PalArch's Journal of Archaeology of Egypt / Egyptology

COMPARISON BETWEEN PROPOSED METHOD FOR WAVELET LASSO AND CLASSICAL PENALIZED METHOD (SIMULATION STUDY)

Dr. Nabeel George Nacy[1], *Alan Ghafur Rahim*[2]

[1,2]Department of Statistics, University of Salahaddin, Erbil, Iraq

E.mail: [1]Nabeel.sulaiman@su.edu.krd, [2]alan.rahim@su.edu.krd

Dr. Nabeel George Nacy, Alan Ghafur Rahim. Comparison Between Proposed Method For Wavelet Lasso And Classical Penalized Method (Simulation Study) -- Palarch's Journal Of Archaeology Of Egypt/Egyptology 19(3), 1695-1711. ISSN 1567-214x

Keywords: Penalized Method; Lasso Regression; Wavelet Shrinkage; Threshold; Wavelet Lasso.

ABSTRACT

We propose an unused strategy for estimation in direct models. The 'Wavelet Lasso' minimizes the remaining sum of squares subject to the entirety of the supreme esteem of the coefficients being less than consistent. Since of the nature of this limitation, it tends to create a few coefficients that are preciseness and consequently give interpretable models.

This paper proposes a simple estimate for tuning parameters based on wavelet shrinkage of penalized method (Lasso) compared with the classic penalized method depending on the tail probability behavior of the response variables and using simulation experiments for (10% and 50%) contamination and real data. The comparing results between the proposed method with a classic penalized method based on the statistical criterion (MAE and MSE). It was concluded that the wavelet shrinkage of penalized method gives the best results and a more accurate classical method for all simulations and real data based on (MAE and MSE) criteria.

INTRODUCTION:

Penalized regression methods for linear regression have been developed over the last few decades to overcome the flaws of ordinary least squares regression with regard to prediction accuracy (Van der Kooij., 2007).

The ordinary least squares (OLS) estimates are obtained by minimizing the residual squared error. There are two reasons why the data analyst is often not satisfied with the OLS estimates. The first is prediction accuracy: the OLS estimates often have low bias but large variance; prediction accuracy can sometimes be improved by shrinking or setting to 0 some coefficients. By doing so we sacrifice a little bias to reduce the variance of the predicted values and hence may improve the overall prediction accuracy. The second reason is interpretation. With a large number of predictors, we often would like to determine a smaller subset that exhibits the strongest effects (Tibshirani., 1996). Multiple regression is often used to estimate a model for predicting future responses or to investigate the relationship between the response variable and the predictor variables. For the first goal, the prediction accuracy of the model is important, for the second goal the complexity of the model is of more interest. Ordinary least squares (OLS) regression is known for often not performing well with respect to both prediction accuracy and model complexity. Several regularized regression methods were developed the last few decades to overcome these flaws of OLS regression, starting with Ridge regression (Hoerl and Kennard 1970a,b), followed by Bridge regression (Frank and Friedman 1993), and the Lasso (Tibshirani1996).

However, the effectiveness of this system is dependent on selecting the tuning parameter that is included in the penalty functions correctly. There are many other approaches for selecting the tuning parameter. They are determined by using a suitable criterion. The desirable selector can be obtained by minimizing this criterion in relation to the tuning parameter. The most well-known current methods are data-driven approaches such as cross-validation (CV) and generalized cross-validation (GCV) (Fan and Li., 2001).

(Donoho and Johnstone., 1995) devised the wavelet threshold approach, which reconstructs signals using thresholding coefficients. The denoising effect of the wavelet threshold approach is determined by the threshold. If the specified threshold is too high, some useful information is filtered out; if the threshold is too low, some noise is preserved. Many academics researched threshold determination approaches in to try to tackle this challenge. (Donoho and Johnstone., 1993) proposed a universal threshold by evaluating a normal Gaussian noise model. The flaw in these systems is that a universal threshold is frequently imposed. The issue in these systems is that the universal threshold is frequently set too high, which might result in excess of relevant information.

All of these methods are based on a specific coefficient distribution, although the distribution may not be applicable to a specific signal. (Donoho and Johnstone.,1994) suggested a new minimax criterion-based threshold technique. However, this method requires prior knowledge about the original signal, which is difficult to obtain in practice. Based on the concept of parameter estimates, Stein's unbiased risk estimate (SURE) criterion and generalized cross validation (GCV) criterion (Jansen and Bultheel.,1999) were presented. SURE criterion is an unbiased estimate of the minimized mean square error (MSE) criterion, and GCV criterion is a biased estimate of the minimized MSE criterion. (Cai and Zhou.,2009) suggested an SURE-based data-driven threshold determination approach. (Autin and von Sachs., 2012) proposed a novel approach by integrating various threshold rules.

