Brazilian Journal of Probability and Statistics
2016, Vol. 30, No. 4, 614–652
DOI: 10.1214/15-BJPS296

Wavelet shrinkage for regression models with random design and correlated errors

Rogério Porto[a], Pedro Morettin[b], Donald Percival[c] and Elisete Aubin[b]

[a]*Bank of Brazil*
[b]*University of São Paulo*
[c]*University of Washington*

Abstract. Extraction of a signal in the presence of stochastic noise via wavelet shrinkage has been studied under assumptions that the noise is independent and identically distributed (IID) and that the samples are equispaced (evenly spaced in time). Previous work has relaxed these assumptions either to allow for correlated observations or to allow for random sampling, but very few papers have relaxed both together. In this paper we relax both assumptions by assuming the noise to be a stationary Gaussian process and by assuming a random sampling scheme dictated either by a uniform distribution or by an evenly spaced design subject to jittering. We show that, if the data are treated as if they were autocorrelated and equispaced, the resulting wavelet-based shrinkage estimator achieves an almost optimal convergence rate. We investigate the efficacy of the proposed methodology via simulation studies and illustrate it by the extraction of the light curve for a variable star.

1 Introduction

A mathematical problem of considerable interest is to approximate a continuous function $f(t)$, $t \in [0, 1]$, based upon samples $f(t_i)$, $i = 1, \dots, n$. We do not observe $f(t_i)$ directly, but only in the presence of correlated zero mean noise $\{\varepsilon(t_1), \dots, \varepsilon(t_n)\}$, which we assume throughout to obey a multivariate Gaussian distribution. The data are $\{(t_1, y(t_1)), \dots, (t_n, y(t_n))\}$, where $y(t_i) = f(t_i) + \varepsilon(t_i)$, for $i = 1, \dots, n$, and our objective is to extract the signal f from the data using an estimator $\hat{f}$ with low integrated mean squared error (IMSE), defined as

$$E\|\hat{f} - f\|_2^2 = \int_0^1 E\big(\hat{f}(x) - f(x)\big)^2\,dx.$$

Wavelet shrinkage methods have been very successful in signal extraction and nonparametric regression, but most methods are focused on a regular design (i.e., equispaced samples over a regular grid $t_i = i/n$) with independent and identically distributed (IID) errors. The assumption of a regular design has been relaxed to handle unequally spaced samples with either a fixed design (Cai and Brown,

Key words and phrases. Autocorrelation, denoising, semi-parametric estimation, smoothing, wavelets, non-parametric regression.
Received June 2014; accepted July 2015.

1998), a uniformly distributed design (Cai and Brown, 1999) or a general random design (Antoniadis and Pham, 1998, Sardy et al., 1999, Delouille, Simoens and von Sachs, 2004, Kerkyacharian and Picard, 2004), but these extensions are restricted to IID errors. More recent work considering random design and IID errors include those from Chesneau (2007), Gaïffas (2009) and Antoniadis, Pensky and Sapatinas (2014). Wavelet shrinkage methods have also been adapted to handle correlated errors, in the context of equispaced samples (Neumann and von Sachs, 1995, Johnstone and Silverman, 1997, von Sachs and Macgibbon, 2000) and of unequally spaced samples with a fixed design (Porto, Morettin and Aubin, 2008). In this paper, we call "design" the set of values $\{t_1, \ldots, t_n\}$. The design can be fixed or random, depending on their values being deterministic or realizations from a random variable. The design can also be regular or irregular depending on whether the difference between any two successive ordered values is constant (equally spaced) in the entire set or not (unequally spaced).

In this paper, we investigate wavelet shrinkage for certain unequally sampled designs in the presence of correlated errors. Although our main result is valid for very general sampling schemes and correlation structure, the random sampling schemes that we explicitly consider are stochastic, where either the sample points t_i are uniformly distributed in $[0, 1]$ or they come from a jittering; that is, $t_i = (2i - 1)/(2n) + j_i$, where j_i are IID uniform $[-1/(2n), 1/(2n)]$ random variables. Stochastic sampling techniques are of interest because they can overcome certain aliasing problems associated with a regular design (Dippé and Wold, 1985). We show that under our assumptions the samples can be treated as if they were equispaced with correlated noise (Johnstone and Silverman, 1997), and hence we can apply the VisuShrink procedure (Donoho and Johnstone, 1994) with level-dependent thresholds.

Two recent papers also address unequally sampled designs with correlated errors using wavelets. Delouille and von Sachs (2005) deal with correlated observations and random design, but use a totally different model (a non-parametric autoregression). The paper by Kulik and Raimondo (2009) is more closely related to our work and is in two ways more general than ours, but with one key difference. The functions they consider belong to Besov classes (we focus on more restrictive Hölder classes), with errors that can be drawn from weak- or strong-dependent stochastic processes (we only consider weak dependence); however, they impose a correlation structure prior to a random distribution of design points, rather than *after* specifying the points, as we do in this paper (see Corollary 1 in Section 3 for details). While mathematically interesting, their approach is not in keeping with traditional time series models, whereas ours is. Our approach is much more closely related to the irregularly observed processes discussed in Brillinger (1996) in the context of linear estimators.

In the literature, general nonparametric regression methods with correlated errors have been studied, but also with settings different from ours. For instance, the works of Efromovich (1999) and Yang (2001) differ from ours for the same reasons as cited for Kulik and Raimondo (2009). Similarly our work differs from those of Hofmann (1999) and Baraud, Comte and Viennet (2001) for the same reasons as cited for Delouille and von Sachs (2005).

The paper is organized as follows. In Section 2, we review some basic properties of wavelets along with earlier research on wavelet shrinkage that we exploit within the text. Our new results on wavelet shrinkage for stochastic sampling schemes with correlated errors are given in Section 3, after which we present some simulation results and apply the methodology to an unequally sampled series of magnitude measurements for a variable star. All the proofs are given in detail in the last section, after the summary and a discussion on results for situations more complex than those considered here.

2 Wavelets and wavelet shrinkage

Consider an orthonormal wavelet basis generated from dilation and translation of a "father" wavelet ϕ (or scaling function) and a "mother" wavelet ψ. We assume that both functions are compactly supported in $[0, N]$ and $[(1-N)/2, (1+N)/2]$ respectively, $\int \phi = 1$, $\int \psi = 0$ and ψ has r vanishing moments. Let

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k) \quad \text{and} \quad \psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$$

so that $\psi_{j,k}$ has support $[2^{-j}((1-N)/2+k), 2^{-j}((1+N)/2+k)]$. For $t \in [0, 1]$, let

$$\phi^p_{j,k}(t) = \sum_{l\in\mathbb{Z}} \phi_{j,k}(t-l) \quad \text{and} \quad \psi^p_{j,k}(t) = \sum_{l\in\mathbb{Z}} \psi_{j,k}(t-l)$$

denote the periodized wavelets, which we use henceforth, but with the superscript "p" suppressed, since it is a standard way of handling boundary conditions even if the signal is not regarded as periodic (see e.g. Ogden (1997), for details). For some coarse scale $j_0 \geq 0$, the collection formed by $\phi_{j_0,k}$, $k = 0, \ldots, 2^{j_0} - 1$, and $\psi_{j,k}$, $j \geq j_0$, $k = 0, \ldots, 2^j - 1$, constitutes an orthonormal basis of $L_2[0, 1]$ (see, e.g., Härdle et al. (1998), for details).

Denote the inner product by $\langle\cdot, \cdot\rangle$. For a given square-integrable function f on $[0, 1]$, let

$$c_{j,k} = \langle f, \phi_{j,k}\rangle \quad \text{and} \quad d_{j,k} = \langle f, \psi_{j,k}\rangle.$$

The function f can be expanded into a wavelet series as

$$f(t) = \sum_{k=0}^{2^{j_0}-1} c_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(t).$$

This expansion decomposes f into components with different resolutions. The coefficients $c_{j_0,k}$ at the coarsest level capture the gross structure of the function f. The detail coefficients $d_{j,k}$ represent finer and finer structures in f as the resolution level j increases.

2.1 Regular design with IID errors

Suppose we have data sampled on a regular grid that obeys the model

$$y_i = f\left(\frac{i}{n}\right) + e_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

where the noise e_i is drawn from some stochastic process. Our task is to formulate an estimator $\hat{f}$ of f with small IMSE. In practice, we do this by transforming y_i into empirical wavelet coefficients and then defining $\hat{f}$ in terms of the inverse transform of wavelet coefficients that have been denoised using wavelet shrinkage. The most widely used shrinkage method is the VisuShrink procedure (Donoho and Johnstone, 1994) described as follows.

An orthonormal wavelet basis has an associated exact orthogonal discrete wavelet transform W that transforms sampled data into discrete wavelet coefficients. Let $y = (y_1, \ldots, y_n)^T$ be the vector of observations, where $n = 2^J$ for some $J \in \mathbb{N}$, and let

$$\tilde{\theta} = Wy = (\tilde{c}_{j_0,0}, \ldots, \tilde{c}_{j_0,2^{j_0}-1}, \tilde{d}_{j_0,0}, \ldots, \tilde{d}_{j_0,2^{j_0}-1}, \ldots, \tilde{d}_{J-1,0}, \ldots, \tilde{d}_{J-1,2^{J-1}-1})^T$$

be the coefficients of the discrete wavelet transform. Define the soft threshold function by

$$\eta_S(d, \lambda) = \operatorname{sgn}(d)\big(|d| - \lambda\big)_+,$$

for some threshold λ, where $x_+ = \max(x, 0)$ (the theoretical results of this paper focus on soft thresholding, but the results of this and of the following sections remain valid for hard thresholding function $\eta_H(d, \lambda) = dI(|d| \geq \lambda)$, with $I(\cdot)$ being the usual indicator function). If the errors e_i, $i = 1, \ldots, n$ are IID $N(0, \sigma^2)$ random variables with known σ^2, the VisuShrink estimator of $\{f(i/n), i = 1, \ldots, n\}$ is constructed by thresholding the wavelet coefficients $\tilde{d}_{j,k}$ with threshold $\lambda = \sigma\sqrt{2\log n}$ and then transforming back. Thus, we define

$$\hat{d}_{j,k} = \eta_S(\tilde{d}_{j,k}, \lambda)$$

and the estimator

$$\hat{f} = W^T\hat{\theta},$$

where

$$\hat{\theta} = (\tilde{c}_{j_0,0}, \ldots, \tilde{c}_{j_0,2^{j_0}-1}, \hat{d}_{j_0,0}, \ldots, \hat{d}_{j_0,2^{j_0}-1}, \ldots, \hat{d}_{J-1,0}, \ldots, \hat{d}_{J-1,2^{J-1}-1})^T.$$

In practice, the transform W and its inverse W^T are carried out by a fast $O(n)$ algorithm. Note that thresholding is restricted to levels j above some user-specified primary resolution level j_0. It is supposed that signal predominates over noise in levels below j_0.

2.2 Uniform design with IID errors

Consider the model

$$y(t_i) = f(t_i) + \varepsilon_i, \qquad i = 1, \dots, n,$$

where t_i are IID uniform [0, 1] random variables, and ε_i are IID $N(0, \sigma^2)$ variables with σ^2 known and independent of t_i. Let $0 \le t_{(1)} < t_{(2)} < \cdots < t_{(n)} \le 1$ be the order statistics of the t_i. Changing the labels accordingly to the order of the t_i, the model can be rewritten as

$$y_i = f(t_{(i)}) + e_i, \qquad i = 1, \dots, n, \tag{2.2}$$

where $y_i \equiv y(t_{(i)})$ and $e_i = y(t_{(i)}) - f(t_{(i)})$ (note that the values e_i represent a reordering of the ε_i). The data consist of observed pairs $\{(t_{(1)}, y_1), (t_{(2)}, y_2), \dots, (t_{(n)}, y_n)\}$. Because the t_i are uniformly distributed on [0, 1], the $t_{(i)}$ are distributed as Beta$(i, n - i + 1)$ and $E(t_{(i)}) = i/(n + 1)$ (Cai and Brown, 1999). Hence, in expectation this is a regular sampled design $(i/(n + 1), y_i)$, and we can apply the VisuShrink procedure directly to the data $y = (y_1, \dots, y_n)^T$. To within a logarithmic factor, this procedure achieves the optimal convergence rate over the range of Hölder classes $\Lambda^\alpha(M)$ with $1/2 \le \alpha \le r$, a result that holds for both hard and soft thresholding (Cai and Brown, 1999). In the case of random uniform design and independent Gaussian errors, the data thus can be treated as if they were sampled in a regular design. An argument in terms of an isometry involving a risk measure with a non-Euclidian norm can be used to justify this practice for other types of nonuniform sampling (Sardy et al., 1999). In fact, Kerkyacharian and Picard (2004) obtained similar asymptotic results but for very general random designs and wider function classes.

2.3 Regular design with correlated errors

Let us consider model (2.1) again, but now suppose that the error vector $e = (e_1, \dots, e_n)^T$ has a multivariate Gaussian distribution with mean 0 and covariance matrix Γ. Also, assume that the errors are stationary so that Γ has entries $\gamma(|r - s|)$, for $r, s = 1, \dots, n$. Let $z = We$ be the wavelet transform of the scaled error vector and let $V = W\Gamma W^T$ be the covariance matrix of z. Neglecting boundary effects, within each level $z_{j,k}$ will be a portion of a stationary process with level-dependent variance $\sigma_j^2 = \mathrm{Var}(z_{j,k})$ (Johnstone and Silverman, 1997), where $z_{j,k}$ is the element at level j and index k of z.

The properties of the wavelet transform have two heuristic consequences. First, for many (but not all) models encountered in practice, the autocorrelation of the $z_{j,k}$ within each level dies away rapidly. Second, there will tend to be little correlation between the wavelet coefficients at different levels (Johnstone and Silverman, 1997). For a process with positively correlated long-range dependence, the wavelet coefficients form series with negligible autocorrelation and cross-correlations.

In view of these facts, a natural extension of the VisuShrink procedure is to apply level-dependent thresholding to the transformed data $\tilde{d}_{j,k}$, $j = j_0, \dots, J-1$, $k = 0, \dots, 2^j - 1$:

$$\hat{d}_{j,k} = \eta_S(\tilde{d}_{j,k}, \lambda_j), \tag{2.3}$$

where $\lambda_j = \sigma_j \sqrt{2 \log n}$, and the estimator is

$$\hat{f} = W^T \hat{\theta},$$

with $\hat{\theta}$ given by (2.1). In practice, the noise variance σ_j^2 is often estimated from the coefficients in each level, through a robust estimator like the median absolute deviation from zero. Note that the number of coefficients at the coarsest level j_0 could be small if j_0 is set too small, resulting in dicey estimates of $\sigma_{j_0}^2$.

