Research Article

Robust Wavelet Estimation to Eliminate Simultaneously the Effects of Boundary Problems, Outliers, and Correlated Noise

Alsaidi M. Altaher and Mohd Tahir Ismail

School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia

Correspondence should be addressed to Alsaidi M. Altaher, assaedi76@yahoo.com

Classical wavelet thresholding methods suffer from boundary problems caused by the application of the wavelet transformations to a finite signal. As a result, large bias at the edges and artificial wiggles occur when the classical boundary assumptions are not satisfied. Although polynomial wavelet regression and local polynomial wavelet regression effectively reduce the risk of this problem, the estimates from these two methods can be easily affected by the presence of correlated noise and outliers, giving inaccurate estimates. This paper introduces two robust methods in which the effects of boundary problems, outliers, and correlated noise are simultaneously taken into account. The proposed methods combine thresholding estimator with either a local polynomial model or a polynomial model using the generalized least squares method instead of the ordinary one. A primary step that involves removing the outlying observations through a statistical function is considered as well. The practical performance of the proposed methods has been evaluated through simulation experiments and real data examples. The results are strong evidence that the proposed method is extremely effective in terms of correcting the boundary bias and eliminating the effects of outliers and correlated noise.

1. Introduction

Suppose a noisy data set $y_1, y_2, \ldots, y_n$ lives on the fixed design model

$$y_i = f\left(\frac{i}{n}\right) + \varepsilon_i; \quad i = 1, 2, \ldots; \ n = 2^j; \ j = 1, 2, \ldots. \tag{1.1}$$

The classical model (1.1) assumes that the unknown function, f, is square integrable on the interval $[0, 1]$. The sequence $\{\varepsilon_i\}_{i=1}^{n}$ is independent and distributed normally with identical

means zero and identical variances σ^2. Wavelet methods have been intensively used over the last two decades for estimating an unknown function observed in the presence of noise, following the pioneering work presented in the seminal papers of Donoho and Johnstone [1, 2], where the concept of wavelet thresholding was introduced to the statistical literature. Researchers have begun looking at the situation where the usual assumptions are no longer satisfied. When the noise contains a certain amount of structure in the form of a correlation, the variances of the wavelet coefficients will depend on the resolution level of the wavelet decomposition but will be constant at each level. As a result, the global thresholding breaks down because of its great difficulty in providing a global threshold value to threshold wavelet coefficients at all the desired levels. In order to overcome this limitation, Johnstone and Silverman [3] recommended using level by level thresholding, that is, not to use only one thresholding value for all coefficients at all resolution levels, but instead, one depending on the levels. Different thresholding methods have been considered, such as Universal of Donoho and Johnstone [1], SURE of Donoho and Johnstone [2], and the translation-invariant denoising algorithm of Coifman and Donoho [4]. Details can be found in Johnstone and Silverman [3]. The level-dependent thresholding method proposed by [3] considered the case where the noise is correlated but stationary. If the noise is correlated but not stationary, the DWT coefficients will be heteroscedastic and correlated. Kovac and Silverman [5] have taken these circumstances into consideration and proposed more general thresholding method, which can be used for both stationary and nonstationary noise. Another method was proposed by Wang and Wood [6], in which the covariance structure of the data has been taken into account. This method is based on estimating the correlation structure of the noise in two different domains, namely, the time domain and the wavelet domain. As to the presence of outliers, a lot of research has been carried out to eliminate the effects of outliers. Outliers in the data typically lead to large coefficients in the wavelet decomposition. These large coefficients will not be removed by the usual thresholding approaches. One of the early robust wavelet thresholding estimators dates back to Kovac [7], where thresholding was applied to the output of a median filter. Additionally, Kovac and Sliverman [5] proposed a robust procedure involving the following steps. First, the problematic outliers are identified using a statistical test. Then, these outliers are removed from the data. The final step is to apply wavelet thresholding to the remaining irregularly spaced data to estimate the underlying function. Sardy et al. [8] proposed a robust M-type wavelet denoising method. Averkamp and Houdré [9] extended the minimax theory for wavelet thresholding to some known symmetric heavy tail noises. In a similar way to Kovac [7], Wei et al. [10] used a two-stage robust denoising approach based on a median filter and wavelet thresholding. At the first stage, the corrupt data is passed through the median filter, in which the outliers are suppressed. At the second stage, the data is shrunk with a soft wavelet threshold, and the ultimate reconstructed signal is attained. More recently, new methods have been proposed by Oh et al. [11], based on the concept of pseudo data, and by Oh et al. [12], based on the concept of robust thresholding within robust estimation. Altaher and Ismail [13] showed that the practical performance of these methods would be improved by taking into consideration the automatic boundary treatment proposed by Oh et al. [14] and Oh and Lee [15].

In all the literature discussed above, Kovac and Silverman [5] were the only ones to take into account the effect of correlated noise and outliers at the same time, but the limitation of this method is the use of classical boundary assumptions. This method performs well when the boundary assumptions are satisfied, but its practical achievement can be greatly improved by using the idea of polynomial wavelet regression (PWR) or local polynomial wavelet regression (LPWR) for automatic boundary corrections. We conclude that there is

a real need to find a proper way to take into simultaneous consideration the effect of the boundary problem, correlated noise, and outliers by one estimator; this is main contribution of the present paper.

The rest of this paper is organized as follows. In Section 2, we give a brief background for automatic boundary treatments in wavelet regression. Section 3 reviews the wavelet estimation of Kovac and Silverman [5] and generalized least squares, followed by our proposed methods. A simulation experiment and real data examples are presented in Section 4. Conclusions are drawn in Section 5.

2. Background

2.1. Boundary Treatment in Wavelet Thresholding

Boundary effects have been taken into account when using classical wavelet thresholding by imposing on f some boundary assumptions such as symmetry or periodicity. Unfortunately, such assumptions are not always valid, and the problem still exists for some cases. The boundary problem can be treated automatically using the principles of either polynomial wavelet regression (PWR) or local polynomial wavelet regression (LPWR).

