Consistency of a least squares orthonormal series estimator for a regression function

Michel Delecroix[a], Camelia Protopopescu[b]

[a]*ENSAI, Rue Blaise Pascal, Campus de Kerr Lann, 35170 Bruz, France, e-mail: delecroi@ensai.fr;* [b]*GREQAM, Centre de la Vieille Charité, 2 rue de la Charité, 13002 Marseille, France, e-mail: camelia@ehess.cnrs-mrs.fr*

Abstract: This paper establishes the almost sure consistency of least squares regression series estimators, in the L^2-norm and the sup-norm, under very large assumptions on the underlying model. Three examples are considered in order to illustrate the general results: trigonometric series, Legendre polynomials and wavelet series estimators. Then optimal choices for the number of functions in the series are discussed and convergence rates are derived. It is shown that for the wavelet case, the best possible convergence rate is attained.

Keywords: Nonparametric regression, orthonormal series estimators, least squares, almost sure consistency, convergence rates, trigonometric series, Legendre polynomials, wavelets.

1 Introduction

Let (X_i, Y_i), $i = 1, ..., n$, be a random sample of independent, identically distributed variables, taking values respectively in $\mathcal{X} \subset I\!R^d$ and $I\!R$. Let g denote the regression function of Y_i given X_i, so that $g(x) = \mathbf{E}(Y_i | X_i = x)$, for $x \in \mathcal{X}$. It is assumed that the distribution of X_i is absolutely continuous with density f. This paper deals with the problem of estimating g, in a nonparametric approach.

Our method relies on what has been called the "projection" method (see e.g. Bosq and Lecoutre [5] or, more recently, Shen and Wong [17] and Shen [16] for the "sieve" approach). It consists, generally speaking, in estimating the projection g^* of g onto a finite dimensional space, the dimension of which is allowed to grow to infinity, with g^* depending on a finite number of real numbers c_j. The estimator $\widehat{g}_n$ is then defined using the least squares estimates of c_j, computed on the basis of the sample.

In the particular case where g^* is a finite linear combination of known functions ($\widehat{g}_n$ is then called a "series-type" estimator), several important results have been recently achieved, a few of which are described below. For this purpose, three types of criteria, previously used in studying the large sample properties of these estimators, are first defined:

- the sample mean squared error: $\frac{1}{n} \sum_{i=1}^{n} [\widehat{g}_n(X_i) - g(X_i)]^2$, (1)
- the integrated squared error (weighted by the marginal density):

$$\int_{\mathcal{X}} [\widehat{g}_n(x) - g(x)]^2 f(x)\, dx, \tag{2}$$

- the maximal absolute deviation: $\sup_{x \in \mathcal{X}} |\widehat{g}_n(x) - g(x)|$. (3)

Newey [14] obtained a consistency theorem for a large class of series estimators and used it to establish the consistency of power series and regression splines estimators. He gave asymptotic bounds in probability for criteria (1) and (3), and showed asymptotic normality for nonlinear functionals of series estimators. Andrews [1] gave an asymptotic distribution theorem: he provided conditions under which $\widehat{g}_n(x)$ is asymptotic normally distributed after being centered at either its expectation or the estimand and normalized by premultiplication by the squared root of its covariance matrix.

Both the previous authors supposed g to belong to a Sobolev space and used a sequence of functions which span it. Related work can be found in Lugosi and Zeger [12]. They used an uniformly bounded sequence of functions $(\psi_j)_{j \geq 1}$ for which the set of all finite linear combinations is assumed to be dense in $L^2(\mu)$, for *any* probability measure μ on $\mathcal{X}$. Then $\widehat{g}_n = \sum_{j=1}^{k_n} \widehat{a}_j \psi_j$, where $(\widehat{a}_j)_{j=\overline{1,k_n}}$ minimizes:

$$\frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \sum_{j=1}^{k_n} a_j \psi_j(X_i)\right]^2 \tag{4}$$

under a boundedness constraint on $\sum_{j=1}^{k_n} |a_j|$. A universal strong consistency is proved for this estimator, namely the almost sure convergence of risk (2) for all distributions of (X_i, Y_i) with $\mathbf{E}|Y_i|^2 < \infty$.

The case of an orthonormal basis on a compact interval $\mathcal{X}$ is analyzed in Andrews [1], for the particular case of trigonometric functions, and in Eubank and Speckman [8], where upper bounds for the expectation of risk (1) are obtained, for trigonometric and polynomial-trigonometric least squares estimators. Antoniadis, Grégoire and McKeague [2] obtained the same type of result, among several others, when using wavelet series estimators.

The estimator defined in this paper is also based on orthonormal bases. The previous authors typically assumed that the density f is bounded from below on $\mathcal{X}$, or an equivalent condition[1]. In the same way, they assumed that g belongs to $L^2(\mathcal{X})$.

In this paper, the asymptotic consistency of the estimator will be established assuming only the following hypothesis, which seemed to provide a more realistic condition on the model:

H0 There exists an increasing sequence of bounded sets K_n such that $\bigcup_{n\geq 1} K_n = \mathcal{X}$ and:

- $g|_{K_n} \in L^2(K_n)$, $\forall n \geq 1$;
- $\alpha_n \overset{\text{def}}{=} \inf_{x\in K_n} f(x) > 0$, $\forall n \geq 1$.

For instance, assumption **H0** is fulfilled if K_n are compact sets, if f is continuous and doesn't vanish on $\mathcal{X}$, and if g is continuous or bounded on $\mathcal{X}$. This kind of regularity condition (with respect to f) imposed on the sequence $(K_n)_{n\geq 1}$ is a possible approach in the case when regression estimator consistency cannot be obtained over the whole space, but only over a suitably increasing sequence of compact sets (see also Bosq [4], in the case of kernel density estimation for discrete time processes).

The regression estimator $\widehat{g}_n$ is then formally defined by:

$$\widehat{g}_n(x) = \sum_{j=1}^{q(n)} \widehat{c}_j^n e_j^n(x)\, \mathbf{I}_{K_n}(x), \tag{5}$$

[1] A boundedness condition on the smallest eigenvalue of a second moment matrix (see also Lemma 1 in the Appendix) is imposed in Andrews [1] and Newey [14].

where

$$\left(\widehat{c}_1^n, ..., \widehat{c}_{q(n)}^n\right)' = \arg\min_{a_j} \sum_{i=1}^{n} \left[Y_i - \sum_{j=1}^{q(n)} a_j e_j^n (X_i) \mathbf{I}_{K_n} (X_i) \right]^2,$$

$\left(e_j^n\right)_{j\geq 1}$ denotes an orthonormal basis of $L^2(K_n)$ and $q(n)$ is a deterministic sequence of integers increasing to infinity with the sample size.

To study the consistency of the proposed estimator, it was considered more natural to use the criteria $\|\widehat{g}_n - g\|_{n,2}$ and $\|\widehat{g}_n - g\|_{n,\infty}$, where $\|\cdot\|_{n,2}$ denotes the usual norm of $L^2(K_n)$ and $\|\cdot\|_{n,\infty}$ denotes the sup-norm on K_n. These two norms correspond to intrinsic distances between two curves, which do not change with the distribution of the regressors. The main goal of this paper is to study their almost surely asymptotic behavior, which, to the best of our knowledge, was not considered until now for regression series estimation.

The remainder of the paper is organized as follows. Section 2 will give general sufficient conditions for the almost sure consistency of the estimator in the previous norms. In Section 3, these results will be particularized in the case where $d = 1$ and $\left(e_j^n\right)_{j\geq 1}$ is the Fourier or Legendre basis. Section 4 focuses on a particular case of wavelet basis in the multidimensional setting. In Section 5 general bounds in probability for the criteria defined above will be obtained, after which the problem of the optimal choice of $q(n)$ will be studied. We shall derive best possible rates of convergence and compare them with the classical results obtained by Stone [18]. Finally, in Section 6, some concluding remarks will be highlighted. An Appendix contains the proofs of the results stated in the text.

2 General consistency results

Let us first introduce some notations. Let $U_i = Y_i - g(X_i)$, $i = \overline{1,n}$, denote the errors of the regression model and $\sigma^2(x) = \mathbf{V}(Y_i|X_i = x) = \mathbf{E}(U_i^2|X_i = x)$ denote the conditional variance of Y_i given $X_i = x$. Let $\sum_{j=1}^{\infty} c_j^n e_j^n$ be the series expansion of g in the basis of $L^2(K_n)$, where $c_j^n = \int_{K_n} g(x) e_j^n(x) dx$. Let $\xrightarrow[n\to\infty]{a.s.}$ denote the almost sure convergence and $\xrightarrow[n\to\infty]{co.}$ the complete convergence of random variables (see e.g. Billingsley [3]). The following assumptions will furthermore be considered:

H1 $\exists\, C_1 > 0$ such that $f(x) \leq C_1, \forall x \in \mathcal{X}$.

H2 $\sup_{x\in\mathcal{X}} \sigma^2(x) < \infty$.

H3 $\exists\, (\omega_n)_{n\geq 1}$ an increasing sequence such that $\sup_{\substack{x\in K_n \\ j=\overline{1,q(n)}}} \left|e_j^n(x)\right| \leq \omega_n, \forall n \geq 1$.

H4 $\exists\, C_2 > 0$ such that $\mathbf{E}\left[|U_i|^p \,|X_i\right] \leq (C_2)^{p-2}\, p!\mathbf{E}\left[|U_i|^2 \,|X_i\right], \forall p \geq 3$.

