The Annals of Statistics
2000, Vol. 28, No. 1, 298–335

ADAPTIVE CONFIDENCE INTERVAL FOR POINTWISE CURVE ESTIMATION

BY DOMINIQUE PICARD AND KARINE TRIBOULEY

Université Paris VI et Université Paris VII, and Université Paris XI

We present a procedure associated with nonlinear wavelet methods that provides adaptive confidence intervals around $f(x_0)$, in either a white noise model or a regression setting. A suitable modification in the truncation rule for wavelets allows construction of confidence intervals that achieve optimal coverage accuracy up to a logarithmic factor. The procedure does not require knowledge of the regularity of the unknown function f; it is also efficient for functions with a low degree of regularity.

1. Introduction. A major advantage of non linear methods in curve estimation (wavelet thresholding, local bandwidth selection...) is their adaptivity to erratic fluctuations in the signal. They enjoy excellent mean squared error properties when they are used for estimating functions containing singularities of different forms (cups, chirps,...). They also have minimax convergence rates that are close to optimal for large function classes and large classes of norms. Of course, these minimax rates reflect the complexity of the estimated function. For instance, a very regular function (with, say, only a few discontinuities in its derivative) can be estimated at a higher rate than a function with a very deep singularity, such as a chirp. For details and discussion see, among others, Donoho, Johnstone, Kerkyacharian and Picard (1995, 1996), Hall and Patil (1995a, b), Hall, Kerkyacharian and Picard (1996), Lepski, Mammen and Spokoiny (1996) and Kerkyacharian, Picard and Tribouley (1996).

There is a well-developed theory for obtaining pointwise confidence intervals for an unknown function, based on kernel estimators with non-random bandwidths. This theory establishes the rates of convergence to nominal values for the coverage probabilities [see Hall (1991) and Hall (1992)]. More recently, confidence intervals have been constructed using kernel estimators with data-driven selected bandwidths [Neumann (1995); see also Fareway (1990) and Fareway and Jhun (1990)].

However, in both situations, there are serious practical limitations. First, a *high degree of regularity* is generally required for the unknown function. (Typically, the function must be three or four times differentiable.) Second, *knowledge of the degree of regularity* is generally needed for construction of the confidence intervals. Even for procedures using data-driven selected bandwidths, the construction generally requires knowledge of a lower bound on the regularity.

Received January 1998; revised May 2000.

AMS 1991 *subject classifications*. 62C20, 62G07, 62G15, 26G30.

Key words and phrases. Adaptive estimation, Confidence interval, Edgeworth expansion, Wavelet methods.

In this paper, our aim is to show that the new adaptive non-linear methods can be used to address these limitations in the construction of confidence intervals. We will primarily focus on the following two features:

Possible low regularity. The function may have singularities or be highly oscillating in small intervals.

Adaptation. The procedure should be able to adapt to the function without requiring knowledge of the regularity of the function, since this information is lacking in most practical situations.

We present results in the case of a white noise model and in the case of a regression model.

The first step in constructing confidence intervals is to obtain the asymptotic distribution of a pivotal quantity based on a nonparametric estimator. Thus our first goal will be to prove asymptotic normality of quantities based on adaptive estimators. For the sake of simplicity, we will mainly focus on thresholding wavelet estimators, although we will use a slight modification of these estimators, inspired by Lepski's method, that is more efficient in the context of confidence intervals. (Interestingly, the modification makes no difference to minimax properties.)

When constructing confidence intervals for the unknown quantity $f(x_0)$, a central issue is the problem of bias of the underlying estimator. There are two common methods for dealing with this bias: undersmoothing and explicit bias correction. In Hall (1991, 1992a) and Neumann (1992b), it is shown that, for a large variety of situations, undersmoothing leads to better coverage of the confidence intervals. We thus adopt undersmoothing as our bias correction method. Moreover, in the wavelet context, undersmoothing takes a very simple form: we simply add j-levels to the wavelet basis. Our confidence interval is thus constructed from a nonlinear estimator similar to thresholding estimators, but with a modified truncation rule. To choose this rule optimally we mimic, in an adaptive way, the results in Hall (1991) and Hall (1992) concerning theoretical bandwidths and Edgeworth corrections that yield coverage accuracy up to a fixed order. Of course, no correction is needed in the white noise setting because we can obtain the exact distribution of the pivotal statistic. For this model, we construct a confidence interval with a coverage accurate to within n^{-s} (up to a logarithmic factor), for any regularity $s > 0$ [s being roughly, the number of derivatives at $f(x_0)$]. In the regression setting, it is shown that the rates of convergence of the coverage accuracy are comparable to the optimal rates in Hall (1991) and Hall (1992) (up to a logarithmic factor). We are even able to reduce the assumptions on the regularity s: for instance, to obtain an accuracy up to third order, we only need to assume that $s \geq N/(1+2N)$, where N is a parameter depending on the wavelet system that is used.

Our method does require some assumptions (in addition to the usual regularity conditions), which can be interpreted in the following way. We do not need to know the regularity of f, but the precision of estimation is linked with the local regularity at x_0 and, somehow, this local regularity must be estimated. This problem has no solution without additional assumptions, unless extremely low rates of convergence are allowed; see Low (1997) and Iouditski and Lepski (1997). This phenomenon can be explained through the following example. There is no difficulty in estimating the local regularity of a function f that is locally given by $|x - x_0|^{1/2}$. However, suppose that the function is, instead, like f only at some particular points accumulating near x_0, but is very smooth elsewhere. If we have no access to these particular points in our observations, there is little hope of sharply estimating the true regularity. Hence, we need to assume that the observations contain enough information, relative to the complexity of the function. In signal or image processing, this assumption is, more or less, tacitly assumed, each time a procedure relying on estimating the local regularity is used [see Arneodo et al. (1997)].

We use a thresholding method to estimate $f(x_0)$. The procedure will usually depend on two crucial constants, κ and M. κ appears as a tuning constant for the thresholding; its choice is delicate and crucial in practice. The constant M is connected with the regularity of f: M is the radius of the ball of the space (Hölder, Sobolev, Besov,...) in which f is assumed to belong. The constant M is important in measuring the performance of the estimation method and bounding the bias; indeed, it may even be an a priori of the method (especially in the density model). Here, our procedure for giving confidence intervals imposes M to be equal to κ. Hence, a practitioner starts the thresholding method with the constant M and the regularity is automatically driven by M.

Although we restrict consideration to a white noise model and to a regression model, similar results can be obtained for any situation in which a central limit theorem and Edgeworth expansions can be established and in which large deviation inequalities for the wavelet coefficients can be evaluated. For instance, the same results can be achieved for estimating a density. As usual, in this case, the length of the confidence interval depends on the estimated value. Hence, there is then the additional task of providing a double estimation, for instance using resampling methods. This will be investigated in a later paper.

This paper is organized as follows. The white noise and the regression models are presented in Section 2. Section 3 discusses the thresholding procedures, bandwidth selection truncation, and optimal rules for undersmoothing. At the end of this section, the adaptive confidence intervals are introduced. Section 4 is devoted to the assumptions that are needed and to discussion of their consequences in terms of the construction of the confidence intervals. Section 5 states the main results. Theorem 1 develops an asymptotic expansion for the distribution of the pivotal statistic based on the thresholding estimators. Theorem 2 considers the construction based on the modified Lepski's estimator. Section 6 is devoted to the proofs. We first present the main tools: concentration inequalities and Edgeworth expansions. Then we prove Theorem 1 and Theorem 2, first for the white noise model where the technical difficulties are less. The Edgeworth expansions are primarily needed to handle the regression model. Standard proofs or technical computations are postponed to appendices. Appendix A deals with the proofs of the Edgeworth expansions; in Appendix B, we prove the technical properties and lemmas; Appendix C is devoted to the concentration inequalities.

2. Models. Let N be a fixed positive constant. Suppose that the scaling function, ϕ, and associated wavelet function, ψ, are compactly supported on $[-N, N]$, and that the q-th moments of the wavelet ψ vanish for $q = 1, \dots N$ [see, e.g., the Daubechies wavelets: Daubechies (1992)]. In the following and for any function g, we define g_{jk} to be the function $2^{j/2}g(2^j u - k)$.

For technical reasons, we need to split the sample into two independent parts. This is easy to do in the regression framework. In the white noise setting, we use the wavelet decomposition to perform this splitting. Throughout the paper, we use the following notations: P_1 and E_1 (respectively, P_2 and E_2) denote probability and expectation with respect to the first part of the sample (respectively, the second part).

2.1. *White noise model.* Consider the model

$$dY_t = f(t)dt + (2n)^{-1/2}dW_t, \;\; t \in [0, 1],$$

where W is Brownian motion, f is the function we want to estimate and n tends to infinity. It is usual to transform the continuous white noise model into the discrete model with coefficients $\hat{\beta}_{jk}$, using the cascade algorithm. Indeed, letting $J = \log_2(2n)$ and $j_0 > 0$, we first construct

$$\hat{\alpha}_{Jk} = \int \phi_{Jk}(t)dY_t.$$

We then can calculate the $\hat{\alpha}_{jk}, \hat{\beta}_{jk}$ for $j_0 \leq j \leq J-1$ by using the cascade Daubechies formulae [Daubechies (1992)]. We divide the $\{\alpha_{Jk}\}_k$ into two parts by considering the odd and even k's. This splitting leads to 2 independent samples $\{\hat{\beta}_{jk}(1),\; \hat{\alpha}_{j_0k'}(1)\}_{k,k',j=j_0,\dots,J-1}$ and $\{\hat{\beta}_{jk}(2),\; \hat{\alpha}_{j_0k'}(2)\}_{k,k',j=j_0,\dots J-1}$. In each case, we have for $l = 1, 2$

$$\begin{cases} \hat{\beta}_{jk}(l) \; = \beta_{jk}(l) + n^{-1/2}\, \varepsilon_{jk}(l), & j = j_0, \dots J-1, \;\; k \in \mathscr{Z}, \\ \hat{\alpha}_{j_0k}(l) = \alpha_{j_0k}(l) + n^{-1/2}\, \eta_{j_0k}(l), & k \in \mathscr{Z}, \end{cases}$$

where $\{\varepsilon_{jk}, \eta_{j_0k'}\}_{j=j_0,\dots,J-1,k,k'}$ are independent Gaussian variables with zero mean and variance equal to σ_2^2.

2.2. *Regression model.* Consider the regression model

$$X_i = f(i/2n) + \varepsilon_i, \quad i = 1, \dots, 2n,$$

where f is again the function to be estimated and where the ε's are independent zero mean random variables with characteristic function χ in L^1, variance equal to σ_2^2 and qth moment equal to σ_q. In addition, assume that one of the following conditions is satisfied:

1. For all $q \geq 2$, the qth moment of the ε's is finite and $\exists \mu > 0$ such that

$$E|\varepsilon|^q \leq q!\, \sigma_2^2 \mu^{q-2}/2.$$

2. The ε's are either bounded by $|\varepsilon|_\infty$ or Gaussian.

We need the first condition in the case of local thresholding and the second condition in the case of block thresholding (see the proof of the concentration inequalities, Appendix C).

As before, we can again split the sample into two independent parts $\{X_1, X_3, \dots, X_{2n-1}\}$ and $\{X_2, X_4, \dots, X_{2n}\}$. As usual, we define the empirical wavelet coefficients, for $l = 1, 2$, by

$$(1)\quad \hat{\alpha}_{jk}(l) = \frac{1}{n}\sum_{i=0}^{n-1} X_{2i+l}\phi_{jk}\left(\frac{2i+l}{2n}\right), \qquad \hat{\beta}_{jk}(l) = \frac{1}{n}\sum_{i=0}^{n-1} X_{2i+l}\psi_{jk}\left(\frac{2i+l}{2n}\right).$$

3. Confidence intervals. As discussed in the introduction, the construction of adaptive confidence intervals will proceed in several steps. First, we present the adaptive estimators of $f(x_0)$ that will be used to develop an asymptotically normal pivotal quantity. Second, we construct adaptive undersmoothing statistics. Finally, we combine these to yield confidence intervals for $f(x_0)$ and give their asymptotic coverage properties.

3.1. *Estimators.* In the first subsection, we briefly introduce thresholding methods. In the second subsection, we provide a new estimator, which, although directly connected to wavelet thresholding, is also very much inspired by Lepski's method of adaptation [see Lepski, Mammen and Spokoiny (1994)].

3.1.1. *Wavelet thresholding.* The idea of thresholding is to keep only the more meaningful empirical wavelet coefficients, setting the others to zero. Let $j_\infty = \log_2(n/\log n)$, $j_0 = (1+2N)^{-1}\log_2(n)$ and consider the empirical wavelet coefficients defined in (1) for the regression model or in (1) for the white noise model. Define the "local thresholding" estimate[Donoho, Johnstone, Kerkyacharian, Picard (1996)]

$$(2)\quad \hat{f}(x_0) = \sum_k \hat{\alpha}_{j_0k}(1)\phi_{j_0k}(x_0) + \sum_{j=j_0}^{j_\infty}\sum_k \hat{\beta}_{jk}(1)1_{\left\{\hat{\beta}_{jk}(1) > M\sqrt{\log n/n}\right\}}\psi_{jk}(x_0)$$

for some $M > 0$. In the double index (j, k), the level index j plays a special role very similar to a bandwidth. Indeed, local thresholding has an action quite comparable to selecting a local bandwidth. However, in the context of confidence intervals, local bandwidth selection has the drawback of being unstable and, as mentioned in Bowman and Härdle (1988), its advantage is rather unclear. For these reasons, we also investigate selecting procedures acting on "blocks" of $\hat{\beta}_{jk}$, covering both global and local choice.

