PalArch's Journal of Archaeology of Egypt / Egyptology

COMPARISON BETWEEN WAVELET BAYESIAN AND BAYESIAN ESTIMATORSTOREMEDYCONTAMINATION IN LINEAR REGRESSION MODEL

Prof. Dr. Taha Hussien Ali[1], Lect. Dlshad Mahmood Saleh[2]

[1]Department of Statistics, University of Salahaddin, Erbil, Iraq. Email: taha.ali@su.edu.krd

[2]Department of Statistics, University of Salahaddin, Erbil, Iraq. Email: dlshad.saleh@su.edu.krd

Prof. Dr. Taha Hussien Ali , Lect. Dlshad Mahmood Saleh , Comparison Between Wavelet Bayesian And Bayesian Estimatorstoremedycontamination In Linear Regression Model , Palarch's Journal Of Archaeology Of Egypt/Egyptology 18(10), 3388-3409. ISSN 1567-214x.

Keywords: Linear Model; Bayesian; Wavelet; contamination; shrinkage.

ABSRTACT:

This paper proposes a new method that depends on the wavelet shrinkage and linking it to a Bayesian estimate and it is called wavelet Bayesian method for reducing the effect of contamination and then relying on the outputs of this method to estimate the multiple regression modelthrough using simulation experiment for several of data contamination and real data. The comparing results between the proposed method with traditional Bayesian methodbased on the statistical criterion(RMSE).It was concluded that the Bayesian estimation using wavelet shirking filter gives the best results and more accurate thantraditional method for all simulations and real data based on RMSE criterion.

1: Introduction

Wavelet regression is a technique which attempts to reduce noise in a sampled function corrupted with noise. This is done by thresholding the small wavelet decomposition coefficients which represent mostly noise. It is well understood that conflicting sources of information may contaminate the inference when the classical normality of errors is assumed. The contamination caused by the light normal tails follows from an undesirable effect: the posterior concentrates in an area in between the different sources with a large enough scaling to incorporate them all. Outliers are observations that appear inconsistent with the rest of the data (Barnett, V. and Lewis, T., 1994). Bayesian wavelet shrinkage has been widely used in several areas to reduce noise in data analysis .The theory of conflict resolution in Bayesian statistics (O'Hagan and Pericchi,2012) recommends addressing this problem by limiting the impact of outliers to obtain conclusions consistent with the bulk of the data. (Philippe G, et al.., 2020) proposed a model with super heavy-tailed errors to achieve this. We prove that it is wholly robust, meaning that the impact of outliers gradually vanishes as they move further and further away from the general trend. The super heavytailed density is similar to the normal outside of the tails, which gives rise to an efficient estimation procedure. In addition, estimates are easily computed. This is highlighted via a detailed user guide, where all steps are explained through a simulated case study.

(Alex R. S. et al., 2021) proposed the use of a zero-contaminated beta distribution with a support symmetric around zero as the prior distribution for the location parameter in the wavelet domain in models with additive Gaussian errors. The hyper parameters of the proposed model are closely related to the shrinkage level, which facilitates their elicitation and interpretation. For signals with a low signal-to-noise ratio, the associated Bayesian shrinkage rules provide significant improvement in performance in simulation studies when compared with standard techniques.

The wavelet theory is one of the modern and important theories with wide and different uses in various theoretical and applied fields. Recently, Wavelet shrinkage estimation has recently become a powerful mathematical technique for de-noising of function estimation, based on thresholding parameter and the choice of this threshold determines, to a great extent the efficiency and success of de-noising. In general, if the threshold parameter set too high, then signal structure will be lost. Alternatively, if it is set too low, then noise will be presented in the estimate.

Since the early days of modern Bayesian inference one central issue has, of course, been the potentially strong dependence of the inferences on the prior. In particular in situations where data is scarce or unreliable, the actual estimate obtained by Bayesian techniques may rely heavily on the shape of prior knowledge, expressed as prior probability distributions on the model parameters.. In this paper dealt with presenting the proposed method that depends on the wavelet shrinkage and linking it to a Bayesian estimate and it is called the wavelet Bayesian method for Reducing the Effect of Outliers in Linear Regression Model.

2: Theoretical part

2.1:Bayesian multiple linear regression model

In this study, we will discuss the idea of estimating the parameters of the multiple linear regression models using the Bayesian method based on the Natural Conjugate Prior (Informative Prior) Probability Distribution (Zellner, A. ,1971) &(Shih-Hsien Tseng, M.S. ,2008):

<u>Informative Prior Probability Distribution</u>

Assume that we have the following model (Harrison.J. and Mike.W., 1989):
$Y = X\beta + \underline{e}$ (1)

Whereas: $\underline{e} \sim \mathrm{N}(0, \sigma^2 \mathrm{I}_n)$

$\therefore (Y \mid X, \beta, \sigma^2) \sim N(X\beta, \sigma^2 I_n)$

(σ^2) is known

Depending on this case, the Prior density function with the normal vector of the parameters (β) representing the normal distribution is known as follows:

$\beta \sim N(\beta_0, \sigma^2 M_0^{-1})$

The kernel of the probability density function is:

$$f(\beta) \propto \exp\left[\frac{-1}{2\sigma^2}(\beta - \beta_0)' M_0 (\beta - \beta_0)\right] \quad -\infty < \beta < \infty, \quad \sigma^2 > 0 \qquad (2)$$

Whereas:

$M_0 = X_0' X_0$: Represents the informative Prior matrix
β_0: Represents the mean of the Prior distribution

$\sigma^2 M_0^{-1}$: Represents the variance and covariance matrix of the Prior distribution.

