Long-Memory versus Structural Breaks: An overview[1]

by

Philipp Sibbertsen
Fachbereich Statistik, Universität Dortmund, D-44221 Dortmund, Germany

Version June 2001

Abstract

We discuss the increasing literature on misspecifying structural breaks or more general trends as long range dependence. We consider tests on structural breaks in the long-memory regression model as well as the behaviour of estimators of the memory parameter when structural breaks or trends are in the data but long-memory is not. It can be seen that it is hard to distinguish deterministic trends from long-range dependence.

KEY WORDS: Long-memory; structural breaks; trends

1 Introduction

Long-memory time series have been a popular area of research in econometrics and statistics in the recent years because of their applicability in many sciences. Long-range dependence or long-memory means that the correlation of a time series decays hyperbolically, not exponentially like for example for ARMA-processes.

Long-range dependence was first observed by the hydrologist Hurst who analyzed the minimal water flow of the Nile River when planning the Aswan dam.

[1]Research supported by Deutsche Forschungsgemeinschaft under SFB 475

Hurst acknowledged that standard forecasting methods fail for this data. Instead of independence or weak correlations between data points far away from each other, he observed strong dependencies. The phenomenon of long-range dependence in water flow data was observed in many other rivers by Mandelbrot/Wallis (1969). Also the Rhine River exhibits long-range dependence (see Lohre/Sibbertsen (2001) and references therein). Additional geophysical applications of long-memory are for instance the temperature data of the northern hemisphere. Other domains of application are Computer Science and Economics. Many economic data sets show a persistent behaviour and therefore it seems natural to apply long-memory models to these economic time series. Beginning with Granger (1966), an intensive discussion about the application of long-range dependence in Economics and its consequences was initiated. But in many situations it is not clear whether the observed dependence structure is real long-memory or an artefact of some other phenomenon such as structural breaks or deterministic trends. Long-memory in the data would have strong consequences. As described in section 3 the valuation of an option on stocks would be changed entirely. The price of the option can in case of long-range dependence double the price when long-memory is neglected. Also for forecasting future events as high water it is important to know whether the data exhibits long-range dependence or if it is an artefact of a deterministic trend.

So far there is no acknowledged method to distinguish long range dependence and structural breaks or more general trends. The purpose of this paper is to review the literature concerning the influences of long-memory to tests on structural breaks and on the other hand the consequences of trends to the estimation of the dependence structure of the observed time series.

The paper is organized as follows. In the next section long-range dependence is defined and the most relevant properties of long-memory models are discussed. In section 3 a motivating example is considered. Section 4 discusses the behaviour of tests on structural breaks in the presence of long-range dependence. Section 5 considers the consequences of trends added to a short-memory noise for the estimation of the memory parameter.

2 Long-memory time series

2.1 Definition of long-memory

Long-memory or long-range dependence means that observations far away from each other are still strongly correlated. The correlations of a long-memory process decay slowly that is with a hyperbolic rate. We have the following definition of long-range dependence:

Definition 2.1 (Long-memory process) *Let X_t be a stationary process with correlation function $\rho(k)$ and let $H \in (1/2, 1)$. Furthermore let c_ρ be a positive constant with*

$$\lim_{k\to\infty} \frac{\rho(k)}{c_\rho k^{2H-2}} = 1. \tag{2.1}$$

Then X_t exhibits long-memory or long range dependence.

It follows from this definition that the correlations of a long-memory process decay with a hyperbolic rate. They are not summable. Instead of the parameter H we use in this paper also in some situations the parameter $d := H - 1/2$. H is called Hurst parameter. The use of the parameter d is standard because it is commonly used in the ARFIMA modeling of long-memory processes discussed below.

An equivalent definition of long-range dependence can be given by using the spectral density $f(\lambda)$ of the process X_t. Note that the long-term behaviour of a process is specified by the small frequencies of the periodogram. A long-memory process has a pole of the spectral density at the origin. We have the following definition:

Definition 2.2 *Let X_t be a stationary process and $H \in (1/2, 1)$ real. Let also be c_f a positive constant such that*

$$\lim_{\lambda\to 0} \frac{f(\lambda)}{[c_f|\lambda|^{1-2H}]} = 1. \tag{2.2}$$

Then X_t is called stationary process with long-memory.

These definitions are equivalent. The details are omitted here. From these definitions of long-range dependence we obtain important properties of long-memory time series:

- the covariances behave asymptotically like a constant c_H times k^{2H-2} for $1/2 < H < 1$;
- the correlations are not summable, that is $\sum_{k=-\infty}^{\infty} \rho(k) = \infty$;
- the spectral density f has a pole at the origin and behaves like a constant c_f times λ^{1-2H} near the origin for $1/2 < H < 1$.
- the variance of the sample mean behaves asymptotically for $t \to \infty$ like a constant c_{var} times t^{2H-2} for $1/2 < H < 1$.

Also some qualitative properties of a typical trajectory of a long-memory process can be enumerated:

- the trajectory has local trends and cycles;
- it is mean stationary, so no overall trends or cycles are observable;
- it is mean reverting;
- it shows a persistent behaviour.

In figure 1 a typical trajectory of a long-memory time series of length $N = 1000$ and memory parameter $H = 0.9$ is given.

The process X_t is stationary and exhibits long-range dependence, if $1/2 < H < 1$. For $0 < H < 1/2$ the process has short-memory. In this situation the spectral density is zero at the origin and the process is said to be antipersistent. For $H = 1/2$ we have independence or standard short-memory. In the case $1 < H < 3/2$ the process is non-stationary but still persistent. For this reason in the literature it is often called non-stationary long-memory. Every other

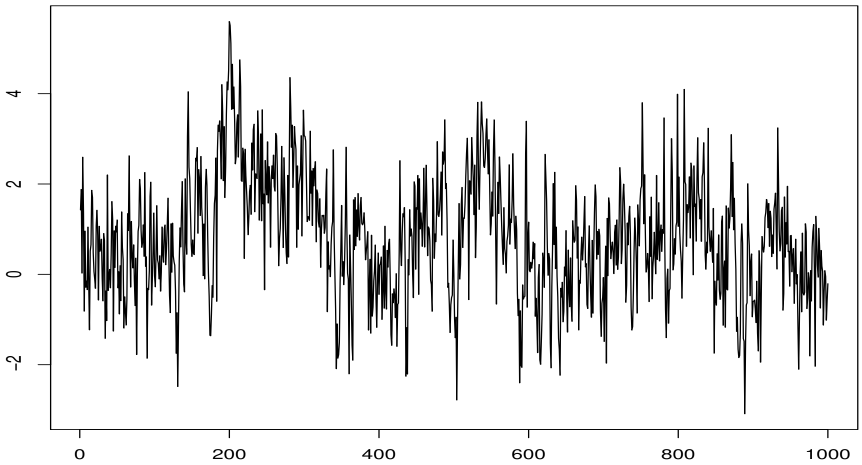

Figure 1: Path of a long-memory time series with $N = 1000$ and $H = 0.9$

situation can be reduced to these cases by differencing the process. In this paper we restrict ourselves to the stationary long-memory case because this is the relevant situation in practise.