The following description of the family of block thresholding estimators comes from Hall, Kerkyacharian and Picard (1996). For all indices $j \in \{j_0, \dots, j_\infty\}$, divide the set of integers between $-N$ and $2^j + N$ into consecutive, nonoverlapping blocks of length l_j (non decreasing in j), say

$$\mathscr{B}_{jK} = \{k : (K-1)\, l_j + 1 \le k \le K l_j\}, \qquad -\infty < K < \infty.$$

Fix $p \geq 2$. Let $\sum_{(K)}$ denote summation over $k \in \mathscr{B}_{jK}$ and define

$$B_{jK} = \sum_{(K)} |\beta_{jk}|^p .$$

This quantity can be estimated by

$$\hat{B}_{jK}(l) = \sum_{(K)} |\hat{\beta}_{jk}(l)|^p, \qquad l = 1, 2,$$

leading to the following estimator of $f(x_0)$:

$$\hat{f}^T(x_0) = \sum_k \hat{\alpha}_{j_0 k}(1)\phi_{j_0 k}(x_0) + \sum_{j=j_0}^{j_\infty} \sum_K \left(\sum_{(K)} \hat{\beta}_{jk}(1)\psi_{jk}(x_0) \right) 1_{\{\hat{B}_{jK}(1) > t_{j,n}\}},$$

for some threshold $t_{j,n} > 0$. In the sequel, we only consider the following particular choices:

1. $l_j = 2^j + 2N, t_{j,n} = M^p l_j \; n^{-p/2}, M > 0$ provides the global thresholding estimator [Kerkyacharian, Picard and Tribouley (1996)]. In this case, $\mathscr{B}_{jK} = \{-2N - 2^j, \dots, 2^j\}$ is the set of all the indices k of the level j. The β_{jk} for each level j are either estimated or thresholded, in a global way. This selection is made by evaluating the global l_p-energy of the level j.
2. Allowing blocks of differing lengths l_j, where $\log(n)^{p/2} \leq l_j \leq 2^j + 2N$, $j = j_0, \dots, j_\infty$ (and still using the threshold $t_{j,n} = M^p l_j \; n^{-p/2}$, $M > 0$) results in the block thresholding estimators [Hall, Kerkyacharian and Picard (1996)]. Smaller block sizes result in more local selection.
3. Finally, the local thresholding estimator $\hat{f}$ [defined in (2)], also belongs to this family with the particular choices $l_j = 1$ and $t_{j,n} = M^p(n^{-1} \; \log n)^{p/2}$, $M > 0$. In this case, $\mathscr{B}_{jK}$ is the single point $\{k\}$. Here p is clearly not a meaningful quantity; thus we take, for instance, $p = 2$.

3.1.2. *Local bandwidth selection.* We introduce the following index of local complexity of the estimated function:

$$\begin{aligned} \hat{j}_1 &= \hat{j}_1(x_0) \\ &= \sup \left\{ j \in \{j_0, \dots, j_\infty\},\; \exists K,\; \sum_{(K)} |\hat{\beta}_{jk}(2)|^p 1_{\{|\psi(2^j x_0 - k)| > m\}} \geq 2^{-2p} t_{j,n} \right\}, \end{aligned} \tag{3}$$

where m is related to an assumption concerning the point x_0 [see Condition $F(x_0)$ in section 4.2]. Set $\hat{j}_1 = j_0$ if, for all indices j and K, the empirical quantity $\sum_{(K)} |\hat{\beta}_{jk}(2)|^p 1_{\{|\psi(2^j x_0 - k)| > m\}}$ is less than $2^{-2p} t_{j,n}$.

The smoother the function to be estimated, the smaller $\hat{j}_1$ is expected to be. While $\hat{j}_1$ is generally a local quantity, its local aspect decreases when the size of the blocks increases. Indeed, $\hat{j}_1$ does not depend on x_0 in the case of global thresholding.

Moreover, as proven in Lepski, Mammen and Spokoiny (1994) for L_p-loss, such an index plays the same role as the optimal bandwidth, up to some logarithmic factor. This leads to the following estimator of $f(x_0)$:

$$\hat{f}^L(x_0) = \sum_k \hat{\alpha}_{j_0k}(1)\phi_{j_0k}(x_0) + \sum_{j=j_0}^{\hat{j}_1} \sum_k \hat{\beta}_{jk}(1)\psi_{jk}(x_0).$$

At first glance (and also for minimax results), this estimator is not very different from the thresholding estimator. The only difference is in the coefficients between the levels j_0 and $\hat{j}_1$, which have a contribution here even if they are below the threshold. We will see that, in the context of confidence intervals, this difference is crucial.

3.2. *Undersmoothing.* We first construct the undersmoothing statistics in a non-adaptive way (i.e., for known regularity of the function) and then explain how to slightly modify the procedure to obtain adaptation.

3.2.1. *Correction when s is known.* In the case of kernel methods, Hall (1992) and Hall (1991) show that the best coverage accuracy to second order (respectively, to third order) is obtained by choosing a bandwidth of order $h_n \sim n^{-1/1+s}$ (respectively, $h_n \sim n^{-1/1+2s/3}$). In the wavelet context, because of the orthogonal projection properties, undersmoothing is equivalent to adding j-levels. More precisely, instead of dealing with $\hat{f}^T(x_0)$ or $\hat{f}^L(x_0)$, we consider $\hat{f}^T(x_0) + \hat{B}(x_0)$ or $\hat{f}^L(x_0) + \hat{B}(x_0)$ with

$$\hat{B}(x_0) = \sum_{\hat{j}_1}^{j'} \sum_k \hat{\beta}_{jk}(1)\psi_{jk}(x_0),$$

where j' is an index to be determined. We focus on the following family of indices $\{j_{\eta s}\}_{\eta \geq 1}$:

$$\begin{aligned}
&\eta = 1, \quad 2^{j_s(n)} \leq \left(\frac{2^8 n}{\log n}\right)^{1/(1+2s)} \leq 2.2^{j_s(n)} \quad \text{for local thresholding,}\\
&\eta = 1, \quad 2^{j_s(n)} \leq \left(2^8 n\right)^{1/(1+2s)} \leq 2.2^{j_s(n)}\\
(4) \qquad & \qquad\qquad\qquad\qquad \text{for block or global thresholding,}\\
&\eta \in \,]1, \infty[\,, \quad 2^{j_{\eta s}(n)} \leq n^{\frac{1}{1+2s/\eta}} \leq 2.2^{j_{\eta s}(n)},\\
&\eta = \infty, \quad 2^{j_\infty(n)} \leq \frac{n}{\log n} \leq 2.2^{j_\infty(n)}.
\end{aligned}$$

In the sequel, we omit the dependence on n of the indices j. From Hall, we can obtain good candidates for j', corresponding to $\eta = 2$ or $\eta = 3$ in the previous family. But these give non-adaptive solutions to the problem.

3.2.2. *Correction when s is unknown.* Our goal is to provide "estimators" of the indices defined in (4), in such a way that they do not depend on the unknown regularity s. Suppose that $\hat{j}_1$, defined in (3), is a "good estimator" of the minimax optimal bandwidth j_s, which corresponds to $\eta = 1$, calculated using the second part of the sample. (We explain, in the next section, what we mean by a "good estimator.") Then it is easy to derive a "good estimator" of $j_{\eta s}$ as follows. First compute

$$\hat{s} = \hat{s}(x_0) = \frac{1}{2}\left(\frac{\log_2 n}{\hat{j}_1} - 1\right),$$

which is an approximation to the local regularity s. Then plug the value $\hat{s}$ into $j_{\eta s}$:

$$\frac{1}{1+2\hat{s}/\eta}\log_2 n - 1 \leq \hat{j}_\eta \leq \frac{1}{1+2\hat{s}/\eta}\log_2 n. \tag{5}$$

Finally, we propose as candidates for the adaptive undersmoothing statistics

$$\forall \eta \geq 1, \quad \hat{B}_\eta(x_0) = \sum_{\hat{j}_1}^{\hat{j}_\eta}\sum_k \hat{\beta}_{jk}(1)\psi_{jk}(x_0).$$

3.3. *Confidence intervals.* Let u_α be the αth quantile of the standard Gaussian distribution. Denote, by $|s_.|_q$, the l_q norm ($q \geq 1$) of the sequence $\{s_l\}_l$. We introduce the following quantities:

(i) In the white noise case,

$$b(n) = \left[\frac{1}{n}\sum_{j_0}^{\hat{j}_\eta}\sum_k \psi_{jk}^2(x_0) + \frac{1}{n}\sum_k \phi_{j_0k}^2(x_0)\right]^{-1/2}; \tag{6}$$

(ii) in the regression model,

$$b(n) = \sigma_2^{-1}|s_.|_2^{-1}, \tag{7}$$

where

$$\begin{aligned} s_{2i+1} &= \frac{1}{n}\sum_{j_0}^{\hat{j}_\eta}\sum_k \psi_{jk}(x_0)\psi_{jk}\left(\frac{2i+1}{2n}\right) + \frac{1}{n}\sum_k \phi_{j_0k}(x_0)\phi_{jk}\left(\frac{2i+1}{2n}\right); \\ |s_.|_q^q &= \sum_{i=0}^{n-1}|s_{2i+1}|^q; \quad |s_.|_\infty = \sup_{i=0,\dots,n-1}|s_{2i+1}|. \end{aligned} \tag{8}$$

We also need to consider correction statistics, T_{qn}, that are needed to determine the accuracy of the confidence intervals proposed in the sequel:

$$\forall q \geq 2, \quad T_{qn} = b(n)^q|s_.|_q^q, \qquad T_{\infty n} = b(n)|s.|_\infty. \tag{9}$$

These quantities are random variables depending on the second part of the sample (because of the way we defined $\hat{j}_\eta$). Their orders of magnitude are evaluated in Section 6.1.2.

We introduce the following confidence intervals at level α:

$$I_\eta^2 = \left[\hat{f}^T(x_0) + \hat{B}_\eta(x_0) - u_{1+\alpha/2} b(n)^{-1}, \hat{f}^T(x_0) + \hat{B}_\eta(x_0) + u_{1+\alpha/2} b(n)^{-1}\right],$$

$$J_\eta^2 = \left[\hat{f}^L(x_0) + \hat{B}_\eta(x_0) - u_{1+\alpha/2} b(n)^{-1}, \hat{f}^L(x_0) + \hat{B}_\eta(x_0) + u_{1+\alpha/2} b(n)^{-1}\right],$$

$$J_\eta^3 = \Big[\hat{f}^L(x_0) + \hat{B}_\eta(x_0) + b(n)^{-1} \left(-u_{1+\alpha/2} + (u_{1+\alpha/2}^2 - 1)\frac{\sigma_3}{3!} T_{3n}\right),$$
$$\hat{f}^L(x_0) + \hat{B}_\eta(x_0) + b(n)^{-1} \left(u_{1+\alpha/2} + (u_{1+\alpha/2}^2 - 1)\frac{\sigma_3}{3!} T_{3n}\right)\Big].$$

Our aim is to prove (under appropriate assumptions on the regularity of f) the following properties about the asymptotic coverage of these confidence intervals. Here, γ_i, $i = 1, 2, 3$, are positive constants that depend on the regularity of f and will be defined later.

1. In both models, the accuracy of the interval associated to $\hat{f}^L$ is always better than that associated to $\hat{f}^T$.
2. In the white noise model, the interval J_η^2 is optimum (in the sense of minimum coverage) for $\eta = \infty$ and the coverage error is given by

$$\left|P(f(x_0) \in J_\infty^2) - \alpha\right| \le C\, n^{-s} (\log n)^{\gamma_1}.$$

3. In the regression model, the interval J_η^2 is optimum (in the sense of second order best coverage accuracy) for $\eta = 2$, and the coverage error is given by

$$\left|P(f(x_0) \in J_2^2) - \alpha\right| \le C\, n^{-\frac{s/2}{1+s}} (\log n)^{\gamma_2}.$$

   The interval J_η^3 is optimum (in the sense of the third order best coverage accuracy) for $\eta = 3$, and the coverage error is given by

$$\left|P(f(x_0) \in J_3^3) - \alpha\right| \le C\, n^{-\frac{2s/3}{1+2s/3}} (\log n)^{\gamma_3}.$$

If we compare these results with those in Hall (1991), Hall (1992) and Neumann (1995), we find that the cost for adaptation is only a logarithmic factor. This is similar to what happens in the minimax setting.

4. Regularity assumptions and results about the estimators $\hat{j}_\eta$.

4.1. *Functional space.* We assume that the function to be estimated has a certain degree of smoothness. Because we focus on adaptation, this smoothness does not affect the construction of the procedure but is, of course, important in the evaluation of procedure.

For $s \in R_+^* - N$ and $M > 0$, we consider the following space of functions:

$$L_s(M) = \{f : [0,1] \to R;\quad \forall (x,y) \in [0,1]^2,$$
$$|f^{([s])}(x) - f^{([s])}(y)| \le (M/c)|x-y|^\alpha, \tag{10}$$
$$\text{where } s = [s] + \alpha,\ 1 \ge \alpha > 0\}.$$

The positive constant c is defined by the following remark: it is known that, if the number of vanishing moments of the wavelet ψ is N, then, for $s < N+1$, the regularity can be determined from the wavelets coefficients. More precisely, there exists a constant c, depending only on ψ, such that, if g has the expansion

$$g = \sum_k \alpha_{0k}\phi_{0k} + \sum_{j=0}^{\infty}\sum_k \beta_{jk}\psi_{jk},$$

then $g \in L_s(M)$ implies

$$\sup_{0\le j}\left(2^{j(s+1/2)}\sup_k|\beta_{jk}|\right) \le c(M/c) = M \tag{11}$$

[see, e.g., Härdle, Kerkyacharian, Picard and Tsybakov (1998), Chapter 8 and notice that the converse is also true].

For $s \in N^*$, we define our functional space $L_s(M)$ by the condition (11) and notice that, in this case, the ball does not exactly coincide with (10), although the difference is small.

4.2. *Assumptions on the point* x_0. The following condition, $F(x_0)$, is very easy to verify in practice. In fact, it usually holds and, if not, a slight change in x_0 will usually fix the problem.

CONDITION $F(x_0)$. $\exists\, m > 0$ such that

$$\inf_{j_0\le j\le j_1}\sup_{k,i}\left|\psi(2^j x_0 + k)\psi\left(2^j\frac{i}{n}+k\right)\right| > m||\psi||_\infty,$$

$$\sup_k|\phi(2^{j_0}x_0+k)| > m.$$

4.3. *Key assumption on the function* f. It is a well known fact that it is impossible to estimate the smoothness of a function or to obtain adaptive confidence bands of reasonable length under the usual regularity conditions [see Low (1997) and Iouditski and Lepski (1997)]. We thus need to add an assumption; this assumption is the most important restriction imposed in the paper, since the following sequence, ρ_n, precisely measures the loss in adaptation.