In this study, wavelet shrinkage for lasso is proposed for effectively handling of these issues. The effectiveness of the proposed methods is examined through simulation studies and applications in the real data.

PENALIZED METHODS:

Penal methods have appeared in recent years and have gained wide popularity among statisticians, as these methods are an important key to performing the selection of variables and estimating parameters simultaneously, so many penalty methods have been proposed through which a penalty constraint is added to the regression models. (Tutz, G. and Ulbricht, J., 2009)The goal of adding the penalty restriction is to control the complexity of the model and provide a criterion for the selection of variables, by introducing some restrictions on the transactions that impose on some transactions that their value is equal to zero (Helwig, N.E., 2017).

The penalty constraint quantity works to balance the variance and bias in the chosen model. When the penalty amount is small, a larger number of explanatory variables are selected with a small bias, but the variance will be Large, on the contrary, a large penalty amount causes few explanatory variables to be selected with a large bias but the variance will be lower. Therefore, a good choice of penalty amount leads to improving the prediction accuracy and ease of understanding and interpretation of the model (Li, Z. and Sillanpää., 2012)

In general, it is known as Penalized Linear Regression (PLR.).
As follows:

$$PLR(\beta;\lambda) = (Y - X\beta)^T(Y - X\beta) + \lambda \sum_{j=1}^{p} P_\lambda(|\beta_j|) \quad (1)$$

where the amount $P_\lambda(|\beta_j|)$ represents the penalty term, which is a function of coefficients, and (λ) represents the tuning parameter, since $(\lambda \geq 0)$, and that the penalty limit depends entirely on The value of (λ) as it controls the amount of shrinkage of parameter values. When the value is $(\lambda = 0)$ then we get the estimations of the Ordinary Least Squares method (OLS). Conversely, as the value of (λ) increases, the number of variables excluded from the model will increase (Wood., 2006).

In partial linear regression, estimates of the model parameters are found using these equation:

$$\hat{\beta}_{PLR} = \operatorname{argmin}(Y - X\beta)^T(Y - X\beta) + \lambda \sum_{j=1}^{p} P_\lambda(|\beta_j|) \quad (2)$$

The two researchers (Jianging Fan and Li., 2001) suggested that a good penalty term should produce an estimator that has three properties, first, (unbiasedness) when the variable is unbiased for large real parameters. Second, (sparsity) makes small estimators exactly zero. Finally, the estimated continuity is (continuous) in the data to avoid instability in the model prediction.

There are many penalized methods that have been proposed and their characteristics studied, including Ridge, Least Absolute Shrinkage and Selection Operator (LASSO), Elastic-Net, Bridge and other methods.

Lasso Regression (Least Absolute Shrinkage and Selection Operator):

The loss functions for the Lasso can be viewed as constrained versions of the ordinary least squares (OLS) regression loss function. In Lasso Regression constrains the sum of the absolute values of the coefficients as follows (Van der Kooij., 2007):

$$L^{lasso}(\beta_1, \dots, \beta_P) = \left\|y - \sum_{j=1}^{P} \beta_j X_j\right\|^2, \text{ subject to} \sum_{j=1}^{P} |\beta_j| \leq t_1 \quad (3)$$

With N the number of observations, P the number of predictor variables, β_j, (j = 1, . . ., P), the regression coefficients, and t_1 the Lasso tuning parameter, and where $\|.\|^2$ denotes the squared Euclidean norm.

This constrains loss functions can also be written as penalized loss functions:
$L^{lasso}(\beta_1, \dots, \beta_P) = \left\|y - \sum_{j=1}^{P} \beta_j X_j\right\|^2 + \lambda_1 \sum_{j=1}^{P} \text{sign}(\beta_j)\beta_j$ (4) the with λ_1 the Lasso penalty, penalizing the sum of the absolute values of the regression coefficients. In matrix notation, the penalized loss functions are written as:

$$L^{lasso}(\beta_1, \dots, \beta_P) = \|y - X\beta\|^2 + \lambda_1 w^T \beta \quad (5)$$

Where the elements wj of (**w**) are either +1 or −1, depending on the sign of the corresponding regression coefficient β_j.

