Optimal Convergence Rates in Nonparametric Regression with Fractional Time Series Errors

Yuanhua Feng[1] and Jan Beran[2]

[1]Department of Actuarial Mathematics and Statistics

The Maxwell Institute for Mathematical Sciences

Heriot-Watt University

[2]Department of Mathematics and Statistics

University of Konstanz

January 16, 2007

Optimal Convergence Rates in Nonparametric Regression with Fractional Time Series Errors

Yuanhua Feng and Jan Beran

Consider the estimation of $g^{(\nu)}$, the ν-th derivative of the mean function, in a fixed-design nonparametric regression model with stationary time series errors ξ_i. We assume that $g \in \mathcal{C}^k$, ξ_i are obtained by applying an invertible linear filter to iid innovations, and the spectral density of ξ_i has the form $f(\lambda) \sim c_f |\lambda|^{-\alpha}$ as $\lambda \to 0$ with constants $c_f > 0$ and $\alpha \in (-1, 1)$. Under regularity conditions, the optimal convergence rate of $\hat{g}^{(\nu)}$ is shown to be n^{-r_ν} with $r_\nu = (1-\alpha)(k-\nu)/(2k+1-\alpha)$. This rate is achieved by local polynomial fitting. Moreover, in spite of including long memory and antipersistence, the required conditions on the innovation distribution turn out to be the same as in nonparametric regression with iid errors.

Keywords: Optimal rate of convergence, nonparametric regression, long memory, antipersistence.

1 Introduction

Consider the estimation of $g^{(\nu)}$, the ν-th derivative of the mean function g in the equidistant-design nonparametric regression model

$$(1.1) \qquad Y_i = g(x_i) + \xi_i,$$

with $x_i = i/n$, $g : [0,1] \to \Re$ a smooth function and ξ_i a linear (second order and strictly) stationary process generated by applying a linear filter to an iid series ε_i. For the autocovariance function $\gamma(k) = \text{cov}(\xi_i, \xi_{i+k})$, it is assumed that $\gamma(k) \to 0$ as $|k| \to \infty$. Equation (1.1) represents a nonparametric regression model with short memory (including iid 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 $\sum \gamma(k) = \infty$. A more specific assumption is that the spectral density $f(\lambda) = (2\pi)^{-1} \sum \gamma(k) \exp(ik\lambda)$ has a pole at the origin of the form

$$(1.2) \qquad f(\lambda) \sim c_f |\lambda|^{-\alpha} \text{ (as } \lambda \to 0)$$

for some $\alpha \in (0, 1)$, 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, (1.2) implies that $\gamma(k) \sim c_\gamma |k|^{\alpha-1}$ $(c_\gamma > 0)$ so that $\sum \gamma(k) = \infty$, i.e. ξ_i has long memory. If (1.2) holds with $\alpha = 0$, then $0 < \sum \gamma(k) < \infty$ and ξ_i is said to have short memory. On the other hand, a stationary process is said to be antipersistent, if (1.2) holds for $\alpha \in (-1, 0)$ implying that $\sum \gamma(k) = 0$.

The aim of this paper is to investigate the minimax optimal convergence rate of a non-parametric estimator of $g^{(\nu)}$ (see e.g. Farrell, 1972, Stone, 1980, 1982 and Hall and Hart, 1990a for related work). For a summary of nonparametric minimax theory we refer the reader to Hall (1989). Hall and Hart (1990a) derived the optimal convergence rate for estimates of g in nonparametric regression with Gaussian stationary short- and long-memory errors. More recently, Li and Xia (2007) investigated the optimal rates of convergence of block thresholded wavelet estimators in nonparametric regression with long-memory errors under weaker smoothness condition. In this paper a unified formula for the optimal convergence rate for estimating $g^{(\nu)}$ in nonparametric regression with short-memory, long-memory and antipersistent errors is given. It is shown that this rate is achieved by local polynomial estimates (Beran and Feng, 2002a). Our finding generalize previous results in Stone (1980) and Hall and Hart (1990a) in several ways. A simple condition under which a sequence n^{-r_ν} forms a lower bound to the convergence rate is given for nonparametric regression with stationary time series errors at any dependence level. Results are obtained for Gaussian and non-Gaussian error processes. The optimal rate of convergence in models with long-memory errors turns out to be lower (i.e. slower) than in the case of short memory, whereas the rate is higher in the presence of antipersistent errors.

The study of antipersistent phenomena has gained increasing attention in recent years. It has been realised that, when estimation of the memory parameter (i.e. $d = \alpha/2$) is discussed, the symmetric range $d \in (-0.5, .5)$ instead of $[0, 0.5)$ only, should be considered (see e.g. Robinson 2005). In a recent publication, Tsai (2006) investigates continuous time fractionally integrated ARMA models with long memory, short memory and antipersistence. In practice, nonparametric regression with antipersistent errors plays an important role in the context of over-differencing of integrated time series. Over-differencing often occurs in the analysis of financial or economic time series (see e.g. Beran and Ocker 2001, Beran et al. 2003, Silverberg and Verspagen 2003). A number of recent publications demonstrate that instead of short or long memory, significant antipersistence can be found in financial returns or residuals of certain financial time series models. This

is in agreement with empirical findings of mean reversion in financial returns. Concrete examples are, for instance, stock price series analyzed in Feng et al. (2006), the German DAX index (Mungo and Härdle, 2006) and high-frequency foreign exchange rate series (Karupiah and Los, 2005).

The estimator and the error process are defined in section 2. Section 3 describes the conditions on the distribution and provides the main results. It turns out that the required regularity conditions on the marginal innovation distribution are the same for all $\alpha \in (-1, 1)$ and hence do not depend on the type of dependence structure. Some auxiliary results that are needed for the proofs are given in section 4. Detailed proofs are given in the appendix.

2   The estimator and the error process

2.1   The local polynomial fitting

Kernel estimation of g in nonparametric regression with short-memory and long-memory errors was considered in Hall and Hart (1990a). Beran and Feng (2002b) extended the results to nonparametric regression with antipersistence. Since kernel estimators are affected by boundary problems, unless corrected by boundary kernels, attractive alternative estimates are obtained by local polynomial fitting introduced by Stone (1977), and Cleveland (1979). Beran and Feng (2002a) studied local polynomial fitting in nonparametric regression with short-memory, long-memory and antipersistent errors. In this paper we will use the approach in Beran and Feng (2002a) to show the achievability of the optimal convergence rate.

