SEMIFAR forecasts, with applications to foreign exchange rates

Jan Beran

and

Dirk Ocker

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

June 2, 1999

Abstract

SEMIFAR models introduced in Beran (1999) provide a semiparametric modelling framework that enables the data analyst to separate deterministic and stochastic trends as well as short- and long-memory components in an observed time series. A correct distinction between these components, and in particular, the decision which of the components may be present in the data have an important impact on forecasts. In this paper, forecasts and forecast intervals for *SEMIFAR* models are obtained. The forecasts are based on an extrapolation of the nonparametric trend function and optimal forecasts of the stochastic component. In the data analytical part of the paper, the proposed method is applied to foreign exchange rates from Europe and Asia.

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

1 Introduction

Prediction of future observations and the calculation of valid prediction intervals strongly depends on the type of model that may be assumed. In particular, many observed time series exhibit apparent trends. Forecasts will differ greatly, depending on how these trends are modelled. A trend may be (1) deterministic, i.e. defined by a deterministic function, (2) purely stochastic or a mixture of (1) and (2). Typical examples of (1) are polynomial functions or functions satisfying certain smoothness assumptions. Typical examples of (2) are nonstationary processes, such as random walk or ARIMA(p, d, q) processes (Box and Jenkins 1976), whose mth difference is stationary ($m = 1, 2, ...$). In addition to nonstationary models, stationary long-memory processes (see e.g. Beran 1994) often exhibit local spurious trends which may be hard to distinguish from deterministic trends or purely stochastic nonstationarity.

Here, 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

$$f(\lambda) \sim c_f |\lambda|^{-\alpha} \ (|\lambda| \to 0) \tag{1}$$

for constants $c_f > 0$ and $\alpha \in (0, 1)$, where " $\sim$ " means that the ratio of the left and right hand side converges to one (see, e.g. Mandelbrot 1983, Cox 1984, Hampel 1987, Künsch 1986, and Beran 1994 and references therein). 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.

While there are many methods in the time series literature that deal with certain types of trends (see e.g. Phillips 1997 and references therein; also see Beran 1994 for references on long-memory processes), the large variety of possible models is often confusing to the applied data analyst. Finding an appropriate model and, in particular, defining realistic forecasts, is therefore a challenging task in practice. A possibility to resolve this problem is to setup a unified framework in which flexible modelling of deterministic and stochastic components is possible, and objective data driven inference can be made to decide which of the components (deterministic trend, stochastic trend, spurious trends, stationary components) may be present. This approach is taken in Beran (1999) by defining so-called *SEMIFAR* models. These models combine parametric modelling of stochastic components by stationary and nonstationary fractional autoregressive models, as defined in Beran (1995), with nonparametric modelling of a deterministic trend function. A semiparametric estimation method can be defined combining maximum likelihood and kernel estimation with data driven bandwidth selection. This is described in detail in Beran (1999). The present paper considers forecasting based on *SEMIFAR* models.

The organization of the paper is as follows. In section 2, a brief summary of the main results in Beran (1999) is given. Formulas for predictions and prediction intervals are derived in section 3. The results are applied to foreign exchange rates in section 4. In particular, the reliability of forecast intervals is examined empirically. Final remarks conclude the paper. Proofs are given in the appendix.

2 *SEMIFAR* models

2.1 Definition of *SEMIFAR* models

Let $g : [0,1] \to R$ be a twice differentiable function such that

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

and

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

for all $x, y \in [0,1]$, constants $C_1, C_2 < \infty$, and some $\beta \in (2,3]$. Here $g^{(j)}$ denotes the jth derivative. Also, assume that for at least one $l \in \{0,1\}$, $g^{(l+1)}$ does not vanish in $[\Delta, 1-\Delta]$ $(0 < \Delta < 0.5)$ and $g^{(l)}$ achieves an absolute maximum or minimum in $[\Delta, 1-\Delta]$. These conditions on g are analogous to Ray and Tsay (1987) and Hall and Hart (1990). Furthermore the following notations will be used: ϵ_i $(i = ..., -1, 0, 1, 2, ...)$ is a sequence of iid zero mean normal random variables with $\sigma_\epsilon^2 = \mathrm{var}(\epsilon_i)$, $\phi(x) = 1 - \sum_{j=1}^{p} \phi_j x^j$ is a polynomial with roots outside the unit circle, B is the backshift operator with $BY_i = Y_{i-1}$. Also, we use the convention $B^j g(t_i) = g(t_{i-j})$, where $t_i = i/n$. Then *SEMIFAR* models are defined by (Beran 1999):

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

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

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

Remarks:

1. Note that for $m = 0$, $Y_i = \mu(t_i) + U_i$ where $\mu(t_i) = g(t_i)$ and U_i is a stationary fractional AR process defined as the stationary solution of

$$\phi(B)(1-B)^{\delta} U_i = \epsilon_i \tag{3}$$

(Granger and Joyeux 1980, Hosking 1981). For $m = 1$, we have $Y_i - Y_1 = \mu(t_i) + U_i$ where $\mu(t_i) = \sum_{j=2}^{i} g(t_j) \sim nG(t_i)$, with $G(t_i) = \int_o^{t_i} g(s)ds$, and the first difference $\tilde{U}_i = U_i - U_{i-1}$ is a stationary fractional AR process. Also, defining $\tilde{Y}_i = (Y_i - Y_{i-1})$, we have $\tilde{Y}_i = g(t_i) + \tilde{U}_i$.