The upper bound imposed on the density by **H1** and the bounded second conditional moment assumption **H2** are quite common in the literature (see e.g. Andrews [1]). Assumption **H3** is a regularity condition imposed on the basis. In particular, it holds with $\omega_n = M$ for the trigonometric basis and $\omega_n = M\sqrt{q(n)}$ in the case of Legendre polynomials. The last assumption, known as "Cramer's condition", is needed in order to write exponential type inequalities to obtain the almost sure consistency. Assumption **H4** is trivially verified if, for example, Y_i is bounded or in the case where U_i is normally distributed.

Then one gets the following consistency result:

Theorem 1 a) *If assumptions* **H0** *to* **H4** *hold and the sequence* $q(n)$ *verifies:*

(i) $$\frac{q(n)^2\,\omega_n^2}{\alpha_n^2}\sum_{j=q(n)+1}^{\infty}\left(c_j^n\right)^2 \underset{n\to\infty}{\longrightarrow} 0$$

(ii) $$\forall\gamma > 0 : \sum_{n=1}^{\infty}\exp\left[-\gamma\frac{n\alpha_n^4}{q(n)^4\,\omega_n^4}\right] < \infty$$

then:

$$\|\widehat{g}_n - g\|_{n,2} \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} 0$$

b) *If furthermore,*

(iii) $$\sup_{x\in K_n}\left|g(x) - \sum_{j=1}^{q(n)} c_j^n e_j^n(x)\right| \underset{n\to\infty}{\longrightarrow} 0$$

then:

$$\|\widehat{g}_n - g\|_{n,\infty} \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} 0$$

Proof. See Appendix.

The assumptions in this theorem are somewhat complicated in this general form. Hypothesis (iii) is a uniformly approximation property that characterizes the basis, whereas the first two conditions can be expressed in terms of convergence rates for the sequences $q(n)$, ω_n, α_n and $\sum_{j=q(n)+1}^{\infty}\left(c_j^n\right)^2$. However, a more simple form for these sufficient conditions can be given if, in each space $L^2(K_n)$, a classical basis $\left(e_j^n\right)_{j\geq 1}$ is chosen (see the next section). In that case, ω_n and coefficients c_j^n can be evaluated if the regression function is supposed to belong to some regularity class. In this way, the sufficient conditions for the almost sure consistency will be expressed only in terms of convergence rates of $q(n)$ to infinity and of α_n to zero or, equivalently, in terms of growth rates of $q(n)$ to infinity and of K_n to $\mathcal{X}$.

Let it be noted that if our study is confined to the classical case analyzed in the literature, in which the density f is bounded from below on $\mathcal{X}$, then all the sufficient conditions will be simplified: in this particular case, the sequence α_n will be bounded away from zero and thus will disappear from all assumptions. In the particular cases presented in Section 3, $q(n)$ thus remains the unique parameter to be chosen in order to obtain consistency.

3 The one-dimensional case: two examples

This section will be focused on the univariate case, to illustrate the application of the general consistency results in a simple classical situation. However, the following discussion can be generalized in a straightforward manner to higher-dimensional settings, using product bases. The two following examples, which are of interest in their own right, will be our main cases of application: the first uses trigonometric bases and the second - Legendre polynomial bases.

For the sake of simplicity, it will be assumed that hypothesis **H0** is fulfilled for a sequence of compact symmetric intervals $K_n = [-k_n, k_n]$, with $(k_n)_{n\geq 1}$ being an increasing sequence of positive numbers. This implies that $\mathcal{X}$ must be a symmetric (bounded or unbounded) interval, an assumption which is not restrictive, because a one-to-one transformation of the original X_i variables can always be performed onto such an interval. One can then evaluate the coefficients c_j^n (see Lemmas 3 and 4 in the Appendix) and obtain relatively simple sufficient conditions for the general results of Section 2, corresponding to a large class of regression functions. More precisely, it is assumed that g belongs to the Sobolev space $W^m(\mathcal{X})$ (the space of functions which are $m-1$ times continuously differentiable on $\mathcal{X}$, and such that their mth derivative is square integrable).

Example 1: Legendre polynomials

Let $\left(e_j^n\right)_{j\geq 1}$ denote the orthonormal basis of Legendre polynomials on K_n:

$$e_{j+1}^n(x) = \sqrt{\frac{2j+1}{2k_n}} \frac{1}{(2k_n)^j\, j!} \frac{d^j}{dx^j}\left[\left(x^2 - k_n^2\right)^j\right] = \sqrt{\frac{2j+1}{2k_n}} P_j\left(\frac{x}{k_n}\right),\ x \in K_n,\ j \geq 0,$$

where $P_j(y) = \dfrac{1}{2^j j!}\dfrac{d^j}{dy^j}\left[\left(y^2-1\right)^j\right]$, $y \in [-1,1]$, is the standard Legendre polynomial of order j. Then assumption **H3** will be verified for $\omega_n = \sqrt{\dfrac{q(n)}{k_n}}$ (see Delecroix and Protopopescu [6]).

One obtains in this case the following consistency result:

Corollary 1 *If assumptions* **H0, H1, H2** *and* **H4** *hold,* $\left(e_j^n\right)_{j\geq 1}$ *denotes the Legendre basis on* K_n*,* $g \in W^m(\mathcal{X})$*, with* $m \geq 3$*, and the sequence* $q(n)$ *verifies:*

(i) $\dfrac{k_n^{2m-1}}{\alpha_n^2 q(n)^{2m-4}} \underset{n\to\infty}{\longrightarrow} 0$

(ii) $\forall \gamma > 0 : \sum_{n=1}^{\infty} \exp\left[-\gamma \dfrac{n\alpha_n^4 k_n^2}{q(n)^6}\right] < \infty$

then:

a) $\|\widehat{g}_n - g\|_{n,2} \overset{a.s.}{\underset{n\to\infty}{\longrightarrow}} 0$

b) $\|\widehat{g}_n - g\|_{n,\infty} \overset{a.s.}{\underset{n\to\infty}{\longrightarrow}} 0$

Proof. It follows from Theorem 1 and Lemma 3 in the Appendix.

As a consequence, if the tail behavior of the marginal density is known, it is possible to express the sufficient conditions above in terms of growth rates for $q(n)$ and k_n. This will be done in two particular cases, which are extreme situations for the behavior of the distributions tails: the standard normal distribution and the Cauchy distribution. Let us note that the results can be presented in a more classical way: the obtained choices are valid for *every* density such that: $f(k_n) \geq C \cdot f_0(k_n)$, where f_0 is the particular chosen density. The regression function is supposed to belong to the regularity class $W^m(\mathcal{X})$, with $m \geq 3$. The obtained sufficient conditions for the almost sure consistency results in Corollary 1, are given in the table below:

$X_i \sim N(0,1)$	$X_i \sim Cauchy$
$\dfrac{q(n)^{2m-1}}{\sqrt{\ln q(n)}} = \mathcal{O}\left(n^{\frac{1}{2+\tau}}\right),\ \tau > 0$	$q(n) = \mathcal{O}\left(n^{\frac{1}{\frac{24m-6}{2m+3}+\tau}}\right),\ \tau > 0$
$k_n \leq \dfrac{\sqrt{\ln q(n)}}{\frac{1}{\sqrt{2m-4}} + \eta},\ \eta > 0$	$k_n = \mathcal{O}\left(q(n)^{\frac{1}{\frac{2m+3}{2m-4}+\eta}}\right),\ \eta > 0$

For the sake of brevity, the proof is left to the reader.

Example 2: Trigonometric functions

Let $\left(e_j^n\right)_{j\geq 1}$ denote the orthonormal basis of trigonometric functions on K_n:

$$e_1^n(x) = \frac{1}{\sqrt{2k_n}},\ \forall x \in K_n\ ;$$

$$e_j^n(x) = \begin{cases} \dfrac{1}{\sqrt{k_n}} \cos\left(\dfrac{j\pi x}{2k_n}\right), & j \text{ even} \\ \dfrac{1}{\sqrt{k_n}} \sin\left[\dfrac{(j-1)\pi x}{2k_n}\right], & j \text{ odd} \end{cases},\ j \geq 2,\ \forall x \in K_n.$$

Then assumption **H3** is fulfilled for $\omega_n = \dfrac{1}{\sqrt{k_n}}$ and we obtain the following result:

Corollary 2 *If assumptions* **H0***,* **H1***,* **H2** *and* **H4** *hold,* $\left(e_j^n\right)_{j\geq 1}$ *denotes the trigonometric basis on* K_n*,* $g \in W^m(\mathcal{X})$*, with* $m \geq 2$*, and verifies for each* n *the periodicity condition* $g^{(i)}(-k_n) = g^{(i)}(k_n)$*, for* $i = \overline{0, m-1}$*, and the sequence* $q(n)$ *verifies:*

(i) $\dfrac{k_n^{2m}}{\alpha_n^2 q(n)^{2m-3}} \underset{n\to\infty}{\longrightarrow} 0$

(ii) $\forall \gamma > 0 : \sum_{n=1}^{\infty} \exp\left[-\gamma \dfrac{n\alpha_n^4 k_n^2}{q(n)^4}\right] < \infty$

then:

a) $\|\widehat{g}_n - g\|_{n,2} \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} 0$

b) $\|\widehat{g}_n - g\|_{n,\infty} \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} 0$

Proof. This follows from Theorem 1 and Lemma 4 in the Appendix.

