Int. J. Nonlinear Anal. Appl. 13 (2022) No. 1, 1855-1862
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.5816

Using the wavelet analysis to estimate the nonparametric regression model in the presence of associated errors

Mohmmed Salh AbduAlkareem Mahdi[a,*], Saad Kadem Hamza[b]

[a]*Ministry of Interior, Directorate of Human Recourses, Iraq*
[b]*Department of Statistics, College of Administration and Economics University of Baghdad, Baghdad, Iraq*

(Communicated by Madjid Eshaghi Gordji)

Abstract

The wavelet shrink estimator is an attractive technique when estimating the nonparametric regression functions, but it is very sensitive in the case of a correlation in errors. In this research, a polynomial model of low degree was used for the purpose of addressing the boundary problem in the wavelet reduction in addition to using flexible threshold values in the case of Correlation in errors as it deals with those transactions at each level separately, unlike the comprehensive threshold values that deal with all levels simultaneously, as (Visushrink) methods, (False Discovery Rate) method, (Improvement Thresholding) and (Sureshrink method), as the study was conducted on real monthly data represented in the rates of theft crimes for juveniles in Iraq, specifically the Baghdad governorate, and the risk ratios about those crimes for the years 2008-2018, with a sample size of (128) (Sureshrink) The study also showed an increase in the rate of theft crimes for juveniles in recent years.

Keywords: Discrete Wavelet Transformation, Threshold Value, Wavelet Shrinkage, Correlated Error

1. Introduction

Regression is perhaps the field of statistics that has received the most attention by researchers in wavelet methods, where wavelet methods are usually used as a form of nonparametric regression, and the techniques take many names such as wavelet reduction, estimation of non-parametric curve

*Corresponding author
Email addresses: mohmmed.salh1201@coadec.uobaghdad.edu.iq (Mohmmed Salh AbduAlkareem Mahdi), saad.hamza@coadec.uobaghdad.edu.iq (Saad Kadem Hamza)

Received: October 2021 *Accepted:* October 2021

or wavelet regression, and generally In general, the nonparametric regression itself constitutes an important and vital field of modern statistics. The general idea of the wavelet regression work is explained according to the following.
Let our observations ($y_i = (y_1, \ldots\ldots\ldots, y_n)$) be given in the following form:

$$y_i = g\left(t_i\right) + \varepsilon_i \tag{1.1}$$

where ($t_i = i/n$), and the objective is to estimate ($g(t_i)$) the unknown function ($t_i \in \{0, 1\}$) using scrambled observations of y_i.
The concept of wavelet shrinkage or wavelet regression was introduced to the statistical literature by researcher[2], and the general idea of applying the discontinuous wavelet transformation to the above model is summarized (1), whereby Mallat algorithm is used.
Let (y_i) represent the observations, (g) represent the unrated function and (e_i) represent the error or noise, and through the discontinuous wavelet transform, the transformation model can be written as follows:

$$d^* = d + \varepsilon \tag{1.2}$$

whereas $d^* = w_y$, $d = w_g$, $\varepsilon = w_e$, w is the wavelet transformation matrix.
Three essential characteristics of successful wavelet reduction

a) The wavelet transforms are adapted to many functions (discrete and heterogeneous smoothing functions).
b) Furthermore, due to the Parsvaal relation, the energy is in the domain of the function $\sum g(t)^2$ It is equal to the sum of the squares of the wavelet coefficients $\sum_{j,k} d_{j,k}^2$ However, if the contrast is taken into account, it means that the energy of the original signal is concentrated in less coefficients and nothing is lost, and therefore for the contrast of noise the vector (d) will not only be scattered but the values themselves are often larger.
c) By (w) resulting from the discrete transformation is an orthogonal matrix. This means that the noise transformation wavelet, which is white noise, remains white noise after the transformation and is spread evenly over all the wavelet coefficients.

Based on the above properties, [2] suggested scaling the following wavelet which we need in estimating the function g(t_i), The basic idea was that the large values of the wavelet coefficients (d^*), were more likely cases consisting of real signal and noise while the small coefficients were due to noise only, and then to estimate (d) successfully I found the threshold idea of estimating (d^*) by removing the coefficients in (d^*) that They are smaller than some threshold and essentially preserve larger coefficients.

