Wavelet Shrinkage for Regression Models with Random Design and Correlated Errors

Rogério F. Porto[*,†], Pedro A. Morettin[†], Donald B. Percival[‡] and Elisete C. Q. Aubin[†]

September 16, 2009

Abstract

This paper presents some results on semi-parametric regression using wavelet methods in the presence of autocorrelated stationary Gaussian errors, and when the explanatory variable follows a uniform distribution or comes from a stochastic sampling like the jittered sampling scheme. The aim is to estimate the signal globally with low risk. It is shown that in these special cases the samples can be treated as if they were equispaced and with correlated noise; i.e., the estimator achieves an almost optimal convergence rate. Some simulation studies compare the cases with and without equal spacings in finite samples. Daily volatility estimatives of a low-traded stock illustrate the usefulness of the method.

Keywords: autocorrelation; semi-parametric estimation; smoothing; wavelets.

2000 Mathematics Subject Classifications: 62M10; 62G08; 91B84.

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 $\{\epsilon(t_1), \dots, \epsilon(t_n)\}$, which we assume throughout to obey a multivariate Gaussian distribution. The data consist of points $\{(t_1, y(t_1)), \dots, (t_n, y(t_n))\}$, where $y(t_i) = f(t_i) + \epsilon(t_i)$, 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

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

[*]e-mail: rdporto@ime.usp.br

[†]Institute of Mathematics and Statistics, University of São Paulo, Brazil.

[‡]Applied Physics Laboratory and Department of Statistics, University of Washington, USA.

Wavelet shrinkage methods have been very successful in signal extraction and nonparametric regression, but most are focused on equispaced samples (i.e., over a regular grid $t_i = i/n$) with independent and identically distributed (IID) errors. The equispaced assumption has been relaxed to handle unequally spaced samples with a fixed design [17], a uniformly distributed design [8] and a general random design [16, 18], but these extensions are restricted to IID errors. Wavelet shrinkage methods have also been adapted to handle correlated errors, but only in the context of equispaced samples [9] and of unequally spaced samples with a fixed design [20].

In this paper, we investigate wavelet shrinkage for certain unequally sampled designs in the presence of correlated errors. We consider stochastic sampling schemes where either the sample points t_i are uniformly distributed in $[0, 1]$ or they come from a jittering; i.e., $t_i = (2i - 1)/(2n) + j_i$, 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 sampling on a regular grid [2]. We show that under our assumptions the samples can be treated as if they were equispaced with correlated noise [9], and hence we can apply the VisuShrink procedure [7] with level-dependent thresholds.

The paper is organized as follows. In Section 2 we review some basic properties of wavelets along with earlier research on wavelet shrinkage. 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 in Section 4 and a financial application in Section 5. We state our conclusions in Section 6 and devote Section 7 to proofs.

2 Wavelets and wavelet shrinkage

An orthonormal wavelet basis is 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 that $\int \phi = 1$. We recall that a wavelet is *r-regular* if it has r vanishing moments and r continuous derivatives. Let

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

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

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

denote the periodized wavelets, which we use henceforth, but with the superscript "p" suppressed. For some coarse scale $j_0 \geq 0$ the collection

$$\phi_{j_0,k}, k = 0, \ldots, 2^{j_0} - 1, \text{ and } \psi_{j,k}, j \geq j_0, k = 0, \ldots, 2^j - 1,$$

constitutes an orthonormal basis of $L_2[0, 1]$.

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

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

The function f can be expanded into a wavelet series as

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

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

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

where the noise e_i is drawn from some stochastic process, and 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 [7] 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

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

for some threshold λ (the theoretical results of this paper focus on soft thresholding, but the results remain valid for hard thresholding funcion $\eta_H(d,\lambda) = dI(|d| \geq \lambda)$). 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 $\tilde{d}_{j,k}$ at threshold $\lambda = \sigma\sqrt{n^{-1} 2 \log n}$ 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

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

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

$$y(t_i) = f(t_i) + \epsilon_i, \qquad i = 1, \ldots, n,$$

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

$$y_i = f(t_{(i)}) + e_i, \qquad i = 1, \ldots, n, \quad (3)$$

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 consists of observed pairs $\{(t_{(1)}, y_1), (t_{(2)}, y_2), \ldots, (t_{(n)}, y_n)\}$. Because the t_i are uniformly distributed on $[0,1]$, the $t_{(i)}$ are distributed as $\text{Beta}(i, n-i+1)$ and $E(t_{(i)}) = i/(n+1)$ [8]. 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 $y = (y_1, \ldots, y_n)^T$. 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 [8]. 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 equispaced design. An isometric argument can be used to justify this practice for other types of nonuniform sampling [16].

2.3 Regular design with correlated errors

Consider model (1) again, but now suppose that the error vector $e = (e_1, \ldots, e_n)^T$ have a multivariate Gaussian distribution with mean 0 and covariance matrix Γ. Also, assume that the errors are stationary so that Γ has entries $\gamma_{|r-s|}$. Let $z = We$ be the wavelet transform of the 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 $\sigma_j^2 = \text{Var}(z_{j,k})$ [9].

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 [9]. 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 $\tilde{d}_{j,k}$, $j = j_0, \ldots, J-1$, $k = 0, \ldots, 2^j - 1$:

$$\hat{d}_{j,k} = \eta_S(\tilde{d}_{j,k}, \lambda_j), \tag{4}$$

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). 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 $(t_1, y(t_1)), (t_2, y(t_2)), \ldots, (t_n, y(t_n))$ from some stochastic sampling scheme with respective order statistics $0 \le t_{(1)} < t_{(2)} < \cdots < t_{(n)} \le 1$ that satisfy

$$\mathrm{Var}\left(t_{(i)}\right) \le \frac{1}{n} \text{ and } \left| E\left(t_{(i)}\right) - \frac{i}{n} \right| \le \frac{1}{\sqrt{n}} \tag{5}$$

for $i = 1, \ldots, n$. Given the data, assume the model

$$y_i = f(t_{(i)}) + e_i, \tag{6}$$

where $y_i \equiv y(t_{(i)})$ and the errors $e_i = e(t_{(i)})$ are such that

$$\mathrm{Cov}\left(e(t_{(i)}), e(t_{(j)})\right) = \gamma(|i-j|) \text{ and } \lim_{n\to\infty} \sum_{u=-(n-1)}^{n-1} |\gamma(u)| < \infty. \tag{7}$$

Let $\hat{f}(t)$ be the estimator of $f(t)$ for all $t \in [0,1]$, where

$$\hat{f}(t) = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} \hat{d}_{j,k}\psi_{j,k}(t); \tag{8}$$

$\hat{d}_{j,k}$ is given by (4); and J' is the largest integer that $2^{J'} \le K\sqrt{n/\log n}$ for some chosen constant $K > 0$. The following theorem states our main result.

Theorem 1 *Suppose that model (6) is valid, the conditions (5) are met and $e_i = e(t_{(i)})$ are stationary Gaussian noise with zero mean satisfying conditions (7). Suppose also that the mother wavelet ψ has r vanishing moments and* *is compactly supported. Then the estimator $\hat{f}$ given by (8) achieves within a logarithmic factor almost the optimal convergence rate over the range of Hölder classes $\Lambda^\alpha(M)$ with $\alpha \in (0, r]$ in the sense that*

$$\sup_{f \in \Lambda^\alpha(M)} E||\hat{f} - f||_2^2 \leq C \left(\frac{\log n}{n}\right)^{2\alpha/(2+2\alpha)}$$

and

$$\sup_{f \in \Lambda^\alpha(M)} \frac{1}{n} \sum E||\widehat{f(t_k)} - f(t_k)||_2^2 \leq C \left(\frac{\log n}{n}\right)^{2\alpha/(2+2\alpha)},$$

for all $M \in (0, \infty)$.

In practice we usually choose the constant K such that $J' \geq J$, and the wavelet thresholding is performed on all the levels beginning at the level j_0.

Conditions (7) occur in diverse applications (see e.g. [12, 10, 13]), and specific cases of interest where conditions (5) also occur are given by the following propositions.