In the particular cases of standard normal distribution and Cauchy distribution, the following sufficient conditions are found:

$X_i \sim N(0,1)$	$X_i \sim Cauchy$
$\dfrac{q(n)^{2m-1}}{\sqrt{\ln q(n)}} = \mathcal{O}\left(n^{\frac{1}{2+\tau}}\right), \ \tau > 0$	$q(n) = \mathcal{O}\left(n^{\frac{1}{\frac{10m-1}{m+2}+\tau}}\right), \ \tau > 0$
$k_n \leq \dfrac{\sqrt{\ln q(n)}}{\frac{1}{\sqrt{2m-3}} + \eta}, \ \eta > 0$	$k_n = \mathcal{O}\left(q(n)^{\frac{1}{\frac{2m+4}{2m-3}+\eta}}\right), \ \eta > 0$

4 The multidimensional case: a wavelet basis

In this section the wavelet version of our estimator is studied, under the general assumptions **H0, H1, H2** and **H4** stated in the first part of the paper. Wavelet estimation is computationally convenient, because the estimator is summarized by relatively few estimated coefficients. Antoniadis, Grégoire and McKeague [2] proved the convergence in expectation of the sample mean squared error, $n^{-1}\sum_{i=1}^{n}\left[g(X_i) - \widehat{g}_n(X_i)\right]^2$, assuming that g belongs to a Sobolev space on $\mathcal{X} = [0,1]$ and verifies the boundary condition $g(0) = g(1)$.

Without loss of generality (see the argument in Section 3), in this section we assume that $\mathcal{X}$ is a centered ball of $\mathbb{R}^d$ and that hypothesis **H0** is verified for a sequence of centered balls K_n of increasing radii. In every space $L^2(K_n)$ we will consider a wavelet orthonormal basis, as described by Jaffard and Meyer [10]. This basis is composed of functions

$$\left\{\psi_{j,k}^n \mid j \geq j_o(n),\ k \in R_j^n\right\},$$

where $j_o(n)$ is an integer depending on the bounded set K_n, and R_j^n is a finite set defined as $R_j^n = \Lambda_{j+1}^n \setminus \Lambda_j^n$, with

$$\Lambda_j^n = \left\{k \in 2^{-j}\mathbb{Z}^d \cap K_n \mid d(k, \partial K_n) \geq (\mu+1)\, 2^{-j}\right\}.$$

Here $\mu \geq 1$ is a fixed integer and $d(\cdot,\cdot)$ is the distance corresponding to the norm $|x| = \max\limits_{i=\overline{1,d}} |x_i|$.

Assumption **H3** is then verified for $\omega_n = C_1^* 2^{dq(n)/2}$, with $C_1^* > 0$ (see the localization formula denoted by (F) in [10]), so we have:

$$\exists C_1^* > 0 \text{ such that } \sup_{\substack{x \in K_n \\ j=\overline{j_o(n), q(n)}, k \in R_j^n}} \left|\psi_{j,k}^n(x)\right| \leq C_1^* 2^{dq(n)/2},\ \forall n \geq 1. \tag{6}$$

The least squares wavelet estimator will be formally defined as:

$$\widehat{g}_n(x) = \sum_{j=j_o(n)}^{q(n)} \sum_{k \in R_j^n} \widehat{c}_{j,k}^n \psi_{j,k}^n(x)\, \mathbf{I}_{K_n}(x), \tag{7}$$

where

$$\left(\widehat{c}_{j,k}^n\right)_{\substack{j=\overline{j_o(n), q(n)} \\ k \in R_j^n}} = \arg\min_{a_{j,k}} \sum_{i=1}^n \left[Y_i - \sum_{j=j_o(n)}^{q(n)} \sum_{k \in R_j^n} a_{j,k} \psi_{j,k}^n(X_i)\, \mathbf{I}_{K_n}(X_i)\right]^2$$

and $q(n) \geq j_o(n)$ is a sequence of integers increasing to infinity.

Since $\mathcal{X}$ is bounded, there exists an integer J_o such that $j_o(n) \geq J_o$, $\forall n \geq 1$ and thus:

$$card\left(R_j^n\right) \leq M 2^{jd},\ \forall n \geq 1,\ \forall j \geq j_o(n),$$

with $M > 0$ constant. Then the number of functions used in the definition of the estimator is bounded by $2M2^{dq(n)}$.

Let $\sum\limits_{j=j_o(n)}^{\infty} \sum\limits_{k \in R_j^n} c_{j,k}^n \psi_{j,k}^n$ denote the series expansion of g in the basis of $L^2(K_n)$, where $c_{j,k}^n = \displaystyle\int_{K_n} g(x)\, \psi_{j,k}^n(x) dx$. A result in Jaffard and Meyer [10, Theorem 1] gives a characterization for the space $C^m(\mathcal{X})$ of Hölder continuous functions of degree m, in terms of coefficients $c_{j,k}^n$. More precisely, if $0 < m < 2\mu - 2$, then g belongs to $C^m(\mathcal{X})$ if and only if

$$\left|c_{j,k}^n\right| \leq C_* 2^{-dj/2} 2^{-jm},\ \forall n \geq 1,\ \forall j \geq j_o(n),\ \forall k \in R_j^n, \tag{8}$$

with $C_* > 0$ a constant depending only on d and μ.

The wavelet estimator inherits the properties proved in Theorem 1. Thus we get the following consistency theorem:

Theorem 2 *If assumptions* **H0, H1, H2,** *and* **H4** *hold,* $\left(\psi_{j,k}^{n}\right)_{j,k}$ *denotes the wavelet basis defined above and* $g \in C^{m}(\mathcal{X})$, *with* $0 < m < 2\mu - 2$, *then for any sequence* $q(n)$ *such that:*

(i) $\dfrac{2^{(3d-2m)q(n)}}{\alpha_n^2} \underset{n\to\infty}{\longrightarrow} 0$

(ii) $\forall \gamma > 0 : \sum_{n=1}^{\infty} \exp\left[-\gamma \dfrac{n\alpha_n^4}{2^{4dq(n)}}\right] < \infty$

we have:

a) $\|\widehat{g}_n - g\|_{n,2} \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} 0$

b) $\|\widehat{g}_n - g\|_{n,\infty} \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} 0$

Proof. See Appendix.

5 Convergence rates

The present section is devoted to the study of mean square and uniform convergence rates for our estimator, in the particular cases considered in the previous sections.

We first state the general result:

Theorem 3 *If assumptions* **H0** *to* **H3** *hold, then:*

a) $\|\widehat{g}_n - g\|_{n,2}^2 = \mathcal{O}_P\left(\dfrac{q(n)}{n\alpha_n}\right) + \mathcal{O}_P\left(\dfrac{1}{\alpha_n}\sum_{j=q(n)+1}^{\infty}\left(c_j^n\right)^2\right)$

b) $\|\widehat{g}_n - g\|_{n,\infty} = \mathcal{O}_P\left(\dfrac{q(n)\,\omega_n}{\sqrt{n\alpha_n}}\right) + \mathcal{O}_P\left(\omega_n\left[\dfrac{q(n)}{\alpha_n}\sum_{j=q(n)+1}^{\infty}\left(c_j^n\right)^2\right]^{1/2}\right) + \mathcal{O}_P\left(\sup_{x\in K_n}\left|g^{q(n)}(x)\right|\right)$

Proof. See Appendix.

In the three cases introduced in Sections 3 and 4, it will be possible to get the values of $q(n)$ minimizing the obtained $\mathcal{O}_P$ bounds. To compare them with Stone's optimal bounds, it will be assumed, as in Stone's paper [18], that density f is bounded away from zero on its support $\mathcal{X}$, which is a compact set of $\mathbb{R}^d$ (to the best of our knowledge, the desire to eliminate this restriction still remains an open question). Therefore assumption **H0** will be replaced by:

H0' $\exists C > 0$ such that $f(x) \geq C$, $\forall x \in \mathcal{X}$ and $g \in L^2(\mathcal{X})$.

In this particular case, the regression estimator can be defined using a single orthonormal basis of $L^2(\mathcal{X})$ (i.e. $e_j^n = e_j$, $\forall n \in I\!N$), obtaining results valid over the whole space $\mathcal{X}$. In the sequel, $\|\cdot\|_{\mathcal{X},2}$ and $\|\cdot\|_{\mathcal{X},\infty}$ will denote respectively the norm of $L^2(\mathcal{X})$ and the sup norm on $\mathcal{X}$.

