Recent Developments in Non- and Semiparametric Regression with Fractional Time Series Errors

Jan Beran and Yuanhua Feng

University of Konstanz

Abstract

This paper summarizes recent developments in non- and semiparametric regression with stationary fractional time series errors, where the error process may be short-range, long-range dependent or antipersistent. The trend function in this model is estimated nonparametrically, while the dependence structure of the error process is estimated by approximate maximum likelihood. Asymptotic properties of these estimators are described briefly. The focus is on describing the developments of bandwidth selection in this context based on the iterative plug-in idea (Gasser et al., 1991) and some detailed computational aspects. Applications in the framework of the SEMIFAR (semiparametric fractional autoregressive) model (Beran, 1999) illustrate the practical usefulness of the methods described here.

Keywords: Nonparametric regression, FARIMA error processes, bandwidth selection, iterative plug-in, SEMIFAR model.

1 Introduction

Nonparametric regression has become a rapidly developing field of statistics in the recent years. Most of the contributions in this area focus on models with independent identically distributed (i.i.d.) (or at least uncorrelated) errors (see the monographs of Müller, 1988, Härdle, 1990, Fan and Gijbels, 1996, Wand and Jones, 1996 and Eubank, 1999 among others). Most literature on nonparametric regression with dependent errors focuses on some special type of short memory cases where asymptotic results are identical to those in the case of i.i.d. errors (see e.g. Bierens, 1983). However, in general cases (in cases with long memory, antipersistence and in most cases with short memory), the finite sample and asymptotic performances of a nonparametric regression estimator will be changed so that theoretical results, computational algorithms developed based on the i.i.d. assumption are no longer applicable (see e.g. Altman, 1990, Hall and Hart, 1990, Hart, 1991, Herrmann et al., 1992, Ray and Tsay, 1997, Beran, 1999 and Beran and Feng, 2002a, b).

This paper is devoted to summarize recent results on equidistant, non- and semiparametric regression with a stationary fractional time series error process, including long memory, short memory and antipersistence. The focus is on the investigating data-driven algorithms based on the iterative plug-in (called IPL in this paper) idea (Gasser et al., 1991). Some related computational aspects will also be discussed. The parameter estimation and applications are carried out in the framework of the SEMIFAR model (Beran, 1999). Research on random design nonparametric regression with strongly dependent errors can be found e.g. in Csörgö and Mielniczuk (1995, 1999). Note in particular that asymptotic results for fixed and random design nonparametric regression are not the same, if the data have long memory, since the effect of the dependence does not play the same roll in the two designs. Another recent review of nonparametric regression with correlated errors may be found in Opsomer et al. (2000), where the emphasis is quite different to the topics discussed here.

The paper is organized as follows. Kernel and local polynomial estimators are described in Section 2 with a brief summary of their asymptotic properties. Section 3 is devoted to the crucial problem of bandwidth selection, where an IPL algorithm and related computational aspects are discussed in detail. Estimation of the unknown parameters is discussed in Section 4. Data examples illustrate the practical usefulness of the SEMIFAR model in Section 5. Final remarks in Section 6 conclude the paper.

2 Nonparametric regression with dependent errors

Research on the topics discussed here began about ten years ago. The model considered is the equidistant nonparametric regression

$$Y_i = g(x_i) + \xi_i, \qquad i = 1, ..., n, \tag{1}$$

where $x_i = i/n$, $g : [0, 1] \to \Re$ is a smooth function and ξ_i is a second order and strict stationary process generated by an (at least) uncorrelated innovation series ε_i through a linear filter. Denote by $\gamma(k) = \text{cov}\,(\xi_i, \xi_{i+k})$ the autocovariances of ξ_i. It is assumed that $\gamma(k) \to 0$ as $|k| \to \infty$. Equation (1) represents a nonparametric regression model with short memory (including i.i.d. ξ_i as a special case), long memory and antipersistence. Here, a stationary process ξ_i is said to have long memory (or long-range dependence), if the spectral density $f(\lambda) = (2\pi)^{-1} \sum \gamma(k) \exp(ik\lambda)$ has a pole at the origin of the form

$$f(\lambda) \sim c_f |\lambda|^{-2\delta} (\text{as } \lambda \to 0) \tag{2}$$

for some $\delta \in (0, 0.5)$, where $c_f > 0$ is a constant and '$\sim$' means that the ratio of the left and the right hand sides converges to one (see Beran, 1994 and references therein). Note that, for $\delta \in (0, 0.5)$, (2) implies that $\gamma(k) \sim c_\gamma |k|^{2\delta-1}$ so that $\sum \gamma(k) = \infty$. Hence now ξ_i has long memory. If (2) holds with $\delta = 0$, then we have $0 < \sum \gamma(k) = 2\pi c_f < \infty$ and ξ_i is said to have short memory, including i.i.d. errors and all causal and invertible Box-Jenkins ARMA processes (Box and Jenkins, 1976) as special cases. On the other hand, a stationary process is said to be antipersistent, if (2) holds with $\delta \in (-0.5, 0)$ implying that $\sum \gamma(k) = 0$. Most of the results described in this paper are valid for all $\delta \in (-0.5, 0.5)$.

Altman (1990) and Hart (1991) proposed kernel estimator of g for errors with short memory. This was extended to the case with long-memory errors by Hall and Hart (1990). Beran (1999) (see also Beran and Feng, 2002a) investigated kernel estimation of g for δ in the whole range $(-0.5, 0.5)$. Note in particular that the derivation of the properties of a kernel estimator with antipersistent errors requires different techniques due to the fact that $\sum \gamma(k) = 0$. The results in Beran (1999) are generalized to local polynomial estimation of $g^{(\nu)}$, the ν-th derivative of g, by Beran and Feng (2002b). Note that the definitions of the kernel and local polynomial estimators under model (1) are the same as for nonparametric regression with i.i.d. errors, since both estimators are linear smoothers. Let $K_{(\nu,k)}(u)$ denote a kernel function of order k for estimating $g^{(\nu)}$ (see e.g. Gasser et al., 1985 and Müller, 1988) with compacted support $[-1, 1]$. For $x \in (0, 1)$, a kernel estimator of $g^{(\nu)}(x)$ (of the Nadaraya-Watson-Type) is given by

$$\hat{g}^{(\nu)}(x; h) = \frac{1}{nh^{\nu+1}} \sum_{i=1}^{n} K_{(\nu,k)}(\frac{x_i - x}{h}) Y_i, \tag{3}$$

where h is the bandwidth. See Nadaraya (1964) and Watson (964) for the original proposal of $\hat{g}$. Note that for equidistant design there is no significant difference between the definition (3) and the Gasser-Müller estimator (Gasser and Müller, 1984).