2. For $g = const$, we obtain the model class defined in Beran (1995). In particular, if $m = 0$, then (2) with constant g reduces to a stationary fractional AR process (Granger and Joyeux 1980, Hosking 1981). Moreover, for $\delta = 0$ (2) with g constant reduces to an ARIMA$(p, d, 0)$ model with $d \in \{0, 1\}$ as defined in Box and Jenkins (1976). Finally note that, if $m = 1$ and g is constant, then $g \neq 0$ implies a linear trend in Y_i.

3. Note that, $d = m + \delta > -0.5$ can assume fractional as well as integer values and both, the integer differencing parameter $m = [d + 0.5]$ (with $[\cdot]$ denoting the integer part) and the fractional differencing parameter $\delta = d - m$, are assumed to be *unknown*.

4. The function g is general, satisfying only the smoothness assumptions given above. Thus, much more than linear trends are included in model (2).

5. For $m = 0$, the trend function in Y_i is bounded and the 'errors' U_i are generated by a stationary process with constant variance. In contrast, for $m = 1$, the trend function in $Y_i - Y_1$ may be unbounded and the variance of the error process U_i is increasing with i.

6. The restriction $m \leq 1$ is imposed for two reasons: a) $m \geq 2$ is rarely encountered in practice; b) for $m \leq 1$, the trend function μ in the original process Y_i (for $m = 0$) and $Y_i - Y_1$ (for $m = 1$) respectively, can be reconstructed from (2), whereas this is no longer the case for $m \geq 2$.

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

2.2 Estimation for *SEMIFAR* models

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 (2) where $d^o = m^o + \delta^o$, $-1/2 < \delta^o < 1/2$ and $m^o \in \{0, 1\}$. Combining maximum likelihood with kernel estimation, the following method for estimating θ^o and the trend function g is obtained in Beran (1999): Let K be a symmetric polynomial kernel (see e.g. Gasser and Müller 1979) defined by $K(x) = \sum_{l=0}^{r} \alpha_l x^{2l}$, $(|x| \leq 1)$, and $K(x) = 0$ if $|x| > 1$, $r \in \{0, 1, 2, ...\}$ and the coefficients α_l such that $\int_{-1}^{1} K(x)dx = 1$. 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

$$\hat{g}(t_i; m) = \frac{1}{nb_n} \sum_{j=1}^{n} K(\frac{t_i - t_j}{b_n})\tilde{Y}_j \tag{4}$$

where $\tilde{Y}_j = (1-B)^m Y_j$ (for $m = 1$, set $\tilde{Y}_1 = 0$). Using equations (2) and (4), define approximate residuals

$$e_i(\eta) = \sum_{j=0}^{i-m-1} a_j(\eta)[c_j(\eta)Y_{i-j} - \hat{g}(t_{i-j}; m)], \tag{5}$$

with coefficients a_j and c_j obtained from (2), and denote by $r_i(\theta) = e_i(\eta)/\sqrt{\theta_1}$ the standardized residuals as a function of a trial value $\theta = (\sigma^2_\epsilon, m + \delta, \phi_1, ..., \phi_p)^t$. Then $\hat{\theta}$ is defined by maximizing the approximate log-likelihood

$$l(Y_1, ..., Y_n; \theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2_\epsilon - \frac{1}{2}n^{-1}\sum_{i=m+2}^{n} r_i^2 \tag{6}$$

with respect to θ and $\hat{g}(t_i)$ is set equal to $\hat{g}(t_i; \hat{m})$.

The asymptotic behaviour of $\hat{g}$ and $\hat{\theta}$ is derived in Beran (1999). As $n \to \infty$, $\hat{g}$ converges in probability to g, the optimal mean squared error of $\hat{g}$ is proportional to $n^{(4\delta-2)/(5-2\delta)}$ and $\sqrt{n}(\hat{\theta} - \theta)$ converges in distribution to a zero mean normal vector with covariance matrix $V = 2D^{-1}$ where

$$D_{ij} = (2\pi)^{-1}[\int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_i}\log f(x)\frac{\partial}{\partial\theta_j}\log f(x)dx]|_{\theta=\theta^o_*} \tag{7}$$

with $\theta_*^o = (\sigma_{\epsilon,o}^2, \eta_*^o)^T = (\sigma_{\epsilon,o}^2, \delta^o, \eta_2^o, ..., \eta_{p+1}^o)^T$. The same result holds if a consistent model choice criterion is used for the estimation of the autoregressive order p. It should be emphasized, in particular, that here both, the integer differencing parameter $m^o = [d^o + 0.5]$ and the fractional differencing parameter $\delta^o = d^o - m^o$ are estimated from the data. Also, the same central limit theorem holds if the innovations ϵ_i are not normal, and satisfy suitable moment conditions. Finally note that the asymptotic convariance matrix does not depend on m^o.

