 473-487.*

Cheung, Y.W. and F.X. Diebold (1994), "On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean." *Journal of Econometrics 62, 317-350.*

Chong, T.T, and G.C. Lui (1999), "Estimating the fractionally integrated process in the presence of measurement errors." *Economics Letters 63, 285-294.*

Chung, C-F. (1996), "A generalized fractionally integrated moving-average process." *Journal of Time Series Analysis, 17, 111-140.*

Cosslett, S.R. (1990), "Semiparametric Estimation." In Eatwell, J., M. Milgate, and P. Newman (eds.), *The New Palgrave Time Series And Statistics*. W.W. Norton & Company.

Creedy, J. and V.L. Martin (1994), *Chaos and Non-Linear Models In Economics: Theory And Applications*. Edward Elgar.

Davies, R.B., and D.S. Harte (1987), "Tests for Hurst effect". *Biometrika 74, 1, 95-101.*

Delgado, M.A. and P.M. Robinson (1994), "New methods for the analysis of long-memory time series: application to Spanish inflation." *Journal of Forecasting 2, 97-107.*

Diebold, F.X., and G.D. Rudebusch (1989), "Long memory and persistence in aggregate output." *Journal of Monetary Economics 24, 189-209.*

Doornik, J. and M. Ooms (1996) "A Package for Estimating, Forecasting and Simulating Arfima Models." Mimeo, Oxford University.

Doornik, J. and M. Ooms (1999) "A Package for Estimating, Forecasting and Simulating Arfima Models: Arfima Package 1.0 for Ox." Mimeo, Oxford University.

Eatwell, J., M. Milgate, and P. Newman (eds.) (1990), *The New Palgrave Time Series And Statistics*. W.W. Norton & Company.

Fox, R., and M.S. Taqqu (1986), "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series." *The Annals of Statistics 2, 517-532.*

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

Gil-Alana, L.A., and P.M. Robinson (1997), "Testing of unit root and other nonstationary hypotheses in macroeconomic time series." *Journal of Econometrics 80, 241-268.*

Granger, C.W.J. (1966), "The typical spectral shape of an economic variable." *Econometrica 34, 150-161.*

Granger, C.W.J. (1980), "Long Memory Relationships And The Aggregation of Dynamic Models." *Journal of Econometrics 14, 227-238.*

Granger, C.W.J. (1990), "Spectral Analysis." In Eatwell, J., M. Milgate, and P. Newman (eds.), *The New Palgrave Time Series And Statistics*. W.W. Norton & Company.

Granger, C.W.J. and M.W. Watson (1984), "Time Series and Spectral Methods". In Griliches, Z. and M.D. Intriligator (eds.) (1984), *The Handbook of Econometrics Volume II*. Elsevier Science.

Gray, H.L., N-F. Zhang, and W.A. Woodward (1989) "On generalized fractional processes." *Journal of Time Series Analysis 3, 233-257.*

Gray, H.L., N-F. Zhang, and W.A. Woodward (1994) "On generalized fractional processes – a correction." *Journal of Time Series Analysis 5, 561-562.*

Griliches, Z. and M.D. Intriligator (eds.) (1984), *The Handbook of Econometrics Volume II*. Elsevier Science.

Gwilym, O. AP, C. Brooks, A. Clare, and S. Thomas (1999) "Tests of non-linearity using LIFFE futures transactions price data." *The Manchester School 2, 167-186.*

Hauser, M.A., R.M. Kunst, and E. Reschenhofer (1994) "Modelling exchange rates: long-run dependence versus conditional heteroskedasticity." *Applied Financial Economics 3, 233-240.*

Harvey, A.C. (1981), *Time Series Models*. Philip Allan.

Hosking, J.R.M. (1981), "Fractional Differencing." *Biometrika 68, 1, 165-76.*

Hosoya, Y. (1997), "A limit theory for long-range dependence and statistical inference on related models." *Annals of Statistics 1, 105-137.*

Hurvich, C.M., and B.K. Ray (1995), "Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes". *Journal of Time Series Analysis 1, 17-42.*

Kanto, A. (1983), *On The Estimation of The Inverse Autocorrelation Function*. Ph.D Thesis, University of Tampere.

