SEMIFAR models

Jan Beran[1], Yuanhua Feng[1], Dirk Ocker[2]
University of Konstanz
Department of Mathematics and Computer Science[1]
and
Department of Economics and Statistics[2]
Universitätsstr. 10, Postfach 5560
78457 Konstanz, Germany

August 17, 1998

SEMIFAR models

By Jan Beran, Yuanhua Feng and Dirk Ocker

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

Abstract

Recent results on so-called *SEMIFAR* models introduced by Beran (1997) are discussed. The nonparametric deterministic trend is estimated by a kernel method. The differencing- and fractional differencing parameters as well as the autoregressive coefficients are estimated by an approximate maximum likelihood approach. A data-driven algorithm for estimating the whole model is proposed based on the iterative plug-in idea for selecting bandwidth in nonparametric regression with long-memory. Prediction for *SEMIFAR* models is also discussed briefly. Two examples illustrate the potential usefulness of these models in practice.

Key words: trend, differencing, long-range dependence, difference stationarity, fractional ARIMA, BIC, kernel estimation, bandwidth, semiparametric models, forecasting.

1 Introduction

SEMIFAR (semiparametic fractional autoregressive) models introduced by Beran (1997) provide a modelling framework that enables us to separate and estimate deterministic and stochastic trends as well as short- or long-memory components in an observed time series. A *SEMIFAR* model is a fractional stationary or non-stationary autoregressive model with a nonparametric trend. This extends Box-Jenkins ARIMA models (Box and Jenkins 1976), by using a fractional differencing parameter $d > -0.5$, and by including a nonparametric trend function g. The trend function can be estimated by the well known kernel method (see e.g. Gasser and Müller 1979). The parameters may be estimated by an approximate maximum likelihood method proposed by Beran (1995). A data-driven algorithm for estimating

SEMIFAR models, which is a mixture of these two approaches, was proposed by Beran (1997).

This paper summarizes recent results on the estimation of *SEMIFAR* models and the application of these models. Most of the paper is based on results in Beran (1997). Some results from Beran and Ocker (1998) and in other preprints are also included. For proofs we refer to these papers.

2 The model

A *SEMIFAR* model is a Gaussian process Y_i with an existing smallest integer $m \in \{0, 1\}$ such that

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

where $t_i = (i/n)$, $\delta \in (-0.5, 0.5)$, g is a smooth function on $[0, 1]$, B is the backshift operator, $\phi(x) = 1 - \sum_{j=1}^{p} \phi x^j$ is a polynomial with roots outside the unit circle and ϵ_i $(i = ..., -1, 0, 1, 2, ...)$ are iid zero mean normal with var $(\epsilon_i) = \sigma_\epsilon^2$. Here, the fractional difference $(1-B)^\delta$ introduced by Granger and Joyeux (1980) and Hosking (1981) is defined by

$$(1-B)^{\delta} = \sum_{k=0}^{\infty} b_k(\delta) B^k \tag{2}$$

with

$$b_k(\delta) = (-1)^k \frac{\Gamma(\delta+1)}{\Gamma(k+1)\Gamma(\delta-k+1)}. \tag{3}$$

The main motivation for introducing fractional autoregressive models (Hosking 1981, Granger and Joyeux 1980) was to model stationary time series with long-range dependence (or long-memory) and to avoid the problem of overdifferencing. Here, long-range dependence is defined as follows (see, e.g. Mandelbrot 1983, Cox 1984, Hampel 1987, Künsch 1986, and Beran 1994 and references therein): A stationary process Y_i with autocovariances $\gamma(k) = \text{cov}(Y_t, Y_{t+k})$ 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

$$f(\lambda) \sim c_f |\lambda|^{-\alpha} \ (|\lambda| \to 0) \tag{4}$$

for a constant $c_f > 0$ and $\alpha \in (0, 1)$, where " $\sim$ " means that the ratio of the left and right hand sides 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. For SEMIFAR models, $Z_i = \{(1-B)^m Y_i - g(t_i)\}$ is a stationary fractional autoregressive process. Thus, the spectral density of Z_i is proportional to $|\lambda|^{-2\delta}$ at the origin so that the process $\{(1-B)^m Y_i - g(t_i)\}$ has long-memory if $\delta > 0$. If $\delta = 0$, Z_i has short-memory. 1 generalizes stationary fractional AR-processes to the nonstationary case, including difference stationarity and deterministic trend. Four special cases of model (1) are:

- (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 determinisitc 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.