2.3 Algorithm

In the derivation of the asymptotic results, only very general conditions are imposed on the sequence of bandwidths b_n. In practice, good bandwith choice for a given data set is essential to distinguish correctly between deterministic trend, stochastic nonstationarity or spurious trends caused by stationary long memory. Minimizing the asymptotic integrated mean squared error, $IMSE = \int_\Delta^{1-\Delta} E\{[\hat{g}(t) - g(t)]^2\}dt$ (for some $0 < \Delta < 1/2$), the asymptotically optimal bandwidth turns out to be given by $b_{opt} = C_{opt} n^{(2\delta^o - 1)/(5 - 2\delta^o)}$ where the constant C_{opt} is a function of $g^{''}$, δ and $c_f = \lim_{\lambda \to 0} |\lambda|^{2\delta^o} f(\lambda)$, with f denoting the spectral density of $(1-B)^{m^o} Y_i - g(t_i)$. This result is obtained in Beran (1999) for the whole range of $\delta^o \in (-0.5, 0.5)$ (and $m^o \in \{0, 1\}$) and generalizes previous results for $\delta^o = 0$ (Chiu 1989, Altman 1990, Hall and Hart 1990, Herrmann, Gasser and Kneip 1992) and $\delta^o \in (0, 0.5)$ (Hall and Hart 1990, Csörgö and Mileniczuk 1995, Ray and Tsay 1997). An algorithm for estimating θ^o and g can then be defined by combining iterative plug-in bandwith selection similar to Ray and Tsay (1997) and Herrmann et al. (1992) with the estimation procedure in the previous section. A detailed description of the algorithm is given in Beran (1999).

3 *SEMIFAR* predictions

3.1 The basic problem

Let $Y_1, ..., Y_n$ be observations generated by a *SEMIFAR* model of order p with parameter vector $\theta = (\sigma_\epsilon^2, d, \phi_1, ..., \phi_p)$ (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

$$Y_{n+k} = \mu(t_{n+k}) + U_{n+k} \tag{8}$$

with

$$\mu(t_{n+k}) = g(t_{n+k}), \quad U_{n+k} = X_{n+k} \tag{9}$$

if $m = 0$, and

$$\mu(t_{n+k}) = Y_n + \sum_{j=1}^{k} g(t_{n+j}) \approx Y_n + nG(t_{n+k}), \quad U_{n+k} = \sum_{j=1}^{k} X_{n+j} \tag{10}$$

if $m = 1$, where $G(t) = \int_1^t g(s)ds$. 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}.

3.2 Extrapolation of the trend function

Since for *SEMIFAR* models only general regularity conditions on g are imposed, the deterministic trend $\mu(t)$ may behave in an arbitrary way in the future. This is in contrast to parametric trend models. However, we may assume that in the neighbourhood of $t_n = 1$, a Taylor expansion of order r can be applied. A natural extrapolation of μ is thus given by

$$\mu(t_{n+k}) = \mu(1 + \Delta) \approx \sum_{j=0}^{r} \mu^{(j)}(1)\Delta^j \tag{11}$$

with $\Delta = k/n$. In general, this approximation is only good for sufficiently small values of k/n. This reflects the well known experience that extrapolation of trends too far into the future may be dangerous. Theoretically, the Taylor extrapolation is valid if k/n tends to zero with increasing n. In the applications below, $r = 0$, i.e. constant extrapolation

$$\mu(t_{n+k}) \approx \mu(1) \tag{12}$$

and $r = 1$, i.e linear extrapolation

$$\mu(t_{n+k}) \approx \mu(1) + \mu^{'}(1)\Delta \tag{13}$$

will be used. There are at least two reasons for choosing a low value of r:

1. Higher order derivatives may be difficult to estimate. In particular, the order of the optimal mean squared error of kernel estimates of the jth derivative increases with increasing j (see e.g. Herrmann et al. 1992, Ray and Tsay 1997, Beran 1999).

2. Higher order polynomials may explode very quickly yielding unrealistic prediction values.

It should be emphasized, however, that higher order Taylor extrapolation may be useful for certain data sets. This question will need to be considered in future research.

In terms of g, linear extrapolation for the two cases, $m = 0$ and $m = 1$, looks as follows. For $m = 0$,

$$\mu(t_{n+k}) = g(t_{n+k}) \approx g(1) + g^{'}(1)\frac{k}{n}. \tag{14}$$

For $m = 1$, $\mu(t) \approx Y_n + nG(t)$ $(t \geq 1)$ so that $\mu^{'}(t) = ng(t)$. Thus,

$$\mu(t_{n+k}) = Y_n + \sum_{j=1}^{k} g(t_{n+j}) \approx Y_n + g(1)k. \tag{15}$$

Constant extrapolation is given by

$$\mu(t_{n+k}) = g(t_{n+k}) \approx g(1) \tag{16}$$

for $m = 0$, and

$$\mu(t) \approx Y_n \tag{17}$$

for $m = 1$.