It is well known that a Kernel estimator is affected by the so-called boundary effect (Gasser and Müller, 1979). A well known estimator with automatic boundary correction is the local polynomial approach introduced by Stone (1977) and Cleveland (1979). For detailed discussion on local polynomial fit see e.g. Ruppert and Wand (1994) and Fan and Gibels (1995, 1996). Let K be a second order kernel with compact support $[-1, 1]$. Let h denote the bandwidth. The local polynomial fit of $g^{(\nu)}$ ($\nu \leq p$) is obtained by solving the weighted least squares problem

$$Q = \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} \beta_j (x_i - x_0)^j \right\}^2 K\left(\frac{x_i - x_0}{h}\right) \Rightarrow \min. \tag{4}$$

Here $\nu!\hat{\beta}_\nu$ estimates $g^{(\nu)}$. Local polynomial fit in nonparametric regression with fractional time series errors is investigated in Beran and Feng (2002b). It is well known that local polynomial fit is asymptotically equivalent to some kernel estimates (see e.g. Müller, 1987 and Hastie and Loader, 1993). This relationship even holds in the boundary area, provided that a corresponding boundary kernel is used (see Feng, 1999). These two estimators are also asymptotically equivalent under model (1) (see Beran and Feng, 2002b). Because of this equivalence, we propose to carry out a local polynomial fit using the bandwidth selected with the corresponding kernel estimator, since a kernel estimator is computationally much simpler than a local polynomial fit and a data-driven procedure based on a kernel estimator runs much faster.

In the following we will therefore restrict attention to the kernel estimator on the interval $[\Delta, 1-\Delta]$, where $\Delta > 0$ is introduced to avoid the boundary effect (see Härdle et al., 1988). The formulae of the asymptotic bias of a kernel estimator do not depend on the dependence structure and will hence be omitted. It is well known that the change in the asymptotic variance of $\hat{g}^{(\nu)}$ is just a constant, if the errors have short memory (Altman, 1990, Hall and Hart, 1990 and Hart, 1991). However, the order of magnitude of $\text{var}\,(\hat{g}^{(\nu)})$ changes, if the errors are long-range dependent or antipersistent (Beran, 1999 and Beran and Feng, 2002a, b). The variance of $\hat{g}^{(\nu)}$ at a point $x \in [\Delta, 1-\Delta]$ is given by

$$\text{var}\,[\hat{g}^{(\nu)}(x)] \doteq (nh)^{2\delta-1} h^{-2\nu} V \tag{5}$$

for all ν and all $\alpha \in (-1, 1)$, where V is a constant. This result shows that the variance of $\hat{g}$ converges slower to zero than for i.i.d. errors, if the errors have long memory and faster, if the errors are antipersistent. This result reduces to the well known formula of the asymptotic variance of a kernel estimator with i.i.d. errors, if $\delta = 0$ and V is replaced by

$$V = \sigma^2 \int_{-1}^{1} K^2_{(\nu,k)}(x)dx. \tag{6}$$

If the uniform kernel is used, then we have

$$V = \frac{2^{2\delta} c_f \Gamma(1-2\delta)\sin(\pi\delta)}{\delta(2\delta+1)} \tag{7}$$

for all $\delta \in (-0.5, 0.5)$ (see corollary 1 in Beran, 1999). Explicit formulae for V in general cases are given in Beran and Feng (2002b). Earlier results for $\nu = 0$ and $\delta > 0$ are given by Hall and Hart (1990).

As a goodness of fit criterion, the MISE (mean integrated squared error) defined on $[\Delta, 1-\Delta]$ will be used. We have

$$\text{MISE} \doteq h^{2(k-\nu)} \frac{I(g^{(k)})\beta^2_{(\nu,k)}}{(k!)^2} + (nh)^{2\delta-1} h^{-2\nu}(1-2\Delta)V, \tag{8}$$

where $\beta_{(\nu,k)}$ denotes the kernel constant and

$$I(g^{(k)}) = \int_{\Delta}^{1-\Delta} [g^{(k)}(x)]^2 dx \,. \tag{9}$$

The asymptotically optimal bandwidth, which minimizes the dominating part of (8) is

$$h_{\mathrm{A}} = C_{\mathrm{opt}} n^{(2\delta-1)/(2k+1-2\delta)} \,, \tag{10}$$

with

$$C_{\mathrm{opt}} = \left[\frac{2\nu+1-2\delta}{2(k-\nu)} \frac{(k!)^2(1-2\Delta)V}{I(g^{(k)})\beta^2_{(\nu,k)}} \right]^{1/(2k+1-2\delta)} \,. \tag{11}$$

This result shows that, compared to i.i.d. data, not only the constant but also the order of magnitude of h_{A} are changed, if the errors are long-range dependent or antipersistent. A bandwidth of larger order is required in the former case and a bandwidth of smaller order in the latter.

Furthermore, let h_{M} denote the optimal bandwidth which minimizes the MISE. It can be shown that the difference between h_{A} and h_{M} is given by

$$(h_{\mathrm{A}} - h_{\mathrm{M}})/h_{\mathrm{M}} \doteq O(h_{\mathrm{M}}^2) = O(n^{2(2\delta-1)/(2k+1-2\delta)}) \tag{12}$$

for all ν, k and $\delta \in (-0.5, 0.5)$ (Beran and Feng, 2002c). For i.i.d. errors with $\nu = 0$, this result reduces to the well known equation:

$$(h_{\mathrm{A}} - h_{\mathrm{M}})/h_{\mathrm{M}} \doteq O(n^{-2/5}) \tag{13}$$

for $k = 2$ and

$$(h_{\mathrm{A}} - h_{\mathrm{M}})/h_{\mathrm{M}} \doteq O(h_{\mathrm{M}}^2) = O(n^{-2/(2k+1)}) \tag{14}$$

for arbitrary (even) k (see e.g. Herrmann and Gasser, 1994).

If a bandwidth of the optimal order $O(n^{(2\delta-1)/(2k+1-2\delta)})$ is used, then the rate of convergence of $\hat{g}^{(\nu)}$ is of order $O_p(n^{(2\delta-1)(k-\nu)/(2k+1-2\delta)})$. Under the condition that ξ_i is a linear process with i.i.d. Gaussian innovations, Hall and Hart (1990) showed that, for $\hat{g}$ with $\alpha \geq 0$, this rate of convergence is optimal in the minimax sense. In a recent paper, Feng (2002a) showed that this is the minimax optimal rate of convergence of a nonparametric regression estimator for all ν, k and $\delta \in (-0.5, 0.5)$. This result also holds for non-Gaussian innovations satisfying some regular distribution conditions.