3 Nonparametric kernel estimation of a trend with long-memory errors

The problem of estimating g from data given by

$$Y_i = g(t_i) + X_i \tag{5}$$

has been considered by various authors for the case where the error process X_t is stationary with (i) short-range dependence, i.e. (4) 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 (1), 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

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

For estimating g by kernel smoothing, symmetric polynomial kernels of the form $K(x) = \{\sum_{l=0}^{r} \alpha_l x^{2l}\} \mathbb{I}_{\{|x| \le 1\}}$ (see e.g. Gasser and Müller 1979) will be used. If (5) holds, then, for a given bandwidth $b > 0$ and $t \in [0, 1]$, the kernel estimate of g is defined by

$$\hat{g}(t) = K_b \diamond y(n) = \frac{1}{nb} \sum_{i=1}^{n} K(\frac{t - t_i}{b}) Y_i \tag{6}$$

where $y(n) = (Y_1, ..., Y_n)$. Let $n_0 = [nt]$, $n_1 = [nb]$ and $0 < \Delta < 0.5$, the following notations will be used:

$$V_n(\theta, b) = (nb)^{-1-2\delta} \sum_{i,j=n_0-n_1}^{n_0+n_1} K\left(\frac{t - t_i}{b}\right) K\left(\frac{t - t_j}{b}\right) \gamma(i - j), \tag{7}$$

$$I(g'') = \int_{\Delta}^{1-\Delta} [g''(t)]^2 dt \tag{8}$$

and

$$I(K) = \int_{-1}^{1} x^2 K(x) dx. \tag{9}$$

The following result is obtained under the assumption that (5) holds and that g is at least twice continuously differentiable.

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 (1) in the interval (-0.5,0.5), we have*

(i) Bias:

$$E[\hat{g}(t) - g(t)] = b_n^2 \frac{g''(t) I(K)}{2} + o(b_n^2) \tag{10}$$

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

(ii)

$$\lim_{n \to \infty} V_n(\theta, b_n) = V(\theta) \tag{11}$$

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

(iii) Variance:

$$(nb_n)^{1-2\delta}\text{var}(\hat{g}(t)) = V(\theta) + o(1) \qquad (12)$$

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

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

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

$$= b_n^4 \frac{I(g'')I^2(K)}{4} + (nb_n)^{2\delta-1}V(\theta) + o(\max(b_n^4, (nb_n)^{2\delta-1})) \qquad (13)$$

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

$$b_{opt} = C_{opt}\ n^{(2\delta-1)/(5-2\delta)} \qquad (14)$$

where

$$C_{opt} = C_{opt}(\theta) = \left(\frac{(1-2\delta)V(\theta)}{I(g'')I^2(K)}\right)^{1/(5-2\delta)}. \qquad (15)$$

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 $\int K(x)x^2dx = 2$ .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):

$$V(\theta) = 2\pi c_f \int_{-1}^{1} K^2(x)dx,\ (\delta = 0)\ , \qquad (16)$$

$$V(\theta) = 2c_f\Gamma(1-2\delta)\sin\pi\delta \int_{-1}^{1}\int_{-1}^{1} K(x)K(y)|x-y|^{2\delta-1}dxdy,\ (\delta > 0). \qquad (17)$$

In order to obtain similar formula for $\delta < 0$, at a point x let $K(y) = \sum_{l=0}^{r}\beta_l(x)(x-y)^l =: K_0(x) + K_1(x-y)$, where $K_0(x) = \beta_0(x)$, $K_1(x-y) = \sum_{l=1}^{r}\beta_l(x)(x-y)^l$. Then we have

$$V(\theta) = 2c_f\Gamma(1-2\delta)\sin(\pi\delta)\int_{-1}^{1} K(x) \times \left\{\int_{-1}^{1} K_1(x-y)|x-y|^{2\delta-1}dy - \int_{|y|>1} K_0(x)|x-y|^{2\delta-1}dy\right\}dx \qquad (18)$$

for $\delta < 0$. For the box-kernel (i.e. $r = 0$), formulas (16), (17) and (18) give the same result

$$V = \frac{2^{2\delta}c_f\Gamma(1-2\delta)\sin(\pi\delta)}{\delta(2\delta+1)} \qquad (19)$$

with $V(0) = \lim_{\delta\to 0} V(\delta) = \pi c_f$ (see corollary 1 in Beran, 1997).

