

W. POPIŃSKI (Warszawa)

ORTHOGONAL SERIES REGRESSION ESTIMATORS FOR AN IRREGULARLY SPACED DESIGN

Abstract. Nonparametric orthogonal series regression function estimation is investigated in the case of a fixed point design where the observation points are irregularly spaced in a finite interval $[a, b]$, $a, b \in \mathbb{R}$. Convergence rates for the integrated mean-square error and pointwise mean-square error are obtained in the case of estimators constructed using the Legendre polynomials and Haar functions for regression functions satisfying the Lipschitz condition.

1. Introduction. Consider the partition of the interval $A = [a, b] \subset \mathbb{R}$ into n subintervals $A_1 = [a_0, a_1]$, $A_i = (a_{i-1}, a_i]$, $i = 2, \dots, n$, where $a = a_0 < a_1 < \dots < a_n = b$. Suppose that the observations y_1, \dots, y_n follow the model $y_i = f(x_i) + \eta_i$, where $f : [a, b] \rightarrow \mathbb{R}$ is an unknown function satisfying certain smoothness conditions specified below, $x_i \in A_i$, $i = 1, \dots, n$, and η_i , $i = 1, \dots, n$, are independent identically distributed random variables with mean zero and finite variance $\sigma_\eta^2 > 0$. For simplicity the dependence of x_i and a_i on n is not indicated in the notation.

Let the functions e_k , $k = 0, 1, \dots$, form a complete orthonormal system in the space $L^2[a, b]$. We assume that the regression function f is an element of this space and consequently it has a representation

$$f = \sum_{k=0}^{\infty} c_k e_k, \quad \text{where} \quad c_k = \int_a^b f(s) e_k(s) ds, \quad k = 0, 1, \dots$$

As an estimator of f we take

$$(1) \quad \hat{f}_N(x) = \sum_{k=0}^N \hat{c}_k e_k(x), \quad \text{where} \quad \hat{c}_k = \sum_{i=1}^n y_i \int_{a_{i-1}}^{a_i} e_k(s) ds.$$

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

Key words and phrases: nonparametric regression, orthogonal series estimator, convergence rates.

This formula can be rewritten in the form

$$\widehat{f}_N(x) = \sum_{i=1}^n y_i \int_{a_{i-1}}^{a_i} K(x, s) ds, \quad \text{where} \quad K(x, y) = \sum_{k=0}^N e_k(x)e_k(y),$$

so it is easy to see that our estimator resembles the Gasser–Müller kernel regression function estimator [4].

Sufficient conditions for almost sure convergence of the above estimator, constructed using the trigonometric functions and Legendre polynomials, were investigated by Rutkowski [9]. Results concerning convergence rates of the integrated mean-square error of the estimator constructed using the trigonometric functions were obtained in [3]. Orthogonal series estimators which attain an optimal IMSE convergence rate were constructed by Rafajłowicz [7].

However, in [3] it is assumed that the observation points satisfy

$$(2) \quad \int_a^{x_i} p(s) ds = (i-1)/n, \quad i = 1, \dots, n,$$

where $p \in L^1[a, b]$ is a density satisfying certain smoothness conditions, and in [7] the observation points coincide with knots of numerical quadratures.

The aim of the present work is to obtain asymptotic results, in particular convergence rates for IMSE and pointwise mean-square error of the estimator (1), for fixed point designs different from the ones considered so far, under milder restrictions on the observation points. This seems to be important since it may be numerically difficult to determine exactly the points x_i satisfying (2) or the knots of appropriate numerical quadratures, especially when their number is large.

The present work also shows that there exist series type regression function estimators with asymptotic properties comparable to or even better than those of the least squares polynomial estimators investigated in [8], [11], which were examined only for special fixed point designs, e.g. when the observation points are equidistant.

