PalArch's Journal of Archaeology of Egypt / Egyptology

USING A PROPOSED METHOD FOR WAVELET SHRINKAGE TO ESTIMATE THE TUNING PARAMETER OF PENALIZED LINEAR REGRESSION

Asst.Prof. Dr. Nabeel George Sulaiman[1], Alan Ghafur Rahim[2]

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

Email: [1]nabeel.sulaiman@su.edu.krd [2]alan.rahim@su.edu.krd

Asst.Prof. Dr. Nabeel George Sulaiman, Alan Ghafur Rahim. Using A Proposed Method for Wavelet Shrinkage to Estimate the Tuning Parameter of Penalized Linear Regression -- Palarch's Journal of Archaeology of Egypt/Egyptology 19(3), 424-439. ISSN 1567-214x

Keywords: Penalized Method; Ridge Regression; Elastic-Net Regression; Wavelet; Threshold.

ABSTRACT

The tuning parameter selection strategy for penalized estimation is crucial to identifying a model that is both interpretable and predictive. However, popular strategies (e.g., wavelet shrinkage is proposed for effectively handle in these issues) tend to select models with more predictors than necessary. This paper proposes a simple estimate for tuning parameters based on wavelet shrinkage of penalized method (Ridge and Elastic-Net) compared with the classic penalized method depending on the tail probability behavior of the response variables and using simulation experiments for (10%) data with 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:

The penalized least squares method has been repeatedly shown to be an appealing regression shrinkage and selection method. This process differs from standard approaches to variable selection in that it identifies significant variables while also estimating regression coefficients. The estimators produced are as efficient as the Oracle estimator. Furthermore, non-significant variables are eliminated by estimating their coefficients as recent related research includes (Van der Kooij, A.J., 2007).

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 extended cross validation (GCV) (T. T. Cai and H. H. Zhou., 2009).

(Donoho and Johnstone., 2011) 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 try to tackle this challenge. Donoho and Johnstone proposed a universal threshold by evaluating a normal Gaussian noise model; Tao et al. (Z. Tao, H.-M. Zhao, X.-J. Zhang, and D. Wu.,2011) enhanced the universal threshold and suggested that it be altered adaptively when the scale changes. 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. Chang et al. (S. G. Chang, B. Yu, and M. Vetterli.,2000) proposed a Bayesian threshold method based on the assumption that the wavelet coefficients followed a generalized Gaussian distribution; (Lu and Loizou Y. Lu and P. C. Loizou., 2011) assumed the coefficients followed a Gaussian distribution and presented a new threshold based on maximum a posteriori probability; Li et al. assuming the coefficients followed a generalized Gamma distribution, we proposed a threshold technique based on Bayesian shrinkage. 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 (M. Jansen and A. 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 a SURE-based data-driven threshold determination approach. (Autin and von Sachs., 2012) proposed a novel approach by integrating various threshold rules.

In this study, penalized methods with wavelet shrinkage are 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,

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

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

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, Simon., 2006).

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

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

The two researchers (2001) (Jianging Fan and Li) 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.

Ridge Regression:

Regression modeling with associated explanatory variables presents a challenging problem when selecting variables and estimating parameters. The reason for this is, in the case of multicollinearity, the data matrix does not have enough information to distinguish the effect of a correlated variable versus a variable another related. In choosing a variable, selection methods tend to choose arbitrarily for one of the variables associated and does not take into account the significance of the specified variable. In addition, the existence of plurality linearity impairs prediction accuracy by amplifying the variance of parameter estimates, which may lead to removing significant coefficients from the model (Van der Kooij, A.J., 2007).

The ridge regression method was proposed by (Hoerl and Kennard) (1970) and it is considered one of the oldest penalty methods, as it received great attention because of its ability to overcome the problem of multicollin- earity without removing the explanatory variables from the regression model. The Ridge Regression method reduces the variance in the coefficient estimates by adding a penalty quantity that follows the rule (L2 - norm) to the sum of the squares of the residuals, as the penalty quantity reduces the regression coefficients.

