SEMIFAR models, with applications to commodities, exchange rates and the volatility of stock market indices

Jan Beran[a,*], Yuanhua Feng[a], Günter Franke[b], Dieter Hess[b], Dirk Ocker[c]

[a]Department of Mathematics and Computer Science

[b]Department of Economics and Statistics

University of Konstanz, D-78457 Konstanz, Germany

[c]Schweizer Verband der Raiffeisenbanken, CH-9001 St.Gallen

Abstract

The distinction between stationarity, difference stationarity, deterministic trends as well as between short- and long-range dependence has a major impact on statistical conclusions, such as confidence intervals for population quantities or point and interval forecasts. In this paper, recent results on so-called *SEMIFAR* models introduced by Beran (1999) are summarized and their potential usefulness for economic time series analysis is illustrated by analyzing several commodities, exchange rates, the volatility of stock market indices and some simulated series. *SEMIFAR* models provide a unified approach that allows for simultaneous modelling of and distinction between deterministic trends, difference stationarity and stationarity with short- and long-range dependence. An iterative data-driven algorithm combines MLE and kernel estimation. Predictions combine stochastic prediction of the random part with functional extrapolation of the deterministic part.

Keywords: *SEMIFAR* models, trend, long-range dependence, fractional ARIMA, kernel estimation, bandwidth selection, semiparametric model.

---

*corresponding author. E-mail: beran@fmi.uni-konstanz.de

1 Introduction

Many economic time series exhibit apparent 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, 1981). 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, (fractional) 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. Statistical inference about population quantities and statistical forecasts are greatly influenced by our decision about the type of the 'trend' generating mechanism. For instance, for a stationary series, forecasts of a conditional expected value converge to the sample mean, with increasing forecasting horizon, and the width of forecast intervals is asymptotically constant. In contrast, for difference stationary series, forecasts converge to the last observation and the width of forecast intervals diverges to infinity. Forecasts for time series with a deterministic trend require reliable trend extrapolation which can usually not be trusted beyond a small forecasting horizon. On a finer scale, the rate at which forecast intervals converge to the asymptotic width (for stationary processes) or diverge to infinity (for difference stationary processes) depends on the fractional differencing parameter (see section 4).

In practical applications, it is often very difficult to find the "right" model and, in particular, to decide whether a series is stationary, has a deterministic or stochastic trend, or whether there may be long-range correlations. In fact, often, a combination of these may be present. To resolve this problem, Beran (1999) introduced the so-called *SEMIFAR* (semiparametric fractional autoregressive) models. These models provide a unified data-driven semiparametric approach that allows for simultaneous modelling of and distinction between deterministic trends, stochastic trends and stationary short- and long-memory components. Within the given framework, the approach helps the data analyst to decide which components are present in the observed data. In this paper, recent results on *SEMIFAR* models (Beran, 1999, Beran and Ocker, 1999, Beran and Feng, 1999) are summarized and their application to economic time series is discussed.

Briefly speaking, a *SEMIFAR* model is a fractional stationary or non-stationary autoregressive model with a nonparametric deterministic 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, for example, by kernel smoothing. The parameters may be estimated by an approximate maximum likelihood method introduced in Beran (1995). Note in particular that, with this method the integer differencing parameter is also estimated from the data. A data-driven algorithm for estimating *SEMIFAR* models, which is a mixture of these two approaches, was introduced in Beran (1999). Clearly, as any statistical method, the analysis by *SEMIFAR* models has to be accompanied by appropriate subject-specific considerations.

The paper is organized as follows. The model is defined in section 2. Estimation issues are discussed in section 3, especially nonparametric estimation of the trend and the method for estimating the parameters characterizing the stochastic component of the process. Forecasting with *SEMIFAR* models is described in section 4. The application of *SEMIFAR* models to economic time series is discussed in section 5. In particular, we discuss modelling and forecasting commodities and exchange rates, and modelling the volatility of stock market indices. Also, four simulated series are analyzed to illustrate the usefulness of the method for cases where the answer is known. (A broader simulation study is reported in Beran, 1999). Some final remarks are given in section 6.

2 The model

2.1 Definition

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 $\mathrm{var}(\epsilon_i) = \sigma_\epsilon^2$. Here, the fractional difference $(1-B)^\delta$ (Granger and Joyeux 1980, Hosking 1981) is defined by

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

with

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

2.2 Intuitive explanation of the definition

The motivation for this definition can be summarized as follows: We wish to have a model that may be decomposed into an arbitrary deterministic (possibly zero) trend and a random component that may be stationary or difference stationary. Moreover, short-range and long-range dependence as well as antipersistence should be included. Here, long-range dependence is defined as follows (see, e.g. Mandelbrot, 1983; Cox, 1984; Hampel, 1987; Künsch, 1986; Beran, 1994 and references therein): A stationary process Y_i with autocovariances $\gamma(k) = \mathrm{cov}(Y_i, Y_{t+k})$ 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

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

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. On the other hand, a stationary process is called antipersistent, if (4) holds with $\alpha \in (-1, 0)$. This implies that the sum of all autocovariances is zero, i.e. $\sum_{k=-\infty}^{\infty} \gamma(k) = 0$. Note that for usual short-memory processes, such as stationary ARMA processes, (4) holds with $\alpha = 0$, and the autocovariances sum up to a nonzero finite value.

The reason for including long-memory and anti-persistence is that for traditional ARIMA models an extreme choice has to be made between taking no or the first difference. The result of this dichotomy is that for many data sets, taking no difference is not enough (i.e. the series seems nonstationary), but taking the first difference leads to overdifferencing. The latter often results in a large negative lag-one correlation for the differenced data. To avoid this and to model slowly decaying correlations, Hosking (1981) and Granger and Joyeux (1980) introduced fractional ARIMA processes. However, there, the differencing parameter d is restricted to the stationarity region $(-1/2, 1/2)$. In a direct extension, Beran (1995) defines an arbitrary differencing parameter $d > -1/2$ such that $(1-B)^m Y_t$ is a stationary fractional ARIMA(p, δ, q) process, $m = [d + 1/2]$ is the integer part of $d + 1/2$ and $\delta = d - m$. This corresponds to equation (1) with a constant function $g \equiv \mu$. Since the integer differencing parameter m assumes integer values only and the fractional differencing parameter δ is in $(-1/2, 1/2)$, both differencing parameters can be recovered uniquely from the 'overall differencing parameter' $d = m + \delta$. If $d > 1/2$, then we have a nonstationary fractional ARIMA process. It should be noted, in particular, that this parameterisation allows for maximum likelihood estimation of d. Thus not only δ, but also m can be estimated from the data and confidence intervals can be given for both differencing parameters (see Beran 1995).

Finally, SEMIFAR models extend the definition of fractional ARIMA models with arbitrary d by including an arbitrary deterministic trend function g. (For simplicity only cases with $q = 0$ (i.e. no moving average terms) are considered. An extension to $q > 0$, which may be called 'SEMIFARIMA models', is obvious.) The definition of SEMIFAR models includes all the desired cases mentioned above. In particular, setting $\delta = 0$ and $g(t) = \mu$, we obtain classical Box-Jenkins ARIMA models. For $g = 0$, and $m = 0$ we have stationary fractional ARIMA models as defined in Hosking (1981) and Granger and Joyeux (1980).

More specifically, 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 $\delta > 0$, antipersistence if $\delta < 0$ and short memory if $\delta = 0$. (1) generalizes stationary fractional AR-processes to the nonstationary case, including difference stationarity and deterministic trend. The following special cases are thus included in (1):

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

(b) Y_t = deterministic trend + stationary process with short- or long-range dependence, or antipersistence;

(c) Y_t =no deterministic trend + difference-stationary process, whose first difference has short- or long-range dependence, or antipersistence;

(d) $Y_t =$ deterministic trend + difference-stationary process, whose first difference has short- or long-range dependence, or antipersistence.

Simulated time series for these special cases are shown in figure 1, where figures 1a to 1d correspond to case (a) trough (d), respectively. A full description of the models used in figures 1a to d is given in section 5.2.

2.3 Some economic motivation

Since the estimation of the SEMIFAR-model is purely data-driven, there exists a danger that the estimated model is inconsistent with economic reasoning. If this happens to be true, then the estimated model and the economic reasoning are called into question. In the following, we will briefly discuss some economic models which can explain short and long-term dependence in time series of prices of commodities and financial securities.

The implications of pricing models necessarily depend on the assumptions made. Many models assume perfect markets and perfectly rational economic agents. A basic requirement for any viable model is that it precludes arbitrage. A market can be arbitrage-free only if all prices for state-contingent claims are positive and finite. Let S_t be the price of some security at date t. For example, consider a stock whose price may be considered the risk-adjusted present value of future dividends. Suppose that there exist exogenous shocks at date 0 which increase (reduce) all future dividends d_t, $(t = 1, 2, \ldots)$ by finite amounts. Then, given sufficiently low discount rates, the stock price would change by an infinite amount. Hence, the market would not satisfy the no-arbitrage requirement (Mandelbrot, 1971). As Mandelbrot points out, fractional Brownian motions, which are typical stochastic models with long-range dependence or antipersistence (in the increment process), do not rule out these cases.

