INTERNATIONAL JOURNAL OF TRANSFORMATIONS IN BUSINESS MANAGEMENT

e-ISSN: 2231-6868, p-ISSN:2454-468X

USING DIFFERENT THRESHOLD VALUES IN WAVELET REDUCTION METHOD TO ESTIMATE THE NONPARAMETRIC REGRESSION MODEL WITH CORRELATION IN ERRORS

*Mohmmed Salh Abdu Alkareem Mahdi, **Dr.Saad Kadem Hamza

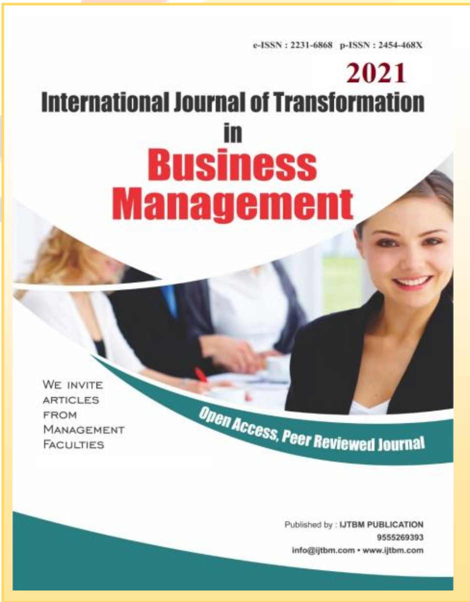
***Ministry of Interior, Directorate of Human Resources, Iraq**

****Department of Statistics, College of Administration and Economics**

University Of Baghdad, Iraq

Paper Received: 20th July, 2021; **Paper Accepted:** 04th September, 2021; **Paper Published:** 05th September, 2021

How to cite the article:
Mohmmed Salh Abdu Alkareem Mahdi, Dr.Saad Kadem Hamza, Using different threshold values in wavelet reduction method to estimate the nonparametric regression model with correlation in errors, IJTBM, July-September 2021, Vol 11, Issue 3; 203-217



ABSTRACT

The wavelet reduction technique is one of the best techniques used in estimating the nonparametric regression function, but it is affected in the event that the errors are related, so (Jonstone) suggested a level-dependent thresholding method to extract the signal from the associated noise. In this paper, a number of types of thresholds will be selected that reduce the risk criterion in estimating the nonparametric regression function and in the presence of a correlation in errors, and these methods are (False Discovery Rate Thresholding), (Bayesshrink Thresholding) and (Universal Thresholding), as simulation experiments were used using Three test and correlation functions of type (AR(1)), sample sizes (64, 128) and different noise ratios. It was found that the best methods were the (False Discovery Rate Thresholding) method, followed by the (Bayesshrink Thresholding) method, while the comprehensive threshold method declined in light of Correlation problem at sample size (128).

LITERATURE REVIEW

Wavelet Regression

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 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 ($y_i = (y_1, \dots \dots \dots, y_n)$) be given in the following form:

$$\boldsymbol{y_i = g(t_i) + \varepsilon_i \dots \dots \dots \dots (1)}$$

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 Donoho 1995, 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 ($\boldsymbol{e_i}$) represent the error or noise, and through the discontinuous wavelet transform, the transformation model can be written as follows:

$$d^* = d + \varepsilon \dots \dots \dots \dots (2)$$

whereas:

$d^* = w_y, d = w_g, \varepsilon = w_\varepsilon$, 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, Dunno and Johnston (1994) 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.

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.

Hard Thresholding

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

$$Thr_{\lambda}^{H}(y,\lambda) = \begin{bmatrix} 0 & if & |y| \leq \lambda \\ y & if & |y| > \lambda \end{bmatrix} \dots\dots\dots(3)$$

where λ is the Thresholding value.

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:

$$Thr_{\lambda}^{s}(y,\lambda) = \begin{bmatrix} 0 & if & |y| \leq \lambda \\ sgn(y) & if & |y-\lambda| > \lambda \end{bmatrix} \dots\dots\dots(4)$$

where λ is the Thresholding value.

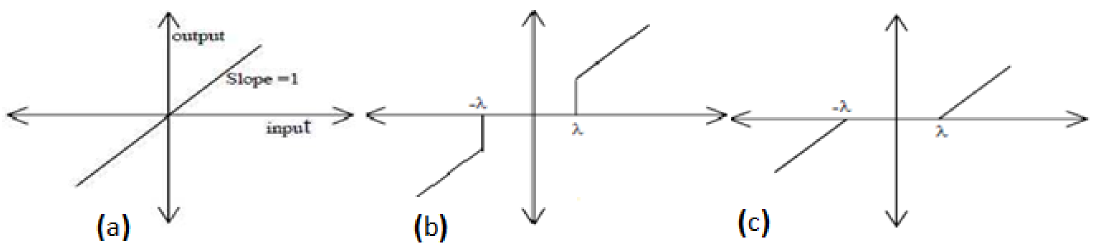


Figure (1) represents the threshold functions of linear (a), solid (b), and elastic (c).