Proposition 1 *Let $\{e(t_{(i)}), i = 1, \ldots, n\}$ be a portion of a continuous-time zero-mean stationary Gaussian process $e(t)$, $t \in (0, 1)$, with the random points being jittered: $t_{(i)} = (2i-1)/(2n) + j_i$, where the j_i are IID uniform $[-1/(2n), 1/(2n)]$. Let $Cov(e(t_{(i)}), e(t_{(j)})) = \sigma^2 e^{-(n+1)\beta|t_{(i)} - t_{(j)}|}$ for some $\beta > 0$, $0 < \sigma^2 < \infty$ and fixed i and j. Then the conditions (5) and (7) hold for all $i, j = 1, \ldots, n$.*

Proposition 2 *Assume the same conditions as in Proposition 1, but now let the random points be such that $t_{(i)} \sim Beta(i, n-i+1)$, that is, the order statistics from independent realizations of a uniform $[0, 1]$ randon variable. Then the conditions (5) and (7) hold for all $i, j = 1, \ldots, n$.*

Two remarks are in order here. First, a sufficient condition for conditions (7) to hold is that $\mathrm{Cov}(e(t_{(i)}), e(t_{(j)})) \leq C\sigma^2 e^{-\beta|i-j|}$ for some positive constant $C < \infty$. Second, the covariance we assume in both propositions is 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.

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 equispaced samples. The package Wavethresh, implemented in R language, was used and the programs used can be obtained from us under request.

We considered three test functions $f(t)$, representing different degrees of spatial variability: sine, Heavisine and Doppler. The formulas for the last two

Table 1: Approximation of the IMSE $R(\hat{f}, f)$ over 200 replications of the test functions, calculated across the sampled times for each realization, from the simulation study. The Daubechies orthonormal compactly supported wavelet of length L=8 [6], least asymmetric family, was used with soft level-dependent thresholding beginning at the level j_0 indicated.

	SNR=5				SNR=7			
n	j_0	Equispaced	Jittered	Uniform	j_0	Equispaced	Jittered	Uniform
Sine								
256	2	0.80	0.82	1.17	2	0.41	0.42	0.72
512	2	0.41	0.41	0.64	2	0.21	0.21	0.38
1024	2	0.21	0.20	0.35	2	0.11	0.11	0.21
2048	2	0.12	0.12	0.20	2	0.06	0.06	0.12
Heavisine								
256	3	1.96	2.00	2.47	3	1.27	1.30	1.63
512	3	1.40	1.41	1.68	4	0.92	0.93	1.13
1024	3	1.01	1.01	1.18	3	0.74	0.75	0.88
2048	4	0.66	0.67	0.76	4	0.41	0.42	0.50
Doppler								
256	5	3.55	3.84	4.50	5	1.87	2.05	2.53
512	5	3.05	3.24	3.92	5	1.61	1.77	2.24
1024	5	2.52	2.59	3.28	6	1.40	1.44	1.79
2048	5	1.97	2.06	2.46	5	0.91	0.94	1.44

functions are in [7]. The sampled functions were normalized such that their standard deviations are equal to 10. We generated three samples of noise, one for each type of design, from the process described at Propostion 1 with $\beta = -\log(0.7)$ and $\sigma^2 = 1$. For the equispaced design, this corresponds to a discrete-time AR(1) process with coefficient $\phi = 0.7$. Then, the noise samples were standardized and added to each respectively sampled function, in order to compare the estimators at two noise levels, one with signal-to-noise ratio SNR=5 and another with SNR=7, where

$$\mathrm{SNR} = \frac{\sqrt{(n-1)^{-1}\sum_{i=1}^{n}\left(f(t_i)-\bar{f}\right)^2}}{\sqrt{Var(\text{noise})}},$$

and $\bar{f} = n^{-1}\sum_{i=1}^{n} f(t_i)$. 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. We take this as an approximation of the IMSE $R(\hat{f}, f)$. We used the Daubechies orthonormal compactly supported wavelet of length L=8 [6], 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 equispaced design with less IMSE and the σ_j values were estimated using the median absolute deviation around zero. The chosen level j_0 happened to be the one with less average MSE for the other designs in almost all cases. The constant $K = 125$ makes $J' \geq J$ in all sample sizes used.

Table 1 shows that the IMSE on random designs is bigger than those on equispaced design in all the cases. The IMSE for jittering fall between those

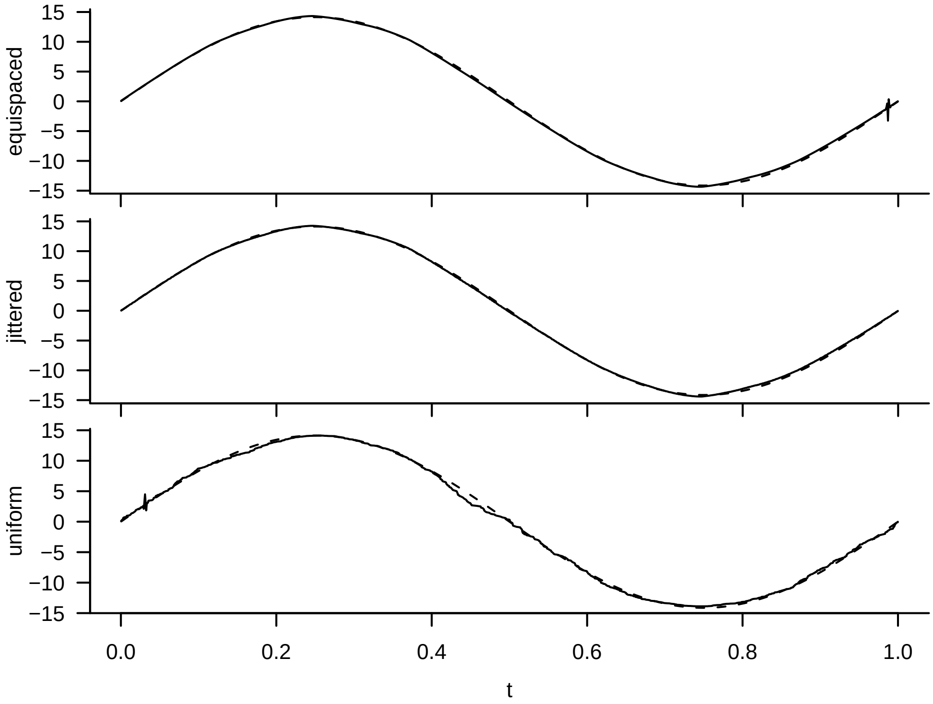

Figure 1: Sine test function and wavelet estimatives based on $n = 1024$ points and SNR=7. A Gaussian correlated noise was added to the test function. The Daubechies orthonormal compactly supported wavelet of length L=8 [6], least asymmetric family, was used with soft level-dependent thresholding beginning at the level $j_0 = 3$.

for uniform and equispaced in almost all the cases. However, the jittered sampling yields almost the same results as the equispaced design so that the efect of small timing errors is small, mainly for bigger sample sizes. Visually, the reconstruction with uniform design is a little more wrinkled than the equispaced and jittered designs. The jittering is visually almost indistinguishable from the equispaced 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 SNR=7.

5 Application

The uniform design can sometimes be useful to estimate the trend or the volatility of financial assets that are not traded every working day, or that are not sufficiently traded. If for a fixed period of time, the trades are uniformly distributed in this interval, then we can apply level-dependent wavelet shrinkage.

