LOCAL POLYNOMIAL FITTING WITH LONG-MEMORY, SHORT-MEMORY AND ANTIPERSISTENT ERRORS

Jan Beran and Yuanhua Feng

University of Konstanz

Abstract

Nonparametric regression with long-range, short-range and antipersistent errors is considered. Local polynomial smoothing is investigated for the estimation of the trend function and its derivatives. It is well known that in the presence of long memory (with a fractional differencing parameter $0 < d < 1/2$), nonparametric regression estimators converge at a slower rate than in the case of uncorrelated or short-range dependent errors ($d = 0$). Here, we show that in the case of anti-persistence ($-1/2 < d < 0$), the convergence rate of a nonparametric regression estimator is faster than for uncorrelated or short-range dependent errors. Moreover, it is shown that unified asymptotic formulas for the optimal bandwidth and the MSE hold for the whole range $-1/2 < d < 1/2$. Also, results on estimation at the boundary are included.

Key words and phrases: Anti-persistence, estimation of derivatives, local polynomial fitting, long-range dependence, nonparametric regression.

1 Introduction

Nonparametric regression with short- or long-range dependent errors has gained increasing attention in the literature. Kernel estimators for observations with short-range dependent errors are considered e.g. by Altman (1990), Hall and Hart (1990), Herrmann, Gasser and Kneip (1992) and Beran (1999). Kernel estimation for observations with long-range dependence is considered e.g. by Hall and Hart (1990), Csörgö and Mielniczuk (1995a) and Beran (1999). Wang (1996) proposed nonparametric regression for data with long-range dependent errors via wavelet shrinkage. The effect of long-range dependent errors in random design nonparametric regression was investigated e.g. by Csörgö and Mielniczuk (1995b). Data-driven procedures for nonparametric regression with short- or long memory errors have also been proposed (see Chiu 1989, Herrmann, Gasser and Kneip 1992 and Ray and Tsay, 1997).

Hall and Hart (1990) show that kernel estimators of the mean function g converge to the true, unknown mean function at the same rate as in the case of uncorrelated errors, if the error process has short memory. On the other hand, they show that, if the error process has long memory, then a nonparametric regression estimator converges at a slower rate than in the case of uncorrelated or short-range dependent errors. Here, a stationary process Y_i with autocovariances $\gamma(k) = \mathrm{cov}\,(Y_i, Y_{i+k})$ is said to have long-range dependence (or long memory), 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

$$f(\lambda) \sim c_f |\lambda|^{-\alpha} (\text{as } \lambda \to 0) \tag{1.1}$$

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

In this paper, the use of local polynomial fitting in nonparametric regression will be investigated for data with short- or long-range dependence as well as anti-persistence. Local polynomial fitting (see Stone 1977 and Cleveland 1979) is known to be an automatic kernel method (see Müller 1987 and Hastie and Loader 1993), which has many exciting statistical properties (see e.g. Ruppert and Wand 1994 and Fan and Gijbels 1996). Local polynomial fitting is adapted to the autoregressive context for modeling nonlinear time series under some mixing conditions e.g. by Masry (1996), Masry and Fan (1997) and Feng and Heiler (1998).

The contributions of our paper are:

1. Asymptotic formulas for local polynomial estimators for the trend function and its derivatives are obtained under short memory, long memory and antipersistence. To our knowledge, *local polynomial fitting* and nonparametric estimation of *derivatives* has not been investigated in the literature for the cases of *long memory* and *antipersistence*. Note that antipersistence implies that the autocorrelations sum

up to zero so that another technique has to be used in the proof than in the case of long-range dependence. We show, in particular, that *unified formulas* can be obtained that are valid for all three cases (short memory, long memory and antipersistence). Note that these three cases correspond to three very different types of dependence structures.

2. In contrast to previous literature on nonparametric regression with long memory, results are obtained not only for interior points but also for *boundary points*. The results hold for kernel estimators (with boundary correction) as well as local polynomial fits.

The paper is organized as follows. The model and the local polynomial estimators for the mean function as well as its derivatives are introduced in section 2. The main results are given in section 3. Section 4 contains some final remarks. Proofs of theorems are listed in the appendix.

2 The model and the estimators

Consider the equidistant design nonparametric regression model

$$Y_i = g(t_i) + X_i, \tag{2.1}$$

where $t_i = (i - 0.5)/n$, $g : [0,1] \to \mathbb{R}$ is a smooth function, $\delta \in (-0.5, 0.5)$, B denotes the backshift operator such that $BY_i = Y_{i-1}$ and X_i is a stationary process having the form

$$(1 - B)^{\delta} X_i = U_i, \tag{2.2}$$

where U_i is a stationary process with short memory so that $\sum_{k=-\infty}^{\infty} \operatorname{cov}(U_i, U_{i+k}) = C$ with $0 < C < \infty$. The parameter δ is called the fractional differencing parameter. The fractional difference $(1 - B)^{\delta}$, introduced by Granger and Joyeux (1980) and Hosking (1981), is defined by

$$(1 - B)^{\delta} = \sum_{k=0}^{\infty} b_k(\delta) B^k \tag{2.3}$$

with

$$b_k(\delta) = (-1)^k \frac{,\,(\delta+1)}{,\,(k+1)\,,\,(\delta-k+1)}\,. \tag{2.4}$$