Penalty linear regression is defined using the ridge term as follows:

$$PLR(\beta;\lambda)^{Ridge} = (Y - X\beta)^T(Y - X\beta) + \lambda \sum_{j=1}^{p} \beta_j^2 \qquad (3)$$

So that the penalty term $(\sum_{j=1}^{p} \beta_j^2)$ represents the (L2 - norm) rule the estimates of the parameters in the penalty regression model can be obtained from equation (1.7) as follows:

$$\hat{\beta}_{PLR}^{Ridge} = (X^T X + \lambda I)^{-1} X^T y \qquad (4)$$

Since I is the identity matrix with capacity P and $\boldsymbol{\lambda}$ is the positive shrinkage parameter, adding $\boldsymbol{\lambda}$ to the main diameter elements in the $(X^T X)$ information matrix reduces the variance of the OLS estimates with the addition of an amount of bias to it. When ($\boldsymbol{\lambda}$=0) the OLS estimators are obtained from Equation (4). Although the resulting estimations from using the ridge term are biased, the ridge regression method improves the prediction accuracy.

In ridge regression, the coefficients are gradually reduced towards zero, but they do not make them equal to zero at all, and then all the variables remain in the model, as a result, it is not possible in the ridge regression method to choose the variables and therefore the resulting linear regression model cannot be easily explained, especially if the number of Large explanatory variables.

Elastic Net Regression:

Elastic net regression combines the penalty terms of ridge and lasso regression. When fitting models with elastic net, we minimize the function. Zou and Hastie (2005) have proposed the Elastic Net and developed an algorithm, called LARS-EN, based on the efficient LARS algorithm, to overcome the Lasso limitations of selecting at most N predictors and of selecting only one predictor from a group of highly correlated predictors. For the Elastic-Net the regression coefficients are estimated as

$$\hat{\beta}_{PLR}^{Enet} = (X^T X + \lambda_2 I)^{-1} \left(X^T y - \frac{\lambda_1}{2}\, sign(B_j^{OLS})\right) \qquad (5)$$

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 .D.L, and Johnostone, I, M., 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, A., 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, M. and Pederiva, R., 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

$$\eta^{U} = \tilde{\sigma}_{(MAD)}\sqrt{2logN} \qquad (6)$$

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

$$\tilde{\sigma}_{(MAD)} = \frac{MAD}{0.6745} \qquad (7)$$

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

$$MAD = median\left[|W_{1,0}|, |W_{1,1}|, \dots, \left|W_{1,\frac{N}{2}-1}\right|\right] \qquad (8)$$

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.

$$P[max(|W_n| \leq \eta^{U})] \to 1 \qquad (9)$$

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[23][47].

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.

$$SURE(\eta, W) = N - 2 \neq \{j: |W_j| \leq \eta\} - \sum_{j=0}^{d} min(|W_j|, \eta) \qquad (10)$$

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

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

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.

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

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 the inverse (DWT) and get denoise data, and then use this data modified for Wavelet shrinkage for Penalized methods (Wavelet for Ridge and Elastic-Net) 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 (6) and put in the place of the tuning parameter from equations (4 and 5), which is as follows:

Tuning parameter estimating by:

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

$$\hat{\beta}_{PLR}^{Ridge} = \left(X^T X + \eta^{U} I\right)^{-1} X^T y \qquad (14)$$

$$\hat{\beta}_{PLR}^{Enet} = \left(X^T X + \eta^{U} I\right)^{-1} \left(X^T y - \frac{\eta^{U}}{2} sign(B_j^{OLS})\right) \qquad (15)$$