CONDITION $H_s(M, x_0)$. There exists some $\rho_n > 0$ such that, for all $n \ge 2$,

$$\exists j^*,\quad j_s - \rho_n \le j^* \le j_s, \exists K,\ \sum_{(K)}|\beta_{j^*k}|^p 1_{\{|\psi(2^{j^*}x_0+k)|>m\}} \ge 2^{3p}M^p 2^{-j_s(s+1/2)p} l_{j^*},$$

where m is defined by Condition $F(x_0)$ and j_s is defined in (4).

The additional assumptions considered here can be interpreted as ensuring that the data contain enough information to estimate the regularity of the function. These assumptions are, more or less, tacitly made each time estimation of local regularity is needed. This is the case in the signal processing community when analyzing, reproducing or coding highly complex signals, in turbulence, seismology, voice reconstruction, etc. [see Arneodo, Jaffard, Levy-Vehel and Meyer (1997)]. It can also be seen as a rough self-similarity property on f in the sense that, if it happens that the β's are large around one point, then they are similarly large at any level. Note that condition $H_s(M, x_0)$ can easily be proved with $s = N$, if $f^{(N+1)}(x_0) \neq 0$ and $f^{(N+1)}$ is continuous at x_0.

Another way of visualizing this condition is by quantifying the idea of self-similarity of the wavelet coefficients. For instance, suppose that, for any j, k,

$$M2^{-j(s+1/2)}\omega(j) \leq |\beta_{jk}| \leq M2^{-j(s+1/2)}.$$

Then, for $\omega(j) = j^{-\beta}$, any $\rho_n > 2^3\beta/(s+1/2)\log\log n$ is suitable.

4.4. *Properties of the indices $j_{\eta s}$ and $\hat{j}_\eta$.* The following property is very easy to obtain from definition (4) and inequality (11).

PROPERTY 1.

$$f \in L_s(M) \Rightarrow \forall j \geq j_s \qquad \forall K,\ B_{jK} \leq 2^{-4p} t_{j,n}.$$

Let us now consider the impact of the previous hypotheses on the behavior of the index $\hat{j}_\eta$ defined in (5). First, observe that $\hat{j}_1$ [defined in (3)] is a function of the second sample; this implies that $\hat{\alpha}$ or $\hat{\beta}$ and $\hat{j}_\eta$, for all $\eta \geq 1$, are independent variables. Moreover, $\hat{j}_\eta$ is a good estimator of $j_{\eta s}$, in the sense of the following Properties 2 and 3. In these properties, the positive function $\gamma(C_0, M)$ is defined in Section 6 and the positive constant $\tilde{R}(p)$ is defined in Appendix C.

PROPERTY 2. *Assume that $f \in L_s(M)$. Under the constraints* (*for block thresholding*) *$M > 8\sigma_p$ for the white noise model and $M > 8\tilde{R}(p)$ for the regression model, there exists some constant $C > 0$ such that, for all n,*

$$P(\hat{j}_\eta > j_{\eta s}) \leq Cn^{-\gamma(1/8, M)}.$$

PROPERTY 3. *Assume that $f \in L_s(M)$ and that the condition $H_s(M, x_0)$ is satisfied for some positive sequence ρ_n. Under the constraints* (*for block thresholding*) *$M > 4\sigma_p$ for the white noise model and $M > 4\tilde{R}(p)$ for the regression model, there exists some constant $C > 0$ such that, for all n,*

$$P(\hat{j}_\eta + (1 + 2N)\,\rho_n < j_{\eta s}) \leq Cn^{-(\gamma(1/4, M)\wedge\gamma(1/8, M))}.$$

The previous large deviation inequalities on $\hat{j}_\eta$ are direct consequences of our assumptions on f; they are proved in the first part of Appendix B.

5. Main results. In this section, we denote by Φ (respectively, ϕ) the distribution function (respectively, the density) of the standard Gaussian distribution.

THEOREM 1. *Assume that f belongs to $L_s(M)$, for some M larger than some absolute constant, and that the conditions $F(x_0)$ and $H_s(M, x_0)$ are satisfied for some positive sequence ρ_n. Then, for all $\delta_n \geq (2N+1)\rho_n$, we can conclude that*:

(i) *in the white noise model*,

$$\forall t,\ P\Big(b(n)\Big[\hat{f}^T(x_0) + \hat{B}_\eta(x_0) - f(x_0)\Big] < t\Big) = \Phi(t) + O(u_n + v_n);$$

(ii) *in the regression model*,

$$\forall t,\ P_1\Big(b(n)\Big[\hat{f}^T(x_0) + \hat{B}_\eta(x_0) - f(x_0)\Big] < t\Big) = \Phi(t) + O(U_n + V_n + W_n(2)),$$

where

$$E_2|U_n| = u_n = \Big(n2^{-j_{\eta s}(2s+1)}2^{\delta_n(2s+1)}\Big)^{1/2},$$

$$E_2|V_n| = v_n = \Big(n2^{j_s - j_{\eta s}}t_{j_s,n}^{2/p}2^{\delta_n}\Big)^{1/2},$$

$$E_2|W_n(2)| = w_n = \left(\frac{2^{j_{\eta s}}}{n}\right)^{1/2} 2^{3\delta_n/2}.$$

THEOREM 2. *Assume that f belongs to $L_s(M)$, for some M larger than some absolute constant, and that the conditions $F(x_0)$ and $H_s(M, x_0)$ are satisfied for some positive sequence ρ_n. Then, for all $\delta_n \geq (2N+1)\rho_n$, we can conclude that*:

(i) *in the white noise model*,

$$\forall t,\ P\Big(b(n)\Big[\hat{f}^L(x_0) + \hat{B}_\eta(x_0) - f(x_0)\Big] < t\Big) = \Phi(t) + O(u_n);$$

(ii) *in the regression model, for $r \geq 2$*,

$$\begin{aligned}\forall t,\quad & P_1\Big(b(n)\Big[\hat{f}^L(x_0) + \hat{B}_\eta(x_0) - f(x_0)\Big] \leq t\Big)\\ &= \Phi(t) + \phi(t)\sum_{k=1}^{r}\left[\sum_{(I_1,\ldots,I_p)\in\mathscr{I}_k} T_{i_1 n}\cdots T_{i_p n}Q_{i_1\ldots,i_p}(t)\right] + O(U_n + W_n(r)),\end{aligned}$$

where:

(a) *the set of indices is defined by*

$$\mathscr{I}_k = \{(i_1,\ldots,i_p),\ p \geq 1, (i_1-2)+\cdots+(i_p-2) = k, 3 \leq i_1 \leq \cdots \leq i_p \leq r\};$$

(b) *the T_{qn}'s are defined in* (9) *and satisfy*

$$\forall (i_1,\ldots i_p) \in \mathscr{I}_k,\ E_2 T_{i_1 n}\cdots T_{i_p n} \leq C\left(\frac{2^{j_{\eta s}}}{n}\right)^{k/2} 2^{\delta_n(k/2+p)};$$

(c) *the $Q_{i_1,\ldots,i_p}$ are polynomials defined later depending only on $\sigma_3,\ldots,\sigma_{r+1}$ and built from integrated Hermite polynomials*;

(d) *in particular, the first term of the expansion is* $T_{3n}(\sigma_3/3!)(1-t^2)$;
(e) *the statistics* U_n *and* $W_n(r)$ *satisfy*

$$E_2|U_n| \le \left(n2^{-j_{\eta s}(2s+1)}2^{\delta_n(2s+1)}\right)^{1/2},$$

$$E_2|W_n(r)| \le \left(\frac{2^{j_{\eta s}}}{n}\right)^{\frac{r-1}{2}} 2^{\delta_n \frac{3(r-1)}{2}} := w_n(r).$$

REMARKS. Theorem 1 (compared to Theorem 2) explains the differences between the thresholding estimator $\hat{f}^T$ and its modification $\hat{f}^L$, inspired by Lepski's adaptation method. If we omit the factor δ_n which we discuss later, the terms u_n and v_n are determining the respective behaviors of the first order coverage accuracy of the confidence intervals built on $\hat{f}^T$ and $\hat{f}^L$, with a bias correction adjusted by the extra parameter η. The relation

$$v_n = u_n n^{\frac{4s^2(1-1/\eta)}{1+2s/\eta}(1+2s)} h_n,$$

where $h_n = l_{j_s}$ (or $\log n$, depending on whether we use block thresholding or local thresholding), immediately shows that the thresholding estimator still has bias that needs to be corrected. This part of the bias is an increasing function of the length of the blocks. One could think of performing a new correction for $\hat{f}^T$ but, if we look at Theorem 2, it is obvious that f^L performs an automatic correction of this part of the bias.

At this stage, without performing any additional correction to improve the coverage accuracy, the previous theorems provide confidence intervals around $f(x_0)$ of length $b(n)^{-1}$. The magnitude of $b(n)$ is directly computable but, if we are concerned with the rates of convergence in various cases, one must evaluate the asymptotic order of $b(n)$. Considering, for instance, the Haar wavelet (or extrapolating from lemma 1) we have

$$b(n) = \left(\sigma_2^2 \frac{2^{\hat{j}_{\eta s}}}{n}\right)^{-1/2}.$$

Because $j_{\eta s} - \delta_n \le \hat{j}_\eta \le j_{\eta s}$ with high probability, the length of the confidence interval is between $n^{1/(1+s/\eta)}$ and $n^{1/(1+s/\eta)}2^{\delta_n}$.

It is also interesting to consider the asymptotic behavior of $\hat{f}^T$ or $\hat{f}^L$, without any bias correction (or undersmoothing). This means considering the case where $\eta = 1$. Then $\lim_n u_n = \infty$. This shows that, even to establish a simple "central limit theorem" around the adaptive estimators, we need to undersmooth. This is also a consequence of the usual phenomena that the rates of convergence of the nonlinear thresholding methods are governed by their bias.

In the white noise model, the interval J_η^2 is optimum (in the sense of minimum coverage) for $\eta = \infty$ and the coverage error is

$$\left|P(f(x_0) \in J_\infty^2) - \alpha\right| \le C\, n^{-s}(2^{\delta_n} \log n)^{s+1/2}.$$

Obviously, no correction is needed to obtain optimal coverage accuracy.

In the regression model, applying Theorem 2, with $r = 2$, yields

$$\begin{aligned} P\Big(b(n)\Big[\hat f^L(x_0) + \hat B_\eta(x_0) - f(x_0)\Big] \le u_{1+\alpha/2}\Big) \\ = E_2 P_1\Big(b(n)\Big[\hat f^L(x_0) + \hat B_\eta(x_0) - f(x_0)\Big] \le u_{1+\alpha/2}\Big) \\ = \frac{1+\alpha}{2} + O\left(u_n + w_n(2)\right). \end{aligned}$$

The best coverage accuracy is obtained for $\eta = 2$ [when u_n and $w_n(2)$ are roughly of the same order]. The best confidence interval of second order is J_2^2, and its coverage error is given by

$$\left|P(f(x_0) \in J_2^2) - \alpha\right| \le C\, n^{-s/2/(1+s)} 2^{\delta_n((s+1/2)\vee 3/2)}.$$

Applying Theorem 2 with $r = 3$, we obtain

$$\begin{aligned} P_1\Big(b(n)\Big[\hat f^L(x_0) + \hat B_\eta(x_0) - f(x_0)\Big] \le u_{1+\alpha/2}\Big) \\ = \frac{1+\alpha}{2} + \phi(u_{1+\alpha/2})(1 - u_{1+\alpha/2}^2) T_{3n}\frac{\sigma_3}{6} + O\left(U_n + W_n(3)\right). \end{aligned}$$

The best coverage accuracy is again obtained when u_n and $w_n(3)$ are of the same order, that is, for $\eta = 3$. Inverting the expansion above, we prove that J_3^3 is optimal at the third order. The coverage error is then given by

$$\left|P(f(x_0) \in J_3^3) - \alpha\right| \le C\, n^{-\frac{2s/3}{1+2s/3}} 2^{\delta_n((s+1/2)\vee 3)}.$$

More generally, by inverting the expansion given in Theorem 2 for $r = m$, it is possible to obtain a confidence interval constructed from the statistics $b(n), T_{3n}, \dots, T_{mn}$. We have to choose $\eta = m$ to obtain the best interval of order m. The coverage error is given by

$$\left|P(f(x_0) \in J_m^m) - \alpha\right| \le C\, n^{\frac{-s(m-1/m)}{1+2s/m}} 2^{\delta_n((s+1/2)\vee(3/2)(m-1))}.$$

If we compare these results with those in Hall (1991) and Hall (1992) in a non-adaptive framework and Neumann (1995), we find that the loss for adaptation depends on the regularity condition $H_s(M, x_0)$. If $\rho_n \sim \log\log n$, the loss for adaptation is only a logarithmic factor. This result is comparable with what is known in the minimax setting.

It is also worth noting that this loss does not depend on the type of thresholding (global, block or local) that was used. This corroborates the rather common opinion that, for confidence intervals, the potential advantages of local bandwidth selection are rather unclear [see Bowman and Härdle (1988)]. It also seems to be linked with the results in the minimax setting, when the loss is measured in the "sup"-norm. In this case, also, all the types of thresholding are equivalent.

However, it is worthwhile to note that, at a finer scale, there is an advantage to local bandwidth selection. Indeed, we can refine our results and see that the condition imposed on f, to be in $L_s(M)$, need only be true for the pairs

j, k such that $\sum_{k\in(K)} I\{\psi_{jk}(x_0) \neq 0\} > 0$. Hence, the more local the procedure is, the sharper the regularity condition is.