2.2 Modeling long-memory processes

A first model for long-memory processes was the fractional Brownian motion introduced by Mandelbrot/van Ness (1968). This approach generalizes standard Brownian motion by using self-similar processes. Here a process X_t is called self-similar with parameter $d \in (-1/2, 1/2)$ if $X_t \stackrel{\mathcal{D}}{=} t^d X_1$. Notice that these equality is only equality in distribution. Self-similarity is not a property of the paths of the process. For the paths the equality above does not hold in general. In what follows fractional Brownian motion is denoted by $B_d(t)$.

Another model class are ARFIMA processes introduced by Granger/Joyeux (1980) and independently by Hosking (1981). They generalize the class of ARIMA models by allowing for a fractional degree of differencing. Denoting with B the Backshift operator, with $\Phi(B)$ and $\Psi(B)$ the AR- and MApolynomials respectively and with ε_t a white noise process, ARFIMA-models are defined as the solution of

$$\Phi(B)(1-B)^d X_t = \Psi(B)\varepsilon_t. \tag{2.3}$$

The operator $(1-B)^d$ can be written as

$$(1-B)^d = \sum_{k=0}^{d} \binom{d}{k} (-1)^k B^{d-k}.$$

The binomial coefficient is defined by terms of the Γ-function

$$\binom{d}{k} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}.$$

Near the origin the spectral density of an ARFIMA process behaves like a constant c_f times $|\lambda|^{-2d}$. Thus these processes exhibit long-range dependence for $0 < d < 1/2$.

2.3 Estimating the memory parameter

There are several methods for estimating the memory parameter of a long-memory process. As this is outside the focus of the present paper we confine ourselves to methods used in below.

The results discussed in this paper are mostly based on the R/S-statistic, a rescaled-range technique. The range of the process X_t is defined by

$$R_N := \max_{1\leq u\leq N}[\sum_{i=1}^{u}(X_i - \bar{X}_N)] - \min_{1\leq u\leq N}[\sum_{i=1}^{u}(X_i - \bar{X}_N)]. \tag{2.4}$$

Let

$$S_N := \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X}_N)^2}, \tag{2.5}$$

where $\bar{X}_N := \frac{1}{N}\sum_{i=1}^N X_i$ be the sample standard deviation. The R/S-statistic is then defined by

$$Q_N := \frac{R_N}{S_N}. \tag{2.6}$$

For the R/S-statistic the following holds:

Theorem 2.1 *Let X_t be a stochastic process with X_t^2 ergodic and $\frac{1}{\sqrt{N}}\sum_{s=1}^N X_s$ converges to a Brownian motion. Then $\frac{1}{\sqrt{N}}Q_N$ converges to a non-degenerated random variable.*

Thus a plot of $\ln Q$ against $\ln N$ scatters around a straight line with slope 1/2 in the case where the central limit theorem holds.

In the case of long-range dependence Mandelbrot (1975) showed the following:

Theorem 2.2 *Let X_t again be a stationary process with X_t^2 ergodic and $N^{-H}\sum_{s=1}^N X_s$ converges to a fractional Brownian motion. Then $N^{-H}Q_N$ converges to a non-degenerated random variable.*

Thus in the case of long-range dependence a plot of $\ln Q$ against $\ln N$ scatters around a straight line with slope H.

Giraitis et al. (2000b) derive a test for long-range dependence based on the R/S-statistic, the so-called V/S-statistic by replacing the range of the partial sums of the process by the estimated variances of the partial sums. This statistic has good power properties. Denoting with $S_k^* := \sum_{i=1}^k (X_i - \bar{X}_N)$ and $\hat{\mathsf{Var}}(S_1^*, \ldots, S_N^*) := \frac{1}{N}\sum_{i=1}^N (S_i^* - \bar{S}_N^*)^2$ the V/S-statistic has the form

$$M_N := N^{-1}\frac{\hat{\mathsf{Var}}(S_1^*, \ldots, S_N^*)}{S_N^2}. \tag{2.7}$$

This test statistic will mainly be used in the last section of this paper.

For other estimators of the memory parameter and for tests for long-memory and further details concerning long-memory processes see for example Beran (1994) or Sibbertsen (1999) and references therein.

3 A motivating example

In economics long-memory is most important for volatilities of stock returns. This has important consequences for the valuation of options.

A standard class of models introduced for modeling volatilities of stock returns consists of the Autoregressive Conditional Heteroscedasticity (ARCH) models (see Engle (1982)). These models assume that the conditional variance depends on the currently known information in a nontrivial way. But they do not allow for modeling long-range dependence, because shocks to the conditional variance decay exponentially and thus have almost no influence for long time optimal forecasts as it is expected due to the persistence property. Empirical findings show for many stock returns that shocks to the conditional variance have a slowly decaying influence to optimal forecasts of the variance. Thus in the recent years long-memory models were used to model the behaviour of the conditional variance of stock returns.

Estimating the dependence structure for daily returns, their absolute values and the squares of daily returns of many German stocks such as BMW, Daimler, Dresdner Bank, Deutsche Bank, Hoechst and BASF show a long-memory behaviour of the absolute values and the squares of daily returns. Figure 2 shows for instance the autocorrelation function of the absolute returns of Deutsche Bank. It clearly seems to point to long-range dependence.

This has important consequences on the valuation of the price of an asset as discussed in Bollerslev/Mikkelsen (1996), by simulating call option prices for the Standard and Poor's 500 composite index. Taking into account a long-memory structure of the volatilities, the price of the call option becomes much higher and in some situations it doubles the price compared with the situation when long-memory is neglected.