4 Maximum likelihood estimation

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 (1) where $d^o = m^o + \delta^o$, $-1/2 < \delta^o < 1/2$ and $m^o \in \{0{,}1\}$. The maximum likelihood estimation of θ^o proposed by Beran (1995) for a constant function $g = \mu$ 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_jc_j are obtained by matching the powers in B. Hence, Y_i admits an infinite autoregressive representation

$$\sum_{j=0}^{\infty} a_j(\eta^o)[c_j(\eta^o)Y_{i-j} - g(t_{i-j})] = \epsilon_i. \tag{20}$$

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

$$\hat{g}(t_i; 0) = K_{b_n} \diamond y(n), \tag{21}$$

and

$$\hat{g}(t_i; 1) = K_{b_n} \diamond Dy(n), \tag{22}$$

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 $\theta = (\sigma^2_\epsilon, m + \delta, \phi_1, ..., \phi_p)^T = (\sigma^2_\epsilon, \eta)^T$, denote by

$$e_i(\eta) = \sum_{j=0}^{i-m-2} a_j(\eta)[c_j(\eta)Y_{i-j} - \hat{g}(t_{i-j}; m)] \tag{23}$$

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 $\sigma^2_{\epsilon,o}$, an approximate maximum likelihood estimator of θ^o is obtained by maximizing the approximate log-likelihood

$$l(Y_1, ..., Y_n; \theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2_\epsilon - \frac{1}{2}n^{-1}\sum_{i=m+2}^{n} r_i^2 \tag{24}$$

with respect to θ and hence by solving the equations

$$\dot{l}(Y_1, ..., Y_n; \theta) = 0 \tag{25}$$

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

$$S_n(\eta) = \frac{1}{n} \sum_{i=m+2}^{n} e_i^2(\eta) \tag{26}$$

with respect to η and setting

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n} \sum_{i=m+2}^{n} e_i^2(\hat{\eta}). \tag{27}$$

The result in Beran (1995) can be extended to *SEMIFAR* models:

Theorem 2 *Let $\hat{\theta}$ be the solution of (26) and (27), 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

$$\Sigma = 2\, D^{-1} \tag{28}$$

where

$$D_{ij} = (2\pi)^{-1} \left[\int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_i} \log f(x) \frac{\partial}{\partial \theta_j} \log f(x) dx \right] |_{\theta=\theta_*^o}. \tag{29}$$

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. The asymptotic convariance matrix does not depend on m. 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 (1) is known. In practise 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 (1) and define*

$$\hat{p} = \arg\min\{AIC_\alpha(p); p = 0, 1, ..., L\} \tag{30}$$

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 SEMIFAR model with autoregressive order p. Moreover, define $\hat{\theta}$ by (26) and (27) 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.

5 A data-driven algorithm

The following algorithm is an adaptation of that in Beran (1995) by replacing $\hat{\mu}$ by a kernel estimate of g. This 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. The optimal bandwidth is estimated by an iterative plugin method similar to the one in Herrmann, Gasser and Kneip (1992) and Ray and Tsay (1997). The steps of the algorithm are defined as follows:

Step 1: Define L =maximal order of $\phi(B)$ that will be tried, and a sufficiently fine grid $G \in (-0.5, 1.5)$. Then, for each $p \in \{0, 1, ..., L\}$, carry out steps 2 through 4.

Step 2: For each $d \in G$, set $m = [d + 0.5]$, $\delta = d - m$, and $U_i(m) = (1 - B)^m Y_i$, and carry out step 3.