The spectral density of X_i in (2.1) is proportional to $|\lambda|^{-2\delta}$ at the origin. In particular, X_i has long memory if $\delta > 0$. In this case, the autocovariances $\gamma(k)$ of X_i are proportional to $k^{2\delta-1}$ and hence are non-summable. If $\delta = 0$, $X_i = U_i$ has short memory and the spectral density $f(\lambda)$ converges to a positive constant $c_f = (2\pi)^{-1}C$ at the origin. If $\delta < 0$, the spectral density f of X_i converges to zero at the origin. This is sometimes called "anti-persistence". In this case we have $\sum_{k=-\infty}^{\infty} \gamma(k) = 0$. The model (2.1) is a special case of SEMIFAR (semiparametric fractional autoregressive models) models introduced by Beran (1999) (see also Beran and Ocker 1999). In what follows we will consider local polynomial estimation of $g^{(\nu)}$, the νth derivative of g. Since (2.1) also includes the case of anti-persistence, the theorems in the next section extend previous results on nonparametric regression to this case, and give formulas for local polynomial estimators that are valid for the whole range $\delta \in (-0.5, 0.5)$.

In the following we will estimate the νth derivative $g^{(\nu)}$ by a local polynomial fit proposed by Stone (1977) and Cleveland (1979). For recent developments in this context we refer the readers to Ruppert and Wand (1994), Wand and Jones (1995), and Fan and Gijbels (1995, 1996) and references therein. Assume that g is at least $(p+1)$-times differentiable at a point t_0. Then $g(t)$ can be approximated locally by a polynomial of order p:

$$g(t) = g(t_0) + g'(t_0)(t-t_0) + ... + g^{(p)}(t_0)(t-t_0)^p/p! + R_p \tag{2.5}$$

for t in a neighbourhood of t_0, where R_p is a remainder term. Let K be a symmetric density (a kernel of order two without boundary correction) having compact support $[-1, 1]$. Giving n observations Y_1, ..., Y_n, we can obtain an estimator of $g^{(\nu)}$ ($\nu \le p$) by solving the locally weighted least squares problem

$$Q = \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} \beta_j (t_i - t_0)^j \right\}^2 K\left(\frac{t_i - t_0}{h}\right) \Rightarrow \min, \tag{2.6}$$

where h is the bandwidth and K is called the weight function. Let $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p)^{\mathrm{T}}$, then it is clear from (2.5) that $\nu!\hat{\beta}_\nu$ estimates $g^{(\nu)}(t_0)$, $\nu = 0, 1, ..., p$. Let

$$\mathbf{X} = \begin{bmatrix} 1 & t_1 - t_0 & \cdots & (t_1 - t_0)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_n - t_0 & \cdots & (t_n - t_0)^p \end{bmatrix}.$$

and let $\mathbf{e}_j$, $j = 1, ..., p+1$, denote the jth $(p+1)\times 1$ unit vector. Also, let $\mathbf{K}$ denote the diagonal matrix with

$$k_i = K\left(\frac{t_i - t_0}{h}\right)$$

as its ith diagonal entry. Finally, let $\mathbf{y} = (Y_1, ..., Y_n)^{\mathrm{T}}$. Then $\hat{g}^{(\nu)}(t_0)$ can be written as

$$\begin{aligned} \hat{g}^{(\nu)}(t_0) &= \nu!\mathbf{e}_{\nu+1}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{K}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{K}\mathbf{y} \\ &=: \{\mathbf{w}^\nu(t_0)\}^{\mathrm{T}}\mathbf{y}, \end{aligned} \tag{2.7}$$

where $\{\mathbf{w}^\nu(t_0)\}^{\mathrm{T}} = \nu!\mathbf{e}_{\nu+1}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{K}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{K}$ is called the weighting system. We see that $\hat{g}^{(\nu)}(t_0)$ is a linear smoother with the weighting system $\mathbf{w}^\nu(t_0) = (w_1^\nu, ..., w_n^\nu)^{\mathrm{T}}$, where $w_i^\nu \neq 0$ only if $|t_i - t_0| \leq h$. The weighting system does not depend on the dependence structure of the errors. For any interior point $t_0 \in [h, 1-h]$ the non-zero part of $\mathbf{w}^\nu(t_0)$ is the same, i.e. $\hat{g}^{(\nu)}$ works as a moving average in the interior. Furthermore, $\mathbf{w}^\nu(t_0)$ satisfies:

$$\sum_{i=1}^n w_i^\nu(t_i - t_0)^\nu = \nu! \text{ and } \sum_{i=1}^n w_i^\nu(t_i - t_0)^j = 0 \text{ for } j = 0, ..., p, j \neq \nu. \tag{2.8}$$

It is the property (2.8) that makes $\hat{g}^{(\nu)}$ exactly unbiased if g is a polynomial of order not larger than p. The property (2.8) also shows the main difference between a local polynomial estimator and a kernel estimator, since for a kernel estimator property (2.8) only holds approximately.