Observe that calculating the integrals occurring in the definition of the Fourier coefficient estimators in (1) is straightforward in the case of trigonometric functions and Haar functions but may be difficult for other orthonormal systems. However, orthogonal polynomials often satisfy recurrence formulae [10] which facilitate calculation of such integrals, e.g. for Legendre polynomials forming an orthonormal system in $L^2[-1, 1]$ we can apply the recurrence formula [10]

$$xe'_k(x) = ke_k(x) + e'_{k-1}(x), \quad \text{where} \quad e'_k(x) = \frac{d}{dx}e_k(x), \quad k = 1, 2, \dots$$

Indeed, for $-1 \leq c < d \leq 1$ we have

$$\int_c^d s e'_k(s) ds = k \int_c^d e_k(s) ds + e_{k-1}(d) - e_{k-1}(c),$$

and integrating by parts gives

$$d e_k(d) - c e_k(c) - \int_c^d e_k(s) ds = k \int_c^d e_k(s) ds + e_{k-1}(d) - e_{k-1}(c),$$

so that finally

$$(k+1) \int_c^d e_k(s) ds = d e_k(d) - c e_k(c) - e_{k-1}(d) + e_{k-1}(c).$$

Thus, to calculate the integrals occurring in (1) in the case of Legendre polynomials one only needs to calculate the point values of such polynomials, which can be accomplished using well known formulae [10].

Recall that according to well known results on uniform polynomial approximation (e.g. Theorem 3.11 of [5]) for any function $f \in C[-1, 1]$ there exists a polynomial q_N of degree $N > 0$ such that

$$(3) \quad \sup_{-1 \leq t \leq 1} |f(t) - q_N(t)| \leq 6\omega(f, 1/N),$$

where $\omega(f, \delta) = \sup_{|t-s| \leq \delta} |f(t) - f(s)|$ for $\delta > 0$. Since for Legendre polynomials we have $|e_k(t)| \leq \sqrt{(2k+1)/2}$, $k = 0, 1, \dots$, for $t \in [-1, 1]$ (see [10]), we easily obtain

$$(4) \quad \sum_{k=0}^N e_k^2(t) \leq \sum_{k=0}^N (2k+1)/2 = (N+1)^2/2 \quad \text{for } t \in [-1, 1].$$

Instead of Legendre polynomials also the Haar functions [6], forming a complete orthonormal system in $L^2[0, 1]$, can be used. Assume that these functions are ordered as follows [6]: $\chi_1(t) \equiv 1$, and for $n = 2^m + k$, $k = 1, \dots, 2^m$, $m = 0, 1, \dots$,

$$\chi_n(t) = \chi_m^k(t) = \begin{cases} \sqrt{2^m} & \text{for } t \in ((2k-2)/2^{m+1}, (2k-1)/2^{m+1}), \\ -\sqrt{2^m} & \text{for } t \in ((2k-1)/2^{m+1}, 2k/2^{m+1}), \\ 0 & \text{otherwise.} \end{cases}$$

If $f \in C[0, 1]$ and

$$f_N(t) = \sum_{j=1}^N c_j \chi_j(t), \quad \text{where } c_j = \int_0^1 f(s) \chi_j(s) ds, \quad j = 1, 2, \dots,$$

denotes the N th partial sum of the Haar series of f , then by [6], we have

$$(5) \quad |f(t) - f_N(t)| \leq 2\omega(f, 1/N)$$

almost everywhere for $t \in [0, 1]$. Moreover, for $t \in [0, 1]$ and $N = 2^m + k$, $k = 1, \dots, 2^m$, $m = 0, 1, \dots$, we have

$$(6) \quad \sum_{j=1}^N \chi_j^2(t) \leq 1 + 2^0 + \dots + 2^m = 2^{m+1} \leq 2(N-1).$$

2. Integrated mean-square error. To investigate the asymptotic properties of our estimator we need the following lemma giving bounds for the variance and squared bias of the Fourier coefficient estimators defined in (1).

LEMMA 2.1. *If $f \in L^2[a, b]$, then the Fourier coefficient estimators \widehat{c}_k , $k = 0, 1, \dots$, given by (1), satisfy*

$$E(\widehat{c}_k - E\widehat{c}_k)^2 \leq \sigma_\eta^2 D_n, \quad |c_k - E\widehat{c}_k|^2 \leq \omega^2(f, D_n)(b-a),$$

where $D_n = \max_{1 \leq i \leq n} (a_i - a_{i-1})$.