Another question is, whether $\hat{g}^{(\nu)}$ is asymptotically normal? Recall that ξ_i is generated by a linear filter of the innovations ϵ_i. The assumption that ϵ_i are uncorrelated $(0, \sigma^2)$

random variables is not sufficient for the asymptotic normality of $\hat{g}^{(\nu)}$ (see e.g. Taqqu, 1975). On the other hand, if ϵ_i are i.i.d. $(0, \sigma^2)$ random variables, i.e. if ξ_i is a linear process, then under regular conditions $\hat{g}^{(\nu)}$ is asymptotically normal (see e.g. Ibragimov and Linnik, 1971). Beran and Feng (2001a) obtained some weaker sufficient conditions for the asymptotic normality of $\hat{g}^{(\nu)}$. As a special case, it is shown that, $\hat{g}^{(\nu)}$ is asymptotically normally distributed, if ϵ_t follow a GARCH model (generalized autoregressive conditional heteroskedastic, Engle, 1982 and Bollerslev, 1986) with finite fourth moments and ξ_t is generated by ϵ_i through a FARIMA model (fractional ARIMA), i.e. if ξ_i is a FARIMA-GARCH model with finite fourth moments (see e.g. Ling and Li, 1997).

3 Bandwidth selection

A key point for the practical implementation of a nonparametric approach is the selection of the bandwidth. Numerous approaches are proposed to perform this in nonparametric regression with i.i.d. errors. Well know traditional methods are the CV (cross-validation, Clark, 1975), the GCV (generalized CV, Graven and Wahba, 1979) and the R-criterion (Rice, 1984) among others. For a survey on traditional proposals see Härdle et al. (1988). It is well known that all of the traditional methods share large sample variation and the very slow rate of convergence $O_p(n^{-1/10})$. In recent years, some *modern bandwidth selectors* are proposed including the IPL approach (Gasser et al., 1991, Herrmann, 1994 and Herrmann and Gasser, 1994), the direct plug-in approach (Ruppert et al., 1995), the double-smoothing method (Gasser et al., 1984, Müller, 1985, Härdle et al., 1992, Heiler and Feng, 1998, Feng, 1999 and Feng and Heiler, 2000) and another approach closely related to the double-smoothing idea (Fan and Gijbels, 1995). Feng (1999) and Feng and Heiler (2000) showed that the double-smoothing bandwidth selection rule can be explained as a criterion obtained by bootstrap in nonparametric regression (Härdle and Bowman, 1988). In a recent paper (Beran, Feng and Heiler, 2000) a bandwidth selector which combines the plug-in and the double-smoothing ideas, was proposed. The key idea of Beran et al. (2000) is: bootstrapping the bias and plugging-in the variance. The IPL idea is also extended to select bandwidth in multivariate nonparametric regression (Herrmann et al., 1995).

In recent years it has been noticed that a bandwidth selector developed for nonparametric regression with i.i.d. error performs very badly, when the errors are correlated (see e.g. Altman, 1990 and Herrmann et al., 1992). A data-driven procedure for nonparametric regression with i.i.d. errors tends to select smaller bandwidths resulting in undersmoothing when the correlations are predominantly positive and larger bandwidths resulting in oversmoothing when negative. For instance for a plug-in method, two main reasons for this phenomenon are: 1. The constant in h_{A} is changed, if the errors are short-range dependent and 2. The method for estimating this constant based on the i.i.d. assumption is wrong, if the errors are correlated. These problems become even worse when the data are long-range dependent or antipersistent due to the strong change of the dependence structure and the change of the order of h_{A}.

This motivates the development of bandwidth selectors for nonparametric regression with correlated errors. Many well known bandwidth selection criteria are adapted to select bandwidth in nonparametric regression with short-range dependent errors, see e.g. Altman (1990) for adaptations of the CV and the GCV, Hart (1991) for a modified R-criterion, Herrmann et al. (1992) for an adapted IPL procedure and Chiu (1989) for another proposal. To adapt the approaches to select bandwidth in nonparametric regression with long-range dependent or antipersistent errors is however more difficult. This is the main topic of the rest part of this section. Hereafter we will focus only on the IPL idea, because of its higher rate of convergence (compared to the CV, the GCV and the R-criterion), stability, computational simplicity and wide applicability.

To our knowledge, a bandwidth selector in nonparametric regression with long memory is proposed first by Ray and Tsay (1997) by modifying the IPL idea of Gasser et al. (1991) and Herrman et al. (1992). This proposal was further modified by Beran (1999) for selecting the bandwidth in the SEMIFAR model, where also the bandwidth selection in nonparametric regression with antipersistent errors is considered. Recent research in this context can be found in Beran and Feng (2002b, c), where the idea in Beran (1999) is discussed in more detail. Note in particular that, in Beran (1999) and Beran and Feng (2002a, c), the trend function, the dependence structure and the nonstationarity in a time series are modelled simultaneously. A large simulation study showed that, the IPL idea works well for nonparametric regression with fractional time series errors.

An IPL bandwidth selector in the current context is obtained based on (10) and (11) by replacing the unknowns δ, V and $I(g^{(k)})$ with proper estimates. The key question is how should these unknowns be estimated? Estimation of δ and V will be discussed in the next section. In the following we will discuss the estimation of $I(g^{(k)})$, provided that approaches for estimating δ and V are given beforehand. A natural estimator of $I(g^{(k)})$ is

$$\hat{I}(g^{(k)}) = n^{-1} \sum_{i=[n\Delta]}^{n-[n\Delta]} [\hat{g}^{(k)}(t_i)]^2 \tag{15}$$

with a bandwidth h_k, where $[\cdot]$ denotes the integer part. Let $L_{(k,l)}$ denote the l-th order kernel for estimating $g^{(k)}$. Then under given conditions we have

$$E[\hat{I}(g^{(k)}) - I(g^{(k)})] \doteq 2h_k^{(l-k)}\frac{\beta_{(k,l)}}{l!}\int_{\Delta}^{1-\Delta} g^{(k)}(t)g^{(l)}(t)dt + (nh_k)^{2\delta-1}h_k^{-2k}(1-2\Delta)V \quad (16)$$

and

$$\text{var}\,[\hat{I}(g^{(k)})] \doteq o[(nh_k)^{(4\delta-2)}h_k^{-4k}] + O(n^{2\delta-1}). \quad (17)$$

These results are given by Gasser et al. (1991), Herrmann and Gasser (1994) and Ruppert et al. (1995) for i.i.d. errors and Beran and Feng (2002b) for fractional time series errors.