Let us now detail what we mean by "M large enough." This condition is obtained by collecting the conditions on the function γ throughout the proofs. It is not difficult to state them in generality (for r and η) but, for the sake of simplicity, we will focus on the most interesting practical case of the optimal confidence interval of third order (regression model, Theorem 2, $\eta = r = 3$). We obtain

$$\gamma(1/8, M) > \frac{16}{1+3/(2s)} \vee \frac{6}{1+2s/3} \vee \frac{s+1/2}{1+2s/3},$$
$$\gamma(1/4, M) > \frac{6}{1+3/(2s)} \vee \frac{6}{1+2s/3} \vee \frac{s+1/2}{1+2s/3}$$

and, in the case of block thresholding,

$$M > 8\tilde{R}(p).$$

Because of the definition of the function γ (see Appendix C), we obtain the sufficient (but not optimal) following conditions on M:

1. In the case of the local thresholding,

$$M \geq 64||\psi||_\infty \left(8\mu \vee \sqrt{2N}\sigma_2\right).$$

2. In the case of block thresholding with Gaussian error,

$$M \geq 8||\psi||\infty(4\sqrt{2}\sigma_2 + \sigma_p);$$

3. in the case of block thresholding with bounded error,

$$M \geq 8\left(4\sigma_2||\psi||_\infty T^{-1/2} + \tilde{R}(p)\right) \vee 8\left(16|\varepsilon|_\infty ||\psi||_\infty T^{-1} + \tilde{R}(p)\right),$$

where σ_2, σ_p, μ are defined in Section 2.2, N is the number of vanishing moments of the wavelet, T is an universal constant (see the Talagrand theorem, Appendix C.2.2.), $\tilde{R}(p)$ is function of the Rosenthal constant $R(p)$ [see the Rosenthal inequality, Appendix C.2.2.; for instance, $R(2) = 1/2$]:

$$\tilde{R}(p) = 2^{1/p}(2N)^{1/2}||\psi||_\infty|\varepsilon|_\infty R(p). \tag{12}$$

6. Proofs of theorems.

6.1. *Preliminaries.*

6.1.1. *Large deviation inequalities.* The following inequality is a consequence of different concentration or large deviation inequalities. The proof is postponed to Appendix C.

PROPOSITION 1. *Let C_0, $M > 0$ be fixed such that, in the case of the block thresholding, $C_0M > \sigma_p$ for the white noise model and $C_0M > \tilde{R}(p)$ for the* *regression model [see the definition of $\tilde{R}(p)$ in (12)]. There exists $C > 0$ such that for all indices j, n and blocks K*

$$P\left(\sum_{(K)} |\hat{\beta}_{jk} - \beta_{jk}|^p > C_0^p t_{j,n}\right) \leq C n^{-\gamma(C_0, M)}$$

where $\gamma(C_0, M)$ is a positive function which properties are detailed in Appendix C.

6.1.2. *Interval lengths and other meaningful rates.* We recall that the crucial quantities for the interval length are $b(n)$ [defined in (6) or in (7)] and the sequence of statistics $\{s_l\}$ in (8) (in the case of the regression model). Combining the following lemmas, it is easy to deduce the order of magnitude of the correction statistics T_{qn} [defined in (9)] via the Cauchy-Schwarz inequality. The proofs of Lemma 1 and Lemma 2 are given in Appendix B.

LEMMA 1. *We assume that f belongs to $L_s(M)$ and that $H_s(M, x_0)$ and $F(x_0)$ are satisfied for some sequence ρ_n. Let $q \geq 1$. Then, for all $\delta_n > \rho_n$, we get*

$$b(n)^q \leq C \left[\left(n 2^{-j_{\eta s}+\delta_n}\right)^{q/2} 1_{\{\hat{j}_\eta + \delta_n \geq j_{\eta s}\}} 1_{\{\hat{j}_\eta \leq j_{\eta s}\}} + n^{q/2} 1_{\{\hat{j}_\eta + \delta_n < j_{\eta s} \text{ or } \hat{j}_\eta > j_{\eta s}\}}\right]$$

and, for all $\delta_n > (2N+1)\rho_n$,

$$E_2 b(n)^q \leq C \left(n 2^{-j_{\eta s}+\delta_n}\right)^{q/2}$$

as soon as

$$\gamma(1/4, M) \wedge \gamma(1/8, M) \geq q/2 \left(\frac{1}{1+2s/\eta} \vee \frac{1}{1+\eta/(2s)}\right).$$

LEMMA 2. *We assume that f belongs to $L_s(M)$ and that $F(x_0)$ is satisfied. Let $q \geq 2$. We get, for all $m \geq 1$,*

$$|s_.|_q^{qm} \leq C \left[\left(\frac{2^{j_{\eta s}}}{n}\right)^{m(q-1)} 1_{\{\hat{j}_\eta \leq j_{\eta s}\}} + 1_{\{\hat{j}_\eta > j_{\eta s}\}}\right],$$

$$|s_.|_\infty^{m} \leq C \left[\left(\frac{2^{j_{\eta s}}}{n}\right)^{m} 1_{\{\hat{j}_\eta \leq j_{\eta s}\}} + 1_{\{\hat{j}_\eta > j_{\eta s}\}}\right],$$

as soon as

$$\gamma(1/8, M) > \frac{(q-1)m}{1+\eta/(2s)}, \qquad E_2 |s_.|_q^{qm} \leq C \left(\frac{2^{j_{\eta s}}}{n}\right)^{m(q-1)}$$

and, as soon as

$$\gamma(1/8, M) > \frac{m}{1+\eta/(2s)}, \qquad E_2 |s_.|_\infty^{m} \leq C \left(\frac{2^{j_{\eta s}}}{n}\right)^{m}.$$

6.1.3. *Edgeworth expansion of the density in the regression model.* We denote $f^1_{b(n)S_n}$ the density, conditionally to the second part of the sample, of the random variable $b(n)S_n$ for

$$S_n = \sum_{i=1}^{n-1} \varepsilon_{2i+1} s_{2i+1} \tag{13}$$

where the s_{2i+1}'s are defined in (8). Let us recall that the kth moment of the errors ε exists, for all k and is denoted σ_k.

PROPOSITION 2. *Let $r \geq 2$. For all $m \geq 1$, we get, conditionally to the second part of the sample,*

$$\forall x, \quad \left| f^1_{b(n)S_n}(x) - \phi(x) - \phi(x) \sum_{k\geq 1}^{r} \sum_{(i_1,\dots,i_p)\in \mathscr{I}_k} T_{i_1 n} \cdots T_{i_p n} P_{i_1 \cdots i_p}(x) \right|$$
$$\leq C \left(T_{(r+1)n} + \left(\sum_{k=3}^{r} T_{kn} \right)^{r-1} + T^m_{\infty n} + (n2^{-\hat{j}_\eta})^{-m} \right)$$

where the T_{qn}'s are defined in (9). *The set of indices is defined by*

$$\mathscr{I}_k = \big\{ (i_1, \dots, i_p),\ \ p \geq 1, (i_1 - 2) + \cdots + (i_p - 2) = k, 3 \leq i_1 \leq \cdots \leq i_p \leq r \big\}, \tag{14}$$

and $P_{i_1 \dots i_p}(x)$ is a real polynomial (*precised in the proof*) *of degree $i_1 + \cdots + i_p$ depending only on $\sigma_{i_1}, \dots, \sigma_{i_p}$ but not on n or r. For instance, for $r \geq 4$, the first terms of the expansion are*

$$T_{3n} \frac{\sigma_3}{3!} H_3(x) \quad (k = 1) \ \ \textit{and} \ \ T^2_{3n} \frac{\sigma_3^2}{3!^2} H_6(x) + T_{4n} \frac{\sigma_4}{4!} H_4(x) \quad (k = 2),$$

where H_3, H_4, H_6 denote the Hermite's polynomials.

6.1.4. *Edgeworth expansion for the distribution in the regression model.* We keep the same notations as above and we denote by $F^1_{b(n)S_n}$ the distribution function associated with the density $f^1_{b(n)S_n}$.

PROPOSITION 3. *Under the same assumptions as in Proposition* 2, *we get, for all $m \geq 1$,*

$$\forall x, \quad \left| F^1_{b(n)S_n}(x) - \Phi(x) - \phi(x) \sum_{k\geq 1}^{r} \sum_{(i_1,\dots,i_p)\in \mathscr{I}_k} T_{i_1 n} \cdots T_{i_p n} Q_{i_1 \cdots i_p}(x) \right|$$
$$\leq C \left(T_{(r+1)n} + \left(\sum_{k=3}^{r} T_{kn} \right)^{r-1} + T^m_{\infty n} + (n2^{-\hat{j}_\eta})^{-m} \right)$$

where $Q_{i_1\cdots i_p}(x)$ *is a real polynomial (precised in the proof) depending only on* $\sigma_{i_1},\dots,\sigma_{i_p}$ *but not on* n *or* r. *For instance, for* $r \geq 4$, *the first term of the expansion of the expansion is* $T_{3n}\frac{\sigma_3}{3!}(1-t^2)$.

6.2. *Proof of Theorem* 1 *in the white noise model.* We have the following expansion:

$$\hat{f}^T(x_0) + \hat{B}_\eta(x_0) - f(x_0) = S_n' - R_{n1} + R_{n2} - R_{n3} + R_{n4} \tag{15}$$

where

$$S_n' = \sum_k (\hat{\alpha}_{j_0k}(1) - \alpha_{j_0k})\phi_{j_0k}(x_0) + \sum_{j_0}^{\hat{j}_\eta}\sum_k (\hat{\beta}_{jk}(1) - \beta_{jk})\psi_{jk}(x_0)$$

$$\begin{aligned} R_{n1} &= \sum_{j_0}^{\hat{j}_1}\sum_K \left[\sum_{(K)} \psi_{jk}(x_0)\hat{\beta}_{jk}(1)\right] 1_{\{\hat{B}_{jK}(1)<t_{j,n}\}}, \\ R_{n2} &= \sum_{\hat{j}_1}^{j_\infty}\sum_K \left[\sum_{(K)} \psi_{jk}(x_0)(\hat{\beta}_{jk}(1) - \beta_{jk})\right] 1_{\{\hat{B}_{jK}(1)>t_{j,n}\}}, \\ R_{n3} &= \sum_{\hat{j}_\eta}^{\infty}\sum_k \psi_{jk}(x_0)\beta_{jk}, \\ R_{n4} &= \sum_{\hat{j}_1}^{j_\infty}\sum_K \left[\sum_{(K)} \psi_{jk}(x_0)\beta_{jk}\right] 1_{\{\hat{B}_{jK}(1)>t_{j,n}\}}. \end{aligned} \tag{16}$$

Because of the regularity of f and using Lemma 1, we get

$$\begin{aligned} b(n)|R_{n3}| &\leq C\,\Big[\left(n2^{-j_{\eta s}+\delta_n}\right)^{1/2}\left(2^{-(j_{\eta s}-\delta_n)s}\right)1_{\{\hat{j}_\eta > j_{\eta s}-\delta_n\}} \\ &\qquad + n^{1/2}2^{-j_0 s}1_{\{\hat{j}_\eta \leq j_{\eta s}-\delta_n \text{ or } \hat{j}_\eta \geq j_{\eta s}\}}\Big] \\ &= C\,\Big[u_n 1_{\{\hat{j}_\eta > j_{\eta s}-\delta_n\}} + n^{1/2}1_{\{\hat{j}_\eta \leq j_{\eta s}-\delta_n\}}\Big]. \end{aligned}$$

Applying Property 3 for $\delta_n > (1+2N)\rho_n$ and Property 2, we deduce

$$\begin{aligned} P(b(n)|R_{n3}| > C\,u_n) &\leq P(\hat{j}_\eta < j_{\eta s} - \delta_n) + P(\hat{j}_\eta > j_{\eta s}) \\ &\leq C\,n^{-(\gamma(1/4,M)\wedge\gamma(1/8,M))}. \end{aligned} \tag{17}$$

and, as soon as

$$\gamma(1/4, M)\wedge\gamma(1/8, M) > \frac{s+1/2}{1+2s/\eta}, \qquad E_2 b(n)|R_{n3}| \leq C\,u_n. \tag{18}$$

Using Lemma 1, observing that at each level j, only a finite number of k's are such that $\psi_{jk}(x_0) \neq 0$, we get for $p' :\ (1/p) + (1/p') = 1$

$$
\begin{aligned}
b(n)|R_{n1}| &\le b(n) \sum_{j_0}^{\hat{j}_1} \sum_K \left[\sum_{(K)} |\psi_{jk}(x_0)|^{p'} \right]^{1/p'} t_{j,n}^{1/p} \\
&\le C \Bigg[\left(n 2^{-j_{\eta s}+\delta_n}\right)^{1/2} \left(2^{j_s} t_{j_1,n}^{2/p}\right)^{1/2} 1_{\{\hat{j}_1 < j_s\}} 1_{\{\hat{j}_\eta \ge j_{\eta s}-\delta_n\}} \\
&\qquad + n^{1/2} \left(2^{j_1} t_{j_1,n}^{2p}\right)^{1/2} 1_{\{\hat{j}_1 > j_s \text{ or } \hat{j}_\eta < j_{\eta s}-\delta_n\}} \Bigg] \\
&= C \left[v_n 1_{\{\hat{j}_1 < j_s\}} + n t_{j_1,n}^{1/p} 1_{\{\hat{j}_1 > j_s \text{ or } \hat{j}_\eta < j_{\eta s}-\delta_n\}} \right].
\end{aligned}
$$

Applying Property 2 and Property 3, we deduce

$$
\begin{aligned}
P(b(n)|R_{n1}| > C\, v_n) &\le P(\hat{j}_1 > j_s) + P(\hat{j}_\eta < j_{\eta s} - \delta_n) \\
&\le C\, n^{-(\gamma(1/4,M)\wedge\gamma(1/8,M))}.
\end{aligned}
\tag{19}
$$

and, as soon as

$$
\gamma(1/4, M) \wedge \gamma(1/8, M) > 2 + \frac{s(1-1/\eta)}{(1+2s)(1+2s/\eta)}, \qquad E_2 b(n)|R_{n1}| \le C\, v_n. \tag{20}
$$

Because of definition (3) of $\hat{j}_1$ and the triangular inequalities, we have

$$
\begin{aligned}
|R_{n2}| &\le \sum_{\hat{j}_1}^{j_\infty} \sum_K \left[\sum_{(K)} |\psi_{jk}(x_0)(\hat{\beta}_{jk}(1) - \beta_{jk})| \right] 1_{\{\hat{B}_{jK}(1) > t_{j,n}\}} 1_{\{\hat{B}_{jK}(2) \le 2^{-2p} t_{j,n}\}} \\
&\qquad \times \left(1_{\{B_{jK} \le 2^{-p} t_{j,n}\}} + 1_{\{B_{jK} > 2^{-p} t_{j,n}\}} \right) \\
&\le \sum_{j_0}^{j_\infty} \sum_K \left[\sum_{(K)} \left|\psi_{jk}(x_0)(\hat{\beta}_{jk}(1) - \beta_{jk})\right| \right] 1_{\{\sum_{(K)} |\hat{\beta}_{jk}(1) - \beta_{jk}|^p \ge 2^{-p} t_{j,n}\}} \\
&\quad + \sum_{j_0}^{j_\infty} \sum_K \left[\sum_{(K)} \left|\psi_{jk}(x_0)(\hat{\beta}_{jk}(1) - \beta_{jk})\right| \right] 1_{\{\sum_{(K)} |\hat{\beta}_{jk}(2) - \beta_{jk}|^p \ge 2^{-2p} t_{j,n}\}}.
\end{aligned}
$$