A natural question is whether the observed phenomenon is long range dependence or if the estimated dependence structure is an artefact of any other phenomenon as for example structural breaks or trends. Granger/Hyung (1999)

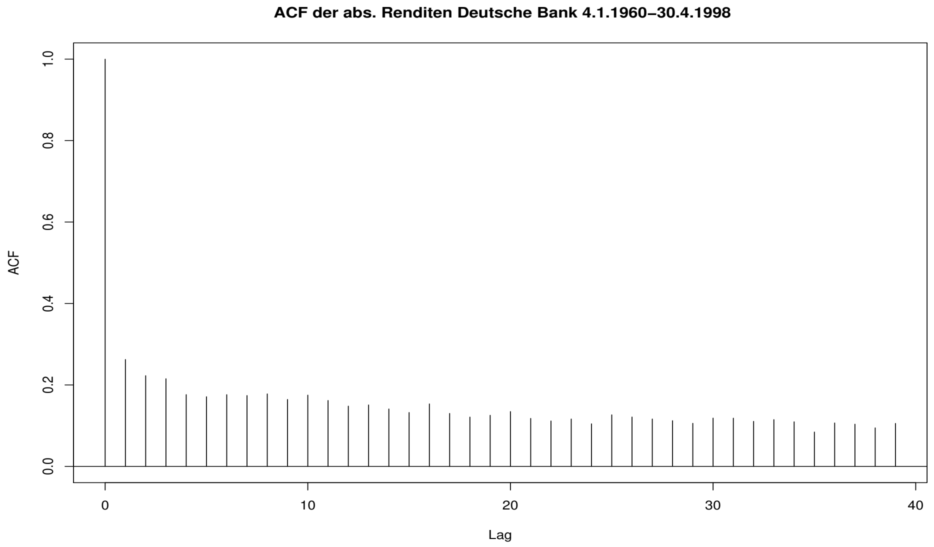

Figure 2: Autocorrelations of the absolute returns of Deutsche Bank from 1960 - 1998.

argue that structural breaks cause the long-memory structure of Standard and Poor's 500 composite index.

Thus long-range dependence and trends including structural breaks can easily be confused. This paper is reviewing the increasing literature concerning this topic. The behaviour of tests on structural breaks in a long-memory model is as well discussed as the behaviour of tests on long-memory if the true underlying process is a weakly dependent process plus a small trend. Distinguishing both of these phenomena is still an open problem.

4 Tests for structural breaks in the presence of long-memory

In this section we consider the behaviour of tests on structural breaks in the linear regression model with long-memory disturbances. Thus the point of departure in this section is the linear regression model

$$y = X\beta + \varepsilon, \tag{4.1}$$

where y is the N-dimensional dependent variable, X is the $N \times k$ matrix of non-stochastic and fixed regressors, β is the k-dimensional unknown parameter vector and here ε is a long-memory zero mean gaussian time series. For the regressors we assume the following

$$\frac{1}{N}\sum_{t=1}^{N} x_t \quad \to \quad c < \infty \quad \text{and} \tag{4.2}$$

$$\frac{1}{N}\sum_{t=1}^{N} x_t x_t' \quad \to \quad Q \text{ (finite, nonsingular).} \tag{4.3}$$

These are standard assumptions in linear regression large sample asymptotics; they exclude trending data, which require separate treatment. But that is not topic of this paper.

The problem is to test the null hypothesis that the parameter vector β is constant for all observations.

For the case of a known breakpoint Hidalgo/Robinson (1996) obtained that a Wald test procedure rejects the null hypothesis with probability tending to one.

Unfortunately in most practical situations the breakpoint is unknown. Thus we focus here on the CUSUM test for structural change. This test has a lack of power but other methods like the optimal test by Andrews/Ploberger (1994) show a similar behaviour as the CUSUM test in the long-memory model. We consider here the CUSUM test because of its more intuitive asymptotics. For an overview about tests on structural change, see for example Stock (1994). Let us first introduce the standard CUSUM test. It is based on recursive residuals and was first introduced by Brown et al. (1975). In detail the standard CUSUM test is defined by

$$\tilde{e}_t \quad = \quad \frac{y_t - x_t'\hat{\beta}^{(t-1)}}{f_t}, \quad \hat{\beta}^{(t-1)} = \left(X^{(t-1)'}X^{(t-1)}\right)^{-1} X^{(t-1)'}y^{(t-1)} \tag{4.4}$$

$$f_t \quad = \quad \left(1 + x_t'(X^{(t-1)'}X^{(t-1)})^{-1}x_t\right)^{\frac{1}{2}} \ (t = K+1, \ldots, N), \tag{4.5}$$

where the superscript $t-1$ means that only observations $1, \ldots, t-1$ are used. It rejects the null hypothesis of no structural break for large values of

$$S_N = \sup_{0<\lambda<1} W_N(\lambda)/(1+2\lambda), \tag{4.6}$$

where

$$W_N(\lambda) := N^{-\frac{1}{2}} \hat{\sigma}_\varepsilon^{-1} \sum_{t=K+1}^{[N\lambda]} \tilde{e}_t. \tag{4.7}$$

Here $\hat{\sigma}_\varepsilon^{-1}$ denotes a consistent estimator for the variance of the error term. In the case of iid or weakly dependent disturbances $W_N(\lambda)$ tends to a standard Brownian motion. The asymptotic behaviour of the test in the long-memory regression model is given in the following theorem:

Theorem 4.1 *In the regression model (4.1), with long-memory disturbances we have*

$$N^{-d} W_N(\lambda) \to B_d(\lambda), \tag{4.8}$$

where $B_d(\lambda)$ denotes fractional Brownian Motion with self-similarity parameter $d \in (0, 1/2)$.

Proof: See Krämer/Sibbertsen (2000).

This theorem shows that the null distribution of the standard CUSUM test tends to infinity in the presence of long-memory disturbances. Likewise the standard CUSUM test has an asymptotic size of unity.