3 Wavelet shrinkage for random design with correlated errors

Consider a sample $(t_1, y(t_1)), (t_2, y(t_2)), \dots, (t_n, y(t_n))$ from some stochastic sampling scheme with corresponding order statistics $0 \leq t_{(1)} < t_{(2)} < \cdots < t_{(n)} \leq 1$. Given the data, assume the model

$$y_i = f(t_{(i)}) + e_i, \tag{3.1}$$

where $y_i \equiv y(t_{(i)})$ and the errors $e_i = e(t_{(i)})$ are Gaussian with mean zero, finite variance and finite covariance. Let $\hat{f}(t)$ be the estimator of $f(t)$ for all $t \in [0, 1]$, where

$$\hat{f}(t) = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k} \phi_{j_0,k}(t) + \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} \hat{d}_{j,k} \psi_{j,k}(t); \tag{3.2}$$

$\hat{c}_{j_0,k} = \tilde{c}_{j_0,k}$, $\hat{d}_{j,k}$ is given by

$$\hat{d}_{j,k} = \eta_S(\tilde{d}_{j,k}, \lambda_{j,k}), \tag{3.3}$$

where $\tilde{c}_{j_0,0}, \dots, \tilde{d}_{J'-1,2^{J'-1}-1}$ are the empirical wavelet coefficients, the threshold $\lambda_{j,k} = \sigma_{j,k} \sqrt{2 \log n}$ is used by the soft threshold function $\eta_S(\tilde{d}_{j,k}, \lambda_{j,k}) = \operatorname{sgn}(\tilde{d}_{j,k})(|\tilde{d}_{j,k}| - \lambda_{j,k})_+$, $\sigma_{j,k}^2$ is the variance of the respective (non observed) empirical wavelet coefficient of the error vector $e = (e_1, \dots, e_n)^T$, and $J' \leq J$ is any cutoff resolution level, $J' \in \mathbb{N}$.

In matrix calculations, W is neither affected by the design nor by the correlated errors and the empirical wavelet coefficients are obtained in the same usual way, ignoring the unequally spaced design, treating the data as if they were equally spaced. However, the threshold is affected by the correlated errors but not by the design.

Our results are for Hölder classes, that is, $f \in \Lambda^\alpha(M)$ which, by definition, if $0 < \alpha \leq 1$, $|f(x) - f(y)| \leq M|x - y|^\alpha$; if $\alpha > 1$, $|f^{(\lfloor\alpha\rfloor)}(x) - f^{(\lfloor\alpha\rfloor)}(y)| \leq M|x - y|^{\alpha'}$ and $|f^{(1)}(x)| \leq M$, where $f^{(m)}(x)$ is the mth derivative of the function f at x, $\lfloor\alpha\rfloor$ is the largest integer less than α, and $\alpha' = \alpha - \lfloor\alpha\rfloor$ (Cai and Brown, 1999, Definition 1). The following theorem states our main result.

Theorem 1. *Suppose that model* (3.1) *is valid, and $e_i = e(t_{(i)})$ are multivariate Gaussian noise with zero mean and finite variance. Suppose also that the "mother" wavelet ψ has r vanishing moments and is compactly supported. Then, over the range of Hölder classes $\Lambda^\alpha(M)$ with $\alpha \in (0, r]$ and $M \in (0, \infty)$, the risk of the estimator $\hat{f}$ given by* (3.2) *is such that*

$$
\begin{aligned}
E\big(\|\hat{f} - f\|_2^2\big) \leq{}& 20C_1 n^{-2s(\alpha)} + 40\frac{C^2_{M,\psi^{(H)}}}{1 - 2^{-2\alpha}}2^{-2\alpha J} \\
&+ 20\frac{M}{n}\sum_{i=1}^{n}\left\{\operatorname{Var}(t_{(i)}) + \left[E\left(t_{(i)} - \frac{i}{n}\right)\right]^2\right\}^{s(\alpha)} \\
&+ \sum_{k=0}^{2^{j_0}-1}\operatorname{Var}(\tilde{r}_{j_0,k}) \\
&+ 8\sum_{(j,k)\in\mathcal{I}_2} d^2_{j,k} + \frac{C^2_{M,\psi}}{1 - 2^{-2\alpha}}2^{-2\alpha J'} \\
&+ 6\log n \sum_{(j,k)\in\mathcal{I}_1}\sigma^2_{j,k} + \frac{2}{n}\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1}\sigma^2_{j,k},
\end{aligned}
$$

where C_1, $C_{M,\psi^{(H)}}$ and $C_{M,\psi}$ are positive constants that do not depend on n; $s(\alpha) = \min(\alpha, 1)$; $\tilde{r}_{j_0,k} = \langle R, \phi_{j_0,k}\rangle$, $R(x) = \sum_{i=0}^{n-1} e_{i+1} I((nx - i) \in (0, 1])$; $\mathcal{I}_1$ is the set of pairs of indexes (j, k) such that $6\sigma^2_{j,k}\log n < 8d^2_{j,k}$ (i.e., the signal is much stronger than the noise); $\mathcal{I}_2$ is the complement of $\mathcal{I}_1$; $j \in \{j_0, \ldots, J' - 1\}$ and $k \in \{0, \ldots, 2^j - 1\}$.

This theorem is very general, and, as special cases, we can deduce and generalize results already stated in the literature for scenarios discussed in Section 2 (and noiseless variations thereof), as follows (proofs for the theorem and the following corollaries and proposition are given in the last section).

The cutoff resolution level J' is a theoretical device used to show that one can discard some of the finest resolution levels and still achieve the almost minimax risk. This also work for theoretical comparison with other works, as we do in the comments following the corollaries. In practice, one can always consider $J' = J$ to achieve the fastest rate of decay but sometimes using $J' < J$ can lead to a better estimator, that is, the same rate of decay but with smaller constants. Selecting the best value J' is not trivial since it depends on α which is usually unknown. For practical situations, we advocate to consider $J' = J$, try some $J' < J$ and make a decision based on the researcher's experience or comparison with other studies.

In what follows, we refer to the following three designs, already cited in Section 1:

1. regular: the points t_i are deterministic and equally sampled, that is, $t_i = t_{(i)} = i/n$;
2. uniform: the n random points t_i are IID uniform $[0, 1]$ random variables, which implies that $t_{(i)} \sim \text{Beta}(i, n - i + 1)$;
3. jittered: the random points are jittered, that is, $t_i = t_{(i)} = (2i - 1)/(2n) + j_i$, where the j_i are IID uniform $[-1/(2n), 1/(2n)]$ random variables.

Corollary 1 (Uniform or jittered design with correlated errors). *Suppose the conditions of Theorem 1 but that the noise is a portion of a Gaussian stochastic process with mean zero, finite variance* $\gamma(0)$, *and finite covariances* $\text{Cov}(e_i, e_j)$, *for* $i, j = 1, \dots, n$. *Suppose also that the stochastic process is stationary and short-memory, where* $\text{Cov}(e(t_{(i)}), e(t_{(j)})) = \gamma(|i - j|)$ *and* $\lim_{n\to\infty} \sum_{u=-(n-1)}^{n-1} |\gamma(u)| = C_\gamma < \infty$, *and that the design is either uniform or jittered. Then the fastest decay of the risk is the same as stated in Corollary 3.*

This result is related to Theorem 4.8 in Kulik and Raimondo (2009). It is less general since they consider Besov classes; however, it is more general since they only consider $\alpha > 1/2$ and known (or estimated) distribution function that generates the design, while we require only its two first moments.

Corollary 2 (Regular design with correlated errors). *Suppose the conditions of Theorem 1 but that the noise is a portion of a Gaussian stochastic process with mean zero, finite variance* $\gamma(0)$, *and finite covariances* $\text{Cov}(e_i, e_j)$, *for* $i, j = 1, \dots, n$. *Suppose also that the design is regular and the stochastic process is stationary and short-memory, where* $\text{Cov}(e(t_{(i)}), e(t_{(j)})) = \gamma(|i - j|)$ *and* $\lim_{n\to\infty} \sum_{u=-(n-1)}^{n-1} |\gamma(u)| = C_\gamma < \infty$. *Then, the fastest decay of the risk is the same as stated in Corollary 4.*

This result is closely related to Theorem 3 in Johnstone and Silverman (1997). It is less general in that they consider both short- and long-memory processes and functions in Besov classes; on the other hand, it is more general in that they only consider $\alpha > 1/2$.

Corollary 3 (Uniform or jittered design with IID errors). *Suppose the conditions of Theorem 1 but that the noise is IID normal with mean zero and finite variance* $\gamma(0)$. *Suppose also that the design is either uniform or jittered. Then the fastest decay of the risk is achieved as*

1. $E(\|\hat{f} - f\|_2^2) = O((\log n/n)^{2\alpha/(1+2\alpha)})$ *for all* $J' \in \{J_1, \dots, J\}$ *if* $1/2 - \log\log n/\log n \leq \alpha \leq r$, *where* $2^{J_1} = O((n/\log n)^{1/(1+2\alpha)})$;

2. $E(\|\hat{f} - f\|_2^2) = O((\log n/n)^{2\alpha/(2+2\alpha)})$ *for all* $J' \in \{J_2, \dots, J\}$ *if* α *is such that* $-\log\log n/\log n \leq \alpha < 1/2 - \log\log n/\log n$, *where* $2^{J_2} = O((n/\log n)^{1/(2+2\alpha)})$;

3. $E(\|\hat{f} - f\|_2^2) = O(n^{-s(\alpha)})$ *when* $J' = J$ *if* $0 < \alpha < -\log\log n/\log n$.

This result expands upon Cai and Brown (1999), which considers only $\alpha \geq 1/2$ and $J' = J$. This result is also closely related to those in Kerkyacharian and Picard (2004). It is less general because they consider functions in Besov classes, but it is more general because they only consider $\alpha > 1/2$, $2^{J'} = \sqrt{n/\log n} \leq 2^{J_1}$ and full knowledge of the design distribution function, while we require only its two first moments.

Corollary 4 (Regular design with IID errors). *Suppose the conditions of Theorem 1 but that the noise is IID normal with mean zero and finite variance* $\gamma(0)$. *Suppose also that the design is regular. Then, the fastest decay of the risk is achieved for all* $J' \in \{J_1, \dots, J\}$, *where* $2^{J_1} = O((n/\log n)^{1/(1+2\alpha)})$, *as*

$$E(\|\hat{f} - f\|_2^2) = O((\log n/n)^{2\alpha/(1+2\alpha)}).$$

This result is classic and can be found in Donoho et al. (1995) and Donoho and Johnstone (1996), for instance.

Corollary 5 (Noiseless observations, uniform or jittered design). *Suppose the conditions of Theorem 1 but that the noise is degenerate (i.e., the variance and covariances are zero) and the design is either uniform or jittered. Then, in both cases, the fastest decay of the risk is achieved when* $J' = J$ *as*

$$E(\|\hat{f} - f\|_2^2) = O(n^{-s(\alpha)}).$$

Here the risk converges to zero slower than in the regular design. The result for the uniform design is already shown in Cai and Brown (1999), at Lemma 3. Although not widely used in statistics, these results can be useful for analyzing antialiasing techniques in signal processing as, for instance, those in Dippé and Wold (1985).

Corollary 6 (Noiseless observations, regular design). *Suppose the conditions of Theorem 1 but that the noise is degenerate (i.e., the variance and covariances are zero) and the design is regular. Then, the fastest decay of the risk is achieved when* $J' = J$ *as*

$$E(\|\hat{f} - f\|_2^2) = O(n^{-2s(\alpha)}).$$

In this case, the risk is exactly the usual approximation error. This result is already known (see, e.g., Lemma 2 and Theorem 1 in Cai and Brown (1998)).

Since Theorem 1 is very general, in practice we use it through its corollaries which, in the case of those correlated errors, use one threshold for each resolution level, as cited in Section 2.3. The proofs of Lemmas 6 and 8 (see the last section) show the threshold affect only $\mathcal{I}_1$ and, consequently, $\mathcal{I}_2$. The proofs of the corollaries for correlated errors show that $\mathcal{I}_1$ can be determined only by a set of resolution levels (do not consider k). Thus, using one threshold for each resolution level is enough in the simulations and the application (where we do not reject the null hypothesis of uniform design).

The noise conditions of Corollaries 2 and 1 occur in diverse applications (see, e.g., Cochrane and Orcutt (1949), Kutner et al. (2004), Qin and Gilbert (2001)), and specific cases of interest where the design conditions of Corollaries 3 and 1 also occur are given by the following proposition.

Proposition 1. *Suppose that model* (3.1) *holds. Let* $\{e_i = e(t_{(i)}), i = 1, \ldots, n\}$ *be a portion of a continuous-time zero-mean stationary process* $e(t)$, *where* $t_{(i)}$ *comes from either a uniform or jittered design. Let*

$$\mathrm{Cov}\big(e(t_{(i)}), e(t_{(j)})\big) = E\big(\sigma^2 e^{-n'\beta|t_{(i)}-t_{(j)}|}\big)$$

for some $\beta > 0$, $0 < \sigma^2 < \infty$ *and fixed* i *and* j, *where* $n' = n + 1$ *for the uniform design and* $n' = n$ *for the jittered design. Then* $\mathrm{Cov}(e(t_{(i)}), e(t_{(j)})) = \gamma(|i - j|)$ *and* $\lim_{n\to\infty} \sum_{u=-(n-1)}^{n-1} |\gamma(u)| = C_\gamma < \infty$.

Two remarks are in order here. First, a sufficient condition for the noise conditions of Corollary 1 to hold is that

$$\big|\mathrm{Cov}\big(e(t_{(i)}), e(t_{(j)})\big)\big| \le C\sigma^2 e^{-\beta|i-j|}$$

for some positive constant $C < \infty$. This exponential decay condition seems to be stronger than the convergence of the sum of covariances but we leave the weakening of this condition for further research. Second, the covariances we assume in the proposition are similar to that for a continuous-time first-order autoregressive (AR(1)) process, but not exactly so. We are essentially mapping a process on the real axis to the $(0, 1)$ interval, so the correlation between two fixed points in this interval must decrease as the sample size increases, whereas it would remain fixed for a true AR(1) process.

We must note that the condition in the covariance matrix is not assumed in the proofs. We only state it is a sufficient condition with the following practical appeal. Suppose we want to estimate the function with an almost minimax risk but our error process is such that the condition is not met for, say, the pair of errors $e(0.42)$ and $e(0.48)$ from a sample of size $n = 100$ using the 42th and 48th observations, that is, at $t_{(42)} = 0.42$ and $t_{(48)} = 0.48$ the absolute value of the covariance between the respective pair of errors is greater than $C\sigma^2 e^{-\beta|48-42|}$. However, the sufficient condition (an upper bound, in fact) says that a sample of size $n = 1000$ could be enough if $|\mathrm{Cov}(e(0.42), e(0.48))| \le C\sigma^2 e^{-\beta|480-420|}$ because there are more observations between these two time points which would weaken the respective correlation. Notice that the upper bound for the pair $e(0.42)$ and $e(0.48)$ with $n = 100$ is the same for the pair $e(0.042)$ and $e(0.048)$ with $n = 1000$ since both are based on $t_{(42)}$ and $t_{(48)}$.