Karr, A.F. (1990), "Martingales." In Eatwell, J., M. Milgate, and P. Newman (eds.), *The New Palgrave Time Series And Statistics*. W.W. Norton & Company.

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

Koopmans, L.H. (1974), *The Spectral Analysis of Time Series*. Academic Press, New York.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin (1992), "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of Econometrics, 54, 159-178.*

Lee, D., and P. Schmidt (1996), "On the power of the KPSS test of stationarity against fractionally-integrated alternatives." *Journal of Econometrics 73, 285-302.*

Li, W.K., and A.I. McLeod (1986), "Fractional time series modelling." *Biometrika 73,1, 217-221.*

Linden, M. (1999), "Time series properties of aggregated AR(1) processes with uniformly distributed coefficients." *Economics Letters 64, 31-36.*

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

Lobato, I., and P.M. Robinson (1996), "Averaged periodogram estimation of long memory." *Journal of Econometrics 73, 303-324.*

Maddala, G.S. and C.R. Rao (eds.) (1996), *Handbook of Statistics, Vol. 14.* Elsevier Science.

Mandelbrot, B.B. (1971), "When can price be arbitraged efficiently? A limit to the validity of the random walk and martingale models." *Review of Economics and Statistics 53, 543-553.*

Mandelbrot, B.B. (1982), *The Fractal Geometry of Nature.* W.H. Freeman And Company.

Mandelbrot, B.B. (1997), *Fractals And Scaling In Finance.* Springer-Verlag.

Martin, V.L., and N.P. Wilkins (1999), "Indirect estimation of ARFIMA and VARFIMA models". *Journal of Econometrics 93, 149-175.*

Mazaheri, A. (1999), "Convenience yield, mean reverting prices, and long memory in the petroleum market." *Applied Financial Economics, 9, 31-50.*

Mills, T.C. (1996), *The Econometric Modeling of Financial Time Series.* Cambridge University Press.

Nelson, C.R. and C.I. Plossner (1982), "Trends and random walks in macroeconomic time series." *Journal of Monetary Economics 10, 139-162.*

Peitgen, H-O., H. Jürgens, and D. Saupe (1992), *Chaos and Fractals: New Frontiers of Science.* Springer-Verlag.

Peters, E.E. (1994), *Fractal Market Analysis.* Wiley.

Peters, E.E. (1996), *Chaos and Order in the Capital Markets. 2nd edition.* Wiley.

Porter-Hudak, S. (1990), "An Application of the Seasonal Fractionally Differenced Model to the Monetary Aggregates." *Journal of the American Statistical Association 410, 338-344.*

Priestley, M.B. (1989), *Non-Linear And Non-Stationary Time Series Analysis*. Academic Press.

Robinson, P.M. (1994), "Semiparametric analysis of long-memory time series." *Annals of Statistics 22, 515-539.*

Robinson, P.M. (1995), "Log-periodogram regression of time series with long range dependence." *Annals of Statistics 3, 1048-1072.*

Robinson, P.M. (1995), "Gaussian semiparametric estimation of long range dependence." *Annals of Statistics 5, 1630-1661.*

Robinson, P.M. and F.J. Hidalgo (1997), "Time series regression with long-range dependence." *Annals of Statistics 1, 77-104.*

Silvapulle, P. (1995) "A Score Test for Seasonal Fractional Integration and Cointegration". Mimeo, University of Iowa.

Smith, J., N. Taylor, and S. Yadav (1997), "Comparing the bias and misspecification in ARFIMA models." *Journal of Time Series Analysis 5, 507-527.*

Sowell, F.B. (1992), "Maximum likelihood estimation of stationary univariate fractionally integrated time series models." *Journal of Econometrics 53, 165-188.*

Sutcliffe, A. (1994), "Time-Series Forecasting Using Fractional Differencing." *Journal of Forecasting, 383-393.*

Taqqu, M.S., and V. Teverovsky (1997), "Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator." *Journal of Time Series Analysis 3, 279-304.*

Taqqu, M.S., and Teverovsky, V. (1998), "On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series." in Adler et al (eds.), (1998) *A Practical Guide To Heavy Tails: Statistical Techniques and Applications*. Birkhauser.

Tse, Y. (1998), "Fractional cointegration tests with GARCH." *Applied Financial Economics 8, 329-332.*