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.

Universal Thresholding

The global threshold method was presented by Donoho and Jonstone and is given according to the following formula:

$$\lambda_{universal} = \sigma\sqrt{2\ log(n)} \dots\dots\dots(5)$$

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.

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:

$H_0 = d_{j,k} = 0$

$H_1 = d_{j,k} \neq 0$

to each $j = 0,1,2,\dots\dots,J-1$, $k = 0,1,2,\dots\dots,2^j - 1$.

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 (nα), 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: ($p_{j,k} = 2(1 - \emptyset(|d^*_{j,k}|/\sigma))$

b. Arrange ($p_{j,k}$) by its size (p(1)≤ $p(2) \dots \le p(m)$) as each (p_i) corresponds to some parameter of ($d_{j,k}$).

c. Let ($i_0 > i$) for ($p_i \le (i/m)q$) $\lambda_{i_0} = \sigma\phi^{-1}(1 - p_{i_0}/2)$

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

Bayesshrink Thresholding

The Bayesian methods are one of the important methods in wavelet reduction, as their work is summarized by reducing the risks (R) in the equation for the bis method:

$R(\lambda) = E(\|\hat{\chi} - \chi\|^2) = E_{\chi} E_{\frac{y}{\chi}}(\|\hat{\chi} - \chi\|^2)$ ....(6)

Where $(\hat{\mathcal{X}} = \boldsymbol{\Theta_s}(\boldsymbol{y},\boldsymbol{\lambda}))$ and $(y/\chi \sim N(0,\sigma^2))$ assuming that (χ) follows the Gaussian Distribution where $(\beta = 2)$.

Accordingly, the optimal threshold value can be obtained by solving the following equation:

$\lambda^*(\sigma_X,\beta) = \min \arg_{\lambda}(R(\lambda))$

Whereas

$$R(\lambda) = E_x E_{y/x}(\|\hat{x} - x\|^2)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\Theta_s(y,\lambda) - x)^2 p(y/x) p(x) dxdy$$

$$= \sigma^2 w(\frac{\sigma_x^2}{\sigma^2}, \frac{\lambda}{\sigma})$$

Where

$$w(\sigma_x^2,\lambda) = \sigma_x^2 + 2(\lambda^2 + 1 - \sigma_x^2)\Phi\left(\frac{\lambda}{\sqrt{1+\sigma_x^2}}\right) - 2\lambda(1+\sigma_x^2)\phi(\lambda, 1+\sigma_x^2)$$

By numerically calculating the approximate threshold, we find that it approaches the optimal threshold so that it is equal to:

$$\lambda = \frac{\sigma^2}{\sigma_x}$$

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 Jonston & Silverman (1997).

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.

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 ($\hat{y}$) is estimated according to the equation below.

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

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

$$e_i = y_i - \hat{y}_i \quad \text{......(8)}$$

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...........(9)

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 ($\hat{g}(t)$) is found by finding the inverse of the discrete wavelet transform (IDWT) according to the following formula:

$$\hat{g}(t) = W^T w^* \quad \text{.......(10)}$$

Test Function

A. Doppler function:

$$f_1(x) = \{x(1-x)\}^{1/2} \sin\{2\pi(1+\varepsilon)/(x+\varepsilon)\}, \varepsilon = 0.05 \quad \text{.........(11)}$$

B. Heavisine function:

$$f_2(x) = 4\sin 4\pi x - \operatorname{sgn}(x-0.3) - \operatorname{sgn}(0.72-x) \quad \text{.........(12)}$$