Applying Proposition 1 for the first part of the sample, we get

$$
\begin{aligned}
E_1 |R_{n2}|^2 &\le C \Bigg(\left(\frac{2^{j_\infty}}{n} n^{-\frac{1}{2}\gamma(\frac{1}{2}, M)} \right) \\
&\qquad + \frac{1}{n} \sum_{j_0}^{j_\infty} \sum_k \psi_{jk}^2(x_0) 1_{\{\sum_{(K)} |\hat{\beta}_{jk}(2) - \beta_{jk}|^p \ge 2^{-2p} t_{j,n}\}} \Bigg).
\end{aligned}
$$

Using the Cauchy-Schwarz inequality, Proposition 1 for the second part of the sample and Lemma 1, we obtain

$$E|b(n)R_{n2}|^2 \leq \left(E_2 b(n)^4\right)^{1/2} \left(E_2(E_1 R_{n2}^2)^2\right)^{1/2}$$
$$\leq C \ \left(n2^{-j_{\eta s}+\delta_n}\right) \left(\frac{2^{j_\infty}}{n} n^{-\frac{1}{2}\gamma(\frac{1}{2},M)-\frac{1}{2}\gamma(\frac{1}{4},M)}\right).$$

It follows that

$$P_1(b(n)|R_{n2}| > Cu_n) \leq C^{-2} u_n^{-2} E_1 |b(n)R_{n2}|^2 = U_{1n} \tag{21}$$

where U_{1n} is a random variable depending on the second part of the sample, such that

$$E_2 U_{1n} \leq C \ \left(n2^{-j_{\eta s}(2s+1)} 2^{\delta_n(2s+1)}\right)^{-1} \left(n^{-\frac{1}{2}\gamma(\frac{1}{2},M)-\frac{1}{2}\gamma(\frac{1}{4},M)} 2^{j_\infty} 2^{-j_{\eta s}+\delta_n}\right)$$
$$\leq C \ u_n$$

as soon as

$$\gamma(1/2, M) + \gamma(1/4, M) > \frac{2s(1-1/\eta)-2}{1+2s\eta}.$$

Obviously, we also deduce

$$P(b(n)|R_{n2}| > Cu_n) \leq C \ u_n. \tag{22}$$

In the same way, we have

$$|R_{n4}| \leq \sum_{j_0}^{j_\infty} \sum_K \left[\sum_{(K)} \left|\psi_{jk}(x_0)\beta_{jk}\right|\right] 1_{\{\sum_{(K)} |\hat{\beta}_{jk}(1)-\beta_{jk}|^p \geq 2^{-p} t_{j,n}\}}$$
$$+ \sum_{j_0}^{j_\infty} \sum_K \left[\sum_{(K)} \left|\psi_{jk}(x_0)\beta_{jk}\right|\right] 1_{\{\sum_{(K)} |\hat{\beta}_{jk}(2)-\beta_{jk}|^p \geq 2^{-2p} t_{j,n}\}}.$$

Using the regularity of f and the same arguments as above, we obtain the same result

$$P_1(b(n)|R_{n4}| > Cu_n) \leq C \ U_{2n} \tag{23}$$

where

$$E_2 U_{2n} \leq C \ \left(n2^{-j_{\eta s}(2s+1)} 2^{\delta_n(2s+1)}\right)^{-1} \left(n^{-\frac{1}{2}\gamma(\frac{1}{2},M)-\frac{1}{2}\gamma(\frac{1}{4},M)} 2^{-2j_0 s} n 2^{-j_{\eta s}+\delta_n}\right)$$
$$\leq C \ u_n,$$

with the same constraint on $\gamma(1/2, M)$ and $\gamma(1/4, M)$. We also get

$$P(b(n)|R_{n4}| > Cu_n) \leq u_n. \tag{24}$$

We now use Lemma 4 for the probability P (see Appendix B): since the variable $b(n)S_n$ is standard Gaussian and because (17), (19), (22) and (24), we obtain

$$\begin{aligned}P\left(b(n)\left[\hat{f}(x_0) - \hat{B}_\eta(x_0) - f(x_0)\right] < t\right) \\ &= P(b(n)S'_n - b(n)R_{n1} + b(n)R_{n2} - b(n)R_{n3} + b(n)R_{n4} < t) \\ &= \Phi(t) + O\Big(u_n + v_n + n^{-\gamma(1/4,M)} + n^{-\gamma(1/8,M)} \\ &\qquad + n^{-1/2\gamma(1/2,M)-1/2\gamma(1/4,M)} n^{\frac{2s}{1+2s/\eta}}\Big)\end{aligned}$$

which leads to the result because M is supposed large enough so we are able to choose $\gamma(., M)$ as prescribed (see the definition of the function γ in Appendix C). □

6.3. *Proof of Theorem* 1 *in the regression model.* We have an analogue of the expansion (15) with S'_n replaced by

$$S_n = \sum_{i=1}^{n-1} \varepsilon_{2i+1} s_{2i+1}$$

where the s_{2i+1}'s are defined in (8). We observe that all the arguments using the wavelet coefficients hold without modifications. We now apply Proposition 3 for $r = 2$. Since $\mathscr{I}_1 = \mathscr{I}_2 = \varnothing$, we find (with $m = 1$)

$$\forall t,\ |F^1_{b(n)S_n}(t) - \Phi(t)| \le C\ \left(T_{3n} + T_{\infty n} + (n^{-1}2^{\hat{j}_\eta})\right).$$

Using the Cauchy-Schwarz inequality and applying Lemma 1 and Lemma 2 for some

$$\gamma(1/4, M) \wedge \gamma(1/8, M) > 4\left(\frac{1}{1+2s\eta} \vee \frac{1}{1+\eta/(2s)}\right),$$

we obtain

$$E_2 T^m_{3n} \le (E_2 b(n)^6 E_2 |s_.|_3^6)^{1/2} \le C\ \left(\frac{2^{j_{\eta s}}}{n}\right)^{\frac{1}{2}} 2^{\delta_n \frac{3}{2}},$$

$$E_2 T_{\infty n} \le (E_2 b(n)^2 E_2 |s_.|_\infty^2)^{1/2} \le C\ \left(\frac{2^{j_{\eta s}}}{n}\right)^{\frac{1}{2}} 2^{\delta_n \frac{1}{2}}.$$

Applying Property 2 gives the bound $E_2|n^{-1}2^{\hat{j}_\eta}| \le C\ n^{-1}2^{j_{\eta s}}$ and we deduce

$$\forall t,\ |F^1_{b(n)S_n}(t) - \Phi(t)| \le O(W_n(2))$$

where $W_n(2)$ is a random variable such that

$$E_2|W_n(2)| \le \left(\left(\frac{2^{j_{\eta s}}}{n}\right)^{\frac{1}{2}} 2^{3\delta_n/2}\right).$$

Choosing $\gamma(1/2, M)$, $\gamma(1/4, M)$, $\gamma(1/8, M)$ as prescribed (which is possible because M is large enough), remembering (18), (20), (21) and (23) and applying Lemma 4 for the probability P_1 (see Appendix B), we get the result.

6.4. *Proof of Theorem* 2 *in the white noise model.* We have the following expansion:

$$\hat{f}^L(x_0) + \hat{B}_\eta(x_0) - f(x_0) = S'_n - R_{n3} \tag{25}$$

where S'_n and R_{n3} are defined in (16). The proof is exactly similar to the proof of Theorem 1 in the white noise setting.

6.5. *Proof of Theorem* 2 *in the regression model.* We have an analogue of the expansion (25) for the variable S_n [defined in (13)]. Thanks to Lemma 4 (see Appendix B), we can restrict our attention to the expansion of the distribution function of the random variable $b(n)S_n$. This will be done using Proposition 3. We need to bound the random quantities appearing in the expansion of Proposition 3. Let $1 \leq k \leq r$ and let $(i_1, \ldots, i_p)$ be an index. Using the Cauchy-Schwarz inequality and applying Lemma 1 and Lemma 2, we obtain, for all $q, m \geq 2$

$$E_2 T_{qn}^m \leq \left(E_2 b(n)^{2qm} E_2 |s_.|_q^{2qm}\right)^{1/2} \leq C \left(\frac{2^{j_{\eta s}}}{n}\right)^{\frac{m(q-2)}{2}} 2^{\delta_n \frac{qm}{2}}.$$

Using the Hölder inequality for $1/a_1 + \cdots + 1/a_p = 1$, we have

$$\begin{aligned} E_2(T_{i_1 n} \cdots T_{i_p n}) &\leq C \prod_{l=1}^{p} \left(E_2 T_{i_l n}^{a_l}\right)^{1/a_l} \\ &\leq C \prod_{l=1}^{p} \left(\frac{2^{j_{\eta s}}}{n}\right)^{(i_l-2)/2} \left(2^{\delta_n\, i_l/2}\right). \end{aligned}$$

We deduce

$$E_2(T_{i_1 n} \cdots T_{i_p n}) \leq C \left(\frac{2^{j_{\eta s}}}{n}\right)^{\frac{k}{2}} 2^{\delta_n (k/2+p)} \qquad \text{if } (i_1, \ldots, i_p) \in \mathscr{I}_k,$$

$$E_2 T_{(r+1)n} \leq C \left(\frac{2^{j_{\eta s}}}{n}\right)^{\frac{r-1}{2}} 2^{\delta_n \frac{(r-1)}{2}},$$

$$E_2 \left(\sum_{k=3}^{r} T_{kn}\right)^{r-1} \leq C \left(\frac{2^{j_{\eta s}}}{n}\right)^{\frac{r-1}{2}} 2^{\delta_n \frac{3(r-1)}{2}},$$

$$E_2 T_{\infty n}^m \leq C \left(\frac{2^{j_{\eta s}}}{n}\right)^{\frac{m}{2}} 2^{\delta_n \frac{m}{2}}.$$

Applying Property 2 gives the bound

$$E_2 \left(n 2^{-\hat{j}_\eta}\right)^{-m} \leq C \left(\frac{2^{j_{\eta s}}}{n}\right)^{m}.$$

Theorem 2 follows from Proposition 3 with m chosen so that $m > r - 1/2$, and $\gamma(1/8, M)$ large enough in such a way that the contribution of $n^{-\gamma(1/8,M)}$ is negligible.

APPENDIX A: PROOF OF THE EDGEWORTH EXPANSIONS

In this section, we work conditionally to the second part of the sample. We follow along the line Feller (Chap XVI). The proofs are standard: first, we establish the Edgeworth expansion for the density (Proposition 2). The construction of the terms in the expansion comes out of this proof. Next, using the previous construction, we establish the expansion for the distribution (Proposition 3). For that, we need to bound the tails of the density (Lemma 3).

A.1. *Preliminary*: *Study of the characteristic function associated to $b(n)S_n$*. Let χ be the characteristic function associated to the distribution of the ε's. We derive the characteristic function associated to $b(n)S_n$

$$\forall x,\ \chi_{b(n)S_n}(x) = \prod_{l=0}^{n-1} \chi(b(n)s_{2l+1}x).$$

Using the identity $b(n) = \sigma_2^{-1}|s_.|_2^{-2}$, we get

$$\begin{aligned}
\chi_{b(n)S_n}(x) &= \exp -\frac{x^2}{2}\left[\exp\left(\sum_{l=0}^{n-1}\log\chi(b(n)s_{2l+1}x) + \frac{x^2}{2}\right)\right] \\
&= \exp -\frac{x^2}{2}\left[\exp\sum_{l=0}^{n-1}\left(\log\chi(b(n)s_{2l+1}x) + \frac{\sigma_2^2(b(n)s_{2l+1}x)^2}{2}\right)\right] \\
&= \exp -\frac{x^2}{2}\left[\exp\sum_{l=0}^{n-1}(\tilde{\chi}(b(n)s_{2l+1}x))\right]
\end{aligned}$$

where

$$\tilde{\chi}(x) = \log\chi(x) + \frac{\sigma_2^2\,x^2}{2}.$$

Thanks to a Taylor expansion of $\tilde{\chi}$ around the origin, we get

$$\chi_{b(n)S_n}(x) = \exp -\frac{x^2}{2}\left[\exp(\alpha(x) + \beta(x))\right]. \tag{26}$$

where

$$\alpha(x) = \sum_{k=3}^{r} T_{kn}(ix)^k\frac{\sigma_k}{k!}\ |\beta(x)| \le C\ T_{r+1n}|x|^{r+1}.$$

A.2. *Proof of Proposition* 2: *Construction of the expansion.* Let H_k denote the kth Hermite polynomial and $\mathscr{F}$ the Fourier transform. Let P denote the following polynomial with real coefficients $p_1, p_2, \dots$:

$$P(x) = \sum_{k=1}^{r-2} \frac{\alpha(-ix)^k}{k!}.$$

Then

$$g = f_{b(n)S_n} - \phi - \phi \sum_k p_k H_k \tag{27}$$

satisfies [because $\mathscr{F}(g) \in L^1$], for all $T > 0$,

$$\begin{aligned}
|g(t)| &= \frac{1}{2\pi} \left| \int e^{-itx} \mathscr{F}(g)(x) dx \right| \\
&\le C \int_{|x| \le T} \left| \mathscr{F}(g)(x) \right| dx + C \int_{|x|>T} \left| \mathscr{F}(f_{b(n)S_n})(x) \right| dx \\
&\quad + C \int_{|x|>T} \left| \mathscr{F}(\phi - \phi \sum_k p_k H_k)(x) \right| dx .
\end{aligned}$$