4 Simulations

We conducted a simulation study to compare the estimator based on unequally spaced samples (with uniform and jittered samples) with the estimator based on a regular design. The package Wavethresh (Nason, Kovac and Maechler, 2006), implemented in R language (R Core Team, 2013), was used (code needed to replicate the results can be obtained from the first author upon request).

We considered three test functions $f(t)$, representing different degrees of spatial variability: sine, Heavisine and Doppler. The formulas for the last two functions are given by Donoho and Johnstone (1994). The sampled functions were normalized such that their standard deviations are equal to 10; that is,

$$\mathrm{SD}_{\mathrm{signal}} = \left(\frac{1}{n-1}\sum_{i=1}^{n}[f(t_i) - \bar{f}]^2\right)^{1/2} = 10,$$

where $\bar{f} = n^{-1}\sum_i f(t_i)$. We generated three samples of noise, one for each type of design, from the process described at Proposition 1 with $\beta = -\log(0.7)$. For the regular design, this corresponds to a discrete-time AR(1) process with coefficient $\phi = 0.7$. Then, the noise samples were standardized to achieve a signal-to-noise ratio (SNR) of either 5 or 7; that is, letting $e_i = e(t_i)$ represent the standardized noise, we have

$$\mathrm{SNR} = \frac{\mathrm{SD}_{\mathrm{signal}}}{\mathrm{SD}_{\mathrm{noise}}} \qquad \text{where } \mathrm{SD}_{\mathrm{noise}} = \left(\frac{1}{n-1}\sum_{i=1}^{n}[e_i - \bar{e}]^2\right)^{1/2} \text{ and } \bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i.$$

For both SNRs, we considered sample sizes from $n = 256$ to 2048.

Table 1 reports the average of the mean-square error (MSE) over 200 replications of the test functions, calculated across the sampled times for each realization, given by

$$\text{average MSE} = \frac{1}{200}\sum_{r=1}^{200}\frac{1}{n}\sum_{i=1}^{n}[\hat{f}_r(t_i) - f(t_i)]^2,$$

where $\hat{f}_r(t_i)$ is the estimative of $f(t_i)$ from the rth simulated replication. We take this as an approximation of the IMSE, as defined in Section 1. We have used the Daubechies orthonormal compactly supported wavelet of length $L = 8$

Table 1 *Average (and standard errors) MSE over* 200 *replications of the test functions, calculated across the sampled times for each realization, from the simulation study with correlated errors. The Daubechies orthonormal compactly supported wavelet of length* $L = 8$ (*Daubechies*, 1992), *least asymmetric family, was used with soft level-dependent thresholding beginning at the level* j_0 *indicated*

	SNR = 5				SNR = 7			
n	j_0	Regular	Jittered	Uniform	j_0	Regular	Jittered	Uniform
Sine								
256	2	0.80	0.82	1.17	2	0.41	0.42	0.72
		(0.34)	(0.37)	(0.37)		(0.18)	(0.18)	(0.21)
512	2	0.41	0.41	0.64	2	0.21	0.21	0.38
		(0.21)	(0.20)	(0.23)		(0.10)	(0.10)	(0.12)
1024	2	0.21	0.20	0.35	2	0.11	0.11	0.21
		(0.11)	(0.11)	(0.13)		(0.05)	(0.06)	(0.06)
2048	2	0.12	0.12	0.20	2	0.06	0.06	0.12
		(0.06)	(0.06)	(0.07)		(0.03)	(0.03)	(0.04)
Heavisine								
256	3	1.96	2.00	2.47	3	1.27	1.30	1.63
		(0.35)	(0.37)	(0.47)		(0.18)	(0.18)	(0.29)
512	3	1.40	1.41	1.68	4	0.92	0.93	1.13
		(0.22)	(0.22)	(0.28)		(0.12)	(0.13)	(0.13)
1024	3	1.01	1.01	1.18	3	0.74	0.75	0.88
		(0.12)	(0.13)	(0.20)		(0.12)	(0.11)	(0.17)
2048	4	0.66	0.67	0.76	4	0.41	0.42	0.50
		(0.10)	(0.09)	(0.11)		(0.08)	(0.08)	(0.08)
Doppler								
256	5	3.55	3.84	4.50	5	1.87	2.05	2.53
		(0.22)	(0.27)	(0.42)		(0.10)	(0.15)	(0.28)
512	5	3.05	3.24	3.92	5	1.61	1.77	2.24
		(0.22)	(0.25)	(0.41)		(0.16)	(0.16)	(0.27)
1024	5	2.52	2.59	3.28	6	1.40	1.44	1.79
		(0.25)	(0.25)	(0.42)		(0.08)	(0.08)	(0.12)
2048	5	1.97	2.06	2.46	5	0.91	0.94	1.44
		(0.32)	(0.35)	(0.35)		(0.13)	(0.14)	(0.21)

(Daubechies, 1992), least asymmetric family, and the wavelet coefficients were soft-thresholded from the indicated level j_0 to the greatest one (finest scale). The chosen level j_0 was the level of the coarsest scale of the regular design which resulted in the smaller average MSE and the σ_j values were estimated using the median absolute deviation from zero. The chosen level j_0 happened to be the one with less average MSE for the other designs in almost all cases.

Table 1 shows that the average MSEs on random designs are bigger than those on regular design in all the cases. The average MSEs for jittering fall between those for uniform and regular designs in almost all the cases. However, the jittered

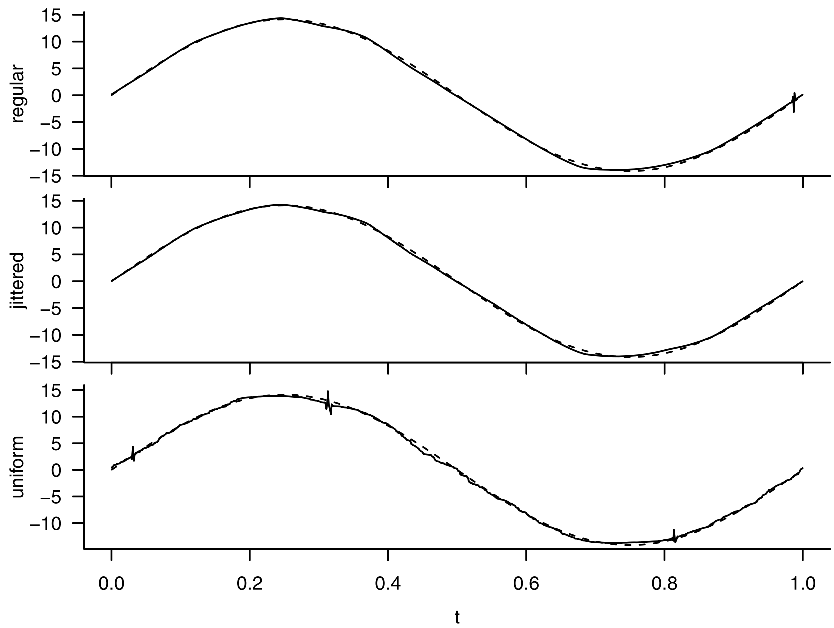

Figure 1 *Sine test function and wavelet estimatives based on* $n = 1024$ *points and* $\mathrm{SNR} = 7$. *Gaussian correlated noise was added to the test function. The Daubechies orthonormal compactly supported wavelet of length* $L = 8$ (*Daubechies*, 1992), *least asymmetric family, was used with soft level-dependent thresholding beginning at the level* $j_0 = 3$.

sampling yields almost the same results as the regular design so that the effect of small timing errors is small, mainly for bigger sample sizes. The average MSEs for uniform design are always greater than those for jittered design. However, as the sample size increases, they decay at approximately the same rate, as prescribed by Corollary 3. Visually, the reconstruction with uniform design is a little more wrinkled than the regular and jittered designs. The jittering is visually almost indistinguishable from the regular design. One realization for the sine, Heavisine and Doppler functions is shown in Figures 1, 2 and 3 respectively, relative to the cases reported in Table 1, with $n = 1024$ and $\mathrm{SNR} = 7$.

5 Application

As an example of the application of our methodology, let us consider the problem of estimating the light curve for the variable star RU Andromeda using data obtained from the American Association of Variable Star Observers (AAVSO) International Database at www.aavso.org (see Sardy et al. (1999) for an earlier attempt to estimate this light curve under the presumption of uncorrelated noise). The data consist of magnitudes of the star measured at irregularly spaced times (the irregular sampling is due to many factors, including blockage of the star by the sun,

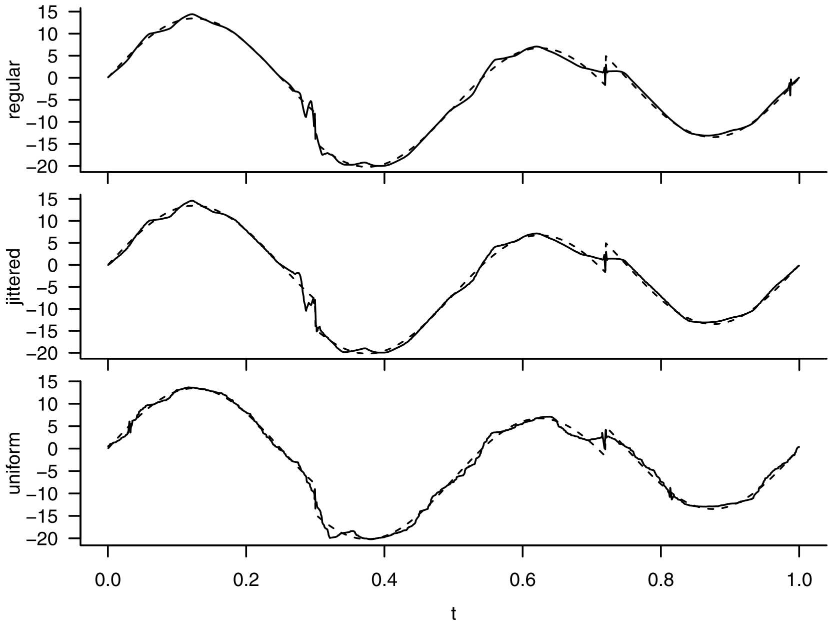

Figure 2 *Heavisine test function and wavelet estimatives based on* $n = 1024$ *points and* SNR $= 7$. *Gaussian correlated noise was added to the test function. The Daubechies orthonormal compactly supported wavelet of length* $L = 8$ (*Daubechies*, 1992), *least asymmetric family, was used with soft level-dependent thresholding beginning at level* $j_0 = 4$.

weather conditions and availability of telescope time). Prior to analysing the data, we eliminated observations reported as upper limits on the star's magnitude (due to atmospheric conditions and the light gathering capabilities of various telescopes). We also replaced multiple observations on the same date by their median value. For our example, we focused on the 256 successive observations recorded from Julian Day 2,440,043 to 2,441,592 (July 5, 1968 to October 1, 1972).

Figure 4(a) shows the RU Andromeda data, along with a light curve estimated using the Haar wavelet (for possible visual comparison with the analysis in Sardy et al. (1999)) and the VisuShrink threshold (see Section 2.1). As noted in Section 2.3, the presence of correlated noise manifests itself as a dependence in the standard deviations σ_j of the wavelet coefficients on the level j.

Figure 4(b) shows σ_j (as estimated by the median absolute deviation from zero) versus j, along with 95% confidence intervals (CIs) obtained by a bootstrapping procedure. The fact that the CIs for σ_4 and σ_7 just barely overlap suggests that we use the threshold of equation (2.3). We used a Kolmogorov–Smirnov test to assess the null hypothesis that the observation times are uniformly distributed, obtaining a p-value of 0.2161. Since we cannot reject the null hypothesis at any reasonable level of significance, we can use Proposition 1 to support using our proposed methodology (this apparent agreement with uniformly distributed sampling times

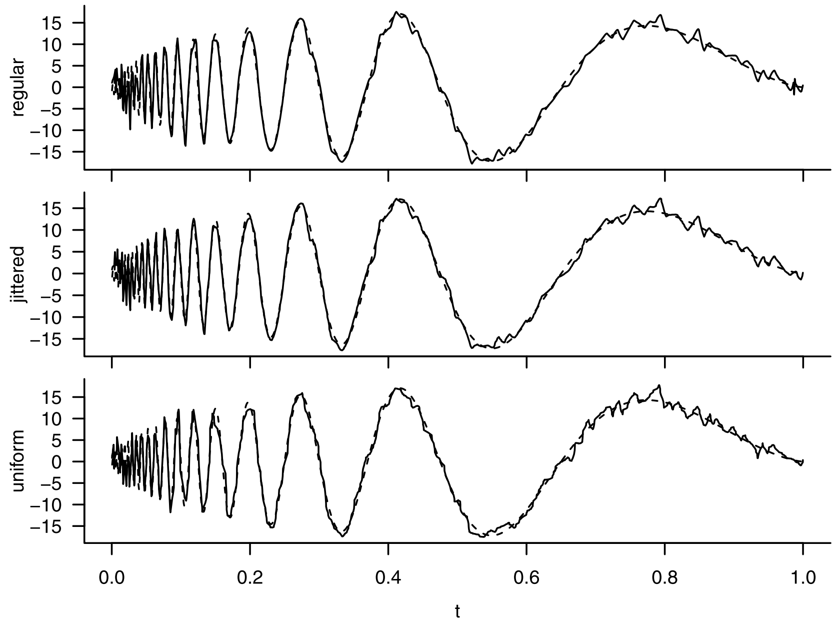

Figure 3 *Doppler test function and wavelet estimatives based on* $n = 1024$ *points and* SNR = 7. *Gaussian correlated noise was added to the test function. The Daubechies orthonormal compactly supported wavelet of length* $L = 8$ (*Daubechies*, 1992), *least asymmetric family, was used with soft level-dependent thresholding beginning at level* $j_0 = 6$.

is one of the reasons we chose this particular subset of the RU Andromeda data). Figure 4(c) shows the estimated light curve using threshold (2.3). We also used $J' = J$ as we did in the simulations and $j_0 = 4$, based on the simulation results and trying to achieve an estimated function visually similar to those in Sardy et al. (1999). Note that this light curve differs from the one in Figure 4(a) at various dates as can be seen, for instance, in the first half of the series, likely due to the autocorrelated errors. Figure 4(d) shows the sample autocorrelation sequence for the residuals from the fitted curve. The fact that this sequence damps down rapidly is an indication that assuming the noise conditions of the Corollary 1 is reasonable here.