PROOF. As $E\widehat{c}_k = \sum_{i=1}^n f(x_i) \int_{A_i} e_k(s) ds$ we immediately see that

$$\begin{aligned} E(\widehat{c}_k - E\widehat{c}_k)^2 &= E\left(\sum_{i=1}^n \eta_i \int_{A_i} e_k(s) ds\right)^2 = \sigma_\eta^2 \sum_{i=1}^n \left(\int_{A_i} e_k(s) ds\right)^2 \\ &\leq \sigma_\eta^2 \sum_{i=1}^n (a_i - a_{i-1}) \int_{A_i} e_k^2(s) ds \leq \sigma_\eta^2 D_n \sum_{i=1}^n \int_{A_i} e_k^2(s) ds \\ &= \sigma_\eta^2 D_n \int_A e_k^2(s) ds = \sigma_\eta^2 D_n, \end{aligned}$$

since $A = \bigcup_{i=1}^n A_i$ and the functions $e_k, k = 0, 1, \dots$, form an orthonormal system in $L^2[a, b]$. For the bias term we easily obtain

$$\begin{aligned} |c_k - E\widehat{c}_k|^2 &= \left[\sum_{i=1}^n \int_{A_i} f(s) e_k(s) ds - \sum_{i=1}^n f(x_i) \int_{A_i} e_k(s) ds \right]^2 \\ &\leq \left[\sum_{i=1}^n \int_{A_i} |f(s) - f(x_i)| \cdot |e_k(s)| ds \right]^2 \\ &\leq \left[\sum_{i=1}^n \left(\int_{A_i} (f(s) - f(x_i))^2 ds \right)^{1/2} \left(\int_{A_i} e_k^2(s) ds \right)^{1/2} \right]^2 \\ &\leq \sum_{i=1}^n \int_{A_i} (f(s) - f(x_i))^2 ds \cdot \sum_{i=1}^n \int_{A_i} e_k^2(s) ds, \end{aligned}$$

and since $\sum_{i=1}^n \int_{A_i} e_k^2(s) ds = 1$ we finally have

$$|c_k - E\widehat{c}_k|^2 \leq \omega^2(f, D_n) \sum_{i=1}^n (a_i - a_{i-1}) = \omega^2(f, D_n)(b - a). \blacksquare$$

It is easy to show the following formula for the IMSE of \widehat{f}_N :

$$R_N = E\|f - \widehat{f}_N\|^2 = \sum_{k=0}^N E(c_k - \widehat{c}_k)^2 + \|f - f_N\|^2,$$

where $f_N = \sum_{k=0}^N c_k e_k$. Lemma 2.1 yields

$$(7) \quad R_N = \sum_{k=0}^N [(c_k - E\widehat{c}_k)^2 + E(E\widehat{c}_k - \widehat{c}_k)^2] + \|f - f_N\|^2 \\ \leq (N + 1)[\sigma_\eta^2 D_n + \omega^2(f, D_n)(b - a)] + \|f - f_N\|^2.$$

Now consider the case where $a = -1, b = 1$ or $a = 0, b = 1$ and the estimator \widehat{f}_N is constructed using the complete orthonormal system of Legendre polynomials in $L^2[-1, 1]$ or Haar functions in $L^2[0, 1]$, respectively.

LEMMA 2.2. *If $f \in C[-1, 1]$ (resp. $f \in C[0, 1]$) satisfies the Lipschitz condition with exponent $0 < \alpha \leq 1$ and $D_n = O(n^{-1})$, then the estimator \widehat{f}_N constructed using Legendre polynomials (resp. Haar functions) satisfies*

$$E\|f - \widehat{f}_N\|^2 \leq \frac{NB_1\sigma_\eta^2}{n} + \frac{NB_2}{n^{2\alpha}} + \frac{B_3}{N^{2\alpha}},$$

where $B_1, B_2, B_3 > 0$.

PROOF. We first give the proof for the polynomial estimator. By assumption, there exist constants $C, L > 0$ such that $D_n \leq Cn^{-1}$ and $\omega(f, \delta) \leq L\delta^\alpha$ for $\delta > 0$. Consequently, (7) shows that

$$R_N \leq (N + 1) \left(\frac{C\sigma_\eta^2}{n} + \frac{2L^2C^{2\alpha}}{n^{2\alpha}} \right) + \|f - f_N\|^2.$$