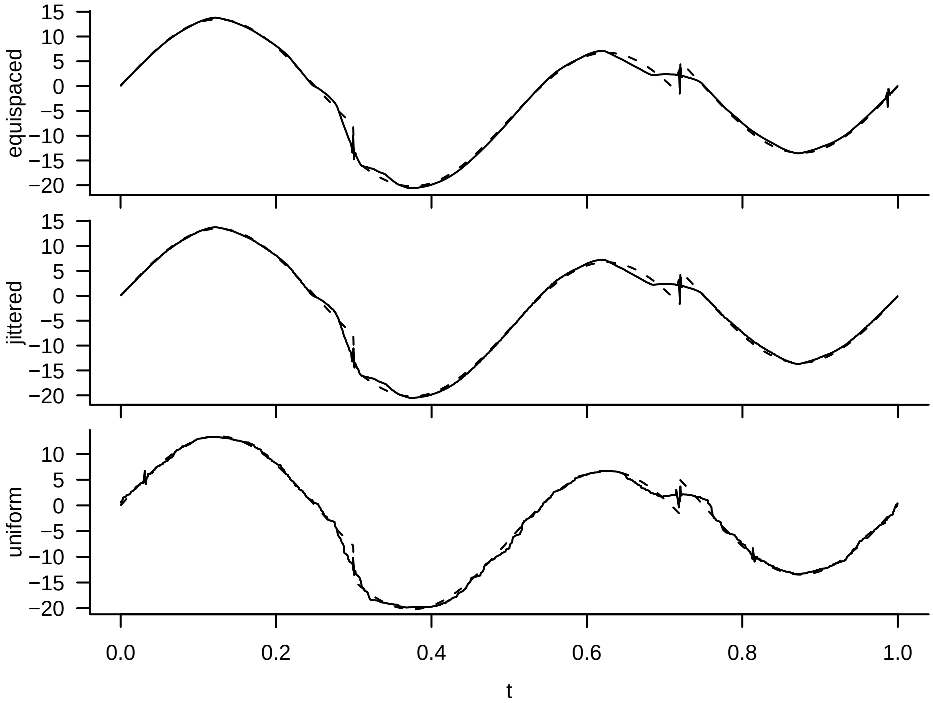

Figure 2: Heavisine test function and wavelet estimatives based on $n = 1024$ points and SNR=7. A Gaussian correlated noise was added to the test function. The Daubechies orthonormal compactly supported wavelet of length L=8 [6], least asymmetric family, was used with soft level-dependent thresholding beginning at level $j_0 = 4$.

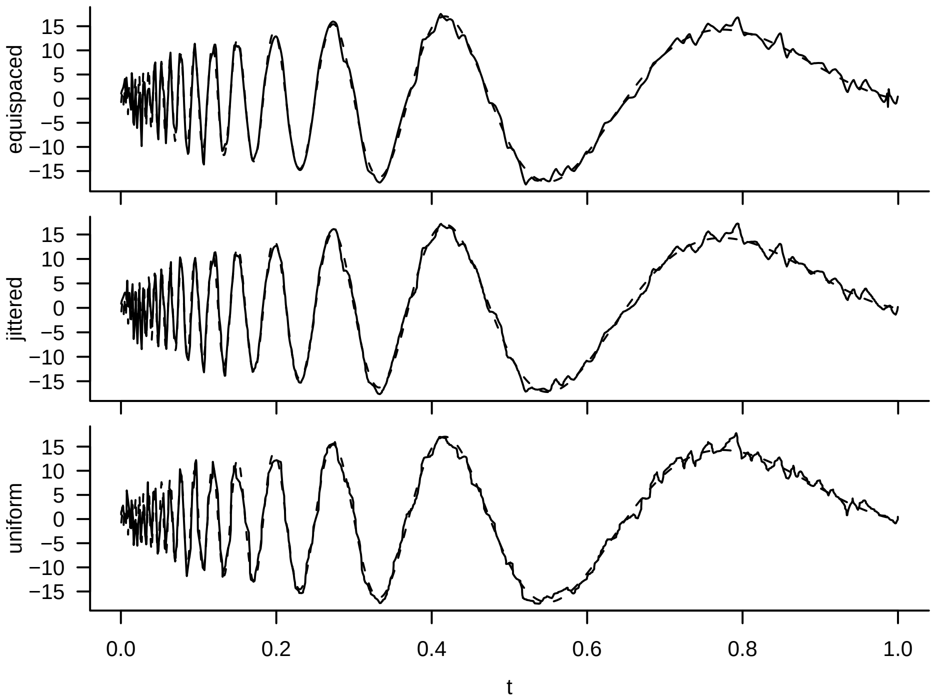

Figure 3: Doppler test function and wavelet estimatives based on $n = 1024$ points and SNR=7. A Gaussian correlated noise was added to the test function. The Daubechies orthonormal compactly supported wavelet of length L=8 [6], least asymmetric family, was used with soft level-dependent thresholding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

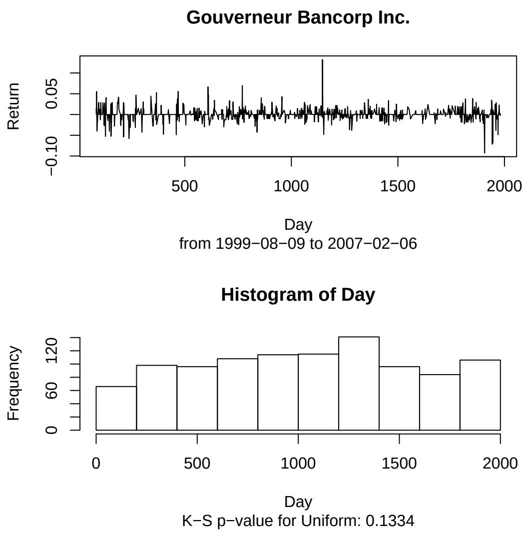

Figure 4: Closing prices simple returns of the Gouverneur Bancorp Inc. stocks and histogram of the days when they occurred. From August 9th, 1999 to February 6th, 2007.

One such case is the closing prices of the Gouverneur Bancorp Inc. stocks from August 9th, 1999 to February 6th, 2007. Its simple returns and an histogram of the days when they occurred are shown if Figure 4. These are not daily returns because the stocks were not traded every day in the period. A Kolmogorov-Smirnov test does not reject the hypothesis of uniform $[0, 1]$ distribution for the days, at 10% level (p-value=0.1334).

The wavelet empirical coefficients of the squared returns are shown in Figure 5 together with a robust variance estimate for each resolution level considered for thresholding. Signals with correlated errors generate wavelet coefficients with different level variances [7]. This seems to be the case here (as can be seen in the autocorrelation check graphic) and thus, we can shrink the coefficients using level-dependent thresholds. The soft thresholded coefficients and the estimated volatility are also shown in this figure.

The returns together with limits of ± 2 times the estimated volatility are shown in Figure 6. This volatility estimate is very rugged and emphasize some known stylized facts about financial returns. Also, the percentage of returns that fall beyond those limits is around 6%, which suggests an unconditional distribution for the returns that has tails a little bit fatter than the Gaussian distribution.

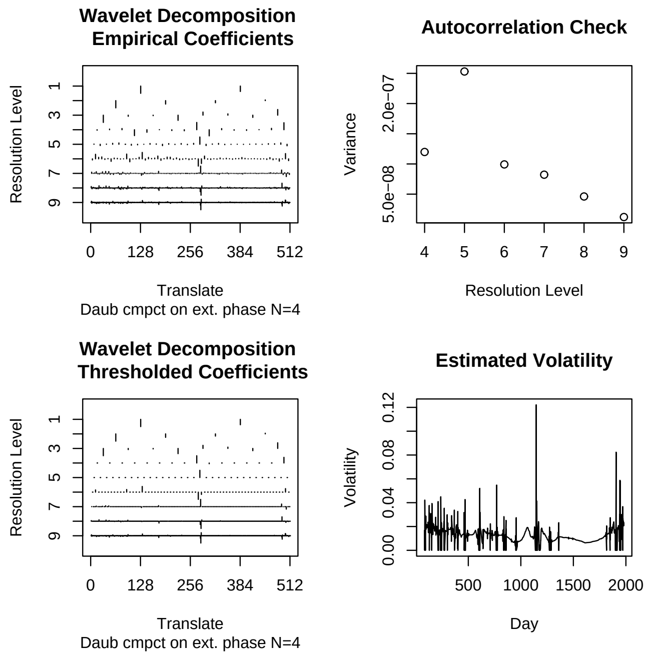

Figure 5: Wavelet analysis graphics for the closing prices simple returns of the Gouverneur Bancorp Inc. stocks of the Figure 4.