Let $k \geq 2$ be a positive integer. The function class considered in this paper is $\mathcal{C}^k(B)$, the collection of all k times differentiable functions g on $[0, 1]$ which satisfy

$$\sup_{0 \leq x \leq 1} \max_{\nu=0,1,\ldots,k} |g^{(\nu)}(x)| \leq B.$$

Let $p = k - 1$. Then g can be locally approximated by a polynomial of order p for x in the neighbourhood of a point x_0:

(2.1) $g(x) = g(x_0) + g'(x_0)(x - x_0) + \ldots + g^{(p)}(x_0)(x - x_0)^p/p! + R_p,$

4

where R_p is a remainder term. Let K be a second order kernel (a symmetric density) having compact support $[-1, 1]$. Given n observations Y_1, ..., Y_n, we can obtain an estimator of $g^{(\nu)}$ ($\nu \leq p$) by solving the locally weighted least squares problem

$$(2.2) \qquad Q = \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} \beta_j (x_i - x_0)^j \right\}^2 K\left(\frac{x_i - x_0}{h}\right) \Rightarrow \min,$$

where h is the bandwidth. Denote by $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p)'$ the solution of (2.2). Then $\hat{g}^{(\nu)}(x_0) := \nu! \hat{\beta}_\nu$ is the local polynomial estimate of $g^{(\nu)}(x_0)$ ($\nu = 0, 1, ..., p$).

2.2 The error process

In this paper, it is assumed that the spectral density of ξ_i has the form (1.2). Hence ξ_i will be called a fractional time series error process. Moreover, ξ_i is assumed to be causal, linear and invertible. That is, ξ_i can be expressed in two ways:

$$(2.3) \qquad \xi_i = \psi(B)\varepsilon_i,$$

and

$$(2.4) \qquad \varepsilon_i = \varphi(B)\xi_i,$$

where the innovations ε_i are iid zero mean random variables with $\text{var}(\varepsilon_i) = \sigma_\varepsilon^2 < \infty$. Here, B is the backshift operator, and $\psi(B) = \sum_{j=0}^{\infty} a_j B^j$ and $\varphi(B) = \sum_{j=0}^{\infty} b_j B^j$ are the characteristic polynomials of the MA and AR representation of ξ_i, respectively, with $a_0 = b_0 = 1$, $\sum a_j^2 < \infty$ and $\sum b_j^2 < \infty$. The causality of ξ_i is assumed for convenience.

Some properties of ξ_i can be understood more easily by means of its inverse process. Following Chatfield (1979), the inverse process of ξ_i, denoted by ξ_i^{I}, is the process with the same innovations ε_i but $\varphi(B)$ and $\psi(B)$ as its characteristic polynomials or the MA and AR representation respectively. Thus, we have

$$(2.5) \qquad \xi_i^{\mathrm{I}} = \varphi(B)\varepsilon_i,$$

and

$$(2.6) \qquad \varepsilon_i = \psi(B)\xi_i^{\mathrm{I}}.$$

This implies (see e.g. Shaman 1975) the spectral density of ξ_i^{I} of the form

$$(2.7) \qquad f^{\mathrm{I}}(\lambda) = \sigma_\varepsilon^4 (2\pi)^{-2} (f(\lambda))^{-1} \sim c_f^{\mathrm{I}} |\lambda|^{-\alpha^{\mathrm{I}}} (\text{as } \lambda \to 0),$$

where $c_f^{\mathrm{I}} = \sigma_\varepsilon^4(2\pi)^{-2}(c_f)^{-1}$ and $\alpha^{\mathrm{I}} = -\alpha$. Equation (2.7) implies that: 1. If ξ_i is a short-memory process, then ξ_i^{I} also has short memory (in particular, the inverse process of an iid process is the process itself); 2. If ξ_i is a long-memory process with $0 < \alpha < 1$, then ξ^{I} is antipersistent with $\alpha^{\mathrm{I}} = -\alpha$, and vice versa.

From (2.3) we see that the autocovariances of ξ_i are $\gamma(k) = \sigma_\varepsilon^2 \sum a_j a_{j+|k|}$. The inverse autocovariances of ξ_i (Cleveland, 1972 and Chatfield, 1979), i.e. the autocovariances of ξ_i^{I}, are given by $\gamma^{\mathrm{I}}(k) = \sigma_\varepsilon^2 \sum_{j=0}^\infty b_j b_{j+|k|}$. Hence we have $\sum \gamma(k) = \sigma_\varepsilon^2 (\sum a_j)^2$ and $\sum \gamma^{\mathrm{I}}(k) = \sigma_\varepsilon^2 (\sum b_j)^2$. This results in $\sum a_j = \infty$, $\sum b_j = 0$ for $\alpha > 0$ and $\sum a_j = 0$, $\sum b_j = \infty$ for $\alpha < 0$. For $\alpha = 0$ we have $0 < \sum a_j < \infty$ and $0 < \sum b_j < \infty$.

An example of processes having property (1.2) is the class of FARIMA(p, d, q) (fractional ARIMA) processes (Granger and Joyeux, 1980 and Hosking 1981), with $d \in (-0.5, 0.5)$ denoting the fractional differencing or memory parameter.

3 Optimal convergence rates

3.1 Assumptions on the innovation distribution

Consider the problem of proving that a given sequence is a lower bound to the convergence rate in nonparametric regression with error process ξ_i. In the following it will be shown that this prove is similar to the case of nonparametric regression with iid errors ε_i. Furthermore, it turns out that the required conditions on the marginal distribution of ε_i under model (1.1) are the same, independently of $\alpha \in (-1, 1)$ i.e. of the type of long-term dependence. In the following we will adapt the regularity conditions of Stone (1980, 1982) to fixed-design nonparametric regression. Assume that $Z(g)$ is a real random variable depending on $g \in \Re$. It is assumed that the density function $f(z, g)$ is strictly positive and that $f(z, g) = f(z - g, 0)$, where g is the mean function of $Z(g)$, i.e.

$$\int z f(z, g) dz = g$$

for all $g \in \Re$. Furthermore, we assume that the equation

$$\int f(z, g) dz = 1$$

can be twice continuously differentiated with respect to g, yielding

$$\int f'(z, g) dz = 0$$

and

$$\int f''(z,g)dz = 0.$$

The iid innovations ε_i are generated by the marginal distribution with density $f(z,0)$, which will be simply denoted by $f(z)$ in the following. Using this notation the density of $Z(g)$ may be represented as $f(z,g) = f(z-g)$. Set $l(z,g) = \log f(z,g)$. It is assumed that there are positive constants τ_0, C and a function $M(z,g)$ such that for $g \in \Re$

$$|l''(z,g+\tau)| \le M(z,g) \ (for \ |\tau| \le \tau_0)$$

and

$$\int M(z,g)f(z,g)dz \le C.$$

Note that the last condition holds, if $l''(z,g)$ is bounded.