C. Blocks function:

$$f_3(x) = \sum h_j k(x - x_j), \quad k(x) = \{1+\operatorname{sgn}(x)\}/2 \quad \text{.........(13)}$$

$$(x_j) = (0.1, 0.13, 0.15, 0.23, 0.25, 0.40, 0.44, 0.65, 0.76, 0.78, 0.81)$$

$$(h_j) = (4, -5, -4, 5, -4.2, 2.1, 4.3, -3.1, 2.1, -4.2)$$

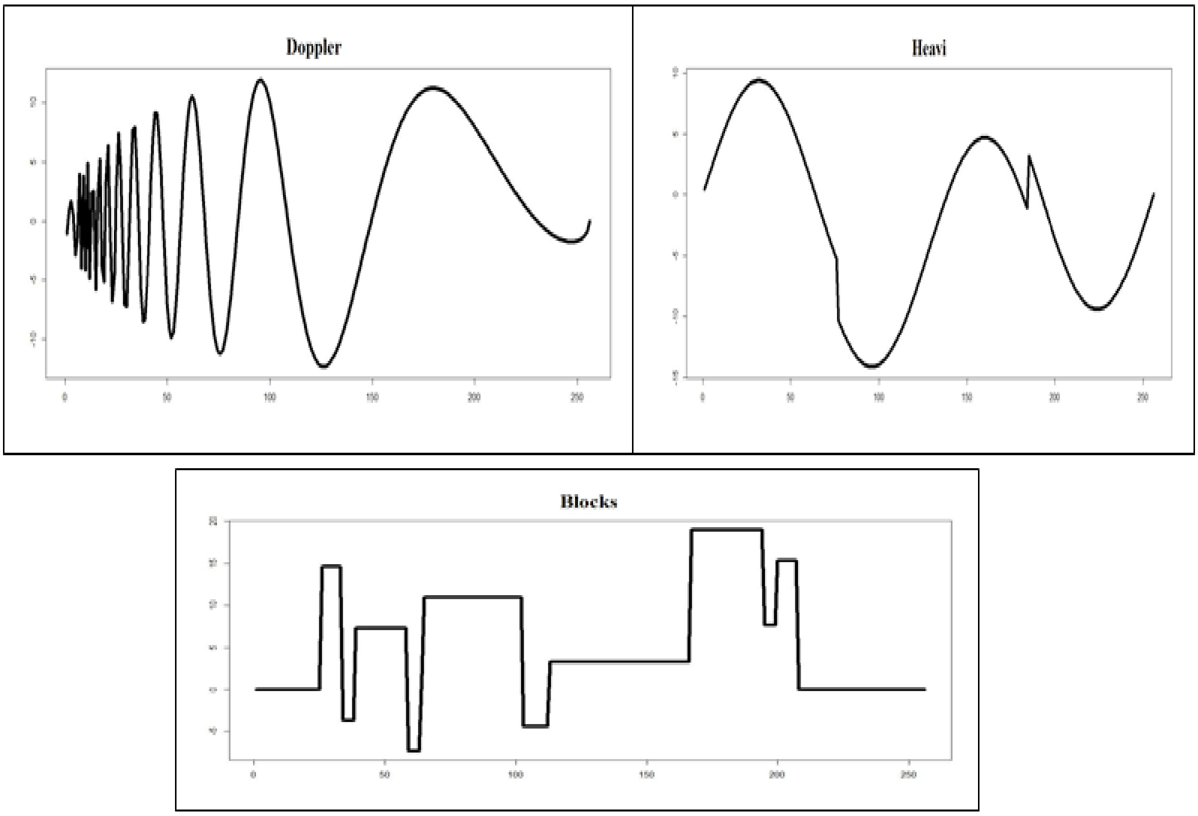


Figure (2) Test functions

Simulation

Three test functions described above were used and a correlation in errors of the type (AR(1)) with two correlation parameters (0.4, 0.7), sample sizes ($2^6 = 64$) and ($2^7 = 128$), two levels of noise were used (SNR =5,10).

CONCLUSIONS:

When using the (Doppler) function and with a correlation parameter (0.4, 0.7) and sample sizes (64, 128) and (SNR = 5, 10), the best estimation methods are the (Universal Thresholding) method, followed by the (False Discovery Rate Thresholding) method, then the (False Discovery Rate Thresholding) method. Bayesshrink Thresholding.

In the case of using the Heavisin function, it was found that the Bayesshrink Thresholding method is highly efficient at (SNR = 5), sample size (64), and correlation (0.40), followed by the (False Discovery Rate Thresholding) method, while the efficiency of the above methods converges. At a correlation (0.70) and a sample size of (64), while at a sample size (128), the (Bayesshrink Thresholding) method gives the best estimate, followed by the (False Discovery Rate Thresholding) method at a correlation (0.40), while the (False Discovery Rate Thresholding) method excels. At a correlation (0.70), followed by (Universal Thresholding). But in the case of using the (Blocks) function, it is clear that the method of (False Discovery Rate Thresholding) is superior to the sample size, correlation parameter, and different disturbance rates.

