Beran, Jan

Working Paper

SEMIFAR Models - A Semiparametric Framework for Modelling Trends, Long Range Dependence and Nonstationarity

CoFE Discussion Paper, No. 99/16

Provided in Cooperation with:
University of Konstanz, Center of Finance and Econometrics (CoFE)

Suggested Citation: Beran, Jan (1999) : SEMIFAR Models - A Semiparametric Framework for Modelling Trends, Long Range Dependence and Nonstationarity, CoFE Discussion Paper, No. 99/16, University of Konstanz, Center of Finance and Econometrics (CoFE), Konstanz, https://nbn-resolving.de/urn:nbn:de:bsz:352-opus-4114

This Version is available at:
https://hdl.handle.net/10419/85181

SEMIFAR models - a semiparametric framework for modelling trends, long-range dependence and nonstationarity

J. Beran
University of Konstanz
Department of Mathematics
and Computer Science
Universitätsstr. 10, Postfach 5560
78457 Konstanz, Germany

June 2, 1999

Abstract

Time series in many areas of application often display local or global trends. Typical models that provide statistical "explanations" of such trends are, for example, polynomial regression, smooth bounded trends that are estimated nonparametrically, and difference-stationary processes such as, for instance, integrated ARIMA processes. In addition, there is a fast growing literature on stationary processes with long memory which generate spurious local trends. Visual distinction between the large variety of possible models, and in particular between deteministic, stochastic and spurious trends, can be very difficult. Also, for some time series, several "trend generating" mechanisms may occur simulateneously. In this paper, a class of semiparametric fractional autoregressive models (*SEMIFAR*) is proposed that includes deterministic trends, difference stationarity and stationarity with short- and long-range dependence. Parameters characterizing stochastic dependence and stochastic trends, including a fractional and an integer differencing parameter, can be estimated by maximum likelihood. Deterministic trends are estimated by kernel smoothing. In combination with automatic model and bandwidth selection, the proposed method allows for flexible modelling of time series and helps the data analyst to decide whether the observed process contains a stationary short- or long-memory component, a difference stationary component, and/or a deterministic trend component. Data examples from various fields of application illustrate the method. Finite sample behaviour is studied in a small simulation study.

Key words: trend, differencing, long-range dependence, anti-persistence, difference stationarity, fractional ARIMA, Box-Jenkins ARIMA, BIC, kernel estimation, bandwidth selection, semiparametric models.

1 Introduction

Time series in many fields of statistical applications exhibit local or global trends. A large number of methods for dealing with trends under specific assumptions are described in the literature (see e.g. standard time series books, such as Diggle 1990, Priestley 1991 and references therein). Essentially, models for trends can be classified as either (1) deterministic or (2) stochastic. A deterministic trend is described by a deterministic function $g(t)$, whereas a stochastic trend is generated by a purely stochastic nonstationary process such as random walk, Brownian motion or an integrated ARIMA process. As a third possibility, local "spurious" trends can be generated by stationary processes with long-range dependence, such as stationary fractional ARIMA models (see below for the definition). For the practitioner, the large variety

of possible models can be quite puzzling, and finding the "right" model is often difficult. Some typical examples are displayed in figures 1a through d where four simulated series are generated respectively by a nonstationary process whose first difference is the sum of a deterministic trend plus a stationary process (Figure 1a), a deterministic trend plus a stationary long-memory process (Figure 1b), a nonstationary process whose first difference is stationary (Figure 1c), and a stationary long-memory process (Figure 1d). A full description of the models used in figures 1 a to d is given in section 7.

This paper grew out of the attempt to provide a unified approach that allows for simultaneous modelling of deterministic trends, stochastic trends and stationary short- and long-memory components. In particular, the approach helps the data analyst to decide which of these components are present in the observed data. The proposed model is a semiparametric generalization of a parametric class of stationary and nonstationary fractional autoregressive models introduced in Beran (1995). The *trend* of the process is modelled *nonparametrically* whereas the dependence structure of the *stochastic component* is modelled *parametrically*. The issue of distingushing long memory and difference stationary noise from a deterministic trend has also been addressed in a recent paper by Deo and Hurvich (1998). There, an alternative approach is taken in that a *parametric (linear) trend* is assumed whereas the fractional differencing parameter of the *noise process* is estimated *semi- or nonparametrically*. For testing stationarity with long-range dependence against trend plus noise also see Künsch (1986a).

A typical example where a decision between stochastic trend, deterministic trend or long-memory is essential, is the global temperture data in figures 2a through c. The foremost question with respect to the global temperature series is whether there is evidence for 'global warming'. Statistically, 'global warming' can be interpreted as an increasing deterministic trend. Fitting a linear trend, for instance, does indeed lead to a significant positive slope, even when taking into account the dependence structure in the residuals (see e.g. Smith 1993, Beran 1994, Deo and Hurvich 1998). It may be argued, however, that a) the assumption of a linear trend is arbitrary, and other (more realistic) trend models may not be significant, and b) the observed 'trend' may have been generated by a purely random (trend-free) mechanism, such as random walk, fractional Brownian motion (at discrete time points) or a stationary long-memory process. In particular, the second argument is often used when pointing out that the observational period is very short and, in the long run, the global temperature may go down again as part of the natural climatic process. The modeling approach introduced in this paper (so-called $SEMIFAR$ models) leads, however, to the conclusion that a significant deterministic trend is found even if the possibilities in a) (*arbitrary deterministic trend*) and b) (*stochastic 'trend' and/or long memory*) are *not exculded a priori*. Thus, $SEMIFAR$ models yield stronger evidence for global warming than more restrictive statistical models used in the past.

The paper is organized as follows. The model is defined in section 2. Results on nonparametric kernel estimation that are needed in the subsequent sections, are obtained in section 3. The method for estimating the trend and the parameters characterizing the stochastic component of the process is discussed in section 4, together with asymptotic results. A simple algorithm is proposed in section 5. Simulation results are presented in section 6. Several data examples illustrate the method in section 7. Final remarks are given in section 8. Proofs are given in the appendix.

2 The model

Recall that a Box-Jenkins ARIMA$(p, m, 0)$ process (Box and Jenkins 1976) is a Gaussian processes Y_i such that

$$\phi(B)\{(1-B)^m Y_i - \mu\} = \epsilon_i \tag{1}$$

holds, where $\mu \in R$, $m \geq 0$ is an integer, ϵ_i $(i = ..., -1, 0, 1, 2, ...)$ are iid zero mean normal with $\sigma_\epsilon^2 = \text{var}(\epsilon_i)$, B denotes the backshift operator such that $BY_i = Y_{i-1}$, and $\phi(x) = \sum_{j=0}^{p} \phi_j x^j$ is a polynomial with $\phi_o = 1$ and roots outside the unit circle. For $m > 1$, Y_i is not stationary, and the integer m is the number of times Y_i has to be differenced in order to achieve stationarity. The mth difference $(1-B)^m Y_i$ is a stationary autoregressive process of order p with expected value μ. Throughout the paper, we will focus on the values of $m \in \{0, 1\}$, and a process Y_i will be called *difference-stationary*, if Y_i is not stationary but its first difference $Y_i - Y_{i-1}$ is stationary in the second order sense.