2. Thresholding Rules

The stage of removing noise in the signal is one of the most important steps for estimating the regression function using wavelet reduction, as the selection of the threshold contributes to removing noise and in turn preserving the coefficients of the original signal because the coefficients of noise are of lower frequency than the frequency of the coefficients of the original signal. Thresholding process can be carried out in several ways, the most important and most common are the soft Thresholding method and Hard Thresholding method.

3. Hard Thresholding

It is a simplified method by zeroing the elements whose absolute value is less than the threshold and is expressed mathematically:

$$\mathrm{Thr}_{\lambda}^{H}(y,\lambda) = \begin{bmatrix} 0 & \text{if} & |y| \leq \lambda \\ & & \\ y & \text{if} & |y| > \lambda \end{bmatrix} \tag{3.1}$$

where λ is the Thresholding value.

4. Soft Thresholding

It is an extension of the previous method and differs from it that after the elements whose absolute value is less than the threshold are zeroed, the non-zero elements are shifted towards zero, and it is expressed mathematically by the following relationship:

$$\mathrm{Thr}_{\lambda}^{s}(y,\lambda) = \begin{bmatrix} 0 & \text{if} & |y| \leq \lambda \\ & & \\ sgn(y) & \text{if} & |y-\lambda| > \lambda \end{bmatrix} \tag{4.1}$$

where λ is the Thresholding value.

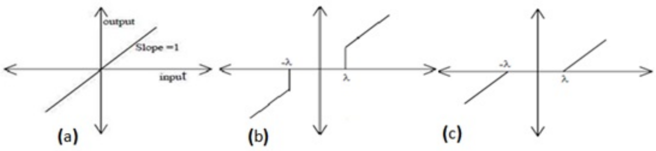

Figure 1: represents the threshold functions of linear (a), solid (b), and elastic (c).

5. Threshold Value

Threshold value (λ) is a very important parameter in the wavelet reduction algorithm to reduce the noise experienced by the signal, as this noise will be directly affected by choosing the appropriate threshold value, there are many types of threshold values.

6. Improvement Thresholding

This method is an improvement of the comprehensive thresholding method proposed by Donoho and Jonstone and given according to the following formula:

$$\lambda_{\text{universal}} = \sigma\sqrt{2\ log\,(n)}/\log_2(j+1) \tag{6.1}$$

whereas
n : the length of the signal.
σ : The standard deviation of the noise level.
Choosing the threshold to be ($\sqrt{2\ log}$) would increase the high probability noise.
With the increase of the decomposition scale (j), the value of the enhanced threshold (λ) gradually decreases to correspond to the noise propagation characteristics at different scales of the wavelet transform to make sure that the noise can be reduced, as the improved threshold method can extract useful information in the signal.

7. Sure Thresholding

Donoho and Jonstone developed an important method for selecting the threshold value, which they called (Sure Shrink). This method is based on Stein's (1981) unbiased risk estimation (SURE) technique, which is an acronym for (Stein Unbiased Risk Estimation) for each wavelet level j. And since the wavelet transformation is orthogonal, so the transformation for noise is also orthogonal, that is, the coefficients ($d_{j,k}$) is also orthogonal, and since the noise is distributed (Gaussian), which means that (d^*) is distributed (Gaussian), and accordingly Stein explained that this estimate is not risk biased.

$$\text{SURE}\ (\lambda_j,\ d_{\text{jk}}) = N - 2\sum_{K=1}^{N} I\left(|d_{\text{jk}}| \leq \lambda_j\right) + \sum_{K=1}^{N} \min\left(|d_{\text{jk}}| \leq \lambda\ \right)^2 \tag{7.1}$$

Therefore, the threshold value for (SURE) can be found from the following formula:

$$\lambda_{k,\ SURE} = \arg\min\nolimits_{0\leq\lambda\leq\sqrt{2\log N}} \text{SURE}\left(\lambda_j,\ d_{\text{jk}}\right) \tag{7.2}$$

8. False Discovery Rate Thresholding

It was presented by Abramovich & Bengamini as one of the methods for selecting the threshold value, as the problem of determining the non-zero distorted wavelet coefficients was formulated here as a multiple hypothesis test problem, for each wavelet parameter we want to test the following hypothesis:

$$H_0 = d_{\text{j.k}} = 0$$
$$H_1 = d_{\text{j.k}} \neq 0$$

to each$j = 0, 1, 2, \ldots, j-1, \quad k = 0, 1, 2, \ldots, 2^j - 1.$
If there is only one hypothesis it will be easy to represent one of the many possible hypothesis tests to make a decision, however since there are many wavelet coefficients the problem is to perform multiple tests. The frequency of the significance test is rarely a single test.
For example, when ($n = 1024$) was ($\alpha = 0.05$), the number of coefficients to be tested is ($n\alpha$), it is equal to 51 coefficients which are assumed to be positive, since sometimes some of these coefficients have a zero sign ($d_{j.k} = 0$) for each (j, k), in other words will be Many coefficients are incorrectly detected as a positive sign.

