

WALDEMAR POPIŃSKI (Warszawa)

ORTHOGONAL SERIES REGRESSION ESTIMATION UNDER LONG-RANGE DEPENDENT ERRORS

Abstract. This paper is concerned with general conditions for convergence rates of nonparametric orthogonal series estimators of the regression function. The estimators are obtained by the least squares method on the basis of an observation sample $Y_i = f(X_i) + \eta_i$, $i = 1, \dots, n$, where $X_i \in A \subset \mathbb{R}^d$ are independently chosen from a distribution with density $\varrho \in L^1(A)$ and η_i are zero mean stationary errors with long-range dependence. Convergence rates of the error $n^{-1} \sum_{i=1}^n (f(X_i) - \hat{f}_N(X_i))^2$ for the estimator $\hat{f}_N(x) = \sum_{k=1}^N \hat{c}_k e_k(x)$, constructed using an orthonormal system e_k , $k = 1, 2, \dots$, in $L^2(A)$, are obtained.

1. Introduction. Consider a random design observation model $Y_i = f(X_i) + \eta_i$, $i = 1, \dots, n$, where the points X_i , $i = 1, \dots, n$, form a random sample from an absolutely continuous distribution X with density $\varrho \in L^1(A)$ on a compact subset A of some euclidean space \mathbb{R}^d , $d \geq 1$, $f \in L^2(A)$ is an unknown regression function and η_i , $i = 1, \dots, n$, are stationary second order errors with zero mean, variance σ_η^2 , and covariance function $E_\eta(\eta_1 \eta_{1+j}) = Cj^{-\gamma}$, $j = 1, 2, \dots$, where $0 < \gamma < 1$ and C is a real constant. We assume that the observation errors $\eta = (\eta_1, \dots, \eta_n)$ are independent of the predictors $\omega = (X_1, \dots, X_n)$.

Let us consider a series type regression function estimator

$$\hat{f}_N(x) = \sum_{k=1}^N \hat{c}_{kN} e_k(x),$$

where the functions e_k , $k = 1, 2, \dots$, form a complete orthonormal system in $L^2(A)$. We assume that the vector $\hat{c}_N = (\hat{c}_{1N}, \dots, \hat{c}_{NN})^T$ of coefficient

2000 *Mathematics Subject Classification*: 62G08, 62G20.

Key words and phrases: nonparametric series regression, least squares method, orthonormal system, convergence rate.

estimators is, for fixed N , obtained by the least squares method, i.e.

$$\widehat{c}_N = \arg \min_{a \in \mathbb{R}^N} \frac{1}{n} \sum_{i=1}^n (Y_i - \langle a, e^N(X_i) \rangle)^2,$$

where $e^N(x) = (e_1(x), \dots, e_N(x))^T$.

The vector \widehat{c}_N can be obtained as a solution of the normal equations

$$G_n \widehat{c}_N = g_n,$$

where

$$G_n = \frac{1}{n} \sum_{i=1}^n e^N(X_i) e^N(X_i)^T, \quad g_n = \frac{1}{n} \sum_{i=1}^n Y_i e^N(X_i),$$

and when $\det G_n \neq 0$ it is uniquely determined. As shown in the author's previous work on a related subject [14], in the case when $\varrho \geq c > 0$ the matrix $G(N) = EG_n$ is nonsingular. Then, if $\det G_n \neq 0$, the relevant estimator can be represented in the form $\widehat{f}_N(x) = \langle h^N(x), \widehat{b}_N \rangle$, where $h^N(x) = G(N)^{-1/2} e^N(x)$, and the vector $\widehat{b}_N = G(N)^{1/2} \widehat{c}_N$ satisfies the normal equations

$$(1) \quad H_n \widehat{b}_N = h_n,$$

where

$$H_n = \frac{1}{n} \sum_{i=1}^n h^N(X_i) h^N(X_i)^T, \quad h_n = \frac{1}{n} \sum_{i=1}^n Y_i h^N(X_i).$$

