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 \( Y_i = f(X_i) + \eta_i, \ i = 1, \ldots, n \), where \( X_i \in A \subset \mathbb{R}^d \) are independently chosen from a distribution with density \( \varrho \in L^1(A) \) and \( \eta_i \) 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 \( e_k, k = 1, 2, \ldots \), in \( L^2(A) \), are obtained.

1. Introduction. Consider a random design observation model \( Y_i = f(X_i) + \eta_i, \ i = 1, \ldots, n \), where the points \( X_i, \ i = 1, \ldots, n \), 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 \( \eta_i, \ i = 1, \ldots, n \), are stationary second order errors with zero mean, variance \( \sigma_{\eta}^2 \), and covariance function \( E_\eta(\eta_i \eta_{1+j}) = C j^{-\gamma}, \ j = 1, 2, \ldots \), where \( 0 < \gamma < 1 \) and \( C \) is a real constant. We assume that the observation errors \( \eta = (\eta_1, \ldots, \eta_n) \) are independent of the predictors \( \omega = (X_1, \ldots, X_n) \).

Let us consider a series type regression function estimator

\[
\hat{f}_N(x) = \sum_{k=1}^{N} \hat{c}_k e_k(x),
\]

where the functions \( e_k, k = 1, 2, \ldots \), form a complete orthonormal system in \( L^2(A) \). We assume that the vector \( \hat{c}_N = (\hat{c}_{1N}, \ldots, \hat{c}_{NN})^T \) of coefficient
estimators is, for fixed $N$, obtained by the least squares method, i.e.
\[ \hat{c}_N = \arg \min_{a \in \mathbb{R}^N} \frac{1}{n} \sum_{i=1}^{n} (Y_i - \langle a, e^N(X_i) \rangle)^2, \]
where $e^N(x) = (e_1(x), \ldots, e_N(x))^T$.

The vector $\hat{c}_N$ can be obtained as a solution of the normal equations
\[ G_n \hat{c}_N = g_n, \]
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 $\hat{f}_N(x) = \langle h^N(x), \hat{b}_N \rangle$, where $h^N(x) = G(N)^{-1/2}e^N(x)$, and the vector $\hat{b}_N = G(N)^{1/2}\hat{c}_N$ satisfies the normal equations
\[ (1) \quad H_n \hat{b}_N = h_n, \]
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 $\hat{f}_N$ using the vector function $h^N$ is convenient, since then $\mathbb{E}h^N(X)h^N(X)^T = I_N$, where $I_N$ is the unit matrix, and consequently $\mathbb{E}\|h^N(X)\|^2 = N$. Moreover, if we put
\[ M_N = \|e^N\|_{\infty} = \sup_{x \in A} \|e^N(x)\| \]
and $\lambda_n$ 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 $P(\det G_n = 0) = P(\det H_n = 0) \leq P(\lambda_n < 1/2) \leq 4NM_N^2 e^{-1}/n$. Thus, the conditions $\varrho \geq c > 0$ and $NM_N^2/n \to 0$ as $n \to \infty$ assure that $P(\det G_n = 0) \to 0$, 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
\[ d_{nN} = \frac{1}{n} \sum_{i=1}^{n} \left( f(X_i) - \hat{f}_N(X_i) \right)^2 \]
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 \( e_k, \ k = 1, 2, \ldots, \) are bounded. In the present work it is shown that the boundedness conditions imposed on \( \varrho, f \) and \( e_k, \ k = 1, 2, \ldots, \) 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 \( \hat{f}_N \). 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 \( \chi_n \) denote the indicator function of the event set \( \{ \lambda_n \geq 1/2 \} \). If \( \varrho \geq c > 0 \), then according to (2) we have
\[ P(\lambda_n < 1/2) \leq 4NM_{N}^{2}c^{-1}/n \]
and the assumption \( NM_{N}^{2}/n \to 0 \) as \( n \to \infty \) implies \( P(\chi_n \neq 1) \to 0 \). Now, consider the mean-square prediction error defined as
\[ R_{nN} = E\omega E\eta \chi_n \frac{1}{n} \sum_{i=1}^{n} \left( f(X_i) - \hat{f}_N(X_i) \right)^2. \]

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 \( \eta_i, \ i = 1, \ldots, n, \) are stationary second order errors with \( E\eta(\eta_1) = 0, \ E\eta(\eta_1^2) = \sigma_n^2 \) and with covariance function \( E\eta(\eta_1 \eta_{1+j}) = Cj^{-\gamma}, \ j = 1, 2, \ldots, \) 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

\[
\frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} |E_\eta(\eta_k\eta_l)| = \frac{1}{n^2} \sum_{k=1}^{n} E_\eta(\eta_k^2) + \frac{2}{n^2} \sum_{k=1}^{n-1} \sum_{j=1}^{n-k} |E_\eta(\eta_k\eta_{k+j})|
\]

\[
\leq \frac{\sigma^2}{n} + \frac{2|C|}{n^2} \sum_{k=1}^{n-1} \sum_{j=1}^{n-k} \frac{1}{j^\gamma} = \frac{\sigma^2}{n} + \frac{2|C|}{n^2} \sum_{j=1}^{n-1} \frac{n-j}{j^\gamma},
\]

