Nonparametric Stochastic Regression with Design-adapted Wavelets

Véronique Delouille, Jürgen Franke and Rainer von Sachs

July 24, 2000

FB Mathematik, Universität Kaiserslautern, D-67653 Kaiserslautern, Germany
e-mail: franke@mathematik.uni-kl.de

and

Institut de Statistique, Université catholique de Louvain, Louvain-la-Neuve, Belgium
e-mail: delouill@stat.ucl.ac.be and rvs@stat.ucl.ac.be

Abstract

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.

Key words: Gaussian approximation, minimax L_2-rate, stochastic design, Unbalanced Haar wavelets, wavelet thresholding.

Financial support from the contract "Projet d'Actions de Recherche Concertées" no. 98/03-217 of the Belgian Government is gratefully acknowledged.

1 Introduction

Wavelet estimation for nonparametric regression has been treated in an almost countless number of papers in various different set-ups. They have been inspired by the seminal work of Donoho and Johnstone ([DJ94], [DJ95], [DJ98] and [DJKP]) which studied minimax wavelet shrinkage methods for the reconstruction of an unknown function $f(x)$ observed in a white noise model. They considered data of the form

$$Y_i \; = \; f(x_i) \; + \; Z_i \quad (i = 1, \ldots, N) \; ,$$

where $x_i = i/N, Z_i$ are i.i.d. $\mathcal{N}(0, \sigma^2)$. Performance of the estimator was measured by its L_2-risk over a ball in a whole smoothness function class characterized by a number of 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, an assumption which has been weakened in different directions again in a high number of subsequent publications. As for the generalisation from the Gaussian distribution we like to refer to [Neu], [NS] and [NvS], as those introduced the techniques also being used in this article. As for the generalisation of the dependence structure of the errors in nonparametric regression, [vSMG] treats non-identically distributed, correlated and possibly non-stationary (non-Gaussian) errors $\{Z_i\}$ by imposing some assumption of local stationarity.

This article, however, is concerned with a completely new method of generalising from the fixed and equidistant design. Again, a long list (e.g., [AGV], [Sar], [KS], [PV], [AF]) of existing work on mostly non-equidistant but non-random design has to be mentioned. Common to almost all of them is the idea of using some preliminary step to get back to the equidistant design situation and then apply the traditional discrete orthogonal wavelet transform algorithms. Methods to achieve this include projecting the data ([CB98]), and in the case of *stochastic* regressors, interpolation ([HT]), or approximating the scaling function by a recursion equation ([AP]). These approaches suffer from being somewhat hybrid (and tailored to the situation at hand), and hence have to deal with an additional approximation error caused by this first pre-processing step. The method of [CB99] finally only treats the case of a random uniform design.

Our new approach overcomes these drawbacks in that it naturally adapts to the situation of the *stochastic* design at hand without needing any pre-processing step. Moreover it allows for a direct parallel treatment of deriving rates of convergence of the L_2-risk over certain function classes of *non-linear*, i.e. shrinkage, wavelet estimators. The key to these objectives is a new construction of a wavelet basis of a weighted space $L_2(d\widehat{F}_N)$ where $\widehat{F}_N$ will be the empirical measure associated to the stochastic design, replacing hence the Lebesgue measure which usually appears in the construction of wavelet bases of $L_2(\mathbb{R})$. In particular, in this paper we will concentrate on using generalized (so-called "Unbalanced") Haar wavelets ([GS]) which are the only ones which allow for an orthonormal basis of a more generally weighted space $L_2(\mu)$ where the measure μ is different from the Lebesgue measure on $\mathbb{R}$. In research in progress we intend to treat more general, i.e. smoother, wavelet bases, for which necessarily orthogonality needs to be given up in favour of, e.g., bi-orthogonality; see [GS]. The aim remains to have a fast, computationally simple and theoretically tractable algorithm to the same extent as to what we provide in this article for the Unbalanced Haar wavelets. Note that this approach differs from the work of [Koh] where the aim of constructing design-adapted bases was achieved by using an orthogonal but only piecewise smooth basis 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 them to be non-Gaussian and show that techniques parallel to [NvS] can be applied to derive the L_2-rate of convergence known from the classical equidistant and fixed design regression. In doing so, we restrict ourselves to present the treatment for a comparatively small smoothness class of regression functions, i.e. $\beta-$Hölder functions with $\beta < 1$. Again, our main intention is to show that with our conceptually very simple methodology we can give the link to the classical situation, without being too ambitious about the size of the function class under consideration. However, our simulations show that our algorithm works fine also for functions with jumps and for a more general error structure, including heteroscedasticity and correlation. This is encouraging in the sense that our algorithm could possibly be applied not only to stochastic regression, but also to nonparametric autoregression where a richer range of interesting practical applications arises. Note that the up-to-date only existing non-linear wavelet algorithm ([Ho]) does unfortunately not allow for an efficient implementation.

In our simulations we actually use a non-decimated (or "translation-invariant") version of our suggested new algorithm. It is well-known from the classical situation (see, e.g., [CD] and [NaSi]) that translation-invariant denoising improves upon the quality of estimation and can be considered as increasing the degree of smoothness of the wavelet reconstruction ([BW]). This is in particular interesting while using non-smooth Haar wavelets.

We finish this introduction by providing how our paper is organised. The next section explains the construction of design-adapted wavelets, in general and in particular for the Unbalanced Haar basis, used in this work. Then, in Section 3, we proceed to specify the assumptions on our stochastic regression model. We state our main result for the non-linear (soft or hard) wavelet shrinkage estimator based on these design-adapted wavelets and indicate the line of its overall proof. In Sections 4 and 5 we explain in detail how our asymptotical considerations for proving the main result work, by first, in Section 4, relating our stochastic design to an accompanying fixed design. This first part deals with the mean-square approximation of the linear wavelet coefficients of both models. Note that in this approximation step, an intermediate third model is used, and note the important fact, that none of these approximating model coefficients is really used in the actual algorithm in practice (which is inherently different from existing methods, such as [HT]). This also applies to the second part on solely treating the non-linear wavelet estimator, in Section 5. This part uses the results of the previous section and furnishes an approximation of the non-Gaussian (i.i.d.) fixed design model by an asymptotic Gaussian (i.i.d.) model, paralleling the mentioned techniques of [Neu] and [NvS]. Section 6 is on some numerical results for the given model and delivers additionally a whole range of interesting generalisations (cf. our discussion above). All proofs are deferred to an appendix section.

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 bi-orthogonal, semi-orthogonal [Dau, CDF]) basis in $L_2(\mathbb{R})$. As will appear in the Section 2.2, for our regression purposes, we will need to consider a more general measure than the Lebesgue measure, i.e. we will use a wavelet basis which is orthonormal in a weighted space $L_2(\mu)$. In [GS], Girardi and Sweldens presented such a type of wavelet basis, which they called Unbalanced Haar, and 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 $\sigma-$field of I and μ is a $\sigma-$finite measure on Σ. This measure μ can be very general; in particular it can be purely non-atomic (i.e. it does not contain point masses), purely atomic (it contains only point masses), or a mixture of both situations.

What we need to do first when building Haar-like wavelets adapted to the space (I, Σ, μ), is to choose a particular partitioning of the interval $I \in \mathbb{R}$ we are working on. This concept of *partitioning* is defined in [GS] and recalled below. It can be thought of as the replacement of the dyadic intervals on the real line.

Definition 2.1 *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) $I_{j,k_1} \cap I_{j,k_2} = \emptyset \; \forall k_1 \neq k_2 \text{ and } \forall j;$

(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$-\sigma-$field of Σ such that the $\mu-$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

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

and as wavelets the functions

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

Note that these functions are all normalized in $L_2(\mu)$. It has been proved in [GS] 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.

If we denote by $s_{j,k}$ the scaling coefficients at level $j, j \in \mathbb{N}$ and location $k, k = 1, \ldots, 2^j$ and by $\alpha_{j,k}$ the wavelet coefficients at the same level and location, then given $s_{j+1,\cdot}$, it is possible to obtain $s_{j,\cdot}$ and $\alpha_{j,\cdot}$ using the following forward transform:

$$\begin{aligned} s_{j,k} &= \left(\frac{\mu(I_{j+1,2k})}{\mu(I_{jk})}\right)^{1/2} s_{j+1,2k} + \left(\frac{\mu(I_{j+1,2k+1})}{\mu(I_{jk})}\right)^{1/2} s_{j+1,2k+1} \\ \alpha_{j,k} &= \left(\frac{\mu(I_{j+1,2k})}{\mu(I_{jk})}\right)^{1/2} s_{j+1,2k+1} - \left(\frac{\mu(I_{j+1,2k+1})}{\mu(I_{jk})}\right)^{1/2} s_{j+1,2k} \end{aligned} \tag{3}$$

Note that, in case of dyadic equispaced intervals $\{I_{jk}\}$ and Lebesgue measure μ, equations (3) represent the usual Haar transform.

The next subsection explains how to use this wavelet transform in the case of nonparametric regression with stochastic design.

2.2 Regression with design-adapted wavelets

We consider the following regression model with stochastic design

$$Y_t = m(X_t) + Z_t \ , \ \ t = 1, \ldots, N, \tag{4}$$

where the design variables $X_1, \ldots, X_N$ are i.i.d. and independent of the residuals $Z_1, \ldots, Z_N$ which themselves are i.i.d. with $\mathcal{E}Z_t = 0, \mathrm{Var}Z_t = \sigma_z^2 < \infty$. Let F_x, f_x denote the distribution function and density of the X_t.

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 $\sigma-$field $\widetilde{\Sigma}$ of Definition 2.1 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

$$I^\circ_{jk} = \left[F_x^{-1}\left(\frac{k-1}{2^j}\right), \ F_x^{-1}\left(\frac{k}{2^j}\right)\right), \tag{5}$$

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 $I^o_{j,k}$ is simply equal to 2^{-j}, that is the points are equispaced in $L_2(dF_x)$. For this partitioning, condition $(P4)$ in Definition 2.1 is fulfilled with $\widetilde{\Sigma} = \mathcal{B}(\mathbb{R})$.

With the quantile partitioning, equations (1) and (2) simplify to

$$\begin{aligned}
\varphi^{\circ}_{jk}(x) &= 2^{j/2}1^{\circ}_{jk}(x), \;\; j=1,\ldots,J, \;\; k=1,\ldots,2^j, && (6)\\
\psi^{\circ}_{jk}(x) &= \frac{1}{\sqrt{2}}\left(\varphi^{\circ}_{j+1,2k}(x)-\varphi^{\circ}_{j+1,2k+1}(x)\right), \\
j &= 1,\ldots,J-1, k=1,\ldots,2^j, && (7)
\end{aligned}$$

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

$$m(x)=\sum_k s^{\circ}_{j_0,k}\varphi^{\circ}_{j_0,k}(x)+\sum_{j\geq j_0,k\in\mathbf{Z}}\alpha^{\circ}_{j,k}\psi^{\circ}_{j,k}(x), \tag{8}$$

where the theoretical coefficients are defined as

$$\alpha^{\circ}_{j,k}=\langle m(x),\psi^{\circ}_{j,k}(x)\rangle_{dF_x}=\int m(x)\psi^{\circ}_{j,k}(x)dF_x(x). \tag{9}$$

and similarly $s^{\circ}_{j_0,k}=\langle m(x),\varphi^{\circ}_{j_0,k}\rangle_{dF_x}$, where j_0 denotes the primary resolution level. Suppose now we have data $\{(X_i,Y_i)\}_{i=1}^N$ generated from a model like (4). To find an estimator of $m(x)$ based on $\{\varphi^{\circ}_{j_0,k}\}\cup\{\psi^{\circ}_{jk}\}$, we first need to estimate the wavelet coefficients from the data. Let $f_{y|x}$ denote the conditional density of Y given X and let $f_{x,y}(x,y)$ be the joint density of the random vector (X,Y). Then we have :

$$\begin{aligned}
\alpha^{\circ}_{jk} &= \int m(x)\psi^{\circ}_{jk}(x)dF_x(x)=\int\mathcal{E}(Y|X=x)\psi^{\circ}_{jk}(x)dF_x(x)\\
&= \int\left\{\int yf_{y|x}(y|x)dy\right\}\psi^{\circ}_{jk}(x)f_x(x)dx\\
&= \mathcal{E}_{(X,Y)}(Y\psi^{\circ}_{jk}(X)),
\end{aligned}$$

and similarly for $s^{\circ}_{j_0k}$. An unbiased estimator for α°_{jk} and $s^{\circ}_{j_0,k}$ is thus the empirical mean:

$$\widehat{\alpha}^{\circ}_{jk}=\frac{1}{N}\sum_{t=1}^N Y_t\psi^{\circ}_{jk}(X_t), \quad \widehat{s}^{\circ}_{j_0,k}=\frac{1}{N}\sum_{t=1}^N Y_t\varphi_{j_0,k}(X_t). \tag{10}$$

Using these coefficients, we could in principle, i.e. if we knew F_x, use either a linear or a nonlinear wavelet estimator. However, we found out that in this case it would not be possible, with our methodology, to develop asymptotic theory for a non-linear threshold estimator. The reason is that, due to the randomness of the design, the variance of $\widehat{\alpha}^{\circ}_{jk}$ and $\widehat{s}^{\circ}_{j_0,k}$ is too high to justify an approximation which is needed to prove the optimality of the non-linear estimator in the setting of non-Gaussian errors in a non-equidistant design.

