Sankhyā : The Indian Journal of Statistics
Special issue on Wavelets
2001, Volume 63, Series A, Pt. 3, pp. 328–366

NONPARAMETRIC STOCHASTIC REGRESSION WITH DESIGN-ADAPTED WAVELETS

By V. DELOUILLE
Université catholique de Louvain, Belgium
J. FRANKE
Universität Kaiserslautern, Germany
and
R. VON SACHS
Université catholique de Louvain, Belgium

SUMMARY. We present a new approach of nonparametric regression with wavelets if the design is stochastic. In contrast to existing approaches we use a new construction of a design-adapted wavelet basis which is constructed given the random regressors. To exemplify the potential of our new methodology we treat the case of using orthogonal design-adapted Haar wavelets for regression with (non-Gaussian) i.i.d. errors. We derive results on the near-optimal rate of convergence of the minimax L_2-risk of non-linear threshold estimators over a certain function class which parallel those of the classical case of fixed equidistant design. We indicate generalisations in various directions and cover parts of those by empirical investigations in our simulation examples.

1. **Introduction**

Wavelet estimation for nonparametric regression has been treated in a large number of papers in various different set-ups. They have been inspired by the seminal work of Donoho and Johnstone (1994, 1995, 1998) and Donoho *et al.* (1995), which studied minimax wavelet shrinkage methods for the reconstruction of an unknown function observed in a white noise

---

Paper received July 2000; revised July 2001.
AMS (1991) *subject classification :* 62G08.
Key words : Gaussian approximation, minimax L_2-rate, stochastic design, Unbalanced Haar wavelets, wavelet thresholding.

model. Performance of the estimator was measured by its L_2-risk over a ball in a function class characterized by some regularity parameters. In the original set-up only a *fixed equidistant design* was considered, and the noise was assumed to be i.i.d. Gaussian.

The assumption of Gaussianity of the errors has been relaxed in various works, amongst which we cite Neumann (1996), Neumann and Spokoiny (1995) and Neumann and von Sachs (1995). Indeed, these papers introduce the techniques that we also use in the present work. The generalisation of the wavelet methods to data corrupted by noise having some dependence structure has been treated in Johnstone and Silverman (1997) for the case of correlated errors, and in von Sachs and MacGibbon (2000) for correlated and possibly non-stationary errors.

This article, however, is concerned with a completely new method of generalising from the fixed and equidistant design. Methods to generalize the wavelet methods for the case of irregular design have been proposed in a large number of papers, see Antoniadis *et al.* (1997), Sardy *et al.* (1999), Kovac and Silverman (2000), Pensky and Vidakovic (2000), Antoniadis and Fan (2000) and Vanraes *et al.* (2001). Apart from Vanraes *et al.* (2001), these works share in common the idea of using some preliminary step to get back to the equidistant design situation in order to then be able to apply the traditional orthogonal wavelet transform. Methods to achieve this include projecting the data (Cai and Brown, 1998), interpolating (Hall and Turlach, 1997), or approximating the scaling function by a recursion equation (Antoniadis and Pham, 1998). These approaches are somewhat hybrid and have to deal with an additional approximation error caused by this first pre-processing step. The method of Cai and Brown (1999) avoids this but only treats the case of a *random uniform* design.

Hence the objective of this paper is to present a new wavelet regression method for irregular design that overcomes these drawbacks. Indeed, our proposed wavelet method directly adapts to the design at hand, and as such avoids any pre-processing steps.

To be more specific, we treat in this paper the following stochastic regression model

$$Y_t = m(X_t) + Z_t \ , \ \ t = 1, \ldots, N, \tag{1}$$

where the X_t's are random variables, (X_t, Y_t) are independent pairs of observations. The residuals $Z_1, \ldots, Z_N$ are i.i.d. with mean zero and finite variance, $E(Z_t) = 0, \mathrm{Var}(Z_t) = \sigma_z^2 < \infty$, but are not necessarily Gaussian. Our target function is the conditional mean $m(x) = E[Y|X = x]$. This set-up also offers the possibility to treat nonlinear autoregression, by setting

$X_t := Y_{t-1}$, and to estimate, for example, autoregressive threshold models. Note that the only existing algorithm (Hoffmann, 1999) that treats autoregressive models using a wavelet estimator does not allow for an efficient implementation.

The main idea behind our procedure is to build a wavelet basis of the weighted space $L_2(d\widehat{F}_N)$, where $\widehat{F}_N$ is the empirical measure associated to the stochastic design given by the $\{X_t\}_{t=1}^N$. In this paper, we consider the simplest example of such a design-adapted wavelet, that is, the Unbalanced Haar wavelet. More precisely, we instantiate the very general construction presented in Girardi and Sweldens (1997) to the space $L_2(d\widehat{F}_N)$, in order to obtain the only possible *orthonormal* wavelet basis of $L_2(d\widehat{F}_N)$. This particular construction allows us to derive quite easily the rates of convergence of the L_2-risk for the nonlinear thresholding wavelet estimator.

The Unbalanced Haar wavelet basis is very simple, but does not provide a smooth reconstruction. However, smoother reconstruction can be obtained with the Lifting Scheme, see Delouille *et al.* (2001). In this process of building smoother wavelets, the Unbalanced Haar basis constitutes the initialising step, and as such it is important to study its approximation properties. Note that this approach differs from the work of Kohler (2000) where the aim of constructing design-adapted bases was achieved by using only piecewise smooth bases of piecewise polynomials.

To keep the technicality of our paper comparatively low, we decided to treat the case of i.i.d. errors only. However, we allow the errors to be non-Gaussian and show that techniques parallel to the ones of Neumann and von Sachs (1995) can be applied to derive the L_2-rate of convergence known from the classical equidistant and fixed design regression. We consider a class of β-Hölder continuous functions having a regularity parameter $\beta < 1$. Indeed, when using Unbalanced Haar wavelets, we cannot prove any optimality results for a class of functions having a regularity parameter larger than one. However, we could have considered *piecewise* Hölder functions, but again this would have made the proofs much more technical.

In the simulations of Section 6 however, we enlarged the scope of practical use of our method. Firstly, we use a non-decimated, also called 'translation-invariant', version of the Unbalanced Haar wavelet transform. It is well-known from the classical situation (Coifman and Donoho, 1995, Nason and Silverman, 1995) that translation-invariant denoising can be considered as increasing the degree of smoothness of the wavelet reconstruction, see Berkner and Wells (1998). This is in particular interesting when using non-smooth Haar wavelets. Secondly, we propose an algorithm that allows for heteroscedastic errors, and show how it works for various error structures, including correlated and heteroscedastic. Our simulations also show that this wavelet algorithm works fine for functions with jumps.

Note that the sample size N can be arbitrary in the case of the Unbalanced Haar wavelets (it does not need to be a power of two). For the non-decimated version of Haar, however, N must be a power of two. Hence to simplify the presentation, we consider in this article sample sizes N which are a power of two. This restriction on the sample size can be dropped when using smoother wavelets built with the Lifting Scheme, see Delouille *et al.* (2001).

The remaining part of this paper is organised as follows. Section 2 presents the construction of a design-adapted wavelet basis. It shows how to use it for regression purpose and presents the resulting nonlinear wavelet estimator. In Section 3, we specify the assumptions on our stochastic regression model and state our main theorem. We indicate the line of its overall proof. This involves two major steps, which are further detailed in Sections 4 and 5, respectively. The algorithm to be used in practice is presented in Section 6, together with our simulation study. Finally, the proofs related to Sections 3 and 4 are deferred to Section 7. The proofs of the propositions in Section 5 being largely similar to the ones of Neumann (1996) and Neumann and Spokoiny (1995), we omit them and refer instead to Delouille *et al.* (2000).

2. Construction of Design-adapted Wavelets

2.1. *Wavelet bases which are orthogonal on a weighted space* $L_2(\mu)$. Wavelets are traditionally built to form an orthonormal (or biorthogonal, semiorthogonal – see Daubechies (1988) and Cohen *et al.* (1992)) basis in $L_2(\mathbb{R})$. For our regression purposes, however, we will need to consider a more general measure than the Lebesgue measure. More precisely, we will use a wavelet basis which is orthonormal in a weighted space $L_2(\mu)$. Girardi and Sweldens (1997) presented such a type of wavelet basis, called Unbalanced Haar basis, which can be built on very general spaces. Here we particularize their construction to the Euclidean space $L_2(I, \Sigma, \mu)$, where I is a subset of $\mathbb{R}$, Σ is the σ-field of I and μ is a σ-finite measure on Σ. This measure μ can be very general; in particular it can be purely non-atomic (it does not contain point masses), purely atomic (it contains only point masses), or a mixture of both situations.

In order to build Haar-like wavelets adapted to the space (I, Σ, μ), we first need to choose a good *partitioning* of the interval $I \in \mathbb{R}$ we are working on. This concept of partitioning is defined in Girardi and Sweldens (1997) and recalled below. It can be thought of as the replacement of the dyadic intervals on the real line.

DEFINITION 1. A partitioning on $I \in \mathbb{R}$ is a subset of intervals $I_{jk}, j \in \mathbb{Z}, k \in \mathbb{Z}$ such that:

(P1) $I = \cup_k \{I_{jk} | j \text{ fixed}, k \in \mathbb{Z}\}, \forall\ j$;

(P2) $I_{j,k_1} \cap I_{j,k_2} = \emptyset\ \forall k_1 \neq k_2$ and $\forall j$;

(P3) $I_{jk} = I_{j+1,2k} \cup I_{j+1,2k+1}$;

(P4) $\sigma(I_{jk} | j \in \mathbb{Z}, k \in \mathbb{Z}) = \widetilde{\Sigma}$, where $\widetilde{\Sigma}$ is the sub-σ-field of Σ such that the μ-completion of $(I, \widetilde{\Sigma})$ is (I, Σ).

Having a nested partitioning of I, we are now ready to build unbalanced Haar wavelets. These have as their basic building blocks the scaling functions $\{\varphi_{jk}\}$, where

$$\varphi_{jk} = \frac{1_{I_{jk}}}{\mu(I_{jk})^{1/2}}, \tag{2}$$

and as wavelets the functions

$$\psi_{jk} = \frac{1}{\mu(I_{jk})^{1/2}} \left(\sqrt{\mu(I_{j+1,2k+1})}\varphi_{j+1,2k} - \sqrt{\mu(I_{j+1,2k})}\varphi_{j+1,2k+1} \right). \tag{3}$$

Note that these functions are all normalized in $L_2(\mu)$. It has been proved in Girardi and Sweldens (1997) that these wavelets $\{\psi_{j,k}\}$ form an orthonormal basis for the space $L_2(I, \Sigma, \mu)$, and moreover have the ability to generate a multiresolution analysis. The next section explains how to use this wavelet basis in the case of nonparametric regression with stochastic design.

2.2. *Regression with design-adapted wavelets.* Consider the model (1) given in Section 1 and let F_x, f_x denote the distribution function and the density of the regressors X_t, respectively. By 'design adapted' wavelets, we mean wavelets which form a basis of the space $L_2(dF_x)$, where dF_x is the Borel-Stieltjes measure associated to F_x. In this space, the measure of an interval $[a, b)$ is given by

$$F_x(b) - F_x(a) = P(X_t \in [a, b[);\ t = 1, \ldots, N.$$

In words, we take into account in the measure of the space the information contained in the distribution of the regressors X_t.

In this paper we assume that F_x is strictly increasing (that is, f_x is bounded away from zero). In this case the σ-field $\widetilde{\Sigma}$ of Definition 1 equals the Borel field on $\mathbb{R}$, i.e. $\mathcal{B}(\mathbb{R})$. To proceed, we need to specify which partitioning to use. Two situations have to be considered, depending on whether or not F_x is known.

If the distribution F_x is considered to be known, it is possible to build what we call a *quantile partitioning*, denoted by I°_{jk}, where the endpoints of the intervals are the quantiles of the distribution F_x. Here and in the following we choose as usual $F_x^{-1}(u) = \inf\{x : \ F_x(x) \geq u\}$ as the quantile function. By continuity of F_x, we always have $F_x(F_x^{-1}(u)) = u$. The elements of the partitioning are then defined as

$$I^{\circ}_{jk} = \left[F_x^{-1}\left(\frac{k-1}{2^j}\right),\ F_x^{-1}\left(\frac{k}{2^j}\right)\right),$$

where $j = 1, \ldots, J,\ \ k = 1, \ldots, 2^j$; with J denoting the finest scale in the decomposition. In this scheme, the measure of an interval $I^{\circ}_{j,k}$ is simply equal to 2^{-j}, that is, the quantiles considered are equispaced in $L_2(dF_x)$. For this partitioning, condition $(P4)$ in Definition 1 is fulfilled with $\widetilde{\Sigma} = \mathcal{B}(\mathbb{R})$. With the quantile partitioning, equations (2) and (3) simplify to a form similar to the classical Haar wavelets

$$\begin{aligned} \varphi^{\circ}_{jk}(x) &= 2^{j/2} 1^{\circ}_{jk}(x),\ \ j = 1, \ldots, J,\ k = 1, \ldots, 2^j, \\ \psi^{\circ}_{jk}(x) &= \frac{1}{\sqrt{2}}\left(\varphi^{\circ}_{j+1,2k}(x) - \varphi^{\circ}_{j+1,2k+1}(x)\right),\ j = 1, \ldots, J-1,\ k = 1, \ldots, 2^j, \end{aligned} \quad (4)$$

with $1^{\circ}_{jk}(x)$ denoting the indicator of I°_{jk}. Using these wavelets, an $L_2(dF_x)$-integrable function can be represented as the infinite series

$$m(x) = \sum_k s^{\circ}_{j_0,k}\varphi^{\circ}_{j_0,k}(x) + \sum_{j \geq j_0, k \in \mathbb{Z}} \alpha^{\circ}_{jk}\psi^{\circ}_{j,k}(x), \quad (5)$$

where j_0 denotes the primary resolution level, and where

$$\alpha^{\circ}_{j,k} = \left\langle m(x)\,, \psi^{\circ}_{jk}(x)\right\rangle_{dF_x} = \int m(x)\psi^{\circ}_{jk}(x) dF_x(x) \quad (6)$$

denotes the wavelet or detail coefficient at level j and location k, and $s^{\circ}_{j_0,k} = \left\langle m(x)\,, \varphi^{\circ}_{j_0,k}\right\rangle$ represents the scaling coefficients at level j_0 and location k. Given an arbitrary design $X_1, \ldots, X_N$, one can always estimate α°_{jk} as

$$\hat{\alpha}^{\circ}_{jk} = \frac{1}{N}\sum_{t=1}^{N} Y_t \psi^{\circ}_{j,k}(X_t) \quad (7)$$

and similarly for $s^\circ_{j_0,k}$. Note that in this operation, the intervals I°_{jk} do not have as endpoints the design points $X_1,\dots,X_N$ and hence some binning of the data occurs.