Representing the estimator \widehat{f}_N using the vector function h^N is convenient, since then $Eh^N(X)h^N(X)^T = I_N$, where I_N is the unit matrix, and consequently $E\|h^N(X)\|^2 = N$. Moreover, if we put

$$M_N = \|e^N\|_\infty = \sup_{x \in A} \|e^N(x)\|$$

and λ_n denotes the smallest eigenvalue of the matrix H_n , then we also have (inequality (3) in [14])

$$(2) \quad E|\lambda_n - 1|^2 \leq \frac{NM_N^2}{cn}.$$

For $\varrho \geq c > 0$ inequality (2) implies $P(\det G_n = 0) = P(\det H_n = 0) \leq P(\lambda_n < 1/2) \leq 4NM_N^2 c^{-1}/n$. Thus, the conditions $\varrho \geq c > 0$ and $NM_N^2/n \rightarrow 0$ as $n \rightarrow \infty$ assure that $P(\det G_n = 0) \rightarrow 0$, i.e. the estimator is uniquely determined with growing probability. The above facts will be used to prove the results of this work, which continues the investigations of asymptotic properties of series type regression estimators, started by the author in [11], [13].

In Section 2 we give some sufficient conditions for convergence rates in probability of the square prediction error

$$d_{nN} = \frac{1}{n} \sum_{i=1}^n (f(X_i) - \widehat{f}_N(X_i))^2$$

for orthogonal series regression estimators and for the observation model considered. They extend the results of [11], [13] where we have only investigated consistency in the sense of that error for estimators constructed using orthonormal systems of univariate analytic functions and i.i.d. observation errors.

In [8] Huang has obtained general conditions for convergence rates in probability of the error d_{nN} for series type estimators in the case of i.i.d. observation errors, assuming that $D \geq \varrho \geq c > 0$, and both the regression function f and the functions e_k , $k = 1, 2, \dots$, are bounded. In the present work it is shown that the boundedness conditions imposed on ϱ , f and e_k , $k = 1, 2, \dots$, can be relaxed.

In Section 3 we examine the asymptotic properties of the mean-square prediction error Ed_{nN} of an estimator obtained by a proper truncation of the estimator \widehat{f}_N . Consistency in the sense of that error and its convergence rates for series estimators in the case of fixed design observation models were investigated in [1], [6], [7], [12].

2. Convergence rates of the square prediction error. Let χ_n denote the indicator function of the event set $\{\lambda_n \geq 1/2\}$. If $\varrho \geq c > 0$, then according to (2) we have

$$P(\lambda_n < 1/2) \leq 4NM_N^2c^{-1}/n$$

and the assumption $NM_N^2/n \rightarrow 0$ as $n \rightarrow \infty$ implies $P(\chi_n \neq 1) \rightarrow 0$. Now, consider the mean-square prediction error defined as

$$R_{nN} = E_\omega E_\eta \chi_n \frac{1}{n} \sum_{i=1}^n (f(X_i) - \widehat{f}_N(X_i))^2.$$

We need the following lemmas to obtain an upper bound for the error R_{nN} of the series type estimators considered.

LEMMA 2.1. *If η_i , $i = 1, \dots, n$, are stationary second order errors with $E_\eta(\eta_1) = 0$, $E_\eta(\eta_1^2) = \sigma_\eta^2$ and with covariance function $E_\eta(\eta_1\eta_{1+j}) = Cj^{-\gamma}$, $j = 1, 2, \dots$, where $0 < \gamma < 1$ and $C \neq 0$, then*

$$\frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n |E_\eta(\eta_k\eta_l)| = O(n^{-\gamma}).$$

Proof. By the assumption $|E_\eta(\eta_1\eta_{1+j})| = |C|j^{-\gamma}$, we obtain

$$\begin{aligned} \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n |E_\eta(\eta_k\eta_l)| &= \frac{1}{n^2} \sum_{k=1}^n E_\eta(\eta_k^2) + \frac{2}{n^2} \sum_{k=1}^{n-1} \sum_{j=1}^{n-k} |E_\eta(\eta_k\eta_{k+j})| \\ &\leq \frac{\sigma_\eta^2}{n} + \frac{2|C|}{n^2} \sum_{k=1}^{n-1} \sum_{j=1}^{n-k} \frac{1}{j^\gamma} = \frac{\sigma_\eta^2}{n} + \frac{2|C|}{n^2} \sum_{j=1}^{n-1} \frac{n-j}{j^\gamma}, \end{aligned}$$

and since the function $g(s) = (n-s)/s^\gamma$ is strictly decreasing on the interval $(0, n]$ (its derivative is negative there) we also have

$$\begin{aligned} \sum_{j=1}^{n-1} \frac{n-j}{j^\gamma} &\leq \int_0^n \frac{n-s}{s^\gamma} ds = \int_0^n \frac{n}{s^\gamma} ds - \int_0^n s^{1-\gamma} ds \\ &= \frac{nn^{1-\gamma}}{1-\gamma} - \frac{n^{2-\gamma}}{2-\gamma} = \frac{n^{2-\gamma}}{(1-\gamma)(2-\gamma)}. \end{aligned}$$