In practice we only have information about the empirical distribution $\widehat{F}_N$ of the given realisation $X_1,\ldots,X_N$. Note that even for $\widehat{F}_N$ the condition $(P4)$ in Definition 2.1 does hold. This leads us to defining 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 end-points of the intervals. Here the random interval $I_{j,k}$ are determined by the data as :

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

where $X_{(1)} \leq \ldots \leq X_{(N)}$ denote the order statistics of the design variables and, as a convention, $X_{(N+1)} := X_{(N)} + (X_{(N)} - X_{(1)})/N$. The wavelets $\psi_{j,k}$ on level j are then defined by

$$\psi_{jk}(x) = \frac{1}{\sqrt{2}}(\varphi_{j+1,2k}(x) - \varphi_{j+1,2k+1}(x)),\ \ j = j_0, \ldots, J-1, k = 1, \ldots, 2^j, \tag{11}$$

with

$$\varphi_{jk}(x) = 2^{j/2}1_{jk}(x),\ \ j = j_0, \ldots, J,\ \ k = 1, \ldots, 2^j, \tag{12}$$

and where $1_{jk}(x)$ denotes the indicator of the random interval I_{jk}. Similarly to the relations (10), we define the corresponding empirical wavelet coefficients as:

$$\widehat{\alpha}_{j,k} = \frac{1}{N}\sum_{t=1}^{N} Y_t\psi_{j,k}(X_t), \qquad \widehat{s}_{j,k} = \frac{1}{N}\sum_{t=1}^{N} Y_t\varphi_{j,k}(X_t). \tag{13}$$

In contrast to the basis $\{\varphi^{\circ}_{j_0,k}\} \cup \{\psi^{\circ}_{jk}\}_{j=j_0}^{\infty}$, the basis $\{\varphi_{j_0,k}\} \cup \{\psi_{jk}\}_{j=j_0}^{J-1}$ defined by (12) and (11) cannot generate an infinite dimensional series representation of $m(x)$ as in (8). This is simply due to the construction of this basis by the random intervals $\{I_{jk}\}$ based on the finite sample $X_1, \ldots, X_N$. Consequently, the reference object for estimation has to be the projection $m_J(x)$ of $m(x)$ onto the space generated by $\{\varphi_{Jk}\}$, with J equal at most to $\log_2(N)$. To find an expression for this projection, consider the decomposition of $\widehat{\alpha}_{jk}$ as

$$\begin{aligned}\widehat{\alpha}_{jk} &= \widehat{\beta}_{jk} + \widehat{\rho}_{jk} \\ &= \frac{1}{N}\sum_{t=1}^{N} m(X_t)\,\psi_{jk}(X_t) + \frac{1}{N}\sum_{t=1}^{N} Z_t\,\psi_{jk}(X_t),\end{aligned} \tag{14}$$

and similarly $\widehat{s}_{j_0,k} = \widehat{b}_{j_0,k} + \widehat{r}_{j_0,k}$. Then the projection $m_J(x)$ is simply

$$m_J(x) = \sum_k \widehat{b}_{j_0,k}\varphi_{j_0,k}(x) + \sum_{jk} \widehat{\beta}_{jk}\psi_{jk}(x). \tag{15}$$

In the next section, we will study in detail the properties of a nonlinear wavelet thresholding estimator based on the wavelet coefficients $\widehat{\alpha}_{jk}$.

3 Estimator and main result

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 integrated mean squared error

(IMSE), i.e. the L_2-risk, between estimate $\widehat{m} = \widehat{m}_N$ and the curve m of interest. Typically, in the literature of non-linear wavelet threshold estimation this is done in a uniform sense over a ball $\mathcal{F} = \mathcal{F}(L)$ in a function class, e.g. a Hölder, Sobolev or Besov class, where L is some positive finite constant. Thus we aim at finding an estimator $\widehat{m}$ which minimises the quantity

$$\sup_{m\in\mathcal{F}} \left\{\mathcal{E}\|\widehat{m} - m\|^2\right\} . \tag{16}$$

In this paper, mainly for reasons of keeping the degree of technicality low we restrict ourselves to treat a Hölder class $\mathcal{F}(L;\beta)$, i.e.

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

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 $x \in \mathring{S}$, $f_x(x) = 0$ for all $x \in \mathbb{R} \setminus \mathring{S}$.

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

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

Note that the function m may well be defined on all $\mathbb{R}$ but due to assumption (A1) we estimate it only on the interval S.

We also need to add a mainly technical assumption on the error distribution which allows to treat the non-Gaussianity of the errors without needing to apply lengthy truncation techniques (see the proof of Lemma 5.1 in Section 5).

(A4) We assume that all moments of the errors $\{Z_i\}$ are uniformly bounded, i.e.

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

The estimator $\widehat{m}$ based on the Unbalanced Haar basis $\{(\psi_{j,k})_{j\geq j_0} \cup (\varphi_{j_0,k})\}$ and the given regressors $X_1, \ldots, X_N$ will have the following form:

$$\widehat{m}(x) = \sum_k \widehat{s}_{j_0k}\ \varphi_{j_0k}(x) \ + \ \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} \delta^{(\cdot)}(\widehat{\alpha}_{jk}, \lambda_{jk})\ \psi_{jk}(x) \ , \tag{17}$$

where $\delta^{(\cdot)}$ denotes soft or hard thresholding based on a threshold λ_{jk}. The range of coefficients is actually determined through $(j,k) \in \mathcal{J}_N$, where the set $\mathcal{J}_N$ is defined as

$$\mathcal{J}_N = \{(j,k)|2^j \leq C\, N^{1-\delta}\} \text{ with } 0 < \delta < 1. \tag{18}$$

For a more precise specification of the range of δ in this definition we refer to equation (23) below.

In order to prove our main result, that is Theorem 3.1 below, we will consider as an intermediate tool the wavelet coefficients, noted $\widehat{\alpha}^*_{jk}$ and defined in equation (30) below, which are estimated from a fixed design composed of the quantiles $q_t = F_x^{-1}(\frac{t}{N})$ and the fixed intervals $I^\circ_{j,k}$ as defined in (5).

Now, to formulate our main result we need to be more precise in view of the actual choice of the individual thresholds λ_{jk} in our estimator of type (17). For this we follow [Neu] and [NvS] and distinguish between two different situations. We call

$$\widetilde{m}(x) = \sum_k \widehat{s}_{j_0k}\ \varphi_{j_0k}(x)\ +\ \sum_{(j,k)\in\mathcal{J}_N} \delta^{(\cdot)}(\widehat{\alpha}_{jk}, \widetilde{\lambda}_{jk})\ \psi_{jk}(x) \tag{19}$$

the estimator that is based on an optimal (non–random) threshold $\widetilde{\lambda}_{jk} = \lambda(N,j,k;\mathcal{F})$, whereas

$$\widehat{m}(x) = \sum_k \widehat{s}_{j_0k}\ \varphi_{j_0k}(x)\ +\ \sum_{(j,k)\in\mathcal{J}_N} \delta^{(\cdot)}(\widehat{\alpha}_{jk}, \lambda_{jk})\ \psi_{jk}(x) \tag{20}$$

is an estimator with (random, as data-driven) individual thresholds λ_{jk}. As usual, their choice is coupled to the (square root of the) variance of the empirical wavelet coefficients, here $\sigma^2_{jk} = \mathrm{Var}(\widehat{\alpha}^*_{jk})$, which has to be of order N^{-1}. Proposition 4.3 below shows that $\sigma^2_{jk} = \sigma^2_z/N$, and even though this is the variance in the approximating model (29), its choice is appropriate even for thresholding in our original model (4). Indeed in this original model the *conditional variance* of the $\widehat{\alpha}_{jk}$ is equal to $\sigma^2_{jk} = \sigma^2_z/N$, and it is precisely this conditional variance which measures the noise variation about the regression function m given the $X_1,\ldots,X_N$, and hence the one which should be taken into account when computing the threshold value.

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 inequality on an event with probability tending to one with a rate as given in [Neu] (condition (A6) therein): $\sigma_{jk}\ \sqrt{2\log(\#\mathcal{J}_N)} \le \lambda_{jk} \le C\ N^{-1/2}\sqrt{\log(N)}$ for a positive constant C. This choice will be fully motivated in Section 5.

Theorem 3.1 *Let $\mathcal{F} = \Lambda^\beta(L)$, with $1/2 < \beta < 1$. Let $\widetilde{m}$ and $\widehat{m}$ be the wavelet threshold estimators given in (19) and (20), respectively. Then,*

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

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

(ii) for thresholds λ_{jk} satisfying $\sigma_{jk}\ \sqrt{2\log(\#\mathcal{J}_N)} \le \lambda_{jk} \le C\ N^{-1/2}\sqrt{\log(N)}$ for any positive constant C,

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

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 (not known in practice, however), whereas for a data–driven threshold we get the "near-optimal" rate which is optimal apart from the log-term. 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 3.1 but before we give in Section 3.2 the overall line of its proof.

3.2 Methodology for proving the main result

Consider the regression function $m(x)$ as represented by the series in (8). Our paradigm here is to provide a theoretical treatment of the non-linear wavelet estimator which parallels the one in the setting of non-Gaussian errors. We would like to transfer to the stochastic regression setting the results established for the fixed design case. To parallel the situation of fixed design, we consider the quadratic risk defined with respect to the measure $L_2(dF_x)$:

$$\begin{aligned} \mathcal{E} & \ \|\widehat{m}(x) - m(x)\|^2_{L_2(dF_x)} \\ \leq & \ 2\,\mathcal{E}\|\widehat{m}(x) - \sum_k s^\circ_{Jk}\varphi^\circ_{Jk}(x)\|^2_{L_2(dF_x)} + 2\sum_{j\geq J,k}(\alpha^\circ_{jk})^2. \end{aligned} \tag{21}$$

To prove Theorem 3.1, we first need to ensure that the second term in the r.h.s. of (21), called the deterministic bias, is of order $o(N^{-\frac{2\beta}{2\beta+1}})$. For this, we need a further specification of the set $\mathcal{J}_N$, i.e. of its parameter $0 < \delta < 1$, given in equation (18). Lemma 3.2 below tells us that $|\alpha^\circ_{jk}|$ decays as $2^{-j(\beta+\frac{1}{2})}$. Hence $\sum_{j\geq J}\sum_k |\alpha^\circ_{jk}|^2 \leq C2^{-2J\beta}$, for $m \in \mathcal{F} = \Lambda^\beta(L)$ and for some generic constant C. So, if, in the definition (18) of the set $\mathcal{J}_N$, we take $2^J = N^{1-\delta}$, the deterministic bias will be of the required order as soon as

$$0 < \delta < \frac{2\beta}{2\beta+1}. \tag{22}$$

Note that this condition automatically entails that J will be greater than the optimal cutting level of a linear wavelet estimator. The set $\mathcal{J}_N$ is thus now defined as

$$\mathcal{J}_N = \{(j,k)|2^j \leq C\,N^{1-\delta}\} \text{ with } 0 < \delta < \frac{2\beta}{2\beta+1}\ . \tag{23}$$

Lemma 3.2 *Let $m \in \Lambda^\beta(L)$ with $0 < \beta < 1$, and define wavelets $\{\psi^\circ_{jk}\}$ on $L_2(dF_x)$ as in equation (7). Assume (A1)-(A3), then*

$$|\langle m, \psi^\circ_{jk}\rangle_{dF_x}| \leq C2^{-j(\beta+\frac{1}{2})},$$

where the constant C is independent of j and k.

Consider now the first term of the r.h.s. in (21). In practice, we only have information at the points $\{X_i\}_{i=1}^N$ and the measure dF_x is not available. Hence, following our discussion at the end of Section 2.2, it is only possible to estimate the projection $m_J(x)$, defined by

(15), and we will rather treat the quadratic risk defined now with respect to the empirical measure $d\widehat{F}_N$

$$\mathcal{E}\|\widehat{m} - m_J\|^2_{L_2(d\widehat{F}_N)}, \tag{24}$$

with $\widehat{m}$ defined by expression (20).