Step 3: Carry out the following iteration:

Step 3a: Let $b_o = \Delta_o \min(n^{(2\delta-1)/(5-2\delta)}, 0.5)$ with $0 < \Delta_o < 1$ and set $j = 1$.

Step 3b: Set $b = b_{j-1}$.

Step 3c: Calculate $\hat{g}(t_i; m)$ using the bandwidth b. Set $\hat{X}_i = U_i(m) - \hat{g}(t_i; m)$.

Step 3d: Set $\tilde{e}_i(d) = \sum_{j=0}^{i-1} b_j(\delta) \hat{X}_{i-j}$, where the coefficients b_j are defined by (4).

Step 3e: Estimate the autoregressive parameters $\phi_1, ..., \phi_p$ from $\tilde{e}_i(d)$ and obtain the estimates $\hat{\sigma}_\epsilon^2 = \hat{\sigma}_\epsilon^2(d; j)$ and $\hat{c}_f = \hat{c}_f(j)$. Estimation of the parameters can be done, for instance, by using the *Splus* functions *ar.burg* or *arima.mle*. If $p = 0$,set $\hat{\sigma}_\epsilon^2$ equal to $n^{-1} \sum \tilde{e}_i^2(d)$ and $\hat{c}_f$ equal to $\hat{\sigma}_\epsilon^2/(2\pi)$.

Step 3f: Set $b_2 = b^{(5-2\delta)/(9-2\delta)}$ and estimate g'' by

$$\hat{g}''(t) = \frac{1}{nb_2^3} \sum_{j=1}^{n} \tilde{K}(\frac{t_j - t}{b_2}) U_j(m)$$

where $\tilde{K} : R \to R$ is a polynomial symmetric kernel such that $\tilde{K}(x) = 0$ for $|x| > 1$, $\int \tilde{K}(x)dx = 0$ and $\int \tilde{K}(x)x^2 dx = 2$ .Calculate $I(\hat{g}'')$.

Step 3g: Calculate V and C_{opt} from δ and the estimated parameters obtained in Step 3f. Set

$$b_j = C_{opt}\ n^{(2\delta-1)/(5-2\delta)}.$$

Step 3h: Increase j by one and repeat steps 3b through 3g 4 times. This yields, for each $d \in G$ separatly, the ultimate value of $\hat{\sigma}_\epsilon^2(d)$, as a function of d.

Step 4: Define $\hat{d}$ to be the value of d for which $\hat{\sigma}_\epsilon^2(d)$ is minimal. This, together with the corresponding estimates of the AR parameters, yields $AIC_\alpha(p)$ (as a function of p) and the corresponding values of $\hat{\theta}$ and $\hat{g}$ for the given order p.

Step 5: Select the order p that minimizes $AIC_\alpha(p)$. This yields the final estimates of θ and g.

The factor $(5-2\delta)/(9-2\delta)$ in step 3f inflates the bandwidth b to a bandwidth b_2, which is optimal for estimating g'' in the case of $\delta = \delta^o$. The estimated parameters, the selected bandwidth $\hat{b}$ as well as the estimated trend $\hat{g}(t)$, $t \in [0, 1]$, by the above algorithm are all consistent. Denote by b_M the true optimal bandwidth that minimizes the IMSE, then we have:

Theorem 4 *Assume that the conditions of theorems 1 to 3 hold and that g is at least four times continuously differentiable, then*

(i) the results for $\hat{\theta}$ as given in theorem 2 hold,

(ii)

$$\hat{b} = b_M\{1 + O_p(n^{2(2\delta^o-1)/(9-2\delta^o)})\}, \tag{31}$$

(iii)

$$\hat{g}(t_i) = g(t_i)\{1 + O_p(n^{2(2\delta^o-1)/(5-2\delta^o)})\}. \tag{32}$$

6 *SEMIFAR* forecasting

Let $Y_1, ..., Y_n$ be observations generated by a *SEMIFAR* model of order p with parameter vector $\theta = (\sigma_\epsilon^2, d, \phi_1, ..., \phi_p)^T$ (where $d = m + \delta$). The aim is to predict a future observation Y_{n+k} for some $k \in \{1, 2, 3, ...\}$. Denote by X_i a zero mean fractional AR process of order p with parameter vector $\theta_* = (\sigma_\epsilon^2, \delta, \eta_2, ..., \eta_{p+1})^T$, and define $t_{n+k} = (n+k)/n = t_n + k/n$. Then