3.3 Linear 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

Theorem 1 *For all integers $r, s > 0$, define*

$$\gamma_r^{(s)} = [\gamma(r+s-1), \gamma(r+s-2), ..., \gamma(r)]^t, \tag{18}$$

$$\tilde{\gamma}_k^{(n)} = \sum_{j=1}^{k} \gamma_j^{(n-1)}, \tag{19}$$

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

i) If $m = 0$,

$$\beta_{opt} = \Sigma_n^{-1} \gamma_k^{(n)}, \tag{20}$$

$$MSE_{opt} = \gamma(0) - [\gamma_k^{(n)}]^t \Sigma_n^{-1} [\gamma_k^{(n)}]; \tag{21}$$

ii) If $m = 1$,

$$\beta_{opt} = \Sigma_n^{-1} \tilde{\gamma}_k^{(n)}, \tag{22}$$

$$MSE_{opt} = \sum_{s=-(k-1)}^{k-1} (k - |s|)\gamma(s) - [\tilde{\gamma}_k^{(n)}]^t \Sigma_n^{-1} [\tilde{\gamma}_k^{(n)}]. \tag{23}$$

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*

$$\nu(\delta) = \frac{2\Gamma(1-2\delta)\sin \pi\delta}{\delta(2\delta+1)} \tag{24}$$

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 $m = 0$,

$$MSE_{opt} \to \gamma(0) = \mathrm{var}(X_i); \tag{25}$$

ii) If $m = 1$,

$$MSE_{opt} \sim c_f \nu(\delta) k^{1+2\delta}. \tag{26}$$

The intuitive interpretation of this result is that, as k tends to infinity, the past values $X_1, ..., X_n$ do not contribute anything to the prediction so that the MSE approaches the variance of U_{n+k}. For $m = 0$, $\mathrm{var}(U_{n+k})$ is equal to $\gamma(0) = \mathrm{var}(X_i)$, independently of the value of δ. For $m = 1$ (i.e. $d > 0.5$), the variance of U_{n+k} is proportional to $k^{2d-1} = k^{1+2\delta}$. Note, in particular, that for classical models such as Box-Jenkins ARIMA models, a discrete choice is made between $d = 0$ (stationarity) and $d = 1$ (unit root). As a result, prediction intervals are either asymptotically constant ($d = 0$) or diverge to infinity at the rate $O(\sqrt{k})$ (for $d = 1$). In contrast, when using fractional models with $d \in (-0.5, 0.5) \cup (0.5, 1.5)$, the length of prediction intervals is of the order $O(k^{\psi/2})$ with $\psi = \max(0, 2d-1)$. As $d \to 0.5$, $\psi \to 0$ and $k^{\psi} \to k^o = 1$. Thus, there is a continuous transition between $O(1)$ and $O(k^{d-1/2})$ (for $d > 0.5$). Due to this flexibility, prediction intervals for fractional models are better adapted to the data. In particular, for $0.5 < d < 1$, confidence intervals based classical models will either be asymptotically too narrow, if $d = 0$ is chosen, or unnecessarily wide (by the factor k^{1-d}), if $d = 1$ is chosen. The examples considered in the next section indicate that values of $m = 1$ and $\delta < 0$ may indeed occur quite frequently.

3.4 Estimated prediction intervals

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

$$\hat{Y}_{n+k} \pm z_{\alpha/2}\sqrt{MSE_{opt}} \tag{27}$$

where $\hat{Y}_{n+k} = \sum_{j=0}^{r} \mu^{(j)}(1)\ (k/n)^r + \beta_{opt}^t X(n)$ and the values of β_{opt} and MSE_{opt} are obtained from theorem 1. If g and θ are estimated, the quantities in (27) are replaced by the corresponding estimated quantities. More specifically, the following procedure can be defined:

1. *Estimation of the parameters:* Estimate p, θ^o and g by applying the estimation method defined in section 2, together with a model selection criterion such as the BIC (Schwarz 1978, Akaike 1979, Beran 1999).

2. *Deterministic extrapolation:* Depending on the values of $\hat{m} = [\hat{\theta}_2 + 0.5]$ and r, use $\hat{g}$ and equations (16) or (17) respectively, to obtain $\hat{\mu}(t_{n+k})$. (For $\hat{m} = 0$ and $r = 1$, estimate $g^{'}$ by a suitable nonparametric method, such as kernel estimation).

3. *Stochastic prediction:* Obtain $\hat{\beta}_{opt}^t$ and the estimated mean squared prediction error $s_{n,k}^2 = MSE(\hat{\theta})$ from theorem 1, using the estimated covariance matrix $\hat{\Sigma}_n = \Sigma_n(\hat{\theta})$.