6 Summary and discussion

In this paper, we have considered the special cases of uniformly distributed and jittered sampling from a signal in the presence of Gaussian stationary errors with summable autocovariances. We proved that, in these special cases, the samples can be treated as if they were equispaced and with correlated noise; that is, we can use a discrete wavelet transform followed by a level-dependent thresholding

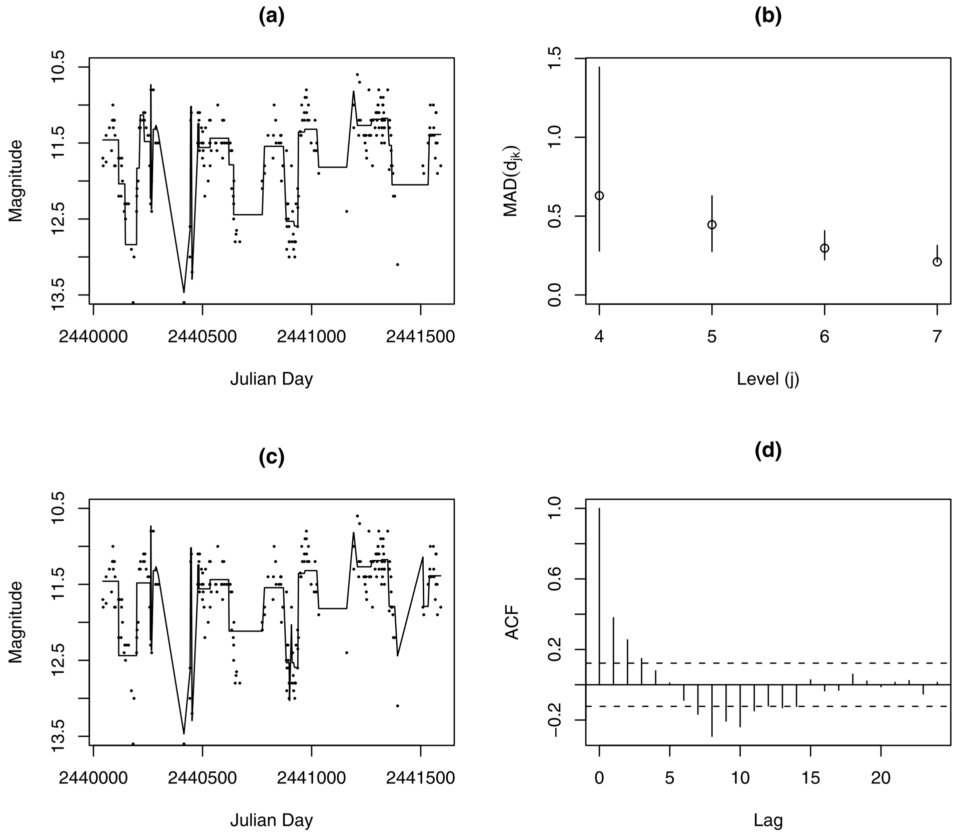

Figure 4 (a) *Data points and estimated light curve through VisuShrink.* (b) *Mean absolute deviation (MAD) from zero of the wavelet coefficients at each resolution level j. Level $j = 7$ is the finest. Endpoints of the error bars are the* 0.025 *and* 0.975 *quantiles of MAD obtained from* 500 *samples (with replacement) of the wavelet coefficients at each resolution level j.* (c) *Data points and estimated light curve considering correlated errors.* (d) *Residuals sample autocorrelation function and* 95% *confidence interval.*

of the wavelet coefficients and an inverse transform to obtain estimators that adaptively achieve—within a logarithmic factor—the optimal convergence rate across a range of Hölder classes. We carried out a brief simulation study to evaluate the finite-sample performance of the proposed methodology. The study found mean-squared errors comparable to those from samples from a regular design (our study involved correlated errors, but Cai and Brown (1999) found the same to be true for uncorrelated errors). We also used our methodology to extract a light curve from unequally spaced observations of a variable star.

The focus of this paper has been on presenting detailed proofs for relatively simple classes of functions, sampling schemes, correlation structures and thresholding rules. Open issues for future research are to consider more complex classes, for example, to consider noise drawn from long-memory stochastic processes and functions that belong to wider Besov classes. Using proofs similar to the ones we have presented, it is straightforward to extend our results to functions that belong to piecewise Hölder classes or Besov classes, but the proofs would be much more technical and notationally dense. Similarly, although our Theorem 1 is valid for very general sampling schemes (we only require that the two first moments are finite) and correlation structures often found in practical applications, it is of interest to consider other random designs and other types of error dependence (e.g., long-range, alternating). Finally, we conjecture that convergence rates can be improved by considering thresholding rules different from the ones we have focused on.

7 Proofs

We begin by restating the problem clearly showing all the assumptions that we are going to need for the proofs of the theorem and the corollaries.

Consider a sample $(t_1, y(t_1)), (t_2, y(t_2)), \ldots, (t_n, y(t_n))$ from some stochastic sampling scheme with respective order statistics $0 \le t_{(1)} < t_{(2)} < \cdots < t_{(n)} \le 1$, $n = 2^J$, $J \in \mathbb{N}$, and suppose that

1. the data are generated by the model $y_i = f(t_{(i)}) + e_i$, where $y_i \equiv y(t_{(i)})$, the function f is unknown, and the errors $e_i \equiv e(t_{(i)})$, for $i = 1, \ldots, n$;
2. the function $f \in L_2[0, 1]$ belongs to a Hölder class $\Lambda^\alpha(M)$, with $M > 0$, $\alpha \in (0, r]$, and known $r > 0$;
3. the errors e_i follow a multivariate normal distribution with mean zero and finite variance $\mathrm{Var}(e_i) = \gamma(0) < \infty$, for $i = 1, \ldots, n$;
4. for some $j_0 \in \mathbb{N}$ fixed, the functions $\phi_{j_0,k}$, $k = 0, \ldots, 2^{j_0} - 1$, and $\psi_{j,k}$, $k = 0, \ldots, 2^j - 1$, $j_0 \le j \in \mathbb{N}$, form a compactly supported orthonormal wavelet basis of $L_2[0, 1]$, where the "mother" wavelet ψ has $r \ge \alpha$ vanishing moments.

Note that we require r vanishing moments, not r-regularity, for the "mother" wavelet.

It is also important to observe that the design has order statistics with finite expected value $E(t_{(i)}) < \infty$, and finite variance $\mathrm{Var}(t_{(i)}) < \infty$, for $i = 1, \ldots, n$, since the order statistics are bounded, that is, $0 \le t_{(1)} < t_{(2)} < \cdots < t_{(n)} \le 1$.

Additionally, we need the following eight lemmas. The first lemma is by Cai and Brown (1999), Lemma 1, which we restate with a proof developed by us, since it is omitted in their article.

Lemma 1 (Cai and Brown (1999), Lemma 1). *Under assumptions* 2 *and* 4, *the wavelet coefficients* $d_{j,k} = \langle f, \psi_{j,k}\rangle$ *are such that*

$$|d_{j,k}| \le C_{M,\psi} 2^{-j(1/2+\alpha)},$$

where $C_{M,\psi}$ *is a positive constant that does not depend on n but only on M and* ψ.

Proof. Since $f \in \Lambda^{\alpha}(M)$, by definition we have that if $0 < \alpha \leq 1$, $|f(x) - f(y)| \leq M|x-y|^{\alpha}$; if $\alpha > 1$, $|f^{(\lfloor\alpha\rfloor)}(x) - f^{(\lfloor\alpha\rfloor)}(y)| \leq M|x-y|^{\alpha'}$ and $|f^{(1)}(x)| \leq M$, where $f^{(m)}(x)$ is the mth derivative of the function f at x, $\lfloor\alpha\rfloor$ is the largest integer less than α, and $\alpha' = \alpha - \lfloor\alpha\rfloor$ (Cai and Brown, 1999, Definition 1). Notice also that since the wavelets are compactly supported, we have that $\int |x|^r |\psi(x)|\, dx < \infty$.

Now we follow closely the proof of Theorem 2.9.1 in Daubechies (1992). Since $\int \psi(x)\, dx = 0$ we have

$$
\begin{aligned}
\langle f, \psi_{j,k}\rangle &= \int \sum_{l\in\mathbb{Z}} \psi_{j,k}(x-l) f(x)\, dx \\
&= \int \sum_{l\in\mathbb{Z}} 2^{j/2}\psi(2^j x - k - l) f(x)\, dx \\
&= \int \sum_{l\in\mathbb{Z}} 2^{j/2}\psi\left(2^j x - \frac{2^j(k+l)}{2^j}\right)\left(f(x) - f\left(\frac{k+l}{2^j}\right)\right) dx;
\end{aligned}
$$

hence if $0 < \alpha \leq 1$

$$
\begin{aligned}
|\langle f, \psi_{j,k}\rangle| &\leq \int \sum_{l\in\mathbb{Z}} 2^{j/2}\left|\psi\left(2^j\left[x - \frac{(k+l)}{2^j}\right]\right)\right| M\left|x - \frac{k+l}{2^j}\right|^{\alpha} dx \\
&\leq \int (N+1)2^{j/2}|\psi(y)| M 2^{-j\alpha}|y|^{\alpha} 2^{-j}\, dy \\
&= 2^{-j(1/2+\alpha)} M(N+1) \int |\psi(y)||y|^{\alpha}\, dy \\
&\leq 2^{-j(1/2+\alpha)} C_{M,\psi},
\end{aligned}
$$

where $N+1$ is the length of the support of ψ.

If $1 < \alpha \leq r$, the Taylor formula let us write

$$
f\left(\frac{k+l}{2^j}\right) - f(x) = \frac{1}{m!}\sum_{m=1}^{\lfloor\alpha\rfloor-1} f^{(m)}(x)\left(\frac{k+l}{2^j} - x\right)^m + R,
$$

where R is the remainder in the Schlömilch form such that

$$
\begin{aligned}
|R| &\leq \frac{|f^{(\lfloor\alpha\rfloor)}(c)|}{\lfloor\alpha\rfloor!}\frac{\lfloor\alpha\rfloor}{\alpha}\left|\frac{k+l}{2^j} - c\right|^{-\alpha'}\left|\frac{k+l}{2^j} - x\right|^{\alpha} \\
&\leq \frac{2^{\lfloor\alpha\rfloor-1}M}{\lfloor\alpha\rfloor!}\left|\frac{k+l}{2^j} - x\right|^{\alpha} \\
&\leq M\left|\frac{k+l}{2^j} - x\right|^{\alpha},
\end{aligned}
$$

for some c strictly between x and $2^{-j}(k+l)$. Thus, since the moments of ψ of orders 1 up to $\lfloor\alpha\rfloor - 1$ vanishes, we just follow the same steps when $0 < \alpha \leq 1$. □

Lemma 2. *Under assumptions* 2 *and* 4, *any estimator* $\hat{f}(x)$ *of* $f(x)$, *where*

$$\hat{f}(x) = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k}\phi_{j_0,k}(x) + \sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1} \hat{d}_{j,k}\psi_{j,k}(x),$$

for all $x \in [0, 1]$, *and for any* $J' \in \mathbb{N}$ *where* $J' \leq J$, *has a risk that is equal to*

$$\begin{aligned} E\big(\|\hat{f} - f\|_2^2\big) = &\sum_{k=0}^{2^{j_0}-1} E\big((\hat{c}_{j_0,k} - c_{j_0,k})^2\big) \\ &+ \sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1} E\big((\hat{d}_{j,k} - d_{j,k})^2\big) \\ &+ \sum_{j=J'}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}^2, \end{aligned}$$

where $\hat{c}_{j_0,k}$ *and* $\hat{d}_{j,k}$ *are any estimators of the respective coefficients* $c_{j_0,k} = \langle f, \phi_{j_0,k}\rangle$, *and* $d_{j,k} = \langle f, \psi_{j,k}\rangle$.

Proof. By the assumption 2, the function f can be expanded into the wavelet basis of the assumption 4 and the risk can be written as

$$\begin{aligned} E\big(\|\hat{f} - f\|_2^2\big) &= E\left(\int_0^1 [\hat{f}(x) - f(x)]^2\, dx\right) \\ &= E\Bigg(\int_0^1 \Bigg[\sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k}\phi_{j_0,k}(x) + \sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1} \hat{d}_{j,k}\psi_{j,k}(x) \\ &\qquad - \sum_{k=0}^{2^{j_0}-1} c_{j_0,k}\phi_{j_0,k}(x) - \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(x)\Bigg]^2 dx\Bigg) \\ &= E\Bigg(\int_0^1 \Bigg[\sum_{k=0}^{2^{j_0}-1} (\hat{c}_{j_0,k} - c_{j_0,k})\phi_{j_0,k}(x) \\ &\qquad + \sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1} (\hat{d}_{j,k} - d_{j,k})\psi_{j,k}(x) \\ &\qquad - \sum_{j=J'}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(x)\Bigg]^2 dx\Bigg). \end{aligned}$$

By the orthogonality of the wavelet basis, this expression is equal to

$$E\left(\sum_{k=0}^{2^{j_0}-1}\int_0^1(\hat{c}_{j_0,k}-c_{j_0,k})^2\phi_{j_0,k}^2(x)\,dx+\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1}\int_0^1(\hat{d}_{j,k}-d_{j,k})^2\psi_{j,k}^2(x)\,dx\right.$$
$$\left.+\sum_{j=J'}^{\infty}\sum_{k=0}^{2^j-1}\int_0^1 d_{j,k}^2\psi_{j,k}^2(x)\,dx\right),$$

and by orthonormality we have

$$E\big(\|\hat{f}-f\|_2^2\big)=\sum_{k=0}^{2^{j_0}-1}E\big((\hat{c}_{j_0,k}-c_{j_0,k})^2\big)+\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1}E\big((\hat{d}_{j,k}-d_{j,k})^2\big)$$
$$+\sum_{j=J'}^{\infty}\sum_{k=0}^{2^j-1}d_{j,k}^2.$$

□

Lemma 3. *For all* $\alpha>0$ *and any* $J_1<J'$, *where* $J_1\in\mathbb{N}$ *and* $J'\in\mathbb{N}$, *under assumptions* 2 *and* 4,

$$\sum_{j=J'}^{\infty}\sum_{k=0}^{2^j-1}d_{j,k}^2\le\frac{C_{M,\psi}^2}{1-2^{-2\alpha}}2^{-2\alpha J'}$$

and

$$\sum_{j=J_1}^{J'-1}\sum_{k=0}^{2^j-1}d_{j,k}^2\le C_{M,\psi}^2\left(\frac{2^{-2\alpha J_1}-2^{-2\alpha J'}}{1-2^{-2\alpha}}\right),$$

where $d_{j,k}=\langle f,\psi_{j,k}\rangle$, *and* $C_{M,\psi}$ *is a positive constant that does not depend on* n *but only on* M *and* ψ.