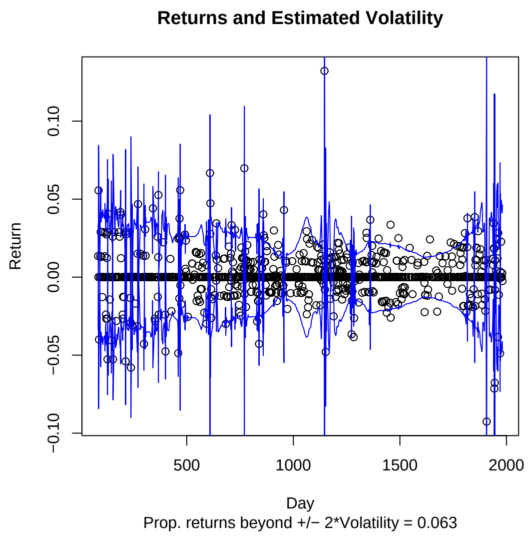

Figure 6: Closing prices simple returns of the Gouverneur Bancorp Inc. stocks (circles), together with limits (lines) of ± 2 times the volatility estimated by wavelet methods.

6 Conclusion

In this paper, we have considered the special cases of uniform and jittered samples 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, discrete wavelet transform followed by a level-dependent threshold policy give us estimators that adaptively achieve within a logarithmic factor of the optimal convergence rate across a range of Hölder classes. Therefore, for samples based on uniform design, there are good and fast algorithms for the estimation procedure.

A brief simulation study was carried in order to evaluate the numerical performance of the method. It was shown that the mean-squared error is comparable to that from samples with truly equispaced designs, as was the case of uncorrelated errors [8]. A financial application illustrated the usefullness of the method.

7 Proofs

7.1 Proof of Theorem 1

We need the following definition.

Definition 1 *Let $f^{(m)}$ denote the m-th derivative of a function f, and let $\lfloor\alpha\rfloor$ denote the largest integer less than α. For any positive real number α, the Hölder class $\Lambda^\alpha(M)$ on $[0,1]$ consists of functions f such that*

1. $|f(x)-f(y)| \le M|x-y|^\alpha$ *if* $0<\alpha\le 1$;
2. $|f^{(\lfloor\alpha\rfloor)}(x)-f^{(\lfloor\alpha\rfloor)}(y)| \le M|x-y|^{\alpha-\lfloor\alpha\rfloor}$ *and* $|f^{(1)}(x)|\le M$ *if* $\alpha>1$.

Let $y_i = f(t_{(i)}) + e_i$, where $e_1,\ldots,e_n$ are drawn from a stationary Gaussian process with $E(e_i)=0$, $\mathrm{Var}(e_i)=\sigma^2$ and $\mathrm{Cov}(e_r,e_s)=\gamma(|r-s|)$, for $i,r,s=1,\ldots,n$. Suppose that $\sum_{u=-\infty}^{\infty}|\gamma(u)|<\infty$. Let $t_{(1)}<\ldots<t_{(n)}$ be such that conditions (5) are valid. Let also $f\in\Lambda^\alpha(M)$ be fixed.

Using Definition 1, $|f(x)-f(y)|\le M|x-y|^{s(\alpha)}$, where $s(\alpha)=\min(\alpha,1)$. Thus, for any $f\in\Lambda^\alpha(M)$ the approximation error

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

$$\leq \frac{M}{n}\sum_{i=1}^{n}\left[\frac{1}{n}+\left(\frac{1}{\sqrt{n}}\right)^2\right]^{s(\alpha)}$$
$$= O(n^{-s(\alpha)}), \tag{9}$$

by Jensen's inequality and conditions (5).

Hereafter, $C_1, C_2, \ldots, C_{16}$ will denote positive constants that do not depend on n. Let

$$\begin{aligned}
\tilde{f}(x) &= \sum_{i=0}^{n-1} n^{-1/2} y_{i+1}\phi_{J,i}(x); && (10)\\
f_n(x) &= \sum_{i=0}^{n-1} n^{-1/2} f\left(\frac{i+1}{n}\right)\phi_{J,i}(x); && (11)\\
f(x) &= \sum_{k=0}^{n-1} c_{J,k}\phi_{J,k}(x) + \sum_{j=J}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(x), && (12)
\end{aligned}$$

where $c_{J,k} = \langle f, \phi_{J,k}\rangle$, $d_{j,k} = \langle f, \psi_{j,k}\rangle$, $n = 2^J$, $f \in \Lambda^\alpha$ as in Definition 1, with $\alpha > 0$.

Rewrite

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

where

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

Note that $A(x)$ is not random while $B(x)$ is random, but depends only on $\{t_i\}_{i=1}^n$. For some $j_0 \geq 0$ and some compactly supported wavelet basis $\{\phi_{j_0,k}, k = 0, \ldots, 2^{j_0}-1\} \cup \{\psi_{j,k}, j \geq j_0, k = 0, \ldots, 2^j-1\}$, ψ with $r \geq \alpha$ vanishing moments, let

$$c_{j_0,k} = \langle f, \phi_{j_0,k}\rangle, \quad \tilde{a}_{j_0,k} = \langle A, \phi_{j_0,k}\rangle, \quad \tilde{b}_{j_0,k} = \langle B, \phi_{j_0,k}\rangle, \quad \tilde{r}_{j_0,k} = \langle R, \phi_{j_0,k}\rangle;$$

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

$$d_{j,k} = \langle f, \psi_{j,k} \rangle, \quad a_{j,k} = \langle A, \psi_{j,k} \rangle, \quad b_{j,k} = \langle B, \psi_{j,k} \rangle, \quad r_{j,k} = \langle R, \psi_{j,k} \rangle;$$

$$\tilde{d}_{j,k} = d'_{j,k} + r_{j,k}, \text{ where } d'_{j,k} = d_{j,k} + a_{j,k} + b_{j,k}.$$

This wavelet basis can be different from the one used in equations (10), (11) and (12). In those equations, we will take the Haar scaling function

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

where $I(\cdot)$ denotes the usual indicator function. Then, for $k = 1, \ldots, n$,

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

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

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

Also let

$$\hat{\tilde{r}}_{j_0,k} = \frac{1}{n}\sum_{i=1}^{n} e_i \phi_{j_0,k}(i/n) \quad \text{and} \quad \hat{r}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} e_i \psi_{j,k}(i/n)$$

be estimators of $\tilde{r}_{j_0,k}$ and $r_{j,k}$, respectively, as given in [1]. Let

$$\hat{c}_{j_0,k} = c_{j_0,k} + \tilde{a}_{j_0,k} + \tilde{b}_{j_0,k} + \hat{\tilde{r}}_{j_0,k};$$

$$\hat{\tilde{d}}_{j,k} = d'_{j,k} + \hat{r}_{j,k}, \quad \hat{d}_{j,k} = \operatorname{sgn}(\hat{\tilde{d}}_{j,k})(|\hat{\tilde{d}}_{j,k} - \lambda|)_+,$$

where $\lambda = \sigma_{j,k}\sqrt{2n^{-1}\log n}$, $n^{-1}\sigma^2_{j,k} = \mathrm{Var}(\hat{r}_{j,k})$ and

$$\begin{aligned}
\mathrm{Var}(\hat{r}_{j,k}) &= \frac{1}{n^2}\mathrm{Cov}\left(\sum_{i=1}^{n} e_i\psi_{j,k}(i/n), \sum_{t=1}^{n} e_t\psi_{j,k}(t/n)\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\psi_{j,k}(i/n)\psi_{j,k}(t/n)Cov(e_i, e_t) \\
&\leq C_1||\psi||^2_\infty \frac{2^j}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}|\gamma(i-t)| \\
&= C_1||\psi||^2_\infty \frac{2^j}{n^2}\sum_{u=-(n-1)}^{n-1}|\gamma(u)|(n - |u|) \\
&\leq C_1||\psi||^2_\infty \frac{2^j}{n}\sum_{u=-(n-1)}^{n-1}|\gamma(u)| \\
&\leq C_2 2^j n^{-1}.
\end{aligned}$$

By an analogous argument $\mathrm{Var}(\hat{\tilde{r}_{j_0,k}}) \leq C_3 2^{j_0} n^{-1}$. Observe that $\hat{\tilde{d}_{j,k}} \sim N(d'_{j,k}, n^{-1}\sigma^2_{j,k})$.