Using equality (26) and the properties of the Hermite polynomials, we obtain

$$\begin{aligned}
|g(t)| \le{}& C \int_{|x| \le T} \left| \exp(\alpha(x) + \beta(x)) - 1 - P(ix) \right| \exp -\frac{x^2}{2} dx \\
&+ C \int_{|x|>T} \left| \mathscr{F}(f_{b(n)S_n})(x) \right| dx + C \int_{|x|>T} (1 + |P(ix)|) \exp -\frac{x^2}{2} dx \\
\le{}& A_1 + A_2 + A_3.
\end{aligned}$$

For the first term, we use the formula

$$\left| e^a - 1 - \sum_1^{r-2} \frac{b^k}{k!} \right| \le e^c \left(|a - b| + \frac{|b|^{r-1}}{r - 1!} \right),$$

valid when $|a| < c$ and $|b| < c$. We take $a := \alpha(x) + \beta(x)$ and $b := \alpha(x)$. We choose

$$\tilde{T}_n = \min\left(\frac{T_{3n}}{T_{4n}}, \dots, \frac{T_{(r-1)n}}{T_{rn}}, T_{3n}^{-1} \right) = \min\left(\frac{T_{(r-1)n}}{T_{rn}}, T_{3n}^{-1} \right)$$

in such a way that $\exp(|\alpha(x)| + |\beta(x)|) \le \exp c_0 \; x^2$ (for $0 < c_0 < 1/2$) when $|x| \le c\tilde{T}_n$ where c is a positive constant to determine. In fact, using again the identity $b(n) = \sigma_2^{-1} |s_.|_2^{-2}$, we remark that

$$T_n = C \; T_{\infty n}^{-1} \le \tilde{T}_n$$

and then we consider T_n instead of $\tilde{T}_n$. We obtain

$$\begin{aligned}
A_1 &\leq C \int_{|x|\leq cT_n} \left(|\beta(x)| + |\alpha(x)|^{r-1}\right) \exp\left(|\alpha(x)| + |\beta(x)|\right) \, \exp -\frac{x^2}{2} dx \\
&\leq C \int \left(T_{(r+1)n}|x|^{r+1} + \left| \sum_{k=3}^{r} T_{kn} x^k \right|^{r-1} \right) \exp -(1/2 - c_0)x^2 \, dx \\
&\leq C \left(T_{(r+1)n} + \left(\sum_{k=3}^{r} T_{kn} \right)^{r-1} \right).
\end{aligned}$$

The last term A_3 is obviously bounded by any power of $(cT_n)^{-1} = C\, T_{\infty n}$. Let us study A_2. Because of Condition $F(x_0)$, there exists $l_0 \in [x_{0-1/n}, x_{0+1/n}]$ such that

$$b(n)|s_{l_0}| \geq b(n)\, m \frac{2^{\hat{j}_\eta}}{n} ||\psi||_\infty = \lambda_n.$$

Let us consider l such that $b(n)|s_l - s_{l_0}| \leq \lambda_n/2$. Then l satisfies $b(n)|s_l| > \lambda_n/2$. Because we obviously have

$$|s_l - s_{l_0}| \leq \frac{2^{2\hat{j}_\eta}}{n^2} |l - l_0| \, ||\psi||_\infty = |l - l_0| \mu_n,$$

we deduce

$$\begin{aligned}
\#\{l;\ b(n)|s_l| > \lambda_n/2\} &\geq \#\{l;\ b(n)|s_l - s_{l_0}| \leq \lambda_n/2\} \\
&\geq \#\{l;\ |l - l_0| \leq C\, \lambda_n/2\, \mu_n^{-1} b(n)^{-1}\} \\
&\geq C\, \lambda_n \mu_n^{-1} b(n)^{-1} \\
&\geq C\, n 2^{-\hat{j}_\eta} = N_n.
\end{aligned}$$

The assumption on the characteristic function of the ε's implies that

$$\lim_{x \longrightarrow \infty} |\chi(x)| = 0.$$

Then there exists some $q < 1$ and some $c_q > 0$ such that

$$\sup_{|x| > c_q} |\chi(x)| < q.$$

Using Lemma 2, there exists some constant C such that

$$b(n)|s_.|_\infty |x| > c \Rightarrow |y| = b(n)|s_l||x| > |s_l||s_.|_\infty^{-1}\, c \geq \left(\frac{m^2}{2} \frac{2^{\hat{j}_\eta}}{n} \right) \left(C\, \frac{2^{\hat{j}_\eta}}{n} \right)^{-1} c.$$

We just have to choose $c > c_q\, C(2/m^2)$ to obtain

$$
\begin{aligned}
A_2 &\le \int_{b(n)|s_.|_\infty |x|>c} \prod_{l\in\{l; b(n)|s_l|>\lambda_n/2\}} \left|\chi(b(n)s_{2l+1}x)\right| \\
&\qquad\qquad \times \prod_{l\in\{l; b(n)|s_l|\le\lambda_n/2\}} \left|\chi(b(n)s_{2l+1}x)\right| dx \\
&\le Cq^{N_n} \int \prod_{l\in\{l; b(n)|s_l|\le\lambda_n/2\}} \left|\chi(b(n)s_{2l+1}x)\right| dx
\end{aligned}
$$

and this tends to zero more rapidly than any power of N_n^{-1}.

To complete the proof and obtain the result given in Proposition 2, it remains to rearrange the coefficients p_k of the polynomial

$$
P(x) = \sum_{k'} \left(\sum_{k=3}^{r} T_{kn} x^k \frac{\sigma_k}{k!} \right)^{k'},
$$

with respect to the order of (14).

A.3. *Proof of Proposition* 3. We follow along the lines the proof of Proposition 2 and we apply the following Lemma [see Feller (1966), volume 2, page 512] when G is the standard Gaussian and F is the distribution associated with the function g defined in (27).

LEMMA 3. *Let F be a probability distribution with vanishing expectation and characteristic function ξ. Suppose that $F-G$ vanishes at $+-\infty$ and that G has a derivative g such that $|g| \le m$. Finally, suppose that g has a continuously differentiable Fourier transform γ such that $\gamma(0) = 1$ and $\gamma'(0) = 0$. Then, for all x and for all $T > 0$, we get*

$$
\left|F(t) - G(t)\right| \le \frac{1}{\pi} \int_{-T}^{T} \left| \frac{\xi(x) - \gamma(x)}{x} \right| dx + \frac{24m}{\pi T}.
$$

Choosing $T \ge cT_n$ and remembering that $T_n = T_{\infty n}^{-1}$, we obtain

$$
\begin{aligned}
\left|F(x) - G(x)\right| \le C \Biggl(&\int_{|x|\le cT_n} \left| \frac{\mathscr{F}(g)(x) - \gamma(x)}{x} \right| dx \\
&+ \int_{cT_n<|x|\le T} \left| \frac{\mathscr{F}(g)(x) - \gamma(x)}{x} \right| dx + T^{-1} \Biggr).
\end{aligned}
$$

The study of the first term (respectively the second) is similar to the study of A_1 (respectively, A_2 and A_3). The choice $T = (n2^{-\hat{j}_\eta})^m$ leads to the result. □

APPENDIX B: TECHNICAL RESULTS AND PROOFS

B.1. *Proofs of Property* 2 *and Property* 3.

B.1.1. *Case* $\eta = 1$. On one hand, because of the definition (3) of $\hat{j}_1$, we have

$$
\{\hat{j}_1 > j_s\} \subset \{\exists j^* > j_s, \exists K^*, \hat{B}_{j^*K^*}(2) > 2^{-2p} t_{j^*,n}\}. \tag{28}
$$

On the other hand, using Property 1, we have

$$f \in L_s(M) \Rightarrow \forall K, B_{jK} < 2^{-4p} t_{j,n} \quad \text{as soon as } j > j_s. \tag{29}$$

By the triangular inequality, we get

$$\sum_{(K)} |\hat{\beta}_{jk}(2)|^p \le 2^{(p-1)} \left(\sum_{(K)} |\hat{\beta}_{jk}(2) - \beta_{jk}|^p + \sum_{(K)} |\beta_{jk}|^p \right). \tag{30}$$

Then combining (28), (29) and (30), we obtain

$$P(\hat{j}_1 > j_s) \le P\left(\sum_{k \in K^*} |\hat{\beta}_{j^*k}(2) - \beta_{j^*k}|^p > 2^{-3p} t_{j^*,n} \right).$$

Using Proposition 1 (large deviation inequality), we obtain Property 2 for $\eta = 1$. We establish Property 3 using $H_s(M, x_0)$ instead of the assumption of Besov regularity. Because of definition (3) of $\hat{j}_1$, we have

$$\{\hat{j}_1 < j_s - \rho_n\} \subset \{\forall j_1 \ge j \ge j_s - \rho_n, \forall K,\ \hat{B}_{jK}(2) \le 2^{-2p} t_{j,n}\}. \tag{31}$$

Now,

$$H_s(M, x_0) \Rightarrow\ \exists j^*, \quad j_1 \ge j^* \ge j_s - \rho_n, \exists K^*, B_{j*K*} > 2^{-p} t_{j^*,n}. \tag{32}$$

By the triangular inequality, we get

$$\sum_{(K)} |\hat{\beta}_{jk}(2) - \beta_{jk}|^p \ge 2^{(1-p)} B_{jK} - \hat{B}_{j*K*}(2). \tag{33}$$

Combining (31), (32) and (33), we obtain

$$P(\hat{j}_1 < j_s - \rho_n) \le P\left(\sum_{(K*)} |\hat{\beta}_{j*k*} - \beta_{j*k*}|^p \ge 2^{-2p} t_{j^*,n} \right).$$

Using Proposition 1 (large deviation inequality), we establish Property 3 for $\eta = 1$.

B.1.2. *Case* $\eta > 1$. Considering now the formulas

$$\hat{s} = \frac{1}{2}\left(\frac{\log_2 n}{\hat{j}_1} - 1 \right) \quad \text{and} \quad \hat{j}_\eta = \frac{1 + 2\hat{s}}{1 + 2\hat{s}/\eta} \hat{j}_1,$$

we easily get that

$$\hat{j}_1 \le j_s \Rightarrow \hat{j}_\eta \le j_{\eta s}$$

and then

$$\forall \eta \ge 1, \qquad P(\hat{j}_\eta > j_{\eta s}) \le P(\hat{j}_1 > j_s)$$

which ends the proof of Property 2. Moreover, let us assume that $\hat{j}_1 > j_s - \rho_n$ and put

$$r_n = \frac{\rho_n}{\log_2 n}.$$

Using the bounds $0 < \hat{s}, s \leq N$, it follows that

$$\hat{j}_\eta = \frac{1+2\hat{s}}{1+2\hat{s}/\eta}\hat{j}_1 \geq \frac{1+2s}{1+2s/\eta} j_s + \left(\frac{1+2\hat{s}}{1+2\hat{s}/\eta} - \frac{1+2s}{1+2s/\eta}\right) j_s - \rho_n \left(\frac{1+2\hat{s}}{1+2\hat{s}/\eta}\right)$$

$$\geq j_{\eta s} + \frac{\rho_n}{r_n}\left(\frac{2(\hat{s}-s)(1-(1/\eta)}{(1+2s)(1+2\hat{s}/\eta)(1+2s/\eta)}\right) - \rho_n\left(\frac{1+2\hat{s}}{1+2\hat{s}/\eta}\right)$$

$$\geq -\rho_n(1+2N) \qquad \text{if } \hat{s} \geq s$$

Then, applying Property 1, we obtain for $\tau_n = (1+2N)\rho_n$,

$$\begin{aligned} P(\hat{j}_1 > j_s - \rho_n) &\leq P(\hat{j}_\eta > j_{\eta s} - \tau_n \text{ and } \hat{j}_1 < j_s) + P(\hat{j}_1 \geq j_s) \\ &\leq P(\hat{j}_\eta > j_{\eta s} - \tau_n) + P(\hat{j}_1 \geq j_s) \\ &\leq P(\hat{j}_\eta > j_{\eta s} - \tau_n) + Cn^{-\gamma(1/8,M)} \end{aligned}$$

We deduce

$$P(\hat{j}_\eta \leq j_{\eta s} - \tau_n) \leq P(\hat{j}_1 \leq j_s - \rho_n) + Cn^{-\gamma(1/8,M)}.$$

which completes the proof of Property 3 for any $\eta \geq 1$. □

B.2. *Probability technical lemma.*

LEMMA 4. *Let X_n be a sequence of random variables admitting the Edgeworth expansion*

$$P(X_n < t) = \Phi(t) + p_n(t)\phi(t) + O(u_n)$$

with some polynomials p_n of bounded order with bounded coefficients. We assume that the sequence Y_n of random variables satisfies

$$P(|Y_n| > v_n) \leq w_n.$$

Then

$$P(X_n + Y_n < t) = P(X_n < t) + O(u_n + v_n + w_n).$$

The proof follows immediately from the inequalities

$$\begin{aligned} P(X_n + Y_n < t) &\leq P(X_n < t + v_n) + P(|Y_n| > v_n) \\ &\geq P(X_n < t + v_n) - P(|Y_n| > v_n) \end{aligned}$$

and the Lipschitz equicontinuity of the functions $\Phi(t) + p_n(t)\phi(t)$.