Another violation of the no-arbitrage requirement is obtained if the short run-autocorrelation of price changes is very high or very low. Then an observed price change would permit an almost riskless forecast of the price change over the next time period which then could be arbitraged against a risk-free asset. Rogers (1997) also proves the existence of arbitrage opportunities in fractional Brownian motions, but he also shows that a slight modification of the model suffices to rule out these opportunities. Note, in particular, that according to Mandelbrot's definition, arbitrage exists for all long-memory processes whereas this is not the case according to Rogers' definition. Thus, the answer to the question whether arbitrage is possible depends on which definition of arbitrage is used.

Even if arbitrage opportunities do not exist, the economist wonders how any short or long range dependencies in price series might be explained. Samuelson (1965) has shown that prices must follow a random walk in a risk neutral world with a non-random risk-free interest rate. For simplicity, consider an asset with an exogenously given random price S_T at date T. Then in a risk neutral world with homogeneous expectations of economic agents, the forward price S_t^f of the asset at date t equals $E_t[S_T]$, i.e. the conditional expectation of S_T. The forward price eliminates, by definition, the discounting effects of the risk-free interest rates. Since the conditional expectation of S_T follows a random walk without drift, any dependencies in forward price changes are ruled out.

In a risk-averse world with a frictionless complete market, there exists a unique forward pricing kernel $\phi_{t,T}(S_T)$ at date t ($t < T$), by which the forward price S_t^f can be derived. We have $E[\phi_{t,T}(S_T)] = 1$ and $S_T^f = E_t[\phi_{t,T}(S_T)S_T]$. Then there still exist cases in which dependencies in forward price changes do not exist. Suppose, for example, that $E_t(S_T)$ follows a standard geometric Brownian motion without drift. Then $ln(S_{t+1}^f/S_t^f)$ follows a standard geometric Brownian motion with drift if and only if the forward pricing kernel has constant elasticity $\eta_{t,T}$, i.e. if $dln\phi_{t,T}/d\ln S_T = \eta_{t,T}, \forall t$ (Franke, Stapleton and Subrahmanyam, 1999). Now suppose that the elasticity depends on S_T, holding the current forward price S_0^f constant. Suppose that $d\eta_{t,T}/dS_T < 0$ which may be thought of as "declining relative risk aversion of the market". Then the variance of the forward price S_t^f increases and the log returns $\ln(S_{t+1}^f/S_t^f)$ are negatively autocorrelated. In the case of increasing elasticity of the pricing kernel, the variance of the forward price declines relative to the constant elasticity case, but autocorrelation of log returns still is negative. The intuition behind this result is straightforward. Whenever the forward price S_t^f is higher or lower, relative to the constant elasticity case which implies zero autocorrelation, then given S_0^f and the distribution of S_T, a lower return in one period must be compensated by a higher return in the residual period. This implies short and long-range negative autocorrelation. Hence, in this framework, antipersistence may exist whereas it is difficult to argue in favor of positive autocorrelation.

Of course, real markets are not perfect. Introducing asymmetric information broadens the spectrum of return processes. Insiders, for example, attempt to exploit their information privilege by strategic trading which leads to a gradual price adjustment and, thus, short range positive autocorrelation of returns. The same autocorrelation is to be expected in the case of positive feedback trading. Then agents observe a price increase and place additional buy orders since they expect a further price increase. Finally, if large investors buy or sell consecutively small portions of a rather illiquid security, this induces positive autocorrelation.

Mandelbrot (1971) suggests that economic agents have a finite foresight horizon. This may imply various dependencies in returns. If agents, for example, ignore effects of a shock on a corporation's profits beyond some horizon, then these effects will gradually be taken into consideration and generate long-range dependent price changes. Alternatively, if agents naively extrapolate growth rates of profits over very long time horizons, this extrapolation error will gradually be corrected with corresponding gradual price changes. Such behavior might explain the well documented winner-loser effect which states that stocks with high returns over the last years tend to generate low returns over the next years and vice versa. Also cyclical macroeconomic factors tend to generate cyclical stock price behavior given either a short foresight horizon or naive extrapolative behavior. Otherwise it would be useless to distinguish between cyclical and noncyclical stocks.

Similar considerations apply to commodities for which short and long range-contracts are traded. The economic analysis of commodity prices becomes more complicated since durability of commodities and side effects of storing commodities summarized in convenience yields come into play.

Finally, long memory in aggregated indices may be a result of aggregation. As was shown, for instance, in Granger (1980), adding up a large number of time series can lead (asymptotically) to a series with long-range dependence, even if the individual series do not exhibit any long memory.

Yet, adding these real world aspects to purely theoretical models should not be understood as providing unlimited freedom to all kinds of short and long-range dependencies in time series of prices. The ultimate purpose of research is to find out price processes that are observable with sufficient reliability and are grounded on solid economic reasoning. *SEMIFAR* models provide a rather general class of models to do the empirical job. The economists should use the empirical insights for developing sensible economic models.

3 Estimation of *SEMIFAR* models - a review

Estimation of *SEMIFAR* models includes (1) nonparametric estimation of the trend component and (2) estimation of the parameters characterizing the stochastic component. This section summarizes theoretical results on the proposed kernel estimator of the trend function and approximate maximum likelihood estimator of the parameters without proofs. See Beran (1999) and Beran and Feng (1999) for details. A data-driven algorithm for estimating the whole model is also briefly described.

3.1 Kernel estimation of the trend function

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 antipersistent case, i.e. $\delta < 0$ so that the spectral density f of Y_i (or $Y_i - Y_{i-1}$ respectively) converges to zero at the origin. 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 $K(x) = \{\sum_{l=0}^{r} \alpha_l x^{2l}\} \mathbb{I}_{\{|x| \leq 1\}}$ (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 (see Beran, 1999 for the proof).

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:

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

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

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

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

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

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

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

where

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

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):

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

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

In order to obtain a 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 (see Beran and Feng, 1999)

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

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

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

with $V(0) = \lim_{\delta\to 0} V(\delta) = \pi c_f$ (see corollary 1 in Beran, 1999).

3.2 Maximum likelihood estimation

The maximum likelihood estimation proposed by Beran (1995) for a constant function $g = \mu$ can be carried over directly to *SEMIFAR* models with time-deterministic trend functions (see Beran 1999). In particular, from the 'overall differencing parameter' $d = m + \delta$ both, the discrete differencing parameter m and the fractional differencing parameter δ can be recovered uniquely, since m can take on integer values only and δ is in $(-1/2, 1/2)$. Moreover, this parameterization allows for maximum likelihood estimation of d (and thus of δ *and* m) along with the autoregressive parameters and the trend function. Moreover, inference about the autoregressive parameters takes into account that m and δ were not known a priori.

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 $m^o \in \{0, 1\}$. Then

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

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

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 standardized 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

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

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 (Beran 1999):

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

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

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 covariance 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 practice p_o needs to be estimated by applying a suitable model choice criterion. It can be shown, however, that 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 (Beran 1999):

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.*

Combining Theorems 1 through 3, It is straightforward to obtain from confidence intervals for the unknown parameter vector θ and the unknown trend function g, as well as for testing hypotheses about θ and g. Note, in particular, that the integer differencing parameter m is also estimated by maximum likelihood ($\hat{m}$ is equal to the integer part of $\hat{d} + 1/2$).

3.3 Estimation of the whole model

For estimating the whole model one needs a semiparametric data-driven algorithm combining the two estimation methods described above. An algorithm for the case where g is assumed to be equal to a constant μ is given in Beran (1995). A data-driven algorithm for estimating the *SEMIFAR* model with a general trend function g was proposed by Beran in 1997 in the original, unpublished paper on the *SEMIFAR* model. What follows is a brief description of this algorithm.

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 plug-in 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 has been extended to fitting *SEMIFAR* models (Beran 1999). A detailed study on the consistency, rates of convergence and comparison of different iterative algorithms for *SEMIFAR* fitting will be given in a forthcoming paper.

4 *SEMIFAR* forecasting

This section describes out-of-sample predictions of *SEMIFAR* processes. 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

$$Y_{n+k} = \mu(t_{n+k}) + U_{n+k} \tag{31}$$

with

$$\mu(t_{n+k}) = g(t_{n+k}), \quad U_{n+k} = X_{n+k} \tag{32}$$

if $m = 0$, and

$$\mu(t_{n+k}) = Y_n + \sum_{j=1}^{k} g(t_{n+j}), \quad U_{n+k} = \sum_{j=1}^{k} X_{n+j} \tag{33}$$

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}.

4.1 Extrapolation of the trend function

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 $j \in \{1, 2, ..., k\}$, for instance by a local constant or a local linear extension of $\hat{g}(t_n)$. $\hat{\mu}(t_{n+k})$ is obtained by inserting $\hat{g}(t_{n+k})$ in (32) or $\hat{g}(t_{n+j})$ for $j \in \{1, 2, ..., k\}$ in (33) (see Beran and Ocker, 1999).