Now let $\hat{f}(x)$ be an estimator of $f(x)$ for all $x \in [0,1]$, where

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

and J' is the largest integer such that $2^{J'} \leq K\sqrt{n/\log n}$, for some chosen constant $K > 0$. Then, the risk function is

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

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

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

and by the orthonormality of the wavelet basis

$$E\left(||\hat{f} - f||_2^2\right) = \sum_{k=0}^{2^{j_0}-1} E\left((\hat{c}_{j_0,k} - c_{j_0,k})^2\right) + \sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1} E\left((\hat{d}_{j,k} - d_{j,k})^2\right) + \sum_{j=J'}^{\infty}\sum_{k=0}^{2^j-1} d^2_{j,k}. \tag{13}$$

By Theorem 2.9.1 in [6] (see also Lemma 1 in [8]),

$$
\begin{aligned}
\sum_{j=J'}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}^2 &\le C_4 \sum_{j=J'}^{\infty} 2^j 2^{2(-j(1/2+\alpha))} = C_4 \sum_{j=J'}^{\infty} 2^{-2j\alpha} \\
&= C_4 \left(\sum_{j=0}^{\infty} 2^{-2j\alpha} - \sum_{j=0}^{J'-1} 2^{-2j\alpha} \right) \\
&= C_4 \frac{(2^{-2\alpha})^{J'}}{1-2^{-2\alpha}} \le C_4 (2^{J'})^{-2\alpha} \\
&= C_4 \left(2^{J'+1-1}\right)^{-2\alpha} = 2^{2\alpha} C_4 \left(2^{J'+1}\right)^{-2\alpha} \\
&\le 2^{2\alpha} C_4 \left(K^2 \frac{n}{\log n}\right)^{-2\alpha/2} = C_5 \left(\frac{\log n}{n}\right)^{\alpha} \\
&\le C_5 \left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2+2\alpha}} \qquad \forall \alpha > 0, n \ge 3.
\end{aligned}
$$

Throughout all the following text, we will use repeatedly a specific application of the (numerical) Hölder inequality: $(a+b)^2 \le 2a^2 + 2b^2$, $a, b \in \mathbb{R}$. Also let $E_1(Y) = E(Y|t_{(1)}, \ldots, t_{(n)})$ for any random variable Y.

We have

$$
\begin{aligned}
E\left((\hat{c}_{j_0,k} - c_{j_0,k})^2\right) &= E\left(E_1\left((\hat{c}_{j_0,k} - c_{j_0,k})^2\right)\right) \\
&= E\left(E_1\left((\hat{\tilde{r}}_{j_0,k} + \tilde{a}_{j_0,k} + \tilde{b}_{j_0,k})^2\right)\right) \\
&= E(\hat{\tilde{r}}_{j_0,k}^{\,2}) + 0 + E\left(E_1\left((\tilde{a}_{j_0,k} + \tilde{b}_{j_0,k})^2\right)\right) \\
&\le E(\hat{\tilde{r}}_{j_0,k}^{\,2}) + E\left(E_1\left(2\tilde{a}_{j_0,k}^2 + 2\tilde{b}_{j_0,k}^2\right)\right) \\
&= E(\hat{\tilde{r}}_{j_0,k}^{\,2}) + 2\tilde{a}_{j_0,k}^2 + 2E(\tilde{b}_{j_0,k}^2) \\
&\le C_3 2^{j_0} n^{-1} + 2\tilde{a}_{j_0,k}^2 + 2E(\tilde{b}_{j_0,k}^2),
\end{aligned}
$$

and then,

$$
\sum_{k=0}^{2^{j_0}-1} E\left((\hat{c}_{j_0,k} - c_{j_0,k})^2\right) \le C_3 2^{2j_0} n^{-1} + 2\sum_{k=0}^{2^{j_0}-1} \tilde{a}_{j_0,k}^2 + 2\sum_{k=0}^{2^{j_0}-1} E(\tilde{b}_{j_0,k}^2). \tag{14}
$$

We also have

$$
\begin{aligned}
E\left((\hat{d}_{j,k} - d_{j,k})^2\right) &= E\left((\hat{d}_{j,k} - d'_{j,k} + a_{j,k} + b_{j,k})^2\right) \\
&\le E\left(2(\hat{d}_{j,k} - d'_{j,k})^2 + 2(a_{j,k} + b_{j,k})^2\right) \\
&= E\left(E_1\left(2(\hat{d}_{j,k} - d'_{j,k})^2\right) + 2(a_{j,k} + b_{j,k})^2\right). \qquad (15)
\end{aligned}
$$

Let $n^{-1}\sigma^2_{j,k;1} = E_1(\hat{r}^2_{j,k})$. Denote $\min(x, y)$ by $x \wedge y$. Using Lemma 4 in [8], we obtain

$$\begin{aligned}
E_1\left((\hat{d}_{j,k} - d'_{j,k})^2\right) &\leq (2(d'_{j,k})^2 + n^{-2}\sigma^2_{j,k;1}) \wedge (2\log n + 1)n^{-1}\sigma^2_{j,k;1} \\
&\leq (2(d'_{j,k})^2 + n^{-2}\sigma^2_{j,k;1}) \wedge (2\log n + \log n + 1/n)n^{-1}\sigma^2_{j,k;1} \\
&\leq (2(d'_{j,k})^2 + n^{-2}\sigma^2_{j,k;1}) \wedge (3n^{-1}\sigma^2_{j,k;1}\log n + n^{-2}\sigma^2_{j,k;1}) \\
&= 2(d'_{j,k})^2 \wedge 3n^{-1}\sigma^2_{j,k;1}\log n + n^{-2}\sigma^2_{j,k;1}.
\end{aligned}$$

Now, use this result in (15):

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

and thus,

$$\begin{aligned}
&E\left((\hat{d}_{j,k} - d_{j,k})^2\right) \\
&\leq 8d^2_{j,k} \wedge 6n^{-1}E(\sigma^2_{j,k;1})\log n + 20a^2_{j,k} + 20E(b^2_{j,k}) + 2n^{-2}E(\sigma^2_{j,k;1}) \\
&= 8d^2_{j,k} \wedge 6n^{-1}\sigma^2_{j,k}\log n + 20a^2_{j,k} + 20E(b^2_{j,k}) + 2n^{-2}\sigma^2_{j,k} \\
&\leq 8d^2_{j,k} \wedge 6n^{-1}2^jC_2\log n + 20a^2_{j,k} + 20E(b^2_{j,k}) + 2n^{-2}2^jC_2.
\end{aligned}$$

Note that

$$\begin{aligned}
&8d^2_{j,k} \wedge 6n^{-1}2^jC_2\log n = 8d^2_{j,k} \\
&\Leftrightarrow\ 8d^2_{j,k} \leq 6n^{-1}2^jC_2\log n \\
&\Leftrightarrow\ d^2_{j,k}/2^j \leq 6/8C_2n^{-1}\log n.
\end{aligned} \tag{16}$$

Since by Theorem 2.9.1 in [6] (see also Lemma 1 in [8]),

$$\frac{d^2_{j,k}}{2^j} \leq \frac{C_4 2^{-j(1+2\alpha)}}{2^j} = C_4 2^{-j(2+2\alpha)}$$

for all $j \geq 0$, then if exist such J_1 that $C_4 2^{-j(2+2\alpha)} \leq 6/8C_2n^{-1}\log n$ for all $j \geq J_1$, then $d^2_{j,k}/2^j \leq C_4 2^{-j(2+2\alpha)} \leq 6/8C_2n^{-1}\log n$ and (16) will be true. To find J_1, note that

$$C_4 2^{-j(2+2\alpha)} \leq 6/8C_2n^{-1}\log n$$

$$\Leftrightarrow \quad 2^{-j} \le \left(6C_2/(8C_4)n^{-1}\log n\right)^{1/(2+2\alpha)} = C_6\left(\frac{\log n}{n}\right)^{1/(2+2\alpha)}$$
$$\Leftrightarrow \quad 2^{j} \ge 1/C_6\left(\frac{n}{\log n}\right)^{1/(2+2\alpha)}.$$