Proof. According to Lemma 1, $|d_{j,k}|\le C_{M,\psi}2^{-j(1/2+\alpha)}$. Thus, for all $\alpha>0$,

$$\sum_{j=J'}^{\infty}\sum_{k=0}^{2^j-1}d_{j,k}^2\le C_{M,\psi}^2\sum_{j=J'}^{\infty}2^j2^{2(-j(1/2+\alpha))}=C_{M,\psi}^2\sum_{j=J'}^{\infty}2^{-2j\alpha}$$
$$=C_{M,\psi}^2\left(\sum_{j=0}^{\infty}2^{-2j\alpha}-\sum_{j=0}^{J'-1}2^{-2j\alpha}\right)$$
$$=C_{M,\psi}^2\frac{(2^{-2\alpha})^{J'}}{1-2^{-2\alpha}},$$

and

$$\sum_{j=J_1}^{J'-1}\sum_{k=0}^{2^j-1}d_{j,k}^2\le C_{M,\psi}^2\sum_{j=J_1}^{J'-1}2^{-2j\alpha}=C_{M,\psi}^2\left(\frac{2^{-2\alpha J_1}-2^{-2\alpha J'}}{1-2^{-2\alpha}}\right).$$

□

Lemma 4. *Consider the assumption* 1 *and let*

$$\tilde{f}(x) = \sum_{i=0}^{n-1} n^{-1/2} y_{i+1} \phi_{J,i}^{(H)}(x),$$

where

$$\phi_{J,i}^{(H)}(x) = 2^{J/2} \phi^{(H)}(2^J x - i) = \sqrt{n} I((nx - i) \in (0, 1]),$$

is the Haar scaling function, for $i = 0, \dots, n-1$, *and* $I(\cdot)$ *denotes the usual indicator function. Let also*

$$f_n(x) = \sum_{i=0}^{n-1} n^{-1/2} f\left(\frac{i+1}{n}\right) \phi_{J,i}^{(H)}(x),$$

and consider the function

$$A(x) = f_n(x) - f(x),$$

which is a deterministic completion of $f(x)$, *the function*

$$B(x) = \sum_{i=0}^{n-1} n^{-1/2} f(t_{(i+1)}) \phi_{J,i}^{(H)}(x) - f_n(x),$$

which is random but depends only on $t_{(1)}, \dots, t_{(n)}$, *and the function*

$$R(x) = \sum_{i=0}^{n-1} n^{-1/2} e_{i+1} \phi_{J,i}^{(H)}(x),$$

which is also random but depends only on $e_1, \dots, e_n$. *Then, under the assumption* 2, $\tilde{f}(x) \in L_2[0, 1]$ *is a piecewise constant function that approximates* $f(x)$ *for all* $x \in [0, 1]$ *such that* $\tilde{f}(k/n) = y_k$, *and conveniently separates the sources of uncertainties as*

$$\tilde{f}(x) = f(x) + A(x) + B(x) + R(x).$$

Proof. The proof is straightforward since by assumption 2, $f(x)$ can be written as

$$f(x) = \sum_{k=0}^{n-1} c_{J,k}^{(H)} \phi_{J,k}^{(H)}(x) + \sum_{j=J}^{\infty} \sum_{k=0}^{2^j-1} d_{j,k}^{(H)} \psi_{j,k}^{(H)}(x),$$

where $\psi_{j,k}^{(H)}$ denotes the respective Haar wavelet function (see, e.g., Härdle et al. (1998), for details), $c_{J,k}^{(H)} = \langle f, \phi_{J,k}^{(H)} \rangle$, and $d_{j,k}^{(H)} = \langle f, \psi_{j,k}^{(H)} \rangle$.

Thus,

$$
\begin{aligned}
\tilde{f}(x) &= f(x) + [f_n(x) - f(x)] + [\tilde{f}(x) - f_n(x)] \\
&= f(x) + [f_n(x) - f(x)] + \left[\sum_{i=0}^{n-1} n^{-1/2}\big(f(t_{(i+1)}) + e_{i+1}\big)\phi_{J,i}^{(H)}(x) - f_n(x)\right] \\
&= f(x) + A(x) + B(x) + R(x).
\end{aligned}
$$

Also, for $k = 1, \ldots, n$,

$$
\tilde{f}(k/n) = \sum_{i=0}^{n-1} y_{i+1}/\sqrt{n}\phi_{J,i}^{(H)}(k/n) = \sum_{i=0}^{n-1} y_{i+1} I((nk/n - i) \in (0, 1]) = y_k,
$$

so that $\tilde{f}(x)$ is a piecewise constant approximation to $f(x)$, based on the observed points $y_1, \ldots, y_n$. Similarly, we also have

$$
f_n\left(\frac{k}{n}\right) = f\left(\frac{k}{n}\right), \qquad R\left(\frac{k}{n}\right) = e_k. \qquad \square
$$

Lemma 5. *Under assumptions* 1 *and* 2, *let* $\tilde{f}(x)$, $A(x)$, $B(x)$ *and* $R(x)$ *be the functions of Lemma* 4. *Consider also the assumption* 4, *and let* $\tilde{r}_{j_0,k} = \langle R, \phi_{j_0,k}\rangle$, $\tilde{a}_{j_0,k} = \langle A, \phi_{j_0,k}\rangle$, $\tilde{b}_{j_0,k} = \langle B, \phi_{j_0,k}\rangle$, *and*

$$
\hat{c}_{j_0,k} = \tilde{c}_{j_0,k} = c_{j_0,k} + \tilde{a}_{j_0,k} + \tilde{b}_{j_0,k} + \tilde{r}_{j_0,k} = \int_0^1 \tilde{f}(x)\phi_{j_0,k}(x)\,dx.
$$

Then, *under the additional assumption* 3,

$$
\sum_{k=0}^{2^{j_0}-1} E\big((\hat{c}_{j_0,k} - c_{j_0,k})^2\big) \le \sum_{k=0}^{2^{j_0}-1} \operatorname{Var}(\tilde{r}_{j_0,k}) + 2\sum_{k=0}^{2^{j_0}-1} \tilde{a}_{j_0,k}^2 + 2\sum_{k=0}^{2^{j_0}-1} E\big(\tilde{b}_{j_0,k}^2\big).
$$

Proof. Let $E_1(Y) = E(Y|t_{(1)}, \ldots, t_{(n)})$ for any random variable Y. Using the (numerical) Hölder inequality $(a+b)^2 \le 2a^2 + 2b^2$, $a, b \in \mathbb{R}$, we have that

$$
\begin{aligned}
E\big((\hat{c}_{j_0,k} - c_{j_0,k})^2\big) &= E\big(E_1\big((\hat{c}_{j_0,k} - c_{j_0,k})^2\big)\big) \\
&= E\big(E_1\big((\tilde{r}_{j_0,k} + \tilde{a}_{j_0,k} + \tilde{b}_{j_0,k})^2\big)\big) \\
&= E\big(\tilde{r}_{j_0,k}^2\big) + 0 + E\big(E_1\big((\tilde{a}_{j_0,k} + \tilde{b}_{j_0,k})^2\big)\big) \\
&\le E\big(\tilde{r}_{j_0,k}^2\big) + E\big(E_1\big(2\tilde{a}_{j_0,k}^2 + 2\tilde{b}_{j_0,k}^2\big)\big) \\
&= E\big(\tilde{r}_{j_0,k}^2\big) + 2\tilde{a}_{j_0,k}^2 + 2E\big(\tilde{b}_{j_0,k}^2\big).
\end{aligned}
$$

In the last expression,

$$
E\big(\tilde{r}_{j_0,k}^2\big) = \operatorname{Var}(\tilde{r}_{j_0,k}) + \big(E(\tilde{r}_{j_0,k})\big)^2 = \operatorname{Var}(\tilde{r}_{j_0,k})
$$

since, by the form of $R(x)$ in Lemma 4 and by assumption 3,

$$E(\tilde{r}_{j_0,k}) = E\left(\int_0^1 \sum_{i=0}^{n-1} n^{-1/2} e_{i+1} \phi_{J,i}^{(H)}(x) \phi_{j_0,k}(x)\, dx\right) = 0.$$

Thus,

$$\sum_{k=0}^{2^{j_0}-1} E\big((\hat{c}_{j_0,k} - c_{j_0,k})^2\big) \le \sum_{k=0}^{2^{j_0}-1} \mathrm{Var}(\tilde{r}_{j_0,k}) + 2 \sum_{k=0}^{2^{j_0}-1} \tilde{a}_{j_0,k}^2 + 2 \sum_{k=0}^{2^{j_0}-1} E\big(\tilde{b}_{j_0,k}^2\big). \qquad \square$$

Lemma 6. *Under assumptions* 1 *and* 2, *let* $A(x)$, $B(x)$ *and* $R(x)$ *be the functions of Lemma* 4. *Consider also the assumption* 4, *and let* $a_{j,k} = \langle A, \psi_{j,k}\rangle$, $b_{j,k} = \langle B, \psi_{j,k}\rangle$, $r_{j,k} = \langle R, \psi_{j,k}\rangle$, $d_{j,k} = \langle f, \psi_{j,k}\rangle$, $d'_{j,k} = d_{j,k} + a_{j,k} + b_{j,k}$, $\tilde{d}_{j,k} = d'_{j,k} + r_{j,k}$, *and* $\hat{d}_{j,k} = \mathrm{sgn}(\tilde{d}_{j,k})(|\tilde{d}_{j,k} - \lambda_{j,k}|)_+$, *where* $\lambda_{j,k} = \sigma_{j,k}\sqrt{2\log n}$ *and* $\sigma_{j,k}^2 = \mathrm{Var}(r_{j,k})$. *Then, under the additional assumption* 3,

$$\begin{aligned}
&\sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} E\big((\hat{d}_{j,k} - d_{j,k})^2\big) \\
&\quad \le 6\log n \sum_{(j,k)\in\mathcal{I}_1} \sigma_{j,k}^2 + 8 \sum_{(j,k)\in\mathcal{I}_2} d_{j,k}^2 + 20 \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} \big(a_{j,k}^2 + E\big(b_{j,k}^2\big)\big) \\
&\qquad + \frac{2}{n} \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} \sigma_{j,k}^2,
\end{aligned}$$

where $\mathcal{I}_1$ *is the set of pairs of indexes* (j,k) *where the signal is much stronger than the noise, such that* $6\sigma_{j,k}^2 \log n < 8d_{j,k}^2$, $\mathcal{I}_2$ *is the complement of* $\mathcal{I}_1$, $j \in \{j_0, \dots, J'-1\}$ *and* $k \in \{0, \dots, 2^j - 1\}$.

Proof. Let $E_1(Y) = E(Y|t_{(1)}, \dots, t_{(n)})$ for any random variable Y, and $\sigma_{j,k;1}^2 = E_1(r_{j,k}^2)$. Using the (numerical) Hölder inequality $(a+b)^2 \le 2a^2 + 2b^2$, $a, b \in \mathbb{R}$, we have that

$$\begin{aligned}
E\big((\hat{d}_{j,k} - d_{j,k})^2\big) &= E\big((\hat{d}_{j,k} - d'_{j,k} + a_{j,k} + b_{j,k})^2\big) \\
&\le E\big(2(\hat{d}_{j,k} - d'_{j,k})^2 + 2(a_{j,k} + b_{j,k})^2\big) \qquad (7.1)\\
&= E\big(E_1\big(2(\hat{d}_{j,k} - d'_{j,k})^2\big) + 2(a_{j,k} + b_{j,k})^2\big).
\end{aligned}$$

Denote $\min(x,y)$ by $x \wedge y$. Using Lemma 4 in Cai and Brown (1999), we obtain

$$\begin{aligned}
E_1\big((\hat{d}_{j,k} - d'_{j,k})^2\big) &\le \big(2(d'_{j,k})^2 + n^{-1}\sigma_{j,k;1}^2\big) \wedge (2\log n + 1)\sigma_{j,k;1}^2 \\
&\le \big(2(d'_{j,k})^2 + n^{-1}\sigma_{j,k;1}^2\big) \wedge (2\log n + \log n + 1/n)\sigma_{j,k;1}^2
\end{aligned}$$

$$= \left(2(d'_{j,k})^2 + n^{-1}\sigma^2_{j,k;1}\right) \wedge \left((3\sigma^2_{j,k;1} \log n) + n^{-1}\sigma^2_{j,k;1}\right)$$
$$= \left(2(d'_{j,k})^2 \wedge (3\sigma^2_{j,k;1} \log n)\right) + n^{-1}\sigma^2_{j,k;1}.$$

Now, use this result in (7.1):

$$\begin{aligned}
&E_1\big(2(\hat{d}_{j,k} - d'_{j,k})^2\big) + 2(a_{j,k} + b_{j,k})^2 \\
&\quad \le E_1\big(2(\hat{d}_{j,k} - d'_{j,k})^2\big) + 4a^2_{j,k} + 4b^2_{j,k} \\
&\quad \le 2\big((2(d'_{j,k})^2 \wedge 3\sigma^2_{j,k;1} \log n) + n^{-1}\sigma^2_{j,k;1}\big) + 4a^2_{j,k} + 4b^2_{j,k} \\
&\quad = 2\big((2(d_{j,k} + a_{j,k} + b_{j,k})^2 \wedge 3\sigma^2_{j,k;1} \log n) + n^{-1}\sigma^2_{j,k;1}\big) + 4a^2_{j,k} + 4b^2_{j,k} \\
&\quad \le 2\big((2(2d^2_{j,k} + 4a^2_{j,k} + 4b^2_{j,k}) \wedge 3\sigma^2_{j,k;1} \log n) + n^{-1}\sigma^2_{j,k;1}\big) + 4a^2_{j,k} + 4b^2_{j,k} \\
&\quad = 2\big(((4d^2_{j,k} + 8a^2_{j,k} + 8b^2_{j,k}) \wedge 3\sigma^2_{j,k;1} \log n) + n^{-1}\sigma^2_{j,k;1}\big) + 4a^2_{j,k} + 4b^2_{j,k} \\
&\quad \le 2\big((4d^2_{j,k} \wedge 3\sigma^2_{j,k;1} \log n) + 8a^2_{j,k} + 8b^2_{j,k} + n^{-1}\sigma^2_{j,k;1}\big) + 4a^2_{j,k} + 4b^2_{j,k} \\
&\quad = (8d^2_{j,k} \wedge 6\sigma^2_{j,k;1} \log n) + 20a^2_{j,k} + 20b^2_{j,k} + 2n^{-1}\sigma^2_{j,k;1},
\end{aligned}$$

and thus,

$$\begin{aligned}
&E\big((\hat{d}_{j,k} - d_{j,k})^2\big) \\
&\quad \le \big(8d^2_{j,k} \wedge 6E(\sigma^2_{j,k;1}) \log n\big) + 20a^2_{j,k} + 20E(b^2_{j,k}) + 2n^{-1}E(\sigma^2_{j,k;1}) \\
&\quad = \big(8d^2_{j,k} \wedge 6\sigma^2_{j,k} \log n\big) + 20a^2_{j,k} + 20E(b^2_{j,k}) + 2n^{-1}\sigma^2_{j,k}.
\end{aligned}$$

Then,

$$\begin{aligned}
&\sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} E\big((\hat{d}_{j,k} - d_{j,k})^2\big) \\
&\quad \le 6 \log n \sum_{(j,k)\in\mathcal{I}_1} \sigma^2_{j,k} + 8 \sum_{(j,k)\in\mathcal{I}_2} d^2_{j,k} + 20 \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} \big(a^2_{j,k} + E(b^2_{j,k})\big) \\
&\qquad + \frac{2}{n} \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} \sigma^2_{j,k}.
\end{aligned}$$

□

Lemma 7. *Under assumptions* 1 *and* 2, *let* $A(x)$ *and* $B(x)$ *be the functions of Lemma* 4. *Then*

$$\|A\|_2^2 = \int_0^1 A(x)^2\, dx \le C_1 n^{-2s(\alpha)} + 2 \sum_{j=J}^{\infty} \sum_{k=0}^{2^j-1} (d^{(H)}_{j,k})^2$$

and

$$
\begin{aligned}
E\|B\|_2^2 &= E\Big(\int_0^1 B(x)^2\,dx\Big) \\
&\le \frac{M}{n}\sum_{i=1}^{n}\Big\{\mathrm{Var}\Big(t_{(i)}-\frac{i}{n}\Big)+\Big[E\Big(t_{(i)}-\frac{i}{n}\Big)\Big]^2\Big\}^{s(\alpha)},
\end{aligned}
$$

where $s(\alpha)=\min(\alpha,1)$, *and the constant* $C_1>0$ *does not depend on* n.