The previous theorem can thus be rewritten as follows:

Corollary 3 *Given assumptions* **H0'**, **H1**, **H2** *and* **H3**, *we have:*

a) $\|\widehat{g}_n - g\|_{\mathcal{X},2}^2 = \mathcal{O}_P\left(\frac{q(n)}{n}\right) + \mathcal{O}_P\left(\sum_{j=q(n)+1}^{\infty} c_j^2\right)$

b) $\|\widehat{g}_n - g\|_{\mathcal{X},\infty} = \mathcal{O}_P\left(\frac{q(n)\,\omega_n}{\sqrt{n}}\right) + \mathcal{O}_P\left(\omega_n\left[q(n)\sum_{j=q(n)+1}^{\infty} c_j^2\right]^{1/2}\right) + \mathcal{O}_P\left(\sup_{x\in\mathcal{X}}\left|g^{q(n)}(x)\right|\right)$

As is common in smoothing methods, antagonistic terms appear in these upper bounds: the bias terms, depending on $\sum_{j=q(n)+1}^{\infty} c_j^2$, which are decreasing functions of $q(n)$, while the variance terms, depending on $q(n)$, increase with the parameter. Achieving a trade-off between the smoothness of the estimator and fidelity to the sample will produce the best minimizing choice for $q(n)$. In the following, the practical choices adapted to the three bases introduced above will be given, then the appropriate decay rates for the two criteria will be calculated:

Example 1: Legendre polynomials
Using the bounds derived in Lemma 3 (see Appendix), we obtain for $\mathcal{X} = [-k, k]$, $k > 0$ and $g \in W^m(\mathcal{X})$, with $m \geq 2$:

$$\begin{aligned}\|\widehat{g}_n - g\|_{\mathcal{X},2}^2 &= \mathcal{O}_P\left(\frac{q(n)}{n}\right) + \mathcal{O}_P\left(\frac{1}{q(n)^{2m-1}}\right) \text{ and} \\ \|\widehat{g}_n - g\|_{\mathcal{X},\infty} &= \mathcal{O}_P\left(\frac{q(n)^{3/2}}{\sqrt{n}}\right) + \mathcal{O}_P\left(\frac{1}{q(n)^{m-3/2}}\right).\end{aligned}$$

When $q(n)$ goes to infinity at the same rate as $n^{\frac{1}{2m}}$, it results that

$$\begin{aligned}\|\widehat{g}_n - g\|_{\mathcal{X},2}^2 &= \mathcal{O}_P\left(n^{-\frac{2m-1}{2m}}\right) \text{ and} \\ \|\widehat{g}_n - g\|_{\mathcal{X},\infty} &= \mathcal{O}_P\left(n^{-\frac{m-3/2}{2m}}\right).\end{aligned} \tag{9}$$

In Stone [18], it has been proven that the best global convergence rates of any nonparametric regression estimator in the class $C^m(\mathcal{X})$ are respectively n^{-r} for the integrated squared error and $(n^{-1}\log n)^r$ for the sup norm, where $r = \frac{2m}{2m+d}$ and d is the dimension of the regressors. Let us note that the convergence rates (9) do not attain Stone's optimal bounds.

Example 2: Trigonometric functions
In this second particular case, if $\mathcal{X} = [-k, k]$, $k > 0$ and $g \in W^m(\mathcal{X})$, $m \geq 2$, satisfies the periodicity condition: $g^{(i)}(-k) = g^{(i)}(k)$, for $i = \overline{0, m-1}$, then:

$$\begin{aligned} \|\widehat{g}_n - g\|^2_{\mathcal{X},2} &= \mathcal{O}_P\left(\frac{q(n)}{n}\right) + \mathcal{O}_P\left(\frac{1}{q(n)^{2m-1}}\right) \text{ and} \\ \|\widehat{g}_n - g\|_{\mathcal{X},\infty} &= \mathcal{O}_P\left(\frac{q(n)}{\sqrt{n}}\right) + \mathcal{O}_P\left(\frac{1}{q(n)^{m-1}}\right). \end{aligned}$$

When $q(n)$ goes to infinity at the same rate as $n^{\frac{1}{2m}}$, we obtain the convergence rates

$$\begin{aligned} &\|\widehat{g}_n - g\|^2_{\mathcal{X},2} = \mathcal{O}_P\left(n^{-\frac{2m-1}{2m}}\right) \text{ and} \\ &\|\widehat{g}_n - g\|_{\mathcal{X},\infty} = \mathcal{O}_P\left(n^{-\frac{m-1}{2m}}\right), \end{aligned} \tag{10}$$

which are again not optimal in the sense of Stone's bounds, nevertheless the second of these improves on the corresponding rate (9) obtained in the previous case.

Example 3: Wavelet basis
If $g \in \mathcal{C}^m(\mathcal{X})$, with $m > 3d/2$, for a fixed ball $\mathcal{X} \subset \mathbb{R}^d$, the results of Section 4 lead to

$$\begin{aligned} \|\widehat{g}_n - g\|^2_{\mathcal{X},2} &= \mathcal{O}_P\left(\frac{2^{dq(n)}}{n}\right) + \mathcal{O}_P\left(\frac{1}{2^{2mq(n)}}\right) \text{ and} \\ \|\widehat{g}_n - g\|_{\mathcal{X},\infty} &= \mathcal{O}_P\left(\frac{2^{3dq(n)/2}}{\sqrt{n}}\right) + \mathcal{O}_P\left(\frac{1}{2^{(m-d)q(n)}}\right) \end{aligned}$$

(see also the bounds derived in the proof of Theorem 2). Then, choosing $q(n)$ so that $2^{q(n)}$ goes to infinity at the same rate as $n^{\frac{1}{2m+d}}$, we obtain the convergence rates

$$\begin{aligned} &\|\widehat{g}_n - g\|^2_{\mathcal{X},2} = \mathcal{O}_P\left(n^{-\frac{2m}{2m+d}}\right) \text{ and} \\ &\|\widehat{g}_n - g\|_{\mathcal{X},\infty} = \mathcal{O}_P\left(n^{-\frac{m-d}{2m+d}}\right). \end{aligned} \tag{11}$$

In this case, the rate for the integrated squared error is optimal, i.e. it reaches Stone's bound on the best obtainable rate.

Obviously, to decide if the series estimator is or is not optimal in the first two cases, it would be necessary to obtain equivalent functions of $\|\widehat{g}_n - g\|^2_{\mathcal{X},2}$ and $\|\widehat{g}_n - g\|_{\mathcal{X},\infty}$, instead of upper bounds. But, the dependence of the two criteria on $q(n)$ is somewhat complicated (see Appendix). As far as we know, the problem remains open.

6 Conclusion

In this paper it has been shown that series estimators can be used for regression problems under large assumptions on the underlying model. They are almost surely consistent for practical choices of the parameters, which are detailed in the text for several classical cases.

In particular, it is not necessary to assume that the density of the regressors is bounded from below on its support.

Finally, it has been deduced from the obtained results that choosing a wavelet basis leads to the optimal decay rate of convergence given by Stone, when assuming his own basic hypothesis.

7 Appendix

This Appendix contains the proofs of the results stated in the text. To give them, some additional definitions and notations are introduced. We define:

$$\begin{aligned} N_n &= \left(n_{ij}^n\right) = \left(e_j^n\left(X_i\right)\mathbf{I}_{K_n}\left(X_i\right)\right)_{i=1,\dots,n,\ j=1,\dots,q(n)}, \\ Y^n &= \left(Y_1,\dots,Y_n\right)'. \end{aligned}$$

In the sequel we will use the following $q(n) \times q(n)$ matrices A_n and B_n, of which the generic entries are respectively

$$a_{ij}^n = \frac{1}{n}\sum_{k=1}^{n} e_i^n\left(X_k\right)e_j^n\left(X_k\right)\mathbf{I}_{K_n}\left(X_k\right)$$

and

$$b_{ij}^n = \mathbf{E}a_{ij}^n = \int_{K_n} e_i^n(x)\, e_j^n(x)\, f(x)\, dx.$$

The above notations are similar to those in Delecroix and Protopopescu [6]: the B_n matrix is in fact identical, but A_n is a little bit different, although its properties are the same.

Then we have $A_n = \dfrac{1}{n}N_n' N_n$ and the coefficients that define the estimator will be given by the matricial formula:

$$\left(\widehat{c}_1^n,\dots,\widehat{c}_{q(n)}^n\right)' = \left(N_n' N_n\right)^- N_n' Y^n,$$

where $(\cdot)^-$ denotes a generalized inverse.

Let $\|A\|_2 \overset{\text{def}}{=} \sup_{\|x\|_2=1} \|Ax\|_2$ denote the matricial norm corresponding to the Euclidean norm for vectors and $\|A\|_\infty \overset{\text{def}}{=} \max_i \left(\sum_j |a_{ij}|\right)$. Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote respectively the minimum and maximum eigenvalue of a symmetric matrix A.

The next two lemmas show some properties of the matrices A_n and B_n, used in proving the results. As a consequence, we can deduce that, under assumptions **H0** and **H1**, B_n is a positive definite matrix. By a similar argument, A_n is only a positive semidefinite matrix (we cannot obtain a nonrandom lower bound for $\lambda_{\min}(A_n)$).

Lemma 1 *Given assumptions* **H0** *and* **H1**, *the matrix* B_n *verifies:*

$$\alpha_n \leq \lambda_{\min}(B_n) \leq \lambda_{\max}(B_n) \leq C_1, \forall n \geq 1.$$

Proof. See Delecroix and Protopopescu [6].