Note that, in general, the mean squared error (MSE) of $\hat{I}(g^{(k)})$ is dominated by the squared bias, i.e.

$$\text{MSE}\{\hat{I}(g^{(k)})\} \doteq \left\{2h_k^{(l-k)}\frac{\beta_{(k,l)}}{l!}\int_{\Delta}^{1-\Delta} g^{(k)}(t)g^{(l)}(t)dt + (nh_k)^{2\delta-1}h_k^{-2k}(1-2\Delta)V\right\}^2.$$

The optimal bandwidth for estimating $I(g^{(k)})$ which minimizes the MSE is of the order $O(n^{(2\delta-1)/(k+l+1-2\delta)})$. This bandwidth is not the same as the optimal one for estimating $g^{(k)}$ itself. If a bandwidth $h_k = O(n^{(2\delta-1)/(k+l+1-2\delta)})$ is used, then we have $\text{MSE}(\hat{I}(g^{(k)})) = O(n^{2(l-k)(2\delta-1)/(k+l+1-2\delta)})$. In the most important special case with $k = 2$, $l = 4$, the optimal choice is $h_k = O(n^{(2\delta-1)/(7-2\delta)})$ which results in $\text{MSE}\big(\hat{I}(g'')\big) = O(n^{4(2\delta-1)/(7-2\delta)})$.

We see that for selecting the bandwidth h we have to at first select a *pilot bandwidth* h_k for estimating $I(g^{(k)})$. This seems to be paradoxical, but it is a problem faced by all modern bandwidth selection rules. The IPL idea is motivated by fixpoint search (Gasser et al, 1991 and Herrmann and Gasser, 1994). Starting with an h_0, $h_{k,j}$ is calculated from h_{j-1} with an inflation method. The original inflation method proposed by Gasser et al. (1991) is $h_{k,j} = h_{j-1} \cdot n^{\alpha}$, called the *multiplicative inflation method* (MIM), where α (the so-called inflation factor) is a suitably chosen constant. This idea was also used in Herrmann et al. (1992) and Ray and Tsay (1997) and was discussed in detail in Herrmann and Gasser (1994). Beran (1999) introduced another inflation method $h_{k,j} = h_{j-1}^{\alpha}$, called the *exponential inflation method* (EIM). This idea is discussed in detail in Beran and Feng (2002b, c). Note that, the rate of convergence of an IPL bandwidth selector using the MIM and the EIM is the same, if corresponding inflation factors α are used (see later). The EIM was introduced to reduce the required number of iterations. It can be shown that, under same conditions, the required number of iterations using the EIM is much smaller then that using the MIM (see Beran and Feng, 2002b for examples). This plays a more important role, if the errors are long-range dependent, since in this case an IPL procedure using the MIM requires too many iterations (see Ray and Tsay, 1997).

The following algorithm was proposed by Beran and Feng (2002b) using the EIM. It works well for nonparametric regression with short- and long-range dependent and antipersistent errors. Hereafter only estimation of g with $k = 2$ and $l = 4$ will be considered for simplicity. The algorithm is defined as follows:

i) Start with the bandwidth $h_0 = \Delta_0 n^{-1/3}$;

ii) For $j = 1, 2, ...$ estimate g using h_{j-1} and let $r_i = y_i - \hat{g}(t_i)$. Estimate δ and V from r_i with an appropriate method;

iii) Set $h_{2,j} = h_{j-1}^{\alpha}$ with $\alpha = (5 - 2\hat{\delta})/(7 - 2\hat{\delta})$ and set

$$h_i = \left(\frac{1 - 2\hat{\delta}}{\beta_{(\nu,2)}^2} \frac{(1 - 2\Delta)\hat{V}}{\hat{I}(g''(t; h_{2,j}))} \right)^{1/(5-2\hat{\delta})} \cdot n^{(2\hat{\delta}-1)/(5-2\hat{\delta})}, \tag{18}$$

vi) Increase j by 1 and repeat steps *ii*) and *iii*) until convergence is reached or until a given maximum number of iterations have been done. And set $\hat{h} = h_j$.

In a semiparametric model, the starting bandwidth has to satisfy the condition $h_0 \to 0$ and $nh_0 \to \infty$ as $n \to \infty$, since the unknown parameters have to be estimated in the first iteration. Here we propose to use $h_0 = \Delta_0 n^{-1/3}$ as a default starting bandwidth, where Δ_0 is a small positive number and $n^{-1/3}$ is the smallest possible order of the optimal bandwidth with any $\delta \in (-0.5, 0.5)$. It is well known that the choice of h_0 does not change the rate of convergence of $\hat{h}$. In the case when δ is known or when there is an estimate of δ, it is preferable to start with an h_0 of order $n^{(2\delta-1)/(5-2\delta)}$ (see the algorithm proposed in the next section).

The bandwidth $\hat{h}$ selected by an IPL method is a fixpoint of this procedure. For many data sets there exists only one fixpoint. In this case, $\hat{h}$ does not depend on h_0. However, sometimes there exist several fixpoints for a given data set (see Herrmann and Gasser, 1994 and Feng, 2002b, c for examples). The "reasonable" bandwidth is the one obtained by starting with a moderate h_0. The same $\hat{h}$ is achieved for all h_0 in a proper interval (which depends however on the data set). Another fixpoint is selected, if h_0 lies outside this interval. We propose to use a default h_0 with the hope that this h_0 lies in the proper interval for almost all practical data sets. In case of doubt, one may run the program with several different h_0's, to find all possible fixpoints for a given data set and then select the optimal bandwidth from these fixpoints by further analysis. For more details see Feng (2002b, c).

