MEMORY DIAGNOSTIC IN TIME SERIES ANALYSIS

A Dissertation Presented for the Degree of *Doktor der Philosophie (Dr. Phil.)*

at the Faculty of Verhaltens- und Empirische Kulturwissenschaften of the

Ruprecht-Karls-Universität Heidelberg

by

Simone L. Braun

Born in 75015 Bretten, Germany

Dean of Faculty: Prof. Dr. Andreas Kruse

Advisor/ First Reviewer: Prof. Dr. Joachim Werner

Second Reviewer: Prof. Dr. Andreas Voß

Oral examination: 2nd of June, 2010

ACKNOWLEDGEMENT

The author wishes to express sincere appreciation to Professor Dr. Joachim Werner and Dr. Tetiana Stadnyska for their assistance in the preparation of this manuscript and their guidance throughout my research. In addition, special thanks to my very dear collaborators and colleagues Dipl. Psych. Esther Stroe-Kunold, Dipl. Math. Antje Gruber, and Astrid Milde for their most valuable input. I owe a special debt of gratitude to Professor Dr. Andreas Voß for granting his expert opinion.

MEMORY DIAGNOSTIC IN TIME SERIES ANALYSIS

The objectives of this thesis is to evaluate the reliability of different periodogram-based estimation techniques and their non-spectral alternatives, implemented in the free software environment for statistical computing and graphics *R*, in distinguishing time series sequences with different memory processes, specifically to discriminate (1) two different classes of persistent signals within fractal analysis, fractional Brownian motions (fBm) and fractional Gaussian noises (fGn) (2) nonstationary and stationary ARFIMA (*p*,*d*,*q*) processes as well as (3) short- and long-term memory properties of the latter, and to assess the accuracy of the corresponding estimates. After a brief introduction into time- and frequency-domain analyzes fundamental concepts such as the ARFIMA methodology and fractal analysis for modeling and estimating long-(LRD) and short-range dependence (SRD) as well as (non)stationary of time series are presented. Furthermore, empirical studies utilizing time series analysis of long memory processes as diagnostic tools within psychological research are demonstrated. Three simulation studies designed to solve the abovementioned methodological problems represent the main field of this thesis, i.e., the reliable identification of different memory as well as specific statistical properties of ARFIMA and fractal time series and the assessment of estimation accuracy of the procedures under evaluation, and thus, based on the empirical findings, recommending the most reliable procedures for the task at hand.

Keywords: time series, time-and frequency domain analyzes, ARFIMA, stationary, long-range dependence, periodogram analyzes.

CONTENTS

1 INTRODUCTION

Glass, Willson, and Gottman (1975), McCleary and Hay (1980), and Gottman (1981) introduced time series procedures to social and behavioral sciences three decades ago and thus challenged the popular view that most psychological phenomena can be viewed as randomly distributed in time around a more or less stable mean. Since then researchers from different fields of psychology have recognized the advantages of time series methods to capture dependence and instability in their empirical data. Persistent autocorrelations in the data generating process indicates long-range dependence (LRD) or, in other words, a process with long memory. Long memory implies statistical dependence between observations separated by a large number of time units (Beran, 1994) as opposed to processes with short-range dependence (SRD), whose autocorrelations decay quickly as the number of observation increases. Gilden et al. (Gilden, 1997, 2001; Gilden & Wilson, 1995a,b; Gilden et al., 1995) demonstrated in experiments including mental rotation, lexical decision, shape and color discrimination or visual search that persistent autocorrelations account for even more variability in the data than most standard manipulations in cognitive psychology. Wagenmakers et al. (2004) confirmed these findings employing the ARFIMA methodology in their analyzes. Van Orden et al. (2003), Wagenmakers et al. (2004) and Ward & Richard (2001) found LRD in automatic cognitive performances such as word naming or simple reaction times. Chen et al. (1997, 2001), Delignières et al. (2004) and Ding et al. (2002) observed persistent correlations in human rhythmic activities such as tapping or other tasks requiring coordination or synchronization of motor and cognitive activities. Delignières, Fortes & Ninot (2004) reported LRD in time series of self-esteem and physical self as well as in human gait (Hausdorff et al, 1999), force production tasks (Pressing, 1999), brain activity (Linkenkaer-Hansen, 2002), heart rate fluctuations or other biological phenomena (Hausdorff & Peng, 1996), demonstrating the prevalence of long memory processes in human science and

the need for reliable diagnostic tools for identifying processes with different memory properties.

Various analyzing techniques for model fitting and parameter estimation of different process types, e.g., different periodogram-based methods and non-spectral alternatives are freely available in the statistical software R. The evaluation of their diagnostic abilities in identifying time series with different memory properties within the ARFIMA (p,d,q) methodology and fractal analysis is the main objective of this thesis.

This paper is divided into six parts. Following the introduction in Chapter 1, Chapter 2 presents two major approaches of the time series paradigm: time- and frequency-domain analyzes and their fundamental concepts. Chapter 3 focuses on time series with long memory, especially on the modeling, identification and estimation of LRD by means of ARFIMA and fractal analysis. Empirical studies within the psychological research implementing time series analysis to distinguish between clinical and normal groups are demonstrated in Chapter 4. However, the foremost work of this thesis can be found in Chapter 5. Three simulation studies evaluating the diagnostic capability of different periodogram-based estimation methods and their non-spectral alternatives to distinguish between stationary and nonstationary LRD processes as well as between stationary SRD and LRD processes are designed to empirically determine the most reliable estimation method for the rigorous discrimination of qualitative different process types.

2 TIME SERIES ANALYSIS: MAJOR APPROACHES

There are two related methods for the analysis of time series data. The first approach includes frequency-domain methods such as harmonic analysis, periodogram analysis, and spectral analysis (Warner, 1998, p. 186). The second approach is a set of time-domain methods formally called Box-Jenkins-ARIMA modeling, a strategy proposed by Box and Jenkins (1970). Both approaches examine time series data from different perspectives. Frequency-domain methods essentially decompose the variance of a time series into variance that is accounted for by a set of sinusoidal cyclic components while time-domain methods detect pattern in the data such as coefficients that describe consecutive elements of the series from specific time-lagged or previous elements. Although pursuing different objectives both approaches are mathematically equivalent. For example, time series that are well explained by certain kinds of second-order autoregressive models with large coefficients in time-domain will also tend to have a rather large and broad peak at the low frequency end of the spectrum in frequency-domain, thus implying a relatively high percentage of the variance in the time series is accounted for by long cycles (Warner, 1998, p. 187).

The objective of this chapter is to provide a brief introduction to the frequency- and time-domain methods. A detailed description of the frequency-domain methods can be found in Bloomfield (2000). Warner (1998) provides a thorough introduction to methods for detecting and describing cyclic patterns in time series data. A detailed comparison of both frequency- and time-domain approaches is provided by Gottman (1981).

2.1 Frequency-Domain Analysis

Exploring cyclical patterns explaining the variance of a time series is the main concern in frequency-domain methods. A couple of well-related statistical methods used to detect cycles in time-series data are harmonic analysis, periodogram analysis, and spectral analysis. The common ground of these methods is the sinusoid for representing cycles, i.e., the waveform of the trigonometric sine or cosine function, and the basis for estimating the *spectral density function* assessing the variance of an observed time series.

2.1.1 Basic Notation and Principles

A periodic function is a function that repeats its values in regular intervals or periods. Trigonometric functions like the *sine or cosine function* are the most prominent representatives of periodic functions. They repeat over intervals of length 2π, and serve as models for cycles. For example, different sine waves can be modeled by varying the mean (μ), the angular frequency (ω), the phase (φ), and the amplitude (A) of the function

$$Y_t = \mu + A\sin(\omega t + \varphi) = \mu + A\sin(2\pi t/\tau + \varphi) = \mu + A\sin(2\pi f t + \varphi)$$

The wave length or *frequency(f)* of a sine or cosine function is typically expressed in terms of the number of cycles per unit time and given by

$$f = \frac{1}{\tau} = \frac{\omega}{2\pi},$$

where τ denotes the period or the length of the cycle, i.e., the distance from one peak to the next. Since the period of a sine or cosine function is defined as the length of time required for one full cycle, it is the reciprocal of the frequency.

Time series with a length equal to a power of 2 can be approximated by a Fourier representation, where the series length (T) determines the number of the frequencies. For series with odd number of observations, there exist (T-1)/2 different frequencies:

$$f_j = \frac{j}{T}, j = 1, 2, 3, \ldots, \frac{(T-1)}{2}$$

corresponding to cycles of period T, $T/2$, $T/3$,…, 2 time units, inferring the fastest detectable frequency is

$$f = 1/2 = 0.5 \text{ or } \omega = f 2\pi = 1/2 \cdot 2\pi = \pi .$$

Furthermore,

$$A\sin(\omega + \varphi) = A(\cos\omega t \sin\varphi + \sin\omega t \cos\varphi) = a\cos\omega t + b\sin\omega t ,$$

as a sum of sine waves can be written as:

$$\sum_j A_j \sin(\omega_j t + \varphi_j) \text{ or } \sum_j (a_j \cos\omega_j t + b_j \sin\omega_j t) ,$$

where

$$A_j = (a_j^2 + b_j^2)^{1/2} \text{ and } \sin^2\varphi_j + \cos^2\varphi_j = 1 .$$

Because sine and cosine functions of the same period are independent from each other any standardized time series can be approximated as a set of orthogonal functions

$$Y_t = \sum_j (a_j \cos\omega_j t + b_j \sin\omega_j t) + u_t ,$$

where u_t ~iid (0,1). The parameters a_j and b_j discriminating different time series can be obtained by least-square estimations

$$\hat{a}_j = \frac{2}{T}\sum_{t=0}^{T-1} Y_t \cos\omega_j t \text{ and } \hat{b}_j = \frac{2}{T}\sum_{t=0}^{T-1} Y_t \sin\omega_j t ,$$

respectively.

2.1.2 Harmonic Analysis

When the cycle length is determined by a previous analysis, harmonic analysis can be used to break down a periodic function into components, each expressed as a sine or cosine function, i.e., to model the cyclic component of a time series by estimating the mean, phase and amplitude of the sinusoid. However, usually several approximate cycle lengths that account for relatively large proportions of the variance in the data are identified, so harmonic analysis has to fit cycles with different periods until a cycle length or period is identified for which the goodness of fit is particularly strong, requiring the cycle length itself does not change over time.

2.1.3 Periodogram

The periodogram is a non-parametric method that can be seen as a generalization of harmonic analysis, where not just the parameters for one, but for a set of frequencies are estimated. The number of frequencies is determined by the length of the series. If the number of observation is even, the periodogram presents the total sum of squares of (T)/2+1 different frequencies. If the number is odd, the sum of squares of (T-1)/2+1 different frequencies are provided with an additional frequency of zero corresponding to an infinite cycle implying an infinite period. The sum of squares at each frequency can be obtained through

$$T\left(\frac{\hat{a}_j^2}{2}+\frac{\hat{b}_j^2}{2}\right) \text{ with } \hat{a}_j=\frac{2}{T}\sum_{t=0}^{T-1}Y_t\cos 2\pi f_j t \text{ and } \hat{b}_j=\frac{2}{T}\sum_{t=0}^{T-1}Y_t\sin 2\pi f_j t\,.$$

The power or variance of the series is computed by the total sum of squares divided by T. Equal spaced time series are inherently limited to detecting frequencies between 0 and 0.5 or 0 and π in radians, respectively. As illustrated in Figure 2.1.1, in a deterministic periodic time series (right upper panel), the peaks of the periodogram occur at corresponding frequencies (right lower panel), while for a completely random series (left upper panel) its variance is

approximately equally distributed across all possible frequencies, so illustrated in Figure 2.1.1 (left lower panel). Both series are characterized by a stable mean and variance, whereas the variance of a process changing over time is not defined, and likewise is its spectrum (Crato, 96). The periodogram can be obtained through

$$I(f) = \frac{1}{4\pi} \sum_{t=0}^{T-1} (\hat{a} \cos 2\pi ft + \hat{b} \sin 2\pi ft)^2 .$$

If we write

$$I(f) = \frac{1}{2T} \left(\hat{\alpha}_f^2 + \hat{\beta}_f^2 \right) 0 \le f \le \frac{1}{2}$$

then *I(f)* is called the *sample spectrum*, i.e., a *Fourier* cosine transformation of the autocovariance function estimate und thus an operation converting functions from time to frequency domain.

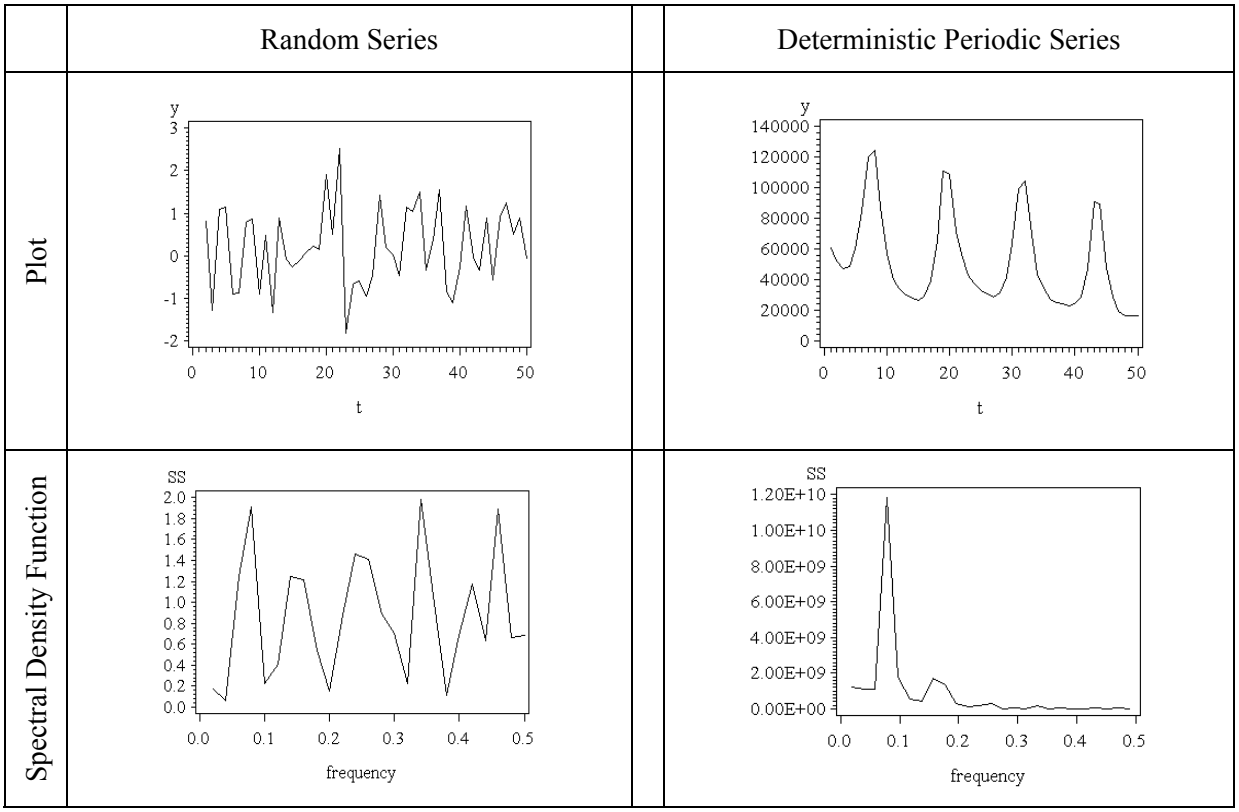


Figure 2.1.1. Spectral density functions (lower panel) of a purely random or white noise-process u_t ~iid(0,1) (left column), a process with a seasonal component and a downward trend (right column), and the plots of their corresponding series (upper panel).

2.1.4 Spectral Analysis

Although the periodogram analysis is a useful technique in assessing whether there is a strong cyclic component in a time series, it may break down when applied to time series with stochastic cycles, because its most serious liability is that the sampling errors associated with the sums of squares estimates are quite large. Hence, different methods of deriving the power spectrum from the periodogram and thus minimizing the estimates sampling error were developed. A common technique is the so called *smoothing,* a process in which each periodogram intensity is replaced by a *weighted average* that includes intensity estimates for a few neighboring frequencies. Different Smoothing procedures can be distinguished by the width of the *window*, i.e., the number of neighboring frequencies included in the weighted average, and the magnitude of the weights used. Some smoothing windows like the *Daniell window* give equal weight to all included frequencies, whereas others give more weight to frequencies near the center of the window than to frequencies near the edges, like the *Tukey-Hamming smoothing*. As a result, the graph of the weighting function may have different shapes. For example, a Daniell window looks like a rectangle, whereas other popular smoothing procedures have a bell shape, like the Tukey-Hamming smoothing. In general, determining the proper width of the window is usually a predicament, because detecting distinct cycles requires the use of narrow windows associated with large sampling errors. With increasing bandwidth more and more values are averaged and thus relatively adjacent peaks will be melded in one.

In general, the main objective of the frequency-domain method is the identification of major cyclic components explaining the variance in an observed time series by estimating the spectral density. Deterministic cycles appear in the periodogram as clear peaks whose height increase with sample size. For time series that are not strictly periodic, random changes of frequencies are typical, hence, the periodogram analysis is associated with problems.

Therefore smoothed versions of the power spectrum provide more reliable estimates of the spectral density function.

2.2 Time-Domain Analysis

The main objective in frequency-domain is the identification of major cyclic components explaining the variance in an observed time series. In time domain, the primary goal is to identify certain patterns to infer to the underlying data-generating process consisting of deterministic and/or stochastic components by means of *autocovariances* and *autocorrelations.* When the process has been identified, the dependency structure of the series can be modeled and forecasts generated.

2.2.1 Basic Notation and Principles

Many time series consist of elements that are serially dependent and the consecutive elements of the series can be described by specific, time-lagged or previous observations. Here, the underlying data generating mechanism is a so called *autoregressive process:* $Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + u_t$ where ϕ are the autoregressive model parameters, p specifies the order of the dependence, and u_t is a sequence of purely independent and identically distributed random variables or innovations. An autoregressive process will only be stable if the autoregressive parameters are within a certain range, else past effects would accumulate and the values of the successive Y_t would move towards infinity.

Furthermore, each observation of a time series can be affected by a past innovation that cannot be accounted for by the autoregressive component, i.e., a series is made up of a random error component and a linear combination of prior random shocks, formally called a *moving average process:* $Y_t = u_t - \theta_1 u_{t-1} - ... - \theta_q u_{t-q}$ where θ are the moving average model parameters, q specifies the order of the dependence, and, once again, u_t is a sequence of

purely independent and identically distributed random variables or innovations. Since a moving average equation can be rewritten or *inverted* into an autoregressive form, the range of the moving average parameters are restricted as well.

Finally, a process can be made up by autoregressive and moving average components as well. Such a process is called a *mixed autoregressive-moving average process:*

$$Y_t = \phi_1 Y_{t-1} + u_t + \theta_1 u_{t-1}.$$

A special type of a discrete stochastic process is an integrated process $Y_t = Y_{t-1} + a_t$, where the random part a_t can be generated by any ARMA process. The term 'integrated' implies that the impact of the random component on the series does not dissipate over time. As a result, the process shows instability in level. That is why the integrated process with $a_t \sim$iid N $(0,\sigma^2)$ is also called *random walk*.

2.2.2 Stationary vs. Nonstationary Processes

Identifying the underlying data generating mechanism of a stochastic process requires that its mean and/or variance is/are constant or independent of time, a quality formally called *stationary*. Before modeling, processes with time-varying mean or a time-varying variance have to be properly transformed and the proper transformation method depends on the cause of the instability. For example, a random walk process can also presented as the sum of random shocks where a particular shock is remembered forever. If a constant is added to the equation $Y_t = \alpha + Y_{t-1} + u_t$, the process is called *random walk with drift* or *stochastic trend* where α is formally called the drift parameter. Mean and Variance of random walks increase over time, a behavior called *nonstationary*. Computing the first differences $\Delta Y_t = Y_t - Y_{t-1}$ results in a *difference stationary* (DS) process. In general, if a time series has to be differenced *d* times to be transformed in a stationary series, the series is called *integrated* of order *d*. If, however, a non-constant term is added to a stochastic process, his behavior will be

determined by a *deterministic time trend* implying a nonstationary process with constant variance but changing mean. Here, stationarity is achieved by polynomial detrending of the data, which is then said to be *trend stationary* (TS).

While integrated time series exhibit an infinite memory, the deviations from the trend line of processes with deterministic trends do not contribute to the long-run development of the series and therefore misclassifications of the series type can lead to unnecessary detrending of stationary series (overdifferencing) or underdifferencing of nonstationary data. Although there has been some debate in the literature arguing that overdifferencing is a less serious error than underdifferencing, all kind of inappropriate transformations are consequential for subsequent statistical analyzes and should therefore be omitted. For an overview, consult Maddala & Kim (1998) or Stadnytska (2009c). Ayat and Burridge (2000), Elder and Kennedy (2001), and Stadnytska (2009c) describe testing strategies for determining whether the nonstationary-causing component of a time series is deterministic or stochastic.

2.2.3 Sample (Partial-) Autocorrelation Function

In the time domain the underlying data-generating mechanism can be identified by the *sample autocorrelation function* (SAC) and the *sample partial autocorrelation function* (SPAC). According to Bowerman and O'Connell (1993, p.442) the SAC of the original time series values $Y_b, Y_{b+1},\ldots,Y_n$ at lag k denoted by r_k is

$$r_k = \frac{\sum_{t=b}^{n-k}(Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{\sum_{t=b}^{n}(Y_t - \overline{Y})^2} \quad \text{where } \overline{Y} = \frac{\sum_{t=b}^{n} z_t}{(n-b+1)}$$

The SAC measures the linear relationship between time series observations separated by a lag of k time units (r_k will always be between –1 and +1).

The SPAC of the original time series values (see Bowerman and O'Connell,1993) p.453) at lag *k* is

$$r_{kk} = \begin{cases} r_1 & \text{if } k = 1 \\ r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j} & \\ r \dfrac{}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} & \text{if } k = 2,3,\ldots \end{cases}$$

The SPAC computes correlations between time series values that are *k* time periods apart after controlling for correlations at intermediate lags, i.e., partial autocorrelation between Y_t and Y_{t-k} after removing the effects of intermediate observations. The r_{kk} also ranges between –1 and +1 with values near ± 1 indicating a strong correlation.

Plotted against the lag length *k,* r_k and r_{kk}, respectively, give the so called *correlogram* of the sample autocorrelation SAC and sample partial autocorrelation function SPAC, respectively. A nonstationary stochastic process, specifically a random walk or integrated process of order one with an autoregressive term ϕ=0.5 and its SAC and SPAC is shown in Figure 2.2.1 (upper panel). If the SAC of nonstationary series exhibits meaningful autocorrelations up to a lag of about one-quarter the length of the time series - as illustrated - instability in the mean due to a stochastic or deterministic trend is indicated. Properly transformed, the SAC of the first differences $\Delta Y_t = Y_t - Y_{t-1}$ (lower panel) exhibits the typical signature of a pure autoregressive processes with no significant autocorrelations.

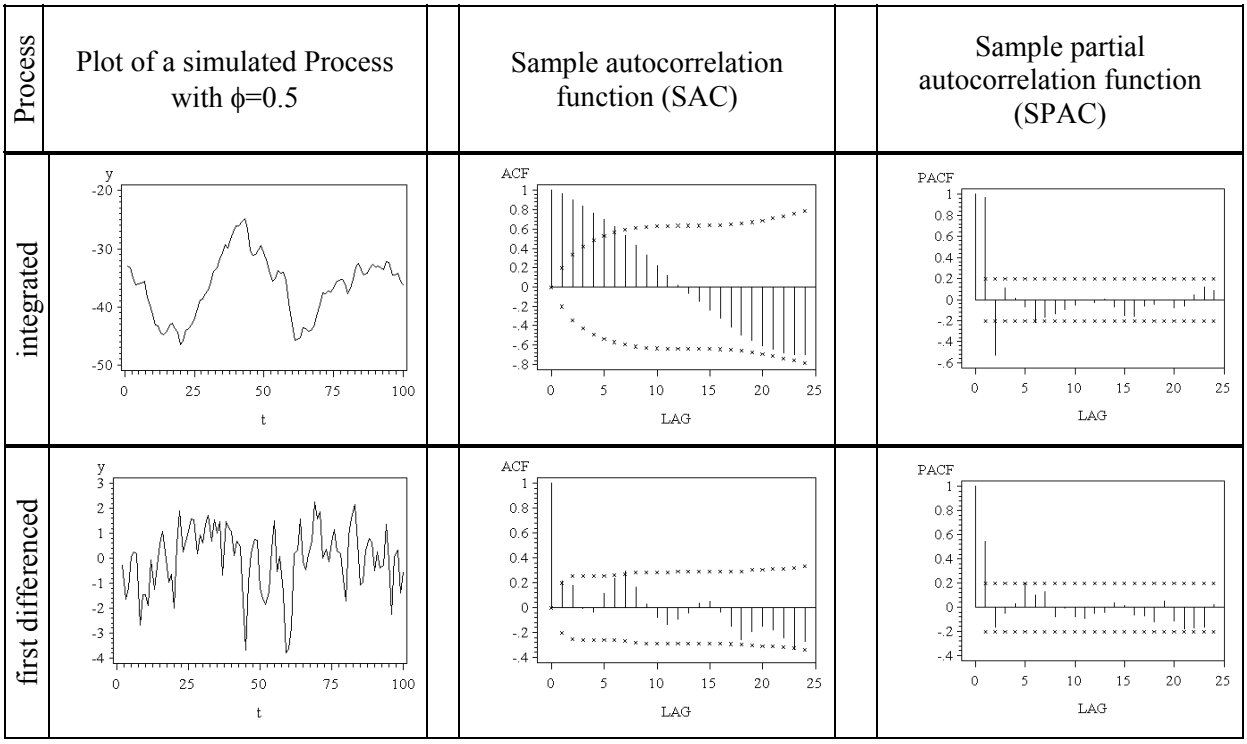


Figure 2.2.1. Simulated integrated process of order one and length *T=100* with one autoregressive term ϕ=0.5 (upper panel), its first differences $\Delta Y_t = Y_t - Y_{t-1}$ (lower panel), and their corresponding correlograms. The dotted line in the correlograms marks two standard errors.

Moving average terms, too, exhibit specific patterns or signatures, as illustrated in Figure 2.2.2. For example, a significant partial autocorrelation at lag *k* is equal to the estimated autoregressive component in an autoregressive process, while a significant autocorrelation is an indication for the presence of a moving average term. Significant correlations in correlograms are called *spikes.* The correlograms of the white noise process illustrated in Figure 2.2.2, exhibit no spikes at all since the underlying data-generating process is random. Hence, it contains neither autoregressive nor moving average terms. The SPAC of the mixed process indicates two moving average terms, one at lag 1 and one at lag 2, while in fact the simulated process is mixed with only one autoregressive and one moving average term. In general, the probability for the successful identification of correlative structures in stochastic processes by means of their SAC and SPAC, respectively, declines with the complexity of the underlying data-generating process.

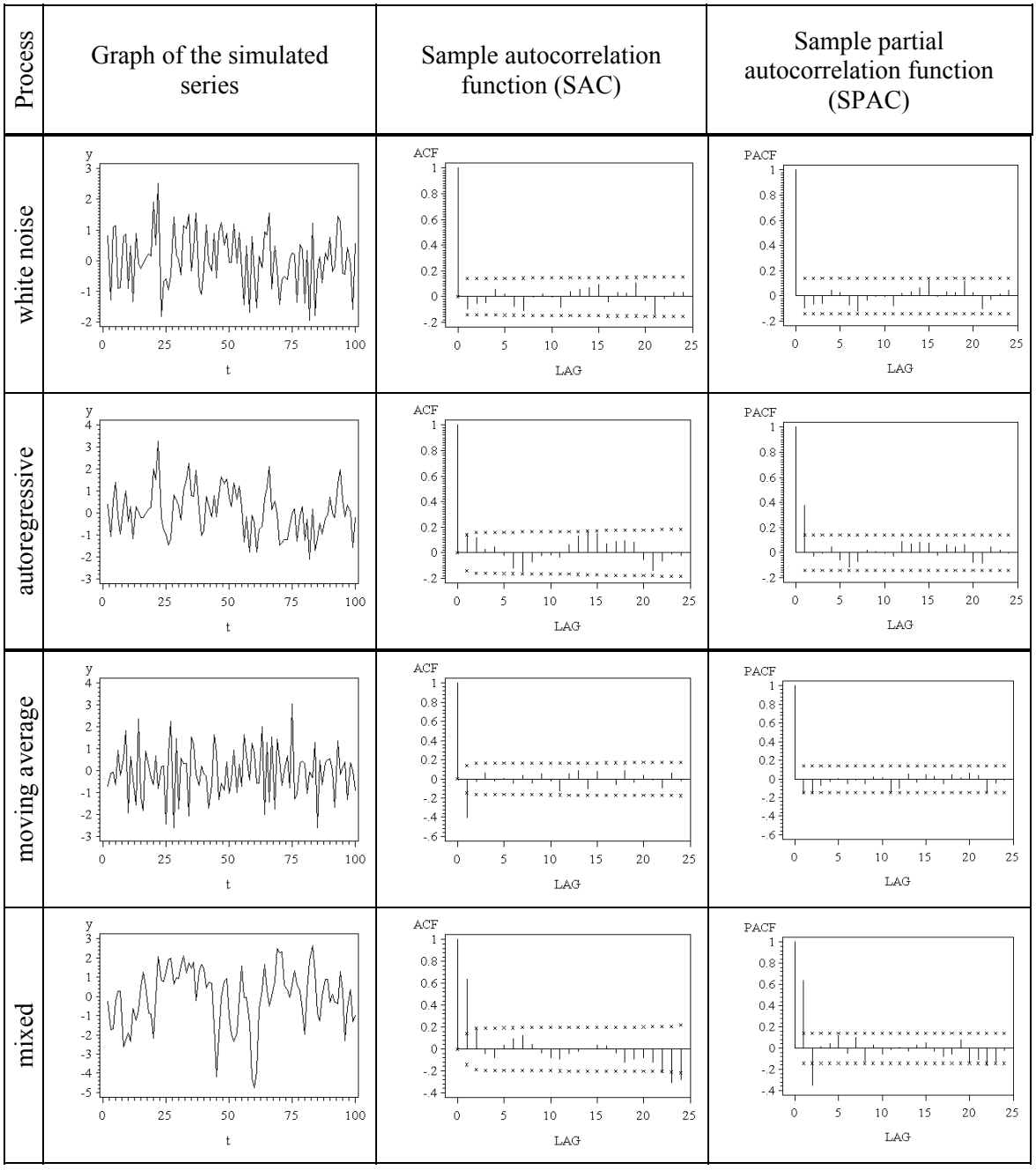


Figure 2.2.2. Simulated stationary processes, from top to bottom: white noise, a simple autoregressive process with ϕ=0.5, a simple moving average process with θ=0.5 and a mixed process with ϕ=0.5 and θ=0.3. The dotted line in the correlograms marks two standard errors.

2.2.4 Box-Jenkins ARIMA Modeling

The ARIMA strategy proposed by Box and Jenkins (1970) offers a large class of models to describe a wide spectrum of time series behavior. Based on a systematic approach a

preliminary model is fitted and validated by statistical tests. The procedure is repeated as necessary until a model is chosen that best fits the data.

There are three basic model components: AR (autoregressive), MA (moving average), and a combined ARMA (autoregressive moving average) part (for autoregressive, moving average and mixed processes see Chapter 2.2.1, p.9). When regular differencing (see Chapter 2.2.2, p.10) is applied together with AR and MA, they are referred to as ARIMA, with an 'I' short for 'integrated', which references the differencing procedure. Additionally, there are model components available describing seasonal phenomena, i.e., the so called SAR (seasonal autoregressive) and SMA (seasonal moving average), as well as for modeling the interdependence of multivariate time series (VAR-models).

In ARIMA(*p,d,q*) models, AR terms are represented by the parameter *p* reflecting the number of preceding observations influencing the current value. The parameter *q* of the MA term assesses how many previous random shocks must be taken into account to capture the dependency present in the time series. Finally, the parameter *d* refers to the order of differencing needed to stabilize the time series. For example, an ARIMA(1,1,1) model represents a nonstationary integrated process of order 1 with one autoregressive and one moving average term, that becomes stationary by taking its first differences $\Delta Y_t = Y_t - Y_{t-1}$.

The *Box-Jenkins Method* is one of the most widely used methodologies for the analysis of time series data. It is popular because of its generality, it can handle any series, stationary or not, with or without seasonal components, and it is implemented in a wide range of statistical software (Maddala & Kim, 1998, p.17). The minimum prerequisites for precise modeling by the Box-Jenkins method are around 200 observations assessed on a five-point Likert-scale, as demonstrated by a simulation study (Braun & Werner in Werner, 2005, p. 165ff). The Box-Jenkins method consists of five consecutive steps, i.e., four stages of the so

called *model building* process plus a forecasting stage. However, a plot of the original data should always be run as a starting point.

Differencing to achieve stationarity: The correlograms of a nonstationary series does not decay (see Chapter 2.2.2). After determining whether the nonstationary-causing component of a time series is deterministic or stochastic (see testing strategy in Stadnytska, 2009c), successive differencing of the original series, providing the cause of the nonstationary is a stochastic trend, can be taken until the correlogram dampens.

Identification of a tentative model: The appropriate orders of the AR and/or MA components are obtained by examining the correlograms. For a basic AR model with just one autoregression term, that is, p=1, the theoretical autocorrelation function decays exponentially or with a damped sine wave or both, while the theoretical partial autocorrelation function exhibits a significant spike at lag 1 and cuts off after that. For a basic MA model with q=1, the theoretical autocorrelation function exhibits a significant spike at lag 1 and cuts off after that, while the theoretical partial autocorrelation function declines exponentially. For mixed models with order p and q both correlograms decline exponentially. When analyzing an empirical series, a tentative or preliminary model is identified by comparing the theoretical correlograms with the behavior of the SAC and SPAC of the empirical data. Obviously, this step involves more of a judgmental procedure than the use of any clear-cut rules.

Estimation of the model: Having identified a preliminary model, the next step is to estimate the corresponding parameters through *conditional least squares* (CLS), *unconditional least squares* (ULS) or *full maximum likelihood* (ML) algorithms by minimizing the differences between the empirical data and the 'theoretical' data generated by the model calculated for the different values of ARMA coefficients. Comparing the performance of the different estimation algorithms for different models, Fang (2005) concludes that tests based on CLS are more reliable than those based on ML or ULS.

Diagnostic checking or identification revisited: Having estimated an ARIMA model, it is necessary to revisit the question of identification to see if the selected model can be improved: Some of the estimated parameters may be insignificant or there may be several competing models. There are two *penalty function criteria* often used that reflect the closeness of fit and the number of parameters estimated. One is the *Akaike Information Criterion* (AIC) and the other is the *Schwarz Bayesian Information Criterion* (BIC). If p is the total number of parameters estimated, we have

$$AIC(p) = n\log\hat{\sigma}^2 + 2p \quad \text{and} \quad BIC(p) = n\log\hat{\sigma}^2 + p\log n.$$

Here n is the sample size and $\hat{\sigma}^2 = SSE/(n-p)$. From several competing models the one with the lowest AIC or BIC is chosen. In addition, the correlation pattern of the residuals of the chosen model have to be checked for serial correlations. Box and Pierce (1970) suggest looking at not just the first order autocorrelation of the residuals but of all orders. They suggest calculating

$$Q = T\sum_{r=1}^{m}\rho_k^2,$$

where ρ_k is the autocorrelation of lag k and T is the number of observations in the series. If autocorrelations up to certain lags are simultaneously equal to zero, the residuals are white noise and the chosen model under investigation is appropriate.

Forecasting: In general, there are two basic approaches to forecasting time series: the self-projecting time series and the cause-and-effect approach. Cause and effect methods' forecasts are based on underlying series that are believed to cause the behavior of the original series. Self-projecting time series use only the observed time series data to generate forecasts. The latter approach is typically less expensive to apply and requires far less data and is useful for short to medium-term forecasting. Gujarati (2003) differentiates five forecasting approaches based on time series data: exponential smoothing, single-equation regression

methods, simultaneous-equation regression models, the Box and Jenkins method, and vector autoregression. The univariate version of the Box and Jenkins forecasting technique is a self-projecting forecasting method. For example, the lead-one forecast is computed by substituting, into the formula defining the model, the observed (past) value with the estimates of the past error term. These errors terms can be estimated recursively: they are simply the past forecast errors. To compute the optimal forecast beyond lead one, forecasts for future observations and zeros for future error terms are simply substituted.

To sum up, the Box and Jenkins methodology represents a flexible class of models that can be used to represent the behavior of a wide class of time series. It is popular because of its generality, it can handle any series, stationary or not, with or without seasonal elements, and it is well-documented by a variety of statistical software. It is perhaps the last factor that contributed most to its popularity. See also Box et al. (1994), Bowerman and O'Connell (1993), Brockwell and Davis (2002), and Makridakis et al. (1998) for a detailed treatment of the Box-Jenkins methodology.

2.2.5 Automated Model Identification

The model identification strategy of the Box and Jenkins methodology, as described in Chapter 2.2.4, consists of examining the behavior of the sample autocorrelation and partial autocorrelation functions and compares it with the theoretical ARIMA patterns, involves more of a judgmental procedure than the use of any clear-cut rules. Not only for accurate forecasting, but also in theory testing, where model identification represents the main objective of the analysis, serious conceptual consequences can be caused by model misspecification. For example, autoregressive processes are characteristic for internal temporal regularity, however, moving average processes predominantly represent systems determined by external and occasional events.

Autoregressive structures are typically found in processes representing addictive behavior. For example, Palaniappan (2005) proposed a second order AR model to discriminate alcoholics using single trial gamma band VEP signals distracted from the Electroencephalogram (EEG) from alcoholic (n=10) and control subjects (n=10). Rosel and Elósegui (1994) examined the daily cigarette-consumption of nine male and 20 female smokers over a 12-week period. 75% of the series assessed were identified as autoregressive models of order one typical for daily smokers when nicotine addiction plays a prominent role in maintaining the behavior. Velicer, Redding, Richmond, Greeley and Swift (1992) provide an excellent example how several popular tobacco-consumption models representing highly addictive smoking behavior are well-described by AR models. Occasional substance use, however, seems to depend more on intermittent and external events, and is predominantly represented by moving average models.

In addition to the model identification technique of the Box and Jenkins methodology, *automated model identification methods* may help the researcher to reduce the ambiguity often experienced by the interpretation of the correlograms. Velicer and Harrop (1983) demonstrated that only 28% of simulated series were correctly classified by highly trained judges using the Box and Jenkins approach. Therefore an abundance of attempts have been made during the last three decades to develop more reliable methods for identifying the underlying pattern of stochastic processes.

Choi (1992) published a survey classifying automated methods into three categories: penalty function methods (e.g., BIC or MINIC of Rissannen, 1978; Schwarz, 1978; AIC of Akaike, 1974), innovation regression methods (e.g., HR of Hannan & Rissanen, 1982; KP of Koreisha & Pukkila, 1990), and pattern identification methods (e.g., Corner Method of Beguin, Gourieroux & Montfort, 1980; ESACF and SCAN of Tsay & Tiao, 1984, 1985).

Stadnytska, Braun, and Werner (2006) evaluated automated model identification methods for stationary processes and provided a review on Monte Carlo simulations employing further identification methods. They compared the performance of three automated procedures available in current versions of SAS for Windows, SCAN, ESACF and MINIC, in identifying the internal structure of the simulated processes. Regardless of model, parameterization and sample size, the automated methods correctly identified the true model in about 60% of trials conducted or selected parsimonious nearly equivalent mathematical representations. In a subsequent study, Stadnytska, Braun and Werner (2006b) examined the performance of SCAN and ESACF for nonstationary processes and found 79% of correct identifications for SCAN and 80% for ESACF. For some models and parameterizations, the accuracy of SCAN and ESACF was disappointing. As a result, an elaborated strategy for model selection combining different techniques was developed and demonstrated on two empirical examples.

3 LONG-RANGE DEPENDENCE

Time series with long-range dependencies are widespread in nature and have been extensively studied in hydrology and geophysics (Hurst, 1965; Mandelbrot et al., 1968, 1975). More recently, time series with so called *long memory* have been observed in DNA sequences (Peng et al., 1992, 1994), cardiac dynamics (Peng et al., 1993; Makikallio, 1999), internet traffic (Willinger et al., 1996), meterology (Koscielny-Bunde et al., 1998; Montanari, 2000), geology (Malamud and Turcotte, 1999) and even ethology (Alados and Huffman, 2000), where behavioral sequences of wild chimpanzees in Tanzania were shown to exhibit long-range correlations. In economics and finance, long-range dependency, too, has a long history (for a review see Baillie and King, 1996; Mandelbrot, 1997). In psychology processes with long memory, i.e., correlated noise were first observed in controlled cognitive experiments by Gilden, Thornton & Mallon (1995), such as mental rotation, lexical decision, shape and color discrimination.

3.1 Definition

According to Leite and Rocha (2007), a stationary process $Y(t)_{t\in z}$ is said to have long-range correlations, if there exist a real number $\gamma \in]0,1[$ and a constant $c_p \triangleright 0$ such that

$$\rho(k) \approx c_\rho |k|^{-\gamma}, k \to \infty, \text{ where } \rho(k) = \frac{\text{cov}[y(t), y(t+1)]}{\text{var}(y(t))}$$

is the autocorrelation function. Alternately, a stationary process $Y(t)_{t\in z}$ is said to have long-range correlations if there exists a real number $\beta \in]0,1[$ and a constant $c_f > 0$ such that $f(\omega) \approx c_f |\omega|^{-\beta}, \omega \to 0$, where $f(.)$ is the spectral density function.

3.2 Modeling Long-Range Dependence

If a series exhibits long memory, there are persistent temporal dependencies between observations widely separated in time resulting in hyperbolically decaying autocorrelations and low-frequency spectral distributions. On the other hand, quickly declining autocorrelations and predominantly high-frequency spectral distributions describe the low-order correlation structure of a series with short-term dependence.

3.2.1 ARFIMA Methodology

In the Box and Jenkins methodology, the ARIMA model parameters *d, p* and *q,* described in Chapter 2.2.4, p.14, are indicative of the correlation pattern of a stochastic process. An ARIMA (0,0,0) or white noise is a random sequence with no memory at all. ARIMA (*p*,0,*q*) models with small values of *p* or *q* represent stationary processes that are predictable only from their immediate past, that is, they possess only short memory. Their autocorrelations decay quickly as the number of intervening observations increases. However, an autoregressive process of order one with ϕ=.99 can generate autocorrelations over hundreds of lags, which is plausible, since an AR(1) model with ϕ=1 can be written as an ARIMA(0,1,0) model representing an integrated process of order 1 that is equivalent to the sum of random terms $Y_t = \sum a_t$. As a consequence, the impact of a particular random term does not dissipate over time resulting in persistent autocorrelations and infinite memory. Finite memory or long-range dependence, however, can be modeled through the differencing parameter *d*, if we allow it to take on continuous values. These processes can be modeled by the *autoregressive fractional integrated moving average* (ARFIMA) methodology suggested by Granger & Joyeux (1980) and Hosking (1981), allowing the simultaneous modeling of short- and long-memory components, where low-lag autocorrelations can be modeled through the parameters *p* and *q* while the fractional differencing parameter *d* captures the long-range dependencies of

a series. The effect of d on distant observations decays hyperbolically, in contrast to the exponential decay due to the effects of the short-range parameters p and q.