The basic idea of this method is to assume that (R) is the number of operands that are not set to zero by some threshold procedure, and (S) is held correctly, (S) is the number of operands ($d_{j,k}$) that are not set to zero, and (v)) are the operands that are kept erroneously (i.e. that (v) of the operands of ($d_{j,k}$) should not be kept) because ($d_{j,k}$) is zero for those operands. And that($R = V + S$) expresses the error in such a procedure ($Q = V/R$) which was wrongly kept out of all the parameters that were kept. If ($R = 0$) this means that ($Q = 0$), here the parameter false discovery rate is defined as expectation Q, The FDR method works assuming m the number of parameters is defined as:

a. For each ($d^*_{j,k}$) the value of (P) is calculated on both sides and ($p_{j,k}$) and then we find: ($p_{j,k} = 2(1 - \varnothing(\left|d^*_{\text{j.k}}\right| / \sigma))$
b. Arrange ($p_{j,k}$) by its size (p(1)$\leq p\,(2) \ldots \leq p(m)$) as each (p_i) corresponds to some parameter of ($d_{j,k}$).
c. Let ($i_0 > i$) for ($p_i \leq (i/m)q$)

$$\lambda_{i_0} = \sigma\phi^{-1}(1 - p_{i_0}/2)$$

d. Threshold of all coefficients at level (λ_{i_0}).

9. Visushrink Thresholding

It was presented by [2] and it can be defined according to the following formula

$$\sigma\sqrt{2\ log\,(N)},$$

where (σ) is the variance of noise and (N) represents the number of observations, and this method is efficient even with the increase in the sample size as it gives a more homogeneous and smooth estimation unlike the global threshold methods threshold, which operates under restrictions, the most important of which is that with an increase in the sample size, it is less preparation for the signal, as it leads to a loss of many wavelet coefficients with noise and therefore does not perform well when there are correlations in the signal. This method has been improved through the following formula:

$$\lambda = \sigma\sqrt{\mathrm{logN}} - \sigma_s^2 \tag{9.1}$$

Where (σ_s^2) is the variance of the signal as the increase in the average value of the noisy signal is removed by subtracting the noise variance from the signal variance, while (σ_n) is the standard deviation of the noise level and it can be found through the following formula:

$$\widehat{\sigma}_n = MAD\ (Y)/0.6745 \tag{9.2}$$

Where MAD is the median absolute deviation for the detailed coefficients.

10. Error link issue

In real situations, the noise structure is often coherent and thus wavelet estimations fail to reconstruct a coherent noise signal, due to the fact that the wavelet transformations of a coherent noise signal provide a series of coherent waveform parameters whose differences in the wavelet coefficients will depend on the level of accuracy in Wavelet analysis, but it will be constant at each level.

As a result, the use of a Global Threshold usually breaks down with great difficulty in providing an appropriate threshold value for the wavelet coefficients at all desired levels [7].

Researchers have begun to study situations for which noise is no longer independent eg Chipman (1998) and Opsomer. Therefore, in theory, the associated noise can affect the performance of wavelet gradient, but it is not clear to what extent the known theoretical results reflect what happens in practical situations.

However, the researchers did not stand idly by in front of this problem, and there were many attempts to overcome this problem and obtain efficient estimates, and the most prominent of these treatments is choosing an appropriate threshold value when estimating using wavelet reduction.

11. Estimation methods

The basic idea in the estimation methods is the most appropriate choice of the threshold value, which has a decisive impact on the accuracy and efficiency of the estimation, especially in the case of the correlation of errors, and as it was clarified in the selection methods in the threshold value, so the general method of estimation will be explained and that the difference in the methods used comes from the different threshold values Used in the packing process, which is the main part of the estimation process using wavelet reduction.