The inflation factor $\alpha = (5 - 2\hat{\delta})/(7 - 2\hat{\delta})$ is chosen in order that the MSE of $\hat{I}(g'')$ is of the optimal order, when convergence is reached. The optimal choice of α is $\alpha =$

$2(1-2\delta)/[(5-2\delta)(7-2\delta)]$ for the MIM (Beran and Feng, 2002b), which is $\alpha = 2/35$ for $\delta = 0$ (see Herrmann and Gasser, 1994). The choice of $\alpha = (1-2\hat{\delta})/(2(5-2\hat{\delta}))$ for the MIM used in Gasser et al. (1991), Herrmann et al. (1992) and Ray and Tsay (1997) is made so that the variance of $\hat{I}(g'')$ given in (17) achieves the order $O(n^{(2\delta-1)})$, which results in a variance term of order $O(n^{(2\delta-1)/2})$ in the selected bandwidth. It is well known that, for $\delta = 0$, $O(n^{-1/2})$ is the lower bound of the rate of convergence of any bandwidth selectors (see Hall and Marron, 1991). We think, for $\delta > 0$, $O(n^{(2\delta-1)/2})$ should be the lower bound of the rate of convergence of any bandwidth selectors. Hence with this choice of α we obtain a most stable bandwidth selector. The most stable choice of α for the EIM is simply $\alpha = 1/2$. Another possibility is to choose α so that $\hat{g}''$ is optimized (see Beran, 1999 and Beran and Feng, 2002a, b). Although the most stable choice of α works well for the MIM. The corresponding choice, i.e. $\alpha = 1/2$ for the EIM does not work well for small n, since now the inflation with the formula $h_{2,j} = h_{j-1}^{1/2}$ is too strong. Hence, for the EIM, the MSE optimal choice of α is both theoretically and practically preferable.

Beran and Feng (2002b) show

$$\hat{h} = h_{\rm M}\left\{1 + O(n^{2(2\delta-1)/(5-2\delta)}) + O_p(n^{2(2\delta-1)/(7-2\delta)})\right\}. \tag{19}$$

The $O(n^{2(2\delta-1)/(5-2\delta)})$ term in (19) is due to the difference between $h_{\rm A}$ and $h_{\rm M}$, which provides a natural bound for the rate of convergence of a plug-in bandwidth selector. Note however that this term is asymptotically negligible compared to the error in $\hat{I}(g'')$.

Remark 1. Note that for an IPL bandwidth selector, only the order of magnitude of the pilot bandwidth is considered. The constant in h_2 is ignored. This ensures the computational simplicity of an IPL algorithm. It is shown by numerous variants of the IPL idea that this simplification works well in practice, since the behaviour of $\hat{h}$ is mainly determined by the order of magnitude of h_2.

Remark 2. In the case when $\int g''(x)g^{(4)}(x)dx < 0$ the MSE($\hat{I}(g'')$), and hence the performance of $\hat{h}$, can be further improved, provided that the constant in h_2 is properly estimated so that the two dominate terms in (16) sum up to zero (see Ruppert et al., 1995, Heiler and Feng, 1998 and Feng, 1999). However, this is not discussed here due to the additional computational requirements for estimating $g^{(4)}$ and the constant in h_2.

4 A semiparametric framework

Note that estimation of δ and V is equivalent to that of δ and c_f in (2), where δ is the long-memory parameter and c_f determines the short-range dependence structure of the error process. These two parameters can be estimated semiparametrically. Ray and Tsay (1997) propose to estimate δ at first nonparametrically using the method introduced by Geweke and Porter-Hudak (1983) based on the log-periodogram, and then estimate c_f parametrically under the parametric assumption that ξ_i follows a FARIMA model. Note that there are some problems for $\hat{\delta}$ obtained based on the log-periodogram (see e.g. Beran, 1994. pp. 96-97). Furthermore, under the assumption that ξ_i follows a FARIMA model, δ can also be estimated semiparametrically following the approximate maximum likelihood proposed by Beran (1995, 1999).

To perform this, Beran (1999) proposed the SEMIFAR model, where the process or its first difference series follow a semiparametric regression model with fractional time series errors. A SEMIFAR model is a process Z_i satisfying

$$\phi(B)(1-B)^{\delta}\{(1-B)^m Z_i - g(x_i)\} = \epsilon_i, \tag{20}$$

where $m \in \{0, 1\}$, $\delta \in (-0.5, 0.5)$, ϵ_i are e.g. i.i.d. normal and $(1-B)^{\delta}$ is the fractional difference operator introduced by Granger and Joyeux (1980) and Hosking (1981) (see also Beran, 1994 and references therein). Set $Y_i = (1-B)^m$ and $\xi_i = (1-B)^{-\delta}\epsilon_i$, provided m is known, we obtain the nonparametric regression model (1). The process ξ_i has the property given in (2). Let σ_ϵ^2 denote the variance of ϵ_i and ϕ_1, ..., ϕ_p denote the unknown coefficients of $\phi(B)$. Then c_f is determined by σ_ϵ^2, ϕ_1, ..., ϕ_p. The other two unknown parameters δ and m can be written as one parameter $d := m + \delta$, since m is either one or zero. The unknown parameter vector $\theta = (\sigma_\epsilon^2, d, \phi_1, ..., \phi_p)^{\mathrm{T}}$ can be estimated from the residuals $\hat{\xi}_i(m) = Y_i - \hat{g}(x_i, m)$ by the approximate maximum likelihood in Beran (1995). The order of the AR part can then be selected using BIC (see Brean, 1999). Following the results in Beran et al. (1998) it can be shown that, $\hat{p}$ selected by the BIC is consistent.

The SEMIFAR model provides not only a tool for estimating θ but also a framework for simultaneously modelling of trend (g), short-range dependence (by means of $\phi(B)$), long-range dependence (δ) and nonstationarity (if $m = 1$). For estimating the SEMIFAR model, we need a data-driven algorithm combining the nonparametric estimation of g and maximum likelihood estimation of θ. The original algorithm proposed by Beran (1999) with some minor improvements is defined as follows:

Step 1: Define L = maximal order of $\phi(B)$ that will be tried, and a sufficiently fine grid

$G \in (-0.5, 1.5) \setminus \{0.5\}$. Then, for each $p \in \{0, 1, ..., L\}$, carry out steps 2 through 4.

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

Step 3: Carry out the following iteration:

Step 3a: Let $h_0 = \Delta_0 n^{(2\delta-1)/(5-2\delta)}$ and set $j = 1$.

Step 3b: Calculate $\hat{g}(t_i; m)$ using the bandwidth h_{j-1}. Set $\hat{\xi}_i = Y_i(m) - \hat{g}(t_i; m)$.

Step 3c: Set $\tilde{e}_i(d) = \sum_{j=0}^{i-1} \beta_j(\delta)\hat{x}i_{i-j}$, where the coefficients β_j are obtained from $\phi(B)(1 - B)^\delta$ by matching the powers in B.