2.1.1. Polynomial Wavelet Regression (PWR)

The polynomial wavelet regression method (PWR) considered by Oh et al. [14] is based on a combination of wavelet functions $\hat{f}_W(x)$ and low order polynomials $\hat{f}_P(x)$. Therefore, the estimator of f, $\hat{f}_{PW}$, is

$$\hat{f}_{PW}(x) = \hat{f}_P(x) + \hat{f}_W(x). \tag{2.1}$$

The optimal order for the polynomial model will enable the polynomial estimator to remove the "nonperiodicity" in the data. Then, the remaining signal can be well estimated using wavelet thresholding with, say, a periodic boundary assumption. Different approaches have been proposed in order to find such choices of the proper polynomial order and threshold value. For the threshold value, the EBayesThresh procedure of [16] is the most popular for its good theoretical properties and good performance in simulations and in practice. To find $\hat{f}_{PW}(x)$, it is proposed to regress the observed data $\{y_i\}_{i=1}^n$ on $\{x, \ldots, x^d\}$ for a fixed order value, d. Once $\hat{f}_P(x)$ has been estimated, the remaining signal is expected to be hidden in the residuals, $\{e_i = y_i - \hat{f}_P(x)\}_{i=1}^n$. For this, the second step is to apply wavelet regression to $\{e_i\}_{i=1}^n$. The final estimate of f will be the summation of $\hat{f}_P(x)$ and $\hat{f}_W(x)$ as in (2.1). For more details, see, for example, Oh et al. [14] or Lee and Oh [17].

The use of PWR for resolving boundary problems works efficiently if the polynomial estimator $\hat{f}_P$ is able to remove the "nonperiodicity" in the data, and this certainly requires using a proper order with an appropriate threshold. Different criteria have been proposed to aid in choosing the order of the polynomial model that should be incorporated into the wavelet functions in order to get successful corrections to the bias at the boundary region. Here we give a brief introduction to the common ones.

(1) Lee and Oh [17] proposed two criteria: the first criterion uses the value of d that maximizes

$$r(d) = \sum_{i=1}^{d} a_i^2 - \frac{2\sigma^2 d}{n}, \quad d = 0, 1, \ldots. \tag{2.2}$$

Here, $\{a_i\}_{i=0}^{d}$ denote the parameter coefficients of the polynomial model. In this method, a model of this order d has been used with the SURE threshold.

(2) The second criterion is based on the Bayesian information criterion (BIC), which is employed to select d by minimizing

$$\text{BIC} = n \log \left[\frac{1}{n} \left\{ \sum_{i=1}^{n} y_i - \hat{f}_P \left(\frac{i}{n} \right) \right\}^2 \right] + d \log(n). \tag{2.3}$$

In this method, a model of order d should be used with EbayeThresh of Johnstone and Silverman [16].

(3) Oh and Kim [18] proposed three different Bayesian methods based on integrated likelihood, conditional empirical Bayes, and reversible jump Markov chain Monte Carlo (MCMC). For the mathematical details regarding these methods, see Oh and Kim [18].

2.1.2. Local Polynomial Wavelet Regression (LPWR)

The local polynomial wavelet regression method was introduced by Oh and Lee [15] as an improved boundary adjustment in wavelet regression. Instead of using the global polynomial fit as in Oh et al. [14], it was proposed to use a local polynomial fit, $\hat{f}_{LP}$. Therefore, the local polynomial wavelet regression estimator, $\hat{f}_{LPWR}(x)$, can be written as follows:

$$\hat{f}_{LPWR}(x) = \hat{f}_{LP}(x) + \hat{f}_W(x). \tag{2.4}$$

As shown in Oh and Lee [15], $\hat{f}_{LPWR}(x)$ can be computed through an iterative algorithm inspired by the back-fitting algorithm, see Johnstone and Silverman [16]. The following steps summarize the key points for finding the final local polynomial wavelet regression estimate, $\hat{f}_{LPWR}(x)$.

(1) Select an initial estimate $\hat{f}_0$ for f and let $\hat{f}_{LPWR} = \hat{f}_0$.

(2) For $j = 1, 2, \ldots$ iterate the following steps.

   (a) Apply wavelet thresholding to the residuals $y_i - \hat{f}_{LPWR}$ and obtain $\hat{f}_W^j$.

   (b) Estimate $\hat{f}_{LP}^j$ by fitting a local polynomial regression to $y_i - \hat{f}_W^j$.

(3) Stop if $\hat{f}_{LPWR} = \hat{f}_{LP}^j + \hat{f}_W^j$ converges.

3. Method

3.1. Wavelet Estimation of Kovac and Silverman [5]

The level-dependent thresholding method proposed by Johnstone and Silverman [3] considers the case where the noise is correlated but stationary. In this case, the variances of the wavelet coefficients differ between levels but are identical, so the coefficients can be thresholded level by level, see Altaher and Ismail [19] for numerical simulation results. On the other hand, if the noise is correlated but not stationary, the discrete wavelet transform (DWT) coefficients will be heteroscedastic and correlated. Kovac and Silverman [5] have taken these circumstances into consideration and proposed more general thresholding method which can be used for both stationary and nonstationary noise.

As shown in Kovac and Silverman [5], an algorithm can be used to find all the variances and within-level covariances in the wavelet table of a sequence with a given covariance structure. If the covariance matrix of the original data is band limited, then the algorithm is linear in the length of the sequence. It also has been shown that the variance-calculation algorithm allows data on any set of independent variable values to be treated as follows: first, the noisy data are interpolated to a new equally spaced grid of a suitable length (the function *makegrid* available in R can be used). Then, wavelet thresholding is applied to the gridded data.

Based on the information given in Kovac and Silverman [5], the key points of this method are as follows.

Given a discrete sequence $y = (y_1, \ldots, y_n)$ observed on $(x_1, \ldots, x_n)$, a new equally spaced grid $t_1, \ldots, t_N$ on $(0, 1)$ is generated, where $t_k = (k+0.5)/N$, $k = 0, 1, \ldots, N-1$, $N = 2^j$, $j = \min\{J \in \mathbb{Z} : 2^J > n\}$.

Based on this grid, the original data values y_i on x_i are linearly interpolated to new data values y^* on the new grid:

$$y_i^* = \begin{cases} y_1; & t_k \leq x_1 \\ y_i + (t_k - x_i)\dfrac{y_{i+1} - y_i}{x_{i+1} - x_i}; & x_i \leq t_k \leq x_{i+1} \\ y_n; & t_k > x_n. \end{cases} \tag{3.1}$$

