Ann. Inst. Statist. Math.
Vol. 53, No. 4, 681–690 (2001)

ON NON-EQUALLY SPACED WAVELET REGRESSION

MARIANNA PENSKY[1] AND BRANI VIDAKOVIC[2*]

[1] *Department of Mathematics, University of Central Florida, Orlando, FL 32816, U.S.A.*
[2] *Institute of Statistics and Decision Sciences, Duke University, Box 90251, Durham, NC 27708-0251, U.S.A.*

(Received December 14, 1998; revised December 13, 1999)

Abstract. Wavelet-based regression analysis is widely used mostly for equally-spaced designs. For such designs wavelets are superior to other traditional orthonormal bases because of their versatility and ability to parsimoniously describe irregular functions. If the regression design is random, an automatic solution is not available. For such non equispaced designs we propose an estimator that is a projection onto a multiresolution subspace in an associated multiresolution analysis. For defining scaling empirical coefficients in the proposed wavelet series estimator our method utilizes a probabilistic model on the design of independent variables. The paper deals with theoretical aspects of the estimator, in particular MSE convergence rates.

Key words and phrases: Irregular design, NES regression, nonparametric statistical procedures, projection estimators, wavelets.

1. Introduction

Function estimation is of fundamental importance in statistics and science in general where it applies to a wide range of problems and has a multitude of different objectives. For the majority of data sets encountered in real life the most appropriate procedures are nonparametric. Wavelet-based non-parametric methods, introduced in statistics by the work of Donoho and Johnstone in early 90's, represent a novel, break-through technology in theory and practice of nonparametric function estimation. The benefit of wavelets is their ability to adapt to unknown smoothness (Donoho *et al.* (1995)), although the adaptivity can be obtained by using traditional kernel estimators (Lepski *et al.* (1997)). For the equally spaced observations (for example, measurements at equal time increments), Donoho and Johnstone developed a simple and adaptive procedure, called *WaveShrink*, based on discrete wavelet transformations. *WaveShrink* is a fast procedure that has very broad asymptotic near-optimal properties.

Generalizations of WaveShrink-type techniques to non-equally spaced (NES) designs impose additional conceptual and calculational burdens. There are several proposals on how to estimate regression function by wavelets when a design is irregular. The simplest proposal to ignore the design and is to carry out the analysis as if the data were equally spaced. This "method" is known as the *coercion to equal spacing*.

Another class of wavelet-based methods applicable on non-equally spaced data utilize interpolations and averaging. Based on the available data, the approximations (inter-

*Now at Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, U.S.A. The research of the second author was partially supported by NSF Grant DMS-9626159.

polations) of the function are done at equally spaced dyadic points. On such approximate values the standard wavelet methods are applied. References on that method include Antoniadis *et al.* (1997), Deslauriers and Dubuc (1989), Foster (1996), Hall and Turlach (1997), and Härdle *et al.* (1998), among others. Sweldens (1995) proposes the "lifting scheme" technique to define wavelet transform of data sampled over a variety of topological objects. Sardy *et al.* (1999) consider the Haar basis and propose four approaches that extend the Haar wavelet transform to a NES data. Each approach is formulated in terms of continuous wavelet functions applied to a piecewise constant interpolation of the observed data, and each approach leads to wavelet coefficients that can be computed via a matrix transform of the original data. Some related approaches can be found in Antoniadis *et al.* (1994), Antoniadis and Pham (1998), Cai and Brown (1998), Delyon and Juditsky (1995), and Hall *et al.* (1998).

In this paper we propose a linear wavelet-based regression estimator and explore some of its large sample properties. The proposed estimator can be re-expressed in a form which reminds the wavelet modification of Nadaraja-Watson estimator (Antoniadis *et al.* (1994)) but conditions on the "pre-estimator" of the design distribution are relaxed. The practical implementations are subject of ongoing simulational study.

2. The estimator and its large sample properties

Let ϕ be a compactly supported scaling function generated by an s-regular multiresolution analysis (MRA) of $\mathbb{L}_2(\mathbb{R})$. Assume that $s \geq 2$, and that the support of ϕ is contained in the interval $[-A, A]$.

Let

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n), \tag{2.1}$$

be a sample of size n. Denote by f the marginal density of X and by $m(x) = E(Y \mid X = x)$ the regression function to be estimated. Assume that $m(x) \in \mathbb{L}^2(-\infty, \infty)$.