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

Step 3e: Set $h_{2,j} = (h_{j-1})^\alpha$ with $\alpha = \alpha_0 = (5 - 2\delta)/(7 - 2\delta)$, improve h_{j-1} by

$$h_j = \left(\frac{1 - 2\delta}{I^2(K)} \frac{(1 - 2\Delta)\hat{V}}{\hat{I}(g''(t; h_{2,j}))} \right)^{1/(5-2\delta)} \cdot n^{(2\delta-1)/(5-2\delta)}. \tag{21}$$

Step 3f: Increase j by one and repeat steps 3b to 3e until convergence is reached or until a given number of iterations has been done. This yields for each $d \in G$ separately, the ultimate value of $\hat{\sigma}_\epsilon^2(d)$, as a function of d.

Step 4: Define $\hat{d}$ to be the value of d for which $\hat{\sigma}_\epsilon^2(d)$ is minimal. This together with the corresponding estimates of the AR parameters, yields an information criterion, e.g. $\text{BIC}(p) = n \log \hat{\sigma}_\epsilon^2(p) + p \log n$, as a function of p and the corresponding values of $\hat{\theta}$ and $\hat{g}$ for the given order p.

Step 5: Select the order p that minimizes $\text{BIC}(p)$. This yields the final estimates of θ^0 and g.

For more details see Beran (1999). It is proposed to use e.g. $\Delta_0 = 0.2$ as a default value. A simulation study and applications show that this algorithm works well in practice (see Beran, 1999, Beran and Feng, 2002a and Beran and Ocker, 1999, 2001).

Beran (1999) and Beran and Feng (2001a) showed that, under given conditions,

- $\hat{\theta}$ is asymptotically normally distributed.

- $\sqrt{n}$-consistent estimator is available.

The drawback of the above algorithm is that the required computing time is very long, in particular when the grid of d is fine. Hence, some fast variants of this algorithm were proposed by Beran and Feng (2002c). Simulation results given in Beran and Feng (2001b, 2002c) show that also these variants work well in practice. In the SEMIFAR packet developed by Beran in S-Plus[1], a variant of the above algorithm is proposed as a standard version of the SEMIFAR model, which is written in an S-Plus function called *SEMIFAR*. This S-Plus *SEMIFAR* function will be used in the next section.

5 Applications

In the following, the SEMIFAR model will be applied to some data examples to show its practical usefulness. These examples are chosen so that the different applicabilities of the SEMIFAR model can be shown. Earlier applications of the SEMIFAR model may be found in Beran (1999) and Beran and Ocker (1999, 2001). Applications of other approaches mentioned in this paper may be found in the cited works (Altman, 1990, Hart, 1991, Herrmann et al., 1992 and Rayand Tsay, 1997).

The first example is a traditional example of long-memory time series, i.e. the yearly minimum water levels in the Nile River at Roda Gauge near Cairo from 622 to 1281 (called the Nile Data). The second example is the transformed series $r_i = |Y_i - Y_{i-1}|^{1/4}$, where Y_i are the observations of the daily S&P500-Index series from Jan. 03, 1994 to Jun. 30, 1999 (called SAPd25). See Ding et al. (1993) and Beran and Ocker (2001) for more discussions on this transformation. As a third example, time series of the daily copper spot price from Jan. 02, 1997 to Sep. 02, 1998, is used (called Copper Price). The last example is the time series of the daily exchange rate between US Dollar and Euro (USD/Euro) from Jan. 03, 1999 to Oct. 19, 2001. The estimated parameters $\hat{m}$, $\hat{h}$ and $\hat{\delta}$ together with a 95% confidence interval for δ are given in Table 1. The answers to the questions, if the estimated long memory parameter is significant and if the estimated trend is significant, are also given in Table 1. Note in particular that the null hypothesis for $\hat{m} = 1$ is of the form $H_0 : g \equiv 0$. The four time series together with the estimated trends are shown in Figures 1(a) to (d). The estimated trends shown in Figures 1(c) and (d) are the cumulative sums of $\hat{g}$. The order of the autoregressive part was selected form

[1] The SEMIFAR packet developed by Beran in S-Plus is now published as a part of *S+FinanceMetrics*. See the web-site of Insightful.

Table 1: The estimation results for the four examples

Time Series	$\hat{m}$	$\hat{\delta}$	95%-CI for δ	$\hat{\delta}$-sig	$\hat{h}$	$\hat{g}$-sig
Nile Data	0	0.369	[0.309, 0.429]	Y	0.155	N
SAPd25	0	0.017	[-0.024, 0.058]	N	0.080	Y
Copper Price	1	-0.173	[-0.247, -0.100]	Y	0.077	Y
USD/Euro	1	0.002	[-0.058, 0.063]	N	0.124	Y

$0, 1, ..., 5$. However, we have $\hat{p} = 0$ for all the four examples implying that there is no clear short memory in these time series. $\hat{p}$ is hence not listed in Table 1.

Results in Table 1 show that:

1. The time series of the minimum water levels of the Nile River seems to be a stationary, purely long-memory time series;

2. The transformed time series r_i from the S&P Index can be modelled by a nonparametric regression model with a significantly increasing trend and i.i.d. errors. This significant trend shows that the difference series of the S&P 500 Index is no more covariance stationary. If the trend is not estimated and adjusted, we will obtain a wrong conclusion, that *there is strong long memory* for this data set due to the nonstationarity.

3. The differences of the daily copper spot price follow a nonparametric regression model with a significantly decreasing trend and antipersistent errors. The antipersistence means that there is an overdifferencing in the first differences of this series.

4. The differences of the daily exchange rates between US Dollar and Euro in the observed period seems to be a nonparametric regression with a significantly decreasing trend and i.i.d. errors, i.e. the original time series seems to be a random walk with a smooth, nonparametric drift.