An ARFIMA process is both stationary and invertible when $-0.5<d<0.5$. For $0.5\geq d\leq 1$, the process is nonstationary with noninvertible ARMA representations. Values of d from the interval (-0.5;0) suggest antipersistence, that is, adjacent values are negatively correlated in contrast to the positively correlated values of persistent processes with $d>0$. Finite long memory can be modeled with $0<d<0.5$. The upper bound $d<0.5$ is needed, because for $d\geq 0.5$ the process is nonstationary, in particular, the usual definition of the spectral density of Y_t would lead to a nonintegrable function, and the case of $d\geq 0.5$ can be reduced to the case $-0.5<d\leq 0.5$ by taking appropriate differences. For example, the first differences of a nonstationary ARFIMA process with d=0.7 result in a stationary solution with d=-0.3 (for mathematical details see Beran, 1994, p.61).

Figure 3.2.1 illustrates the behavior of long memory processes without short-term components. From top to bottom, the SAC of different ARFIMA(0,d,0) processes with T=400 and varying values of d demonstrates the typical characteristics of series with long-range dependence: hyperbolically decaying autocorrelation and an increasing spectral density function as the frequency tends to zero are more prominent the higher the value of d is, i.e., the stronger the long-range dependence is.

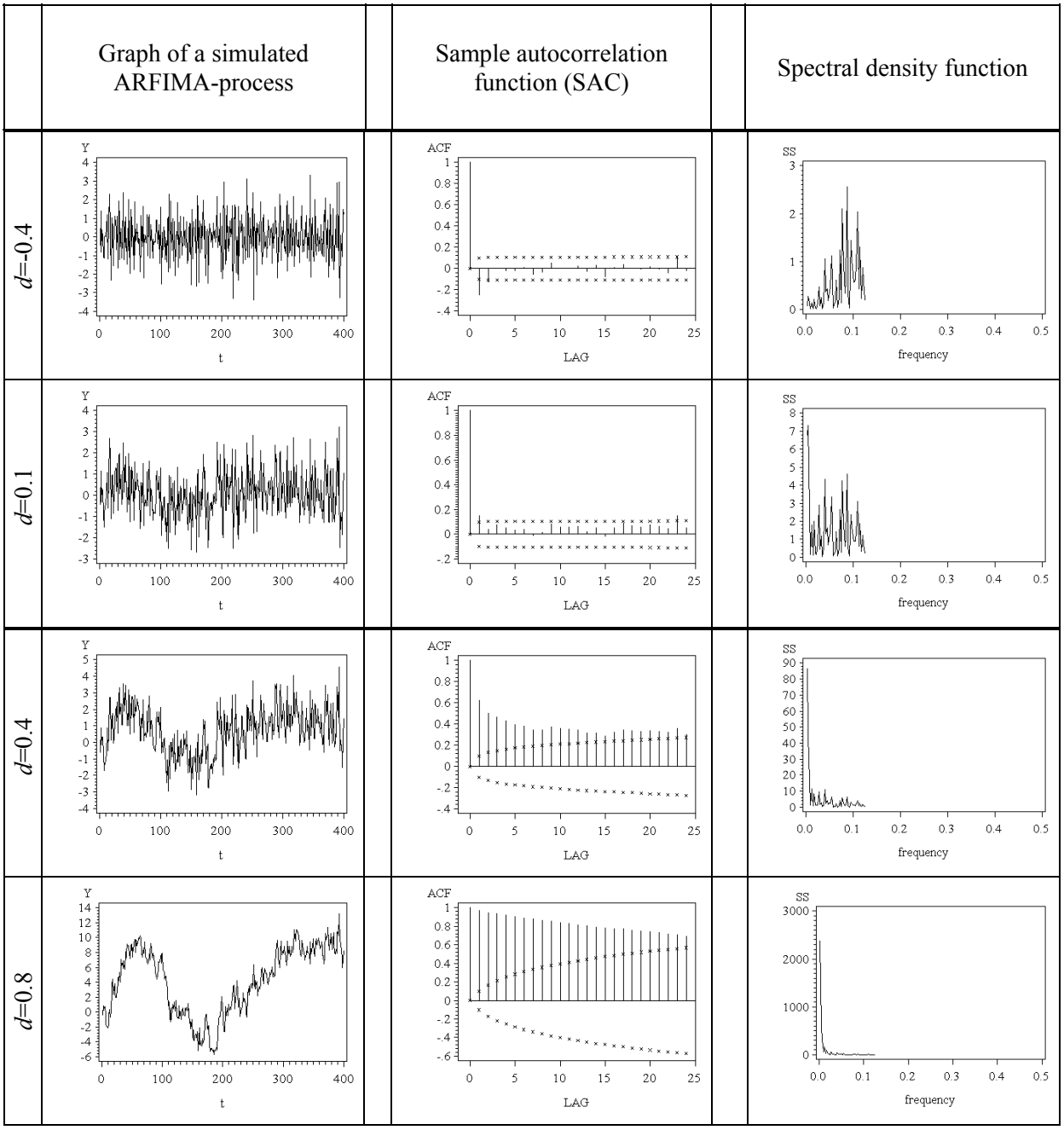


Figure 3.2.1. Plots of simulated stationary ARFIMA (0,d,0) process with varying *d* and *T*=400 (left column), their autocorrelation (center column) and spectral density function (right column). The dotted line in the autocorrelation function marks two standard errors.

3.2.2 Fractal Analysis

Alternatively, persistent processes may be modeled within the framework of Fractal Geometry, which is rooted in the works of the late 19th and early 20th century mathematicians, who found their fancy in generating complex geometrical structures from simple objects like a

line, triangle or square, the so called *initiators,* by applying a simple rule or transformation, the so called *generator,* in an infinite number of iterative steps resulting in a complex structure rich in detail at every scale of observation (Eke et al., 2002, p.82), a concept commonly known as self-similarity. Self-similar processes were first introduced by Kolmogorov (1941), but statisticians did not seem to have been aware of the existence or statistical relevance of such processes, until Mandelbrot and van Ness (1968) introduced them into statistics. Self-similarity, where the scaling is identical in all directions, needs to be distinguished from self-affinity, where in one direction the proportions between the enlarged pieces are different from those in the other (Eke et al., 2002, p.84). For example, physiological time series are self-affine temporal structures due to a restriction of range or observation time or by the physical nature of the processes (Eke et al, 2000, p.413). For ideal mathematical fractal processes, however, the range over which there is self-similarity is infinite.

In this context, self-similarity is defined in terms of the distribution of the process (Beran, 1994, p.48): Let Y_t be a stochastic process with continuous time parameter t. Y_t is called self-similar with the self-similar parameter H, if for any positive stretching factor c, the rescaled process with time scale ct, $c^{-H}Y_{ct}$ is equal in distribution to the original process Y_t. Thus, typical sample paths of a self-similar process look qualitatively the same, irrespective of the distance they are looked at.

The self-similar parameter H or Hurst exponent representing the probability that an event in a time series is followed by a similar event deviates from 0.5, can be any real number in the range (0;1). It is named after the famous hydrologist Hurst (1965) investigating the question how to regulate the flow of the Nile River. For each value of $H \in (0,1)$, there is exactly one Gaussian process X_i that is the stationary increment of a self-similar process Y_t. X_i is called *fractional Gaussian noise* or fGn, the corresponding but nonstationary self-similar

process Y_t is called *fractional Brownian motion* or fBm (Beran, 1994, p.55). Differencing fBm creates fGn and summing fGn produces fBm, i.e., the cumulating summation of an fGn signal results in an fBm signal. Hence, the related processes are characterized by the same Hurst exponent. FGn and fBm are fractal processes that can be persistent (H>0.5) or anti-persistent (H<0.5). Anti-persistence implies negative correlations and persistence positive correlations between the successive increments of a time series. Mandelbrot and van Ness (1968) introduced fBm as a generalization of ordinary *Brownian motion* with H= 0.5, which is a well-known stochastic process that can be represented as the random movement of a single particle along a straight line. Mathematically, Brownian motion is the integral of Gaussian noise.

Figure 3.3.2. illustrates the different change in behavior of stationary fGn (upper panel) and nonstationary fBm (lower panel) processes and T=400 depending on the magnitude of H by means of simulated data. Both signal types become unstable, the more H approaches one. However, if instability in levels is disregarded, fGn signals with large H coefficients are rougher, whereas fBm signals become smoother the more H approaches zero. For an overview on recent developments in fractal analyzes in psychological and behavioral research, see Delignières et al. (2005).

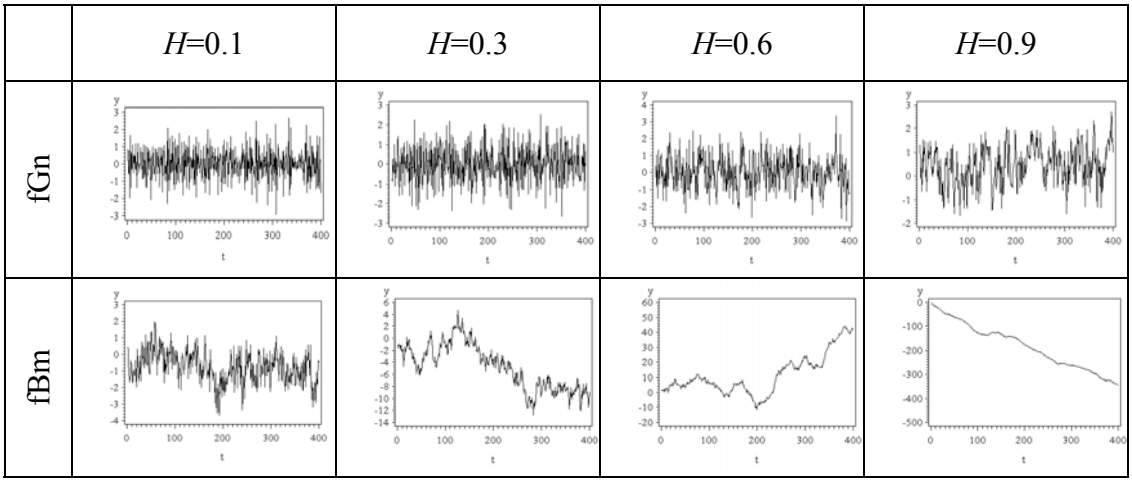


Figure 3.2.2. Illustration of the change in behavior of fractal signals as H approaches one. Simulated fGn processes (upper panel), simulated fBm processes (lower panel), each with T=400 observations.

3.3 Identifying long-range dependence

The common aim of ARFIMA or fractal analysis is to identify the typical signatures like a hyperbolically decaying autocorrelation function, a spectral density increasing without limit as the frequency tends to zero and self-similarity. Fractal methods are diverse, but their approaches have one thing in common: they employ a power law relationship on fitting their proposed model to data pairs of log frequency versus log power for finding the scaling exponent (see Eke, 2002) from the regression slope. The working hypothesis is, that behind the seemingly random but in fact complex fluctuations of a signal one may find a time-invariant mechanism that one wishes to describe with the smallest possible numbers of parameters.

3.3.1 Time Domain Methods

In time domain, a common technique, as described in Chapter 2.2.4, p.14, for identifying short-range dependence is by means of examining the correlograms. For example, exponentially decaying sample autocorrelation functions and significant spikes are indicative of autoregressive components in a process. Long memory is characterized by a slow decay of the correlations proportional to k^{2H-2}. For $0.5<H<1$, the SAC should therefore exhibit a slow decay, in particular, follow a hyperbolic curve proportional to k^{2H-2}. However, it is difficult to distinguish between hyperbolic or exponential curves especially for different values of *H*. For example, the closer *H* gets to one, the slower the autocorrelation plot decays. Another difficulty is, that long memory is an asymptotic notion and correlations at high lags can not be estimated reliably. A more suitable plot can be obtained by taking the logarithms on both sides of the correlogram. If the asymptotic decay is hyperbolic, for large lags the points in the plot should be scattered around a straight line with a negative slope approximately equal to $2H-2$. In contrast, for short memory processes, the log-log correlogram should show

divergence to minus infinity at a rate that is at least exponential. According to Beran (1994, p. 92), autocorrelation correlograms are mainly useful in cases where long-range dependence is strong or for very long time-series. A detailed description of the usefulness of the sample correlations and partial correlations for detecting long memory can be found in Beran (1994, p. 89).

A very different approach, introduced by Peng et al. (1994), is the so called *Detrended Fluctuation Analysis,* a modified root mean square analysis of highly nonstationary data, by removing nonstationary trends from long-term correlated time series. First, the signal is summed and the mean is subtracted, then the local trend is estimated in non-overlapping boxes of equal length, using a least square fit on the data. For a given box-size n the fluctuation is determined as a variance upon the local trend, i.e., the root mean square of its differences, and calculates the fluctuation

$$F_n = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(Y_j - Y_{j,n})^2}.$$

For self-similar fBm processes of length N with non-overlapping windows of size n, the fluctuation F depends on n in a power law fashion, where the slope of the line relating to log $F(n)$ to log n determines the scaling exponent α. If the original series is an fGn signal, then Y_j will be an fGn signal and in this case $\alpha = H$, whereas if the original series is an fBm signal, Y_j will be a summed fBm and $\alpha = H+1$.

Many more heuristic methods to detect long-range dependence were suggested. Well known is the *R/S* statistic Q, which was first proposed by Hurst (1965). The *R/S* statistic is computed and plotted against the lag k. For long memory processes, the points in the *R/S* plot should be scattered randomly around a straight line with a slope $H > 0.5$, for sufficiently large lags k. According to Beran (1994), the estimate of the *R/S* statistic can be misleading if there

is a slowly decaying trend in a nonstationary time series with only short-term memory, that is, the values of Q are the same as in a process with long memory.

Another method is the log-log plot of $\mathrm{var}(\overline{X}_n)$ versus *n,* the so called *variance plot*, based on one of the properties of long memory processes that the variance of the sample mean converges slower to zero than n^{-1}. For long memory processes the points in the plot for large lags are expected to be scattered around a straight line with a negative slope approximately equal to *2H-2.* In the case of short-range dependence or independence the slope is steeper. In principle, the weaknesses of the variance plot are the same as for the *R/S* plot and the log-log correlogram.

3.3.2 Frequency Domain

In the frequency domain, the power spectrum density analysis (PSD) models self-similarity of long memory structures by means of the power exponent β. Time series with long-range dependence have self-similar power-spectra with a spectral density proportional to the reciprocal of the frequency $S(f) \propto 1/f^{\beta}$, where f is the frequency and $S(f)$ the corresponding squared amplitude. Plotted on a log-log scale log $S(f)$ versus $log(f)$, the power exponent β is estimated by calculating the negative slope $-\beta$. However, the signal has to be preprocessed before applying the fast Fourier transform *FFT*, that is, the mean has to be subtracted and each value of the series has to be multiplied with a parabolic window (so called windowing)

$$W(j) = 1 - \left(\frac{2j}{N+1} - 1\right)^2 \text{ for } j=1,2,\ldots N.$$

Thirdly, bridge detrending (so called endmatching) is performed by subtracting from the data the line connecting the first and last point of the series. Finally the fitting of β excludes the high-frequency power estimates (f>7/8 of maximal frequency), termed low. This method was proven by Eke et al. (2000) to provide more reliable estimates of the spectral index or power

exponent β, and was designated as ${}^{low}PSD_{we}$. Combining the various transformation leads to different results (see Chapter 5, p.52, for a thorough investigation).

Illustrated in Figure 3.3.1 (left column) is the power spectrum of a white noise process. It is an approximately straight line with a slope of zero, thus, β=0. The spectral density of a random walk, otherwise known as Brownian or brown noise because it is the kind of signal produced by Brownian motion, is proportional to $1/f^2$, thus, β=2 (center column). Time series with β=1 are called pink, flicker, burst or $1/f$ noise (right column). In general, differencing increases the power spectrum slope by 2, and integrating decreases the slope by 2.

To summarize, for a process without memory β=0, with infinite memory β=2, and in the case of long memory, β can vary from 0.5 to 1.5. For short memory processes, the log-log power spectrum in not a straight line because random variation at the lower frequencies leads to a breakdown of the linear relationship and results in a flat plateau at frequencies close to zero

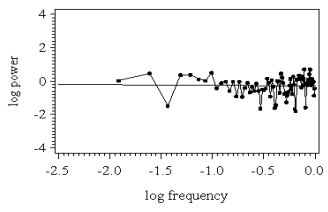
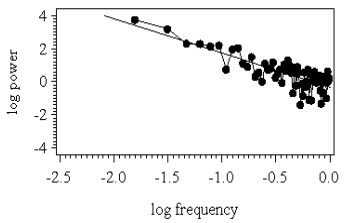
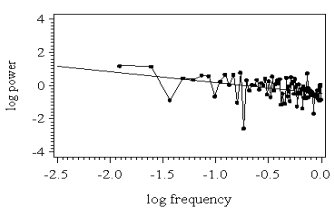
ARIMA (0, 0, 0) WHITE NOISE no memory	ARIMA (0, 1, 0) RANDOM WALK infinite memory	ARFIMA (0, 0.4, 0) long memory
		

Figure 3.3.1. Log-log power spectra for series with T=400.

Eke et al. (2000) and Delignières et al. (2006) recommended to use the periodogram analysis for distinguishing fGn and fBm signals since it can be applied to both stationary fractal noises and nonstationary motions. Signals with $-1<\beta<+1$ and $1<\beta<+3$ closely resemble

fGn and fBm signals, respectively. For a detailed description of the periodogram method see Beran (1994), Gilden (2001), Handel and Chung (1993), Kasdin (1995), and Pilgram and Kaplan (1998).

3.3.3 Relation between Measures

Chapters 3.3.1, p.27, and 3.3.2, p.29, demonstrated the diversity of methods modeling and identifying fractal time series. A detailed description of the relationship between exact fractal time series generated according to the dichotomous fGn/ fBm model of random signals and its fractal descriptors, H and β can be found in Eke et al. (2000). For example, fractal noises and motion of the same H have β values that differ by two, leading to the equations

$$H = \frac{\beta + 1}{2} \text{ for fGn and } H = \frac{\beta - 1}{2} \text{ for fBm.}$$

The relationship of H and d is given by Hosking (1984), who demonstrated that self-similar hyperbolically decaying autocorrelations of fractional series can be parsimoniously modeled by means of the differencing parameter d of the Box-Jenkins ARIMA methodology. The ARIMA (0,d,0) process can be thought of as the result of applying fractional differencing to a white noise series. The process is a kind of fractional noise, and its properties are similar to those of fractional Gaussian noise. If the H parameter of fGn satisfies $H=d+0.5$, then both fGn and the ARIMA (0,d,0) process have correlations ρ_k that behave asymptotically as k^{2d-1} (Hosking, 1984, p.1900). Since Gaussian white noise, the process without memory, is the ARFIMA (0,0,0), and the discrete analogue of ordinary Brownian motion is the ARFIMA (0,1,0), the following relationships hold:

$$\beta = 2d \text{ and } H = \frac{2d + 1}{2} \text{ in the fGn case.}$$

The *Detrended Fluctuation Analysis* proposed by Peng et al. (1993), as introduced in Chapter 3.4.1, p.39, measures the extent of long-range correlations in time series by

calculating the slope of a double logarithmic plot of the average fluctuation F as a function of the interval size n. The relation between α and the Hurst coefficient is $H=\alpha$ for fGn and $H=\alpha$-1 for fBm, while $\beta = 2d = 2\alpha$ - 1 for both fractal signals.

Figures 3.3.2 and 3.3.3 illustrate the typical behavior of the SAC, the periodogram and the log-log power spectrum of simulated fractal signals with T=400 and varying H. The series in Figure 3.3.2 are stationary fractal noises. For H=0 the series is a pure random noise und thus, ARIMA (0,0,0) with d=0 and a slope close to zero, that is, β=0. H=0.5 separates anti-persistent (H<0.5) and persistent (H>0.5) series, as demonstrated by corresponding negative or positive spikes in the SAC (left column). Extreme values of H are associated with large negative and positive slopes, respectively, thus a power exponent close to $|1|$ due to a concentration of power in high or low frequencies. Finally, as H grows larger, the stationary series approximate the border of nonstationarity, i.e., d=0.5, resulting in persistent autocorrelations and infinite memory. The simulated series illustrated in Figure 3.3.3. are nonstationary fBm signals with infinite memory, a SAC dying down extremely slowly with predominating low frequencies in the spectral density plot, especially the closer H gets to 1. Brownian motion with H=0.5, that is, the Random Walk or otherwise known as an ARIMA(0,1,0) process separates anti-persistent and persistent fBms.

For H=0.5, the observations of an fBm are not independent, as it is in the fGn case, but its successive increments $\Delta Y = Y_t - Y_{t-1} = u_t$ are uncorrelated. Brownian motion with H = 0.5 separates anti-persistent and persistent fBms, but the pattern of the SAC does not allow the identification of positive or negative correlations between adjacent values, as in the fGn case, since the SAC of a nonstationary series dies down extremely slowly.

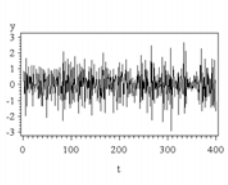
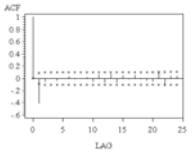
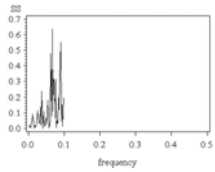
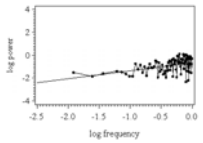
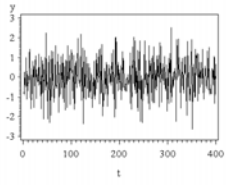
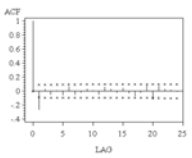
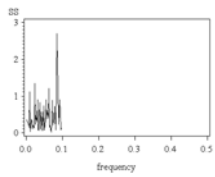
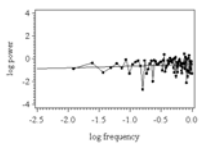
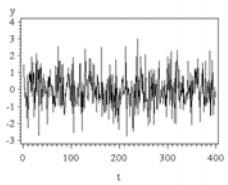
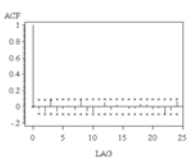
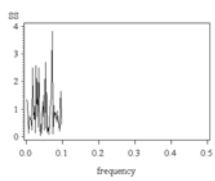
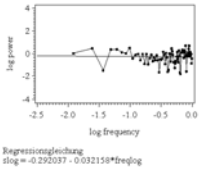
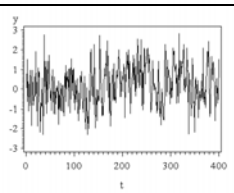
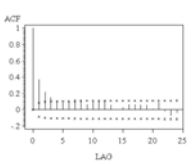
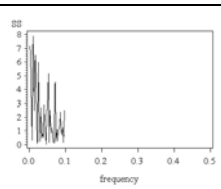
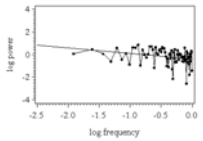
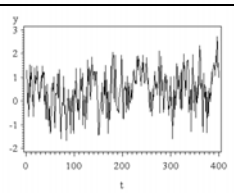
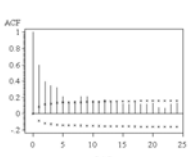
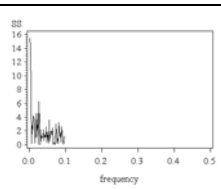
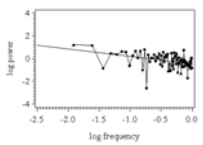
H	Simulated fGn-process	Sample autocorrelation function (SAC)	Spectral density function	Log-Log power spectrum
H=α=0.1 d=-0.4 *β=-0.8*				
H=α=0.3 d=-0.2 *β=-0.4*				
H=α=0.5 d=0 *β=0*				
H=α=0.7 d=0.2 *β=0.4*				
H=α=0.9 d=0.4 *β=0.8*				

Figure 3.3.2. Graph, SAC, spectral density and log-log power spectra with slope for simulated stationary fGn signals with T=400.

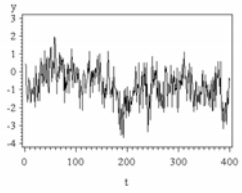
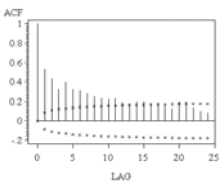
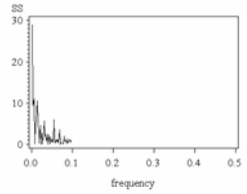
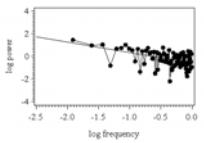
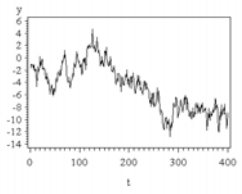
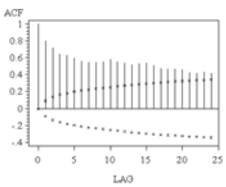
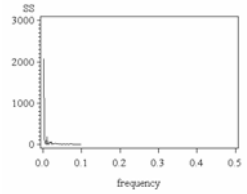
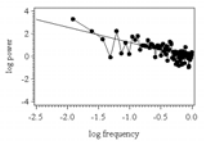
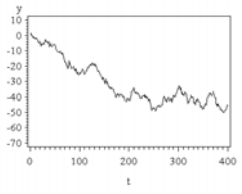
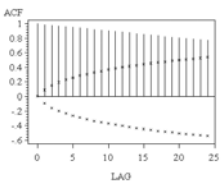
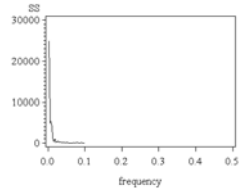
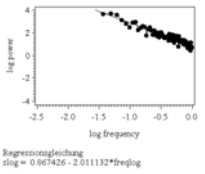
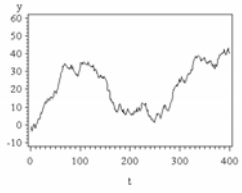
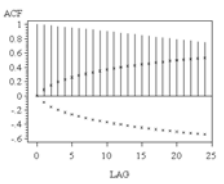
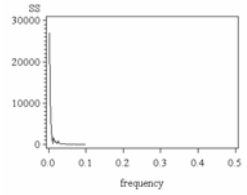
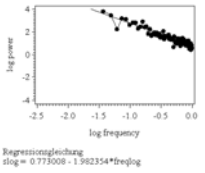
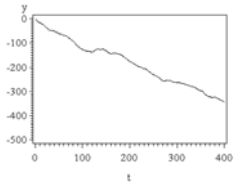
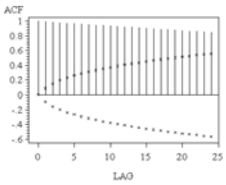
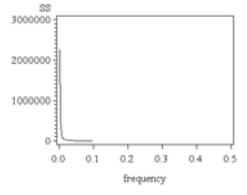
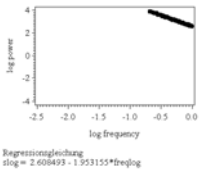
H	Simulated fBm - process	Sample autocorrelation function (SAC)	Spectral density function	Log-Log power spectrum
H=0.1 α=1.1 d= 0.6 *β=1.2*				
H=0.3 α=1.3 d= 0.8 *β=1.6*				
H=0.5 α=1.5 d=1.0 *β=2.0*				
H=0.7 α=1.7 d=1.2 *β=2.4*				
H=0.9 α=1.9 d= 1.3 *β=2.8*				

Figure 3.3.3. Graph, SAC, spectral density and log-log power spectra with slope for simulated nonstationary fBm signals with T=400.

Table 3.3.1 summarizes the relations between α, β, d, and H. Table 3.3.2 shows the parameter values and ranges desribing (fractional) Brownian Motions and (fractional) Gaussian Noises.

Table 3.3.1. Theoretical relationships between parameters capturing long-range dependence.

	d	α	β	H_{fGn}	H_{fBm}
d	–	$\alpha - .5$	$\frac{\beta}{2}$	$H_{fGn} - .5$	$H_{fBm} + .5$
α	$d + .5$	–	$\frac{\beta + 1}{2}$	H_{fGn}	$H_{fBm} + 1$
β	$2d$	$2\alpha - 1$	–	$2H_{fGn} - 1$	$2H_{fBm} + 1$
H_{fGn}	$d + .5$	α	$\frac{\beta + 1}{2}$	–	–
H_{fBm}	$d - .5$	$\alpha - 1$	$\frac{\beta - 1}{2}$	–	–

Table 3.3.2. Fractional processes and their specific parameter value and range.

	α	β	H	d
Random Walk = Ordinary Brownian Motion	1.5	2	.5	1
White Noise= Ordinary Gaussian Noise	.5	0	.5	0
fGn	[0; 1] ✓ [0; .5) AntiP ✓ (.5; 1] Pers	[-1; 1] ✓ [-1; 0)AntiP ✓ (0; 1] Pers	[0; 1] ✓ [0; .5)AntiP ✓ (.5; 1] Pers	[-.5; .5] ✓ [-.5; 0)AntiP ✓ (0; .5] Pers
fBm	[1; 2] ✓ [1; 1.5) AntiP ✓ (1.5; 2] Pers	[1; 3] ✓ [1; 2)AntiP ✓ (2; 3] Pers	[0;1] ✓ [0;.5)AntiP ✓ (.5; 1] Pers	[.5; 1.5] ✓ [.5; 1)AntiP ✓ (1; 1.5] Pers

3.4 Estimating long-range dependence

Numerous procedures for estimating long memory parameters H and d have been developed during the last years (Beran, 1994). Wagenmakers et al. (2004) demonstrated by simulation

that the most popular measure for long-range dependence used in psychology, i.e., the slope of the log–log power spectrum, may not be conclusive, since it is unduly affected by processes that are short-range dependent (for more information, see Chapter 5, p.53). Because spurious elevation in the presence of short memory is a general weakness of most long memory measures, the authors proposes, that in order to establish the presence of long-range dependence in human cognition, it is necessary to test against the hypothesis of short-range dependence, e.g., a first-order AR model plus additive white noise. They advocate the use of ARFIMA(*p,d,q*) time series modeling as a principle method that allows for inferential testing of long memory and the simultaneous estimation of both short- and long-term coefficients.

Furthermore, Caccia et a. (1997), Cannon et al. (1997), Delignières et al. (2006), and Eke et al. (2000, 2002) systematically evaluated different classical fractal analysis methods for estimating *H*, such as the rescaled range analysis, the scaled windowed variance method or the dispersional analysis. Taqqu et al. (1995) and Taqqu and Teverovsky (1998) described and compared various procedures for estimating *d* and *H* such as the *Whittle* estimator, the *aggregated variance*, or the *absolute value method*. Reisen, Abraham, and Toscano (2000) evaluated parametric and semiparametric estimators of the fractional differencing parameter *d* for stationary ARFIMA models. Stadnytska and Werner (2006) compared the performance of the *conditional sum of square procedure* for estimating *d* proposed by Chung (1996) with the exact maximum likelihood approach of Sowell (1992). The most prevalent estimators for capturing long-range dependence are listed below in alphabetical order.

The ***Detrended Fluctuation Analysis*** DFA proposed by Peng et al. (1993) reveals the extent of long-range correlations in time series by means of the scaling exponent α. As implied by its name, it was conceived as a method for detrending variability in a sequence of events. Initially, the series has to be integrated and divided into intervals of equal length *n*. In each interval a least squares regression is fit to the data. Next, the series is transformed by

subtracting the local trend. The root-mean-square fluctuation of the series is calculated. This computation is repeated over all possible interval length to characterize the relationship between the average fluctuation $F(n)$ and the interval size. For long memory processes, $F(n)$ increases with interval length. Expecting the power law $F(n) \propto n^{\alpha}$, the DFA procedure estimates α by calculating the slope of a double logarithmic plot of F as a function of n. α is converted into H according to $H = \alpha$ for fGn and $H = \alpha - 1$ for fBm. PSD and DFA are methods for estimating H and can be applied to both fGn and fBm series.

The ***Geweke and Porter-Hudak*** (1983) algorithm GPH estimates d using the linear regression of the log periodogram on a deterministic regressor. The GPH is the ordinary least squares estimator of the slope parameter in this regression considering only the lowest frequency ordinates of the log periodogram. Unlike some ML procedures constraining estimates of d to be not greater than 0.5, the Sperio and Geweke and Porter-Hudak methods are not restricted in this way. Therefore, they can be directly applied to nonstationary data.

The ***hurstSpec*** method estimates H via spectral regression by a modification of the periodogram method. It compensates for the fact that on a log-log plot most of the frequencies fall on the far right. Thus, they exert a very strong influence on the least-squared line fitted to the periodogram. The frequency axis is divided into logarithmically equally spaced boxes. Afterwards, the periodogram values corresponding to the frequencies inside the box are averaged. According to Taqqu et al. (1995) several of the values at very low frequencies are left untouched as there are so very few of them to begin with.

The ***Maximum Likelihood*** Methods estimate d optimizing the fit of the assumed (p,d,q) model to the autocovariance function of the data. The exact maximum likelihood approach was proposed by Sowell (1992). There exist a number of faster approximate algorithms. The ML estimator of d implemented in R is the approximate method of Haslett and Raftery (1989). In contrast to the techniques described above, this procedure additionally

provides short-range dependency estimates of p and q. The range of d is confined to (0; 0.5] considering only stationary persistent ARFIMA series.

The ***Power Spectral Density*** method models self-similarity of long memory structures by means of the power exponent β. Time series with long-range dependence have self-similar power-spectra with a spectral density proportional to the reciprocal of the frequency $S(f) \propto 1/f^{\beta}$. The power spectrum *S(f)* determines how much power (i.e., variance or amplitude) is accounted for by each frequency (f) in the series. For ordinary Gaussian noise, β is 0. Brownian motion is characterized by $\beta=2$. Time series with $\beta=1$ are called pink, flicker, burst or $1/f$ noise. The original PSD procedure estimates β by calculating the negative slope of the line relating log S(f) and log f. The high-frequency spectral estimates are usually excluded from fitting for the spectral slope, and only frequencies close to zero are employed for the analysis. Fougere (1985), Eke et al. (2000), Delignières et al. (2006) and Stadnytska et al. (2009c) demonstrated that transformations like bridge detrending or windowing can improve estimation.

The ***Sperio*** method proposed by Reisen (1994) estimates the memory parameter d by means of the regression equation employing the smoothed periodogram function for modeling the spectral density.

The ***Whittle*** method is like PSD based on the spectral density function and estimates H by minimizing a function containing the periodogram and the spectral density at a specified frequency. For details, see Taqqu & Teverovsky (1998) and Beran (1994). This procedure can handle either fGn or fBm data.

3.4.1 Software

The free software environment for statistical computing and graphics R[1] allows the estimation of the memory parameters *d*, *H* and α while allowing the user to add additional functionality by defining new functions. Being under constant development, various packages are available in R, covering a very wide range of modern statistics. Within the R framework, ARFIMA analysis can be performed by means of the *fracdiff* package whereas the estimators of fractal analysis are implemented in the package *fractal*.

The following *d* estimators are available in the package *fracdiff*: (1) the Geweke-Porter-Hudak estimator GPH , (2) the approximate maximum likelihood estimator by Haslett and Raftery fracdiff, (3) the smoothed periodogram approach by Reisen Sperio, (4) the Whittle estimator Whittle, (5) the modified periodogram method hurstspec, and (8) the DFA approach are implemented in the *fractal* package. Descriptions of the packages and procedures are provided at http://ftp5.gwdg.de/pub/misc/cran/). Whittle, Sperio, GPH, and fracdiff estimate *d*, DFA estimates α, while hurstSpec and provides and estimate of *H*.

In their recent study Stroe-Kunold, Stadnytska, Werner and Braun (2009) compared the accuracy of different estimators of long memory parameters implemented in R for fGn signals. As a result, Whittle was the most accurate estimator of *d*, independent of parameterizations and sample sizes. For persistent series, the performance of fracdiff was comparable to that of Whittle. For estimating *H*, hurstSpec was the best method. The DFA approach can be used a priori to distinguish between fGn and fBm. A comprehensive study evaluating the performance of estimators implemented in R concerning their ability to distinguish between fractional Brownian motions and fractional Gaussian noises, stationary and nonstationary processes, and short and long memory series can be found in Chapter 5.

[1] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

4 TIME SERIES RESEARCH IN PSYCHOLOGY

A number of recent research works apply fractal methods to psychological or behavioral variables, that were previously conceived as highly stable and fluctuations in successive measurements were considered as randomly distributed and uncorrelated in time. As such, a sample of repeated measures was assumed to be normally distributed around its mean value, and noise could be discarded by averaging. This methodological standpoint was implicitly adopted in most classical psychological researches (for a deeper analysis, see Gilden, 2001; Slifkin & Newell, 1998). In other words, temporal ordering of data points was ignored and the possible correlation structure of fluctuations was clearly neglected. The application of fractal methods, nevertheless, often remains rudimentary: analyzes are limited to the use of a unique method, the collected series are sometimes too short for a valid assessment, and more generally the theoretical background of fractals and related methods is not fully exploited. The recent theoretical and methodological refinements of fractal analyzes (see, e.g., Eke et al., 2000; 2002) appear largely unknown in the psychological community.

4.1 Review of Empirical Findings

Long memory processes are characterized by the power exponent β of the spectral density function varying from 0.5 to 1.5. Time series with β=1 are called *pink* or *1/f noise*, a fractal process resembling a wide range of natural phenomena such as heart beat rhythmus, brain activity or human coordination. A bibliography currently containing more than 1400 interdisciplinary publications on $1/f$ noise can be found under: http://www.nslij-genetics.org/wli/1fnoise/. In psychological research fractal noise was initially detected in *controlled cognitive performances* (Gilden, Thornton & Mallon, 1995). Using spectral

analysis techniques, Gilden and his colleagues demonstrated in experiments including mental rotation, lexical decision, shape and color discrimination or visual search that persistent autocorrelations account for even more variability in the data than most standard manipulations in cognitive psychology (Gilden, 1997, 2001; Gilden & Wilson, 1995; Gilden et al., 1995).

Van Orden et al. (2003), Wagenmakers et al. (2004), and Ward and Richard (2001) found that long-range dependencies in *automatic cognitive performances,* such as word naming or simple reaction times, are of lower magnitude than in tasks requiring cognitive control, and therefore less empirical support for the existence of long-term dependence in automatic tasks compared to controlled cognitive performances (Wagenmakers et al., 2005).

Chen et al. (1997, 2001), Delignières et al. (2004), and Ding et al. (2002) observed persistent correlations in human *rhythmic activities* such as tapping or other tasks requiring the coordination or synchronization of motor and cognitive activities with estimated power exponents ranging from 0.5 to 1.7 and a perfectly straight line reflecting pink noise in the log-log power spectra of most fitted series. Since long-range dependence is a characteristic property of successively produced time intervals, such as in un-paced or continuous tapping, Madison and Delignières (in press) examined the effect of auditory feedback on long-range correlation in ISIP series under conditions of eliminated or diminished sensory feedback (ISIP is short for isochronous serial interval production, commonly referred to as finger tapping or continuation tapping, in which a participant attempts to function as a metronome in the absence of external temporal cues). They hypothesized that serial dependence in such tasks could be related to a close-loop regulation process, in which the current interval is determined by preceding ones. They found that the quality of sensory feedback affects the serial dependence, e.g., diminished sensory information tends to increase the Hurst exponent for short inter-onset intervals, but decreases it for long intervals.

Aks and Sprott (2003) detected 1/*f* noise in *visual perception*, where the timing of perceptual reversals of Necker cubes served as an independent variable. Spectral analysis of 40 series detected 1/*f* noise in 80% of the cases with regression slopes varying from –0.6 to –0.9 assuming a stabilizing function of binocular disparity in perception. Disparity may either filter out extraneous information or signal the system to rely more on previous percepts. However, more studies and methodologies employing rigorous testing of alternative hypotheses to 1/*f* noise are necessary to verify this assumption.

Delignières, Fortes and Ninot (2004) reported long-range dependencies in time series of self esteem and physical self while employing different methods. Long-range dependencies were also found in human gait (Hausdorff et al., 1997, 1999); force production tasks (Pressing, 1999); brain activity (Linkenkaer-Hansen, 2002); heart rate fluctuations or other biological phenomena (Hausdorff & Peng, 1996). Delignieres and Torre (in press) reassessed Hausdorf's data indicating altered fractal dynamics of human gait, that is, reduced stride-interval correlations with aging and Huntington's disease (Hausdorff et al., 1997), and confirmed the presence of genuine fractal correlations in stride interval series in self-paced conditions using ARFIMA/ARIMA modeling. However, in contrast with Hausdorff (1997), the correlations did not appear when walking is paced by a metronome, i.e., the source of 1/*f* noise is still at work in this condition, but expressed differently under the influence of a continuous coupling process.

In social psychology, Correll (2008) postulated that latencies in simple computer tasks typically reveal 1/*f* noise, but the magnitude of the noise decreases as tasks become more challenging. He hypothesizes a correspondence between 1/*f* noise and effort, leading to the prediction that increasing effort would reduce 1/*f* noise. In two experiments the author examined the relationship between an individual's attempts to avoid bias (measured in Study 1, manipulated in Study 2) and 1/*f* noise in implicit measures of stereotyping and prejudice. In

each study, participants making an effort to modulate the use of racial information showed less 1/*f* noise than participants who made little or no effort.

In any case, the origin of the fluctuations remains in question. There have been considerations that this very specific kind of fluctuation plays an essential role in the stability of behavior as well as in the adaptability and flexibility of organisms. A number of hypotheses have been proposed for accounting for this phenomenon, however, currently, two categories of explanations can be discerned: The first one seeks a general explanation for the existence of 1/*f* noise. For example, Kello, Beltz, Van Orden, and Turvey (2007) consider 1/*f* fluctuations as the natural outcome of self-organization processes in complex systems. Cognitive functions are conceived as metastable patterns of neural and behavioral activity, and this metastability generates intrinsic fluctuations that universally exhibit the 1/*f* fluctuations. According to this point of view, 1/*f* noise is supposed to manifest in all aspects of behavior, as long as the same behavior is repeated consistently with minimal perturbation (Beltz & Kello, 2006).

A second line of reasoning seeks for domain-specific explanations. Here, 1/*f* noise is supposed to emerge from specific underlying processes within the system, and consequently local models should be proposed that take the serial properties of behavior into account, as well as their alteration under various experimental conditions (Delignières, Torre & Lemoine, 2008; West & Scaffeta, 2003).

The following studies may be exemplary for the domain specific approach. They demonstrate that specific underlying processes, e.g., reaction time sequences or certain EEG brain waves of healthy and clinical groups are represented by different models and thus demonstrating the use of time series analysis as diagnostic tools for the discrimination or identification of clinical groups.

4.2 Response Variability in Attention-Deficit Disorder

An empirical study of Gilden and Hancock (2007) provides an excellent example how time-series methods can be used in assessing dysfunctions, using the example of attention-deficit/hyperactivity disorder (ADHD). ADHD is one of the most common neurobehavioral disorders of childhood and can persist through adolescence and into adulthood. Currently the causes are unknown. A person with ADHD has a chronic level of inattention, impulsive hyperactivity, or both, such that daily functioning is compromised. The symptoms of the disorder must be present at levels that are higher than expected for a person's developmental stage and must interfere with the person's ability to function in different settings (e.g., in school and at home). A person with ADHD may struggle in important areas of life, such as peer and family relationships, and school or work performance. The American Psychiatric Association's Diagnostic and Statistical Manual-IV, Text Revision (DSM-IV-TR), estimates that 3%-7% of U.S. American children suffer from ADHD. Some studies have estimated higher rates in community samples. ADHD is diagnosed approximately three times more often in boys than in girls. Three types of ADHD have been established according to which symptoms are strongest in the individual: (1) the *predominantly inattentive type,* to whom it is hard to organize or finish a task, to pay attention to details, or to follow instructions or conversations. The person is easily distracted or forgets details of daily routines; (2) the *predominantly hyperactive-impulsive type*, where a person is fidging and talking a lot. It is hard to sit still for long (e.g., for a meal or while doing homework). Smaller children may run, jump or climb constantly. The individual feels restless and has trouble with impulsivity. Someone who is impulsive may interrupt others a lot, grab things from people, or speak at inappropriate times. It is hard for the person to wait their turn or listen to directions. A person with impulsiveness may have more accidents and injuries than others; and (3) the *combined type*, where symptoms of the above two types are equally predominant in a person.