Minimization of the constrained loss function is more complicated. The regression coefficients are estimated as

$$\beta^{lasso}(\beta_1, \dots, \beta_P) = (X^T X)^{-1}(X^T y + \frac{\lambda_1}{2} w) \quad (6)$$

Wavelets Shrinkage:

Wavelets are functions that satisfy certain mathematical requirements and are used in representing data or other functions. This idea is not new. Approximation using superposition of functions has existed since the early 1800’s, when Joseph Fourier discovered that he could superpose sines and cosines to represent other functions. However, in wavelet analysis, the scale that we use to look at data plays a special role. Wavelet algorithms process data at different scales or resolutions. If we look at a signal with a large “window,” we would notice gross features. Similarly, if we look at a signal with a small “window,” we would notice small features (Antoniadis A.,2007) The result in wavelet analysis is to see both the forest and the trees, so to speak. This makes wavelets interesting and useful. For many decades, scientists have wanted more appropriate functions than the sines and cosines which comprise the bases of Fourier analysis, to approximate choppy signals. By their definition, these functions are non-local (and stretch out to infinity) (Donoho and Johnostone., 1995). They therefore do a very poor job in approximating sharp spikes. But with wavelet analysis, we can use approximating functions that are contained neatly in finite domains. Wavelets are well-suited for approximating data with sharp discontinuities. The wavelet analysis procedure is to adopt a wavelet prototype function, called an analyzing wavelet or mother wavelet. Temporal analysis is performed with a contracted, high-frequency version of the prototype wavelet, while frequency analysis is performed with a dilated, low-frequency version of the same wavelet. Because the original signal or function can be represented in terms of a wavelet expansion (using coefficients in a linear combination of the wavelet functions), data operations can be performed using just the corresponding wavelet coefficients. And if you further choose the best wavelets adapted to your data, or truncate the coefficients below a threshold, your data is sparsely represented. This sparse coding makes wavelets an excellent tool in the field of data compression. Other applied fields that are making use of wavelets include astronomy, acoustics, nuclear engineering, sub-band coding, signal and image processing, neurophysiology, music, magnetic resonance imaging, speech discrimination, optics, fractals, turbulence, earthquake-prediction, radar, human vision, and pure mathematics applications such as solving partial differential equations.

Daubechies Wavelets:

Ingrid Daubechies invented what are called compactly supported orthonormal wavelets , one of the brightest stars in the world of wavelet research, thus making discrete wavelet analysis practicable. The Daubechies family wavelets are written as dbN, where N is the order, db is the family name of the wavelet (Dhamija., 2013).

Advantages:

a) The Daubechies wavelets are orthogonal in nature which is energy preserving.
b) compactly-supported, orthogonal wavelets.

Fejer-Korovkin:

Fejer-Korovkin It's a wavelet filter more symmetric than the Daubechies filters, but less soft. This filter has a wide application on the approximation theory, and a frequency response adequate as the support increases (Varanis, and Pederiva., 2017).

Thresholding:

Thresholding is the simplest method of non-linear wavelet denoising, in which sub dividing the wavelet coefficient in to two sets, one of which represents signal while the other represents noise (Hamad.A.S., 2010) .

There are different rules to apply the thresholds of the wavelet coefficients, and several different methods for choosing a threshold value exist such as:

Universal Threshold:

(Donoho and Johnstone.,1994) proposed universal threshold, which is given by

$$\eta^{U} = \tilde{\sigma}_{(MAD)}\sqrt{2\log N} \qquad (7)$$

Where N is the data length series, and $\tilde{\sigma}_{(MAD)}$ is the estimator of standard deviation of details coefficients, which is estimated as:

$$\tilde{\sigma}_{(MAD)} = \frac{MAD}{0.6745} \qquad (8)$$

MAD is the median absolute deviation of the wavelet coefficients at the finest scale, defined to be.

$$MAD = \text{median}\left[\left|W_{1,0}\right|, \left|W_{1,1}\right|, \dots, \left|W_{1,\frac{N}{2}-1}\right|\right] \qquad (9)$$

So that $W_{1,t}$ represents the element of the W_1 while the constant is the median of the standard normal distribution.

"For a sequence of independently and identically distribution (IID) $N(0, \sigma^2)$ random variables, as $N \to \infty$, so the universal threshold shrank all noise coefficients to zero with high probability, but part of the real underlying might also be lost. Thus, the universal threshold tends to over smoothing.

$$P[\max(|W_n| \leq \eta^{U})] \to 1 \qquad (10)$$

Note, that the combination of the universal threshold and soft thresholding is suggested by Donoho and Johnstone under the name Visu Shrink.

An important feature of visu shrink is that it "guarantees" a noise- free reconstruction although by doing so it usually under fits the data by setting the threshold too height.

SURE Threshold

The sure threshold proposed by (Donoho and Johonstone ,1995), which based upon the minimization of stein's unbiased risk estimator.

In sure threshold specifies a threshold value of η_j for each level j of the wavelet coefficients, then for the soft threshold estimator we have.

$$SURE(\eta, W) = N - 2 \neq \{j: |W_j| \leq \eta\} - \sum_{j=0}^{d} \min(|W_j|, \eta) \qquad (11)$$

Where $\{W_j: j = 1,2, \dots, d\}$ be a wavelet coefficients in the j^{th} level, and

Then, select η^S that minimizes SURE (η, W).

$$\eta^S = \arg\min SURE\,(\eta, W)$$

Donoho and Johonstone (1995) recommended that the SURE threshold is in fact hybrid thresholding approach, utilissing both the universal and SURE threshold. The set of coefficients is judged to be sparsely represented, then the universal threshold is used, otherwise the SURE threshold is used to select a threshold level.