6 Final remarks

In this paper recent developments in the area of non- and semiparametric regression with fractional time series errors were summarized. The focus was on computational aspects, in particular the selection of the bandwidth, semiparametric estimation of the parameters

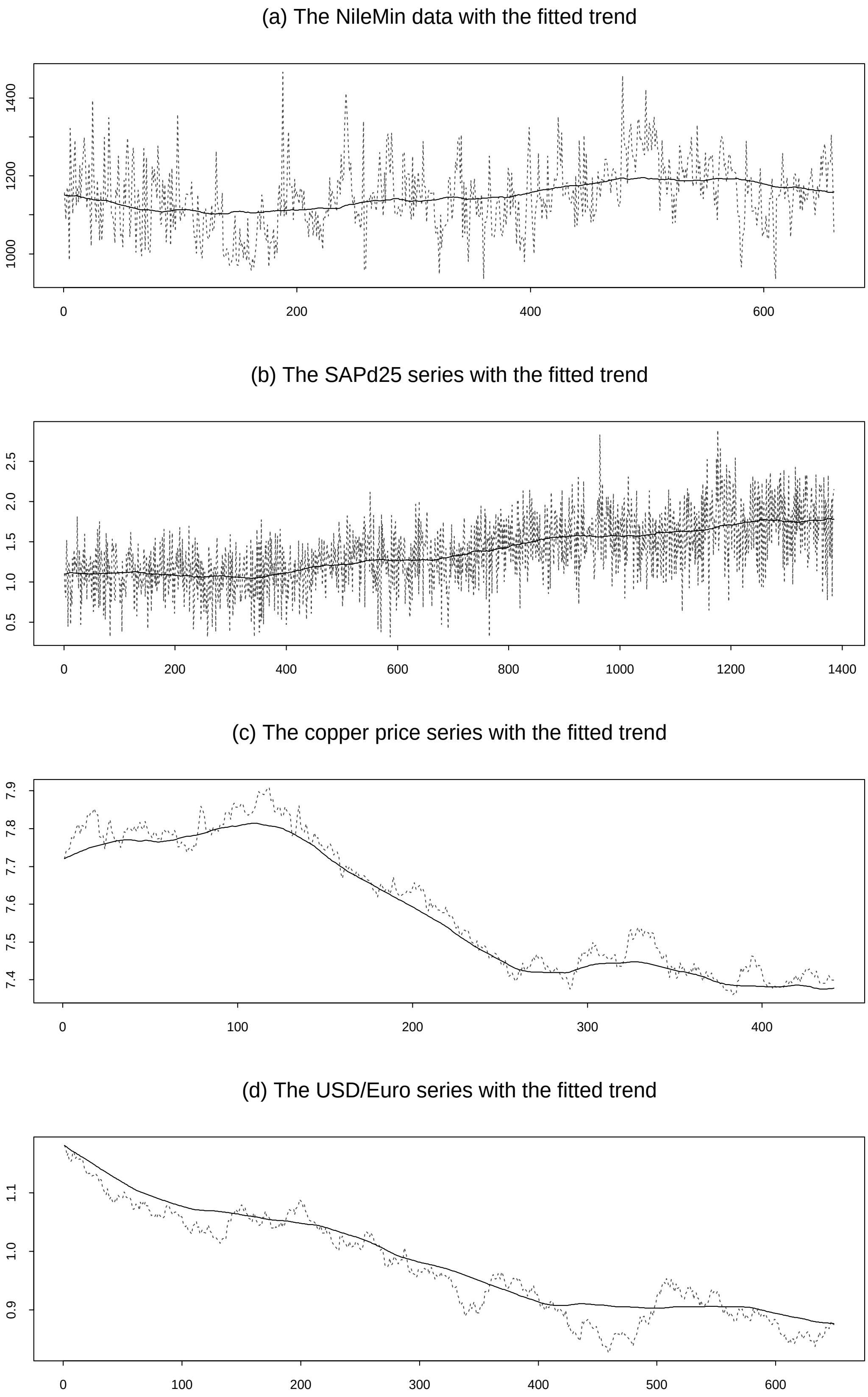

Figure 1: The four time series together with fitted trends.

and so on. Theoretical results were discussed only briefly. Readers who are interested in these topics are referred to Hall and Hart (1990), Beran (1999), Beran and Feng (2001, 2002a, b) and Feng (2002a). New applications of the IPL idea in a related context are proposed by Feng (2002b, c), where IPL bandwidth selectors for decomposing seasonal time series (Feng, 2002b) and for estimating the scale change in nonparametric regression with heteroskedastic time series errors (Feng, 2002c) are proposed. Furthermore, only results of kernel and local polynomial estimation in fixed design nonparametric regression are discussed here. Research on other related topics such as smoothing-splines, wavelet methods and estimation in random design nonparametric regression may be found e.g. in Csörgö and Mielniczuk (1995, 1999), Wang (1996) and numerous references in Opsomer et al. (2000).

7 Acknowledgements

This paper was partly supported by an NSF SBIR grant to MathSoft Inc and partly by the *Center of Finance and Econometrics* (CoFE) at the University of Konstanz. This paper is written based on a talk given at *the* 2$^{\text{nd}}$ *Euroworkshop on Statistical Modelling - Nonparametric Models* held between November 1 - 4, 2001, Bernried, near Munich, Germany. We are grateful to the organizers, especially Dr. Göran Kauermann, University of Glasgow, for their excellent organization. We would also like to thank Dr. Dirk Ocker, Swiss Union of Raiffeisenbannks, for providing us the time series of the copper stock price. The the exchange rates data of USD/Euro was downloaded from the *Statistical Release* of the US *Federal Reserve Bank of St. Louis* on the web.

References

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

Beran, J. (1994). *Statistics for Long-Memory Processes.* Chapman & Hall, New York.

Beran, J. (1995). Maximum likelihood of estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *J. Roy. Statist. Soc. Ser. B* **57** 659–672.

Beran, J. (1999). SEMIFAR models – A semiparametric framework for modelling trends, long range dependence and nonstationarity. Discussion paper No. 99/16, CoFE (Center of Finance and Econometrics), University of Konstanz.

Beran, J., Bhansali, R.J. and Ocker, D. (1998). On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. Biometrika, **85**, 921-934.

Beran, J. and Feng, Y. (2001a). Local polynomial estimation with a FARIMA-GARCH error process. *Bernoulli*, **7**, 733–750.

Beran, J. and Feng, Y. (2001b). Supplement to the Paper "Iterative plug-in algorithms for SEMIFAR models - definition, convergence and asymptotic properties" – Detailed simulation results. Discussion Paper No. 01/12, CoFE, University of Konstanz.

Beran, J. and Feng, Y. (2002a). SEMIFAR models - A semiparametric framework for modelling trends, long range dependence and nonstationarity. *Comput. Statist. & Data Anal.* (in press).

Beran, J. and Feng, Y. (2002b). Local polynomial fitting with long-memory, short-memory and antipersistent errors. *The Ann. Instit. Statist. Math.* (in press).

Beran, J. and Feng, Y. (2002c). Iterative plug-in algorithms for SEMIFAR models - definition, convergence and asymptotic properties. To appear in *J. Comput. and Graph. Statist.*.