Since $f_N = \sum_{k=0}^N c_k e_k$, where c_0, c_1, \dots, c_N are the Fourier coefficients of f , we have $\|f - f_N\|^2 \leq \|f - q_N\|^2$ for any polynomial q_N of degree not greater than N and so $\|f - f_N\|^2 \leq 72L^2/N^{2\alpha}$ by (3). Thus, we finally obtain the desired bound

$$R_N \leq N \left(\frac{B_1\sigma_\eta^2}{n} + \frac{B_2}{n^{2\alpha}} \right) + \frac{B_3}{N^{2\alpha}}.$$

The proof for the Haar series estimator is analogous with the use of (5). \blacksquare

In view of Lemma 2.2 one can easily see that the polynomial estimator $\widehat{f}_{N(n)}$ will be consistent in the sense of IMSE if the sequence of natural numbers $N(n)$ satisfies $N(n) \rightarrow \infty$, $N(n)/n \rightarrow 0$ for $1/2 < \alpha \leq 1$, or

$N(n) \rightarrow \infty$, $N(n)/n^{2\alpha} \rightarrow 0$ for $0 < \alpha \leq 1/2$. The same is true for the Haar series estimator.

Assume now that $N(n) \sim n^\beta$, $\beta > 0$ (i.e. $r_1 \leq n^{-\beta}N(n) \leq r_2$ for $r_1, r_2 > 0$). Lemma 2.2 easily yields IMSE convergence rates for the polynomial or Haar series estimator $\widehat{f}_{N(n)}$.

THEOREM 2.1. *If the assumptions of Lemma 2.2 hold, then the estimator \widehat{f}_N constructed using Legendre polynomials (or Haar functions) satisfies*

- (a) $E\|f - \widehat{f}_{N(n)}\|^2 = O(n^{-4\alpha^2/(2\alpha+1)})$ for $0 < \alpha \leq 1/2$ and $N(n) \sim n^{2\alpha/(2\alpha+1)}$,
- (b) $E\|f - \widehat{f}_{N(n)}\|^2 = O(n^{-2\alpha/(2\alpha+1)})$ for $1/2 < \alpha \leq 1$ and $N(n) \sim n^{1/(2\alpha+1)}$.

3. Pointwise mean-square error. In this section we examine the pointwise mean-square error of the estimator \widehat{f}_N and obtain convergence rates for the estimators constructed using Legendre polynomials and Haar functions. If the Fourier series of f converges to $f(x)$ at some $x \in [a, b]$, then

$$E(f(x) - \widehat{f}_N(x))^2 = E\left(\sum_{k=0}^N (c_k - \widehat{c}_k)e_k(x)\right)^2 + r_N^2(x) + 2r_N(x) \sum_{k=0}^N (c_k - E\widehat{c}_k)e_k(x),$$

where $r_N(x) = f(x) - \sum_{k=0}^N c_k e_k(x)$. From the Cauchy-Schwarz inequality it further follows that

$$E(f(x) - \widehat{f}_N(x))^2 \leq \sum_{k=0}^N [(c_k - E\widehat{c}_k)^2 + E(\widehat{c}_k - E\widehat{c}_k)^2] \sum_{k=0}^N e_k^2(x) + 2|r_N(x)| \left(\sum_{k=0}^N (c_k - E\widehat{c}_k)^2\right)^{1/2} \left(\sum_{k=0}^N e_k^2(x)\right)^{1/2} + r_N^2(x),$$

and in view of Lemma 2.1 we finally obtain the inequality

$$(8) \quad E(f(x) - \widehat{f}_N(x))^2 \leq (N + 1)[\sigma_\eta^2 D_n + \omega^2(f, D_n)(b - a)] \sum_{k=0}^N e_k^2(x) + r_N^2(x) + 2|r_N(x)|(N + 1)^{1/2}\omega(f, D_n)(b - a)^{1/2} \left(\sum_{k=0}^N e_k^2(x)\right)^{1/2}.$$

For $f \in C[-1, 1]$ satisfying the Lipschitz condition with exponent $1/2 < \alpha \leq 1$ we can obtain a uniform upper bound for the pointwise mean-square error of the estimator \widehat{f}_N constructed using Legendre polynomials.

LEMMA 3.1. *If $f \in C[-1, 1]$ satisfies the Lipschitz condition with exponent $1/2 < \alpha \leq 1$ and $D_n = O(n^{-1})$, then the estimator \widehat{f}_N constructed using Legendre polynomials satisfies*

$$E(f(x) - \widehat{f}_N(x))^2 \leq \frac{G_1 N^3}{n} + \frac{G_2 N^{2-\alpha}}{n^\alpha} + \frac{G_3}{N^{2\alpha-1}},$$

where $x \in [-1, 1]$ and $G_1, G_2, G_3 > 0$ are constants.