Proof. For the first part, we can expand $f(x)$ using the Haar wavelet basis, and obtain

$$
\begin{aligned}
\|A\|_2^2 &= \int_0^1 A(x)^2\,dx = \int_0^1 [f_n(x)-f(x)]^2\,dx \\
&= \int_0^1 \Bigg[\sum_{i=0}^{n-1} n^{-1/2} f\Big(\frac{i+1}{n}\Big)\phi_{J,i}^{(H)}(x) - \sum_{i=0}^{n-1} c_{J,i}^{(H)}\phi_{J,i}^{(H)}(x) \\
&\quad - \sum_{j=J}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}^{(H)}\psi_{j,k}^{(H)}(x)\Bigg]^2 dx \\
&= \int_0^1 \Bigg[\sum_{i=0}^{n-1}\Big(n^{-1/2} f\Big(\frac{i+1}{n}\Big) - c_{J,i}^{(H)}\Big)\phi_{J,i}^{(H)}(x) \\
&\quad - \sum_{j=J}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}^{(H)}\psi_{j,k}^{(H)}(x)\Bigg]^2 dx \\
&\le \int_0^1 2\Bigg[\sum_{i=0}^{n-1}\Big(n^{-1/2} f\Big(\frac{i+1}{n}\Big) - c_{J,i}^{(H)}\Big)\phi_{J,i}^{(H)}(x)\Bigg]^2 dx \\
&\quad + \int_0^1 2\Bigg[\sum_{j=J}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}^{(H)}\psi_{j,k}^{(H)}(x)\Bigg]^2 dx \\
&= 2\sum_{i=0}^{n-1}\Big(n^{-1/2} f\Big(\frac{i+1}{n}\Big) - c_{J,i}^{(H)}\Big)^2 + 2\sum_{j=J}^{\infty}\sum_{k=0}^{2^j-1}(d_{j,k}^{(H)})^2 \\
&\le C_1 n^{-2s(\alpha)} + 2\sum_{j=J}^{\infty}\sum_{k=0}^{2^j-1}(d_{j,k}^{(H)})^2,
\end{aligned}
$$

by Lemma 2(i) in Cai and Brown (1998), where the constant $C_1>0$ does not depend on n. Note that Lemma 2(i) in Cai and Brown (1998) is presented with-out proof, but it is straightforward and uses Theorem 2.9.1 in Daubechies (1992), which requires r vanishing moments, not regularity, for the "mother" wavelet.

For the second part,

$$
\begin{aligned}
E\|B\|_2^2 &= E\big(E_1\|B\|_2^2\big)\\
&= E\left(E_1\int_0^1 B(x)^2\,dx\right)\\
&= E\left(E_1\int_0^1\left[\sum_{i=0}^{n-1} n^{-1/2} f(t_{(i+1)})\phi_{J,i}^{(H)}(x) - f_n(x)\right]^2 dx\right)\\
&= E\left(E_1\int_0^1\left[\sum_{i=0}^{n-1} n^{-1/2} f(t_{(i+1)})\phi_{J,i}^{(H)}(x)\right.\right.\\
&\qquad \left.\left.-\sum_{i=0}^{n-1} n^{-1/2} f\left(\frac{i+1}{n}\right)\phi_{J,i}^{(H)}(x)\right]^2 dx\right)\\
&= E\left(E_1\int_0^1\left[\sum_{i=0}^{n-1} n^{-1/2}\left\{f(t_{(i+1)}) - f\left(\frac{i+1}{n}\right)\right\}\phi_{J,i}^{(H)}(x)\right]^2 dx\right)\\
&= E\left(E_1\left(\sum_{i=0}^{n-1}\frac{1}{n}\left[f(t_{(i+1)}) - f\left(\frac{i+1}{n}\right)\right]^2\right)\right)\\
&= E\left(E_1\left(\frac{1}{n}\sum_{i=1}^{n}\left[f(t_{(i)}) - f\left(\frac{i}{n}\right)\right]^2\right)\right),
\end{aligned}
$$

using the orthogonality of the Haar wavelet basis.

Using the Definition 1 from Cai and Brown (1999), we have that $|f(x) - f(y)| \le M|x-y|^{s(\alpha)}$, where $s(\alpha) = \min(\alpha, 1)$. Thus, the approximation error

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} E\left(\left(f(t_{(i)}) - f\left(\frac{i}{n}\right)\right)^2\right)\\
&\qquad \le \frac{M}{n}\sum_{i=1}^{n} E\left(\left(t_{(i)} - \frac{i}{n}\right)^{2s(\alpha)}\right)\\
&\qquad \le \frac{M}{n}\sum_{i=1}^{n}\left[E\left(\left(t_{(i)} - \frac{i}{n}\right)^2\right)\right]^{s(\alpha)}\\
&\qquad = \frac{M}{n}\sum_{i=1}^{n}\left\{\operatorname{Var}\left(t_{(i)} - \frac{i}{n}\right) + \left[E\left(t_{(i)} - \frac{i}{n}\right)\right]^2\right\}^{s(\alpha)}
\end{aligned}
$$

by Jensen's inequality and observing that the design has order statistics with finite expected value $E(t_{(i)}) < \infty$, and finite variance $\mathrm{Var}(t_{(i)}) < \infty$, for $i = 1, \dots, n$, since the order statistics are bounded, that is, $0 \le t_{(1)} < t_{(2)} < \cdots < t_{(n)} \le 1$. $\square$

Lemma 8. *Under assumptions* 1 *and* 2, *let* $R(x)$ *be the function of Lemma* 4, *and, considering also the assumption* 4, *let* $r_{j,k} = \langle R, \psi_{j,k}\rangle$. *Then,*

$$\mathrm{Var}(r_{j,k}) \le \frac{\gamma(0)}{n} + \frac{2^j\|\psi\|_\infty^2}{n^2} \sum_{i=0}^{n-1} \sum_{t=0,t\neq i}^{n-1} \big|\mathrm{Cov}(e_{i+1}, e_{t+1})\big|,$$

and, if $\mathrm{Cov}(e_{i+1}, e_{t+1}) = \gamma(|t-i|)$, *for* $i, t = 0, 1, \dots, n-1$,

$$\mathrm{Var}(r_{j,k}) \le \frac{2C_{\phi,\psi} \sum_{u=0}^{n-1} |\gamma(u)|}{n},$$

where $C_{\phi,\psi}$ *is a positive constant that does not depend on* n. *Inequalities for* $\mathrm{Var}(\tilde r_{j_0,k})$, *where* $\tilde r_{j_0,k} = \langle R, \phi_{j_0,k}\rangle$, *are similar but replacing* $2^j\|\psi\|_\infty^2$ *and* $C_{\phi,\psi}$ *by* $2^{j_0}\|\phi\|_\infty^2$ *and* C_ϕ, *respectively, where* C_ϕ *is also a positive constant that does not depend on* n.

Proof.

$$\begin{aligned}
r_{j,k} = \langle R, \psi_{j,k}\rangle &= \int_0^1 R(x)\psi_{j,k}(x)\,dx \\
&= \int_0^1 \sum_{i=0}^{n-1} n^{-1/2} e_{i+1}\phi_{J,i}^{(H)}(x)\psi_{j,k}(x)\,dx \\
&= \sum_{i=0}^{n-1} \int_0^1 n^{-1/2} e_{i+1}\phi_{J,i}^{(H)}(x)\psi_{j,k}(x)\,dx \\
&= \sum_{i=0}^{n-1} \int_{i/n}^{(i+1)/n} n^{-1/2} e_{i+1}\phi_{J,i}^{(H)}(x)\psi_{j,k}(x)\,dx \\
&= \sum_{i=0}^{n-1} \int_{i/n}^{(i+1)/n} n^{-1/2} e_{i+1} n^{1/2}\psi_{j,k}(x)\,dx \\
&= \sum_{i=0}^{n-1} \int_{i/n}^{(i+1)/n} e_{i+1}\psi_{j,k}(x)\,dx,
\end{aligned}$$

such that

$$\begin{aligned}
&\mathrm{Var}(r_{j,k}) \\
&\quad = \mathrm{Cov}\left(\sum_{i=0}^{n-1} \int_{i/n}^{(i+1)/n} e_{i+1}\psi_{j,k}(x)\,dx, \sum_{i=0}^{n-1} \int_{i/n}^{(i+1)/n} e_{i+1}\psi_{j,k}(x)\,dx\right)
\end{aligned}$$

$$\leq \sum_{i=0}^{n-1}\left(\int_{i/n}^{(i+1)/n} \psi_{j,k}(x)\,dx\right)^2 \gamma(0)$$

$$+ \sum_{i=0}^{n-1}\sum_{\substack{t=0\\t\neq i}}^{n-1}\left(\int_{i/n}^{(i+1)/n} \psi_{j,k}(x)\,dx\right)\left(\int_{t/n}^{(t+1)/n} \psi_{j,k}(y)\,dy\right)\mathrm{Cov}(e_{i+1}, e_{t+1}).$$

By the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\sum_{i=0}^{n-1}\left(\int_{i/n}^{(i+1)/n} \psi_{j,k}(x)\,dx\right)^2 \gamma(0)\\
&\quad\leq \sum_{i=0}^{n-1}\left(\int_{i/n}^{(i+1)/n} \psi_{j,k}^2(x)\,dx\right)\left(\int_{i/n}^{(i+1)/n} dx\right)\gamma(0)\\
&\quad= \sum_{i=0}^{n-1}\int_{i/n}^{(i+1)/n} \psi_{j,k}^2(x)\,dx\,\frac{\gamma(0)}{n}\\
&\quad= \frac{\gamma(0)}{n}\int_0^1 \psi_{j,k}^2(x)\,dx\\
&\quad= \frac{\gamma(0)}{n}.
\end{aligned}$$

Now, notice that

$$\left|\int_{i/n}^{(i+1)/n} \psi_{j,k}(x)\,dx\right| \leq \int_{i/n}^{(i+1)/n} \left|\psi_{j,k}(x)\right|dx \leq \int_{i/n}^{(i+1)/n} 2^{j/2}\|\psi\|_\infty\,dx = \frac{2^{j/2}\|\psi\|_\infty}{n},$$

for all $i = 0, 1, \ldots, n-1$.

Then,

$$\mathrm{Var}(r_{j,k}) \leq \frac{\gamma(0)}{n} + \frac{2^j\|\psi\|_\infty^2}{n^2}\sum_{i=0}^{n-1}\sum_{t=0,t\neq i}^{n-1}\left|\mathrm{Cov}(e_{i+1}, e_{t+1})\right|.$$

If $\mathrm{Cov}(e_{i+1}, e_{t+1}) = \gamma(|t-i|)$, for $i, t = 0, 1, \ldots, n-1$, $\mathrm{Var}(r_{j,k})$ is equal to

$$\mathrm{Cov}\left(\sum_{i=0}^{n-1}\int_0^1 n^{-1/2}e_{i+1}\phi_{J,i}^{(H)}(x)\psi_{j,k}(x)\,dx,\right.$$
$$\left.\sum_{i=0}^{n-1}\int_0^1 n^{-1/2}e_{i+1}\phi_{J,i}^{(H)}(x)\psi_{j,k}(x)\,dx\right),$$

which is also equal to

$$\sum_{i=0}^{n-1}\sum_{t=0}^{n-1} n^{-1}\gamma\big(|t-i|\big)\int_0^1 \phi_{J,i}^{(H)}(x)\psi_{j,k}(x)\,dx \int_0^1 \phi_{J,i}^{(H)}(y)\psi_{j,k}(y)\,dy.$$

We are going to use the following inequality, obtained by using $y = 2^j(x-l)-k$.

$$\begin{aligned}\int f(x)\psi_{j,k}(x)\,dx &= \int \sum_{l\in\mathbb{Z}} f\left(\frac{y+k+2^j l}{2^j}\right)2^{j/2}\psi(y)\frac{dy}{2^j}\\ &\le 2^{-j/2}\int \sum_{l\in\mathbb{Z}} \|f\|_\infty |\psi(y)|\,dy\\ &\le 2^{-j/2}(N+1)\|f\|_\infty\|\psi\|_1.\end{aligned}$$

Then we have that

$$\begin{aligned}&|\operatorname{Var}(r_{j,k})|\\ &\quad\le \sum_{u=-(n-1)}^{n-1} n^{-1}|\gamma(|u|)|\\ &\qquad\times \sum_{t=0}^{n-1-|u|}\int_0^1 |\phi_{J,t}^{(H)}(x)||\psi_{j,k}(x)|\,dx \left|\int_0^1 \phi_{J,t+|u|}^{(H)}(y)\psi_{j,k}(y)\,dy\right|\\ &\quad\le \sum_{u=-(n-1)}^{n-1} n^{-1}|\gamma(|u|)|\\ &\qquad\times \sum_{t=0}^{n-1-|u|}\int_0^1 |\phi_{J,t}^{(H)}(x)||\psi_{j,k}(x)|\,dx|2^{-J/2}\|\phi^{(H)}\|_1 2^{j/2}\|\psi\|_\infty 2|\\ &\quad= \sum_{u=-(n-1)}^{n-1} n^{-1}|\gamma(|u|)|\\ &\qquad\times \sum_{t=0}^{n-1-|u|}\int_{t/n}^{(t+1)/n} 2^{J/2}|\phi^{(H)}(2^J x-t)||\psi_{j,k}(x)|\,dx\\ &\qquad\times \big(2^{-J/2}\|\phi^{(H)}\|_1 2^{j/2}\|\psi\|_\infty 2\big)\\ &\quad= \sum_{u=-(n-1)}^{n-1} n^{-1}|\gamma(|u|)|\big(2^{-J/2}\|\phi^{(H)}\|_1 2^{j/2}\|\psi\|_\infty 2\big)\\ &\qquad\times \int_0^{(n-|u|)/n} 2^{J/2}|\phi^{(H)}(x)||\psi_{j,k}(x)|\,dx\end{aligned}$$

$$\leq \sum_{u=-(n-1)}^{n-1} n^{-1}|\gamma(|u|)| \|\phi^{(H)}\|_1 2^{j/2}\|\psi\|_\infty 2\int_0^1 |\phi^{(H)}(x)||\psi_{j,k}(x)|\,dx$$

$$\leq \sum_{u=-(n-1)}^{n-1} n^{-1}|\gamma(|u|)| \|\phi^{(H)}\|_1 2^{j/2}\|\psi\|_\infty 2 2^{-j/2}\|\psi\|_1 \|\phi^{(H)}\|_\infty (N+1)$$

$$= 2(N+1)\|\phi^{(H)}\|_1 \|\psi\|_\infty \|\psi\|_1 \|\phi^{(H)}\|_\infty \sum_{u=-(n-1)}^{n-1} n^{-1}|\gamma(|u|)|$$

$$= \frac{2C_{\phi,\psi}\sum_{u=0}^{n-1}|\gamma(u)|}{n},$$

where $C_{\phi,\psi}$ is a positive constant that does not depend on n.