In practice we only have information about the empirical distribution $\widehat{F}_N$ of the given realisation $X_1,\dots,X_N$. Since even for $\widehat{F}_N$ the condition $(P4)$ in Definition 1 does hold, we can define wavelets which are orthonormal in $L_2(d\widehat{F}_N)$. More precisely, we define the *empirical quantile partitioning*, noted I_{jk}, where we use the empirical quantiles instead of the theoretical quantiles in order to define the endpoints of the intervals. Here the random intervals $\{I_{j,k}\}$ are determined by the data as :

$$I_{jk} = [X_{((k-1)n_j+1)},\ X_{(kn_j+1)})\ ,\ \ k = 1,\dots,2^j\ ,\ n_j = N2^{-j},$$
$$j_0 \le j \le J = \log_2(N/M) \ \text{ for some integer } 1 \le M < N,$$

where $X_{(1)} \le \dots \le X_{(N)}$ denote the order statistics of the design variables and, as a convention, $X_{(N+1)} := X_{(N)} + (X_{(N)} - X_{(1)})/N$. The measure of I_{jk} in $L_2(d\hat{F}_N)$ is also equal to 2^{-j}, and the corresponding Haar wavelets ψ_{jk} are given by

$$\psi_{jk}(x) = \frac{1}{\sqrt{2}}(\varphi_{j+1,2k}(x) - \varphi_{j+1,2k+1}(x)),\ \ j = j_0,\dots,J-1, k = 1,\dots,2^j, \tag{8}$$

with

$$\varphi_{jk}(x) = 2^{j/2}1_{jk}(x),\ \ j = j_0,\dots,J,\ \ k = 1,\dots,2^j, \tag{9}$$

where $1_{jk}(x)$ denotes the indicator of the random interval I_{jk}. The corresponding empirical wavelet coefficients are defined as:

$$\widehat{\alpha}_{j,k} = \frac{1}{N}\sum_{t=1}^{N} Y_t\psi_{j,k}(X_t), \qquad \widehat{s}_{j,k} = \frac{1}{N}\sum_{t=1}^{N} Y_t\varphi_{j,k}(X_t). \tag{10}$$

The forward wavelet transform corresponding to (8)-(9) proceeds as follows. First initialise $\hat{s}_{J,k}$ as $N^{-1}\sum_{t=1}^{N} Y_t\varphi_{J,k}(X_t)$. Then for each level $j = J-1,\dots,j_0$, do:

$$\hat{s}_{j,k} = \frac{1}{\sqrt{2}}\hat{s}_{j+1,2k} + \frac{1}{\sqrt{2}}\hat{s}_{j+1,2k+1} \quad ; \quad \hat{\alpha}_{j,k} = \frac{1}{\sqrt{2}}\hat{s}_{j+1,2k+1} - \frac{1}{\sqrt{2}}\hat{s}_{j+1,2k}\ .$$

These filters are the usual Haar filters. In other words, when using this last basis, we treat the data as if they were on a regular grid. This is justified in the case of Haar wavelets, since we are only aiming at estimating functions which are β-Hölder continuous with $0 < \beta < 1$. To recover smoother functions, however, it is necessary to consider the irregularity of the grid. Indeed, one can show (Cai, 1996) that in many situations, the rate of convergence of the estimator is suboptimal if nonequispaced data are simply treated as equispaced. Using the Lifting Scheme, it is possible to take into account the irregularity of the grid while providing a smoother reconstruction, see Delouille *et al.* (2001) and Vanraes *et al.* (2001).

Given a data set $(X_t, Y_t)_{t=1}^N$, our nonlinear wavelet thresholding estimator based on the Unbalanced Haar basis (8)-(9) is given by

$$\widehat{m}(x) = \sum_k \widehat{s}_{j_0k}\ \varphi_{j_0k}(x)\ +\ \sum_{j=j_0}^{J-1}\sum_{k=1}^{2^j} \delta^{(\cdot)}(\widehat{\alpha}_{jk}, \lambda_{jk})\ \psi_{jk}(x)\ , \qquad (11)$$

where $\delta^{(\cdot)}(w, \lambda)$ denotes either 'hard' or 'soft' threshold nonlinearities:

$$\delta^H(w,\lambda) = w 1_{\{|w|>\lambda\}}\ ; \quad \delta^S(w,\lambda) = \text{sign}(w)(|w| - \lambda)_+\ .$$

In expression (11), we only take into account the wavelet coefficients of indices (j, k) belonging to a certain set $\mathcal{J}_N$, defined for the purpose of deriving the near-optimal rate of convergence. This set $\mathcal{J}_N$ is such that

$$\mathcal{J}_N = \{(j,k)|2^j \leq C\, N^{1-\delta}\} \text{ with } 0 < \delta < 1. \qquad (12)$$

For a more precise specification of the range of δ in this definition we refer to equation (18).

In the next section, we study in detail the properties of the nonlinear wavelet thresholding estimator given in (11).

3. Properties of the Unbalanced Haar Wavelet Estimator

3.1. *Assumptions and main theorem.* One possibility of measuring the asymptotic performance of some non-parametric curve estimation procedure is to evaluate rates of convergence of the L_2-risk, or integrated mean squared error (IMSE) between the estimate $\widehat{m}$ and the curve m of interest. In the literature of non-linear wavelet threshold estimation this is usually done in a uniform sense over a ball $\mathcal{F}$ in a function class such as a Hölder, Sobolev or Besov class. In this paper, we restrict ourselves to treat a Hölder class $\mathcal{F}(L;\beta)$:

$$\mathcal{F}(L;\beta) = \Lambda^\beta(L) = \{m \in L_2(\mathbb{R}) : |m(x) - m(y)| \leq L\,|x-y|^\beta\}\ , \quad 0 < \beta < 1\ .$$

More precisely, throughout the paper we will need the following assumptions on the regression function m and on the density f_x.

(A1) The density function f_x is continuous and bounded away from zero on the interior of a compact interval S: $f_x(x) \geq c > 0$ for all $x \in \overset{\circ}{S}$, $f_x(x) = 0$ for all $x \in \mathbb{R} \setminus \overset{\circ}{S}$.

(A2) The function m is Hölder continuous on S, with exponent $1/2 < \beta < 1$, that is, $m \in \Lambda^{\beta}(L)$, $\beta > 1/2$.

(A3) The function m is of bounded total variation over the interval S.

For an easier treatment of the models with non-Gaussian error distribution, we also need a technical assumption on the error distribution:

(A4) All moments of the errors $\{Z_i\}$ are uniformly bounded:

$$E[|Z_i|^K] \ \leq \ C_K \quad \text{for all } K \in \mathbb{N} \text{ and for some fixed constants } C_K < \infty.$$

To define our L_2-risk, we consider as reference function the projection $m_J(x)$ of $m(x)$ onto the space generated by $\{\varphi_{J,k}\}$, with J defined as the finest level that belongs to the set $\mathcal{J}_N$ in (12). This projection $m_J(x)$ can be developed in the basis (8)-(9) as $m_J(x) = \sum_k \widehat{b}_{j_0,k}\varphi_{j_0,k}(x) + \sum_{jk} \hat{\beta}_{jk}\psi_{jk}(x)$, where

$$\widehat{b}_{j_0,k} = \frac{1}{N} \sum_{t=1}^{N} m(X_t)\, \varphi_{j_0,k}(X_t) \ ; \ \widehat{\beta}_{j,k} = \frac{1}{N} \sum_{t=1}^{N} m(X_t)\, \psi_{j,k}(X_t). \tag{13}$$

Our aim is, given a stochastic design $X_1, \ldots, X_N$, to find an estimator that comes 'as close as possible', in an IMSE sense, to the projected function $m_J(x)$, and this uniformly for all functions $m \in \Lambda^{\beta}(L)$. Thus, we define our risk in $L_2(d\hat{F}_N)$ and look for the minimizer $\widehat{m}$ of

$$\sup_{m \in \mathcal{F}} \left\{ E\|\widehat{m} - m_J\|^2_{L_2(d\hat{F}_N)} \right\}. \tag{14}$$

In our main result, stated in Theorem 1, we consider only soft thresholding. Now, to be more precise about the choice of the thresholds λ_{jk} present in the estimator (11), we follow Neumann (1996) and Neumann and von Sachs (1995) and distinguish between two different situations. Firstly, we call

$$\widetilde{m}(x) = \sum_k \widehat{s}_{j_0k}\ \varphi_{j_0k}(x) \ + \sum_{(j,k)\in\mathcal{J}_N} \delta^S(\widehat{\alpha}_{jk}, \widetilde{\lambda}_{jk})\ \psi_{jk}(x) \tag{15}$$

the estimator that is based on an optimal, non-random, threshold $\widetilde{\lambda}_{jk}$ which depends on the sample size N, the scale j, the location k, and the smoothness of the underlying function, thus $\widetilde{\lambda}_{jk} = \lambda(N, j, k; \mathcal{F})$. Secondly, an estimator $\hat{m}(x)$, with data-driven, and hence random, threshold λ_{jk} is given by

$$\widehat{m}(x) = \sum_k \widehat{s}_{j_0 k}\ \varphi_{j_0 k}(x)\ +\ \sum_{(j,k)\in\mathcal{J}_N} \delta^S(\widehat{\alpha}_{jk}, \lambda_{jk})\ \psi_{jk}(x) \tag{16}$$

A proper choice of the random threshold λ_{jk} in view of the asymptotic behaviour of $\widehat{m}(x)$ is a threshold which fulfills the following inequalities on an event with probability tending to one with a rate as given in Neumann (1996, condition A6):

$$\sigma_{jk}\ \sqrt{2\log(\#\mathcal{J}_N)} \leq \lambda_{jk} \leq C\ N^{-1/2}\sqrt{\log(N)}\ , \tag{17}$$

for a positive constant C. In the inequalities (17), σ^2_{jk} is the conditional variance of the empirical wavelet coefficients $\hat{\alpha}_{jk}$, that is, their variance given the value of the regressors $X_1, \ldots, X_N$.

THEOREM 1. *Let $\mathcal{F} = \Lambda^\beta(L)$, with $1/2 < \beta < 1$. Let $\widetilde{m}$ and $\widehat{m}$ be the wavelet threshold estimators given in (15) and (16), respectively. Then,*

(i) for an optimal choice of the non-random threshold $\widetilde{\lambda}_{jk} = \lambda(N, j, k; \mathcal{F})$,

$$\sup_{m\in\mathcal{F}} \left\{ E\,\|\widetilde{m}\ -\ m_J\|^2_{L_2(d\hat{F}_n)} \right\}\ =\ O\left(N^{-2\beta/(2\beta+1)}\right),$$

(ii) for thresholds λ_{jk} satisfying

$$\sigma_{jk}\ \sqrt{2\log(\#\mathcal{J}_N)} \leq \lambda_{jk} \leq C\ N^{-1/2}\sqrt{\log(N)}$$

for any positive constant C,

$$\sup_{m\in\mathcal{F}} \left\{ E\,\|\widehat{m}\ -\ m_J\|^2_{L_2(d\hat{F}_n)} \right\}\ =\ O\left((\log(N)/N)^{2\beta/(2\beta+1)}\right).$$

We have attained the 'classical' rate $N^{-2\beta/(2\beta+1)}$ for the L_2-risk for functions in the class $\Lambda^\beta(L)$. This rate is attained for the optimal threshold, unknown in practice, whereas for a data-driven threshold the usual additional logarithmic term appears, giving a near-optimal rate of convergence.

An important example for such a data-driven threshold is the *universal* threshold $\widehat{\lambda}_{jk} = \widehat{\sigma}_{jk}\ \sqrt{2\log(\#\mathcal{J}_N)}$, based on an estimator $\widehat{\sigma}^2_{jk}$ of the unknown variance σ^2_{jk} of the empirical wavelet coefficients. In Section 6, we present some ways to estimate σ^2_{jk}. Sections 4 and 5 develop the tools needed to prove Theorem 1 but before we give in Section 3.2 the overall line of its proof.

3.2. *Methodology for proving the main result.* Consider the design $q_1, \ldots, q_N$ made of theoretical and equispaced (in $L_2(dF_x)$) quantiles $q_t = F_x^{-1}(t/N)$. For this particular design, $I_{jk} = I^\circ_{jk}$ and similarly for the wavelet basis. Using the infinite representation of $m(x)$ defined in (5), it is possible to show that the IMSE-optimal cutting level j_1 of a *linear* wavelet estimator is such that $2^{j_1} \asymp N^{1/(2\beta+1)}$. The corresponding linear wavelet estimator has a rate of convergence of order $N^{-\frac{2\beta}{2\beta+1}}$, see Proposition 8 in Section 7.1 for a complete proof.

As $q_1, \ldots, q_N$ are also coming from the distribution function $F_x(x)$, in the same way as $X_1, \ldots, X_N$ are, the order of a linear wavelet estimator based on $X_1, \ldots, X_N$ and having an optimal cutting level, can not be better than $O(N^{-\frac{2\beta}{2\beta+1}})$. As we consider in this article a Hölder class of functions, a linear wavelet estimator, with optimal cutting level, is the best way to estimate the function. The problem, however, is to estimate in practice this optimal cutting level. Cross-validation could be used (Nason, 1999), or the regularity parameter β could be estimate. This is a difficult task, however.

This explains why we choose to adopt the nonlinear wavelet thresholding estimator as defined in (11). The set $\mathcal{J}_N$ in (12) must be such that the finest level J belonging to $\mathcal{J}_N$ is greater than j_1. For this to hold, we add a condition on δ

$$0 < \delta < \frac{2\beta}{2\beta + 1}. \tag{18}$$

Consider now the risk in $L_2(d\hat{F}_N)$ as defined in (14). This is quite similar to the risk used by Donoho and Johnstone in Donoho and Johnstone (1994). Using Parseval's relation, we could then use directly their techniques to derive, with the help of an oracle, some upper and lower bounds on the l_2-risk of the wavelet coefficients. Our aim however, is to obtain the precise rate of convergence of the wavelet estimator (16), and this even in the case where the errors are not Gaussian. This is the reason why we will rather use the methodology developed in Neumann (1996) and Neumann and von Sachs (1995).