On the other hand, Hosking (1981) and Granger and Joyeux (1980) define a fractional autoregressive process with fractional differencing parameter $\delta \in (-1/2, 1/2)$ (Granger and Joyeux 1980, Hosking 1981) to be a stationary process for which

$$\phi(B)(1-B)^\delta\{Y_i - \mu\} = \epsilon_i \tag{2}$$

holds. Here, the fractional difference $(1-B)^\delta$ is defined by

$$(1-B)^\delta = \sum_{k=0}^{\infty} b_k(\delta) B^k \tag{3}$$

with

$$b_k(\delta) = (-1)^k \frac{\Gamma(\delta+1)}{\Gamma(k+1)\Gamma(\delta-k+1)}. \tag{4}$$

The main motivation for introducing these processes was to model stationary time series with long-range dependence (or long memory) and to avoid the problem of overdifferencing which is often encountered in the usual Box-Jenkins setting. Here, long-range dependence is defined as follows (see, e.g. Mandelbrot 1983, Cox 1984, Hampel 1987, Künsch 1986b, and Beran 1994

and references therein): A stationary process Y_i with autocovariances $\gamma(k) = \text{cov}(Y_i, Y_{t+k})$ is said to have long-range dependence, if the spectral density $f(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \exp(ik\lambda)\gamma(k)$ has a pole at the origin of the form

$$f(\lambda) \sim c_f |\lambda|^{-\alpha} \ (|\lambda| \to 0) \tag{5}$$

for a constant $c_f > 0$ and $\alpha \in (0, 1)$, where " $\sim$ " means that the ratio of the left and right hand side converges to one. In particular, this implies that, as $k \to \infty$, the autocovariances $\gamma(k)$ are proportional to $k^{\alpha-1}$ and hence their sum is infinite. The spectral density of Y_i defined by (2) is proportional to $|\lambda|^{-2\delta}$ at the origin so that a fractional AR process has long-memory if $\delta > 0$. In contrast, for (1), the mth difference $(1-B)^m Y_i$ has short-range dependence (or short memory) in the sense that $\sum \gamma(k) < \infty$, and (5) holds with $\alpha = 0$ (i.e. $f(\lambda)$ converges to a positive constant at the origin). A unified class of processes that includes (1) and (2) as special cases can be defined as in Beran (1995) by

$$\phi(B)(1-B)^{\delta}\{(1-B)^m Y_i - \mu\} = \epsilon_i. \tag{6}$$

For $\delta = 0$, this reduces to (1), whereas for $m = 0$, we obtain stationary fractional autoregressive processes defined by (2). This definition generalizes traditional AR models by including an (unknown) fractional differencing parameter $\delta \in (-0.5, 0.5)$. On the other hand, (6) generalizes stationary fractional AR processes to nonstationarity by including the (unknown) integer differencing parameter m. The parameters of the process, including m and δ, can be estimated by maximum likelihood (Beran 1995). Note that, since m is an integer, m and δ correspond to one parameter $d = m + \delta$ only, through $m = [d + 0.5]$ and $\delta = d - m$, where $[.]$ denotes the integer part.

With respect to modelling trends, the model class defined by (6) includes stochastic trends (for $m > 0$) as well as "spurious" trends generated by long-range dependence ($m = 0$, $\delta > 0$). Moreover, if $m = 1$ and $\mu \neq 0$, then a polynomial trend is included. Thus, using (6), data are modelled either by a stationary process with *no trend* or an integrated process whose mth difference is stationary and that may also include a *polynomial trend* of order m. Excluded are, however, the possibility of a deterministic trend with stationary errors ($m = 0$) and other than polynomial trends. Thus, for instance, fitting (6) to a process that is in fact generated by a deterministic monotonous trend plus stationary noise will usually lead to the conclusion that m is equal to one, although in reality the stochastic part of the series *is* already stationary.

The idea of the model proposed below is thus to build in a trend component into (6) in order to include the following four possibilities:

- (a) no deterministic trend + stationary process with short- or long-range dependence;

- (b) deterministic trend + stationary process with short- or long-range dependence;
- (c) no deterministitc trend + difference-stationary process, whose first difference has short- or long-range dependence;
- (d) deterministic trend + difference-stationary process, whose first difference has short- or long-range dependence.

Model (6) includes (a) and (c). Moreover, (d) is also included, however with a polynomial trend only. In contrast, no assumptions on the trend function, except for general regularity assumptions, will be made in the following.

A number of results on nonparametric estimation of a bounded and sufficiently smooth trend function exist in the literature for the cases of short- or long-range dependent errors (see e.g. Chiu 1989, Altman 1990, Hall and Hart 1990 and Herrmann, Gasser and Kneip 1992 for short-memory errors; Hall and Hart 1990, Csörgö and Mielniczuk 1995, Ray and Tsay 1997 for long-memory errors). This motivates the following extension: Let $g : [0,1] \to R$ be a function such that g is twice differentiable. Also, denoting by $g^{(j)}$ the jth derivative, assume

$$\sup_{0<x<1} \max_{j=0,1,2} |g^{(j)}(x)| \leq C_1 < \infty$$

and

$$|g^{''}(x) - g^{''}(y)| \leq C_2 \cdot |x-y|^{\beta}$$

for all $x, y \in [0,1]$, constants $C_1, C_2 < \infty$, and some $\beta \in (2,3]$. Moreover, assume that for at least one $l \in \{0,1\}$, $g^{(l+1)}$ does not vanish in $[\Delta, 1-\Delta]$ ($0 < \Delta < 0.5$) and $g^{(l)}$ achieves an absolute maximum or minimum in $[\Delta, 1-\Delta]$. These conditions on g correspond to those in Ray and Tsay (1987) and Hall and Hart (1990). Also, let ϵ_i ($i = ..., -1, 0, 1, 2, ...$) be a sequence of iid zero mean normal random variables with $\sigma_\epsilon^2 = \mathrm{var}(\epsilon_i)$, and $\phi(x) = \sum_{j=0}^{p} \phi_j x^j$ a polynomial with $\phi_o = 1$ and roots outside the unit circle. Finally, let $BY_i = Y_{i-1}$ and use the convention $B^j g(t_i) = g(t_{i-j})$, where $t_i = i/n$. Then the following class of models is defined.

Definition 1 *A Gaussian process Y_i is called a semiparametric fractional autoregressive model (or SEMIFAR model), if there exists a smallest integer $m \in \{0,1\}$ such that*

$$\phi(B)(1-B)^{\delta}\{(1-B)^m Y_i - g(t_i)\} = \epsilon_i \tag{7}$$

where $\delta \in (-0.5, 0.5)$.

The following remarks clarify the meaning of this defintion:

1. For $g = \mu$, we obtain (6), and thus the cases (a) (if $m = 0$), (c) (if $\mu = 0$ and $m = 1$) and (d) with a linear trend (if $\mu \neq 0$ and $m = 1$).

2. If g is not constant, then we obtain (b) if $m = 0$ and (d) with a general smooth trend function if $m = 1$.

3. Note that for $m = 0$, $Y_i = g(t_i) + X_i$ where $t_i = i/n$ and X_i is a stationary fractional autoregressive process. Thus, Y_i is generated by a bounded continuous trend function and stationary noise (with at most "spurious" trends).

   On the other hand, if $m = 1$ and $g \not\equiv 0$, a trend in Y_i is generated simulateneously by a deterministic and a stochastic trend. More specifically, for $m = 1$, $Y_i - Y_{i-1} - g(t_i)$ is a stationary fractional autoregressive process and

$$E[Y_i - Y_1 | Y_1] = \sum_{j=2}^{i} g(t_j) \sim n \int_o^{t_i} g(u)du = nG(t_i).$$

   Thus, in general, the deterministic trend function in the observed process Y_i is of the order $O(n)$. Note that the variance $\mathrm{var}(Y_i - Y_1 | Y_1)$ also increases with n.

4. The restriction of m to the set $\{0, 1\}$ is not needed theoretically. However, in practise, m larger than 1 is unlikely to occur. Moreover, for $m = 1$, the trend function in the original process can be reconstructed, upto an additive constant by $nG(t_i)$, whereas this is no more possible for $m \geq 2$.