Remark 1. It is easy to show that the conditions hold, if $Z(g)$ is Gaussian with

$$f(z,g) = (2\pi\sigma_\varepsilon^2)^{-\frac{1}{2}}\exp\{-\frac{1}{2}(z-g)^2/\sigma_\varepsilon^2\} \ (-\infty < z,g < \infty)$$

or if the innovations ε_i are t_m-distributed with $m \ge 3$, i.e.

$$f_m(z,g) = \frac{\Gamma[(m+1)/2]}{\Gamma(m/2)\sqrt{m\pi}}\left(1+\frac{(z-g)^2}{m}\right)^{-(m+1)/2} \ (-\infty < z,g < \infty).$$

Remark 2. Note that some of the other distributions considered in Stone (1980), such as the exponential distribution, do not satisfy the regularity conditions above. If ε_i are iid exponentially distributed with $E(\varepsilon_i) = 0$ and var $(\varepsilon_i) = \lambda$, then the density function of $Z(g)$ is given by

$$f(z,g) = \lambda^{-1}\exp\{-(z+\lambda-g)/\lambda\}, \ (-\infty < g < \infty, g - \lambda \le z < \infty)$$

and zero otherwise. Thus, the support of $f > 0$ depends on g.

3.2 Lower bounds to convergence rates

For the minimax optimal convergence rate we will use the following definition (see e.g. Farrell, 1972, Stone, 1980 and Hall and Hart, 1990a). Let $\nu < k$ be a nonnegative integer, r_ν a positive number, and let $\tilde{g}_n^{(\nu)}$ denote a generic nonparametric estimator of $g^{(\nu)}$ based on $(Y_1, ..., Y_n)$. The sequence n^{-r_ν} is called a *lower bound to the convergence rate* at x_0 if

(3.1) $$\liminf_n \sup_{g \in \mathbb{C}^k} P(|\tilde{g}_n^{(\nu)}(x_0) - g^{(\nu)}(x_0)| > c_\nu n^{-r_\nu}) > 0$$

for c_ν sufficiently small. Moreover, n^{-r_ν} is called an *achievable convergence rate* if there is a sequence of estimators $\hat{g}_n^{(\nu)}$ such that

$$(3.2) \qquad \lim_{c_\nu \to \infty} \limsup_n \sup_{g \in \mathcal{C}^k} P(|\hat{g}_n^{(\nu)}(x_0) - g^{(\nu)}(x_0)| > c_\nu n^{-r_\nu}) = 0.$$

Also, the sequence n^{-r_ν} is called the *optimal convergence rate* if it is an achievable lower bound to the convergence rate. The optimal convergence rate for a nonparametric regression estimator of $g^{(\nu)}$ with iid errors is $n^{-(k-\nu)/(2k+1)}$ (Stone, 1980). Moreover, $n^{-(k-\nu)/(2k+1)}$ is also the optimal convergence rate for estimating $g^{(\nu)}$ in nonparametric regression with short-memory errors (results for $\nu = 0$ may be found in Hall and Hart, 1990a). In the case with $0 < \alpha < 1$, Hall and Hart (1990a) showed that the optimal convergence rate for estimating g is $n^{-(1-\alpha)k/(2k+1-\alpha)}$. In this paper, we will show that n^{-r_ν} with $r_\nu = (1-\alpha)(k-\nu)/(2k+1-\alpha)$ is the optimal convergence rate for estimating $g^{(\nu)}$, uniformly for $\alpha \in (-1, 1)$. The following theorem shows at first that n^{-r_ν} is a lower bound to the convergence rate, i.e. n^{-r_ν} satisfies (3.1).

Theorem 1 *Consider model (1.1) with $g \in \mathcal{C}^k$. Let $x_0 \in (0, 1)$ be an interior point of the support of g. Furthermore, let $\nu < k$ and $r_\nu = (1-\alpha)(k-\nu)/(2k+1-\alpha)$. Assume that the regularity conditions on the marginal distribution of the innovations given in section 3.1 hold. Then n^{-r_ν} is a lower bound to the convergence rate for estimating $g^{(\nu)}(x_0)$.*

The proof of Theorem 1 is given in the appendix.

Theorem 1 extends previous results as obtained by Stone (1980) and Hall and Hart (1990a) in different ways. The results in Stone (1980) are extended to nonparametric regression with fractional time series errors. The results by Hall and Hart (1990a) are generalized in the following way: 1. Theorem 1 is derived for all $\alpha \in (-1, 1)$ including the antipersistent case; 2. The results are obtained for non-Gaussian error processes satisfying regularity conditions on the marginal innovation distribution; 3. Estimation of derivatives is also considered.

Remark 3. The sequence n^{-r_ν} as defined in Theorem 1 is of course also a lower bound to the convergence rate for the estimation at the two boundary points $x_0 = 0$ or $x_0 = 1$, since the set of all measurable functions of observations at $x_0 = 0$ (rep. $x_0 = 1$) under the restriction that there are no observations on the left (right) hand side is a subset of all measurable functions.

Remark 4. In the proof of Theorem 1 a two-point discrimination argument is used. It will be shown that the probability on the right hand side of (3.1) can be made arbitrarily close to $\frac{1}{2}$. If a more sophisticated multi-point discrimination argument is used as in Stone (1980), then it can be shown that

$$(3.3) \qquad \lim_{c_\nu \to 0} \liminf_n \sup_{g \in \mathcal{C}^k} P(|\tilde{g}_n^{(\nu)}(x_0) - g^{(\nu)}(x_0)| > c_\nu n^{-r_\nu}) = 1.$$

Remark 5. Results of Theorem 1 are in general not available for random design nonparametric regression or density estimation with dependent observations, since the effect of dependence in such cases tends to weaker than in the model discussed here (see e.g. Hall and Hart 1990b and Csörgö and Mielniczuk, 1995a, b).

3.3 Achievability

Beran and Feng (2002a) showed that for $g \in \mathcal{C}^k$ with $k - \nu$ even, the uniform convergence rate of the local polynomial estimator $\hat{g}^{(\nu)}$ is of order n^{-r_ν} for all $x \in [0, 1]$, if a bandwidth of the optimal order $n^{-(1-\alpha)/(2k+1-\alpha)}$ is used, with r as defined in Theorem 1 (see Theorem 2 in Beran and Feng, 2002a). Similar results hold for the function class $\mathcal{C}^k$ with $k - \nu > 0$ odd. This result can be used to show the achievability of the lower bound to the convergence rate as defined in Theorem 1. In other words, (3.2) holds for the local polynomial estimator $\hat{g}^{(\nu)}$ with n^{-r_ν}, including the two boundary points $x_0 = 0$ and $x_0 = 1$. We thus have

Theorem 2 *Let $x_0 \in [0, 1]$. Under the conditions of Theorem 1, n^{-r_ν} is the optimal convergence rate for estimating $g^{(\nu)}(x_0)$.*

The detailed proof of Theorem 2 is straightforward (as outlined here) and is therefore omitted to save space.