More precisely, consider the difference

$$D = \widehat{m}(x) - m_J(x) = \sum_k \widehat{s}_{j_0 k}\ \varphi_{j_0 k}(x) + \sum_{(j,k)\in\mathcal{J}_N} \delta^{(s)}(\widehat{\alpha}_{jk}, \lambda_{jk})\psi_{jk}(x) - \left(\sum_k \widehat{b}_{j_0,k}\varphi_{j_0,k}(x) + \sum_{j,k}\widehat{\beta}_{jk}\psi_{jk}(x)\right) . \quad (19)$$

To compute the corresponding L_2-risk, it is useful to add and subtract $\sum_{jk}\alpha^\circ_{jk}\psi_{jk}(x)$, where α°_{jk} is defined in (6), to the equality (19). This allows us to write

$$D = \sum_k(\widehat{s}_{j_0,k} - \widehat{b}_{j_0,k})\varphi_{j_0,k}(x) + \sum_{(j,k)\in\mathcal{J}_N}(\delta^S(\widehat{\alpha}_{jk},\lambda_{jk}) - \alpha^\circ_{jk})\psi_{jk}(x) + \sum_{jk}(\alpha^\circ_{jk} - \widehat{\beta}_{jk})\psi_{jk}(x) .$$

Using orthonormality of the basis $\{\varphi_{j_0,k},\psi_{jk}\}$ and Parseval's equality, we obtain:

$$\begin{aligned} E\,\|\widehat{m}-m_J\|^2_{L_2(d\hat{F}_N)} &\le 2\left(\sum_k E[(\widehat{s}_{j_0,k} - \widehat{b}_{j_0,k})^2] + \sum_{jk} E[(\delta^S(\widehat{\alpha}_{jk},\lambda_{jk}) - \alpha^\circ_{jk})^2] + \sum_{jk} E[(\alpha^\circ_{jk} - \widehat{\beta}_{jk})^2]\right) \\ &= T_1 + T_2 + T_3 . \end{aligned} \quad (20)$$

It is straightforward to prove that $T_1 = o(N^{-\frac{2\beta}{2\beta+1}})$, see Proposition 9 in Section 7.1. The main difficulty lies in proving that $T_2 = O(\log N/N)^{\frac{2\beta}{2\beta+1}}$. Our proposed method consists in reducing the stochastic regression to the set-up of a fixed design composed of the equispaced quantiles $\{q_t\}$. Consider the empirical coefficients $\widehat{\alpha}^*_{jk} = N^{-1}\sum_{t=1}^N Y^*_t\psi^\circ_{jk}(q_t)$ corresponding to data generated by the model (22) below. Using the monotonicity of the soft threshold rule δ^S in its first argument, the term T_2 can be bounded as:

$$E[(\delta^S(\widehat{\alpha}_{jk},\lambda_{jk}) - \alpha^\circ_{jk})^2] \le 2\,E[(\widehat{\alpha}_{jk} - \widehat{\alpha}^*_{jk})^2] + 2\,E[(\delta^S(\widehat{\alpha}^*_{jk},\lambda_{jk}) - \alpha^\circ_{jk})^2] . \quad (21)$$

With this decomposition, it is sufficient to prove the two following results.

(R1) $T_{2,1} := \sum_{(j,k)\in\mathcal{J}_N} E[(\widehat{\alpha}_{jk} - \widehat{\alpha}^*_{jk})^2] = o(N^{-\frac{2\beta}{2\beta+1}})$, for a set $\mathcal{J}_N$ that satisfies (12) and that will be further specified in Proposition 5.

(R2) $T_{2,2} := \sum_{jk} E[(\delta^S(\widehat{\alpha}^*_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})^2] = O\left(\frac{\log N}{N}\right)^{\frac{2\beta}{2\beta+1}}$. The result (R1) tells us that the approximation made by replacing $\widehat{\alpha}_{jk}$ by $\widehat{\alpha}^*_{jk}$ is smaller than the rate of convergence of our wavelet estimator. In Section 4 the tools needed to prove (R1) are developed. These will also allow us to deduce that $T_3 = o(N^{-\frac{2\beta}{2\beta+1}})$. The corresponding proofs are given in Section 7. Finally, Section 5 explains how to prove (R2).

4. Asymptotic Approximation of the Wavelet Coefficients

4.1. *Reducing to the simple set-up of fixed design.* We consider our original regression model (1) given in Section 1. As announced before, for theoretical purposes, we also consider the following twin model

$$Y^*_t = m(q_t) + Z'_t\,,\ t = 1, \ldots, N, \tag{22}$$

where $Z'_1, \ldots, Z'_N$ are appropriately chosen i.i.d. random variables with the same distribution as the residuals Z_t of the model (1), but the design $q_1, \ldots, q_N$ is deterministic with $q_t = F_x^{-1}(\frac{t}{N})$, $t = 1, \ldots, N$. To simplify notation, we set $q_0 = F_x^{-1}(0) = -\infty$ as a convention.

We assume $N = M2^J$ for some integers M and J, that is, $M = N^\delta$ with δ satisfying the inequalities in (18). To simplify notation, we set $\alpha_{j_0-1,k} := s_{j_0,k}$, $\psi_{j_0-1,k} := \varphi_{j_0,k}$ and similarly for $\hat{s}_{j_0,k}$ and $\varphi^\circ_{j_0,k}$. The coefficients $\widehat{\alpha}^*_{jk}$ based on the deterministic design model (22) are

$$\widehat{\alpha}^*_{jk} = \frac{1}{N}\sum_{t=1}^{N} Y^*_t\, \psi^\circ_{jk}(q_t),\ \ j = j_0 - 1, \ldots, J-1, k = 1, \ldots, 2^j. \tag{23}$$

In this section our main goal is to show (R1), i.e.:

$$\textstyle\sum_{(j,k)\in\mathcal{J}_N} E[(\widehat{\alpha}^*_{jk} - \widehat{\alpha}_{jk})^2] = o(N^{-\frac{2\beta}{2\beta+1}}). \tag{24}$$

To study the effect of noise on the wavelet coefficients in both designs (random $\{X_t\}$ and deterministic $\{q_t\}$), we have to select an appropriate realisation of the model (22) with noise variables coupled to model (1). To make this precise, let $X_1, \ldots, X_N, Z_1, \ldots, Z_N$ and $Y_t = m(X_t) + Z_t$, $t = 1, \ldots, N$, be chosen as in (1). Then, we choose $Z'_t = Z_s$ iff $X_{(t)} = X_s$, $t, s = 1, \ldots, N$. As $Z_1, \ldots, Z_N$ are both i.i.d. and independent of the $X_1, \ldots, X_N$, $Z'_1, \ldots, Z'_N$ are i.i.d., too, with the same distribution as the Z_s.

($Z_1', \ldots, Z_N'$ is just an independent reordering of $Z_1, \ldots, Z_N$.) Then, we set $Y_t^* = m(q_t) + Z_t'$, $t = 1, \ldots, N$, which is a particular realization of the model (22). To show the relation (24), consider the following decomposition of $\widehat{\alpha}_{jk}$:

$$\widehat{\alpha}_{jk} = \widehat{\beta}_{jk} + \widehat{\rho}_{jk} = \frac{1}{N} \sum_{t=1}^{N} m(X_t)\, \psi_{jk}(X_t) + \frac{1}{N} \sum_{t=1}^{N} Z_t\, \psi_{jk}(X_t), \tag{25}$$

and similarly

$$\widehat{\alpha}_{jk}^* = \widehat{\beta}_{jk}^* + \widehat{\rho}_{jk}^* = \frac{1}{N} \sum_{t=1}^{N} m(q_t)\, \psi_{jk}^\circ(q_t) + \frac{1}{N} \sum_{t=1}^{N} Z_t'\, \psi_{jk}^\circ(q_t)\ , \tag{26}$$

where the $\hat{\rho}_{jk}$ and $\widehat{\rho}_{jk}^*$ depend on the noise and the remainder terms do not. With these decompositions, it is easier to show (24) using the following relationship:

$$\begin{aligned} E[(\widehat{\alpha}_{jk}^* - \widehat{\alpha}_{jk})^2] &\leq 2\{E[(\widehat{\rho}_{jk} - \widehat{\rho}_{jk}^*)^2] + E[(\widehat{\beta}_{jk} - \widehat{\beta}_{jk}^*)^2]\} \\ &= 2\{E[(\widehat{\rho}_{jk} - \widehat{\rho}_{jk}^*)^2] + \mathrm{Var}(\widehat{\beta}_{jk}) + (E[\widehat{\alpha}_{jk} - \widehat{\alpha}_{jk}^*])^2\} \end{aligned} \tag{27}$$

and to consider each term separately. Section 4.2 evaluates each term of (27) individually, and in Section 4.3 we show that the sum (24) over all coefficients in $\mathcal{J}_N$ has the desired order.

4.2. *Bias and variance of an individual coefficient.* Consider, as a useful intermediate, the coefficients $\widehat{\alpha}_{jk}^\circ$ from equation (7), which are built with random design X_t and fixed partitioning I_{jk}°. Note that these wavelet coefficients are just introduced as a practical tool in the proofs because they are unbiased estimators of the theoretical coefficients α_{jk}° given by (6), see Proposition 1 below. These values $\{\alpha_{jk}^\circ\}$ will constitute our reference point when computing the expectation of $\widehat{\alpha}_{jk}$ and $\widehat{\alpha}_{jk}^*$, see Proposition 3. Similarly to what has been done above, we decompose $\widehat{\alpha}_{jk}^\circ$ into two parts:

$$\widehat{\alpha}_{jk}^\circ = \widehat{\beta}_{jk}^\circ + \widehat{\rho}_{jk}^\circ = \frac{1}{N} \sum_{t=1}^{N} m(X_t)\, \psi_{jk}^\circ(X_t) + \frac{1}{N} \sum_{t=1}^{N} Z_t\, \psi_{jk}^\circ(X_t)\ . \tag{28}$$

PROPOSITION 1. *The coefficients $\widehat{\alpha}_{jk}^\circ, \alpha_{jk}^\circ$ given by (7) and (6), respectively, fulfill uniformly over (j,k) in $\mathcal{J}_N$:*

(i) $E(\widehat{\alpha}_{jk}^\circ) = \alpha_{jk}^\circ$.

(ii) $\mathrm{Var}(\widehat{\alpha}_{jk}^\circ) = \frac{1}{N}\left(\sigma_z^2 + \mathrm{Var}(m(X_1)\psi_{jk}^\circ(X_1))\right)$

. $\leq \frac{1}{N}\left(\sigma_z^2 + \sup\{m^2(F^{-1}(u));\ \frac{k-1}{2^j} \leq u < \frac{k}{2^j}\}\right)$

The behaviour of the noise-dependent part of $E[(\widehat{\alpha}_{jk}-\widehat{\alpha}^*_{jk})^2]$ is described below.

PROPOSITION 2. *Consider the coefficients $\widehat{\rho}_{jk}$ and $\widehat{\rho}^*_{jk}$ as defined in* (25) *and* (26) *respectively. Then uniformly over (j,k) in $\mathcal{J}_N$, we have:*

$$E[(\widehat{\rho}_{jk}-\widehat{\rho}^*_{jk})^2]=O\left(\frac{2^j}{N^{3/2}}\right) .$$

Let $n_j = N2^{-j}$ and, in the following proposition, let $V(g;a,b)$ denote the total variation of the real-valued function g over the finite interval $[a,b]$. Proposition 3 evaluate the bias and variance of the wavelet coefficients.

PROPOSITION 3. *Assume (A1) through (A3). Define α°_{jk}, $\widehat{\alpha}_{jk}$ and $\widehat{\alpha}^*_{jk}$ as in* (6), (10) *and* (23) *respectively. Then, uniformly over $(j,k)\in\mathcal{J}_N$, we have the following assertions.*

(i) $E[\widehat{\alpha}^*_{jk}]=\alpha^\circ_{jk}+V(m;q_{(k-1)n_j},q_{kn_j})\ O(\frac{2^{j/2}}{N})$,

(ii) $E[\widehat{\alpha}_{jk}]=\alpha^\circ_{jk}+O\left(\sqrt{\frac{\log N}{n_j}}(\frac{n_j}{N}+\sqrt{\frac{\log N}{N}})^\beta\right)$,

(iii) $\mathrm{Var}(\widehat{\alpha}^*_{jk})=\frac{\sigma_z^2}{N}$,

Finally, we need to establish the order of the variance of $\widehat{\beta}_{jk}$.

PROPOSITION 4. *Consider the coefficient $\widehat{\beta}_{jk}$ as defined in (25). Under assumptions (A1) and (A2) we have, uniformly over $(j,k)\in\mathcal{J}_N$:*

$$\mathrm{Var}(\widehat{\beta}_{jk})=O\left(\frac{n_j}{N}\left(\frac{\log N}{N}\right)^\beta\right) .$$

4.3. *Total error of approximation.* The next proposition provides us with the total error of approximation made when approximating $\widehat{\alpha}_{jk}$ by $\widehat{\alpha}^*_{jk}$ and summing over the indices (j,k) belonging to a well-defined set $\mathcal{J}_N$.

PROPOSITION 5. *Assume (A1) through (A3). Consider the set $\mathcal{J}_N$ as defined by expression (12), with δ such that*

$$\frac{\beta}{2\beta+1}+\frac{1}{4}<\delta<\frac{2\beta}{2\beta+1} .$$

We then obtain the following rates of convergence:

(i) $\sum_{(j,k)\in\mathcal{J}_N}E[(\widehat{\rho}_{jk}-\widehat{\rho}^*_{jk})^2]=o(N^{-\frac{2\beta}{2\beta+1}})$,

(ii) $\sum_{(j,k)\in\mathcal{J}_N}(E[\widehat{\alpha}_{jk}-\widehat{\alpha}^*_{jk}])^2 = o(N^{-\frac{2\beta}{2\beta+1}})$,

(iii) $\sum_{(j,k)\in\mathcal{J}_N}\mathrm{Var}(\widehat{\beta}_{jk}) = o(N^{-\frac{2\beta}{2\beta+1}})$,

from which it follows immediately that

$$T_{2,1} = \sum_{(j,k)\in\mathcal{J}_N} E[(\widehat{\alpha}_{jk}-\widehat{\alpha}^*_{jk})^2] \;=\; o\left(N^{-\frac{2\beta}{2\beta+1}}\right)\ .$$

Using the tools developed to prove Proposition 5, we obtain the following estimate on the order of the term T_3 in the decomposition (20).

PROPOSITION 6. *Under the assumptions of Proposition 5, we have:*

$$\sum_{(j,k)\in\mathcal{J}_N} E[(\alpha^{\circ}_{jk}-\widehat{\beta}_{jk})^2] = o\left(N^{-\frac{2\beta}{2\beta+1}}\right)\ .$$

5. Asymptotic Gaussian Approximation

5.1. *Introduction.* We now consider only the model (22) with deterministic design, and we aim at proving the result (R2). This is relatively easy since we are back to the situation of a regular design q_t. This allows us to use the techniques developed in Neumann (1996) and Neumann and von Sachs (1995) to show that our estimator attains the 'classical' rates of convergence of the L_2-risk.