Instead of estimating $m(x)$ directly we estimate its projection on a multiresolution subspace V_J, $m_J(x) = \mathrm{Proj}_{V_J}\, m(x) = \sum_{k \in \mathbb{Z}} c_{J,k}\phi_{J,k}(x)$, where $\phi_{J,k}(x) = 2^{J/2}\phi(2^J x - k)$. By the properties of MRA, $\cup_j V_j$ is dense in $\mathbb{L}_2(\mathbb{R})$ and the linear approximation $m_J(x)$ converges to $m(x)$ uniformly on compacts, when $J \to \infty$.

The coefficients are

$$\begin{aligned} c_{J,k} &= 2^{J/2}\int_{-\infty}^{\infty} \phi(2^J x - k)\, m(x) dx \\ &= \int_{-\infty}^{\infty} \frac{2^{J/2}\phi(2^J x - k)\, m(x)}{f(x)} f(x) dx \\ &= \mathbb{E}\left[\frac{\phi_{J,k}(X)\, Y}{f(x)}\right]. \end{aligned} \tag{2.2}$$

Consider an estimator $\hat{f}_n(x)$ of $f(x)$ constructed on basis of $n-1$ observations,

$$\hat{f}_n(x) = \hat{f}_n(x \mid Z_1, \ldots, Z_{n-1})$$

which is symmetric with respect to $Z_1, \ldots, Z_{n-1}$, that is,

$$\hat{f}_n(x \mid Z_1, \ldots, Z_{n-1}) = \hat{f}_n(x \mid Z_{i_1}, \ldots, Z_{i_{n-1}}).$$

Let $\hat{m}_n(x)$ be an estimator of $m(x)$, more precisely, of $\mathrm{Proj}_{V_J} m(x)$,

$$\hat{m}_n(x) = \sum_k \hat{c}_{J,k}\phi_{J,k}(x), \tag{2.3}$$

with $\hat{c}_{J,k}$ motivated by (2.2),

$$\hat{c}_{J,k} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\phi_{J,k}(X_i)Y_i \cdot \mathbf{1}(\hat{f}_n(X_i \mid X_{-i}) \geq \delta_n)}{\hat{f}_n(X_i \mid X_{-i})}\right], \tag{2.4}$$

where X_{-i} is the sample with i-th observation excluded, i.e., $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$. We will call estimator in (2.3) the NES linear wavelet estimator. Estimator similar to (2.3) was recently proposed by Pinheiro (1997). He proved that under mild regularity conditions on $\hat{f}_n$, $\hat{c}_{J,k} \to c_{J,k}$, *a.s.*

To derive quantitative results, we now discuss the MSE convergence rates of the NES estimator. Denote $v(x) = \mathbb{E}(Y^2 \mid X)$ and notice that $v(x) - m^2(x) \geq 0$. Let us consider the class of linear estimators of the density $f(x)$, that is, estimators of the form

$$\hat{f}_n(x \mid X_{-i}) = \frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n} \mathbb{K}_n(x, X_j), \tag{2.5}$$

where $\mathbb{K}_n(x,y)$ is a bounded kernel, symmetric in its arguments. Assume the following conditions hold:

A1. $\|f\|_c = \sup_x f(x) < \infty$;

A2. $\sup_x(\frac{v(x)}{f(x)}) < V_1 < \infty$, $\sup_x(\frac{m^2(x)}{f^2(x)}) < V_2 < \infty$;

A3. the kernel $\mathbb{K}_n(x,y)$ satisfy the conditions

$$\text{(a)}\ \sup_{x,y}|\mathbb{K}_n(x,y)| \leq C_1 n^{1/(2\gamma+1)}, \qquad \text{(b)}\ \sup_x \mathbb{E}\left[\mathbb{K}_n^2(x, X_1)\right] \leq C_2 n^{1/(2\gamma+1)};$$

A4. the estimator (2.5) achieves the following convergence rate

$$\sup_x \mathbb{E}(\hat{f}_n(x) - f(x))^2 dx \leq C_3 n^{-2\gamma/(2\gamma+1)}, \ \sup_x \mathbb{E}(\hat{f}_n(x) - f(x))^4\, dx \leq C_4 n^{-(4\gamma-1)/(2\gamma+1)},$$

with $\gamma \geq 1$;

A5. $m(x)$ is s times continuously differentiable with $|m^{(s)}(x)| \leq m_s < \infty$.

Denote $\|\phi\|_c = \sup_x \phi(x)$, $\|\phi\|_L = \int |\phi(x)| dx$. The main result of this paper is expressed in the following theorem.