The level j is considered to be sparse if

$$W_{SS}(\eta) \leq 1 + \frac{\left(\log N_j\right)^{\frac{3}{2}}}{\sqrt{N_j}} \qquad (12)$$

Where N_j is the number of wavelet coefficients in the level j, and $W_{SS}(\eta)$ is the sum of square of wavelet coefficients.

$$W_{SS}(\eta) = \sum W_{j,t}^2 \qquad (13)$$

Minimax Threshold

The optimal minimax threshold proposed by Donoho (1995) as an improvement to the universal threshold, Minimax is based on an estimator $\tilde{f}$ that attains to the minimax risk.

$$\tilde{R}(F) = in_{\tilde{f}}fsup_{f\in\tilde{R}(F)}R(\tilde{f}, f) \qquad (14)$$

Where

$$R(\tilde{f}, f) = \frac{1}{N}\sum_{i=1}^{N} E\,[\tilde{f} - f]^2 \qquad (15)$$

Where $f = f(x_i)$ and$\tilde{f} = \tilde{f}(x_i)$, denote the vectors of true and estimated sample values.

The threshold minimax estimator is different from universal counter parts, in which the minimax threshold is concentration on reducing the over all mean square error but the estimates are not over-smoothing.

As already mentioned, when the optimal minimax threshold is incorporated into the soft thresholding rules, it is known the risk shrink.

Proposed Method:

The proposed method is use of wavelet shrinkage for estimate tuning parameter in Penalized linear regression, which depends on the small wave filter after treating it with a threshold rule, and then using the outputs to find (DWT) and get denoise data, and then use this data modified for Wavelet shrinkage for Penalized methods (Wavelet Lasso) in estimating a multiple linear regression model when heavy-tailed distributions and de-noising values are present parameters and calculating (MSE and MAE) comparing it with the classical Penalized methods.

By shrinking the detail coefficients, the inverse DWT is applied to the shrunken set of coefficients. Wavelet shrinkage for each level, we will have a threshold. The Fixed form threshold (i.e.; Universal threshold) technique is considered from equation (7) and put in the place of the tuning parameter from equations (6), then for SURE and Minimax equations (13 and 14) put in the place of the tuning parameter from equations (6) which is as follows:

Tuning parameter estimating by:

Universal threshold $\eta^{U} = \lambda$

$$\beta^{lasso}(\beta_1, \dots, \beta_P) = (X^T X)^{-1}(X^T y + \frac{\eta^U}{2} w) \qquad (16)$$

SURE threshold $W_{SS}(\eta) = \lambda$

$$\beta^{lasso}(\beta_1, \dots, \beta_P) = (X^T X)^{-1}(X^T y + \frac{W_{SS}(\eta)}{2} w) \qquad (17)$$

Minimax threshold $\widetilde{R}(F) = \lambda$

$$\beta^{lasso}(\beta_1, \dots, \beta_P) = (X^T X)^{-1}(X^T y + \frac{\widetilde{R}(F)}{2} w) \qquad (18)$$

Depending on the wavelet matrix such as (db1) and (fk4), we get the values of (observations of the processed response variable), which will be used with the independent variable in estimating the parameters of the multiple linear regression model.

Finally, as shown in picture (2), the methodologies utilized to estimate and compare Penalized linear regression performance in Wavelet shrinkage for penalized methods (wavelet Lasso) will be described:

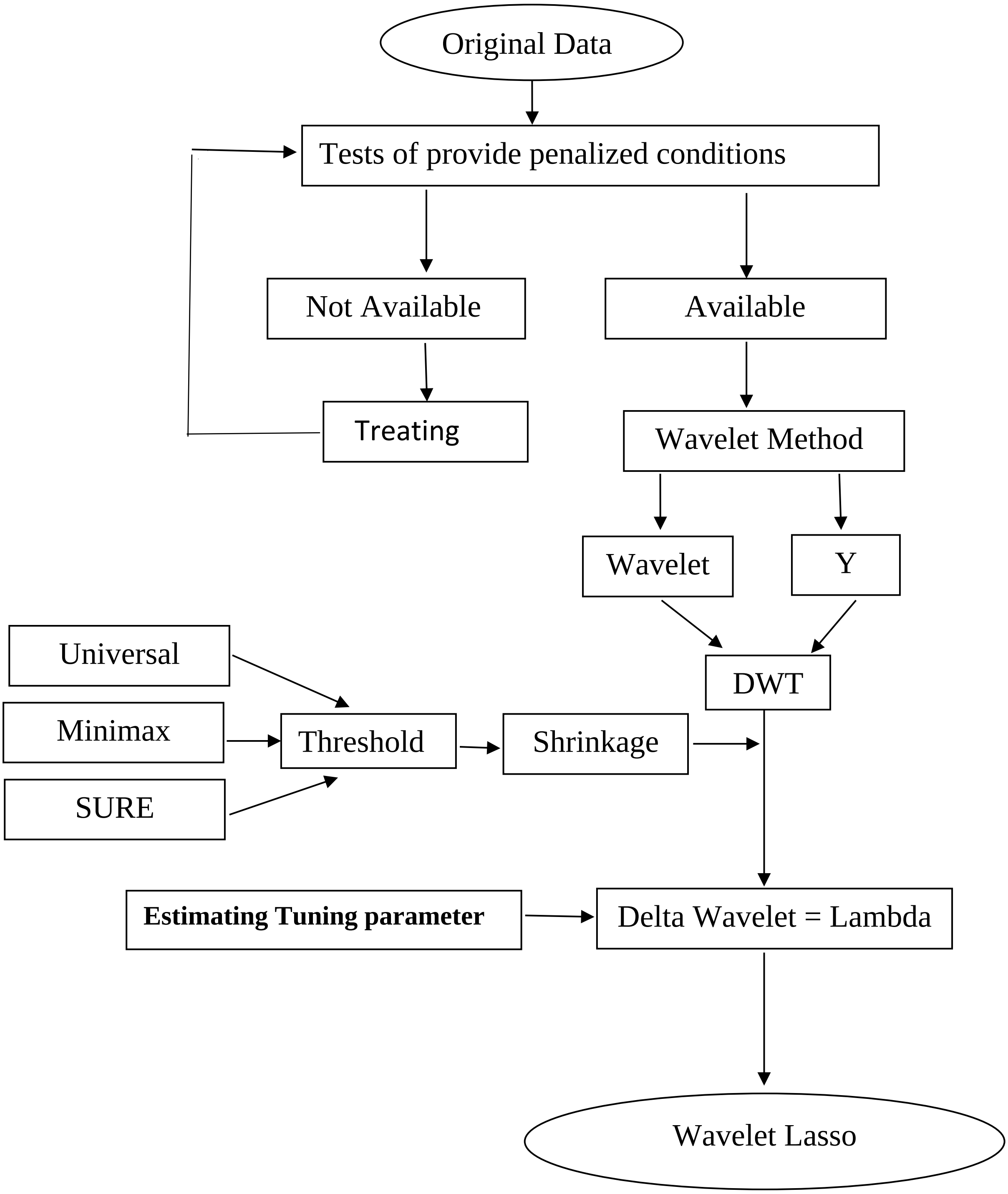

Diagram (1): Proposed Method (Wavelet shrinkage for penalized methods)

Application Part:

This Part included a practical comparison of the methodologies employed in the estimation process represented by Wavelet shrinkage for penalized methods and classical penalized methods. The relative efficiency, which is represented by the mean square of error and mean absolute error, was determined to present with a review of the most essential strategy of regularization for coefficients regression.

Simulation Study:

To implement the simulation experiments, different levels of the following factors were used sample sizes n, Where two sample sizes were used, namely, the simulation experiment included many cases, as two sizes of samples were used, which are (100 and 200) when the number of parameters (P) is equal to(11), and another two sample sizes are (100 and 300) when the number of parameters (P) is equal to(51), and we contaminate of (e_i) vector without modifying explanatory variables such that this contaminated values can cause outliers. Here original (e_i) values are taken from a standard normal distribution with (zero mean and standard deviation equal to 1and 3) and generated (10%, 50%, and 100%) values from the Laplace distribution with (location =2, scale=1). These values produce outliers and contaminate the data by (Hawkins.,1980) using this formula $f(x) = (1 - p) * f_1(x) + p * f_2(x)$. The explanatory variables are independent of a normal distribution (with a mean equal to zero and a standard deviation equal to one). When the number of parameters (P) is equal to (3 -5 0 0 -0.5 0 0 0.5 5 0 0) where **q=5** are numbers of non- zero coefficients, and the second case (P) equal to (2 4 0 -6 0 3 0 1 0 0.5 0 -8 5 0 3 -0.5 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -5) where **q=11** are numbers of non-zero coefficients. For the frequency of (1000) iterations of the assumed regression model and each of the cases shown in tables (1, 2, 3, 4) a comparison was made between the methods used in the estimation process represented by a method Wavelet shrinkage for Penalized methods (Wavelet Lasso) with Classic Penalized methods (Lasso) and parameters can now be defined for the default model. The comparison was made to calculate the relative efficiency, which represents the mean square of error (MSE) and mean absolute error (MAE).