The method described in Neumann (1996) and Neumann and von Sachs (1995) consists in showing that the L_2-risk is asymptotically equivalent to the same risk in an accompanying Gaussian noise model. To achieves this, a large deviation principle is used to show the asymptotic normality of the wavelet coefficients $\widehat{\alpha}^*_{jk}$, see equation (31) below. Then, with this Gaussian approximation, we can use known results in Gaussian regression (Donoho and Johnstone, 1998, 1994) to obtain the exact analogue of our main Theorem 1, formulated now for the deterministic design model (22) with wavelet coefficients $\widehat{\alpha}^*_{jk}$. This is done in Section 5.3, where we also indicate how to chose a data-driven threshold. The proofs of the results in this section being largely similar to the ones of Neumann (1996), Neumann and von Sachs (1995) and Dahlhaus and Neumann (2000), we do not include them in this paper, but instead refer to Delouille *et al.* (2000).

5.2. *Methodology to prove the Gaussian approximation.* We present in this section the techniques needed to prove the Gaussian approximation. For sake of clarity, we repeat the conditions of Neumann and von Sachs (1995, Section 2.1) necessary for our setting of independent, but not necessarily homoscedastic errors. Generalisations of this technique to the correlated case, can be found in Neumann and von Sachs (1995, Section 2.2), and in von Sachs and MacGibbon (2000).

In the fixed design model $Y_i^* = m(q_i) + Z_i^{'}$, $i = 1, \ldots, N$, we assume the following.

(B1) $\int_{-\infty}^{q_i} f_x(t)\, dt \;=\; i/N$.

(B2) The errors $Z_i^{'}$ are independently, not necessarily identically distributed with $C_1 \;\leq\; \sigma_i^2 \;:=\; \mathrm{Var}(Z_i^{'}) \;\leq\; C_2$.

(B3) $E[|Z_i^{'}|^K] \;\leq\; C_K$ for all $K \in \mathbb{N}$ and for some fixed constants $C_K < \infty$.

(B4) At level j_0, uniformly in k, $\varphi^{\circ}_{j_0,k}$ and $\psi^{\circ}_{j_0,k}$ are of bounded total variation on I.

For our asymptotic Gaussian approximation, we consider the model

$$G_i \;=\; m(q_i) \;+\; \xi_i, \quad i = 1, \ldots, N, \tag{29}$$

where the ξ_i's are independent with $\xi_i \;\sim\; N(0, \sigma_i^2)$. For simplicity of notation, let again $\psi_{j_0-1,k} := \varphi_{j_0,k}$ and $\alpha_{j_0-1,k} := s_{j_0,k}$, where j_0 is the primary resolution level.

Assume (B1) through (B4) and define the quantity

$$\overline{\alpha}_{jk} \;=\; \sum_i (q_i - q_{i-1})\; w_{jk}(i)\; G_i, \tag{30}$$

where $w_{jk}(i) \;=\; \frac{\psi^{o}_{jk}(q_i)}{N\ (q_i - q_{i-1})}$. Then, on an appropriate probability space, it is possible to couple $\widehat{\alpha}^*_{jk}$ and $\overline{\alpha}_{jk}$, that is, to show that $P((|\widehat{\alpha}^*_{jk} \;-\; \overline{\alpha}_{jk}| > N^{\gamma-\delta/2-1/2}) = O(N^{-\lambda})$, for any $\gamma > 0$, $\lambda < \infty$, and for any coefficients $(j,k) \;\in\; \mathcal{J}_N \;=\; \{(j,k) \mid\; 2^j \leq N^{1-\delta}\}$. Now, let $\Phi(x)$ denote the standard normal distribution function. Using the bias and variance behaviour of $\widehat{\alpha}^*_{jk}$ as described in Proposition 3, the coupling of $\widehat{\alpha}^*_{jk}$ and $\overline{\alpha}_{jk}$, and the Corollary 4 in Sakhanenko (1991), we obtain the asymptotic normality of coefficients $\widehat{\alpha}^*_{jk}$ as follows.

Proposition 7. *For arbitrary large $\lambda \;>\; 0$, we have, with $\sigma^2_{jk} \;:=$ $\mathrm{Var}(\widehat{\alpha}^*_{jk})$, that*

$$P\left(\pm(\widehat{\alpha}^*_{jk} \;-\; \alpha^{\circ}_{jk})/\sigma_{jk} \geq x\right) = (1 - \Phi(x))(1 + o(1)) \;\; + O(N^{-\lambda}) \tag{31}$$

holds uniformly in $(j,k) \in \mathcal{J}_N$ *and uniformly over* $x \in \mathbb{R}$.

The strong form of asymptotic normality (31) proven, we can justify the l_2-risk equivalence between our model and the accompanying Gaussian model (29). Written in the wavelet domain, the model (29) becomes

$$\overline{\alpha}_{jk} = \alpha^{\circ}_{jk} + \varepsilon_{jk}, \ (j,k) \in \mathcal{J}_N,$$

where $\varepsilon_{jk} \sim N(0, \sigma^2_{j,k})$. The following theorem parallels Theorem 4.1 of Dahlhaus and Neumann (2000). It states the l_2-risk equivalence between Gaussian and non-Gaussian model.

THEOREM 2. *Let* α°_{jk}, $\widehat{\alpha}^{*}_{jk}$, *and* $\overline{\alpha}_{jk}$ *be defined in (6), (23), and (30), respectively, and let* $\delta^S(., \lambda_{jk})$ *denote soft thresholding, with scale and location dependent threshold* λ_{jk}. *Then, with the set* $\mathcal{J}_N$ *defined in* (12), *we have*

$$\sum_{(j,k)\in\mathcal{J}_N} E\left(\delta^S(\widehat{\alpha}^{*}_{jk}, \lambda_{jk}) - \alpha^{\circ}_{jk}\right)^2$$
$$= \ (1 + o(1)) \sum_{(j,k)\in\mathcal{J}_N} E\left(\delta^S(\overline{\alpha}_{jk}, \lambda_{jk}) - \alpha^{\circ}_{jk}\right)^2 + O(N^{-1}) \ .$$

Theorem 2 allows us to replace the term $T_{2,2}$ in (R2) by

$$\sum_{(j,k)\in\mathcal{J}_N} E[\delta^S(\overline{\alpha}_{jk}, \lambda_{jk}) - \alpha^{\circ}_{jk}]^2 \ (1 + o(1)) + O(N^{-1}) \ ,$$

which completes the proof of the Gaussian approximation.

5.3. *Optimality of the unbalanced Haar wavelet estimator.* It remains to complete the discussion of the proof of our main Theorem 1 and to be more precise about the threshold choice. In the accompanying Gaussian model, it is known how to choose the random threshold λ_{jk} that will lead to a rate $(\log N/N)^{\frac{2\beta}{2\beta+1}}$ for the risk of the estimator $\widehat{m}$. Note that, by Lemma 1 of Donoho and Johnstone (1994), the following relation holds:

$$E[\delta^S(\overline{\alpha}_{jk}, \lambda_{jk}) - \alpha^{\circ}_{jk}]^2$$
$$\leq \ C \sum_{(j,k)\in\mathcal{J}_N} \left(\sigma^2_{jk}\left(\frac{\lambda_{jk}}{\sigma_{jk}} + 1\right)\varphi\left(\frac{\lambda_{jk}}{\sigma_{jk}}\right) + \min\{{\alpha^{\circ}_{jk}}^2, \lambda^2_{jk}\}\right),$$

where φ denotes the standard normal density. As stated in condition (A5) of Neumann (1996), if the individual thresholds λ_{jk} satisfy the condition

$$\begin{aligned} \textstyle\sum_{(j,k)\in\mathcal{J}_N}\left(\frac{\lambda_{jk}}{\sigma_{jk}}+1\right)\varphi\left(\frac{\lambda_{jk}}{\sigma_{jk}}\right) &= O(N^{1/(2\beta+1)}) \\ \max_{(j,k)\in\mathcal{J}_N}\{\lambda_{jk}\} &= O(N^{-1/2}\sqrt{\log N}) \end{aligned} \tag{32}$$

then the L_2-risk of $\widehat{m}$ attains the rate $(\log N/N)^{\frac{2\beta}{2\beta+1}}$. Condition (32) however, does not give a practical rule to choose the λ_{jk} from the data. Hence Neumann provided a sufficient condition (see Neumann, 1996, Condition A6) for random thresholds which ensures the desired rate for the estimator. This condition essentially says that the thresholds λ_{jk} must fulfill the inequality $\sigma_{jk}\sqrt{2\log(\sharp\mathcal{J}_N)} \le \lambda_{jk} \le CN^{-1/2}\sqrt{\log N}$, on an event with probability tending to 1 as $N \to \infty$. This gives both a practical rule for a data-driven choice of the threshold λ_{jk}, and provides the completion of the proof of Theorem 1 by exactly following the proof of Theorem 5.2 in Neumann (1996).

6. Numerical Results

6.1. *Introduction.* In Section 4, we fully justify the approximation of our initial model (1) by model (22) only in the case of i.i.d. errors. Now, the present section reports the results of simulations made to evaluate the robustness of the Unbalanced Haar method with respect to a change in the error model and to compare this method with the binned wavelet smoother of Antoniadis and Pham (1998). We thereby show through some empirical evidence that our estimator $\widehat{m}(x)$ based on random partitioning I_{jk} is performing well for three error models. In the first model, the errors $\epsilon_t^{(1)}$'s are i.i.d. $N(0,\sigma^2)$. In the second model, the $\epsilon_t^{(2)}$'s follow a stationary dependent $AR(1)$ model, $\epsilon_t^{(2)} + 0.9\ \epsilon_{t-1}^{(2)} = e_t$, e_t i.i.d. $N(0,1)$. In the third model, the errors are heteroscedastic: $\epsilon_t^{(3)} = \sigma(X_t)e_t$, with e_t i.i.d. $N(0,1)$.

The noise variance in the wavelet domain is classically estimated by a median absolute deviation from the median (called MAD estimator) of the detail coefficients, either at the highest level (for i.i.d. errors) or at each level (for stationary dependent errors), see Johnstone and Silverman (1997). In the heteroscedastic model, we need to estimate the conditional variance $\sigma_{jk}^2 := \mathrm{Var}(\widehat{\alpha}_{jk}|X_1,\ldots,X_N)$. Indeed, writing $\widehat{\alpha}_{jk}$ as $\widehat{\alpha}_{jk} = \frac{1}{N}\sum_t m(X_t)\psi_{j,k}(X_t) + \frac{1}{N}\sum_t \sigma(X_t)e_t\psi_{j,k}(X_t) = A + B$, it is clear that we should threshold on the basis of $\mathrm{Var}(B)$ only. The asymptotic variance of B,

$AVar(B) = AVar(\widehat{\alpha}_{jk}) = \frac{1}{N}\int \sigma^2(x)\psi_{j,k}^2(x)dF_x$ gives us a straightforward way to estimate σ_{jk}. Simply taking its plug-in estimator gives

$$\widehat{\sigma}_{jk}^2 = \widehat{AVar}(\widehat{\alpha}_{jk}) = \frac{2^j}{N^2}\sum_{X_t \in I_{j,k}} \widehat{\sigma}^2(X_t). \tag{33}$$

We will make use of this estimator in the computation of the threshold λ_{jk}.

In the next section we present our denoising algorithm. Then in Section 6.3 the set-up of the simulation study is introduced. Finally, Section 6.4 presents and comments the results.

6.2. *Denoising algorithm.* As has already been mentioned in the global introduction, the Unbalanced Haar wavelets are non smooth and hence produce a rough, piecewise constant estimator. To improve the visual quality of the reconstruction we therefore used the discrete non-decimated wavelet transform (NDWT), as introduced for example in Coifman and Donoho (1995) and Nason and Silverman (1995), based on the Unbalanced Haar basis. The principle of the NDWT is to take the average of N shifted decimated wavelet transform (DWT). To allow for non-periodic functions on the interval, we left out the shifts at the boundaries on a length equal to the support of the wavelets present at the boundaries. To obtain a non-linear wavelet estimator, we used hard-thresholding. The primary resolution level j_0 is always taken equal to two.

We propose here an algorithm, which is in principle tailored to the heteroscedastic regression, but which can be used for the three error models presented above. Theorem 1 tells us it is optimal (in a minimax sense) to use a threshold $\lambda_{j,k}$ such that $\sigma_{jk}\sqrt{2\log(\#\mathcal{J}_N)} \leq \lambda_{jk} \leq C\, N^{-1/2}\sqrt{\log(N)}$ with $\sigma_{j,k}^2 = \mathrm{Var}(\widehat{\alpha}_{jk}|X_1,\ldots,X_N)$. Based on this observation and expression (33), we propose an algorithm which proceeds in three steps.

STEP 1. Obtain a pilot estimator of $m(x)$ using a linear non-decimated wavelet estimator, where the cut-off scale j_1 is chosen by a rule of thumb: take $j_1 := \lfloor \log_2(N)/2 \rfloor$. Denote this pilot estimator by $\widehat{m}_{(0)}(x)$. Let $i := 0$.

STEP 2. Take residuals $\eta_t := Y_t - \widehat{m}_i(X_t)$ and estimate $\sigma^2(.)$ from the data set $(X_t, \eta_t^2), t = 1,\ldots,N$ using a linear NDWT, with the same cut-off level as in step one. Call this estimate $\widehat{\sigma}_{(i)}^2(x)$. Estimate m by non-linear thresholding estimator, again based on a NDWT, with a threshold $\lambda_{j,k}$ depending on the location and scale. We compute this threshold as $\lambda_{j,k} = \widehat{\sigma}_{j,k}\sqrt{2\log N}$, where $\widehat{\sigma}_{j,k}$ is given by (33): $\frac{2^j}{N^2}\sum_{X_t \in I_{j,k}} \widehat{\sigma}_{(i)}^2(X_t)$. Call the resulting estimator $\widehat{m}_{(i+1)}(x)$.

STEP 3. Repeat step 2 for $i := 1$ and take as final estimator $\widehat{m}_{(2)}(x)$.

Naturally, other possibilities are presented in the literature to estimate a heteroscedastic model, see for example von Sachs and MacGibbon (2000) and Kovac and Silverman (2000).