4. *Prediction of Y_{t+k}:* Set $\hat{X}_i(n) = Y_i - \hat{\mu}(t_i)$ if $\hat{m} = 0$ and $\hat{X}_i = Y_i - Y_{i-1} - \hat{g}(t_i)$ if $\hat{m} = 1$, respectively, and define

$$\hat{Y}_{n+k} = \hat{\mu}(t_{n+k}) + \hat{\beta}_{opt}^t \hat{X}(n). \tag{28}$$

5. *Prediction interval:* Define the prediction interval $\hat{Y}_{n+k} \pm z_{\alpha/2} s_{n,k}$ where $P(|Z| > z_{\alpha/2}) = \alpha$ for a standard normal random variable Z.

4 *SEMIFAR* forecasts for foreign exchange rates

To illustrate SEMIFAR predictions, we consider prediction of foreign exchange rates. The exchange rates chosen here are: daily data (between January 1993 and May 1997) for the Swiss Franc (CHF), German Mark (DM), British Pound (GBP) and Japanese Yen (JPY). The data are displayed in figures 1a through d. Given is the value in US dollars per unit of the foreign currency.

According to current knowledge, prediction of foreign exchange rates appears to be particularly difficult. There has been a controversial discussion in how far foreign exchange rates can be predicted at all (see e.g. Frankel and Rose 1995 and references therein; Meese and Singleton 1982, Meese and Rogoff 1983, Baillie and Bollerslev 1989, Diebold and Nason 1990, Liu and He 1991, Meese and Rose 1991, Cheung 1993, Brooks 1997). Proposed models include for instance random walk, fractional ARIMA models with long memory, macroeconomic models, nonlinear function models, GARCH models and nonparametric prediction models. In spite of the large variety of models, the success in forecasting the future development of exchange rates seems to have been rather limited. In particular, Frankel and Rose (1995) come to the conclusion that the simple random walk model appears to provide short- to medium term forecasts that are as good as or even better than more sophisticated models. It should be noted, however, that the predominant criterion for judging the performance of predictions has been the accuracy of *point* predictions, measured for instance by the mean squared prediction error. This is not be the only informative criterion. The purpose of a statistical prediction is not only to obtain a point estimate but also to have a confidence interval. Thus, an important criterion for assessing the usefulness of a statistical prediction is whether prediction intervals are realistic, i.e. neither too short nor too long. The results below indicate that, for the exchange rates considered here, prediction intervals based on random walk may indeed be unnecessarily large, even if no other information than the observed time series is used.

For a given currency, denote by Y_t the exchange rate at time t. To check

the performance of SEMIFAR predictions, we consider for each exchange rate series, 31 partially overlapping subseries $Y_1^{(j)}, ..., Y_n^{(j)}$ where $n = 250$, and $Y_i^{(j)}$ are the rescaled log-values $Y_i^{(j)} = log(Y_{i+25j}/Y_{1+25j})$ $(j = 0, ..., 30)$. Note that $n = 250$ corresponds to a period of about one year. A SEMIFAR(p) model is fitted to each subseries, where p, g and θ are estimated and forecasts (point and interval estimates) are calculated for $Y_{n+k}^{(j)}$ $(k = 1, 2, ..., 10, 15, 20, 30)$ as described above. Two types of extrapolations for the deterministic trend are used: (1) Taylor extrapolation with $r = 0$ (constant extrapolation) and (2) Taylor extrapolation with $r = 1$ (linear extrapolation). A typical prediction with $r = 0$ and $r = 1$ is plotted in figures 2a and b respectively, together with the actually observed future values.

Table 1 summarizes the essential features of the fitted models. The only chosen orders were $\hat{p} = 0$ and 1, and in most cases $\hat{p}$ was equal to zero. The first question is in how far the nominal levels of the prediction intervals are correct. Table 2 gives the percentage of "future" observations contained in the 95%-prediction intervals for $k-$steps ahead SEMIFAR forecasts. Table 3 summarizes the same for random walk predictions. As a cautionary remark, it should be noted that the percentages are obtained as an average of 31 (dependent) indicator variables. Thus, in particular, if a future value was outside the prediction interval for one of the subseries, then the observed coverage probability drops from 100 to 97%. For comparison, the coverage probabilities of random walk forecast intervals are also given. Table 2 indicates that, generally, for short-term forecasts, the observed coverage percentages are close to the nominal ones for both, $r = 0$ and $r = 1$. For long-term forecasts, linear extrapolation ($r = 1$) appears to be less reliable, whereas constant predictions yield realistic prediction intervals even upto $k = 30$ (except perhaps for the Japanese Yen where long-term forecasts seem difficult). Apparently, the deterministic trend of foreign exchange rates is far from linear, with rapidly changing slopes. A linear Taylor expansion is thus valid only in a very small neighbourhood and the best long-term "guess" of the deterministic function is the last observed value. Also note that the coverage probabilities of random walk predictions given in table 3 are similar to those of SEMIFAR predictions with $r = 0$.