Table 1: The average (MAE and MSE) values for classic and proposed methods.
Where **(σ = 1) and (P=11)**

Wavelet	Criteria	**n=100**					
		Proposed Method					
		10% Contaminate			**50%**Contaminate		
		Universal	SURE	Minimax	Universal	SURE	Minimax
db1	MAE	0.8844	0.9026	0.9010	1.2075	1.2283	1.2271
	MSE	1.4827	1.5475	1.5417	2.5384	2.6410	2.6353
	q	3.9980			3.9980		
FK4	MAE	**0.8842**	0.9027	0.9009	**1.2074**	1.2286	1.2275
	MSE	**1.4812**	1.5479	1.5417	**2.5380**	2.6427	2.6372
	q	3.9980			3.9980		
Classic Lasso							
MAE		3.1985			4.0002		
MSE		12.9643			20.5071		

q		5			5		
n=200							
db1	MAE	**0.9049**	0.9085	0.9083	1.2402	1.2449	1.2448
	MSE	1.4670	1.4829	1.4821	2.5063	2.5321	2.5317
	q	5			5		
FK4	MAE	**0.9049**	0.9085	0.9082	**1.2401**	1.2448	1.2447
	MSE	**1.4664**	1.4828	1.4819	**2.5061**	2.5322	2.5315
	q	3.9990			3.9990		
Classic Lasso							
MAE		3.2036			4.0046		
MSE		12.326			19.473		
q		5			5		

Table 2: The average (MAE and MSE) values for classic and proposed methods.

Where **(σ = 3) and (P=11)**

Wavelet	Criteria	**n=100**					
		Proposed Method					
		10% Contaminate			**50%** Contaminate		
		Universal	SURE	Minimax	Universal	SURE	Minimax
db1	MAE	2.3138	2.3222	2.3211	2.4425	2.4538	2.4530
	MSE	9.4824	9.5506	9.5420	10.5497	10.6437	10.6378
	q	4			4		
FK4	MAE	**2.3137**	2.3221	2.3210	**2.4424**	2.4537	2.4528
	MSE	**9.4806**	9.5490	9.5410	**10.5472**	10.6463	10.6367
	q	4			4		
Classic Lasso							
MAE		3.5954			4.27		
MSE		21.0316			28.5702		
q		5			5		
n=200							
db1	MAE	**2.3841**	2.3859	2.3858	**2.5135**	2.5164	2.5163
	MSE	9.4718	9.4875	9.4866	10.5075	10.5334	10.5329
	q	3.9920			3.9920		
FK4	MAE	**2.3841**	2.3859	2.3858	**2.5135**	2.5164	2.5163

	MSE	**9.4713**	9.4874	9.4863	10.5373	10.5333	**10.5324**
	q	3.9920			3.9920		
Classic Lasso							
MAE		3.6372			4.3053		
MSE		20.4142			27.5666		
q		5			5		

Table 3: The average (MAE and MSE) values for classic and proposed methods Where **(σ = 1) and (P=51)**

Wavelet	Criteria	**n=100**					
		Proposed Method					
		10% Contaminate			**50%** Contaminate		
		Universal	SURE	Minimax	Universal	SURE	Minimax
db1	MAE	**0.7095**	1.1532	1.0898	**0.9244**	1.3571	1.3027
	MSE	**1.6768**	4.5550	4.0486	**2.7375**	6.0497	5.5569
	q	9.9980			9.9980		
FK4	MAE	0.7101	1.1677	1.0986	0.9250	1.3711	1.3127
	MSE	1.6800	4.6611	4.1030	2.7419	6.1586	5.6241
	q	9.9980			9.9980		
Classic Lasso							
MAE		2.2049			2.9986		
MSE		11.4156			20.9249		
q		11			11		
n=300							
db1	MAE	**0.8542**	0.8659	0.8628	**1.1582**	1.1745	1.1721
	MSE	**1.4668**	1.5175	1.5056	**2.5087**	2.5794	2.5687
	q	10			10		
FK4	MAE	**0.8542**	0.8658	0.8626	**1.1582**	1.1745	1.1721
	MSE	**1.4668**	1.5169	1.5050	**2.5087**	2.5791	2.5686
	q	10			10		
Classic Lasso							
MAE		2.2086			3.0039		
MSE		7.2934			2.5087		