6.3. *Parameters of the simulation study and performance criterion.* In this simulation study, the regressors $\{X_t\}$ will always be independent and normally distributed random variables, $X_t \sim N(0.5, (0.1)^2)$. We use the design-adapted basis given in (8)-(9). The three error models presented in Section 6.1 are considered. For each type of error, we consider three different functions: a smooth sine function (Sine), a function with one jump (Jump) and a piecewise-constant function (Blocks). These are represented in Figure 1. The sample sizes selected were $N = 256$ (intermediate size) and $N = 512$ (large sample size). In order to specify the variance σ_ϵ^2 of the errors, we define the signal-to-noise ratio as SNR:= $\mathrm{sd}(m)/\sigma_\epsilon$, where

$$\mathrm{sd}(m)^2 = \frac{1}{N-1}\sum_{t=1}^{N}(m(x_t) - \overline{m(x)})^2, \tag{34}$$

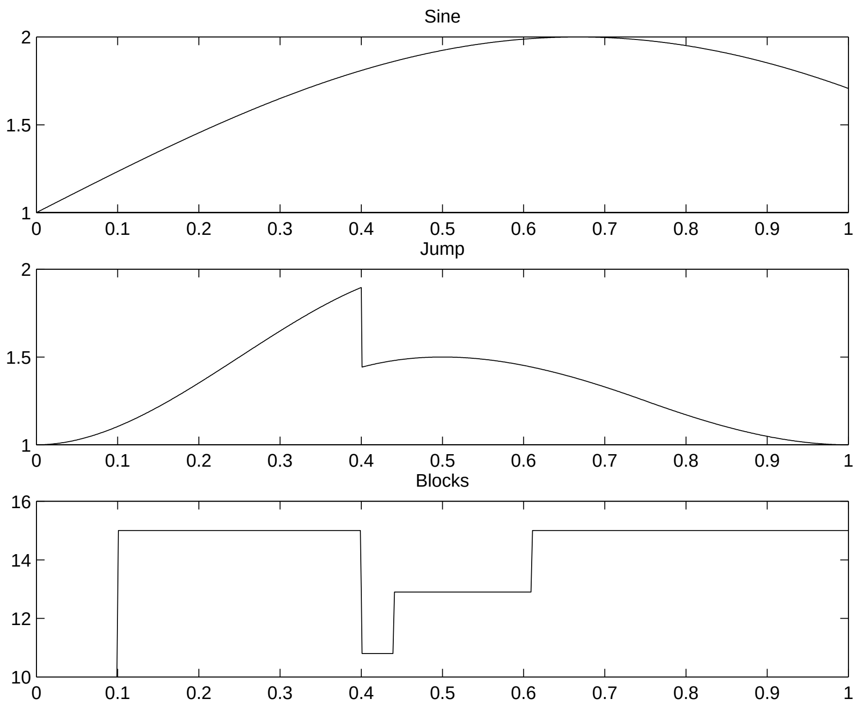

Figure 1. Underlying signals: smooth sine function, function containing a jump and blocks function.

for a realization $\{x_t\}_{t=1}^N$ of the random variable $X \sim N(0.5, 0.01)$, and where $\overline{m(x)}$ is equal to the mean value of the $m(x_t)$, $t = 1, \ldots, N$. We set the SNR equal to one, which allows us to specify σ_ϵ. In the first error model, we took, $\sigma_\epsilon := \sigma$. For the second model, we generated Gaussian autoregressive errors $\epsilon_t^{(2)} \sim AR(1)$ and took $\sigma_\epsilon := \text{sd}(\epsilon_t^{(2)})$, with sd(.) computed as in (34). Finally, for the heteroscedastic model, $\sigma_\epsilon := \int \sigma(x)dx$, and this integral is approximated by averaging $\sigma(x)$ over discrete x. The function $\sigma(x)$ used in the simulations is $\sigma(x) = C(0.001 + (0.3 - 1.5(x - 0.515)^2)^2)$, (see Gao, 1997), where the constant C is chosen to have SNR $= 1$. This function $\sigma(x)$ introduces a larger variance around the middle of the interval $[0, 1]$ and a smaller variance at the boundaries. To evaluate the performance of the methods on one given data set, we use as a criterion the relative mean square error:

$$\text{MSE} = \frac{1}{N} \sum_{t=1}^{N} \left(\frac{\widehat{m}(x_t) - m(x_t)}{m(x_t)} \right)^2, \tag{35}$$

where $\widehat{m}(x_t)$ is the estimated function and $m(x_t)$ is the real underlying function computed at x_t. When computing equation (35) for a large number of, say B, data sets, we need to summarize the information given by these B runs. We choose to compute the median, first and third quartiles over the B relative MSE values. We then took the square root of these three quantities.

For the binned wavelet smoother of Antoniadis and Pham (1998) (hereafter called BinWav), we took B-splines of order 3. For each data set, we choose the best cutting level (or, in other words, the best binwidth) from four values, using as a criterion the relative MSE defined in (35). We are indebted to A. Antoniadis for supplying the BinWav code.

6.4. *Simulation results.* Table 1 gives the results of the simulation study for the proposed NDWT algorithm based on Haar, using the above three-steps procedure, as well as for the BinWav estimator.

The BinWav estimator uses smooth B-splines, hence it gives very good results for a smooth function like the sine function and outperforms our estimator in this case. Figure 2(a) shows, in case of dependent errors, where this difference lies: it is mainly at the boundaries, where there are few data points, that the BinWave estimator gives better results. Note that we tried to illustrate the results of Table 1 by giving the graphs of estimates having a value of the square root of the MSE, denoted RMSE, approximately equal to the median RMSE value reported in Table 1.

Table 1. Results of the simulation study for the three errors models (I.I.D., dependent and heteroscedastic), using the method described in Section 6.2 and the method of Antoniadis and Pham (denoted by AP). For each function and each sample size, the first line presents the square root of the median of the relative MSE (called 'median RMSE'). The numbers between brackets give the square root of the inter-quartile intervals of the MSE. Values are expressed in percent.

	IID	IID (AP)	Dep.	Dep.(AP)	Hete.	Hete. (AP)
Sine,	1.40	1.02	1.53	0.47	2.18	1.71
$N = 256$	[1.17; 1.72]	[0.81; 1.25]	[1.28; 1.81]	[0.41; 0.55]	[1.92; 2.51]	[1.34; 2.12]
Sine,	1.09	0.84	1.03	0.33	2.08	1.36
$N = 512$	[0.91; 1.30]	[0.69; 1.03]	[0.77; 1.25]	[0.28; 0.39]	[1.85; 2.31]	[1.11; 1.61]
Jump,	2.89	3.62	2.76	2.63	4.23	4.61
$N = 256$	[2.56; 3.17]	[3.37; 3.88]	[2.55; 2.94]	[2.47; 2.84]	[3.71; 4.83]	[4.19; 4.83]
Jump,	2.35	2.99	2.36	2.00	3.54	3.80
$N = 512$	[2.12; 2.66]	[2.83; 3.23]	[2.12; 2.57]	[1.94; 2.07]	[3.19; 3.92]	[3.5; 4.12]
Blocks,	2.91	4.48	2.48	3.38	5.76	5.97
$N = 256$	[2.61; 3.27]	[4.21; 4.81]	[2.20; 2.88]	[3.21; 3.53]	[5.21; 6.41]	[5.54; 6.51]
Blocks,	2.18	3.74	1.87	2.65	4.69	4.83
$N = 512$	[1.96; 2.42]	[3.54; 3.93]	[1.69; 2.08]	[2.5; 2.79]	[4.17; 5.27]	[4.48; 5.21]

For the jump function, our estimates give better RMSE values, except in the case of dependent errors, where the BinWav estimator performs slightly better. Figure 2(b) illustrates how the estimators behave for one typical data set with heteroscedastic errors. Due to the high variance of the noise, some wrinkles appear in the estimators. It is also clear that the BinWav estimator smoothes out the jump more than our estimator.

The case of the block function is the least favourable for the BinWav method, whereas our estimator copes pretty well with jumps. This explains the better results of our method for that function. In Figure 2(c) we see the artifacts introduced by the BinWav estimator. Note that for functions with jumps, in the situation of heteroscedastic errors no estimator will perform as well as for the less challenging error structures such as i.i.d. or stationary dependent errors.

To conclude, this study shows that the algorithm proposed in this paper gives reasonably good results for smooth or piecewise smooth functions, and for a large variety of error models.

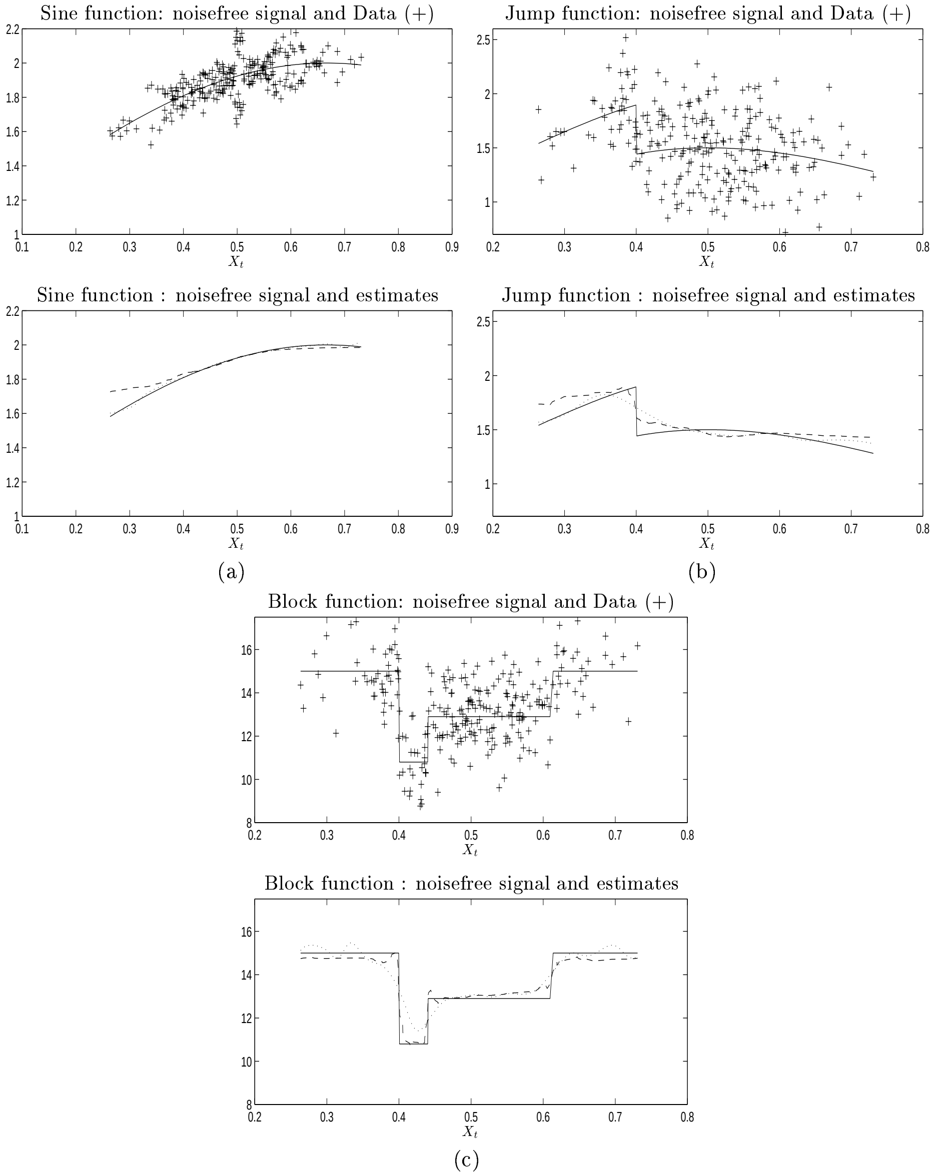

Figure 2. Estimation of the Sine (a), Jump (b), and Block (c) functions with $N = 256$ data points. The first part of the graphs represents the underlying signal and the data (represented by '+'), generated with (a) stationary dependent errors, (b) independent heteroscedastic errors and (c) i.i.d. errors. The second part shows the underlying signal (plain line), our estimator (dashed line) and the BinWav estimator (dotted line). The RMSE values were, respectively for both methods, equal to (a) 1.47% and 0.44%, (b) 4.32% and 4.55% and (c) 2.88% and 4.51%.

7. Proofs Related to Sections 3 and 4

7.1. *Results needed to complete the proof of Theorem 1 of Section 3.*

Proposition 8. *Consider the model $Y_t = m(q_t) + Z_t$, where $q_t = F^{-1}(t/N)$, $E(Z_t) = 0$ and $Var(Z_t) = \sigma_z^2$. Let $m \in \Lambda^\beta(L)$ with $0 < \beta < 1$, and define the wavelet basis ψ°_{jk} in the space $L_2(dF_x)$ as in equation (4). Assume (A1)-(A3). Let $2^{j_1^{opt}} \asymp N^{1/(2\beta+1)}$. Then the linear wavelet estimator*

$$\widehat{m}^{lin}(x) = \sum_k \hat{s}_{j_0,k}\varphi^\circ_{j_0,k}(x) + \sum_{j=j_0}^{j_1^{opt}} \sum_k \hat{\alpha}_{jk}\psi^\circ_{jk}(x) \ ,$$

with $\hat{s}_{j_0,k} = \frac{1}{N}\sum_{t=1}^N Y_t\varphi^\circ_{j_0,k}(q_t)$ and $\hat{\alpha}_{j,k} = \frac{1}{N}\sum_{t=1}^N Y_t\psi^\circ_{j,k}(q_t)$, attains the optimal rate of convergence of $N^{-\frac{2\beta}{2\beta+1}}$ using the $L_2(dF_x)$-risk.

Proof. Using the expansion (5) of $m(x)$, the definition of $s^\circ_{j_0,k}$ and α°_{jk} as in (6), and Parseval's equality, our L_2-risk can be re-written as

$$\begin{aligned} &E\left\|\widehat{m}^{\text{lin}}(x) - m(x)\right\|^2_{L_2(dF_x)} \\ &= \sum_k E\left|\hat{s}_{j_0,k} - s^\circ_{j_0,k}\right|^2 + \sum_{j=j_0}^{j_1}\sum_k E\left|\hat{\alpha}_{j,k} - \alpha^\circ_{j,k}\right|^2 + \sum_{j=j_1+1}^{\infty}\sum_k \left|\alpha^\circ_{jk}\right|^2 \\ &= S_1 + S_2 + S_3 \ . \end{aligned}$$

If we prove the two following relations:

$$E[(\hat{\alpha}_{jk} - \alpha^\circ_{jk})^2] = \frac{\sigma_z^2}{N} + o(N^{-1}) \tag{36}$$

$$|\alpha^\circ_{jk}| := \left|\int m(x)\psi^\circ_{jk}(x)dF_x\right| \le C2^{-j(\beta+1/2)} \ , \tag{37}$$

it will then follow that $S_1 + S_2 \le 2^{j_1}\ (\frac{\sigma_z^2}{N} + o(N^{-1}))$ and $S_3 \le C2^{-2\beta j_1}$. The cutting level that minimises the L_2-risk is such that the two terms $2^{j_1}\sigma_z^2/N$ and $2^{-2\beta j_1}$ are of the same order. Hence we obtain

$$2^{-j_1^{\text{opt}}} = C''(\frac{\sigma_z^2}{N})^{\frac{1}{1+2\beta}} = O(N^{1/(1+2\beta)}) \tag{38}$$

Replacing (38) in the expression of the L_2-risk gives $E\left\|\hat{m} - m\right\|^2_{L_2(dF_x)} \le C'(\frac{\sigma_z^2}{N})^{\frac{2\beta}{1+2\beta}}$.

