

WAVELET SMOOTHING FOR DATA WITH AUTOCORRELATED ERRORS

**ROGÉRIO F. PORTO, JOÃO R. SATO,
ELISETE C. Q. AUBIN and PEDRO A. MORETTIN**

Institute of Mathematics and Statistics
University of São Paulo, Rua do Matão 1010
05508-090-São Paulo-SP, Brazil
e-mail: rdporto@ime.usp.br

Abstract

This paper presents an alternative approach to wavelet smoothing procedure for a time series model of signal added to autocorrelated stationary errors. The aim is to estimate the signal globally with near minimum risk. The usual approach to this problem is to threshold the wavelet coefficients with different thresholds in each level. In this paper, the autocorrelation is taken account in a parametric way letting the wavelet methods for the function estimation only. Thus, an iterative semi-parametric method is proposed. This iterative method borrows some ideas from the Cochrane-Orcutt procedure. The simulation results show that the proposed method is at least as good as other benchmark wavelet methods independent of the type of autocorrelation present in the error term.

1. Introduction

Let y_1, y_2, \dots, y_n be a sequence of observations from the model

$$Y_t = f(t/n) + \varepsilon_t, \quad (1)$$

2000 Mathematics Subject Classification: 62M10, 62G05.

Keywords and phrases: autocorrelation, semi-parametric estimation, smoothing, wavelets.

This work was partially supported by CNPq and FAPESP (grant 03/10105-2).

Received February 7, 2007

where $t = 1, 2, \dots, n$, Y_t is a random variable, f is an unknown function to be estimated, and the sequence $\{\varepsilon_t\}_{t=1}^n$ is a stationary zero mean Gaussian process. This process can be autocorrelated or not.

The goal is to estimate f globally. One measure of the performance of an estimator \hat{f} is the global squared L_2 norm risk,

$$R(\hat{f}, f) = E\left(\int_0^1 (\hat{f}(t) - f(t))^2 dt\right),$$

where the integral must be well defined.

There are some methods for estimation of f in this context using kernels, splines and wavelets. Wavelet methods are appropriate mainly when the function f is non stationary with transients and fractal-like structures [10]. The function f is estimated by a wavelet series expansion where the coefficients are estimated using shrinkage or threshold techniques. These techniques were pioneered in the wavelet context by Donoho and Johnstone [3, 4].

Although these techniques were developed for non-correlated errors, wavelet methods were extended to the situation when they are correlated [8]. In this article, it is developed an alternative approach to wavelet smoothing for the model with autocorrelated stationary errors.

The paper is organized as follows. Section 2 reviews some basics on wavelets and wavelets for correlated errors. Section 3 describes the proposed method and Section 4 shows some simulation results. In Section 5 the proposed method is applied to a real data set and in Section 6 some aspects of the proposed method are discussed and some suggestions for future research are given.

2. Wavelets and Wavelet Approximations

Some basic wavelet results are summarized in this section for the sake of notation.

An orthonormal wavelet basis is generated from dilations and translations of a “father” wavelet ϕ (or scaling function) and a “mother”

wavelet ψ . These functions are assumed to be compactly supported in $[0, N]$ and ϕ satisfies $\int \phi = 1$. A wavelet is *r-regular* if it has r vanishing moments and r continuous derivatives.

Let \mathbb{Z} be the set of all integers and

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k), \quad \psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \quad j, k \in \mathbb{Z}.$$

It follows that $\psi_{j,k}$ has support $[2^{-j}k, 2^{-j}(N+k)]$. Denote the periodized wavelets by

$$\phi_{j,k}^p(t) = \sum_{l \in \mathbb{Z}} \phi_{j,k}(t-l), \quad \psi_{j,k}^p(t) = \sum_{l \in \mathbb{Z}} \psi_{j,k}(t-l),$$

for $t \in [0, 1]$. These are the wavelets to be used in this paper and so the superscript “ p ” will be suppressed thereafter. The collection

$$\phi_{j_0,k}, k = 1, \dots, 2^{j_0}; \quad \psi_{j,k}, j \geq j_0, k = 1, \dots, 2^j$$

constitutes an orthonormal basis of $L_2[0, 1]$, for some coarse scale j_0 .

An orthonormal wavelet basis has an associated exact orthogonal discrete wavelet transform that transforms sampled data into the wavelet coefficient domain. To see this denote inner product by $\langle \cdot, \cdot \rangle$. For a given square-integrable function f on $[0, 1]$, denote

$$c_{j,k} = \langle f, \phi_{j,k} \rangle, \quad d_{j,k} = \langle f, \psi_{j,k} \rangle. \quad (2)$$

So the function f can be expanded into a wavelet series:

$$f(t/n) = \sum_{k=1}^{2^{j_0}} c_{j_0,k} \phi_{j_0,k}(t/n) + \sum_{j=j_0}^{\infty} \sum_{k=1}^{2^j} d_{j,k} \psi_{j,k}(t/n). \quad (3)$$

Wavelets completely characterize many traditional smoothness spaces like Hölder, Sobolev and Besov spaces but this will not be of concern here.

We can rewrite this discrete wavelet transform W in matrix form applied to Y which is the n -vector of observations Y_1, \dots, Y_n . Suppose

that $n = 2^J$ for some natural number J and take $J_0 = 0$, then write

$$d_{j,k} = (WY)_{j,k}, \quad j = 0, 1, \dots, J-1, \quad k = 1, \dots, 2^j, \quad (4)$$

where the remaining element $c_{j_0,1}$ is generically labeled as $d_{-1,1}$. Let

$\theta = Wf$ be the wavelet transform of the signal $f = (f(t/n))_{t=1}^n$ and $z = W\varepsilon$ be the wavelet transform of the noise.