In consequence,

$$\frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n |E_\eta(\eta_k\eta_l)| \leq \frac{\sigma^2}{n} + \frac{2|C|}{n^\gamma(1-\gamma)(2-\gamma)} = O(n^{-\gamma}). \blacksquare$$

LEMMA 2.2. *If $\varrho \geq c > 0$ and $d_N = \sum_{k=1}^N c_k e_k$, $c_1, \dots, c_N \in \mathbb{R}$, then*

$$R_{nN} \leq \int_A (f(x) - d_N(x))^2 \varrho(x) dx + \frac{2N}{n^2} \sum_{k=1}^n \sum_{l=1}^n |E_\eta(\eta_k\eta_l)|.$$

Proof. The standard squared bias plus variance decomposition with respect to the η variable yields

$$\begin{aligned} R_{nN} &= E_\omega \chi_n \frac{1}{n} \sum_{i=1}^n (f(X_i) - E_\eta \hat{f}_N(X_i))^2 \\ &\quad + E_\omega \chi_n \frac{1}{n} \sum_{i=1}^n E_\eta (\hat{f}_N(X_i) - E_\eta \hat{f}_N(X_i))^2. \end{aligned}$$

Taking into account (1) together with $\chi_n \lambda_n^{-1} \leq 2$ and $E\|h^N(X)\|^2 = N$, we easily obtain

$$\begin{aligned} &E_\omega \chi_n \frac{1}{n} \sum_{i=1}^n E_\eta (\hat{f}_N(X_i) - E_\eta \hat{f}_N(X_i))^2 \\ &= E_\omega \chi_n \frac{1}{n} \sum_{i=1}^n E_\eta \left\langle h^N(X_i), H_n^{-1} \frac{1}{n} \sum_{k=1}^n \eta_k h^N(X_k) \right\rangle^2 \\ &= E_\omega \chi_n \frac{1}{n^3} \sum_{i=1}^n \sum_{k=1}^n \sum_{l=1}^n E_\eta(\eta_k\eta_l) h^N(X_k)^T H_n^{-1} h^N(X_i) h^N(X_i)^T H_n^{-1} h^N(X_l) \end{aligned}$$

$$\begin{aligned}
 &= E_\omega \chi_n \frac{1}{n^3} \sum_{k=1}^n \sum_{l=1}^n \sum_{i=1}^n E_\eta(\eta_k \eta_l) h^N(X_k)^T H_n^{-1} h^N(X_i) h^N(X_i)^T H_n^{-1} h^N(X_l) \\
 &= E_\omega \chi_n \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n E_\eta(\eta_k \eta_l) h^N(X_k)^T H_n^{-1} h^N(X_l) \\
 &\leq E_\omega \chi_n \lambda_n^{-1} \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n |E_\eta(\eta_k \eta_l)| \cdot \|h^N(X_k)\| \cdot \|h^N(X_l)\| \\
 &\leq \frac{2}{n^2} \sum_{k=1}^n \sum_{l=1}^n |E_\eta(\eta_k \eta_l)| E_\omega \|h^N(X_k)\| \cdot \|h^N(X_l)\| \\
 &\leq \frac{2}{n^2} \sum_{k=1}^n \sum_{l=1}^n |E_\eta(\eta_k \eta_l)| (E_\omega \|h^N(X_k)\|^2)^{1/2} (E_\omega \|h^N(X_l)\|^2)^{1/2} \\
 &\leq \frac{2N}{n^2} \sum_{k=1}^n \sum_{l=1}^n |E_\eta(\eta_k \eta_l)|.
 \end{aligned}$$

Thus,

$$R_{nN} \leq E_\omega \chi_n \frac{1}{n} \sum_{i=1}^n (f(X_i) - E_\eta \hat{f}_N(X_i))^2 + \frac{2N}{n^2} \sum_{k=1}^n \sum_{l=1}^n |E_\eta(\eta_k \eta_l)|.$$