Table No. (1) shows the average sum of mean squares error (MSE) using the Doppler function

	SNR=5		
	n=64		
	UNIPW	**Bayes PW**	**FDR PW**
0.40	0.031238870	0.044072604	0.034252372
0.70	0.03070347	0.042059405	0.038353180
SNR=5			
n=128			
0.40	0.036249572	0.047403645	0.039540183
0.70	0.036619573	0.048198886	0.041463855
SNR=10			
n=64			
0.40	0.03400737	0.04534398	0.03808286
0.70	0.033615483	0.04872265	0.04227809
SNR=10			
n=128			
0.40	0.036007128	0.046753602	0.04255514
0.70	0.034171310	0.044821539	0.039538667

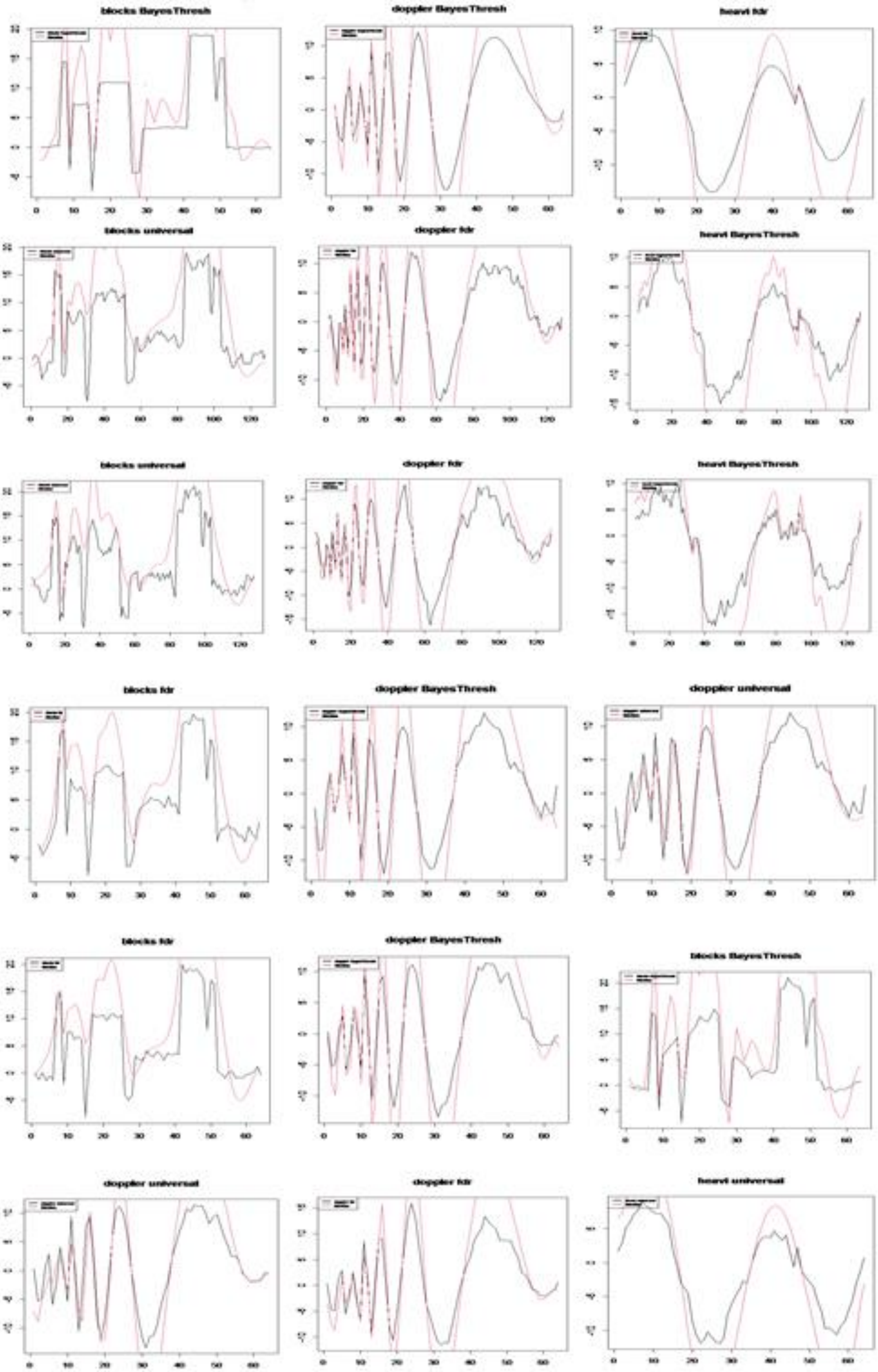
Table No. (2) shows the average sum of mean squares error (MSE) using the Heavisin function