3 Main results

3.1 Assumptions

In this section we discuss the asymptotic properties of the estimators proposed in the last section. Pointwise asymptotic bias and variance will be given for interior points and for boundary points, i.e. $t \in [0, h) \cup (1-h, 1]$, separately.

It is well known that a local polynomial estimator is asymptotically equivalent to a certain kernel estimator, called an (asymptotically) equivalent kernel estimator of the local polynomial estimator. Hence, the asymptotic properties of a local polynomial estimator are the same as those of the equivalent kernel estimator. Although it is shown in Ruppert and Wand (1994) that one can analyze local polynomial fitting directly as a weighted least squares estimator rather than as an approximate kernel estimator, the asymptotic results given in this section will be proved by means of the asymptotically equivalent kernel function. For the asymptotic results given below we need the following assumptions:

A1. g is an at least k_0 times continuously differentiable function on $[0, 1]$ with $k_0 \geq \nu+2$ and $k_0 - \nu$ even.

A2. The weight function $K(u)$ is a symmetric density (i.e. a kernel of order two) with compact support $[-1, 1]$ having the polynomial form

$$K(u) = \sum_{l=0}^{r} \alpha_l u^{2l} \mathbb{I}_{[-1,1]}(u)$$

(see e.g. Gasser and Müller 1979).

A3. The bandwidth satisfies: $h \to 0$, $(nh)^{1-2\delta} h^{2\nu} \to \infty$ as $n \to \infty$.

A4. A local polynomial fit of order p with $\nu \leq p < k_0$ is used. Let $k = p + 1$, if $p - \nu$ is odd and $k = p + 2$, if $p - \nu$ is even.

Under the assumptions A1 and A4, we have $k \leq k_0$. It will be shown that $\hat{g}^{(\nu)}$ converges to $g^{(\nu)}$ at the same rate in the interior as well as at the boundary, if $p - \nu$ is odd. However, the convergence rate of $\hat{g}^{(\nu)}$ at the boundary is slower than in the interior, if $p-\nu$ is even. Hence, a local polynomial fit with $p-\nu$ odd is preferable. If $p-\nu$ is odd, using $p-1$ for $t\in[h,1-h]$ and p for $t\in[0,h)\cup(1-h,1]$ also results in the same overall convergence rate, because now the local polynomial fits in the interior with p and $p-1$ are equivalent.

3.2 Results for interior points

First, the pointwise asymptotic behaviour of $\hat{g}^{(\nu)}$ at an interior point t will be discussed. For a weight function K (i.e. a symmetric density with support $[-1,1]$), the asymptotically equivalent kernel function, $K_{(\nu,p)}$, can be defined as follows. Let

$$\mathbf{N}_p=\begin{bmatrix} 1 & \mu_1 & \cdots & \mu_p \\ \mu_1 & \mu_2 & \cdots & \mu_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_p & \mu_{p+1} & \cdots & \mu_{2p} \end{bmatrix}, \tag{3.1}$$

where $\mu_j=\int u^j K(u)du$ is the jth moment of K. For $i,j=1,...,p+1$, let $(\alpha_{ij})=\mathbf{N}_p^{-1}$ and define

$$K_{(\nu,p)}(u)=\nu!Q_{(\nu,p)}(u)K(u), \tag{3.2}$$

where

$$Q_{(\nu,p)}(u)=\sum_{j=1}^{p+1}\alpha_{\nu+1,j}u^{(j-1)}.$$

It is easily established that the function $K_{(\nu,p)}$ defined in (3.2) satisfies

$$\int u^j K_{(\nu,p)}(u)du=\begin{cases} 0, & j=0,...,\nu-1,\nu+1,...,k-1, \\ \nu!, & j=\nu, \\ \beta_{(\nu,p)}, & j=k, \end{cases} \tag{3.3}$$

where $\beta_{(\nu,p)}$ is a non-zero constant. Therefore, $K_{(\nu,p)}$ is a kernel of order k for estimation of the νth derivative, which we will call an "equivalent kernel". It is the same as defined by Gasser, Müller and Mammitzsch (1985) up to a $(-1)^\nu$ sign. It is clear that $K(u)$ and the equivalent kernel $K_{(\nu,p)}(u)$ are both polynomial Lipschitz-continuous kernels.

Let $\tilde{\mathbf{w}}^\nu(t) = (\tilde{w}_1^\nu, ..., \tilde{w}_n^\nu)^{\mathrm{T}}$ denote the weighting system of a kernel estimator with $K_{(\nu,p)}$. Then we obtain a kernel estimator of $g^{(\nu)}$:

$$\tilde{g}^{(\nu)}(t) = \sum_{i=1}^{n} \tilde{w}_i^\nu Y_i, \tag{3.4}$$

where

$$\tilde{w}_i^\nu = \frac{1}{nh^{\nu+1}} K_{(\nu,p)}\left(\frac{t_i - t}{h}\right). \tag{3.5}$$

See e.g. Müller (1987) and Ruppert and Wand (1994). In the case of equidistant design the definition (3.4) is asymptotically equivalent to that given by Gasser and Müller (1984). Following Müller (1987) we have:

Lemma 1. *Under the assumptions A1 to A4 the pth order local polynomial estimator* $\hat{g}^{(\nu)}$ *at an interior point* t *is asymptotically equivalent to the* k*th order kernel estimator* $\tilde{g}^{(\nu)}$ *defined in* (3.4) *in the sense that*

$$\lim_{n\to\infty} \sup_{1\le i\le n} \left|\frac{w_i^\nu}{\tilde{w}_i^\nu} - 1\right| = 0, \quad \text{defining } \frac{0}{0} = 1. \tag{3.6}$$

In the following we denote $x_i = (t_i - t)/h$, $y_j = (t_j - t)/h$. Let $n_0 = [nt + 0.5]$, $b = [nh]$, where $[\cdot]$ denotes the integer part. The notation

$$V_n(h) = (nh)^{-1-2\delta} \sum_{i,j=n_0-b}^{n_0+b} K_{(\nu,p)}(x_i) K_{(\nu,p)}(y_i)\gamma(i-j) \tag{3.7}$$

will be used for convenience. The following result is obtained under the assumptions A1 to A4.

Theorem 1. *Let the assumptions A1 to A4 hold, and let* t *be an interior point. Then for* $\delta \in (-0.5, 0.5)$*, we have*

i) Bias:

$$E[\hat{g}^{(\nu)} - g^{(\nu)}] = h^{(k-\nu)} \frac{g^{(k)}(t)\beta_{(\nu,p)}}{k!} + o(h^{(k-\nu)}); \tag{3.8}$$

ii):

$$\lim_{n\to\infty} V_n(h) = V, \tag{3.9}$$

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

iii) Variance:

$$(nh)^{1-2\delta} h^{2\nu} var(\hat{g}^{(\nu)}(t)) = V + o(1). \tag{3.10}$$

Note that the order of the bias in (3.8) is determined by k and not by p.

3.3 Results for boundary points

Similar results as given in Lemma 1 and Theorem 1 can be obtained for boundary points in the interval $[0, h) \cup (1-h, 1]$. The discussion will only be carried out for the left boundary region $[0, h)$, since it is symmetric for the right boundary. Note, however, that any fixed point $t \in (0, 1)$ will asymptotically not be a boundary point, since $h \to 0$ as $n \to \infty$. A standard definition of a left boundary point is $t = ch$ with $0 \le c < 1$. We define the truncated kernel $K_c(u)$ with $0 \le c < 1$ as

$$K_c(u) = \left(\int_{-c}^{1} K(x)dx \right)^{-1} K(u) \mathbb{1}_{[-c,1]}(u) \,. \tag{3.11}$$

A special case of truncated kernel is the one used for the estimation at the left end point with $c = 0$, where $K_0(u) = 2K(u)\mathbb{1}_{[0,1]}$. Let $\mu_{jc} = \int_{-c}^{1} u^j K_c(u) du$ be the jth moment of K_c. And let $\mathbf{N}_{pc}$ be the same as $\mathbf{N}_p$ but with each μ_j replaced by μ_{jc}. An important difference between $\mathbf{N}_{pc}$ and $\mathbf{N}_p$ is that $\mu_j = 0$ for j odd, but in general, any μ_{jc} is non-zero, if $c \neq 1$. A $(p{+}1)$th order boundary kernel $K_{(\nu,p,c)}$ can be defined similarly as $K_{(\nu,p)}$ given in (3.2). Let $(\alpha_{ijc}) = \mathbf{N}_{pc}^{-1}$. Then

$$K_{(\nu,p,c)}(u) = \left(\sum_{j=1}^{p+1} \alpha_{\nu+1,j,c} u^{(j-1)} \right) K_c(u) \tag{3.12}$$

is a $(p{+}1)$th order boundary kernel, which satisfies

$$\int_{-c}^{1} u^j K_{(\nu,p,c)}(u) du = \begin{cases} 0, & j = 0, ..., \nu-1, \nu+1, ..., p, \\ \nu!, & j = \nu, \\ \beta_{(\nu,p,c)}, & j = p+1, \end{cases} \tag{3.13}$$

where $\beta_{(\nu,p,c)}$ is a non-zero constant. An asymptotically equivalent kernel estimator at $t = ch$ is then defined by $\tilde{g}^{(\nu)}(t) = \sum_{i=1}^{n} \tilde{w}_{ci}^{\nu} Y_i$, where $\tilde{w}_{ci}^{\nu} = (nh^{\nu+1})^{-1} K_{(\nu,p,c)}(x_i)$. Under this definition, Lemma 1 also holds at a boundary point.

Again, let $n_0 = [nt + 0.5]$, $b = [nh]$, $b_c = [nch]$. Using the notation

$$V_n(c,h) = (nh)^{-1-2\delta} \sum_{i,j=n_0-b_c}^{n_0+b} K_{(\nu,p,c)}(x_i) K_{(\nu,p,c)}(y_i) \gamma(i-j)\,, \tag{3.14}$$

we obtain

Theorem 2. *Let the assumptions A1 to A4 hold, and let $t = ch$ with $0 \le c < 1$ be a left boundary point. Then for $\delta \in (-0.5, 0.5)$, we have*

i) Bias:

$$E[\hat{g}^{(\nu)} - g^{(\nu)}] = h^{(p+1-\nu)} \frac{g^{(p+1)}(t)\beta_{(\nu,p,c)}}{(p+1)!} + o(h^{(p+1-\nu)})\,; \tag{3.15}$$

ii):

$$\lim_{n\to\infty} V_n(h,c) = V(c), \tag{3.16}$$

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

iii) Variance:

$$(nh)^{1-2\delta} h^{2\nu} var(\hat{g}^{(\nu)}(t)) = V(c) + o(1)\,. \tag{3.17}$$