Lemma 2 *Given assumptions* **H0** *and* **H3**, *if the sequence* $q(n)$ *verifies:*

$$\forall \gamma > 0 : \sum_{n=1}^{\infty} q(n)^2 \exp\left[-\gamma \frac{n\alpha_n^2}{q(n)^2 \omega_n^4}\right] < \infty,$$

then the following complete convergence results hold:

a) $\frac{1}{\alpha_n} \|A_n - B_n\|_\infty \underset{n\to\infty}{\overset{co.}{\longrightarrow}} 0$ *and*

b) $\alpha_n \|A_n^-\|_2 \underset{n\to\infty}{\overset{co.}{\longrightarrow}} 1.$

Proof. a) This is similar to the demonstration given in Delecroix and Protopopescu [6] (the single difference lies in the sample size used in the definition of the matrix A_n).
b) For any $\varepsilon > 0$ we can write:

$$P\left[\alpha_n \|A_n^-\|_2 - 1 > \varepsilon\right] = P\left[\lambda_{\max}(A_n^-) > \frac{\varepsilon+1}{\alpha_n}\right]$$
$$= P\left[\frac{1}{\underset{i=1,q(n)}{\min}\left[\lambda_i(A_n)\,|\lambda_i(A_n) > 0\right]} > \frac{\varepsilon+1}{\alpha_n}\right] = P\left[\underset{i=1,q(n)}{\min}\left[\lambda_i(A_n)\,|\lambda_i(A_n) > 0\right] < \frac{\alpha_n}{\varepsilon+1}\right].$$

Using the inequality between eigenvalues of matrices A_n and B_n (see, for example, Lascaux and Théodor [11]),

$$|\lambda_i(A_n) - \lambda_i(B_n)| \leq \|A_n - B_n\|_\infty \tag{12}$$

we can write:

$$\underset{i=1,q(n)}{\min}\left[\lambda_i(A_n)\,|\lambda_i(A_n) > 0\right] \geq \lambda_{\min}(B_n) - \|A_n - B_n\|_\infty \overset{\text{Lemma 1}}{\geq} \alpha_n - \|A_n - B_n\|_\infty$$

and therefore

$$P\left[\alpha_n \|A_n^-\|_2 - 1 > \varepsilon\right] \leq P\left[\alpha_n - \|A_n - B_n\|_\infty < \frac{\alpha_n}{\varepsilon+1}\right] = P\left[\frac{1}{\alpha_n}\|A_n - B_n\|_\infty > \frac{\varepsilon}{\varepsilon+1}\right].$$

Using part a), the last quantity is the general term of a convergent series, so this proves the result. ■

Proof of Theorem 1.

a) The first part of the theorem is established as follows. We denote $c^n = \left(c_1^n, ..., c_{q(n)}^n\right)'$ and $\widehat{c}^n = \left(\widehat{c}_1^n, ..., \widehat{c}_{q(n)}^n\right)'$. Then we have:

$$\|\widehat{g}_n - g\|_{n,2}^2 = \int_{K_n} [\widehat{g}_n(x) - g(x)]^2 \, dx = \sum_{j=q(n)+1}^{\infty} \left(c_j^n\right)^2 + \sum_{j=1}^{q(n)} \left(\widehat{c}_j^n - c_j^n\right)^2.$$

Let us now establish the following majorization:

$$\begin{aligned} \|\widehat{g}_n - g\|_{n,2}^2 \ \leq \ & \sum_{j=q(n)+1}^{\infty} \left(c_j^n\right)^2 + \left\|\left(N_n' N_n\right)^{-1} N_n' Y^n - c^n\right\|_2^2 \mathbf{I}\left[\lambda_{\min}(A_n) > 0\right] \\ & + \left\|\left(N_n' N_n\right)^{-} N_n' Y^n - c^n\right\|_2^2 \mathbf{I}\left[\lambda_{\min}(A_n) = 0\right]. \end{aligned} \tag{13}$$

To prove this, it suffices to write that:

$$\sum_{j=1}^{q(n)} \left(\widehat{c}_j^n - c_j^n\right)^2 = \|\widehat{c}^n - c^n\|_2^2 = \left\|\left(N_n' N_n\right)^{-} N_n' Y^n - c^n\right\|_2^2.$$

The first term of (13), which is non random, converges to zero by hypothesis (i). The third term converges to zero almost surely, because for any $\varepsilon > 0$ we have

$$\begin{aligned} & P\left[\left\|\left(N_n' N_n\right)^{-} N_n' Y^n - c^n\right\|_2^2 \mathbf{I}\left[\lambda_{\min}(A_n) = 0\right] > \varepsilon\right] \leq P\left[\lambda_{\min}(A_n) = 0\right] \\ & \overset{\text{Lemma 1}}{\leq} P\left[0 \geq \alpha_n - \|A_n - B_n\|_\infty\right] = P\left[\frac{1}{\alpha_n}\|A_n - B_n\|_\infty \geq 1\right] \end{aligned}$$

and $\frac{1}{\alpha_n}\|A_n - B_n\|_\infty \xrightarrow[n\to\infty]{co.} 0$ by Lemma 2 a). Finally, the second term of (13) can be majorized as following:

$$\begin{aligned} & \left\|\left(N_n' N_n\right)^{-1} N_n' Y^n - c^n\right\|_2^2 \mathbf{I}\left[\lambda_{\min}(A_n) > 0\right] \\ & = \left\|A_n^{-1} \cdot \frac{1}{n} N_n' \left(Y^n - N_n c^n\right)\right\|_2^2 \mathbf{I}\left[\lambda_{\min}(A_n) > 0\right] \\ & \leq \left[\alpha_n \|A_n^{-1}\|_2\right]^2 \frac{1}{n^2 \alpha_n^2} \|N_n' \left(Y^n - N_n c^n\right)\|_2^2 \mathbf{I}\left[\lambda_{\min}(A_n) > 0\right]. \end{aligned}$$

According to Lemma 2 b), it remains to show that

$$\frac{1}{n^2 \alpha_n^2} \|N_n' \left(Y^n - N_n c^n\right)\|_2^2 \xrightarrow[n\to\infty]{a.s.} 0.$$

To see this, we shall use the following notations:

$$\begin{aligned} g_n &= \left(g(X_1)\mathbf{I}_{K_n}(X_1), ..., g(X_n)\mathbf{I}_{K_n}(X_n)\right)' \\ g_n^r &= \left(g^{q(n)}(X_1), ..., g^{q(n)}(X_n)\right)', \end{aligned}$$

where $g^{q(n)}(x) = g(x)\mathbf{I}_{K_n}(x) - \sum\limits_{j=1}^{q(n)} c_j^n e_j^n(x)$. Then $g_n = N_n c^n + g_n^r$ and thus

$$\frac{1}{n^2\alpha_n^2}\left\|N_n'(Y^n - N_n c^n)\right\|_2^2 \leq \frac{1}{n^2\alpha_n^2}\left\|N_n'(Y^n - g_n)\right\|_2^2 + \frac{1}{n^2\alpha_n^2}\left\|N_n' g_n^r\right\|_2^2. \tag{14}$$

We will prove the almost sure convergence to zero for each term of this formula. Let us first compute:

$$\begin{aligned}\left\|N_n' g_n^r\right\|_2^2 &\leq \left[\sum_{k=1}^{n}\left(g^{q(n)}(X_k)\right)^2\right]\cdot trace\left(N_n' N_n\right)\\ &\leq nq(n)\left[\sum_{k=1}^{n}\left(g^{q(n)}(X_k)\right)^2\right]\cdot\lambda_{\max}(A_n)\end{aligned}$$

and, therefore, using (12) and Lemma 1, this gives the following majorization:

$$\frac{1}{n^2\alpha_n^2}\left\|N_n' g_n^r\right\|_2^2 \leq \frac{q(n)}{n\alpha_n^2}\left[\sum_{k=1}^{n}\left(g^{q(n)}(X_k)\right)^2\right]\left[\left\|A_n - B_n\right\|_\infty + C_1\right].$$

By Lemma 2 a), $\left\|A_n - B_n\right\|_\infty \xrightarrow[n\to\infty]{a.s.} 0$, therefore it remains to be proven that

$$\frac{q(n)}{n\alpha_n^2}\left[\sum_{k=1}^{n}\left(g^{q(n)}(X_k)\right)^2\right] \xrightarrow[n\to\infty]{a.s.} 0, \tag{15}$$

to get the almost sure convergence to zero for the second term of (14).

Let us now denote $\zeta_k^n = \left(g^{q(n)}(X_k)\right)^2$, $\forall k = \overline{1,n}.$ Then

$$\mathbf{E}\zeta_k^n = \int\limits_{K_n}\left(g^{q(n)}(x)\right)^2 f(x)\,dx \leq C_1\sum_{j=q(n)+1}^{\infty}\left(c_j^n\right)^2. \tag{16}$$

Hence we obtain:

$$\frac{q(n)}{n\alpha_n^2}\sum_{k=1}^{n}\left(g^{q(n)}(X_k)\right)^2 \leq \frac{q(n)}{n\alpha_n^2}\sum_{k=1}^{n}\left(\zeta_k^n - \mathbf{E}\zeta_k^n\right) + C_1\frac{q(n)}{\alpha_n^2}\sum_{j=q(n)+1}^{\infty}\left(c_j^n\right)^2.$$

The second term converges to zero by (i) and for $\varepsilon > 0$ we have by Hoeffding's inequality:

$$P\left[\frac{q(n)}{n\alpha_n^2}\left|\sum_{k=1}^{n}\left(\zeta_k^n - \mathbf{E}\zeta_k^n\right)\right| > \varepsilon\right] \leq 2\exp\left[-\frac{n\alpha_n^4\varepsilon^2}{2q(n)^2\left(\sup\limits_{x\in K_n}\left|g(x) - \sum\limits_{j=1}^{q(n)} c_j^n e_j^n(x)\right|\right)^4}\right].$$

Using (iii), there exists a constant $\gamma > 0$ such that

$$P\left[\frac{q(n)}{n\alpha_n^2}\left|\sum_{k=1}^{n}\left(\zeta_k^n - \mathbf{E}\zeta_k^n\right)\right| > \varepsilon\right] \leq 2\exp\left[-\gamma\frac{n\alpha_n^4}{q(n)^2}\right],$$

which, combined with (ii) and Borel-Cantelli's lemma, yields the statement (15).