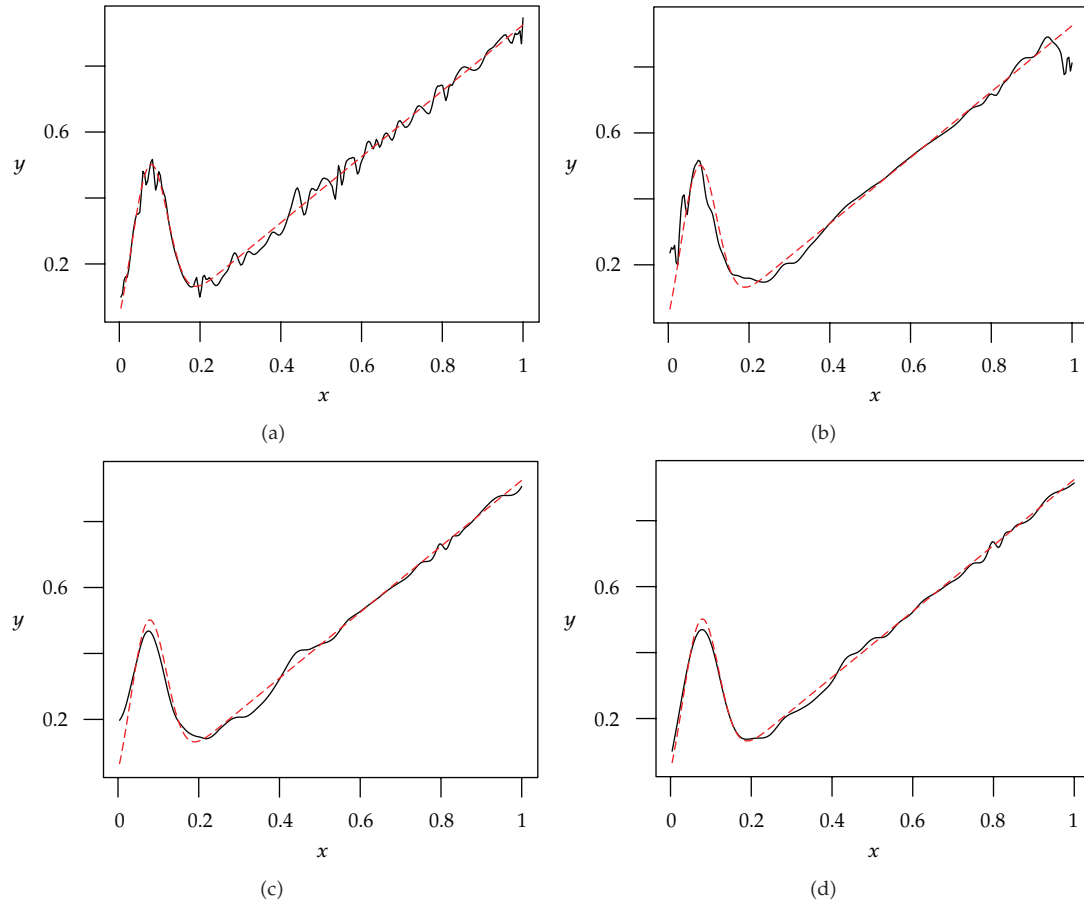This interpolation can be written in the matrix form as follows:

$$y^* = Ry, \tag{3.2}$$

where the matrix R is the linear transformation matrix. The first step of wavelet regression is to apply the discrete wavelet transform, which can be written in the matrix form as follows:

$$d = Wy^*, \tag{3.3}$$

where W is the $N \times N$ wavelet matrix. Since the original observations $\{y_i\}$ are correlated, the covariance matrix of y^* is given by

$$\Sigma = R\Sigma_y R' = \{\sigma_{jk}\}. \tag{3.4}$$

Figure 1: Curve estimation of data with correlated noise: (a) fitted by level-dependent thresholding, (b) fitted by the method of Kovac and Silverman [5], (c) fitted by the first proposed method, and (d) fitted by the second proposed method. The black line is the fitted curve, while the red is the true function.

For all j and k, consider thresholds of the form

$$\lambda_{j,k} = \lambda \hat{\sigma}_{jk}, \tag{3.5}$$

where $\lambda = \sqrt{2\log(n)}$, the universal threshold.

Although this method works fine for many different kinds of functions, only the classical treatment for the boundary problem is considered (such as assumptions of periodicity or symmetry). Now, as a motivation for our proposed methods which will be described later, we present here a very popular function, called fg1, of Donoho and Johnstone [1] to show how badly this traditional method behaves when the regression function is neither periodic nor symmetric. Then, we will show how we could improve the estimation quality if we combine this method with a polynomial model or a local polynomial model. Figure 1 depicts the weakness and the improvement in the estimation method of [5]. In this figure, it can be seen that the best fit is given by our proposed methods, that is, when

we combine the thresholding method of Kovac and Silverman [5] with polynomial or local polynomial models.

Details for these methods will be given next, but first we will introduce the method of generalized least squares which will be used in the first stage of our proposed method.

3.2. Ordinary Least Squares versus Generalized Least Squares

Consider the standard linear model:

$$Y = X\beta + \epsilon, \tag{3.6}$$

where Y is the $n \times 1$ response vector; X is an $n \times p$ model matrix; β is a $p \times 1$ vector of parameters to be estimated; ϵ is an $n \times 1$ vector of errors. The error term ϵ could be constant and uncorrelated such $\epsilon \sim N(0, \sigma^2 I_n)$ or correlated, such as $\epsilon \sim N(0, \Omega)$, where Ω is the variance-covariance matrix of the errors. Based on these two assumptions, we can distinguish two formulas for the parameter estimates.

If $\epsilon \sim N(0, \sigma^2 I_n)$, then

$$\hat{\beta}_{\text{OLS}} = (X'X)^{-1}X'Y, \qquad \text{var}\left(\hat{\beta}_{\text{OLS}}\right) = \sigma^2(X'X)^{-1}. \tag{3.7}$$

The presence of correlated noise has several effects on OLS. Although the regression coefficient estimates are still unbiased, they are no longer minimum-variance estimates. In the case of positively correlated noise, the residual mean squared error may seriously underestimate the error variance, σ^2. Consequently the standard errors of the regression coefficients may be too small. As a result, the confidence intervals are shorter than they should be, see Montgomery et al. [20].

An alternative solution to estimate the regression coefficients in the presence of correlated noise is to use the following generalized least square estimates.

If $\epsilon \sim N(0, \Omega)$, then

$$\hat{\beta}_{\text{GLS}} = \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}Y, \qquad \text{var}\left(\hat{\beta}_{\text{GLS}}\right) = \left(X'\Omega^{-1}X\right)^{-1}. \tag{3.8}$$

Using the Choleski decomposition, the matrix Ω can be written in terms of a triangular matrix, S, such that $\Omega = S'S$. The model in (3.6) can be written as follows:

$$S^{-1}Y = S^{-1}X\beta + S^{-1}\epsilon, \tag{3.9}$$

$$Y^* = X^*\beta + \epsilon^*, \tag{3.10}$$

where $Y^* = S^{-1}Y$ and $X^* = S^{-1}X$. In the new model (3.10) one can easily show that $\text{var}(\epsilon^*) = \text{var}(S^{-1}\epsilon) = I_n$.