5. The normality assumption on the innovations ϵ_i can be relaxed to obtain non-Gaussian *SEMIFAR* models.

3 Nonparametric kernel estimation with independent, long-memory or anti-persistent errors

The problem of estimating g from data given by

$$Y_i = g(t_i) + X_i \tag{8}$$

has been considered by various authors for the case where the error process X_t is stationary with (i) short-range dependence, i.e. (5) holds with $\alpha = 0$ (see e.g. Chiu 1989, Altman 1990, Hall and Hart 1990 and Herrmann, Gasser and Kneip 1992) or (ii) long-range dependence, i.e. $0 < \alpha < 1$ (see e.g. Hall and Hart 1990, Csörgö and Mielniczuk 1995 and Ray and Tsay 1997). For *SEMIFAR* models defined by (7), the cases (i) and (ii) are obtained by

setting $m = 0$ and $\delta = \alpha/2 = 0$ (case (i)), or $m = 0$ and $\delta \in (0, 1/2)$ (case (ii)) respectively. For $m = 1$, the same is true for the first difference $Y_i - Y_{i-1}$. (Note, however, that for *SEMIFAR* models, $m \in \{0, 1\}$ is an unknown parameter.) In addition to cases (i) and (ii), definition (7) also includes the case where δ is negative so that the spectral density f of Y_i (or $Y_i - Y_{i-1}$ respectively) converges to zero at the origin. This case is sometimes called "anti-persistence". The theorem below extends previous results on kernel estimation to the anti-persistent case, and gives formulas for the mean squared error and the optimal bandwidth that are valid for the whole range $\delta \in (-0.5, 0.5)$.

Before stating the result, the following definitions are needed: For estimating g by kernel smoothing, symmetric polynomial kernels (see e.g. Gasser and Müller 1979) will be used, defined by

$$K(x) = \sum_{l=0}^{r} \alpha_l x^{2l}, \ (|x| \leq 1), \tag{9}$$

and $K(x) = 0$ if $|x| > 1$. Here $r \in \{0, 1, 2, ...\}$ and the coefficients α_l are such that $\int_{-1}^{1} K(x)dx = 1$. Note that considering polynomial kernels is not a serious restriction, since bandwith choice is known to be more important than the choice of the kernel and, also, other typical kernel functions can be approximated with arbitrary accuracy by polynomials. Moreover, at least in the case of independent errors, optimal kernels in the sense of Gasser and Müller (1979) are known to be polynomials.

If (8) holds, then, for a given bandwidth $b > 0$ and $t \in [0, 1]$, the kernel estimate of g is defined by

$$\hat{g}(t) = K_b \diamond y(n) = \frac{1}{nb} \sum_{i=1}^{n} K(\frac{t - t_i}{b}) Y_i \tag{10}$$

where $y(n) = (Y_1, ..., Y_n)$. Furthemore, the following notations will be used:

$$V_n(\theta, b) = (nb)^{-1-2\delta-2l_1-2l_2} \sum_{l_1, l_2=0}^{r} \alpha_{l_1} \alpha_{l_2} \sum_{i,j=-nb}^{nb} i^{2l_1} j^{2l_2} \gamma(i - j), \tag{11}$$

$$I(g^{''}) = \int_{\Delta}^{1-\Delta} [g^{''}(t)]^2 dt; \tag{12}$$

and

$$I(K) = \int_{-1}^{1} x^2 K(x) dx. \tag{13}$$

The following result is obtained under the assumption that (8) holds.

Theorem 1 *Let $b_n > 0$ be a sequence of bandwidths such that $b_n \to 0$ and $nb_n \to \infty$. Then, under the stated assumptions and δ in (7) in the interval (-0.5,0.5), we have*

(i) Bias:

$$E[\hat{g}(t) - g(t)] = b_n^2 \frac{g''(t) I(K)}{2} + o(b_n^2) \tag{14}$$

uniformly in $\Delta < t < 1 - \Delta$;

(ii)

$$\lim_{n \to \infty} V_n(\theta, b_n) = V(\theta) \tag{15}$$

where $0 < V(\theta) < \infty$ *is a constant;*

(iii) Variance:

$$(nb_n)^{1-2\delta} \mathrm{var}(\hat{g}(t)) = V(\theta) + o(1) \tag{16}$$

uniformly in $\Delta < t < 1 - \Delta$;

(iv) IMSE: The integrated mean squared error in $[\Delta, 1 - \Delta]$ *is given by*

$$\int_{\Delta}^{1-\Delta} E\{[\hat{g}(t) - g(t)]^2\} dt = IMSE_{asympt}(n, b_n) + o(\max(b_n^4, (nb_n)^{2\delta - 1}))$$

$$= b_n^4 \frac{I(g'') I^2(K)}{4} + (nb_n)^{2\delta-1} V(\theta) + o(\max(b_n^4, (nb_n)^{2\delta-1})) \tag{17}$$

(v) Optimal bandwidth: The bandwidth that minimizes the asymptotic IMSE is given by

$$b_{opt} = C_{opt}\ n^{(2\delta-1)/(5-2\delta)} \tag{18}$$

where

$$C_{opt} = C_{opt}(\theta) = [\frac{(1 - 2\delta) V(\theta)}{I(g'') I^2(K)}]^{1/(5-2\delta)}. \tag{19}$$

Similar results can be obtained for kernel estimates of derivatives of g. For instance, the second derivative can be estimated by $\hat{g}''(t) = n^{-1} b^{-3} \sum K((t_j - t)/b) Y_j$ where K is a symmetric polynomial kernel such that $\int K(x) dx = 0$ and $\int K(x) x^2 dx = 2$. By analogous arguments, the optimal bandwidth is then of the order $O(n^{(2\delta-1)/(9-2\delta)})$.

Simple explicit formulas for $V(\theta)$ can be given for $\delta = 0$ and $\delta > 0$ as follows (see e.g. Hall and Hart 1990):

$$V(\theta) = 2\pi c_f \int_{-1}^{1} K^2(x) dx,\ (\delta = 0), \tag{20}$$

$$V(\theta) = 2 c_f \Gamma(1 - 2\delta) \sin \pi\delta \int_{-1}^{1} \int_{-1}^{1} K(x) K(y) |x - y|^{2\delta - 1} dx dy,\ (\delta > 0). \tag{21}$$

These formulas can not be carried over directly to the case $\delta < 0$. In particular, the integral on the right hand side of (21) is infinite for $\delta < 0$. For specific kernels, simple closed form formulas that are valid for the whole range $\delta \in (-0.5, 0.5)$ can be obtained however by explicit calculation. For instance, for the box-kernel (i.e. $r = 0$), we obtain

Corollary 1 *Let* $K(x) = \frac{1}{2}1\{x \in [-1,1]\}$. *Define*

$$\nu(\delta) = \frac{2^{2\delta}\Gamma(1-2\delta)\sin(\pi\delta)}{\delta(2\delta+1)} \tag{22}$$

with $\nu(0) = \lim_{\delta\to 0}\ \nu(\delta) = \pi$. *Then, under the assumptions of theorem 1, we have*

(i) Bias:

$$E[\hat{g}(t) - g(t)] = b_n^2 \frac{g^{''}(t)}{6} + o(b_n^2); \tag{23}$$

(ii) Variance:

$$\mathrm{var}(\hat{g}(t)) = (nb_n)^{2\delta-1}\nu(\delta)c_f + o((nb_n)^{2\delta-1}); \tag{24}$$

(iii) IMSE:

$$\int_{\Delta}^{1-\Delta} E\{[\hat{g}(t) - g(t)]^2\}dt = b_n^4 \frac{I(g^{''})}{36} + (nb_n)^{2\delta-1}\nu(\delta)c_f + o(\max(b_n^4, (nb_n)^{2\delta-1})). \tag{25}$$

(iv) Optimal bandwidth:

$$b_{opt} = C_{opt}\ n^{(2\delta-1)/(5-2\delta)} \tag{26}$$

with

$$C_{opt} = [\frac{9(1-2\delta)\nu(\delta)c_f}{I(g^{''})}]^{1/(5-2\delta)} \tag{27}$$

4 Maximum likelihood estimation with estimated trend

Let $\theta^o = (\sigma^2_{\epsilon,o}, d^o,\ \phi^o_1, ..., \phi^o_p)^t = (\sigma^2_{\epsilon,o}, \eta^o)^t$ be the true unknown parameter vector in (7) where $d^o = m^o + \delta^o$, $-1/2 < \delta^o < 1/2$ and $m^o \in \{0,1\}$. For a constant function $g = \mu$, maximum likelihood estimation of θ^o, based on the autoregressive representation of the process, is considered in Beran (1995) and generalized to processes with GARCH components in Ling and Li (1997). This approach can be carried over directly to *SEMIFAR* models, since

$$\phi(B)(1-B)^{\delta^o}\{(1-B)^{m^o}Y_i - g(t_i)\} = \sum_{j=0}^{\infty} a_j(\eta^o)B^j[c_j(\eta^o)Y_i - g(t_i)]$$

$$= \sum_{j=0}^{\infty} a_j(\eta^o)[c_j(\eta^o)Y_{i-j} - g(t_{i-j})],$$

where the coefficients a_j and $a_j c_j$ are obtained by matching the powers in B. Hence, Y_i admits an infinite autoregressive representation

$$\sum_{j=0}^{\infty} a_j(\eta^o)[c_j(\eta^o)Y_{i-j} - g(t_{i-j})] = \epsilon_i. \tag{28}$$