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

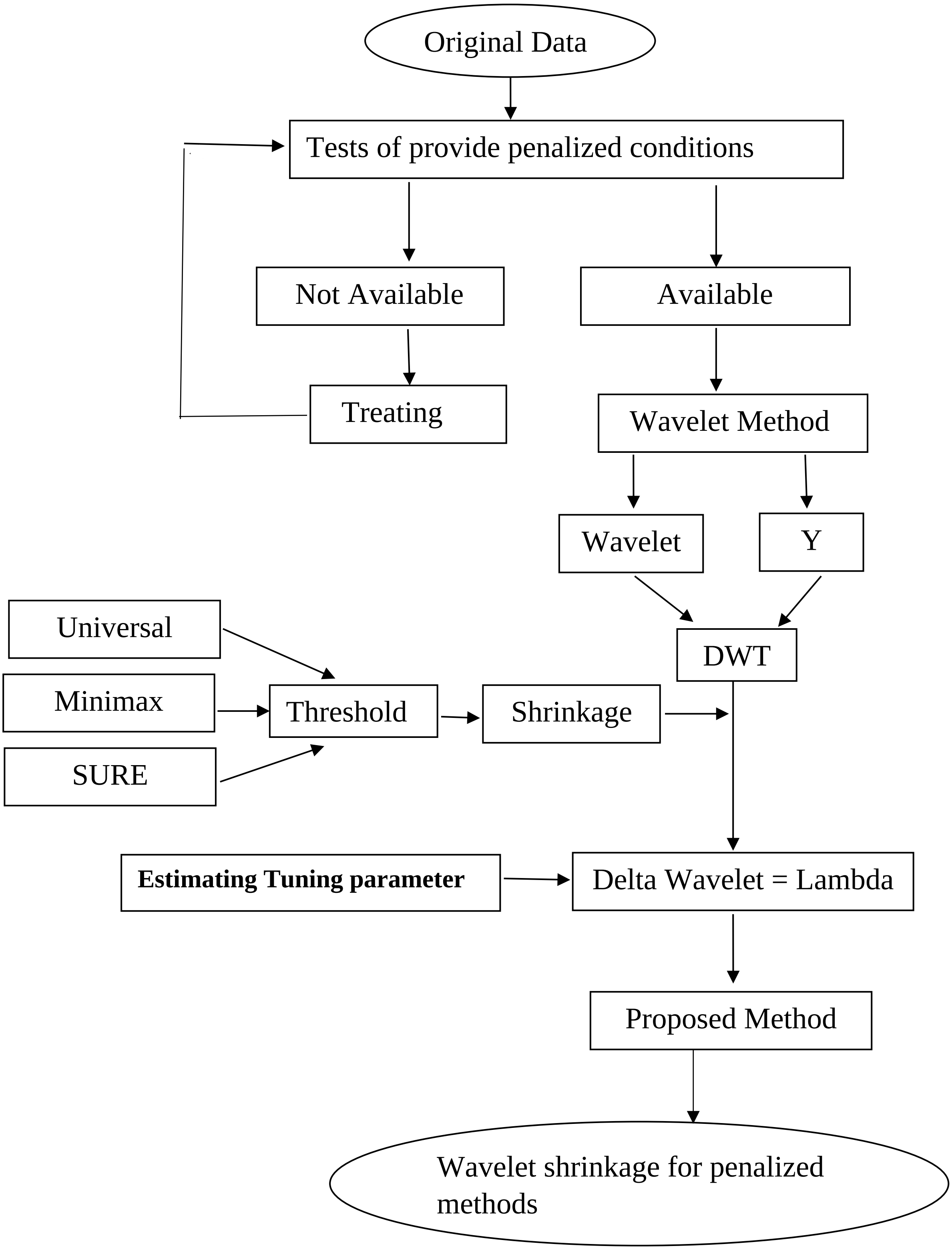
$$\hat{\beta}_{PLR}^{Ridge} = \left(X^T X + Wss(\eta) * \mathrm{I}\right)^{-1} X^T y \qquad (16)$$

$$\hat{\beta}_{PLR}^{Enet} = \left(X^T X + Wss(\eta) * \mathrm{I}\right)^{-1} \left(X^T y - \frac{Wss(\eta)}{2} sign(B_j^{OLS})\right) \qquad (17)$$