By analogous arguments we can bound $\mathrm{Var}(\tilde{r}_{j_0,k})$ similarly, with $2^{j_0}\|\phi\|_\infty^2$ and C_ϕ instead of $2^j\|\psi\|_\infty^2$ and $C_{\phi,\psi}$, respectively, where C_ϕ is also a positive constant that does not depend on n. □

Now we are ready to prove the Theorem 1, its corollaries, and the Proposition 1.

7.1 Proof of Theorem 1

We first use Lemma 2 to decompose the risk function $E(\|\hat{f}-f\|_2^2)$ and follow by applying Lemmas 4, 5 and 6, in order to bound the risk from above as

$$\begin{aligned} E(\|\hat{f}-f\|_2^2) \leq{}& 2\sum_{k=0}^{2^{j_0}-1}\tilde{a}_{j_0,k}^2 + 20\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1} a_{j,k}^2 \\ &+ 2\sum_{k=0}^{2^{j_0}-1} E(\tilde{b}_{j_0,k}^2) + 20\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1} E(b_{j,k}^2) \\ &+ \sum_{k=0}^{2^{j_0}-1}\mathrm{Var}(\tilde{r}_{j_0,k}) \\ &+ 8\sum_{(j,k)\in\mathcal{I}_2} d_{j,k}^2 + \sum_{j=J'}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}^2 \\ &+ 6\log n\sum_{(j,k)\in\mathcal{I}_1}\sigma_{j,k}^2 + \frac{2}{n}\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1}\sigma_{j,k}^2. \end{aligned}$$

Now, by the orthogonality of the wavelet basis (assumption 4), we have that

$$2\sum_{k=0}^{2^{j_0}-1}\tilde{a}_{j_0,k}^2 + 20\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1} a_{j,k}^2 \leq 20\|A\|_2^2$$

and

$$2\sum_{k=0}^{2^{j_0}-1} E(\tilde{b}^2_{j_0,k}) + 20\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1} E(b^2_{j,k}) \le 20E\|B\|_2^2,$$

which, together with a direct application of the results of Lemmas 7 and 3, permit us to write that

$$\begin{aligned} E(\|\hat{f}-f\|_2^2) \le {} & 20C_1 n^{-2s(\alpha)} + 40\frac{C^2_{M,\psi^{(H)}}}{1-2^{-2\alpha}}2^{-2\alpha J} \\ & + 20\frac{M}{n}\sum_{i=1}^{n}\left\{\mathrm{Var}(t_{(i)}) + \left[E\left(t_{(i)}-\frac{i}{n}\right)\right]^2\right\}^{s(\alpha)} \\ & + \sum_{k=0}^{2^{j_0}-1}\mathrm{Var}(\tilde{r}_{j_0,k}) \\ & + 8\sum_{(j,k)\in\mathcal{I}_2} d^2_{j,k} + \frac{C^2_{M,\psi}}{1-2^{-2\alpha}}2^{-2\alpha J'} \\ & + 6\log n\sum_{(j,k)\in\mathcal{I}_1}\sigma^2_{j,k} + \frac{2}{n}\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1}\sigma^2_{j,k}. \end{aligned}$$

7.2 Proof of Corollary 1

In order to make the proof easier to follow, we rewrite the result of Theorem 1 as $E(\|\hat{f}-f\|_2^2) \le T_1 + T_2 + \cdots + T_8$, where

$$T_1 = 20C_1 n^{-2s(\alpha)}, \qquad T_2 = 40\frac{C^2_{M,\psi}}{1-2^{-2\alpha}}2^{-2\alpha J},$$

$$T_3 = 20\frac{M}{n}\sum_{i=1}^{n}\left\{\mathrm{Var}(t_{(i)}) + \left[E\left(t_{(i)}-\frac{i}{n}\right)\right]^2\right\}^{s(\alpha)},$$

$$T_4 = \sum_{k=0}^{2^{j_0}-1}\mathrm{Var}(\tilde{r}_{j_0,k}), \qquad T_5 = 8\sum_{(j,k)\in\mathcal{I}_2} d^2_{j,k}, \qquad T_6 = \frac{C^2_{M,\psi}}{1-2^{-2\alpha}}2^{-2\alpha J'},$$

$$T_7 = 6\log n\sum_{(j,k)\in\mathcal{I}_1}\sigma^2_{j,k} \quad \text{and} \quad T_8 = \frac{2}{n}\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1}\sigma^2_{j,k}.$$

For the uniform design, $E(t_{(i)}) = i/(n+1)$ implies that

$$\left|E(t_{(i)}) - \frac{i}{n}\right| = \frac{i}{n(n+1)} \le \frac{1}{\sqrt{n}}$$

and

$$\mathrm{Var}(t_{(i)}) = \frac{(n+1)i - i^2}{(n+1)^2(n+2)} < \frac{1}{n}.$$

For the jittered design, $E(t_{(i)}) = (2i-1)/(2n)$, such that

$$\left| E(t_{(i)}) - \frac{i}{n} \right| = \left| -\frac{1}{2n} \right| \le \frac{1}{\sqrt{n}}$$

and

$$\mathrm{Var}(t_{(i)}) = E(j_i) = \frac{1}{12n^2} < \frac{1}{n}.$$

In both cases we have $T_3 \le 20n^{-1} M \sum_{i=1}^{n} (2/n)^{s(\alpha)} = 20M(2/n)^{s(\alpha)}$.

Since by Lemma 1 $d_{j,k}^2 \le C_{M,\psi}^2 2^{-j(1+2\alpha)}$, we have that T_5 is less than or equal to $8 \sum_{(j,k)\in\mathcal{I}_2} C_{M,\psi}^2 2^{-j(1+2\alpha)}$.

Consider the stationary short-memory case, where $\mathrm{Cov}(e(t_{(i)}), e(t_{(j)})) = \gamma(|i - j|)$ and $\lim_{n\to\infty} \sum_{u=-(n-1)}^{n-1} |\gamma(u)| = C_\gamma < \infty$. Then by Lemma 8

$$\mathrm{Var}(r_{j,k}) \le \frac{2C_{\phi,\psi} \sum_{u=0}^{n-1} |\gamma(u)|}{n} \le \frac{2C_{\phi,\psi} C_\gamma}{n},$$

and similarly, $\mathrm{Var}(r_{j,k}) \le n^{-1} 2C_\phi C_\gamma$, where $C_{\phi,\psi}$ and C_ϕ are positive constants that do not depend on n. Thus, we can write

$$T_4 \le 2^{j_0} \frac{2C_\phi C_\gamma}{n}, \qquad T_7 \le 6 \log n \sum_{(j,k)\in\mathcal{I}_1} \left(\frac{2C_{\phi,\psi} C_\gamma}{n} \right)$$

and

$$T_8 \le \frac{2}{n} \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} \left(\frac{2C_{\phi,\psi} C_\gamma}{n} \right) \le \frac{2^{J'}}{n} \frac{4C_{\phi,\psi} C_\gamma}{n}.$$

Compare the previous results for T_5 and T_7. The set of levels j where $6 \log n (2C_{\phi,\psi} C_\gamma)/n < 8C_{M,\psi}^2 2^{-j(1+2\alpha)}$ is delimited by J_1, which is the largest integer such that $2^{J_1} < C_3 (n/\log n)^{1/(1+2\alpha)}$, where

$$C_3 = \{8C_{M,\psi}^2 / [6(2C_{\phi,\psi} C_\gamma)]\}^{1/(1+2\alpha)}.$$

We may have $J_1 \le J'$ or $J_1 > J'$.

In the first case, that is, $J_1 \le J'$,

$$T_5 \le 8 \sum_{(j,k)\in\mathcal{I}_2} C_{M,\psi}^2 2^{-j(1+2\alpha)} \le 8 \sum_{j=J_1}^{J'-1} \sum_{k=0}^{2^j-1} C_{M,\psi}^2 2^{-j(1+2\alpha)} \le 8C_{M,\psi}^2 \frac{2^{-2\alpha J_1}}{1 - 2^{-2\alpha}}$$

and

$$T_7 \leq 6\log n \sum_{(j,k)\in\mathcal{I}_1} \left(\frac{2C_{\phi,\psi}C_\gamma}{n}\right)$$
$$\leq 6\log n \sum_{j=j_0}^{J_1-1}\sum_{k=0}^{2^j-1} \frac{2C_{\phi,\psi}C_\gamma}{n}$$
$$\leq 6\log n \frac{2C_{\phi,\psi}C_\gamma}{n} 2^{J_1}.$$

Since

$$(C_3-1)\left(\frac{n}{\log n}\right)^{1/(1+2\alpha)} < C_3\left(\frac{n}{\log n}\right)^{1/(1+2\alpha)} - 1 < 2^{J_1} < C_3\left(\frac{n}{\log n}\right)^{1/(1+2\alpha)},$$

then $2^{-2\alpha J_1} < (C_3-1)^{-2\alpha}(n/\log n)^{-2\alpha/(1+2\alpha)}$ and

$$2^{J_1}\frac{\log n}{n} < C_3\left(\frac{n}{\log n}\right)^{1/(1+2\alpha)}\frac{\log n}{n} = C_3\left(\frac{\log n}{n}\right)^{2\alpha/(1+2\alpha)}.$$

Thus, a little algebra shows that for $r \geq \alpha \geq 1/2 - (\log n)^{-1}\log\log n$, the rate of decay of the risk is dominated by the terms with J_1 for all J' in the interval $\{J_1, \ldots, J\}$. For $0 < \alpha < 1/2 - (\log n)^{-1}\log\log n$, the rate of decay is dominated by the term with $s(\alpha)$ and is the fastest when $J' = J$.

In the second case (when $J' < J_1$), $T_5 = 0$ and

$$T_7 \leq 6\log n \sum_{j=j_0}^{J'}\sum_{k=0}^{2^j-1}\frac{2C_{\phi,\psi}C_\gamma}{n} \leq 6\log n\frac{2C_{\phi,\psi}C_\gamma}{n}2^{J'+1}.$$

Thus, as before, a little algebra shows that for $1/2 - (\log n)^{-1}\log\log n \leq \alpha \leq r$, the fastest rate of decay occurs when $J' = J_1 - 1$ and it is in the order of $(\log n/n)^{2\alpha/(1+2\alpha)}$. For $0 < \alpha < 1/2 - (\log n)^{-1}\log\log n$, the rate of decay is dominated by the term with $s(\alpha)$ and is the fastest also when $J' = J_1 - 1$.

In any case, notice that for $0 < \alpha < -(\log n)^{-1}\log\log n$,

$$\frac{1}{n^{s(\alpha)}} = \frac{1}{n^\alpha} > \left(\frac{\log n}{n}\right)^{2\alpha/(2+2\alpha)}.$$

Considering also that

$$\left(\frac{\log n}{n}\right)^{2\alpha/(1+2\alpha)} < \left(\frac{\log n}{n}\right)^{2\alpha/(2+2\alpha)},$$

we conclude that, when $-(\log n)^{-1}\log\log n \leq \alpha < 1/2 - (\log n)^{-1}\log\log n$, the rate of decay of the risk can be improved for all J' in the interval $\{J_2, \ldots, J\}$, where J_2 is the largest integer such that

$$2^{J_2} < \left(8C_{M,\psi}^2/(6\gamma(0))\right)^{1/(2+2\alpha)}(n/\log n)^{1/(2+2\alpha)}.$$

Thus, the rate of decay is $O(n^{-\alpha})$ only for $0 < \alpha < (\log n)^{-1} \log\log n$. A rate of convergence in the order of $(\log n/n)^{2\alpha/(2+2\alpha)}$ can also be obtained by using the first result in Lemma 8.

7.3 Proof of Corollary 2

Follow the proof of Corollary 1 but note that $T_3 = 0$.

Thus, if $J' \geq J_1$, the rate of decay of the risk is dominated by T_7 for all J' in the interval $\{J_1, \dots, J\}$. If $J' < J_1$, the rate of decay of the risk is also dominated by T_7, which is minimized when $J' = J_1 - 1$. In both cases, T_7 is of order $O((\log n/n)^{2\alpha/(1+2\alpha)})$.

7.4 Proof of Corollary 3

Follow the proof of Corollary 1 but with $C_\gamma = \gamma(0)$.

7.5 Proof of Corollary 4

Follow the proof of Corollary 2 but with $C_\gamma = \gamma(0)$.

7.6 Proof of Corollary 5

Follow the proof of Corollary 1 but note that T_4, T_5, T_7 and T_8 are all equal to zero.

In the case when $J' = J$, and $0 < \alpha \leq 1$, T_1, T_2 and T_6 are of order $O(n^{-2\alpha})$. When $J' = J$ and $1 < \alpha \leq r$, T_1 is of order $O(n^{-2})$ and T_2 and T_6 are of order $O(n^{-2\alpha})$. In all these cases, T_3 is of order $O(n^{-s(\alpha)})$. Thus, for $0 < \alpha \leq r$, the rate of convergence is dominated by T_3 and it is on the order of $O(n^{-s(\alpha)})$.