In a review article on the neuroscience of ADHD, Castellanos and Tannock (2002, p.624) remarked that high response variability (HV) is the one ubiquitous finding in ADHD research across a variety of speeded reaction time tasks, laboratories, and cultures. A common laboratory practice in speeded response paradigms is to use young adults displaying vigilance, stamina, and speed. In this case, the RTs are considered to reflect a chain of processes: perceptual analysis, response mapping, and response execution (Pashler & Johnston, 1998). Although HV data might also arise from the chaining of these processes, it seems likely that the frequent large RTs in such data are an indication of intrusions associated with loss of vigilance (becoming lost or distracted). If the episodes of inattention are randomly interspersed in the trial sequence, then they might not distort the underlying patterns created by the normal execution chain. In terms of cell means, data with episodes of inattention might then be characterized as simply being slower and more variable than data without such episodes. However, random processing glitches would potentially have a more deleterious effect on the autocorrelation function, and for this reason, the power spectrum might be a more powerful tool than distributional statistics for characterizing HV data.

Gilden and Hancock (2007) constructed an experiment with the single purpose of collecting data with extreme individual differences in RT variability. The task chosen was randomly drawn letters from a set of four (R, Q, G, and F), presented at different angles of rotation. On half of the trials the letter was mirror-reflected. The observer's task was to determine if the letter was mirror-reflected or not (pressing '1' on the keyboard if it was not, and '2' if it was). Each observer completed a single block of 480 trials. Initially, students with little or no ADHD symptoms attending psychology classes at the University of Texas were recruited, who tended to generate low variability data (LV). In order to find reliable sources of HV data, young adults diagnosed with ADHD in alcohol recovery (i.e., members of

Alcoholics Anonymous, AA) were recruited, since alcohol abuse and ADHD symptoms go hand in hand.

Figure 4.2.1 shows example RT histories for a task that involved determining if the second and fourth letters in a five-place letter string were the same or different. In this RT study, the observers were instructed to respond as quickly as possible without making too many errors.

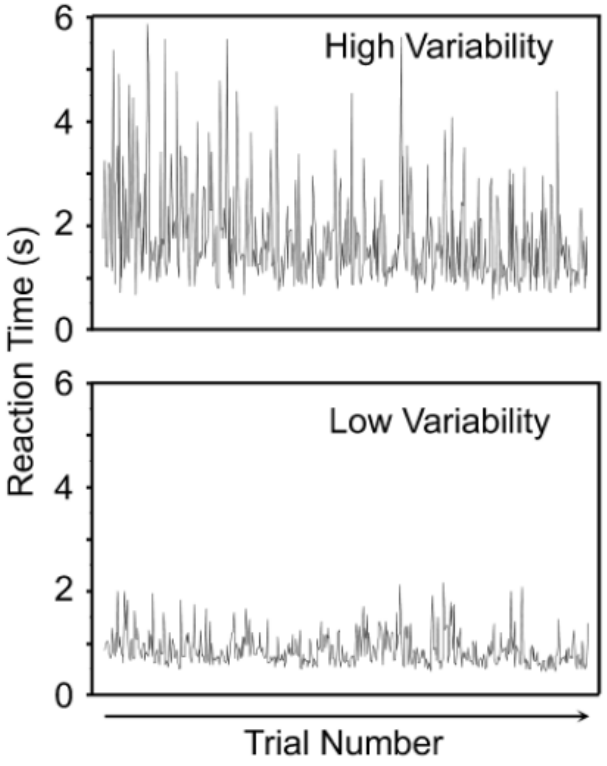


Figure 4.2.1. Reaction time sequences from high- and low-variability observers in the mental rotation task. These sequences were taken from the two observers at the median variability for their respective groups (Gilden and Hancock, 2007, p.798)

The bottom panel is an example of typical data from an undergraduate, and the top panel shows the data produced by an unmediated adult diagnosed with the combined type of ADHD. The standard deviation of the ADHD data is three times larger with many extremely high peaks. This observer seems to be cycling through some kind of process that is continuously interrupted by large and random perturbations.

As illustrated in Figure 4.2.2, the waves running through the RT data were of different amplitudes in the two groups, implying the two groups generated RT histories with different forms of autocorrelation. The low variability (LV) power spectrum is quite similar to what is generically produced by normal adults in mental rotation and other choice RT tasks (Gilden, 1997, 2001). The high variability (HV) spectrum, however, does not resemble any published spectrum for RT sequences.

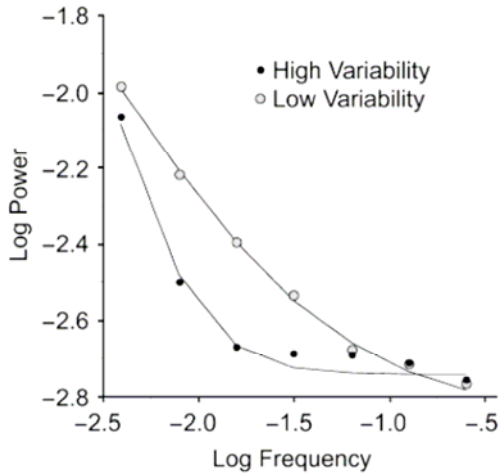


Figure 4.2.2. Average power spectra of reaction time sequences for the high and low-variability groups (Gilden and Hancock, 2007, p.800).

A dual source model was fitted that specified how white and correlated noises are mixed together to produce the bowed spectra illustrated in Figure 4.2.2. As a result, the LV group generated RT noise was about 28% 1/*f* and 72% white noise. The HV parameter were quite different, i.e., the signal was mostly white noise (92%) and 8% resembling a random walk.

To summarize, the spectral analysis of the RT data revealed an entirely different correlation structure for the two groups, that HV data are not simply LV data with higher gain, as might be concluded from the display of mean trends. HV data contain random walk, created by a process that perseveres, that is, a process in which new states are built from their immediate predecessors. The principal difference between data derived from people who can

maintain vigilance and data from those who cannot is literally in the noise. Normal undergraduates produce large amounts of 1/*f* noise as a natural consequence of decision making. People with attention deficits generate an erratic signal that develops from the intrinsic pressure of being asked to make a speeded response. This finding must serve as a caution to researchers who wish to use speeded judgment to test theories of attention dysfunction.

Similar findings have been confirmed by Johnson et al. (2008). 128 children with and without ADHD were asked to perform a deliberately boring computer task, to see how often the children made mistakes and drifted off task. The children were grouped according to whether they possessed two copies or one/no copies of a genetic risk marker for ADHD located on a gene that codes for the neurotransmitter dopamine's D4 receptor. Interestingly, the children with no copy of the risk marker and the clinical diagnosis of ADHD performed significantly more poorly on the task than the children with ADHD and at least one copy of the risk marker, but only on the elements of the task that reflected slow drifting attention. On the measures that reflected moment-to-moment control of attention, there was no effect of the genetic marker, just a clinical group difference: children with ADHD performed more poorly than the typically developing children, and were significantly more variable in the slow-frequency domain than the control children. These findings led the researchers to propose that there may be a genetic basis to this variance and propose a new theory about the role of the dopamine D4 receptor in controlling the release of the neurotransmitter noradrenalin in the prefrontal cortex, the area of the brain heavily involved in attention.

4.3 Long-Range Temporal Correlations and Major Depression

Neuroimaging has revealed robust large-scale patterns of high neuronal activity in the human brain in the classical eyes-closed wakeful rest condition, pointing to the presence of a baseline of sustained endogenous processing in the absence of stimulus-driven neuronal activity. This baseline state has been shown to differ in major depressive disorder. More recently, several studies have documented that despite having a complex temporal structure, baseline oscillatory activity is characterized by long-range temporal (auto-)correlations (LRTC) that are highly replicable within and across subjects.

An empirical study of Linkenkaer-Hansen et al. (2005) recorded neuromagnetic activity in patients with a major depressive disorder and in healthy control subjects during eyes-closed wakeful rest and quantified the long-range temporal correlations in the amplitude fluctuations of different frequency bands. In a balanced design (n =20), the ongoing brain activity of unmediated and unipolar depressed (according to DSM-IV, 1994) outpatients as well as healthy comparison subjects were measured with magnetoencephalography during eyes-closed wakeful rest for 16 min. The resulting data was separately analyzed using the Detrended Fluctuation Analysis DFA proposed by Peng et al. (1993) (see Chapter 3.4, p.36), i.e., the amplitude fluctuations of the ongoing activity in three frequency ranges: 3–7, 7–13, and 15–29 Hz, referred to as theta, alpha, and beta oscillations, respectively. The cross-section analysis of group differences of the oscillation amplitudes by a two-way ANOVA with the between-factor group (controls, patients) and within-factor location (occipitoparietal and left and right temporocentral regions), did not yield a main effect of group nor of location, that is, there was no significant main effect in the mean amplitude of theta-, alpha-, or beta-frequency bands. However, theta oscillations were significantly smaller in amplitude in the occipitoparietal and right temporocentral regions in patients with a major depressive disorder compared with the healthy control subjects.

The analysis of the long-range temporal correlations of the DFA coefficient by means of ANOVA yielded a main effect of group, caused by a larger DFA exponent in the control subjects, and location, caused by a larger DFA exponent of theta oscillations in the occipitoparietal region (see Table 4.3.1). Larger DFA coefficients means a slower decay in temporal correlations. Furthermore, there was a significant main effect for alpha-oscillations in the occipitoparietal region, and a significant main effect for beta oscillations and group. The DFA exponents of the theta oscillations in the depressive patients were very close to the theoretical value of 0.50 for uncorrelated data.

Table 4.3.1. Group differences in DFA exponents and oscillation amplitudes.

		Theta		Alpha		Beta	
		Patient	Control	Patient	Control	Patient	Control
DFA	LSM	$0.56 \pm 0.01^{**}$	0.62 ± 0.02	0.70 ± 0.03	0.71 ± 0.04	0.60 ± 0.02	0.67 ± 0.03
	RSM	$0.55 \pm 0.01^{**}$	0.61 ± 0.02	0.67 ± 0.03	0.70 ± 0.04	0.62 ± 0.02	0.68 ± 0.03
	OP	$0.61 \pm 0.02^{*}$	0.68 ± 0.03	0.72 ± 0.02	0.80 ± 0.05	0.61 ± 0.02	0.64 ± 0.03
Amp	LSM	2.20 ± 0.23	2.44 ± 0.30	5.18 ± 0.94	4.08 ± 0.80	3.67 ± 0.39	3.28 ± 0.43
	RSM	$1.96 \pm 0.14^{*}$	2.56 ± 0.25	4.45 ± 0.54	3.87 ± 0.85	3.53 ± 0.42	3.04 ± 0.33
	OP	$1.77 \pm 0.14^{*}$	2.48 ± 0.28	5.65 ± 0.71	6.83 ± 1.31	3.18 ± 0.37	2.87 ± 0.36

The significance levels from the ANOVA are indicated as follows: *p<0.05; **p<0.01. Amp, Amplitude; LSM, left temporocentral; OP, occipitoparietal; RSM, right temporocentral (Linkenkaer-Hansen et al. , 2005).

Furthermore, the authors tested whether the LRTC was related to the severity of depression in the patients. A strong linear correlation was observed between the DFA exponents of theta oscillations detected over the left temporocentral region and the score in the Hamilton Depression Rating Scale. Thus, the more depressed the patient, the less 'autocorrelated' are the amplitude fluctuations in the theta-frequency band , however not for the occipitoparietal and right temporocentral cortices. The correlation of DFA exponents and Hamilton scores was contrasted by a lack of correlation between DFA exponents and the theta amplitudes or between amplitudes and the Hamilton score, thus, it is the temporal dynamics

of these oscillations rather than the ability to generate them that is adversely affected by the major depressive disorder.

Figure 4.3.1 displays the amplitude envelope of theta oscillations in the left temporocentral region for epochs of 100s in three representative depressive patients and three control subjects. The exact fluctuation patterns are distinct from subject to subject, but it is not possible to discriminate the patients from the control subjects after visual inspection. In this sense, the DFA has indeed revealed a "hidden" difference in the complex structure of the amplitude fluctuations of ongoing oscillations.

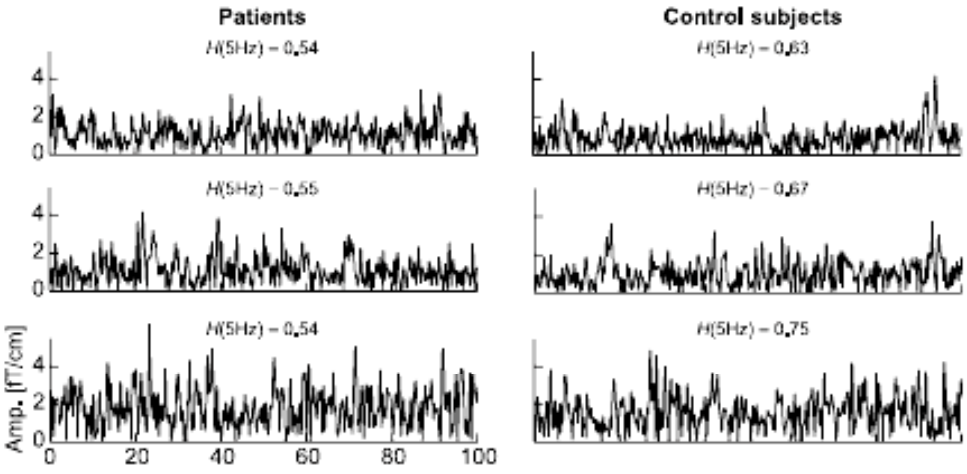


Figure 4.3.1. Amplitude fluctuations of the theta oscillations in the left temporocentral region over the course of 100s are displayed for three representative patients (left column) and control subjects (right column) (Linkenkaer-Hansen et al., 2005).

Recent reports suggest that many but not all functional abnormalities found during a depressive episode recover after pharmacological or psychotherapeutic treatment (Castren, 2005). Thus, future studies should seek to establish whether the long-range temporal autocorrelations in theta oscillations increase with the recovery from a depressive episode, or whether the absence of LRTC represents trait abnormalities.

5 SIMULATION STUDIES

The review of the empirical findings (see Chapter 4.1, p.40) revealed several methodological issues to be clarified. While long memory, i.e., the 1/*f* noise phenomenon, is one of the topics dominating the current intraindividual psychological research (Delignières et al., 2004; Ding et al., 2002; Gilden, 1997, 2001; Gilden & Hancock, 2007; Torre et al., 2007b; Van Orden et al., 2003; Wagenmakers et al., 2004), the following issues determine the actual methodological discussion within this research field:

1. *The reliable classification of fractal signals as fractional Gaussian noises or fractional Brownian motions.* Caccia et al. (1997), Cannon et al. (1997), Delignières et al. (2006) and Eke et al. (2000; 2002) systematically evaluated different classical fractal analysis methods for estimating *H* such as rescaled the range analysis (R/S), the scaled windowed variance method (SWV), or the dispersional analysis (Disp). These studies revealed that most of the methods performed well within a given class, that is fGn or fBm, but led to inconsistent results for the other. Consequently, the accurate estimation of the fractal parameter *H* (see chapter 3.2.2, p.24) requires the identification of the data-generating signal before the application of fractal analysis. Eke et al. (2000) and Delignières et al. (2006) recommended to use of the power exponent β of the periodogram analysis (see Chapter 3.3.2, p.29) for distinguishing fGn and fBm signals since it can be applied to both stationary fractal noises and nonstationary motions. For example, signals with $-1<\beta<+1$ and $+1<\beta<+3$ are almost identical with fGn and fBm signals, respectively. For β close to one, a zone of uncertainty is to be expected, as well as for the region around $d=0.5$ corresponding to the border of nonstationary within the ARFIMA analysis. However, the majority of

methods for estimating *d* can handle only stationary processes. This leads to the evaluation of different periodogram-based estimators as reliable classification tools of fractal signals.

2. *The reliable discrimination of stationary and nonstationary processes.* A process is said to be stationary if its mean, variance and covariance are stable over time. In the ARFIMA framework, this is the case for processes with $d<0.5$, where the impact of each and every innovation (see Chapter 2.2.1, p.9) is always final. With $d\geq0.5$ ARFIMA processes become unstable due to a stochastic trend. In this case, the impact of the innovations become extremely persistent, but not yet permanent. Innovations with a permanent impact are found in the case of ARIMA processes with $d=1$, hence, they are called processes with infinite memory. Obviously, the value of the fractional differencing parameter *d* does not only determine whether a series is stationary or not, but also its underlying memory. For $d=0$, a process displays short memory, while finite long memory is characteristic for processes with $0<d<0.5$. However, Rangarajan and Ding (2000), Thornton and Gilden (2005), and Wagenmakers et al. (2004; 2005) observed that short memory time series may mimic the statistical properties of long memory processes. To explore this problem further, determining the memory property of processes within the ARFIMA framework will be the third main issue.

3. *The accurate discrimination of short-term and long-term dependence in time series.* In the ARFIMA framework, allowing the simultaneous modeling of short- and long-term dependencies, the memory property of a process crucially depends on the value of the fractional differencing parameter *d*. The process displays short memory for $d=0$ and the finite long memory for $0<d<0.5$. To solve this problem, different methods for identifying long-run developments have been proposed. Naturally, they are functions of their employed estimators, thus their performance is resting on the quality of the

estimation of the fractional differencing parameter *d*. A region of uncertainty is expected around the border of *d*=0, as well as an impact of the number and magnitude of the short memory coefficients.

The overall objective of this chapter therefore is, to evaluate the performance of different periodogram-based estimators and non-spectral alternatives within the fractal analysis and ARFIMA framework as diagnostic tools for reliably distinguishing between stationary and nonstationary fractal signals and ARFIMA processes, as well as between ARFIMA series with short and long memory. Strategies for the reliable classification of fractal signals, (non)stationary processes and processes with different memory properties will be developed based on the empirical results. Furthermore, the most accurate estimation techniques for the task at hand will be determined.

5.1 Study 1: Distinguishing Fractal Signals

5.1.1 Introduction

Currently, there are two main fields determining the discussion on persistent autocorrelation, that is, the search for explanations of the phenomena of long-range dependence (Van Orden et al., 2003; Wagenmakers et al., 2005), as well as the development of methods for a precise estimation of the persistency parameters and the reliable discrimination of series with long-lasting autocorrelations from those with small and transient ones (Delignières et al., 2006; Farrell et al., 2006b; Thornton & Gilden, 2005; Torre et al, 2007a; Wagenmakers et al., 2004). Aiming at providing a contribution to the latter, Caccia et al. (1997), Cannon et al. (1997), Eke et al. (2000, 2002), and Delignières et al. (2006) systematically evaluated different classical fractal analysis methods for estimating *H* (for a detailed description of the fractal parameter *H* see Chapter 3.2.2, p.24). Revealing that most estimators performed well only within an appropriate class (fGn or fBm), the authors concluded, that the accurate estimation

of *H* requires the identification of the class of the series before the application of fractal analysis. According to Eke et al. (2002), until recently researchers were not aware of the necessity of this step, resulting in a number of questionable empirical analysis and theoretical implications.

Taqqu et al. (1995) and Taqqu and Teverovsky (1998) described and compared various procedures for estimating *d* and *H* such as the Whittle estimator, the aggregated variance, or the absolute value method. Reisen, Abraham, and Toscano (2000) evaluated parametric and semiparametric estimators of the fractional differencing parameter *d* for stationary ARFIMA models. Stadnytska and Werner (2006) compared the performance of the conditional sum of square procedure (CSS) for estimating *d* with the exact maximum likelihood approach of Sowell (1992).

5.1.2 Background

Mandelbrot (1975) introduced the term 'fractal' (from the latin fractus, meaning 'broken') to characterize spatial or temporal phenomena that are continuous but not differentiable. Every attempt to split a fractal into smaller pieces results in the resolution of more structure, and therefore displays 'self-invariant' properties. Within the fractal geometry there are two classes of fractal signals, fractional Brownian motion (fBm) and fractional Gaussian noise (fGn). Both can be characterized by the same Hurst exponent $H \in (0,1)$ (see Figures 3.3.2 and 3.3.3). While the successive increments of fractal processes with H=0.5 are uncorrelated, for H within the range of [0; 0.5) they are antipersistent implying negative correlations, and for H within the range of (0.5; 1], they are persistent implying positive correlations between successive increments. For H=0.5 fGn corresponds to ordinary Gaussian noise, a stationary process with constant mean and variance, while fBm corresponds to ordinary Brownian motion, a nonstationary process with stationary increments. The main difference between fBm

and ordinary Brownian motion is that, while both are nonstationary processes, the increments in Brownian motion are independent Gaussian noise, while in fractional Brownian motion they are dependent.

5.1.3 Modifications of Estimation Methods

Numerous procedures for estimating persistency have been developed both in fractal analysis and within the ARFIMA framework. This study focuses on methods able to classify fractal signals as fGn or fBm directly by estimating the power exponent β of the regression slope within spectral analysis, indirectly by computing the scaling exponent α of the DFA method proposed by Peng et al. (1993), and the fractional differencing parameter d in the time domain.

The power exponent of the Power Spectral Density method PSD typically is $-1 \leq \beta < 1$ for fGn processes, while $1 < \beta \leq 3$ represents fBm processes. The DFA allows the discrimination of fractal signals, as the scaling exponent α varying from 0 to1 implies fGn, and exponents from 1 to 2 represent fBm processes. Note that $\beta = 2\alpha - 1$ for both fractal signals, thus, both parameter allow a definitive classification of a series as fGn or fBm signal simply by the magnitude of the coefficient, provided precise parameter estimation of the respective method. In this study, the following modifications estimating β, α and H were evaluated:

Modification of the PSD Method

Eke et al. (2000) and Delignières et al. (2006) advocated to use the periodogram analysis *PSD* for fractal signals since it can be applied to both stationary fractal noises and nonstationary motions. It estimates the so-called power exponent β of a series to determine the signal class, since persistent processes have self-similar power-spectra with a spectral density proportional to the reciprocal of the frequency $S(f) \propto 1/f^{\beta}$. The power spectrum $S(f)$ determines how much

power (i.e., variance or amplitude) is accounted for by each frequency *(f)* in the series. For ordinary Gaussian noise, β is 0, while Brownian motion is characterized by β=2. The power exponent of an fGn process can be any real value in the range of (-1;1), while estimates of $\hat{\beta} > 1$ suggest fBm or nonstationary processes.

The original PSD procedure estimates $\hat{\beta}$ by calculating the negative slope of the line relating log *S(f)* and log *f*. The high-frequency spectral estimates are usually excluded from fitting for the spectral slope, and only frequencies close to zero are employed for the analysis. For example, Taqqu and Teverovsky (1996) used the lowest 10% of the frequencies for their calculations. Different modifications of the periodogram method have been suggested to improve estimation. Fougere (1985) showed in simulation experiments that end matching or bridge detrending (subtracting from the data the line connecting the first and last points), and applying a parabolic window before analysis, improves the consistency of estimates for fBm signals. The parabolic window for a series of length *T* is a function that multiplies each value in the series and is given by:

$$W(j) = 1 - \left(\frac{2j}{T+1} - 1 \right)^2 \text{ for } j=1,\ldots,T.$$

According to Eke et al. (2000), estimation accuracy depends on the order of the transformation steps: better results were achieved if parabolic window preceded the step of end matching. Therefore they proposed the method designated as ${}^{low}PSD_{we}$ consisting of the following operations: subtracting the mean of the series from each value; applying a parabolic window to the data (w), performing bridge detrending (e) and estimating β excluding 7/8 of the high frequency power estimates (low).

Delignières et al. (2006) compared the performance of ${}^{low}PSD_{we}$ with the original periodogram algorithms. Both methods were able to distinguish between fGn and fBm and

showed comparable performance in series with a true *H* exponent ranging from 0.3 to 0.7. For low and high coefficients, however, the results were ambiguous. For instance, the original PSD clearly outperformed $^{low}PSD_{we}$ in short fBm series with H=0.2 and was significantly inferior in fGn series with H=0.9. Adopting the notation introduced by Eke et al. (2002), $^{low}PSD_{we}$ implies the following order of operations: (1) subtracting the mean of the series from each value, (2) applying a parabolic window to the data (w), (3) performing end matching (e), and finally (4) estimating β excluding 7/8 of high frequency power estimates (low). In this study, the operations described were combined to 8 different PSD versions yielding different results in a preliminary study: ^{low}PSD, $^{low}PSD_e$, $^{low}PSD_w$, $^{low}PSD_{we}$, PSD_{we}, PSD_{ew}, PSD_w and PSD_e. The R-code for $^{low}PSD_{we}$ is attached.

Modifications of the DFA Method

The *Detrended Fluctuation Analysis* DFA proposed by Peng et al. (1993) reveals the extent of long-range correlations in time series by means of the scaling exponent α. As implied by its name, it was conceived as a method for detrending variability in a sequence of events. Initially, the series has to be integrated and divided into intervals of equal length *n*. In each interval a least squares regression is fit to the data. Next, the series is transformed by subtracting the local trend. The root-mean-square fluctuation of the series is calculated. This computation is repeated over all possible interval length to characterize the relationship between the average fluctuation $F(n)$ and the interval size. For long memory processes, $F(n)$ increases with *n*. Expecting the power law $F(n) \propto n^{\alpha}$, the DFA procedure estimates α by calculating the slope of a double logarithmic plot of *F* as a function of *n* (see also Chapter 3.4.1).

Own modifications of the DFA method implemented in R were renamed by adding the initials jw (after its author, Prof. Joachim Werner). Furthermore, the DFAjw2 and DFAjw8

modifications specify the successive augmentation of the interval size by 2 and 8, respectively. Further DFA modifications are DFAjwL2 and DFAjwL6, were the ratio of the successive scales was set to 2 and 6, respectively. These modifications led to 5 different DFA variations to be evaluated in the current study: DFAbridge, DFAjw2, DFAjw8, DFAjwL2, DFAjwL6

Modifications of th HurstSpec Method

The HurstSpec method, implemented in R, is a function to estimate the Hurst parameter *H* of a time series by linear regression of the log spectrum versus log frequency (see Chapter 3.4, p.35). However, this approach is a modification of the periodogram method compensating for the fact that on a log-log plot most of the frequencies fall on the far right, and thus exerting a strong influence on the least-squared line fitted to the periodogram while dividing the frequency axis into logarithmically equally spaced boxes. In the 'smoothed' version of hSpec, here called hSpecsm, the periodogram values corresponding to the frequencies inside the box are averaged. According to Taqqu et al. (1995) several of the values at very low frequencies are left untouched. Additionally, the 'wosa' or Welch's overlapped segment averaging method, here called hSpecwo, as well as the 'multitaper' method, here called hSpecmu, was used. The latter is based on multitaper spectrograms with averaged spectrum estimates obtained by first windowing the time series with a collection of orthogonal taper functions. All hSpec modifications are already implemented in R. For this study, all modifications were complemented with the return function for outputting $\hat{\beta}$ and were renamed accordingly by adding the prefix 'exp' resulting in ExphSpec, ExphSpecsm, ExphSpecwo and ExphSpecmu.

To summarize, stationary fractal noises and nonstationary motions can be both characterized by the same Hurst exponent, therefore other methods like the power exponent β in frequency domain or the scaling exponent α and the fractional differencing parameter *d* in

time domain have to be used for classifying fractal signals, which can be then converted into *H* by known relations (see Chapter 3.3.3, p.31). The objectives of this study is to evaluate the above mentioned methods by computing the percentage of signal misclassification and determining the accuracy of estimation to develop a strategy for reliably classifying fractal signals and estimating their corresponding parameter.

5.1.4 Method

The reliability of the abovementioned methods as well as the Whittle and Sperio method (see Chapter 3.4, p.38) is tested by simulated fGn and fBm processes with known Hurst coefficient by means of the option lmSimulate (the R-code is attached) of the R package fractal (for details, consult the R documentation at http://ftp5.gwdg.de/pub/misc/cran/). The procedure is based on the Davies-Harte technique (1987) developed for the generation of exact fractional Gaussian noise. *The Fractal Package* of R also includes the standard versions of the DFA and hSpec method. Descriptions of the packages and procedures are provided at http://ftp5.gwdg.de/pub/misc/cran/). The R code is also available at http://www.stat.osu.edu/~pfc/software/ (see Craigmile, 2003).

For this study, fGn and fBm signals with *H* variations from 0.1 to 0.9, each replicated a 1000 times, are used. Manipulated are the following **independent variables**:

- class of signal: fGn, fBm
- value of *H*: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9
- length of series *T*. 128, 265, 512, 1024, 2048
- estimation methods:
  - 8 PSD versions estimating $\hat{\beta}$: ^{low}PSD, ^{low} PSD_e, ^{low} PSD_w, ^{low} PSD_{we}, PSD_e, PSD_{ew}, PSD_w, $PSD_{we;}$
  - 4 ExphSpec versions estimating $\hat{\beta}$: ExhSpec, ExhSpecsm, ExhSpecwo, ExhSpecmu;

- 5 DFA versions estimating $\hat{\alpha}$: DFAbridge, DFAjw2, DFAjw8, DFAjwL2, DFAjwL6
- 2 methods estimating $\hat{d}$ *:* Whittle, Sperio

As quality criterion the following **dependent variables** were computed:

- percentage of signal misclassification (MISCLASS) of $\hat{\beta}$, $\hat{\alpha}$ and $\hat{d}$. $\hat{\beta}$ >1 (ExphurstSpec and PSD methods), $\hat{\alpha}$ >1 (DFA methods) and $\hat{d}$ >0.5 stand for fBm, $\hat{\beta}$ <1, $\hat{\alpha}$ <1 and $\hat{d}$ <0.5 stand for fGn series, respectively.
- mean (*M*), standard error (*SE*) of $\hat{H}$ received from the transformed values of $\hat{\beta}$, $\hat{\alpha}$ and $\hat{d}$ (see Table 3.3.1, p.39) for better comparability.
- minimum (*MIN*), maximum (*MAX*) from $\hat{\beta}$, $\hat{\alpha}$ and $\hat{d}$, respectively.

Table 3.3.1 (see Chapter 3.3, p.39) presents the transformation rules of the scaling exponent $\hat{\alpha}$, power exponent $\hat{\beta}$ and fractional differencing parameter $\hat{d}$ into the Hurst exponent $\hat{H}$ (For comparability, all received estimations had to be transformed into $\hat{H}$ before computing the dependent variables *M* and *SE*, while the computation of *MIN* and *MAX* are based on the untransformed estimations as given by the various methods. A true fGn signal was misclassified if estimated $\hat{\beta}>1$ by the PSD and exphSpec versions, $\hat{\alpha} > 1$ by the DFA methods and $\hat{d}$ >0.5 by Whittle and Sperio. Correspondingly, an fBm signal was misclassified by $\hat{\beta}<1$, $\hat{\alpha} < 1$ and $\hat{d} < 0.5$, respectively.

All computations were performed with R version 2.7.2.

5.1.5 Results

This chapter investigates the accuracy of the estimation methods under evaluation first by comparing the estimates' standard error and bias. The assessment of the estimators' ability of reliably classifying fractal signals is considered thereafter, taking the percentage of misclassifications into account.

Since in empirical settings the true structure is never known, a strategy for the estimation of the long memory parameter is recommended to identify the estimator(s) with the smallest number of false decisions regardless of signal type, parameterization and sample size. An additional investigation of the minimal and maximal estimates received from those procedures performing best will be done, so it may give the researcher an idea what to expect in the case of a single analysis of an empirical series.

Accuracy of Estimation

The analysis of the accuracy of the procedures under evaluation requires the transformation of the estimated scaling exponent $\hat{\alpha}$, power exponent $\hat{\beta}$ and the fractional differencing parameter $\hat{d}$, given by the DFA, PSD and ExphSpec methods as well as Sperio and Whittle, respectively, into the Hurst exponent $\hat{H}$ first (see Chapter 3.3, p.35).

The quality of parameter estimation is assessed by computing the averaged standard error and bias of each estimation (over 1000 replications) for each level of the independent variable, signal type, *H* and time series length *T* (see Table 3.3.1, p.35). Hereby special consideration will be given to the sign of the bias, since a negative bias for fBm series with small *H* coefficients as well as a positive bias for fGn signals with large coefficients may lead to misinterpretation of the true signal. Therefore the bias on the edge of the 1/*f* boundary, that is, of fGn series with *H*=0.9 and fBm series with *H*=0.1 will be given special attention.

Standard Error

The mean standard error (for the previously transformed values) of $\hat{H}$ for fGn and fBm series for the different sample sizes is shown in Table 5.1.2. The values presented in the table are the *SE*s averaged over all *H* and additionally averaged over all sample sizes (right column). Overall, there is only little effect of sample size on standard error for both signal types and the

SE decreases about one fourth up to one half from the smallest to the largest sample size. The range between the smallest and largest *SE* – averaged over all parameters and sample sizes – is about one third larger in fBm series.

Independently of signal type, parameterization and sample size, the largest SEs are delivered by the PSD estimates employing only frequencies close to zero in the analysis as well as Sperio, while the PSD modifications including the high-frequency spectral estimates show distinctively less variability. They are the estimates with the smallest *SE*s in the fBm case, followed by Whittle and the DFA methods. In fGn series the DFA variations and Whittle, followed by the PSD methods including the high-frequency spectral estimates deliver the smallest standard errors, regardless of parameterization and sample size. Independently of signal type, the ExphSpec methods are slightly superior compared to the ^{low}PSD modifications but show more variability than the DFA methods, Whittle and the PSD methods just applying a parabolic window and /or bridge detrending to the data.

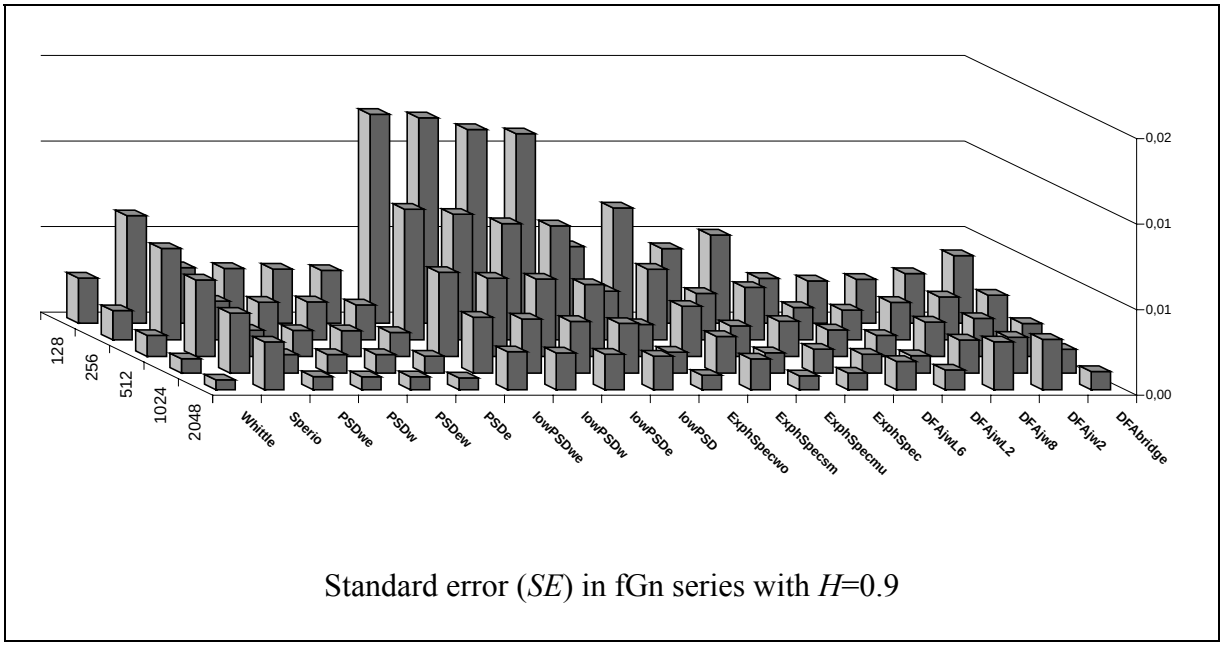
Table 5.1.2 Mean standard error (*SE*) for fGn and fBm series at *T*= 128, 512, 1024, 2048, regardless of parameterization (parameter averaged *SE*), and regardless of parameterization and sample size (parameter and sample size averaged *SE*). The data is sorted in ascending order by the parameter and sample size averaged *SE*.

Signal Type	Method	Mean Standard Error *SE*					
		T (Parameter Averaged *SE*)					Parameter and Sample Size Averaged *SE*
		128	256	512	1024	2048	
fGn	DFAjwL2	0.0018	0.0012	0.0009	0.0007	0.0006	0.0010
	DFAjwL6	0.0019	0.0014	0.0011	0.0008	0.0008	0.0012
	Whittle	0.0026	0.0017	0.0012	0.0008	0.0006	0.0014

	DFAjw8	0.0018	0.0015	0.0014	0.0014	0.0014	0.0015
	DFAjw2	0.0020	0.0017	0.0015	0.0015	0.0015	0.0016
	DFAbridge	0.0031	0.0021	0.0015	0.0011	0.0008	0.0017
	PSD_w	0.0033	0.0022	0.0015	0.0010	0.0007	0.0017
	PSD_{ew}	0.0033	0.0023	0.0016	0.0012	0.0008	0.0018
	PSD_e	0.0037	0.0025	0.0018	0.0013	0.0009	0.0020
	ExphSpecmu	0.0041	0.0026	0.0018	0.0012	0.0008	0.0021
	ExphSpecwo	0.0041	0.0027	0.0018	0.0012	0.0008	0.0021
	PSD_{we}	0.0040	0.0027	0.0019	0.0014	0.0010	0.0022
	ExphSpec	0.0050	0.0032	0.0021	0.0014	0.0009	0.0025
	ExphSpecsm	0.0066	0.0042	0.0029	0.0022	0.0018	0.0036
	Sperio	0.0062	0.0050	0.0041	0.0033	0.0027	0.0043
	^{low}PSD	0.0110	0.0068	0.0044	0.0029	0.0020	0.0054
	$^{low}PSD_w$	0.0120	0.0074	0.0048	0.0032	0.0022	0.0059
	$^{low}PSD_e$	0.0141	0.0096	0.0067	0.0049	0.0035	0.0078
	$^{low}PSD_{we}$	0.0151	0.0103	0.0073	0.0053	0.0038	0.0084
fBm	PSD_e	0.0031	0.0021	0.0014	0.0010	0.0007	0.0017
	PSD_w	0.0034	0.0023	0.0015	0.0011	0.0007	0.0018
	PSD_{ew}	0.0034	0.0023	0.0015	0.0011	0.0007	0.0018
	DFAjwL2	0.0029	0.0022	0.0017	0.0014	0.0012	0.0019
	Whittle	0.0029	0.0022	0.0018	0.0015	0.0014	0.0020
	PSD_{we}	0.0039	0.0026	0.0018	0.0012	0.0008	0.0020
	DFAjwL6	0.0031	0.0022	0.0021	0.0014	0.0017	0.0021
	DFAbridge	0.0044	0.0032	0.0024	0.0018	0.0015	0.0027
	DFAjw8	0.0033	0.0029	0.0027	0.0026	0.0026	0.0028
	ExphSpec	0.0049	0.0033	0.0024	0.0019	0.0015	0.0028
	ExphSpecmu	0.0055	0.0036	0.0023	0.0015	0.0011	0.0028

	DFAjw2	0.0037	0.0032	0.0030	0.0028	0.0027	0.0031
	ExphSpecwo	0.0067	0.0043	0.0026	0.0016	0.0011	0.0033
	ExphSpecsm	0.0062	0.0042	0.0031	0.0026	0.0022	0.0037
	Sperio	0.0060	0.0048	0.0040	0.0032	0.0027	0.0042
	$^{low}PSD_e$	0.0106	0.0070	0.0046	0.0031	0.0021	0.0055
	$^{low}PSD_w$	0.0111	0.0075	0.0049	0.0033	0.0022	0.0058
	$^{low}PSD_{we}$	0.0111	0.0072	0.0054	0.0038	0.0026	0.0060
	^{low}PSD	0.0110	0.0074	0.0054	0.0038	0.0026	0.0061

Figures 5.1.1 and 5.1.2 show the standard error (*SE*), averaged over 1000 replications, of all DFA, ExphSpec and PSD methods as well as Whittle and Sperio on the 1/*f* border, that is, for fGn series with *H*=0.9 (Figure 5.1.1) and fBm series with *H*=0.1 (Figure 5.1.2), both for *T*=128, 265, 512, 1024 and 2048.



Standard error (*SE*) in fGn series with *H*=0.9

Figure 5.1.1. *SE* for the different PSD, ExphSpec and DFA modifications as well as Whittle and Sperio (presented in alphabetical order) in fGn series with *H*=0.9 at *T*=128, 256, 512, 1024 and 2048.

Overall, the standard error for fGn series with *H*=0.9 ranges from 0.0025 (DFAjwL2) to 0.0122 ($^{low}PSD_{we}$) at *T*=128 and from 0.0006 (Whittle) to 0.0029 (DFAjw2) at *T*=2048.

There is no clear effect of sample size on standard error for DFAjw2 and DFAjw8 with a larger *SE* at *T*=2048 than at *T*=1024 for both procedures. Overall, the smallest *SE*s for fGn series with *H*=0.9 are obtained by the Whittle method as well as the PSD modifications including the high-frequency spectral estimates in their analysis. They are followed by the ExphSpec methods which outperform the DFA modifications at *H*=0.9. The ^{low}PSD modifications exhibit distinctly large standard errors in short series but perform much better in long series with even less variability than Sperio, DFAjw2 and DFAjw8.