THEOREM 2.1. *Let the conditions* A1–A5 *hold, and* δ_n *be such that*

$$\delta_n = \delta_0\, n^{-\gamma/(2\gamma+1)} \ln n, \tag{2.6}$$

with $\delta_0 \geq 4C_1$. *Then*

$$\sup_x \mathbb{E}(\hat{m}_n(x) - m(x))^2 \leq \begin{cases} (C_5 + C_7)n^{-4\gamma s/(2\gamma+1)(2s+1)}(\ln n)^{4s/(2s+1)}(1 + o(1)), \\ \qquad\qquad \text{if} \quad 2\gamma \geq 2s+1 \\ (C_6 + C_7)n^{-(2\gamma-1)/(2\gamma+1)}(\ln n)^{-2}(1 + o(1)), \\ \qquad\qquad \text{if} \quad 2\gamma < 2s+1, \end{cases}$$

provided

$$2^J \sim \begin{cases} n^{2\gamma/(2\gamma+1)(2s+1)}(\ln n)^{-2/(2s+1)}, & \text{if } \ 2\gamma \geq 2s+1 \\ n^{(2\gamma-1)/(2s(2\gamma+1))}(\ln n)^{1/s}, & \text{if } \ 2\gamma < 2s+1. \end{cases} \tag{2.7}$$

Here $C_5 = 24(2A+1)^2\|\phi\|_c^2(C_4+C_1^4)V_2\delta_0^{-2}\|\phi\|_c^2$, $C_6 = 27/2\cdot(2A+1)^2\|\phi\|_L^2V_2^2\delta_0^2$, *and* $C_7 = (2A+1)^2\|\phi\|_c^2\|\phi\|_L^2(s!)^{-2}m_s^2 2^{2s}A^{2s}$.

The proof of Theorem 2.1 is given in Section 3.

It is easy to see that conditions A3 and A4 are satisfied if $\hat{f}_n(x)$ is a kernel density estimator or a linear wavelet estimator provided $f(x)$ is γ times continuously differentiable with $\sup_x |f^{(\gamma)}(x)| < \infty$. The choice of the parameter δ_n is determined by γ and C_1 where C_1 depends on $K_n(x,y)$ which is known.

From Theorem 2.1 it follows that the rate of convergence of the estimator $\hat{m}_n(x)$ is determined significantly by γ. If γ is fixed, then the rate of convergence of $\hat{m}_n(x)$ is $O(n^{-(2\gamma-1)/(2\gamma+1)}(\ln n)^{-2})$ for any $s > \gamma - 0.5$. Hence, even infinite grows of s does not improve convergence rate. If s is fixed, the rate of convergence grows slowly with the increase of γ and reaches $O(n^{-(4\gamma s)/[(2\gamma+1)(2s+1)]}(\ln n)^{4s/(2s+1)})$ as $\gamma \to \infty$. Therefore, the estimator (2.3) should be applied only if we expect $f(x)$ to be reasonably smooth, that is, when γ is sufficiently large.

3. Proof of Theorem 2.1

We give now a proof of Theorem 2.1. This proof is based on a series of auxiliary lemmas. To simplify the notations, in what follows we will denote $\hat{f}_n(X_i) \equiv \hat{f}_n(X_i \mid X_{-i})$, $\hat{f}_n(x) \equiv \hat{f}_n(x \mid X_{-n})$. Also, the index n in J_n will be suppressed and we will simply write J. However, the level J is a function of a sample size as conditioned in (2.6).

LEMMA 1. *Under the conditions* A1–A4

$$\begin{aligned} &\sup_{X_1,X_2} \mathbb{E}[(\hat{f}_n(X_1)-f(X_1))^2 \mid X_1,X_2] \leq 2(C_3+C_1^2)n^{-2\gamma/(2\gamma+1)}(1+o(1)), \\ &\sup_{X_1,X_2} \mathbb{E}[(\hat{f}_n(X_1)-f(X_1))^4 \mid X_1,X_2] \leq 8(C_4+C_1^2)n^{-(4\gamma-1)/(2\gamma+1)}(1+o(1)). \end{aligned} \tag{3.1}$$