For the construction of the estimator, define the hard threshold function

$$\eta_H(d, \lambda) = dI(|d| \geq \lambda),$$

where $I(\cdot)$ denotes the indicator function. The threshold λ is commonly chosen to be $\lambda = \sigma\sqrt{2 \log n}$ when the errors ε_i are independently identically distributed (i.i.d.) $N(0, \sigma^2)$ random variables with known σ^2 .

In practice one may replace the hard threshold function by a soft one

$$\eta_S(d, \lambda) = \text{sign}(d)(|d| - \lambda).$$

Soft thresholding is generally smoother than hard thresholding and has better and smoother visual presentation.

The Donoho-Johnstone estimator [3] is then constructed by thresholding the wavelet coefficients (4) at threshold λ and then transforming back. Thus we define $\hat{\theta}$ by

$$\hat{\theta}_{j,k} = \eta(d_{j,k}, \lambda)$$

and the estimator \hat{f} by

$$\hat{f} = W^T \hat{\theta}, \quad (5)$$

where η denotes soft or hard thresholding, or some compromise between the two, and T denotes transpose.

Now, suppose that the errors ε_i have a multivariate normal distribution with mean 0 and covariance matrix Γ_n , as in model (1). Since we have assumed that the errors are stationary, Γ_n has entries

$\eta_{|s-t|}$, say. Let $V_n = \text{Cov}(z)$ be the covariance matrix of the wavelet transform of the error vector so that it can be obtained by the following orthogonal transform:

$$V_n = W\Gamma_n W^T.$$

It can be shown [8, 5] that since the filters used to construct the discrete wavelet transform are time invariant, within each level the distribution of the $z_{j,k}$ will be stationary, and the variance of $z_{j,k}$ will depend only on the level j . We write

$$\sigma_j^2 = \text{Var}(z_{j,k}),$$

for each j .

In view of this, a natural extension of the usual wavelet thresholding method is to apply level-dependent thresholding to the transformed data d . Let λ_j be a sequence of thresholds to be applied to the coefficients at level j , and define $\hat{\theta}$ to be the estimator

$$\hat{\theta}_{j,k} = \eta(d_{j,k}, \sigma_j \lambda_j),$$

with the corresponding estimator of θ denoted by $\hat{\theta}$, and set

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

In practice, one may wish to threshold only coefficients at higher levels, where there is a considerable number of coefficients at each level and the signal $\theta_{j,k}$ can be assumed to be sparse. Also, the noise variance σ_j^2 can be estimated from the data [3], usually through the use of robust estimators.

Finally, the choice of $\lambda = \sigma\sqrt{2 \log n}$ is attractive from certain theoretical perspectives since it is conservative for the following reason. If Z_1, \dots, Z_n are normally distributed random variables with mean 0 and variances σ_i^2 , then

$$P(\max_{1 \leq i \leq n} |Z_i/\sigma_i| > \sqrt{2 \log n}) \rightarrow 0 \tag{6}$$

as $n \rightarrow \infty$ whether or not the variables are independent [8].

This consistency property will be useful for the development of our estimator, which will be discussed in the next section.

3. Proposed Iterative Method

In this section it is described the proposed method in an intuitive way and it is given some justification for its performance.

The proposed estimator is obtained as a result of an iterative process which borrow some ideas from the Cochrane-Orcutt procedure [1]. The useful consistency property of the thresholding policy described earlier is also similar to what is obtained from least squares. Time series regression, using time as explanatory variable and ordinary least squares, leads to unbiased and consistent estimators when the time series is stationary, since they generally satisfy the Grenander conditions [6]. As least squares estimators, the use of wavelet thresholding ignoring error correlation gives non-optimal estimators. Thus, some improvement can be achieved through level-thresholding algorithms.

Instead of applying level-thresholding algorithms, the proposed method is to apply simple wavelet thresholding and estimate the error covariance matrix from the resulting residuals. This process would be iterated until a convergence criterion be achieved.

Suppose we postulate model (1) and use estimator (5) as a primary estimator. Then we obtain the residuals

$$e_t = y_t - \hat{f}(t/n), \quad t = 1, \dots, n.$$

Since \hat{f} is a consistent estimator of the true f function ($R(\hat{f}, f) \xrightarrow{n \rightarrow \infty} 0$), the residuals e_t are good predictors of the errors ε_t , $t = 1, \dots, n$. Since the sequence $\{\varepsilon_t\}_1^n$ is a stationary zero mean Gaussian process, it can be modeled as an autoregressive process of sufficient large order p , briefly $AR(p)$. This can be achieved in an automatic way using any information criterium like Akaike (AIC) or Schwarz's BIC . We propose to choose p as the order that minimizes the AIC of the AR model adjusted to the residual sequence $\{e_t\}_1^n$. Denote this model by

$$e_t = \mu + \phi_1 e_{t-1} + \dots + \phi_p e_{t-p} + a_t, \quad (7)$$

where a_t are i.i.d. $N(0, \sigma_\varepsilon^2)$ random variables, for $t = 1, \dots, n$.