Depending on the wavelet matrix such as (db1) and (fk4), we get the values of (observations of the processed dependent 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 Ridge and Elastic-Net) 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 three sample sizes were used, namely, the simulation experiment included many cases, as three sizes of samples were used, which are (50, 100, 200) when the number of parameters (k) is equal to(11), and another three sample sizes are (100, 150, 300) when the number of parameters (k) is equal to(51), and we contaminate (10) 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 8) and generated (10%) values from the Laplace distribution with (location =2, scale=1). These values produce outliers and contaminate the data by 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 (k) is equal to (3 -5 0 0 -0.5 0 0 0.5 5 0 0) where q=5 are numbers of non- zero parameters, and the second case (K) 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 parameters. 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 for Ridge and Elastic-Net) with Classic Penalized methods (Ridge and Elastic-Net) 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.

Wavelet	Method	Threshold Method	**(σ = 1) (k=11) (10% Contaminate)**					
			n =50		**n =100**		**n =200**	
			MAE	MSE	MAE	MSE	MAE	MSE
db1	Rd	Universal	0.9616	1.9541	0.9483	1.6843	0.9277	1.5413
		SURE	1.0022	2.1151	0.9993	1.8554	0.9698	1.6763
	EN	Universal	0.9950	2.1143	0.9536	1.7141	0.9284	1.5478
		SURE	1.1058	2.5982	1.0203	1.9513	0.9751	1.7024
Fk4	Rd	Universal	**0.9609**	**1.9519**	**0.9467**	**1.6782**	**0.9254**	**1.5340**
		SURE	1.0022	2.1145	0.9998	1.8572	0.9697	1.6760
	EN	Universal	0.9934	2.1062	0.9516	1.7066	0.9259	1.5398
		SURE	1.1050	2.5891	1.0209	1.9536	0.9751	1.7020

Classic						
Ridge	1.4945	4.5948	1.1857	2.5760	1.0528	1.9488
Elastic Net	3.1967	14.6203	3.1985	12.9662	3.2035	12.3272

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

Wavelet	Method	Threshold Method	**(σ = 8) (k=11) (10% Contaminate)**					
			n =50		**n =100**		**n =200**	
			MAE	MSE	MAE	MSE	MAE	MSE
db1	Rd	Universal	5.7266	66.3029	6.0630	64.9929	6.2352	64.6483
		SURE	5.7420	66.6571	6.0747	65.2331	6.2432	64.8121
	EN	Universal	5.7416	66.6204	6.0660	65.0576	6.2358	64.6631
		SURE	5.7812	67.4997	6.0818	65.3860	6.2449	64.8497
Fk4	Rd	Universal	**5.7262**	**66.2880**	**6.0624**	**64.9802**	**6.2350**	**64.6449**
		SURE	5.7435	66.6834	6.0747	65.2350	6.2432	64.8142
	EN	Universal	5.7404	66.5932	6.0655	65.0426	6.2356	64.6592
		SURE	5.7821	67.5266	6.0820	65.3868	6.2449	64.8521
Classic								
Ridge			5.8509	69.0303	6.0979	65.7469	6.2533	65.0051
Elastic Net			6.3155	79.8423	6.6088	76.6624	6.7668	75.8330