Thus, let J_1 be the smallest integer such that

$$2^{J_1} \ge 1/C_6\left(\frac{n}{\log n}\right)^{1/(2+2\alpha)}.$$

Then, since $J_1 \le J'$ for sufficiently large n,

$$\begin{aligned}
&\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1} E\left((\hat{d}_{j,k}-d_{j,k})^2\right) \\
\le\ & \frac{6C_2\log n}{n}\sum_{j=j_0}^{J_1-1}\sum_{k=0}^{2^j-1}2^j + 8\sum_{j=J_1}^{J'-1}\sum_{k=0}^{2^j-1}d_{j,k}^2 + 20\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1}a_{j,k}^2 + E(b_{j,k}^2) \\
&+\frac{2C_2}{n^2}\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1}2^j \\
\le\ & \frac{6C_2\log n}{n}\sum_{j=j_0}^{J_1-1}2^{2j} + 8C_4\sum_{j=J_1}^{J'-1}2^j2^{-j(1+2\alpha)} + 20\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1}a_{j,k}^2 + E(b_{j,k}^2) \\
&+\frac{2C_2}{n^2}\sum_{j=j_0}^{J'-1}2^{2j} \\
\le\ & \frac{6C_2\log n}{n}\left(\frac{2^{2J_1}-2^{2j_0}}{2^2-1}\right) + 8C_4\left(\frac{2^{-2\alpha J'}-2^{-2\alpha J_1}}{2^{-2\alpha}-1}\right) \\
&+20\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1}a_{j,k}^2 + E(b_{j,k}^2) + \frac{2C_2}{n^2}\left(\frac{2^{2J'}-2^{2j_0}}{2^2-1}\right) \\
\le\ & \frac{6C_2\log n}{n}2^{2J_1} + 8C_4\left(\frac{2^{-2\alpha J'}-2^{-2\alpha J_1}}{2^{-2\alpha}-1}\right) + 20\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1}a_{j,k}^2 + E(b_{j,k}^2) \\
&+\frac{2C_2}{n^2}2^{2J'} \\
=\ & \frac{6C_2\log n}{n}2^{2J_1} + 2^{2\alpha}8C_4\left(\frac{2^{-2\alpha J'}-2^{-2\alpha J_1}}{1-2^{2\alpha}}\right) + 20\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1}a_{j,k}^2 + E(b_{j,k}^2) \\
&+\frac{2C_2}{n^2}2^{2J'} \\
=\ & \frac{6C_2\log n}{n}2^{2J_1} + C_72^{-2\alpha J_1} - C_72^{-2\alpha J'} + 20\sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1}a_{j,k}^2 + E(b_{j,k}^2) + \frac{2C_2}{n^2}2^{2J'}.
\end{aligned}$$

In the last expression

$$\frac{6C_2 \log n}{n} 2^{2J_1} = \frac{6C_2 \log n}{n} 2^2 2^{2(J_1-1)} = \frac{24C_2 \log n}{n} (2^{J_1-1})^2$$

$$\leq \frac{24C_2 \log n}{n} \frac{1}{C_6^2} \left(\frac{n}{\log n}\right)^{\frac{2}{2+2\alpha}} = C_8 \left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2+2\alpha}},$$

$$C_7 2^{-2\alpha J_1} \leq C_7 \left[\frac{1}{C_6}\left(\frac{n}{\log n}\right)^{\frac{1}{2+2\alpha}}\right]^{-2\alpha} = C_9 \left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2+2\alpha}},$$

and

$$\frac{2C_2}{n^2} 2^{2J'} = \frac{2C_2}{n^2} \frac{Kn}{\log n} \leq \frac{2C_2 K}{n} \leq \frac{2C_2 K}{n^{\frac{2\alpha}{2+2\alpha}}} \leq 2C_2 K \left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2+2\alpha}}.$$

Thus

$$\sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} E\left((\hat{d}_{j,k} - d_{j,k})^2\right) \tag{17}$$

$$\leq \quad C_8 \left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2+2\alpha}} + C_9 \left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2+2\alpha}} + 20 \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} a_{j,k}^2 + E(b_{j,k}^2) + 2C_2 K \left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2+2\alpha}}$$

$$= \quad C_{10} \left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2+2\alpha}} + 20 \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} a_{j,k}^2 + E(b_{j,k}^2) \tag{18}$$

Now, collecting the second terms in the right hand side of the inequalities (14) and (18),

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

where

$$||A||_2^2$$

$$= \int_0^1 A(x)^2 dx = \int_0^1 [f_n(x) - f(x)]^2 dx$$

$$= \int_0^1 \left[\sum_{i=0}^{n-1} n^{-1/2} f\left(\frac{i+1}{n}\right) \phi_{J,i}(x) - \sum_{i=0}^{n-1} c_{J,i}\phi_{J,i}(x) - \sum_{j=J}^{\infty} \sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(x)\right]^2 dx$$

$$= \int_0^1 \left[\sum_{i=0}^{n-1} \left(n^{-1/2} f\left(\frac{i+1}{n}\right) - c_{J,i}\right) \phi_{J,i}(x) - \sum_{j=J}^{\infty} \sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(x)\right]^2 dx$$

$$\begin{aligned}
&\le \int_0^1 2\left[\sum_{i=0}^{n-1}\left(n^{-1/2}f\left(\frac{i+1}{n}\right)-c_{J,i}\right)\phi_{J,i}(x)\right]^2 dx + \int_0^1 2\left[\sum_{j=J}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(x)\right]^2 dx \\
&\le 2\sum_{i=0}^{n-1}\left(n^{-1/2}f\left(\frac{i+1}{n}\right)-c_{J,i}\right)^2 + 2\sum_{j=J}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}^2 \\
&= 2\sum_{i=0}^{n-1}\left(n^{-1/2}f\left(\frac{i+1}{n}\right)-c_{J,i}\right)^2 + O(n^{-2\alpha}) \le C_{11}n^{-2(1/2+s(\alpha))},
\end{aligned}$$

by Lemma 2(i) in [17] and equation (11) in [8].

Similarly, collecting the third terms in the right hand side of the inequalities (14) and (18),

$$\begin{aligned}
&\sum_{k=0}^{2^{j_0}-1} E(\tilde{b}_{j_0,k}^2) + \sum_{j=j_0}^{J'-1}\sum_{k=0}^{2^j-1} E(b_{j,k}^2) \\
&\le E||B||_2^2 = E(E_1||B||_2^2) \\
&= E\left(E_1\int_0^1 B(x)^2 dx\right) \\
&= E\left(E_1\int_0^1\left[\sum_{i=0}^{n-1} n^{-1/2} f\left(t_{(i+1)}\right)\phi_{J,i}(x) - f_n(x)\right]^2 dx\right) \\
&= E\left(E_1\int_0^1\left[\sum_{i=0}^{n-1} n^{-1/2} f\left(t_{(i+1)}\right)\phi_{J,i}(x) - \sum_{i=0}^{n-1} n^{-1/2} f\left(\frac{i+1}{n}\right)\phi_{J,i}(x)\right]^2 dx\right) \\
&= E\left(E_1\int_0^1\left[\sum_{i=0}^{n-1} n^{-1/2}\left\{f\left(t_{(i+1)}\right) - f\left(\frac{i+1}{n}\right)\right\}\phi_{J,i}(x)\right]^2 dx\right) \\
&= E\left(E_1\left(\sum_{i=0}^{n-1}\frac{1}{n}\left[f\left(t_{(i+1)}\right) - f\left(\frac{i+1}{n}\right)\right]^2\right)\right) \\
&= E\left(E_1\left(\frac{1}{n}\sum_{i=1}^{n}\left[f\left(t_{(i)}\right) - f\left(\frac{i}{n}\right)\right]^2\right)\right) \\
&\le C_{12}n^{-s(\alpha)},
\end{aligned}$$

using the orthogonality of the wavelet basis and the result (9).