Let b_n $(n \in N)$ be a sequence of positive bandwidths such that $b_n \to 0$ and $nb_n \to \infty$ and define $\hat{g}(t_i) = \hat{g}(t_i; m)$ by

$$\hat{g}(t_i; 0) = K_{b_n} \diamond y(n), \tag{29}$$

and

$$\hat{g}(t_i; 1) = K_{b_n} \diamond Dy(n), \tag{30}$$

with $Dy(n) = (Y_2 - Y_1, Y_3 - Y_2, ..., Y_n - Y_{n-1})$. Consider now ϵ_i as a function of η. For a chosen value of $\theta = (\sigma_\epsilon^2,\, m + \delta,\, \phi_1, ..., \phi_p)^t = (\sigma_\epsilon^2,\, \eta)^t$, denote by

$$e_i(\eta) = \sum_{j=0}^{i-m-2} a_j(\eta)[c_j(\eta)Y_{i-j} - \hat{g}(t_{i-j}; m)] \tag{31}$$

the (approximate) residuals and by $r_i(\theta) = e_i(\eta)/\sqrt{\theta_1}$ the standadized residuals. Assuming that $\{\epsilon_i(\eta^o)\}$ are independent zero mean normal with variance $\sigma_{\epsilon,o}^2$, an approximate maximum likelihood estimator of θ^o is obtained by maximizing the approximate log-likelihood

$$l(Y_1, ..., Y_n; \theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma_\epsilon^2 - \frac{1}{2}n^{-1}\sum_{i=m+2}^{n} r_i^2 \tag{32}$$

with respect to θ and hence by solving the equations

$$\dot{l}(Y_1, ..., Y_n; \theta) = 0 \tag{33}$$

where $\dot{l}$ is the vector of partial derivatives with respect to θ_j $(j = 1, ..., p+2)$. More explicitly, $\hat{\eta}$ is obtained by minimizing

$$S_n(\eta) = \frac{1}{n}\sum_{i=m+2}^{n} e_i^2(\eta) \tag{34}$$

with respect to η and setting

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n}\sum_{i=m+2}^{n} e_i^2(\hat{\eta}). \tag{35}$$

Note that the solution of (34) can also be obtained by solving

$$\sum_{i=m+2}^{n} [\frac{\partial}{\partial \eta_j} e_i(\eta)] e_i(\eta) = 0,\ \ j = 1, ..., p+1. \tag{36}$$

For the case where g is known to be constant, it follows from Beran (1995) and Ling and Li (1997) that, if the constant $g = \mu$ is estimated consistently, then, as $n \to \infty$, $\hat{\theta}$ converges in probability to θ^o, and $\sqrt{n}(\hat{\theta} - \theta^o)$ converges in distribution to a normal random variable with zero mean vector and covariance matrix equal to the inverse Fisher-Information matrix. This result can be extended to *SEMIFAR* models:

Theorem 2 *Let $\hat{\theta}$ be the solution of (34) and (35), and define $\theta_*^o = (\sigma_{\epsilon,o}^2, \eta_*^o)^T = (\sigma_{\epsilon,o}^2, \delta^o, \eta_2^o, ..., \eta_{p+1}^o)^T$. This means that, $\theta_2^o = d = m^o + \delta^o$ is replaced by $\theta_{2,*}^o = \delta^o$. Then, as $n \to \infty$,*

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

(ii) $n^{\frac{1}{2}}(\hat{\theta} - \theta^o)$ converges in distribution to a normal random vector with mean zero and covariance matrix

$$\Sigma = 2D^{-1} \tag{37}$$

where

$$D_{ij} = (2\pi)^{-1}[\int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_i} \log f(x) \frac{\partial}{\partial\theta_j} \log f(x) dx]|_{\theta=\theta_*^o} \tag{38}$$

Remarks:

1. It should be noted that in theorem 2, both, the fractional differencing parameter δ and the integer differencing parameter m are estimated from the data.

2. The asymptotic convariance matrix does not depend on m.

3. Theorem 2 can be generalized to the case where the innovations ϵ_i are not normal, and satisfy suitable moment conditions.

Theorem 2 is derived under the assumption that the order $p = p_o$ of the autoregressive polynomial in (7) is known. This can not be assumed in practise. Thus, p_o needs to be estimated by applying a suitaible model choice criterion. In a recent paper, Beran et al. (1998) showed that, for the case where g is equal to a constant μ, consistency properties of model choice criteria, such as the BIC (Schwarz 1978, Akaike 1979) and the HIC (Hannan and Quinn 1979), are analogous to the case of stationary short-memory autoregressive processes, provided that a consistent estimate of μ is used. By analogous arguments, theorem 2 can be extended to the case where p_o is estimated:

Theorem 3 *Under the assumptions of theorem 2, let p_o be the true order of the polynomial ϕ in (7) and define*

$$\hat{p} = \text{argmin}\{AIC_\alpha(p); p = 0, 1, ..., L\} \tag{39}$$

where L is a fixed integer, $AIC_\alpha(p) = n \log \hat{\sigma}_\epsilon^2(p) + \alpha \cdot p$ and $\hat{\sigma}_\epsilon^2(p)$ is the maximum likelihood estimate of the innovation variance $\sigma_{\epsilon,o}^2$ using a SEMI-FAR model with autoregressive order p. Moreover, define $\hat{\theta}$ by (34) and (35) 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 theorem 2 hold.

Remarks:

1. Under the assumption that the series is generated by a *SEMIFAR* process, the results in theorems 1, 2, 3, and corollary 1, can be used for testing a) whether there is a significant *deterministic trend* component; b) whether the random component of the process is *stationary* or *difference stationary* ($m = 0$ or $m = 1$); and c) whether there is *long-memory, short-memory or antipersistence* in the stationarized random component ($\delta > 0$, $\delta = 0$, $\delta < 0$). The formal tests (and confidence intervals) for b) and c) follow directly from theorems 2 and 3. For the trend component, testing is more difficult, since we are dealing with a function (instead of a finite number of parameters). If $\hat{m} = 0$, then we test the null hypothesis $H_o : g \equiv constant$. If the box-kernel is used, then an approximate pointwise acceptance region at significance level α is given by the horizontal lines $\bar{y} \pm z_{\alpha/2}(nb_n)^{\hat{\delta}-\frac{1}{2}}[\nu(\hat{\delta})c_f(\hat{\theta})]^{\frac{1}{2}}$. If $\hat{m} = 1$, then the null hypothesis is $H_o : g \equiv 0$ and an approximate pointwise acceptance region is given by the horizontal lines $\pm z_{\alpha/2}(nb_n)^{\hat{\delta}-\frac{1}{2}}[\nu(\hat{\delta})c_f(\hat{\theta})]^{\frac{1}{2}}$. One-sided acceptance regions can be defined analogously. Clearly, this test is only pointwise and can therefore only be used as a guideline. The construction of simulateneous acceptance bands is more complicated, and will be treated in detail in a forthcoming paper.

2. The proposed method can also be used in an exploratory way, in particular to identify the essential shape of a potential trend function g. Subsequently, a model with a suitable parametric trend can be fitted and a test for significance of the trend parameters can be performed. Such a test again has to be taken as a guideline only, because of the 'preselection' of the parametric model by the *SEMIFAR* fit.