4.2 Prediction of the stochastic component

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 $\gamma(k) = \mathrm{cov}(X_i, X_{i+k})$ 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 (Beran and Ocker, 1999)

Theorem 4 *For all integers $r, s > 0$, define*

$$\gamma_r^{(s)} = [\gamma(r+s-1), \gamma(r+s-2), ..., \gamma(r)]^T, \tag{34}$$

$$\tilde{\gamma}_k^{(n)} = \sum_{j=1}^{k} \gamma_j^{(n-1)}, \tag{35}$$

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

i) If $m = 0$,

$$\beta_{opt} = \Sigma_n^{-1} \gamma_k^{(n)}, \tag{36}$$

$$MSE_{opt} = \gamma(0) - [\gamma_k^{(n)}]^T \Sigma_n^{-1} [\gamma_k^{(n)}]; \tag{37}$$

ii) If $m = 1$,

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

$$MSE_{opt} = \sum_{s=-(k-1)}^{k-1} (k - |s|)\gamma(s) - [\tilde{\gamma}_k^{(n)}]^T \Sigma_n^{-1} [\tilde{\gamma}_k^{(n)}]. \tag{39}$$

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 (Beran and Ocker, 1999)

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*

$$\nu(\delta) = \frac{2\Gamma(1 - 2\delta) \sin \pi\delta}{\delta(2\delta + 1)} \tag{40}$$

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 $m = 0$,

$$MSE_{opt} \to \gamma(0) = \mathrm{var}(X_i); \tag{41}$$

ii) If $m = 1$,

$$MSE_{opt} \sim c_f \nu(\delta) k^{1+2\delta}. \tag{42}$$

Note in particular that, for $m = 1$ and $\delta < 0$, the MSE_{opt} diverges to infinity at a slower rate than in the case of a random walk (with $\delta = 0$). Similarly, for $m = 1$ and $\delta > 0$, the MSE_{opt} diverges faster to infinity.

4.3 Prediction intervals

Results in theorem 4 and corollary 1 can be used to obtain prediction intervals for Y_{n+k} with $k \geq 1$. For known values of g and θ a $100(1 - \alpha)$ percent prediction interval for Y_{n+k}, is given by

$$\hat{Y}_{n+k} \pm z_{\alpha/2} \sqrt{MSE_{opt}} \tag{43}$$

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 (43) are replaced by the corresponding estimated quantities.

5 Examples

In this section we provide some insight into the empirical validity of the *SEMIFAR* models by analyzing some price series and some index volatility series. Moreover, some simulation exercises demonstrate the model's capacity to find out the true properties of a time series.

5.1 Commodities and exchange rates

The data (figure 2) include daily spot prices for copper (between January 2, 1997, and September 2, 1998, n=421), a monthly price series for cocoa beans (between January 1971 and September 1996, n=310), and two daily nominal exchange rates (between September 17, 1997, and August 4, 1998, n=221). The currencies are the Swiss Franc (chf) and the European Currency Unit (xeu). The data are expressed in US dollars per unit of the corresponding series. The log-transformation (natural logarithm) was applied to each series.

First, we fit *SEMIFAR* models to the observed series. Note in particular that, instead of continuously compounded returns (first difference in natural logarithm of the closing price for consecutive trading days/months), the *original series of observed* (log-)prices is considered. Thus, in contrast to the traditional approach, it is not assumed *a priori* that the first integer difference has to be taken to make the series stationary. Instead, the possibilities of stationarity, difference stationarity, deterministic trend, short memory, long memory and antipersistence are left open. It is then decided based on the data which combination of these components may be present.

There has been some discussion in the recent literature about possible unit root behaviour or long memory in financial time series. In view of this, it is interesting to see which hypothesis may be supported by fitting *SEMIFAR* models. Table 1 summarizes the essential features of the fitted models. The corresponding 95%-confidence intervals are given in brackets. The models were selected using the BIC.

Table 1: *Estimation results*

series	$\hat{d}$	95%-c.i. d	$\hat{\phi}_1$	95%-c.i. ϕ_1	significant trend
cocoa	.897	[.682,1.112]	.394	[.142, .646]	no
copper	.780	[.705,.855]	-	-	yes
chf	.913	[.810,1.016]	-		no
xeu	.870	[.767,.973]	-	-	no

The estimated value of d and the confidence intervals suggest that all series are nonstationary ($d > 1/2$). In addition, the unit roots hypothesis ($d = 1$) can not be rejected for cocoa and chf. On the other hand, for copper and the European Currency Unit, $d = 1$ is not contained in the 95%-confidence interval. Thus, for these data, taking the first (integer) difference would lead to overerdifferencing. Furthermore, there is substantial short-term dependence in the cocoa series in form of a strong AR(1) term.

Since in all cases the estimated value of m was one, testing the presence/absence of a deterministic trend can be done by testing $H_o : g \equiv 0$ against $H_a : g \not\equiv 0$. (Note that for $m = 1$, g is the trend function for the first difference.) The only series where H_o was rejected (at the 5% level) was copper. As one may expect (at least a posteriori), for this series, a significant trend is detected due to the relatively long descent in the middle part of the observed time period. The starting and end point of the time interval where $\hat{g}$ exceeded the critical bound are marked in figure 2b by two vertical lines. Note in particular that fitting a global linear trend would not be appropriate here. For the other three series, apparent local trends do not persist long enough, and can therefore be 'explained' as purely stochastic.

The satisfactory fits of the models are demonstrated by the q-q-plots and correlograms of the residuals in figures 3 and 4. Slight departure from normality (for the residuals) can be noticed for the exchange rate data. (Note, however, that normality of ϵ_t is not required for the theoretical results described above to hold.) Also, there is no strong evidence for ARCH (autoregressive conditional heteroskedasticity) errors in the correlograms of the squared residuals (figures 4e through h).

Second, we explore the reliability of forecasts. The $k-$steps ahead out-of-sample forecasts and 95%- and 99%-forecast intervals for $k = 1, 2, ..., 20$, using constant extrapolation of g, are displayed in Figure 5. Overall, every future value was inside the 95% prediction interval. Observe also the weak US dollar in the exchange rate data during the last quarter of the period under consideration. Despite this sudden development, the future values were within the 95% prediction intervals. It should also be noted that, for $1/2 < \hat{d} < 1$, the width of forecast intervals diverges to infinity at a slower rate than under the unit root hypothesis $d = 1$. Thus, shorter forecast intervals are obtained than with unit-root models, such as a random walk. For a detailed discussion see Beran and Ocker (1999). Clearly, as always with forecasting, sudden extreme structural changes in the behaviour of the data that have not occurred in the past cannot be foreseen (except perhaps with the help of additional information).

5.2 Volatility of stock market indices

Figure 6a shows daily values of the DAX and the FTSE300 between January 2, 1992 and November 10, 1995 (weekdays only, excluding holidays). The first differences are given in figure 6b. Let I_t be the original index. To study volatility, we analyze the transformed absolute differences $Y_t = |I_t - I_{t-1}|^{\frac{1}{4}}$. The reason for taking the fourth root of the increments is that the marginal distribution of the resulting series is very close to normal (see the normal probability plots in figures 6c and d). A similar transformation approach is used, for instance, by Ding, Granger and Engle (1993), Ding and Granger (1996) and Granger and Ding (1996). Ding and Granger found long range dependence in several volatility series that were defined in a similar way. The correlograms of Y_t in figures 6e and f do indeed indicate slowly decaying autocorrelations.

Applying the *SEMIFAR* method yields $\hat{p} = 0$ for both series, $\hat{d} = -0.02$ $[-0.07, 0.03]$ for the DAX and $\hat{d} = 0.05$ $[0.003, 0.100]$ for the FTSE300. In both cases, a significant deterministic trend is found. Figure 7 shows the two Y_t series with the fitted trends and upper and lower 5% critical limits for testing significance of the trends. The result indicates that there are relatively long periods where volatility is high/low systematically for both series. This extends and is comparable to results by Ding and Granger in the following sense. For stationary long-memory processes, long-term behaviour is determined by the fractional parameter d. SEMIFAR models include, apart from d, a deterministic (and essentially arbitrary) trend function as an additional building block that can 'explain' long-term fluctuations. A smooth deterministic function can be interpreted as an even stronger (and more systematic) degree of temporal dependence than stationarity with slowly decaying correlations. The significant trends fitted to the volatility series of DAX and FTSE300 thus indicate that there may be even stronger 'long memory' in volatility than suggested by a stationary model with long-range dependence.

A more sophisticated analysis of volatility may be obtained by applying GARCH-type extensions of *SEMIFAR* models to the original series I_t. The mathematical theory necessary for such extensions is the subject of current research. For fractional GARCH models that do not include deterministic trend functions see e.g. Baillie, Bollerslev and Mikkelsen (1996), Ding and Granger (1996), Granger and Ding (1996), Ling and Li (1997). In particular, Ling and Li (1997) extend the maximum likelihood method of Beran (1995) to fractional GARCH models.

5.3 Simulated examples

In this subsection *SEMIFAR* models are fitted to some simulated series. The series ($n = 400$) are shown in figures 1a through d, which are:

Figure 1a: $Y_i = X_i$ where X_i is a fractional autoregressive process of oder $p_0 = 0$ with $d^0 = 0.4$.