$$Y_{n+k} = \mu(t_{n+k}) + U_{n+k} \tag{33}$$

with

$$\mu(t_{n+k}) = g(t_{n+k}), \quad U_{n+k} = X_{n+k} \tag{34}$$

if $m = 0$,and

$$\mu(t_{n+k}) = Y_n + \sum_{j=1}^{k} g(t_{n+j}), \quad U_{n+k} = \sum_{j=1}^{k} X_{n+j} \tag{35}$$

if $m = 1$. Thus, to predict Y_{n+k} from $Y_1, ..., Y_n$, two problems need to be solved:

1. *extrapolation* of the function $\mu(t)$ to $t = t_{n+k}$;
2. *prediction* of the stochastic component U_{n+k}.

Since for *SEMIFAR* models only general regularity conditions on g are imposed, the deterministic trend $g(t)$ may behave in an arbitrary way in the future. This is in contrast to parametric trend models. However, we may obtain the predictions of $\hat{g}(t_{n+j})$ for $j \in \{1, 2, ..., k\}$ by a local constant or a local linear extesion of $\hat{g}(t_n)$. $\hat{\mu}(t_{n+k})$ is obtained by inserting $\hat{g}(t_{n+k})$ in (34) or $\hat{g}(t_{n+j})$ for $j \in \{1, 2, ..., k\}$ in (35).

Note that $X_i = U_i = Y_i - g(t_i)$ for $m = 0$,and $X_i = U_i - U_{i-1} = Y_i - Y_{i-1} - g(t_i)$ for $m = 1$ .Let $\gamma(k) = \text{cov}(X_i, X_{i+k})$ denote the autocovariances of X_i. Using the mean square criterion, the best linear predictor of U_{n+k} based on $Y_1, ..., Y_n$ is defined by $\hat{U}_{n+k} = \beta_{opt}^T X(n)$ where $X(n) = (X_1, ..., X_n)^T$ and the vector $\beta_{opt} = (\beta_1, ..., \beta_n)^T$ minimizes the mean squared prediction error $MSE = E[(U_{n+k} - \hat{U}_{n+k})^2]$. The values of β_{opt} and the corresponding optimal mean squared prediction error MSE_{opt} are given by

Theorem 5 *For all integers $r, s > 0$, define*

$$\gamma_r^{(s)} = [\gamma(r+s-1), \gamma(r+s-2), ..., \gamma(r)]^T, \tag{36}$$

$$\tilde{\gamma}_k^{(n)} = \sum_{j=1}^{k} \gamma_j^{(n-1)}, \tag{37}$$

and denote by $\Sigma_n = [\gamma(i-j)]_{i,j=1,...,n}$ the covariance matrix of $X(n)$. Then, the following holds.

i) If $m = 0$,

$$\beta_{opt} = \Sigma_n^{-1} \gamma_k^{(n)}, \tag{38}$$

$$MSE_{opt} = \gamma(0) - [\gamma_k^{(n)}]^T \Sigma_n^{-1} [\gamma_k^{(n)}]; \tag{39}$$

ii) If $m = 1$,

$$\beta_{opt} = \Sigma_n^{-1} \tilde{\gamma}_k^{(n)}, \tag{40}$$

$$MSE_{opt} = \sum_{s=-(k-1)}^{k-1} (k - |s|)\gamma(s) - [\tilde{\gamma}_k^{(n)}]^T \Sigma_n^{-1} [\tilde{\gamma}_k^{(n)}]. \tag{41}$$

Note in particular that, as $k \to \infty$,the MSE tends to a finite constant in the case of a stationary stochastic component ($m = 0$), whereas it diverges to infinity in the case of a nonstationary stochastic component ($m = 1$) .More specifically we have