Now, since for fixed observation points $X_i, i = 1, \dots, n$, we have

$$\chi_n \frac{1}{n} \sum_{i=1}^n (f(X_i) - E_\eta \hat{f}_N(X_i))^2 \leq \frac{1}{n} \sum_{i=1}^n (f(X_i) - d_N(X_i))^2$$

for any linear combination $d_N = \sum_{k=1}^N c_k e_k, c_1, \dots, c_N \in \mathbb{R}$, we immediately obtain

$$\begin{aligned}
 R_{nN} &\leq \frac{1}{n} \sum_{i=1}^n E_\omega (f(X_i) - d_N(X_i))^2 + \frac{2N}{n^2} \sum_{k=1}^n \sum_{l=1}^n |E_\eta(\eta_k \eta_l)| \\
 &= \int_A (f(x) - d_N(x))^2 \varrho(x) dx + \frac{2N}{n^2} \sum_{k=1}^n \sum_{l=1}^n |E_\eta(\eta_k \eta_l)|. \blacksquare
 \end{aligned}$$

Now, we can prove the following theorems on convergence in probability of the square prediction error for the regression estimators considered.

THEOREM 2.1. *Assume that $\varrho \geq c > 0$, the sequence of natural numbers $N(n), n = 1, 2, \dots$, satisfies*

$$\lim_{n \rightarrow \infty} N(n) = \infty, \quad \lim_{n \rightarrow \infty} \frac{N(n)M_{N(n)}^2}{n} = 0,$$

and for $N=1, 2, \dots$, there exist $g_N \in \text{span}\{e_1, \dots, e_N\}$ such that $\|f - g_N\|_\infty = O(N^{-\alpha})$, where $\alpha > 0$. Then for the orthogonal series estimator $\widehat{f}_{N(n)}$, we have

$$\frac{1}{n} \sum_{i=1}^n (f(X_i) - \widehat{f}_{N(n)}(X_i))^2 = O_p(N(n)/n^\gamma + N(n)^{-2\alpha}).$$

Proof. Since the assumption $N(n)M_{N(n)}^2/n \rightarrow 0$ assures that $P(\chi_n \neq 1) \rightarrow 0$, the assertion of the theorem follows by applying Lemmas 2.1 and 2.2 with $d_N = g_N$. ■

In the case when the regression function is square integrable the following theorem holds.

THEOREM 2.2. *Assume that $D \geq \varrho \geq c > 0$, the sequence of natural numbers $N(n)$, $n = 1, 2, \dots$, satisfies*

$$\lim_{n \rightarrow \infty} N(n) = \infty, \quad \lim_{n \rightarrow \infty} \frac{N(n)M_{N(n)}^2}{n} = 0,$$

and f_N is the orthogonal projection of the regression function $f \in L^2(A)$ onto the subspace $\text{span}\{e_1, \dots, e_N\}$. Then for the orthogonal series estimator $\widehat{f}_{N(n)}$, we have

$$\frac{1}{n} \sum_{i=1}^n (f(X_i) - \widehat{f}_{N(n)}(X_i))^2 = O_p(N(n)/n^\gamma + \|f - f_{N(n)}\|^2).$$

Proof. Observe that $\int_A (f - f_N)^2 \varrho \leq D\|f - f_N\|^2$ and follow the proof of Theorem 2.1. In fact we can even prove that

$$E\chi_n \frac{1}{n} \sum_{i=1}^n (f(X_i) - \widehat{f}_{N(n)}(X_i))^2 = O(N(n)/n^\gamma + \|f - f_{N(n)}\|^2). \quad \blacksquare$$