The estimates of ϕ_1, \dots, ϕ_p in (7) can be used to construct an estimate of the covariance matrix Γ_n described in Section 2 under *AR* setting. Being $\{e_t\}_1^n$ a real stationary zero mean process and autocovariance function $\gamma_\tau = E(e_t e_{t+\tau})$, then if e_t follows an *AR(p)* process, it can be shown [10] that for $\tau \geq 0$,

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1 \rho_1 - \dots - \phi_p \rho_p}, \quad \text{for } \tau = 0, \quad (8)$$

and

$$\gamma_\tau = \phi_1 \gamma_{\tau-1} + \phi_2 \gamma_{\tau-2} + \dots + \phi_p \gamma_{\tau-p}, \quad \text{for } \tau > 0. \quad (9)$$

The idea is that the estimates $\hat{\gamma}_\tau \approx \eta_{|s-t|}$ for $\tau = |s-t|$, where $\eta_{|s-t|}$ are the entries of the covariance matrix Γ_n giving rise to an estimated covariance matrix $\hat{\Gamma}_n$.

Once we obtain $\hat{\Gamma}_n$, we are ready to decorrelate the data in order to obtain an efficient wavelet smoother of the function f . Denote by

$$\hat{\Gamma}_n = \hat{S}_n \hat{S}_n'$$

the Cholesky decomposition of the estimated covariance matrix since it should be a positive definite matrix. To decorrelate the original errors, simply do

$$Y' = \hat{S}_n^{-1} Y,$$

then obtain an efficient wavelet estimator through simple thresholding, \hat{f}' , say.

The final step in this procedure is to apply back \hat{S}_n to \hat{f}' putting back all the covariance structure that was estimated in order to get a new estimate for the true function f . Let us call this estimate $\hat{f}^{(1)}$.

This process can be iterated producing estimates $\hat{f}^{(2)}, \dots, \hat{f}^{(s)}, \dots$ until a selected convergence criterion is achieved. Since the main concern of wavelet thresholding methods is the minimization of the risk $R(\hat{f}, f)$, a suitable choice is to minimize the mean square error (*MSE*) defined by

$$MSE(s) = \frac{1}{n} \sum_{t=1}^n (\hat{f}^{(s)}(t/n) - Y_t)^2, \quad (10)$$

where $\hat{f}^{(s)}(t/n)$ is the estimate obtained from the s -th iteration through the application of an inverse discrete wavelet transform. The iterative process goes on until $\min_s MSE(\hat{f}^{(s)}, Y)$ is achieved. Simulations show that convergence is achieved in a few iterations, commonly 4 or 5, and the entire process is very fast even for large n , e.g., $n = 2048$.

In short, the suggested steps are:

1. postulate model (1) and use estimator (5), $\hat{f}^{(0)} = W^T \hat{\theta}$, as an initial estimator and calculate $MSE(0)$;
2. calculate the residuals $e_t = y_t - \hat{f}^{(0)}(t/n)$, $t = 1, \dots, n$;
3. fit an autoregressive model $AR(p)$ to the residuals choosing the order p that minimizes *AIC* and do $s = 0$;
4. use the estimates of ϕ_1, \dots, ϕ_p in the fitted model to construct an estimate of the covariance matrix Γ_n through the difference equation (9);
5. find the Cholesky decomposition of the estimated error covariance matrix denoted by \hat{S}_n , and decorrelate the residuals by letting $Y' = \hat{S}_n^{-1}Y$;
6. apply a wavelet thresholding algorithm to the decorrelated data Y' in order to obtain an estimator \hat{f}' ;
7. calculate the residuals $e_i^{(s)} = y_i - \hat{f}^{(s)}(i/n)$, $i = 1, \dots, n$;
8. do $s = s + 1$; multiply \hat{S}_n to \hat{f}' and get a new estimate $\hat{f}^{(s)}$ to the true function f ;

9. calculate $MSE(s)$ as in (10):

9a. if $MSE(s-1) - MSE(s)$, for $s > 0$, is greater than the convergence criterion, go back to step 4;

9b. if $MSE(s-1) - MSE(s)$, for $s > 0$, is less than or equal to the convergence criterion, stop.

An important point is that the order of the $AR(p)$ process must be chosen only once at the beginning of the iterative process, in order to achieve convergence of the algorithm.

4. Simulations

In order to evaluate the performance of the proposed iterative method, some simulations were done along other two procedures, using universal thresholding [3, 4] and *SURE* thresholding by level [8]. *SURE* is a label for “Stein Unbiased Risk Estimate”. The conservative properties of the universal threshold (6) come at the price of high threshold levels: in terms of L^2 -loss, better performance should be obtained with smaller thresholds that can be achieved by *SURE* thresholds by level [8], in the case of correlated errors. A data-based *SURE* threshold choice can then be obtained simply by minimizing an unbiased estimate for the mean squared error with respect to a possible threshold over the range $[0, \sigma\sqrt{2 \log n}]$.

The simulation was done with 3 different functions. For each function $f(t/n)$ it was added a noise that could be a Gaussian white noise, an autoregressive of order 1 with coefficient $\phi_1 = 0.8$ or $\phi_1 = -0.8$. Each of these nine simulations was done with 3 sample sizes ($n = 64, 128$ and 256 points) and replicated 1000 times. The errors were scaled so that the simulations were done with a signal-to-noise ratio $SNR = 1$, where

$$SNR = \frac{(n-1)^{-1} \sum_t (f(t/n) - \bar{f})^2}{Var(\text{noise})}$$

and

$$\bar{f} = n^{-1} \sum_t f(t/n).$$

The functions used in the simulations were, for $t = 1, \dots, n$:

- Gaussian density: $f(t/n) = 100 \times (n/8 (\phi(t) + n/2))$, where $\phi(\cdot)$ denotes the standard Normal density;

- Heavisine function:

$$f(t/n) = 4 \sin(4\pi t/n) - \text{sign}(t/n - 0.3) - \text{sign}(0.72 - t/n).$$

- Doppler function:

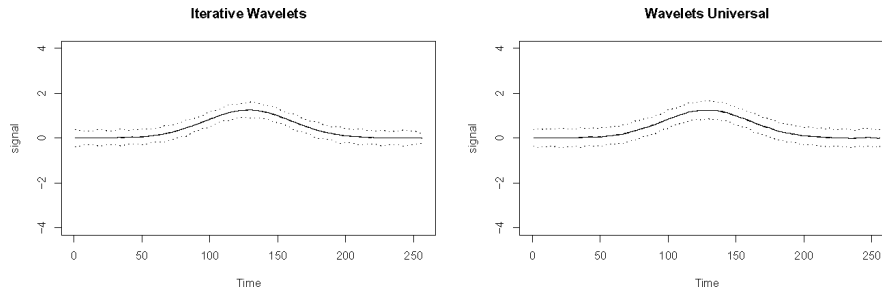
$$f(t/n) = \sqrt{(1 - t/n)t/n} * \sin(2\pi(1 + 0.05)/(t/n + 0.05)).$$

These functions are shown with dashed lines in Figures 1, 2 and 3, respectively.

At the end, some statistics were computed that allow useful comparisons and analysis. The statistics computed are sample cases of squared bias, average of the *MSE* (10) and standard error of *MSE*. Comments on these statistics are given in next subsection. All the simulations were done using the *R* language [15]. The wavelet calculations were done in *R* with the *Waveslim* package [18]. The routines are available upon request to the authors.

4.1. Results

In this section the results of the simulations are discussed. For the sake of simplicity, we refer the estimation methods simply as “universal”, “*SURE*” and “iterative”, instead of “wavelet estimator using the universal threshold”, for example.



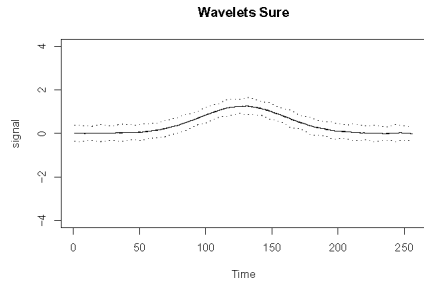


Figure 1. Simulation results for the Gaussian density function, autoregressive coefficient = 0.8, sample size = 256; estimator (continuous line), true function (dashed line) and estimator ± 1 standard error (dotted lines).

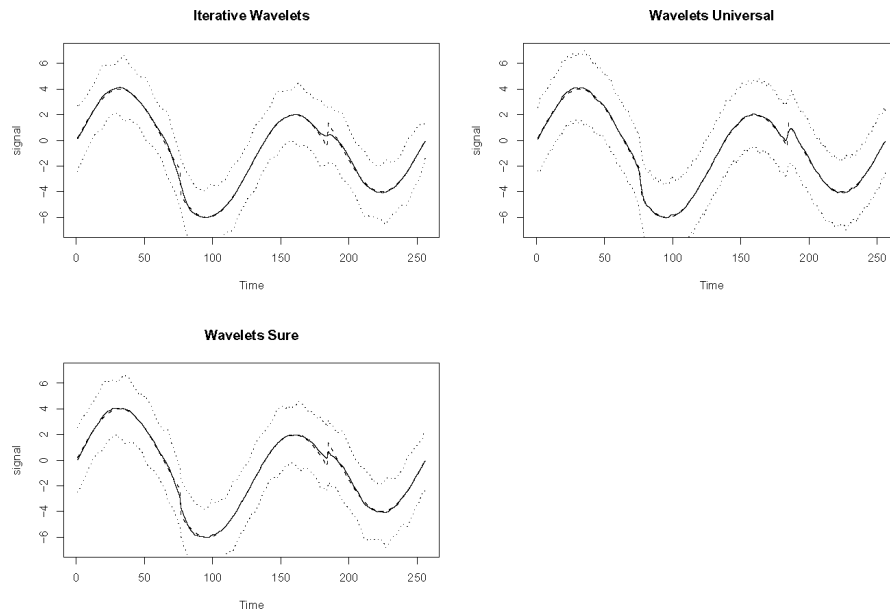


Figure 2. Simulation results for the Heavisine function, autoregressive coefficient = 0.8, sample sizes = 256; estimator (continuous line), true function (dashed line) and estimator ± 1 standard error (dotted lines).

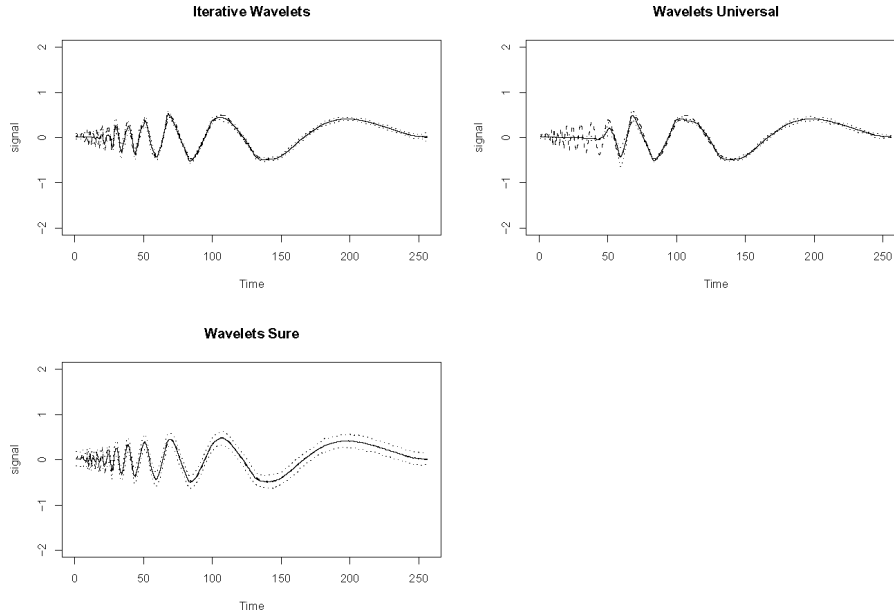


Figure 3. Simulation results for the Doppler function, autoregressive coefficient = -0.8 , sample sizes = 256; estimator (continuous line), true function (dashed line) and estimator ± 1 standard error (dotted lines).