Overall, table 2 indicates that *SEMIFAR* forecasts yield (approximately)

valid prediction intervals for the foreign exchange rates considered here. A second question is how the precision of *SEMIFAR* point predictions compares to random walk predictions. This question is interesting in view of the apparently good performance of random walk predictions for foreign exchange rates reported in the literature (see the references given above). Table 4 compares the mean squared prediction errors (MSE) of SEMIFAR forecasts (for $r = 0$ and $r = 1$) with those of random walk forecasts. Clearly, in terms of the MSE, random walk forecasts are competitive for the forecasting horizon $k \leq 10$. *SEMIFAR* forecasts with $r = 0$ are slightly better with respect to the MSE, but overall, in terms of the MSE not much is gained by using the more complicated *SEMIFAR* models. Consider, however, the length of *SEMIFAR* prediction intervals. Figure 3 displays the ratio of the average length of *SEMIFAR* prediction intervals (with $r = 0$) divided by the length of random walk prediction intervals, plotted against k. *SEMIFAR* intervals turn out to be considerably shorter, in particular for larger lags. The most dramatic improvement is achieved for the British Pound. At the same time, a comparison with table 3 indicates that, even for long-term forecasts, the coverage probability of random walk prediction intervals is not generally higher, although they are wider. In this sense, *SEMIFAR* predictions with $r = 0$ outperform random walk forecasts.

5 Final remarks

The main advantages of *SEMIFAR* models for forecasting may be summarized as follows:

1. A unified framework for simultaneous modelling of stationary and non-stationary short- and long-memory components as well as smooth deterministic trends is provided. This allows, in particular, for automatic data driven modelling of and distinction between (almost) arbitrary deterministic trends as well as stochastic and spurious trends. All estimated components are incorporated in the forecast.

2. In "stationarity 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 length 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 *SEMIFAR* models, prediction intervals are of the order $O(k^{\psi/2})$ with ψ varying in a continuous range, including $\psi = 0$ and $\psi = 1$ as special cases. The value of $\psi = \max(0, 2d - 1)$ is estimated from the data by maximum likelihood. As a result, prediction intervals are better adapted to the observed data, and often shorter. The extreme decision between $O(1)$ and $O(k^{1/2})$ is avoided.

Clearly, as always with forecasting, structural changes in the behaviour of the data that have not occurred in the past cannot be foreseen. Thus, for instance, sudden extreme drops in a particular exchange rate are usually due to an "artificial" intervention that can hardly be predicted from the one observed time series only. More practical experience will be needed to explore the potential of *SEMIFAR* models.

Also, a number of open problems remain. For instance, other ways of extrapolating the nonparametric trend function may be better for certain data sets. Other smoothing methods, such as local polynomials or wavelets may be useful in this context. Another open problem is the comparison of different model choice criteria. Here, the BIC was used for choosing the order p. In view of results for stationary short-memory processes (see e.g. Shibata 1978) the question may be raised whether other model choice criteria, such as the AIC (Akaike 1971), may lead to improved forecasts. Also, incorporating possible deviations from normality and in particular long-tailed distributions may improved the relialability of forecasts. Finally, better and faster algorithms for semiparametric estimation and forecasting may be developed. In particular, the forecasting formulae in theorem 1 require calculation of autocovariances for a large number of lags and the inversion of a possibly very large autocovariance matrix. In our numerical calculations, inversion of Σ_n and calculation of β_{opt} turned out to require very precise evaluation of $\gamma(k)$ for all lags. After trying a number of approaches, Monte Carlo calculation of the covariances turned out to be most reliable. Interpreting $\gamma(k) = \int_{-\pi}^{\pi} f(x) \cos kx dx$ as $2\pi E[f(U) \cos kU]$ where U is uniformly distributed on $[-\pi, \pi]$, $\gamma(k)$ was obtained by simulation from this distribu-

tion. For the models fitted to the data examples, 100000 simulations turned out to yield reliable approximations of $\gamma(k)$.

6 Acknowledgements

We would like to thank the referees for their constructive comments that lead to a better presentation of the results. Also, we would like to thank Sucharita Ghosh (WSL, Switzerland) for helpful discussions that lead to a better implementation of the prediction formulae in Splus. The data were obtained from the Web-site of *PACIFIC* (Policy Analysis Computing & Information Facility in Commerce) at the University of British Columbia, Vancouver, Canada.

7 Appendix

Proof of theorem 1:

Case 1: $m = 0$: Note that $U_{n+k} = X_{n+k}$. Let $\hat{X}_{n+k} = \beta^t X(n)$ with $X(n) = (X_1, ..., X_n)^t$. Then

$$MSE = E[(X_{n+k} - \hat{X}_{n+k})^2] = \gamma(0) + \beta^t \Sigma_n^{-1} \beta - 2\beta^t \gamma_k^{(n)}.$$

The result then follows by setting the partial derivatives with respect to β_j equal to zero.

Case 2: $m = 1$: Note that $U_{n+k} = \sum_{j=1}^{k} X_{n+k}$. Then

$$MSE = E[(U_{n+k} - \hat{X}_{n+k})^2] = \sum_{s=-(k-1)}^{k-1} (k - |s|)\gamma(s) + \beta^t \Sigma_n^{-1} \beta - 2\beta^t \tilde{\gamma}_k^{(n)}.$$

As above, the result then follows by setting the partial derivatives with respect to β_j equal to zero.