At this stage, we could in principle use conditional expectations (given $X_1, \ldots, X_N$), and treat the problem similarly to what, e.g., Sardy *et al* do in [Sar]. However, one of our aims is to show near-optimality of our non-linear estimator in the case of non-Gaussian errors. To do so, we first have to approximate $\widehat{\alpha}_{jk}$ by $\widehat{\alpha}^*_{jk}$ and come back to the situation of fixed design composed of theoretical quantiles; we again refer to model (29) below. This explains why, using expression (15) and (20) for $m_J(x)$ and $\widehat{m}(x)$, respectively, we decompose the risk in $L_2(d\widehat{F}_N)$ as follows.

$$\begin{aligned}
\mathcal{E}||\widehat{m} - m_J||^2_{L_2(d\widehat{F}_N)} &= \mathcal{E}||\sum_k \widehat{s}_{j_0k}\ \varphi_{j_0k}(x) + \sum_{(j,k)\in\mathcal{J}_N} \delta^{(\cdot)}(\widehat{\alpha}_{jk}, \lambda_{jk})\psi_{jk}(x) \\
- &\sum_k s^\circ_{j_0,k}\varphi_{j_0,k}(x) + \sum_k s^\circ_{j_0,k}\varphi_{j_0,k}(x) - \sum_{jk} \alpha^\circ_{jk}\psi_{jk}(x) \\
+ &\sum_{jk} \alpha^\circ_{jk}\psi_{jk}(x) - \sum_k \widehat{b}_{j_0,k}\varphi_{j_0,k}(x) - \sum_{j,k} \widehat{\beta}_{jk}\psi_{jk}(x)||^2_{L_2(d\widehat{F}_N)} \\
\leq\ &2(\sum_k \mathcal{E}[(\widehat{s}_{j_0,k} - s^\circ_{j_0,k})^2] + \sum_{jk} \mathcal{E}[(\delta^{(\cdot)}(\widehat{\alpha}_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})^2] && (25) \\
+ &\sum_k \mathcal{E}[(s^o_{j_0,k} - \widehat{b}_{j_0,k})^2] + \sum_{jk} \mathcal{E}[(\alpha^\circ_{jk} - \widehat{\beta}_{jk})^2]). && (26)
\end{aligned}$$

Considering the second term of (25), we have the following individual bounds which hold due to the monotonicity of the threshold rules $\delta^{(\cdot)}$ in its first argument:

$$\mathcal{E}[(\delta^{(\cdot)}(\widehat{\alpha}_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})^2] \ \leq\ 2\ \ \mathcal{E}[(\widehat{\alpha}_{jk} - \widehat{\alpha}^*_{jk})^2]\ +\ 2\ \mathcal{E}[(\delta^{(\cdot)}(\widehat{\alpha}^*_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})^2]. \tag{27}$$

Using inequality (27), the quadratic risk can be bounded from above as follows.

$$\begin{aligned}
\mathcal{E}\|\widehat{m} - m_J\|^2_{L_2(d\widehat{F}_N)} &\leq C\{\sum_{jk} \mathcal{E}[(\widehat{\alpha}_{jk} - \widehat{\alpha}^*_{jk})^2] + \sum_{jk} \mathcal{E}[(\delta^{(\cdot)}(\widehat{\alpha}^*_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})^2] \\
+ &\sum_k \mathcal{E}[(\widehat{s}_{j_0,k} - s^\circ_{j_0,k})^2] + \sum_k \mathcal{E}[(s^\circ_{j_0,k} - \widehat{b}_{j_0,k})^2] + \sum_{jk} \mathcal{E}[(\alpha^\circ_{jk} - \widehat{\beta}_{jk})^2]\} \\
= &\ T_1 + T_2 + T_3 + T_4 + T_5. && (28)
\end{aligned}$$

Using this last decomposition, our strategy to prove the main theorem proceeds in two major steps.

In the first step, we develop tools to prove that the term T_1 is of the required order, i.e.,

$$T_1 = \textstyle\sum_{(j,k)\in\mathcal{J}_N} \mathcal{E}[(\widehat{\alpha}_{jk} - \widehat{\alpha}^*_{jk})^2] = o(N^{-\frac{2\beta}{2\beta+1}}),$$

for a set $\mathcal{J}_N$ that satisfies (18) and that will be further specified in Section 4 (in order to obtain the desired result). With the tools developed to bound T_1, it will then be easy to prove that each of the terms T_3, T_4 and T_5 are also of order $o(N^{-\frac{2\beta}{2\beta+1}})$, for the same set $\mathcal{J}_N$ as before.

The second important step is to prove that the term T_2 is such that

$$T_2 = \sum_{jk} \mathcal{E}[(\delta^{(\cdot)}(\widehat{\alpha}^*_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})^2] = O(\frac{\log N}{N})^{\frac{2\beta}{2\beta+1}}.$$

Here we are back to the situation with fixed non-equispaced design and deterministic quantile intervals. This is similar enough to standard wavelet regression such that we can use the corresponding appropriately adapted arguments for justifying our estimates. Using techniques developed in [Neu] and [NvS], we attain the 'classical' rates of convergence of the L_2-risk estimators as T_2 will be asymptotically equivalent to the same risk in an accompanying Gaussian noise model .

This will be shown by a strong form of asymptotic normality of the empirical wavelet coefficients $\widehat{\alpha}^*_{jk}$, which puts emphasis on moderate and large deviations; see equation (36) below. Once this equivalence has been established, we can apply the (theoretical) thresholding methods originally developed for the case of Gaussian i.i.d. noise, leading to our main Theorem 3.1. This is explained in full length in Section 5. The proofs are deferred to Section 7.

4 Asymptotic approximations: first part

4.1 Reducing to the simple set-up of fixed design

We consider our original regression model (4) given in Section 2.2. As announced before, for theoretical purposes, we also consider the following twin model

$$Y^*_t = m(q_t) + Z'_t\,,\ t = 1, \ldots, N, \tag{29}$$

where $Z'_1, \ldots, Z'_N$ are appropriately chosen i.i.d. random variables with the same distribution as the residuals Z_t of (4), 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 (22). The wavelets $\psi_{j,k}$ and scaling functions $\varphi_{j,k}$ are defined in (11) and (12), respectively, where, as before,

$$I_{jk} = [X_{((k-1)n_j+1)},\ X_{(kn_j+1)})\ ,\ \ k = 1, \ldots, 2^j\,,\ n_j = N2^{-j}.$$

The corresponding empirical wavelet and scaling coefficients of model (4) are given by $\widehat{\alpha}_{jk}$ and $\widehat{s}_{j_0,k}$ as defined in (13). To simplify notation, we set $\alpha_{j_0-1,k} := s_{j_0,k}$ and $\psi_{j_0-1,k} := \varphi_{j_0,k}$. The analogs $\widehat{\alpha}^*_{jk}$ of (13) on the basis of the deterministic design model (29) are defined by

$$\widehat{\alpha}^*_{jk} = \frac{1}{N}\sum_{t=1}^N Y^*_t\, \psi^\circ_{jk}(x_t),\ \ j = j_0 - 1, \ldots, J-1, k = 1, \ldots, 2^j. \tag{30}$$

In this section our main goal is to show that

$$T_1 = \sum\nolimits_{(j,k)\in\mathcal{J}_N} \mathcal{E}[(\widehat{\alpha}^*_{jk} - \widehat{\alpha}_{jk})^2] = o(N^{-\frac{2\beta}{2\beta+1}}). \tag{31}$$

By definition, the wavelet coefficients $\widehat{\alpha}_{jk}$ are based on random design X_t and random intervals I_{jk}, whereas the coefficients $\widehat{\alpha}^*_{jk}$ are defined from a deterministic design q_t and corresponding fixed intervals I°_{jk}. To study the effect of noise on the wavelet coefficients in both designs we have to select an appropriate realisation of the model (29) with noise variables coupled to model (4). 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 (4). Then, we choose

$$Z'_t = Z_s \quad \text{iff} \quad X_{(t)} = X_s\,,\ \ t,\, s = 1,\ldots,N.$$

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 model (29). To show (31), consider the decomposition of $\widehat{\alpha}_{jk}$ into $\widehat{\rho}_{jk}$ and $\widehat{\beta}_{jk}$ as given in (14). Similarly, we decompose $\widehat{\alpha}^*_{jk}$ into two parts

$$\begin{aligned}
\widehat{\alpha}^*_{jk} &= \widehat{\beta}^*_{jk} + \widehat{\rho}^*_{jk} \\
&= \frac{1}{N}\sum_{t=1}^{N} m(q_t)\,\psi^\circ_{jk}(q_t) + \frac{1}{N}\sum_{t=1}^{N} Z'_t\,\psi^\circ_{jk}(q_t)\ ,
\end{aligned} \tag{32}$$

where the $\widehat{\rho}^*_{jk}$ depends on the noise and the remainder term does not. With this decomposition, it is easier to show (31) using the following relationship:

$$\begin{aligned}
\mathcal{E}[(\widehat{\alpha}^*_{jk} - \widehat{\alpha}_{jk})^2] &\le 2\{\mathcal{E}[(\widehat{\rho}_{jk} - \widehat{\rho}^*_{jk})^2] + \mathcal{E}[(\widehat{\beta}_{jk} - \widehat{\beta}^*_{jk})^2]\} \\
&= 2\{\mathcal{E}[(\widehat{\rho}_{jk} - \widehat{\rho}^*_{jk})^2] + \mathrm{Var}(\widehat{\beta}_{jk}) + (\mathcal{E}[\widehat{\alpha}_{jk} - \widehat{\alpha}^*_{jk}])^2\} \\
&= 2\,(A + B + C),
\end{aligned} \tag{33}$$

and to consider each term separately.

4.2 Bias and Variance of an individual coefficient

Consider, as a useful intermediate, the coefficients $\widehat{\alpha}^\circ_{jk}$ from equation (10), 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 (9), see Proposition 4.1 below. These coefficients α°_{jk} will constitute our reference point when computing the expectation of $\widehat{\alpha}_{jk}$ and $\widehat{\alpha}^*_{jk}$, see Proposition 4.3. Similarly to what has been done above, we decompose $\widehat{\alpha}^\circ_{jk}$ into two parts:

$$\begin{aligned}
\widehat{\alpha}^\circ_{jk} &= \widehat{\beta}^\circ_{jk} + \widehat{\rho}^\circ_{jk} \\
&= \frac{1}{N}\sum_{t=1}^{N} m(X_t)\,\psi^\circ_{jk}(X_t) + \frac{1}{N}\sum_{t=1}^{N} Z_t\,\psi^\circ_{jk}(X_t).
\end{aligned}$$

Proposition 4.1 *The coefficients $\widehat{\alpha}^{\circ}_{jk}, \alpha^{\circ}_{jk}$ given by (10) and (9), respectively, fulfill uniformly over (j,k) in $\mathcal{J}_N$:*

$$\begin{array}{rcl} (i) \quad \mathcal{E}\widehat{\alpha}^{\circ}_{jk} & = & \alpha^{\circ}_{jk} \\ (ii) \quad Var\,\widehat{\alpha}^{\circ}_{jk} & = & \frac{1}{N}\left(\sigma_z^2 + Var(m(X_1)\psi^{\circ}_{jk}(X_1))\right) \\ & \le & \frac{1}{N}\left(\sigma_z^2 + \sup\{m^2(F^{-1}(u)); \frac{k-1}{2^j} \le u < \frac{k}{2^j}\}\right) \end{array}$$

We now come to the evaluation of the noise-dependent part of $\mathcal{E}[(\widehat{\alpha}_{jk}-\widehat{\alpha}^*_{jk})^2]$. Proposition 4.2 describes its behavior.

Proposition 4.2 *Consider the coefficients $\widehat{\rho}_{jk}$ and $\widehat{\rho}^*_{jk}$ as defined in (14) and (32) respectively. Then*

$$\mathcal{E}[(\widehat{\rho}_{jk} - \widehat{\rho}^*_{jk})^2] = O(\frac{2^j}{N^{3/2}}), \tag{34}$$

uniformly over (j,k) in $\mathcal{J}_N$.

In order to formulate the next proposition, let $V(g;a,b)$ denote the total variation of the real-valued function g over the finite interval $[a,b]$.

Proposition 4.3 *Assume (A1) through (A3). Define α°_{jk}, $\widehat{\alpha}_{jk}$ and $\widehat{\alpha}^*_{jk}$ as in (9), (13) and (30) respectively. Then we have the following assertions.*

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

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

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

uniformly over (j,k) in $\mathcal{J}_N$.

Finally, the next proposition establishes the order of $\mathrm{Var}(\widehat{\beta}_{jk})$.

Proposition 4.4 *Consider the coefficient $\widehat{\beta}_{jk}$ as defined in (14). Under assumptions (A1) and (A2) we have uniformly over $(j,k) \in \mathcal{J}_N$:*

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

4.3 Total error of approximation

The next proposition provides us with the total error of approximation of $\widehat{\alpha}_{jk}$ by $\widehat{\alpha}^*_{jk}$ while summing over (j,k) in a well-defined set $\mathcal{J}_N$.

Proposition 4.5 *Assume (A1) through (A3). Consider the set $\mathcal{J}_N$ as defined by expression (18), with δ such that*

$$\frac{\beta}{2\beta+1}+\frac{1}{4}<\delta<\frac{2\beta}{2\beta+1}\ . \tag{35}$$

We then have the following equalities:

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

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

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

from which it follows immediately that

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

Using the tools developed to prove Proposition 4.5, we obtain the following estimates on the order of the terms T_3, T_4 and T_5 of decomposition (28).

Proposition 4.6 *Under the assumptions of Proposition 4.5, the following holds.*

(i) $T_3=\sum_k\mathcal{E}[(\widehat{s}_{j_0,k}-s^\circ_{j_0,k})^2]=o(N^{-\frac{2\beta}{2\beta+1}})$,

(ii) $T_5=\sum_{(j,k)\in\mathcal{J}_N}\mathcal{E}[(\alpha^\circ_{jk}-\widehat{\beta}_{jk})^2]=o(N^{-\frac{2\beta}{2\beta+1}})$,

(iii) $T_4=\sum_k\mathcal{E}[(s^\circ_{j_0,k}-\widehat{b}_{j_0,k})^2]=o(N^{-\frac{2\beta}{2\beta+1}})$.