The standard CUSUM test has bad properties in the case of structural breaks occuring at the end of the observation period. For this reason Ploberger/Krämer (1992) modified the standard CUSUM test by replacing the recursive residuals by standard OLS residuals. This OLS-based CUSUM test is sensitive to structural breaks at the end of the data. The test statistic is defined by

$$TS := \sup_{0<\lambda<1} |C_N(\lambda)|, \quad \text{where} \tag{4.9}$$

$$C_N(\lambda) := N^{-\frac{1}{2}} \hat{\sigma}_\varepsilon^{-1} \sum_{t=1}^{[N\lambda]} e_t, \tag{4.10}$$

and where $e_t := y_t - x_t' \hat{\beta}$ are the OLS–residuals from (4.1). In the case of iid or short-memory disturbances $C_N(\lambda)$ converges to a standard Brownian bridge. In our situation we obtain the following limiting null distribution:

Theorem 4.2 *In the regression model (4.1), with long-memory disturbances we have*

$$N^{-d} C_N(\lambda) \to B_d(\lambda) - c' Q^{-1} \xi(\lambda), \tag{4.11}$$

where $B_d(\lambda)$ is fractional Brownian Motion with self-similarity parameter $d \in (0, 1/2)$ and $\xi(\lambda) \sim N(0, \lambda \sigma_\varepsilon^2 Q)$.

Proof: See Krämer/Sibbertsen (2000).

The process on the right hand side of (4.11) is called fractional Brownian bridge. For $d = 0$ it is standard Brownian bridge.

Thus also the test statistic of the OLS-based CUSUM test tends in probability to infinity under the null hypothesis of no structural break. Both, the standard CUSUM test as well as the OLS-based CUSUM test is extremely non-robust to long-memory disturbances, in the sense that long-range dependence is easily mistaken for structural change when conventional critical values are employed.

The reason for these results is the bad rate of convergence of the OLS-estimator in the long-memory regression model. In the case of long-range dependent error terms the least squares estimator has a rate of convergence of $N^{1/2-d}$, where d is the memory parameter instead of $N^{1/2}$ in the case of independent or short-memory disturbances. But both types of the CUSUM test depend on the least-squares estimation of the parameter vector β. Because this is the optimal rate of convergence for estimates of β also tests which are optimal in the case of independent or short-memory regressors have similar properties in the long-memory regression model.

Sibbertsen (2000) generalized these results to robust CUSUM-M tests because also outlier can cause the size of the test tending to one. Therefore M-estimation of the parameter vector is used instead of least-squares estimation. The results for robust tests are similar to the non-robust case. Details are omitted here.

5 Long-memory versus trends

A more general problem than distinguishing long-memory and structural breaks is in a way the question if general trends in the data can cause the Hurst effect. So in this section we consider tests of long-memory and their behaviour when trends are present in the data generating process. In the next subsection we restrict the considerations to monotonic trends. Thereafter general trends are considered. At the end of this section SEMIFAR-models are introduced. They allow for modeling trends and long-range dependence simultaneously.

5.1 Long-memory and monotonic trends

The first paper dealing with this problem is Bhattacharya et al (1983). They show that adding a deterministic trend to a short memory process can cause spurious long-memory. They consider the model

$$X_t = f(t) + Y_t. \tag{5.1}$$

At first $f(t)$ is a deterministic trend of the form

$$f(t) = k(m+t)^{\beta}, \tag{5.2}$$

where m is nonnegative and k does not equal zero. The exponent β is assumed to be in the interval $(-1/2, 0)$. The parameter m can be interpreted as a location parameter. Notice that the trend is decreasing for k positive and increasing for negative values of k.

The process Y_t in (5.1) is assumed to have short-memory in the following sense. We say a stationary process Y_t has short-memory, if the covariances are absolutely summable, that is

$$\sum_{k=-\infty}^{\infty} |Cov(Y_k, Y_0)| < \infty$$

and the functional central limit theorem holds, that is

$$N^{-1/2} \sum_{j=1}^{[nt]} Y_j \longrightarrow \sigma B(t).$$

Here σ denotes the variance of the process and $B(t)$ denotes the standard Brownian motion. This is a quite general definition of short-memory following Giraitis et al. (2000a). It includes standard models like ARMA(p,q)-models as well as GARCH(p,q)-processes. Using rescaled-range techniques Bhattacharya et al. (1983) show that a trend of type (5.2) produce long-range dependence of order $1+\beta$. Thus a weak monotonic trend of form (5.2) can be confused with long-range dependence of order $1+\beta$. Following the notation of Bhattacharya et al. (1983) we denote

$$p_t = \frac{1}{t} \sum_{n=1}^{t} f(n), \qquad p_0 := 0, \quad t = 1, 2, \ldots, N$$

and

$$\mu_N(t) = t(p_t - p_N), \qquad t = 0, 1, \ldots, N.$$

Using this notation and R_N as defined in (2.4) the following holds:

Theorem 5.1 *Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of random variables of the form (5.1). Let*

$$\triangle_N := \frac{1}{\sqrt{N}}[\max_t \mu_N(t) - \min_t \mu_N(t)].$$

Then $(1/N^H)R_N$ *converges in distribution as* $N \to \infty$ *to a limit almost surely not* 0 *with* $H > 1/2$ *if and only if*

$$\triangle_N \sim cN^{H-1/2},$$

where $\sim$ *denotes asymptotic equality as* $N \to \infty$ *and* c *denotes a positive constant.*

Proof: See Bhattacharya et al. (1983).

Thus the theorem says that R_N/N^H converges in probability to the positive constant c. So this theorem gives a necessary and sufficient condition for trends to produce spurious long-memory.

We also have the following result:

Theorem 5.2 *Let again* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of random variables of the form (5.1). If* $\triangle_N = o(1)$ *as* $N \to \infty$ *then* $(1/\sqrt{N})R_N$ *converges in distribution to a limit which is almost surely not* 0.

Proof: See Bhattacharya et al. (1983).

Note that these results of course depend strongly on the use of R/S methodology. As we see in the following other methods lead to different results.

Künsch (1986) proposed a method for distinguishing monotonic trends and long-memory by considering the periodogram instead of rescaled-range methodology. He proved that the periodogram behaves different in case of a deterministic monotonic trend function compared to long-memory. Define the periodogram of the process X as usual by

$$I_X(j) = \frac{2\pi}{N}|\sum_{n=1}^{N} X_n \exp(-in2\pi j/N)|^2, \qquad 0 < j < \frac{N}{2}. \tag{5.3}$$

He obtained the following result concerning the periodogram:

Theorem 5.3 *Under model (5.1) $4\pi I_X(j)$ has a non-central χ_2^2-distribution with non-centrality parameter*

$$\tau(j,N)^2 = \frac{2}{N}|\sum_{n=1}^{N} f(n)\exp(-in2\pi j/N)|^2.$$

Proof: See Künsch (1986).