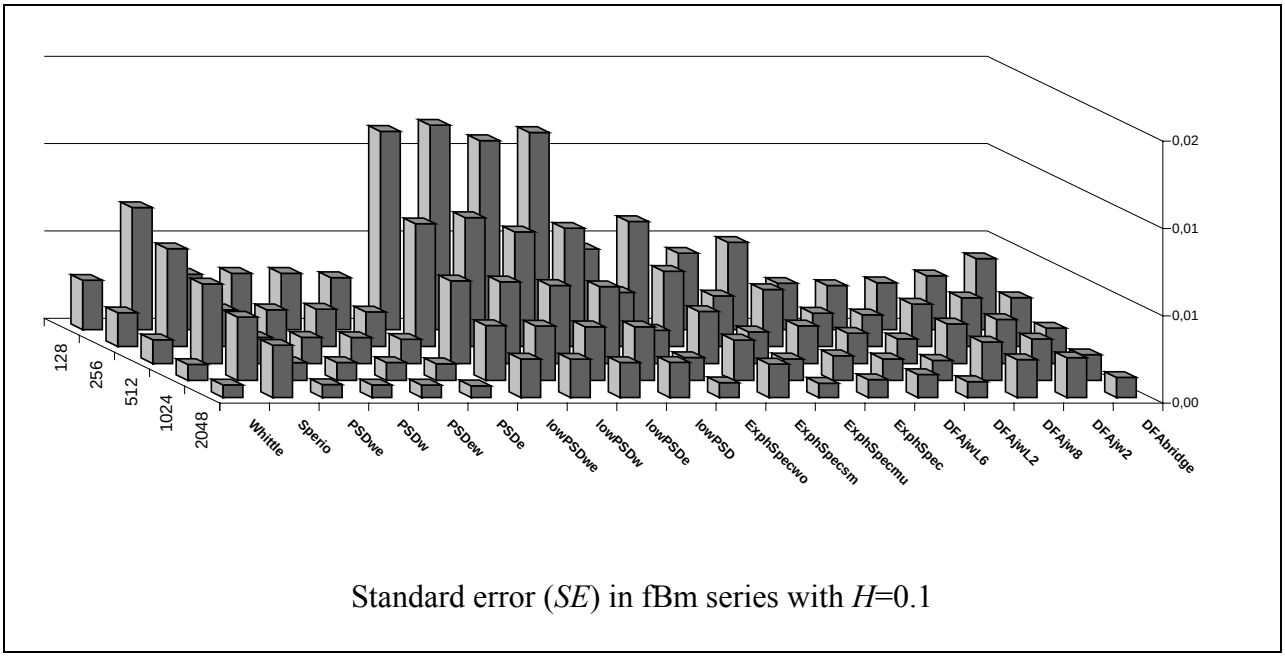


Figure 5.1.2. *SE* for the different PSD, ExphSpec and DFA modifications as well as Whittle and Sperio in fBm series (presented in alphabetical order) with *H*=0.1 at *T*=128, 256, 512, 1024 and 2048.

The standard error for fBm series with *H*=0.1 ranges from 0.0025 (DFAjwL2) to 0.0117 ($^{low}PSD_{we}$) at *T*=128 and from 0.0007 (PSD_e , PSD_{ew} , Whittle, PSD_w , PSD_{we}) to 0.0030 (Sperio) at *T*=2048 and is therefore quite similar to the fGn case at *H*=0.9. On the 1/*f* boundary, there is a small effect of sample size on standard error for all procedures under evaluation. As in the fGn case, the smallest standard errors are delivered by the PSD modifications including the high-frequency spectral estimates in their analysis and the Whittle

method, followed by the ExphSpec procedures. The largest variability in fBm series with H=0.1 is, as in fGn series with H=0.9, demonstrated by the ^{low}PSD modifications and Sperio.

To summarize, the superiority of the DFA methods in delivering the smallest standard errors in fGn series does not hold on the edge of the 1/f boundary. There, the PSD estimates using all frequencies for the fitting of the spectral slope and the Whittle estimates have demonstrated the least variability of all procedures under evaluation, independently of signal type and sample size and deliver the smallest standard errors of all procedures under evaluation. In general, those PSD estimates excluding the high-frequency spectral estimates as well as Sperio are associated with large standard errors, independently of signal type, parameterization and time series length. However, accuracy of estimation requires not only low variability but also a little bias ($\mathrm{Mean}(\hat{H}) - H$). Latter will be investigated in the following section.

Bias

While estimates with little or no bias are a necessary prerequisite for accurate estimation, large negative biases for fGn as well as large positive biases for fBm series, especially on the edge of the 1/f boundary, may even foster the reliable identification of fractal signals. Therefore not only the bias but also the sign of the bias will be given special consideration.

First, we start with the bias modulus ($\left| \mathit{Mean}(\hat{H}) - H \right|$) for fGn and fBm series at T= 128, 256, 512, 1024, 2048, independently of parameterization, and averaged over all parameter and sample sizes (right column), as shown in Table 5.1.3. There is a clear effect of sample size on bias for both signal types. In the fGn case, distinctly small biases, independently of sample size, are delivered by ^{low}PSD. Coefficients of fGn series with low values of H are generally underestimated by both procedures, whereas estimations of large values are positively biased. The largest bias modulus of all procedures under evaluation in the fGn case is delivered by ^{low}PSD$_e$ and ^{low}PSD$_{we}$. Both methods overestimate the true

parameter, regardless of parameterization and sample size. Fairly good results are delivered by the DFA variations (not DFAbridge) and Sperio, which are also part of the least biased methods in fBm series after $^{low}PSD_e$. The performance of the Exphspec methods worsens in fBm series with a bias modulus about twice the size than in the fGn case. DFAbridge and Whittle are distinctively biased, independently of signal type. In fGn series with small coefficients Whittle generally underestimates the true parameter but delivers a positive bias for series with $H \geq 0.6$, whereas all parameterizations in fBm series are clearly underestimated. All DFA estimates are positively biased, except for fBm series with $H=0.1$.

Table 5.1.3. Bias Modulus ($|Mean(\hat{H}) - H|$) for fGn and fBm series and T= 128, 256, 512, 1024, 2048, averaged over all parameters, and additionally averaged over all sample sizes (right column). The data is sorted in ascending order by the parameter and sample size averaged bias modulus.

Signal Type	Method	Mean Bias Modulus					
		T (Parameter Averaged)					Parameter and Sample Size Averaged
		128	256	512	1024	2048	
fGn	^{low}PSD	0.0171	0.0126	0.0130	0.0130	0.0087	0.0129
	$^{low}PSD_w$	0.0342	0.0202	0.0140	0.0140	0.0086	0.0182
	DFAjw8	0.0257	0.0197	0.0235	0.0235	0.0325	0.0249
	DFAjw2	0.0279	0.0286	0.0329	0.0329	0.0373	0.0319
	Sperio	0.0565	0.0386	0.0273	0.0273	0.0156	0.0330
	DFAjwL6	0.0624	0.0671	0.0177	0.0177	0.0123	0.0355
	DFAjwL2	0.0657	0.0468	0.0337	0.0337	0.0204	0.0401
	PSD_w	0.0461	0.0438	0.0439	0.0439	0.0416	0.0439
	PSD_{ew}	0.0766	0.0551	0.0383	0.0383	0.0259	0.0469
	PSD_e	0.0867	0.0640	0.0427	0.0427	0.0273	0.0527
	ExphSpecwo	0.0688	0.0567	0.0502	0.0502	0.0392	0.0530

	ExphSpecsm	0.0961	0.0571	0.0440	0.0440	0.0329	0.0548
	ExphSpec	0.0750	0.0599	0.0507	0.0507	0.0397	0.0552
	ExphSpecmu	0.0758	0.0594	0.0513	0.0513	0.0395	0.0555
	PSD_{we}	0.1103	0.0808	0.0542	0.0542	0.0284	0.0656
	Whittle	0.0781	0.0712	0.0688	0.0688	0.0655	0.0705
	DFAbridge	0.1628	0.1313	0.1095	0.1095	0.0820	0.1190
	$^{low}PSD_{e}$	0.4807	0.4023	0.3077	0.3077	0.1780	0.3353
	$^{low}PSD_{we}$	0.5783	0.4744	0.3611	0.3611	0.2052	0.3960
fBm	$^{low}PSD_{e}$	0.0695	0.0078	0.0175	0.0201	0.0155	0.0261
	DFAjwL2	0.0636	0.0526	0.0436	0.0364	0.0310	0.0454
	$^{low}PSD_{w}$	0.0718	0.0824	0.0634	0.0441	0.0300	0.0583
	DFAjwL6	0.0751	0.0710	0.0728	0.0387	0.0569	0.0629
	$^{low}PSD_{we}$	0.0665	0.0932	0.0986	0.0773	0.0527	0.0777
	Sperio	0.1259	0.0945	0.0786	0.0687	0.0707	0.0877
	DFAjw8	0.1253	0.0997	0.0939	0.0858	0.0842	0.0978
	^{low}PSD	0.1110	0.1251	0.1237	0.0929	0.0617	0.1029
	DFAjw2	0.1273	0.1117	0.0993	0.0904	0.0871	0.1032
	ExphSpec	0.0994	0.0946	0.0993	0.1106	0.1183	0.1044
	PSD_{we}	0.1171	0.1060	0.1014	0.1035	0.1083	0.1073
	PSD_{ew}	0.0958	0.1041	0.1098	0.1136	0.1149	0.1077
	PSD_{w}	0.1052	0.1073	0.1092	0.1132	0.1148	0.1099
	ExphSpecsm	0.1696	0.1000	0.0929	0.0949	0.0927	0.1100
	PSD_{e}	0.1141	0.1146	0.1160	0.1170	0.1168	0.1157
	DFAbridge	0.1583	0.1369	0.1165	0.1016	0.0895	0.1206
	Whittle	0.1431	0.1317	0.1279	0.1272	0.1277	0.1315
	ExphSpecmu	0.2631	0.1831	0.1217	0.0848	0.0605	0.1426
	ExphSpecwo	0.3167	0.2219	0.1439	0.0947	0.0648	0.1684

Figures 5.1.3 and 5.1.4 show the bias ($Mean(\hat{H}) - H$), averaged over 1000 replications, of all DFA, ExphSpec and PSD methods as well as Whittle and Sperio on the edge of the 1/*f* boundary, i.e., for fGn series with *H*=0.9 (Figure 5.1.3) and fBm series with *H*=0.1 (Figure 5.1.4), both for *T*=128, 265, 512, 1024, 2048. Except for some DFA methods and Sperio, there is a clear effect of signal type on the sign of the bias. FGn series with *H*=0.9 are generally overestimated, while fBm series with *H*=0.1 are underestimated.

Obvious is the distinct overestimation of DFAbridge in the fGn case that may lead to many erroneous decisions of the series as fBm signals. The remaining DFA methods (except DFAjwL2) deliver fairly small negative biases. $^{low}PSD_w$ and ^{low}PSD are least biased in the parameter estimation of fGn series, whereas DFAbridge, ExphSpec, $^{low}PSD_{we}$ and Whittle clearly overestimate *H* with the largest biases of all procedures under evaluation. For the Whittle method, the bias even becomes larger with increasing sample size. Fairly good results are likewise obtained for *H*=0.9 just as in the analysis of the bias regardless of parameterization for the remaining DFA modifications (not DFAbridge) and Sperio.

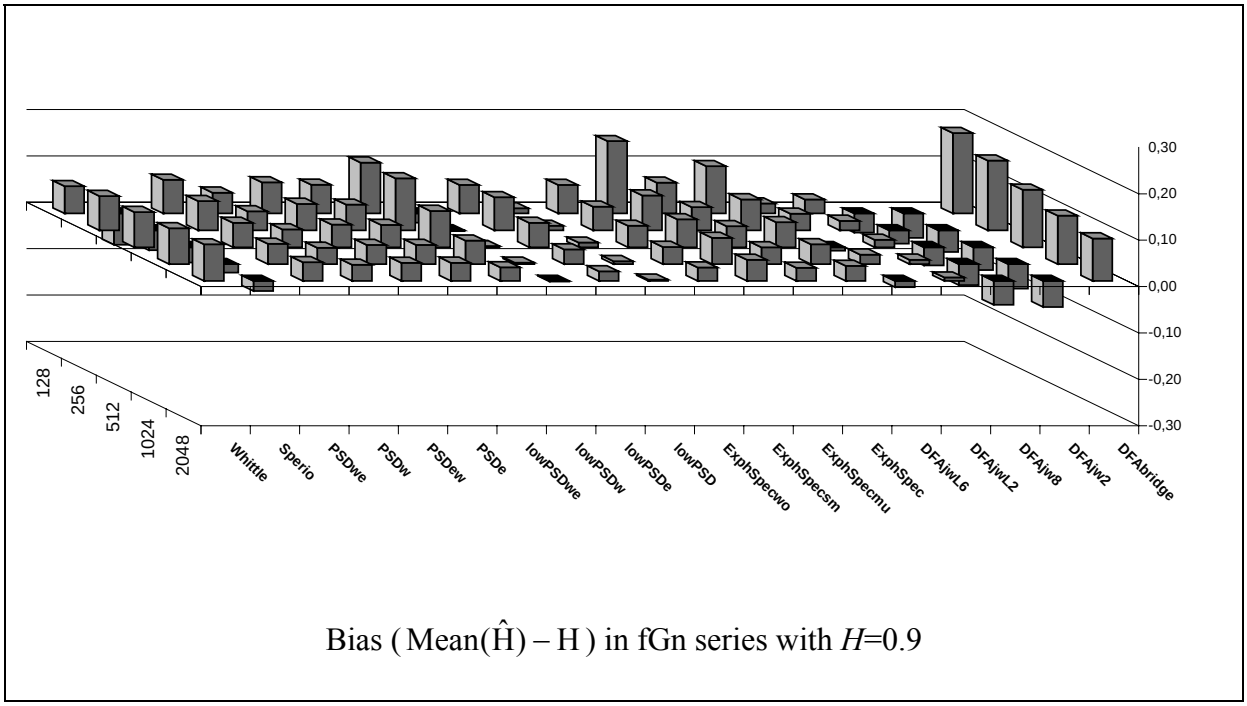


Figure 5.1.3. Bias for the different PSD, ExphSpec and DFA modifications and Whittle and Sperio in fGn series (presented in alphabetical order) with H=0.9 at T=128, 256, 512, 1024 and 2048.

The ExphSpec methods' estimations are less biased in fBm series with H=0.1 with ExphSpecsm delivering the third smallest bias of all estimators under evaluation, surpassed and followed by the PSD methods excluding the high-spectral estimates for the fitting of the spectral slope. $^{low}PSD_{we}$ performing second best is the only method (next to ExphSpecsm at T=128) overestimating the true parameter in short series (for $T \leq 512$), while all other estimates are negatively biased, regardless of sample size.

Distinctly underestimated, with little or no effect of sample size on bias, is the Hurst coefficient in fBm series with H=0.1 by the PSD procedures employing the high-spectral estimates in their analysis as well as Whittle and the DFA modifications (except DFAbridge).

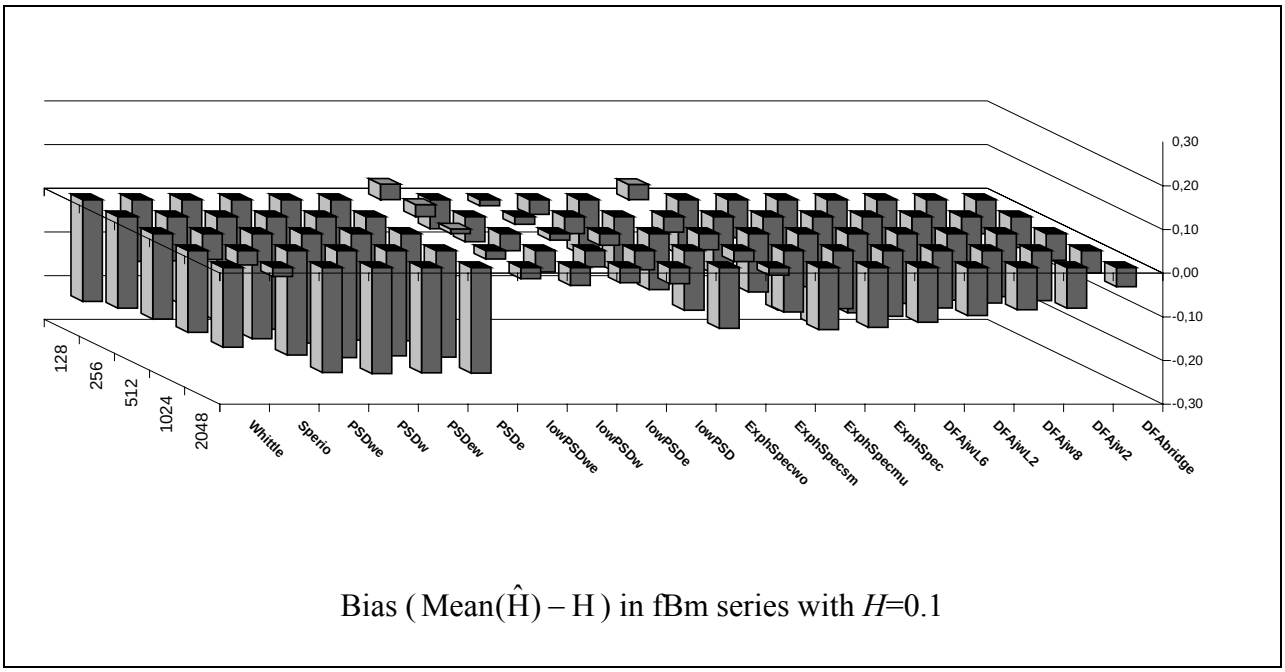


Figure 5.1.4. Bias for the different PSD, ExphSpec and DFA modifications and Whittle and Sperio in fBm series (presented in alphabetical order) with H=0.1 at T=128, 256, 512, 1024 and 2048.

To summarize, the PSD estimations employing all frequencies in their analysis, the Whittle as well as the DFA estimates show the least variability of all methods under evaluation, regardless of sample size and parameterization. On the edge of the 1/f boundary, the PSD estimates including the high-frequency spectral estimates, perform best. However, the PSD methods excluding the high-frequency spectral estimates, as well as the DFA

methods and Sperio are least biased regardless of sample size and parameterization. On the edge of the 1/*f* boundary, the true parameter is overestimated in fGn and underestimated in fBm series. Here, the PSD methods employing only the frequencies close to zero in their analysis deliver satisfying results for both signal types. Fairly small biases are also obtained by the DFA methods (not DFAbridge) in the fGn case and ExphSpecsm in fBm series.

In conclusion, the PSD estimates employing only the frequencies close to zero are least biased but highly variable. Associated with somewhat larger biases but small standard errors are the DFA methods, except DFAbridge, which is clearly biased for both signal types.

Classification of fractal signals

The analysis of the reliable classification of the procedures under evaluation is based on the non-transformed estimations given by the procedures under evaluation, that is, the power exponent $\hat{\beta}$ of the PSD and ExphSpec methods, the scaling exponent $\hat{\alpha}$ of the DFA procedures and the fractional differencing parameter $\hat{d}$ given by Whittle and Sperio. Recall, for the PSD and ExphSpec methods, fGn typically exhibits a power exponent $-1 \leq \beta < +1$, while $+1 > \beta \leq +3$ represents fBm processes. For DFA, the scaling exponent α varying from $0 < \alpha < 1$ implies fGn, and exponents from 1 to 2 represent fBm processes, for Whittle and Sperio $d < +1$ represent fGn and $d > +1$represent fBm processes. Note that $H = \alpha$ for fGn and $H = \alpha - 1$ for fBm, $\alpha = d + 0.5$, $d = \beta/2$ as well as $\beta < 1$ for fGn and $\beta > 1$ for fBm, so α, β and d allow a definitive classification of a series as fGn or fBm simply by the magnitude of the parameter estimation.

First, to determine the reliability of the procedures under evaluation the percentage of misclassification of each estimator is computed in a fully-crossed factorial design, i.e., for all levels of *H, T* and signal type. Second, minimum and maximum value of the non-transformed estimations of the most reliable tools are presented to give the researcher an idea what to expect in a single analysis.

Percentage of misclassifications

The analysis of the classification of fractal signals shows a considerable effect of parameterization and signal type on misclassification. All fGn series with $H \leq 0.6$, independent of sample size, are properly classified as fGn signals by all procedures under evaluation, except by the PSD methods employing only the frequencies close to zero in the analysis, which may be due to the high variability of this group (see Table 5.1.2). Thus only the percentage of misclassifications for $H \geq 0.7$ will be tabulated.

Table 5.1.4 shows the percentage of misclassifications of fGn series with H=0.7, 0.8 and 0.9 at T=128, 265, 512, 1024 and 2048 and in addition the percentage of misclassifications averaged over all sample sizes. Overall, there is a clear effect of sample size on the percentage of wrong decisions. At H=0.7, only DFAjwL6 and Whittle correctly identify all fGn series as such in 100% of all cases. Only few false decisions for series with $T \leq 256$ are obtained by the remaining DFA methods with the exception of DFAbridge. All PSD estimates employing only the frequencies close to zero in their analysis perform distinctly poor with a percentage of misclassifications up 48% for short series with T=128. For fGn series with H=0.8, again, Whittle and the DFA group (except DFAbridge) deliver the least amount of misclassifications, followed by the PSD modifications including the high-frequency spectral estimates in their analysis. Still, more than half of all estimation methods under evaluation manage to correctly classify fGn series with T>512 and DFAjw8 correctly identifies practically all series independently of sample size.

On the edge of the 1/f boundary, that is for fGn series with H=0.9, DFAjw8 and DFAjw2 still manage to correctly identify in more than 90% of all trials the true fGn signal. The DFA modifications' results is in sharp contrast to that of DFAbridge, which delivers an average of about 60% of wrong decisions which is the worst performance of all procedures

under evaluation. The PSD variations including all frequencies for the fitting of the spectral slope perform better in identifying the true fGn signal than the ^{low}PSD group. However, for PSD_{we} (19%) and $^{low}PSD_{w}$(24%) the results aren't distinctly different.

Table 5.1.4. Percentage of misclassification for fGn series with *H*=0.7, 0.8, and 0.9, at *T*= 128, 256, 512, 1024, 2048, and averaged for all sample sizes. The data is sorted in ascending order by the results averaged independently of sample size (right column).

% of misclassifications for fGn series							
H	Method	T					sample sized averaged
		128	256	512	1024	2048	
0.7	DFAjwL6	0	0	0	0	0	0
	Whittle	0	0	0	0	0	0
	DFAjw8	0	0.1	0	0	0	0.02
	DFAjwL2	0.1	0	0	0	0	0.02
	DFAjw2	0.1	0.1	0	0	0	0.04
	PSD_w	0.5	0	0	0	0	0.1
	ExphSpecwo	0.8	0	0	0	0	0.16
	PSD_e	0.9	0	0	0	0	0.18
	ExphSpecmu	1.0	0.1	0	0	0	0.22
	PSD_{ew}	1.4	0	0	0	0	0.28
	PSD_{we}	2.2	0.1	0	0	0	0.46
	Sperio	2.9	0.8	0.4	0.1	0	0.84
	ExphSpec	4.9	0.2	0	0	0	1.02
	DFAbridge	10.6	0.8	0.1	0	0	2.3
	ExphSpecsm	11.0	1.0	0	0	0	2.4
	^{low}PSD	17.5	7.4	1.3	0	0	5.24
	$^{low}PSD_w$	18	6.8	2.1	0.1	0	5.4

	$^{low}PSD_e$	38.7	27.6	12.8	2.3	0.2	16.32
	$^{low}PSD_{we}$	48.4	38.2	20.9	5.1	0.4	22.6
0.8	DFAjw8	1.4	0.3	0.3	0	0	0.40
	Whittle	2.5	0.4	0	0	0	0.58
	DFAjw2	2.3	0.8	0.4	0.1	0	0.72
	DFAjwL2	5.0	0.4	0.1	0	0	1.10
	PSD_w	6	0.4	0	0	0	1.28
	PSD_e	7.9	1.1	0	0	0	1.80
	DFAjwL6	6.9	2.1	0.2	0	0	1.84
	PSD_{ew}	9.8	1	0	0	0	2.16
	ExphSpecmu	10.1	1.9	0.2	0	0	2.44
	ExphSpecwo	10.0	2.2	0.2	0	0	2.48
	PSD_{we}	12.1	1.8	0	0	0	2.78
	Sperio	9.1	5.5	3.7	2	0.4	4.14
	ExphSpec	20.4	7.2	1.4	0	0	5.80
	^{low}PSD	26.4	17.9	6.6	1.4	0	10.46
	ExphSpecsm	35.3	14.2	3.5	0.6	0.1	10.74
	$^{low}PSD_w$	26.2	17.8	8	2.8	0	10.96
	DFAbridge	40.7	20.4	7.9	0.3	0	13.86
	$^{low}PSD_e$	40.2	31.6	20.2	7	0.8	19.96
	$^{low}PSD_{we}$	47.6	39.7	29.1	12.3	2.3	26.20
0.9	DFAjw8	10.6	9.6	6.4	5.3	4.8	7.3
	DFAjw2	12.3	10.2	6.8	5.6	4.8	7.9
	DFAjwL2	28.2	14.9	6.0	2.0	0	10.2
	DFAjwL6	26.4	22.7	5.2	5.3	1.3	12.2
	PSD_w	30.1	20.3	9	2.9	0.5	12.56
	PSD_e	37	25.4	12.4	2.9	0.3	15.6

	Sperio	20.1	19.9	17.5	13.8	8.2	15.9
	PSD_{ew}	38.2	27.8	14.2	4.8	0.6	17.12
	ExphSpecmu	42.9	28.7	16.5	5.3	0.3	18.7
	ExphSpecwo	41.4	30.1	18.5	5.6	0.5	19.2
	PSD_{we}	41.1	32.5	16.2	5.6	0.7	19.22
	$^{low}PSD_{w}$	36.7	33.8	24.5	18.4	6.8	24.04
	^{low}PSD	41.1	33.9	25	15.6	5.8	24.28
	ExphSpec	49.4	37.3	23.4	10.3	1.5	24.4
	Whittle	32.3	33.5	26.7	20.2	12.9	25.12
	$^{low}PSD_{e}$	45.3	46.5	36.5	24.2	11.5	32.8
	ExphSpecsm	64.1	44.7	34.8	27.0	18.1	37.7
	$^{low}PSD_{we}$	51.7	54.1	44	33.2	15.7	39.74
	DFAbridge	72.3	75.4	67.9	54.7	40.4	62.14

Table 5.1.5 shows the percentage of misclassifications of fBm series with H=0.1, 0.2 and 0.3 at T=128, 265, 512, 1024 and 2048 and in addition the percentage of misclassifications averaged over all sample sizes and parameters. Overall, there is a clear effect of sample size on the amount of false decisions, except for fBm series with H=0.1. There, the percentage of misclassification even increases for longer series for some Exphspec methods, the PSD modifications including all frequencies for the fitting of the spectral slope, and Whittle.

For H=0.1 ExphSpecsm delivers surprisingly good results due to only about 9% of misclassifications in long series with T=2048, whereas the PSD modifications including all frequencies in their analysis clearly outperform the ^{low}PSD group in the identification of the true fGn signals and vice versa in the fBm case, at least for low values of H. However, the superiority of the DFA group in the *fG*n case can't be repeated in fBm series. The PSD

modification including all frequencies as well as Whittle completely fail in the classification of fractal signals if presented an fBm series with H=0.1, regardless of sample size, because they misclassify the true fBm signals in almost 100%.

For fBm series with H=0.2 the ^{low}PSD group is having more trouble in identifying the true signal. However, the ExphSpec group is doing surprisingly well, only surpassed by DFAbridge with an average of 1.6% of false decisions. Series with T>256 are all correctly classified. Even Whittle is delivering good results with only about 6.7% of false decisions. The DFA group (except DFAbridge) and the PDS modifications including all frequencies in their analysis perform worst but still deliver only up to a third of wrong decisions compared to up to 100% of misclassifications in fBm series with H=0.1.

With increasing parameter value the amount of misclassification decreases dramatically. The worst performance at H=0.3 are about 7% of false identifications of the true fBm signal. The largest amount of erroneous decisions are obtained by the ^{low}PSD group, DFAjw8 and DFAjw2. There is a clear deterioration in the performance of the ^{low}PSD group, compared to the performance of the remaining estimators with increasing coefficient value. Again, DFAbridge (almost up to 100% of correct decisions), the ExphSpec methods and Whittle deliver surprisingly good results with practically no false decisions for sample sizes with T>256.

Table 5.1.5. Percentage of misclassification for fBm series with H=0.1, 0.2, and 0.3, at T= 128, 256, 512, 1024, 2048, and averaged over all sample sizes. The data is sorted in ascending order by the results averaged over all sample sizes (right column).

% of misclassifications for fBm series							
H	Method	T					sample sized averaged
		128	256	512	1024	2048	
0.1	ExphSpecsm	22.6	30.7	22.9	14.6	8.8	19.92
	$^{low}PSD_{we}$	33.8	28.1	21.2	19.5	13.9	23.3
	DFAbridge	41.8	35.3	27.1	16.1	7.1	25.48
	$^{low}PSD_{e}$	38.5	33.6	27.8	24.6	14.6	27.82
	^{low}PSD	39.2	37.7	28.8	28.2	16.9	30.16
	$^{low}PSD_{w}$	44.2	40.4	32.4	28.9	18.9	32.96
	Sperio	54.7	48.6	37.2	25.9	19.4	37.16
	ExphSpec	47.1	54.6	61.3	76.9	86.6	65.3
	DFAjw2	83.6	75.9	63.8	56.5	47.2	65.4
	DFAjw8	89.7	80.6	68.6	58.3	48.3	69.1
	ExphSpecwo	57.1	60.9	69.5	79.9	93.6	72.2
	ExphSpecmu	56.5	62.3	72.1	84.1	95	74
	DFAjwL2	87.7	87.2	84	73.7	57.4	78
	DFAjwL6	90.6	96.2	85.4	82.8	65.7	84.14
	PSD_{we}	92.0	98.3	99.9	100	100	98.04
	Whittle	94.4	97.4	99	99.9	100	98.14
	PSD_{ew}	93.3	98.7	100	100	100	98.4
	PSD_{w}	94.1	99.3	100	100	100	98.68
	PSD_{e}	95.2	99.3	99.9	100	100	98.88
0.2	DFAbridge	7.0	1.0	0	0	0	1.60
	ExphSpecsm	5.9	2.6	0.9	0.2	0.2	1.96

	ExphSpecmu	7.6	3.5	0.2	0.1	0	2.28
	ExphSpecwo	8.8	3.3	0.4	0.1	0	2.52
	ExphSpec	9.4	3.6	0.5	0.1	0	2.72
	Whittle	23.9	7.8	1.7	0	0	6.68
	DFAjwL2	29.4	16.9	5.6	1.2	0	10.62
	^{low}PSD	25.8	16.3	8.5	2.2	0.3	10.62
	$^{low}PSD_{we}$	25.2	17.9	8.5	2.2	0.4	10.84
	$^{low}PSD_{e}$	28.0	19.1	9.0	2.6	0.1	11.76
	$^{low}PSD_{w}$	29.4	23.9	10.5	3.6	0.8	13.64
	Sperio	32.7	20.5	12.6	4.1	2.9	14.56
	DFAjwL6	37.6	29.1	16.8	2.1	4.2	17.96
	DFAjw8	41.4	27.8	17.1	15.8	12.7	22.96
	DFAjw2	38.8	29.8	19.7	17.3	13.9	23.90
	PSD_{we}	40.0	35.0	29.2	24.5	12.2	28.18
	PSD_{ew}	38.8	37.6	29.7	23.5	12.9	28.50
	PSD_{e}	41.9	40.2	30.2	24.3	12.4	29.80
	PSD_{w}	43.7	41.4	34.9	26.0	14.3	32.06
0.3	DFAbridge	0.2	0	0	0	0	0.04
	ExphSpec	0.6	0.3	0	0	0	0.2
	ExphSpecmu	0.7	0.2	0	0	0	0.2
	ExphSpecwo	0.7	0.2	0	0	0	0.2
	Whittle	1.6	0	0	0	0	0.32
	ExphSpecsm	0.9	0.9	0.1	0	0	0.4
	PSD_{we}	3.5	1.0	0.1	0	0	0.92
	PSD_{w}	4.9	1.1	0	0	0	1.2
	PSD_{ew}	4.8	1.3	0.1	0	0	1.24
	PSD_{e}	5.2	1.1	0	0	0	1.26

	DFAjwL2	5.7	0.9	0.1	0	0	1.3
	DFAjwL6	9.8	2.5	1.3	0	0	2.7
	^{low}PSD	14.8	7.6	1.7	0.2	0	4.86
	Sperio	14.8	7.5	2.9	0.9	0.4	5.3
	^{low}PSD$_{we}$	17.7	7.3	1.7	0.2	0	5.38
	DFAjw8	12.5	5.2	4.1	2.2	3.1	5.4
	^{low}PSD$_{w}$	17.1	9.3	2.6	0.2	0	5.84
	^{low}PSD$_{e}$	19.7	8.9	2.0	0.2	0	6.16
	DFAjw2	14.9	8	5.9	4.1	3.6	7.3

In addition, the percentage of misclassification for fractal signals averaged over all parameter at T=128, 265, 512, 1024, and 2048, and averaged over all parameter and sample sizes is shown in Table 5.1.6. The parameter averaged results for fGn series confirm the findings shown in Table 5.1.4 for H=0.7, 0.8, and 0.9. The DFA methods (not DFAbridge) and the PSD modifications including the high-frequency spectral estimates in their analysis are the most reliable estimators for correctly identifying fGn signals, with even less than 1% of misclassifications for DFAjw8 and DFAjw2, independently of parameterization and time series length. In contrast, the ^{low}PSD and ExphSpec methods as well as DFAbridge are clearly less qualified for the identification of an fGn series, especially ^{low}PSD$_e$ and ^{low}PSD$_{we}$ with up to 30% of false decisions.

Interestingly, those procedures with the poorest performance in the fGn case are the most powerful in the identification of fBm series. ExphSpec and DFAbridge correctly identify in more than 95% of all cases the true fBm signal, regardless of parameterization and sample size. The remaining DFA methods and the PSD modifications including the high-frequency spectral estimates in their analysis are the least reliable estimators in the fBm case. The

performance of ExphSpec, ExphSpecwo and ExphSpecmu as well as Whittle and Sperio are mediocre at best for both fractal signals.

Table 5.1.6. Percentage of misclassification of fractal signals averaged over all parameter at *T*=128, 256, 512, 1024, and 2048, and averaged over all parameter and sample sizes. The data is sorted in ascending order by the results averaged over all sample sizes and parameter (right column).

Signal Type	Method	% of misclassification					
		T (averaged over all parameter)					averaged over parameter and sample size
		128	256	512	1024	2048	
fGn	DFAjw8	1.33	1.11	0.74	0.59	0.53	0.86
	DFAjw2	1.63	1.23	0.80	0.63	0.53	0.97
	DFAjwL2	3.70	1.70	0.68	0.22	0.00	1.26
	PSD_w	4.07	2.30	1.00	0.32	0.06	1.55
	DFAjwL6	3.70	2.76	0.60	0.59	0.15	1.56
	PSD_e	5.09	2.94	1.38	0.32	0.03	1.95
	PSD_{ew}	5.50	3.20	1.58	0.53	0.07	2.18
	Sperio	3.67	2.92	2.40	1.77	0.96	2.34
	ExphSpecmu	6.00	3.41	1.86	0.59	0.03	2.38
	ExphSpecwo	5.80	3.59	2.08	0.62	0.06	2.43
	PSD_{we}	6.17	3.82	1.80	0.62	0.08	2.50
	Whittle	3.87	3.77	2.97	2.24	1.43	2.86
	ExphSpec	8.34	4.97	2.76	1.14	0.17	3.48
	^{low}PSD	11.94	7.09	3.70	1.89	0.64	5.05
	$^{low}PSD_w$	12.22	7.08	3.88	2.37	0.76	5.26
	ExphSpecsm	12.51	6.66	4.26	3.07	2.02	5.70
	DFAbridge	13.84	10.73	8.43	6.11	4.49	8.72

	$^{low}PSD_e$	47.90	38.17	19.73	5.76	1.39	22.59
	$^{low}PSD_{we}$	56.93	47.67	28.94	10.81	2.10	29.29
fBm	ExphSpecsm	3.37	3.80	2.66	1.64	1.00	2.49
	DFAbridge	5.44	4.03	3.01	1.79	0.79	3.01
	$^{low}PSD_{we}$	11.38	6.48	3.52	2.43	1.59	5.08
	^{low}PSD	11.12	7.33	4.39	3.40	1.91	5.63
	$^{low}PSD_e$	13.77	7.73	4.36	3.04	1.63	6.11
	$^{low}PSD_w$	13.23	8.91	5.13	3.63	2.19	6.62
	Sperio	12.53	8.78	5.93	3.43	2.52	6.64
	ExphSpec	6.38	6.50	6.87	8.56	9.62	7.58
	ExphSpecwo	7.41	7.16	7.77	8.89	10.40	8.32
	ExphSpecmu	7.21	7.33	8.03	9.36	10.56	8.50
	DFAjwL2	13.73	11.67	9.97	8.32	6.38	10.01
	DFAjw8	16.47	12.70	10.04	8.50	7.14	10.97
	DFAjw2	16.09	12.93	10.07	8.71	7.21	11.00
	Whittle	13.32	11.69	11.19	11.10	11.11	11.68
	DFAjwL6	15.56	14.21	11.50	9.43	7.77	11.69
	PSD_{we}	15.14	14.92	14.36	13.83	12.47	14.14
	PSD_{ew}	15.28	15.29	14.42	13.72	12.54	14.25
	PSD_e	15.88	15.62	14.46	13.81	12.49	14.45
	PSD_w	15.93	15.76	14.99	14.00	12.70	14.68

When analyzing empirical time series, the researcher may have some hypotheses about the signal type and internal dependency structure, however, only the series length is known for sure. Hence, the percentage of misclassification, independent of signal type and parameterization, for the different sample sizes T=128, 265, 512, 1024, and 2048, as well as averaged over signal type, paramterization and sample size, in shown in Table 5.1.7.

Table 5.1.7. Percentage of misclassification of fractal signals averaged over signal type and parameter, at *T*=128, 512, 1024, 2048, and averaged over signal type, parameter and sample size. The data is sorted ascending by the percentage of misclassification independent of parameterization, sample size and signal type.

Method	% of misclassification					
	T (averaged over signal type and parameter)					averaged over signal type, parameter, and sample size
	128	256	512	1024	2048	
ExphSpecsm	7.94	5.23	3.46	2.36	1.51	4.10
Sperio	8.10	5.85	4.17	2.60	1.74	4.49
ExphSpecwo	6.61	5.37	4.92	4.76	5.23	5.38
ExphSpecmu	6.61	5.37	4.94	4.97	5.29	5.44
^{low}PSD	11.53	7.20	4.04	2.64	2.01	5.49
ExphSpec	7.36	5.73	4.81	4.85	4.89	5.53
DFAjwL2	8.72	6.68	5.32	4.27	3.19	5.64
DFAbridge	9.64	7.38	5.72	3.95	2.64	5.87
DFAjw8	8.90	6.91	5.39	4.54	3.84	5.92
^{low}PSD$_w$	12.73	7.99	4.51	3.00	1.47	5.94
DFAjw2	8.86	7.08	5.43	4.67	3.87	5.98
DFAjwL6	9.63	8.48	6.05	5.01	3.96	6.63
Whittle	8.59	7.73	7.08	6.67	6.27	7.27
PSD$_w$	10.00	9.03	7.99	7.16	6.38	8.11
PSD$_e$	10.48	9.28	7.92	7.07	6.26	8.20
PSD$_{ew}$	10.39	9.24	8.00	7.13	6.31	8.21
PSD$_{we}$	10.66	9.37	8.08	7.23	6.27	8.32
^{low}PSD$_e$	30.83	22.95	12.04	4.40	1.51	14.35
^{low}PSD$_{we}$	34.16	27.07	14.23	7.62	2.84	17.19

Averaged over signal type, parameter, and sample size, ExphSpecsm and Sperio deliver the least amount of misclassifications with less than 5% false decisions for the identification of fractal signals with $T \geq 512$, suggesting that time series with long-term dependence structures should consist of a minimal of around 500 observations. Table 5.1.7 also shows that Exphspecsm renders the least percentage of false decisions independently of sample size and should therefore be the first choice of all procedures under evaluation – regardless of time series length – for the classification of a fractional series as motion or noise. However, if the analysis suggests an fGn process, a reanalysis with DFAjw8 or DFAjw2 is suggested because of their superiority in the fGn case.

Figure 5.1.5 visualizes the interdependence of signal type and estimator performance. While ExphSpecsm is clearly outperforming Sperio in the fBm case, the percentage of false decisions in fGn signals is obviously much smaller for Sperio. There is little or no effect of sample size on the correlation of signal type and estimator performance, but a clear effect on the amount of false decisions resulting in only few misspecifications on the edge of the $1/f$ boundary.

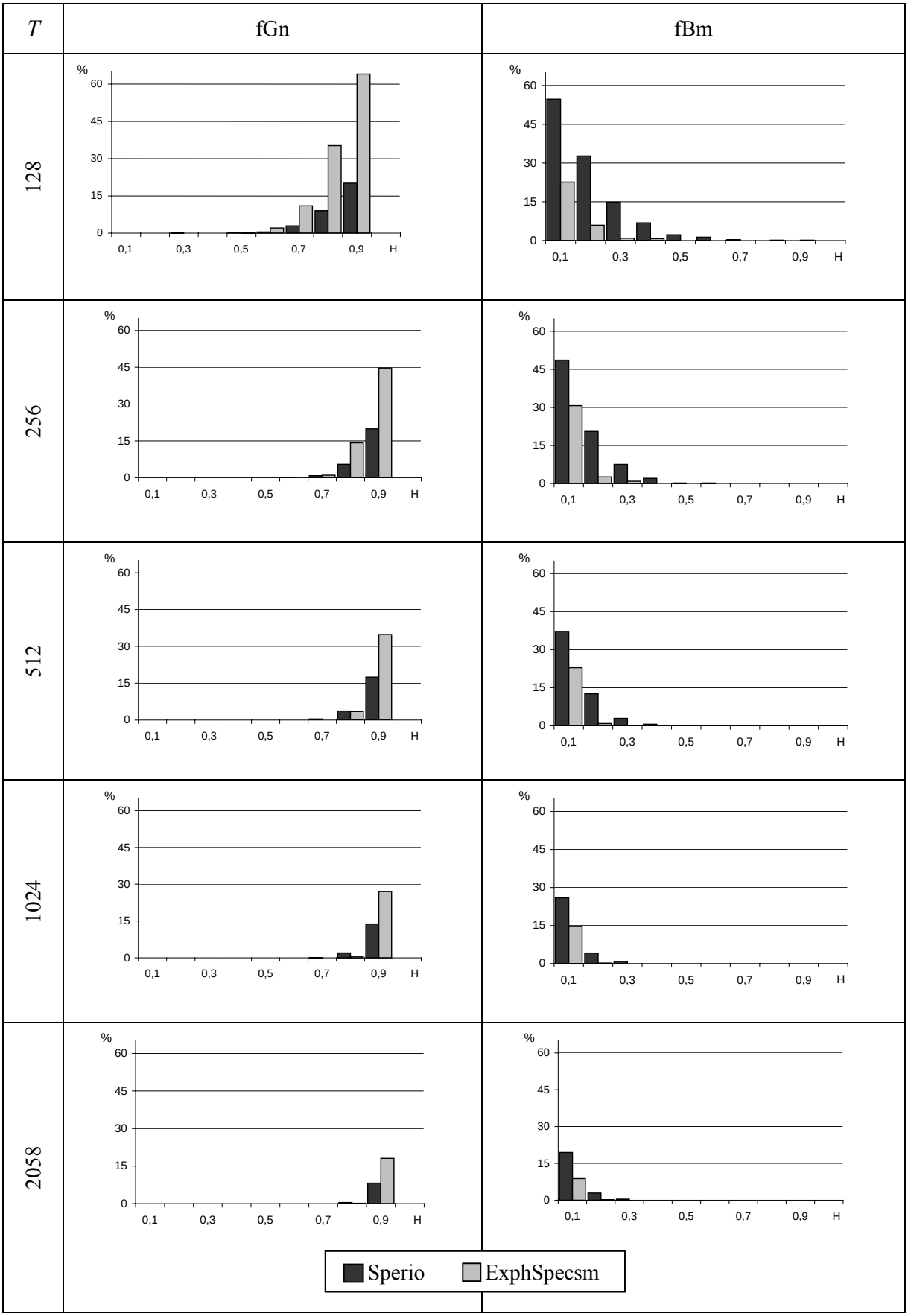


Figure 5.1.5. Percentage of misclassifications received from ExphSpecsm and Sperio for fGn and fBm signals of H=0.1-0.9 with T=128, 256, 512, 1024, and 2048.