5 Computational aspects

An algorithm for the case where g is assumed to be equal to a constant μ is given in Beran (1995). The algorithm makes use of the fact that d is the only additional parameter, in addition to the autoregressive parameters, so that a systematic search with respect to d can be made. This algorithm can be adapted to the case where g is an unknown function, by replacing $\hat{\mu}$ by a kernel estimate of g. The optimal bandwidth can be estimated by an iterative plugin method similar to the one in Herrmann, Gasser and Kneip (1992) and Ray and Tsay (1997). These authors consider the case of stationary errors, i.e. m is known to be equal to zero. The algorithm in Ray and Tsay (1997) is as follows: 1. an initial bandwidth is defined; 2. a preliminary estimate of g is computed and subtracted from the observations; 3. the relevant parameters of the error process are estimated from the residuals; 4. the bandwidth is updated. Steps 2 to 4 are repeated until the change in the bandwidth

is below a predefined threshold. This algorithm can readily be extended to fitting *SEMIFAR* models. A detailed study and comparison of iterative algorithms for *SEMIFAR* fitting will be given in a forthcoming paper.

6 Simulations

For $n = 200, 500$ and $d^o \in \{-0.3, 0, 0.3, 0.7, 1, 1.3\}$, $N = 100$ series of a *SEMIFAR* model with $p_o = 0, \sigma_\epsilon^2 = 1$ and $g(t) = 0.5(t-0.5)+0.5\exp[-100(t-0.5)^2]$ were generated. For each simulated series, p_o, g and θ^o were estimated using the BIC (evaluated for $p = 0, 1, 2, 3, 4, 5$) as described above. The following quantities in table 1 summarize the results:

1. N_1 =number of simulations with $\hat{m} = m^o$ and $\hat{p} = p_o$;
2. $\bar{d}$ =sample mean of simulated estimates $\hat{d}$;
3. $\tilde{d}$ =sample mean of simulated estimates with $\hat{m} = m^o$ and $\hat{p} = p_o$;
4. theoretical standard deviation of $\hat{d}$ obtained from the asymptotic formula $\sigma_{asympt} = \sqrt{\Sigma_{22}/n}$, where Σ is the covariance matrix of $\hat{\theta}$ defined in theorem 2;
5. s =simulated standard deviation of $\hat{d}$.
6. $\tilde{s}$ =simulated standard deviation of $\hat{d}$ for simulations with $\hat{m} = m^o$ and $\hat{p} = p_o$;
7. theoretical integrated mean squared error $IMSE_{asympt}$ obtained from the asymptotic formulas (25), (26) and (27);
8. simulated integrated mean squared error $IMSE_o$ obtained from the simulations with $\hat{m} = m^o$ and $\hat{p} = p_o$.

A comparison of $\bar{d}$, $\tilde{d}$, s and $\tilde{s}$ in table 1 shows that misspecification of p_o or m^o can lead to a serious bias and high variability in $\hat{d}$. Although asymptotically the probability of correctly estimating m^o and p_o tends to one for all values of d^o, p_o and m^o may be misspecified more often for small sample sizes, if there is strong long memory in the data. It should be noted at this point that, when assuming that the differencing parameter $d \in (0.5, 1.5)$ (and thus $m \in \{0, 1\}$ and $\delta \in (-0.5, 0.5)$) and the autoregressive order p_o are unknown, a sample size of 200 is rather small. For $n = 500$, N_1 is however already quite close to 100 even for $d = 1.3$. Also, for $n = 500$, the values of $\tilde{d}$ and $\tilde{s}$ are reasonably close to d^o and σ_{asympt} respectively. Note however that, in the presence of long memory ($\delta = 0.3$, i.e. $d = 0.3, 1.3$), convergence to the asymptotic results seems to be slower. For $n = 200$ and $\delta = 0.3$,

the simulated standard deviations are about twice as large as those given by the asymptotic formula. In constrast, for $\delta = -0.3$ the asymptotic formula appears to be applicable already for $n = 200$. Similar comments apply to the mean squared error.

7 Data examples

7.1 Temperature data for the northern hemisphere

Figure 2a displays, for the years 1856-1989 and the northern hemisphere, yearly averages of monthly deviations of the observed temperature from monthly averages obtained from the time period 1950-1979. These data are obtained from spatial averaging of temperatures measured over land *and* sea. Figure 2b displays the corresponding series for land mass measurements only, and figure 2c for the Atlantic. The series 2a and b seem to exhibit an increasing trend which is generally interpreted as "global" warming. This does not seem to be the case for the Atlantic data.

Fitting *SEMIFAR* models of orders $p = 0, 1, 2, 3, 4, 5$, the BIC turns out to have a distinct minimum at $p = 0$ for the first two data sets. The estimated parameters are: $\hat{d} = 0.33$ (95%-confidence interval $[0.19, 0.46]$) for land-and-sea temperatures, and $\hat{d} = 0.21$ ($[0.08, 0.34]$) for land only temperatures. The good quality of the fit is illustrated by the correlogram and normal probability plot of the residuals (figures 2d through g). Moreover, in both cases (land-and-sea and land data), the trend exceeds the 5% critical bounds.

Thus, the conjecture of global warming is supported. For the land-and-sea data, a significant trend was also found by Smith (1993) and Deo and Hurvich (1998). In these papers, a linear trend is fitted and tested for significance. In contrast, here, no assumptions on the shape of the trend are made. Moreover, the possibility of a nonstationary noise component with $m = 1$ is included whereas Smith (1993) assumes stationary noise a priori. This is an important extension, since trend like behaviour can also be generated by stochastic trends. In spite of the more general model used here, global warming for the land-and-sea and the land data turned out to be significant. This provides an even stronger argument for the existence of global warming.

Figures 2a and b also show that the significance of the trend is clearer for the land-only data. A possible reason can be seen by fitting a *SEMIFAR* model to the Atlantic series: Here, the smallest values of the BIC are achieved at $p = 1$ and $p = 0$, and in both cases the trend is not significant (at the 5% level). The quality of the fit is essentially the same for both choices of p, and, for $p = 1$, the 95%-confidence interval for ϕ_1 ($[-1.05, -0.77]$) includes the value -1, which would correspond to a unit root. Therefore, the model

with $p = 0$ is a natural choice. The estimated value of d is then equal to $\hat{d} = 0.54$ ($[0.40, 0.67]$). This means, in particular, that $\hat{m} = [\hat{d} + 1/2] = 1$, and the 95%-confidence set for m is $\{0, 1\}$. Thus, the Atlantic series may be nonstationary (in the sense of pure difference stationarity without trend), but the evidence is not significant formally at the 5% level. More importantly, however, there is no significant deterministic trend.

In conclusion, there is evidence for global warming in land mass temperatures, but no evidence for global warming in Atlantic temperatures. The evidence for global warming is stronger for land only temperatures as compared to land-and-sea data. The statistical reason is that land-and-sea data are obtained by spatial averaging of land and Atlantic data. Since the Atlantic series does not have any trend, the increasing trend in the land mass data is blurred by adding the Atlantic temperatures. From a climatological point of view, a smaller or no increase in temperatures over the oceans is plausible, since air pollution - which mainly takes place over land - is believed to be one of the main factors responsible for global warming. A further observation is that the estimated value of d for the land-and-sea data is higher than the one for the land only data. This is also caused by adding the Atlantic data which have a much higher value of d.

7.2 Exchange rate between German Mark and US dollar

Figure 3a displays the logarithm of the daily exchange rate between the German Mark (DM) and the US dollar, between September 1985 and August 1990 ($n = 1287$). More specifically, the logarithm of the value of 100 DM in US dollars, divided by a baseline value, is plotted. There has been some discussion in the recent literature about possible unit root behaviour or long memory in foreign exchange rates (see e.g. Cheung 1993, Liu and He 1991, and references therein). Deviations from random walk behaviour may have major implications for economic theory, since real markets may not be arbitrage-free (for a discussion, see e.g. Mandelbrot 1971, Rogers 1997). In view of this, it is interesting to see which hypothesis may be supported by fitting *SEMIFAR* models. Using the BIC, we obtain $\hat{p} = 0$, with $\hat{d} = 0.96$ and a 95%-confidence interval for d of $[0.91, 1.00]$. Thus, d appears to be slightly below 1 though the value of 1 (unit root) is just in the confidence interval. Moreover, there is an apparent deterministic trend function. For the difference, the estimated function $\hat{g}$ (see figure 3b) is almost always positive and exceeds the acceptance region (for a significance level of 5%), indicating a predominantly increasing trend in the original series. (The fitted curve is slightly wiggly, because the box-kernel was used here, and the first difference of the data shows high variability.) Almost no, or even a negative, trend can be observed between about observations 600 to 800. Compared to the

random variability, the trend in the differenced series may appear negligible (figure 3b). However, for the original data, it is cumulated so that the deterministic trend function is the dominating component (see figures 3a). The good fit of the model is demonstrated by figures 3c and d where the sample autocorrelations and the histogram of the residuals are displayed.