Note that for different indices j the periodogram values are independent.

In comparison with the results of Bhattacharya et al. (1983) we consider for example the specific trend

$$f(t) \sim kt^{-\beta} \qquad 0 < \beta < 1$$

and k is a constant. Following Bhattacharya et al. (1983) this trend produces a Hurst effect of order $1-\beta$. On the other hand it can be shown that for such a trend the non-centrality parameter $\tau(j,N)^2$ tends uniformly to zero in regions $\varepsilon N^\gamma \leq j \leq \frac{N}{2}$ for any $\varepsilon > 0$ and $\gamma > \frac{\frac{1}{2}-\beta}{1-\beta}$. This means that the proportion of frequencies of the periodogram effected by such a trend tends to zero fast.

If we consider a process exhibiting real long-range dependence its spectral density is of the form

$$g(\lambda) = k|\lambda|^{\alpha-1}, \qquad 0 < \alpha < 1 \tag{5.4}$$

as mentioned in the introduction. Such a process has long-memory with memory parameter $H = 1 - \frac{\alpha}{2}$. The spectral density of a long-memory process has a pole at the origin and thus standard results concerning the periodogram do not hold in this situation. Künsch (1986) proves that the periodogram of a long-memory process follows a multiple of a χ_2^2 distributed random variable. In detail we have:

Theorem 5.4 *Let X_t be a Gaussian process with long-range dependence and thus having a spectral density $g(\lambda)$ of the form (5.4) and let $j_N(1) < \ldots < j_N(k)$ be a sequence of frequencies with $j_N(1)N^{-1/2} \longrightarrow \infty$ and $j_N(k)N^{-1} \longrightarrow 0$. Then for $1 \leq i \leq k$ the $I_X(j_N(i))(\frac{j_N(i)}{N})^{1-\alpha}$ are asymptotically iid, each being distributed like a constant multiple of a χ^2_2 distributed random variable.*

Proof: See Künsch (1986).

Hence it can be seen that the periodogram converges to different distributions in case of trends and long-range dependence. This enables us to distinguish between monotonic trends and long-range dependence.

5.2 Long-range dependence and non-monotonic trends

Unfortunately in most practical situations trends are not monotonic. A natural question is which shapes of non-monotonic trends added to a short-memory process cannot be distinguished from long-memory. This problem is considered for example by Giraitis et al. (2000a). Teverovsky/Taqqu (1997) considered the behaviour of a variance-type estimator of the memory parameter by adding a model with shifts in the mean or a slowly decaying trend of the same type as in Bhattacharya et al. (1983) and Künsch (1986) to the noise process. These trends are special cases of the model of Giraitis et al. (2000a) and thus of the considerations below. Again the point of departure is model (5.1). But from now on $f(t)$ is a general deterministic trend fulfilling only some weak regularity conditions. Assume for the trend the following:

Assumption T1: $[f^{(N)}(k)]_{k=1,\ldots,N}$, $N \geq 1$, is an array of real numbers for which there exists a positive sequence p_N and a function h on $[0,1]$, which is not identically zero, such that for $N \to \infty$

$$p_N^{-1} \sum_{k=1}^{[Nt]} f^{(N)}(k) \to h(t)$$

and

$$\frac{p_N}{N^{1/2}} \to a,$$

where $a \in [0, \infty]$. We further on assume for the trend

Assumption T2: There exists a positive sequence $r_N \to \infty$ and a number $0 < b < \infty$, such that as $N \to \infty$

$$r_N^{-1} \sum_{k=1}^{N} [f^{(N)}(k)]^2 \to b,$$

$$\sum_{k=1}^{N-1} |f^{(N)}(k) - f^{(N)}(k+1)| k^{1/2} = O(r_N^{1/2}),$$

$$\sum_{k=1}^{q_N} |f^{(N)}(k)|^2 = o(r_N)$$

for any $q_N = o(N)$,

$$|f^{(N)}(k)|^2 = O(r_N/N)$$

for $k \sim N$ and

$$\frac{p_N^2}{N r_N} \to b^* < \infty.$$

Note that these trends include as well the change point model considered in the previous section as monotonic trends. Giraitis et al. (2000a) use for their analysis the V/S-statistic. It turns out that the V/S-statistic rejects the null hypothesis of a short-memory structure of the data with a probability tending to one, if the trend decreases with a rate higher than $N^{-\frac{1}{2}}$. Otherwise the added trend has no influence to the test statistic. This again shows that R/S-based methods are not able to distinguish between long-range dependence and "large" trends independently of their shape. They detect spurious long-memory.

Denote in what follows $\alpha := a - b^* h^2(1)$, $h^0(t) := h(t) - th(1)$ and $B^0(t) = B(t) - tB(1)$ is a standard Brownian Bridge. Define

$$\hat{s}_{N,q}^2 = \frac{1}{N} \sum_{j=1}^{N} (X_j - \bar{X}_N)^2 + 2 \sum_{j=1}^{q_N} (1 - \frac{j}{(q_N+1)} \hat{\gamma}_j),$$

where $\hat{\gamma}_j$ denote the empirical covariances.

We have the following theorem describing the behaviour of the V/S statistic:

Theorem 5.5 *Suppose the process $(X_n)_{n\in\mathbb{N}}$ is given by model (5.1) and assumptions (T1) and (T2) for the trend hold. Let $r_N \to \infty$, $q_N \to \infty$, $q_N/N \to 0$ and there exists the limit $q_N r_N N^{-1} \to c \in [0,\infty]$. Then*

$$(T_N U_N)^{-1} \frac{V_N}{\hat{s}^2_{N,q}} \quad \xrightarrow{d} \quad \frac{\int_0^1 (Z_a(t))^2 dt - (\int_0^1 Z_a(t)dt)^2}{V(c)},$$

where

$$Z_a(t) = \begin{cases} \sigma B^0(t) + a h^0(t) & \text{if} \quad a < \infty \\ h^0(t) & \text{if} \quad a = \infty \end{cases}$$

$$V(c) = \begin{cases} \alpha c + \sigma^2 & \text{if} \quad c < \infty \\ \alpha & \text{if} \quad c = \infty \end{cases}$$

$$T_N = \begin{cases} 1 & \text{if} \quad a < \infty \\ p_N^2 N^{-1} & \text{if} \quad a = \infty \end{cases}$$

$$U_N = \begin{cases} 1 & \text{if} \quad c < \infty \\ \frac{N}{q_N r_N} & \text{if} \quad c = \infty \end{cases}$$

Proof: See Giraitis et al. (2000a).