PROOF OF LEMMA 1. Let us prove the first assertion. Partitioning $\hat{f}_n(X_1)$ into the part containing and not containing X_2 and applying the inequalities $(a+b)^2 \leq 2a^2+2b^2$ and $(n-2)/(n-1) < 1$, we obtain

$$\begin{aligned} \mathbb{E}[(\hat{f}_n(X_1)-f(X_1))^2 \mid X_1,X_2] \leq 2\int\cdots\int\left(\sum_{k=3}^{n}\frac{\mathbb{K}_n(X_1,x_k)}{n-2}-f(X_1)\right)^2\prod_{k=3}^{n}f(x_k)dx_k \\ +\frac{2}{(n-1)^2}\int\cdots\int(\mathbb{K}_n(X_1,X_2)-f(X_1))^2\prod_{k=3}^{n}f(x_k)dx_k. \end{aligned}$$

Therefore, $\mathbb{E}[(\hat{f}_n(X_1)-f(X_1))^2 \mid X_1,X_2] \leq 2C_3(n-1)^{-2\gamma/(2\gamma+1)}+2(n-1)^{-2}(C_1^2+\|f\|_c^2)$, which implies (3.1). The second inequality in Lemma 1 can be derived in a similar manner.

LEMMA 2. *If assumptions* A1–A4 *and* (2.6) *are valid, then*

$$\sup_x \mathbb{P}(|\hat{f}_n(x) - f(x)| > \delta_n) \le 2n^{-1}.$$

PROOF OF LEMMA 2. According to condition A4, $\sup_x |\mathbb{E}\hat{f}_n(x) - f(x)| \le \sqrt{C_3}n^{-\gamma/(2\gamma+1)}$, which implies that for every x

$$\mathbb{P}(|\hat{f}_n(x) - f(x)| > \delta_n) \le \mathbb{P}(|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| > (\delta_0 \ln n - \sqrt{C_3})n^{-\gamma/(2\gamma+1)}).$$

To majorize the right-hand side of the last inequality, recall Bernstein's inequality (see Pollard (1984)). If $Z_1, Z_2, \ldots, Z_M$ are i.i.d. bounded random variables such that $\mathbb{E}Z_i = 0$, $\mathbb{E}Z_i^2 = \sigma^2$, and $|Z_i| \le \|Z\|_\infty < \infty$, then for any $\lambda > 0$

$$\mathbb{P}\left(\left|M^{-1}\sum_{i=1}^{M} Z_i\right| > \lambda\right) \le 2\exp\left(-\frac{M\,\lambda^2}{2\,\sigma^2 + (2/3)\,\lambda\|Z\|_\infty)}\right). \tag{3.2}$$

Apply (3.2) with $Z_i = \mathbb{K}_n(x, X_i) - \mathbb{E}\mathbb{K}_n(x, X_i)$, $M = n-1$, $\sigma^2 \le C_2 n^{-\gamma/(2\gamma+1)}$, $\|Z\|_\infty \le 2C_1 n^{-\gamma/(2\gamma+1)}$, and $\lambda = (\delta_0 \ln n - \sqrt{C_3})n^{-\gamma/(2\gamma+1)}$. Taking into account that for large n the following inequalities hold: $C_2 + (2/3)C_1(\delta_0 \ln n - \sqrt{C_3})n^{-(\gamma-1)/(2\gamma+1)} \le C_1\delta_0 \ln n$ and $n^{-1}(n-1)(\delta_0 \ln n - \sqrt{C_3})^2 \ge 0.5\delta_0^2 \ln^2 n$, we obtain

$$\mathbb{P}(|\hat{f}_n(x) - f(x)| > \delta_n) \le 2\exp\{-\delta_0 \ln n/(4C_1)\},$$

which completes the proof.

LEMMA 3. *Under the assumptions* A1–A4 *and* (2.6)

$$\sup_k \mathbb{E}(\hat{c}_{J,k} - c_{J,k})^2 \le n^{-2\gamma/(2\gamma+1)}[C_8 n^{1/(2\gamma+1)}2^{-J}(\ln n)^{-2} + C_9(\ln n)^2](1 + o(1)), \tag{3.3}$$

where $C_8 = 24(C_4 + C_1^4)V_2\delta_0^{-2}\|\phi\|_L^2$ *and* $C_9 = 27/2V_2^2\delta_0^2$.

PROOF OF LEMMA 3. Denote by $\Delta = \mathbb{E}(\hat{c}_{J,k} - c_{J,k})^2$ and observe that $\Delta \le 3(\Delta_1 + \Delta_2 + \Delta_3)$ where

$$\Delta_1 = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\phi_{J,k}(X_i)Y_i\left(\frac{1}{\hat{f}_n(X_i)} - \frac{1}{f_n(X_i)}\right)\mathbf{1}(\hat{f}_n(X_i) \ge \delta_n)\right]^2,$$

$$\Delta_2 = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\phi_{J,k}(X_i)\,Y_i}{f(X_i)}\mathbf{1}(\hat{f}_n(X_i) \ge \delta_n)\right]^2,$$

$$\Delta_3 = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\phi_{J,k}(X_i)\,Y_i}{f(X_i)} - c_{J,k}\right]^2.$$

In the proof of the theorem we will bound from above Δ_1, Δ_2, and Δ_3, separately.