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

\[
\sum_{j=1}^{n-1} \frac{n-j}{j^\gamma} \leq \int_0^n \frac{n-s}{s^\gamma} ds = \int_0^n \frac{n}{s^\gamma} ds - \int_0^n s^{1-\gamma} ds
\]

\[
= \frac{nn^{1-\gamma}}{1-\gamma} - \frac{n^{2-\gamma}}{2-\gamma} = \frac{n^{2-\gamma}}{(1-\gamma)(2-\gamma)}.
\]

In consequence,

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

Lemma 2.2. If \(\varrho \geq c > 0\) and \(d_N = \sum_{k=1}^{N} c_k e_k, c_1, \ldots, c_N \in \mathbb{R}\), then

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

Proof. The standard squared bias plus variance decomposition with respect to the \(\eta\) variable yields

\[
R_{nN} = E_\omega \chi_n \frac{1}{n} \sum_{i=1}^{n} (f(X_i) - E_\eta \hat{f}_N(X_i))^2
\]

\[
+ E_\omega \chi_n \frac{1}{n} \sum_{i=1}^{n} E_\eta(\hat{f}_N(X_i)) - E_\eta \hat{f}_N(X_i))^2.
\]

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

\[
E_\omega \chi_n \frac{1}{n} \sum_{i=1}^{n} E_\eta(\hat{f}_N(X_i)) - E_\eta \hat{f}_N(X_i))^2
\]

\[
= E_\omega \chi_n \frac{1}{n} \sum_{i=1}^{n} E_\eta \left\langle h^N(X_i), H_n^{-1} \frac{1}{n} \sum_{k=1}^{n} \eta_k h^N(X_k) \right\rangle ^2
\]

\[
= E_\omega \chi_n \frac{1}{n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} E_\eta(\eta_k\eta_l) h^N(X_k)^T H_n^{-1} h^N(X_i) h^N(X_i)^T H_n^{-1} h^N(X_l)
\]
= E_\omega \chi_n 1/n^2 \sum_{k=1}^{n} \sum_{l=1}^{n} \sum_{i=1}^{n} \eta_i h^n(X_k) h^n(X_l) h^{-1}_n h^n(X_i)

= E_\omega \chi_n 1/n^2 \sum_{k=1}^{n} \sum_{l=1}^{n} \eta_i h^n(X_k) h^{-1}_n h^n(X_l)

\leq E_\omega \chi_n \lambda_n^{-1} 1/n^2 \sum_{k=1}^{n} \sum_{l=1}^{n} |\eta_i| \cdot \|h^n(X_k)\| \cdot \|h^n(X_l)\|

\leq \frac{2}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} |\eta_i| (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} |\eta_i|.

Thus,

R_{nN} \leq E_\omega \chi_n 1/n \sum_{i=1}^{n} (f(X_i) - \widehat{f}_N(X_i))^2 + \frac{2N}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} |\eta_i|.

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

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

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

$R_{nN} \leq \frac{1}{n} \sum_{i=1}^{n} E_\omega (f(X_i) - d_N(X_i))^2 + \frac{2N}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} |\eta_i|$

\[= \int_A (f(x) - d_N(x))^2 \, dx + \frac{2N}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} |\eta_i| \quad \blacksquare \]

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 $\rho \geq c > 0$, the sequence of natural numbers $N(n)$, $n = 1, 2, \ldots$, satisfies

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

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

**Proof.** Since the assumption $N(n)M^2_{N(n)}/n \to 0$ assures that $P(\chi_n \neq 1) \to 0$, 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 $D \geq p \geq c > 0$, the sequence of natural numbers $N(n)$, $n = 1, 2, \ldots$, satisfies

$$\lim_{n \to \infty} N(n) = \infty, \quad \lim_{n \to \infty} \frac{N(n)M^2_{N(n)}}{n} = 0,$$

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

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

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

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

For many orthonormal systems the following bound holds true:

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

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 $N(n)M^2_{N(n)}/n \to 0$ is satisfied if $N(n)^2/n \to 0$ as $n \to \infty$. 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 $N/n^\gamma$ 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 $N(n)^2/n \to 0$.

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 $n^{-1}\sum_{i=1}^{n} (f(X_i) - \hat{f}_{N(n)}(X_i))^2 = o_p(1)$, 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 $\tilde{f}_N = \chi_n \hat{f}_N$, that is,

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

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 $N(n)$, $n = 1, 2, \ldots$, satisfies

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

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

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

**Proof.** According to the definition of $\tilde{f}_N$, we have

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

$$= E\chi_n \frac{1}{n} \sum_{i=1}^{n} (f(X_i) - \hat{f}_{N(n)}(X_i))^2 + E(1 - \chi_n) \frac{1}{n} \sum_{i=1}^{n} f^2(X_i)$$

$$\leq E\chi_n \frac{1}{n} \sum_{i=1}^{n} (f(X_i) - \hat{f}_{N(n)}(X_i))^2 + P(\lambda_n < 1/2) \|f\|^2.$$  

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

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

**Corollary 3.1.** Assume that $D \geq \varrho \geq c > 0$, the orthogonal system $e_k$, $k = 1, 2, \ldots$, 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 $\text{span}\{e_1, \ldots, e_N\}$. 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

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

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

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

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

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

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 $\lambda_n$ 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, $P(\tilde{f}_N \neq \hat{f}_N) \leq P(\lambda_n < 1/2) \to 0$ as $n \to \infty$, 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 $\varrho$ 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 \( \varrho \) and \( N \leq n \) (see Theorem 3.1 in [13] for the proof):

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

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 \( \varrho \) 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 \( \alpha \) 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


[8] J. Z. Huang, *Projection estimation in multiple regression with application to func-


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)