Remark 6. The convergence rate n^{-r_ν} as defined in Theorem 1 may be achieved under much weaker conditions. It is clear that, (3.2) holds, if $\hat{g}^{(\nu)}$ is asymptotically normal. Sufficient conditions under which $\hat{g}^{(\nu)}$ is asymptotically normal are given, for instance, in Beran and Feng (2001). These conditions are much weaker than those in section 3.1.

4 Auxiliary results

4.1 Notations

Note that $r_\nu < 1$ for all $\alpha \in (-1, 1)$ and that the interpolation error is of order n^{-1}, and is hence negligible. We thus may assume without loss of generality that x_0 is of the form i_0/n. It is notationally convenient to have $x_0 = i_0/n = 0$, so that we will consider the shifted model

$$Y_i = g(i/n) + \xi_i, i = -n, ..., -1, 0, 1, ..., n,$$

and estimate $g^{(\nu)}$ at the origin. Moreover, we shall assume that both, the infinite past and the infinite future, are given, i.e. we observe

(4.1) $$Y_i = g(i/n) + \xi_i, -\infty < i < \infty.$$

Model (4.1) is assumed only for notational convenience, which helps us to save symbols for distinguishing finite and infinite sample paths. It turns out that the extra information is of negligible benefit for the derivation of a lower bound to the convergence rate.

The main idea to prove Theorem 1 is to construct two sequences of functions. If these two sequences are "hard to distinguish", then their difference will form a lower bound to the convergence rate. If they are "far apart" at the same time, then their difference will form an achievable convergence rate, and we will therefore obtain the optimal convergence rate. Following Stone (1980) and Hall and Hart (1990a), let $\Psi \geq 0$ be a $k + 1$-differentiable function on $(-\infty, \infty)$, vanishing outside $(-1, 1)$ and satisfying $\Psi^{(\nu)}(0) > 0$ for $\nu = 0, 1, ..., k$. Set

$$B' = \sup_{0 \leq x \leq 1} \max_{\nu = 0, 1, ..., k} |\psi^{(\nu)}(x)|,$$

and choose $a > 0$ so small that $aB' < B$. Moreover, let $0 < s < 1$, set $h = n^{-s}$ and define

(4.2) $$g_\theta(x) = \theta a h^k \Psi(x/h).$$

Then $g_\theta(x)$ for $\theta \in \{0, 1\}$ are two sequences of functions in $\mathcal{C}^k$.

In the following we will denote the limits $\lim_{n \to \infty} \prod_{i=-n}^{n}$ and $\lim_{n \to \infty} \sum_{i=-n}^{n}$ by $\prod$ and $\sum$ for simplicity. For $-\infty < i < \infty$, define the doubly infinitive column vectors $\boldsymbol{\xi} = (\xi_i)$, $\boldsymbol{\varepsilon} = (\varepsilon_i)$, $\mathbf{g} = (g_1(i/n))$, and the doubly infinite matrices $\boldsymbol{\Sigma} = (\gamma(i - j))$, $\boldsymbol{\Gamma} = (\gamma^I(i - j))$,

$\mathbf{\Omega} = (a_{i-j})$ and $\mathbf{\Lambda} = (b_{i-j})$, where $a_i = b_i = 0$ for $i < 0$. The explicit form of $\mathbf{\Lambda}$ is given in (A.6) in the appendix. Then we have $\boldsymbol{\xi} = \mathbf{\Omega}\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon} = \mathbf{\Lambda}\boldsymbol{\xi}$. For $\mathbf{Y} = (Y_i)$ we have $\mathbf{Y}_\theta = \theta\mathbf{g} + \boldsymbol{\xi}$. Also, defining $\mathbf{X}_\theta = \mathbf{\Lambda}\mathbf{Y}_\theta = \theta\boldsymbol{\eta} + \boldsymbol{\varepsilon}$, with $\boldsymbol{\eta} = \mathbf{\Lambda}\mathbf{g}$, we have $\mathbf{X}_0 = \boldsymbol{\varepsilon}$, $\mathbf{Y}_0 = \boldsymbol{\xi}$. Moreover, $\mathbf{X}_1$ is a sequence of independent random variables.

4.2   The likelihood functions and the error probabilities

Let L_0 and $\mathcal{L}_0$ denote the likelihood functions of $\mathbf{X}_0 = \boldsymbol{\varepsilon}$ and $\mathbf{Y}_0 = \boldsymbol{\xi}$, respectively. Observe that $L_0(x) = \prod f(x_i)$, where $x = (..., x_{-1}, x_0, x_1, ...)'$ is a doubly infinite vector and f is the marginal density function of ε_i. The following lemma gives the relationship between these two likelihood functions.

Lemma 1 *For the fractional time series process defined by (2.3) and (2.4), and a doubly infinite real vector y we have*

$$(4.3) \qquad\qquad \mathcal{L}_0(y) = L_0(x) = \prod_{i=-\infty}^{\infty} f(x_i),$$

where $x = \mathbf{\Lambda}y$ with $x_i = \sum_{j=-\infty}^{\infty} b_j y_{i-j}$, $-\infty < i < \infty$, and f is the marginal density function of ε_i.

The proof of Lemma 1 is given in the appendix. Lemma 1 shows that $\mathcal{L}$ is uniquely determined by L. Note that, the opposite is also true, i.e. L is uniquely determined by $\mathcal{L}$. Following Lemma 1 estimation of the likelihood function of an invertible stationary time series is equivalent to that of the corresponding iid innovations. The idea behind this lemma plays a very important role for the derivation of asymptotic results in non-parametric regression with dependent errors. Discussions on asymptotic results in this case may therefore often be reduced to those for models with iid errors after a suitable transformation. Note finally that Lemma 1 only holds for causal processes.

Let L_1 and $\mathcal{L}_1$ denote the likelihood functions of $\mathbf{X}_1 = \boldsymbol{\varepsilon} + \boldsymbol{\eta}$ and $\mathbf{Y}_1 = \boldsymbol{\xi} + \mathbf{g}$, respectively. To prove Theorem 1 we need to estimate $P(\mathcal{L}_0 < \mathcal{L}_1|\theta = 0)$ and $P(\mathcal{L}_0 > \mathcal{L}_1|\theta = 1)$. The following corollary of lemma 1 reduces the estimation of these error probabilities to the case of independent sequences $\mathbf{X}_\theta$.