5 Asymptotic Gaussian approximation

We now consider only the model (29) with deterministic design, and we proceed using the bias and variance behaviour of the empirical wavelet coefficients $\widehat{\alpha}^*_{jk}$ in the fixed regression, fixed interval set-up, as given in Proposition 4.3. For the approximation of

$$\sum_{(j,k)\in\mathcal{J}_N}\mathcal{E}[(\delta^{(.)}(\widehat{\alpha}^*_{jk},\lambda_{jk})-\alpha^\circ_{jk})^2]$$

to a Gaussian model, we basically follow [NvS], Section 2.1. Using Proposition 4.3, and by Corollary 4 in [Sak], we first show that, with $\Phi(x)$ denoting the standard normal distribution function, and with $\sigma^2_{jk}:=\mathrm{Var}(\widehat{\alpha}^*_{jk})=\sigma^2_z/N$,

$$P\left(\pm(\widehat{\alpha}^*_{jk}\ -\ \alpha^\circ_{jk})/\sigma_{jk}\geq x\right)\ =\ (1\ -\ \Phi(x))(1\ +\ o(1)) \tag{36}$$

holds uniformly in $(j,k) \in \mathcal{J}_N$ and uniformly on some interval $-\infty < x \leq \Delta_N$, $\Delta_N \asymp N^\eta$ for some $\eta > 0$. Second, we can generalise this to hold uniformly in $x \in \mathbb{R}$ if we allow for an arbitrary polynomically fast decaying remainder of order $N^{-\lambda}$. With the relation (36) we derive the desired asymptotic risk equivalence. With this, we obtain, from known results in Gaussian regression (see, e.g., [DJ98, DJ94]), the exact analogue of our main Theorem 3.1, formulated now for the model (29) with deterministic design coefficients $\widehat{\alpha}^*_{jk}$.

5.1 Outline of proving the Gaussian approximation

For the reader's convenience we briefly summarize the technique of how to prove the Gaussian approximation. In order to be more general as needed we repeat the conditions of [NvS], Section 2.1, for the Gaussian approximation to hold in fixed design models with independent, not necessarily homoscedastic errors. Generalisations of this technique to the correlated case, can be found in [NvS], Section 2.2, and in [vSMG].

In the fixed design model

$$Y_i^* = m(q_i) + Z_i', \; i = 1, \ldots, N,$$

we suppose the design points q_i are equispaced quantiles of F_x that is,

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

To proceed we need the following assumptions which are implied by our assumptions (A1)–(A4).
The errors Z_i' are independently, not necessarily identically distributed with

(B2) $C_1 \;\leq\; \sigma_i^2 \;:=\; \mathrm{Var}(Z_i') \;\leq\; C_2.$

Again for simplicity, we assume that all moments are uniformly bounded, i.e.

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

Further, we assume that, at the coarsest level j_0, uniformly in k,

(B4) $\varphi^\circ_{j_0,k}$ and $\psi^\circ_{j_0,k}$ are of bounded total variation on I.

We start with showing (36). For our asymptotic approximation to (29) we consider the model

$$G_i \;=\; m(q_i) \,+\, \xi_i, \quad i = 1, \ldots, N, \tag{37}$$

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

First we state the following intermediate lemma.

Lemma 5.1 *Assume (B1) through (B4). Define the quantity*

$$\overline{\alpha}_{jk} \;=\; \sum_i (q_i - q_{i-1})\; w_{jk}(i)\; G_i, \tag{38}$$

where

$$w_{jk}(i) \;=\; N^{-1}\psi^{o}_{jk}(q_i)\;/(q_i - q_{i-1}).$$

Then, on an appropriate probability space,

$$P\left(|\widehat{\alpha}^{*}_{jk} \;-\; \overline{\alpha}_{jk}| > N^{\gamma-\delta/2}N^{-1/2} \quad \textit{for any} \quad (j,k)\in\mathcal{J}_N\right) \;=\; O(N^{-\lambda}) \tag{39}$$

for any $\gamma > 0$, $\lambda < \infty$, *where* $\mathcal{J}_N = \{(j,k) \mid\; 2^j \leq N^{1-\delta}\}$ *for some* $\delta > 0$.

Using this Lemma 5.1, we prove the asymptotic normality of coefficients $\widehat{\alpha}^{*}_{jk}$ as follows.

Proposition 5.2 *For arbitrary large $\lambda > 0$, we have, with $\sigma^2_{jk} := Var(\widehat{\alpha}^{*}_{jk})$, that*

$$P\left(\pm(\widehat{\alpha}^{*}_{jk} \;-\; \alpha^{\circ}_{jk})/\sigma_{jk} \geq x\right) = (1-\Phi(x))(1+o(1)) \;\; + O(N^{-\lambda}) \tag{40}$$

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

The strong form of asymptotic normality (40) proven, we can justify the l_2-risk equivalence between our model and the accompanying Gaussian model defined previously, now written in the coefficient domain as

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

where $\varepsilon_{jk} \sim N(0,\sigma^2_{j,k})$. The following theorem paralleling Theorem 4.1 of [DaNeu] states the l_2-risk equivalence between Gaussian and non-Gaussian model. For sake of completeness, we also reproduce the proof.

Theorem 5.3 *Let α°_{jk}, $\widehat{\alpha}^{*}_{jk}$, and $\overline{\alpha}_{jk}$ be defined in (9), (30), and (38), respectively, and let $\delta^{(\cdot)}(.,\lambda_{jk})$ denote either soft or hard thresholding, with scale and location dependent threshold λ_{jk}. Then we have*

$$\sum_{(j,k)\in\mathcal{J}_N} \mathcal{E}\left(\delta^{(\cdot)}(\widehat{\alpha}^{*}_{jk},\lambda_{jk}) - \alpha^{\circ}_{jk}\right)^2 \;=\; (1\;+\;o(1)) \sum_{(j,k)\in\mathcal{J}_N} \mathcal{E}\left(\delta^{(\cdot)}(\overline{\alpha}_{jk},\lambda_{jk}) - \alpha^{\circ}_{jk}\right)^2 \;+\; O(N^{-1}), \tag{41}$$

where $\mathcal{J}_N$ is defined in (18).

We observe now that by this Theorem 5.3, and in view of equation (27), the term T_2 of our original decomposition (28) can be replaced by

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

which completes the proof of the overall approximation including the Gaussian one.

5.2 Concluding remarks

It remains to complete the discussion of the total proof of our main Theorem 3.1 including the announced remarks on 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 [DJ94], the following relation holds:

$$\mathcal{E}[\delta^{(\cdot)}(\overline{\alpha}_{jk},\lambda_{jk})-\alpha^{\circ}_{jk}]^2 \le C \sum_{(j,k)\in\mathcal{J}_N}\left(\sigma^2_{jk}\left(\frac{\lambda_{jk}}{\sigma_{jk}}+1\right)\varphi\left(\frac{\lambda_{jk}}{\sigma_{jk}}\right)+\min\{\alpha^{\circ}_{jk}{}^2,\lambda^2_{jk}\}\right), \qquad (42)$$

where φ denotes the standard normal density. As stated in condition (A5) of [Neu], if the individual thresholds satisfy the condition

$$\begin{aligned}\textstyle\sum_{(j,k)\in\mathcal{J}_N}\left(\frac{\lambda_{jk}}{\sigma_{jk}}+1\right)\varphi\left(\frac{\lambda_{jk}}{\sigma_{jk}}\right) &= O(N^{1/(2\beta+1)})\\ \max_{(j,k)\in\mathcal{J}_N}\{\lambda_{jk}\} &= O(N^{-1/2}\sqrt{\log N})\end{aligned} \qquad (43)$$

then the L_2-risk of $\widehat{m}$ attains the rate $(\log N/N)^{\frac{2\beta}{2\beta+1}}$. Condition (43) however, does not give a practical rule to choose the λ_{jk} from the data. Hence Neumann provided in [Neu] a sufficient condition for random thresholds which ensures the desired rate for the estimator. This condition is recalled below.

(i) $\sum_{(j,k)\in\mathcal{J}_N}P\left(\lambda_{jk}<\alpha_N\sigma_{jk}\sqrt{2\log(C(\sharp\mathcal{J}_N))}\right)=O(N^{\nu})$ for any $C>0,\nu<1/(2\beta+1)$ and $\alpha_N\to 1$,

(ii) $\sum_{(j,k)\in\mathcal{J}_N}P(\lambda_{jk}>CN^{-1/2}\sqrt{\log N})=O(N^{-2\beta/(2\beta+1)})$ for any $C<\infty$.

In other word, the thresholds λ_{jk} must fulfill the inequality

$$\sigma_{jk}\sqrt{2\log(\sharp\mathcal{J}_N)}\le\lambda_{jk}\le CN^{-1/2}\sqrt{\log N}, \qquad (44)$$

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 3.1 by exactly following the proof of Theorem 5.2 of [Neu].

6 Numerical results

6.1 Introduction

In Section 4, we fully justify the approximation of our initial model (4) by model (29) only in the case of i.i.d. errors. Once we stick to model (29), however, results similar to the one of Section 5 can be obtained for a more general setting, that is, for errors which are stationary dependent and for independent but heteroscedastic errors (see Chapter 2.2 of [Neu]). This 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 [AP]. We thereby show through some empirical evidence that our estimator $\widehat{m}(x)$ based on random partitioning I_{jk} is performing well for the three following error models:

(a) $Y_t = m(X_t) + \epsilon_t^{(1)}$, with $\epsilon_t^{(1)}$ i.i.d. $N(0, \sigma^2)$.

(b) $Y_t = m(X_t) + \epsilon_t^{(2)}$, with $\epsilon_t^{(2)}$ following a stationary dependent $AR(1)$ model.

(c) $Y_t = m(X_t) + \epsilon_t^{(3)}$, with $\epsilon_t^{(3)} = \sigma(X_t)e_t$, e_t i.i.d $N(0,1)$. (Heteroscedastic regression).

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 if we assume we have errors of type (a), or at each level if we assume dependent errors as in (b), see [JS]. In the heteroscedastic model (c), in order to estimate the variance $\sigma_{jk}^2 := \mathrm{Var}(\widehat{\alpha}_{jk})$, it is of interest to look at the asymptotic variance of the detail coefficients. Writing $\widehat{\alpha}_{jk}$ as

$$\begin{aligned} \widehat{\alpha}_{jk} &= \frac{1}{N}\sum_t m(X_t)\psi_{j,k}(X_t) + \frac{1}{N}\sum_t \sigma(X_t)e_t\psi_{j,k}(X_t) \\ &= A + B, \end{aligned}$$

it is clear that we should threshold on the basis of $\mathrm{Var}(B)$ only. The asymptotic variance of B,

$$AVar(B) = AVar(\widehat{\alpha}_{jk}) = \frac{1}{N}\int \sigma^2(x)\psi_{j,k}^2(x)dF_x \tag{45}$$

gives us a straightforward way to estimate σ_{jk}. Simply taking the plug-in estimator of (45) gives

$$\widehat{\sigma}_{jk}^2 := \widehat{AVar}(\widehat{\alpha}_{jk}) = \frac{2^j}{N^2}\sum_{X_t \in I_{j,k}} \widehat{\sigma}^2(X_t). \tag{46}$$

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 [CD, NaSi], 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.

Our aim is to present an algorithm which is robust with respect to a change in the specification of the error models. Hence 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 3.1 tells us it is optimal (in a minimax sense) to use a threshold $\lambda_{j,k}$ such that

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

with $\sigma_{j,k}^2 = Var(\widehat{\alpha}_{jk})$. Based on this observation and expression (46), we propose to obtain an estimator of $m(x)$ in three steps.

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 $i := 0$.

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 follows:
$$\lambda_{j,k} = \widehat{\sigma}_{j,k}\sqrt{2\log N}$$
where $\widehat{\sigma}_{j,k}$ is given by (46): $\frac{2^j}{N^2}\sum_{X_t \in I_{j,k}} \widehat{\sigma}_{(i)}^2(X_t)$. Call the resulting estimator $\widehat{m}_{(i+1)}(x)$.

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. We mention here von Sachs and MacGibbon [vSMG] who proposed a level dependent threshold, which is also location dependent. On each given level, they find blocks of coefficients with approximately the same variance and apply a different threshold in each block.