As for the likelihood function, it is:

$$L(\beta, \sigma^2) \propto \exp\left[\frac{-1}{2\sigma^2}(\beta - \hat{\beta}_{OLS})'(X'X)(\beta - \hat{\beta}_{OLS})\right] \qquad (3)$$

Through Bayes' theorem, the function (2) can be combined with the function (3) to obtain the posterior probability density function of the parameter vector β as follows:

$$f(\beta \mid Y, \sigma) \propto f(\beta) L(\beta, \sigma^2)$$

$$\propto \exp\left[\frac{-1}{2\sigma^2}(\beta-\hat{\beta}_{OLS})'(X'X)(\beta-\hat{\beta}_{OLS})+(\beta-\beta_0)'M_0(\beta-\beta_0)\right]$$

$$\propto \exp\left[\frac{-1}{2\sigma^2}(\beta-\hat{\beta}_{Bayse})'(X'X+M_0)(\beta-\hat{\beta}_{Bayse})\right],\ -\infty<\beta<\infty\ , \qquad (4)$$

Whereas:

$$\hat{\beta}_{Bayse}=(X'X+M_0)^{-1}(X'Y+M_0\beta_0) \qquad (5)$$

The function (4) represents the kernel of a multivariate normal distribution with mean ($\hat{\beta}_{Bayse}$), which represents a Bayes estimator for the parameter vector (β) and that the variance and covariance matrix can be defined as follows:

$$V-Cov(\hat{\beta}_{Bayse})=\sigma^2(X'X+M_0)^{-1} \qquad (6)$$

2.2: Contaminations

Contamination caused by outliers is inevitable in data analysis, and robust statistical methods are often needed kanamori, T. &fujisawa, H., (2015). The data come from two types of distributions the first is called Distribution Basic that generates good data While the second type, which is called Distribution Contaminating, which generates Outliers. This can be explained mathematically, if it was $f_1(x)$ is the probability density function of the basic distribution and $f_2(x)$ represents the probability of the contaminant distribution, then the distribution of any observations will be (Hawkins, 1980):

$$f(x)=(1-p)*f_1(x)+p*f_2(x) \qquad (7)$$

Where the standard normal distribution $f_1(x)$ is used for the regular data points, and $f_2(x)$ is used for the contamination is chosen as the mixture distributions or Heavy tailed distribution.

2.3: Wavelet Analysis

The main goal of using filters in regression analysis of linear models is to exclude noise. The recent trends in the analysis of regression models are based on the use of wavelet filters instead of the usual filters, which are of course better and more efficient than the usual filters(Hamad.A.S, 2010).

Using the discrete Wavelet Transform coefficients as filters for the contaminate observations presented by Researchers (Morris J. M. and Peravali R., 1999).

A filter can be considered as an operator on ℓ_2 (discrete form of $L^2(R)$) in to itself, a filter applied to a signal contaminated with noise to isolate the signal or to extract the noise. In general, an observed (discrete) observation g is represented by a sequence $\{g_i\}$ $i \in z$, assuming $g \in \ell^2(z)$.

A filter (A) can be presented by a $\ell^2(z)$ sequence $\{ai\} i \in z$. The filtering process consists of a discrete convolution of the filter sequence with the observation.

Applying the filter (A) to the observation (g) is written as:

$$(Ag)_i = \sum_{\ell \in z} a_{\ell - 2i} g\ell$$

Yielding a new observation indexed by (i), which ranges over (z).

The DWT is based on filter H and G defined respectively by $\{h_i\} i \in z$ and $\{g_i\} i \in z$ that were derived from the multiresolution analysis, the filters must be satisfied the following conditions:

1- The stability of coefficient h_i.

$$\sum_{i=0}^{n-1} h_i = 2 \qquad (8)$$

2- Requiring the convergence of wavelet expansion forces the condition

$$\sum_{i=0}^{n-1} (-1)^i r^m h_i = 0 \qquad m = 0,1,\ldots,\frac{n}{2}-1 \qquad (9)$$

3- The orthogonally of wavelets requires the condition

$$\sum_{i=0}^{n-1} h_i h_{i+2m} = 0 \qquad m = 0,1,\ldots,\frac{n}{2}-1 \qquad (10)$$

4- Finally if the scaling function is required to be orthogonal.

$$\sum_{i=0}^{n-1} h_i^2 = 2 \qquad (11)$$

2.3.1: Discrete Wavelet Transform

Discrete Wavelet Transform is coefficients that summarize the information of all observations with a smaller number and are located in the domain of time and frequency. Discrete Wavelet Transformation (DWT) is used in many different areas of life, especially when there is contaminate or noise in the data. (DWT) decomposes a signal by using scaled and shifted versions of a compact supported basis function (WalkerJ.S., 1999).& (Abramovich F., et al., 2000)

Given a vector of a signal (X) consisting of 2^j observation. The (DWT) of X is

$$W = wX \qquad (12)$$

Where W is a $(n*1)$ vector comprising both discrete scaling and wavelet coefficients. The vector of wavelet coefficients can by organized into $j+1$ vectors.

$$W = [W1, W2, \dots, Wj_0, Vj_0]^T$$

Where Wj is a length $N_j = N/2^j$ vector of wavelet coefficients (Details) associated with changes on a scale of length $\lambda_j = 2^{j-1}$ symbol as CD, and(V_{jo})is a length $N_{j0} = N/2^j$ vector of scaling coefficients (approximation or smoothing) associated with average on a scale of length $\lambda_{J_0} = 2^{J_0}$ symbol as CA, and w is an orthonormal $N*N$ matrix associated with the orthonormal wavelet basis chosen(Antoniadis,A.,2007)(Gencay,R.,et al.,2002).