3.3. Proposed Robust Estimation Method for Data with Outliers and Correlated Noise

In this section, we will propose some two-stage combined methods which can be used in the presence of outliers and correlated noise. The classical way in PWR and LPWR is to use OLS to get the residuals. Due to the serious effects of the correlated noise on OLS, we suggest using the method of generalized least squares (GLSs) as an effective alternative solution to get rid of (or at least reduce) these effects. A primary step for removing outlying observation is employed before applying GLS.

At the second stage, we employ the wavelet thresholding estimator illustrated in Section 3.1 for removing the remaining noise. Therefore, the final estimate will be the summation of the function estimate from the polynomial model (or local polynomial) at the first stage and the estimate of the Kovac and Silverman [5] estimator at the second stage.

The following two algorithms describe our two proposed methods. The first algorithm is for PWR and the second is for LPWR.

Algorithm 3.1. Let $\{x_i, y_i\}$ be the original noisy data sample. The following few steps describe the first proposed method.

(1) Remove any outliers in $\{y_i\}$. The function *get.outliers* available in R can be used to determine probable problematic outliers.

(2) Interpolate the remaining data to a new equally spaced design and use them in the next steps. The function *makegrid* (available in R) can be used.

(3) Apply the generalized least squares method and select the optimal polynomial order to be used for getting $\hat{f}_P$ and then the residuals $\hat{e}_i$. The function *gls* is used to apply GLS; it is available in R in the *nlme* package.

(4) Apply the wavelet thresholding procedure of [5] to the residuals and find $\hat{f}_W$.

(5) The final estimate will be $\hat{f}_P + \hat{f}_W$.

Algorithm 3.2. In a similar way to Algorithm 3.1, the second proposed method involves the following steps.

(1) Remove any outliers in $\{y_i\}$. The function *get.outliers* is used.

(2) The function *makegrid* is used to interpolate the remaining data to a new equally spaced design and use them in the next steps.

(3) Apply a local polynomial estimator using the interpolated data to get $\hat{f}_{LP}$ and then the residuals.

(4) Apply the wavelet thresholding procedure of Kovac and Silverman [5] to the residuals and find $\hat{f}_W$.

(5) The final estimate will be $\hat{f}_{LP} + \hat{f}_W$.

4. Evaluating the Practical Performance

4.1. Simulation Experiment

The R statistical package was used to carry out a simulation study in order to compare the numerical performance of the following three methods:

Table 1: Autoregressive models used in the simulations.

AR(1)	AR(2)	MA(1)
	Parameter	
0.3	(0.7, 0.2)	3.8
0.5	(0.5, 0.2)	2.5
0.7	(0.2, 0.7)	1.8
0.9	(0.1, 0.5)	0.8

(1) The wavelet regression method of Kovac and Silverman [5], labelled as (I).

(2) The proposed method based on a polynomial model, labelled as (II).

(3) The proposed method based on a local polynomial model, labelled as (III).

We simulated 100-time series data of length 128 using three common correlated process models with different parameters. These included first- and second-order autoregressive models and a first-order moving average model, see Table 1. For each model, we added four different kinds of noise, similar to Gelper et al. [21]. Then the three wavelet regression methods (I, II, and III) were applied, and, finally, the global and local mean squared errors were computed for comparison. The four kinds of noise were as follows.

(1) Pure normal noise with 0 mean and variance equal to 1 (clean data, CD)

(2) Mixture normal noise: $0.95\, N(n, 0, 1) + 0.05\, N(n, 0, 10)$ (symmetric outliers, SQ).

(3) Mixture normal noise: $0.95\, N(n, 0, 1) + 0.05\, N(n, 10, 1)$ (asymmetric outliers, AQ).

(4) $t(3)$ noise (fat tail noise).

Generating time series data will ensure that we have data with correlated noise. The noises mentioned above will play the role of having outliers. For example, in the second noise setting, a small fraction of the clean observations is replaced by symmetric outliers (SOs), and, in the third noise setting, it is replaced by asymmetric outliers (AOs). The last noise is a heavy tail noise, a common example of non-Gaussian noise with outliers.

The numerical results for the global mean squared error and local mean squared error are given in Tables 3, 4, and 5. Having examined these simulation results, the following empirical remarks can be made.

(i) In terms of the global mean squared error:

(1) In the case of clean data, the global mean squared error shows that the proposed method (III) outperforms the other two methods for all three autoregressive models with their four different parameters. Among the other two methods, the proposed method (I) outperforms the traditional method of [5]. The only exception was for AR(2) with parameter (0.2, 0.7).

(2) In the case of symmetric outliers, the global mean squared error shows that the proposed method (II) outperforms the other two methods for all three autoregressive models and their four different parameters. The only exception was for MA(1) with parameter (3.8). In this case, the proposed method (III) provided the best results.

(3) In the case of asymmetric outliers, the global mean squared error shows that the proposed method (II) outperforms the other two methods for all three autoregressive models and their four different parameters.

(4) In the case of fat tail data, the global mean squared error shows that the proposed method (III) performs much better than the other two methods for all three autoregressive models with their four different parameters.

(ii) In terms of the local mean squared error:

(1) In the case of clean data, the local mean squared error shows that the proposed method (III) outperforms the remaining two methods for all three autoregressive models and their four different parameters. The only exception was for AR(1) with parameter (0.3) and AR(2) with parameter $(0.1, 0.5)$.

(2) In the case of symmetric outliers, the local mean squared error shows that the two proposed methods II and III outperform the traditional method of Kovac and Silverman [5]. For AR(1) and AR(2), the proposed method (II) achieved better results except for the parameters (0.3) and $(0.1, 0.5)$.

(3) In the case of asymmetric outliers, the local mean squared error shows that the proposed method (II) outperforms the other two methods except for AR(1) with parameter (0.9) and MA(1) with parameter (3.8).

(4) In the case of long tail data, the local mean squared error shows that the proposed method (III) performs better than the other two methods.

(5) Generally, in terms of the global mean squared error, out of the 48 different models with different parameters used in this simulation, it has been found that the first proposed method won 25 times, the second proposed method won 23 times, while the traditional method of Kovac and Silverman [5] did not win even once. In terms of the local mean squared error, out of the 48 different models with different parameters used in the simulation, it has been found that the first proposed method won 19 times, the second proposed method won 22 times, while the traditional method of Kovac and Silverman [5] won only 7 times.