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, $X_t \sim N(0.5, (0.1)^2)$. The partitioning used to define the estimator is the empirical quantile partitioning $\{I_{j,k}\}$. 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 a signal-to-noise ratio as SNR:= sd$(m)/\sigma_\epsilon$, where

$$sd(m)^2 = \frac{1}{N-1}\sum_{i=1}^{N}(m(x_i) - \overline{m(x)})^2, \tag{47}$$

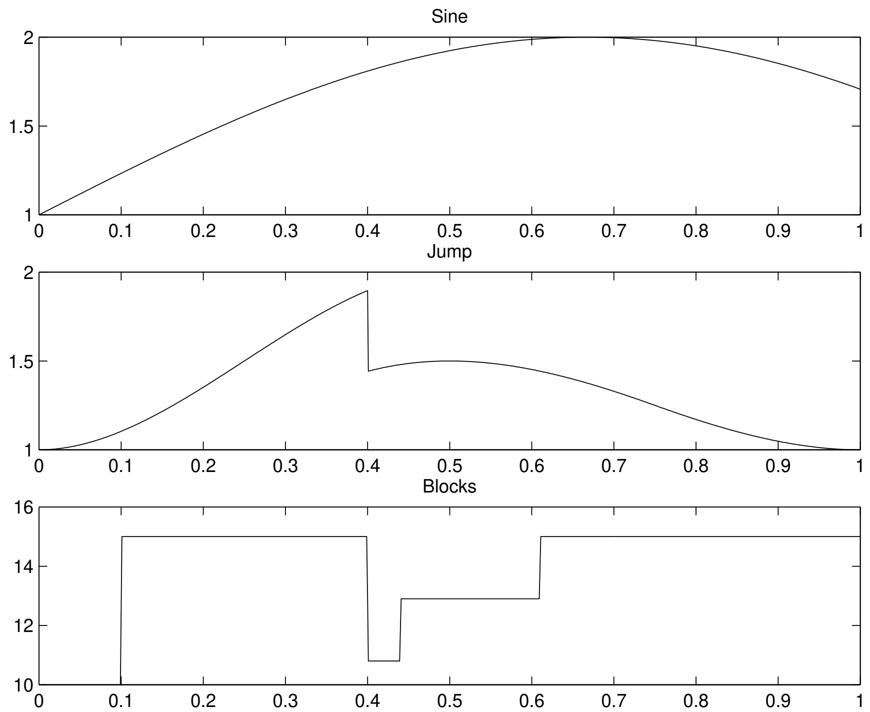

Figure 1: Underlying signals : smooth sine function, function containing a jump and blocks function.

for a realization $\{x_i\}_{i=1}^N$ of the random variable $X \sim N(0.5, 0.01)$, and $\overline{m(x)}$ is equal to the mean value of the $m(x_i)$, $i = 1, \ldots, N$. We set the SNR equal to one, which allows us to specify σ_ϵ. In the error model (a), we took, $\sigma := \sigma_\epsilon$. For model (b), we generated Gaussian autoregressive errors $\epsilon_t^{(2)} \sim AR(1)$ with parameter 0.9, and took $\sigma_\epsilon := \mathrm{sd}(\epsilon_t^{(2)})$, with $sd(.)$ computed as in (47). Finally, for model (c), $\sigma_\epsilon := \int \sigma(x)dx$, and this integral is approximated by averaging $\sigma(x)$ over discrete x. The function $\sigma(x)$ used in the simulations is the following (see [Gao]):

$$\sigma(x) = C(0.001 + (0.3 - 1.5(x - 0.515)^2)^2),$$

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:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \{\frac{\widehat{m}(x_i) - m(x_i)}{m(x_i)}\}^2, \tag{48}$$

where $\widehat{m}(x_i)$ is the estimated function and $m(x_i)$ is the real underlying function computed at x_i. When computing equation (48) 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. Hence we obtained a performance criterion which can be expressed as a "percentage of loss error": a value of 1% expresses an error in the reconstruction of 1%.

For the binned wavelet smoother of Antoniadis and Pham [AP] (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 (48). We are indebted to A. Antoniadis for supplying the BinWav code.

6.4 Simulation results

Table 1 gives all 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 inter-quartile intervals are never excessively large, which means that both methods are quite robust with respect to both the randomness of the design ('horizontal' variability) and the variance of the noise part ('vertical' variability).

The BinWav estimator uses smooth B-splines, hence it gives very good results for a smooth function like the sine function and outperforms our proposed estimator in this case. Figure 2 shows, in case of dependent errors, where this difference lies: it is mainly at the boundaries, where we have few data points, that the BinWave estimator gives better results. For this data set, the square root of the relative MSE (noted RMSE) is equal to 1.47% and 0.44%, for our method and the BinWav method, respectively. Note that we tried to illustrate the results of Table 1 by giving the graphs of estimates having a value of the RMSE approximately equal to the median RMSE value reported in this table.

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 3 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 less favorable for the BinWav method of Antoniadis and Pham, whereas our estimator copes pretty well with jumps. This explains the better results of our method for that function. In Figure 4 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 summarize, 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.

We also carried out some simulations using a level-dependent thresholding estimator based on MAD to estimate the variance σ_{jk}, as prescribed in [JS], for data sets having i.i.d. or dependent errors. In case of i.i.d. errors, the results were of the same order as the one given in Table 1, whereas in the case of stationary dependent errors, the results were significantly better than with our three steps procedure. Indeed, a level-dependent thresholding is the recommended procedure to deal with stationary dependent errors (see [JS]).

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. The numbers between brackets give the square root of the inter-quartile intervals of the MSE. Values are expressed in percent.

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

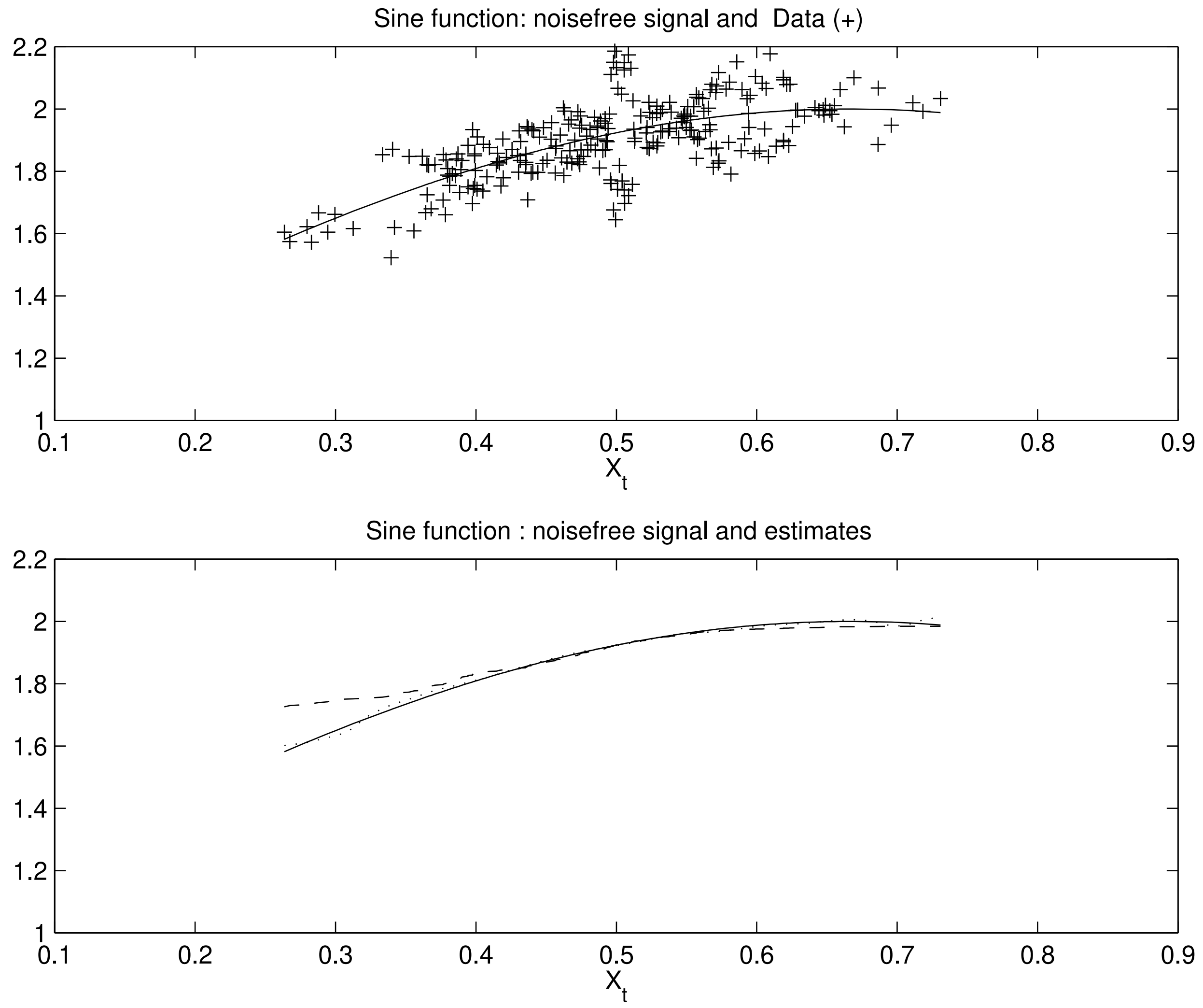

Figure 2: Estimation of the sine function with $N = 256$ data points. The first part of the graph represents the underlying signal and the data (represented by '+'), generated with stationary dependent errors. For comparison purpose, the second part shows the underlying signal (plain line), our estimator (dashed line) and the BinWav estimator (dotted line). The RMSE values were, respectively for both methods, equal to 1.47% and 0.44%.

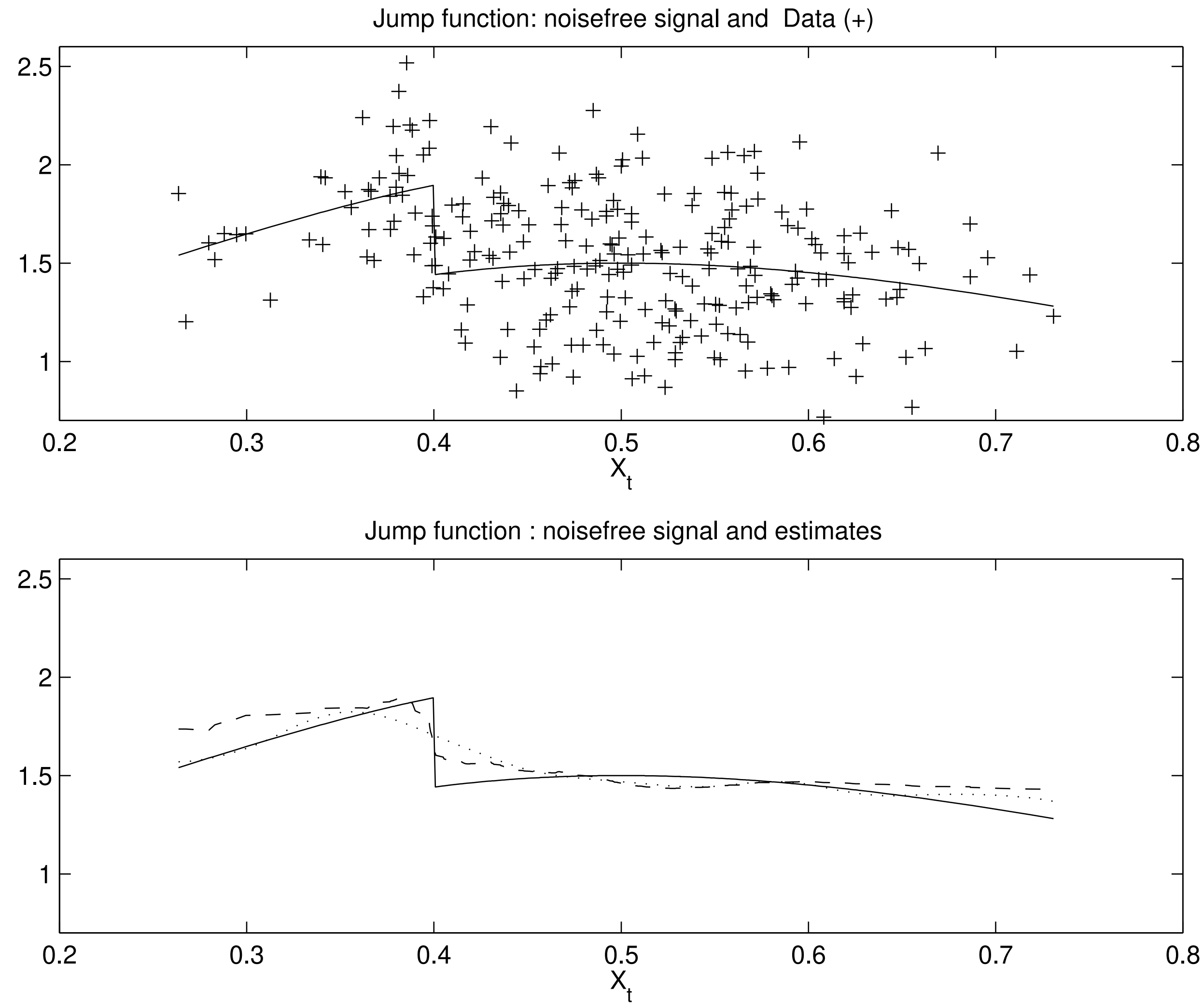

Figure 3: Estimation of the jump function with $N = 256$ data points. The first part of the graph represents the underlying signal and the data (represented by '+'), generated with independent heteroscedastic errors. The second part shows the underlying signal (plain line), our estimator (dashed line) and the BinWav estimator (dotted line). The RMSE values were, respectively for both methods, equal to 4.32% and 4.55%.