After each DWT, the approximation coefficients are divided into bands using the same filter as before, with the result that the details are appended with the details of the latest decomposition, at each level, the signal can be reconstructed of the de-noise signal by the inverse transform.

$$X = W\,w^T = \sum_{j=1}^{j_0} W_j^T W_j + V_{j_0}^T V_{j_0} \qquad (13)$$

2.3.2: Wavelet Shrinkage

One of the primary applications that are made to the view after its analysis by wavelet transformation is the noise or contaminates removal of the observation. And by using Discrete Wavelet Transformation (DWT), the scientists found that the noise produced after the conversion has a frequency lower than the frequency of the original observation. In order to estimate the noisy signal, Shrinkage is usually used to reduce the risk level or reduce the noise or outlier by relying on Thresholding,any appropriate frequency threshold setting such that this threshold cancels the noise coefficients and maintains the original observation coefficients. This represents the simplest non-linear reduction of the wavelet coefficients introduced by (Donoho&Johnston) and thus obtaining a summary of the significant transformation coefficients that pass the threshold cut as a test for them so that the coefficients are zero if their absolute value is less than a certain threshold cut level, attempts to recover a signal $g(t)$ from noisy anobservation $x(i)$.

$$x(i) = g(i) + v(i) \qquad i = 0,1,2,\dots, N-1 \qquad (17)$$

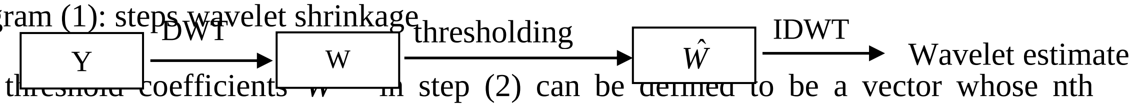
Therefore, the following basic steps must be taken, which represent a summary of the wavelet shrinkage method:

1- The data are transformed to a different representation called wavelet coefficients by the DWT. They are multiplied by an orthogonal matrix W.

2- The wavelet coefficients are modified using a thresholding rule. Reducing the number of coefficients is the basic of wave shrink.

3- The inverse discrete wavelet transformation (IDWT) is applied to the modified coefficients to obtain an estimate of the signal. So the resulting three-step wavelet shrinkage procedure can be summarized by the following diagram:

Diagram (1): steps wavelet shrinkage



The threshold coefficients $\hat{W}$ in step (2) can be defined to be a vector whose nth element $W_n^{(i)}$ is such that

$$W_n^{(i)} = \begin{bmatrix} 0 & if \ |W_n| \le \eta \\ Some\,non\,zero\,value & otherwise \end{bmatrix}$$

The wavelet shrinkage has several good properties that gained this popularity in statistics (Donoho .D.L, and Johnostone, I, M., 1995): nearly minmax for a wide range of loss function and for general function classes; simple, practical and fast; adaptable to spatial and frequency in homogeneities; readily extendable to high dimensions; applicable to various problems such as density estimation and inverse problems.

2.4: Proposed methods

The proposed method is use of wavelet shrinkage in multiple linear regression model, which depends on the small wave filter after treating it with a threshold, and then using the outputs to find the inverse of the (DWT) and get denoisedata, and then use this data modified to estimating parameters Bayesian multiple linear regression and calculating RMSE and comparing it with the aforementioned classical methods.

For the purpose of isolating contaminate from the values of observations of the dependent variable, one of the types of threshold such as(hard or soft) is usually used by shrinkage the detail coefficients, which we can get from re-covering the original observations and splitting them into two components using wavelets. The first represents the sum of the coefficients Details, while the second represents the smoothing parameters based on Multiple Re-Resolution Analysis (MRA), that is:

$$\boldsymbol{Y} = \boldsymbol{w}^T\boldsymbol{W} = \sum_{j=1}^{J_o} \boldsymbol{w}_j^T\boldsymbol{W}_j + \boldsymbol{v}_{J_o}^T\boldsymbol{V}_{J_o} \qquad (18)$$

The threshold level is estimated by one of the well-known methods, including the fixed from threshold method at level j = 1 only(W_1). And then using the soft threshold in treating the DWT coefficients and by returning the remaining coefficients to the vector elements (w) we can get the DWT coefficients of the modified wavelet usually symbolized by (W'), through which it is possible to re-cover the observations of the treating dependent variable, i.e.:

$$\tilde{Y} = w^T W' \tag{19}$$

Depending on the wavelet matrix such as (db7) and (bior1.1), 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 modelsdepending on the methods i.e.:

$$\underline{\hat{\beta}}_{wavelet\ baysian} = (X'X + M_0)^{-1}(X'\tilde{Y} + M_0\beta_0) \tag{20}$$

Finally, the methods used will be summarized (Bayesian, wavelet Bayesian) in analyzing the multiple linear regression models and comparing its efficiency with the presence of outlier and noise values through the following diagram:

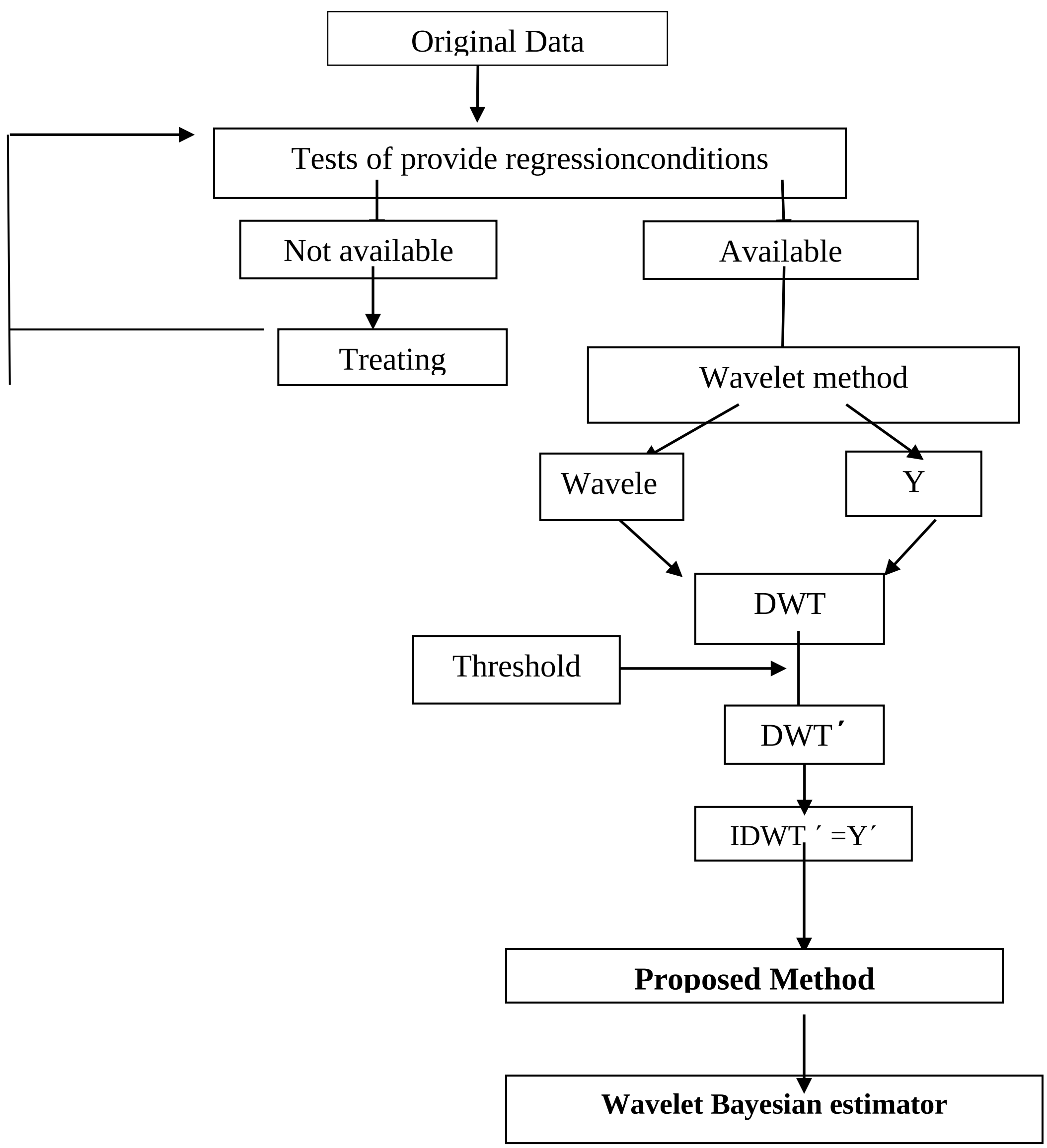


Diagram (2): for proposed method (Wavelet Bayesian)