Finally, from (13) and the calculations that follow it, we have

$$E\left(||\hat{f}-f||_2^2\right) \le C_3 2^{2j_0}/n + C_{13}/n^{2(1/2+s(\alpha))} + C_{14}/n^{s(\alpha)} + C_{10}(\log n/n)^{2\alpha/(2+2\alpha)} + C_5(\log n/n)^{2\alpha/(2+2\alpha)}.$$

But

$$C_3 2^{2j_0}/n \le C_3 2^{2j_0}/n^{2\alpha/(2+2\alpha)} \le C_{15}(\log n/n)^{2\alpha/(2+2\alpha)},$$

and for every $\alpha > 0$, $s(\alpha) = \min\{\alpha, 1\} \geq 2\alpha/(2+2\alpha)$ and

$$C_{14}/n^{s(\alpha)} \leq C_{14}(\log n/n)^{2\alpha/(2+2\alpha)}.$$

Also, since $s(\alpha) \geq 0$,

$$C_{13}/n^{2(1/2+s(\alpha))} \leq C_{13}/n \leq C_{13}(\log n/n)^{2\alpha/(2+2\alpha)}.$$

Thus,

$$E\left(||\hat{f} - f||_2^2\right) \leq C_{16}(\log n/n)^{2\alpha/(2+2\alpha)}$$

for $\alpha \geq 0$ and sufficiently large n.

7.2 Proof of Proposition 1

It is straightfoward to prove that conditions (5) hold. Since $t_{(i)} = t_i = (2i-1)/(2n) + j_i$, where j_i are IID uniform $[-1/(2n), 1/(2n)]$, then for all $n \geq 1$,

$$E\left(t_{(i)}\right) = E\left(t_i\right) = \frac{2i-1}{2n},$$

such that

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

and

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

Now let us prove that conditions 7 hold. Since $\mathrm{Cov}(e(r), e(s)) = \sigma^2 e^{-(n+1)\beta|r-s|}$, for some $\beta > 0$, $0 < \sigma^2 < \infty$ and fixed r and s, then,

$$\begin{aligned}
\mathrm{Cov}(e(t_{(r)}), e(t_{(s)})) &= E\left(E\left(e(t_{(r)})e(t_{(s)})|t_{(1)}, \ldots, t_{(n)}\right)\right) = E\left(\sigma^2 e^{-(n+1)\beta|t_{(r)}-t_{(s)}|}\right) \\
&= E\left(\sigma^2 \exp\left(-(n+1)\beta\left|\frac{r}{n+1} + j_r - \frac{s}{n+1} - j_s\right|\right)\right) \\
&= E\left(\sigma^2 \exp\left(-(n+1)\beta\left|\frac{r-s}{n+1} + j_r - j_s\right|\right)\right) = \gamma(|r-s|).
\end{aligned}$$

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

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

where $u = r - s$. Then,

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

7.3 Proof of Proposition 2

Since $\{e(t_{(i)}), i = 1, \ldots, n\}$ is a portion of a continuous-time zero-mean stationary Gaussian AR(1)-like process $e(t)$, $t \in (0,1)$ and the random points $0 < t_{(1)} < \ldots < t_{(n)} < 1$ are such that $t_{(i)} \sim \text{Beta}(i, n-i+1)$, then conditions (5) hold. In fact, $E\left(t_{(i)}\right) = 1/(n+1)$ implies that

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

and

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

Now, since $\text{Cov}(e(t_{(i)}), e(t_{(j)})) = \sigma^2 e^{-(n+1)\beta|t_{(i)}-t_{(j)}|}$, for some $\beta > 0$, $0 < \sigma^2 < \infty$ and fixed i and j, then conditions 7 also hold.

To see this, note that ([15], p.217):

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

Then,

$$\begin{aligned}
\text{Cov}(e(t_{(i)}), e(t_{(j)})) &= E\left(E\left(e(t_{(i)})e(t_{(j)})|t_{(1)}, \ldots, t_{(n)}\right)\right) = E\left(\sigma^2 e^{-(n+1)\beta|t_{(i)}-t_{(j)}|}\right) \\
&= \sigma^2 \sum_{k=0}^{\infty} (-1)^k \frac{(n+1)^k \beta^k}{k!} E\left(\left|t_{(i)} - t_{(j)}\right|^k\right) \\
&= \sigma^2 \sum_{k=0}^{\infty} (-1)^k \frac{(n+1)^k \beta^k}{k!} \frac{(|i-j|+k-1)!n!}{(|i-j|-1)!(n+k)!} = \gamma(|i-j|).
\end{aligned}$$

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

$$\begin{aligned}
|\gamma(|i-j|)| &= \left|E\left(\sigma^2 e^{-(n+1)\beta|t_{(i)}-t_{(j)}|}\right)\right| \\
&= E\left(\sigma^2 e^{-(n+1)\beta|t_{(i)}-t_{(j)}|}\right) = \gamma(|i-j|).
\end{aligned}$$

Note also that

$$\begin{aligned}
\sum_{u=1}^{n-1} \frac{(u+k-1)!}{(u-1)!} &= \sum_{v=0}^{n-2} \frac{(v+k)!}{v!} \\
&= k! \sum_{v=0}^{n-2} \frac{(v+k)!}{v!k!} \\
&= k! \binom{k+(n-2)+1}{k+1} \qquad (19) \\
&= \frac{k!(n+k-1)!}{(k+1)!(n-2)!} \\
&= \frac{(n+k-1)!}{(k+1)(n-2)!},
\end{aligned}$$

where (19) comes from the equation 0.151.1 in [19]. Then using these facts,

$$\sum_{u=1}^{n-1}|\gamma(u)| = \sum_{u=1}^{n-1}\gamma(u) = \sum_{u=1}^{n-1}\sigma^2\sum_{k=0}^{\infty}(-1)^k\frac{(n+1)^k\beta^k}{k!}\frac{(u+k-1)!n!}{(u-1)!(n+k)!} \quad (20)$$

$$= \sigma^2\sum_{k=0}^{\infty}(-1)^k\frac{(n+1)^k\beta^k n!}{k!(n+k)!}\sum_{u=1}^{n-1}\frac{(u+k-1)!}{(u-1)!} \quad (21)$$

$$= \sigma^2\sum_{k=0}^{\infty}(-1)^k\frac{(n+1)^k\beta^k n!}{k!(n+k)!}\frac{(n+k-1)!}{(k+1)(n-2)!}$$

$$= \sigma^2\sum_{k=0}^{\infty}(-1)^k\frac{(n+1)^k\beta^k}{k!}\frac{n(n-1)}{(n+k)(k+1)}. \quad (22)$$

One way to justify the switch of the summations from (20) to (21) is the following. In (20), let

$$g_k(u) = \sigma^2(-1)^k\frac{(n+1)^k\beta^k}{k!}\frac{(u+k-1)!n!}{(u-1)!(n+k)!},$$

and note that

$$\begin{aligned}
|g_k(u)| &= \sigma^2\frac{(n+1)^k\beta^k}{k!}\frac{(u+k-1)!n!}{(u-1)!(n+k)!}\\
&< \sum_{k=0}^{\infty}\sigma^2\frac{(n+1)^k\beta^k}{k!}\frac{(u+k-1)!n!}{(u-1)!(n+k)!}\\
&= \sigma^2\sum_{k=0}^{\infty}\frac{(n+1)^k\beta^k}{k!}\frac{(|i-(i+u)|+k-1)!n!}{(|i-(i+u)|-1)!(n+k)!}\\
&= \sigma^2\sum_{k=0}^{\infty}\frac{(n+1)^k\beta^k}{k!}E\left(\left|t_{(i)}-t_{(i+u)}\right|^k\right)\\
&= E\left(\sigma^2 e^{(n+1)\beta|t_{(i)}-t_{(i+u)}|}\right)\\
&\le \sigma^2 e^{(n+1)\beta},
\end{aligned}$$

for $u = 1,\ldots,n-1$, and for all $k \ge 0$. Let u be the counting measure so that

$$\begin{aligned}
\sum_{u=1}^{n-1}\sum_{k=0}^{\infty}g_k(u) &= \sum_{u=1}^{n-1}\lim_{\kappa\to\infty}\sum_{k=0}^{\kappa}g_k(u)\\
&= \int_1^{n-1}\lim_{\kappa\to\infty}\sum_{k=0}^{\kappa}g_k(u)du\\
&= \lim_{\kappa\to\infty}\int_1^{n-1}\sum_{k=0}^{\kappa}g_k(u)du \qquad (23)
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{\kappa\to\infty} \sum_{k=0}^{\kappa} \int_1^{n-1} g_k(u)du \\
&= \sum_{k=0}^{\infty} \sum_{u=1}^{n-1} g_k(u),
\end{aligned}
$$

where the Dominated Convergence Theorem was used in (23).