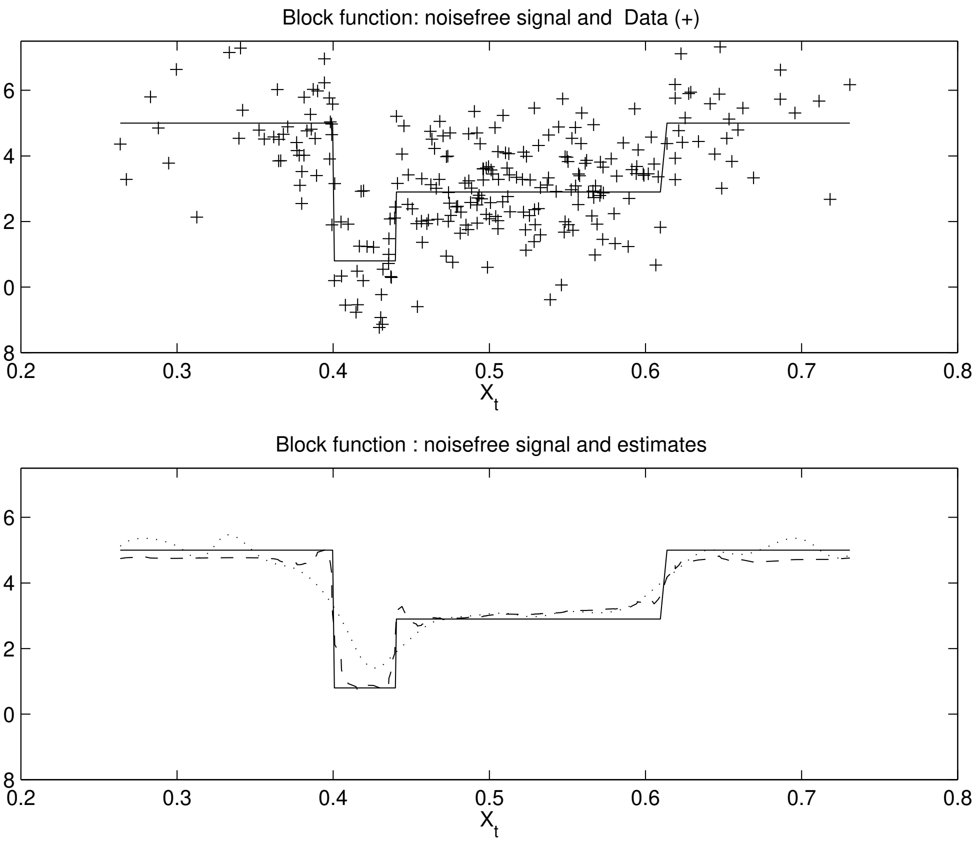

Figure 4: Estimation of the block function with $N = 256$ data points. The first part of the graph represents the underlying signal and the data (represented by '+'), generated with i.i.d. errors. The second part shows the underlying signal (plain line), our estimator (dashed line) and the BinWav estimator (dotted line). The RMSE values were, respectively for both methods, equal to 2.88% and 4.51%.

7 Appendix

Proof of Lemma 3.2:
Let q_α denote the $\alpha-$quantile of F_x. As $\int \psi^\circ_{jk} dF_x = 0$, we have:

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

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

$$|\langle m, \psi^\circ_{jk}\rangle_{dF_x}| \leq M \int |x - x_0|^\beta 2^{j/2} |1_{I^\circ_{j+1,2k}}(x) - 1^\circ_{I_{j+1,2k+1}}(x)|dF_x(x). \tag{49}$$

Let us concentrate on the term

$$\int |x - x_0|^\beta 1_{I^\circ_{j+1,2k}}(x)dF_x(x), \tag{50}$$

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 equation (50) can be re-expressed as:

$$\int_{F_x(q_{(2k-1)/2^{j+1}})}^{F_x(q_{2k/2^{j+1}})} |F_x^{-1}(y) - F_x^{-1}(y_0)|^\beta dy. \tag{51}$$

As by (A1) $f_x(x)$ is bounded away from zero on the support S, $F_x^{-1}(y)$ is differentiable with $F_x^{-1}(y_0)' = f_x(x_0)^{-1}$, so by a Taylor approximation

$$F_x^{-1}(y) - F_x^{-1}(y_0) = (y - y_0)\frac{1}{f_x(x_0)} + o(|y - y_0|). \tag{52}$$

Inserting (52) into (51) and integrating out, we obtain that (51) is approximated, up to a term of higher order, by

$$\frac{1}{f_x(x_0)^\beta(\beta+1)} \left| F_x(q_{\frac{2k}{2^{j+1}}}) - F_x(q_{\frac{2k-1}{2^{j+1}}}) \right|^{\beta+1} = \frac{2^{-(j+1)(\beta+1)}}{f_x(x_0)^\beta(\beta+1)}. \tag{53}$$

Inserting (53) into (49), we get:

$$\begin{aligned}
|\langle m, \psi^\circ_{jk}\rangle_{dF_x}| &\leq M 2^{j/2} \frac{2}{(\beta+1)f(x_0)^\beta} 2^{-(j+1)(\beta+1)} \\
&= C 2^{-j(\beta+\frac{1}{2})}
\end{aligned}$$

which ends the proof. $\square$

7.1 Proofs of the results of Section 4

Proof of Proposition 4.1:
(i) By (4), $\mathcal{E}Z_t = 0$ and the identical distribution of $X_1, \ldots, X_N$:

$$\mathcal{E}\widehat{\alpha}^\circ_{jk} = \mathcal{E}m(X_1)\psi^\circ_{jk}(X_1) = \alpha^\circ_{jk}.$$

(ii) By the independence of the X_t, Z_t, and as $\mathcal{E}Z_t = 0$ and ψ°_{jk} are orthonormal in $L^2(dF_x)$, we have

$$\begin{aligned}
\operatorname{Var}\widehat{\alpha}^\circ_{jk} &= \frac{1}{N}\operatorname{Var}(Y_1\psi^\circ_{jk}(X_1)) \\
&= \frac{1}{N}\left(\sigma_z^2\mathcal{E}(\psi^\circ_{jk}(X_1))^2 + \operatorname{Var}(m(X_1)\psi^\circ_{jk}(X_1))\right) \\
&= \frac{1}{N}\left(\sigma_z^2 + \operatorname{Var}(m(X_1)\psi^\circ_{jk}(X_1))\right).
\end{aligned}$$

As $|\psi^\circ_{jk}| = |\varphi^\circ_{jk}|$, we have, using the definition of I°_{jk} as a quantile interval of F_x-mass 2^{-j} :

$$\begin{aligned}
\operatorname{Var}(m(X_1)\psi^\circ_{jk}(X_1)) &\le \mathcal{E}(m(X_1)\psi^\circ_{jk}(X_1))^2 \\
&= 2^j\int_{I^\circ_{jk}} m^2(x)dF_x(x) \\
&\le \sup\{m^2(x); x \in I^\circ_{jk}\}.
\end{aligned}$$

□

Proof of Proposition 4.2:
Using $\widehat{\rho}^\circ_{jk}$ as an intermediate coefficient, we write:

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

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 have

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

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

$$\begin{aligned}
V_1 &= \frac{1}{Nn_j}\sum_{t=1}^{N} \mathcal{E}(1^{\circ}_{jk}(X_{(t)}) - 1^{\circ}_{jk}(q_t))^2 \\
&\leq \frac{1}{Nn_j}\sum_{t=1}^{N} \mathcal{E}\, 1_{\Delta jk}(X_{(t)}) + \frac{2}{Nn_j}
\end{aligned}$$

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 7.1 below. Using that Lemma, we have

$$\begin{aligned}
V_1 &= \frac{1}{Nn_j}\sum_{t=1}^{N} \mathcal{E}1_{\Delta jk}(X_t) + \frac{2}{Nn_j} \\
&= \frac{1}{Nn_j}\sum_{t=1}^{N} \mathrm{pr}(X_t \in \Delta_{jk}) + \frac{2}{Nn_j} = O(\frac{1}{n_j\sqrt{N}})
\end{aligned}$$

uniformly in j, k. As V_2 is of exactly the same form, we have the desired result for $\mathcal{E}(\widehat{\rho}^{\circ}_{jk} - \widehat{\rho}^{*}_{jk})^2$.
(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 $I_{jk},\ k = 1, \ldots, 2^j$, is specified by the order statistics $X_{(1)} \leq \ldots \leq X_{(N)}$, we get, conditioning on the latter and using $n_j 2^j = N$:

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

as $Z_1, \ldots, Z_N$ are zero-mean i.i.d. random variables independent of the i.i.d. variables $X_1, \ldots, X_N$. By Lemma 7.1, with Δ_{jk} denoting the symmetric difference of I_{jk} and I_{jk}°, we conclude:

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

□

Lemma 7.1 *Let $X_1, \ldots, X_N$ 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 $n_j 2^j = N$, $k = 1, \ldots, 2^j$, 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,*

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

uniformly in j, k.

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

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

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

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:

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

as ranks and order statistics of an i.i.d. sample are independent and, by symmetry, $\mathrm{pr}(R_1 = s) = \frac{1}{N}$ 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., [Ser], Chapter 2.4) we have, using $F(q_{kn_j}) = \frac{kn_j}{N}$:

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

Therefore,

$$\begin{aligned}
\mathrm{pr}(q_{kn_j} \leq X_1 < X_{(kn_j+1)}) &= \frac{1}{N} \sum_{s=1}^{kn_j} \sum_{i=0}^{s-1} \mathrm{pr}(B_{jk} = i) \\
&= \frac{1}{N} \sum_{i=0}^{kn_j} (kn_j - i)\ \mathrm{pr}(B_{jk} = i) \\
&= \frac{1}{N}\, \mathcal{E}(kn_j - B_{jk})_+ \\
&\leq \frac{1}{N}\, \mathcal{E}|B_{jk} - kn| \\
&\leq \frac{1}{N} \sqrt{\mathrm{Var} B_{jk}} = \frac{1}{\sqrt{N}} \sqrt{\frac{kn_j}{N}(1 - \frac{kn_j}{N})}
\end{aligned}$$

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

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

□

Proof of Proposition 4.3:
(i) Let $g_{jk}(u) = m(F_x^{-1}(u))\psi^\circ_{jk}(F_x^{-1}(u))$. Using this abbreviation, we have by definition of α°_{jk} and of q_t for $1 < k < 2^j$:

$$\begin{aligned}
|\mathcal{E}\widehat{\alpha}^*_{jk} - \alpha^\circ_{jk}| &= |\frac{1}{N} \sum_{t=1}^{N} m(q_t)\psi^\circ_{jk}(q_t) - \int_{-\infty}^{\infty} m(x)\psi^\circ_{jk}(x) dF_x(x)| \\
&= |\frac{1}{N} \sum_{t=1}^{N} g_{jk}(\frac{t}{N}) - \int_0^1 g_{jk}(u) du| \\
&\leq \frac{1}{N} V(g_{jk}; 0, 1)
\end{aligned}$$

using an inequality of Pólya and Szegö for functions of bounded variation in Chapter 2.1 of [PS].

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

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

(ii) By Proposition 4.1, $\mathcal{E}(\widehat{\alpha}_{jk}) - \alpha^\circ_{jk} = \mathcal{E}(\widehat{\alpha}_{jk} - \widehat{\alpha}^\circ_{jk})$. We then have:

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

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 $\tau = (k - \frac{1}{2})n_j$ such that q_τ is in the center of I°_{jk} dividing the interval into two halves of equal F_x-mass. We have

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

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 $U_t = F_x(X_t),\ t = 1, \ldots, N$, i.i.d. uniformly on $[0,1]$. Let $\chi_{jk}(u) = 1_{jk}(F_x^{-1}(u)),\ \chi^\circ_{jk}(u) = 1^\circ_{jk}(F_x^{-1}(u))$ be the indicator functions in terms of the U_t, i.e.