Finally, it remains to prove the almost sure convergence for the first term of (14). For any $\varepsilon > 0$ we have

$$\begin{aligned} P\left[\frac{1}{n^2\alpha_n^2}\left\|N_n'\left(Y^n - g_n\right)\right\|_2^2 > \varepsilon\right] &= P\left[\sum_{i=1}^{q(n)}\left[\sum_{k=1}^{n} e_i^n\left(X_k\right)\left(Y_k - g\left(X_k\right)\mathbf{I}_{K_n}\left(X_k\right)\right)\right]^2 > n^2\alpha_n^2\varepsilon\right] \\ &\leq \sum_{i=1}^{q(n)} P\left[\left|\sum_{k=1}^{n} e_i^n\left(X_k\right)\left(Y_k - g\left(X_k\right)\right)\right| > \frac{n\alpha_n\sqrt{\varepsilon}}{\sqrt{q(n)}}\right]. \end{aligned}$$

The random variables in the last sum are centered and their variance is bounded:

$$\begin{aligned} & \mathbf{V}\left[e_i^n\left(X_k\right)\left(Y_k - g\left(X_k\right)\right)\right] \\ =\ & \mathbf{VE}\left[e_i^n\left(X_k\right)\left(Y_k - g\left(X_k\right)\right)|X_k\right] + \mathbf{EV}\left[e_i^n\left(X_k\right)\left(Y_k - g\left(X_k\right)\right)|X_k\right] \\ =\ & \mathbf{E}\left[\left(e_i^n\left(X_k\right)\right)^2\mathbf{V}\left(Y_k|X_k\right)\right] = \mathbf{E}\left[\left(e_i^n\left(X_k\right)\right)^2\sigma^2\left(X_k\right)\right] \\ \leq\ & \left[\sup_{x\in\mathcal{X}}\sigma^2(x)\right]\int_{K_n}\left(e_i^n(x)\right)^2 f(x)\,dx \leq C_1\left[\sup_{x\in\mathcal{X}}\sigma^2(x)\right]. \end{aligned}$$

Combined with Bernstein's inequality and assumption **H4**, this gives

$$\begin{aligned} & P\left[\frac{1}{n^2\alpha_n^2}\left\|N_n'\left(Y^n - g_n\right)\right\|_2^2 > \varepsilon\right] \leq \\ & \leq 2q(n)\exp\left[-\frac{n^2\alpha_n^2\varepsilon}{q(n)}\cdot\frac{1}{4nC_1\left[\sup_{x\in\mathcal{X}}\sigma^2(x)\right] + 2C_2\omega_n\frac{n\alpha_n\sqrt{\varepsilon}}{\sqrt{q(n)}}}\right]. \end{aligned}$$

Thus, there exists a constant $\gamma > 0$ such that

$$P\left[\frac{1}{n^2\alpha_n^2}\left\|N_n'\left(Y^n - g_n\right)\right\|_2^2 > \varepsilon\right] \leq 2q(n)\exp\left[-\gamma\frac{n\alpha_n^2}{q(n)\,\omega_n}\right],$$

and the last expression is the general term of a convergent series, by hypothesis (ii).

b) For the second part of the theorem, we can write:

$$\begin{aligned} \left\|\widehat{g}_n - g\right\|_{n,\infty} &= \sup_{x\in K_n}\left|\widehat{g}_n(x) - g(x)\right| = \sup_{x\in K_n}\left|\sum_{j=1}^{q(n)}\left(\widehat{c}_j^n - c_j^n\right)e_j^n(x) - g^{q(n)}(x)\right| \\ &\leq \sup_{x\in K_n}\left|g^{q(n)}(x)\right| + \omega_n\sqrt{q(n)}\left[\sum_{j=1}^{q(n)}\left(\widehat{c}_j^n - c_j^n\right)^2\right]^{\frac{1}{2}}. \end{aligned} \quad (17)$$

The first term converges to zero by hypothesis (iii) and it remains to prove that

$$\omega_n^2 q(n) \sum_{j=1}^{q(n)} \left(\widehat{c}_j^n - c_j^n\right)^2$$

converges to zero almost surely. The result can be easily verified by arguments similar to those in the first part of this demonstration. ■

Lemma 3 *If* $\left(e_j^n\right)_{j\geq 1}$ *denotes the Legendre basis on* $K_n = [-k_n, k_n]$ *and* $g \in W^m(\mathcal{X})$, $m \geq 2$, *then:*

a) $\displaystyle\sum_{j=q(n)+1}^{\infty} \left(c_j^n\right)^2 = \mathcal{O}\left(\frac{k_n^{2m}}{q(n)^{2m-1}}\right)$ *as* $n \to \infty$

b) $\displaystyle\sum_{j=1}^{q(n)} \left(c_j^n\right)^2 = \mathcal{O}\left(k_n^{2m}\right)$ *as* $n \to \infty$

c) $\displaystyle\sup_{x\in K_n} \left|\sum_{j=q(n)+1}^{\infty} c_j^n e_j^n(x)\right| = \mathcal{O}\left(\frac{k_n^{m-\frac{1}{2}}}{q(n)^{m-1}}\right)$ *as* $n \to \infty$

Proof. For parts **a)** and **b)** see Delecroix and Protopopescu [6].
c) Given the assumptions above, we can prove that (see again Delecroix and Protopopescu [6])

$$\left|c_{j+1}^n\right| = \mathcal{O}\left(\frac{k_n^m}{j^m}\right)$$

and thus we obtain

$$\begin{aligned}\sup_{x\in K_n} \left|\sum_{j=q(n)+1}^{\infty} c_j^n e_j^n(x)\right| &\leq \sup_{x\in K_n} \sum_{j=q(n)}^{\infty} \left|c_{j+1}^n e_{j+1}^n(x)\right| = \mathcal{O}\left(\sqrt{\frac{q(n)}{k_n}} \sum_{j=q(n)}^{\infty} \frac{k_n^m}{j^m}\right) \\ &= \mathcal{O}\left(\frac{k_n^{m-\frac{1}{2}}}{q(n)^{m-1}}\right).\end{aligned}$$ ■

Lemma 4 *If* $\left(e_j^n\right)_{j\geq 1}$ *denotes the trigonometric basis on* $K_n = [-k_n, k_n]$, $g \in W^m(\mathcal{X})$, $m \geq 2$, *and verifies for each* n *the periodicity condition*

$$g^{(i)}(-k_n) = g^{(i)}(k_n), \text{ for } i = \overline{0, m-1}, \tag{18}$$

then:

a) $\sum_{j=q(n)+1}^{\infty} \left(c_j^n\right)^2 = \mathcal{O}\left(\frac{k_n^{2m+1}}{q(n)^{2m-1}}\right)$ *as* $n \to \infty$

b) $\sum_{j=1}^{q(n)} \left(c_j^n\right)^2 = \mathcal{O}\left(k_n^{2m+1}\right)$ *as* $n \to \infty$

c) $\sup_{x \in K_n} \left| \sum_{j=q(n)+1}^{\infty} c_j^n e_j^n(x) \right| = \mathcal{O}\left(\frac{k_n^m}{q(n)^{m-1}}\right)$ *as* $n \to \infty$

Proof. For parts **a)** and **b)** see Delecroix and Protopopescu [6].
c) Under the assumptions above, we can prove that (see again Delecroix and Protopopescu [6])

$$\left|c_j^n\right| = \mathcal{O}\left(\frac{k_n^{m+\frac{1}{2}}}{j^m}\right)$$

and thus we obtain

$$\sup_{x \in K_n} \left| \sum_{j=q(n)+1}^{\infty} c_j^n e_j^n(x) \right| = \mathcal{O}\left(\frac{1}{\sqrt{k_n}} \sum_{j=q(n)+1}^{\infty} \frac{k_n^{m+\frac{1}{2}}}{j^m}\right) = \mathcal{O}\left(\frac{k_n^m}{q(n)^{m-1}}\right). \quad \blacksquare$$

Proof of Theorem 2.
The proof of consistency for the wavelet version of the estimator follows the proof of Theorem 1 given in the general case of an arbitrary orthonormal basis, with simple modifications. Essentially, one replaces in the demonstration the number of terms in the series, $q(n)$, by $2^{dq(n)}$ (here the constant has been omitted) and uses the uniform bound given by $\omega_n = C_1^* 2^{dq(n)/2}$ and formula (8).