Proof of corollary 1:

First note that, as $k \to \infty$, the components of $\gamma_k^{(n)}$ and $\tilde{\gamma}_k^{(n)}$ converge to zero. Hence, $MSE_k \sim \gamma(0)$ for $m = 0$, and for $m = 1$

$$MSE \sim \sum_{s=-(k-1)}^{k-1} (k - |s|)\gamma(s).$$

Now, for $\delta = 0$,

$$\sum_{s=-(k-1)}^{k-1} (k - |s|)\gamma(s) \sim n \sum_{s=-\infty}^{\infty} \gamma(s) = n2\pi c_f = nc_f\nu(0).$$

For $0 < \delta < 0.5$,

$$\sum_{s=-(k-1)}^{k-1} (k - |s|)\gamma(s) \sim 4c_f\Gamma(1 - 2\delta)\sin\pi\delta \sum_{s=1}^{k-1}(k - |s|)|s|^{2\delta-1}$$

(see e.g. Beran 1994, p.43) and the result follows by interpreting the sum as $k^{2\delta+1}$ times a Rieman sum. For $0 < \delta < 0.5$, $\sum_{s=-\infty}^{\infty} \gamma(s) = 0$ so that

$$\sum_{s=-(k-1)}^{k-1} (k - |s|)\gamma(s) = 2k \sum_{s=k}^{\infty} \gamma(s) - \sum_{s=-(k-1)}^{k-1} |s|\gamma(s).$$

The result then again follows by interpreting the sum as $k^{2\delta+1}$ times a Rieman sum.

BIBLIOGRAPHY

Akaike, H. (1971) Information theory and an extension of the maximum likelihood priciple. In 2nd Int. Symposium on Information Theory, B.N. Petrov and F. Csaki (eds.), pp. 267-281, Akademia Kiade, Budapest.

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.

Baillie, R.T. and Bollerslev, T. (1989) The message in daily exchange rates: a conditional-variance tail. Journal of Business and Economic Statistics, **7, No. 3**, 297-305.

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**, 659-672.

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

Brooks, C. (1997) Linear and non-linear (non-)forecastability of high-frequency exchange rates. Journal of Forecasting, **16**, 125-145.

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

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.

Diebold, F.X. and Nason, J.A. (1990) Nonparametric exchange rate prediction. Journal of International Economics, **315-332**.

Frankel, J.A. and Rose, A.K. (1995) A survey of empirical research on nominal exchange rates. CIDER, University of California, Berkeley, Working Paper No. C95-051.

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.

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

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

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

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

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.

Meese, R.A. and Rogoff, K. (1983) Empirical evidence rate models of the seventies - Do they fit out of sample? Journal of International Economics, **14**, 3-24.

Meese, R.A. and Rose, A.K. (1991) An empirical assessment of nonlinearities in models of exchange rate determination. Review of Economic Studies, **58**, 603-619.

Meese, R.A. and Singleton, K.J. (1982) On unit roots and the empirical modeling of exchange rates. The Journal of Finance, **XXXVII, No.4**, 1029-1035.

Phillips, P.C.B. (1997) Unit-root tests. In *Encyclopedia of Statistical Sciences*, S. Kotz, C.R. Read, D.L. Banks (editors), New York: Wiley, pp. 531-542.

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.

Shibata, R. (1976) Selection of the order of an autoregressive model. Ann. Statist., **8**, 147-164.

Table 1: Sample means, standard deviations and ranges of $\hat{d}$ obtained from *SEMIFAR*-fits to 31 (partially overlapping) subseries of four foreign exchange rates (CHF, DEM, GBP, JPY). The values in brackets are the corresponding quantities for the subseries with $\hat{p} = 0$ and $\hat{p} = 1$ respectively. Other values of $\hat{p}$ did not occur. The number of subseries with $\hat{p} = 0$ was equal to 24 for CHF, DEM and GPB and 29 for JPY.

Currency	mean	standard deviation	range
CHF	0.83 (0.97, 0.35)	0.30 (0.06, 0.32)	[-0.04, 1.09] ([0.87, 1.09], [-0.04, 0.65])
DEM	0.81 (0.92, 0.42)	0.28 (0.10, 0.36)	[-0.02, 1.04] ([0.60, 1.04], [-0.02, 0.97])
GBP	0.75 (0.88, 0.39)	0.29 (0.05, 0.38)	[-0.09, 1.01] ([0.80, 1.01], [-0.09, 0.93])
JPY	0.87 (0.89, 0.59)	0.11 (0.09, 0.003)	[0.59, 1.02] ([0.67, 1.02], [0.59, 0.59]

Table 2: Empirical coverage percentages of the $100(1-\alpha)\%-$ $k-$step ahead prediction intervals with $1-\alpha = 0.95$ and 0.99. The percentages were obtained from 31 *SEMIFAR* predictions with $r = 0$ and $r = 1$, each based on one of 31 (partially overlapping) subseries of four foreign exchange rates (CHF, DEM, GBP, JPY).