Figure 1b: $Y_i = g(t) + X_i$ where X_i is a fractional autoregressive process of oder $p_0 = 0$ with $d^0 = 0.4$ (but not the same realization as in figure 1a) and $g(t) = 1.75 * (1/(1 + e^{4-8t}) - \sin(2\pi t))$.

Figure 1c: $Y_i - Y_{i-1} = X_i$ where X_i is a fractional autoregressive process of oder $p_0 = 0$ with $d^0 = -0.3$.

Figure 1d: $Y_i - Y_{i-1} = g(t) + X_i$ where X_i is the same fractional autoregressive process as in figure 1c and $g(t) = 0.2 * (t - 0.5)$.

All of these simulated series were generated by S-Plus with the "error" series X_i generated by the 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 $\hat{\eta} = \hat{d}$ (because of $p_0 = 0$) together with 95%-confidence intervals, obtained by fitting *SEMIFAR* models for $p = 0, 1, 2, 3, 4, 5$ and choosing p based on the BIC, are given in table 2. Also given are the 95%-confidence intervals for $d^o = [d^o + 0.5]$ and the results of the testing whether there is a significant trend g in the data.

Table 2: *Estimates of* p_o, d^o *and* $m^o = [d^o + 0.5]$ *for the four simulated examples in figures 1a through 1d. The true values of* p_o, d^o *and* m^o *are given in brackets. Also given are the 95%-confidence intervals for* d^o *and the results of the testing on whether there is a significant trend* g *in the data.*

Figure	$\hat{p}(p_o)$	$\hat{m}(m^o)$	$\hat{d}(d^o)$	95%-C.I. for d^o	testing on g
Fig. 1a	0(0)	0(0)	0.425(0.4)	[0.348, 0.502]	not significant
Fig. 1b	0(0)	0(0)	0.329(0.4)	[0.252, 0.406]	significant
Fig. 1c	0(0)	1(1)	0.764(0.7)	[0.687, 0.841]	not significant
Fig. 1d	0(0)	1(1)	0.762(0.7)	[0.685, 0.839]	significant

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. Similarly, regarding the presence of a deterministic trend component, the results give correct indications. Hence the proposed models provide a way to distinguish stochastic trends, deterministic trends, long- and short memory or mixtures of these. It can be expected that more refined smoothing methods, such as local bandwidth choice (see e.g. Brockmann 1993), may lead to even better estimates of g. This will be pursued elsewhere.

5.4 Comparison between SEMIFAR and AR

In this section a brief comparison between the *SEMIFAR* model and the well known AR model will be made using the four examples in section 5.1. Using the S-PLUS function *arima.mle* and the AIC criterion, an AR(2) model $y_t = 0.3392y_{t-1} - 0.0896y_{t-2} + \epsilon_t$ with $\hat{\sigma}^2 = 0.00402$ was obtained for the **cocoa** data. The AR model obtained for the **xeu** data was $y_i = -0.1178y_{t-1} + \epsilon_t$ of order 1 with $\hat{\sigma}^2 = 0.000027$. The other two data sets **copper** and **chf** were shown to be white noises. The ratios between the widths of prediction intervals for the k-step forecasting obtained be the fitted *SEMIFAR* model and the AR one are given in table 3.

Table 3: *Ratios of prediction intervals by SEMIFAR model and AR one*

k	1	2	3	4	5	6	7	8	9	10	15	20
cocoa	1.00	0.97	0.98	0.99	0.99	0.99	0.98	0.98	0.97	0.97	0.94	0.92
copper	0.98	0.88	0.82	0.77	0.74	0.71	0.69	0.67	0.66	0.64	0.59	0.56
chf	1.00	0.96	0.93	0.91	0.89	0.88	0.87	0.86	0.86	0.85	0.82	0.80
xeu	1.00	0.99	0.96	0.94	0.92	0.91	0.89	0.88	0.87	0.86	0.82	0.79

Note in particular that, in 'stationary versus unit root' approaches, a decision has to be made between $d = 0$ and $d = 1$. A wrong decision has an extreme impact on forecast intervals, since the width of forecast intervals is asymptotically constant for $d = 0$ whereas it diverges to infinity at the rate $\sqrt{k}$ for $d = 1$. In contrast, for FARIMA models, prediction intervals are of order $O(k^{\tau/2})$ with τ varying in a continuous range, including $\tau = 0$ and $\tau = 1$ as special cases. The value of $\tau = max\{0, 2d-1\}$ is estimated from the data by maximum likelihood and the extreme decision between $O(1)$ and $O(k^{.5})$ is avoided. As a result, prediction intervals are better adapted to the observed data, and often shorter if there is antipersistent. This is, in particular, often the case for foreign exchange rates. Consider for example the results in Beran and Ocker (1999) on forecasting nominal exchange rates. There, the most dramatic improvement was achieved for the British Pound. Already for $k = 20$, the average interval was shorter by a factor of about 0.7, while the coverage probability of the interval appeared to be correct. Similar results were obtained in a recent PhD. thesis of Ocker (1999), who found, in comparison, shorter prediction intervals for eight (out of eight) nominal foreign exchange rates. Many of them were shorter by a factor clearly smaller than 0.9 for $k = 20$ (see Ocker, 1999). Further evidence of antipersistence in financial time series can also be found in Ocker (1999). In contrast to foreign exchange rates (and commodities), Ocker (1999) found that traditional Box-Jenkins ARIMA forecast intervals are typically too optimistic (i.e. too short) if the degree of persistence is strong, such as for nominal stock market indices.

6 Final remarks

In this paper, we summarized recent results on so-called *SEMIFAR* models for time series that incorporate stochastic trends, deterministic trends, long-range dependence and short-range dependence. The potential usefulness of this model for economic time series analysis is illustrated by several data examples. In particular, the proposed 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 Definition (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 algorithms (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 equation (1) by a polynomial $\phi_{seas}(B) = \sum \phi_{j,seas} B^{sj}$ where $s \in N$ is the seasonal period. Other extensions, such as inclusion of parametric and nonparametric explanatory variables, other seasonal components and nonlinearities in the stochastic part of the process, are the subject of current research.

7 Acknowledgements

This research was supported in part by an NSF grant to MathSoft, Inc. (Seattle), and by the Center of Finance and Econometrics, University of Konstanz, Germany. The data for the exchange rates were obtained from the Web-site of PACIFIC (Policy Analysis Computing & Information Facility in Commerce) at the University of British Columbia, Vancouver, Canada; the prices of copper is from the homepage of the London Metal Exchange, and the cocoa price series from the ICCO's (International Cocoa Organization) Web-page. We would like to thank the authors of these data sets for making their data publicly available. Finally, we would like to thank Dr. Elke M. Hennig (Citibank, Frankfurt) for the stock market series.

References

Akaike, H., 1979, A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. Biometrika 66, 237-242.

Altman, N.S., 1990, Kernel smoothing of data with correlated errors. Journal of the American Statistical Association 85, 749-759.

Bailie, R.T., Bollerslev, T, and Mikkelsen, H.O. 1996, Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 74, 3-30.

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

Beran, J., 1995, Maximum likelihood estimation of the differencing parameter for invertible short- and long-memory ARIMA models. Journal of the Royal Statistical Society, series B 57, 695-672.

Beran, J. 1999, SEMIFAR models – a semiparametric fractional framework for modelling trends, long-range dependence and nonstationarity. Preprint, University of Konstanz.

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

Beran, J., Feng, Y., 1999, Local polynomial fitting with long-memory, short-memory and antipersistent errors. Preprint, University of Konstanz.

Beran, J., Ocker, D., 1999, SEMIFAR forecasts, with applications to foreign exchange rates. To appear in the Journal of Statistical Planning and Inference.

Box, G.E., 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. Journal of the American Statistical Association 88, 1302-1309.

Cheung, Y.W., 1993, Long memory in foreign exchange rates. Journal of Business and Economic Statistics 11, 93-101.

Chiu, S.T., 1989, Bandwidth selection for kernel estimates with correlated noise. Statistics and Probability Letters 8, 347-354.

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

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

Diggle, P.J., 1990, Time Series - a biostatistical introduction. Oxford University Press, Oxford.

Ding, Z., and Granger, C.W.J. 1996, Modeling volatility persistence of speculative returns: A new approach Journal of Econometrics 73, 185-215.

Ding, Z., Granger, C.W.J., and Engle, R.F. 1993, A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83-106.

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

Franke, G., Stapleton, R., and Subrahmanyan, M.G. 1999, When are options overpriced? The Black-Shcholes model and alternative characterisations of the pricing kernel. Forthcoming in European Finance Review.

Gasser, T., Müller, H.G., 1979, Kernel estimation of regression functions. In T. Gasser, M. Rosenblatt (editors), *Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics 757, pp. 23-68. Springer, New York.

Gasser, T., Kneip, A., Köhler, W., 1991, A flexible and fast method for automatic smoothing. Journal of the American Statistical Association 86, 643-652.

Granger, C.W.J. 1980, Long memory relationships and the aggregation of dynamic models. Journal of Econometrics 14, 227- 238.