Proof. By assumption, there exist constants $C, L > 0$ such that $D_n \leq Cn^{-1}$ and $\omega(f, \delta) \leq L\delta^\alpha$ for $\delta > 0$, so by (4) and (8) we have

$$\begin{aligned} E(f(x) - \widehat{f}_N(x))^2 &\leq (N + 1)^3 \left(\frac{C\sigma_\eta^2}{2n} + \frac{L^2 C^{2\alpha}}{n^{2\alpha}} \right) \\ &\quad + 2|r_N(x)|(N + 1)^{3/2} \frac{LC^\alpha}{n^\alpha} + r_N^2(x). \end{aligned}$$

If f satisfies the Lipschitz condition with exponent $\alpha > 1/2$, then its Fourier-Legendre series converges uniformly and $|r_N(x)| \leq B(\alpha)/N^{\alpha-1/2}$ for $x \in [-1, 1]$ (see Theorem 4.9 of [12]). Hence,

$$\begin{aligned} E(f(x) - \widehat{f}_N(x))^2 &\leq (N + 1)^3 \left(\frac{C\sigma_\eta^2}{2n} + \frac{L^2 C^{2\alpha}}{n^{2\alpha}} \right) \\ &\quad + \frac{2(N + 1)^{3/2} B(\alpha) LC^\alpha}{N^{\alpha-1/2}} \frac{1}{n^\alpha} + \frac{B^2(\alpha)}{N^{2\alpha-1}}, \end{aligned}$$

which can be rewritten as

$$E(f(x) - \widehat{f}_N(x))^2 \leq N^3 \left(\frac{F_1 \sigma_\eta^2}{n} + \frac{F_2}{n^{2\alpha}} \right) + \frac{G_2 N^{2-\alpha}}{n^\alpha} + \frac{G_3}{N^{2\alpha-1}},$$

and since $2\alpha > 1$ the lemma follows. ■

For the Haar series estimator the following lemma holds.

LEMMA 3.2. *If $f \in C[0, 1]$ satisfies the Lipschitz condition with exponent $0 < \alpha \leq 1$ and $D_n = O(n^{-1})$, then the estimator \widehat{f}_N constructed using Haar functions satisfies*

$$E(f(x) - \widehat{f}_N(x))^2 \leq N^2 \left(\frac{J_1}{n} + \frac{J_2}{n^{2\alpha}} \right) + \frac{J_3 N^{1-\alpha}}{n^\alpha} + \frac{J_4}{N^{2\alpha}}$$

almost everywhere in $[0, 1]$, where $J_1, J_2, J_3, J_4 > 0$ are constants.

Proof. The proof is analogous to that of Lemma 3.1 except that now we use inequality (6) and the bound (5), which yield $|r_N(x)| \leq 2L/N^\alpha$ almost everywhere in $[0, 1]$. ■

Now, we can formulate and prove the main results of this section concerning the convergence rates of the pointwise mean-square error of the relevant estimators.

THEOREM 3.1. *If $f \in C[-1, 1]$ satisfies the Lipschitz condition with exponent $1/2 < \alpha \leq 1$, $D_n = O(n^{-1})$ and $N(n) \sim n^{1/(2(\alpha+1))}$, then the estimator \widehat{f}_N constructed using Legendre polynomials satisfies*

$$E(f(x) - \widehat{f}_{N(n)}(x))^2 = O(n^{-(2\alpha-1)/(2(\alpha+1))}),$$

uniformly for $x \in [-1, 1]$.