Currency	r	$1-\alpha$	$k=1$	2	3	4	5	6	7	8	9	10	15	20	30
CHF	0	0.95	97	97	97	97	100	94	97	94	90	90	97	94	90
		0.99	97	100	100	100	100	94	97	94	97	97	97	100	97
	1	0.95	97	94	94	97	97	94	90	87	90	87	87	87	87
		0.99	97	100	97	100	100	97	97	94	94	90	94	94	97
DEM	0	0.95	97	94	94	100	100	94	94	94	94	94	97	100	100
		0.99	97	100	100	100	100	97	97	94	97	97	97	100	100
	1	0.95	97	94	87	94	94	87	94	87	87	87	90	90	84
		0.99	97	97	100	100	100	94	94	90	90	97	97	97	87
GBP	0	0.95	100	84	94	90	90	97	94	90	90	94	87	90	87
		0.99	100	100	97	100	100	97	97	94	97	100	94	97	97
	1	0.95	100	87	90	90	94	94	94	87	87	94	84	84	77
		0.99	100	97	97	100	100	94	100	90	97	100	94	94	90
JPY	0	0.95	97	94	97	90	97	94	90	84	87	87	87	94	87
		0.99	100	100	100	100	100	97	97	94	97	90	94	94	94
	1	0.95	94	90	94	90	94	90	90	84	84	84	81	90	81
		0.99	97	100	100	94	100	97	97	90	90	90	87	90	84

Table 3: Empirical coverage percentages of the $100(1-\alpha)\%-$ $k-$step ahead prediction intervals with $1-\alpha = 0.95$ and 0.99, and the random walk model. The percentages were obtained from 31 random walk predictions, each based on one of 31 (partially overlapping) subseries of four foreign exchange rates (CHF, DEM, GBP, JPY).

Currency	$1-\alpha$	$k=1$	2	3	4	5	6	7	8	9	10	15	20	30
CHF	0.95	97	94	94	97	97	97	97	90	90	87	90	97	94
	0.99	97	97	97	100	100	97	97	94	94	94	97	100	100
DEM	0.95	97	94	90	97	97	94	94	87	94	90	97	97	94
	0.99	97	97	100	100	100	97	97	94	97	100	100	100	100
GBP	0.95	100	90	97	97	100	94	94	94	97	100	94	97	94
	0.99	100	100	100	100	100	97	100	97	97	100	100	100	97
JPY	0.95	94	100	97	94	94	97	97	90	90	90	87	94	90
	0.99	97	100	100	100	100	100	97	97	97	97	94	97	94

Table 4: Empirical mean squared prediction errors of $k-$step ahead prediction intervals using *SEMIFAR* forecasts with $r = 0$ and $r = 1$, and random walk respectively. The results were obtained from 31 (partially overlapping) subseries of four foreign exchange rates (CHF, DEM, GBP, JPY). The values of the MSE are multiplied by 10^4.

Currency	model	$k=1$	2	3	4	5	6	7	8	9	10	15	20	30
CHF	random walk	0.5	1.0	1.6	1.9	1.6	3.2	4.2	6.7	7.2	6.5	8.1	7.7	13.1
	SEMIFAR $r=0$	0.5	0.9	1.5	1.8	1.5	3.1	4.0	6.2	6.8	6.3	8.1	7.8	13.3
	SEMIFAR $r=1$	0.6	1.0	1.8	2.3	2.2	4.0	5.1	7.7	8.3	8.3	11.9	13.4	26.1
DEM	random walk	0.5	0.9	1.2	1.6	1.3	2.6	3.3	4.7	5.1	4.9	6.0	6.7	8.5
	SEMIFAR $r=0$	0.5	0.9	1.2	1.6	1.3	2.5	3.2	4.5	5.0	4.8	6.3	7.6	9.6
	SEMIFAR $r=1$	0.5	0.9	1.3	1.8	1.6	2.9	3.6	4.8	5.1	5.3	7.2	10.5	18.2
GBP	random walk	0.2	0.7	0.8	1.2	1.1	1.7	1.6	2.4	2.4	2.0	2.9	2.9	4.8
	SEMIFAR $r=0$	0.2	0.7	0.9	1.2	1.1	1.7	1.8	2.6	2.6	2.3	3.2	3.5	5.4
	SEMIFAR $r=1$	0.2	0.7	0.8	1.2	1.1	1.8	1.8	2.6	2.7	2.5	3.8	4.9	8.8
JPY	random walk	0.4	0.9	1.5	2.0	1.7	2.1	3.1	4.8	5.3	5.3	10.6	8.0	14.7
	SEMIFAR $r=0$	0.4	0.9	1.5	2.0	1.8	2.3	3.3	5.1	5.7	5.9	11.3	8.8	14.8
	SEMIFAR $r=1$	0.5	1.1	1.6	2.5	2.5	3.2	3.9	5.4	6.4	6.8	14.6	13.3	29.0