Corollary 1 *Define $c_f = \lim_{\lambda \to 0} |\lambda|^{2\delta} f(\lambda)$ where f is the spectral density of X_i, and let*

$$\nu(\delta) = \frac{2\Gamma(1-2\delta)\sin \pi\delta}{\delta(2\delta+1)} \tag{42}$$

for $0 < |\delta| < 0.5$ and $\nu(0) = \lim_{\delta \to 0} \nu(\delta) = 2\pi$. Then, as $k \to \infty$,the following holds:

i) If $m = 0$,

$$MSE_{opt} \to \gamma(0) = \text{var}(X_i); \tag{43}$$

ii) If $m = 1$,

$$MSE_{opt} \sim c_f \nu(\delta) k^{1+2\delta}. \tag{44}$$

Moreover, for known values of g and θ a $(1-\alpha)100\%-$prediction interval for Y_{n+k}, is given by

$$\hat{Y}_{n+k} \pm z_{\alpha/2}\sqrt{MSE_{opt}} \tag{45}$$

where $\hat{Y}_{n+k} = \mu(t_{n+k} + \beta_{opt}^T X(n)$ and the values of β_{opt} and MSE_{opt} are obtained from theorem 1. If g and θ are estimated, the quantities in (45) are replaced by the corresponding estimated quantities.

7 Data examples

7.1 Exchange rate German Mark/US dollar

Figure 1a 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). 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 $\hat{g}$ (figure 1c) is almost always positive, indicating a predominantly increasing trend in the original series. 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 1b). However, for the original data, it is cumulated so that the deterministic trend function is the dominating component (see figures 1a and d). Note however that no formal test was applied here to test for significance of the trend. Formal procedures for doing so are currently under investigation. The good fit of the model is demonstrated by figures 1e and f 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 may be caused by the presence of a (slight) deterministic trend (which is another type of long memory) instead of a stochastic long-memory component.

7.2 Temperature data for the northern hemisphere

Figure 2a displays, for the years 1854-1989 and the northern hemisphere, yearly averages of monthly deviations of the observed temperature from monthly averages obtained from the time period 1950-1979. The series seems to exhibit an increasing S-shaped trend which is generally understood as "global" warming. The question arises, whether, instead of a deterministic trend (global warming), this increase may be explained by a stochastic or spurious trend.

Fitting *SEMIFAR* models of orders $p = 0$,$1, 2, 3, 4, 5$, the BIC turns out to have a distinct minimum at $p = 0$ .The satisfactory autocorrelations and histogram of the residuals (figures 2b and c), confirm that $p = 0$ provides a reasonable fit. The estimated function g is an S-shaped function and the estimated value of d is 0.27 ($[0.14, 0.41]$), indicating stationarity with long-range dependence in the stochastic part of the process. In conclusion, within the given framework, the most plausible model for the temperature data appears to be an S-shaped deterministic trend that shows a pronounced increase in the middle part, plus stationary long-memory deviations. This supports, the conjecture of global warming.

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. Estimation of these models was discussed in details. 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. In cases where one has sufficient a priori knowledge about the type of trend (e.g. linear, exponential etc.), parametric trend estimation is likely to provide more accurate results. This can be done simply by replacing the general function g in (1) by the corresponding parametric function.

Further refinements of the method, such as local polynomial fitting of g, 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. 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 (1) by a polynomial $\phi_{seas}(B) = \sum \phi_{j,seas} B^{sj}$ where $s \in N$ is the seasonal period. For example, 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.

9 Acknowledgements

This research was supported in part by an NSF Phase I SBIR grant to MathSoft and inpart by the Center for Finance and Econometrics at the University of Konstanz. We would like to thank Professor Richard Smith (University of North Carolina, Chapel Hill, USA) for providing us with the temperature data from the data base of the Climate Research Unit of the University of East Anglia, Norwich, England. The exchange rate data were extracted from the data base of the Journal of Applied Econometrics.

BIBLIOGRAPHY

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. (1997) Estimating trends, long-range dependence and nonstationarity. Preprint.

Beran, J., Bhansali, R.J. and Ocker, D. (1998) On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. To appear in Biometrika.

Beran, J. and Ocker, D. (1998). SEMIFAR forecasts, with applications to foreign exchange rates. To appear in Journal of Statistical Planning and Inference.

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.

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

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.

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.

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

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

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

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

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

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

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