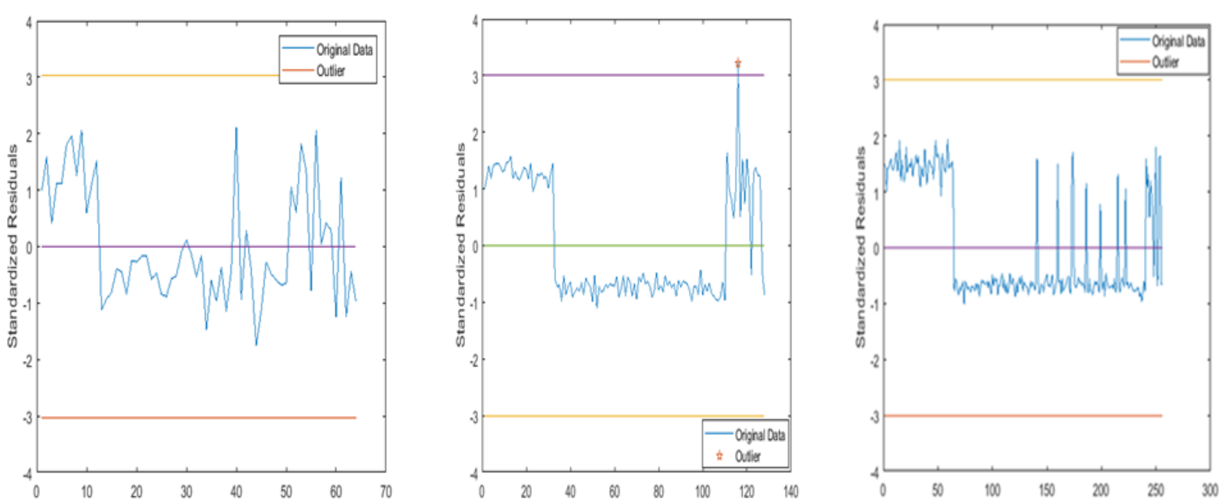
3: Application Part

In this part, a practical comparison was conducted between the methods used in the estimation process represented by wavelet Bayesian methodandtraditional Bayesian method. To present with a review of the most important approach of minimizing contaminated values in the data, the relative efficiency, which is represented by the root mean square of error, was calculated.

3.1: Simulation experiment

To implement the simulation experiments, different levels of the following factors were used sample sizes n, Where three sample sizes were used, namely $2^6 = 64$, $2^7 = 128$, $2^8 = 256$ The sample size here should be $n = 2^j$ whereas(j) a positive integer. When the number of parameters (k) is equal to (2, 5, 10) and we contaminate (10%& 40 %) of (ei) vector without modifying explanatory variables such that this contaminated values can cause outliers. Here original (ei) values are taken from standard normal distribution with (zero mean and standard and generated (10%)&(40%) values from Cauchy distribution. Obviously, these values produce outliers and contaminate in the data by using this formula (7).The explanatory variables independently from a normal distribution (with a mean equal to zero and standard deviation equal to one). For the frequency of (1000) replications of the assumed regression model and for each of the cases shown in the tables (1, 2, 3, 4, 5, 6).

If n=(64 ,128,256) and (40%)contaminate .



If n=(64 ,128,256) and (10%)contaminate .

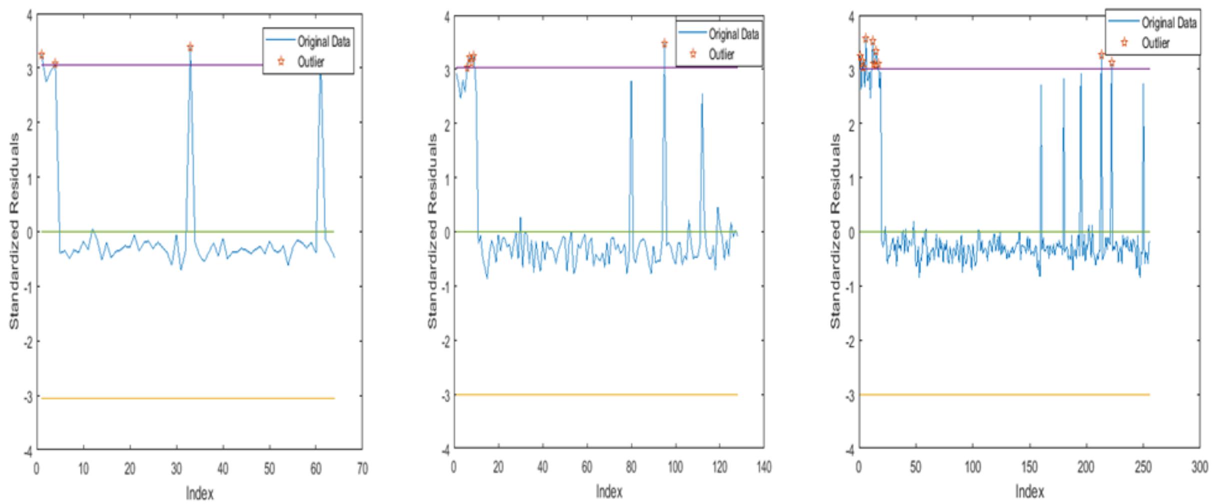


Figure 1. Standadirazed Residual plot

Method Estimation	RMSE %10 contaminate	RMSE %40 contaminate
n=64		
Bayesian	3.2256	5.1041
Wavelet Bayesian -bior1.1	2.4271	4.1172
Wavelet Bayesian -db7	**2.2961**	**3.3961**
n=128		
Bayesian	2.9986	3.6660
Wavelet Bayesian -bior1.1	2.2305	3.6573
Wavelet Bayesian -db7	**1.9761**	**3.4092**
n=256		
Bayesian	2.6913	3.6587
Wavelet Bayesian -bior1.1	1.9074	3.1454
Wavelet Bayesian -db7	**1.7659**	**2.8753**

Table 1: RMSE values for different estimation methods when (σ = 1, k=2)

Method Estimation	RMSE %10 contaminate	RMSE %40 contaminate
n=64		
Bayesian	5.2086	7.7043
Wavelet Bayesian -bior1.1	4.0428	**6.0413**
Wavelet Bayesian -db7	**4.0304**	6.4470
n=128		
Bayesian	4.1984	5.8854
Wavelet Bayesian -bior1.1	3.2687	5.9119
Wavelet Bayesian -db7	**3.1463**	**5.7411**
n=256		
Bayesian	4.2774	4.0228
Wavelet Bayesian -bior1.1	3.2932	3.6964
Wavelet Bayesian -db7	**3.2013**	**3.5056**

Table 2: RMSE values for different estimation methods when (σ = 1, k=5)