B.3. *Approximation of a Riemann integral with a sum.*

LEMMA 5. *Let $(g^l)_{l=1,\dots,q}$ be q compactly supported (support included in $[-N, N]$) functions in $B_{s_l,\infty,\infty}$ for $s_l > 0$. Let us denote*

$$\sigma_l = \min(s_l, 1), \quad G(x) = \prod_{l=1}^q g^l_{j_l k_l}(x), \quad ||G|| = \prod_{l=1}^q ||g^l||_\infty.$$

For $j_1 \geq \sup(j_l,\ l = 2, \ldots, q)$, *we have*

$$I = |G(x_0)| \left| \int_0^1 G(x)dx - \frac{1}{n} \sum_{i=0}^{n-1} G(\frac{2i+1}{2n}) \right|$$
$$\leq 2N||G||^2 \prod_{l=2}^{q} 2^{j_l} \sum_{l=1}^{q} \left(\frac{2^{j_l}}{n}\right)^{\sigma_l} \frac{||g^l||_{\sigma_l,\infty,\infty}}{||g^l||_\infty}.$$

PROOF. Let $l = 1, \ldots, q$. To simplify the notation, we put

$$\omega^l(u) = \left| g^l_{j_l k_l}(u) - g^l_{j_l k_l}\left(\frac{2i+1}{2n}\right) \right|$$

(ignoring the index i). Let us first remark that, if $g \in L_2$, then

$$\sum_{i=0}^{n} \int_{\frac{2i+1}{2n}}^{\frac{2i+3}{2n}} |g(u)|\omega^l(u)du \leq 2^{j_l/2} \left(\frac{2^{j_l}}{n}\right)^{\sigma_l} ||g^l||_{\sigma_l\infty\infty} \int_0^1 |g|. \tag{34}$$

We decompose I into q terms

$$I \leq |G(x_0)| \sum_{i=0}^{n-1} \int_{\frac{2i+1}{2n}}^{\frac{2i+3}{2n}} \left| G\left(\frac{2i+1}{2n}\right) - G(u) \right| du$$
$$\leq |G(x_0)| \sum_{i=0}^{n-1} \int_{\frac{2i+1}{2n}}^{\frac{2i+3}{2n}} \sum_{p=1}^{q} \left| \prod_{l=1}^{p-1} g^l_{j_l k_l}(u)\ \omega^p(u) \prod_{l=p+1}^{q} g^l_{j_l k_l}\left(\frac{2i+1}{2n}\right) \right| du$$
$$= \sum_{l=1}^{q} I_q.$$

The terms $I_2, \ldots, I_q$ can be bounded using the same argument. For example, we study the last one. Using the inequality (34) and noticing that $\int_0^1 |g^1_{j_1 k_1}(u)|du \leq 2^{-j_1/2} \|g^1\|_\infty$, we get

$$I_q \leq 2^{j_2+\cdots+j_q} ||g^1||^2_\infty| \cdots ||g^q||_\infty \left(\frac{2^{j_1}}{n}\right)^{\sigma_1} ||g^1||_{\sigma_1\infty\infty}.$$

The first term requires a slightly different treatment:

$$I_1 \leq 2^{j_2+\cdots+j_q} ||g^1||^2_\infty \cdots ||g^{q-1}||^2_\infty \left(\frac{2^{j_q}}{n}\right)^{\sigma_q}$$
$$\times ||g^q||_{\sigma_q\infty\infty} 2^{j_q} ||g^q||_\infty n^{-1} \# \left\{ i / g^q_{j_q k_q}(x_0) g^q_{j_q k_q}\left(\frac{2i+1}{n}\right) \neq 0 \right\}$$

Observing that

$$\# \left\{ i / g^q_{j_q k_q}(x_0) g^q_{j_q k_q}\left(\frac{2i+1}{n}\right) \neq 0 \right\} \leq 2Nn2^{-j_q},$$

we obtain the result. □

B.4. *Proofs of Lemma* 1 *and Lemma* 2.

B.4.1. *White noise model.* Because of Condition $F(x_0)$, we get

$$b(n)^q = \left[\frac{1}{n}\sum_{j_0}^{\hat{j}_\eta}\sum_k \psi^2_{jk}(x_0) + \frac{1}{n}\sum_k \phi^2_{j_0k}(x_0)\right]^{-q/2} \left(1_{\{j_{\eta s}-\delta_n \le \hat{j}_\eta\}} + 1_{\{\hat{j}_\eta < j_{\eta s}-\delta_n\}}\right)$$

$$\le C\left(\frac{2^{j_{\eta s}-\delta_n}}{n}\right)^{-q/2} 1_{\{j_{\eta s}-\delta_n \le \hat{j}_\eta\}} + \left(\frac{2^{j_0}}{n}\right)^{-q/2} 1_{\{\hat{j}_\eta < j_{\eta s}-\delta_n\}}.$$

Moreover, assuming Condition $H_s(M, x_0)$, we apply Property 3 for $\delta_n > (2N+1)\rho_n$ and for some

$$\gamma(1/4, M) \wedge \gamma(1/8, M) \ge \frac{q/2}{1+2s/\eta}.$$

We obtain

$$E_2 b(n)^q \le C\left(\frac{2^{j_{\eta s}-\delta_n}}{n}\right)^{-q/2}$$

which completes the proof of Lemma 1. □

B.4.2. *Regression model.* Let $q \ge 2$. For sake of simplicity, we denote $\psi_{j_0-1,k} = \phi_{j_0k}$. We use the approximation of a sum by a Riemann integral

$$\sum_{i=1}^n |s_i|^q = \sum_{i=1}^n \left|\frac{1}{n}\sum_{j=j_0-1}^{\hat{j}_\eta}\sum_k \psi_{jk}(x_0)\psi_{jk}\left(\frac{i}{n}\right)\right|^q$$

$$\le \frac{1}{n^{q-1}}\sum_{l_1,\dots,l_q=j_0-1}^{\hat{j}_\eta}\sum_{k_1,\dots,k_q}\prod_{p=1}^q |\psi_{l_pk_p}(x_0)|\left(\frac{1}{n}\sum_{i=1}^n\prod_{p=1}^q\left|\psi_{l_pk_p}\left(\frac{i}{n}\right)\right|\right)$$

$$\le \frac{1}{n^{q-1}}\sum_{l_1,\dots,l_q=j_0-1}^{\hat{j}_\eta}\sum_{k_1,\dots,k_q}\prod_{p=1}^q |\psi_{l_pk_p}(x_0)|\left(\int_0^1\prod_{p=1}^q|\psi_{l_pk_p}(u)|du\right) + R_n$$

where

$$R_n = \frac{1}{n^{q-1}}\sum_{l_1,\dots,l_q=j_0}^{\hat{j}_\eta}\sum_{k_1,\dots,k_q} R_{l_1,\dots,l_q}\left(1_{\{\hat{j}_\eta < j_{\eta s}\}} + 1_{\{\hat{j}_\eta \ge j_{\eta s}\}}\right)$$

for

$$\sum_{k_1,\dots,k_q} R_{l_1,\dots,l_q} = \sum_{k_1,\dots,k_q}\prod_{p=1}^q|\psi_{l_pk_p}(x_0)|$$

$$\times\left(\frac{1}{n}\sum_{i=1}^n\prod_{p=1}^q\left|\psi_{l_pk_p}\left(\frac{i}{n}\right)\right| - \int_0^1\prod_{p=1}^q|\psi_{l_pk_p}(u)|du\right).$$

Using Lemma 5 taking $l_1, \dots, l_q$ arbitrary and $g^1 = \cdots = g^q = \psi$, we obtain

$$\sum_{k_1,\dots,k_q} |R_{l_1,\dots,l_q}| \le C \, \frac{2^{l_1+l_2+\cdots+l_q}}{n}.$$

Then we get

$$|R_n| \le C \left[\left(\frac{2^{j_{\eta s}}}{n} \right)^q 1_{\{\hat{j}_\eta < j_{\eta s}\}} + \left(\frac{2^{j_\infty}}{n} \right)^q 1_{\{\hat{j}_\eta \ge j_{\eta s}\}} \right]$$

It comes that, for all $m \ge 2$,

$$\begin{aligned} \left(\sum_{i=1}^n |s_i|^q \right)^m &= \Bigg(\frac{1}{n^{q-1}} \sum_{l_1,\dots,l_q=j_0-1}^{\hat{j}_\eta} \sum_{k_1,\dots,k_q} \prod_{p=1}^q |\psi_{l_p k_p}(x_0)| \Big(\int_0^1 \prod_{p=1}^q |\psi_{l_p k_p}(u)| du \Big) \\ &\qquad \times \big(1_{\{\hat{j}_\eta < j_{\eta s}\}} + 1_{\{\hat{j}_\eta \ge j_{\eta s}\}} \big) + R_n \Bigg)^m \\ &\le C \left[\Big(\frac{2^{j_{\eta s}}}{n} \Big)^{q-1} 1_{\{\hat{j}_\eta < j_{\eta s}\}} + \Big(\frac{2^{j_\infty}}{n} \Big)^{q-1} 1_{\{\hat{j}_\eta \ge j_{\eta s}\}} + R_n \right]^m . \end{aligned}$$

Moreover, applying Property 2 for some

$$\gamma(1/8, M) \ge \frac{m(q-1)}{1 + \eta/(2s)},$$

we obtain

$$E_2 |s_.|_q^{mq} \le C \left(\frac{2^{j_{\eta s}}}{n} \right)^{m(q-1)}$$

which ends the proof of Lemma 2. The proof is similar when $q = \infty$. We immediately get, for all $m \ge 2$,

$$\begin{aligned} |s_.|_\infty^m &= \sup_l \left| \frac{1}{n} \sum_{j=j_0}^{\hat{j}_\eta} \sum_k \psi_{jk}(x_0) \psi_{jk}\Big(\frac{2l+1}{2n}\Big) \right|^m \\ &\le C \left[\left(\frac{2^{j_{\eta s}}}{n} \right)^m 1_{\{\hat{j}_\eta < j_{\eta s}\}} + \left(\frac{2^{j_\infty}}{n} \right)^m 1_{\{\hat{j}_\eta \ge j_{\eta s}\}} \right] \end{aligned}$$

which leads to Lemma 2 for some

$$\gamma(1/8, M) \ge \frac{m}{1 + \eta/(2s)}.$$

Using Condition $F(x_0)$, Condition $H_s(M, x_0)$ and taking some $\delta_n > (2N+1)\rho_n$, it comes

$$\left(\sum_{i=1}^{n}|s_i|^2\right)^{-q/2} = \left(\frac{1}{n}\sum_{l_1,l_2=j_0-1}^{\hat{j}_\eta}\sum_{k_1,k_2}\prod_{p=1}^{2}\psi_{l_pk_p}(x_0)\left(\int_0^1\prod_{p=1}^{2}\psi_{l_pk_p}(u)du\right)\right.$$
$$\left.\times\left(1_{\{\hat{j}_\eta > j_{\eta s}-\delta_n\}} + 1_{\{\hat{j}_\eta \le j_{\eta s}-\delta_n\}}\right) + R_n\right)^{-q/2}$$
$$\le C\left[\left(\frac{2^{j_{\eta s}-\delta_n}}{n}\right) + \left(\frac{2^{j_0}}{n}\right)1_{\{\hat{j}_\eta \le j_{\eta s}-\delta_n\}} + R_n\right]^{-q/2}$$
$$\le C\left[\left(\frac{2^{j_{\eta s}-\delta_n}}{n}\right)^{-q/2} 1_{\{\hat{j}_\eta \le j_{\eta s}\}}1_{\{\hat{j}_\eta > j_{\eta s}-\delta_n\}}\right.$$
$$+\left(\left(\frac{2^{j_{\eta s}-\delta_n}}{n}\right) - \left(\frac{2^{j_\infty}}{n}\right)^2\right)^{-q/2} 1_{\{\hat{j}_\eta > j_{\eta s}\}}1_{\{\hat{j}_\eta > j_{\eta s}-\delta_n\}}$$
$$+\left(\left(\frac{2^{j_0}}{n}\right) - \left(\frac{2^{j_{\eta s}}}{n}\right)^2\right)^{-q/2} 1_{\{\hat{j}_\eta \le j_{\eta s}\}}1_{\{\hat{j}_\eta \le j_{\eta s}-\delta_n\}}$$
$$\left.+\left(\left(\frac{2^{j_0}}{n}\right) - \left(\frac{2^{j_\infty}}{n}\right)^2\right)^{-q/2} 1_{\{\hat{j}_\eta > j_{\eta s}\}}1_{\{\hat{j}_\eta \le j_{\eta s}-\delta_n\}}\right]$$

Hence, if $\hat{j}_\eta \le j_{\eta s}$ and $\hat{j}_\eta > j_{\eta s} - \delta_n$, we get

$$\left(\sum_{i=1}^{n}|s_i|^2\right)^{-q/2} \le C\left(\frac{2^{j_{\eta s}-\delta_n}}{n}\right)^{-q/2}.$$

Moreover, applying Property 3 for some

$$\gamma(1/4, M) \wedge \gamma(1/8, M) \ge q/2\left(\frac{1}{1+\eta/(2s)} \vee \frac{1}{1+2s/\eta}\right),$$

we obtain

$$E_2\left(\sum_{i=1}^{n}|s_i|^2\right)^{-q/2} \le C\left(\frac{2^{j_{\eta s}-\delta_n}}{n}\right)^{-q/2}.$$

Because of definition (7) of $b(n)$, it ends the proof of Lemma 1 in the case of the regression model.

APPENDIX C: PROOF OF THE EXPONENTIAL INEQUALITIES

Let $\gamma(C_0, M)$ be a positive function satisfying:

(i) in the case of the local thresholding, white noise model:

$$\gamma(C_0, M) \le (MC_0)^2/(2\sigma_2^2);$$

(ii) in the case of the local thresholding, regression model:

$$\gamma(C_0, M) \le \inf(C_0^2 M^2/(8N\sigma_2^2||\psi||_\infty^2), C_0 M/(4||\psi||_\infty \mu));$$

(iii) in the case of blocks thresholding, if $\log(n)^{p/2} l_j^{-1} = O(1)$, white noise model:

$$\gamma(C_0, M) \le (C_0 M - \sigma_p)^2/(2\sigma_2^2);$$

(iv) in the case of block thresholding, regression model with Gaussian error:

$$\gamma(C_0, M) \le (MC_0 - \sigma_p||\psi||_\infty)^2/(2\sigma_2^2||\psi||_\infty^2);$$

(v) in the case of block thresholding, regression model with bounded error:

$$\gamma(C_0, M) \le \inf\left(T(MC_0 - \tilde{R}(p))^2/(\sigma_2^2||\psi||_\infty^2), T(MC_0 - \tilde{R}(p))/(|\varepsilon|_\infty||\psi||_\infty)\right).$$

C.1. *Proof of Proposition* 1 *in the case of the local thresholding.*

C.1.1. *White noise model.* We just have to use the following inequality:

$$\forall\, \lambda > 0, \quad P(|\eta| \ge \lambda) \le 2\exp -\frac{\lambda^2}{2},$$

where η is a standard Gaussian variable. Then Proposition 1 is true for $0 < \gamma(C_0, M) \le (MC_0)^2/2$.