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

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

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

By Lemma 7.2 below, we have

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

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

$$\mathrm{pr}(|U_{(s)} - \frac{s}{N}| \leq \sqrt{\frac{2\log N}{N}} \text{ for all } s = 1, \dots, N) = 1$$

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

$$\mathcal{E}|\chi_{jk}(U_t) - \chi^{\circ}_{jk}(U_t)| \leq 4(\sqrt{\frac{2\log N}{N}} + \frac{1}{N}).$$

Under assumptions (A1) through (A3), Lemma 7.3 below shows that $|g(u) - g(v)| \leq \tilde{H}|u-v|^{\beta}$ 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

$$|u - \frac{\tau}{N}| \leq \frac{n_j}{N} + \sqrt{\frac{2\log N}{N}}$$

for all u with $\chi_{jk}(u) - \chi^{\circ}_{jk}(u) \neq 0$, we get

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

The same inequality holds for E_2, and this ends the proof of (ii).
(iii) By definition of $\widehat{\alpha}^*_{jk}$, we have

$$\mathrm{Var}(\widehat{\alpha}^*_{jk}) = \mathrm{Var}\left(\frac{1}{N}\sum_{t=1}^{N} Z'_t \psi^{\circ}_{jk}(q_t)\right) = \frac{\sigma_z^2}{N}\frac{1}{N}\sum_{t=1}^{N} \psi^{\circ 2}_{jk}(q_t),$$

and there are only n_j elements different from zero in this sum. Hence we obtain:

$$\mathrm{Var}(\widehat{\alpha}^*_{jk}) = \frac{\sigma_z^2}{N}\frac{2^j n_j}{N} = \frac{\sigma_z^2}{N}.$$

□

Lemma 7.2 *Let $U_1, \dots, U_N$ be i.i.d. with common distribution function G, and let $U_{(1)} \leq \dots \leq U_{(N)}$ denote the order statistics.*

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

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

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

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

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

Proof: (i) This result is essentially stated in the proof of Lemma C of Serfling ([Ser], Chapter 2.5.4), and it is an immediate consequence of Hoeffding's Lemma ([Ser], 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 7.3 *Assume (A1) through (A3). Let j be fixed. Then for N large enough the Hölder condition*

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

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 $u < v$

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

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. □

Proof of Proposition 4.4:
Similarly to the proof of Proposition 4.3, we decompose $\widehat{\beta}_{jk}$ as follows:

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

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 $I_{j,k}$:

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

by the Cauchy-Schwarz inequality. Let $g(u), U_t$ be as in the proof of Proposition 4.3. We have, for $(k-1)n_j + 1 \leq t \leq kn_j$,

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

by Lemma 7.3. Now by Lemma 7.2 above, we have

$$\tilde{H}^2\mathcal{E}[(U_{(t)} - \frac{t}{N})^2] \leq \tilde{H}^2\left(\frac{2\log N}{N}\right)^{\beta},$$

for N large enough such that $|U_{(s)} - \frac{s}{N}| \leq \sqrt{\frac{2\log N}{N}}$ for all s with probability 1. Therefore,

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

□

Proof of Proposition 4.5:
(i) By Proposition 4.2, we have

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

Summing over j, k, we obtain,

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

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 required 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

$$\begin{aligned} \textstyle\sum_{(j,k)\in\mathcal{J}_N} (\mathcal{E}[\widehat{\alpha}^*_{jk} - \widehat{\alpha}^\circ_{jk}])^2 &= o(N^{-\frac{2\beta}{2\beta+1}}) && (54) \\ \textstyle\sum_{(j,k)\in\mathcal{J}_N} (\mathcal{E}[\widehat{\alpha}_{jk} - \widehat{\alpha}^\circ_{jk}])^2 &= o(N^{-\frac{2\beta}{2\beta+1}}), && (55) \end{aligned}$$

the result is proved. We first prove (54). By Proposition 4.3(i), $\mathcal{E}[\widehat{\alpha}^*_{jk} - \widehat{\alpha}^\circ_{jk}] \le C_1\ 2^{j/2}/N$, 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,

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

Second, by the proof of Proposition 4.3(ii), we have

$$\mathcal{E}[\widehat{\alpha}_{jk} - \widehat{\alpha}^\circ_{jk}] \le \tilde{H}\ C_2 \sqrt{\frac{\log N}{N2^{-j}}} \left(\frac{1}{2^j} + \sqrt{\frac{\log N}{N}}\right)^\beta . \tag{56}$$

In our assumptions, we have $\beta > 1/2$. Hence taking $\beta = 1/2$ in (56) will overestimate the bound of the error of approximation; in this case the square of (56) is of order $\frac{\log N}{N} + 2^j \left(\frac{\log N}{N}\right)^{3/2}$. The first term is not of sufficiently small order, whereas the second term is of sufficiently small order in order to have the result (55), provided that

$$\delta > \frac{2\beta}{4\beta+2} + \frac{1}{4} + \epsilon,$$

for $\epsilon > 0$ arbitrarily small.

Let us now concentrate on the term $\sqrt{(\log N)/(N2^{-j})}(1/2^j)^\beta$ in equation (56). For $\beta > 1/2$, we have:

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

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

$$\textstyle\sum_{(j,k)\in\mathcal{J}_N} (\mathcal{E}[\widehat{\alpha}_{jk} - \widehat{\alpha}^\circ_{jk}])^2 = \frac{\log N}{N}\, 2^{2J(1-\beta)} \tag{57}$$

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 (57) will be of order $o(N^{-\frac{2\beta}{2\beta+1}})$ if

$$-1 + \epsilon + (2 - 2\delta)(1-\beta) < -\frac{2\beta}{2\beta+1},$$

i.e. if

$$\delta > (1 - \frac{4\beta^2}{2\beta+1} + \epsilon)\frac{1}{2(1-\beta)}$$

and this condition is automatically fullfilled when

$$\beta > 1/2 \text{ and } \delta > \frac{2\beta}{4\beta+2} + \frac{1}{4}.$$

(iii) By Proposition 4.4, we have:

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

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

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

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

Proof of Proposition 4.6:
(i) We decompose the mean square error in bias and variance:

$$\mathcal{E}[(\widehat{s}_{j_0,k} - s^{\circ}_{j_0,k})^2] = \mathrm{Var}(\widehat{s}_{j_0,k}) + (\mathcal{E}[\widehat{s}_{j_0,k} - s^{\circ}_{j_0,k}])^2.$$

By reproducing the proof of Proposition 4.4, we obtain

$$\mathrm{Var}(\widehat{s}_{j_0,k}) = O\left(2^{-j_0}\left(\frac{\log N}{N}\right)^{\beta}\right)$$

and hence $\sum_k \mathrm{Var}(\widehat{s}_{j_0,k})$ is of the required order for $\beta > 1/2$.

To treat the bias term, we do the same as in the proof of Proposition 4.3(ii) and obtain

$$\mathcal{E}[\widehat{s}_{j_0,k}] - s^{\circ}_{j_0,k} = O\left(\sqrt{\frac{\log N}{N2^{-j_0}}}(2^{-j_0} + \sqrt{\frac{\log N}{N}})^{\beta}\right),$$

which, as j_0 is fixed, gives us the desired result.
(ii) Note that

$$\begin{aligned}
\mathcal{E}[(\widehat{\beta}_{jk} - \alpha^{\circ}_{jk})^2] &= (\mathcal{E}[\widehat{\beta}_{jk} - \alpha^{\circ}_{jk}])^2 + \mathrm{Var}(\widehat{\beta}_{jk}) \\
&= (\mathcal{E}[\widehat{\alpha}_{jk} - \alpha^{\circ}_{jk}])^2 + \mathrm{Var}(\widehat{\beta}_{jk}).
\end{aligned}$$

Proposition 4.3(ii) gives us the order or $\mathcal{E}[\widehat{\alpha}_{jk} - \alpha^{\circ}_{jk}]$ and, using part of the demonstration in Proposition 4.5(ii), we obtain that $\sum_{(j,k)\in\mathcal{J}_N}(\mathcal{E}[\widehat{\alpha}_{jk} - \alpha^{\circ}_{jk}])^2 = o(N^{-\frac{2\beta}{2\beta+1}})$. Finally, by Proposition 4.5(iii) we have $\sum_{(j,k)\in\mathcal{J}_N}\mathrm{Var}(\widehat{\beta}_{jk}) = o(N^{-\frac{2\beta}{2\beta+1}})$.
(iii) The proof is completely similar to the one of item (ii). $\square$

7.2 Proofs of the results of Section 5

Proof of Lemma 5.1:
We can first write the empirical coefficients as

$$\widehat{\alpha}^*_{jk} = \sum_i (q_i - q_{i-1}) \; w_{jk}(i) \; Y^*_i .$$

To obtain a connection between $\widehat{\alpha}^*_{jk}$ and $\overline{\alpha}_{jk}$, we consider the partial sum processes

$$S_j = \sum_{i \le j} Z_i \; (q_i - q_{i-1})$$

and

$$T_j = \sum_{i \le j} \xi_i \; (q_i - q_{i-1}).$$

Under assumptions (B1) through (B3), we infer by Corollary 4 in [Sak] that there exists a joint probability space on which

$$P\left(\max_{1 \le i \le N} \{|S_i - T_i|\} > N^{\gamma - 1} \right) = O(N^{-\lambda}) \tag{58}$$

holds for arbitrarily small $\gamma > 0$ and arbitrarily large $\lambda < \infty$.

As $\psi^\circ_{j_0,k}(x)$ is of bounded total variation, we have that:

$$\sum_{i=1}^{N} |\psi^\circ_{j_0,k}(q_i) - \psi^\circ_{j_0,k}(q_{i-1})| + \psi^\circ_{j_0,k}(q_N) \le C.$$

Hence for the wavelet function at scale j, we have:

$$\sum_{i=1}^{N} |\psi^o_{j,k}(q_i) - \psi^o_{j,k}(q_{i-1})| + \psi^o_{j,k}(q_N) \le 2^{j/2} C. \tag{59}$$

Finally,

$$\begin{aligned} &\sum_{i=1}^{N} |w_{jk}(i) - w_{jk}(i+1)| + w_{jk}(N) \\ \le\; & \frac{1}{N \min_{1 \le i \le N} |q_{i+1} - q_i|} \{ \sum_{i=1}^{N} |\psi^o_{j,k}(q_i) - \psi^o_{j,k}(q_{i+1})| + \psi^o_{j,k}(q_N) \} \end{aligned} \tag{60}$$

If the q_i's are the quantiles of a uniform distribution $U(0,1)$, we have $q_{i+1} - q_i = N^{-1}$. This implies that, even in the non-uniform design case, we still keep the exact order $\mathrm{O}(N^{-1})$, i.e., $aN^{-1} \le \min_{1 \le i \le N} |q_{i+1} - q_i| \le bN^{-1}$. This is a consequence of our assumption (A1) on the design density $f_x(x)$ to be bounded away from zero and infinity, and being continuous. Combining equations (59) and (60), we obtain that

$$\sum_{i=1}^{N} |w_{jk}(i) - w_{jk}(i+1)| + w_{jk}(N) = O(2^{j/2}). \tag{61}$$

Now, equation (61) yields by summation by parts that

$$
\begin{aligned}
|\widehat{\alpha}^*_{jk} - \overline{\alpha}_{jk}| &= \sum_{i=1}^{N-1} |w_{jk}(i)(Z_i - \xi_i)(q_i - q_{i-1})| && (62)\\
&= \left| \sum_{i=1}^{N-1} (w_{jk}(i) - w_{jk}(i+1))(S_i - T_i) + w_{jk}(N)(S_N - T_N) \right| \\
&\leq C\, 2^{j/2} \max_{1\leq i\leq N} \{|S_i - T_i|\}. && (63)
\end{aligned}
$$

Finally, equations (58) and (63) yield that

$$
P(|\widehat{\alpha}^*_{jk} - \overline{\alpha}_{jk}| > 2^{j/2} N^{\gamma-1}) = O(N^{-\lambda}), \text{ for } \gamma > 0, \lambda < \infty. \tag{64}
$$

As, by definition of $\mathcal{J}_N, 2^j \leq N^{1-\delta}$, necessarily $N^{-\delta/2-1/2}$ is larger than $2^{j/2}N^{-1}$, and hence the desired result (39) follows directly from (64). □

Proof of Proposition 5.2:
We first show the following intermediate result:
In Lemma 5.1, let γ be chosen such that $0 < \gamma < \delta/2$. Then,

$$
P\left(\pm(\widehat{\alpha}^*_{jk} - \alpha^\circ_{jk})/\sigma_{jk} \geq x\right) = (1 - \Phi(x))(1 + o(1)) , \tag{65}
$$

uniformly in $(j, k) \in \mathcal{J}_N$ and uniformly over $\infty < x \leq \Delta_N$, where $\Delta_N := N^{\delta/2-\epsilon}$ for some ϵ with $\gamma < \epsilon < \delta/2$.

To show this we concentrate on expression (65) with the 'plus' sign. The 'minus' sign can be treated in a similar way. We rewrite (65) as

$$
P(A_N + B_N + C_N \geq x),
$$

where

$$
A_N = (\widehat{\alpha}^*_{jk} - \overline{\alpha}_{jk})/\sigma_{j,k} \ ,
$$

$$
B_N = (\overline{\alpha}_{jk} - \mathcal{E}\overline{\alpha}_{jk})/\sigma_{j,k} \ ,
$$

and

$$
C_N = (\mathcal{E}\overline{\alpha}_{jk} - \alpha^\circ_{jk})/\sigma_{j,k}.
$$

Note that C_N describes the (normalized) bias in estimating α°_{jk} which is the same in both model (37) and (29), i.e. with Proposition 4.3,

$$
\mathcal{E}\overline{\alpha}_{jk} = \mathcal{E}\widehat{\alpha}^*_{jk} = \alpha^\circ_{jk} + O(2^{j/2}N^{-1}) = \alpha^\circ_{jk} + O(N^{-1/2-\delta/2}) \ ,
$$

as the two models differ only by the distribution of their errors. Note that $C_N = O(N^{-\delta/2})$ as $\sigma_{j,k}$ is exactly of order $N^{-1/2}$.