We see that, the order of the kernel $K_{(\nu,p,c)}$ (for $c < 1$), and hence the order of the bias at a boundary point, is determined by p and not by k. If $p - \nu$ is even, the bias term is of a lower order at a boundary point than at an interior point. In particular, a zero order local polynomial estimator for g is just a kernel estimator. In this case, it was shown by Gasser and Müller (1979), that the integrated squared bias over the interior is of the order $O(h^4)$, and the squared bias integrated over [0, 1] is of the order $O(h^3)$. That is, the bias in the boundary region dominates the bias in the interior and it will cause a slower convergence rate of the MISE (mean integrated squared error). This is the so-called boundary effect of nonparametric regression estimators. It can be shown that, for a general local polynomial estimator $\hat{g}^{(\nu)}$ the following holds: 1. For $p-\nu$ even, the integrated squared bias over the interior is of the order $O(h^{2(k-\nu)})$, but that over [0, 1] is of the order $O(h^{2(k-\nu)-1})$; 2. For $p-\nu$ odd, the integrated squared bias over the interior and over [0, 1] is of the same order $O(h^{2(k-\nu)})$. In the second case we have $\text{MISE}([0,1]) = \text{MISE}([h, 1-h])(1+O(h))$. Hence, in local polynomial fitting $p-\nu$ is often taken to be odd in order that the MISE can be calculated over the whole support $[0,1]$ of g. Similar to the terminology in Gasser and Müller (1984), we call a $(\nu+1)$-th local polynomial estimator $\hat{g}^{(\nu)}$ with $k=\nu+2$ a standard local polynomial estimator. So, standard local polynomial estimators are those with lowest polynomial order such that $p-\nu$ is odd. The first three standard local polynomial estimators are the local linear estimate for $\hat{g}$, the local quadratic estimate for $\hat{g}'$ and the local cubic estimate for $\hat{g}''$.

3.4 The MISE

Let $p-\nu$ be odd, and let

$$I(g^{(p+1)}) = \int_0^1 [g^{(p+1)}(t)]^2 dt\,. \tag{3.18}$$

The following result holds:

Theorem 3. *Under the assumptions A1 to A4 and for $\delta \in (-0.5, 0.5)$, we have*

i) The mean integrated squared error (MISE) of $\hat{g}^{(\nu)}$ is given by

$$\begin{aligned}
&\int_0^1 E\{[\hat{g}^{(\nu)}(t) - g^{(\nu)}(t)]^2\}dt \\
&= \text{MISE}_{\text{asympt}}(n,h) + o(\max(h^{2(p+1-\nu)}, [(nh)^{2\delta-1}h^{-2\nu}])) \\
&= h^{2(p+1-\nu)}\frac{I(g^{(p+1)})\beta^2_{(\nu,p)}}{(p+1)!} + (nh)^{2\delta-1}h^{-2\nu}V \\
&+o(\max(h^{2(p+1-\nu)}, [(nh)^{2\delta-1}h^{-2\nu}]))\,;
\end{aligned} \tag{3.19}$$

ii) The optimal bandwidth that minimizes the asymptotic MISE is given by

$$h_{\text{opt}} = C_{\text{opt}} n^{(2\delta-1)/(2p+3-2\delta)}\,, \tag{3.20}$$

where

$$C_{\rm opt} = \left[\frac{2\nu+1-2\delta}{2(p+1-\nu)}\,\frac{[(p+1)!]^2 V}{I(g^{(p+1)})\beta^2_{(\nu,p)}}\right]^{1/(2p+3-2\delta)}, \tag{3.21}$$

where it is assumed that $I(g^{(p+1)}) > 0$.

Note that by inserting $h_{\rm opt}$ in (3.19), Theorem 3 implies that for $p-\nu$ odd the optimal MISE is of the order

$$\int_0^1 E\{[\hat{g}^{(\nu)}(t) - g^{(\nu)}(t)]^2\}dt = O(n^{2(2\delta-1)(k-\nu)/(2k+1-2\delta)}). \tag{3.22}$$

The following remarks clarify the results given above.