Corollary 1 *Let $\mathbf{X}_\theta$ and $\mathbf{Y}_\theta$ are defined above. Let y is a doubly infinite real vector. Then, under the assumptions of Lemma 1, we have*

$$P(\mathcal{L}_0(y) < \mathcal{L}_1(y)|\theta = 0) = P(L_0(x) < L_1(x)|\theta = 0)$$

11

and

$$P(\mathcal{L}_0(y) > \mathcal{L}_1(y)|\theta = 1) = P(L_0(x) > L_1(x)|\theta = 1),$$

where $x = \mathbf{\Lambda}y$.

The proof of Corollary 1 is given in the appendix. Following Corollary 1, a method for estimating the error probability developed for nonparametric regression with iid errors could be adapted to the current case. We will do this, adapting the methodology proposed by Stone (1980). Note that $\boldsymbol{\eta}$, the deterministic part of $\mathbf{X}_1$, does not necessarily have the same smoothness properties as $\mathbf{g}$, the deterministic part of $\mathbf{Y}_1$. However, this does not affect the estimation of the error probability.

4.3   A sufficient condition

Let $\Upsilon_n = \frac{1}{2}g_1(0) = \frac{1}{2}a\Psi(0)h^k = c_0h^k$, where $c_0 = \frac{1}{2}a\Psi(0)$. Moreover, set $\Upsilon_n^\nu = c_\nu h^{(k-\nu)}$, where $c_\nu = \frac{\nu!}{2}a\Psi^{(\nu)}(0)$ for $\nu < k$. If ξ_i in model (1.1) are iid, then, following Stone (1980), it can be shown that a sufficient condition, under which Υ_n^ν is a lower rate of convergence for estimating $g^{(\nu)}$, is that there is an $M > 0$ such that $\sum g^2(i/n) < M$ (see equation (2.1) in Stone, 1980). The following lemma gives a simple extension of this result to the case where ξ_i are fractional stationary time series errors defined by (2.3) and (2.4).

Lemma 2 *Let ξ_i be defined by (2.3) and (2.4). Consider the estimation of $g^{(\nu)}$. Then Υ_n^ν is a lower rate of convergence, if there is an $M > 0$ such that*

$$(4.4) \qquad \sum_{i=-\infty}^{\infty} \eta_i^2 = \mathbf{g}'\mathbf{\Lambda}'\mathbf{\Lambda}\mathbf{g} < M,$$

where η_i are the elements of $\boldsymbol{\eta} = \mathbf{\Lambda}\mathbf{g}$.

The proof of Lemma 2 is given in the appendix. Note that $g_0 \equiv 0$ and hence $\mathbf{g}$ is the difference sequence between the two functions g_0 and g_1. Lemma 2 shows that this sequence will form a lower rate of convergence for estimating g, if the transformed difference sequence $\boldsymbol{\eta}$ is square summable. From Lemma 2 we can also see that, if Υ_n is a lower rate of convergence for estimating g, then Υ_n^ν, the sequence of the ν-th derivative Υ_n, is a lower rate of convergence for estimating $g^{(\nu)}$ provided that $\Psi^{(\nu)}(0) > 0$.

It is easy to show that condition (4.4) is equivalent to

$$(4.5) \qquad\qquad\qquad \mathbf{g}'\mathbf{\Gamma}\mathbf{g} < \sigma_\varepsilon^2 M$$

and further equivalent to

$$(4.6) \qquad \mathbf{g}'\mathbf{\Sigma}^{-1}\mathbf{g} < \sigma_\varepsilon^{-2}M.$$

Proofs of (4.5) and (4.6) are given in the appendix. These two representations are easy to understand. Equation (4.6) directly shows the change in this sufficient condition caused by the dependence structure. The following remarks clarify the results further.

Remark 7. For iid errors $\xi_i = \varepsilon_i$ we have $\mathbf{\Lambda} = \mathbf{I}$, $\mathbf{\Gamma} = \sigma_\varepsilon^2\mathbf{I}$ and $\mathbf{\Sigma}^{-1} = \sigma_\varepsilon^{-2}\mathbf{I}$, where $\mathbf{I}$ denotes the doubly infinite identity matrix. In this case we have simply $\sum g^2(i/n) < M$. Note that $D = \sqrt{\sum g^2(i/n)}$ is the L^2-norm of $\mathbf{g}$. Lemma 2 implies that any method of deciding between $\theta = 0$ and $\theta = 1$, i.e. of deciding between the vector $\mathbf{g}$ and the zero vector, must have overall positive error probability, if the norm of $\mathbf{g}$ is bounded.

Remark 8. Assume that ε_i are normal. Following Hall and Hart (1990a) it can be shown that, the overall error probability of any estimator of θ based on $\mathbf{Y}$ is at least

$$(4.7) \qquad P_a = 1 - \Phi\left(\left(\mathbf{g}'\mathbf{\Sigma}^{-1}\mathbf{g}\right)^{1/2}\right),$$

where Φ is the standard normal distribution function. The error probability P_a will be positive, if $\mathbf{g}'\mathbf{\Sigma}^{-1}\mathbf{g}$ is finite. P_a in (4.7) can be made arbitrarily close to $\frac{1}{2}$ by choosing the constant a in (4.2) so that $a \to 0$ and hence $\mathbf{g}'\mathbf{\Sigma}^{-1}\mathbf{g} \to 0$.

Appendix: Proofs

Proof of Lemma 1. It is well known that, under standard conditions, the likelihood functions of two random vectors forming a reciprocal one-to-one mapping are uniquely determined by each other (see e.g. Theorem 2 of Section 4.4 in Rohatgi and Saleh, 2001, pp. 127ff). Note that this result can be extended to doubly infinite random vectors. To prove Lemma 1 it is therefore sufficient to check that all conditions of this theorem hold. At first, $\boldsymbol{\varepsilon} = \boldsymbol{\Lambda}\boldsymbol{\xi}$ form a doubly infinite dimensional reciprocal one-to-one-mapping with the inverse transformation $\boldsymbol{\xi} = \boldsymbol{\Omega}\boldsymbol{\varepsilon}$, where both, the original function and the inverse transformation are linear. Hence, conditions (a) to (c) of Theorem 2 of Section 4.4 in Rohatgi and Saleh (2001) hold. Furthermore, $\boldsymbol{\Lambda}$ is also the matrix of the partial derivatives of $\boldsymbol{\varepsilon}$ with respect to $\boldsymbol{\xi}$. Also, the Jacobian J of the inverse transformation is the determinant $|\boldsymbol{\Lambda}| = 1$, since $\boldsymbol{\Lambda}$ is a (doubly infinite) lower triangle matrix, whose diagonal elements are identically equal to one. The relationship between $\mathcal{L}_0$ and L_0 as given in Lemma 1 holds. $\diamond$