An upper bound for Δ_1. Observe that

$$\Delta_1 = \mathbb{E}\left[\frac{1}{n^2}\sum_{i=1}^{n}\phi_{J,k}^2(X_i)Y_i^2\frac{(\hat f_n(X_i)-f(X_i))^2}{\hat f_n^2(X_i)f^2(X_i)}\mathbf{1}(\hat f_n(X_i)\ge\delta_n)\right]$$
$$+\mathbb{E}\left[\frac{1}{n^2}\sum_{\substack{i,j=1\\ i\ne j}}^{n}\frac{\phi_{J,k}(X_i)\phi_{J,k}(X_j)(\hat f_n(X_i)-f(X_i))(\hat f_n(X_j)-f(X_j))Y_iY_j}{\hat f_n(X_i)f(X_i)\hat f_n(X_j)f(X_j)}\right.$$
$$\left.\cdot\mathbf{1}(\hat f_n(X_i)\ge\delta_n)\mathbf{1}(\hat f_n(X_j)\ge\delta_n)\right]$$
$$=\Delta_{11}+\Delta_{12}.$$

Since (X_i,Y_i), $i=1,\ldots,n$, are identically distributed, it is easy to see that

$$\Delta_{11}=\frac{1}{n}\int_{-\infty}^{\infty}\frac{\phi_{J,k}^2(x)v(x)}{f^2(x)}\mathbb{E}\left[\frac{(\hat f_n(x)-f(x))^2}{\hat f_n^2(x)}\mathbf{1}(\hat f_n(X_i)\ge\delta_n)\right]f(x)dx$$
$$\le V_1C_3(n\delta_n^2)^{-1}n^{-2\gamma/(2\gamma+1)}=o(n^{-2\gamma/(2\gamma+1)}).$$

Similar considerations lead to the following upper bound for Δ_{12}:

$$\Delta_{12}\le\frac{n-1}{n}\mathbb{E}\left\{\frac{|\phi_{J,k}(X_1)||\phi_{J,k}(X_2)||m(X_1)||m(X_2)|}{f(X_1)f(X_2)}\right.$$
$$\cdot\mathbb{E}\left[\frac{|\hat f_n(X_1)-f(X_1)|}{\hat f_n(X_1)}\frac{|\hat f_n(X_2)-f(X_2)|}{\hat f_n(X_2)}\right.$$
$$\left.\left.\cdot\mathbf{1}(\hat f_n(X_1)\ge\delta_n)\mathbf{1}(\hat f_n(X_2)\ge\delta_n)|X_1,X_2\right]\right\}.$$

Representing $\hat f_n^{-1}$ as $f^{-1}+(f-\hat f_n)(\hat f_n\ f)^{-1}$ we majorize Δ_{12} by the sum of two terms: $\Delta_{12}<\Delta_{121}+\Delta_{122}$. Here

$$\Delta_{121}=\mathbb{E}\left\{\prod_{i=1}^{2}\left[\frac{|\phi_{J,k}(X_i)||m(X_i)|}{f^2(X_i)}\right]\right.$$
$$\left.\cdot\mathbb{E}[|\hat f_n(X_1)-f(X_1)||\hat f_n(X_2)-f(X_2)|\mid X_1,X_2]\right\},$$

$$\Delta_{122}=\mathbb{E}\left\{\prod_{i=1}^{2}\left[\frac{|\phi_{J,k}(X_i)||m(X_i)|}{f^2(X_i)\hat f_n(X_i)}\right]\right.$$
$$\left.\cdot\mathbb{E}[(\hat f_n(X_1)-f(X_1))^2(\hat f_n(X_2)-f(X_2))^2\mid X_1,X_2]\right\}.$$