The general steps for assessment are summarized as follows:

a. A second-degree polynomial model is used for the purpose of addressing the boundary problem, which is a general problem that nonparametric estimations suffer from, including wavelet estimations, as the function ($\widehat{y}$) is estimated according to the equation below.

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 \tag{11.1}$$

b. Find the residuals (e_i) using the following formula:

$$e_i = y_i - \hat{y}_i \tag{11.2}$$

c. The wavelet reduction method is applied to the residuals through the following:

- Finding the values of the wavelet coefficients (w) by performing the discrete wavelet transform over the residuals (e_i).

$$W = we_i \tag{11.3}$$

  Since (w) is a wavelet transformation matrix with an orthogonal wavelet base.

- The appropriate wavelet coefficients are selected by passing them through the soft threshold and using a threshold value from one of the threshold values shown in the theoretical side to find the threshold coefficients (w^*).
- The estimation of the regression function ($\widehat{g}(t)$) is found by finding the inverse of the discrete wavelet transform (IDWT) according to the following formula:

$$\hat{g}(t) = W^T w^* \tag{11.4}$$

12. The practical aspect of real data

The real data represented the rate of theft crimes for the category of juveniles, as the sample was taken from 2008 to 2018, the eighth month of the monthly rates of theft crimes in the Juvenile Police Department and a sample size of 128, where it was found through the statistics that most of the juveniles arrested for theft crimes were mostly about Through the news, whether the news is direct or indirect. As for the direct method, it is to inform the security forces present in the areas of any crime of theft witnessed by the citizen, so the force moves to arrest the thieves and refer them to the competent courts according to the law.

As for the indirect method, it is a phone call in the event of any theft witnessed by the citizen or reported to the police through the phone numbers of the immediate response to any case of a report of theft, where the information is confirmed by the informant while maintaining the confidentiality of the informant and heading immediately to the place of occurrence the crime.

13. model building

One of the most important stages of building the model is to determine the dependent variable and the explanatory variable based on the nature and type of data under study, which will be clarified as below:

A. The dependent variable (y) (represents the monthly theft rates).
B. The explanatory variable (x) (represents the monthly risk ratios).

14. model analysis

Before starting the estimation process, a (Durbin Watson) test was conducted for the correlation of errors, as it was found that the errors are associated with a correlation of type AR (1) and with a parameter of (0.57). Accordingly, the nonparametric regression function was estimated using the methods described above. The (Visu Polynomial Wavelet) method was followed later by the (Sure Polynomial Wavelet) method, and then the (Pdr Polynomial Wavelet) method, as shown in the table below:

Table 1: shows the values of the mean squares of error (MSE) for the rate of theft crimes for juveniles

Estimation Method	***Sure Polynomial Wavelet***	***Ath Uni Polynomial Wavelet***	***Pdr Polynomial Wavelet***	***Visu Polynomial Wavelet***
MSE	***0.018746***	***0.0202556***	***0.01914485***	***0.007059137***