Method Estimation	RMSE %10 contaminate	RMSE %40 contaminate
n=64		
Bayesian	4.6845	6.1221
Wavelet Bayesian -bior1.1	**3.1489**	4.3711
Wavelet Bayesian -db7	3.2658	**3.9746**
n=128		
Bayesian	4.5548	5.0123
Wavelet Bayesian -bior1.1	3.0445	4.3623
Wavelet Bayesian -db7	**2.9937**	**4.1569**
n=256		
Bayesian	4.3763	5.0136
Wavelet Bayesian -	**2.9301**	3.8699

bior1.1		
Wavelet Bayesian -db7	2.9588	**3.7126**

Table 3: RMSE values for different estimation methods when (σ = 1, k=10)

Method Estimation	RMSE %10 contaminate	RMSE %40 contaminate
n=64		

Bayesian	5.1212	6.8818
Wavelet Bayesian -	**4.0083**	**5.9054**

Table 4: RMSE values for different estimation methods when (σ = 5, k=2)

Method Estimation	RMSE %10 contaminate	RMSE %40 contaminate
n=64		
Bayesian	5.5316	7.6691
Wavelet Bayesian - bior1.1	4.2156	**6.3495**
Wavelet Bayesian -db7	**4.0412**	6.4866
n=128		
Bayesian	5.3765	6.7349
Wavelet Bayesian - bior1.1	4.1178	5.9663
Wavelet Bayesian -db7	**3.9329**	**6.2285**
n=256		
Bayesian	5.4918	5.3023
Wavelet Bayesian - bior1.1	4.0354	4.4214
Wavelet Bayesian -db7	**3.9791**	**4.2590**

Table 5: RMSE values for different estimation methods when (σ = 5, k=5)

bior1.1		
Wavelet Bayesian -db7	**4.0708**	**5.7766**
n=256		
Bayesian	5.4554	5.3816
Wavelet Bayesian -bior1.1	4.1958	4.3577
Wavelet Bayesian -db7	**4.1473**	**4.1617**

Table 6: RMSE values for different estimation methods when (σ = 5, k=10)

The Simulation Results

The results in Tables (1, 2, 3, 4, 5 and 6) indicate that the allRMSE values of the models estimated using the wavelet Bayesian method is less than the RMSE value of the models estimated using the Bayesian method and they gave the best estimate of the model used in the (db-7) and (bior1.1) wavelet. This result means that the models estimated using thewavelet Bayesian method is better than the Bayesian method.

3.2: The Real data

The data set consists of seven variables for (32) countries. Gunst and Mason (1980, Appendix A), indicate that these data are a subset of a larger data set (data set 41 of Loether et al., 1974). We use the same nomenclature as Gunst and Mason (1980), from which the data were taken, namely,

X_1:Infant deaths per 1000 live births (INFD).

X_2:Number of inhabitants per physician (PHYS).

X_3:Population per square kilometer (DENS).

X_4:Population per 1000 hectares of agricultural land (AGDS).

X_5:Percentage literate of population aged 15 years and over (LIT).

X_6:Number of students enrolled in higher education per 100,000 populations (HIED).

Y: gross national product per capita, 1957 U.S. dollars (GNP).

Let us fit a linear model relating the GNP to the remaining six variables plus a constant column, that is,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon_i$$

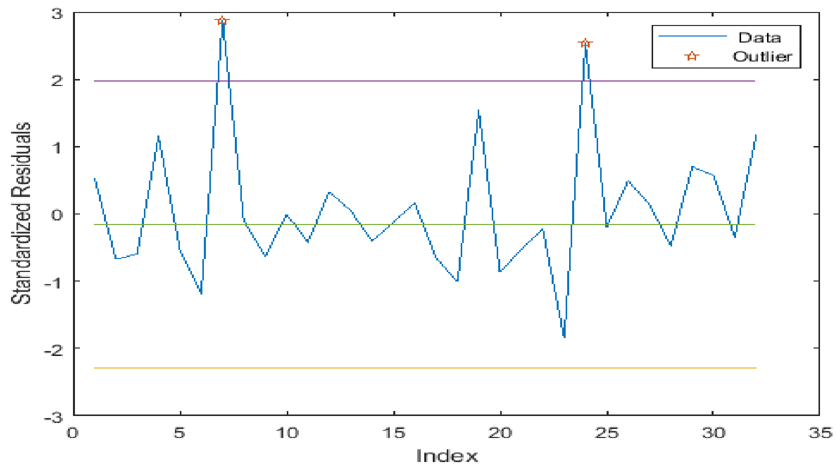


Figure 2. Standadirazed Residual plot

Table 7: RMSE values for different estimation methods

Method Estimation	RMSE
Bayesian	205.3620
Wavelet Bayesian -bior1.1	59.8052
Wavelet Bayesian -db7	72.2054

From Table (7) that the proposed method (Bayesian wavelet- bior1.1) method shows root mean squares of error (RMSE) less than the method (Bayesian).

4: Counclusion

1- The proposed method(wavelet Bayesian) was better than the Bayesian method for multiplelinear regression based on RMSE.
2- The proposed method reduced noise while also solving the problem of data contamination.
3- Using real data, the proposed method (wavelet Bayesian bior1.1) outperforms the (wavelet Bayesian db7) based on RMSE.