Granger, C.W.J., and Ding, Z. 1996, Varieties of long memory models. Journal of Econometrics 73, 61-77.

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

Hall, P., Hart, J., 1990, Nonparametric regression with long-range dependence. Stochastic Processes and Their Applications 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., Quinn, B.G., 1979, The determination of the order of an autoregression. Journal of the Royal Statistical Society, series B 41, 190-195.

Herrmann, E., Gasser, T., 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.

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

Liu, C.Y., 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. The Review of Economics and Statistics, 53, 225 – 236.

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

Ocker, D. (1999). Stationary and nonstationary fractional ARIMA models - model choice, forecasting, aggregation and intervention. Unpublished PhD thesis. University of Konstanz.

Priestley, M.B., 1981, Spectral Analysis and Time Series. Academic Press, London.

Ray, B.K., Tsay, R.S., 1997, Bandwidth selection for kernel regression with long-range dependence. Biometrika 84, 791–802.

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

Samuelson, P.A., 1965, Proof that properly anticipated prices fluctuate randomly. Industrial Management Review, VI, 41–49.

Schwarz, G., 1978, Estimating the dimension of a model. Annals of Statistics 6, 461-464.

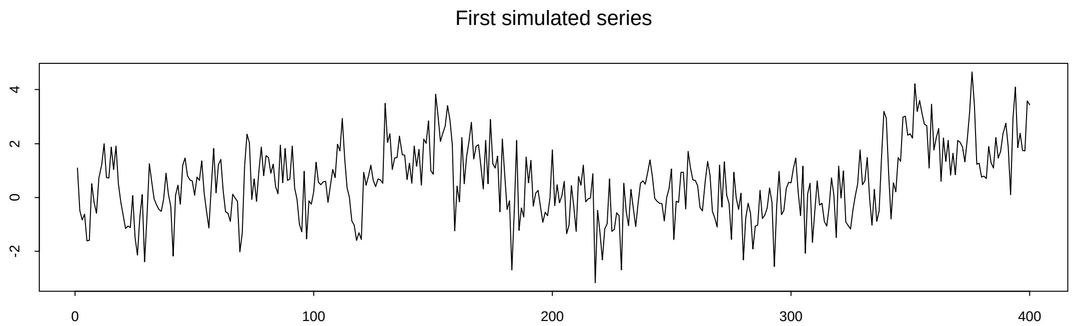

Figure 1a

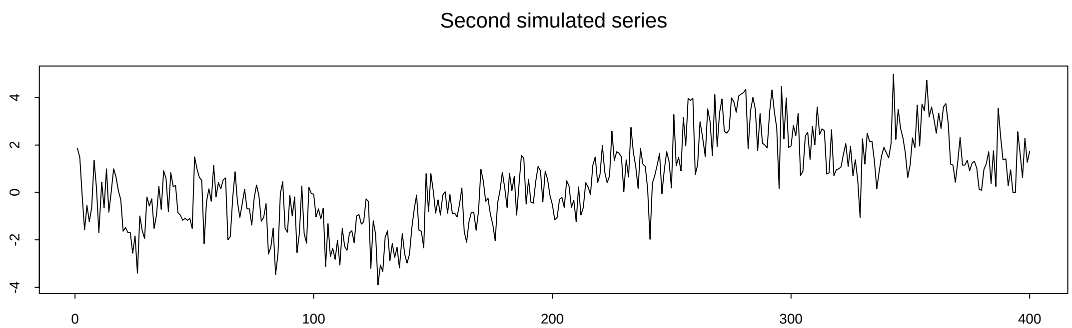

Figure 1b

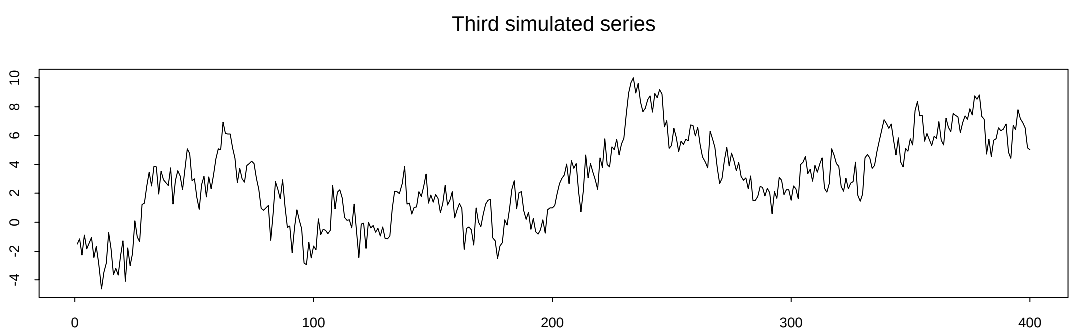

Figure 1c

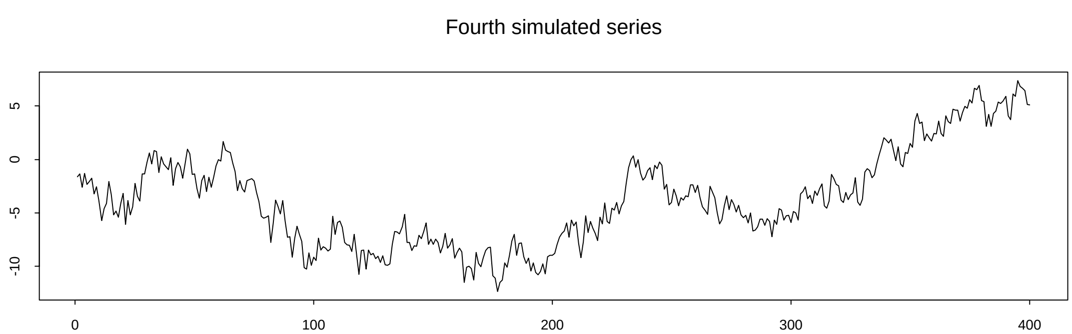

Figure 1d

Figure 1: Simulated series

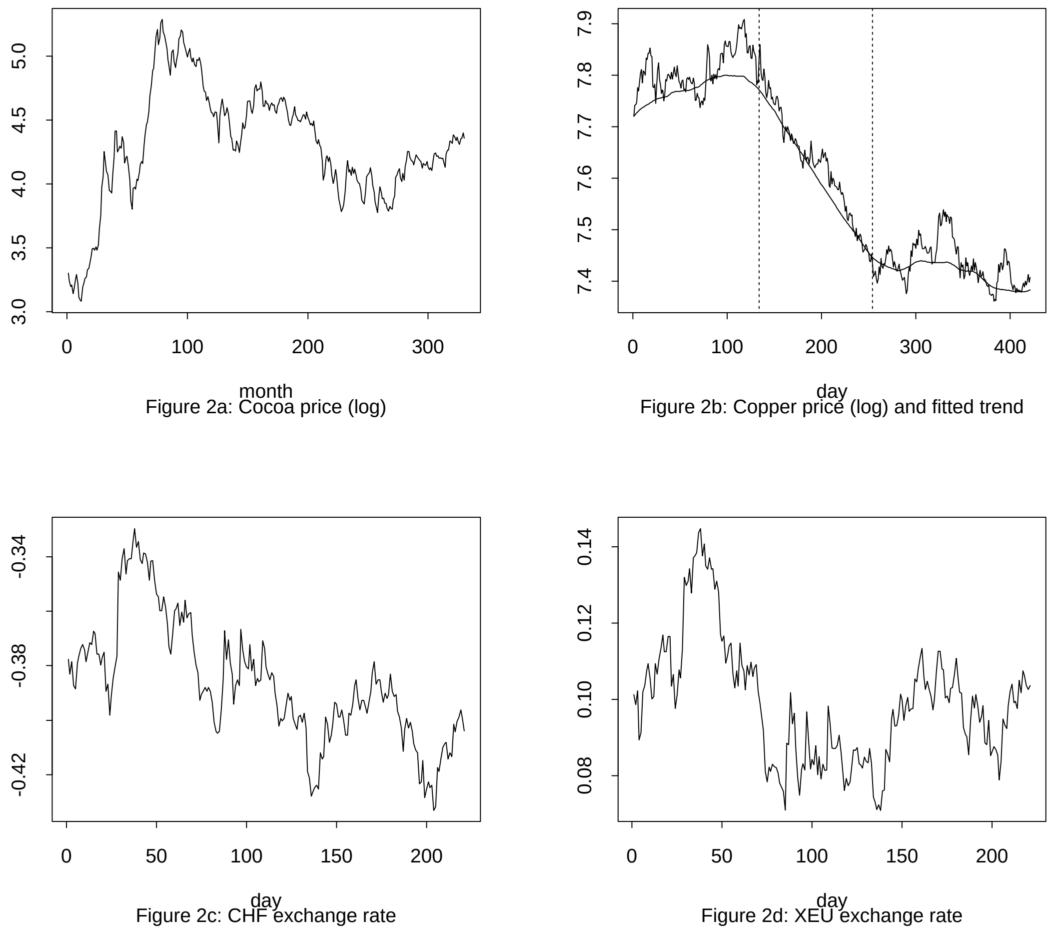

Figure 2: Monthly prices for cocoa beans (Jan. 1971 - Sept. 1996, n=310), daily prices for copper (Jan. 2, 1997 - Sept. 2, 1998, n=421), daily nominal exchange rates for the Swiss Franc (log(USD/CHF)) and the European Currency Unit (log(USD/XEU)) (Sept. 17, 1997 - Aug. 4, 1998, n=221).