Proof. By Lemma 3.1, putting $N(n) \sim n^{1/(2(\alpha+1))}$, we have for $x \in [-1, 1]$,

$$E(f(x) - \widehat{f}_{N(n)}(x))^2 \leq H_1 n^{-\frac{2\alpha-1}{2(\alpha+1)}} + H_2 n^{-\frac{2\alpha(\alpha+1)+\alpha-2}{2(\alpha+1)}} + H_3 n^{-\frac{2\alpha-1}{2(\alpha+1)}},$$

where $H_1, H_2, H_3 > 0$. In other words

$$E(f(x) - \widehat{f}_{N(n)}(x))^2 \leq (H_1 + H_3) n^{-\frac{2\alpha-1}{2(\alpha+1)}} + H_2 n^{-\frac{2\alpha-1}{2(\alpha+1)}} n^{-\frac{2\alpha^2+\alpha-1}{2(\alpha+1)}}$$

for $x \in [-1, 1]$, and since $2\alpha^2 + \alpha - 1 = 2(\alpha + 1)(\alpha - 1/2) > 0$ for $\alpha > 1/2$ the assertion follows. ■

THEOREM 3.2. *If $f \in C[0, 1]$ satisfies the Lipschitz condition with exponent $0 < \alpha \leq 1$, $D_n = O(n^{-1})$ and the estimator \widehat{f}_N is constructed using Haar functions, then*

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

$$E(f(x) - \widehat{f}_{N(n)}(x))^2 = O(n^{-2\alpha^2/(\alpha+1)}) \quad \text{almost everywhere in } [0, 1],$$

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

$$E(f(x) - \widehat{f}_{N(n)}(x))^2 = O(n^{-\alpha/(\alpha+1)}) \quad \text{almost everywhere in } [0, 1].$$

Proof. (a) Observe that the assumptions of Lemma 3.2 hold. Consequently, for $0 < \alpha \leq 1/2$ and $N(n) \sim n^{\alpha/(\alpha+1)}$ we have

$$E(f(x) - \widehat{f}_{N(n)}(x))^2 \leq L_1 n^{-\frac{2\alpha^2}{\alpha+1}} + L_2 n^{-\frac{\alpha(\alpha+1)+\alpha(\alpha-1)}{\alpha+1}} + L_3 n^{-\frac{2\alpha^2}{\alpha+1}},$$

almost everywhere in $[0, 1]$, where $L_1, L_2, L_3 > 0$. Otherwise,

$$E(f(x) - \widehat{f}_{N(n)}(x))^2 \leq (L_1 + L_2 + L_3) n^{-2\alpha^2/(\alpha+1)}$$

almost everywhere in $[0, 1]$, which proves (a).

To prove (b) observe that for $2\alpha > 1$ and $N(n) \sim n^{1/(2(\alpha+1))}$ Lemma 3.2 gives

$$E(f(x) - \widehat{f}_{N(n)}(x))^2 \leq K_1 n^{-\frac{2\alpha}{2(\alpha+1)}} + K_2 n^{-\frac{2\alpha(\alpha+1)+\alpha-1}{2(\alpha+1)}} + K_3 n^{-\frac{2\alpha}{2(\alpha+1)}}$$

almost everywhere in $[0, 1]$, where $K_1, K_2, K_3 > 0$. In other words,

$$E(f(x) - \widehat{f}_{N(n)}(x))^2 \leq (K_1 + K_3) n^{-\frac{2\alpha}{2(\alpha+1)}} + K_2 n^{-\frac{2\alpha}{2(\alpha+1)}} n^{-\frac{2\alpha^2+\alpha-1}{2(\alpha+1)}}$$

almost everywhere in $[0, 1]$, and again (b) follows. ■

4. Conclusions. In [11] it was proved that the pointwise mean-square error of the polynomial estimator obtained by the least squares method in the case of equidistant observation points converges uniformly to zero if the regression function f satisfies the Lipschitz condition with exponent $1/2 < \alpha \leq 1$ and $N^2(n)/n \rightarrow 0$ for $n \rightarrow \infty$. Earlier Rafajłowicz [8] obtained sufficient conditions for uniform consistency of that estimator in the sense of pointwise mean-square error for other observation point designs, but they are satisfied only by functions of higher smoothness. In the present work we go one step further showing that there exist orthogonal series estimators uniformly consistent in the sense of pointwise mean-square error, and obtained the convergence rate for that error, under the same smoothness conditions on the regression function as in [11] and less restrictive assumptions on the observation points. Theorem 3.2 shows that using Haar functions makes it possible to obtain the convergence rate of the pointwise mean-square error (valid almost everywhere) for a slightly larger regression function class than in the case of Legendre polynomials. Moreover, for regression functions satisfying the Lipschitz condition with exponent $1/2 < \alpha \leq 1$, that convergence rate is faster for the Haar series estimator.