Table 3: The average (MAE and MSE) values for classic and proposed methods

Wavelet	Method	Threshold Method	**(σ = 1) (k=51) (10% Contaminate)**					
			n =100		**n =150**		**n =300**	
			MAE	MSE	MAE	MSE	MAE	MSE
db1	Rd	Universal	**0.6729**	**1.4908**	**0.7735**	**1.4803**	**0.8555**	**1.4673**
		SURE	0.6748	1.4993	0.7762	1.4899	0.8579	1.4745
	EN	Universal	0.7142	1.6987	0.7772	1.5046	0.8551	1.4692
		SURE	1.1378	4.4286	0.8836	1.9670	0.8691	1.5260
Fk4	Rd	Universal	**0.6729**	1.4909	0.7736	1.4804	**0.8555**	**1.4673**
		SURE	0.6749	1.4994	0.7762	1.4899	0.8579	1.4746
	EN	Universal	0.7148	1.7017	0.7773	1.5046	0.8551	1.4692
		SURE	1.1523	4.5345	0.8798	1.9501	0.8690	1.5256
Classic								
Ridge			3.6320	42.2253	2.9381	20.5330	1.9262	7.0710
Elastic Net			2.2043	11.4712	2.2042	8.8484	2.2091	7.3172

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

Wavelet	Method	Threshold Method	**(σ = 8) (k=51) (10% Contaminate)**					
			n =100		**n =150**		**n =300**	
			MAE	MSE	MAE	MSE	MAE	MSE
db1	Rd	Universal	**4.4992**	65.2412	**5.2248**	**65.0795**	5.8381	64.5659
		SURE	4.5076	65.4817	5.2338	65.2986	5.8425	64.6662
	EN	Universal	4.5098	65.5536	5.2263	65.1184	5.8383	64.5695
		SURE	4.6429	69.5224	5.2584	65.2931	5.8460	64.7325
Fk4	Rd	Universal	**4.4992**	**65.2410**	**5.2248**	65.0800	**5.8380**	**64.5656**
		SURE	4.5076	65.4829	5.2338	65.2977	5.8425	64.6672
	EN	Universal	4.5098	65.5557	5.2263	65.1192	5.8383	64.5692
		SURE	4.6425	69.5056	5.2575	65.8959	5.8461	64.7347
Classic								
Ridge			6.0924	119.1600	6.0655	87.6014	6.1057	70.5770
Elastic Net			4.8877	76.8281	5.5427	73.1492	6.1133	70.6422

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

A- Show the case (10%) 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).
C- Show the proposed method for Penalized method (Ridge) is better than (Elastic-Net) according to the criterion of (MAE) and (MSE).
D- They found the result increased the sample sizes then decreased the values of (MAE and MSE).
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 = 1) and (k=51) the wavelet type (db1) is better than the wavelet type (FK4).

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 et al. (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,

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

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

Term	Proposed Method (Ridge)		Classic	Proposed Method (Elastic -Net)		Classic
	db1	FK4		db1	FK4	
Intercept	4.75	2.96	1.03	-0.17	-0.25	1.07
lcavol	0.47	0.45	0.40	0.38	0.39	0.34
lweight	0.22	0.19	0.55	0.41	0.45	0.21
age	-0.04	-0.03	-0.03	-0.02	-0.02	-0.14
lbph	0.04	0.22	0.17	0.17	0.17	0.22
svi	-0.1	-0.16	-0.29	-0.32	-0.3	-0.06
lcp	-0.06	-0.07	-0.09	-0.01	-0.08	-0.05
gleason	-0.05	-0.04	0.28	0.22	-	0.17
pgg45	-	0.01	0.07	-	-	0.01
MAE	0.4509	**0.4238**	5.2838	0.4238	**0.4096**	0.7826
MSE	0.4289	**0.3784**	35.6354	0.3784	**0.3420**	0.9777