According to Lemma 1, as $n\to\infty$

$$\Delta_{121}\le\mathbb{E}\left\{\prod_{i=1}^{2}\left[\frac{|\phi_{J,k}(X_i)||m(X_i)|}{f^2(X_i)}\cdot\sqrt{\mathbb{E}[(\hat f_n(X_i)-f(X_i))^2\mid X_1,X_2]}\right]\right\}$$

$$\leq 2(C_3 + C_1^2)\left[\int_{-\infty}^{\infty} \frac{|\phi_{J,k}(x)||m(x)|}{f^2(x)} f(x)dx\right]^2 n^{-2\gamma/(2\gamma+1)}(1+o(1))$$
$$\leq 2(C_3 + C_1^2)V_2\|\phi\|_L^2 n^{-2\gamma/(2\gamma+1)}2^{-J}(1+o(1)).$$

For Δ_{122} we can construct a similar upper bound

$$\Delta_{122} < \delta_n^{-2}\prod_{i=1}^{2}\left\{\mathbb{E}\left[\frac{|\phi_{J,k}(X_i)||m(X_i)|}{f^2(X_i)}\cdot\sqrt{\mathbb{E}[(\hat f_n(X_i) - f(X_i))^4 \mid X_1, X_2]}\right]\right\}$$
$$\leq 8(C_4 + C_1^4)V_2\|\phi\|_L^2 n^{-(4\gamma-1)/(2\gamma+1)}2^{-J}\delta_n^{-2}(1+o(1)).$$

Plugging in the value of δ_n and combining Δ_{11}, Δ_{121} and Δ_{122}, we derive that

$$\Delta_1 = (1/3)C_8 n^{-(2\gamma-1)/(2\gamma+1)}(\ln n)^{-2}2^{-J}. \tag{3.4}$$

An upper bound for Δ_2. Note that

$$\Delta_2 = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\phi_{J,k}(X_i)\ Y_i}{f(X_i)}\mathbf{1}(\hat f_n(X_i) \leq \delta_n)\right)^2 = \Delta_{21} + \Delta_{22},$$

where the first term Δ_{21} has the form

$$\Delta_{21} = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\left(\frac{\phi_{J,k}(X_i)Y_i^2}{f^2(X_i)}\mathbf{1}(\hat f_n(X_i) \leq \delta_n)\right)$$
$$\leq \frac{1}{n}\mathbb{E}\left(\frac{\phi_{J,k}(X_1)Y_1^2}{f^2(X_1)}\right)$$
$$= \frac{1}{n}\int_{-\infty}^{\infty}\frac{\phi_{J,k}(x)v(x)}{f(x)}\,dx \leq \frac{V_1}{n}.$$

For the second term the following relations are valid

$$\Delta_{22} = \frac{1}{n^2}\sum_{\substack{i,j=1;\\ i\neq j}}^{n}\mathbb{E}\left[\frac{\phi_{J,k}(X_i)\phi_{J,k}(X_j)Y_iY_j}{f(X_i)f(X_j)}\mathbf{1}(\hat f_n(X_i)\leq\delta_n)\mathbf{1}(\hat f_n(X_j)\leq\delta_n)\right]$$
$$\leq \mathbb{E}\left[\frac{\phi_{J,k}(X_1)\phi_{J,k}(X_2)m(X_1)m(X_2)}{f(X_1)f(X_2)}\mathbf{1}(\hat f_n(X_1)\leq\delta_n)\mathbf{1}(\hat f_n(X_2)\leq\delta_n)\right]$$
$$\leq \left\{\mathbb{E}\left[\frac{\phi_{J,k}^2(X_1)m^2(X_1)}{f^2(X_1)}\mathbf{1}(\hat f_n(X_1)\leq\delta_n)\right]\right\}^2.$$

Observe that $\mathbf{1}(\hat f_n(x) \leq \delta_n) \leq \mathbf{1}(f(x) \leq 1.5\delta_n) + \mathbf{1}(|\hat f_n(x) - f(x)| > 0.5\delta_n)$. Therefore,

$$\Delta_{22} \leq 2\Delta_{221} + 2\Delta_{222}, \tag{3.5}$$

where

$$\Delta_{221} = \left\{\mathbb{E}\left[\frac{\phi_{J,k}^2(X_1)m^2(X_1)}{f^2(X_1)}\mathbf{1}(f(X_1)\leq 1.5\delta_n)\right]\right\}^2$$

$$= \left\{ \int_{-\infty}^{\infty} \frac{\phi_{J,k}^2(x) m^2(x)}{f^2(x)} \mathbf{1}(f(x) \le 1.5\delta_n)\ f(x)dx \right\}^2$$

$$\le (1.5\delta_n V_2)^2 = 2.25 V_2^2 \delta_0^2 n^{-2\gamma/(2\gamma+1)} (\ln n)^2,$$