Figure 1 and Table 1 show results for the Gaussian case. Visually, all the three methods give equivalent results in terms of mean and variability. For smaller sample sizes, the iterative method seems to have less variability and less bias when the errors are correlated. For uncorrelated errors, the *SURE* estimators seem to have larger average *MSE*.

For the Heavisine function, the results can be seen in Figure 2 and Table 2. The same general pattern observed for the Gaussian case was observed here. It is important to note the higher variability supposedly due to the discontinuities in the function. The superiority of the iterative method can be observed better for autoregressive errors with negative coefficients and small sample size. For large sample sizes, both iterative and universal have similar performance, while *SURE* has larger average *MSE* for all sample sizes.

Table 1. Simulations: Gaussian function

		Iterative			Universal			Sure by level		
<i>N</i>	<i>AR</i>	<i>Bias</i> ²	<i>MSE</i>	s.e.(<i>MSE</i>)	<i>Bias</i> ²	<i>MSE</i>	s.e.(<i>MSE</i>)	<i>Bias</i> ²	<i>MSE</i>	s.e.(<i>MSE</i>)
64	0	0.218	0.728	0.011	0.218	0.739	0.011	0.249	1.434	0.021
	0.8	0.018	2.544	0.030	0.006	2.898	0.025	0.251	2.376	0.030
	-0.8	0.032	0.323	0.014	0.352	0.503	0.010	0.324	1.253	0.031
128	0	0.002	0.082	0.002	0.002	0.078	0.002	0.001	0.292	0.005
	0.8	0.001	0.495	0.005	<0.001	0.650	0.003	0.001	0.503	0.005
	-0.8	0.002	0.031	0.002	0.002	0.026	0.002	0.001	0.217	0.007
256	0	<0.001	0.016	<0.001	<0.001	0.015	<0.001	<0.001	0.064	0.001
	0.8	<0.001	0.104	0.001	<0.001	0.155	<0.001	<0.001	0.123	0.001
	-0.8	<0.001	0.005	<0.001	<0.001	0.004	<0.001	<0.001	0.048	0.002

Table 2. Simulations: Heavisine function

		Iterative			Universal			Sure by level		
<i>N</i>	<i>AR</i>	<i>Bias</i> ²	<i>MSE</i>	s.e.(<i>MSE</i>)	<i>Bias</i> ²	<i>MSE</i>	s.e.(<i>MSE</i>)	<i>Bias</i> ²	<i>MSE</i>	s.e.(<i>MSE</i>)
64	0	1.556	3.824	0.033	1.138	3.548	0.030	2.779	6.125	0.060
	0.8	0.220	8.833	0.083	0.045	8.640	0.077	2.281	8.565	0.088
	-0.8	0.271	1.486	0.049	2.958	3.414	0.027	2.924	5.433	0.088
128	0	0.072	5.742	0.057	0.025	7.534	0.037	0.126	5.931	0.052
	0.8	0.129	0.574	0.019	0.266	0.557	0.020	0.089	2.657	0.084
	-0.8	0.040	0.263	0.010	0.048	0.227	0.010	0.016	2.241	0.075
256	0	0.045	0.780	0.011	0.045	0.736	0.010	0.021	2.937	0.047
	0.8	0.034	4.915	0.037	0.015	7.178	0.021	0.025	5.775	0.038
	-0.8	0.040	0.263	0.010	0.048	0.227	0.010	0.016	2.241	0.075

Table 3. Simulations: Doppler function

		Iterative			Universal			Sure by level		
N	AR	$Bias^2$	MSE	$s.e.(MSE)$	$Bias^2$	MSE	$s.e.(MSE)$	$Bias^2$	MSE	$s.e.(MSE)$
64	0	0.027	0.054	<0.001	0.023	0.051	<0.001	0.026	0.068	0.001
	0.8	0.005	0.088	0.001	0.004	0.089	0.001	0.020	0.090	0.001
	-0.8	0.013	0.036	0.001	0.047	0.055	<0.001	0.027	0.056	0.001
128	0	0.013	0.027	<0.001	0.013	0.026	<0.001	0.005	0.043	<0.001
	0.8	0.005	0.070	<0.001	0.002	0.078	<0.001	0.004	0.070	<0.001
	-0.8	0.007	0.014	<0.001	0.017	0.022	<0.001	0.003	0.030	0.001
256	0	0.007	0.016	<0.001	0.007	0.016	<0.001	0.003	0.035	<0.001
	0.8	0.004	0.057	<0.001	0.002	0.071	<0.001	0.002	0.063	<0.001
	-0.8	0.003	0.008	<0.001	0.009	0.012	<0.001	0.001	0.023	0.001

The last function used in the simulations was the Doppler function and the results are shown in Figure 3 and Table 3. Still the proposed iterative method has the same general behavior as before. For small sample sizes the iterative and universal performs about in the same way concerning biases and average MSE , while for larger sample sizes $SURE$ has smaller biases but larger average MSE . Overall $SURE$ presents larger average MSE .