Proof of Corollary 1. Observe that $\mathbf{X}_1 = \mathbf{X}_0 + \boldsymbol{\eta}$ and $\mathbf{Y}_1 = \mathbf{Y}_0 + \mathbf{g}$. Hence we have, $L_1(x) = L_0(x - \boldsymbol{\eta})$ and $\mathcal{L}_1(y) = \mathcal{L}_0(y - \mathbf{g})$. From Lemma 1, it follows that, for any doubly infinite dimensional real vectors y and $\mathbf{g}$,

$$
\begin{aligned}
\mathcal{L}_1(y) &= \mathcal{L}_0(y - \mathbf{g}) \\
&= L_0(x - \boldsymbol{\eta}) \\
&= L_1(x) = \prod_{i=-\infty}^{\infty} f(x_i - \eta_i),
\end{aligned}
$$
(A.1)

where $x = \boldsymbol{\Lambda}y$, $\boldsymbol{\eta} = \boldsymbol{\Lambda}\mathbf{g}$ and f is the marginal density function of ε_i. Equations (4.3) and (A.1) together imply that $\mathcal{L}_0(y) < \mathcal{L}_1(y)$ (or $\mathcal{L}_0(y) > \mathcal{L}_1(y)$, or $\mathcal{L}_0(y) = \mathcal{L}_1(y)$), if and only if $L_0(x) < L_1(x)$ (or $L_0(x) > L_1(x)$, or $L_0(x) = L_1(x)$), where $x = \boldsymbol{\Lambda}y$. Corollary 1 then follows. $\diamond$

The proofs given in the following are related to those in Stone (1980) and Hall and Hart (1990a). Some details will therefore be omitted to save space (we refer the reader to the proofs in these papers). We also refer the reader to Theorem 1 in Hall (1989) and its proof. Note that the definition of α used here differs from Hall and Hart (1990a).

Proof of Lemma 2. Let Υ_n^{ν} be as defined in Lemma 2. Note that

$$
\sup_{g \in \mathbb{C}^k} P_g\{|\tilde{g}_n^{(\nu)}(0) - g^{(\nu)}(0)| \geq \Upsilon_n^{\nu}\} \geq \max_{\theta=0,1} P_{\theta}\{|\tilde{g}^{(\nu)}(0) - g_{\theta}^{(\nu)}(0)| \geq \Upsilon_n^{\nu}\}.
$$

Let $\tilde{\theta}_n = 0$ or 1 depending on which value minimizes $|\tilde{g}_n^{(\nu)}(0) - g_{\tilde{\theta}}^{(\nu)}(0)|$. Then $\tilde{\theta}_n \neq \theta$ implies $|\tilde{g}_n^{(\nu)}(0) - g_\theta^{(\nu)}(0)| \geq \Upsilon_n^\nu$, and hence

$$
\begin{aligned}
\max_{\theta=0,1} P_\theta\{|\tilde{g}^{(\nu)}(0) - g_\theta^{(\nu)}(0)| \geq \Upsilon_n^\nu\} \quad &\geq \quad \max_{\theta=0,1} P_\theta(\tilde{\theta} \neq \theta) \\
&\geq \quad \frac{1}{2}\{P_0(\tilde{\theta} = 1) + P_1(\tilde{\theta} = 0)\} \\
&\geq \quad \frac{1}{2}\{P_0(\hat{\theta} = 1) + P_1(\hat{\theta} = 0)\},
\end{aligned}
$$

(A.2)

where $\hat{\theta}$ is the maximum likelihood estimator of θ (or the likelihood ratio discriminator) in the two-parameter problem. The last inequality follows from the Neyman-Pearson lemma. From Corollary 1 we have

$$
\begin{aligned}
\max_{\theta=0,1} P_\theta\{|\tilde{g}^{(\nu)}(0) - g_\theta^{(\nu)}(0)| \geq \Upsilon_n^\nu\} \quad &\geq \quad \frac{1}{2}(P_0(\mathcal{L}_0 < \mathcal{L}_1) + P_1(\mathcal{L}_1 < \mathcal{L}_0)) \\
&= \quad \frac{1}{2}(P_0(L_0 < L_1) + P_1(L_1 < L_0)).
\end{aligned}
$$

(A.3)

Let L_R denote the likelihood ratio L_1/L_0. By calculations similar to those given on pages 1352 - 1353 of Stone (1980), it can be shown under the regularity conditions on the marginal distribution of ε_i given in Section 3.1, that there is a positive constant M_1 such that

(A.4)
$$
E_0|\log(L_R)| < M_1
$$

and

(A.5)
$$
\lim_{a \to 0} E_0|\log(L_R)| = 0.
$$

Similar formulas as given in (A.4) and (A.5) hold for the expectation under $\theta = 1$ with another positive constant M_2. Let $M_0 = \max(M_1, M_2)$. Then we can find an integer $K \geq 2$ and $0 < \tau < \frac{1}{2}$ such that if $L_R > (1 - \tau)/\tau$ or $L_R < \tau/(1 - \tau)$, then $|\log(L_R)| \geq KM_0$. Following the Markov inequality, we obtain

$$
P_0\left(\frac{\tau}{1 - \tau} \leq L_R \leq \frac{1 - \tau}{\tau}\right) > \frac{K - 1}{K}
$$

and

$$
P_1\left(\frac{\tau}{1 - \tau} \leq L_R \leq \frac{1 - \tau}{\tau}\right) > \frac{K - 1}{K}.
$$

Assigning priori probabilities 1/2 to $\theta = 0$ and $\theta = 1$, we have

$$
P(\theta = 1|\mathbf{Y}) = \frac{\frac{1}{2}L_1}{\frac{1}{2}L_1 + \frac{1}{2}L_0} = \frac{L_R}{L_R + 1}
$$

15

and

$$P(\tau \le P(\theta = 1|\mathbf{Y}) \le 1 - \tau) = P\left(\tau \le \frac{L_{\mathrm{R}}}{L_{\mathrm{R}} + 1} \le 1 - \tau\right)$$

$$= P\left(\frac{\tau}{1 - \tau} \le L_{\mathrm{R}} \le \frac{1 - \tau}{\tau}\right)$$

$$= \frac{1}{2}P_0\left(\frac{\tau}{1 - \tau} \le L_{\mathrm{R}} \le \frac{1 - \tau}{\tau}\right)$$

$$+ \frac{1}{2}P_1\left(\frac{\tau}{1 - \tau} \le L_{\mathrm{R}} \le \frac{1 - \tau}{\tau}\right)$$

$$\ge \frac{K - 1}{K}.$$

Thus, the error probability of $\hat{\theta}$ is at least $\frac{K-1}{K}\tau$. Note that $\frac{K-1}{K}\tau$ can be made arbitrarily close to $\frac{1}{2}$ as $\delta \to 0$ by choosing K sufficiently large and τ sufficiently close to $\frac{1}{2}$ at the same time. $\diamond$