Beran, J., Feng, Y. and Heiler, S. (2000). Modifying the double smoothing bandwidth selection in nonparametric regression. Discussion paper No. 00/37, CoFE, University of Konstanz. Submitted.

Beran, J. and Ocker, D. (1999). SEMIFAR forecasts, with applications to foreign exchange rates. *J. Statist. Plann. Infer.*, **80**, 137–153.

Beran, J. and D. Ocker (2001). Volatility of Stock Market Indices - An Analysis based on SEMIFAR models. *J. Busin. Econ. Statist.*, **19**, 103-116.

Bierens, H.J. (1983). Uniform consistent of kernel estimators of a regression function under generalized conditions. *J. Amer. Statist. Assoc.*, **78**, 69 –707.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, **31**, 307–327.

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

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

Clark, R.M. (1975). A calibration curve for radiocarbon dates. *Antiquity*, **49**, 251–266.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.*, **74**, 829–836.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, **31**, 377–403.

Csörgö, S. and Mielniczuk, J. (1995). Nonparametric regression under long-range dependent normal errors. Ann. Statist., **23**, 1000–1014.

Csörgö, S. and Mielniczuk, J. (1999). Random design regression under long-range dependent errors. *Bernoulli*, **5**, 209–224.

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

Engel, R.F. (1982). Autoregressive conditional heteroskedasticity with estimation of U.K. inflation, *Econometrica*, **50**, 987–1008.

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing.* Marcel Dekker, New York.

Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, 998–1004.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. *J. Roy. Statist. Soc. Ser. B*, **57**, 371–394.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications.* Chapman & Hall, London.

Feng, Y. (1999). *Kernel- and Locally Weighted Regression – with Applications to Time Series Decomposition.* Verlag für Wissenschaft und Forschung, Berlin.

Feng, Y. (2002a). Optimal Rates of Convergence for Nonparametric Regression with Fractional Time Series Errors. Discussion paper No. 02/01, CoFE, University of Konstanz. Submitted.

Feng, Y. (2002b). An iterative plug-in algorithm for nonparametric modelling of seasonal time series. Discussion paper No. 02/04, CoFE, University of Konstanz. Submitted.

Feng, Y. (2002c). Simultaneously Modelling Conditional Heteroskedasticity and Scale Change. Discussion paper No. 02/12, CoFE, University of Konstanz. Submitted.

Feng, Y. and Heiler, S. (2000). Bandwidth selection for local polynomial fits based on bootstrap idea. Preprint, University of Konstanz. Submitted.

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

Gasser, T. and Müller, H.G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds.) 23–68. Springer-Verlag, Heidelberg.

Gasser, T. and Müller, H.G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, **11**, 171–185.

Gasser, T., Müller, H.G., Köhler, W., Molinari, L. and Prader, A. (1984). Nonparametric regression analysis of growth curves. *Ann. Statist.*, **12**, 210–229.

Gasser, T., Müller, H.G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B*, **47**, 238–252.

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

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, New York.

Härdle, W. and Bowman, A.W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.*, **83**, 102–110.

Härdle, W., Hall, P. and Marron, J.S. (1988). How far are automatically chosen regression smoothing parameters from their optimum (with discussion)? *J. Amer. Statist. Assoc.*, **83**, 86–99.

Härdle, W., Hall, P. and Marron, J.S. (1992). Regression smoothing parameters that are not far from their optimum. *J. Amer. Statist. Assoc.*, **87**, 227–233.

Hall, P. and Hart, J.D. (1990). Nonparametric regression with long-range dependence. *Stochastic Process. Appl.*, **36**, 339–351.

Hall, P. and Marron, J.S. (1991). Lower bounds for bandwidth selection in density estimation. *Probability Theory and Related Fields*, **90**, 149–173.

Hart, J.D. (1991). Kernel regression estimation with time series errors. *J. R. Statist.Soc. Ser. B*, **53**, 173–188.

Hastie, T. and Loader, C. (1993). Local regression: Automatic kernel carpentry (with discussion). *Statistical Science*, **8**, 120–143.

Heiler, S. and Feng, Y. (1998). A simple root n bandwidth selector for nonparametric regression. *J. Nonpar. Statist.*, **9**, 1–21.

Herrmann, E. (1994). Asymptotic distribution of bandwidth selectors in kernel regression estimation. *Statistical Papers*, **35**, 17–26.

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

Herrmann, E. and Gasser, T. (1994). Iterative plug-in algorithm for bandwidth selection in kernel regression estimation. Preprint, Darmstadt Institute of Technology and University of Zürich.

Herrmann, E., Wand, M.P., Engle, J. and Gasser, T. (1995). A bandwidth selector for bivariate kernel regression. *J. Roy. Statist. Soc. Ser. B*, **57**, 171–180.

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

Ibragimov, I.A. and Linnik, Yu.V. (1971). *Independent and Stationary Sequences of Variables Random.* Wolters-Noordhoff Publishing Groningen, The Netherlands.

Ling, S. and Li, W.K. (1997). On fractional integrated autoregressive moving-average time series models with conditional heteroskedasticity. *J. Amer. Statist. Assoc.*, **92**, 1184–1194.

Müller, H.G. (1985). Empirical bandwidth choice for nonparametric kernel regression by means of pilot estimators. *Statist. Decisions*, Supp. Issue **2**, 193–206.

Müller, H.G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *J. Amer. Statist. Assoc.*, **82**, 231–238.

Müller, H.G. (1988). *Nonparametric Analysis of Longitudinal Data*, Springer-Verlag, Berlin.

Nadaraya, E.A. (1964). On estimating regression. *Theory of Probab. Appl.*, **9**, 141–142.

Opsomer, J., Wang, Y. and Yang, Y. (2000). Nonparametric regression with correlated errors. Preprint, Iowa State University.

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

Rice, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.*, **12**, 1215–1230.

Ruppert, D., Sheather, S.J. and Wand, M.P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.*, **90**, 1257–1270.

Ruppert, D. and Wand, M.P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.*, **22**, 1346–1370.

Stone, C.J. (1977). Consistent nonparametric regression (with discussion). *Ann. Statist.*, **5**, 595–620.

Taqqu, M.S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt processes. *Z. Wahrsch. verw. Geb.*, **31**, 287–302.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*, Chapman & Hall, London.

Wang, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *Ann. Statist.*, **24**, 466–484.

Watson, G.S. (1964). Smooth regression analysis. *Sankhyā* A, **26**, 359–372.