The above results can be better analyzed through Figures 4 to 6. Figure 4 shows simulation results for the Doppler function with autoregressive errors (coefficient $\phi_1 = 0.8$). The average of the mean square error - MSE (10) for each of three sample sizes are displayed along with line segments representing its sample standard deviation. As expected, the average MSE gets smaller with increasing sample size, leading to better fits. Since there is positive autocorrelation, $SURE$ generally gives better results than universal. This also occurs with other values for the autoregressive coefficient and to other functions (see tables).

Fixing the autoregressive coefficient at $\phi_1 = 0.8$ and the sample size at $n = 256$, then Figure 5 shows that the iterative method has smaller average MSE than universal and than $SURE$, independently of the curve

homogeneity. It should be noted the large variability of the MSE for the Heavisine function, again supposedly due to discontinuities.

Finally, taking the Doppler function and fixing the sample size at 256 observations, Figure 6 makes clear a fact superficially touched when the graphics of Figures 1, 2 and 3 were analyzed. When the errors are correlated, $SURE$ would be expected to give smaller average MSE than universal, since the variance of the wavelets coefficients is different at each level. Nevertheless the simulations showed that this is true for autoregressive coefficients greater than zero; for autoregressive coefficients less than zero, universal had a better performance in

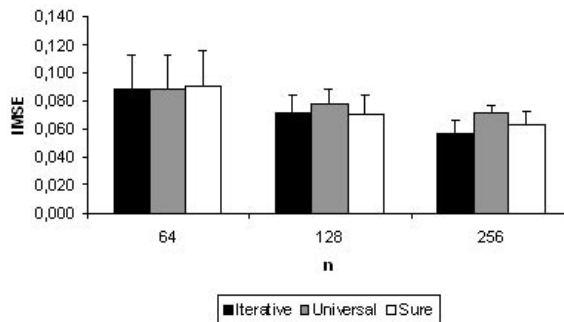


Figure 4. Simulation results for the Doppler function, autoregressive coefficient = 0.8, three sample sizes: average of the mean square errors - MSE (10), with line segments representing its sample standard deviation.

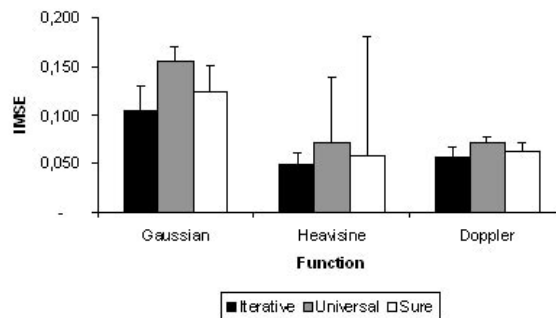


Figure 5. Simulation results for three functions, autoregressive coefficient = 0.8, sample size = 256: average of the mean square errors - MSE (10) with line segments representing its sample standard deviation.

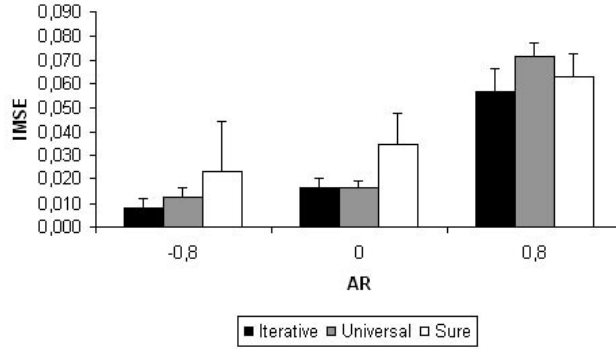


Figure 6. Simulation results for the Doppler function, sample size = 256, three values for autoregressive coefficient: average of mean square errors - MSE (10) with line segments representing its sample standard deviation.

terms of average MSE . Besides that, the proposed iterative method gives better results in any case. This adaptability of the iterative method is appreciated since in practice we may find cases where there is no information about the nature of the autocorrelation.

In the next subsection some brief discussion will be carried out on the results described above.

4.2. Comments

The last subsection discussed a number of results from the simulation study that was carried out. In this subsection we make some further comments. These will be given in an intuitive way and no formal proofs will be given. Some reasons for the validity of the main results will be only conjectured.

Universal is consistent as sample size goes to infinity. It has also almost optimal rates of convergence for a wide range of function classes. The same occurs with $SURE$. Since the iterative method uses universal thresholding it is reasonable that all the three methods are equivalent for large sample sizes. For small sample sizes, the iterative method has better performance apparently because it uses the information better than $SURE$. Maybe a hybrid method [4] could give yet better results.

For negative autocorrelations, the variance of the wavelet coefficients still vary among levels. However, they vary less than for high positive autocorrelations. Besides that, negative autocorrelations do not generate local trends apparently causing *SURE* to take the errors as i.i.d. For these cases, universal had better performance. By the way, the iterative method estimates the covariance structure with moment estimators independently if the autocorrelation is negative or positive. Maybe this is the reason for the iterative method to have better fit even for small samples.

Since the error covariance term is incorporated in the model through a moment estimator, and the signal part is estimated by wavelet thresholding, the pros and cons of each seems to be balanced giving rise to a better method by the use of iterations in order to achieve minimum *MSE* for finite samples, independently of the signal homogeneity or the autocorrelation signal.

Finally, although the presented results were done only for $SNR = 1$, the iterative method has a little better relative performance when $SNR < 1$. This advantage decreases for $SNR > 1$ but it still continues to be close to the best method.

In summary, when there is positive autocorrelation, *SURE* has better performance but the iterative method has similar performance. When there is negative autocorrelation, the iterative method has performance near to universal which is the best method. When there is no autocorrelations, it seems that universal and the iterative methods are barely better than *SURE*, maybe because of signal sparsity in some levels. Thus, in any case, the proposed iterative method behaves at least as well as the benchmark.