In case that the random term in (5.1) exhibits long-range dependence and a deterministic trend is present the V/S-statistic tends to infinity with the same rate as in the short-memory case with trend. This rate is faster than in the case of long-range dependence without deterministic trend. For the following considerations denote with $B_d(t)$ the fractional Brownian motion with parameter d. In the case of long-range dependence Giraitis et al. (2000b) obtain the following behaviour for the V/S-statistic:

$$(\frac{q_N}{N})^{2d} \frac{V_N}{\hat{s}^2_{N,q}} \quad \xrightarrow{d} \quad \int_0^1 (B_d^0(t))^2 dt - (\int_0^1 B_d^0(t)dt)^2. \tag{5.5}$$

This means that under the alternative of long-memory $V_N/\hat{s}^2_{N,q} \xrightarrow{P} \infty$ with the rate $(N/q_N)^{2d}$.

Considering now the situation where also trends are present that is considering model (5.1) with Y_t exhibiting long-range dependence the following result holds.

Theorem 5.6 *Suppose that again the series $(X_n)_{n\in\mathbb{N}}$ is given by model (5.1) but now the Y_n exhibit long-range dependence. The trend $f^{(N)}(n)$ is assumed to fulfill assumptions (T1) and (T2). Let in addition $r_N \to \infty$, $q_N \to \infty$, $q_N/N \to 0$ and there exists the limit $q_N^{1-2d} r_N N^{-1} \to c^* \in [0,\infty)$. Then*

$$(T_N U_N)^{-1} \frac{V_N}{\hat{s}^2_{N,q}} \xrightarrow{d} \frac{\int_0^1 (Z_a(t))^2 dt - (\int_0^1 Z_a(t)dt)^2}{V(c^*)},$$

where

$$Z_a(t) = \begin{cases} c_d B_d^0(t) + a h^0(t) & \text{if} \quad a < \infty \\ h^0(t) & \text{if} \quad a = \infty \end{cases}$$

$$V(c^*) = \begin{cases} \alpha c^* + c_d^2 & \text{if} \quad c^* < \infty \\ \alpha & \text{if} \quad c^* = \infty \end{cases}$$

$$T_N = \begin{cases} N^{2d} & \text{if} \quad a < \infty \\ p_N^2 N^{-1} & \text{if} \quad a = \infty \end{cases}$$

$$U_N = \begin{cases} q_N^{-2d} & \text{if} \quad c^* < \infty \\ \frac{N}{q_N r_N} & \text{if} \quad c^* = \infty \end{cases}$$

and c_d is a positive number.

Proof: See Giraitis et al. (2000a).

Note that T_N, U_N, Z_a and V are different than in Theorem 5.5 and here depend on the memory parameter d.

These results generalize also the findings of Diebold/Inoue (1999). They show the behaviour of the variance of a process generated as in (5.1) under various models of structural breaks that is of shifts in the mean. The work of Diebold/Inoue (1999) in a way initialized the discussion about confusing long range dependence and trends. But their findings are special cases of the more general work of Giraitis et al. (2000a). Thus we decided to discuss only the results of Giraitis et al. (2000a) here in detail.

5.3 Modeling long-memory and trends

To model long-memory and deterministic trends Beran et al. (1998) (schon erschienen ????) introduced so called SEMIFAR-models. SEMIFAR-models extend ARFIMA-models by allowing for a non-constant deterministic mean function. In detail a SEMIFAR model is a Gaussian process Y_i fulfilling the following equation:

$$\Phi(B)(1-B)^{\delta}\{(1-B)^{m}Y_i - g(t_i)\} = \varepsilon_i, \tag{5.6}$$

where B denotes again the Backshift operator, $m \in \{0, 1\}$, $\delta \in (-0.5, 0.5)$, $g(t)$ is a smooth function on $[0, 1]$, $t_i = \frac{i}{n}$, $\Phi(x)$ is a polynomial with roots outside the unit circle defining the autoregressive part of the model and the $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ are iid random variables. This model includes long-memory, short-memory, deterministic trends and no trends that is a constant mean. We have short-memory if $m = 0$ and $\delta \in (-0.5, 0]$. Long-memory can be modeled by $m = 0$ and $\delta \in (0, 0.5)$ and for $m = 1$ we have difference stationary processes that is the first differences $Y_i - Y_{i-1}$ exhibit short- or long-memory respectively for $\delta \in (-0.5, 0]$ and $\delta \in (0, 0.5)$. To each of these processes a deterministic trend can be added. For a constant mean function we obtain a standard ARFIMA(p, d, 0)-model. Note, that SEMIFAR-models consists only of an autoregressive part for modeling the short-term behaviour of the series because of simplicity.

To fit a SEMIFAR-model to a series the order p of the autoregressive model, the memory parameter $d = m + \delta$ and the trend function have to be estimated simultaneously. Thus before considering properties of SEMIFAR-models in detail we focus on the problem of estimating the trend function. Nonparametric trend estimation has been considered by many authors in several situations. For an overview in the case of short-memory or independent errors see for example Fan/Gijbels (1996). In the case of long-memory errors see for example Csörgö/Mielniczuk (1995) or Beran/Feng (1999). For our purpose we describe only the results of Beran et al. (2000, 2001). Here robust kernel estimators are considered but the results include also standard kernel smoothers. Because robustness is not the purpose of this paper we give the results for the non-robust case. To define a local polynomial estimator let K be a positive symmetric kernel with support $[-1,1]$ and $\int_{-1}^{1} K(u)du = 1$. In addition let $t \in [0,1]$ and $b \in (0,1)$ be a positive bandwidth, and denote by $p \in \mathbb{N}$ the degree of the local polynomial. Then a local polynomial estimator of the trend function $f^{(N)}(t)$ is defined by $\hat{f}^{(N)}(t) = z^T(t)\hat{\beta}(t)$, where $z(t) = (1, t, t^2, \ldots, t^p) \in \mathbb{R}^{p+1}$, and $\hat{\beta}(t) \in \mathbb{R}^{p+1}$ solves the system of $p+1$ equations

$$\frac{1}{Nb}\sum_{i=1}^{N} K(\frac{t_i - t}{b})(Y_i - z^T\hat{\beta}(t))z_j(t) = 0, \qquad j = 0, 1, \ldots, p. \tag{5.7}$$