In conclusion, for the observed period, the daily DM/US $ exchange rate is described in good approximation by a process whose first difference consists of a deterministic trend plus a fractional autoregressive process with fractional differencing parameter $\delta = -0.04$. Since $d = m + \delta = 1$ is just at the border of the 95%-confidence interval, a simpler, and perhaps acceptable, model for the *stochastic part* of the first difference may be iid normal observations. Note that formal tests in Fong and Ouliaris (1995), reject the hypothesis of random walk (i.e. $d = 1$ and $p_o = 0$) for the DM/US $ exchange rate. Fong and Ouliaris conjecture that this may be due to long-range dependence. Our results suggest that rejection of the random walk hypothesis could instead be caused by the presence of a deterministic trend (which is another type of long memory) instead of a stochastic long-memory component.

7.3 Boston robberies

Figure 4a displays seasonally adjusted logarithms of monthly numbers of robberies in Boston between January 1966 - October 1975 (source: McCleary et al. 1980). A *SEMIFAR*-fit yields the following solution: $\hat{p} = 0$, $\hat{d} = 0.665$ ($[0.52, 0.81]$), and a significant upward trend (see figure 4a). Thus, over the years there was a significant nonrandom increase in the number of robberies. The fluctuations around this trend are nonstationary, but not random walk (unit root). Instead, the increments of the residual process are antipersistent. Figures 4c and d, with the sample autocorrelations and the histogram of the residuals, illustrate that the model fits the data reasonably well.

7.4 Tree rings

Figure 5a displays 10-year averaged of treering widths of a bristlecone pine in California between the years 4141 B.C. to 1963 A.D. (source C. W. Ferguson, E. Schulman, H.C. Fritts). Long memory in treerings has been reported by many authors (see in particular Mandelbrot 1983 and references therein). Some caution is however needed, since treerings may also have trends (e.g. age trends) that could be confounded with long-range depencence. The results obtained with the $SEMIFAR-$fit confirms however that, for this particular tree, a stationary fractional ARIMA(0,d,0) with $0 < d < 0.5$ (i.e. stationarity with long-range dependence) provides a good model. More specifically, the results are: $\hat{p} = 0$, $\hat{d} = 0.28$ ($[0.22, 0.35]$), and no significant

trend. Figures 5b and c with the sample autocorrelations and the histogram of the residuals confirm the good fit.

7.5 Simulated examples

The simulated series ($n = 200$) in figures 1a through d are:

Figure 1a: $Y_i - Y_{i-1} = g(t_i) + X_i$, where X_i are iid standard normal variables and $g(t) = -0.5(t - 0.5) - 0.5\exp[-100(t - 0.5)^2]$.

Figure 1b: $Y_i = g(t_i) + X_i$ where X_i is a fractional autoregressive process of order $p_o = 0$ with $d^o = 0.3$ and g is as above.

Figure 1c: $Y_i - Y_{i-1} = X_i$ where X_i is a fractional autoregressive process of order $p_o = 0$ with $d^o = -0.2$.

Figure 1d: $Y_i = X_i$ where X_i is a fractional autoregressive process of order $p_o = 1$ with $d^o = 0.4$ and $\phi_1^o = -0.5$.

The "error" series X_i were all generated by the *S-Plus* function *arima.fracdiff.sim*. Since a visual assessment of the time series plots appears to be difficult, it is interesting to see in how far the proposed method provides better information. The estimates $\hat{p}$ and $\hat{\eta} = (d, \phi_1, ..., \phi_p)$ together with 95%-confidence intervals, obtained by fitting *SEMIFAR* models for $p = 0, 1, 2, 3, 5$ and choosing p based on the BIC, are given in table 2. Also given is the estimated value of $m^o = [d^o + 0.5]$.

The values of $\hat{m}$ and $\hat{p}$ are correct for all four series. Thus, in particular, the method yields the correct answer to the question whether differencing is needed, i.e. whether the observed series has a stochastic trend component. Moreover, the estimates $\hat{\eta}$ are very close to the true values and the true values are always in the confidence intervals. Similarily, regarding the presence of a deterministic trend component, the results give correct indications. For the third and fourth series, the estimated trends turned out to be almost horizontal and close to zero. For the other two examples, the estimated trends are reasonably close to the true trend functions, in particular in view of the small data size of $n = 200$. Figure 6b displays the first difference of series 1a, together with the estimated (full line) and the true (dotted line) trend function. Figure 6a shows the same for the original (i.e. not differenced) series. Figure 6c shows simulated series 1b together with the estimated (full line) and the true (dotted line) trend.

8 Final remarks

In this paper, we introduced a semiparametric method for time series modelling that incorporates stochastic trends, deterministic trends, long-range dependence and short-range dependence. The trend function is modelled nonparametrically. In particular, the method helps the data analyst to answer the question which of these components are present in the observed series. How well the different components can be distinguished depends on the specific process and, in particular, on the shape of the trend function. Therefore, in order that the proposed method is effective in general, the observed series must not be too short. As any 'omnibus method' in statistics, it is recommended to use the method in conjunction with available subject specific knowledge. In cases where sufficient a priori knowledge about the type of trend (e.g. linear, exponential etc.) is available, parametric trend estimation should be used, since is likely to provide more accurate results. This can be done simply by replacing the general function g in (7) by the corresponding parametric function.

The most difficult part of the problem addressed here is the distinction between stochastic and deterministic trend without any a priori knowledge (except smoothness). It can, easily be demonstrated by simulations that, for instance, the sample path of a random walk process over a short time period very often ressembles the sample path of a monotonous deterministic trend plus noise. An additional source of uncertainty is long-range dependence that can generate spurious trends. The method proposed here *formalizes how difficult it is to separate these components*. It is therefore natural that for relatively short series, results based on *SEMIFAR* models show a high variability (see the simulation results above). This is not due to the specific estimation method, but rather reflects the *principal uncertainty* about the *decision between stochastic and deterministic trends, and between short and long memory*, in situations where no assumptions can be made about 1. the existence and shape of a possible deterministic trend and 2. stationarity or nonstationarity of the stochastic component.

Further refinements of the method, such as local bandwidth choice (see e.g. Brockmann 1993), bootstrap confidence intervals, faster algortihms (see Gasser et al. 1991) or other smoothing methods, etc., will be worth pursuing in future. A detailed study of several algorithms for fitting *SEMIFAR* models, including the one used here, is given in a forthcoming paper. Recent results on local polynomial smoothing for the case of stationary errors will be extended to the case where errors may be nonstationary.

Also, various extensions of *SEMIFAR* models are possible. For instance, as for classical ARIMA models, stochastic seasonal components can be included by multiplying the left hand side of (7) by a polynomial $\phi_{seas}(B) = \sum \phi_{j,seas} B^{sj}$ where $s \in N$ is the seasonal period. For instance, for monthly

data, s is typically equal to 12. Other extensions, such as inclusion of parametric and nonparametric explanatory variables, other seasonal components and nonlinearities in the stochastic part of the process, are subject of current research.

Finally, it should be noted that the acceptance regions used here for testing significance of a trend, are only pointwise regions. The problem of simultaneous acceptance bands is subject to current research and will be discussed elsewhere. At this point, the given acceptance bands can be used as rough guidelines. Also, the proposed method can be used in an exploratory way. It can be used in particular to identify the essential shape of a potential trend function g. Subsequently, a model with a suitable parametric trend may be used.