Remark 1. For bandwidth selection with the plug-in method one has to calculate the value of V. Simple explicit formulas for V can be given as follows:

$$V = 2\pi c_f \int_{-1}^{1} K^2_{(\nu,p)}(x)dx \tag{3.23}$$

for $\delta = 0$ and

$$V = 2c_f,\, (1-2\delta)\sin\pi\delta \int_{-1}^{1}\int_{-1}^{1} K_{(\nu,p)}(x)K_{(\nu,p)}(y)|x-y|^{(2\delta-1)}dxdy \tag{3.24}$$

for $\delta > 0$ (Hall and Hart 1990), where $c_f = (2\pi)^{-1}C$. The explicit form of V for $\delta < 0$ is more complex, since the integral $\int_{-1}^{1} K_{(\nu,p)}(y)|x-y|^{2\delta-1}dy$ does not exist. However, at any point x the kernel $K_{(\nu,p)}(y)$ may be written as $K_{(\nu,p)}(y) = \sum_{l=0}^{r}\beta_l(x)(x-y)^l =: K_0(x) + K_1(x-y)$, where r is an integer, $K_0(x) = \beta_0(x)$ and $K_1(x-y) = \sum_{l=1}^{r}\beta_l(x)(x-y)^l$. Note that, in the case of anti-persistence it holds $\sum_{k=-\infty}^{\infty}\gamma(k) = 0$. We have, for $\delta < 0$,

$$\begin{aligned} V \;=\; & 2c_f,\, (1-2\delta)\sin(\pi\delta)\int_{-1}^{1} K_{(\nu,p)}(x)\times \\ & \left\{\int_{-1}^{1} K_1(x-y)|x-y|^{2\delta-1}dy - K_0(x)\int_{|y|>1}|x-y|^{2\delta-1}dy\right\}dx. \end{aligned} \tag{3.25}$$

If g is estimated by a first order local polynomial with the uniform kernel as the weight function, then we have, in the interior, $K_{(0,1)}(x) = K(x) = \mathbb{I}_{\{|x|\le 1\}}/2$. In this case we have $K_0(x) = \mathbb{1}_{\{|x|\le 1\}}/2$ and $K_1 \equiv 0$. The formulas (3.23), (3.24) and (3.25) give the same result

$$V = \frac{2^{2\delta} c_f, \, (1-2\delta)\sin(\pi\delta)}{\delta(2\delta+1)} \tag{3.26}$$

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

Remark 2. Theorems 1 to 3 extend previous results (see Altman 1990, Hall and Hart 1990, Herrmann, Gasser and Kneip 1992, Csörgö and Mielniczuk 1995a and Beran 1999) in several ways by including $\delta < 0$, estimation of derivatives, estimation with higher order kernels and the pointwise asymptotic behaviour in the boundary region.

Remark 3. Theorem 1 also holds for kernel estimators $\tilde{g}^{(\nu)}$. Theorems 2 and 3 hold for kernel estimators with boundary correction. For kernel estimators without boundary correction Theorem 3 holds with the whole interval $[0,1]$ replaced by $[\Delta, 1-\Delta]$.

Remark 4. The estimation of g'' is essential for selecting the optimal bandwidth with the plug-in method. If g'' is estimated by a standard local polynomial estimator, i.e. with $p = 3$, the optimal bandwidth for $\hat{g}''$ is of the order $O(n^{(2\delta-1)/(9-2\delta)})$.

4 Final remarks

It is clear that the bias of a nonparametric regression estimator for observations with dependent errors is the same as that for uncorrelated errors if the estimator is a linear smoother and the error process is stationary. However, the variance of a linear smoother depends on how the errors are correlated. If the errors have short memory with $\delta = 0$, then only the constant V is influenced. In particular, V is larger than $\gamma(0)\int K^2_{(\nu,p)}(u)du$, when the dependence structure is dominated by positive correlations. In this case, the optimal bandwidth is larger than in the case of independent errors (see Herrmann, Gasser and Kneip 1992). If the error process has long memory, i.e. $\delta > 0$, then not only the constant V but also the order of the variance is changed. The variance converges at a slower rate to zero than in the case of short memory. On the other hand, if $\delta < 0$, the data are mostly negatively correlated and $\sum_{k=-\infty}^{\infty}\gamma(k) = 0$. In this case, the variance of $\hat{g}^{(\nu)}$ converges to zero at a higher rate than under independence or short-range dependence.

In order to use the estimator $\hat{g}^{(\nu)}$ effectively, one also needs a suitable data-driven bandwidth selection procedure for nonparametric regression with short- or long memory. This will be discussed elsewhere. Bandwidth selection procedures for data with short memory may be found in e.g. Chiu (1989) and Herrmann, Gasser and Kneip (1992). The iterative plug-in method proposed by Herrmann, Gasser and Kneip (1992) is adapted to bandwidth selection for data with long-range dependent errors in Ray and Tsay (1997). This procedure was used in the data-driven algorithm for estimating so-called SEMIFAR models in Beran (1999). Similar data-driven procedures can be developed for $\hat{g}^{(\nu)}$ proposed in this paper.

Acknowledgements

This paper was supported in part by the Center of Finance and Econometrics at the University of Konstanz and by an NSF (SBIR, phase 2) grant to MathSoft, Inc.

Appendix: Proofs of the theorems

Proof of Lemma 1. See Müller (1987). The weighting systems do not depend on the dependence structure and noting that the nonparametric regression model considered here follows an equidistant design. Hence, theorem 1 in Müller (1987) applies to $\hat{g}^{(\nu)}$ for $t \in [h, 1-h]$.

Proof of Theorem 1.

(i): The proof of the pointwise bias is standard.