A lot of thresholding methods has been developed beyond Universal and *SURE*. Recently, some attention has been paid to a method named EbayesThresh [9]. It uses empirical Bayes methods to take advantage from the sparsity of the wavelet coefficients and to achieve better thresholds. Simulation results (not shown) with this method did not change the general conclusions above, since EbayesThresh had performances a little bit better than Universal (which is worse than

SURE in some cases) but still worse than the proposed iterative method.

In fact, the proposed iterative method presented here uses Universal thresholding just for simplicity and popularity. Any other method could easily substitute Universal. Actually, the proposed method is an alternative to leveldependent thresholding; not a new thresholding method.

5. An Application

In this section the proposed iterative method is applied to financial data to illustrate its usefulness. A sample of size 2048 from Nasdaq Composite daily index from 5-May-1998 to 23-June-2006 was used to motivate the analysis. The respective log-returns have zero mean and an apparent conditional heteroskedasticity. The series and the returns are presented in Figures 7(a) and 7(b), respectively.

An autoregressive model of order 13 was fitted to these log-returns. The squared residuals from this fit are shown in Figure 7(c). A *GARCH*(2,1) model, with two *ARCH* terms and 1 *GARCH* term, was fitted to the residuals from the *AR*(13) fit in order to model the conditional volatility as common practice prescribes. Residual analysis of this *GARCH* fit indicates that the final *AR-GARCH* model is adequate (Figure 7(d)). A *GARCH* model implies a deterministic model for volatility which cannot always be the case in practice. This means that the volatility should be modelled stochastically. Such modelling is often carried in practice by a stochastic volatility model, but this supposes independent errors in the volatility equation, and independent of the errors in the mean equation. The *GARCH* fit to the data of Figure 7(c) is shown in Figure 8(a).

Thus, the proposed iterative method seems to be appropriate in this situation. The iterative method was applied to the squared residuals from the *AR*(13) model fitted to the Nasdaq log-returns. In this case, model (1) was applied to the volatility, represented by the squared residuals. Convergence occurred in only two iterations. The *GARCH* fit of Figure 8(a) is smoother, showing local trends and smaller values when compared to the fit obtained from the iterative method (Figure 8(b)).

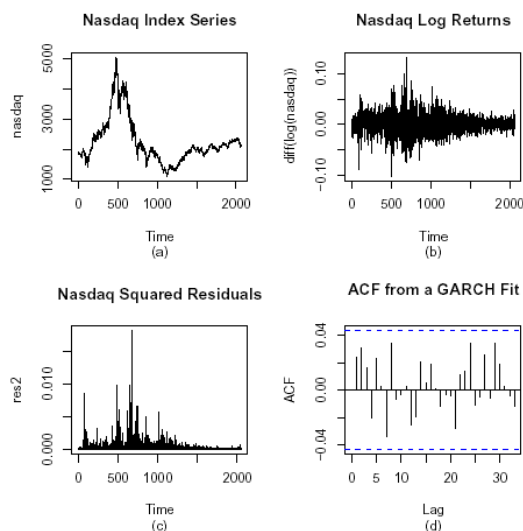


Figure 7. (a) 2048 Nasdaq daily index series. (b) Log-returns from the series in (a). (c) Squared residuals from an $AR(2)$ fit to the series in (b). (d) Sample autocorrelation function of the residuals from a $GARCH(2,1)$ fit to the series in (c).

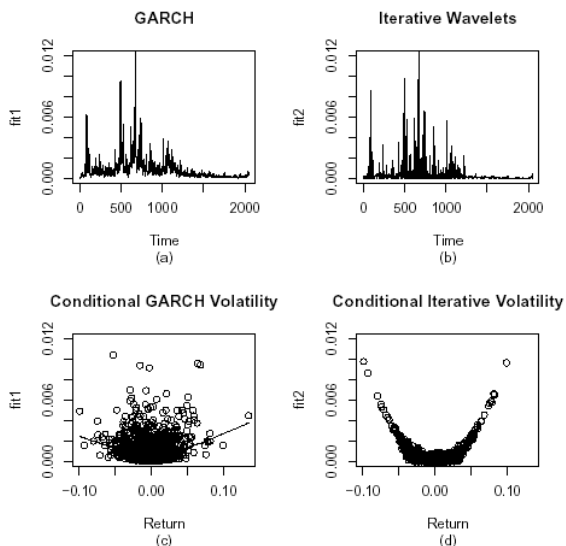


Figure 8. (a) $GARCH$ fit to the squared log-returns of Figure 7(c). (b) Fit to the data of Figure 7(c) obtained with the proposed iterative method. (c) Conditional volatility from the fitted $GARCH(2,1)$ model and smoothed lowess curve. (d) Conditional volatility from the fitted model using the proposed iterative method.

Conditional heteroskedasticity or volatility is important for macroeconomic [12] and financial modeling [16], mainly in the areas of derivative pricing, risk management and asset allocation [17]. In Finance, asset returns are usually assumed to be normally distributed but extensive research shows that these are not normal. Stylized facts about return distributions include negative skewness and excess kurtosis [7, 13, 14]. Specifically, excess kurtosis (roughly the fourth moment of the distribution) makes extreme observations more likely than in the normal case.

Plots of the estimated volatility against returns commonly show some curvature (“smile” in Finance jargon) [19] indicating excess kurtosis and making market asymmetries clearer. Volatility smile and smirk effects are closely related to the presence of excess kurtosis and negative skewness in the underlying asset returns distribution [2]. Also, skewness affects the persistence in variance and have different impacts on the volatility because it increases uncertainty and risk.