For many orthonormal systems the following bound holds true:

$$(3) \quad M_N^2 = \|e^N\|_\infty^2 = \sup_{x \in A} \sum_{k=1}^N e_k^2(x) \leq KN,$$

where K is a constant [2]. It clearly holds for uniformly bounded systems (e.g. the trigonometric system in $L^2([0, 2\pi]^d)$) but also for strongly localized systems (e.g. splines, piecewise polynomials) and leveled localized systems (e.g. compactly supported wavelets in $L^2[0, 1]$) [2], [17]. For such systems the condition $N(n)M_{N(n)}^2/n \rightarrow 0$ is satisfied if $N(n)^2/n \rightarrow 0$ as $n \rightarrow \infty$. Then, if $\|f - f_N\| = O(N^{-\alpha})$, where $\alpha > 0$, Theorem 2.2 allows one to obtain convergence rates in probability of the square prediction error. Indeed, the two terms N/n^γ and $N^{-2\alpha}$ go to zero at the same rate when $N(n) \sim n^{\gamma/(1+2\alpha)}$ (i.e. $r_1 \geq N(n)n^{-\gamma/(1+2\alpha)} \geq r_2$, $r_1, r_2 > 0$), and then

the convergence rate of the square prediction error will be $n^{-2\alpha\gamma/(1+2\alpha)}$, provided $\alpha > \max\{\gamma - 1/2, 0\}$, which assures that $N(n)^2/n \rightarrow 0$.

It is easy to see that in the case of i.i.d. observation errors the best obtainable convergence rate $n^{-2\alpha/(1+2\alpha)}$ can be attained [16], provided $\alpha > 1/2$. Moreover, in that case the assumptions of Theorem 2.2 imply that for orthogonal systems satisfying (3) and $f \in L^2(A)$ we also have $n^{-1} \sum_{i=1}^n (f(X_i) - \widehat{f}_{N(n)}(X_i))^2 = o_p(1)$, i.e. the orthogonal series estimator considered is consistent in the sense of the square prediction error.

3. Asymptotic mean-square prediction error. For the estimator $\widetilde{f}_N = \chi_n \widehat{f}_N$, that is,

$$\widetilde{f}_N(x) = \begin{cases} \widehat{f}_N(x) & \text{if } \lambda_n \geq 1/2, \\ 0 & \text{otherwise,} \end{cases}$$

the following result concerning its mean-square prediction error holds true.

THEOREM 3.1. *Assume that $D \geq \varrho \geq c > 0$, the sequence of natural numbers $N(n)$, $n = 1, 2, \dots$, satisfies*

$$\lim_{n \rightarrow \infty} N(n) = \infty, \quad \lim_{n \rightarrow \infty} \frac{N(n)M_{N(n)}^2}{n} = 0,$$

and f_N is the orthogonal projection of the bounded regression function $f \in L^2(A)$ onto $\text{span}\{e_1, \dots, e_N\}$. Then for the estimator $\widetilde{f}_{N(n)}$, we have

$$E \frac{1}{n} \sum_{i=1}^n (f(X_i) - \widetilde{f}_{N(n)}(X_i))^2 = O\left(\frac{N(n)}{n^\gamma} + \|f - f_{N(n)}\|^2 + \frac{N(n)M_{N(n)}^2}{n}\right).$$

Proof. According to the definition of \widetilde{f}_N , we have

$$\begin{aligned} & E \frac{1}{n} \sum_{i=1}^n (f(X_i) - \widetilde{f}_{N(n)}(X_i))^2 \\ &= E \chi_n \frac{1}{n} \sum_{i=1}^n (f(X_i) - \widehat{f}_{N(n)}(X_i))^2 + E(1 - \chi_n) \frac{1}{n} \sum_{i=1}^n f^2(X_i) \\ &\leq E \chi_n \frac{1}{n} \sum_{i=1}^n (f(X_i) - \widehat{f}_{N(n)}(X_i))^2 + P(\lambda_n < 1/2) \|f\|_\infty^2. \end{aligned}$$

As remarked in the proof of Theorem 2.2 the first term on the right-hand side of the above inequality is $O(N(n)/n^\gamma + \|f - f_{N(n)}\|^2)$ and the second is bounded by $4\|f\|_\infty^2 N(n)M_{N(n)}^2 c^{-1}/n$, which completes the proof. ■

We can also prove the following corollary on the convergence rate of the mean-square prediction error of the estimator \widetilde{f}_N .