(iii): (iii) holds if (ii) holds, since

$$
\begin{aligned}
&(nh)^{1-2\delta} h^{2\nu} \operatorname{var}(\hat{g}^{(\nu)}(t)) \\
&\quad = (nh)^{1-2\delta} h^{2\nu} \sum_{i,j=n_0-b}^{n_0+b} w_i^\nu w_j^\nu \gamma(i-j) \\
&\quad = (nh)^{1-2\delta} h^{2\nu} \sum_{i,j=n_0-b}^{n_0+b} \tilde{w}_i^\nu \tilde{w}_j^\nu \gamma(i-j)[1+o(1)] \\
&\quad = (nh)^{-1-2\delta} \sum_{i,j=n_0-b}^{n_0+b} K_{(\nu,p)}(x_i) K_{(\nu,p)}(y_j) \gamma(i-j)[1+o(1)] \\
&\quad = V_n(h)[1+o(1)]
\end{aligned}
$$

(ii): The results will be proved separately for $\delta = 0$, $0 < \delta < 0.5$ and $-0.5 < \delta < 0$. The formulas (3.23), 3.24) and (3.25) will be obtained immediately.

a) $\delta = 0$. Observing that, for $\delta = 0$, $\sum_{k=-\infty}^{\infty} \gamma(k) = C$ and

$$V_n(h) = (nh)^{-1} \sum_{i=n_0-b}^{n_0+b} K_{(\nu,p)}(x_i) \sum_{j=n_0-b}^{n_0+b} K_{(\nu,p)}(y_j) \gamma(i-j).$$

Let b' be an integer such that $b' = o(b)$, $b' \to \infty$ as $n \to \infty$ (e.g. $b' = [\sqrt{b}]$). For $n_0 - b + b' < i < n_0 + b - b'$ we have

$$\sum_{j=n_0-b}^{n_0+b} K_{(\nu,p)}(y_j)\gamma(i-j) = \sum_{|i-j|<b'} K_{(\nu,p)}(y_j)\gamma(i-j) + \sum_{|i-j|\geq b'} K_{(\nu,p)}(y_j)\gamma(i-j).$$

For $|i-j| < b'$ we have $|x_i - y_j| = o(1)$ and hence $K(y_j) = K(x_i)[1+o(1)]$, because $b' = o(b)$ and $K_{(\nu,p)}$ is Lipschitz-continuous. Then we obtain

$$\sum_{|i-j|<b'} K_{(\nu,p)}(y_j)\gamma(i-j) = C K_{(\nu,p)}(x_i)[1+o(1)],$$

because $\sum_{|i-j|<b'} \gamma(i-j) = C + o(1)$. Furthermore,

$$\sum_{|i-j|\geq b'} K_{(\nu,p)}(y_j)\gamma(i-j) = o(1),$$

because $K_{(\nu,p)}$ is bounded and $\sum_{|i-j|\geq b'} \gamma(i-j) = o(1)$.

For $i = n_0 - b$ or $i = n_0 + b$ we have, analogously,

$$\begin{aligned} \sum K_{(\nu,p)}(y_j)\gamma(i-j) &= [(C+\gamma(0))/2]K_{(\nu,p)}(x_i)(1+o(1)) \\ &= O(K_{(\nu,p)}(x_i)). \end{aligned}$$

It is clear that $\sum K_{(\nu,p)}(y_j)\gamma(i-j) = O(K_{(\nu,p)}(x_i))$ holds for all i such that $n_0 - b < i \leq n_0 - b + b'$ or $n_0 + b - b' < i \leq n_0 + b$. Since $b' = o(b)$, we have

$$\begin{aligned} V_n(h) &= (nh)^{-1} \sum_{i=n_0-b}^{n_0+b} K^2_{(\nu,p)}(x_i)[C + o(1)] \\ &= C \int_{-1}^{1} K^2_{(\nu,p)}(x)dx + o(1). \end{aligned} \tag{A.1}$$

b) $\delta > 0$. In this case we have $\gamma(k) \sim c_\gamma |k|^{2\delta-1}$ with $c_\gamma = 2c_f$, $(1-2\delta)\sin\pi\delta > 0$ (see Beran 1994, pp. 61-63).

$$\begin{aligned} V_n(h) &= (nh)^{-1-2\delta} \sum_{i,j=n_0-b}^{n_0+b} K_{(\nu,p)}(x_i)K_{(\nu,p)}(y_j)\gamma(i-j) \\ &\doteq c_\gamma (nh)^{-1-2\delta} \sum_{\substack{i,j=n_0-b \\ i\neq j}}^{n_0+b} K_{(\nu,p)}(x_i)K_{(\nu,p)}(y_j)|i-j|^{2\delta-1} \\ &= c_\gamma (nh)^{-2} \sum_{\substack{i,j=n_0-b \\ i\neq j}}^{n_0+b} K_{(\nu,p)}(x_i)K_{(\nu,p)}(y_j)|x_i-y_j|^{2\delta-1} \\ &\doteq c_\gamma \int_{-1}^{1}\int_{-1}^{1} K_{(\nu,p)}(x)K_{(\nu,p)}(y)|x-y|^{2\delta-1}dxdy \end{aligned} \tag{A.2}$$