9 Acknowledgements

This research was supported in part by an NSF SBIR grant to MathSoft, Inc. (Seattle), and by the Center for Finance and Econometrics, University of Konstanz. Dirk Ocker and Yuanhua Feng (both University of Konstanz) helped me to obtain some of the data sets by internet. I would like to thank the authors of these data sets for making their data publicly available. Professor Richard Smith (University of North Carolina, Chapel Hill, USA) provided me with the temperature data from the data base of the Climate Research Unit of the University of East Anglia, Norwich, England. Finally, I would like to thank Don Percival (University of Washington and MathSoft) and MathSoft for generous computer support.

10 Appendix

Proof of theorem 1:

(i) The proof for the bias is standard.

(ii) and (iii): We have $K(x) = \sum_{l=0}^{r} \alpha_l x^{2l} 1\{x \in [-1,1]\}$, and

$$n^{1-2\delta}\mathrm{var}(\hat{g}(t)) = (nb)^{-1-2\delta} \sum_{l_1,l_2=0}^{r} \alpha_{l_1}\alpha_{l_2} \sum_{i,j=-nb}^{nb} (\frac{i}{nb})^{2l_1} (\frac{j}{nb})^{2l_2} \gamma(i-j)$$

$$= (nb)^{-1-2\delta} \sum_{l_1,l_2=0}^{r} \alpha_{l_1}\alpha_{l_2} V(l_1,l_2;nb).$$

Since r is fixed, it is sufficient to show for each l_1, l_2 separatly that

$n^{-1-2\delta}V(l_1,l_2;nb)$ converges to a finite value $V(l_1,l_2)\neq 0$. Now,

$$V(l_1,l_2;nb)=\sum_{s=0}^{2l_2}\binom{2l_2}{s}\sum_{i=-nb}^{nb}(\frac{i}{nb})^{2l_1+2l_2-s}\sum_{k=i-nb}^{i+nb}(-1)^s(\frac{k}{nb})^s\gamma(k)$$

$$=\sum_{s=0}^{2l_2}\binom{2l_2}{s}V_s(l_1,l_2;nb).$$

Again, it is sufficient to show that, for each s, $n^{-1-2\delta}V_s(l_1,l_2;nb)\to V_s(l_1,l_2)$ with $0<|V_s(l_1,l_2)|<\infty$. Note that, as $|k|\to\infty$, $\gamma(k)\sim c_\gamma|k|^{2\delta-1}$ where $c_\gamma>0$ is a constant. For $s\neq 0$, we then have

$$V_s(l_1,l_2;bn)\sim c_\gamma(nb)^\delta\sum_{i=-nb}^{nb}(\frac{i}{nb})^{2l_1+2l_2-s}\int_{i/nb-1}^{i/nb+1}x^s|x|^{2\delta-1}dx\sim c_\gamma n^{2\delta+1}\cdot const.$$

Moreover, for $s=0$, since $\sum_{k=-\infty}^{\infty}\gamma(k)=0$,

$$V_s(l_1,l_2;bn)=-\sum_{i=-nb}^{nb}(\frac{i}{bn})^{2l_1+2l_2}[\sum_{k=i+nb+1}^{\infty}\gamma(k)+\sum_{k=-\infty}^{i-nb-1}\gamma(k)]$$

$$\sim -c_\gamma n^\delta\sum_{i=-nb}^{nb}(\frac{i}{nb})^{2l_1+2l_2}[\int_{i/nb+1}^{\infty}x^{2\delta-1}dx-\int_{1-i/nb}^{\infty}x^{2\delta-1}dx]$$

$$\sim n^{2\delta+1}\frac{c_\gamma}{2\delta}\int_{-1}^{1}x^{2l_1+2l_2}[(1+x)^{2\delta}-(1-x)^{2\delta}]dx.$$

This concludes the proof of (ii) and (iii).

(iv) follows from (i), (ii) and (iii).

(v) follows from (iv) by maximimzing the asymptotic expression for the IMSE with respect to b.

Proof of corollary 1:

(i) follows from theorem 1.

(ii) For $\delta\geq 0$, the result follows by evaluating the asymptotic expressions (20) and (21) respectively.

For $\delta<0$, the result is obtained as follows:

Without loss of generality, we may assume $t=i_o/n$ for some integer i_o. Then, for n large enough, $\hat{g}(t)=(2nb)^{-1}\sum_{i=-nb}^{nb}Y_i$ and

$$\mathrm{var}(\hat{g}(t))=\frac{1}{4}(nb)^{-2}\sum_{k=-(2nb-1)}^{2nb-1}(2nb-|k|)\gamma(k)$$

$$= \frac{1}{2}(nb)^{-1} \sum_{k=-(2nb-1)}^{2nb-1} \gamma(k) - \frac{1}{4}(nb)^{-2} \sum_{k=-(2nb-1)}^{2nb-1} |k|\gamma(k).$$

Recalling that $\sum_{k=-\infty}^{\infty} \gamma(k) = 0$ and $\gamma(k) \sim c_\gamma |k|^{2\delta-1}$ where c_γ is a constant, we have

$$\sum_{k=-(2nb-1)}^{2nb-1} \gamma(k) = -2c_\gamma (nb)^{2\delta} \int_2^\infty x^{2\delta-1} dx = (nb)^{2\delta} c_\gamma \delta^{-1} 2^{2\delta}.$$

Moreover,

$$\sum_{k=-(2nb-1)}^{2nb-1} |k|\gamma(k) \sim 2c_\gamma (nb)^{2\delta+1} \int_o^2 x^{2\delta} dx = (nb)^{2\delta+1} c_\gamma (2\delta+1)^{-1} 2^{2\delta+2}.$$

Finally, note that $c_\gamma = 2c_f \Gamma(1-2\delta) \sin \pi\delta$ (see e.g. Beran 1994, p. 43). Putting these results together, we obtain

$$(nb)^{1-2\delta} \mathrm{var}(\hat{g}(t)) \to \frac{2^{2\delta} \Gamma(1-2\delta) \sin \pi\delta}{\delta(2\delta+1)} c_f.$$

(iii) and (iv) follow directly from (i) and (ii).

Proof of theorem 2:

(i) Let $\theta^o = (\sigma^2_{\epsilon,o}, d^o, \phi^o_1, ..., \phi^o_p)^t$ be the true parameter vector and, in particular, $m^o = [d^o + 0.5]$ and $\delta^o = d^o - m^o$ the true integer and fractional differencing parameter respectively. For the case where m^o is known, consistency of the maximum likelohood estimator of $\theta^o_* = (\sigma^2_{\epsilon,o}, \delta^o, \phi^o_1, ..., \phi^o_p)^t$ follows directly from the consistency of $\hat{g}$ and known results on maximum likelihood estimation for stationary fractional ARIMA processes (see e.g. Fox and Taqqu 1986, Giraitis and Surgailis 1990, Dahlhaus 1989, Yajima 1985). In theorem 2, m^o is assumed to be unknown. Consistency of $\hat{\theta}$ follows, if $\hat{m}$ can be shown to converge to m^o in probability.

We consider the two possible cases, $m^o = 0$ and $m^o = 1$ separatly:

Case 1 ($m^o = 0$) : Since $m^o = 0$, Y_i as well as $Y_i - Y_{i-1}$ are stationary. Hence, (29) converges in probability to $g(t_i)$ and (30) converges to $\lim_{n\to\infty} g(t_i) - g(t_{i-1}) = 0$. Consistency of $\hat{m}$ and hence of $\hat{\theta}$ then follows as in Beran (1995) and Ling and Li (1997).

Case 2 ($m^o = 1$) : The estimate (30) converges in probability to $g(t_i)$ and, $S_n(\eta)$ to a finite value $\sigma^2_\epsilon(\eta)$ which is minimal for $\eta = \eta^o$. On the other hand, for trial values of θ with $m = 0$, the variance of (29) does not converge to zero, and the variance of e_i defined by (31) and $S_n(\eta)$ diverge to infinity, in probability. Consistency then follows.