Furthermore, it may be interesting to know what to expect in a single analysis, that is, to know the extreme values given by ExphSpec and Sperio received over 1000 replications. Therefore the minimum and maximum values of estimation as well as the percentage of misclassification received by ExphSpecsm and Sperio are presented in Table 5.1.8 for fGn signals, in Table 5.1.9 for fBm signals. In the fGn case, there is a positive association between the magnitude of the extremes and percentage of misclassifications, i.e., the more extreme the minimum or maximum values, the larger the number of false decisions. In the fBm case a negative association between the magnitude of the extreme values and the number of false decisions can be observed: the smaller the percentage of misclassification, the more extreme are the minimum and maximum values.

Table 5.1.8. Minimum and Maximum value of estimations ($\bar{\beta}, \hat{d}$) and percentage of misclassifications received for fGn series by ExphSpecsm and Sperio.

		ExphSpecsm ($\bar{\beta}$)				Sperio ($\hat{d}$)		
T	H_{true}	Min	Max	% misclass		Min	Max	% misclass
128	0.1	-1.0557	0.4284	0		-1.0947	0.1647	0
	0.2	-1.0987	0.6580	0		-1.0993	0.2637	0
	0.3	-1.0290	0.7012	0		-0.9958	0.5767	0.1
	0.4	-0.6264	0.9022	0		-0.7724	0.3843	0
	0.5	-0.3683	1.0218	0.1		-0.7554	0.6475	0.3
	0.6	-0.5656	1.1296	2.1		-0.7381	0.5957	0.5
	0.7	-0.0421	1.3254	11.0		-0.7581	0.7084	2.9
	0.8	-0.0778	1.3902	35.3		-0.7038	0.7856	9.1
	0.9	0.1716	1.6291	64.1		-0.3840	0.9217	20.1
256	0.1	-0.5805	0.3160	0		-0.8793	0.1076	0
	0.2	-0.7122	0.4506	0		-0.9384	0.1648	0

	0.3	-0.2769	0.5835	0		-0.9555	0.1731	0
	0.4	-0.2710	0.7160	0		-0.5944	0.2761	0
	0.5	-0.0399	0.8446	0		-0.5472	0.4374	0
	0.6	0.0879	0.9494	0		-0.5203	0.5034	0.1
	0.7	0.0317	1.0697	1.0		-0.3879	0.6206	0.8
	0.8	0.3108	1.2256	14.2		-0.4340	0.6982	5.5
	0.9	0.4688	1.3067	44.7		-0.3233	0.8505	19.9
512	0.1	-0.3951	0.2326	0		-0.8074	-0.0217	0
	0.2	-0.1749	0.3592	0		-0.6640	0.0308	0
	0.3	-0.1791	0.5134	0		-0.7268	0.1683	0
	0.4	-0.0044	0.6611	0		-0.5965	0.3301	0
	0.5	0.0554	0.7366	0		-0.4609	0.3469	0
	0.6	0.3189	0.8492	0		-0.3860	0.4579	0
	0.7	0.3192	0.9326	0		-0.5033	0.5761	0.4
	0.8	0.4558	1.1334	3.5		-0.4646	0.6401	3.7
	0.9	0.5005	1.1793	34.8		-0.1092	0.7518	17.5
1024	0.1	-0.2970	0.2092	0		-0.7430	-0.0494	0
	0.2	-0.1949	0.3380	0		-0.6370	-0.0123	0
	0.3	-0.0171	0.4630	0		-0.5238	0.1038	0
	0.4	0.1362	0.6033	0		-0.6488	0.2189	0
	0.5	0.2501	0.6628	0		-0.3731	0.3070	0
	0.6	0.3625	0.8238	0		-0.2766	0.3965	0
	0.7	0.4227	0.9024	0		-0.2111	0.5037	0.1
	0.8	0.5353	1.0289	0.6		-0.0623	0.6024	2
	0.9	0.7229	1.1299	27.0		-0.0327	0.7993	13.8
2048	0.1	-0.1843	0.1973	0		-0.7128	-0.1478	0
	0.2	-0.0585	0.3334	0		-0.5776	-0.0583	0

	0.3	0.0614	0.4069	0		-0.5488	0.0529	0
	0.4	0.1574	0.5295	0		-0.4053	0.1275	0
	0.5	0.2918	0.6931	0		-0.4499	0.2805	0
	0.6	0.4131	0.7553	0		-0.2604	0.3406	0
	0.7	0.4890	0.8713	0		-0.1082	0.4987	0
	0.8	0.6175	1.0014	0.1		-0.0373	0.5577	0.4
	0.9	0.7429	1.0940	18.1		0.0957	0.6069	8.2

Table 5.1.9. Minimum and Maximum value of estimations ($\bar{\beta}, \hat{d}$) and percentage of misclassifications received for fBm series by ExphSpecsm and Sperio.

		ExphSpecsm ($\bar{\beta}$)				Sperio ($\hat{d}$)		
T	H_{true}	MIN	MAX	% MISCLASS		MIN	MAX	% MISCLASS
128	0.1	-0.847	2.3534	22.6		-0.4107	1.1422	54.7
	0.2	-0.1306	2.9986	5.9		-0.2097	1.1632	32.7
	0.3	0.1664	3.3014	0.9		-0.1410	1.3668	14.8
	0.4	0.3712	3.3086	0.7		0.0671	1.4410	6.8
	0.5	1.0502	3.5642	0		0.1429	1.5212	2.2
	0.6	1.207	3.773	0		0.2292	1.4893	1.3
	0.7	1.1622	4.2308	0		0.2769	1.6182	0.3
	0.8	0.9754	4.0122	0.1		0.5046	1.5020	0
	0.9	0.8492	4.5458	0.1		0.5107	1.7156	0
256	0.1	0.0258	1.8608	30.7		-0.1554	1.0527	48.6
	0.2	0.4362	2.2644	2.6		0.0522	1.1142	20.5
	0.3	0.3804	2.5496	0.9		0.1198	1.2595	7.5
	0.4	1.1146	2.7162	0		0.3178	1.3534	2

	0.5	1.1268	2.9296	0		0.3767	1.4371	0.2
	0.6	1.209	3.0826	0		0.2562	1.5291	0.2
	0.7	1.4114	3.3932	0		0.5299	1.5071	0
	0.8	1.4004	3.6272	0		0.6658	1.5278	0
	0.9	1.5414	3.6612	0		0.8013	1.6991	0
512	0.1	0.4646	1.7284	22.9		-0.0376	0.9570	37.2
	0.2	0.5216	1.992	0.9		0.2099	1.1011	12.6
	0.3	0.7142	2.2706	0.1		0.2713	1.1600	2.9
	0.4	1.2024	2.4992	0		0.3538	1.2258	0.6
	0.5	1.4686	2.6084	0		0.3892	1.3353	0.1
	0.6	1.6264	2.8906	0		0.6171	1.4016	0
	0.7	1.7222	3.0232	0		0.6013	1.4468	0
	0.8	1.5858	3.1258	0		0.5351	1.5941	0
	0.1	1.7762	3.412	0		0.8288	1.6589	0
1024	0.1	0.53	1.5404	14.6		0.1138	0.9357	25.9
	0.2	0.9318	1.9268	0.2		0.2117	1.0147	4.1
	0.3	1.1488	2.2114	0		0.3604	1.1014	0.9
	0.4	1.3916	2.368	0		0.5713	1.2310	0
	0.5	1.5962	2.5132	0		0.5939	1.2781	0
	0.6	1.8776	2.7518	0		0.7023	1.3947	0
	0.7	1.7874	2.9508	0		0.7858	1.4407	0
	0.8	1.8152	3.0756	0		0.8166	1.5678	0
	0.9	1.9534	3.2968	0		0.9352	1.5269	0
2048	0.1	0.7592	1.5038	8.8		0.1067	0.8782	19.4
	0.2	0.901	1.9006	0.2		0.3086	0.9832	2.9
	0.3	1.137	2.1518	0		0.4525	1.0963	0.4
	0.4	1.5556	2.292	0		0.5760	1.1722	0

	0.5	1.748	2.431	0		0.6967	1.2470	0
	0.6	1.8678	2.593	0		0.7499	1.3219	0
	0.7	1.786	2.7964	0		0.8076	1.3822	0
	0.8	1.837	3.0982	0		0.8856	1.4683	0
	0.9	1.8222	3.1766	0		0.9647	1.5072	0

5.1.6 Conclusions

In this study the performance of the 8 PSD, 4 ExphSpec and 5 DFA versions as well as the Whittle and Sperio methods as classification tools for fractal signals has been empirically evaluated by means of simulated fGn and fBm signals with H variations of 0.1 to 0.9, and T variations of 128, 265, 512, 1024, 2048, each replicated a 1000 times.

Regardless of signal type, parameterization and sample size, ExphSpecsm and Sperio perform best followed by the remaining ExphSpec variations and ^{low}PSD. On the average, all procedures under evaluation correctly identify the true signal in more than 85% of all series with $T \geq 512$ (more than 90% of series with $T \geq 1024$ and 96% of series with $T=2048$).

ExphSpecsm and Sperio correctly classify the true signal in more than 95% for $T \geq 512$ suggesting both as adequate tools for a preliminary classifications of time series as Gaussian noises or Brownian motions in fractal signals with about 500 observations.

However, if the result of an empirical analysis suggests an fGn signal, a reanalysis with DFAjw8 and DFAjw2 is suggested, since both misclassify fGn signals in less than 1%, independently of sample size. If the preliminary empirical analysis suggests an fBm signal, a reanalysis with DFAbridge will surely confirm the results, since both ExphSpecsm and DFAbridge perform best in classifying fBm signals regardless of sample size with around 3% or less false decisions.

Nevertheless, if a single analysis of an empirical time series of T≥128 with ExphSpecsm should result in an estimated beta somewhere around one, there is still a chance of misclassification. The analysis of the extreme values has demonstrated. that ExphSpecsm produces estimates greater than one in the case of true fGn series of H≥0.6 even at T≥1024 as well as estimates smaller than one in the case of true fBm series of H≤0.2 and T≥1024. Therefore a reanalysis with Expecmu, ExphSpecwo or DFAbridge might be in order, since these procedures perform almost as well as ExphSpecsm.

However, the underlying structure of a time series may not be known, the number of observation surely is. Hence, after the preliminary classification the researcher may look into Table 5.1.6 to determine the best procedure for the suggested signal type and given time series length. For example, in short fGn series with T≤256, DFAjw8 and DFAjw2 outperform DFAjwL2, whereas the latter is superior in classifying longer fGn series with T≥512.

If the signal type is adequately determined, an accurate parameter estimate of an fGn series may be received through the DFA variations DFAjwL2 and/or DFAjwL6, which have delivered estimates with the smallest standard errors of all procedures under evaluation in the fGn case. Estimates with the smallest standard error in the fBm case are best obtained by the PSD methods including the high-frequency spectral estimates for the fitting of the spectral slope as well as DFAjwL2 or Whittle, leaving DFAjwL2 as the procedure delivering an estimation of the fractional parameter with the least variability of all procedures under evaluation. However, on the edge of the 1/f boundary, the smallest SEs for fGn series with H=0.9 and fBm series with H=0.1, are received by the PSD methods including the high-frequency spectral estimates as well as the Whittle method.

Although more variable, the PSD estimates excluding the high-frequency spectral estimates, as well as the DFA methods and Sperio are least biased, independently of signal

type, sample size and parameterization. On the edge of the 1/*f* boundary, the PSD methods employing only the frequencies close to zero in their analysis deliver the least biased estimates for both signal types.

To summarize, the most reliable preliminary classification tools of fractal signals are ExphSpecsm and Sperio. If an fGn signal is identified, a reanalysis with DFAjw8 and DFAjw2 is suggested. Furthermore, all DFA estimates (not DFAbridge) have proven to be least variable and biased, independently of signal type, parameterization and sample size. The most accurate estimator for the respective signal type and parameterization as well as given sample size may be obtained by using the results of Tables 5.1.2 and 5.1.3 as guidelines, since the estimation accuracy highly depends on signal type, parameterization and sample size.

5.2 Study 2: Distinguishing (Non-)Stationary Processes

5.2.1 Introduction

One of the main concepts of time series analysis is the stationarity of processes, that is, to determine if a time series' mean, variance and covariance are stable over time. While the memory of stationary series is always final, that of nonstationary processes may be not depending on the cause of the nonstationarity. In the ARFIMA framework, processes with differencing parameters smaller than 0.5 are stationary, while models with $d \geq 0.5$ are unstable due to a stochastic trend and usually require differencing as a stabilizing transformation. Therefore, misclassifications of the series type could lead to unnecessary transformation of stationary processes, that is, overdifferencing or underdifferencing of nonstationary data. There has been some debate in the literature arguing that overdifferencing is a less serious error than underdifferencing. Nevertheless, all kind of inappropriate transformations are consequential for subsequent statistical analyzes and should be omitted. For an overview, consult Maddala & Kim (1998) or Stadnytska (2009c).

Stationary and nonstationary ARFIMA processes exhibit different memory properties. Stationary series with $0<d<0.5$ have autocovariances that decay much more slowly than those of ARMA processes and therefore possess long memory. In nonstationary series with $0.5\leq d<1$ the impact of the series innovations are very persistent but not permanent, however, in processes with $d=1$ the impact of random shocks persist forever. Hence, they possess an infinite memory.

The distinction between stationary and nonstationary ARFIMA processes is also important within the cointegration framework, since cointegration is commonly defined as a stationary relation between nonstationary variables (Granger, 1981, 1986).

Special procedures called *unit root tests* were developed to prove stationarity conditions. Test versions available in popular statistical packages like R, SAS or EViews typically check the null hypothesis $d=1$ against $d=0$, thus they are not appropriate for differentiating between fractionally integrated series. Implementation of approaches specially designed for fractional data like Lagrange Multiplier tests by Robinson (1994) or Tanaka (1999) demand some additional programming. A computationally simple testing technique proposed by Dolado et al. (2002) requires a reasonable pre-estimate of d. In the fractional case it is also possible to test for stationarity employing Wold type tests by obtaining a point estimate of the memory parameter and building confidence intervals around it assuming a normal distribution (Geweke & Porter-Hudak, 1983; Robinson, 1992; Sowell, 1992). Test properties in this case, however, strongly depend on procedures used for estimating d.

Different evaluation studies for various estimators of the fractional differencing parameter d have revealed that (1) the most accurate procedures, like the exact maximum likelihood approach ML proposed by Sowell (1992), can handle only stationary processes, i.e., d is constrained by the ML optimization algorithm to be not greater than 0.5, (2) methods

that can be directly applied to nonstationary data are often distinctly biased, (3) the magnitude of the bias strongly depends on parameterizations, and (4) semiparametric procedures (GPH, Sperio) tend to yield large confidence intervals, whereas the precision of parametric methods (Whittle, ML) depends on the correct specification of the model (Hauser et al., 1999; Reisen et al., 2001; Smith et al., 1997; Stadnytska & Werner, 2006; Stroe-Kunold et al., 2009; Taqqu et al., 1995; Taqqu & Teverovsky, 1996).

To sum up, stationary and nonstationary processes exhibit different memory characteristics. These properties can be identified by the magnitude of the ARFIMA differencing parameter *d*, and thus allowing the proper transformation of the data for further statistical analysis as well as the choice of the suitable estimator, since some estimation methods are constraint to stationary processes with strongly biased results for nonstationary series.

The main goal of the following study is to evaluate the ability of estimators implemented in R as well as own modifications to reliably classify stationary and nonstationary ARFIMA series. Not only fractal noises but also more complex models with either one AR- or MA-term or both will be considered to investigate to what extend the (additional) presence of short-term component(s) may affect the reliability of the procedures under evaluation. Since the fractional differencing parameter *d* and the scaling exponent α of the DFA methods and power exponent β of the ExphSpec and PSD variations (see Chapters 3.4.3 and 5.1.3) are related by $\beta = 2d = 2\alpha - 1$, the DFA, ExphSpec and PSD methods evaluated in Chapter 5.1 (see Chapter 5.1.3, p.56) are also considered next to Whittle, Sperio, and GPH (see Chapter 3.4.3) as direct estimators of *d*.

In addition to the investigation to what extend the above mentioned procedures may qualify as diagnostic tools for the preliminary classification of stationary or nonstationary

ARFIMA series, the accuracy of the fractional differencing parameter estimates of the methods under evaluation will be assessed, too.

5.2.2 Method

The reliability of the abovementioned methods is tested by simulated ARFIMA processes with known differencing parameter *d* by employing the command *fracdiff.sim* of the R package *fracdiff* (for details, consult the R documentation at http://ftp5.gwdg.de/pub/misc/cran/). The R-code for the different stationary and nonstationary ARFIMA models is attached. For this study, ARFIMA series with *d* variations from 0.1 to 0.9, each replicated a 1000 times, are used. Manipulated are the following **independent variables**:

- value of *d*: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9;
- Model: ARFIMA (0,*d*,0); ARFIMA (1,*d*,0) with autoregressive parameters ranging from +/–0.2 to +/–0.8 by step of 0.2; ARFIMA (0,*d*,1) with moving average coefficients from +/–0.2 to +/–0.8 by step of 0.2; ARFIMA (1,*d*,1) with the same autoregressive parameterizations as in the (1,*d*,0) case combined with the moving average coefficient θ=0.3;
- length of series *T*= 128, 265, 512, 1024, 2048;
- 20 estimation methods:
  - estimating $\hat{d}$: Whittle, Sperio, GPH;
  - estimating $\hat{\beta}$: ExphSpec, ExphSpecsm, ExphSpecwo, ExphSpecmu, ^{low}PSD, ^{low}PSD$_e$, ^{low}PSD$_w$, ^{low}PSD$_{we}$, PSD$_e$, PSD$_{ew}$, PSD$_w$, PSD$_{we}$;
  - estimating $\hat{\alpha}$: DFAbridge, DFAjw2, DFAjw8, DFAjwL2, DFAjwL6;

As a quality criterion the following **dependent variables** were computed:

- percentage of signal misclassification (MISCLASS) of the nontransformed estimations $\hat{d}$, $\hat{\beta}$ and $\hat{\alpha}$: $\hat{d} \geq 0.5$ (Whittle, Sperio, GPH), $\hat{\beta} \geq 1$ (ExphurstSpec and PSD methods) and $\hat{\alpha} \geq 1$ (DFA methods) stand for nonstationarity, $\hat{d} < 0.5$, $\hat{\beta} < 1$ and $\hat{\alpha} < 1$ stand for stationarity series, respectively.
- Mean (*M*), standard error (*SE*) of the transformed estimates (all estimates are transformed prior for comparison, see Table 3.3.1, p.35) , minimum (*MIN*), maximum (*MAX*) of the original estimations $\hat{d}$, $\hat{\beta}$ and $\hat{\alpha}$, respectively.

5.2.3 Results

In the following chapter the accuracy of the estimation methods under evaluation is studied first by assessing the estimations' standard error and bias. The procedures' capability of reliably distinguishing stationary and nonstationary processes are considered thereafter while taking the percentage of misclassifications into account. Since in empirical settings, i.e., with non-simulated data, the true structure is never known, a strategy for reliably classifying (non)stationary ARFIMA processes, regardless of process type, model or parameterization, will be developed. An additional investigation of the extreme parameter estimates of those procedures performing best will be done, so it may give the researcher an idea what to expect in the case of a single analysis of an empirical time series.

Accuracy of Estimation

The analysis of the accuracy of the procedures under evaluation requires the transformation of the estimated scaling exponent $\hat{\alpha}$ and power exponent $\hat{\beta}$, given by the DFA and PSD and ExphSpec methods respectively, into the fractional differencing parameter $\hat{d}$ (see Chapter 5.2.2 and Table 5.2.1).

The quality of parameter estimation is assessed by computing the mean standard error and bias of each estimation (over 1000 replications) for each level of the independent variables, i.e., model, process type (stationary or nonstationary), value of *d,* and time series length *T* (see Chapter 5.2.2). Special consideration will be given to the sign of the bias, as even a small positive bias in stationary processes with $d<0.5$, or a small negative bias for nonstationary processes with $d>0.5$, around the border of nonstationary, may lead to misclassification of a series as nonstationary or stationary, respectively.

Standard Error

Overall, there is only a small effect of sample size on standard error (*SE*), independently of model type, process type and *d*. Generally, the standard errors at *T*=128 are about two to four times the magnitude compared to *T*=2048, and are smallest for Whittle and the PSD methods including the high-frequency spectral estimates for the fitting of the spectral slope, and largest for GPH and Sperio. Since the standard error always decreases with increasing sample size, only *SE*s at *T*=2048 are tabulated. Table 5.2.2 shows the procedures with the first two smallest and largest *SE*s, i.e., those estimates with the smallest and largest variability, for ARFIMA (0,*d*,0) series at *T*=2048, received from 1000 replications.

Table 5.2.2. Procedures with the first two smallest and largest standard errors for ARFIMA(0,*d*,0) series at *T*=2048, received from 1000 replications.

Model	*d*	Smallest *SE*			Largest *SE*	
0d0	0.1	Whittle .0005	PSD_{ew}, PSD_{w} .0007		GPH .0035	$^{low}PSD_{we}$ .0030
	0.4	Whittle .0006	PSD_{e}, PSD_{w}, PSD_{ew}, PSD_{we} .0007		GPH .0034	Sperio .0029
	0.5	Whittle .0006	PSD_{e},PSD_{w}, PSD_{ew} .0007		GPH .0034	Sperio .0029

	0.9	PSD_e, PSD_w, PSD_{ew}, PSD_{we} .0007	Whittle .0008		DFAjw2 .0038	DFAjw8 .0037

Regardless of parameterization, the smallest standard errors for estimating the fractional differencing parameter *d* in fractal noises are obtained by Whittle and the PSD procedures also including the high frequencies in their analysis, the largest SEs by GPH, Sperio and the DFA variations. As demonstrated in Table 5.2.2, there is only little variability in the procedures' performances within the model albeit distinctively different process types, that is, stationary and nonstationary processes.

For the more complex models, the additional presence of short-term components has no effect on the standard error's behavior, i.e., extreme values and range of the *SE* is almost identical in fractional noise with and without short-term dependencies. Hence, only the methods receiving the first two smallest and largest *SE*s of the ARFIMA(1,*d*,0), ARFIMA(0,*d*,1) and ARFIMA(1,*d*,1) series (all at *d*=0.4 and *T*=2048) are shown in Table 5.2.3.

Table 5.2.3. Procedures with the first two smallest and largest standard errors for ARFIMA(0,*d*,0), ARFIMA(0,*d*,1), and ARFIMA(1,*d*,1) series at *d*=0.4 and *T*=2048, received from 1000 replications.

Model	ϕ, θ	Smallest *SE*			Largest *SE*	
(0,*d*,1)	θ=0.2	PSD_e, PSD_w, PSD_{ew}, PSD_{we} .0007	ExphSpecmu, ExphSpecwo .0008		GPH .0036	DFAjw2 .0029
	θ=0.4	PSD_e, PSD_w, PSD_{ew}, PSD_{we},Whittle .0007	ExphSpecmu .0008		GPH .0035	Sperio .0029
	θ=0.6	PSD_e, PSD_w, PSD_{ew} .0007	PSD_{ew}, ExphSpecmu .0007		GPH .0035	Sperio .0029
	θ=0.8	PSD_w .0007	PSD_{ew}, PSD_e, ExphSpecmu .0008		GPH .0036	$^{low}PSD_{we}$ .0034

(1,*d*,0)	ϕ=0.2	Whittle .0006	PSD_e, PSD_w, PSD_{ew}, PSD_{we} .0007		GPH .0036	Sperio .0029
	ϕ=0.4	Whittle, PSD_e .0006	,PSD_{ew}, PSD_{we}, PSD_w .0007		GPH .0035	Sperio .0029
	ϕ=0.6	Whittle, PSD_e, PSD_{ew}, PSD_{we} .0007	PSD_w, ExphSpecmu, ExphSpecwo .0008		GPH .0034	Sperio .0029
	ϕ=0.8	Whittle, PSD_e, PSD_{ew}, PSD_{we} .0007	ExphSpecmu .0008		GPH .0034	Sperio .0028
(1,*d*,1)	ϕ=0.2 θ=0.3	Whittle .0006	PSD_e, PSD_{ew}, PSD_w .0007		GPH .0035	DFAjw2 .0029
	ϕ=0.4 θ=0.3	Whittle .0007	PSD_e, PSD_{we}, PSD_{ew}, , PSD_w .0008		GPH .0035	DFAjw2 .0029
	ϕ=0.6 θ=0.3	Whittle .0006	PSD_e, PSD_{we}, PSD_{ew}, PSD_w .0007		GPH .0034	DFAjw2 .0028
	ϕ=0.8 θ=0.3	Whittle .0007	PSD_e, PSD_{we}, PSD_{ew}, PSD_w .0008		GPH .0034	Sperio .0027

Obviously, there is no effect of model type and parameterization on the variability of parameter estimation. Whittle as well as the PSD methods including only windowing and/or endmatching deliver the smallest standard errors over all model types and parameter values. GPH and Sperio as well as the DFA variations perform likewise consistent, yet their estimates show the largest variability of all procedures under evaluation, regardless of model type and parameterization.

Bias

Another measure of accuracy next to the standard error is the bias ($Mean(\hat{d}) - d$) of the estimate. While a small standard error will surely promote the correct identification of the true data-generating process, a smaller bias may in some cases lead to more false decisions than a larger deviation. For example, a small positive bias overestimating the fractional differencing parameter of a stationary process on the border of nonstationary leads to misclassification as a

nonstationary series, while a large but negative bias would have lead to the right conclusion, even though the estimate would have been distinctively biased. Hence, not only the absolute magnitude but also the sign of the bias is relevant and has to be investigated.

To determine the accuracy of estimation, first, the bias modulus of all procedures under evaluation at T=2048 will be compared, followed by a more refined analysis of the sign of the bias at the different levels of d, ϕ, and θ.

For ARFIMA (0,d,0) series, there is no clear effect of sample size and parameterization on bias for all procedures. With no short-term memory coefficients present, the sample size and parameter averaged bias modulus in ARFIMA (0,d,0) series is smallest for $^{low}PSD_w$ (0.0096), Whittle (0.0098), and GPH (0.0189), and largest for DFAbridge (0.0980) and $^{low}PSD_w$ (0.0917).

Table 5.2.4. shows the sample size and parameter averaged bias modulus of ARFIMA (0,d,0) series considering stationary and nonstationary ARFIMA (0,d,0) processes separately. For stationary series, the averaged bias modulus is smallest for ^{low}PSD (0.0054), ExphSpecmu (0.0073), and GPH (0.0075), and largest for $^{low}PSD_{we}$ (0.1543) and $^{low}PSD_e$ (0.1142). For nonstationary series with $d\geq0.5$, the averaged bias modulus is smallest for Whittle (0.0083) and $^{low}PSD_w$ (0.0104). Clearly, the PSD estimates excluding the high frequencies from their analysis are less biased in nonstationary processes, except for ^{low}PSD with an exceptionally small bias in stationary noises. Furthermore, the superiority of the ExphSpec methods in stationary series does not hold in the nonstationary case.

Table 5.2.4. Sample size and parameter averaged bias modulus for stationary and nonstationary ARFIMA(0,*d*,0) series, received from 1000 replications. The data is sorted in ascending order.

Type of Model	Stationary series			Nonstationary series	
	Estimator	Sample size and parameter averaged bias modulus		Estimator	Sample size and parameter averaged bias modulus
(0,*d*,0)	^{low}PSD	0.0054		Whittle	0.0083
	ExphSpecmu	0.0073		$^{low}PSD_w$	0.0116
	GPH	0.0075		$^{low}PSD_e$	0.0151
	ExphSpecwo	0.0076		Sperio	0.0258
	$^{low}PSD_w$	0.0077		DFAjwL2	0.0283
	Whittle	0.0121		GPH	0.0291
	DFAjwL2	0.0148		^{low}PSD	0.0315
	ExphSpec	0.0165		$^{low}PSD_{ew}$	0.0328
	PSD_{ew}	0.0194		DFAjwL6	0.0408
	PSD_e	0.0200		$^{low}PSD_{we}$	0.0433
	DFAjwL6	0.0217		ExphSpecmu	0.0525
	PSD_{we}	0.0227		ExphSpec	0.0594
	PSD_w	0.0257		ExphSpecwo	0.0614
	ExphSpecsm	0.0375		PSD_{we}	0.0628
	Sperio	0.0402		PSD_{ew}	0.0656
	DFAjw8	0.0581		PSD_e	0.0699
	DFAjw2	0.0663		PSD_w	0.0715
	DFAbridge	0.0914		DFAjw8	0.0784
	$^{low}PSD_{ew}$	0.1123		DFAjw2	0.0835
	$^{low}PSD_e$	0.1142		DFAbridge	0.1075
	$^{low}PSD_{we}$	0.1543		ExphSpecsm	0.1220

Table 5.2.5 shows the sample size and parameter averaged bias modulus for stationary and nonstationary ARFIMA (0,*d*,1), ARFIMA(1,*d*,0), and ARFIMA(1,*d*,1) series. While the Whittle and the ^{low}PSD estimates are least biased in fractional noise, the Whittle estimates are distinctively biased in the presence of additional short memory parameters. Stationary and nonstationary ARFIMA (0,*d*,1) series are estimated best by GPH and the PSD methods excluding the high-frequency spectral estimates for the fitting of the spectral slope. In the autoregressive case, i.e., in nonstationary and stationary ARFIMA(1,*d*,0) series, their performance is surpassed by Sperio and some of the DFA variations while $\hat{d}$ in mixed ARFIMA(1,*d*,1) series is least biased if estimated by Sperio and GPH.

Table 5.2.5. Sample size and parameter averaged bias modulus for stationary and nonstationary ARFIMA(0,*d*,1), ARFIMA(1,*d*,0), and ARFIMA(1,*d*,1) series, received from 1000 replications. The data is sorted in ascending order.

Type of Model	Stationary series			Nonstationary series	
	Estimator	Sample size and parameter averaged bias modulus		Estimator	Sample size and parameter averaged bias modulus
(0,*d*,1)	GPH	0.0651		$^{low}PSD_{ew}$	0.0543
	^{low}PSD	0.0882		GPH	0.0577
	$^{low}PSD_{w}$	0.0990		$^{low}PSD_{we}$	0.0581
	Sperio	0.1083		$^{low}PSD_{e}$	0.0590
	$^{low}PSD_{ew}$	0.1663		^{low}PSD	0.0741
	ExphSpecsm	0.1822		$^{low}PSD_{w}$	0.0838
	$^{low}PSD_{e}$	0.1893		Sperio	0.0874
	DFAbridge	0.1997		ExphSpecsm	0.1416
	$^{low}PSD_{we}$	0.2448		DFAbridge	0.2074
	DFAjw2	0.2483		DFAjw2	0.2377
	DFAjw8	0.2545		DFAjw8	0.2556

	DFAjwL2	0.2753		ExphSpec	0.2572
	DFAjwL6	0.2949		ExphSpecwo	0.2688
	ExphSpec	0.3013		ExphSpecmu	0.2746
	ExphSpecwo	0.3163		DFAjwL2	0.3031
	ExphSpecmu	0.3169		DFAjwL6	0.3402
	PSD_{we}	0.3279		Whittle	0.3519
	PSD_{e}	0.3414		PSD_{we}	0.4243
	PSD_{ew}	0.3467		PSD_{ew}	0.4305
	Whittle	0.3726		PSD_{e}	0.4333
	PSD_{w}	0.3986		PSD_{w}	0.4419
(1,*d*,0)	Sperio	0.0546		Sperio	0.0508
	GPH	0.0740		DFAjw2	0.0655
	$^{low}PSD_{w}$	0.0805		DFAjw8	0.0772
	^{low}PSD	0.0889		GPH	0.0833
	DFAjw2	0.0945		DFAjwL2	0.1692
	DFAjw8	0.1141		DFAjwL6	0.1836
	$^{low}PSD_{e}$	0.1513		$^{low}PSD_{e}$	0.1915
	$^{low}PSD_{ew}$	0.1564		$^{low}PSD_{w}$	0.2052
	$^{low}PSD_{we}$	0.1808		$^{low}PSD_{ew}$	0.2099
	DFAjwL2	0.2323		$^{low}PSD_{we}$	0.2143
	DFAjwL6	0.2481		^{low}PSD	0.2180
	ExphSpecsm	0.2523		ExphSpecsm	0.2949
	ExphSpecwo	0.3204		ExphSpec	0.3163
	ExphSpecmu	0.3210		DFAbridge	0.3193
	ExphSpec	0.3323		PSD_{e}	0.3391
	PSD_{w}	0.3480		PSD_{w}	0.3394
	DFAbridge	0.3612		PSD_{ew}	0.3435

	PSD_e	0.3613		PSD_{we}	0.3452
	PSD_{ew}	0.3622		Whittle	0.3629
	PSD_{we}	0.3674		ExphSpecmu	0.3723
	Whittle	0.4279		ExphSpecwo	0.3826
(1,*d*,1)	Sperio	0.0558		Sperio	0.0494
	GPH	0.0638		GPH	0.0723
	$^{low}PSD_w$	0.0738		DFAjw2	0.0747
	^{low}PSD	0.0779		DFAjw8	0.0804
	DFAjw2	0.0904		$^{low}PSD_e$	0.0804
	DFAjw8	0.0962		$^{low}PSD_w$	0.0863
	$^{low}PSD_e$	0.1562		$^{low}PSD_{ew}$	0.1010
	$^{low}PSD_{ew}$	0.1590		^{low}PSD	0.1037
	DFAjwL2	0.1592		$^{low}PSD_{we}$	0.1097
	DFAjwL6	0.1631		DFAjwL2	0.1218
	ExphSpecsm	0.1834		DFAjwL6	0.1330
	$^{low}PSD_{we}$	0.1893		PSD_w	0.1729
	PSD_w	0.1927		PSD_e	0.1731
	PSD_e	0.1969		PSD_{ew}	0.1749
	PSD_{ew}	0.1977		PSD_{we}	0.1755
	PSD_{we}	0.1997		Whittle	0.2048
	ExphSpecwo	0.2177		ExphSpec	0.2197
	ExphSpecmu	0.2178		ExphSpecsm	0.2365
	ExphSpec	0.2234		DFAbridge	0.2377
	Whittle	0.2300		ExphSpecmu	0.2521
	DFAbridge	0.2483		ExphSpecwo	0.2606

These findings lead to the conclusion, that $\hat{d}$ is least biased in fractional processes without short-term dependencies if estimated by Whittle and ^{low}PSD, whereas in the presence of an additional moving average coefficient, GPH estimates of *d* are less distorted. However, in the presence of an additional short-term autoregressive parameter, Sperio's estimates are least biased.

Since the Sperio method delivers rather small biases in the autoregressive and mixed case, its behavior will be visulaized more detailed. Figure 5.2.1 shows the bias of the Sperio's estimates for the different values of *d* for ARFIMA (0,*d*,0) series, when plotted separately for *T*=128, 256, 512, 1024, and 2048.

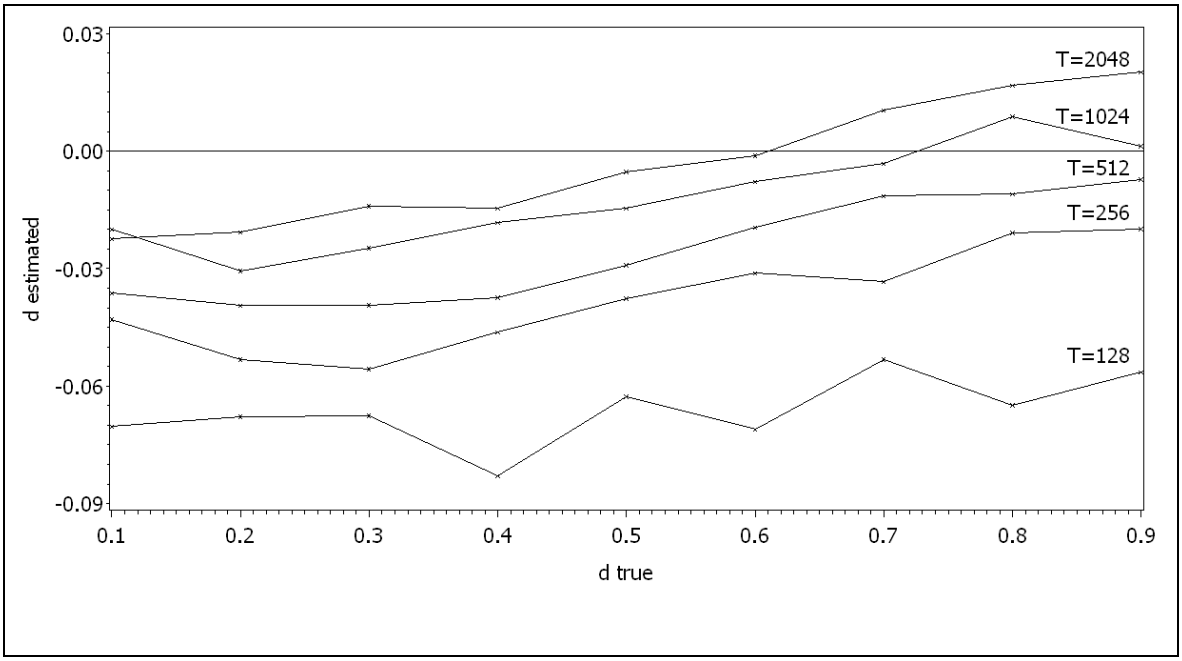


Figure 5.2.1. Bias ($Mean(\hat{d}) - d$) of the Sperio estimates for ARFIMA (0,*d*,0) series with *d*=0.1-0.9, received from 1000 replications each, with separate lines for each step of sample size.

Overall, Sperio underestimates the true *d* with only few overestimations in long nonstationary series. As Figure 5.2.1 demonstrates, extending a time series length does not necessarily lead to more accurate estimations of *d*. Sperio's estimates at $d \geq 0.7$ are clearly less biased for smaller sample sizes (*T*=512 and 1024) than for long series with *T*=2048, and even of equal

magnitude as some short series' estimates (T=256 and d≥0.8). Since Sperio and GPH deliver the least biased estimates in series with additional short memory parameter present, and the Whittle estimates are quite accurate in the estimation of d in fractional noise, their biases, depending on model type and parameterization, is shown in the following figures.

Figure 5.2.2 shows the bias ($Mean(\hat{d}) - d$) received by Sperio, GPH and Whittle in long ARFIMA (0,d,0) series with T=2048 for all steps of the fractional differencing parameter. Regardless of parameterization, the Whittle procedure's estimates are clearly least biased with only slight underestimation of the true parameter in stationary and more pronounced overestimation of d in nonstationary series.

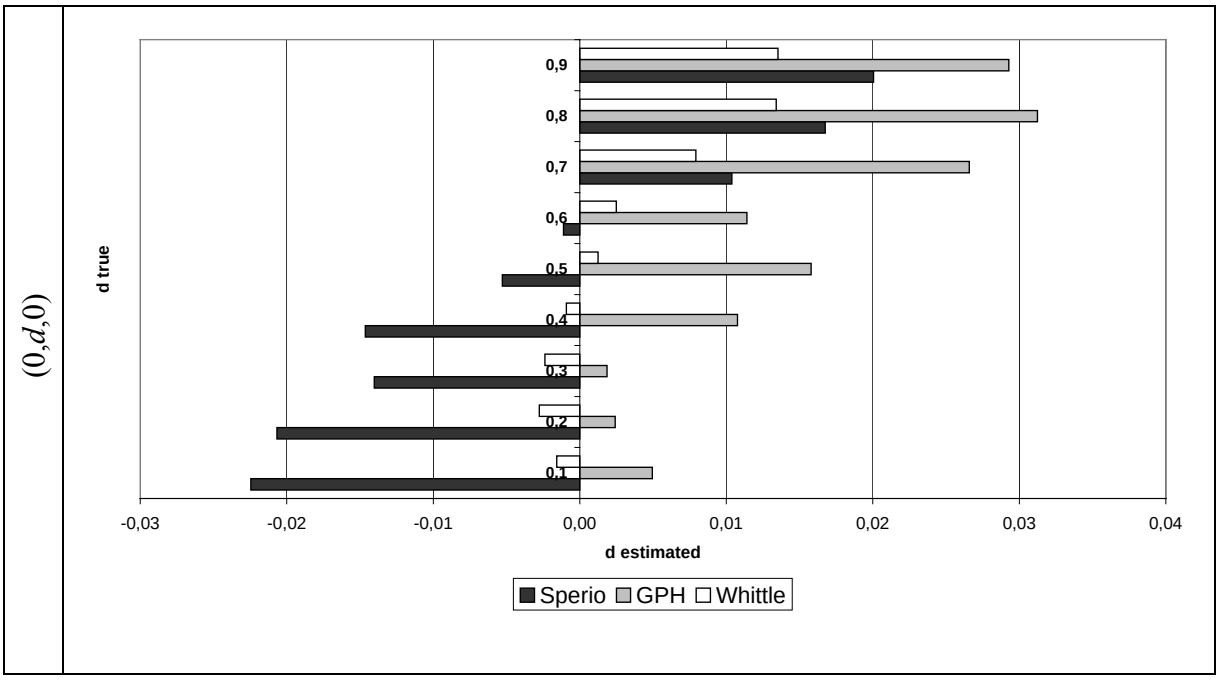


Figure 5.2.2. Bias ($Mean(\hat{d}) - d$) of the Sperio, GPH and Whittle estimates in ARFIMA (0,d,0) series with d=0.1 to 0.9, at T=2048, each over 1000 replications.

The course of the Sperio estimate is similar to that of Whittle but with larger biases, especially in the stationary case. GPH's biases are exclusively positive and increase with growing parameter value. Hence, Sperio performs satisfactorily only within the region of 0.5≤d≤0.6, GPH within the region of 0.2≤d≤0.3, while the Whittle estimates are quite accurate over a

much broader range, i.e., over all stationary series as well as nonstationary series with small coefficients.

The visualization of the bias' behavior for the models with additional short-term memory parameter(s) will not take the Whittle method into account, since it delivers large biases in the estimation of *d* for ARFIMA models with one autoregressive and/or moving-average parameter present. Hence, only the results of the Sperio and GPH procedures for ARFIMA(0,*d*,0) series are shown in Figure 5.2.3.