Analysis of this conditional volatility again highlights the smaller values obtained from the *GARCH* model. In Figure 8(c), a smoothed curve was obtained by the “lowess” method. It shows little curvature and no skewness for the conditional volatility values from the *GARCH* model. Meanwhile, the proposed iterative method shows that a polynomial of higher order could be useful since Figure 8(d) shows much more curvature for the conditional volatility obtained from the iterative method. Also, the values from the iterative method are less variable for returns near zero than its correspondents from the *GARCH* model. Little or no skewness was observed in this plot, indicating that probably asymmetric models are not needed.

Thus, the proposed iterative method shows some important facts about the conditional volatility that were hidden by the *GARCH* model. This probably occurs because of the deterministic nature of the variance in the *GARCH* model (given all the past information), and because of the separated treatment given to the covariance structure of the volatility in the iterative method.

6. Conclusion

Correlation is a common factor in practical applications. This correlation can have important consequences on the statistical properties of the estimator.

In this article a method for function estimation using wavelets in the presence of correlated errors was proposed. Similar ideas were briefly touched by some authors [8] for the case where the correlation structure was known and comparisons were advocated as an interesting topic.

A simulation study was carried out in order to evaluate risk and bias of the proposed iterative procedure. The results were reported and compared with two other wavelet methods. To illustrate the usefulness of the method, it was applied to a financial situation where the signal-to-noise is usually rather low. It was tentatively shown that the proposed method is at least as good as other benchmark wavelet methods independently of the type of autocorrelation of the error term.

The proposed iterative method uses Universal but any other threshold could easily substitute it. Actually, the proposed method is an alternative to leveldependent thresholding; not a new thresholding method.

Research to develop theoretical properties for the proposed iterative method is in progress. It seems that order of convergence and asymptotic distributional properties are not hard to obtain under mild conditions. For results regarding kernel - and splines - based methods under correlation see [11] for a review.

References

- [1] D. Cochrane and G. H. Orcutt, Applications of least squares regression to relationships containing autocorrelated errors, *J. Amer. Statist. Assoc.* 34 (1949), 32–61.
- [2] S. Das and R. Sundaram, Of smiles and smirks: a term structure perspective, *Journal of Financial and Quantitative Analysis*, 34 (1999) 211–239.
- [3] D. L. Donoho and I. M. Johnstone, Ideal spatial adaptation via wavelet shrinkage, *Biometrika* 81 (1994), 425–455.

- [4] D. L. Donoho and I. M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, *J. Amer. Statist. Assoc.* 90 (1995), 1200–1224.
- [5] P. Flandrin, Time-scale analyses and self-similar stochastic processes, J. S. Byrnes, ed., *Wavelets and Their Applications*, Kluwer Academic, Dordrecht, 1994, pp. 121–142.
- [6] W. A. Fuller, *Introduction to Statistical Time Series*, 2nd ed., Wiley, New York, 1996.
- [7] C. R. Harvey and A. Siddique, Autoregressive conditional skewness, *Journal of Financial and Quantitative Analysis* 34 (1999), 465–487.
- [8] I. M. Johnstone and B. W. Silverman, Wavelet threshold estimators for data with correlated noise, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 59 2 (1997), 319–351.
- [9] I. M. Johnstone and B. W. Silverman, Empirical Bayes selection of wavelet thresholds, *The Annals of Statistics* 33 4 (2005), 1700–1752.
- [10] P. A. Morettin, *Waves and Wavelets*, Univ. of São Paulo Press, São Paulo, 1999.
- [11] J. Opsomer, Y. Wang and Y. Yang, Nonparametric regression with correlated errors, *Statistical Science* 16 2 (2001), 134–153.
- [12] A. Pagan and A. Ullah, *Nonparametric Econometrics*, Cambridge University Press, Cambridge, 1999.
- [13] A. Peiró, Skewness in financial returns, *Journal of Banking and Finance* 23 (1999), 847–862.
- [14] G. Premaratne and A. K. Bera, Modeling Asymmetry and Excess Kurtosis in Stock Return Data, Working Paper, Department of Economics, University of Illinois, 2001.
- [15] R Development Core Team, 2005, *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available online at: <http://www.R-project.org> (accessed 3 July 2006).
- [16] C. W. Smithson, C. W. Smith Jr. and D. S. Wilford, *Managing Financial Risk*, Irwin, New York, 1995.
- [17] R. S. Tsay, *Analysis Of Financial Time Series*, Wiley, New York, 2005.
- [18] B. Whitcher, 2005, *Waveslim: basic wavelet routines for one-,two- and three-dimensional signal processing*. R package version 1.5. Available online at: <http://www.image.ucar.edu/staff/whitcher/> and <http://www.image.ucar.edu/staff/whitcher/book/> (accessed 3 July 2006).
- [19] F. A. Ziegelmann, Nonparametric and semi-parametric methods in time series: the kernel smoothing approach. Mini-course presented at the 10th ESTE - Escola de Séries Temporais e Econometria, São Pedro, SP, Brazil, 11-13 August 2003.



Kindly return the proof after correction to:

*The Publication Manager
Pushpa Publishing House
Vijaya Niwas
198, Mumfordganj
Allahabad-211002 (India)*

along with the print charges*
by the fastest mail

***Invoice enclosed**

Proof read by:

Signature:

Date:

Tel:

Fax:

E-mail:

Number of additional reprints required

.....

Cost of a set of 25 copies of additional reprints @ U. S. Dollars 15.00 per page.
(25 copies of reprints are provided to the corresponding author ex-gratis)