It is straightforward to prove (36) since $E(\widehat{\alpha}_{jk}-\alpha^{\circ}_{jk})$ is of order $O(N^{-1})$ and $\mathrm{Var}(\hat{\alpha}_{jk})=\frac{\sigma_z^2}{N}$ by definition of the Haar wavelets $\psi^{\circ}_{jk}(x)$. We now prove the relation (37), that is, the decay of the coefficient α°_{jk}. As $\int\psi^{\circ}_{jk}dF_x=0$,

$$\begin{aligned}\left|\langle m,\psi^{\circ}_{jk}\rangle_{dF_x}\right| &= \left|\int m(x)\psi^{\circ}_{jk}(x)dF_x(x)-m(x_0)\int\psi^{\circ}_{jk}(x)dF_x(x)\right|\\ &\leq \int|m(x)-m(x_0)||\psi^{\circ}_{jk}(x)|dF_x(x)\ ,\end{aligned}$$

where we take $x_0=q_{\frac{2k}{2^{j+1}}}$. As $m\in\Lambda^{\beta}(L)$, we have

$$|\langle m,\psi^{\circ}_{jk}\rangle_{dF_x}|\leq M\int|x-x_0|^{\beta}2^{j/2}|1_{I^{\circ}_{j+1,2k}}(x)-1_{I^{\circ}_{j+1,2k+1}}(x)|dF_x(x). \quad (39)$$

Let us concentrate on the term $\int|x-x_0|^{\beta}1_{I^{\circ}_{j+1,2k}}(x)dF_x(x)$, as the other one will be treated in the same way. Consider the change of variables $y=F_x(x)$ and let $y_0=F_x(x_0)$. The previous integral can be re-expressed as:

$$\int_{F_x(q_{(2k-1)/2^{j+1}})}^{F_x(q_{2k/2^{j+1}})}|F_x^{-1}(y)-F_x^{-1}(y_0)|^{\beta}dy. \quad (40)$$

As $f_x(x)$ is bounded away from zero on the support S by assumption (A1), $F_x^{-1}(y)$ is differentiable with $F_x^{-1}(y_0)'=f_x(x_0)^{-1}$. Hence we can use a Taylor approximation for $F_x^{-1}(y)-F_x^{-1}(y_0)$. Integrating out this Taylor expansion, we obtain that (40) is approximated, up to a term of higher order, by $\frac{1}{f_x(x_0)^{\beta}(\beta+1)}\left|F_x(q_{\frac{2k}{2^{j+1}}})-F_x(q_{\frac{2k-1}{2^{j+1}}})\right|^{\beta+1}=\frac{2^{-(j+1)(\beta+1)}}{f_x(x_0)^{\beta}(\beta+1)}$. Inserting this last equation into (39), we get:

$$|\langle m,\psi^{\circ}_{jk}\rangle_{dF_x}|\leq M2^{j/2}\frac{2}{(\beta+1)f(x_0)^{\beta}}2^{-(j+1)(\beta+1)}=C2^{-j(\beta+\frac{1}{2})}\ ,$$

which ends the proof.

PROPOSITION 9. *Consider the coefficients* $\hat{s}_{j_0,k}$ *and* $\hat{b}_{j_0,k}$ *defined in* (10) *and* (13), *respectively. Then*

$$\sum_k E(\widehat{s}_{j_0,k}-\widehat{b}_{j_0,k})^2 \;=\; o\left(N^{-\frac{2\beta}{2\beta+1}}\right)\ .$$

PROOF. Let $\widehat{A}_{j_0,k}:=\hat{s}_{j_0,k}-\hat{b}_{j_0,k}$ and $E(\widehat{A}^2_{j_0,k})=E(\widehat{A}_{j_0,k})^2+\mathrm{Var}(\widehat{A}_{j_0,k})$. Using the conditional expectation, we have

$$E(\widehat{A}_{j_0,k})=E\left(\frac{1}{N}\sum_{t=1}^{N}E(Y_t|X_1,\ldots,X_N)\psi_{jk}(X_t)-\frac{1}{N}\sum_{t=1}^{N}m(X_t)\psi_{jk}(X_t)\right)=0\,.$$

It follows that the variance $\mathrm{Var}(\hat{A}_{j_0,k})$ is equal to $E(\mathrm{Var}(\hat{A}_{j_0,k}|X_1,\ldots,X_N))$ and

$$\mathrm{Var}(\hat{A}_{j_0,k}|X_1,\ldots,X_N) = \mathrm{Var}(\hat{s}_{j_0,k}|X_1,\ldots,X_N) = \frac{1}{N^2}\sum_{t=1}^{N} \sigma_z^2 \varphi_{j_0,k}^2(X_t) \ .$$

Taking expectation, we have

$$E(\mathrm{Var}(\hat{s}_{j_0,k}|X_1,\ldots,X_N)) = \frac{\sigma_z^2}{N}\left(\frac{1}{N}\sum_{t=1}^{N} 2^{j_0} E(1_{j_0,k}(X_t))\right) = \frac{\sigma_z^2}{N} \ ,$$

since $\sum_{t=1}^{N} E(1_{j_0,k}(X_t)) = N2^{-j_0}$. Since j_0 is taken smaller than $\log_2(N)$, we obtain

$$\sum_{k=1}^{2^{j_0}} E(\hat{s}_{j_0,k} - \hat{b}_{j_0,k})^2 = 2^{j_0}\frac{\sigma_z^2}{N} = o\left(N^{-\frac{2\beta}{2\beta+1}}\right) \ .$$

7.2. *Proofs of the results of Section 4.*

7.2.1. *Proof of Proposition 1.* (i) As $E(Z_t) = 0$ and the regressors are i.i.d. in (1), $E(\widehat{\alpha}_{jk}^{\circ}) = E(m(X_1)\psi_{jk}^{\circ}(X_1)) = \alpha_{jk}^{\circ}$.

(ii) As $E(Z_t) = 0$, by the independence of the X_t, Z_t, and by orthonormality of ψ_{jk}° in $L^2(dF_x)$, we have

$$\begin{aligned}
\mathrm{Var}\ \widehat{\alpha}_{jk}^{\circ} &= \frac{1}{N}\mathrm{Var}(Y_1\psi_{jk}^{\circ}(X_1)) \\
&= \frac{1}{N}\left(\sigma_z^2 E(\psi_{jk}^{\circ}(X_1))^2 + \mathrm{Var}(m(X_1)\psi_{jk}^{\circ}(X_1))\right) \\
&= \frac{1}{N}\left(\sigma_z^2 + \mathrm{Var}(m(X_1)\psi_{jk}^{\circ}(X_1))\right) .
\end{aligned}$$

As $|\psi_{jk}^{\circ}| = |\varphi_{jk}^{\circ}|$, we have, using the definition of I_{jk}° as a quantile interval of F_x-mass 2^{-j} :

$$\begin{aligned}
\mathrm{Var}(m(X_1)\psi_{jk}^{\circ}(X_1)) &\le E(m(X_1)\psi_{jk}^{\circ}(X_1))^2 \\
&= 2^j\int_{I_{jk}^{\circ}} m^2(x)dF_x(x) \ \le \ \sup\{m^2(x); x\in I_{jk}^{\circ}\}.
\end{aligned}$$

7.2.2. *Proof of Proposition 2.* Using $\widehat{\rho}_{jk}^{\circ}$, defined in (28) as an intermediate coefficient, we write:

$$E[(\widehat{\rho}_{jk}^{*} - \widehat{\rho}_{jk})^2] \ \le \ 2\{E[(\widehat{\rho}_{jk}^{*} - \widehat{\rho}_{jk}^{\circ})^2] + E[(\widehat{\rho}_{jk} - \widehat{\rho}_{jk}^{\circ})^2]\}$$

and show that both terms are of order $\sigma_z^2\, O(\frac{2^j}{N^{3/2}})$.

(i) Using the particular choice of $Z'_1, \ldots, Z'_N$, as described in Section 4.1, we have: $\widehat{\rho}^\circ_{jk} - \widehat{\rho}^*_{jk} = \frac{1}{N}\sum_{t=1}^N Z_t \psi^\circ_{jk}(X_{(t)}) - Z'_t \psi^\circ_{jk}(q_t)$. As $Z'_1, \ldots, Z'_N$ are i.i.d. mean-zero random variables with values independent of the X_s, we obtain, with $n_j = N2^{-j}$:

$$\begin{aligned} E(\widehat{\rho}^\circ_{jk} - \widehat{\rho}^*_{jk})^2 &= \frac{\sigma_z^2}{N^2} \sum_{t=1}^N \mathrm{Var}(\psi^\circ_{jk}(X_{(t)}) - \psi^\circ_{jk}(q_t)) \\ &\leq \frac{2\sigma_z^2}{N n_j} \sum_{t=1}^N \big\{ E(1^\circ_{j+1,2k}(X_{(t)}) - 1^\circ_{j+1,2k}(q_t))^2 \\ &\qquad + E(1^\circ_{j+1,2k+1}(X_{(t)}) - 1^\circ_{j+1,2k+1}(q_t))^2 \big\} \\ &= 2\sigma_z^2 (V_1 + V_2). \end{aligned}$$

Changing notation from $(j+1, 2k)$ to (j,k), V_1 is of the form

$$V_1 = \frac{1}{N n_j} \sum_{t=1}^N E(1^\circ_{jk}(X_{(t)}) - 1^\circ_{jk}(q_t))^2 \leq \frac{1}{N n_j} \sum_{t=1}^N E\left(1_{\Delta jk}(X_{(t)})\right) + \frac{2}{N n_j}$$

as $1^\circ_{jk}(q_t) = 1_{jk}(X_{(t)})$ for all $t \notin \{(k-1)n_j, kn_j\}$ by definition of I_{jk} and I°_{jk}, where Δ_{jk} is defined as in Lemma 1 below. Using that Lemma, we have

$$\begin{aligned} V_1 &= \frac{1}{N n_j} \sum_{t=1}^N E\left(1_{\Delta jk}(X_t)\right) + \frac{2}{N n_j} \\ &= \frac{1}{N n_j} \sum_{t=1}^N P(X_t \in \Delta_{jk}) + \frac{2}{N n_j} \quad = \quad O\left(\frac{1}{n_j \sqrt{N}}\right) \end{aligned}$$

uniformly in j,k. As V_2 is of exactly the same form, we have the desired result for $E(\widehat{\rho}^\circ_{jk} - \widehat{\rho}^*_{jk})^2$.

(ii) For the second term we have

$$\begin{aligned} (\widehat{\rho}_{jk} - \widehat{\rho}^\circ_{jk})^2 &\leq 2\frac{2^j}{N^2} \left\{ \sum_{t=1}^N Z_t (1_{j+1,2k}(X_t) - 1^\circ_{j+1,2k}(X_t)) \right\}^2 \\ &\quad + 2\frac{2^j}{N^2} \left\{ \sum_{t=1}^N Z_t (1_{j+1,2k+1}(X_t) - 1^\circ_{j+1,2k+1}(X_t)) \right\}^2 \\ &= 2(S_1^2 + S_2^2). \end{aligned}$$

We only discuss the first term S_1^2 in detail. For sake of convenience, we change from $(j+1, 2k)$ to (j,k) in the definition of S_1^2. As the random partition $I_{jk},\ k=1,\dots,2^j$, is specified by the order statistics $X_{(1)} \le \dots \le X_{(N)}$, we get, conditioning on the latter and using $n_j 2^j = N$:

$$\begin{aligned} E(S_1^2) &= \frac{1}{Nn_j} E\left(E\left\{ \left(\sum_{t=1}^N Z_t (1_{jk}(X_t) - 1_{jk}^\circ(X_t)) \right)^2 |X_{(1)}, \dots, X_{(N)} \right\} \right) \\ &= \frac{1}{Nn_j} E\left(\sum_{t=1}^N \sigma_z^2\, E\left\{ \left(1_{jk}(X_t) - 1_{jk}^\circ(X_t)\right)^2 |X_{(1)}, \dots, X_{(N)} \right\} \right) \\ &= \frac{\sigma_z^2}{n_j} E\left((1_{jk}(X_1) - 1_{jk}^\circ(X_1))^2 \right) \ . \end{aligned}$$

By Lemma 1, with Δ_{jk} denoting the symmetric difference of I_{jk} and I_{jk}°, we conclude:

$$E\, S_1^2 = \frac{\sigma_z^2}{n_j} P(X_1 \in \Delta_{jk}) = \sigma_z^2\, O\left(\frac{1}{n_j \sqrt{N}} \right).$$

LEMMA 1. *Let $X_1, \dots, X_N$ be i.i.d. with distribution function F_x and density f_x. Let $I_{jk} = [X_{((k-1)n_j+1)}, X_{(kn_j+1)})$ and $I_{jk}^\circ = [q_{(k-1)n_j}, q_{kn_j})$ with $q_t = F_x^{-1}(\frac{t}{N})$ be defined as before with $n_j 2^j = N,\ k = 1, \dots, 2^j$, and let $\Delta_{jk} = (I_{jk} \setminus I_{jk}^\circ) \cup (I_{jk}^\circ \setminus I_{jk})$ denote the symmetric difference of the random quantile interval I_{jk} and the fixed quantile interval I_{jk}°. Then,*

$$P(X_1 \in \Delta_{jk}) = O\left(\frac{1}{\sqrt{N}} \right)$$

uniformly in j,k.