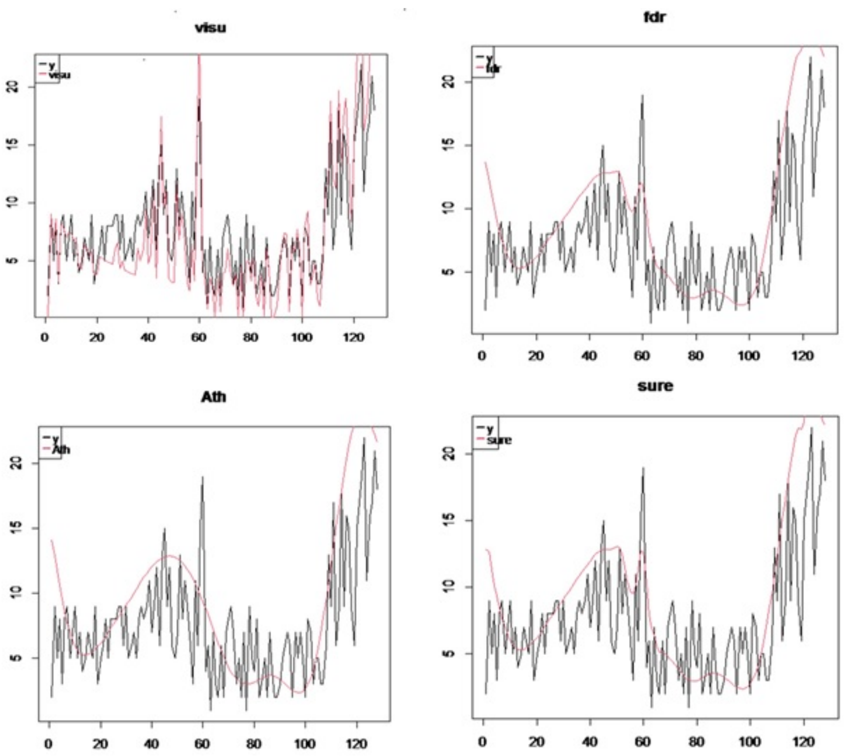

Figure 2: shows the real and estimated data for the dependent variable y (theft crime rate) for the estimation methods used

15. Conclusions

It was found that the behavior of the real data is close to the test function (Doppler), that is, the data signal has different frequencies and associated errors of the type (AR(1)) and a correlation parameter of (0.57) according to the (Durbin Watson) test, and that the best estimation methods are in the presence of errors correlation The (Visushrink) method was followed by the (Sureshrink) method and then the (PDR) method. The reason for this is due to the high flexibility of these methods in dealing with the signal at each level. Also, the use of a hybrid model consisting of a polynomial function with a wavelet regression function It reduces the dimensionality problems resulting from reducing the wavelength in particular and the nonparametric estimates in general. As for the efficiency of the model in estimating the rate of theft crimes in juveniles, we notice a convergence between the real value and the estimated value through the depreciation of the (MSE) value, and it is expected to increase the rate of those crimes than what should be taken necessary measures to prevent this.

References

[1] F. Abramovich and Y. Benjamini, *Adaptive thresholding of wavelet coefficients*, Comput. Stat. Data Anal. 22(4) (1996) 351–361.

[2] L.D. Donoho and M.I. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika 81(3) (1994) 425–55.

[3] C. He, J. Xing, J. Li, Q. Yang and R. Wang,*A New wavelet threshold determination method considering intercale correlation in signal denoising*, Math. Prob. Engin. 2015 (2015) Article ID 280251.

[4] I. Johnstone, *Wavelet shrinkage for correlated data and inverse problems: A daptivity results*, Stat. Sinica 9 (1999) 51–83.

[5] S.C. Mallat, *A Theory for multiresolution signal De composition : The wavelet Representation*, IEEE Trans. Pattern Machine Intell. 11(7) (1989).

[6] S. Mupparaju, B. Naga and V.S. Jahuvvi, *Comparison of varians thresholding techinues of image denoising*, Int. J. Engin. Res. Technol. 2(9) (2013).

[7] G.P. Nason and B.W. Silverman, *Wavelet for Regression and Other Statistical Problems*, School of Mathematics, University of Bristol, 1997.

[8] P.G. Nason, *Choice of The Threshold Parameter in Wavelet Function Estimation*, Springer, 1995.

[9] S.H. Oh and M.C.T. Lee, *Hybrid local polynomial wavelet shrinkage: Wavelet regression with automatic boundary adjustment*, Comput. Stat. Data Anal. 48 (2005) 809–819.

[10] S.H. Oh, P. Naveau and G. Lee, *Polynomial boundary treatment for wavelet regression*, Biometrika 88(1) (2001) 291–298.

[11] R. Porto, P.A. Moretin, D.B. Percival and E.C. Aubin, *Wavelet shrinkage for regression models with random design and correlated errors*, Brazil. J. Probab. Stat. 30(4) (2016) 614–652.

[12] R. Porto and P. Morettin, *Wavelet regression with correlated errors on apiecewise Holder class*, Stat. Probab. Lett. 78(16) (2008) 2739–2743.

[13] O. Shestakov, *Wavelet thresholding risk estimate for the model with random samples correlated noise*, Math. 8 (3) (2020) 377.

[14] F. Xiao and Y. Zhang, *A Comparative study thresholding methods in wavelet-based image denoising*, Procedia Engin. 15 (2011) 3998–4003.

[15] L.J. Yi, L. Hong, Y. Dong and Z.Y. Sheng, *A new wavelet threshold function and denoising application*, Math. Prob. Engin. 2016 (2016) Article ID 3195492.