Overall, this simulation experiment revealed that the two proposed methods are preferable to that of Kovac and Silverman [5] in fitting data with correlated noise and outliers.

4.2. Evaluation through Real Data Examples

In this section, we will apply the traditional method of Kovac and Silverman [5] and the two new proposed methods to two data sets.

The first data are the US monthly influenza mortality data taken from Shumway and Stoffer [22]. The response variable denotes the mortality caused by pneumonia and influenza at month t. Figure 2 is a plot of the monthly pneumonia and influenza deaths per 10,000 in the US 11 years from 1968 to 1978. Many interesting features can be seen in this data. The number of deaths typically tends to increase more slowly than it decreases. There is no reason to think that this time series follows a linear Gaussian process. If that was the case, we would not have seen such large bursts of positive and negative changes occurring periodically.

The second data are also from Shumway and Stoffer [22], the average global temperature deviations for $n = 64$ years in degrees centigrade (1880–1943); see Figure 3.

We first study the correlation structure to make sure that we have data with correlated noise. Figure 4 shows the autocorrelation functions and their corresponding partial correlation functions. Then we follow up by computing the Durbin-Watson statistics for
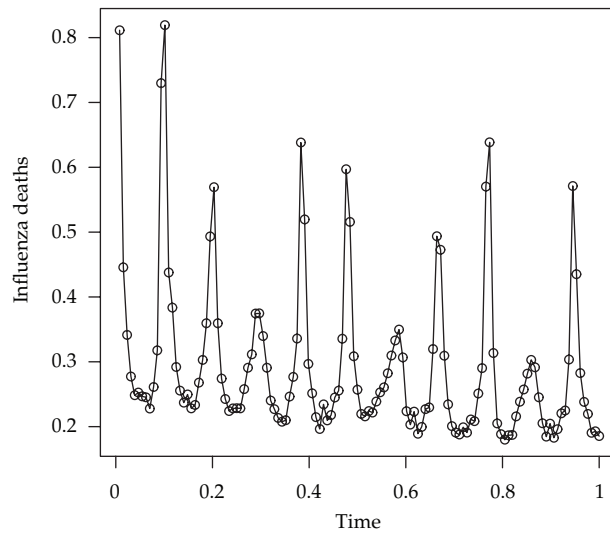
Figure 2: U.S. monthly pneumonia and influenza deaths per 10,000 over 11 years from 1968 to 1978.
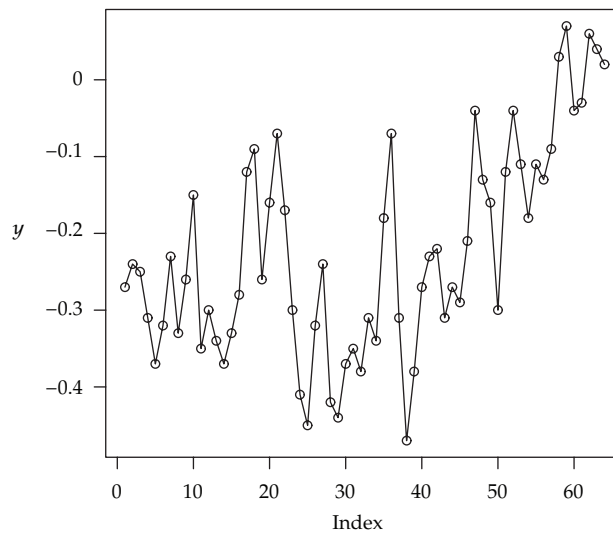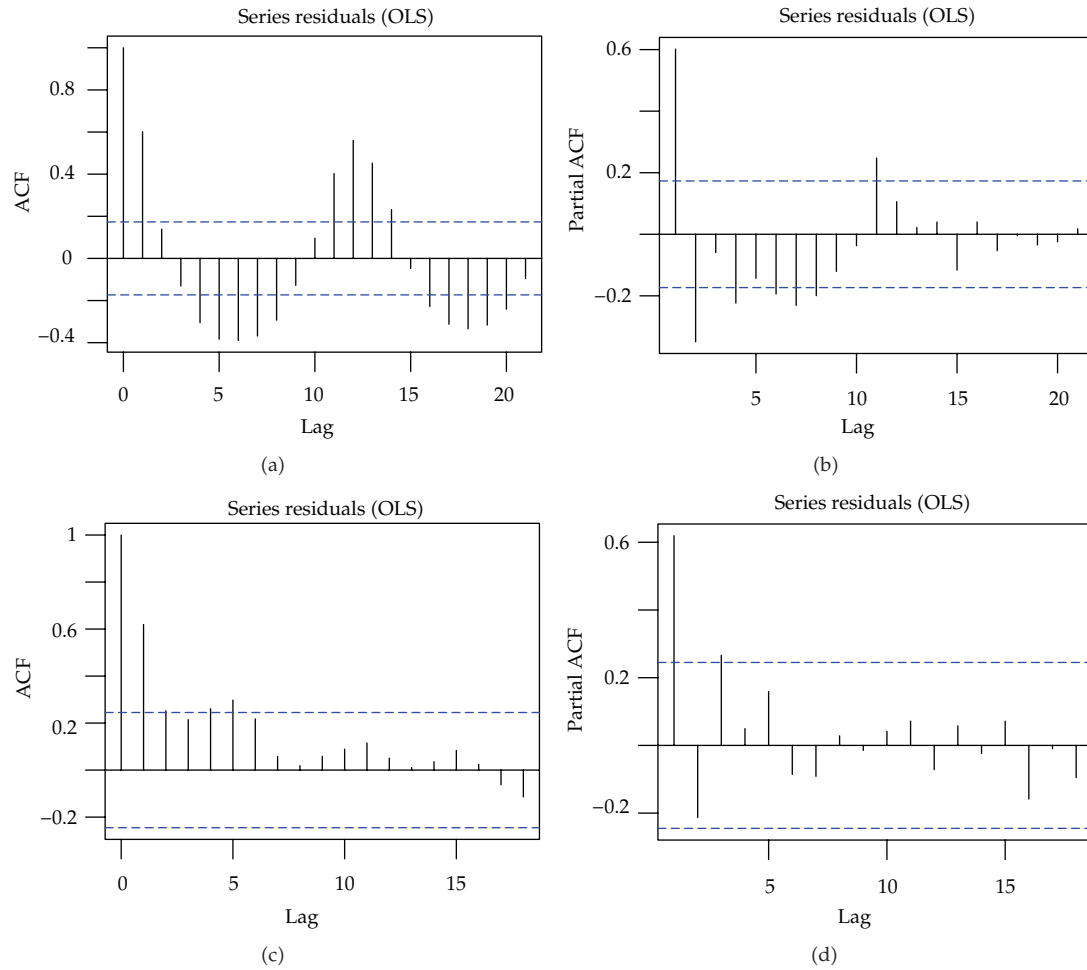

Figure 3: Average global temperature deviations for $n = 64$ years in degrees centigrade (1880–1943).

the OLS regression estimates, see Table 2. The results are clear evidence for significant correlations at lags 1, 4, and 5 for both data sets.