Proof of equations (4.5) and (4.6). The explicit form of the matrix $\mathbf{\Lambda}$ is

$$(A.6) \qquad \mathbf{\Lambda} = \begin{pmatrix} \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \\ \cdots & 1 & 0 & 0 & \cdots & 0 & 0 & \cdots \\ \cdots & b_1 & 1 & 0 & \cdots & 0 & 0 & \cdots \\ \cdots & b_2 & b_1 & 1 & \cdots & 0 & 0 & \cdots \\ \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \cdots \\ \cdots & b_{n-1} & b_{n-2} & b_{n-3} & \cdots & 1 & 0 & \cdots \\ \cdots & b_n & b_{n-1} & b_{n-2} & \cdots & b_1 & 1 & \cdots \\ \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{pmatrix}.$$

Ω has an analogous form but with b_{i-j} being replaced by a_{i-j}. Following the definition of $\gamma^{\mathrm{I}}(i - j)$ we have $\mathbf{\Gamma} = \sigma_\varepsilon^2 \mathbf{\Lambda}\mathbf{\Lambda}'$. Furthermore, it can be shown that $\mathbf{\Lambda}\mathbf{\Lambda}' = \mathbf{\Lambda}'\mathbf{\Lambda}$. The equivalence between (4.4) and (4.5) follows from this fact. The equivalence between the two conditions (4.5) and (4.6) is due to the fact that $\mathbf{\Sigma}^{-1} = \sigma_\varepsilon^{-4}\mathbf{\Gamma}$ in the sense that $\mathbf{\Sigma}\mathbf{\Gamma}/\sigma_\varepsilon^4 = \mathbf{I}$ (see e.g. Shaman 1975 and Beran 1994, pp. 109 ff.). $\diamond$

Proof of Theorem 1. Without loss of generality we will assume that $\sigma_\varepsilon^2 = 1$ for convenience. For $\nu = 0$ let $\Upsilon_n = c_0 h^k$ equal to the rate $c_0 n^{-r_0}$, where $r_0 = (1 - \alpha)k/(2k + 1 - \alpha)$ is as defined in Theorem 1. Then we have $h = n^{-s}$ with $s = (1 - \alpha)/(2k + 1 - \alpha)$. Following Lemma 2, in order to see that $c_\nu n^{-r_\nu}$ is a lower rate of convergence for estimating $g^{(\nu)}$, we have to show that the sequence $\mathbf{g}$ under this choice of h satisfies e.g. the condition $\sum \eta_i^2 = \mathbf{g}'\mathbf{\Lambda}'\mathbf{\Lambda}\mathbf{g} < \infty$,

Let $m = [nh]$ be the integer part of nh. Let $v_i = \Psi(i/m)$, $-\infty < i < \infty$, and let $\boldsymbol{v} = (v_i)$ denote the corresponding doubly infinite vector. Then we have

$$\mathbf{g'\Lambda'\Lambda g} = \frac{1}{4}h^{2k}\Psi^2(0)\boldsymbol{v'\Lambda'\Lambda v}.$$

Observe that $v_i = 0$ for $i < -m$ or $i > m$. We have

$$
\boldsymbol{v'\Lambda'\Lambda v} = \sum_{j=-\infty}^{\infty}\left(\sum_{i=-m}^{m} v_i b_{i+j}\right)^2
$$

(A.7)
$$
= (2m+1)\sum_{k=-2m}^{2m}\gamma^{\mathrm{I}}(k)\frac{1}{2m+1}\sum_{j=-m}^{m}\Psi(j/m)\Psi\{(j+k)/m)\}.
$$

Equation (A.7) can also be obtained by directly analyzing $\boldsymbol{v'\Gamma v}$.

Based on (A.7) we can obtain results for the cases $\alpha = 0$, $0 < \alpha < 1$ and $-1 < \alpha < 0$ separately. Note that the methodology used in the proof of Theorem 3.1 in Hall and Hart (1990a) for the case with $0 < \alpha < 1$ is based on the assumption $b_{-i} = b_i$ for $i = 1, 2, ...$, and is hence not suitable for the causal error process in this paper, since now we have $b_i = 0$ for $i < 0$. The methodology used in the following is developed based on the property (1.2) of a fractional time series, which does not involve the exact structure of b_i.

Assume that $\alpha = 0$. In this case $\sum\gamma(k)^{\mathrm{I}} > 0$ and $\sum|\gamma(k)^{\mathrm{I}}| < \infty$. From (A.7) we have

$$\boldsymbol{v'\Lambda'\Lambda v} \doteq (2m+1)\left(\sum\gamma^{\mathrm{I}}(k)\right)\int_{-1}^{1}\Psi^2(u)du.$$

Note that $h = n^{-1/(2k+1)}$ and $m = nh = n^{2k/(2k+1)} = h^{-2k}$ for $\alpha = 0$, whence

$$\frac{1}{4}h^{2k}\Psi^2(0)\boldsymbol{v'\Lambda'\Lambda v} < \infty.$$

In the case with $0 < \alpha < 1$ the inverse process ξ^{I} is an antipersistent process with the parameter $-1 < \alpha^{\mathrm{I}} = -\alpha < 0$ in (2.7). Hence, for $|k|$ sufficiently large, we have $\gamma^{\mathrm{I}}(k) \sim c_\gamma^{\mathrm{I}}|k|^{-\alpha-1}$, where $c_\gamma = 2c_f^{\mathrm{I}}\Gamma(1-\alpha^{\mathrm{I}})\sin(\pi\alpha^{\mathrm{I}}/2) < 0$ (see Beran, 1994 and Beran and Feng, 2002a). This implies that $\gamma^{\mathrm{I}}(k)$ are ultimately negative for $|k|$ sufficiently large. Furthermore, we have $\sum\gamma^{\mathrm{I}}(k) = 0$ and hence $\sum_{k=-m}^{m}\gamma^{\mathrm{I}}(k) = -2\sum_{k>m}\gamma^{\mathrm{I}}(k) = O(m^{-\alpha})$. From (A.7) we then have

$$
\begin{aligned}
\boldsymbol{v'\Lambda'\Lambda v} &= (2m+1)\sum_{k=-2m}^{2m}\gamma^{\mathrm{I}}(k)\frac{1}{2m+1}\sum_{j=-m}^{m}\Psi(j/m)\Psi\{(j+k)/m)\} \\
&\leq (2m+1)\sum_{k=-m}^{m}\gamma^{\mathrm{I}}(k)\frac{1}{2m+1}\sum_{j=-m}^{m}\Psi(j/m)\Psi\{(j+k)/m)\} \\
&= (2m+1)O\left(\sum_{k=-m}^{m}\gamma^{\mathrm{I}}(k)\right) = O(m^{1-\alpha}).
\end{aligned}
$$