COROLLARY 3.1. *Assume that $D \geq \varrho \geq c > 0$, the orthogonal system e_k , $k = 1, 2, \dots$, satisfies (3) and $\|f - f_N\| = O(N^{-\alpha})$, where $\alpha > 0$ and f_N*

is the orthogonal projection of the bounded regression function $f \in L^2(A)$ onto $\text{span}\{e_1, \dots, e_N\}$. Then

(a) for $0 < \gamma \leq (2\alpha + 1)/(2\alpha + 2)$ and $N(n) \sim n^{\gamma/(2\alpha+1)}$,

$$E \frac{1}{n} \sum_{i=1}^n (f(X_i) - \tilde{f}_{N(n)}(X_i))^2 = O(n^{-2\alpha\gamma/(2\alpha+1)}),$$

(b) for $(2\alpha + 1)/(2\alpha + 2) < \gamma < 1$ and $N(n) \sim n^{1/(2\alpha+2)}$,

$$E \frac{1}{n} \sum_{i=1}^n (f(X_i) - \tilde{f}_{N(n)}(X_i))^2 = O(n^{-\alpha/(\alpha+1)}).$$

Proof. Theorem 3.1 and inequality (3) imply

$$E \frac{1}{n} \sum_{i=1}^n (f(X_i) - \tilde{f}_{N(n)}(X_i))^2 = O(N(n)/n^\gamma + N(n)^2/n + N(n)^{-2\alpha}).$$

Now, for $0 < \gamma \leq (2\alpha + 1)/(2\alpha + 2)$ and $N(n) \sim n^{\gamma/(2\alpha+1)}$ we have

$$\begin{aligned} N(n)/n^\gamma + N(n)^{-2\alpha} &= O(n^{-2\gamma\alpha/(2\alpha+1)}), \\ N(n)^2/n &= O(n^{-1+2\gamma/(2\alpha+1)}), \end{aligned}$$

and $2\gamma\alpha/(2\alpha + 1) \leq 1 - 2\gamma/(2\alpha + 1)$.

Similarly, for $(2\alpha + 1)/(2\alpha + 2) < \gamma < 1$ and $N(n) \sim n^{1/(2\alpha+2)}$ we have

$$N(n)^2/n + N(n)^{-2\alpha} = O(n^{-2\alpha/(2\alpha+2)}), \quad N(n)/n^\gamma = O(n^{-\gamma+1/(2\alpha+2)}),$$

and $2\alpha/(2\alpha + 2) \leq \gamma - 1/(2\alpha + 2)$. ■

As one can easily see, in the case of i.i.d. observation errors the convergence rate of the mean-square prediction error of the estimator \tilde{f}_N is $n^{-\alpha/(\alpha+1)}$, i.e. the same as in the case of long-range dependent errors with $(2\alpha + 1)/(2\alpha + 2) < \gamma < 1$.

Since λ_n is the minimal eigenvalue of the matrix H_n which is not used in computations we cannot verify directly whether the condition $\lambda_n \geq 1/2$ is satisfied. However, under the assumptions of Theorem 3.1, $P(\hat{f}_N \neq \tilde{f}_N) \leq P(\lambda_n < 1/2) \rightarrow 0$ as $n \rightarrow \infty$, so the estimators \hat{f}_N and \tilde{f}_N are identical with growing probability.

4. Conclusions. As proved in [11], in the case when we use multivariate orthonormal systems constructed as tensor products of univariate analytic functions forming an orthonormal system in $L^2([a, b])$ (e.g. multivariate trigonometric functions or multivariate polynomials), the normal equations matrix G_n is almost surely positive definite for any density ϱ and $N \leq n$. Thus, in that case the estimators considered are uniquely defined with probability one. Furthermore, in the case of i.i.d. observation errors

the following inequality is valid for an arbitrary density ϱ and $N \leq n$ (see Theorem 3.1 in [13] for the proof):

$$E_\omega E_\eta \frac{1}{n} \sum_{i=1}^n (f(X_i) - \widehat{f}_N(X_i))^2 \leq \int_{[a,b]^d} (f(x) - d_N(x))^2 \varrho(x) dx + \sigma_\eta^2 \frac{N}{n}.$$

Thus, if the error of uniform approximation of f by linear combinations of the first N basis functions is of order $N^{-\alpha}$, where $\alpha > 0$, or its mean-square approximation error is of this order and the density ϱ is bounded, one can show that the convergence rate of the mean-square prediction error is $n^{-2\alpha/(2\alpha+1)}$, i.e. the optimal convergence rate for that error is attained [15], [16].

The exponent α defining the decrease rate of the uniform approximation error of the regression function is related not only to the smoothness of the regression function but also to the dimension of X . For example, if f is s times continuously differentiable on $[-1, 1]^d$, then in the case of polynomial approximation we have $\alpha = s/d$ according to Lorentz [10].