PROOF. By definition of I_{jk} and I_{jk}°, $X_1 \in \Delta_{jk}$ if either

$$X_{((k-1)n_j+1)} \le X_1 < q_{(k-1)n_j} \quad \text{or} \quad q_{(k-1)n_j} \le X_1 < X_{((k-1)n_j+1)}$$

$$\text{or} \quad X_{(kn_j+1)} \le X_1 < q_{kn_j} \quad \text{or} \quad q_{kn_j} \le X_1 < X_{(kn_j+1)}.$$

The probabilities of these four events can be treated in the same manner, and we discuss only the last case in detail. We decompose this event with respect to the values of the rank R_1 of X_1 in the sample. As we have

$X_1 = X_{(R_1)} < X_{(kn_j+1)}$, R_1 is at most kn_j:

$$\begin{aligned} P(q_{kn_j} \leq X_1 < X_{(kn_j+1)}) &= \sum_{s=1}^{kn_j} P(R_1 = s,\ q_{kn_j} \leq X_{(s)}) \\ &= \frac{1}{N} \sum_{s=1}^{kn_j} P(X_{(s)} \geq q_{kn_j}) \end{aligned}$$

as ranks and order statistics of an i.i.d. sample are independent and, by symmetry, $P(R_1 = s) = \frac{1}{N}$ for all s. Let B_{jk} denote a binomial random variable with parameter $(N, \frac{kn_j}{N})$. From the well-known distribution of the order statistic $X_{(s)}$ (compare, e.g., Serfling, 1980, Chapter 2.4) we have, using $F(q_{kn_j}) = \frac{kn_j}{N}$:

$$P(X_{(s)} \geq q_{kn_j}) = P(B_{jk} \leq s-1).$$

Therefore,

$$\begin{aligned} P \quad & (q_{kn_j} \leq X_1 < X_{(kn_j+1)}) = \frac{1}{N} \sum_{s=1}^{kn_j} \sum_{i=0}^{s-1} P(B_{jk} = i) \\ = \quad & \frac{1}{N} \sum_{i=0}^{kn_j} (kn_j - i)\ P(B_{jk} = i) = \frac{1}{N} E(kn_j - B_{jk})_+ \\ \leq \quad & \frac{1}{N} E|B_{jk} - kn| \leq \frac{1}{N} \sqrt{\operatorname{Var} B_{jk}} = \frac{1}{\sqrt{N}} \sqrt{\frac{kn_j}{N}\left(1 - \frac{kn_j}{N}\right)} \end{aligned}$$

by Jensen's inequality, using $E(B_{jk}) = kn_j$. As $x(1-x) \leq \frac{1}{4}$ for all $x \in [0,1]$, we have uniformly in j, k :

$$P(q_{kn_j} \leq X_1 < X_{(kn_j+1)}) \leq \frac{1}{2\sqrt{N}}$$

7.2.3. *Proof of Proposition 3.* (i) Let $g_{jk}(u) = m(F_x^{-1}(u))\psi_{jk}^{\circ}(F_x^{-1}(u))$. Using this abbreviation, we make the change of variables $u = F_x(x)$ and we get, by definition of α_{jk}°, q_t and for $1 < k < 2^j$:

$$|E\widehat{\alpha}_{jk}^* - \alpha_{jk}^{\circ}| \;=\; \left| \frac{1}{N} \sum_{t=1}^{N} g_{jk}\left(\frac{t}{N}\right) - \int_0^1 g_{jk}(u)du \right| \;\leq\; \frac{1}{N} V(g_{jk}; 0, 1)\ ,$$

using an inequality of Pólya and Szegö for functions of bounded variation in Part II, Chapter 1, §2, n° 9 of Pólya and Szerö (1979). As $|\psi^\circ_{jk}(x)| = 0$ for $x \notin I^\circ_{jk}$ and $|\psi^\circ_{jk}(x)| = 2^{j/2}$ for $x \in I^\circ_{jk} = [q_{(k-1)n_j}, q_{kn_j})$, we have

$$\begin{aligned} V(g_{jk};0,1) &\leq 2^{j/2}\sup\Big\{\sum_{i=1}^{l} |m(x_i) - m(x_{i-1})|; \\ &\qquad\qquad l \geq 1, q_{(k-1)n_j} \leq x_0 < \ldots x_l \leq q_{kn_j}\Big\} \\ &= 2^{j/2} V(m; q_{(k-1)n_j}, q_{kn_j}). \end{aligned}$$

(ii) By Proposition 1, $E(\widehat{\alpha}_{jk}) - \alpha^\circ_{jk} = E(\widehat{\alpha}_{jk} - \widehat{\alpha}^\circ_{jk})$. We then obtain, with $n_j = N2^{-j}$:

$$\begin{aligned} E\quad(\widehat{\alpha}_{jk} - \widehat{\alpha}^\circ_{jk}) &= \frac{1}{N}\sum_{t=1}^{N} E\left(m(X_t)(\psi_{jk}(X_t) - \psi^\circ_{jk}(X_t))\right) \\ &= \frac{1}{\sqrt{Nn_j}}\sum_{t=1}^{N} E(m(X_t)(1_{j+1,2k}(X_t) - 1^\circ_{j+1,2k}(X_t) \\ &\qquad - (1_{j+1,2k+1}(X_t) - 1^\circ_{j+1,2k+1}(X_t)))) \\ &= E_1 - E_2. \end{aligned}$$

We study E_1 only, as E_2 can be treated in exactly the same manner. Again, we change notation from $(j+1, 2k)$ to (j,k) for sake of simplicity. Let $\tau = (k - \frac{1}{2})n_j$ be such that q_τ is in the center of I°_{jk} dividing the interval into two halves of equal F_x-mass. We have

$$\begin{aligned} E_1 &= \frac{1}{\sqrt{Nn_j}}\sum_{t=1}^{N} E\left(m(X_t)(1_{jk}(X_t) - 1^\circ_{jk}(X_t))\right) \\ &= \frac{m(q_\tau)}{\sqrt{Nn_j}} E\left(\sum_{t=1}^{N}(1_{jk}(X_t) - 1^\circ_{jk}(X_t))\right) \\ &+ \frac{1}{\sqrt{Nn_j}}\sum_{t=1}^{N} E\left((m(X_t) - m(q_\tau))(1_{jk}(X_t) - 1^\circ_{jk}(X_t))\right) . \end{aligned}$$

The first term on the right-hand side vanishes as $1_{jk}(X_t) = 1$ for exactly n_j values of t by definition of I_{jk} and as $\sum_{t=1}^{N} 1^\circ_{jk}(X_t)$ is binomially distributed with parameter $(N, \frac{n_j}{N})$ by definition of I°_{jk}.

Let $U_t := F_x(X_t),\ t = 1, \ldots, N$. Hence $U_t \sim U(0,1)$. Let

$$\begin{aligned} \chi_{jk}(u) &= 1_{jk}(F_x^{-1}(u)), \\ \chi^\circ_{jk}(u) &= 1^\circ_{jk}(F_x^{-1}(u)) \end{aligned}$$

be the indicator functions in terms of the U_t, i.e.

$$\begin{aligned} \chi_{jk}(u) = 1 \quad &\text{iff} \quad U_{((k-1)n_j+1)} \le u < U_{(kn_j+1)}, \\ \chi^\circ_{jk}(u) = 1 \quad &\text{iff} \quad \frac{(k-1)n_j}{N} \le u < \frac{kn_j}{N}, \end{aligned}$$

and let $g(u) = m(F_x^{-1}(u))$. Then, we may rewrite E_1 as

$$E_1 = \frac{1}{\sqrt{Nn_j}} \sum_{t=1}^{N} E(g(U_t) - g(\frac{\tau}{N}))(\chi_{jk}(U_t) - \chi^\circ_{jk}(U_t)).$$

By Lemma 2 below, we have

$$\begin{aligned} & P\left(|U_{(s)} - \frac{s}{N}| > \sqrt{\frac{2\log N}{N}} \text{ for any } s = 1, \ldots, N\right) \\ \le & \sum_{s=1}^{N} P\left(|U_{(s)} - \frac{s}{N}| > \sqrt{\frac{2\log N}{N}}\right) \\ \le & \frac{N}{N^4} \quad = \quad \frac{1}{N^3} \end{aligned}$$

which is summable over N. Therefore, the Borel-Cantelli Lemma implies that

$$P(|U_{(s)} - \frac{s}{N}| \le \sqrt{\frac{2\log N}{N}} \text{ for all } s = 1, \ldots, N) = 1$$

for all N large enough. For those N, we therefore have $\chi_{jk}(u) - \chi^\circ_{jk}(u) = 0$ with probability 1 for all $u \in [0,1]$ with $|u - \frac{(k-1)n_j}{N}| > \sqrt{\frac{2\log N}{N}} + \frac{1}{N}$ and $|u - \frac{kn_j}{N}| > \sqrt{\frac{2\log N}{N}} + \frac{1}{N}$, which implies by the uniformity of the U_t that

$$E|\chi_{jk}(U_t) - \chi^\circ_{jk}(U_t)| \le 4\left(\sqrt{\frac{2\log N}{N}} + \frac{1}{N}\right).$$

Under assumptions (A1) through (A3), Lemma 3 below shows that $|g(u) - g(v)| \le \tilde{H}|u - v|^\beta$ for all $u, v \in \left[\frac{(k-1)n_j-1}{N} - \sqrt{\frac{2\log N}{N}}, \frac{kn_j+1}{N} + \sqrt{\frac{2\log N}{N}}\right]$

for N large enough. Using this Hölder condition for g and the fact that $|u - \frac{\tau}{N}| \le \frac{n_j}{N} + \sqrt{\frac{2\log N}{N}}$ for all u with $\chi_{jk}(u) - \chi^\circ_{jk}(u) \neq 0$, we get

$$\begin{aligned} |E_1| &\le \frac{1}{\sqrt{Nn_j}} N\ \tilde{H}(\frac{n_j}{N} + \sqrt{\frac{2\log N}{N}})^\beta\, 4(\sqrt{\frac{2\log N}{N}} + \frac{1}{N}) \\ &= \tilde{H} \cdot O\left(\sqrt{\frac{\log N}{n_j}}(\frac{n_j}{N} + \sqrt{\frac{\log N}{N}})^\beta\right). \end{aligned}$$

The same inequality holds for E_2, and this ends the proof of (ii).

(iii) By definition of $\widehat{\alpha}^*_{jk}$, we have $\mathrm{Var}(\widehat{\alpha}^*_{jk}) = \mathrm{Var}\left(\frac{1}{N}\sum_{t=1}^N Z'_t \psi^\circ_{jk}(q_t)\right) = \frac{\sigma_z^2}{N}\frac{1}{N}\sum_{t=1}^N \psi^{\circ 2}_{jk}(q_t) = \frac{\sigma_z^2}{N}$, since there are only $N2^{-j}$ elements different from zero in this sum.

Lemma 2. *Let $U_1, \ldots, U_N$ be i.i.d. with common distribution function G, and let $U_{(1)} \le \ldots \le U_{(N)}$ denote the order statistics.*

(i) For all $\varepsilon > 0$ and for all v with $0 < G(v) < 1,\ \ s = 1, \ldots, N$:

$$P(|U_{(s)} - v| > \varepsilon) \le e^{-2N\delta^2}$$

with $\delta = \min(G(v+\varepsilon) - \frac{s}{N},\ \frac{s}{N} - G(v-\varepsilon))$.

(ii) If $U_1, \ldots, U_N$ are uniformly distributed on $[0,1]$, then, for $0 < v < 1,\ s = 1, \ldots, N$:

$$P\left(|U_{(s)} - \frac{s}{N}| > \sqrt{\frac{2\log N}{N}}\right) \le \frac{1}{N^4}$$

Proof. (i) This result is essentially stated in the proof of Lemma C of Serfling (1980, Chapter 2.5.4), and it is an immediate consequence of Hoeffding's Lemma (Serfling, 1980, Chapter 2.3.2).

(ii) We choose $v = \frac{s}{N}$ and $\varepsilon = \sqrt{\frac{2\log N}{N}}$. Then, the assertion follows from part (i) as now $G(u) = u$ and, therefore, $\delta = \varepsilon$.

Lemma 3. *Assume (A1) through (A3). Let j be fixed. Then for N large enough the Hölder condition*

$$|m(F_x^{-1}(u)) - m(F_x^{-1}(v))| \le \tilde{H}|u - v|^\beta$$

is satisfied for all $u, v \in \mathcal{I} = \left[\frac{(k-1)n_j - 1}{N} - \sqrt{\frac{2\log N}{N}}, \frac{kn_j + 1}{N} + \sqrt{\frac{2\log N}{N}}\right]$.

PROOF. By the Hölder regularity of $m(.)$, we have for all $u < v$

$$\begin{aligned}
&|m(F_x^{-1}(u)) - m(F_x^{-1}(v))| \\
&\quad\leq H|F_x^{-1}(u) - F_x^{-1}(v)|^\beta \\
&\quad= H\frac{1}{f(F_x^{-1}(\omega))^\beta}|u-v|^\beta \quad \text{for some } \omega \in (u,v) \\
&\quad\leq \frac{H}{C^\beta}|u-v|^\beta = \tilde{H}|u-v|^\beta,
\end{aligned}$$

as long as $[F_x^{-1}(u), F_x^{-1}(v)] \subseteq S$. If on the other hand $[F_x^{-1}(u), F_x^{-1}(v)] \subseteq S^c$, then $|m(F_x^{-1}(u)) - m(F_x^{-1}(v))| = 0$. Finally, $F_x^{-1}(u) \notin S, F_x^{-1}(v) \in S$ or vice-versa may happen only for two values of k. Hence we have shown that

$$|m(F_x^{-1}(u)) - m(F_x^{-1}(v))| \leq \tilde{H}|u-v|^\beta$$

for all $u, v \in \mathcal{I} = \left[\frac{(k-1)n_j-1}{N} - \sqrt{\frac{2\log N}{N}}, \frac{kn_j+1}{N} + \sqrt{\frac{2\log N}{N}}\right]$ for all but at most two values of k.