17

Now $h = n^{-(1-\alpha)/(2k+1-\alpha)}$ and $m = nh = n^{2k/(2k+1-\alpha)}$, so that $m^{1-\alpha} = h^{-2k}$ and

$$\frac{1}{4}h^{2k}\Psi^2(0)\boldsymbol{v}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\boldsymbol{v} = h^{2k}O(h^{-2k}) < \infty.$$

If $-1 < \alpha < 0$, the inverse process ξ^{I} is a long-memory process with the parameter $0 < \alpha^{\mathrm{I}} = -\alpha < 1$ in (2.7). Hence, as $|k| \to \infty$, $\gamma^{\mathrm{I}}(k) \sim c_\gamma^{\mathrm{I}}|k|^{-\alpha-1}$ with $c_\gamma = 2c_f^{\mathrm{I}}\Gamma(1 - \alpha^{\mathrm{I}})\sin(\pi\alpha^{\mathrm{I}}/2) > 0$, so that $\gamma^{\mathrm{I}}(k) > 0$ for $|k|$ sufficiently large. Furthermore, we have $\sum \gamma^{\mathrm{I}}(k) = \infty$ with $\sum_{-2m}^{2m} \gamma^{\mathrm{I}}(k) = O(m^{-\alpha})$. Note that Ψ can be chosen so that, for large k, $\sum_{j=-m}^{m} \Psi(j/m)\Psi\{(j+k)/m)\} < \sum_{j=-m}^{m} \Psi^2(j/m)$. Hence we have

$$
\begin{aligned}
\boldsymbol{v}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\boldsymbol{v} &= (2m+1)\sum_{k=-2m}^{2m} \gamma^{\mathrm{I}}(k)\frac{1}{2m+1}\sum_{j=-m}^{m} \Psi(j/m)\Psi\{(j+k)/m)\} \\
&\leq (2m+1)\sum_{k=-2m}^{2m} \gamma^{\mathrm{I}}(k)\frac{1}{2m+1}\sum_{j=-m}^{m} \Psi^2(j/m) \\
&\doteq (2m+1)\sum_{k=-2m}^{2m} \gamma^{\mathrm{I}}(k)\int_{-1}^{1} \Psi^2(u)du \\
&= O(m^{1-\alpha}).
\end{aligned}
$$

In summary, we have

$$\boldsymbol{v}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\boldsymbol{v} = O(m^{1-\alpha})$$

uniformly for $\alpha \in (-1, 1)$. Note, however, that the derivation of this result was different for the three cases. From $h = n^{-(1-\alpha)/(2k+1-\alpha)}$ and whence $m^{1-\alpha} = h^{-2k}$, we now have

$$\frac{1}{4}h^{2k}\Psi^2(0)\boldsymbol{v}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\boldsymbol{v} = h^{2k}O(h^{-2k}) < \infty,$$

and the proof of Theorem 1 is completed. $\diamond$

References

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

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

Beran, J. and Feng, Y. (2002a). Local polynomial fitting with long-memory, short-memory and antipersistent errors. Ann. Inst. Stat. Math., 54, 291-311.

Beran, J. and Feng, Y. (2002b). SEMIFAR models - A semiparametric framework for modelling trends, long-range dependence and nonstationarity. Comp. Stat. Data Anal., 40, 393-419.

Beran, J., Feng, Y., Franke, G., Hess, D. and Ocker, D. (2003). Semiparametric modeling of stochastic and deterministic trends and fractional stationarity. In: Processes with Long-Range Correlations — Theory and Applications, (G. Rangarajan, M. Ding, editors), Lecture Notes in Physics, Vol. 621, Springer, New York, pp. 225 – 250.

Beran, J. and Ocker, D. (2001). Volatitily of stock market indices - an analysis based on SEMIFAR models. Journal of Business & Economic Statistics, 19, No. 1, 103–116.

Chatfield, C. (1979). Inverse autocorrelations. J. R. Statist. Soc.. ser. A, 142, 363–377.

Cleveland, W.S. (1972). The inverse autocorrelations of a time series and their applications (with discussion). Technometrics, 14, 277–298.

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

Csörgö, S. and Mielniczuk, J. (1995a). Density Estimation Under Long-Range Dependence. Ann. Statist., Vol. 23, No. 3, 990–999.

Csörgö, S. and Mielniczuk, J. (1995b). Nonparametric Regression Under Long-Range Dependent Normal Errors. Ann. Statist., Vol. 23, No. 3, 1000-1014.

Farrell, R.H. (1972). On the best obtainable asymptotic rates of convergence in estimation of density function at a point. Ann. Math. Statist., 43, 170–180.

Feng, Y., Beran, J. and Yu, K. (2006). Modelling financial time series with SEMIFAR-GARCH model. Preprint, Heriort-Watt University.

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., 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. (1989). On convergence rates in nonparametric problems. Intern. Statiat. Review, 57, 45–58.

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

Hall, P. and Hart, J.D. (1990b). Convergence rates in density estimation for data from infinite-order moving average processes. Probab. Theory Rel. Fields, 87, 253–274.

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

Karuppiaha, J. and Losb, C.A. (2005). Wavelet multiresolution analysis of high-frequency Asian FX rates. Intern. Rev. Finan. Anal., 14 (Summer 1997), 211–246.

Li, L. and Xiao, Y. (2007). On the minimax optimality of block thresholded wavelet estimators with long memory data. J. Stat. Plann. Inf., in press.

Mungo, J. and Härdle, W. (2006). Long memory persistence in the factor of implied volatility dynamics. Preprint, Humboldt-University zu Berlin.

Rangarajan, G. and Ding, M. (eds) (2003). Processes with Long Range Correlations: Theory and Applications. Springer, New York.

Robinson, P.M. (2005). Robust Covariance Matrix Estimation: HAC Estimates with Long Memory/Antipersistence Correction. Economet. Th., 21, 171-180.

Shaman, P. (1975). An approximate inverse for the covariance matrix of moving average and autoregressive processes. Ann. Statist., 3, 532–538.

Silverberg, G. and Verspagen, B. (2003). Long Memory and Economic Growth in the World Economy Since the 19th Century. In: Processes with Long-Range Correlations — Theory and Applications, (G. Rangarajan, M. Ding, editors), Lecture Notes in Physics, Vol. 621, Springer, New York, p. 270-285.

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

Stone, C.J. (1980). Optimal rates of convergence for nonparametric estimators. Ann. Statist., 8, 1348–1360.

Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. Ann. Statist., 10, 1040–1053.

Tsai, H.S. (2006). On continuous time autoregressive fractionally integrated moving average processes. Preprints, Academia Sinica.