Figure 5 shows the estimated curve using the two proposed methods (panels (c) and (d)) and the traditional method (panel (b)) of Kovac and Silverman [5]. Visual evaluation is not sufficient for determining which method is more accurate than the others, so we used the mean squared error for judgement. The mean squared error values are (0.01127487, 0.01046806, and 0.004596079) for the traditional method, the first proposed method, and the second proposed method, respectively. This means that the best fit is given by the second proposed method, then by the first proposed method, and finally by the traditional method.

(a)

(b)

(c)

(d)

Figure 4: Autocorrelation and partial-autocorrelation functions for the residuals from the OLS method. The first row is for first data, while the second row is for the second data.

Table 2: Summary for Durbin-Watson test.

lag	Autocorrelation	D-W Statistic	P value
		First data	
1	0.6883450	0.6168636	0.000
2	0.1922665	1.5993563	0.140
3	−0.1685699	2.3187448	0.254
4	−0.3652775	2.6990538	0.010
5	−0.3673240	2.6521103	0.008
		Second data	
1	0.6883450	0.6168636	0.000
2	0.1922665	1.5993563	0.162
3	−0.1685699	2.3187448	0.274
4	−0.3652775	2.6990538	0.022
5	−0.3673240	2.6521103	0.010

Table 3: Simulation results of global and local mean squared using AR(1) with different parameters. The first half of the table displays the global mean squared error.

ρ	I	II	III
		CD	
0.3	1.032684	1.031848	0.9912508
0.5	1.190477	1.142319	0.8841418
0.7	1.454564	1.417338	0.8683088
0.9	1.186370	1.117227	0.7043864
		SQ	
0.3	1.603859	1.070997	1.672791
0.5	1.646231	1.084239	1.700433
0.7	1.829821	1.050726	1.625967
0.9	1.981135	1.121463	1.589858
		AQ	
0.3	1.430155	0.8245817	1.563744
0.5	1.921071	1.3050117	1.888181
0.7	1.774816	1.0581883	1.713439
0.9	1.917869	1.5325202	1.702284
		Fat tail data	
0.3	1.0792571	0.9587506	0.9144366
0.5	0.9066489	0.8696987	0.7690691
0.7	1.3491463	1.2460089	0.8146345
0.9	1.5744066	1.4142226	0.7023962
		mse at boundary	
		CD	
0.3	1.959192	1.993277	1.9739248
0.5	1.698549	1.601896	1.2203974
0.7	1.161763	1.426407	0.9157507
0.9	1.647253	1.808840	1.1332881
		SQ	
0.3	2.045994	1.6759544	2.374387
0.5	1.891612	1.2883680	2.514418
0.7	1.714374	0.8980506	2.353474
0.9	2.196825	0.9884269	1.968370
		AQ	
0.3	1.796641	1.135821	2.304636
0.5	2.336372	1.676010	2.936859
0.7	1.752143	1.256421	1.807269
0.9	2.542747	2.410233	2.061829
		Fat tail data	
0.3	1.3175606	1.3487222	1.0755325
0.5	0.6829416	0.6869467	0.8313126
0.7	1.0077950	1.0629237	0.8260527
0.9	1.0001727	1.3372696	0.6346054