Notice that $t_i = i/N$. In case of $p = 0$, that is local constant estimation, (5.7) is the standard Nadaraya-Watson type kernel estimator. For the consideration of the asymptotic bias and variance of local polynomials we use for simplicity only the rectangular kernel $K(u) = \frac{1}{2}1_{\{-1\le u\le 1\}}$. We have the following result:

Theorem 5.7 *Let $\hat{\beta}$ be the solution of (5.7). Define the following $(p+1)\times(p+1)$ matrices: $M_N = (m_{ij})_{i,j=1,\ldots,p+1}$ with $m_{ij} = Cov(\hat{\beta}_{i-1}(t), \hat{\beta}_{j-1}(t))$, $P = (p_{ij})_{i,j=1,\ldots,p+1}$ with $p_{ij} = 0$ for $i+j$ odd and $p_{ij} = \frac{\sqrt{(2j-1)(2l-1)}}{(j+l-1)}$ for $i+j$ even, $\kappa_{ij}(d) = \sqrt{(2i-1)(2l-1)}\frac{\Gamma(1-2d)}{[\Gamma(d)\Gamma(1-d)]}$, $Q = (q_{ij})_{i,j=1,\ldots,p+1}$ with*

$$q_{ij} = \kappa_{ij}(d)\int_{-1}^{1}\int_{-1}^{1} x^{i-1}y^{i-1}|x-y|^{2d-1}dxdy,$$

$D_N = (d_{ij}(N))_{i,j=1,\ldots,p+1}$ with $d_{ij} = 0$ for $i \neq j$ and $d_{jj} = \frac{2(Nb)^j}{(2j-1)}$. Then as $N \to \infty$, $b \to 0$, $Nb \to \infty$,

$$(2Nb)^{-2d}D_N M_N D_N \to 2\pi c_f P^{-1}QP^{-1}.$$

Proof: See Beran et al. (2001).

For the bias we obtain:

Theorem 5.8 *Denote with* $J(K) = \int_{-1}^{1} x^{(p+1)} K^*_{(0,p)}(x)dx$*, where* $K^*_{(0,p)}(x)$ *is the so-called equivalent kernel (see Beran/Feng (1999)). Let* $0 < \triangle < \frac{1}{2}$ *be a small positive number. Then*

$$E[\hat{f}^{(N)}(t) - f^{(N)}(t)] = b^{p+1} \frac{f^{(N)(p+1)}(t) J(K)}{k!} + o(b^{p+1})$$

uniformly in $\triangle < t < 1 - \triangle$.

Proof: See Beran et al. (2001).

Using these results we have for the asymptotic integrated mean squared error (IMSE)

$$\int_0^1 E\{[\hat{f}^{(N)}(t) - f^{(N)}(t)]^2\}dt \approx b^{2(p+1)} \frac{[f^{(N)(p+1)}(t)]^2 J^2(K)}{[(p+1)!]^2} + (Nb)^{2d-1} \int_0^1 v(t)dt. \quad (5.8)$$

Here $v(t)$ denotes the limit of the variance of the local polynomial estimator. For an explicit formula see Beran/Feng (1999). The bandwidth that minimizes the asymptotic IMSE is thus given by

$$b_{opt} = C_{opt} N^{(2d-1)/(2p+3-2d)},$$

where

$$C_{opt} = \left\{ \frac{(1-2d)[(p+1)!]^2 \int_0^1 v(t)dt}{2(p+1) J(f^{(N)(p+1)}) I^2(K)} \right\}^{1/(2p+3-2d)}.$$

Here $J(f^{(N)(p+1)}) = \int_0^1 [f^{(N)(p+1)}(t)]^2 dt$.

Note that the formula of the asymptotic IMSE is given on the interval $[0, 1]$, since a local polynomial estimator adapts automatically at the boundary. For further details concerning kernel estimators in the long-memory setup see also Beran et al. (2000).

The memory parameter as well as the parameters of the autoregressive part in SEMIFAR-models are estimated by Maximum Likelihood estimation. In the case of a constant mean function Maximum Likelihood estimation of the parameters is considered in Beran (1995). The same methodology can be used also in the case of non-constant trend functions. Starting with model (5.6) denote with $\theta^0 = (\sigma^2_{\epsilon,0}, d^0, \phi^0_1, \ldots, \phi^0_p)^T = (\sigma^2_{\epsilon,0}, \eta^0)^T$ the true unknown parameter vector. The process Y_i in (5.6) admits the infinite autoregressive representation

$$\sum_{j=0}^{\infty} a_j(\eta^0)[c_j(\eta^0)Y_{i-j} - g(t_{i-j})] = \epsilon_i. \tag{5.9}$$

Let now $(b_n)_{n\in\mathbf{N}}$ be a sequence of positive bandwidths with $b_n \to 0$ and $Nb_n \to \infty$ and denote with $K_b \diamond y(N) := \frac{1}{Nb}\sum_{i=1}^{N} K(\frac{t-t_i}{b})Y_i$, where $y(N) = (Y_1, \ldots, Y_N)$ and K is a kernel. Define

$$\hat{g}(t_i, 0) = K_{b_n} \diamond y(N)$$

and

$$\hat{g}(t_i, 1) = K_{b_n} \diamond Dy(N)$$

with $Dy(N) = (Y_2 - Y_1, Y_3 - Y_2, \ldots, Y_N - Y_{N-1})$. For a chosen value of $\theta = (\sigma^2_\epsilon, \eta)^T$ denote by

$$e_i(\eta) = \sum_{j=0}^{i-m-2} a_j(\eta)[c_j(\eta)Y_{i-j} - \hat{g}(t_{i-j}, m)]$$

the residuals and by $r_i(\theta) = e_i(\eta)/\sqrt{\theta_1}$ the standardized residuals.