q	11	11

Table 4: The average (MAE and MSE) values for classic and proposed methods

Where **(σ = 3) and (P=51)**

Wavelet	Criteria	**n=100**					
		Proposed Method					
		10% Contaminate			**50%** Contaminate		
		Universal	SURE	Minimax	Universal	SURE	Minimax
db1	MAE	**1.7318**	1.9856	1.9427	**1.8249**	2.1014	2.0608
	MSE	**9.6642**	12.7377	12.181	**10.7327**	14.2694	13.7102
	q	9.9890			9.9890		
FK4	MAE	1.7321	1.9864	1.9442	1.8252	2.1053	2.0648
	MSE	9.6667	12.7491	12.1957	10.7371	14.3222	13.7538
	q	9.9890			9.9890		
Classic Lasso							
MAE		2.564			3.2104		
MSE		19.7432			29.2596		
q		11			11		
n=300							
db1	MAE	**2.2347**	2.2409	2.2395	**2.3568**	2.3650	2.3638
	MSE	**9.4746**	9.5264	9.5144	**10.5236**	10.596	10.5851
	q	9.9990			9.9990		
FK4	MAE	**2.2347**	2.2411	2.2394	**2.3568**	2.3651	2.3637
	MSE	**9.4746**	9.5275	9.5137	**10.5236**	10.5963	10.5843
	q	9.9990			9.9990		
Classic Lasso							
MAE		2.8884			3.4688		
MSE		15.3297			21.3667		
q		11			11		

Results Interpretation from Tables (1,2,3, and 4):

A- Show the cases (10% and 50%) that contaminate the proposed method for wavelet types (db1 and Fk4) the average (MAE) and (MSE) is less than the classic method.

B- Noted the proposed method of threshold method (Universal) for the average of (MAE) and (MSE) less than from the case of the proposed method for threshold methods (SURE and Minimax).
C- Show the proposed method for number Non-Zero coefficients is better than Classic method according to the criteria of (q).

D- They found the result increased the rate of contaminate and then increased the values of (MAE and MSE) for all cases.

E- In most cases for sample sizes, the wavelet type (FK4) is shown to be better than the wavelet type (db1) according to the average (MAE and MSE) except in the case of (σ = 1and 3) and (P=51) the wavelet type (db1) is better than the wavelet type (FK4) when sample size equal (100).

Application for Real Data:

To take advantage of the proposed penalized methods data related to studies was used by The prostate cancer data come from a study by Stamey (1989) that examined the correlation between the level of prostate specific antigen and a number of clinical measures in men who were about to receive a radical prostatectomy. The study had a total of 64 observations of male patients aged from 41 to 79 years. The response variable is lpsa - the logarithm of prostate-specific antigen. The covariates are as follows:

1. lcavol - log (cancer volume)
2. lweight - log (prostate weight)
3. age
4. lbph - log (benign prostatic hyperplasia amount)
5. svi - seminal vesicle invasion
6. lcp - log (capsular pEnetration)
7. gleason - Gleason score
8. pgg45 - percentage Gleason scores 4 or 5 As Let us fit linear model relating the log of PSA (lpsa) to the remaining eight variables plus a constant column, that is,

$$\mathrm{lbph} = \beta_0 + \beta_1 \mathrm{lcavol} + \cdots + \beta_8 \mathrm{pgg45} + \epsilon_i$$

Table 5: Estimated coefficients and (MAE and MSE) values for classic and proposed methods of threshold method (Universal) for prostate cancer.

Term	Proposed Method (Lasso)		Classic
	db1	FK4	
Intercept	-0.99	-0.76	1.07
lcavol	0.24	0.24	0.34
lweight	0.57	0.54	0.21
age	-0.01	-0.01	-0.13
lbph	0.17	0.17	0.21
svi	-0.1	-0.1	-0.06

lcp	0.00	0.00	-0.05
gleason	0.01	0.01	0.17
pgg45	0.00	0.00	0.01
MAE	0.4039	**0.4037**	0.7810
MSE	0.3243	**0.3238**	0.9742

Table 6: Estimated coefficients and (MAE and MSE) values for classic and proposed methods of threshold method (SURE) for prostate cancer.

Term	Proposed Method (Lasso)		Classic
	db1	FK4	
Intercept	-1.01	-1.01	1.07
lcavol	0.4	0.4	0.34
lweight	0.07	0.07	0.21
age	-0.02	-0.02	-0.13
lbph	0.17	0.17	0.21
svi	-0.03	-0.03	-0.06
lcp	-0.09	-0.09	-0.05
gleason	0.03	0.03	0.17
pgg45	0.00	0.00	0.01
MAE	**0.4040**	**0.4040**	0.7810
MSE	**0.3246**	**0.3246**	0.9742

Table 7: Estimated coefficients and (MAE and MSE) values for classic and proposed methods of threshold method (Minimax) for prostate cancer.