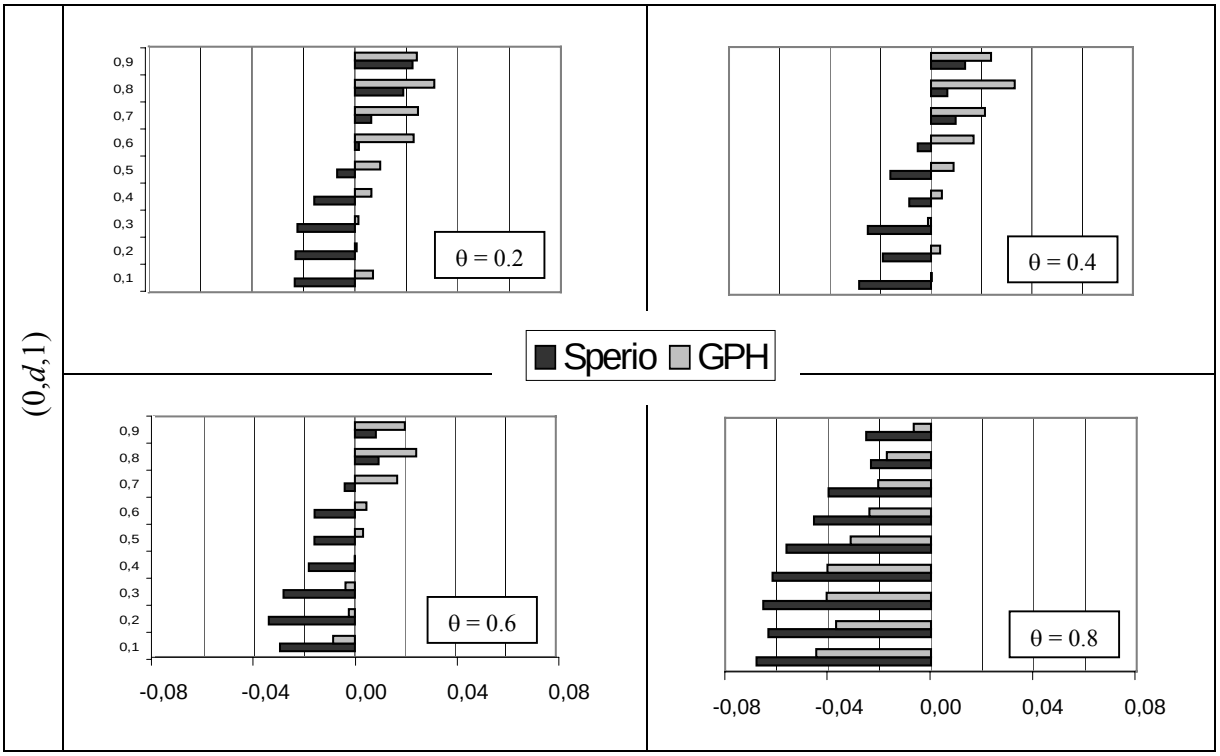


Figure 5.2.3. Bias ($Mean(\hat{d}) - d$) of the Sperio and GPH estimates for all levels of *d* of ARFIMA (0,*d*,1) series, at *T*=2048.

Figure 5.2.4 shows the bias ($Mean(\hat{d}) - d$) of the Sperio and GPH estimates for all levels of *d* of ARFIMA (1,*d*,0) and ARFIMA (1,*d*,1) series, all at *T*=2048.

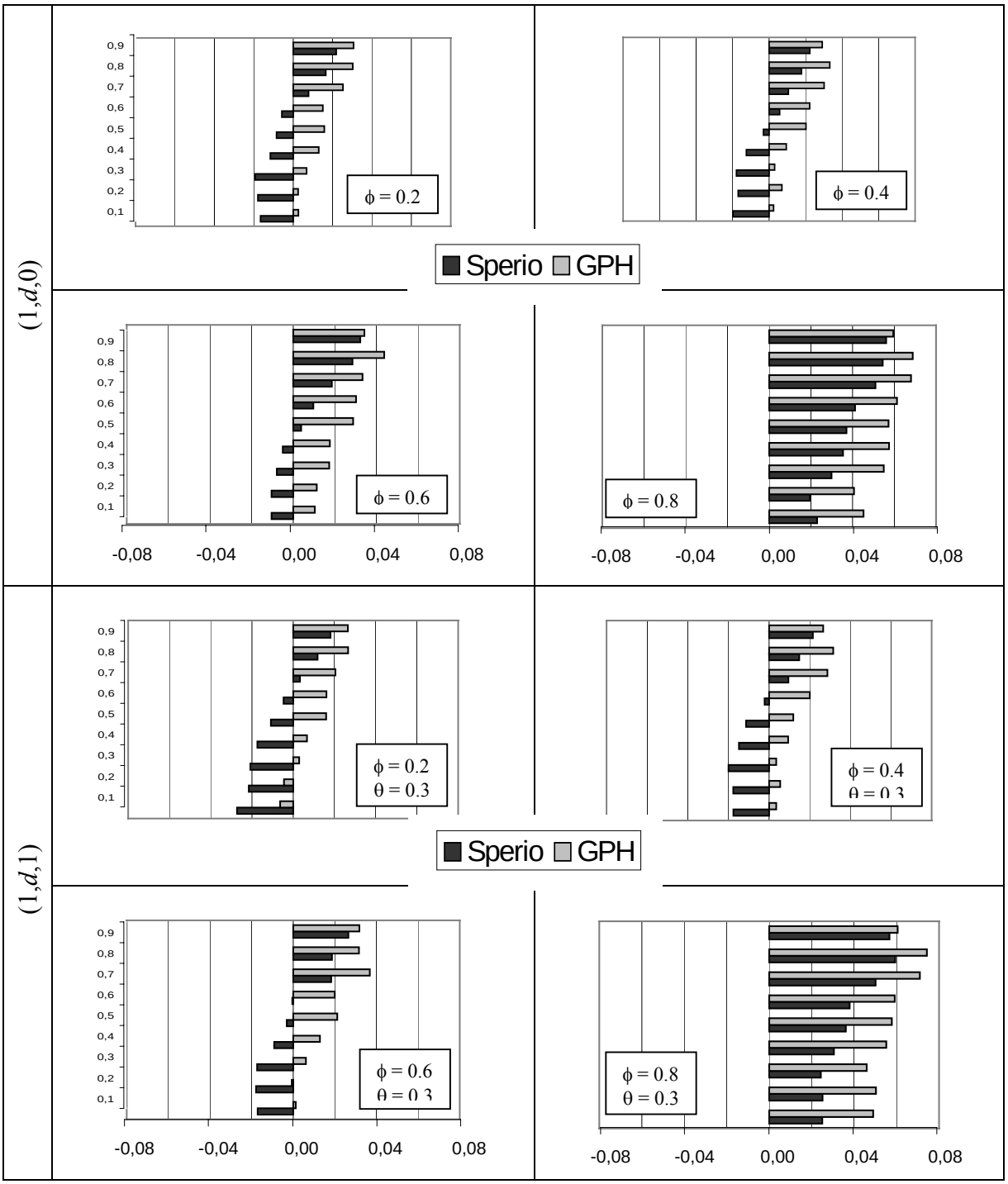


Figure 5.2.4. Bias ($Mean(\hat{d}) - d$) of the Sperio and GPH estimates for all levels of *d* of ARFIMA (1,*d*,0) and ARFIMA (1,*d*,1) series, all at *T*=2048.

In the case of the additional presence of large short-term memory coefficients (θ,ϕ=0.8), the bias of the Sperio and GPH estimates is alike for the different values of *d*. Moving average series with large coefficients are exclusively underestimated, while series

dominated with a large autoregressive coefficient are exclusively overestimated, with smaller biases for GPH in MA-, and for Sperio in AR-processes.

Since processes with a large moving average coefficient are clearly overestimated and processes with a large autoregressive parameter present are distinctively underestimated by both procedures, it is expected, that nonstationary series dominated by large moving average parameters may be more often falsely identified as stationary series, whereas stationary series dominated by large autoregressive parameters may be more frequently classified as nonstationary

Classification of (non)stationary series

The analysis of the classification of series as stationary and nonstationary is based on the non-transformed estimates of the different procedures. Recall, that a true stationary ARFIMA series is correctly identified as such if the fractional differencing parameter, as estimated by Whittle, Sperio and GPH, is smaller than 0.5. Likewise for the PSD and ExphSpec methods, if $\hat{\beta}$ <1, and for the DFA procedures if $\hat{\alpha}$ <1. Correspondingly, an true nonstationary ARFIMA series is correctly identified if $\hat{d} \geq 0.5$, $\hat{\beta} \geq 1$, and $\hat{\alpha} \geq 1$, respectively.

First, the reliability of each procedure under evaluation is determined by computing the percentage of misclassifications in a fully-crossed factorial design, i.e., for all levels of model type, *d* and *T*. Then the percentage of wrong decisions averaged over parameter, sample size, and process types (stationary and nonstationary), and additionally averaged over both stationary and nonstationary series, is computed, to ascertain the probability of each procedure under evaluation to correctly identify the underlying process type in the case of a single analysis of an empirical time series. Finally, the minimum and maximum values of the non-transformed estimates of the procedures performed best are tabulated, so that the researcher may have an idea what to expect in the case of a single analysis.

Percentage of Misclassification

The analysis of the misclassification of fractional ARFIMA processes shows a considerable effect of parameterization, sample size and model type on the amount of false decisions. Table 5.2.6 shows the percentage of misclassifications of ARFIMA (0,*d*,0) series at *d*=0.1-0.9 and averaged over all parameters, at *T*=2048, received from 1000 replications. All stationary series with $d \leq 0.3$ and nonstationary series with $d \geq 0.8$ are correctly identified as such in more than 95% of all trials by all methods under evaluation. However, the behavior of the different methods is highly variable. Procedures like ExphSpec and two of its modifications, DFAjwL2 as well as Whittle manage to correctly classify all stationary and nonstationary series regardless of parameterization in 100% of all cases, except for series with *d*=0.5, while there are at least some erroneous decisions for almost all parameter values for GPH and ${}^{low}PSD_{we}$. The largest amount of false decisions around the border of nonstationary (*d*=0.5) is delivered by the PSD methods using just windowing and/or endmatching for the fitting of the spectral slope (with about twice the amount of false decisions than the PSD variations using only the frequencies close to zero for their analysis), albeit these methods perform perfectly in the identification of stationary series.

Table 5.2.6. Percentage of misclassifications for ARFIMA(0,*d*,0) series with *d*=0.1-0.9, and averaged over all parameters (right column), at *T*=2048, received from 1000 replications. The data is assorted in ascending order by % of misclassification averaged over all parameters.

% of misclassifications for ARFIMA(0,*d*,0) series										
	d_{true}									
Method	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Averaged over all parameters
DFAbridge	0	0	0	14.1	2.9	0	0	0	0	1.9
ExphSpecsm	0	0	0	11	19.7	0.6	0.1	0	0	3.5
ExphSpec	0	0	0	0	38.2	0	0	0	0	4.2
ExphSpecwo	0	0	0	0	41.9	0	0	0	0	4.7
ExphSpecmu	0	0	0	0	44	0	0	0	0	4.9
Whittle	0	0	0	0	46.8	0	0	0	0	5.2
^{low}PSD	0	0	0	4.8	46.1	4.3	0	0	0	6.1
$^{low}PSD_e$	0	0	0.4	9.3	40.9	4.9	0	0	0	6.2
$^{low}PSD_{ew}$	0	0	0.9	11.4	38.7	5.5	0.5	0	0	6.3
$^{low}PSD_{we}$	0.1	0.3	1.5	15.1	35.2	4.7	0.3	0	0	6.4
$^{low}PSD_w$	0	0	0	6.1	47.5	6.5	0.3	0.1	0	6.7
DFAjwL2	0	0	0	0	64	0	0	0	0	7.1
Sperio	0	0	0.5	10.8	51.5	13.4	1.3	0.3	0	8.6
DFAjwL6	0	0	0	0	72	8	0	0	0	8.9
GPH	0	0.1	3.5	20.6	43.4	16.2	1.3	0.7	0	9.5
PSD_{we}	0	0	0	0	97.3	2.8	0	0	0	11.1
PSD_{ew}	0	0	0	0	97.9	3.2	0	0	0	11.2
PSD_e	0	0	0	0	98.4	3	0	0	0	11.3
PSD_w	0	0	0	0	98.6	3.7	0	0	0	11.4
DFAjw8	0	0	0	4	74	28	14	0	0	13.3
DFAjw2	0	0	0	6	74	30	14	0	0	13.8

Clearly, the superior method in the classification of fractional noise with d=0.5 and averaged over all parameters is DFAbridge, with more than 98% of correct identifications, independently of parameterization.

Table 5.2.7 shows the percentage of misclassifications for ARFIMA(0,d,1) series with θ=0.2 and 0.8, at d=0.1-0.9, and averaged over all parameters, at T=2048, received from 1000 replications. The pattern of misclassification for the different procedures in moving average processes is similar to that identified in the case of fractional noise. Practically no false decisions are observed for ARFIMA(0,d,1) series at $0.3 \leq d \geq 0.8$. The biggest portion of misclassification for all series is ascertained for d=0.5 and 0.6, with best results for DFAbridge with only about one third of false decisions at d=0.5. Again, the PSD methods including all frequencies in the analysis falsely identify almost all series at the border of nonstationary (d=0.5, 0.6).

With an increasing moving average coefficient, the portion of misclassifications increases particularly for nonstationary series. All nonstationary ARFIMA(0,d,1) series, regardless of parameterization, at θ=0.8, are misclassified as stationary by the PSD methods including all frequencies in their analysis. The true d of fractional series with small moving average parameter values is, as in the fractional noise case, best classified by DFAbridge and ExphSpec, while series with large moving average coefficients are best identified by Sperio and GPH.

Table 5.2.7. Percentage of misclassifications for ARFIMA(0,*d*,1) series with θ=0.2 and 0.8, at *d*=0.1-0.9 and averaged over all parameters, at *T*=2048, received from 1000 replications. The data is assorted in ascending order by % of misclassification averaged over all parameters.

% of misclassifications for ARFIMA(0,*d*,1) series											
		d_{true}									
Method	θ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Averaged over all parameters
DFAbridge	0.2	0	0	0	0.2	36.9	0	0	0	0	4.1
ExphSpecsm		0	0	0	4	35.4	1.8	0	0	0	4.6
$^{low}PSD_{ew}$		0	0	0.4	11.3	43.8	4.5	0.6	0	0	6.7
^{low}PSD		0	0	0	3.8	51.3	5.7	0.1	0	0	6.8
$^{low}PSD_{e}$		0	0	0.9	12.3	46.4	5	0.3	0	0	7.2
$^{low}PSD_{we}$		0	0.5	2.3	16.8	42	4.6	0.3	0	0	7.4
$^{low}PSD_{w}$		0	0	0.1	5.9	54.1	6.6	0.6	0	0	7.5
Sperio		0	0	0.4	8.6	51.5	14.9	2.1	0.1	0	8.6
GPH		0	0.2	3.1	21	44.4	12.6	2.9	0.7	0.2	9.5
DFAjwL2		0	0	0	0	90	10	0	0	0	11.1
ExphSpecwo		0	0	0	0	99	5.7	0	0	0	11.6
ExphSpecmu		0	0	0	0	99.6	6.2	0	0	0	11.8
ExphSpec		0	0	0	0	98.2	7.8	0	0	0	11.8
DFAjw2		0	0	0	0	78	24.1	6	4.4	0	12.5
DFAjw8		0	0	0	0	80	24	6.1	4	0	12.7
DFAjwL6		0	0	0	0	96	20.1	6.3	0	0	13.6
Whittle		0	0	0	0	100	81.8	0	0	0	20.2
PSD_{we}		0	0	0	0	100	100	39.1	0	0	26.6
PSD_{e}		0	0	0	0	100	100	39.7	0	0	26.6
PSD_{ew}		0	0	0	0	100	100	39.8	0	0	26.6

PSD_w		0	0	0	0	100	100	43.3	0	0	27.0
GPH		0	0.1	0.9	8.6	59.6	25.5	6.1	1.2	0	11.3
Sperio		0	0	0.3	3.3	73.3	27.2	5.8	0.2	0	12.2
$^{low}PSD_{we}$		6	4.2	3.6	5.6	91	74.1	30.9	4.8	0.1	24.5
$^{low}PSD_{ew}$		0	0	0.1	0.2	98.5	82.9	34.9	5.2	0.2	24.7
$^{low}PSD_e$		1.7	1.2	1	1.7	95.1	83.2	36.8	5	0.1	25.1
^{low}PSD		0	0	0	0	100	94.1	42.1	4	0	26.7
DFAjw2		2.3	1.8	0.1	0	100	81.8	44.2	13.9	2.1	27.4
$^{low}PSD_w$		0	0	0	99.9	93.7	47.2	8.1	0.3	0	27.7
DFAjw8		1.4	1.2	0	0	100	88	58.1	17.9	2	29.8
ExphSpecsm		0	0	0	0	100	99	65.7	8.9	0	30.4
DFAbridge	0.8	0	0	0	0	100	100	100	90.5	7.2	44.2
DFAjwL6		0	0	0	0	100	100	100	84	20	44.9
DFAjwL2		0	0	0	0	100	100	100	96	20	46.2
ExphSpec		0	0	0	0	100	100	100	100	92.7	54.7
Whittle		0	0	0	0	100	100	100	100	98.9	55.4
ExphSpecmu		0	0	0	0	100	100	100	100	100	55.6
ExphSpecwo		0	0	0	0	100	100	100	100	100	55.6
PSD_e		0	0	0	0	100	100	100	100	100	55.6
PSD_{ew}		0	0	0	0	100	100	100	100	100	55.6
PSD_w		0	0	0	0	100	100	100	100	100	55.6
PSD_{we}		0	0	0	0	100	100	100	100	100	55.6

Table 5.2.8 shows the percentage of misclassifications for ARFIMA(1,*d*,0) series with ϕ=0.2 and 0.8, at *d*=0.1-0.9, and averaged over all parameters, at *T*=2048. Overall, fractional

series with an additional autoregressive parameter present are less often misclassified than fractional moving average processes.

Table 5.2.8. Percentage of misclassifications for ARFIMA(1,*d*,0) series with ϕ=0.2 and 0.8, at *d*=0.1-0.9 and averaged over all parameters, at *T*=2048, received from 1000 replications. The data is assorted in ascending order by % of misclassification averaged over all parameters.

% of misclassifications for ARFIMA(1,*d*,0) series											
		d_{true}									
Method	ϕ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Averaged over all parameters
ExphSpecmu	0.2	0	0	0	10.7	0.1	0	0	0	0	1.2
ExphSpecwo		0	0	0	11.4	0.3	0	0	0	0	1.3
ExphSpec		0	0	0	17.1	1	0	0	0	0	2
DFAjwL2		0	0	0	0	25.1	0	0	0	0	2.8
PSD_w		0	0	0	26.4	0.1	0	0	0	0	2.9
PSD_e		0	0	0	30.6	0.1	0	0	0	0	3.4
PSD_{ew}		0	0	0	34.1	0.1	0	0	0	0	3.8
PSD_{we}		0	0	0	36.3	0	0	0	0	0	4.0
DFAjwL6		0	0	0	2	33.6	3.4	0	0	0	4.3
ExphSpecsm		0	0	0	26.2	12.3	0.4	0	0	0	4.3
$^{low}PSD_e$		0	0	0.5	11	40.5	4.3	0	0	0	6.3
$^{low}PSD_{ew}$		0	0	0.7	13.4	39	4.4	0.1	0	0	6.4
^{low}PSD		0	0	0	7	46.5	4.6	0.1	0	0	6.5
$^{low}PSD_w$		0	0	0	7	47	6.4	0.2	0	0	6.7
$^{low}PSD_{we}$		0	0.1	2.3	17	36.6	4.7	0.2	0	0	6.8
DFAbridge		0	0	0.4	70.8	0.3	0	0	0	0	7.9
Sperio		0	0	0.3	11.4	53.1	16	2.1	0.1	0	9.2
GPH		0	0.1	3	21.3	43.5	15	2.7	0.4	0.2	9.6

Whittle		9.2	0	0	98.6	0	0	0	0	0	11
DFAjw8		0	0	0	4	65.1	28	7.8	21	0	11.9
DFAjw2		0	0	0	4	69.2	30.8	9.9	4.1	0	13.1
Sperio		0	0	2.1	23.7	33.3	6.7	1	0	0	7.4
GPH		0.1	0.5	9	35.7	30.2	7.6	1.1	0.3	0.2	9.4
DFAjw8		0	0	3	37.3	30.7	14.8	2.2	0	0	9.8
DFAjw2		0	0	2.1	29.9	35.8	17.2	8	0	0	10.3
$^{low}PSD_w$		0	6.1	49.6	92.4	0	0	0	0	0	16.5
^{low}PSD		0	4.2	49.6	94.8	0.1	0	0	0	0	16.5
$^{low}PSD_e$		0.4	14.7	59.3	96.1	0	0	0	0	0	18.9
$^{low}PSD_{ew}$		0.4	12.2	63.6	96.1	0	0	0	0	0	19.1
$^{low}PSD_{we}$		2	19.5	67.5	96.4	0.2	0	0	0	0	20.6
DFAjwL6		2.1	22.9	90	100	0	0	0	0	0	23.9
ExphSpecsm	0.8	2.8	65.4	97.7	100	0	0	0	0	0	29.5
DFAjwL2		3.2	75.8	100	100	0	0	0	0	0	31.0
DFAbridge		99.3	100	100	100	0	0	0	0	0	44.4
ExphSpec		100	100	100	100	0	0	0	0	0	44.4
ExphSpecmu		100	100	100	100	0	0	0	0	0	44.4
ExphSpecwo		100	100	100	100	0	0	0	0	0	44.4
Whittle		100	100	100	100	0	0	0	0	0	44.4
PSD_e		100	100	100	100	0	0	0	0	0	44.4
PSD_{ew}		100	100	100	100	0	0	0	0	0	44.4
PSD_w		100	100	100	100	0	0	0	0	0	44.4
PSD_{we}		100	100	100	100	0	0	0	0	0	44.4

Furthermore, the pattern of misclassification for fractional autoregressive processes appears to be diametrically opposed to that of moving average processes. When the short-term

parameter value increases, most estimators, especially the PSD methods including all frequencies in the analysis, misclassify stationary series in the autoregressive and nonstationary series in the moving average case. However, as in the moving average case, the fractional parameter d of purely fractional autoregressive series is best classified by Sperio and GPH, except when the short memory parameters are low.

Table 5.2.8 shows the percentage of misclassifications for ARFIMA(1,d,1) series with ϕ =0.2 and 0.8, and θ=0.3, at d=0.1-0.9, and averaged over all parameters for series, at T=2048. In mixed fractional processes with small short memory coefficients, DFAbridge and ExphSpec deliver the smallest amount of false decisions, just as in the fractional noise and moving average case with small values of θ. However, as the autoregressive parameter increases while at the same time the moving average parameter value remains small, the identification pattern of the mixed process resembles that of the purely fractional autoregressive process with high amounts of misclassification for stationary series. As in the pure autoregressive case, Sperio and GPH deliver (almost) best results in the mixed case, however, they are surpassed by DAFjw8 in ARFIMA(1,d,1) series with ϕ=0.8 and θ=0.3.

Again, Sperio and GPH deliver best results in mixed series when the autoregressive coefficient is large. However, in the mixed case DAFjw8 is performing exceptionally well and even surpasses GPH.

Table 5.2.9. Percentage of misclassifications for ARFIMA(1,*d*,1) series with ϕ=0.2, 0.8, and θ=0.3, at *d*=0.1-0.9 and averaged over all parameters, at *T*=2048, received from 1000 replications. The data is assorted in ascending order by % of misclassification averaged over all parameters.

% of misclassifications for ARFIMA(1,*d*,1) series											
		d_{true}									
Method	ϕ,θ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Averaged over all parameters
DFAbridge	ϕ = 0.2, θ = 0.3	0	0	0	1.3	17.8	0	0	0	0	2.1
ExphSpecsm		0	0	0	5.2	30.7	1.3	0	0	0	4.1
^{low}PSD		0	0	0	4.9	47.7	6.4	0.2	0	0	6.6
^{low}PSD$_{ew}$		0	0	0.5	10.3	42.3	6.2	0.1	0	0	6.6
^{low}PSD$_{e}$		0	0.1	1.1	9	43.1	6.2	0.2	0	0	6.6
^{low}PSD$_{we}$		0	0.4	1.9	14.9	38.4	5.6	0	0	0	6.8
^{low}PSD$_{w}$		0	0	0.1	6.2	51.8	8.2	0.2	0	0	7.4
Sperio		0	0	0.7	8.9	53.3	15.2	2.5	0	0	9
GPH		0	0.1	2.9	19.4	42.9	15.4	2.9	0.2	0	9.3
ExphSpec		0	0	0	0	90.9	2.7	0	0	0	10.4
ExphSpecwo		0	0	0	0	93.6	1.2	0	0	0	10.5
ExphSpecmu		0	0	0	0	95.5	1.3	0	0	0	10.8
DFAjwL2		0	0	0	0	94.1	5.9	0	0	0	11.1
Whittle		0	0	0	0	100	3.3	0	0	0	11.5
DFAjwL6		0	0	0	0	93	25.1	1.9	0	0	13.1
DFAjw8		0	0	0	2	79.7	35.3	13.7	1.4	1	14.8
DFAjw2		0	0	0	2.1	80.1	34.8	14.6	1.3	1.1	14.9
PSD$_{we}$		0	0	0	0	100	86.4	0.4	0	0	20.8
PSD$_{ew}$		0	0	0	0	100	86.9	0.4	0	0	20.8
PSD$_{w}$		0	0	0	0	100	87.3	0.3	0	0	20.8
PSD$_{e}$		0	0	0	0	100	89.5	0.3	0	0	21.1

DFAjw8	$\phi = 0.8, \theta = 0.3$	0	0	3.8	25.8	24.2	13	2	1	0	7.8
Sperio		0	0	2.3	22.9	35.7	7.8	1.2	0	0	7.8
GPH		0	1.2	7.7	33.8	29.9	7.1	1.5	0	0.1	9
DFAjw2		0	0	4	26.1	29.9	20.2	2.8	1	0	9.3
$^{low}PSD_w$		0.2	4.3	45.7	91.1	0.2	0	0	0	0	15.7
^{low}PSD		0.1	3.1	47.4	92.8	0.3	0	0	0	0	16.0
DFAjwL6		0	8.1	52	97.9	0	0	0	0	0	17.6
$^{low}PSD_{ew}$		0.5	10.6	58.5	93.8	0.6	0	0	0	0	18.2
$^{low}PSD_e$		0.4	10	59.8	93.8	0.1	0	0	0	0	18.2
$^{low}PSD_{we}$		1.9	16.7	62.6	93.7	0.1	0	0	0	0	19.4
DFAjwL2		0	16.2	91.8	100	0	0	0	0	0	23.1
ExphSpecsm		0.4	36.1	91.6	99.6	0	0	0	0	0	25.3
DFAbridge		58	99.8	100	100	0	0	0	0	0	39.8
PSD_w		68.7	100	100	100	0	0	0	0	0	41.0
PSD_{ew}		80.1	100	100	100	0	0	0	0	0	42.2
PSD_{we}		81.4	100	100	100	0	0	0	0	0	42.4
PSD_e		81.5	100	100	100	0	0	0	0	0	42.4
ExphSpec		94.6	100	100	100	0	0	0	0	0	43.8
ExphSpecmu		96.4	100	100	100	0	0	0	0	0	44
ExphSpecwo		96.1	100	100	100	0	0	0	0	0	44
Whittle		100	100	100	100	0	0	0	0	0	44.4

Although the underlying structure of a time series may be merely hypothesized, the sample size is always known. Hence, the probability of a given time series length of being correctly classified as stationary or nonstationary is analyzed by computing the percentage of misclassifications for each sample size, averaged over model type and parameterization.

The portions of misclassification for the different models, independent of parameterization, for ARFIMA(0,*d*,0), ARFIMA(0,*d*,1), ARFIMA(1,*d*,0), and ARFIMA(1,*d*,1) series, and the different sample sizes is given in Table 5.2.10. Also listed are the portions of misclassification averaged over model type and parameterization, and the total portion of misclassification averaged over model type, parameterization, and sample size. The data is sorted in ascending order by the model, parameterization, and sample size averaged portion of misclassification.

The tabulated data may give the researcher the opportunity to choose the procedure performing best depending on the information at hand. If only sample size is known, Sperio and GPH deliver the smallest portion of misclassification, closely followed by ExphSpecsm. However, if the model type has been already determined, a more precise information may be obtained by looking up the best results in the appropriate column. For example, the chances of misclassification of a fractional noise process of sample size T=1024 is less than 4% if the fractional differencing parameter d is estimated by DFAbridge, compared to almost 11% if estimated by Sperio, assuming that the underlying structure of the series to be analyzed corresponds indeed to an ARFIMA(0,*d*,0) process.

In summary, short ARFIMA(0,*d*,0) series are best classified as stationary or nonstationary by Whittle and the ExphSpec methods, whereas DFAbridge performs better in long series with $T \geq 1024$. Regardless of length, the number of false decisions in fractional series with additional short-term memory parameters are least for Sperio and GPH.

Table 5.2.10. Percentage of misclassifications for the parameter averaged ARFIMA(0,*d*,0), ARFIMA(0,*d*,1), ARFIMA(1,*d*,0), and ARFIMA(1,*d*,1) series at *T*=128, 256, 512, 1024, and 2048, received from 1000 replications. Also listed are the portions of misclassification averaged over model type and parameterization, and the total portion of misclassification averaged over model type, parameterization, and sample size. The data is sorted in ascending order by the model, parameter and sample size averaged portion of misclassification.

% of misclassifications							
Method	T	(0,*d*,0)	(0,*d*,1)	(1,*d*,0)	(1,*d*,1)	Model and parameter averaged	Model, parameter and sample size averaged
Sperio	128	20	28.8	21.2	21.2	22.8	14.9
	256	16.1	22.3	16.2	16.7	17.8	
	512	13.1	16.5	12.9	12.9	13.9	
	1024	10.7	12.8	10.1	10	10.9	
	2048	8.6	10	8.3	8.6	8.9	
GPH	128	21.3	26.8	24.9	24.2	24.3	16
	256	17.2	20.7	19.5	19.5	19.2	
	512	13.7	16.5	15.3	15.1	15.2	
	1024	11	12.5	11.7	11.5	11.7	
	2048	9.5	10.1	9.3	9.2	9.5	
ExphSpecsm	128	15.3	23.5	30.4	25.9	23.8	16.6
	256	10.4	22.8	25.2	20.7	19.8	
	512	6.3	20.1	20.5	16.8	15.9	
	1024	4.8	16.6	17.1	13.6	13.0	
	2048	3.5	14.4	13.4	10.6	10.5	
DFAjw8	128	16	41.8	22.1	21.5	25.4	18.9
	256	15.6	36	17.8	17.1	21.6	
	512	15.3	29.8	13.6	14.5	18.3	
	1024	14	24.9	10.8	12.3	15.5	

	2048	13.3	19.6	10.5	11.3	13.7	
DFAjw2	128	17.3	39.9	21.5	20.6	24.8	19
	256	16.9	34.8	17	17.6	21.6	
	512	16.7	28	13.5	15.2	18.4	
	1024	14.7	23.8	11.7	13.3	15.9	
	2048	13.8	18.6	11.8	12.1	14.1	
DFAjwL2	128	11.6	41.2	27.8	23.3	26.0	19.8
	256	8.7	37.8	23.6	20.3	22.6	
	512	8.9	33.9	19.4	17.4	19.9	
	1024	6.7	30.5	16.7	13.1	16.8	
	2048	7.1	25.6	12.1	10.8	13.9	
^{low}PSD	128	24.8	27.6	26.4	26.2	26.3	19.9
	256	17.7	21.2	19.5	19.4	19.4	
	512	11.8	17.0	14.1	13.8	14.2	
	1024	8.4	14.5	10.5	10.4	11.0	
	2048	6.1	13.3	8.3	8.2	9.0	
ExphSpecwo	128	10.6	31.9	28.4	22.8	23.4	20.8
	256	7.1	31.1	26.7	20.6	21.4	
	512	4.7	31.5	25	19.7	20.2	
	1024	4.3	32.2	23.5	18.8	19.7	
	2048	4.7	32.5	22.2	18.5	19.5	
ExphSpecmu	128	10	32	28.7	22.9	23.4	20.9
	256	6.8	31.8	26.8	20.6	21.5	
	512	4.8	32.2	24.9	19.6	20.4	
	1024	4.4	32.6	23.4	18.9	19.8	
	2048	4.9	32.8	22.2	18.5	19.6	
DFAbridge	128	12.5	32.3	35.7	26.9	26.9	21.1

	256	8.1	29.4	32.7	23.9	23.5	
	512	6	26.7	29.7	21.3	20.9	
	1024	3.8	24.1	26.3	18.9	18.3	
	2048	1.9	21.5	23.5	16.6	15.9	
ExphSpec	128	11.9	30.7	30.5	24.6	24.4	
	256	7.7	30.4	27.6	21.5	21.8	
	512	5.4	31.1	25.8	19.9	20.6	21.1
	1024	4.1	31.5	23.9	19	19.6	
	2048	4.2	31.9	22.6	18.6	19.3	
$^{low}PSD_w$	128	26.8	29.7	27.6	27.3	27.9	
	256	19.8	23.1	20.6	20.8	21.1	
	512	13.1	18.6	15.0	14.6	15.3	21.4
	1024	9.5	15.6	11.1	11.0	11.8	
	2048	6.7	13.9	8.6	8.7	9.5	
DFAjwL6	128	14.4	43.7	28.7	23.9	27.7	
	256	12.2	42.6	28.9	22.8	26.6	
	512	12.4	36.2	18.6	16.1	20.8	21.7
	1024	7.1	33.8	19.6	16.1	19.2	
	2048	8.9	27.4	10.2	10.8	14.3	
Whittle	128	9.2	42.2	31.6	22.1	26.3	
	256	7.2	40.5	31.8	21.6	25.3	
	512	5.9	39.2	32	21.2	24.6	24.9
	1024	5.5	38.5	32.2	20.6	24.2	
	2048	5.2	38.2	32.6	20	24.0	
$^{low}PSD_{ew}$	128	33.6	42.6	31.7	33.2	35.3	
	256	25.6	33.8	24.5	25.5	27.4	
	512	15.1	23.3	17.6	17.8	18.5	25.7

	1024	9.6	15.5	12.4	12.4	12.5	
	2048	6.3	12.4	9.3	9.2	9.3	
$^{low}PSD_e$	128	34.8	43.4	32.0	34.0	36.1	26.8
	256	26.5	36.2	24.9	25.9	28.4	
	512	17.0	27.6	17.8	18.6	20.2	
	1024	10.1	18.2	12.4	12.6	13.3	
	2048	6.2	12.8	9.3	9.1	9.3	
$^{low}PSD_{we}$	128	35.6	43.7	33.5	35.1	37.0	28.6
	256	28.7	37.5	26.8	28.0	30.3	
	512	19.2	29.8	19.9	20.9	22.4	
	1024	11.3	20.3	13.7	14.0	14.8	
	2048	6.4	12.9	10.0	9.8	9.8	
PSD_w	128	13.9	46.3	30.3	22.6	28.3	33.7
	256	12.3	46.2	29.0	21.4	27.2	
	512	11.9	46.1	28.3	20.8	26.8	
	1024	11.9	45.7	27.8	20.3	26.4	
	2048	11.4	45.6	27.6	20.1	26.2	
PSD_{ew}	128	12.8	45.7	31.9	23.5	28.5	33.9
	256	11.4	45.9	30.3	22.0	27.4	
	512	11.1	45.8	29.4	21.3	26.9	
	1024	11.5	45.6	28.6	20.7	26.6	
	2048	11.2	45.5	28.2	20.5	26.4	
PSD_{we}	128	12.4	45.3	32.5	24.0	28.5	34
	256	10.9	45.7	30.8	22.4	27.5	
	512	10.9	45.7	29.7	21.4	26.9	
	1024	11.3	45.5	28.9	20.8	26.6	
	2048	11.1	45.5	28.4	20.6	26.4	

PSD_e	128	13.0	46.6	31.8	23.6	28.7	34.1
	256	11.6	46.3	30.3	22.1	27.6	
	512	11.6	46.0	29.3	21.3	27.1	
	1024	11.3	45.7	28.6	20.9	26.6	
	2048	11.3	45.6	28.2	20.6	26.4	

Independent of parameterization, model type and sample size, GPH and Sperio have shown to be the most reliable methods in distinguishing (non-)stationary fractional processes with or without additional long-memory parameters. Therefore their portions of misclassification for all levels of d as well as autoregressive and moving average parameter are visualized in Figures 5.2.5, 5.2.6, and 5.2.7.

Overall, the largest amount of misclassification for both procedures is at the border of nonstationary. There, the number of false decisions increases along with a growing moving average coefficient, but decreases with a growing autoregressive influence, suggesting that reliable classification in the presence of additional short-term paramter(s) is hampered only if the short memory component is a moving average term. Furthermore, stationary series and series at the border of nonstationary with d=0.5 are most reliably identified by the Sperio method, whereas GPH performs better in nonstationary series.

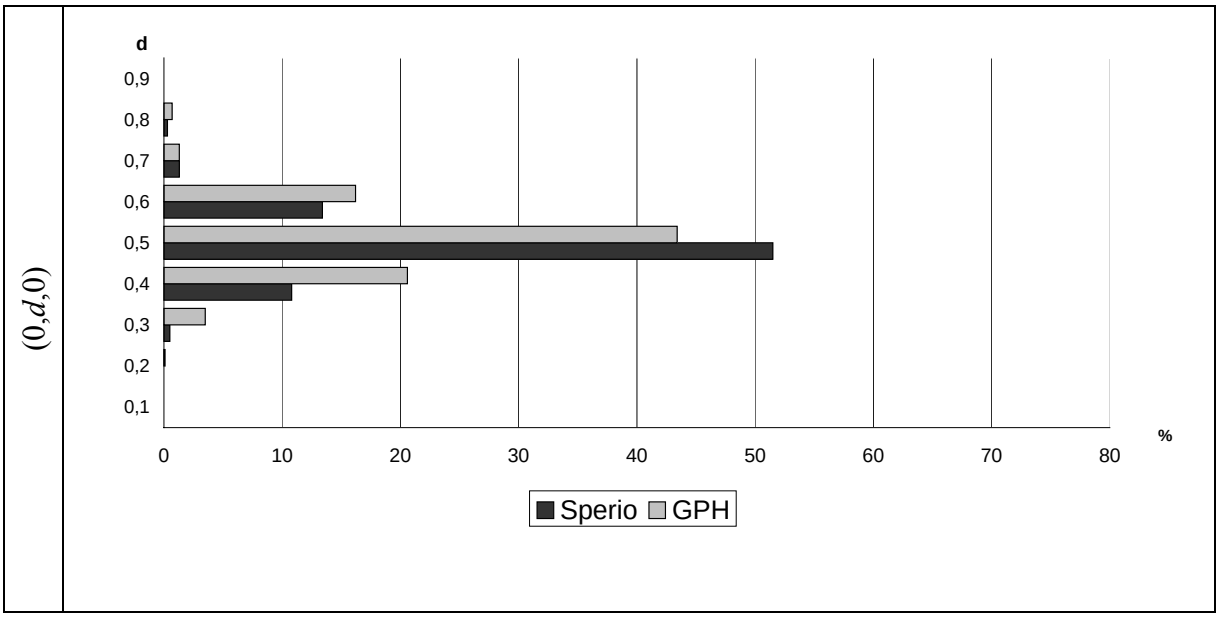


Figure 5.2.5. Percentage of misclassifications for Sperio and GPH of ARFIMA (0,*d*,0) series with *d*=0.1 to 0.9, at *T*=2048.

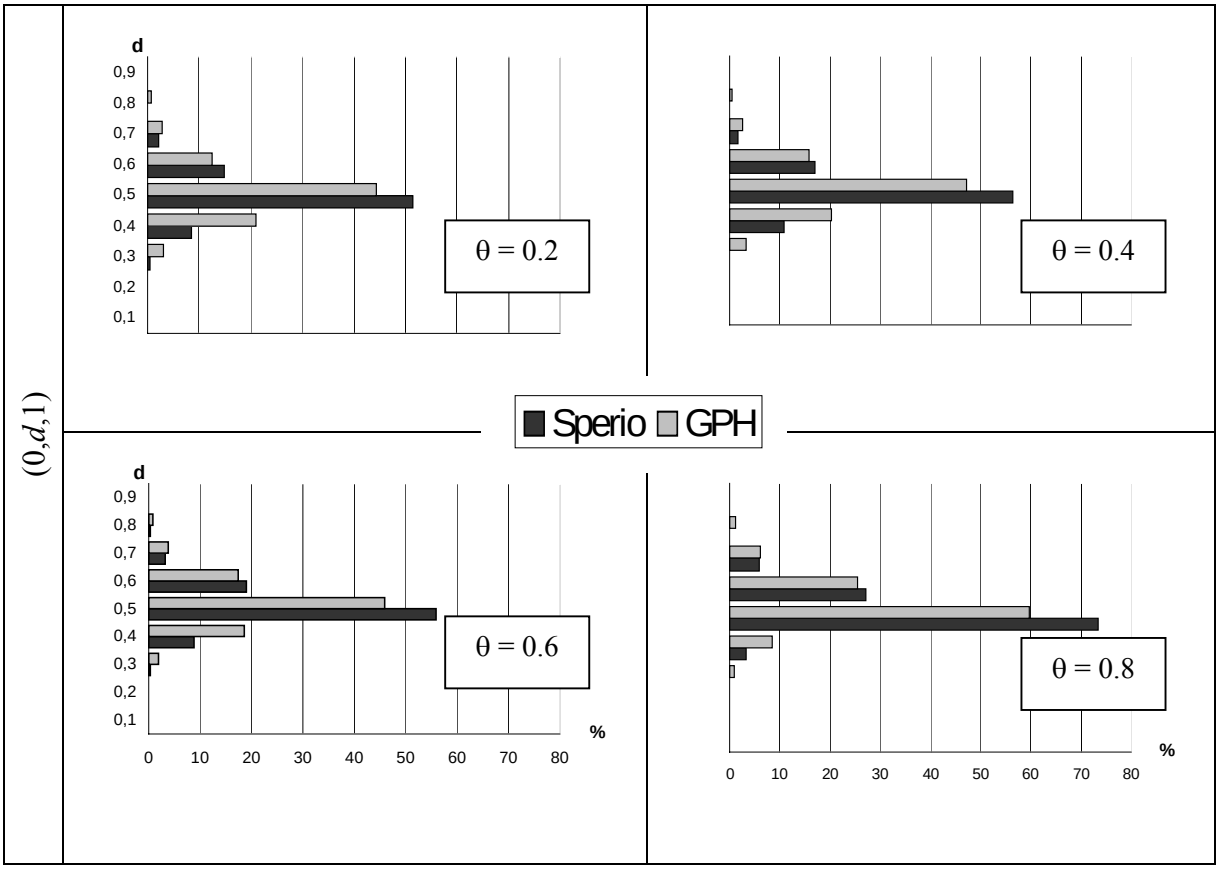


Figure 5.2.6. Percentage of misclassifications for Sperio and GPH of ARFIMA (0,*d*,1) series with *d*=0.1 to 0.9, at *T*=2048.