First, the matrices A_n and B_n, defined like in the case of general orthonormal bases,

$$A_n = \left(a^n_{i\lambda_1, j\lambda_2}\right), \text{ with } a^n_{i\lambda_1, j\lambda_2} = \frac{1}{n}\sum_{k=1}^{n} \psi^n_{i,k_1}(X_k)\, \psi^n_{j,k_2}(X_k)\, \mathbf{I}_{K_n}(X_k)$$

$$B_n = \left(b^n_{i\lambda_1, j\lambda_2}\right), \text{ with } b^n_{i\lambda_1, j\lambda_2} = \int_{K_n} \psi^n_{i,k_1}(x)\, \psi^n_{j,k_2}(x)\, f(x)\, dx,$$

satisfy the convergence properties of Lemma 2, for any sequence $q(n)$ that verifies assumption (ii) of Theorem 2.
a) The following similar formula holds for the integrated squared error criterion:

$$\|\widehat{g}_n - g\|^2_{n,2} = \sum_{j=q(n)+1}^{\infty} \sum_{k \in R_j^n} \left(c^n_{j,k}\right)^2 + \sum_{j=j_o(n)}^{q(n)} \sum_{k \in R_j^n} \left(\widehat{c}^n_{j,k} - c^n_{j,k}\right)^2. \tag{19}$$

The first term (which is non random) converges to zero because, using (8), we have:

$$\sum_{j=q(n)+1}^{\infty} \sum_{k \in R_j^n} \left(c^n_{j,k}\right)^2 \leq \sum_{j=q(n)+1}^{\infty} card\left(R_j^n\right) \cdot \left(C_* 2^{-dj/2} 2^{-jm}\right)^2 \tag{20}$$

$$\leq MC_*^2 \sum_{j=q(n)+1}^{\infty} 2^{-2jm} = \frac{MC_*^2}{\left(2^{2m}-1\right) 2^{2mq(n)}} \underset{n\to\infty}{\longrightarrow} 0.$$

By following a similar computation as that given in the proof of Theorem 1, the second term of (19) is bounded by:

$$\sum_{j=j_o(n)}^{q(n)} \sum_{k\in R_j^n} \left(\widehat{c}_{j,k}^n - c_{j,k}^n\right)^2 \leq \left[\alpha_n \left\|A_n^{-1}\right\|_2\right]^2 \frac{1}{n^2\alpha_n^2} \left\|N_n'\left(Y^n - N_n c^n\right)\right\|_2^2 \mathbf{I}\left[\lambda_{\min}\left(A_n\right) > 0\right] + \left\|\left(N_n' N_n\right)^- N_n' Y^n - c^n\right\|_2^2 \mathbf{I}\left[\lambda_{\min}\left(A_n\right) = 0\right]. \quad (21)$$

Using the results of Lemma 2 a) and b) as well as the proof of Theorem 1, in order to get its almost sure consistency it is sufficient to show that

$$\frac{1}{n^2\alpha_n^2} \left\|N_n'\left(Y^n - N_n c^n\right)\right\|_2^2 \xrightarrow[n\to\infty]{a.s.} 0.$$

With the notations of Theorem 1, we can write in a similar fashion:

$$\frac{1}{n^2\alpha_n^2} \left\|N_n'\left(Y^n - N_n c^n\right)\right\|_2^2 \leq \frac{1}{n^2\alpha_n^2} \left\|N_n'\left(Y^n - g_n\right)\right\|_2^2 + \frac{2M2^{dq(n)}}{n\alpha_n^2} \left[\sum_{k=1}^n \left(g^{q(n)}\left(X_k\right)\right)^2\right] \left[\left\|A_n - B_n\right\|_\infty + C_1\right].$$

Since $\left\|A_n - B_n\right\|_\infty \xrightarrow[n\to\infty]{a.s.} 0$ (by Lemma 2 b)), for the second term of the right hand side it remains to prove that

$$\frac{2^{dq(n)}}{n\alpha_n^2} \left[\sum_{k=1}^n \left(g^{q(n)}\left(X_k\right)\right)^2\right] \xrightarrow[n\to\infty]{a.s.} 0.$$

Let denote $\zeta_k^n = \left(g^{q(n)}\left(X_k\right)\right)^2$, $\forall k = \overline{1,n}$. Then

$$\begin{aligned}\frac{2^{dq(n)}}{n\alpha_n^2} \left[\sum_{k=1}^n \left(g^{q(n)}\left(X_k\right)\right)^2\right] &\leq \frac{2^{dq(n)}}{n\alpha_n^2} \sum_{k=1}^n \left(\zeta_k^n - E\zeta_k^n\right) + C_1 \frac{2^{dq(n)}}{\alpha_n^2} \sum_{j=q(n)+1}^{\infty} \sum_{k\in R_j^n} \left(c_{j,k}^n\right)^2 \\ &\leq \frac{2^{dq(n)}}{n\alpha_n^2} \sum_{k=1}^n \left(\zeta_k^n - E\zeta_k^n\right) + C_1 M C_*^2 \frac{2^{dq(n)}}{\alpha_n^2\left(2^{2m}-1\right)2^{2mq(n)}}.\end{aligned}$$

The second term converges to zero by (i) and for $\varepsilon > 0$ we have by Hoeffding's inequality:

$$P\left[\frac{2^{dq(n)}}{n\alpha_n^2} \left|\sum_{k=1}^n \left(\zeta_k^n - E\zeta_k^n\right)\right| > \varepsilon\right] \leq 2\exp\left[-\frac{n\alpha_n^4\varepsilon^2}{2\cdot 2^{2dq(n)} \left(\sup_{x\in K_n} \left|g(x) - \sum_{j=j_o(n)}^{q(n)} \sum_{k\in R_j^n} c_{j,k}^n \psi_{j,k}^n(x)\right|\right)^4}\right].$$

But

$$\sup_{x\in K_n}\left|g(x)-\sum_{j=j_o(n)}^{q(n)}\sum_{k\in R_j^n}c_{j,k}^n\psi_{j,k}^n(x)\right| \le \sup_{x\in K_n}\left|\sum_{j=q(n)+1}^{\infty}\sum_{k\in R_j^n}c_{j,k}^n\psi_{j,k}^n(x)\right|$$
$$\le \sum_{j=q(n)+1}^{\infty}\sum_{k\in R_j^n}C_*2^{-dj/2}2^{-jm}\cdot C_1^*2^{dj/2} \le C_*C_1^*\sum_{j=q(n)+1}^{\infty} card\left(R_j^n\right)2^{-jm}$$
$$\le MC_*C_1^*\sum_{j=q(n)+1}^{\infty}2^{-(m-d)j} = \frac{MC_*C_1^*}{\left(2^{m-d}-1\right)2^{(m-d)q(n)}}.$$

Notice that assumption (i) implies that $3d-2m<0$ and thus $m-d>0$. In conclusion, there exists a constant $\gamma>0$ such that

$$P\left[\frac{2^{dq(n)}}{n\alpha_n^2}\left|\sum_{k=1}^{n}\left(\zeta_k^n-E\zeta_k^n\right)\right|>\varepsilon\right]\le 2\exp\left[-\gamma\frac{n\alpha_n^4}{2^{(6d-4m)q(n)}}\right]\le 2\exp\left[-\gamma\frac{n\alpha_n^4}{2^{4dq(n)}}\right],$$

which, combined with (ii), proves the required statement.

Finally, it remains to prove that $\frac{1}{n^2\alpha_n^2}\left\|N_n'\left(Y^n-g_n\right)\right\|_2^2 \xrightarrow[n\to\infty]{a.s.} 0$. For $\varepsilon>0$ we have by Bernstein's inequality:

$$P\left[\frac{1}{n^2\alpha_n^2}\left\|N_n'\left(Y^n-g_n\right)\right\|_2^2>\varepsilon\right]$$
$$\le \sum_{j=q(n)+1}^{\infty}\sum_{k\in R_j^n}P\left[\left|\sum_{k=1}^{n}\psi_{i,k}^n\left(X_k\right)\left(Y_k-g\left(X_k\right)\mathbf{I}_{K_n}\left(X_k\right)\right)\right|>\frac{n\alpha_n\sqrt{\varepsilon}}{\sqrt{q(n)}}\right]$$
$$\le 4M2^{dq(n)}\exp\left[-\frac{n\alpha_n^2\varepsilon}{8MC_1\left[\sup_{x\in\mathcal{X}}\sigma^2(x)\right]2^{dq(n)}+2C_2C_1^*\sqrt{2M}2^{dq(n)}\alpha_n\sqrt{\varepsilon}}\right]$$
$$\le 4M2^{dq(n)}\exp\left[-\gamma\frac{n\alpha_n^2}{2^{dq(n)}}\right],\text{ with }\gamma>0,$$

which is the general term of a convergent series, by hypothesis (ii).

b) As in the case of a general orthonormal basis, we obtain:

$$\left\|\widehat{g}_n-g\right\|_{n,\infty} \le \sup_{x\in K_n}\left|g^{q(n)}(x)\right|+C_1^*2^{dq(n)/2}\sum_{j=j_o(n)}^{q(n)}\sum_{k\in R_j^n}\left|\widehat{c}_{j,k}^n-c_{j,k}^n\right|$$
$$\le MC_*C_1^*\frac{1}{2^{(m-d)(1+q(n))}}+C_1^*\sqrt{2M}2^{dq(n)}\left[\sum_{j=j_o(n)}^{q(n)}\sum_{k\in R_j^n}\left(\widehat{c}_{j,k}^n-c_{j,k}^n\right)^2\right]^{\frac{1}{2}}.$$

The first term converges to zero since $m-d>0$, so it remains to prove that the second one converges to zero almost surely. It is easy to see that we only have to repeat the arguments of the first part of the demonstration. Indeed, the almost sure convergence of the two terms of (21) multiplied by $2^{2dq(n)}$ will be verified under our hypothesis. ■