	SNR=5		
	n=64		
	UNIPW	**Bayes PW**	**FDR PW**
0.40	0.05221770	0.05090477	0.05121784
0.70	0.06059649	0.06043290	0.06007991
SNR=5			
n=128			
0.40	0.05694865	0.05343519	0.05375606
0.70	0.05765072	0.05940961	0.05738160
SNR=10			
n=64			
0.40	0.05429333	0.05384871	0.05376025
0.70	0.044154296	0.043331087	0.043295833
SNR=10			
n=128			
0.40	0.05141954	0.05126225	0.05072455
0.70	0.05136628	0.05168991	0.05090908

Table No. (3) shows the average sum of mean squares

error (MSE) using the Blocks function

	SNR=5		
	n=64		
	UNIPW	**Bayes PW**	**FDR PW**
0.40	0.07317259	0.06900870	0.06498236
0.70	0.07031079	0.07235679	0.06401480
SNR=5			
n=128			
0.40	0.06852533	0.07305147	0.06141250
0.70	0.06057767	0.07338275	0.06138189
SNR=10			
n=64			
0.40	0.07162636	0.07157032	0.06607583
0.70	0.08076263	0.08103836	0.07739078
SNR=10			
n=128			
0.40	0.06760367	0.07660869	0.06474073
0.70	0.05993789	0.07221573	0.06074824


blocks BayesThresh
doppler BayesThresh
heavi fdr
blocks universal
doppler fdr
heavi BayesThresh
blocks universal
doppler fdr
heavi BayesThresh
blocks fdr
doppler BayesThresh
doppler universal
blocks fdr
doppler BayesThresh
blocks BayesThresh
doppler universal
doppler fdr
heavi universal

REFERENCES

1. Abramovich, F. , Benjamini, Y. (1996), "Adaptive Thresholding of Wavelet Coefficients" Computational Statistics and Data Analysis .
2. Donoho, L. D. , Johnstone, M. I. , (1994), "Ideal spatial adaptation by wavelet shrinkage", Biometrika, 81, 3, pp. 425-55 .
3. He, C. , Xing, J. , Li, J. , Yang, Q. , Wang, R. (2015), "A New wavelet Threshold Determination Method Considering intercale correlation in signal Denoising " mathematical problems in Engineering, Vol 2015, Article ID 280251 .
4. Johnstone, I. (1999), "Wavelet Shrinkage for Correlated Data and Inverse Problems : A Daptivity Results" Statistica Sinica Vol 9 P 51-83 .
5. Mallat, S. C. (1989), "A Theory for multiresolution signal De composition : The wavelet Representation" IEEE transactions on pattern and Machine intelligence vol. 11. NO. 7.
6. Mupparaju, S. , Naga, B. , Jahuvvi, V. S. (2013), "Comparison of Varians Thresholding Techinues of Image Denoising" International Journal of Engineering Research and Technology . ISSN : 2278-0181, VOL. 2 Issue 9 .
7. Nason, G. P. , Silverman, B. W. (1997), "Wavelet for regression and other statistical problems" school of mathematics, University of Bristol.
8. Nason, P. G. , (1995), "Choice of The Threshold Parameter in Wavelet Function Estimation", Springer, VOL, 103 .
9. Oh, S. H. , Lee, M. C. T. , (2005), "Hybrid Local Polynomial Wavelet Shrinkage : Wavelet Regression With Automatic Boundary Adjustment", Computational Statistics & Data Analysis 48. 809-819 .
10. Oh, S. H. , Naveau, P. , Lee, G. , (2001), "Polynomial Boundary Treatment for Wavelet Regression", Biometrika, 88, 1, pp. 291-298 .
11. Porto, R. , Moretin, P. A. , Percival, D. B. , Aubin, E. C. (2016), "Wavelet Shrinkage for Regression Models with Random Design and Correlated Errors" Brazilian Journal of Probability and Statistics .
12. Porto, R. , Morettin, P. (2008), "Wavelet Regression with Correlated Errors on Apiecewise Holder Class" Statistics and Probability Letters .
13. Shestakov, O. (2020), "wavelet Thresholding Risk Estimate for the model with Random Samples correlated Noise" Faculty of computational Mathematics and Cybernetics, M. V. Lomonosav Moscow state university .

14. Xiao, F. , Zhang, Y. (2011), “A Comparative study thresholding methods in wavelet – based image denoising” procedia engineering 15(2011) 3998-4003 .

15. Yi, L. J. , Hong, L. , Dong, Y. , Sheng, Z. Y. (2016), “A New wavelet Threshold Function and Denoising Application” mathematical problems in Engineering, http: // dx. Doi. Org / 10.1155/2016/3195492 .