$$\Delta_{222} = \left\{ \mathbb{E} \left[\frac{\phi_{J,k}^2(X_1)\ m^2(X_1)}{f^2(X_1)}\ \mathbf{1}(|\hat{f}_n(x) - f(x)| > 0.5\ \delta_n) \right] \right\}^2$$

$$= \left\{ \mathbb{E} \left[\frac{\phi_{J,k}^2(X_1) m^2(X_1)}{f^2(X_1)} \mathbb{P}(|\hat{f}_n(x) - f(x)| > 0.5\delta_n \mid X_1) \right] \right\}^2$$

$$\le \left\{ \mathbb{E} \left[\frac{\phi_{J,k}^2(X_1)\ m^2(X_1)}{f^2(X_1)} \cdot \frac{2}{n} \right] \right\}^2 \le \frac{4}{n^2} V_2^2,$$

according to Lemma 2. Combining the upper bounds for Δ_{221} and Δ_{222} in (3.5) and taking into account that $\Delta_{21} = O(n^{-1})$ and $\Delta_{222} = o(n^{-1})$, we obtain

$$\Delta_{22} \le (1/3) C_9 n^{-2\gamma/(2\gamma+1)} (\ln n)^2 (1 + o(1)). \tag{3.6}$$

An upper bound for Δ_3.

$$\Delta_3 = \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\phi_{J,k}(X_i)\ Y_i}{f(X_i)} - c_{J,k} \right)^2$$

$$= \mathrm{Var} \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\phi_{J,k}(X_i)\ Y_i}{f(X_i)} \right) \le \frac{1}{n} \mathbb{E} \left(\frac{\phi_{J,k}^2(X)\ Y^2}{f^2(X)} \right)$$

$$\le \frac{1}{n} \int_{-\infty}^{\infty} \frac{\phi_{J,k}^2(x)\ m^2(x)}{f^2(x)} f(x)dx \le \frac{V_2}{n},$$

so that $\Delta_3 = o(n^{-2\gamma/(2\gamma+1)})$. Combination of the last result, (3.4) and (3.6) completes the proof.

PROOF OF THEOREM 1. Partitioning

$$\hat{m}_n(x) - m(x) = \sum_k (\hat{c}_{J,k} - c_{J,k}) \phi_{J,k}(x) + \sum_k c_{J,k} \phi_{J,k}(x) - m(x),$$

we obtain that

$$\mathbb{E}(\hat{m}_n(x) - m(x))^2 = \mathbb{E} \left(\sum_k (\hat{c}_{J,k} - c_{J,k}) \phi_{J,k}(x) \right)^2 + \left(\sum_k c_{J,k} \phi_{J,k}(x) - m(x) \right)^2$$

$$= R_1 + R_2^2.$$

Denote $\mathcal{K}_x = \{k \mid k \in \mathbb{Z}, 2^J x - A \le k \le 2^J x + A\}$ and observe that $\mathcal{K}_x$ contains at most $(2A+1)$ elements. Then

$$R_1 = \sum_{k,l} 2^J \mathbb{E}(\hat{c}_{J,k} - c_{J,k})(\hat{c}_{J,l} - c_{J,l}) \phi(2^J x - k) \phi(2^J x - l)$$

$$\leq 2^J \sum_{k,l\in\mathcal{K}_x} \sqrt{\mathbb{E}(\hat{c}_{J,k}-c_{J,k})^2}\sqrt{\mathbb{E}(\hat{c}_{J,l}-c_{J,l})^2}|\phi(2^Jx-k)||\phi(2^Jx-l)|$$
$$\leq (2A+1)^2\|\phi\|_c^2\, 2^J \sup_k \mathbb{E}(\hat{c}_{J,k}-c_{J,k})^2,$$

where $\|\phi\|_c = \sup_z |\phi(z)|$. Thus, from Lemma 3 it follows that

(3.7) $R_1 \leq (2A+1)^2\|\phi\|_c^2 n^{-2\gamma/(2\gamma+1)}[C_8 n^{1/(2\gamma+1)}(\ln n)^{-2} + C_9 2^J(\ln n)^2](1+o(1))$.