5: Recommendations

1- Estimation of nonlinear models using the same technique.
2- A comparison of several types of thresholding with wavelets in Bayesian multiple regression estimation
3- Applying the proposed method tominimize the problem of data contamination and estimate the Bayesian model for multiple regression.

References

1- Abramovich F., Bailey T. C and Sapatinas T. (2000) "Wavelet analysis and its statistical applications", The Statistician, Vol.49 (1), pp.I-29.

2- Alex R.S., Nancy L., Brani V., (2021)"Bayesian wavelet shrinkage with beta priors". Computational Statistics36, 1341–1363.

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

4- Barnett, V. and Lewis, T. (1994). Outliers in Statistical Data. John Wiley.

5- Daubechies I., (1992), Ten Lectures on Wavelet, Philadelphia: SIAM, pp.17-52, pp.53-106.

6- Daubechies I., (1988), Orthonormal bases of Company Support Wavelets, Comm. Applied Mathematics, 41: 909-996.

7- 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.

8- Donoho, D. L. and Johnstone, I. M., (1994b) Ideal spatial adaptation by wavelet shrinkage. Biometrika, 81, 425–455.

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

10- 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.

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

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

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

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

15- 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.

16- Harrison, P. J. and Mike.W. (1989): "Bayesian Forecasting and Dynamic Models" springer verlag, New York.

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

18- J.-C. Feauveau A. Cohen, Ingrid Daubechies. Biorthogonal bases of compactly supported wavelets. Communications on Pure and Applied Mathematics, 45(5):485–560, 1992.

19- kanamori, T. &fujisawa, H., (2015) "Robust estimation under heavy contamination using un-normalized models"Biometrika (2015), 102, 3, pp. 559–572.

20- Morris J. M .and Peravali R. (1999) "Minimum-bandwidth discrete-time wavelets", Signal Processing, Vo1.76 (2), pp.181-193.

21- NorbertR. (2012)"contributiontoBayesianwaveletshrinkage"In Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the School of Industrial and Systems Engineering. Georgia Institute of Technology.

22- OHagan, A. and Pericchi, L. (2012). “Bayesian heavy-tailed models and conflict resolution”: a review, Brazilian Journal of Probability and Statistics, 26, 4, 372–401.

23- Philippe G, Alain D. and Myl`ene B. (2020) “A New Bayesian Approach to Robustness Against Outliers in Linear Regression”International Society for Bayesian Analysis (2020).pp. 389–414.

24- Rousseeuw, P. J., & Leroy, A. M. (1987). Robust regression and outlier detection (Vol. 1): Wiley Online Library.

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

26- 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.

27- Zellner, A. (1971a): "An Introduction to Bayesian Inference in Econometrics", John Wiley and Sons, Inc.

28- Zellner, A. (1971b) "Bayesian and Non-Bayesian Analysis of Log-normal Distribution and Log-Normal Regression", Jasa, Vol.66, No.334, P327-330.