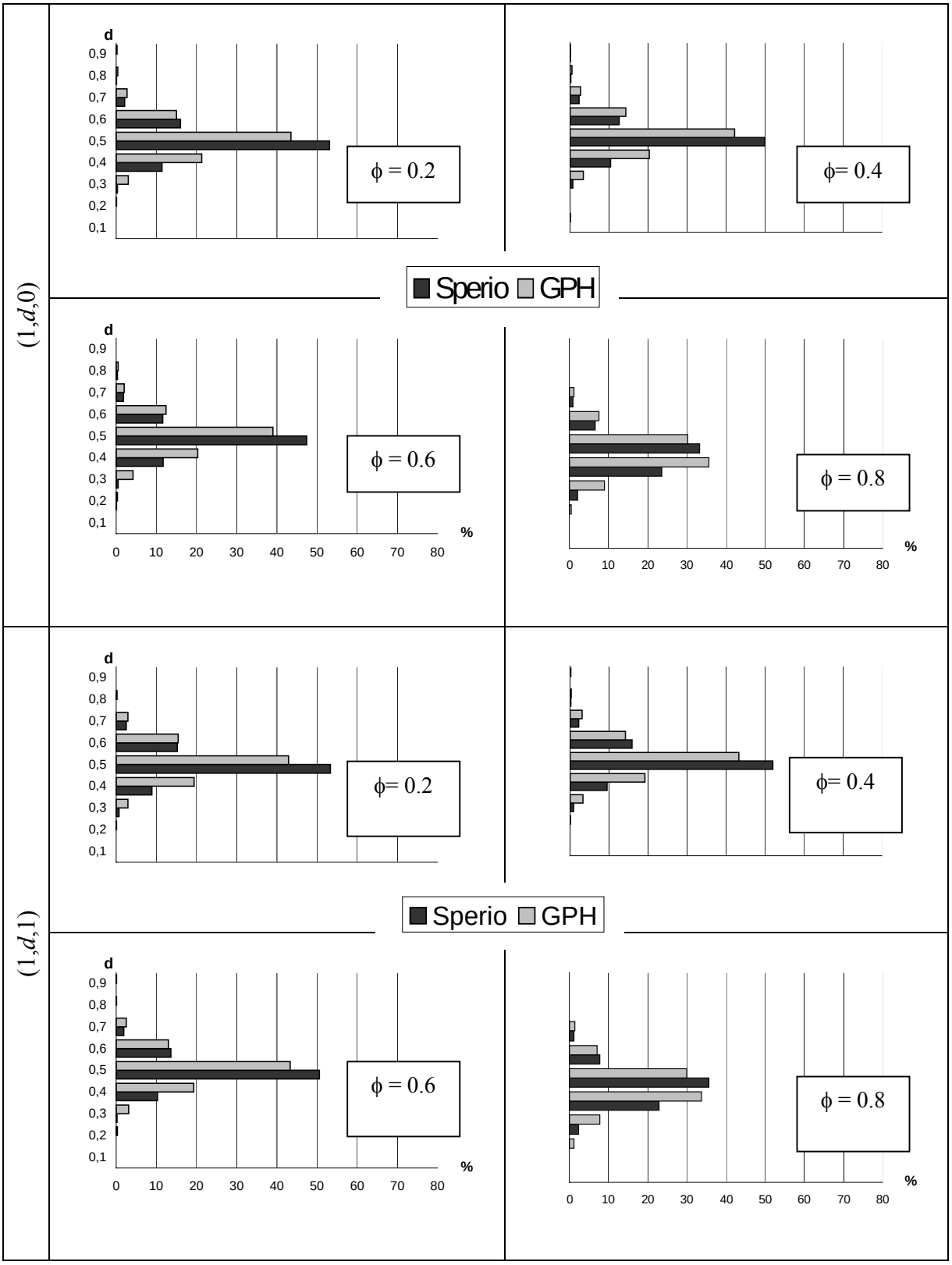


Figure 5.2.7. Percentage of misclassifications for Sperio and GPH of ARFIMA (1,*d*,0) and ARFIMA (1,*d*,1) series with *d*=0.1 to 0.9, at *T*=2048.

After having ascertained Sperio and GPH as the most reliable method of all procedures under evaluation for distinguishing ARFIMA series as stationary or nonstationary by means of the fractional differencing parameter *d*, it may be interesting to know what to expect in the case of a single analysis of an empirical series, that is to know the minimum and maximum values of the estimates. Hence, the extreme estimates of Sperio and GPH, i.e., their minimum and maximum estimates at d_{true}=0.5 and T=2048, are shown in Table 5.2.11.

Table 5.2.11. Minimum and maximum value of the fractional differencing parameter estimate $\hat{d}$ at d_{true}=0.5 and T=2048, received by Sperio and GPH .

Model type	ϕ,θ	Sperio		GPH	
		MIN	*MAX*	*MIN*	*MAX*
(0,*d*,0)		0.1821	0.7570	0.1986	0.9297
(0,*d*,1)	0.2	0.1325	0.7378	0.0408	0.8488
	0.4	0.1870	0.7465	0.1623	0.7817
	0.6	0.1107	0.7788	0.0117	0.8364
	0.8	0.1513	0.8714	0.0917	0.8090
(1,*d*,0)	0.2	0.1128	0.7875	0.0950	0.8259
	0.4	0.1606	0.7948	0.1430	0.9370
	0.6	0.2194	0.7801	0.1672	0.8283
	0.8	0.2621	0.8154	0.1428	0.8573
(1,*d*,1)	0.2	0.2013	0.7484	0.0944	0.8072
	0.4	0.1469	0.7249	0.1816	0.8503
	0.6	0.1811	0.7614	0.1623	0.8818
	0.8	0.2304	0.7705	0.1990	0.8891

Obviously, Sperio's range is smaller than that of the GPH method, at least in most cases, with an average range of 0.6 for Sperio and 0.72 for the GPH procedure, independent of model type and parameterization.

5.2.4 Conclusions

The performance of the 8 PSD, 4 ExphSpec and 5 DFA variations as well as GPH, Sperio and the Whittle method as diagnostic tools to distinguish between stationary and nonstationary ARFIMA processes has been empirically evaluated by means of Monte Carlo simulations.

Although DFAbridge performs best in correctly distinguishing stationary and nonstationary fractional noises, the estimator fails if there are additional short-memory parameters present.

Overall, 88% of all ARFIMA(0,*d*,0), 69% of all ARFIMA(0,*d*,1), 78% of all ARFIMA(1,*d*,0) and 81% of all ARFIMA(1,*d*,1) series were correctly identified, with an average of 79% of correct decisions independent of model type, parameterization, sample size, and procedure.

Over all model types and parameterizations, the Sperio method proposed by Reisen (1994) and the Geweke and Porter-Hudak (1983) algorithm have proven to be the most reliable classification tools of all procedures under evaluation, followed by the smoothed version of the HurstSpec method (ExphSpecsm).

Independently of model type, parameterization and sample size, Sperio and GPH manage to correctly distinguish stationary from nonstationary fractional ARFIMA series in around 90%, if $T \geq 1024$, with superior performance of Sperio in stationary (96%) and Sperio in nonstationary series (87%). Both methods (and likewise all other procedures under evaluation) are distinctively more successful if the short memory parameter present is

autoregressive and not moving average, suggesting that there is an effect of the underlying structure on the bias and not just model complexity on the number of false decisions. Small parameter values of *d* are associated with negative, large parameter values of *d* with positive biases (both advantageous for the correct identification of stationary and nonstationary series, respectively), except in series with large short memory parameter values. Here, the bias is either exclusively negative, as in the pure moving average model, or exclusively positive as in the pure autoregressive or in the mixed model with a large autoregressive and small moving average coefficient. Hence, nonstationary series dominated by large moving average parameters are more often falsely identified as stationary, whereas series dominated by autoregressive parameters are more frequently classified as nonstationary.

Overall, the Whittle and the PSD (including all frequencies) estimates of *d* show the least variability. Furthermore, $\hat{d}$ is least biased in fractional noise if estimated by Whittle and ^{low}PSD, whereas in the case of the presence of an additional moving average parameter GPH's estimates are least biased. If the fractional series, however, are dominated by a short-term autoregressive parameter, Sperio's estimates deliver the smallest bias of all procedures under evaluation.

To summarize, the Sperio or GPH procedure is recommended for a preliminary determination if an empirical process is stationary or nonstationary. However, if an empirical analysis suggests fractional noise without short memory, a reanalysis using DFAbridge is suggested. If confirmed, $\hat{d}$ may be most accurate if estimated by the Whittle method. However, if the empirical analysis suggests additional short memory components, GPH may as well be used for the estimation of *d* in the moving average and Sperio in the autoregressive case. Since only autoregressive dominated mixed models have been investigated in the current study, it can only be speculated that the superiority of GPH in moving average models will

uphold in mixed models dominated by large MA parameters, as it has been the case with Sperio in the autoregressive dominated mixed model.

5.3 Study 3: Distinguishing Short and Long Memory

5.3.1 Introduction

Since some statistical properties of time series with short-term dependence components may mimick those of long memory processes (Rangarajan and Ding, 2000; Thornton and Gilden, 2005; Wagenmakers et al., 2004, 2005), several techniques for specification or classification of processes with different memory properties have been proposed. Wagenmakers et al. (2004, 2005) developed a method in which the ARMA model, describing short-term memory processes, is competitively tested against the ARFIMA model, representing long memory processes. The authors suggested determining the maximum likelihood of a time series under the ARMA and ARFIMA models, respectively, and then selecting the appropriate representation using the Akaike's Information Criterion (AIC). Alternatively, a spectral classifier procedure was proposed by Thornton and Gilden (2005), in which the likelihood of a time series is estimated by comparing its power spectrum with the spectra of the competing memory model. In simulation experiments conducted by Farrell et al. (2006a, b), the spectral classifier method of Thornton and Gilden (2005) was compared with the ARFIMA approach of Wagenmakers et al. (2004). Both procedures proved to be equally effective in discriminating between long and short memory series.

Farrell and colleagues advocated the ARFIMA method because it is widely available in statistical packages such as R (a freely available software increasingly used in the social and behavioral sciences), it is easily extendable to different sample sizes and more complex models, and the theoretical properties of the ARFIMA models are well known, whereas those of e.g., spectral classifiers, have yet to be explored.

An evaluation of the ARFIMA procedure by Torre et al. (2007a) found out that this method presented a bias favoring the detection of long-range dependence: pure ARMA series were falsely identified as ARFIMA with d different from 0. An important rate of false identifications of long memory models was observed, especially when autoregressive or moving average coefficients were low. Employing the Bayes Information Criterion (BIC) instead of the AIC provided better results for some parameterizations.

To sum up, the value of the fractional differencing parameter d of an ARFIMA process determines the memory property of the series. It possesses short memory for d=0 and finite long memory for $0<d<0.5$. Methods proposed for distinguishing series with different long-run developments are functions of the employed estimators, thus their performance is determined by the quality of estimation of the fractional differencing parameter.

Therefore, the aim of the following Monte Carlo study is to evaluate the performance of those estimators delivering the smallest bias and standard error in Chapter 5.2, i.e., the Whittle and Sperio estimators as well as the Geweke- Porter-Hudak (GPH) algorithm. In addition, the approximate ML algorithm of Haslett and Raftery fracdiff (for a description, see Chapter 3.4, p.37) will be evaluated. In contrast to Whittle, Sperio and GPH, this procedure additionally provides short-range dependency estimates of p and q. The range of d is confined to [0;0.5] considering only stationary persistent ARFIMA series.

5.3.2 Method

The reliability of the abovementioned methods is tested by simulated AR(FI)MA processes with with varying long- and short term parameters by employing the command *fracdiff.sim* of the R package *fracdiff* (for details, consult the R documentation at http://ftp5.gwdg.de/pub/misc/cran/). The R-code for generating AR(FI)MA processes is

attached. For this study, only stationary AR(FI)MA series with *d* variations from 0 to 0.4, each replicated a 1000 times, are used. Manipulated are the following **independent variables**:

- value of *d*: 0, 0.1, 0.2, 0.3, 0.4;
- model: AR(FI)MA(1,*d*,0) with autoregressive parameters ranging from +/–0.2 to +/–0.8 by step of 0.2; ARFIMA(0,*d*,1) with moving average coefficients from +/–0.2 to +/–0.8 by step of -0.2; ARFIMA(1,*d*,1) with the same autoregressive parameterization as in the (1,*d*,0) case combined with the moving average value θ=-0.3;
- length of series *T*= 128, 256, 512, 1024, 2048;
- 4 procedures estimating $\hat{d}$: Whittle, Sperio, GPH, fracdiff;

As a quality criterion the following **dependent variables** were computed:

- percentage of deviations from the true parameter ($\%\left|\hat{d} - d\right| > 0.1$),
- Mean (*M*), standard error (*SE*), mean square error (*MSE*), minimum (*MIN*) and maximum (*MAX*) values;

5.3.3 Results

In the following chapter the accuracy of the estimation methods under investigation is studied first by assessing the percentage of estimates deviating more then +/-0.1 from the corresponding true parameter. Special consideration will be given to the zone of uncertainty [0;0.1]. The behavior of the procedures delivering the smallest portion of deviations, independent of parameterization, and independent of parameterization and sample size will be computed. Finally, Mean (*M)*, standard error (*SE*), mean square error (*MSE*), minimum (*MIN*), and maximum (*MAX*) values of the most accurate estimates, i.e., most reliable method for distinguishing short and long memory structures in ARFIMA series will be tabulated.

Deviations from the true parameter

In general, all deviations from the true parameter, regardless of procedure, become larger with increasing short memory parameter value and decreasing sample size. Figure 5.3.1 and 5.3.2 show the percentage of estimates for each method deviating more than +/-0.1 from the corresponding true parameter, at T=128, 265, 512, 1024 and 2048 for ARFIMA(0,d,1), ARFIMA(1,d,0), and ARFIMA(1,d,1) series with d=0 and d=0.1.

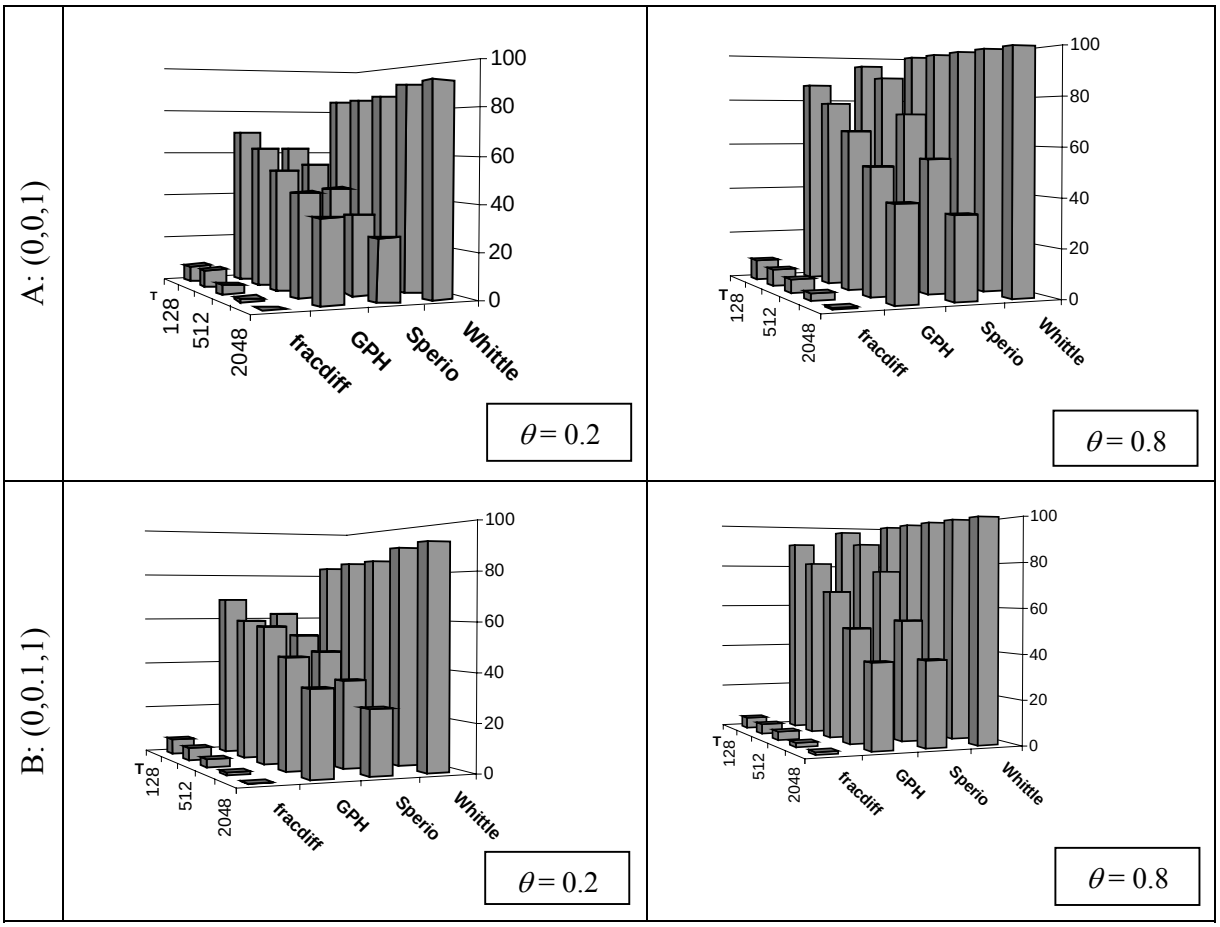


Figure 5.3.1. Percentage of deviations ($\%|\hat{d}-d|>0.1$) for fracdiff, GPH, Sperio, and Whittle estimates in ARFIMA(0,d,1) series with d=0 and d=0.1.

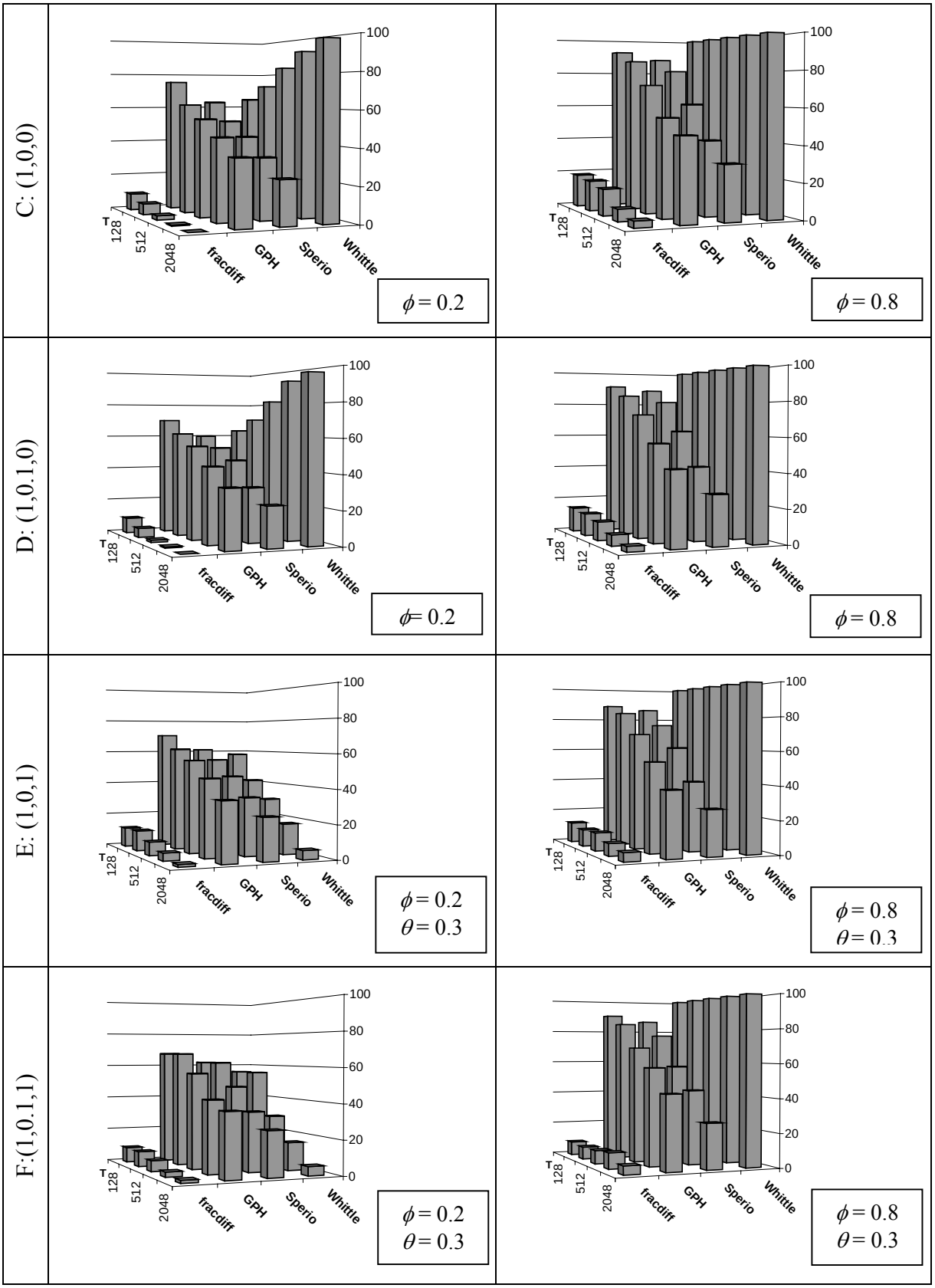


Figure 5.3.2. Percentage of deviations ($\%|\hat{d} - d| > 0.1$) for fracdiff, GPH, Sperio, and Whittle estimates in ARFIMA(1,*d*,0) and ARFIMA(1,*d*,1) series with *d*=0 and *d*=0.1.

The panel clearly demonstrates the superiority of the fracdiff method regardless of model type and parameterization. In ARFIMA(1,0,0) and (1,0.1,0) series, both with ϕ=0.2 and T=2048, all fracdiff estimates $\hat{d}$ deviate less than +/-0.1 from 0 and 0.1, respectively. Overall, the performance of fracdiff is in sharp contrast to that of the Whittle method with deviations between 80 and 100% in most cases. In purely (fractional) autoregressive series, Whittle's portion of deviations even increases with growing sample size. However, Whittle outperforms Sperio and GPH in the mixed model with ϕ=0.2 and should therefore be investigated further.

Figures 5.3.3 and 5.3.4 show the standard error and bias of fracdiff, GPH, Sperio and Whittle estimates for mixed ARFIMA(1,0,1) series with ϕ=0.2, θ=0.3 and ϕ=0.8, θ=0.3, respectively. Obviously, the superiority of Whittle over Sperio and GPH in the mixed model with ϕ=0.2 and θ=0.3 is due to the large standard errors delivered by Sperio and GPH which cannot be compensated by smaller biases, whereas in the ϕ=0.8, θ=0.3 parameter combination the deviations of the Whittle estimates from the true parameter are too large.

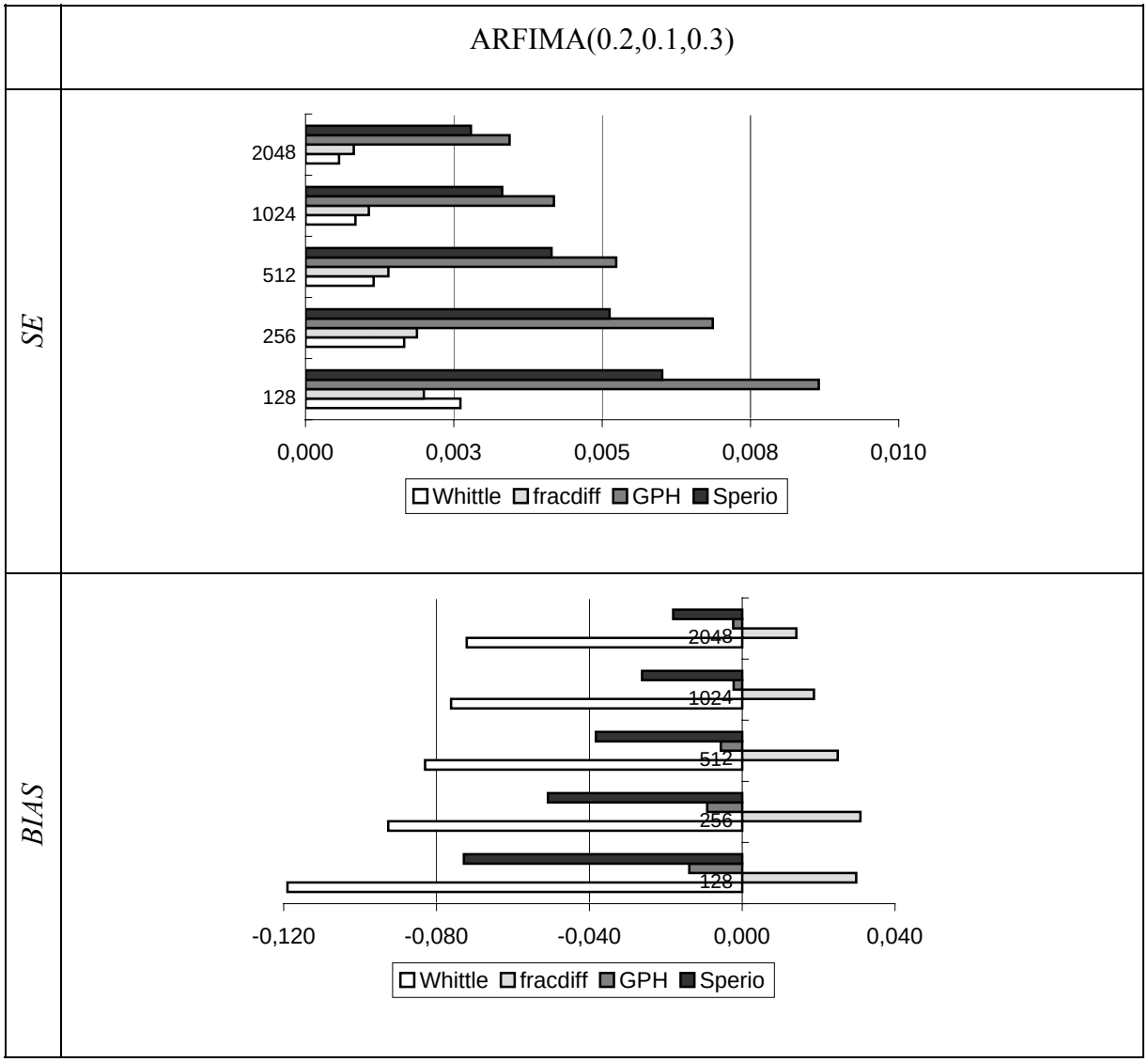


Figure 5.3.3. Standard deviation (upper panel) and bias (lower panel) for fracdiff, GPH, Sperio and Whittle estimates in ARIMA(1,0,1) series with ϕ=0.2 and θ=0.3.

Irrespectively of the magnitude of the short memory coefficient, the fracdiff estimates display the least variability and smallest bias and may therefore qualify as a reliable method for distinguishing between series with short and long memory, hence the behavior of fracdiff for the different values of *d* will be investigated more detailed.

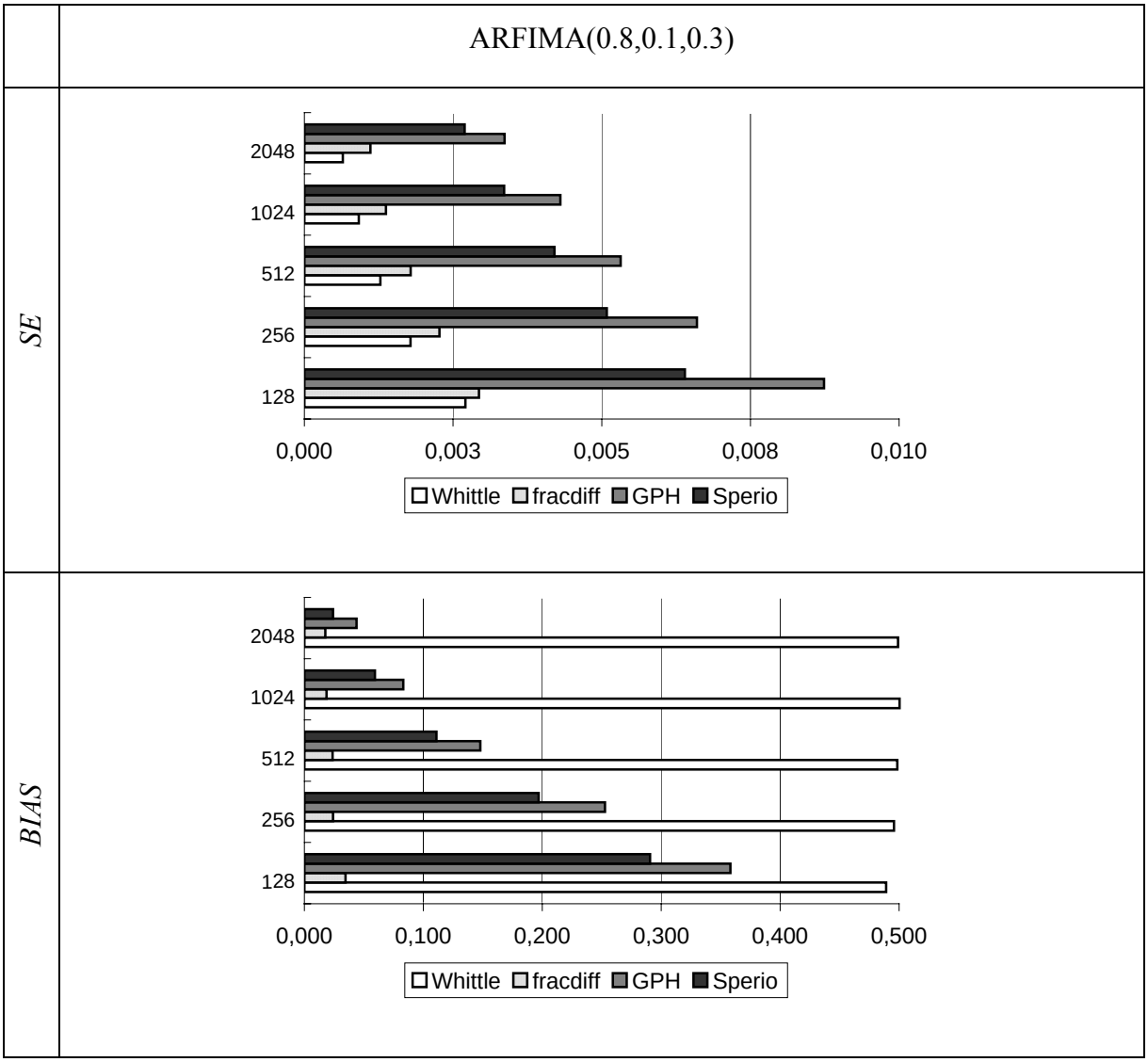


Figure 5.3.4. Standard deviation (upper panel) and bias (lower panel) for fracdiff, GPH, Sperio, and Whittle estimates in ARIMA(1,0,1) series with ϕ=0.8 and θ=0.3.

Figure 5.3.5 shows the bias ($Mean(\hat{d}) - d$) of fracdiff for ARIMA (0,*d*,1), (1,*d*,0) and (1,*d*,1) series for all values of *d* (0-0.4) and θ,ϕ (0.2, 0.4, 0.6, and 0.8). Apparently, fracdiff overestimates ARMA (*d*=0) and underestimates ARFIMA (*d*>0.1) series with an increasing negative bias as *d* gets closer to the border of nonstationary. There is a clear effect of parameter value only in pure moving average ARFIMA and autoregressive ARMA series. In pure moving average series, underestimation of the fractional differencing parameter *d* in ARFIMA series increases with growing moving average coefficient.

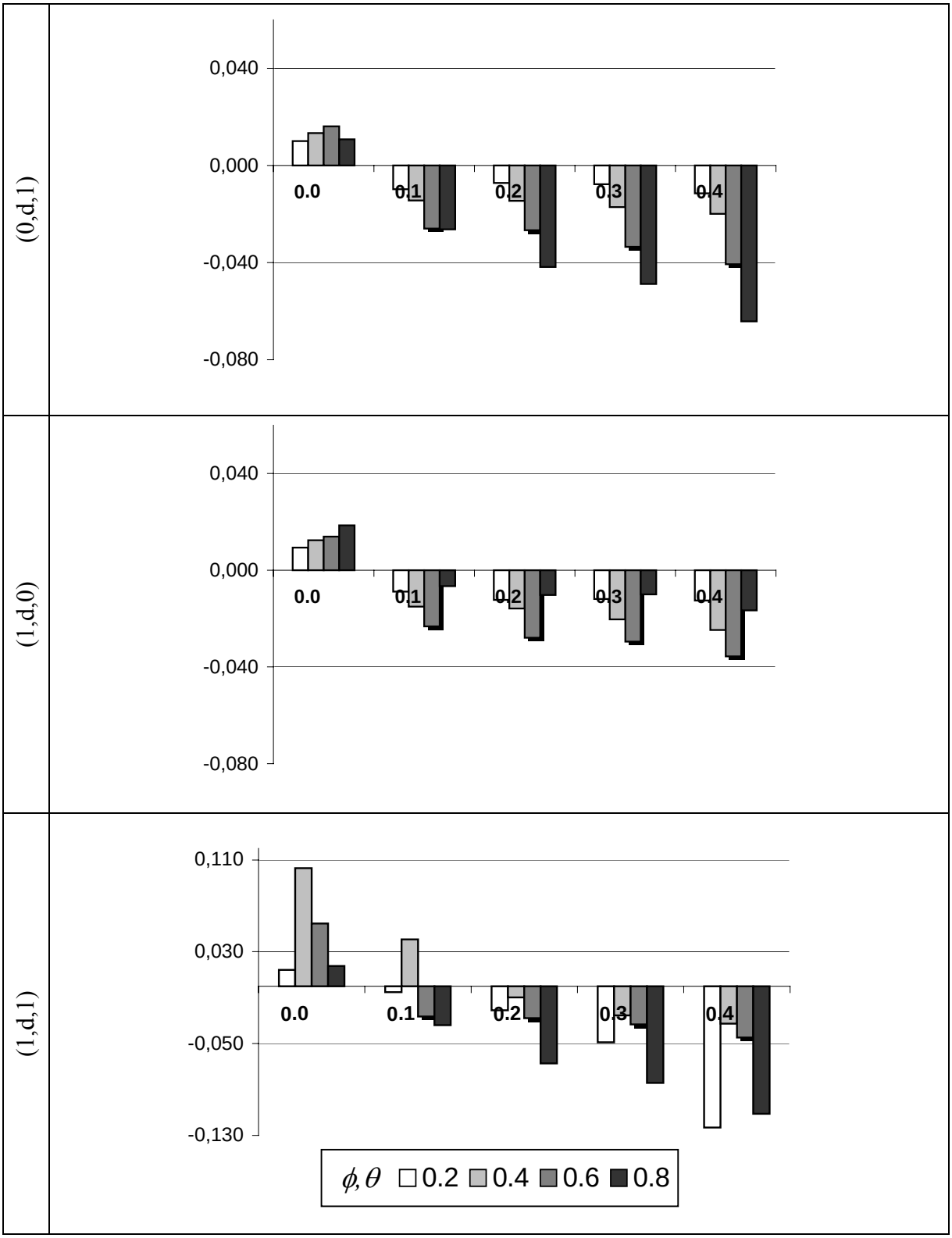


Figure 5.3.5. Bias ($Mean(\hat{d}) - d$) of fracdiff for ARIMA(0,*d*,1), (1,*d*,0), and (1,*d*,1) series with different values of *d* (0-0.4), and short memory parameter (0.2, 0.4, 0.6, and 0.8). Note, that in ARFIMA(1,d,1) series (lower panel), there is an additional θ=0.3 for all steps of ϕ.

In the pure autoregressive case, the overall magnitude of the bias is smaller, and more distinct only for series with medium sized autoregressive coefficients. In mixed ARFIMA series, only the fractional differencing parameter of series with extreme parameter values of ϕ are distinctly underestimated. Hence, the estimates of d in pure ARFIMA processes are least biased if the short memory parameter values are low (except in the autoregressive case with rather small biases for series with ϕ=0.8). The estimation of d in mixed series is least biased in ARMA series with extreme and ARFIMA series with medium sized autoregressive parameter values.

Tables 5.3.1-5.3.3 provide detailed results for fracdiff in ARFIMA(0,d,1), (1,d,0) and (1,d,1) series at T=2048. The biases shown in Figure 5.3.5 for fracdiff correspond to the percentage of estimates deviating more then +/-0.1 from the corresponding true paramete. For example, the bias in pure moving average series is most profound for large values of d and θ, and so is the percentage of deviation (18% at d=0.2, θ=0.8; 22% at d=0.3, θ=0.8; 27% at d=0.4, θ=0.8). Large portions of the deviation $\%\left|\hat{d} - d\right| > 0.1$ correspond to large mean square errors in all model types.

Table 5.3.1. Results for fracdiff of estimating the fractional differencing parameter d in ARFIMA(0,d,1) series, at T=2048, 1000 replications each.

ARFIMA (0,d,1)							
d	θ	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$Min_{\hat{d}}$	$Max_{\hat{d}}$	$\%\lvert\hat{d}-d\rvert > 0.1$
0	0.2	0.0100	0.0005	0.0004	0.0000	0.1161	0.1
	0.4	0.0134	0.0008	0.0009	0.0000	0.4700	1.1
	0.6	0.0161	0.0011	0.0014	0.0000	0.3242	2.8
	0.8	0.0108	0.0007	0.0006	0.0000	0.1199	0.6
0.1	0.2	0.0903	0.0011	0.0013	0.0000	0.2158	0.1
	0.4	0.0856	0.0014	0.0021	0.0000	0.2275	0.4
	0.6	0.0740	0.0018	0.0038	0.0000	0.2845	2.7
	0.8	0.0737	0.0017	0.0035	0.0000	0.2354	1.2
0.2	0.2	0.1928	0.0011	0.0012	0.0921	0.2945	0.4
	0.4	0.1855	0.0013	0.0020	0.0580	0.3087	2.6
	0.6	0.1734	0.0018	0.0040	0.0080	0.3633	11.9
	0.8	0.1582	0.0021	0.0061	0.0000	0.3314	18.3
0.3	0.2	0.2923	0.0011	0.0012	0.1718	0.4038	0.5
	0.4	0.2829	0.0013	0.0021	0.1516	0.4043	2.9
	0.6	0.2665	0.0018	0.0043	0.1064	0.4164	12.4
	0.8	0.2512	0.0022	0.0072	0.0000	0.4369	21.5
0.4	0.2	0.3886	0.0010	0.0011	0.2905	0.4682	0.2
	0.4	0.3801	0.0012	0.0018	0.2520	0.4799	1.9
	0.6	0.3594	0.0016	0.0043	0.1963	0.4888	13.2
	0.8	0.3358	0.0024	0.0097	0.0280	0.4833	27

Table 5.3.2. Results for fracdiff of estimating the fractional differencing parameter d in ARFIMA(1,d,0) series, at T=2048, 1000 replications each.

ARFIMA(1,d,0)							
d	ϕ	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$Min_{\hat{d}}$	$Max_{\hat{d}}$	$\%\lvert\hat{d}-d\rvert > 0.1$
0	0.2	0.0094	0.0005	0.0004	0.0000	0.0957	0
	0.4	0.0124	0.0007	0.0006	0.0000	0.1228	0.1
	0.6	0.0139	0.0008	0.0008	0.0000	0.1564	1.3
	0.8	0.0185	0.0010	0.0014	0.0000	0.2182	3.7
0.1	0.2	0.0912	0.0011	0.0012	0.0000	0.1806	0
	0.4	0.0850	0.0014	0.0022	0.0000	0.2093	0.2
	0.4	0.0768	0.0018	0.0038	0.0000	0.2711	1.5
	0.8	0.0934	0.0017	0.0028	0.0000	0.3578	3.2
0.2	0.2	0.1877	0.0012	0.0015	0.0320	0.2962	0.9
	0.4	0.1842	0.0015	0.0026	0.0001	0.3412	4.9
	0.6	0.1721	0.0021	0.0051	0.0000	0.3556	15.7
	0.8	0.1898	0.0017	0.0029	0.0212	0.3964	5.6
0.3	0.2	0.2881	0.0012	0.0015	0.1436	0.3913	1.4
	0.4	0.2797	0.0016	0.0030	0.0000	0.4044	6.2
	0.6	0.2705	0.0020	0.0049	0.0224	0.4432	14.9
	0.8	0.2900	0.0015	0.0024	0.1500	0.4708	3.8
0.4	0.2	0.3875	0.0011	0.0014	0.1806	0.4727	1.1
	0.4	0.3753	0.0015	0.0029	0.0538	0.4760	5.9
	0.6	0.3644	0.0019	0.0049	0.1312	0.4839	12.8
	0.8	0.3834	0.0015	0.0026	0.2482	0.4996	3.3

Table 5.3.3. Results for fracdiff of estimating the fractional differencing parameter *d* in ARFIMA(1,*d*,1) series, at *T*=2048, 1000 replications each.

ARFIMA(1,*d*,1)							
d	ϕ, θ	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$Min_{\hat{d}}$	$Max_{\hat{d}}$	$\%\lvert\hat{d}-d\rvert > 0.1$
0	0.2. 0.3	0.0142	0.0008	0.0009	0.0000	0.2036	1.5
	0.4. 0.3	0.1028	0.0020	0.0147	0.0000	0.2668	58.4
	0.6. 0.3	0.0546	0.0035	0.0151	0.0000	0.3954	15.2
	0.8. 0.3	0.0176	0.0011	0.0015	0.0000	0.2009	5.3
0.1	0.2. 0.3	0.0948	0.0015	0.0021	0.0000	0.2814	1.7
	0.4. 0.3	0.1408	0.0028	0.0094	0.0000	0.3464	28.3
	0.6. 0.3	0.0736	0.0018	0.0039	0.0000	0.2769	1.1
	0.8. 0.3	0.0661	0.0023	0.0063	0.0000	0.2967	5.1
0.2	0.2. 0.3	0.1790	0.0018	0.0037	0.0000	0.3451	10.7
	0.4. 0.3	0.1902	0.0022	0.0049	0.0000	0.3917	16.6
	0.6. 0.3	0.1721	0.0021	0.0054	0.0000	0.3769	16.9
	0.8. 0.3	0.1328	0.0031	0.0143	0.0000	0.4191	41.9
0.3	0.2. 0.3	0.2512	0.0024	0.0082	0.0001	0.4300	25.9
	0.4. 0.3	0.2745	0.0019	0.0041	0.0463	0.4405	11.3
	0.6. 0.3	0.2667	0.0022	0.0058	0.0000	0.4171	16.5
	0.8. 0.3	0.2157	0.0037	0.0209	0.0000	0.4579	43.5
0.4	0.2. 0.3	0.2768	0.0026	0.0218	0.0552	0.4751	60
	0.4. 0.3	0.3673	0.0017	0.0039	0.1432	0.4826	10.6
	0.6. 0.3	0.3552	0.0022	0.0067	0.0000	0.4836	18.5
	0.8. 0.3	0.2888	0.0043	0.0307	0.0000	0.4984	44.8

Independent of parameterization, the average rate of deviations of the fracdiff estimates ($\%\left|\hat{d}-d\right|>0.1$), at T=2048, for pure autoregressive series, is 4%, for pure moving average series 6%, and for the mixed series 22%. Overall, the most accurate parameter estimation of d at T=2048 is obtained in pure autoregressive series with d=0 and 0.1, each at ϕ=0.2, with no deviation at all larger than +/-0.1. The poorest results are obtained for mixed ARFIMA series with ϕ=0.2, θ=0.3 at d=0.4, with only 40% of deviation smaller than +/-0.1.

5.3.4 Conclusions

The performance of the Whittle method, the semiparametric estimator of Reisen (Sperio), the Geweke and Porter-Hudak (GPH) and the approximate Maximum Likelihood Algorithm of Haslett and Raftery (1989) (fracdiff), all available in the R package *fracdiff*, as diagnostic tools estimating the fractional differencing parameter d to distinguish between short-and long-term dependency structures, has been empirically evaluated by means of Monte Carlo simulations. The parameter d of an ARFIMA process determines the memory property of the series. It possesses short memory for d=0 and finite long memory for $0<d<0.5$. Since methods proposed for distinguishing series with different long-run developments by means of $\hat{d}$, their performance is solely determined by the accuracy of estimation.

Overall, the results are highly depended on estimation techniques, parameterization and sample size. Only fracdiff has qualified as an accurate estimation technique for the classification of stationary ARMA and ARFIMA series due to exceptionally small biases. Although fracdiff succeeds in delivering estimates deviating less than 5% +/-0.1 from d_{true} in short ARFIMA series (θ=0.2, d=0.1) with only T=128 observations, pure autoregressive or moving average ARMA or ARFIMA series should consist of at least 500 observations. For distinguishing mixed AR(FI)MA series, however, processes should be at least 1000

observations long. Depending on parameterization, fracdiff's estimates deviate even less than 5% +/-0.1 from the true parameter in series with $T \geq 1024$.