To construct an upper bound for R_2 recall that $\operatorname{supp}\phi(y) \subset [-A, A]$. Hence, using Taylor's expansion and the fact that $|k-2^Jx| \leq A$ and $|y| \leq A$ imply $|2^{-J}(y+k)-x| \leq 2^{1-J}A$, we derive that

$$R_2 = \sum_k \phi(2^Jx-k)\int_{-\infty}^{\infty} 2^J\phi(2^Jz-k)\ m(z)dz - m(x)$$
$$= \sum_{k\in\mathcal{K}_x} \phi(2^J-k)\int_{-A}^{A} \phi(y)[m(2^{-J}(y+k))-m(x)]dy$$
$$= \sum_{k\in\mathcal{K}_x} \phi(2^J-k)\left\{\sum_{l=0}^{s-1}\int_{-A}^{A}\phi(y)\left[2^{-J}(y+k)-x\right]^l \frac{m^{(l)}(x)}{l!}dy\right.$$
$$\left.\int_{-A}^{A}\phi(y)\left[2^{-J}(y+k)-x\right]^s \frac{m^{(s)}\left[x+\xi(2^{-J}(y+k)-x)\right]}{s!}dy\right\},$$

where $0 \leq \xi = \xi(y) \leq 1$. Since $\int_{-\infty}^{\infty} y^i\phi(y) = 0$, $i = 1,\ldots,s$, all the integrals $\int_{-A}^{A}\phi(y)[2^{-J}(y+k)-x]^l dy = 0$. Therefore, we obtain

$$|R_2| \leq (2A+1)m_s 2^s A^s (s!)^{-1}\|\phi\|_c\|\phi\|_L 2^{-Js} = \sqrt{C_7}2^{-Js}. \tag{3.8}$$

Combination of (3.7) and (3.8) results in

$$\sup_x \mathbb{E}(\hat{m}_n(x)-m(x))^2 \leq [C_5 n^{(1-2\gamma)/(2\gamma+1)}(\ln n)^{-2} + C_6 2^J n^{-2\gamma/(2\gamma+1)}(\ln n)^2 + C_7 2^{-2Js}]$$
$$\cdot(1+o(1)).$$

To complete the proof choose J according to (2.7).

Acknowledgements

Authors thank the editor and two anonymous referees for their constructive suggestions and some references.

References

Antoniadis, A. and Pham, D-T. (1998). Wavelet regression for random or irregular design, *Comput. Statist. Data Anal.*, **28**, 353–369.

Antoniadis, A., Gregoire, G. and McKeague, I. (1994). Wavelet methods for curve estimation, *J. Amer. Statist. Assoc.*, **89**, 1340–1353.

Antoniadis, A., Gregoire, G. and Vial, P. (1997). Random designs wavelet curve smoothing, *Statist. Probab. Lett.*, **35**, 235–232.

Cai, T. and Brown, L. (1998). Wavelet shrinkage for nonequispaced samples, *Ann. Statist.*, **26**, 1783–1799.

Delyon, B. and Juditsky, A. (1995). *Estimating Wavelet Coefficients*, Lecture Notes in Statist., (eds. A. Antoniadis and G. Oppenheim), No. 103, 151–168, Springer, New York.

Deslauriers, G. and Dubuc, S. (1989). Symmetric iterative interpolation processes, *Constr. Approx.*, **5**, 49–68.

Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: Asymptopia? (with discussion), *J. Roy Statist. Soc. Ser. B*, **57**, 301–337.

Foster, G. (1996). Wavelets for period analysis of unequally sampled time series, *Astronomical Journal*, **112**, 1709–1729.

Hall, P. and Turlach, B. A. (1997). Interpolation methods for nonlinear wavelet regression with irregularly spaced design, *Ann. Statist.*, **25**, 1912–1925.

Hall, P., Park, B. U. and Turlach, B. A. (1998). A note on design transformation and binning in nonparametric curve estimation, *Biometrika*, **85**, 469–476.

Härdle, W., Kerkyacharian, G., Pickard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*, Lecture Notes in Statist., No. 129, Springer, New York.

Lepski, O., Mammen, E. and Spokoiny, V. (1997). Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors, *Ann. Statist.*, **25**, 929–947.

Pinheiro, A. (1997). Orthonormal bases and statistical applications; Multiresolution analysis function estimation and representation of Gaussian random fields with discontinuities, Doctoral Dissertation, Department of Statistics, University of North Carolina at Chapel Hill.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer, New York.

Sardy, S., Percival, D. B., Bruce, A. G., Gao, H.-Y., and Stuetzle, W. (1999). Wavelet shrinkage for unequally spaced data, *Statistics and Computing*, **9**, 65–75.

Sweldens, W. (1995). The lifting scheme: A new philosophy in biorthogonal wavelet constructions, *SPIE Proceedings, Signal and Image Processing III*, Vol. 2569, 69–79, San Diego, Carifornia, July 1995.