In the present work it is also shown that the estimators considered can attain, at least for regression functions satisfying the Lipschitz condition with exponent $1/2 < \alpha \leq 1$, the same optimal IMSE convergence rate as the least squares polynomial estimators considered in [8] and orthogonal series estimators based on Legendre polynomials and numerical quadratures [7]. Our Haar series estimator attains the same IMSE convergence rate as the Haar series estimator for the equidistant point design considered in [2], where only the case of regression functions satisfying the Lipschitz condition with exponent $\alpha = 1$ was examined. Hence, the asymptotic properties of our estimator seem to be competitive in comparison to series regression estimators based on orthogonal polynomials and Haar functions, considered so far.

If the regression function f satisfies the Dini–Lipschitz condition, i.e. $\lim_{\delta \rightarrow 0} \ln(1/\delta)\omega(f, \delta) = 0$, then the same technique shows that the estimators constructed using Legendre polynomials or Haar functions are consistent in the sense of integrated mean-square error provided $N(n) \rightarrow \infty$ and $N(n) = O(\ln^2 n)$. Moreover, for such regression functions the pointwise mean-square error of the Haar series estimator converges uniformly to zero almost everywhere in $[0, 1]$ if $N(n) \rightarrow \infty$ and $N(n) = O(\ln n)$. In the case of the estimator based on Legendre polynomials the known results on the uniform convergence of Legendre series (e.g. Theorem 4.7 of [12]) only ensure that the error converges uniformly to zero on any interval $[-1+\varepsilon, 1-\varepsilon]$, $\varepsilon > 0$, if $N(n) \rightarrow \infty$ and $N(n) = O(\ln^{2/3} n)$. In [11] Sharapudinov proved an analogous result for the least squares polynomial estimator in the case of an equidistant point design, provided $N(n) = O(n^{1/2})$.

Since the Haar series estimators are also wavelet estimators, the results obtained give some idea of properties of wavelet estimators in the case where the observation points are irregularly spaced. Asymptotic properties of wavelet estimators for a random uniform design, i.e. a design with irregular spacing of observation points, were recently investigated in [1].

References

- [1] T. T. Cai and L. D. Brown, *Wavelet estimation for samples with random uniform design*, Statist. Probab. Lett. 42 (1999), 313–321.
- [2] J. Engel, *A simple wavelet approach to nonparametric regression from recursive partitioning schemes*, J. Multivariate Anal. 49 (1994), 242–254.
- [3] R. L. Eubank, J. D. Hart and P. Speckman, *Trigonometric series regression estimators with an application to partially linear models*, *ibid.* 32 (1990), 70–83.
- [4] T. Gasser and H. G. Müller, *Kernel estimation of regression functions*, in: *Smoothing Techniques for Curve Estimation*, T. Gasser and M. Rosenblatt (eds.), Lecture Notes in Math. 757, Springer, Heidelberg, 1979, 23–68.
- [5] G. V. Milovanović, D. S. Mitrinović and T. M. Rassias, *Topics on Polynomials: Extremal Problems, Inequalities, Zeros*, World Sci., Singapore, 1994.
- [6] I. Novikov and E. Semenov, *Haar Series and Linear Operators*, Math. Appl. 367, Kluwer, Dordrecht, 1997.
- [7] E. Rafajłowicz, *Nonparametric orthogonal series estimators of regression: a class attaining the optimal convergence rate in L_2* , Statist. Probab. Lett. 5 (1987), 219–224.
- [8] —, *Nonparametric least-squares estimation of a regression function*, Statistics 19 (1988), 349–358.
- [9] L. Rutkowski, *Orthogonal series estimates of a regression function with application in system identification*, in: W. Grossmann et al. (eds.), *Probability and Statistical Inference*, Reidel, 1982, 343–347.
- [10] G. Sansone, *Orthogonal Functions*, Interscience, New York, 1959.
- [11] I. I. Sharapudinov, *On convergence of least-squares estimators*, Mat. Zametki 53 (1993), 131–143 (in Russian).
- [12] P. K. Suetin, *Classical Orthogonal Polynomials*, Nauka, Moscow, 1976 (in Russian).

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

*Received on 27.5.1999;
 revised version on 5.1.2000*