Proof of Theorem 3.
This is a refinement of the corresponding proof in Newey [14, Theorem 1]. To get the result, the majorization (13) in the proof of Theorem 2 will be used. When n is large, the third term of (13) becomes negligible with respect to the first, i.e. it is an $\mathcal{O}_P\left(\sum_{j=q(n)+1}^{\infty}\left(c_j^n\right)^2\right)$, since for a given $\varepsilon > 0$, we can write

$$P\left\{\left[\sum_{j=q(n)+1}^{\infty}\left(c_j^n\right)^2\right]^{-1}\left\|\left(N_n'N_n\right)^{-}N_n'Y^n - c^n\right\|_2^2 \mathbf{I}\left[\lambda_{\min}\left(A_n\right)=0\right] > \varepsilon\right\}$$
$$\leq P\left[\lambda_{\min}\left(A_n\right)=0\right] \leq P\left[\frac{1}{\alpha_n}\left\|A_n - B_n\right\|_\infty \geq 1\right]$$

and the last expression is the general term of a convergent series. Therefore,

$$\left\|\widehat{g}_n - g\right\|_{n,2}^2 \leq \mathcal{O}_P\left(\sum_{j=q(n)+1}^{\infty}\left(c_j^n\right)^2\right) + \left\|\left(N_n'N_n\right)^{-1}N_n'Y^n - c^n\right\|_2^2 \mathbf{I}\left[\lambda_{\min}\left(A_n\right)>0\right]. \quad (22)$$

Using the notations in the proof of Theorem 1, the second term in the latter formula can be rewritten as

$$\begin{aligned}
&\left\|\left(N_n'N_n\right)^{-1}N_n'Y^n - c^n\right\|_2^2 \mathbf{I}\left[\lambda_{\min}\left(A_n\right)>0\right]\\
=\ &\left\|\left(N_n'N_n\right)^{-1}N_n'\left(Y^n - N_nc^n\right)\right\|_2^2 \mathbf{I}\left[\lambda_{\min}\left(A_n\right)>0\right]\\
\leq\ &\left\{\left\|\left(N_n'N_n\right)^{-1}N_n'\left(Y^n - g_n\right)\right\|_2^2 + \left\|\left(N_n'N_n\right)^{-1}N_n'g_n^r\right\|_2^2\right\}\mathbf{I}\left[\lambda_{\min}\left(A_n\right)>0\right]. \quad (23)
\end{aligned}$$

Let us now compute

$$\begin{aligned}
\left\|\left(N_n'N_n\right)^{-1}N_n'g_n^r\right\|_2^2 &= g_n^{r\prime}N_n\left(N_n'N_n\right)^{-1}\left(N_n'N_n\right)^{-1}N_n'g_n^r\\
&\leq \lambda_{\max}\left(N_n'N_n\right)^{-1}g_n^{r\prime}N_n\left(N_n'N_n\right)^{-1}N_n'g_n^r\\
&\leq \lambda_{\max}\left(N_n'N_n\right)^{-1}\lambda_{\max}\left[N_n\left(N_n'N_n\right)^{-1}N_n'\right]\sum_{k=1}^{n}\left(g^{q(n)}\left(X_k\right)\right)^2.
\end{aligned}$$

Since the matrix $N_n\left(N_n'N_n\right)^{-1}N_n'$ is idempotent, this gives

$$\left\|\left(N_n'N_n\right)^{-1}N_n'g_n^r\right\|_2^2 \leq \frac{1}{n}\left\|A_n^{-1}\right\|_2\sum_{k=1}^{n}\left(g^{q(n)}\left(X_k\right)\right)^2$$
$$\leq \left[\alpha_n\left\|A_n^{-1}\right\|_2\right]\frac{1}{n\alpha_n}\sum_{k=1}^{n}\left(g^{q(n)}\left(X_k\right)\right)^2 \overset{\text{Lemma 2 b)}}{=} \mathcal{O}_P(1)\frac{1}{n\alpha_n}\sum_{k=1}^{n}\left(g^{q(n)}\left(X_k\right)\right)^2$$

and, using (16), we obtain

$$\left\|\left(N_n'N_n\right)^{-1}N_n'g_n^r\right\|_2^2 \mathbf{I}\left[\lambda_{\min}\left(A_n\right)>0\right] = \mathcal{O}_P\left(\frac{1}{\alpha_n}\sum_{j=q(n)+1}^{\infty}\left(c_j^n\right)^2\right). \quad (24)$$

Finally, for the first term of (23) we can write

$$\begin{aligned}
&\left\|(N_n'N_n)^{-1} N_n' (Y^n - g_n)\right\|_2^2 \leq (Y^n - g_n)' N_n (N_n'N_n)^{-1} (N_n'N_n)^{-1} N_n' (Y^n - g_n) \\
&\quad \leq \lambda_{\max} (N_n'N_n)^{-1} (Y^n - g_n)' N_n (N_n'N_n)^{-1} N_n' (Y^n - g_n) \\
&\quad = [\alpha_n \|A_n^{-1}\|_2] \frac{1}{n\alpha_n} (Y^n - g_n)' N_n (N_n'N_n)^{-1} N_n' (Y^n - g_n) \qquad (25) \\
&\quad = \mathcal{O}_P \left(\frac{1}{n\alpha_n}\right) (Y^n - g_n)' N_n (N_n'N_n)^{-1} N_n' (Y^n - g_n).
\end{aligned}$$

Let us now denote $X^n = (X_1, ..., X_n)'$. Then

$$\begin{aligned}
& \mathbf{E}\left[(Y^n - g_n)' N_n (N_n'N_n)^{-1} N_n' (Y^n - g_n) | X^n\right] \\
= \; & \mathbf{E}\left\{trace\left[N_n (N_n'N_n)^{-1} N_n' (Y^n - g_n)(Y^n - g_n)'\right] | X^n\right\} \\
= \; & trace\left\{N_n (N_n'N_n)^{-1} N_n' \mathbf{E}\left[(Y^n - g_n)(Y^n - g_n)' | X^n\right]\right\} \\
\leq \; & \left[\sup_{x\in\mathcal{X}} \sigma^2(x)\right] trace\left\{N_n (N_n'N_n)^{-1} N_n'\right\} \leq q(n) \left[\sup_{x\in\mathcal{X}} \sigma^2(x)\right]
\end{aligned}$$

and, combined with (25), this gives

$$\left\|(N_n'N_n)^{-1} N_n' (Y^n - g_n)\right\|_2^2 \mathbf{I}[\lambda_{\min}(A_n) > 0] = \mathcal{O}_P\left(\frac{q(n)}{n\alpha_n}\right). \qquad (26)$$

From inequalities (23), (24) and (26), we obtain finally

$$\left\|(N_n'N_n)^{-1} N_n' Y^n - c^n\right\|_2^2 \mathbf{I}[\lambda_{\min}(A_n) > 0] = \mathcal{O}_P\left(\frac{1}{\alpha_n} \sum_{j=q(n)+1}^{\infty} (c_j^n)^2\right) + \mathcal{O}_P\left(\frac{q(n)}{n\alpha_n}\right)$$

and, comparing with (22), this proves the first part of the theorem.

The $\mathcal{O}_P$ bound of $\|\widehat{g}_n - g\|_{n,\infty}$ follows then straightforwardly, using inequality (17). ■

References

[1] D. W. K. Andrews, Asymptotic normality of series estimators for nonparametric and semiparametric regression models, *Econometrica* **59** (1991), 307-345.

[2] A. Antoniadis, G. Gregoire & I. W. McKeague, Wavelet methods for curve estimation, *J. Amer. Statist. Assoc.* **89** (1994), 1340-1353.

[3] P. Billingsley, "Convergence of Probability Measures", Wiley, 1968.

[4] D. Bosq, "Nonparametric Statistics for Stochastic Processes", Springer Verlag, New York, 1996.

[5] D. Bosq & J. P. Lecoutre, "Théorie de l'estimation fonctionnelle", Economica, 1987.

[6] M. Delecroix & C. Protopopescu, Regression Series Estimators: the MISE Approach, preprint GREMAQ, Université Toulouse I, 1998.

[7] R. L. Eubank, "Spline Smoothing and Nonparametric Regression", Marcel Dekker, New York, 1988.

[8] R. L. Eubank & P. Speckman, Convergence rates for trigonometric and polynomial-trigonometric regression estimators, *Statist. and Prob. Letters* **11** (1991), 119-124.

[9] W. Härdle, "Applied Nonparametric Regression", Cambridge Univ. Press, Cambridge, 1990.

[10] S. Jaffard & Y. Meyer, Bases d'ondelettes dans des ouverts de $\mathbb{R}^n$, *J. Math. Pure et Appl.* **68** (1989), 95-108.

[11] P. Lascaux & R. Théodor, "Analyse numérique matricielle appliquée à l'art de l'ingénieur", Tome 1- Méthodes directes, 1993.

[12] G. Lugosi & K. Zeger, Nonparametric estimators via empirical risk minimization, *IEEE Trans. Inform. Theory* **41** (1995), 677-687.

[13] Y. Meyer, "Ondelettes", Hermann, 1990.

[14] W. Newey, Convergence rates and asymptotic normality for series estimators, *J. of Econometrics* **79** (1997), 147-168.

[15] G. Sansone, "Orthogonal Functions", Interscience Publishers, Inc., New York, 1959.

[16] X. Shen, On Methods of Sieves and Penalization, *Ann. Statist.* **25** (1997), 2555-2591.

[17] X. Shen & W. H. Wong, Convergence rates of sieve estimates, *Ann. Statist.* **22** (1994), 580-615.

[18] C. J. Stone, Optimal global rates of convergence for nonparametric regression, *Ann. Statist.* **10** (1982), 1040-1053.

[19] G. Szegö, "Orthogonal Polynomials", American Mathematical Society Colloquium Publications, Vol. XXIII, New York, 1939.