By definition of the approximating Gaussian model, we have:

$$
P(B_N \leq x) = \Phi(x) \ ,
$$

hence

$$P(B_N + C_N \le x) = P(B_N \le x - C_N) = \Phi(x - C_N).$$

By the lemma of Slutsky, we have the following relationships:

$$P(B_N \le x - \epsilon) - P(|A_N| \ge \epsilon) \le P(A_N + B_N \le x) \le P(B_N \le x + \epsilon) + P(|A_N| \ge \epsilon).$$

By Lemma 5.1, with $\lambda = \lambda_1$, we know that:

$$P\left(\frac{|\widehat{\alpha}^*_{jk} - \overline{\alpha}_{jk}|}{N^{-1/2}} > N^{\gamma-\delta/2}\right) = O(N^{-\lambda_1}),$$

for any $(j,k) \in \mathcal{J}_N$. As γ is chosen such that $\gamma < \delta/2$, we have $N^{\gamma-\delta/2} = N^{-\alpha}$, for some positive constant α and $\sigma_{j,k}$ is exactly of order $N^{-1/2}$, which means that we can write $\sigma_{j,k} = CN^{-1/2}$. It follows that

$$P\left(\frac{|\widehat{\alpha}^*_{jk} - \overline{\alpha}_{jk}|}{\sigma_{j,k}} > CN^{-\alpha}\right) = O(N^{-\lambda_1}) \qquad \text{for any } (j,k) \in \mathcal{J}_N,$$

that is, the probability $P(|A_N| \ge CN^{-\alpha}) = O(N^{-\lambda_1})$. Taking $\epsilon_N = CN^{-\alpha}$ in Slutsky's lemma, we get, on one hand:

$$\begin{aligned} P(A_N + B_N + C_N \le x) &\le P(B_N + C_N \le x + \epsilon_N) + P(|A_N| \ge \epsilon_N) \\ &= \Phi(x - C_N + \epsilon_N) + P(|A_N| \ge \epsilon_N). \end{aligned}$$

On the other hand, we have:

$$\begin{aligned} P(A_N + B_N + C_N \le x) &\ge P(B_N \le x - C_N - \epsilon_N) - P(|A_N| \ge \epsilon_N) \\ &= \Phi(x - C_N - \epsilon_N) - P(|A_N| \ge \epsilon_N). \end{aligned}$$

It remains to treat the convergence of $\Phi(x - C_N \pm \epsilon_N)$ to $\Phi(x)$, which has to be uniform over $\infty < x \le \Delta_N$. This can be treated similarly to [Neu], proof of Theorem 4.1. Let, w.l.o.g., $\rho_N := \epsilon_N - C_N$ (the other case works similarly). We actually show that

$$\frac{1 - \Phi(x + \rho_N)}{1 - \Phi(x)} = 1 + o(1) ,$$

uniformly over $\infty < x \le \Delta_N$. For this it remains to show that

$$\eta_N := \left|\frac{1 - \Phi(x + \rho_N)}{1 - \Phi(x)} - 1\right| = \left|\frac{\Phi(x + \rho_N) - \Phi(x)}{1 - \Phi(x)}\right| = o(1).$$

Let, w.l.o.g., $\rho_N \ge 0$ (otherwise we would estimate $|\frac{1-\Phi(x)}{1-\Phi(x+\rho_N)} - 1|$ in the same manner as below). It is clear that η_N tends to zero as $N \to \infty$ if $x \le c$. For $c < x \le \Delta_N$ we obtain, similarly to [Neu], that

$$\eta_N \le \frac{\rho_N \exp\{-x^2/2\}/\sqrt{2\pi}}{(1 - x^{-2})x^{-1} \exp\{-x^2/2\}/\sqrt{2\pi}} \le \frac{\rho_N x}{1 - c^{-2}} = o(1) ,$$

as $\rho_N\, x \le \rho_N\, \Delta_N = o(1)$. (Recall that $\epsilon_N\, \Delta_N = O(N^{\gamma-\epsilon})$ and that $C_N\, \Delta_N = O(N^{-\delta/2}\, N^{\delta/2-\epsilon}.)$

That is, for an arbitrary large $\lambda_1 > 0$, we have shown that

$$P(A_N + B_N + C_N \ge x) = (1 - \Phi(x))(1 + o(1)) \ + O(N^{-\lambda_1}) \ ,$$

uniformly in $(j,k) \in \mathcal{J}_N$ and uniformly over $\infty < x \le \Delta_N$.

To get the result uniform over all $x \in \mathbb{R}$, we use the following two facts:
First, for $x \ge \Delta_N$, $P(|A_N + B_N + C_N| \ge x) \le P(|A_N + B_N + C_N| \ge \Delta_N)$ and

$$P(|A_N + B_N + C_N| \ge \Delta_N) \le C\ (1 - \Phi(\Delta_N)) \le C\ \Delta_N^{-1} \exp\{-\Delta_N^2/2\}/\sqrt{2\pi} = O(N^{-\lambda_2}) \ ,$$

uniformly in $x \ge \Delta_N$, for an arbitrary large λ_2.
Second, again uniformly in $x \ge \Delta_N$, $1 - \Phi(x) \le 1 - \Phi(\Delta_N) \le O(N^{-\lambda_2})$ (see above).

Hence, for arbitrary large $\lambda > 0$,

$$P\left(\pm(\widehat{\alpha}^*_{jk}\ -\ \alpha^\circ_{jk})/\sigma_{jk} \ge x\right) = (1 - \Phi(x))(1 + o(1)) \ + O(N^{-\lambda}) \ ,$$

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

Proof of Theorem 5.3:
As the function $\delta^{(.)}$ is monotonic in its first argument, we have that

$$(\delta^{(.)}(\widehat{\alpha}^*_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})^2 > x$$

if and only if

$$\widehat{\alpha}^*_{jk} - \alpha^\circ_{jk} \ge / > f_{jk}(x) \quad \text{ or } \quad \widehat{\alpha}^*_{jk} - \alpha^\circ_{jk} \le / < g_{jk}(x),$$

for appropriate functions f_{jk} and g_{jk}, where the equality sign has to be included if and only if $\delta^{(.)}(., \lambda_{jk})$ is not left- or right-continuous, respectively, at the corresponding points x. Hence we obtain from Proposition 5.2 that, uniformly in $0 \le x$,

$$\begin{aligned} & P((\delta^{(.)}(\widehat{\alpha}^*_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})^2 > x) \\ = \ & P((\delta^{(.)}(\overline{\alpha}_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})^2 > x)(1 + o(1)) \ + \ O(N^{-\lambda}). \end{aligned} \tag{66}$$

Let $\rho_{jk} = \sup\{x : P((\delta^{(.)}(\widehat{\alpha}^*_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})^2 > x) \ge N^{-\lambda}\}$. Then

$$P((\delta^{(.)}(\widehat{\alpha}^*_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})^2 > \rho_{jk}) \le N^{-\lambda}, \tag{67}$$

which yields by the Cauchy-Schwarz inequality that

$$\int_{\rho_{jk}}^\infty P(A^2 > x)dx \le \mathcal{E}_Y[A^2 1_{\{A^2 > \rho_{jk}\}}] \le \sqrt{\mathcal{E}(A^4)}\sqrt{P(A^2 > \rho_{jk})},$$

where $A := (\delta^{(.)}(\widehat{\alpha}^*_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})$. In order to bound $\mathcal{E}(A^4)$, let us first remark that

$$|\delta^{(.)}(\widehat{\alpha}^*_{jk}, \lambda_{jk}) - \alpha^\circ_{jk}| \le |\widehat{\alpha}^*_{jk} - \alpha^\circ_{jk}| + 2|\alpha^\circ_{jk}|,$$

hence $\mathcal{E}(A^4) \leq \mathcal{E}([|\widehat{\alpha}^*_{jk} - \alpha^\circ_{jk}| + 2|\alpha^\circ_{jk}|]^4)$. This is bounded due to our assumption (A3) of bounded moments of arbitrary order. Together with equation (67), we obtain

$$\int_{\rho_{jk}}^{\infty} P(A^2 > x)dx \leq O(1)\; O(N^{-\lambda/2}).$$

Writing the expectation of A^2 as

$$\mathcal{E}(A^2) = \int P(A^2 > x)dx = \int_0^{\rho_{jk}} P(A^2 > x)dx + \int_{\rho_{jk}}^{\infty} P(A^2 > x)dx$$

and using (66) to obtain for $0 < x < \rho_{jk}$ that

$$\int_0^{\rho_{jk}} P(A^2 > x)dx = \int_0^{\rho_{jk}} P((\delta^{(\cdot)}(\overline{\alpha}_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})^2 > x)(1 + o(1))dx \; + \; O(N^{-\lambda})$$

allows us to conclude that

$$\mathcal{E}(\delta^{(\cdot)}(\widehat{\alpha}^*_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})^2 = (1 + o(1))(\mathcal{E}(\delta^{(\cdot)}(\overline{\alpha}_{jk}, \lambda_{jk}) - \alpha^\circ_{jk})^2) + O(N^{-\lambda/2})$$

and finally, choosing the arbitrary λ to be $\lambda = 4$,

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

$\square$

ACKNOWLEDGEMENTS

The authors would like to thank C. Almeida, A. Antoniadis, A. Kneip, M. Neumann and W. Sweldens for helpful discussions. 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

[AF] Antoniadis, A. and Fan, J. (2000) Regularization of wavelets approximations. Technical report, IMAG, Grenoble, France.

[AGV] Antoniadis, A., Grégoire, G. and Vial, P. (1997). Random design wavelet curve smoothing. *Statistics and Probability Letters*, **35**, 225–232.

[AP] Antoniadis, A. and Pham, D.T. (1998). Wavelet regression for random or irregular design. *Computational Statistics and Data Analysis*, **28**, 353–369.

[BW] Berkner, K. and Wells, R. (1998). Smoothness estimates for soft-threshold denoising via translation invariant wavelet transforms. Technical report, Department of Mathematics, Rice University.

[CB98] Cai, T. and Brown, L. D. (1998) Wavelet Shrinkage for nonequispaced samples. *The Annals of Statistics*, **26**, 1783–1799.

[CB99] Cai, T. and Brown, L. D. (1999) Wavelet regression for random uniform design. *Statistics and Probability Letters*, **42**, 313–321.

[CDF] Cohen, A., Daubechies, I. and Feauveau, J. (1992). Bi-orthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **45**, 485–560.

[CD] Coifman, R. R. and Donoho D. L. Translation-invariant de-noising. in: A. Anestis, G. Oppenheim (eds), "Wavelets and Statistics", *Lect. Notes Statist.*, **103**, 125–150. Springer-Verlag: New York.

[DaNeu] Dahlhaus, R. and Neumann, M.H. (2000). Locally adaptive fitting of semiparametric models to nonstationary time series. Beiträge zur Statistik 56, Universität Heidelberg.

[Dau] Daubechies, I. (1988) Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, **41**, 909–996.

[DJ94] Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–455.

[DJ95] Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90**, 1200–1224.

[DJ98] Donoho, D. L. and Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.* **26**, 879–921.

[DJKP] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995) Wavelet shrinkage: asymptopia? (with discussion). *J. R. Statist. Soc. B*, **57**, 301–337.

[Gao] Gao, H.-Ye (1997) Wavelet shrinkage estimates for heteroscedastic regression models. Technical report. MathSoft, Inc. Seattle.

[GS] 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. *J. Fourier Anal. Appl.*, **3**, 457–474.

[HT] Hall, P. and Turlach, B. A. (1997) Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *Annals of Statistics*, **25**, 1912–925.

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

[JS] Johnstone, I.M. and Silverman, B.W. (1997) Wavelet threshold estimators for data with correlated noise. *J. R. Statist. Soc. B* **59**, 319–351.

[Koh] Kohler, M. (2000). Nonlinear orthogonal series estimates for random design regression. Technical Report, Department of Mathematics, Universität Stuttgart.

[KS] 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.

[NaSi] Nason, G.P. and Silverman, B. (1995) The stationary wavelet transform and some statistical applications. in: A. Anestis, G. Oppenheim (eds), "Wavelets and Statistics", *Lect. Notes Statist.*, **103**, 281–299. Springer-Verlag: New York.

[Neu] Neumann, M. H. (1996) Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *J. Time Ser. Anal.* **17**, 601–633.

[NS] Neumann, M. H. and Spokoiny, V. G. (1995). On the efficiency of wavelet estimators under arbitrary error distributions. *Math. Methods of Statist.* **4**, 137–166..

[NvS] Neumann, M. H. and von Sachs, R. (1995) Wavelet thresholding: Beyond the Gaussian i.i.d. situation. in: A. Antoniadis, G. Oppenheim (eds), ''Wavelets and Statistics", *Lect. Notes Statist.*, **103**, 301–329. Springer-Verlag: New York.

[PV] Pensky, M. and Vidakovic, B. (2000) On non-equally spaced wavelet regression. To appear in *Annals of Institute of Statistical Mathematics.*

[PS] Pólya,G. and Szegö, G. (1979) *Aufgaben und Lehrsätze aus der Analysis I.* Springer-Verlag, Berlin.

[vSMG] von Sachs, R. and MacGibbon, B. (2000) Nonparametric curve estimation by wavelet thresholding with locally stationary errors. *Scand. J. Statist.*, to appear.

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

[Sar] Sardy, S., Percival, D. B., Bruce, A. G., Gao, H.-Y. and Stuetzle, W. (1998) Wavelet shrinkage for unequally spaced data. Mathsoft research report 41, to appear in *Journal of Computational and Graphical Statistics.*

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