Evaluating the summation in (22), we have that

$$
\begin{aligned}
&\sum_{k=0}^{\infty} \frac{(-1)^k (n+1)^k \beta^k}{k!} \frac{1}{(n+k)(k+1)} \\
&= \sum_{k=0}^{\infty} \frac{(-1)^k (n+1)^k \beta^k}{k!} \frac{\Gamma(n+k)}{\Gamma(n+k+1)} \frac{\Gamma(k+1)}{\Gamma(k+2)} \\
&= \sum_{k=0}^{\infty} \frac{(-1)^k (n+1)^k \beta^k}{k!} \frac{\frac{\Gamma(n+k)}{\Gamma(n)}\Gamma(n)}{\frac{\Gamma(n+k+1)}{\Gamma(n+1)}\Gamma(n+1)} \frac{\frac{\Gamma(k+1)}{\Gamma(1)}\Gamma(1)}{\frac{\Gamma(k+2)}{\Gamma(2)}\Gamma(2)} \\
&= \sum_{k=0}^{\infty} \frac{(-1)^k (n+1)^k \beta^k}{k!} \frac{(n)_k (1)_k}{(n+1)_k (2)_k} \frac{1}{n} \\
&= \frac{1}{n} {}_2F_2(n, 1; n+1, 2; (-1)(n+1)\beta), \qquad (24)
\end{aligned}
$$

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

$$
\begin{aligned}
&\frac{1}{b-a}\left(b\ {}_1F_1(a, a+1, z) - a\ {}_1F_1(b, b+1, z)\right) \\
&= \frac{1}{b-a}\left(b \sum_{k=0}^{\infty} \frac{(a)_k}{(a+1)_k} \frac{z^k}{k!} - a \sum_{k=0}^{\infty} \frac{(b)_k}{(b+1)_k} \frac{z^k}{k!}\right) \\
&= \frac{1}{b-a}\left(b \sum_{k=0}^{\infty} \frac{b+k}{b+k} \frac{(a)_k}{(a+1)_k} \frac{z^k}{k!} - a \sum_{k=0}^{\infty} \frac{a+k}{a+k} \frac{(b)_k}{(b+1)_k} \frac{z^k}{k!}\right) \\
&= \frac{1}{b-a}\left(\sum_{k=0}^{\infty} \frac{(b+k)(b)_k}{(b+1)_k} \frac{(a)_k}{(a+1)_k} \frac{z^k}{k!} - \sum_{k=0}^{\infty} \frac{(a+k)(a)_k}{(a+1)_k} \frac{(b)_k}{(b+1)_k} \frac{z^k}{k!}\right) \\
&= \frac{1}{b-a}\left(b \sum_{k=0}^{\infty} \frac{(b)_k}{(b+1)_k} \frac{(a)_k}{(a+1)_k} \frac{z^k}{k!} - a \sum_{k=0}^{\infty} \frac{(a)_k}{(a+1)_k} \frac{(b)_k}{(b+1)_k} \frac{z^k}{k!}\right) \\
&= \frac{b-a}{b-a} \sum_{k=0}^{\infty} \frac{(b)_k}{(b+1)_k} \frac{(a)_k}{(a+1)_k} \frac{z^k}{k!} \\
&= {}_2F_2(a, b; a+1, b+1; z),
\end{aligned}
$$

and applying this result to equation (24),

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

From equation 9.236.4 in [19], applying

$${}_1F_1(a,a+1,z) = a(-z)^{-a}\left(\Gamma(a) - \Gamma(a,-z)\right) \tag{25}$$

to the last expression we have that

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

where $\Gamma(n,a) = \int_a^{\infty} t^{n-1}e^{-t}dt$ denotes the incomplete gamma function. Using (25) we also have that

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

Thus, for $n > 1$,

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

where the incomplete gamma function

$$\Gamma(n,\beta(n+1)) = \int_{\beta(n+1)}^{\infty} t^{n-1}e^{-t}dt$$

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

when n is an integer. Applying these results in (22), we get that for every $n > 1$,

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

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

Acknowledgments

The authors acknowledge the partial support of CNPq, FAPESP (grant 03/10105-2). RFP was also partially supported by Bank of Brazil.

References

[1] MORETTIN, P. A. **Waves and Wavelets**. University of São Paulo Press: São Paulo, 1999.

[2] DIPPÉ, M.A.Z. and Wold, E.H. Antialiasing through stochastic sampling. **ACM SIGGRAPH Computer Graphics**, v.19, nr. 3, p.69–78, 1985.

[3] PAPOULIS, A. **Probability, Random Variables, and Stochastic Processes**. McGraw-Hill Book Company, Inc.: New York, 1965.

[4] MOOD, A.M., GRAYBILL, F.A. and BOES, D.C. **Introduction to the Theory of Statistics**, 3rd ed. McGraw-Hill: Singapore, 1974.

[5] FELLER, W. **An Introduction to Probability Theory and Its Application**, Vols. 1 and 2, 3rd ed. Wiley: New York, 1968.

[6] DAUBECHIES, I. **Ten Lectures on Wavelets**. CBMS-NSF Regional Coference Series in Applied Mathematics – SIAM:Philadelphia, Pennsylvania, 1992.

[7] DONOHO, D. L. and JOHNSTONE, I. M. Ideal spatial adaptation via wavelet shrinkage. **Biometrika**, v.81, p.425–455, 1994.

[8] CAI, T.T. and BROWN, L. Wavelet estimation for samples with random uniform design. **Statistics and Probability Letters**, v.42, p.313–321, 1999.

[9] JOHNSTONE, I. M. and SILVERMAN, B. W. Wavelet threshold estimators for data with correlated noise. **JRSS-B**, v.59, nr. 2, p.319–351, 1997.

[10] KUTNER, M. H., NACHTSHEIM, C. J., NETER, J. and LI, W. **Applied Linear Statistical Models**, 5th ed. McGraw-Hill/Irwin: New York, 2004.

[11] PYKE, R. Spacings. **JRSS-B**, v.27, nr. 3, p.395–449, 1965.

[12] COCHRANE, D. and ORCUTT, G. Application of least squares regression to relationships containing autocorrelated error terms. **JASA**, v.44, p.32–61, 1949.

[13] QIN, D. and GILBERT, C. L. The error term in the history of time series econometrics. **Econometric Theory**, v.17, nr. 2, p.424–450, 2001.

[14] DAVID, H.A. **Order Statistics**, 2ed., Wiley:New York, 1981.

[15] JOHNSON, N.L., KOTZ, S. and BALAKRISHNAN, N. **Continuous Univariate Distributions**, v.2, 2ed., Wiley:New York, 1995.

[16] SARDY, S., PERCIVAL, D.B., BRUCE, A.G., GAO, H. and STUETZLE, W. Wavelet shrinkage for unequally spaced data. **Statistics and Computing**, v.9, p.65–75, 1999.

[17] CAI, T. T. and BROWN, L. Wavelet shrinkage for nonequispaced samples. **The Annals of Statistics**, v.26, nr. 5, p.1783–1799, 1998.

[18] KERKYACHARIAN, G. and PICARD, D. Regression in random design and warped wavelets. **Bernoulli**, v.10, nr.6, p.1053–1105, 2004.

[19] GRADSHTEĬN, I.S., RYZHIK, I.M., JEFFREY, A. and ZWILLINGER, D. **Table of Integrals, Series, and Products**, 7ed, Academic Press:Amsterdan, 2007.

[20] PORTO, R.F., MORETTIN, P.A. and AUBIN, E.C.Q. Wavelet regression with correlated errors on a piecewise Hölder class. **Statistics and Probability Letters**, v.78, p.2739–2743, 2008.