(ii) Since θ^o, and in particular m^o, are estimated consistently, the central limit theorem follows by Taylor expansion from known results on maximum likelihood estimation as in Beran (1995) and Ling and Li (1997).

Proof of theorem 3:

For $\alpha = O(2c \log\log n)$ with $c > 1$, we have $\hat{p} \to p$ in probability. The result then follows from theorem 2.

Akaike, H. (1979) A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. Biometrika, **26**, 237-242.

Altman, N.S. (1990) Kernel smoothing of data with correlated errors. J. Am. Statist. Assoc., **85**, 749-759.

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

Beran, J. (1995) Maximum likelihood estimation of the differencing parameter for invertible short- and long-memory ARIMA models. J. Roy. Statist. Soc. **B 57, No. 4**, 695-672.

Beran, J., Bhansali, R.J. and Ocker, D. (1998) On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. Biometrika, **Vol. 85, No. 4**, 921-934.

Box, G.E. and Jenkins, G.M. (1976) Time Series Analysis: Forecasting and Control. Holden Day, San Francisco.

Brockmann, M. (1993) Locally adaptive bandwidth choice for kernel regression estimators. J. Amer. Statist. Assoc., **88**, 1302-1309.

Cheung, Y.W. (1993) Long memory in foreign exchange rates. J. Bus. Econ. Statist., **11**, 93-101.

Chiu, S.T. (1989) Bandwidth selection for kernel estimates with correlated noise. Statist. Probab. Lett., **8**, 347-354.

Cox, D.R. (1984) Long-range dependence: a review. In: Statistics: An Appraisal. Proceedings 50th Anniversary Conference, H.A. David and H.T.David (eds.), pp. 55-74, The Iowa State University Press.

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

Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes. Ann. Statists **17** 1749-1766.

Deo, R.S., and Hurvich, C. M. (1998) Linear trend with fractionally integrated errors. J. Time Series Analysis, 19, No.4, 379–397.

Diggle, P.J. (1990) Time series - a biostatistical introduction. Oxford: Oxford University Press.

Fong, W.M. and Ouliaris, S. (1995) Spectral Tests of the Martingale Hypothesis for Exchange Rates. Journal of Applied Econometrics, Vol. 10, 255-271.

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

Gasser, T. and Müller, H.G. (1979) Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation*, T. Gasser and M. Rosenblatt (editors), New York: Springer, pp. 23-68.

Gasser, T., Kneip, A. and Köhler, W. (1991) A flexible and fast method for automatic smoothing. J. Amer. Statist. Assoc., **86**, 643-652.

Giraitis,L., Surgailis,D. (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its application to the asymptotic normality of Whittle's estimate. Probab. Th. Rel. Fields **86** 87-104.

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

Hall, P. and Hart, J. (1990) Nonparametric regression with long-range dependence. Stoch. Processes Appl., **36**, 339-351.

Hampel, F.R. (1987) Data analysis and self-similar processes. In: Proceedings of the 46th Session of ISI, Tokyo, Book 4, 235-254.

Herrmann, E., Gasser, T. and Kneip, A. (1992) Choice of bandwidth for kernel regression when residuals are correlated. Biometrika, **79**, 783-795.

Hannan, E.J. and Quinn, B.G. (1979) The determination of the order of an autoregression. J. Royal. Statist. Soc., **B 41**, 190-195.

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

Künsch, H (1986a) Statistical aspects of self-similar processes. Invited paper, Proc. First World Congress of the Bernoulli Society, Tashkent, Vol. 1, 67-74.

Künsch, H (1986b) Discrimination between monotonic trends and long-range dependence. J. Appl. Probab., 23, 1025-1030.

Ling, S.-Q. und Li, W.K. (1997). Fractional ARIMA-GARCH Time Series Models. JASA, **92**, 1184-1194.

Liu, C.Y. and He, J. (1991) A variance-ratio test of random walks in foreign exchange rates. Journal of Finance, **36**, 773-785.

Mandelbrot, B.B., 1971, When can price be arbitraged efficiently? A limit to the validity of the random walk and martingale models. Rev. Econ. Statist., LIII, 225-236.

Mandelbrot, B.B. (1983) The fractal geometry of nature. New York: Freeman.

McCleary, R., Hay, R.A., Jr, Meidinger, E.E., and McDowall, D. (1980) Applied time series analysis for the social sciences. London: Sage.

Priestley, M.B. (1981) Spectral analysis and time series. London: Academic Press.

Ray, B.K. and Tsay, R.S. (1997) Bandwidth selection for kernel regression with long-range dependence. Biometrika. (in press).

Rogers, L.C.G., 1997, Arbitrage with fractional Brownian motion. Mathematical Finance 7, No. 1, 95-105.

Schwarz, G. (1978) Estimating the dimension of a model. Ann. Statist., **6**, 461-464.

Smith, R. (1993) Long-ange dependence and global warming. In Statistics for the Environment (V. Barnett and K.F. Turkman, eds.), Wiley, New York, pp. 141-161.

Yajima Y. (1985) On estimation of long-memory time series models. Austral. J. Statist., **27**, 303-320.

Table 1: Simulated values of N_1, $\bar{d}$, $\tilde{d}$, s, $\tilde{s}$, *and* $IMSE_o$, *as defined in section 6. For comparison, also given are the standard deviation* σ_{asympt}, *and the integrated mean squared error* $IMSE_{asympt}$ *obtained from the asymptotic formulas. The results are based on* $N = 100$ *simulations of a SEMIFAR model with* $p_o = 0$ *where* $g(t) = 0.5(t - 0.5) + 0.5\exp[-100(t - 0.5)^2]$, $n = 200,\ 500,$ *and* $d^o = -0.3,\ 0,\ 0.3,\ 0.7,\ 1$ *and* 1.3 *respectively.*

d^o	n	N_1	$\bar{d}$	$\tilde{d}$	σ_{asympt}	s	$\tilde{s}$	$IMSE_{asympt}$	$IMSE_o$
-0.3	200	100	-0.34	-0.34	0.055	0.065	0.065	0.0078	0.0070
-0.3	500	100	-0.32	-0.32	0.035	0.041	0.041	0.0027	0.0027
0	200	96	-0.1	-0.08	0.055	0.121	0.103	0.0262	0.0209
0	500	98	-0.05	-0.05	0.035	0.065	0.056	0.0126	0.0121
0.3	200	83	0.12	0.23	0.055	0.249	0.112	0.2419	0.1574
0.3	500	95	0.26	0.28	0.035	0.142	0.050	0.1733	0.1085
0.7	200	93	0.68	0.71	0.055	0.112	0.058	0.0078	0.0082
0.7	500	99	0.69	0.69	0.035	0.048	0.035	0.0027	0.0026
1	200	95	0.89	0.91	0.055	0.133	0.097	0.0262	0.0203
1	500	96	0.93	0.95	0.035	0.144	0.053	0.0126	0.0103
1.3	200	85	1.08	1.23	0.055	0.421	0.104	0.2419	0.1667
1.3	500	87	1.14	1.26	0.035	0.374	0.054	0.1733	0.1175

Table 2: Estimates of p_o, d^o, ϕ_1^o (for $p_o = 1$), and $m^o = [d^o + 0.5]$ for the four simulated examples in figures 1a through d. The true values of p_o, m^o, d^o and ϕ^o are given in brackets. Also given are the 95%-confidence intervals for d^o and ϕ_1^o (if $p_o = 1$).

Fig.	$\hat{p}$ (p_o)	$\hat{m}$ (m^o)	$\hat{d}$ (d^o)	95%-C.I. for d^o	$\hat{\phi}_1$ (ϕ_1^o)	95%-C.I. for ϕ_1^o
Fig. 1a	0 (0)	1 (1)	1.05 (1)	[0.94, 1.15]	-	-
Fig. 1b	0 (0)	0 (0)	0.29 (0.3)	[0.18, 0.39]	-	-
Fig. 1c	0 (0)	1 (1)	0.67 (0.7)	[0.56, 0.77]	-	-
Fig. 1d	1 (1)	0 (0)	0.25 (0.4)	[-0.14, 0.64]	-0.67 (-0.5)	[-1,-0.30]