C.1.2. *Regression model.* Let us remark that

$$\begin{aligned}\hat{\beta}_{jk} - \beta_{jk} &= \left(\hat{\beta}_{jk} - E\hat{\beta}_{jk}\right) - \left(E\hat{\beta}_{jk} - \beta_{jk}\right)\\ &= \left(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \psi_{jk}(\frac{i}{n})\right) + \left(\frac{1}{n}\sum_{i=1}^{n} f(\frac{i}{n})\psi_{jk}(\frac{i}{n}) - \int_0^1 f(u)\psi_{jk}(u)du\right).\end{aligned}$$

First, we bound the bias term

$$\begin{aligned}|E\hat{\beta}_{jk} - \beta_{jk}| \le &\sum_{i=1}^{n}\int_{\frac{i-1}{n}}^{\frac{i}{n}} |\psi_{jk}(u)||f(u) - f\left(\frac{i}{n}\right)|du\\ &+ \sum_{i=1}^{n}\int_{\frac{i-1}{n}}^{\frac{i}{n}} |f\left(\frac{i}{n}\right)||\psi_{jk}(u) - \psi_{jk}\left(\frac{i}{n}\right)|du\\ &\le n^{-(s\wedge 1)}||f||_{s\wedge 1,\infty,\infty} 2^{-j/2}||\psi||_\infty\end{aligned}$$

$$+||f||_\infty 2^{j/2}||\psi'||_\infty \frac{2^j}{n} n^{-1} \# \left\{ i=1,\dots,n, \psi_{jk}\left(\frac{i}{n}\right) \neq 0 \right\}$$

$$\leq C\left(n^{-(s\wedge 1)} 2^{-j_0/2} + \frac{2^{j_1/2}}{n} \right) \leq C\sqrt{\frac{\log n}{n}} \left(n^{\frac{N}{1+2N}-(s\wedge 1)} + (\log n)^{-1} \right)$$

because the number of terms in the sum is

$$m = \# \left\{ i=1,\dots,n,\ \psi_{jk}\left(\frac{i}{n}\right) \neq 0 \right\} \leq 2N2^{-j}n.$$

Hence, this term is negligible as soon as $s \geq N/(1+2N)$. For the stochastic term, we apply the following inequality to

$$Y_i = \psi_{jk}\Big(\frac{i}{n}\Big)\varepsilon_i.$$

BERNSTEIN'S INEQUALITY [Petrov (1995), page 57]. *Let* $Y_1,\dots,Y_m$ *be* m *independent random variables such that* $EY_i = 0$, $EY_i^2 \leq \mu_i^2 < \infty$ $(i - 1,\dots m)$, $B = \sum_1^m \mu_i^2$. *Suppose there exists a positive constant C such that*

$$E|Y_i|^q \leq \frac{q!}{2}\,\mu_i^2 C^{p-2}, \quad i=1,\dots,m$$

for all integers $q \geq 2$. *Then*

$$\forall \lambda \geq \frac{B}{C}, \quad P\left(|\sum_{i=1}^m Y_i| \geq \lambda \right) \leq 2\exp\Big(-\frac{\lambda}{4C}\Big),$$

$$\forall 0 \leq \lambda \leq \frac{B}{C}, \quad P\left(|\sum_{i=1}^m Y_i| \geq \lambda \right) \leq 2\exp\Big(-\frac{\lambda^2}{4B}\Big).$$

Here, the quantities μ_i, m and B are easily evaluable

$$\mu_i^2 \leq ||\psi||_\infty^2\ \sigma_2^2\ 2^j, \quad m \leq 2Nn2^{-j}, \quad B \leq 2N||\psi||_\infty^2\ \sigma_2^2\ n\,.$$

Because of the assumption on the $q-$moments of the ε's, we take

$$C = ||\psi||_\infty \mu\ 2^{j/2}.$$

Then Proposition 1 is true for

$$0 < \gamma(C_0, M) \leq \inf\left(\frac{C_0^2 M^2}{8N||\psi||_\infty^2\ \sigma_2^2}, \frac{C_0 M}{4||\psi||_\infty \mu} \right).$$

C.2. *Proof of Proposition* 1 *in the other cases.*

C.2.1. *White noise model.* The duality equality

$$\sup_{\left\{a_{jk}, \sum_{(K)} |a_{jk}|^{p'} \leq 1\right\}} \sum_{(K)} a_{jk}\varepsilon_{jk} = \left(\sum_{(K)} |\varepsilon_{jk}|^p \right)^{1/p}, \qquad \frac{1}{p'} + \frac{1}{p} = 1 \tag{35}$$

allows us to apply the following proposition [cf. C'irelson, Ibragimov and Sudakov (1976)].

CIRES'LON-IBRAGIMOV-SUDAKOV INEQUALITY. *Let* $(\eta_t, t \in T)$ *be a Gaussian process. Let* H *and* V *be defined by*

$$H = E \sup_{t \in T} \eta_t, \qquad V = \sup_{t \in T} \operatorname{Var}(\eta_t).$$

Then

$$\forall \lambda > 0, \quad P\left(\sup_{t \in T} \eta_t \geq H + \lambda\right) \leq \exp -\frac{\lambda^2}{2V}.$$

First we have to evaluate H. By equality (35) and Jensen's inequality, we have

$$H = E\left(\sum_{(K)} |\varepsilon_{jk}|^p\right)^{1/p} \leq \left(\sum_{(K)} E|\varepsilon_{jk}|^p\right)^{1/p} = \sigma_p\, l_j^{1/p}.$$

Next, we calculate

$$V = \sup_{\left\{a_{jk}, \sum_{(K)} |a_{jk}|^{p'} \leq 1\right\}} \operatorname{Var}\left(\sum_{(K)} a_{jk}\varepsilon_{jk}\right) = \sup_{\left\{a_{jk}, \sum_{(K)} |a_{jk}|^{p'} \leq 1\right\}} \sigma_2^2 \sum_{(K)} a_{jk}^2 \leq \sigma_2^2.$$

Under the constraint $C_0 M - \sigma_p > 0$, we take $\lambda = (C_0 M - \sigma_p) l_j^{1/p}$. Hence, Proposition 1 is true for

$$0 < \gamma(C_0, M) \leq \frac{(C_0 M - \sigma_p)^2}{2\sigma_2^2} \quad \text{if } \log(n)^{p/2} l_{j_0}^{-1} = O(1),$$

$$0 < \gamma(C_0, M) \quad \text{if } \log(n)^{p/2} l_{j_0}^{-1} = o(1).$$

C.2.2. *Regression model.* In the same way as in the previous section (case of the local thresholding in the regression model), we decompose $\hat{\beta}_{jk} - \beta_{jk}$ into a bias term and a stochastic term. We bound the bias term as before as soon $s > N/(1 + 2N)$. For the stochastic term, we consider two different cases: the case where the $\varepsilon's$ are Gaussian and the case where they are bounded.

(i) The Gaussian case is close to the previous section: we only need to replace ε_{jk} by $n^{-1/2} \sum_{i=1}^n \psi_{jk}(i/n)\varepsilon_i$. This leads to bound H by $||\psi||_\infty l_j^{1/p}$ and V by $\sigma_2^2 ||\psi||_\infty^2$. Hence, under the constraint $C_0 M - \sigma_p ||\psi||_\infty > 0$, Proposition 1 is true for

$$0 < \gamma(C_0, M) \leq \frac{(C_0 M - \sigma_p ||\psi||_\infty)^2}{2\sigma_2^2 ||\psi||_\infty^2} \quad \text{if } \log(n)^{p/2} l_{j_0}^{-1} = O(1),$$

$$0 < \gamma(C_0, M) \quad \text{if } \log(n)^{p/2} l_{j_0}^{-1} = o(1).$$

(ii) If the ε's are bounded, we use the duality equality with $1/p + 1/p' = 1$

$$\sup_{\{a_{jk}, \sum_{(K)} |a_{jk}|^{p'} \leq 1\}} n^{-1} \sum_{(K)} a_{jk} \sum_{i=1}^{n} \psi_{jk}(\frac{i}{n})\varepsilon_i = \left(\sum_{(K)} |\hat{\beta}_{jk} - E\hat{\beta}_{jk}|^p \right)^{1/p} \tag{36}$$

and apply the following inequality to

$$Y_i = \psi_{jk}\left(\frac{i}{n}\right)\varepsilon_i.$$

TALAGRAND'S THEOREM [cf. Talagrand (1994)]. *Let $Y_1, \ldots, Y_n$ be n independent identically distributed random variables. Let $\mathscr{F}$ be a family of uniformly bounded (by some constant B) functions. Let H and V be defined by*

$$H \geq E\left(\sup_{h \in \mathscr{F}} |\sum_{i=1}^{n} h(Y_i)| \right), \qquad V = \sup_{h \in \mathscr{F}} \mathrm{Var}(h(Y_1)).$$

Then there exists a universal constant T such that

$$\forall \lambda > 0, \quad P\left(\sup_{h \in \mathscr{F}} \sum_{i=1}^{n} (h(Y_i) - Eh(Y_i)) \right) \geq \lambda + H) \leq C\ \exp\left(-T(\frac{\lambda^2}{nV} \wedge \frac{\lambda}{B}) \right).$$

Under the assumption that the ε's are bounded by $|\varepsilon|_\infty$, we get

$$B \leq 2^{j/2} |\varepsilon|_\infty ||\psi||_\infty.$$

As previously, the evaluation of V is bounded by

$$V \leq ||\psi||_\infty^2 \sigma_2^2.$$

In view to evaluate H, we use equality (36) and Jensen's inequality

$$H = E\left(\sum_{(K)} |\hat{\beta}_{jk} - E\hat{\beta}_{jk}|^p \right)^{1/p} \leq \left(\sum_{(K)} E|\hat{\beta}_{jk} - E\hat{\beta}_{jk}|^p \right)^{1/p}.$$

Let us now apply the following inequality.

ROSENTHAL'S INEQUALITY [cf. Petrov (1995), page 59]. *Let $Y_1, \ldots, Y_m$ be m independent centered random variables having moments up to $p \geq 2$. Then there exists some constant $R(p)$ such that*

$$E\left| \sum_{1}^{m} Y_i \right|^p \leq R(p)^p \left(\sum_{1}^{m} E|Y_i|^p + \left(\sum_{1}^{m} EY_i^2 \right)^{p/2} \right).$$

We obtain for all indices j, n and K,

$$H^p \leq (\#K)\ E|\sum_{1}^{m} Y_i|^p \leq \tilde{R}^p(p)\, l_j\, n^{p/2}.$$

where

$$\tilde{R}(p) = 2^{1/p}(2N)^{1/2}||\psi||_\infty|\varepsilon|_\infty R(p).$$

Applying now Talagrand's theorem, under the constraint $C_0M - \tilde{R}(p) > 0$, we take $\lambda = (C_0M - \tilde{R}(p))l_j^{1/p}n^{1/2}$. Hence, Proposition 1 is true for

$$0 < \gamma(C_0, M) \le \inf\left(T\frac{(C_0M - \tilde{R}(p))^2}{\sigma_2^2||\psi||_\infty^2}, \ T\frac{C_0M - \tilde{R}(p)}{|\varepsilon|_\infty||\psi||_\infty}\right).$$

Acknowledgments. The authors would like to thank the referees and an Associate Editor for their critical remarks that led to a considerable improvement of this paper.

REFERENCES

ARNEODO, A., JAFFARD, S., LEVY-VEHEL, J. and MEYER, Y. (1997). Méthodes d'ondelettes pour l'analyse de fonctions fractales et multifractales. Conférence CNRS ENS Cachan.

BOWMAN, A. W. and HARDLE, W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.* **83** 102–110.

C'IRELSON, B. S., IBRAGIMOV, I. and SUDAKOV, V. (1976). Norm of Gaussian sample functions. In *Proceedings of the 3rd Japan-URSS Symposium on Probability Theory* (G. Maruyama amd J.V. Prohorov, eds.) Lecture Notes in Math. **550** 20–41. Spring, New York.

DAUBECHIES, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia.

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. AND PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539.

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. AND PICARD, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). *J. Roy. Statist. Soc. Ser. B* **57** 301–369.

FARAWAY, T. (1990). Bootstrap selection of bandwidth and confidence bands for nonparametric regression. *J. Statist. Comput. Simulation* **37** 37–44.

FELLER, W. (1966). *An introduction to Probability Theory and Its Applications.* Wiley, New York.

HALL, P. (1991). Edgeworth expansions for nonparametric density estimators, with applications. *Statistics* **22** 215–232.

HALL, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.* **20** 675–694.

HALL, P., KERKYACHARIAN, G. AND PICARD, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods. *Ann. Statist.* **26** 922–642.

IOUDITSKI, A. AND LEPSKII, O. V. (1997). Personal communication.

KERKYACHARIAN, G., PICARD, D. AND TRIBOULEY, K. (1996). L_p adaptive density estimation. *Bernoulli* **2** 229–247.

LEPSKII, O. V. (1990). On a problem of adaptive estimation in white Gaussian noise. *Teor. Veoryatnost. i Primenen.* **35** 459–470 (in Russian). [Translated in (1991) *Theory Probab. Appl.* **35** 454–466.]

LEPSKII, O. V. (1991). Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** 682–697.

LEPSKII, O. V., MAMMEN, E. AND SPOKOINY, V. G. (1994). Ideal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selection. Technical report.

MEYER, Y. (1990). *Ondelettes.* Hermann, Paris

NEUMANN, M. (1995). Automatic bandwidth choice and confidence intervals in nonparametric regression. *Ann. Statist.* **23** 1937–1959.

PETROV, V. V. (1995) *Limit Theorems of Probability Theory.* Clarendon Press, Oxford.
RAIMONDO, M. (1998). Minimax estimation of sharp change points. *Ann. Statist.* **26** 1379–1397.
TALAGRAND, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22** 28–76.

LABORATOIRE DE PROBABILITÉS ET MODÈLES ALÉATOIRES
CNRS-UMR 7599
UNIVERSITÉ PARIS VI ET UNIVERSITÉ PARIS VII
4, PLACE JUSSIEU
TOUR 56 3EME ÉTAGE, F-75252
PARIS CEDEX 05
FRANCE

URA CNRS 743
UNIVERSITÉ PARIS XI
BÂTIMENT 425
91405 ORSAY CEDEX
FRANCE
E-MAIL: Karine@Tribouley
math.u-psud.fr