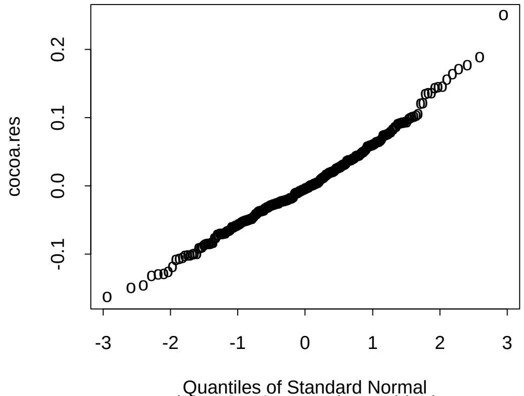

Figure 3a: Cocoa price residuals

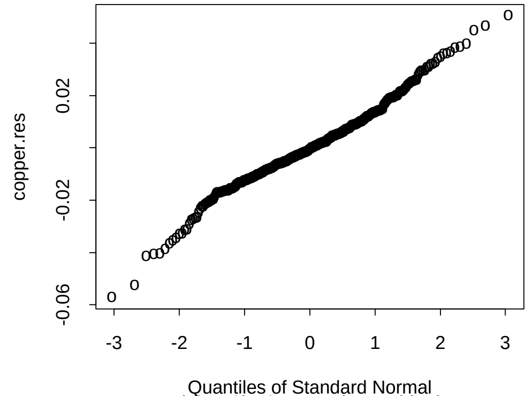

Figure 3b: Cocoa price residuals

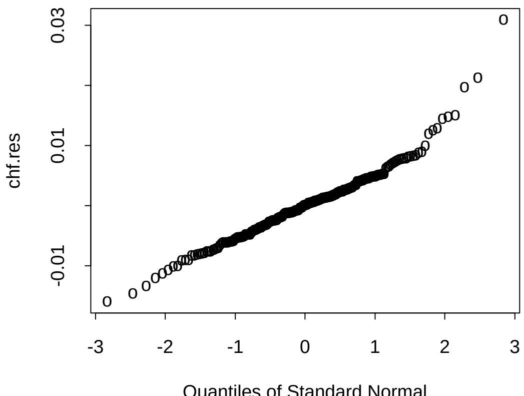

Figure 3c: CHF residuals

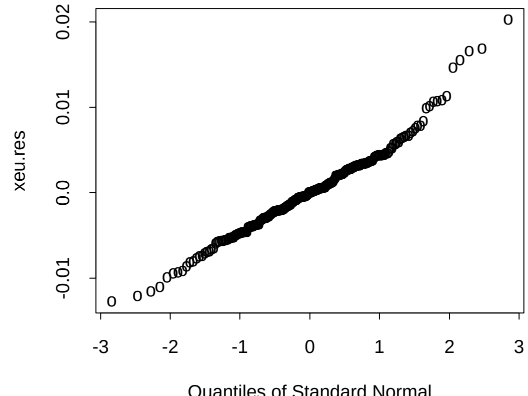

Figure 3f: XEU residuals

Figure 3: Normal probability plots of SEMIFAR-residuals for the examples in figures 2a through d.

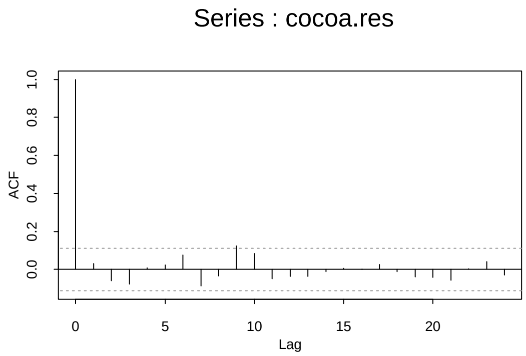

Figure 4a: Cocoa price residuals

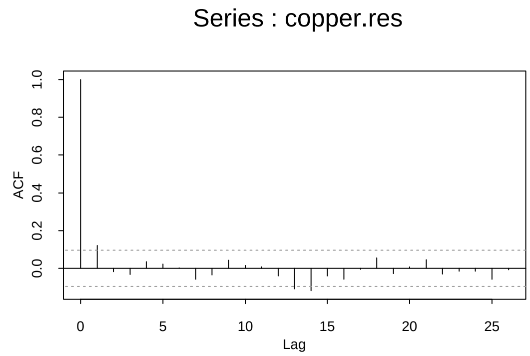

Figure 4b: Cocoa price residuals

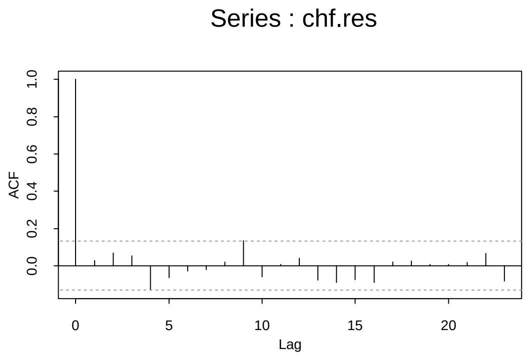

Figure 4c: CHF residuals

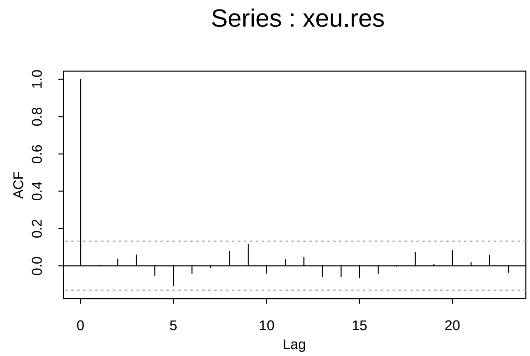

Figure 4d: XEU residuals

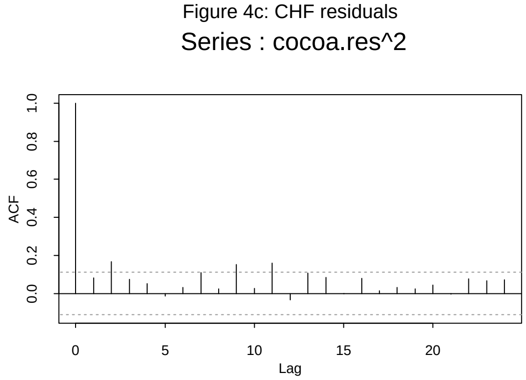

Figure 4e: Cocoa price residuals**2

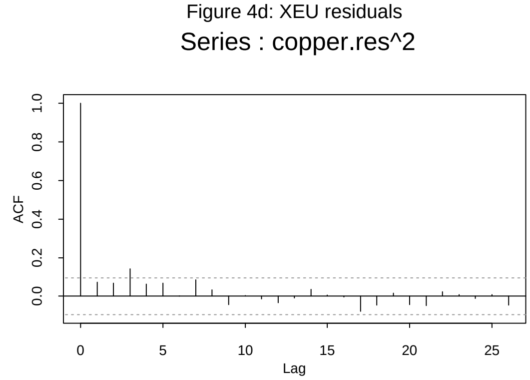

Figure 4f: Cocoa price residuals**2

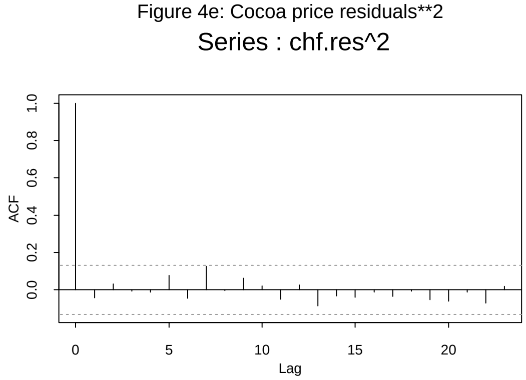

Figure 4g: CHF residuals**2

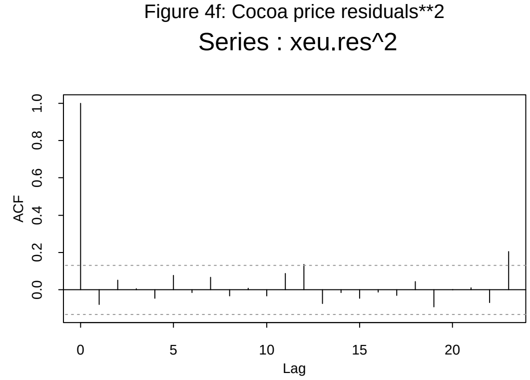

Figure 4h: XEU residuals**2

Figure 4: Autocorrelations of SEMIFAR-residuals (figures 4a through d) and of the squared residuals (figures 4e through h) for the examples in figures 2a through d.