Table 4: Simulation results of global and local mean squared using AR(2) with different parameters. The first half of the table displays the global mean squared error.
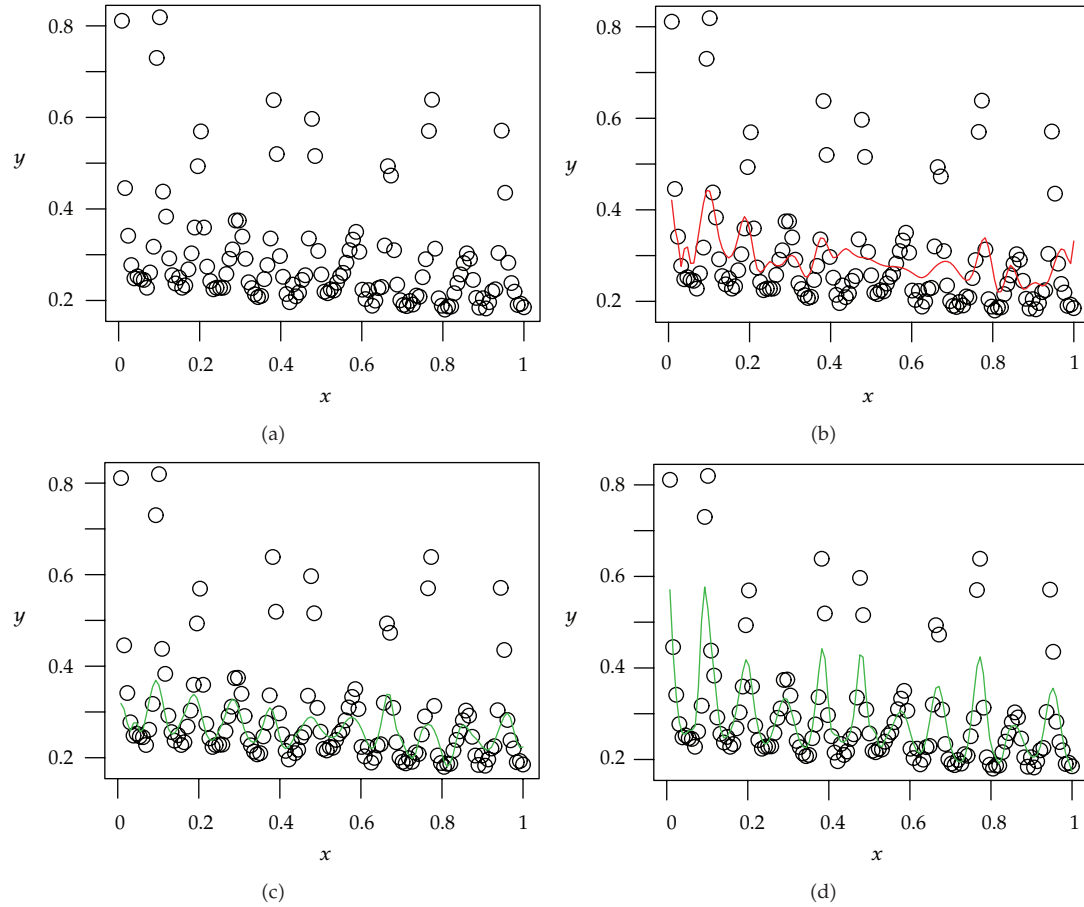
Parameter	I	II	III
		CD	
(0.7, 0.2)	1.1674119	1.1430405	0.6588707
(0.5, 0.1)	1.1753915	1.1456239	0.8297044
(0.2, 0.7)	1.4271601	1.5023487	1.0724085
(0.1, 0.5)	0.9575155	0.9564091	0.8454300
		SQ	
(0.7, 0.2)	1.778795	1.1063826	1.570100
(0.5, 0.1)	1.693634	1.0449833	1.736956
(0.2, 0.7)	1.882562	1.2680705	1.862326
(0.1, 0.5)	1.596086	0.9444905	1.669918
		AQ	
(0.7, 0.2)	1.898400	1.235252	1.742275
(0.5, 0.1)	1.771281	1.129160	1.694660
(0.2, 0.7)	1.904999	1.289010	1.979608
(0.1, 0.5)	1.860164	1.297206	1.929902
		Fat tail data	
(0.7, 0.2)	1.1154769	0.9981609	0.4987108
(0.5, 0.1)	0.9402973	0.8866934	0.6799741
(0.2, 0.7)	1.1724818	1.2601606	1.1130583
(0.1, 0.5)	0.9557715	0.9238237	0.8206673
		mse at boundary	
		CD	
(0.7, 0.2)	3.3196129	3.3166944	1.1779039
(0.7, 0.2)	1.0757030	1.1300048	0.6423341
(0.7, 0.2)	1.0365045	2.0357543	0.9276551
(0.7, 0.2)	0.6378157	0.7400542	0.6481636
		SQ	
(0.7, 0.2)	3.1439208	2.5169418	2.102952
(0.7, 0.2)	0.9689782	0.3868747	1.548121
(0.7, 0.2)	1.8243869	1.0661310	1.839487
(0.7, 0.2)	1.8154754	1.2451924	2.386717
		AQ	
(0.7, 0.2)	1.576587	0.9909815	1.885401
(0.7, 0.2)	2.258808	1.1307186	2.460695
(0.7, 0.2)	1.872568	1.0907302	2.251098
(0.7, 0.2)	1.532332	1.2593857	2.012984
		Fat tail data	
(0.7, 0.2)	3.1254564	2.5821331	0.5579562
(0.7, 0.2)	0.7235822	0.7457132	0.7487358
(0.7, 0.2)	0.6287745	0.7475777	0.7258418
(0.7, 0.2)	1.2695745	1.2063235	0.9366066

Table 5: Simulation results of global and mean squared using MA(1) with different parameters. The first half of the table displays the global mean squared error.

Parameter	I	II	III
CD			
3.8	13.536209	13.381343	12.409326
2.5	6.493029	6.250129	3.902021
1.8	3.632426	3.476316	2.471782
0.8	1.209791	1.207804	1.093832
SQ			
3.8	11.937604	11.322564	10.695495
2.5	6.600532	6.249817	6.266773
1.8	3.320055	2.938204	3.127264
0.8	1.977752	1.474812	2.017470
AQ			
3.8	10.178566	9.860996	10.013281
2.5	6.419729	5.897459	5.992016
1.8	4.255425	3.715337	4.192729
0.8	2.090731	1.512278	2.047738
Fat data			
3.8	13.016700	12.559023	9.863463
2.5	5.898913	5.943974	5.090725
1.8	3.065080	3.040102	2.481520
0.8	1.436546	1.366507	1.207752
mse at boundary			
CD			
3.8	5.519441	5.366619	3.964631
2.5	5.020789	4.724176	3.088903
1.8	3.580416	3.968404	1.913348
0.8	1.723382	1.708755	1.546141
SQ			
3.8	10.195461	10.449241	9.949788
2.5	8.098829	7.855652	7.235322
1.8	5.466819	5.554564	4.674587
0.8	2.176779	1.743749	2.731832
AQ			
3.8	15.197856	15.222380	15.307931
2.5	4.591728	3.899142	4.743319
1.8	2.667287	2.348781	3.388881
0.8	2.589583	2.201528	2.374601
Fat tail data			
3.8	14.576058	14.671735	13.338548
2.5	3.395124	3.348310	3.594896
1.8	5.401277	5.372652	4.416008
0.8	1.209738	1.214468	1.478539

(a)

(b)

(c)

(d)

Figure 5: Curve estimation for the first data: (a) noisy data, (b) the fitting by [5], (c) the fitting by first proposed method, and (d) the fitting by the second proposed method.
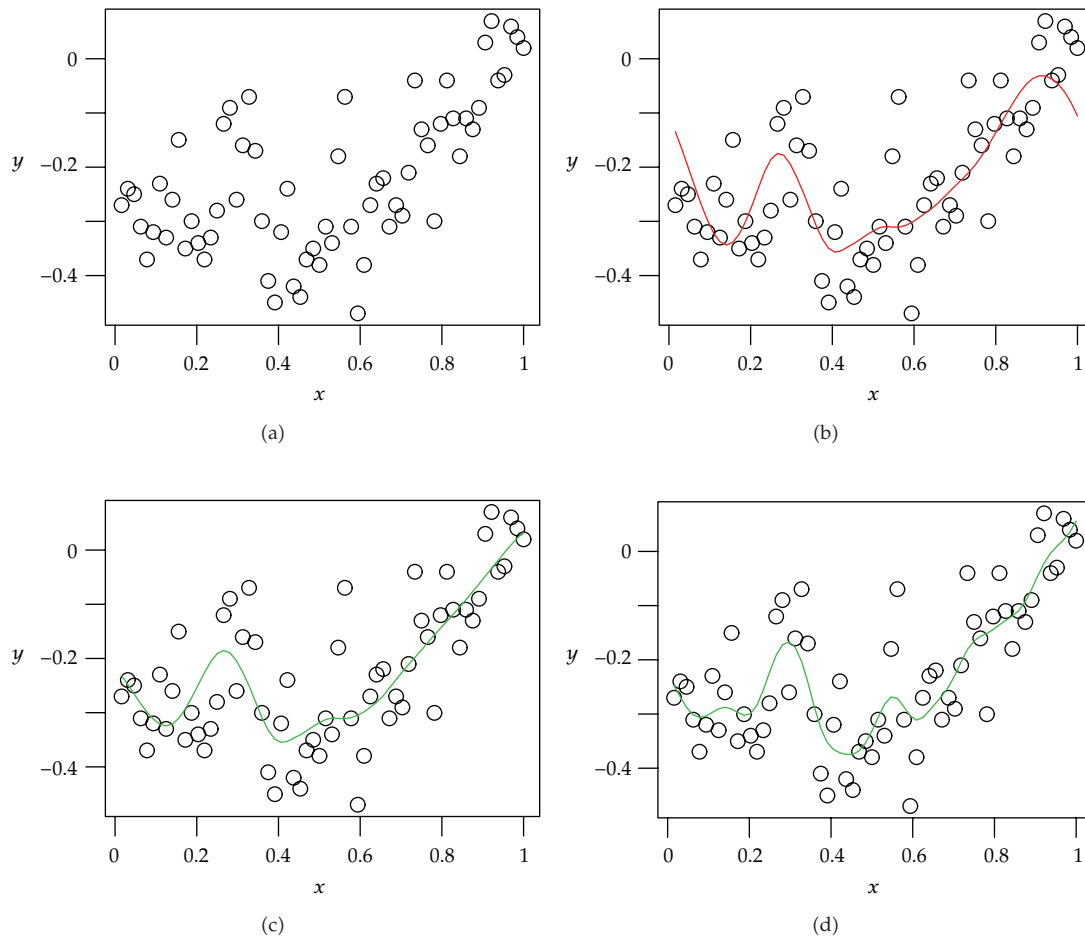
The fitting of the second data set is depicted in **Figure 6**. The mean squared error values are (0.008085288, 0.006734333, and 0.005019866). This means that the best fitting is given by the second proposed method, then by the first proposed method, and finally by the traditional method.