c). The proof for $\delta < 0$ is based on the decomposition of the equivalent kernel and the property $\sum_{k=-\infty}^{\infty} \gamma(k) = 0$, where $\gamma(k) \sim c_\gamma |k|^{2\delta-1}$ for large k with $c_\gamma = 2c_f$, $(1-2\delta)\sin(\pi\delta) < 0$ in the case of anti-persistence (see Beran, 1994). Put $x_i = (t_i - t)/b$, $y_j = (t_j - t)/b$. Define $n_0 = [nt]$ and $n_1 = [nb]$ as before. We have, for given i, $\sum_{j=n_0-n_1}^{n_0+n_1} \gamma(i-j) = -\sum_{|j-n_0|>n_1} \gamma(i-j)$. Recall that the equivalent kernel has the form $K_{(\nu,p)}(x) = \sum_{l=0}^{r} \alpha_l x^l \mathbb{1}_{\{|x|\leq 1\}}$. At a point x_i, $K_{(\nu,p)}(y)$ can be rewritten as $K_{(\nu,p)}(y) = \sum_{l=0}^{r} \beta_l(x_i)(x_i - y)^l =: K_0(x_i) + K_1(x_i - y)$, where $K_0(x_i) = \beta_0(x_i)$, $K_1(x_i - y) = \sum_{l=1}^{r} \beta_l(x_i)(x_i - y)^l$.

Observing that t is an interior point we have

$$V_n = (nb)^{-1-2\delta} \sum_{i=n_0-n_1}^{n_0+n_1} K_{(\nu,p)}(x_i) \sum_{j=n_0-n_1}^{n_0+n_1} K_{(\nu,p)}(y_j)\gamma(i-j).$$

$$\begin{aligned}
&\sum_{j=n_0-n_1}^{n_0+n_1} K_{(\nu,p)}(y_j)\gamma(i-j) \\
&= \sum_{j=n_0-n_1}^{n_0+n_1} K_1(x_i - y_j)\gamma(i-j) - K_0(x_i) \sum_{|j-n_0|>n_1} \gamma(i-j) \\
&\doteq c_\gamma \left\{ \sum_{\substack{j=n_0-n_1 \\ j \neq i}}^{n_0+n_1} K_1(x_i - y_j)|i-j|^{2\delta-1} - K_0(x_i) \sum_{|j-n_0|>n_1} |i-j|^{2\delta-1} \right\} \\
&= c_\gamma (nb)^{2\delta-1} \left\{ \sum_{\substack{j=n_0-n_1 \\ j \neq i}}^{n_0+n_1} K_1(x_i - y_j)|x_i - y_j|^{2\delta-1} - K_0(x_i) \sum_{|j-n_0|>n_1} |x_i - y_j|^{2\delta-1} \right\} \\
&\doteq c_\gamma (nb)^{2\delta} \left\{ \int_{-1}^{1} K_1(x_i - y)|x_i - y|^{2\delta-1} dy - K_0(x_i) \int_{|y|>1} |x_i - y|^{2\delta-1} dy \right\}
\end{aligned}$$

Finally, we obtain

$$\begin{aligned}
V_n &\doteq c_\gamma (nb)^{-1} \sum_{i=n_0-n_1}^{n_0+n_1} K_{(\nu,p)}(x_i) \left\{ \int_{-1}^{1} K_1(x_i - y)|x_i - y|^{2\delta-1} dy \right. \\
&\quad \left. - K_0(x_i) \int_{|y|>1} |x_i - y|^{2\delta-1} dy \right\} \\
&\doteq c_\gamma \int_{-1}^{1} K_{(\nu,p)}(x) \left\{ \int_{-1}^{1} K_1(x-y)|x-y|^{2\delta-1} dy \right. \\
&\quad \left. - K_0(x) \int_{|y|>1} |x-y|^{2\delta-1} dy \right\} dx.
\end{aligned}$$

This concludes the proof. □

Proof of Theorem 2. The proof of Theorem 2 is similar to that of Theorem 1 and is hence omitted. Especially, note that the proof in "(ii): c)" can also be carried out for $t = 0$ or $t = 1$. □

Proof of Theorem 3. Theorem 3 follows from Theorem 1 and Theorem 2, since in the case that $p - \nu$ is odd the MISE on the boundary area is asymptotically negligible. □

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). Estimating trends, long-range dependence and nonstationarity. Preprint, University of Konstanz.

Beran, J. and Ocker, D. (1999). SEMIFAR forecasts, with applications to foreign exchange rates. J. Statistical Planning and Inference. To appear.

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

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

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.) 55-74. The Iowa State University Press.

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

Csörgö, S. and Mielniczuk, J. (1995b). Random-design regression under long-range dependent errors. Preprint, University of Michigan.

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. and Heiler, S. (1998). Locally weighted autoregression. In *Econometrics in Theory and Practice* (R. Galata and H. Küchenhoff, eds.) 101–117. Physica-Verlag, Heidelberg.

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. 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.

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

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

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

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.

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

Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *J. Time Series Analysis* **17** 571–599.

Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scand. J. Statst.* **24** 165–179.

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

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

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.

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.

Address of both authors:

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

Email address of Prof. Jan Beran:

jberan@iris.rz.uni-konstanz.de
or
beran@fmi.uni-konstanz.de

Email address of Yuanhua Feng:

Yuanhua.feng@uni-konstanz.de