Term	Proposed Method (Lasso)		Classic
	db1	FK4	
Intercept	-1.01	-1.01	1.07
lcavol	0.4	0.4	0.34
lweight	0.57	0.57	0.21
age	-0.02	-0.02	-0.13
lbph	0.17	0.17	0.21
svi	-0.01	-0.01	-0.06
lcp	-0.09	-0.09	-0.05
gleason	0.00	0.00	0.17
pgg45	0.00	0.00	0.01
MAE	**0.4040**	**0.4040**	0.7810
MSE	**0.3246**	**0.3246**	0.9742

From Tables (5, 6, and 7) where (k=9) for sample sizes (64) we note the following:

A- Show the proposed method for wavelet types (db1 and Fk4) the average means absolute error (MAE) and mean square error (MSE) is less than the classic method.

B- Shows the proposed method of threshold method (Universal) for wavelet type (Fk4) the average of mean absolute error (MAE) and mean square error (MSE) less than from threshold methods (SURE and Minimax).

C- Shows the proposed method of threshold method (Universal and Minimax) for wavelet type (db1 and Fk4) they were selected (6) variables but the classic method selected all variable, it's meant proposed method more efficiency than classic method.

CONCLUSION:

First: Through Simulation study:

1- In the all cases (10% and 50%) contaminates where (σ =1 and 3) and (P=11) for sample sizes (100 and 200) and (P=51) sample sizes (100 and 300) the proposed method for wavelet types (db1 and Fk4) is better than the classic method according to the criterion of (MAE), (MSE) and (q).

2- In most of the case (10% and 50%) contaminate shows the proposed method of threshold method (Universal) for wavelet type (Fk4) according to the criterion of (MAE) and (MSE) less than from (SURE and Minimax) of wavelet type (db1).

3- In most cases for all sample sizes, the wavelet type (FK4) is shown to be better than the wavelet type (db1) according to the average (MAE and MSE).

Second: Through Real Data:

The proposed method better than the classic method according to the criterion for (MAE and MSE) and variables selection.

REFERENCES

Antoniadis A. (2007) "Wavelet methods in statistic: Some recent development and their applications", Statistics surveys, France, Vol.1, pp. (24-28).

Autin, F., Freyermuth, J.M. and von Sachs, R., 2012. Combining thresholding rules: a new way to improve the performance of wavelet estimators. Journal of Nonparametric Statistics, 24(4), pp.905-922.

Dhamija, A., 2013. A brief study of various wavelet families and compression techniques. Journal of Global Research in Computer Science, 4(4), pp.43-49.

Donoho, D. L., (1993) "Nonlinear wavelet methods of recovery for signals, densities, and spectra from indirect and noisy data. in Proceedings of Symposia in Applied Mathematics", volume 47, American Mathematical Society, Providence: RI.

Donoho, D. L. and Johnstone, I. M., (1994a) Ideal denoising in an orthonormal basis chosen from a library of bases. Compt. Rend. Acad. Sci. Paris A, 319, 1317–1322.

Donoho, D. L. and Johnstone, I. M., (1995) Adapting to unknown smoothness via wavelet shrinkage. J.Am. Statist. Ass. 90, 1200–1224.

Fan, J. and Li, R. (2001), Variable selection via nonconcave penalized likelihood and its oracle properties, J. Amer. Statist. Assos., 96, 1348-1360.

Frank, L.E. and Friedman, J.H., 1993. A statistical view of some chemometrics regression tools. Technometrics, 35(2), pp.109-135.

Hamad.A.S. (2010) "Using some thresholding rules in wavelet shrinkage to denoise signals for simple regression with application in Rezgary hospital – Erbil", PhD. dissertation in statistics, college of Administration and Economic, university of Sulaimania, Iraq.

Hawkins, D. M. (1980). Identification of outliers (Vol. 11): Springer.

Helwig, N.E., 2017. Adding bias to reduce variance in psychological results: A tutorial on penalized regression. The Quantitative Methods for Psychology, 13(1), pp.1-19.

Hoerl, A.E. and Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), pp.55-67.

Li, Z. and Sillanpää, M.J., 2012. Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection. Theoretical and applied genetics, 125(3), pp.419-435.

Jansen, M. and Bultheel, A., 1999. Multiple wavelet threshold estimation by generalized cross validation for images with correlated noise. IEEE transactions on image processing, 8(7), pp.947-953.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58, 267–288.

T. T. Cai and H. H. Zhou, "A data-driven block thresholding approach to wavelet estimation," The Annals of Statistics, vol. 37, no. 2, pp. 569–595, 2009.

Tutz, G. and Ulbricht, J., 2009. Penalized regression with correlation-based penalty. Statistics and Computing, 19(3), pp.239-253.

Van der Kooij, A.J., 2007. Prediction accuracy and stability of regression with optimal scaling transformations. Leiden University.

Varanis, M. and Pederiva, R., 2017. The influence of the wavelet filter in the parameters extraction for signal classification: An experimental study. Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, 5(1).

Wood, S.N., 2006. On confidence intervals for generalized additive models based on penalized regression splines. Australian & New Zealand Journal of Statistics, 48(4), pp.445-464.