7.2.4. *Proof of Proposition 4.* Similarly to the proof of Proposition 3, we decompose $\widehat{\beta}_{jk}$ as follows:

$$\widehat{\beta}_{jk} = \frac{1}{Nn_j}\sum_{t=1}^{N}(1_{j+1,2k}(X_t) - 1_{j+1,2k+1}(X_t)) = D_1 - D_2$$

and use $\mathrm{Var}(\widehat{\beta}_{jk}) \leq 2(\mathrm{Var}(D_1)+\mathrm{Var}(D_2))$. Changing notation from $(j+1, 2k)$ to (j,k), we have by definition of $I_{j,k}$:

$$\begin{aligned}
\mathrm{Var}(D_1) &= \frac{1}{Nn_j}\mathrm{Var}\left(\sum_{j=1}^{N} m(X_t)1_{jk}(X_t)\right) \\
&= \frac{1}{Nn_j}\sum_{s,t=(k-1)n_j+1}^{kn_j}\mathrm{Var}\left(m(X_{(t)})\ m(X_{(s)})\ \right) \\
&\leq \frac{1}{Nn_j}\sum_{s,t=(k-1)n_j+1}^{kn_j}\sqrt{\mathrm{Var}(m(X_{(t)}))\ \ \mathrm{Var}(m(X_{(s)}))},
\end{aligned}$$

by the Cauchy-Schwarz inequality. Let $g(u), U_t$ be as in the proof of Proposition 3. We have, for $(k-1)n_j + 1 \leq t \leq kn_j$,

$$\mathrm{Var}(m(X_{(t)})) = \mathrm{Var}(g(U_{(t)})) \leq E[(g(U_{(t)}) - g(\frac{t}{N}))^2] \leq \tilde{H}^2E[(U_{(t)} - \frac{t}{N})^{2\beta}],$$

by Lemma 3. Now by Lemma 2 above, we have $\tilde{H}^2 E[(U_{(t)} - \frac{t}{N})^2] \le \tilde{H}^2 \left(\frac{2\log N}{N}\right)^\beta$, for N large enough such that $|U_{(s)} - \frac{s}{N}| \le \sqrt{\frac{2\log N}{N}}$ for all s with probability 1. Therefore,

$$\mathrm{Var}(D_1) \le \frac{1}{Nn_j} n_j^2 \tilde{H}^2 \left(\frac{2\log N}{N}\right)^\beta = O\left(\frac{n_j}{N}\left(\frac{\log N}{N}\right)^\beta\right).$$

7.2.5. *Proof of Proposition 5.* (i) By Proposition 2, we have $E[(\widehat{\rho}_{jk} - \widehat{\rho}^*_{jk})^2] = O(\frac{2^j}{N^{3/2}})$. Summing over (j,k), we obtain $\sum_{j,k} E[(\widehat{\rho}_{jk} - \widehat{\rho}^*_{jk})^2] = O\left(\frac{2^{2J}}{N^{3/2}}\right)$ and we want this to be of order $o(N^{-\frac{2\beta}{2\beta+1}})$. By definition of the set $\mathcal{J}_N$, $2^J = N^{1-\delta}$. Thus the condition can be written as $N^{2-2\delta}N^{-3/2} = o(N^{-\frac{2\beta}{2\beta+1}})$, and is fulfilled when $\delta > \frac{2\beta}{4\beta+2} + \frac{1}{4}$, as it is required in the assumptions.

(ii) If we show that

$$\sum_{(j,k)\in\mathcal{J}_N} (E[\widehat{\alpha}^*_{jk} - \widehat{\alpha}^\circ_{jk}])^2 = o\left(N^{-\frac{2\beta}{2\beta+1}}\right) \quad (41)$$

$$\sum_{(j,k)\in\mathcal{J}_N} (E[\widehat{\alpha}_{jk} - \widehat{\alpha}^\circ_{jk}])^2 = o\left(N^{-\frac{2\beta}{2\beta+1}}\right), \quad (42)$$

the result is proved. We first prove (41). By Proposition 3(i), $E[\widehat{\alpha}^*_{jk} - \widehat{\alpha}^\circ_{jk}] \le C_1\, 2^{j/2}/N$, as by assumption (A3), $V(m; q_{(k-1)n_j}, q_{kn_j})$ is uniformly bounded over k and j. With $2^J = N^{1-\delta}$ and $\delta > \frac{\beta}{2\beta+1}$ by assumption,

$$\sum_{(j,k)\in\mathcal{J}_N} (E[\widehat{\alpha}^*_{jk} - \widehat{\alpha}^\circ_{jk}])^2 = O\left(\frac{2^{2J}}{N^2}\right) = o\left(N^{-\frac{2\beta}{2\beta+1}}\right).$$

We now show (42). By the proof of Proposition 3(ii), we have

$$E[\widehat{\alpha}_{jk} - \widehat{\alpha}^\circ_{jk}] \le \tilde{H}\ C_2 \sqrt{\frac{\log N}{N2^{-j}}}\left(\frac{1}{2^j} + \sqrt{\frac{\log N}{N}}\right)^\beta . \quad (43)$$

In our assumptions, we have $\beta > 1/2$. Hence taking $\beta = 1/2$ in (43) will overestimate the bound of the error of approximation; in this case the square of (43) is of order $\frac{\log N}{N} + 2^j\left(\frac{\log N}{N}\right)^{3/2}$. The term $\frac{\log N}{N}$ is not of sufficiently small order, whereas the second term $2^j\left(\frac{\log N}{N}\right)^{3/2}$ is of sufficiently small order in order to have the result (42), provided that $\delta > \frac{2\beta}{4\beta+2} + \frac{1}{4} + \epsilon$, for $\epsilon > 0$ arbitrarily small. We thus concentrate only on the term $\sqrt{(\log N)/(N2^{-j})}(1/2^j)^\beta$ in equation (43), since it is the term of lower order. For $\beta > 1/2$, we have:

$$(E[\widehat{\alpha}_{jk} - \widehat{\alpha}^\circ_{jk}])^2 = O\left(\frac{\log N}{N2^{-j}} 2^{-2j\beta}\right) = O\left(\frac{\log N}{N} 2^{-2(\beta-\frac{1}{2})j}\right).$$

Summing over $k = 1, \ldots, 2^j$ and then over $j = 1, \ldots, 2^J$, we have:

$$\textstyle\sum_{(j,k)\in\mathcal{J}_N} (E[\widehat{\alpha}_{jk} - \widehat{\alpha}^\circ_{jk}])^2 = O\left(\frac{\log N}{N} \, 2^{2J(1-\beta)}\right) . \tag{44}$$

The term $\log N/N$ can be written as $N^{-1+\epsilon}$ for $\epsilon > 0$ arbitrarily small. By definition of the set $\mathcal{J}_N$, $2^J = N^{1-\delta}$ and hence the sum (44) will be of order $o(N^{-\frac{2\beta}{2\beta+1}})$ if $\epsilon - 1 + (2-2\delta)(1-\beta) < -\frac{2\beta}{2\beta+1}$, i.e. if $\delta > (1 - \frac{4\beta^2}{2\beta+1} + \epsilon)\frac{1}{2(1-\beta)}$ and this condition is automatically fulfilled when $\beta > 1/2$ and $\delta > \frac{2\beta}{4\beta+2} + \frac{1}{4}$.
(iii) By Proposition 4, we have:

$$\operatorname{Var}(\widehat{\beta}_{jk}) = O\left(\frac{1}{2^j}\left(\frac{\log N}{N}\right)^\beta\right).$$

Summing over k and j in $\mathcal{J}_N$ we obtain

$$\textstyle\sum_{(j,k)\in\mathcal{J}_N} \operatorname{Var}(\widehat{\beta}_{jk}) = J\left(\frac{\log N}{N}\right)^\beta = \frac{\log(N^{1-\delta})(\log N)^\beta}{N^\beta}$$

and this is of order $o(N^{-\frac{2\beta}{2\beta+1}})$ for all $\delta \in (0,1)$ as long as $\beta > 1/2$.

7.2.6. *Proof of Proposition 6.* Note that $E[(\widehat{\beta}_{jk} - \alpha^\circ_{jk})^2] = (E[\widehat{\beta}_{jk} - \alpha^\circ_{jk}])^2 + \operatorname{Var}(\widehat{\beta}_{jk}) = (E[\widehat{\alpha}_{jk} - \alpha^\circ_{jk}])^2 + \operatorname{Var}(\widehat{\beta}_{jk})$. Proposition 3(ii) gives us the order or $E[\widehat{\alpha}_{jk} - \alpha^\circ_{jk}]$ and, using part of the demonstration in Proposition 5(ii), we obtain that

$$\textstyle\sum_{(j,k)\in\mathcal{J}_N} (E[\widehat{\alpha}_{jk} - \alpha^\circ_{jk}])^2 = o\left(N^{-\frac{2\beta}{2\beta+1}}\right).$$

Finally, by Proposition 5(iii) we finally obtain that

$$\textstyle\sum_{(j,k)\in\mathcal{J}_N} \operatorname{Var}(\widehat{\beta}_{jk}) = o\left(N^{-\frac{2\beta}{2\beta+1}}\right).$$

Acknowledgements. The authors would like to thank C. Almeida, A. Antoniadis, T. Cai, A. Kneip, M. Neumann and W. Sweldens for helpful discussions, and the anonymous referee for his comments that helped to improve the presentation of the paper. This research was mainly done during the visit of J. Franke at the Institut de statistique, Université catholique de Louvain (UCL), Louvain-la-Neuve, which was financed by the visitors' program of Pure and Applied Mathematics (MAPA) of the UCL. V. Delouille was funded as a Research Assistant of the National Fund for Scientific Research, Belgium (FNRS). Partial support from the contract 'Projet d'Actions de Recherche Concertées' ARC 98/03-217 of the Belgian Government is as well gratefully acknowledged.

References

ANTONIADIS, A. and FAN, J. (2000). Regularization of wavelets approximations. *Technical report*, IMAG, Grenoble, France. To appear in *Journal of the American Statistical Association*, with discussion, September 2001.

ANTONIADIS, A., GRÉGOIRE, G. and VIAL, P. (1997). Random design wavelet curve smoothing, *Statistics and Probability Letters*, **35**, 225–232.

ANTONIADIS, A. and PHAM, D.T. (1998). Wavelet regression for random or irregular design, *Computational Statistics and Data Analysis*, **28**, 353–369.

BERKNER, K. and WELLS, R. (1998). Smoothness estimates for soft-threshold denoising via translation invariant wavelet transforms. *Technical Report*, Department of Mathematics, Rice University. To appear in *Applied and Computational Harmonic Analysis*.

CAI, T. (1996). Nonparametric function estimation via wavelets. *Ph.D. Thesis*, Cornell University.

CAI, T. and BROWN, L. D. (1998). Wavelet Shrinkage for nonequispaced samples, *Annals of Statistics*, **26**, 1783–1799.

———— (1999). Wavelet regression for random uniform design, *Statistics and Probability Letters*, **42**, 313–321.

COHEN, A., DAUBECHIES, I. and FEAUVEAU, J. (1992). Bi-orthogonal bases of compactly supported wavelets, *Communications on Pure and Applied Mathematics*, **45**, 485–560.

COIFMAN, R. R. and DONOHO D. L. (1995). Translation-invariant de-noising. In *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, eds., Lecture Notes in Statistics, **103**, Springer-Verlag, New York, 125–150.

DAHLHAUS, R. and NEUMANN, M.H. (2000). Locally adaptive fitting of semiparametric models to nonstationary time series. Beiträge zur Statistik 60, Universität Heidelberg. To appear in *Stochastic Processes and Their Applications.*

DAUBECHIES, I. (1988). Orthonormal bases of compactly supported wavelets, *Communications on Pure and Applied Mathematics*, **41**, 909–996.

DELOUILLE, V., FRANKE, J. and VON SACHS, R. (2000). Nonparametric stochastic regression with design-adapted wavelets. *Technical Report*, Institut de Statistique, UCL, Belgium; http://www.stat.ucl.ac.be/ISpersonnel/delouille/recherche.html.

DELOUILLE, V., SIMOENS, J. and VON SACHS, R. (2001). Smooth Design-adapted Wavelets for Nonparametric Stochastic Regression. *Technical Report* 0117, Institut de Statistique, UCL, Belgium.

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, **81**, 425–455.

DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Assocication*, **90**, 1200–1224.

– – –– (1998). Minimax estimation via wavelet shrinkage, *Annals of Statistics*, **26**, 879–921.

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet shrinkage: asymptopia? (with discussion), *Journal of the Royal Statistical Society, Series B*, **57**, 301–337.

GAO, H.-YE (1997). Wavelet shrinkage estimates for heteroscedastic regression models. *Technical report*. MathSoft, Inc. Seattle.

GIRARDI, M. and SWELDENS, W. (1997). A new class of unbalanced Haar wavelets that form an unconditional basis for L_p on general measure spaces, *Journal of Fourier Analysis and Applications*, **3**, 457–474.

HALL, P. and TURLACH, B. A. (1997). Interpolation methods for nonlinear wavelet regression with irregularly spaced design, *Annals of Statistics*, **25**, 1912–925.

HOFFMANN, M. (1999). On nonparametric estimation in nonlinear AR(1)-models, *Statistics and Probability Letters*, **44**, 29–45.

JOHNSTONE, I. M. and SILVERMAN, B. W. (1997). Wavelet threshold estimators for data with correlated noise, *Journal of the Royal Statistical Society, Series B*, **59**, 319–351.

KOHLER, M. (2000). Nonlinear orthogonal series estimates for random design regression. *Technical Report*, Department of Mathematics, Universität Stuttgart.

KOVAC, A. and SILVERMAN, B. W. (2000). Extending the scope of wavelet regression methods by coefficient-dependent thresholding, *Journal of the American Statistical Association*, **95**, 172–183.

NASON, G.P. (1999). Fast cross-validatory choice of wavelet smoothness, primary resolution and threshold in wavelet shrinkage using the Kovac-Silverman algorithm. *Technical Report*, University of Bristol, Bristol, UK.

NASON, G.P. and SILVERMAN, B. (1995). The stationary wavelet transform and some statistical applications. In *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, eds., Lecture Notes in Statistics, **103**, Springer-Verlag: New York, 281–299.

NEUMANN, M. H. (1996). Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series, *Journal of Time Series Analysis*, **17**, 601–633.

NEUMANN, M. H. and SPOKOINY, V. G. (1995). On the efficiency of wavelet estimators under arbitrary error distributions, *Mathematical Methods of Statistics*, **4**, 137–166.

NEUMANN, M. H. and VON SACHS, R. (1995). Wavelet thresholding: Beyond the Gaussian i.i.d. situation. In *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, eds., Lecture Notes in Statistics, **103**, 301–329. Springer-Verlag, New York.

PENSKY, M. and VIDAKOVIC, B. (2000). On non-equally spaced wavelet regression. To appear in *Annals of Institute of Statistical Mathematics.*

PÓLYA, G. and SZEGÖ, G. (1979). *Aufgaben und Lehrsätze aus der Analysis I.* Springer-Verlag, Berlin.

VANRAES, E., JANSEN, M. and BULTHEEL, A. (2001). Stabilized wavelet transforms for non-equispaced data smoothing. *Technical Report*, Department of Computer Science, K.U. Leuven, Belgium

VON SACHS, R. and MACGIBBON, B. (2000). Nonparametric curve estimation by wavelet thresholding with locally stationary errors, *Scandinavian Journal of Statistics*, **27**, 475–499.

Sakhanenko, A. I. (1991). On the accuracy of normal approximation in the invariance principle, *Siberian Advances in Mathematics*, **1**, 58–91.

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

Serfling, R. J. (1980). *Approximations Theorems of Mathematical Statistics.* Wiley, New York.

V. Delouille
Institut de Statistique
Université catholique de Louvain
Louvain-la-Neuve, Belgium
E-mail: delouill@stat.ucl.ac.be

J. Franke
Universität Kaiserslautern
D-67653 Kaiserslautern, Germany
E-mail: franke@mathematik.uni-kl.de

R. von Sachs
Institut de Statistique
Université catholique de Louvain
Louvain-la-Neuve, Belgium
E-mail: rvs@stat.ucl.ac.be