Appendix

```
clc
clear all
o=1;
while o<=1000
n=256;
k=2;
xe=randn(n,k);
xa1=xe(:,1);
xb2=xe(:,2);
xii=[ones(size(xa1)) xa1 xb2 ]
ri1=randn(n,1)
rri=ri1*5
pdf1 = makedist('tLocationScale','mu',30,'sigma',1,'nu',3)
noises1 = random(pdf1,n,1);
pdf2 = makedist('tLocationScale','mu',0,'sigma',8,'nu',4)
pdf3 = makedist('tLocationScale','mu',22,'sigma',1,'nu',4)
noises3 = random(pdf3,n,1);
rr2=noises3(1:n/4)
s=size(rr2);z=size(rri)% (find size of both)
rd1=[rr2;zeros(z(1)-s(1),1)] % concatenate the smaller with zeros
rei1=[rri+rd1]
yy=xii*[2;-2;0.5]+rei1
beta1=inv(xii'*xii)*(xii'*yy)
yyhad1=xii*beta1
eei=yy-yyhad1
p=1:n;
[TF,L,U,C] = isoutlier(eei);
plot(p,eei,p(TF),eei(TF),'p',p,L*ones(1,n),p,U*ones(1,n),p,C*ones(1,n))
legend('Original Data','Outlier','LowerThreshold','UpperThreshold','Center Value')
clear WST
[ca,cd] = dwt(yy,'db7',n);
%estimate of delta(level of thresholding)for haar wavelet filter
MAD=median(abs(cd));
sigmaMAD=MAD/0.6745;
D=sigmaMAD*((2*log(n))^0.5);
%soft thresholding for haar wavelet filter
W=[ca;cd];
for i=1:n
if W(i)>0
signW(i)=1;
else if W(i)==0
signW(i)=0;
else
signW(i)=-1;
end
```

```
end
end
signW';
for i=1:n
plus(i)=abs(W(i))-D;
if plus(i)<0
plus(i)=0;
else
plus(i)=plus(i);
end
end
plus';
for i=1:n
WST(i)=signW(i)*plus(i);
end
 WST=WST';
cd=WST(n/2+1:n);
  yh1 = idwt(ca,cd,'db7',n)
wavedl=fitlm(xe,yh1)
robwav=fitlm(xe,yh1,'RobustOpts','cauchy')
ols=fitlm(xe,yy)
rob=fitlm(xe,yy,'RobustOpts','cauchy')
   %baysian method
n1=(1:(n/2));
n2=((n/2+1):n);
y1=yy(1:n/2);
y2=yy((n/2+1):n);
x1 = xii(1:n/2, :);
x2 = xii((n/2+1):n, :);
%('Informative Prior Information');
n1=length(x1);
k1= 2 %No. of Prior variables;
v=(n1-k1-1);
bI=inv(x1'*x1)*(x1'*y1);
SSEI=(y1'*y1)-(bI'*x1'*y1);
MSEI=SSEI/(n1-k1-1);

%disp('Bayesian Method by using Non-Informative Prior');
n2=length(x2);
k2= 2;
v1=(n2-k2-1);
bn=inv(x2'*x2)*(x2'*y2);
SSEn=(y2'*y2)-bn'*x2'*y2;
MSEn=SSEn/(n2-k2-1);
sigma_bni =sqrt((v1*MSEn)/(v1+2));
%disp('Bayesian Method by using Informative Prior');
```

```
f=n2+v;
A=(x1'*x1);
b_Bayse=inv(x2'*x2+A)*(x2'*y2+A*bI)
sst_B=(y2'*y2)-(n2*mean(y2)^2);
sse_B=(y2'*y2)-(b_Bayse'*x2'*y2);
f_S2B=SSEI*SSEn;
Sigma2_Bayse=(f_S2B)/(f+2);
Sigma_Bayse=sqrt((f_S2B)/(f+2));
ssr_B=sst_B-sse_B;
R2B=ssr_B/sst_B;
% wavelet baysian method
n1=(1:n/2);
n2=((n/2+1):n);
y1w=yh1(1:n/2);
y2w=yh1((n/2+1):n);
x1 = xii(1:n/2, :);
x2 = xii((n/2+1):n, :) ;
%('Informative Prior Information');
n1=length(x1);
k1= 2 %No. of Prior variables;
v=(n1-k1-1);
bIw=inv(x1'*x1)*(x1'*y1w);
SSEIw=(y1w'*y1w)-(bIw'*x1'*y1w);
MSEIw=SSEIw/(n1-k1-1);

%disp('Bayesian Method by using Non-Informative Prior');
n2=length(x2);
k2= 2;
v1=(n2-k2-1);
bnw=inv(x2'*x2)*(x2'*y2w);
SSEnw=(y2w'*y2w)-(bnw'*x2'*y2w);
MSEnw=SSEnw/(n2-k2-1);
sigma_bniw =sqrt((v1*MSEnw)/(v1+2));

%disp('Bayesian Method by using Informative Prior');
f=n2+v;
A=(x1'*x1);
b_Baysew=inv(x2'*x2+A)*((x2'*y2w+A*bIw))
sst_Bw=(y2w'*y2w)-(n2*(mean(y2w))^2);
sse_Bw=(y2w'*y2w)-(b_Baysew'*x2'*y2w);
f_S2Bw=SSEIw*SSEnw;
Sigma2_Baysew=(f_S2Bw)/(f+2);
Sigma_Baysew=sqrt((f_S2Bw)/(f+2));
ssr_Bw=sst_Bw-sse_Bw;
R2Bw=(ssr_Bw)/(sst_Bw);
mserobwve(o)=robwav.RMSE;
```

```
msewave(o)=wavedl.RMSE;
mserob(o)=rob.RMSE;
mseols(o)=ols.RMSE;
mse_bz(o)=sqrt(sse_B /(n-k-1));
mse__bzwv(o)=sqrt(sse_Bw /(n-k-1));
o=o+1;
end
Mmse_robwv=mean(mserobwve)
Mmse_wv=mean(msewave)
Mmse_rob=mean(mserob)
Mmse_ols=mean(mseols)
Mmse_bays=mean(mse_bz)
Mmse_bayswv=mean(mse__bzwv)
```

(Demographic Data)

Obs.	X1	X2	X3	X4	X5	X6	Y
1	19.5	860	1	21	98.5	856	1316
2	37.5	695	84	1720	98.5	546	670
3	60.4	3000	548	7121	91.1	24	200
4	35.4	819	301	5257	96.7	536	1196
5	67.1	3900	3	192	74	27	235
6	45.1	740	72	1380	85	456	365
7	27.3	900	2	257	97.5	645	1947
8	127.9	1700	11	1164	80.1	257	379
9	78.9	2600	24	948	79.4	326	357
10	29.9	1400	62	1042	60.5	78	467
11	31	620	108	1821	97.5	398	680
12	23.7	830	107	1434	98.5	570	1057
13	76.3	5400	127	1497	39.4	89	219
14	21	1600	13	1512	98.5	529	794
15	27.4	1014	83	1288	96.4	667	943
16	91.9	6400	36	1365	29.4	135	189
17	47.6	650	108	1370	97.5	258	490
18	22.4	840	2	79	98.5	445	572
19	225	5200	138	2279	19.3	220	73
20	30.5	100	40	598	98.5	362	550
21	48.7	746	164	2323	87.5	362	516
22	58.7	4300	143	3410	77	42	316
23	37.7	930	254	7563	98	750	306
24	31.5	910	123	2286	96.5	36	1388
25	68.9	6400	54	2980	38.4	475	356
26	38.3	980	1041	8050	57.6	142	377
27	69.5	4500	352	4711	51.8	14	225
28	77.7	1700	18	296	50	258	262

29	16.5	900	346	24855	98.5	923	836
30	22.8	700	9	170	98.5	839	1310
31	71.7	2800	10	824	38.4	110	160
32	20.2	946	11	3420	98.5	258	1130