7.7 Proof of Corollary 6

Following the proof of Corollary 1, note that T_3, T_4, T_5, T_7 and T_8 are all null.

In the case when $J' = J$, and $0 < \alpha \leq 1$ the remaining terms are of order $O(n^{-2\alpha})$. When $J' = J$ and $1 < \alpha \leq r$, T_1 is of order $O(n^{-2})$ and the other terms are of order $O(n^{-2\alpha})$. Thus, for $0 < \alpha \leq r$, the rate of convergence is dominated by T_1 and it is on the order of $O(n^{-2s(\alpha)})$.

7.8 Proof of Proposition 1

Consider the jittered design. Since $\mathrm{Cov}(e(r), e(s)) = E(\sigma^2 e^{-n\beta|r-s|})$, for some $\beta > 0$, $0 < \sigma^2 < \infty$ and fixed r and s, then,

$$\begin{aligned}
&\mathrm{Cov}\big(e(t_{(r)}), e(t_{(s)})\big)\\
&\quad = E\Big(\sigma^2 \exp\Big(-n\beta\Big|\frac{r-s}{n} + j_r - j_s\Big|\Big)\Big)\\
&\quad = \gamma\big(|r-s|\big).
\end{aligned}$$

Replacing the random variables j_r and j_s by their maximum and minimum, respectively, this expression turns to be less than or equal to

$$E\left(\sigma^2 \exp\left(-n\beta\left|\frac{r-s}{n}+\frac{2}{2n}\right|\right)\right)=\sigma^2 e^{-\beta|u+1|},$$

where $u=r-s$. Then,

$$\lim_{n\to\infty}\sum_{u=-(n-1)}^{n-1}|\gamma(u)| \le \sigma^2 e^{-\beta}\lim_{n\to\infty}\sum_{u=-(n-1)}^{n-1} e^{-\beta u}<\infty.$$

Consider the uniform design such that

$$\operatorname{Cov}\big(e(t_{(i)}),e(t_{(j)})\big)=E\big(\sigma^2 e^{-(n+1)\beta|t_{(i)}-t_{(j)}|}\big),$$

for some $\beta>0$, $0<\sigma^2<\infty$ and fixed i and j.

Now, note that (Johnson, Kotz and Balakrishnan, 1995, page 217):

$$E\big((t_{(i)}-t_{(j)})^k\big)=\frac{\Gamma(|i-j|+k)\Gamma(n+1)}{\Gamma(|i-j|)\Gamma(n+1+k)} = \frac{(|i-j|+k-1)!n!}{(|i-j|-1)!(n+k)!}.$$

Then,

$$\operatorname{Cov}\big(e(t_{(i)}),e(t_{(j)})\big) =\sigma^2\sum_{k=0}^{\infty}(-1)^k\frac{(n+1)^k\beta^k}{k!}E\big(|t_{(i)}-t_{(j)}|^k\big) =\gamma\big(|i-j|\big).$$

To evaluate $\lim_{n\to\infty}\sum_{u=1}^{n-1}|\gamma(u)|$, note first that

$$\big|\gamma\big(|i-j|\big)\big|=\big|E\big(\sigma^2 e^{-(n+1)\beta|t_{(i)}-t_{(j)}|}\big)\big|=\gamma\big(|i-j|\big).$$

Note also that

$$\sum_{u=1}^{n-1}\frac{(u+k-1)!}{(u-1)!}=k!\binom{k+(n-2)+1}{k+1}=\frac{(n+k-1)!}{(k+1)(n-2)!}, \tag{7.2}$$

where (7.2) comes from the equation 0.151.1 in Gradshteĭn and Ryzhik (2007). Then using these facts,

$$\begin{aligned}\sum_{u=1}^{n-1}|\gamma(u)|=\sum_{u=1}^{n-1}\gamma(u)&=\sigma^2\sum_{k=0}^{\infty}(-1)^k\frac{(n+1)^k\beta^k n!}{k!(n+k)!}\frac{(n+k-1)!}{(k+1)(n-2)!}\\ &=\sigma^2\sum_{k=0}^{\infty}(-1)^k\frac{(n+1)^k\beta^k}{k!}\frac{n(n-1)}{(n+k)(k+1)}.\end{aligned} \tag{7.3}$$

Evaluating the summation in (7.3), we have that

$$\begin{aligned}\sum_{k=0}^{\infty} & \frac{(-1)^k (n+1)^k \beta^k}{k!} \frac{1}{(n+k)(k+1)} \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k (n+1)^k \beta^k}{k!} \frac{(n)_k (1)_k}{(n+1)_k (2)_k} \frac{1}{n} \\ &= \frac{1}{n} {}_2F_2\big(n, 1; n+1, 2; (-1)(n+1)\beta\big),\end{aligned} \tag{7.4}$$

where $\Gamma(n)$ denotes the gamma function, the Pochhammer symbol $(a)_k = \Gamma(a+k)/\Gamma(a)$, and ${}_2F_2(a, b; c, d; z)$ denotes a generalized hypergeometric function.

Denoting by ${}_1F_1(a, b, z)$ the confluent hypergeometric function of the first kind, we have that (http://functions.wolfram.com/07.25.03.0005.01)

$$\begin{aligned}\frac{1}{b-a} & \big(b\, {}_1F_1(a, a+1, z) - a\, {}_1F_1(b, b+1, z)\big) \\ &= {}_2F_2(a, b; a+1, b+1; z),\end{aligned}$$

and applying this result to equation (7.4),

$$\begin{aligned}\frac{1}{n} & {}_2F_2\big(n, 1; n+1, 2; (-1)(n+1)\beta\big) \\ &= \frac{1}{n} \frac{1}{1-n} \big({}_1F_1\big(n, n+1, (-1)(n+1)\beta\big) \\ &\quad - n\, {}_1F_1\big(1, 2, (-1)(n+1)\beta\big)\big).\end{aligned}$$

From equation 9.236.4 in Gradshteĭn and Ryzhik (2007), applying

$${}_1F_1(a, a+1, z) = a(-z)^{-a}\big(\Gamma(a) - \Gamma(a, -z)\big) \tag{7.5}$$

to the last expression we have that

$$\begin{aligned}\sum_{k=0}^{\infty} & \frac{(-1)^k (n+1)^k \beta^k}{k!} \frac{1}{(n+k)(k+1)} \\ &= \frac{1}{(n+1)(n-1)} \\ &\quad \times \left(\frac{1}{[(n+1)\beta]^n}\big[-(n+1)\Gamma(n) + (n+1)\Gamma\big(n, (n+1)\beta\big)\big]\right) \\ &\quad + \frac{{}_1F_1(1, 2, (-1)(n+1)\beta)}{n-1},\end{aligned}$$

where $\Gamma(n,a) = \int_a^\infty t^{n-1}e^{-t}\,dt$ denotes the incomplete gamma function. Using (7.5), we also have that

$$\begin{aligned}
&\frac{{}_1F_1(1,2,(-1)(n+1)\beta)}{n-1}\\
&\qquad= \frac{1}{\beta(n^2-1)}\left[1 - \int_{\beta(n+1)}^\infty t^{1-1}e^{-t}\,dt\right]\\
&\qquad= \frac{1}{\beta(n^2-1)}\left[1 - e^{-\beta(n+1)}\right].
\end{aligned}$$

Thus, for $n > 1$,

$$\begin{aligned}
&\sum_{k=0}^\infty (-1)^k \frac{(n+1)^k\beta^k}{k!(n+k)(k+1)}\\
&\qquad= \frac{[-(n+1)\Gamma(n) + (n+1)\Gamma(n,\beta(n+1))]}{(n^2-1)[\beta(n+1)]^n}\\
&\qquad\quad+ \frac{1-e^{-\beta(n+1)}}{\beta(n^2-1)},
\end{aligned}$$

where the incomplete gamma function

$$\begin{aligned}
\Gamma\bigl(n,\beta(n+1)\bigr) &\le (n-1)!e^{-\beta(n+1)}\sum_{k=0}^\infty \beta^k(n+1)^k/k!\\
&= \Gamma(n),
\end{aligned}$$

when n is an integer. Applying these results in (7.3), we get that for every $n > 1$,

$$\begin{aligned}
&\sum_{u=1}^{n-1}|\gamma(u)|\\
&\qquad\le \sigma^2\left\{\frac{n(n-1)[-(n+1)\Gamma(n) + (n+1)\Gamma(n)]}{(n^2-1)[\beta(n+1)]^n} + \frac{1}{\beta}\right\}\\
&\qquad= \frac{\sigma^2}{\beta}.
\end{aligned}$$

Thus, $\lim_{n\to\infty}\sum_{u=1}^{n-1}|\gamma(u)| \le \sigma^2/\beta < \infty$.

Acknowledgements

The authors thank You-Gan Wang (Commonwealth Scientific and Research Organization—CSIRO Mathematics, Informatics and Statistics) and José Carlos

Simon de Miranda (University of São Paulo, Institute of Mathematics and Statistics) for helpful comments and acknowledge with thanks the variable star observations from the American Association of Variable Star Observers (AAVSO) International Database contributed by observers worldwide and used in this research.

R. Porto supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq. P. Morettin supported by Fundação de Amparo à Pesquisa do Estado de São Paulo—Fapesp (Grant 2008/51097-6).

References

Antoniadis, A., Pensky, M. and Sapatinas, T. (2014). Nonparametric regression estimation based on spatially inhomogeneous data: minimax global convergence rates and adaptivity. *ESAIM Probab. Stat.* **18**, 1–41. MR3143732

Antoniadis, A. and Pham, D. T. (1998). Wavelet regression for random or irregular design. *Comput. Statist. Data Anal.* **28**, 353–369. MR1659207

Baraud, Y., Comte, F. and Viennet, G. (2001). Adaptive estimation in autoregression or β-mixing regression via model selection. *Ann. Statist.* **29**, 839–875. MR1865343

Brillinger, D. R. (1996). Some uses of cumulants in wavelet analysis. *J. Nonparametr. Stat.* **6**, 93–114. MR1383046

Cai, T. T. and Brown, L. (1998). Wavelet shrinkage for nonequispaced samples. *Ann. Statist.* **26**, 1783–1799. MR1673278

Cai, T. T. and Brown, L. (1999). Wavelet estimation for samples with random uniform design. *Statist. Probab. Lett.* **42**, 313–321. MR1688134

Chesneau, C. (2007). Regression with random design: A minimax study. *Statist. Probab. Lett.* **77**, 40–53. MR2339017

Cochrane, D. and Orcutt, G. H. (1949). Applications of least squares regression to relationships containing autocorrelated errors. *J. Amer. Statist. Assoc.* **44**, 32–61.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. *CBMS-NSF Regional Conf. Ser. in Appl. Math.* Philadelphia, Pennsylvania: SIAM. MR1162107

Delouille, V., Simoens, J. and von Sachs, R. (2004). Smooth design-adapted wavelets for nonparametric stochastic regression. *J. Amer. Statist. Assoc.* **99**, 643–658. MR2090900

Delouille, V. and von Sachs, R. (2005). Estimation of nonlinear autoregressive models using design-adapted wavelets. *Ann. Inst. Statist. Math.* **57**, 235–253. MR2160649

Dippé, M. A. Z. and Wold, E. H. (1985). Antialiasing through stochastic sampling. *ACM SIGGRAPH Computer Graphics* **19**, 69–78.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **81**, 425–455. MR1311089

Donoho, D. L. and Johnstone, I. M. (1996). Neo-classical minimax problems, thresholding and adaptive function estimation. *Bernoulli* **2**, 39–62. MR1394051

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet Shrinkage: Asymptopia? *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57**, 301–369. MR1323344

Efromovich, S. (1999). How to overcome the curse of long-memory errors. *IEEE Trans. Inform. Theory* **45**, 1735–1741. MR1699909

Gaïffas, S. (2009). Uniform estimation of a signal based on inhomogeneous data. *Statist. Sinica* **19**, 427–447. MR2514170

Gradshteĭn, I. S. and Ryzhik, I. M. (2007). *Table of Integrals, Series, and Products*, 7th ed. Amsterdan: Elsevier/Academic Press. Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger, With one CD-ROM (Windows, Macintosh and UNIX). MR2360010

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*. *Lecture Notes in Statistics* **129**. New York: Springer. MR1618204

Hofmann, M. (1999). On nonparametric estimation in nonlinear AR(1)-models. *Statist. Probab. Lett.* **44**, 29–45. MR1706307

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions, Vol. 2*, 2nd ed. New York: Wiley. MR1326603

Johnstone, I. M. and Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **59**, 319–351. MR1440585

Kerkyacharian, G. and Picard, D. (2004). Regression in random design and warped wavelets. *Bernoulli* **10**, 1053–1105. MR2108043

Kulik, R. and Raimondo, M. (2009). Wavelet regression in random design with heteroscedastic dependent errors. *Ann. Statist.* **37**, 3396–3430. MR2549564

Kutner, M. H., Nachtsheim, C. J., Neter, J. and Li, W. (2004). *Applied Linear Statistical Models*, 5th ed. New York: McGraw-Hill/Irwin.

Nason, G., Kovac, A. and Maechler, M. (2006). Wavethresh—Software to perform wavelet statistics and transforms. R package version 2.2-9. http://www.R-project.org.

Neumann, M. H. and von Sachs, R. (1995). Wavelet thresholding: Beyond the Gaussian IID situation. In *Wavelets and Statistics* (A. Antoniadis and G. Oppenheim, eds.). *Lecture Notes in Statistics* **103**, 301–329. New York: Springer. MR1364677

Ogden, R. T. (1997). *Essential Wavelets for Statistical Applications and Data Analysis*, 1st ed. Boston: Birkhäuser. MR1420193

Porto, R. F., Morettin, P. A. and Aubin, E. C. Q. (2008). Wavelet regression with correlated errors on a piecewise Hölder class. *Statist. Probab. Lett.* **78**, 2739–2743. MR2465116

Qin, D. and Gilbert, C. L. (2001). The error term in the history of time series econometrics. *Econometric Theory* **17**, 424–450. MR1863725

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

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

von Sachs, R. and MacGibbon, B. (2000). Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scand. J. Stat.* **27**, 475–499. MR1795776

Yang, Y. (2001). Nonparametric regression with dependent errors. *Bernoulli* **7**, 633–655. MR1849372

R. Porto
Bank of Brazil
Restructuring of Operational Assets Directorship
SBS Ed. Sede III
Brasília, DF 70073-901
Brazil
E-mail: rogerio@bb.com.br

P. Morettin
E. Aubin
Institute of Mathematics and Statistics
University of São Paulo
R. Matão 1010
São Paulo, SP 05508-090
Brazil
E-mail: pam@ime.usp.br; aubin@ime.usp.br

D. Percival
Applied Physics Laboratory and Department of Statistics
University of Washington
1013 NE 40th Street
Seattle, Washington 98105-6698
USA
E-mail: dbp@apl.washington.edu