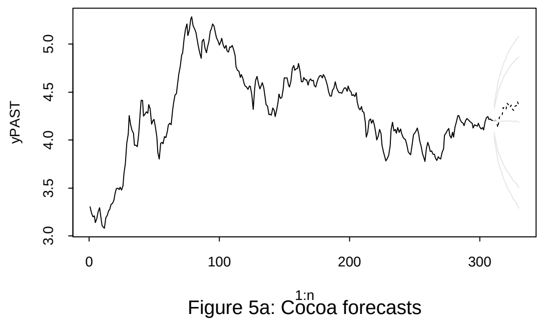

Figure 5a: Cocoa forecasts

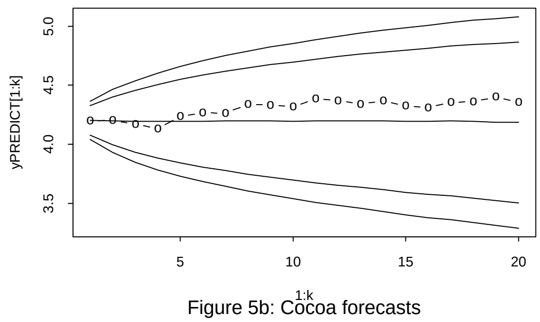

Figure 5b: Cocoa forecasts

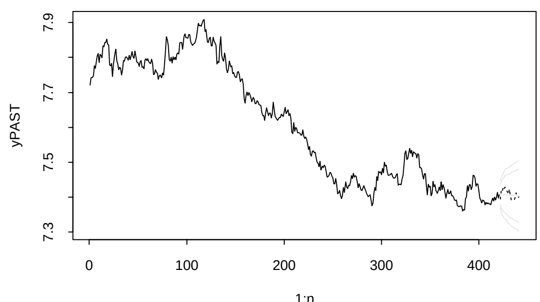

Figure 5c: Copper forecasts

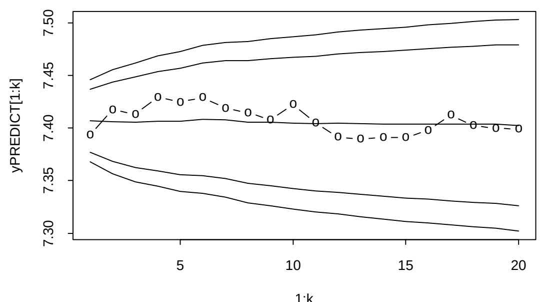

Figure 5d: Copper forecasts

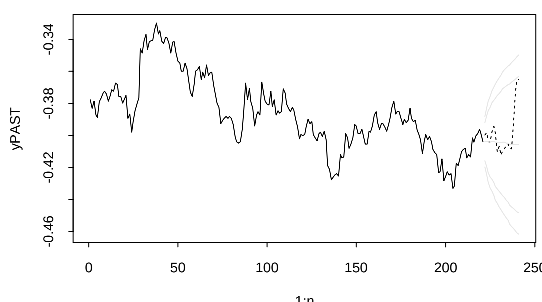

Figure 5e: CHF forecasts

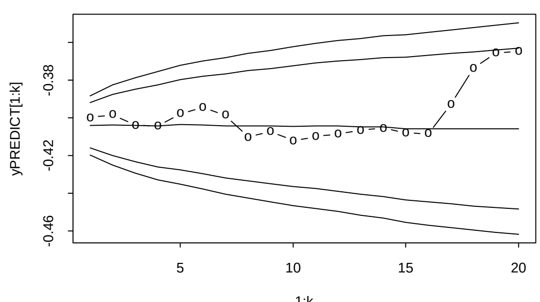

Figure 5f: CHF forecasts

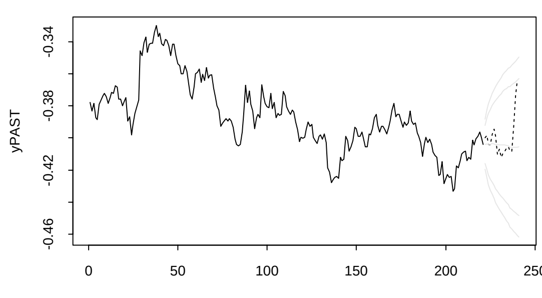

Figure 5g: XEU forecasts

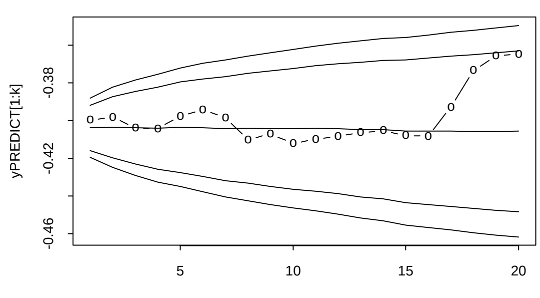

Figure 5h: XEU forecasts

Figure 5: Observed values with $k-$step ahead SEMIFAR forecasts and 95%- and 99%-forecast intervals for the examples in figures 2a through d. Figures 5b, d, f and h display close-ups of the forecasts and forecast intervals in figures 1a, c, e and g.

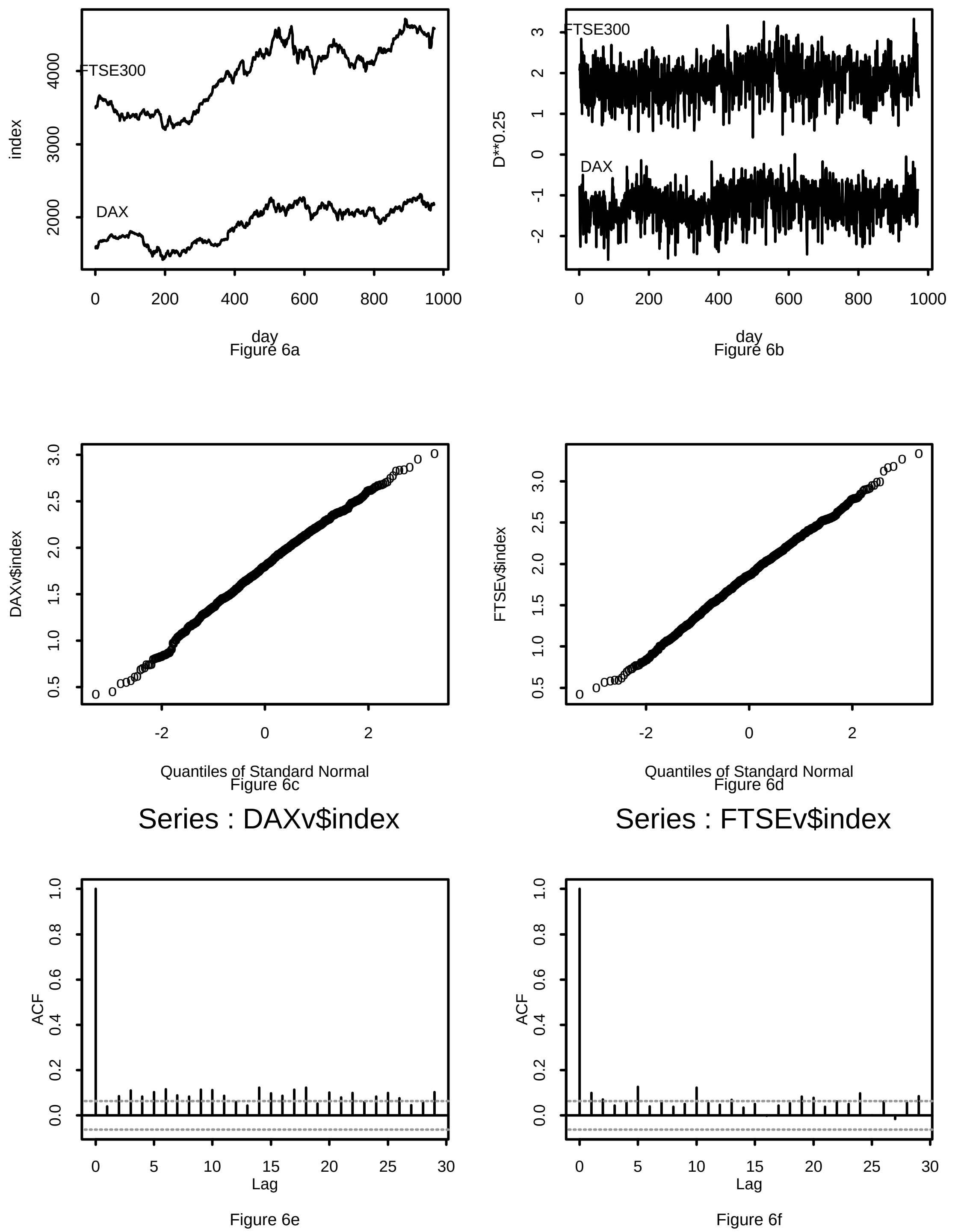

Figure 6: Daily DAX and FTSE300 values I_t (figures 6a), first difference (figures 6b), normal probability plots of $Y_t = |I_t - I_{t-1}|^{\frac{1}{4}}$ (figures 6c,d) and autocorrelations of Y_t (figures 6e,f).