Note that $(\epsilon_i(\eta^0))$ are independent zero mean normal with variance $\sigma^2_{\epsilon,0}$, an approximate Maximum Likelihood estimator of θ^0 is obtained by maximizing the approximate log-likelihood

$$l(Y_1, \ldots, Y_N, \theta) = -\frac{N}{2}\log 2\pi - \frac{N}{2}\log \sigma^2_\epsilon - \frac{1}{2}N^{-1}\sum_{i=m+2}^{N} r_i^2$$

with respect to θ. Denote with $f(x)$ the spectral density of the process (5.6). The specific form of the spectral density is not of interest here. It is only needed for the asymptotic properties of the estimator. We have in detail

Theorem 5.9 *Let $\hat{\theta}$ be the Maximum Likelihood estimator of θ and define $\theta_*^0 := (\sigma_{\epsilon,0}^2, \eta_*^0)^T = (\sigma_{\epsilon,0}^2, \delta^0, \eta_2^0, \ldots, \eta_{p+1}^0)^T$. This means that $\theta_2^0 = d = m^0 + \delta^0$ is replaced by $\theta_{2,*}^0 = \delta^0$. Then for $N \to \infty$*

(1) $\hat{\theta}$ converges in probability to the true value θ^0;

(2) $N^{1/2}(\hat{\theta} - \theta^0)$ converges in distribution to a normal random vector with mean zero and covariance matrix

$$\Sigma = 2D^{-1}$$

where

$$D_{ij} = (2\pi)^{-1} \left[\int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_i} \log f(x) \frac{\partial}{\partial \theta_j} \log f(x) dx \right] |_{\theta=\theta_*^0}.$$

Proof: see Beran et al. (1998).

To obtain an appropriate fit also the degree of the autoregressive polynomial has to be estimated. Thus the behaviour of model choice criteria such as the AIC has to be considered. Here the following holds:

Theorem 5.10 *Under the assumptions of the above theorem let p_0 denote the true order of Φ in (5.6) and define*

$$\hat{p} := \arg\min(AIC_\alpha(p), p = 0, 1, \ldots, L),$$

where L is a fixed integer, $AIC_\alpha(p) = N \log \hat{\sigma}_\epsilon^2(p) + \alpha p$ and $\hat{\sigma}_\epsilon^2(p)$ is the Maximum Likelihood estimate of the innovation variance $\sigma_{\epsilon,0}^2$ using a SEMIFAR model with autoregressive order p. Moreover $\hat{\theta}$ is the Maximum Likelihood estimator as defined above with p set equal to $\hat{p}$. Suppose furthermore that α is at least of the order $O(2c \log\log N)$ for some $c > 1$. Then the results of the above theorem hold.

Proof: See Beran et al. (1998).

This theorem says that consistency and asymptotic normality of the Maximum Likelihood estimator still hold when the autoregressive order is estimated. An algorithm for fitting SEMIFAR models to a time series can be found in Beran et al. (1998).

References

Andrews, D. W. K., Ploberger, W. (1994): "Optimal tests when a nuisance parameter is present only under the alternative." *Econometrica* **62**, 1383 - 1414.

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

Beran, J. (1995): " Maximum likelihood estimation of the differencing parameter for invertible short- and long-memory ARIMA models." *Journal of the Royal Statistical Society B* **57, No. 4**, 659 - 672.

Beran, J., Feng, Y. (1999): "Local polynomial fitting with long-memory errors." *Working paper, University of Konstanz.*

Beran, J., Feng, Y., Ocker, D. (1998): "SEMIFAR models" *Working paper, University of Konstanz.*

Beran, J., Ghosh, S., Sibbertsen, P. (2000): "Nonparametric M-estimation with long-memory errors." *Technical Report 36 / 00, SFB 475, University of Dortmund.*

Beran, J., Feng, Y., Ghosh, S., Sibbertsen, P. (2001): "On robust local polynomial estimation with long-memory errors." *International Journal of Forecasting, forthcoming.*

Bhattacharya, R. N., Gupta, V. K., Waymire, E. (1983): "The Hurst Effect under Trends." *Journal of Applied Probability* **20**, 649–662.

Bollerslev, T., Mikkelsen, H. O. (1996): "Modeling and pricing long memory in stock market volatility." *Journal of Econometrics* **73**, 151 - 184.

Brown, R. L., Durbin, J. and Evans, J. M. (1975): "Techniques for testing the constancy of regression relationships over time." *Journal of the Royal Statistical Society B* **37**, 149 - 163.

Csörgö, S., Mielniczuk, J. (1995): "Nonparametric regression under long-range dependent normal errors." *Annals of Statistics* **23**, 1000 - 1014.

Diebold, F. X., Inoue, A. (1999): "Long-memory and structural change." *Working paper.*

Engle, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of U. K. inflation." *Econometrica* **50**, 987 - 1008.

Fan, J., Gijbels, I. (1996): *Local polynomial modeling and its applications.* Chapman & Hall, London.

Giraitis, L., Kokoszka, P., Leipus, R. (2000a): "Testing for long memory in the presence of a general trend." *Working paper.*

Giraitis, L., Kokoszka, P., Leipus, R. (2000b): "Rescaled variance and related tests for long memory in volatilities and levels." *Working paper.*

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

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., Hyung, N. (1999): "Occasional structural breaks and long-memory." *Discussion paper 99-14, University of California, San Diego.*

Hidalgo, J. and Robinson, P.M. (1996): "Testing for structural change in a long–memory environment." *Journal of Econometrics* **70**, 159 - 174.

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

Krämer, W., Sibbertsen, P. (2000): "Testing for structural change in the presence of long-memory." *Technical Report 31 / 2000, SFB 475, University of Dortmund.*

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

Lohre, M., Sibbertsen, P. (2001): "Modeling the water flow of the Rhine River using seasonal long-memory." *Working Paper, Universität Dortmund.*

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

Mandelbrot, B. B., Wallis, J. R. (1969): Some long–run properties of geophysical records. *Water resources research* **5**, 321–340.

Ploberger, W. and Krämer, W. (1992): "The CUSUM–test with OLS–residuals." *Econometrica* **60**, 271 - 286.

Sibbertsen, P. (1999): *Robuste Parameterschätzung im linearen Regressionsmodell.* VWF, Berlin.

Sibbertsen, P. (2000): "Robust CUSUM-M-test in the presence of long-memory disturbances." *Technical Report 19 / 2000, SFB 475, University of Dortmund.*

Stock, J. H. (1994): "Unit roots and trend breaks." In R. F. Engle and D. Mc Fadden (eds) *Handbook of Econometrics, Volume IV*, North-Holland, Amsterdam.

Teverovsky, V., Taqqu, M. (1997): "Testing for long-range dependence in the presence of shifting means or a slowly declining trend using a variance-type estimator." *Journal of Time Series Analysis* **18**, 279 - 304.