However, when applying fracdiff for determining the specific long-run development of an empirical time series, the researcher should bear in mind that fracdiff overestimates ARMA ($d=0$) and underestimates ARFIMA ($d>0.1$) series with an increasing negative bias as d gets closer to the border of nonstationary, and that regardless of model type.

.

6 GENERAL DISCUSSION

The emphasis of this thesis is certainly on methodological issues. The primary research tasks has been the evaluation of the diagnostic ability of different methods within the ARFIMA and fractal analysis for revealing the nature of time series data, in particular, to distinguish (1) two different classes of persistent processes: fractional Brownian motions (fBm) and fractional Gaussian noises (fGn) by means of the Hurst coefficient *H* within the fractal analysis, (2) stationary and nonstationary ARFIMA (*p*,*d*,*q*) processes, and (3) processes with short and long memory by means of the parameter *d* within the autoregressive fractionally integrated moving average ARFIMA (*p*,*d*,*q*) framework. Chapter 4 demonstrated samples of the many possibilities of time series analysis techniques to deal with *dynamical* psychological phenomena containing internal temporal regularity that can be distinguished from unstable systems depending on external and occasional events. Even complex behavior such as the balancing between preservation and adaptation can be represented by means of rather simple and flexible time series models making processes with different memory properties distinguishable such as the autoregressive moving average fractionally integrated processes (ARFIMA) constituting as an intermediary between ARMA and ARIMA processes.

The application of fractal methods, however, often remains rudimentary: analyzes are limited to the use of a unique method, the collected series are sometimes too short for a valid assessment, and more generally the theoretical background of fractals and related methods is not fully exploited. Thus, recent theoretical and methodological refinements of fractal analyzes (see Eke et al., 2000, 2002) appear largely unknown in the psychological community, although it is by now well established, that, e.g., reaction time sequences in normal adults often show evidence of a long-term memory process known as 1/*f* noise (Gilden, 2001; Thornton & Gilden, 2005) in that part of the data generally regarded as

unexplained variance, i.e., the trial-to-trial residual variability. In typical cognitive tasks, the 1/f noise component may account for 30 to 40% of the variance when the treatment effects explain only about 10% (Gilden, 1997, 2001). As demonstrated in Chapter 4.2, p.44 , both healthy adults and adults diagnosed with ADHD produced data that displayed trends typical of mental rotation, although the two groups reaction time sequences had very different autocorrelation functions resembling 1/*f* noise for healthy adults, while the clinical group data showed substantial traces of random walk, a difference that could have been not detected by an ANOVA.

Utilizing time series analysis as diagnostic tools to distinguish between qualitatively different processes requires adequate analyzing techniques that have been tested for their reliability for distinguishing qualitatively different processes, as well as accurately estimating the corresponding parameters. Hence, this thesis aimed to test the diagnostic capability of different methods within the ARFIMA and fractal analysis, in particular different periodogram-based methods and their non-spectral alternatives to distinguish between stationary fGn and nonstationary fBm signals, stationary and nonstationary ARFIMA processes and stationary ARMA and ARFIMA series, i.e., processes with short and long memory within the ARFIMA framework.

Since the true data generating process of the series used to evaluate the diverse analyzing techniques has to be known, all methods under evaluation were tested by means of simulated data generated by the option lmSimulate of the R package fractal and the command fracdiff.sim of the R package fracdiff (all computations were performed with R version 2.7.2., for details, consult R documentation at http://ftp5.gwdg.de/pub/misc/cran/). The R code is also available at http://www.stat.osu.edu/~pfc/software/ (see also Craigmile, 2003).

In fractal analysis the adequate estimation of memory characteristics usually requires a preliminary classification of a time series as fractional Brownian motion (fBm) or fractional

Gaussian noise (fGn). Regardless of signal type and parameterization, the 'smoothed' version of the HurstSpec method (ExhpSpecsm) and the semiparametric estimator of Reisen (Sperio) correctly classify the true signal in more than 95% of all series with $T \geq 512$, suggesting both as adequate tools for a preliminary classifications of fractal series as Gaussian noise or Brownian motion in samples with about 500 observations (Figure 6.1 summarizes the results). If an empirical analysis suggests an fGn signal, a reanalysis with the non-spectral alternatives DFAjw8 and DFAjw2 is proposed, since both identify fGn signals in more than 99% of all trials, independent of sample size. If the preliminary empirical analysis suggests an fBm signal, a reanalysis with DFAbridge will surely confirm the results, since both ExphSpecsm and DFAbridge perform best in classifying fBm signals regardless of sample size with around 3% or less false decisions.

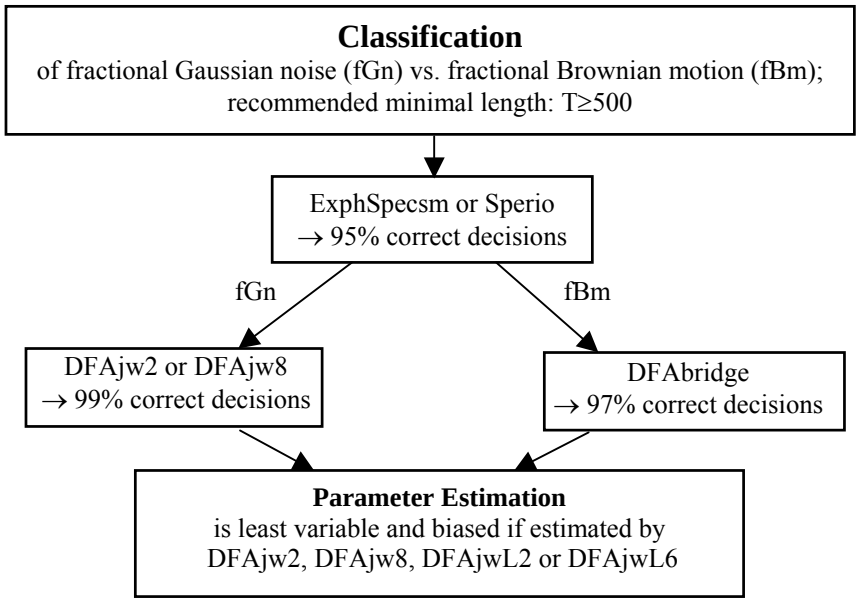


Figure 6.1. Recommendations for the reliable classification and accurate parameter estimation of fractional Gaussian noises and fractional Brownian motions.

Furthermore, all DFA estimates (except DFAbridge) have excelled in being least biased and are associated with the smallest standard errors, independently of signal type, parameterization and sample size (for detailed information see Tables 5.1.2 and 5.1.3 as guidelines, since estimation accuracy highly depends on signal type, parameterization and sample size).

The findings are in agreement with Delignières et al. (2006), who evaluated the ability of PSD, $^{low}PSD_{we}$, DFA, and Signal Summation Conversion (SSC) to distinguish fGn and fBm series, where both PSD versions were able to distinguish between fGn and fBm series with true *H* exponent ranging from 0.3 to 0.7. For low and high coefficients, however, the results were ambiguous. For instance, the original PSD clearly outperformed $^{low}PSD_{we}$ in short fBm series with H=0.2 and was significantly inferior in fGn series with H=0.9.

In their recent study Stroe-Kunold et al. (2009) compared the accuracy of different estimators of long memory parameters implemented in R for simple fGn cases, where the most precise method for *H* was hurstSpec, whereas Whittle was the best procedure for *d*. These findings are consistent with the results shown in Table 5.1.2.

Within the ARFIMA framework, the semiparametric estimator of Reisen (Sperio) and the Geweke and Porter-Hudak estimator (GPH) are superior in distinguishing stationary and nonstationary ARFIMA series. Both procedures are based on the regression equation, however, Sperio uses the smoothed periodogram function as an estimate of the spectral density. Sperio's and GPH's algorithms manage to correctly separate stationary from nonstationary fractional ARFIMA series with $T \geq 1024$ in around 90% of all trials with superior performance of GPH in nonstationary and Sperio in stationary series, regardless of model type and parameterization. The results are summarized in Figure 6.2.

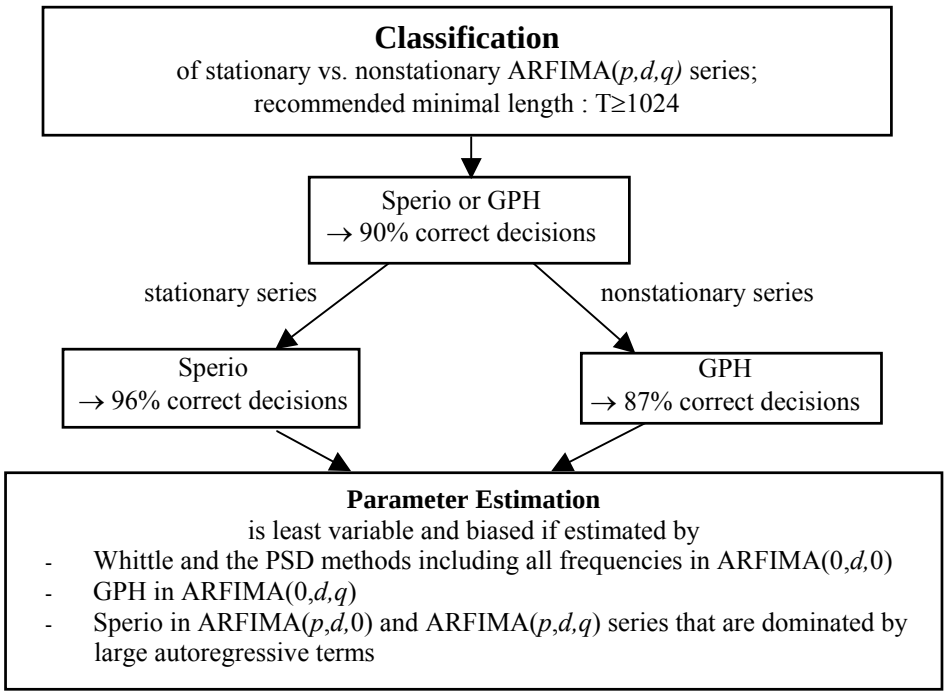


Figure 6.2. Recommendations for the reliable classification and accurate parameter estimation of stationary and nonstationary ARFIMA(*p,d,q*) series.

No clear effect of model complexity on the rate of misclassification has been observed, but autoregressive series are distinctively more often correctly classified than moving average processes. Furthermore, the accuracy of estimating the fractional differencing parameter *d* highly depends on parameterization and sample size. Overall, Whittle and the PSD estimates using only windowing and/or endmatching show the least variability. Additionally, the Whittle estimates of *d* are least biased in fractional noise, the GPH estimates are least biased in moving average series, and the Sperio estimates in fractional series with one additional autoregressive term or in the mixed model with large autoregressive and small moving average coefficients.

The Sperio, GPH and Whittle estimates of the fractional differencing parameter *d*, however, cannot be recommended for the classification of stationary AR(FI)MA processes

with short and long memory dependency structures. Only the approximate Maximum Likelihood Algorithm of Haslett and Raftery (1989) fracdiff, available in the R package *fracdiff*, qualifies as an accurate estimation technique for the classification of stationary ARMA and ARFIMA series due to exceptionally small biases (averaged over all model types, parameterizations and sample sizes). Regardless of model type, the procedure somewhat overestimates ARMA (d=0) and underestimates ARFIMA (d>0.1) series with an increasing negative bias as d gets closer to the border of nonstationary. For reliable classification, pure (fractional) autoregressive or moving average series should consist of at least 500, mixed series of at least 1000 observations. Note, that in contrast to Sperio, GPH and Whittle, the range of d for fracdiff is confined to $0 \leq d < 0.5$. Hence, the method well qualified for distinguishing long and short memory processes cannot be utilized for the classification of (non)stationary series. The results are summarized in Figure 6.3.

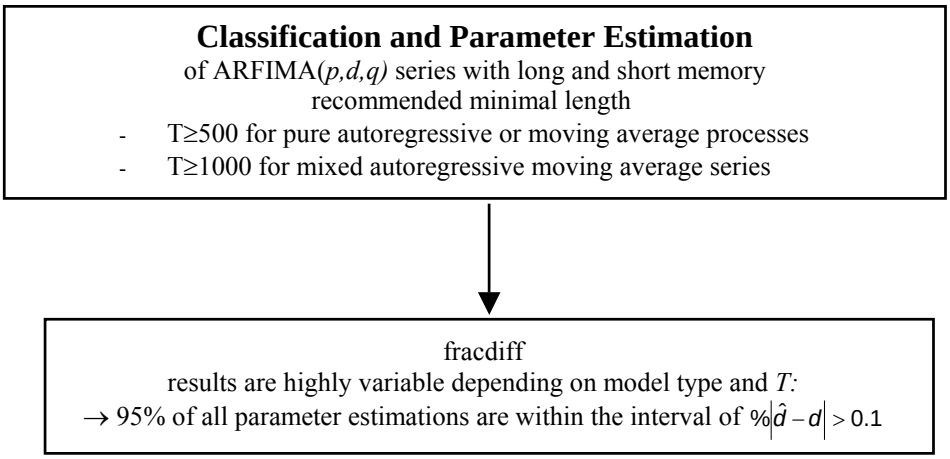


Figure 6.3. Recommendations for the reliable classification and accurate parameter estimation of short and long memory ARFIMA(p,d,q) series.

To summarize, despite the variability in the performance of the procedures under evaluation, i.e., depending on process type, parameterization and time series length, the findings reported in this paper may have distinguished specific algorithms superior in the identification of different process properties and estimation of the corresponding parameters. While some of the HurstSpec methods and Sperio are superior in distinguishing fractal signals, Sperio, GPH and Whittle are best in the classification of (non)stationary ARFIMA series but not in distinguishing stationary AR(FI)MA processes with different memory properties, for which only the approximate Maximum Likelihood Algorithm of Haslett and Raftery (1989) fracdiff qualified.

However, the author does not recommend, e.g., the Sperio method as the most reliable 'multi-purpose' tool, as the procedure has been successful in fractal analysis as well as in the ARFIMA framework. Due to the variability in the performance of the different methods depending on process type, parameterization and sample size, the detailed findings should be considered when parsimonously planning a study by looking into the relevant Tables of Chapter 5, where sample size requirements as well as the optimal analyzing technique for the respective task may be chosen from.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716-723.

Aks, D., J., & Sprott, J. C. (2003). The role of depth and 1/f dynamics in perceiving reversible figures. *Nonlinear Dynamics, Psychology, and Life Sciences, 7*, 161-180.

Alados, C. L.& Huffman, M.A. (2000). Fractal Long-Range Correlations in Behavioral Sequences of Wild Chimpanzees: a Non-Invasive Analytical Tool for the Evaluation of Health. *Ethology 106*, 105-116.

Ayat, L., & Burridge, P. (2000). Unit root tests in the presence of uncertainty about the non-stochastic trend. *Journal of Econometrics, 95*, 71-96.

Baillie, R. T. and King, M.L. (1996). Fractional differences and long memory processes. *Special Issue of Journal of Econometrics*, 73.

Beguin, J. M., Gourieroux, C., & Montfort, A. (1980). Identification of mixed autoregressive-moving average process: the corner method. In O. D Anderson (Eds.). *Time Series* (pp. 423-436). Amsterdam: North Holland.

Beltz, B. C. & Kello, C. T. (2006). Scale Free Network Structure in Orthographic and Phonological Word Forms. *47th Annual Meeting of the Psychonomic Society.* Houston, Texas.

Beran, J. (1994). *Statistics for long-memory processes.* New York: Chapman & Hall.

Beran, J., & Ocker, D. (1999). SEMIFAR forecasts, with applications to foreign exchange rates. *Journal of Statistical Planning & Inference, 80,* 137-153.

Beran, J., & Feng, Y. (2002). SEMIFAR models – a semiparametric approach to modelling trends, long-range dependence and nonstationarity. *Computational Statistics & Data Analysis, 40,* 393-419.

Bloomfield, P. (2000). *Fourier analysis of time series*. New York: Willey.

Bowerman, B. L. & O'Connell, R. T. (1993). *Forecasting and time series: An applied approach.* Belmont: Duxbury Press.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis, forecasting and control*. San Francisco, CA: Holden Day.

Box, G. E. P., & Pierce, W. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of American Statistical Association, 65*, 1509-1526.

Brent, R. (1973). *Algorithms for minimization without derivatives.* Prentice-Hall.

Brockwell, P. J., & Davis, R. A. (1991). *Time series: Theory and methods (2nd Ed.).* New York: Springer.

Caccia, D. C., Percival, D., Cannon, M. J., Raymond, G., & Bassingthwaighte, J. B.(1997). Analyzing exact fractal time series: evaluating dispersional analysis and rescaled range methods. *Physica A, 246*, 609-632.

Cannon, M. J., Percival, D., Caccia, D. C., Raymond, G., & Bassingthwaighte, J. B.(1997). Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series. *Physica A, 241*, 606-626.

Castellanos, F. X., and Tannock, R. (2002). Neuroscience of attentiondeficit/

hyperactivity disorder: The search for endophenotypes. *Nature Reviews Neuroscience, 3,* 617-628.

Castren, E. (2005). Is mood chemistry? *Nature Reviews Neuroscience, 6*, 241–246.

Chen, Y., Ding, M., & Kelso, J. A. S. (1997). Long memory processes ($1/f^{\alpha}$ type) in human coordination. *Physical Review Letters, 79,* 4501-4504.

Chen, Y., Ding, M., & Kelso, J. A. S. (2001). Origins of time errors in human sensorimotor coordination. *Journal of Motor Behavior, 33*, 3-8.

Choi, B. S. (1992). *ARMA Model Identification*. New York: Springer.

Chung, C. F. (1996). A generalized fractionally integrated ARMA process. *Journal of Time Series Analysis, 2,* 111-140.

Correll, J. (2008). 1/f noise and effort on implicit measures of bias. *Journal of Personality and Social Psychology*, *94(1)*, 48-59.

Craigmile, P. F. (2003). Simulating a class of stationary Gaussian processes using the Davies-Harte algorithm, with application to long memory processes. *Journal of Time Series Analysis, 24,* 505-511.

Crato, N. (1969). Some results on the spectral analysis of nonstationary time series. *Portugale Mathematica, 53,* 179-186.

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

Delignières, D., Fortes, N., & Ninot, G. (2004). The fractal dynamics of self-esteem and physical self. *Nonlinear Dynamics in Psychology and Life Sciences, 8*, 479-510.

Delignières, D., Lemoine, L., & Torre, K. (2004). Time intervals production in tapping oscillatory motion. *Human Movement Science, 23*, 87-103.

Delignières, D., Ramdani, S., Lemoine, L., & Torre, K. (2006). Fractal analyzes for short time series: a re-assessment of classical methods. *Journal of Mathematical Psychology, 50,* 525-544.

Delignières, D., Torre, K. & Lemoine, L. (2005). Methodological Issues of monofractal analyzes in psychological and behavioral research. *Nonlinear Dynamics in Psychology and Life Sciences, 9*, 435-462.

Delignières, D., Torre, K., & Lemoine, L. (2008). Fractal models for event-based and dynamical timers. *Acta Psychologica, 127,* p. 382-397.

Delignières, D., Torre, K. (in press). Fractal dynamics of human gait : a reassessment of Hausdorff et al. (1996)'s data. *Journal of Applied Physiology*.

Diagnostic and Statistical Manual of Mental Disorders (1994). Fourth Edition (DSM-IV). American Psychiatric Association: Washington D.C.

Ding, M., Chen, Y., & Kelso, J. A. S. (2002). Statistical analysis of timing errors. *Brain and Cognition, 48*, 98-106.

Dolado, J. J., Gonzalo, J., & Mayoral, L. (2002). A fractional Dickey-Fuller test for unit roots. Econometrica, 70, 1963-2006.

Eke, A., Herman, P., Bassingthwaighte, J. B., Raymond, G., Percival, D., Cannon, M. J., Balla, I., & Ikenyi, C. (2000). Physiological time series: distinguishing fractal noises from motions. *Pflügers Archives, 439*, 403-415.

Eke, A., Herman, P., Kocsis, L., & Kozak, L. R. (2002). Fractal characterization of complexity in temporal physiological signals. *Physiological Measurement, 23*, R1-R38.

Elder, J., & Kennedy, P. E. (2001). Testing for unit roots: what should students be taught? *Journal of Economic Education*, 32, 137-146.

Fang, Y. (2005). The effect of the estimation on Goodness-of-Fit tests in time series models. *Journal of Time Series Analysis, 26,* 527-541.

Farrell, S., Wagenmakers, E-J., & Ratcliff, R. (2006a). 1/f noise in human cognition: Is it ubiquitous, and what does it mean? *Psychonomic Bulletin & Review*, 13 (4), 737-41.

Farrell, S., Wagenmakers, E-J., & Ratcliff, R. (2006b). Methods for detecting 1/*f* noise. *Unpublished manuscript*.

Fougere, P. F. (1985). On the Accuracy of Spectrum Analysis of Red Noise Processes Using Maximum Entropy and Periodogram Methods: Simulation Studies and Application to Geophysical Data, *Journal of Geophysics. Res.*, 90(A5), 4355–4366.

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

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

Gilden, D. L. (1997). Fluctuations in the time required for elementary decisions. *Psychological Science, 8*, 296-301.

Gilden, D. L. (2001). Cognitive emissions of 1/*f* noise. *Psychological Review, 108*, 33-56.

Gilden, D. L. & Hancock, H. (2007). Response variability in attention-deficit disorders.. *Psychological Science,18*, 796-802.

Gilden, D. L., & Wilson, S. G. (1995a). On the nature of streaks in signal detection. *Cognitive Psychology, 28*, 17-64.

Gilden, D. L., & Wilson, S. G. (1995b). Streaks in skilled performance. *Psychonomic Bulletin & Review, 2,* 260-265.

Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). 1/*f* noise in human cognition. *Science, 267*, 1837-1839.

Glass, G. V., Willson, V. L., & Gottman, J. M. (1975). *Design and analysis of time-series experiments*. Boulder, CO: Colorado Associated University Press.

Gottman, J. M. (1981). *Time-Series Analysis*. New York: Cambridge University Press.

Granger, C. W. J., & Joyeux, R. (1980). An introduction to long-range time series models and fractional differencing. *Journal of Time Series Analysis, 1*, 15-30.

Granger, C. W. J. (1981). Some Properties of Time Series Data and Their Use in Econometric Model Specification. *Journal of Econometrics*, 16, 121-130.

Granger, C. W. J. (1986). Developments in the Study of Co-integrated Economic Variables. *Oxford Bulletin of Economics and Statistics, 48*, 213-228.

Gujarati, D. (2003). *Basic econometrics*. (4th ed.). New York: McGraw-Hill.

Handel, P. H., & Chung, A. L. (Eds.) (1993). *Noise in physical systems and 1/f fluctuations*. New York: American Institute of Physics.

Hannan, E. J., & Rissanen, J. (1982). Recursive estimation of mixed autoregressive moving average order. *Biometrika, 69*, 81-94.

Haslett, J., & Raftery, A. E. (1989). Space-time Modelling with long-memory dependence: assessing Ireland's wind power resource. *Applied Statistics 38,* 1–50.

Hausdorff, J. M., & Peng, C. K. (1996). Multiscaled randomness: a possible source of 1/f noise in biology. *Physical Review E, 54*, 2154-2157.

Hausdorff, J. M., Mitchell, S. L., Firtion, R., Peng, C. K., Cudkowicz, M. E., Wei, J. Y., & Goldberger, A. L. (1997). Altered fractal dynamics of gait: reduced stride-interval correlations with aging and Huntington's disease. *Journal of Applied Physiology, 82*, 262-269.

Hausdorff, J. M., Zemany, L., Peng, C. K., & Goldberger, A. L. (1999). Maturation of gait dynamics: stride-to-stride variability and its temporal organization in children. *Journal of applied Physiology, 86*, 1040-1047.

Hauser, A., Pötscher, B. M., & Roschenhofer, E. (1999). Measuring persistence in aggregate output: ARMA models, fractionally integrated ARMA models and nonparametric procedures. *Empirical Economics, 24*, 243-269.

Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D, 31,* 277–283.

Holden, J.G. (2005). Gauging the fractal dimension of response times from cognitive tasks. In M.A. Riley, & G.C. Van Orden (Eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 268-318). [Online] Retrieved October 7, 2008, from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika, 68*, 165-176.

Hosking, J. R. M. (1984). Modeling persistence in hydrological time seris using fractional differencing. *Water Resources Research, Vol. 20, 12*, 1898-1908.

Hurst, H. E. (1965). *Long-term storage: an experimental study.* London: Constable.

Johnson, K. A., Kelly, S. P., Robertson, I. H., Barry, E., Mulligan, A., Daly, M., Lambert, D., McDonnell, C., Connor, T. J., Hawi, Z., Gill, M., Bellgrove, M. A. (2008). Absence of the 7-Repeat Variant of the DRD4 VNTR Is Associated With Drifting Sustained Attention in Children With ADHD But Not in Controls. *American Journal of Medical Genetics Part B 147B*, 927-937.

Kasdin, N. J, (1995). Discrete simulation of colored noise and stochastic processes and $1/f^{\alpha}$ power low noise generation. *Proceedings of the IEEE, 83*, 802-827.

Kello, C. T., Anderson, G. C., Holden, J. G., Van Orden, G. C. (2008). The pervasiveness of 1/*f* scaling in speech reflects the metastable basis of cognition. *Cognitive Science: A Multidisciplinary Journal, 32*, 1217-1231.

Kello, C. T., Beltz, B. C., Holden, J. G., Van Orden, G. C. (2007). The emergent coordination of cognitive function. *Journal of Experimental Psychology: General, 136*, 551-568.

Kolmogorov, A. N. (1941). Local structure of turbulence in fluid for very large Reynold numbers. Transl. In Turbulence. *S. K. Friedlander and L. Topper (eds.), 1961, Inerscience Publishers*, 151-155.

Koreisha, S. G., & Pukkila, T. M. (1990). A generalized least-squares approach for estimation of autoregressive moving-average models. *Journal of Time Series Analysis, 11*, 139-151.

Koscielny-Bunde, E., Bunde, A., Havlin, S. , Roman, H. E. , Goldreich, Y., Schellenhuber, H.-J. (1998). Indication of a Universal Persistence Law Governing Atmospheric Variability. *Physical Review Letters* 81, 729.

Leite, A., Rocha, A. P. (2007). Long-Range Dependence in Heart Rate Variability Data: ARFIMA Modeling vs Detrended Fluctuation Analysis. *Computers in Cardiology*, *34,* 21−24.

Linkenkaer-Hansen, K. (2002). *Self-organized criticality and stochastic resonance in the human brain.* Dissertation. Helsinki University of Technology Laboratory of Biomedical Engineering.

Linkenkaer-Hansen, K., Monto, S., Rytsälä, H., Suominen, K., Isometsä, E., Kähkönen, S. (2005). Breakdown of long-range temporal correlations in theta oscillations in patients with major depressive disorder. *Journal of Neuroscience, 25,* 10131–10137.

Maddala, G. S., & Kim, I. (1998). *Unit Roots, Cointegration and Structural Change*. Cambridge: University Press.

Madison, G. (2004). Fractal modeling of human isochronous serial interval production. *Biological Cybernetics, 90,* 105-112.

Madison, G., Delignières, D. (in press). Auditory feedback affects the long-range correlation of isochronous serial interval production: support for a closed-loop or memory model of timing. *Experimental Brain Research.*

Makikallio, T.-H., Koistinen, J., Jordaens, L., Tulppo, M. P., Wood, N., Golosarsky, B. B., Peng, C.-K., Goldberger, A. L., Huikuri, H.V. (1999). Heart rate dynamics before spontaneous onset of ventricular fibrillation in patients with healed myocardial infarcts. *American Journal of Cardiology,* 83, 880-884.

Malamud, B. D., Turcotte, D. L.(1999). Self-affine time series: Measures of weak and strong persistence. *Journal of Statistical Planning and Inference,* 80, 173-196.

Mandelbrot, B. B. (1975). Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 31,* 271-285.

Mandelbrot, B. B.(1997). *Fractals and Scaling in Finance* Berlin: Springer.

Mandelbrot, B. B., and van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review, 10*, 422-437.

Mandelbrot, B. B., and Wallis, J. R. (1969). Comuter experiments with fractional Gaussian noises. *Water Resources Research, 5,* 228-267.

Marks-Tarlow, T. (1999). The self as a dynamical system. *Nonlinear Dynamics, Psychology and Life Sciences, 3*, 311-345.

McCleary, R., & Hay, R. A. Jr. (1980). *Applied time series analysis for the social sciences.* Beverly Hills, CA: Sage Publications.

Molenaar, P. C. M. (2007).Psychological methodology will change profoundly due to the necessity to focus on intra-individual Variation. *Integrative Psychological and Behavioral Science, 41*, 35-40.

Montanari, A., Rosso, R., M.S. Taqqu, M. S. (2000). *A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan.* Water Resources Res. 1249.

Palaniappan, R. (2005). *Discrimination of Alcoholic Subjects using Second Order Autoregressive Modeling of Brain Signals Evoked during Visual Stimulus Perception*. IEC (Prague) 2005. 282-287.

Pashler, H., & Johnston, J. C. (1998). Attentional limitations in dual task performance. In H. Pashler (Ed.), *Attention* (pp. 155–189). East Sussex, England: Psychology Press.

Peng, C.-K., Buldyrev, A. L., Goldberger, Havlin, S. Sciortino, F., Simons, S., Stanley, H.E. (1992). A growth model for DNA evolution. *Nature 356,* 168.

Peng, C.-K., Mietus, J., Hausdorff, J. M., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1993). Long-range anticorrelations and non-Gaussian behavior of the heartbeat. *Physical Review Letters, 70,* 1343-1346.

Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., Goldberger, A.L. (1994). Mosaic organization of DNA nucleotides. *Phys. Review E*, 49.

Pilgram, B., & Kaplan, D. T. (1998). A comparison of estimator for 1/f noise. *Physica D, 114,* 108-122.

Pressing, J. (1999). S*ources for 1/f noise effects in human cognition and performance.* Proceedings of the 4th Conference of the Australasian Cognitive Science Society, Newcastle.

Rangarajan, G., & Ding, M. (2000). First passage time distribution for anomalous diffusion. *Physics Letters A, 273,* 322-330.

Reisen, V. A. (1994). Estimation of the fractional difference parameter in the ARFIMA(p, d, q) model using the smoothed periodogram. *Journal Time Series Analysis, 15,* 335–350.

Reisen, V. A., Abraham, B., & Toscano, E. M. M. (2000). Parametric and semiparametric estimations of stationary univariate ARFIMA model. *Brazilian Journal of Probability and Statistics 14,* 185–206.

Reisen , V. A., Abraham, B., & Lopes, S. (2001). Estimation of parameters in ARFIMA processes. A simulation study. *Communications in Statistics, Simulation and Computation, 30*, 787-803.

Riley, M. A., & Van Orden, G. C. (Eds.) (2005). *Tutorials in contemporary nonlinear methods for the behavioral sciences.* [Online] Retrieved October 7, 2008, from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp.

Rissannen, J. (1978). Modeling the shortest data description, *Automatica, 14*, 465-471.

Robinson, P. M. (1992). Semiparametric analysis of long-memory time series. *The Annals of Statistics, 22, 515-539.*

Robinson, P. M. (1994). Efficient tests of nonstationarity hypotheses. J*ournal of the American Statistical Association, 89,* 1420-1437.

Rosel, J., & Elósegui, E. (1994). Daily and weekly smoking habits: A Box Jenkins analysis. *Psychological Reports, 75*, 1639-1648.

Schwarz, G. (1978). Estimation the dimension of a model. *Annals of Statistics, 6*, 461-469.

Shumway, R. H., & Stoffer, D. S. (2006). Time series analysis and its applications: with R examples. New York, NY: Springer.

Simons-Morton, B. G. & Chen, R**.** (2006). *Over time relationships between early adolescent and peer substance use. Addictive Behavior, 31*, 1211-1223.

Slifkin, A.B., & Newell, K.M. (1998). Is variability in human performance a reflection of system noise? *Current Directions in Psychological Science, 7*, 170-177.

Smith, J., Taylor, N, & Yadav, S. (1997). Comparing the bias and misclassification in ARFIMA models. *Journal of Time Series Analysis, 18*, 507-527.

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

Stadnytska, T., Braun, S., & Werner, J. (2006a). Comparison of Automated Procedures for ARMA Model Identification. *Behavior Research Methods, 40,* 250-262.

Stadnytska, T., & Werner, J. (2006). Sample size and accuracy of estimation of the fractional differencing parameter. *Methodology, 4*, 135-144

Stadnytska, T., Braun, S., & Werner, J. (2006b). Model identification of integrated ARMA processes. *Multivariate Behavior Research, 43,* 1-28.

Stadnytska, T., Braun, S. & Werner, J. (2009a). Distinguishing Persistent Signals by Means of Periodogram Analyzes. Manuscript under review: *Journal of Mathematical Psychology.*

Stadnytska, T., Braun, S. & Werner, J. (2009b). Analyzing Fractal Dynamics Employing R. Manuscript under review: *Nonlinear Dynamics, Psychology and Life Sciences.*

Stadnytska, T. (2009c). Deterministic or Stochastic Trend: Decision on the Basis of the Augmented Dickey-Fuller Test. *Methodology (in press).*

Stroe-Kunold, E., Stadnytska, T., Werner, J. & Braun, S. (2009). Estimating Long-Range Dependence in Time Series: An Evaluation of Estimators Implemented in R. *Behavior Research Methods*, *41,909-923.*

Tanaka. K. (1999). The nonstationary fractional unit root. *Econometric Theory, 15*, 549-582.

Taqqu, M. S., Teverovsky, V., & Willinger, W. (1995). Estimators for long-range dependence: an empirical study. *Fractals, 3,* 785–798.

Taqqu, M. S., & Teverovsky, V. (1996). Robustness of Whittle-type estimators for time series with long range dependence. *Preprint form*.

Taqqu, M. S., & Teverovsky, V. (1998). On estimating the intensity of long-range dependence in finite and infinite variance time series, in *A practical guide to heavy tails: statistical techniques and applications*, pp. 177–217. Boston: Birkhauser.

Thornton, T. L., & Gilden, D. L. (2005). Provenance of correlation in psychological data. *Psychonomic Bulletin & Review, 12*, 409-441.

Torre, K., Delignières, D., & Lemoine, L. (2007a). Detection of long-range dependence and estimation of fractal exponents through ARFIMA modeling. *British Journal of Mathematical and Statistical Psychology, 60,* 85-106.

Torre, K., Delignières, D., & Lemoine, L. (2007b). $1/f^{\beta}$ fluctuations in bimanual coordination: an additional challenge for modeling. *Experimental Brain Research, 183, 225-234.*

Tsay, R. S., & Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of American Statistical Association, 79*, 84-96.

Tsay, R. S., & Tiao, G. C. (1985). Use of canonical analysis in time series model identification. *Biometrika, 72*, 299-315.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of experimental psychology, 3*, 331-350.

Velicer, W. F. & Harrop, J. W. (1983). The reliability and accuracy of time series model identification. *Evaluation Review,* 7, 551-560.

Velicer, W. F., Redding, C. A., Richmond, R. L., Greeley, J., Swift, W. (1992). A time series investigation of three nicotine regulation models. *Addictive behaviors, 17(4*), 325-45.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation an interpretation of $1/f^{\alpha}$ noise in human cognition. *Psychonomic Bulletin & Review, 11,* 579-615.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2005). Human cognition and a pile of sand: A discussion on serial correlations and self-organized criticality. *Journal of Experimental Psychology: General, 134,* 108-116.

Ward, L. M., & Richard, C. M. (2001). *$1/f^{\alpha}$ noise and decision complexity*. Unpublished manuscript, University of British Columbia, Vancouver, British Columbia, Canada.

Warner, R. M. (1998). *Spectral analysis of time- series data*. New York: Guilford.

West, B. J., & Shlesinger, M. F. (1990). The noise in natural phenomena. *American Scientist, 78*, 40-45.

West, B. J. & Scafetta, N. (2003). Nonlinear dynamical model of human gait. *Physical Review E*, 67.

Werner, J. (2005), *Zeitreihenanalysen mit Beispielen aus der Psychologie*, Berlin: Logos.

Willinger, W. Taqqu, M. S., Erramilli A. (1996). *The changing nature of network traffic: scaling phenomena.* In: F.P. Kelly, S. Zachary, I. Ziedins (eds.), Stochastic Networks: Theory and Applications, Clarendon Press, Oxford, 339-366.

APPENDIX

R-code for $^{low}PSD_{we}$

```
##############################################################################
LowPSDwe_875 <- function(y, ...)
{
n <- length(y)
x <- y - mean(y)
t=seq(1,n,1)
xw=x * (1-((2*t/(n+1)) - 1)**2)
yph <- c(x[1], x[n])
xph <- c(1,n)
reg <- lm(yph ~ xph )
param <- coef(reg)
xt=xw-(param[1]+param[2]*t)
spec = spectrum(xt)
nr=length(spec$freq)
nn =nr* 0.125
specfreq <- spec$freq[1:nn]
specspec <- spec$spec[1:nn]
logfreq <- log(specfreq)
logspec <- log(specspec)
lmb <- lm(logspec ~ logfreq)
b <- coef(lmb)
return(-b[2])
                              }
##############################################################################
```

Simulation of 1000 series of fractional Brownian Motion with H=0.5, T=512

```
################################################################################
lm.Control <- function(mod, HG, HB, N, ...)
{
library(fractal)
y <- lmSimulate(lmModel(mod, HG=HG, HB=HB, ...), n.sample=N,...)
write.csv2(y,"y.csv")
y=read.csv2("y.csv")
y=y[,2]
 v <- matrix(c(y), ncol=1)
 write.table(v, file="sim_fbm.txt", append=TRUE, row.names=FALSE,
                              col.names=FALSE, sep="    ")
HG=.01
hb <- c(.5)
for (HB in hb){
N <- c( 512)
nrep <- 1000
for(i in N) {
          rep <- 0
          for (j in 1:nrep){
                    rep = rep + 1
                    cat ("REP", rep)
                    cat ("N", i)
                    if(HB==.01){cat("HG", HG)} else{cat("HB", HB)}
                    lm.Control("fgn", HG=HG, HB=HB, i)
                    if(rep > nrep) next              }}
                                        }

################################################################################
```

Simulation of 1000 series of fractional Gaussian Noise with H=0.5, T=512

```
##############################################################################
lm.Control <- function(mod, HG, HB, N, ...)
{
library(fractal)
y <- lmSimulate(lmModel(mod, HG=HG, HB=HB, ...), n.sample=N,...)
write.csv2(y,"y.csv")
y=read.csv2("y.csv")
y=y[,2]
 v <- matrix(c(y), ncol=1)
 write.table(v, file="sim_fgn.txt", append=TRUE, row.names=FALSE,
                              col.names=FALSE, sep="    ")
HB=.01
hg <- c(.5)
for (HG in hg){
N <- c( 512)
nrep <- 1000
for(i in N) {
          rep <- 0
          for (j in 1:nrep){
                    rep = rep + 1
                    cat ("REP", rep)
                    cat ("N", i)
                    if(HB==.01){cat("HG", HG)} else{cat("HB", HB)}
                    lm.Control("fgn", HG=HG, HB=HB, i)
                    if(rep > nrep) next              }}
                                        }

##############################################################################
```

Simulation of 1000 series of stationary ARFIMA (0,d,0) processes with d=0.3, T=500

```
##############################################################################
fracdiff.Control  <- function(N, ar, ma, d)
{
library(fracdiff)
y <- fracdiff.sim(n=N, ar=ar, ma=ma, d=d)
v <- matrix(c(y$series), ncol=1)
write.table(v, file="sim_stat0d0.txt", append=TRUE, row.names=FALSE,
                        col.names=FALSE, sep="    ")
}
D <- c(.3)
for (d in D){
N<- c(500)
nrep <- 1000
for(i in N) {
        rep <- 0
        for (j in 1:nrep){
                rep = rep + 1
                fracdiff.Control(i, ar=NULL, ma=NULL, d)
                if(rep > nrep) next             }}
                                                  }

##############################################################################
```

Simulation of 1000 series of nonstationary ARFIMA (0,d,0) processes with d=0.7, T=500

```
################################################################################
fracdiff.Control  <- function(N, ar, ma, d)
{
library(fracdiff)
r <- fracdiff.sim(n=N, ar=ar, ma=ma, d=d)
x<-r$series
y<-cumsum(x)
v <- matrix(c(y), ncol=1)
write.table(v, file="sim_instat0d0.txt", append=TRUE, row.names=FALSE,
                        col.names=FALSE, sep="    ")
}
D <- c(-.3)
for (d in D){
N<- c(500)
nrep <- 1000
for(i in N) {
        rep <- 0
        for (j in 1:nrep){
                rep = rep + 1
                fracdiff.Control(i, ar=NULL, ma=NULL, d)
                if(rep > nrep) next             }}
                                                  }
################################################################################
```

Simulation of 1000 series of simple stationary ARFIMA (1,d,0) processes with ϕ=0.6, d=0.4, T=500

```
##############################################################################
fracdiff.Control  <- function(N, ar, ma, d)
{
library(fracdiff)
y <- fracdiff.sim(n=N, ar=ar, ma=ma, d=d)
v <- matrix(c(y$series), ncol=1)
write.table(v, file="sim_stat1d0.txt", append=TRUE, row.names=FALSE,
                        col.names=FALSE, sep="    ")
}
AR<-c(.6)
for (ar in AR){
D <- c(.4)
for (d in D){
N<- c(500)
nrep <- 1000
for(i in N) {
        rep <- 0
        for (j in 1:nrep){
                rep = rep + 1
                fracdiff.Control(i, ar=ar, ma=NULL, d)
                if(rep > nrep) next             }}
                                                  }
                                                  }
##############################################################################
```

Simulation of 1000 series of simple stationary ARFIMA (0,d,1) processes with θ=0.6, d=0.4, T=500

```
##############################################################################
fracdiff.Control  <- function(N, ar, ma, d)
{
library(fracdiff)
y <- fracdiff.sim(n=N, ar=ar, ma=ma, d=d)
v <- matrix(c(y$series), ncol=1)
write.table(v, file="sim_stat1d0.txt", append=TRUE, row.names=FALSE,
                        col.names=FALSE, sep="    ")
}
MA<-c(.6)
for (ma in MA){
D <- c(.4)
for (d in D){
N<- c(500)
nrep <- 1000
for(i in N) {
        rep <- 0
        for (j in 1:nrep){
                rep = rep + 1
                fracdiff.Control(i, ar=NULL, ma=ma, d)
                if(rep > nrep) next             }}
                                                  }
                                                  }
##############################################################################
```

Simulation of 1000 series of mixed stationary ARFIMA (1,d,1) processes with ϕ=0.6, θ=0.3, d=0.4, T=500

```
##############################################################################
fracdiff.Control  <- function(N, ar, ma, d)
{
library(fracdiff)
y <- fracdiff.sim(n=N, ar=ar, ma=ma, d=d)
v <- matrix(c(y$series), ncol=1)
write.table(v, file="sim_stat1d1.txt", append=TRUE, row.names=FALSE,
                        col.names=FALSE, sep="    ")
}
AR<-c(.6)
for (ar in AR){
MA<-c(.3)
D <- c(.4)
for (d in D){
N<- c(500)
nrep <- 1000
for(i in N) {
        rep <- 0
        for (j in 1:nrep){
                rep = rep + 1
                fracdiff.Control(i, ar=ar, ma=ma, d)
                if(rep > nrep) next             }}
                                                  }
                                                  }
##############################################################################
```