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

Term	Proposed Method (Ridge)		Classic	Proposed Method (Elastic -Net)		Classic
	db1	FK4		db1	FK4	
Intercept	9.45	9.23	1.03	0.8	0.79	1.07
lcavol	0.78	0.78	0.40	0.24	0.24	0.34
lweight	-0.34	-0.32	0.55	-0.07	-0.07	0.21
age	-0.09	-0.09	-0.03	0.02	0.02	-0.14
lbph	0.55	0.55	0.17	0.13	0.13	0.22
svi	0.74	0.74	-0.29	-0.51	-0.51	-0.06

lcp	0.07	0.07	-0.09	0.02	0.02	-0.05
gleason	-0.34	-0.32	0.28	-	-	0.17
pgg45	0.06	0.06	0.07	0.01	-	0.01
MAE	0.6554	**0.6551**	5.2838	0.5219	**0.5205**	0.7826
MSE	0.5944	**0.5939**	35.6354	0.4825	**0.4802**	0.9777

From Table (5and 6) 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) for (Elastic-Net) less than from wavelet type (db1 and FK4) for (Ridge).

CONCLUSION:
First: Through Simulation study:

1- In the case (10%) contaminates where (σ =1 and 8) and (k=11 and k=51) for all cases sample sizes (50, 100, 200) and (100, 150, 300) the proposed method for wavelet types (db1 and Fk4) is better than the classic method according to the criterion of (MAE) and (MSE).
2- In most of the case (10%) 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) of wavelet type (db1).
3- The results explained that whenever increasing the sample size leads to an increase in the values of (MAE) and a decrease in the values of (MSE). As increasing the standard deviation values lead to an increase in the values of (MAE and MSE).
4- Most of the results showed that the (Ridge) method is better than (Elastic-Net) according to the criterion for (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. As well as (Elastic-Net) method is the best method from (Ridge).

REFERENCES
Abramovich F., Bailey T. C and Sapatinas T. (2000) "Wavelet analysis and its statistical applications", The Statistician, Vol.49 (1), pp.I-29.
Antoniadis A. (2007) "Wavelet methods in statistic: Some recent development and their applications", Statistics surveys, France, Vol.1, pp. (24-28).
B. Le, Z. Liu, and T. Gu, “Weak LFM signal detection based on wavelet transform modulus maxima denoising and other techniques,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 8, no. 2, pp. 313–326, 2010
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. 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.

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

Donoho, D. L., (1993b) "Unconditional bases are optimal bases for data compression and statistical estimation". App. Comp. Harm. Anal., 1, 100–115.

Donoho, D. L., (1995a) "De-noising by soft-thresholding" IEEE Trans. Inf. Th., 41, 613–627

Donoho, D. L., (1995b) "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition" App. Comp. Harm. Anal., 2, 101–26.

Fan, J. and Li, R. (2006), Statistical challenges with high dimensionality: Feature selection in knowledge discovery, Proceedings of the International Congress of Mathematicians, M. Sanz-Sole, J. Soria, J.L. Varona, and J. Verdera, eds., Vol. III, 595-622. [8] Golub, G.H., Van Loan, C.F.

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

Gencay, R.,Selcuk.,F.,andWhithcher ,b. (2002),"An Introduction to Wavelet and other Filtering Methods in Finance and Economics ",Turkey.

G. Gupta, “Algorithm for image processing using improved median filter and comparison of mean, median and improved median filter,” *International Journal of Soft Computing and Engineering*, vol. 1, no. 5, pp. 2231–2307, 2011

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.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58, 267–288. Tibshirani, R. J., Taylor, J. E., Lockhart, R. A. and Tibshirani, R. (2014) Exact postselection inference for sequential regression procedures. Journal of the American Statistical Association, 111, 600 – 620.

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.

Vidakovic, B., (1999) Statistical Modeling by Wavelets. Wiley, New York.

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).

Walker, J. S., 1999. "A Primer on Wavelets and Their Scientific Applications", 1st Edition.Studies in Advanced Mathematics. CRC Press LLC, 2000 N. W. Corporate Blvd., Boca Raton, Florida, U.S.A.

Wood, Simon N. "On confidence intervals for generalized additive models based on penalized regression splines." *Australian & New Zealand Journal of Statistics* 48, no. 4 (2006): 445-464

Zou, H. and Hastie, T., 2005. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), pp.301-320.