A similar conclusion is reached when we examine the local mean squared error. The local mean squared error values using the first data set are 0.01767953, 0.01677273, and 0.00516326 and, for the second data set, 0.007946142, 0.002644399, and 0.002241301.

5. Conclusion

In this paper, the boundary problem and the effects of correlated noise and outliers have been considered for wavelet thresholding estimation. Two combined methods have been proposed to eliminate simultaneously the effect of the boundary problem, the outliers, and the correlated noise. The proposed methods have combined the thresholding estimator of [5] with either a local polynomial model or a polynomial model, through generalized

Figure 6: Curve estimation for the second data: (a) noisy data, (b) the fitting by [5], (c) the fitting by first proposed method, and (d) the fitting by the second proposed method. The green line is the fitting.

least squares estimation. Simulation experiment and real data examples have been used to evaluate the practical performance of the proposed methods. Overall, the results revealed that the two proposed methods are preferable to that of Kovac and Silverman [5] in fitting data with correlated noise and outliers.

Acknowledgment

References

[1] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[2] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.

[3] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *Journal of the Royal Statistical Society Series B*, vol. 59, no. 2, pp. 319–351, 1997.

[4] R. Coifman and D. Donoho, *Translation-Invariant de-Noising*, vol. 103 of *Lecture Notes in Statistics*, Springer, 1995.

[5] A. Kovac and B. Silverman, "Extending the scope of wavelet regression methods by coefficient-dependent thresholding," *Journal of the American Statistical Association*, vol. 95, pp. 172–183, 2000.

[6] X. Wang and A. T. A. Wood, "Wavelet estimation of an unknown function observed with correlated noise," *Communications in Statistics*, vol. 39, no. 2, pp. 287–304, 2010.

[7] A. Kovac, *Wavelet Thresholding for Unequally Spaced Data*, University of Bristol, 1998.

[8] S. Sardy, P. Tseng, and A. Bruce, "Robust wavelet denoising," *IEEE Transactions on Signal Processing*, vol. 49, pp. 1146–1152, 2001.

[9] R. Averkamp and C. Houdré, "Wavelet thresholding for non-necessarily Gaussian noise: idealism," *The Annals of Statistics*, vol. 31, no. 1, pp. 110–151, 2003.

[10] G. Wei, T. Jian, and R. Qiu, "Signal denoising via median filter & wavelet shrinkage based on heavy-tailed modeling," in *Proceedings of the 8th International Conference on Signal Processing*, vol. 1, 2006.

[11] H.-S. Oh, D. W. Nychka, and T. C. M. Lee, "The role of pseudo data for robust smoothing with application to wavelet regression," *Biometrika*, vol. 94, no. 4, pp. 893–904, 2007.

[12] H.-S. Oh, D. Kim, and Y. Kim, "Robust wavelet shrinkage using robust selection of thresholds," *Statistics and Computing*, vol. 19, no. 1, pp. 27–34, 2009.

[13] A. M. Altaher and M. T. Ismail, "Robust estimation for boundary correction in wavelet regression," *Journal of Statistical Computation and Simulation*, vol. 82, pp. 1531–1544, 2012.

[14] H.-S. Oh, P. Naveau, and G. Lee, "Polynomial boundary treatment for wavelet regression," *Biometrika*, vol. 88, no. 1, pp. 291–298, 2001.

[15] H.-S. Oh and T. C. M. Lee, "Hybrid local polynomial wavelet shrinkage: wavelet regression with automatic boundary adjustment," *Computational Statistics & Data Analysis*, vol. 48, no. 4, pp. 809–819, 2005.

[16] I. M. Johnstone and B. W. Silverman, "Empirical Bayes selection of wavelet thresholds," *The Annals of Statistics*, vol. 33, no. 4, pp. 1700–1752, 2005.

[17] T. C. M. Lee and H.-S. Oh, "Automatic polynomial wavelet regression," *Statistics and Computing*, vol. 14, no. 4, pp. 337–341, 2004.

[18] H.-S. Oh and H.-M. Kim, "Bayesian automatic polynomial wavelet regression," *Journal of Statistical Planning and Inference*, vol. 138, no. 8, pp. 2303–2312, 2008.

[19] A. M. Altaher and M. Ismail, "Hybrid local polynomial wavelet shrinkage for stationary correlated data," *Informatics Engineering and Information Science*, vol. 253, pp. 262–273, 2011.

[20] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*, Wiley-Interscience, Hoboken, NJ, USA, 2008.

[21] S. Gelper, R. Fried, and C. Croux, "Robust forecasting with exponential and Holt-Winters smoothing," *Journal of Forecasting*, vol. 29, no. 3, pp. 285–300, 2010.

[22] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*, Springer, New York, NY, USA, 2nd edition, 2006.