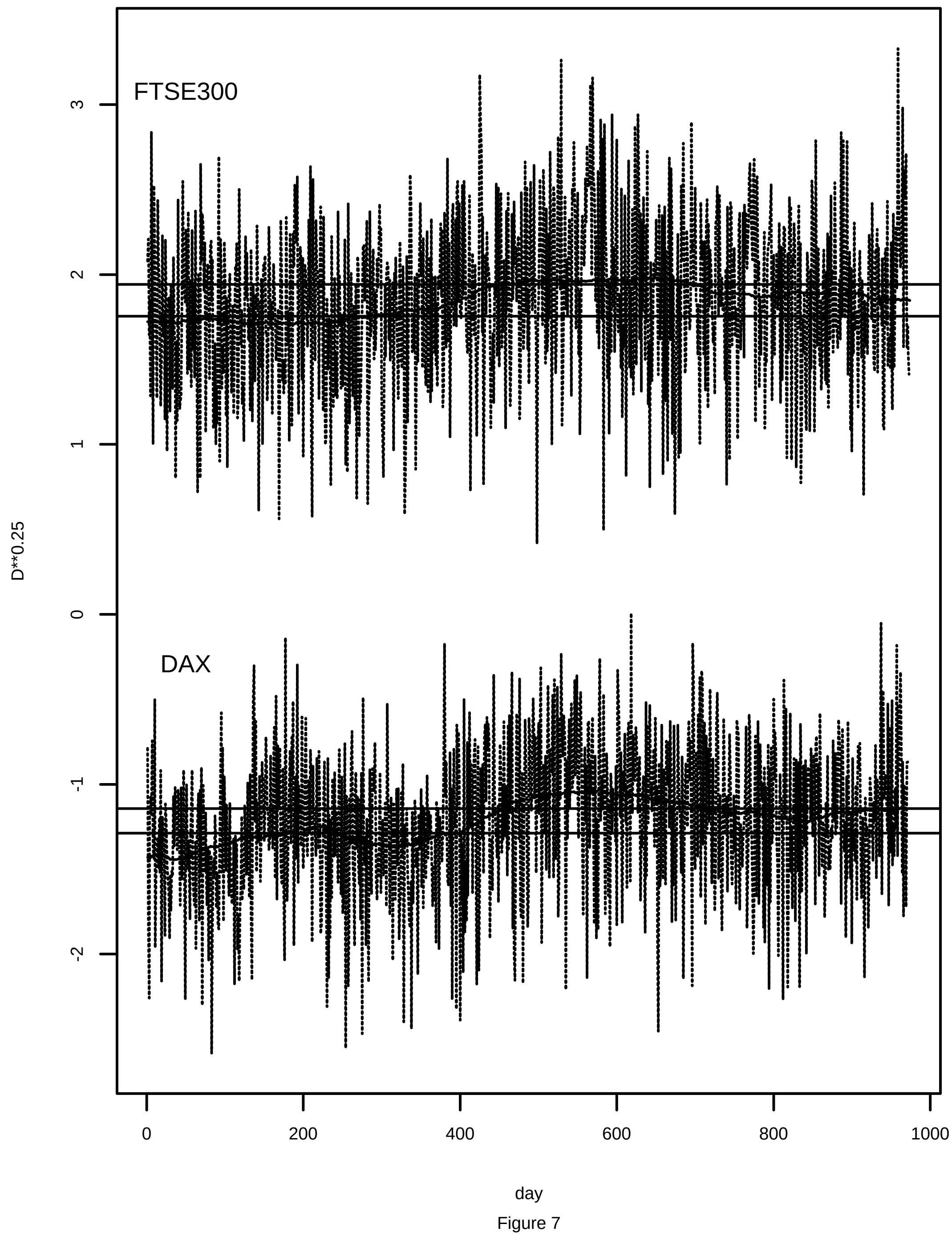

Figure 7: Trends fitted by the SEMIFAR method to $Y_t = |I_t - I_{t-1}|^{\frac{1}{4}}$ where $I_t = DAX$ and FTSE300 respectively. Also given are the 5% rejection limits for testing where the trend is significant.

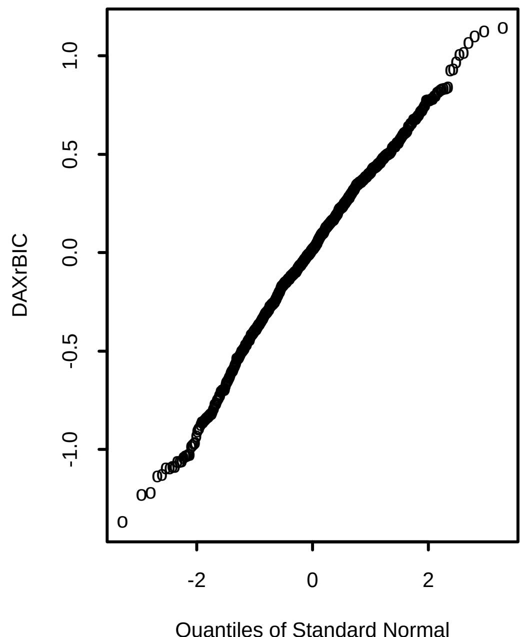

Figure 8a

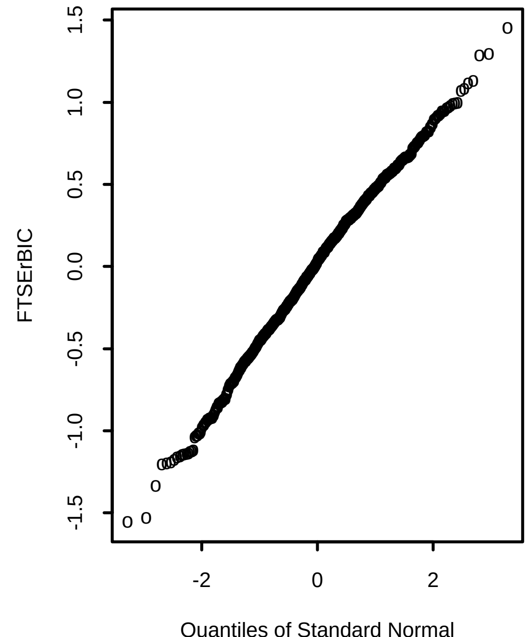

Figure 8b

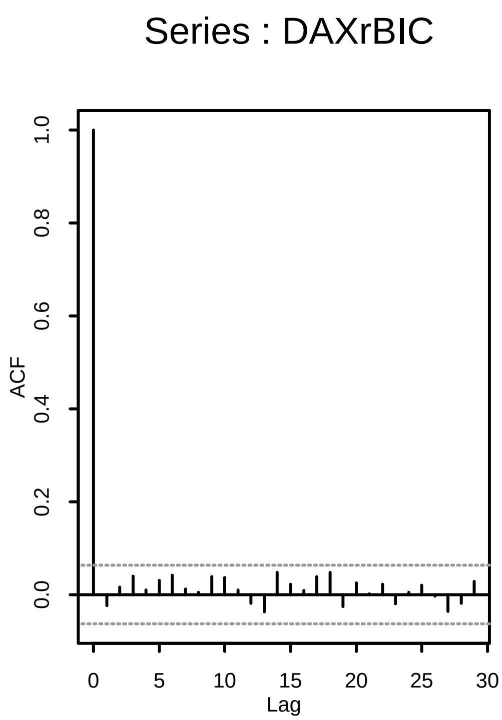

Figure 8c

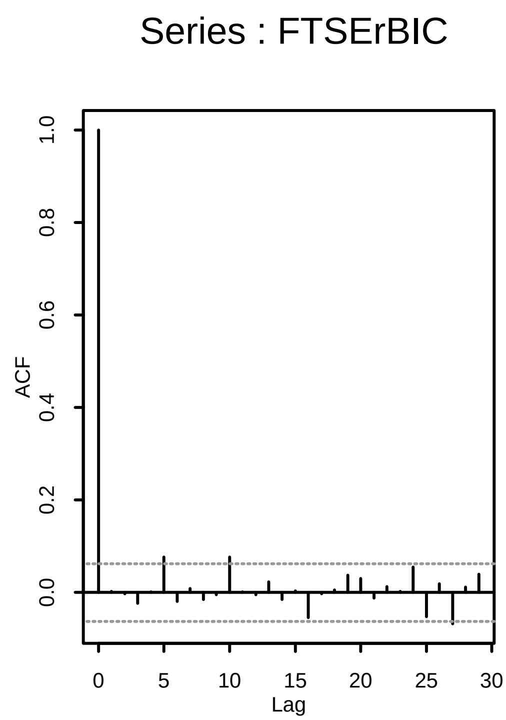

Figure 8d

Figure 8: Normal probability plots (figures 8a,b) and correlogram (figures 8c,d) for the residuals obtained from SEMIFAR-fits to method to $Y_t = |I_t - I_{t-1}|^{\frac{1}{4}}$ where $I_t = DAX$ and FTSE300 respectively.