Convergence rates of uniform and mean-square approximation of univariate functions from a Sobolev space on $[-\pi, \pi]$ by trigonometric functions or from a Sobolev space on $[-1, 1]$ by Legendre polynomials are derived in [4], [6].

Let $\text{Lip}^m(A)$ for $m > 0$ denote the space of Hölder-continuous functions of order m (i.e. $[m]$ times continuously differentiable and such that the $[m]$ th order derivatives satisfy the Lipschitz condition with exponent $m - [m]$, where $[m]$ denotes the integer part of m). For multivariate functions belonging to this class, convergence rates of uniform and mean-square approximation by appropriate orthonormal wavelet bases can be derived using the results of Jaffard and Meyer [9], [4].

It was shown by the author [14] that results concerning consistency and convergence rates in probability of the errors $E_X(f(X) - \widehat{f}_N(X))^2$ and $\|f - \widehat{f}_N\|_\infty$ for orthogonal series regression estimators can be obtained in the case of weakly dependent observation errors. Applying Lemma 2.1 and the techniques of [14] we can also obtain such results for zero mean stationary second order observation errors with long-range dependence. Another approach to estimating regression from dependent observations by the least squares method, which enables obtaining similar results, is presented in [3].

Efromovich [5] has recently shown that there exist trigonometric series estimators for the random design observation model considered that are robust and minimax, and discussed the results obtained both theoretically and via Monte Carlo simulations, indicating also possible extensions.

References

- [1] A. Antoniadis, G. Grégoire and I. W. McKeague, *Wavelet methods for curve estimation*, J. Amer. Statist. Assoc. 89 (1994), 1340–1353.
- [2] L. Birgé and P. Massart, *Minimum contrast estimators on sieves: exponential bounds and rates of convergence*, J. Bernoulli Soc. 4 (1998), 329–375.
- [3] P. Burman, *Regression function estimation from dependent observations*, J. Multivariate Anal. 36 (1991), 263–279.
- [4] M. Delecroix and C. Protopopescu, *Regression series estimators: the MISE approach*, J. Nonparametric Statist. (2000), forthcoming.
- [5] S. Efromovich, *How to overcome the curse of long-memory errors*, IEEE Trans. Inform. Theory IT-45 (1999), 1735–1741.
- [6] R. L. Eubank and B. R. Jayasuryia, *The asymptotic average square error for polynomial regression*, Statistics 24 (1993), 311–319.
- [7] R. L. Eubank and P. Speckman, *Convergence rates for trigonometric and polynomial-trigonometric regression estimators*, Statist. Probab. Lett. 11 (1991), 119–124.
- [8] J. Z. Huang, *Projection estimation in multiple regression with application to functional ANOVA models*, Ann. Statist. 26 (1998), 242–272.
- [9] S. Jaffard et Y. Meyer, *Bases d'ondelettes dans des ouverts de R^n* , J. Math. Pures Appl. 68 (1989), 95–108.
- [10] G. G. Lorentz, *Approximation of Functions*, Holt, Rinehart & Winston, New York, 1966.
- [11] W. Popiński, *On least squares estimation of Fourier coefficients and of the regression function*, Appl. Math. (Warsaw) 22 (1993), 91–102.
- [12] —, *Least-squares trigonometric regression estimation*, *ibid.* 26 (1999), 121–131.
- [13] —, *A note on orthogonal series regression function estimators*, *ibid.* 26 (1999), 281–291.
- [14] —, *Convergence rates of orthogonal series regression estimators*, *ibid.* 27 (2000), 445–454.
- [15] P. Speckman, *Spline smoothing and optimal convergence rates in nonparametric regression models*, Ann. Statist. 13 (1985), 970–983.
- [16] C. J. Stone, *Optimal global rates of convergence for nonparametric regression*, *ibid.* 10 (1982), 1040–1053.
- [17] G. Viennet, *Least-square estimation for regression on random design for absolutely regular observations*, Statist. Probab. Lett. 43 (1999), 13–23.

Department of Survey Design
 Central Statistical Office
 Al. Niepodległości 208
 00-925 Warszawa, Poland
 E-mail: w.popinski@stat.gov.pl

Received on 12.4.2001;
 revised version on 5.11.2001

(1574)