A proposed hybrid method for Multivariate Linear Regression Model and Multivariate Wavelets (Simulation study)

Amira Wali Omer[1] *Asst. Prof. Dr. Bekhal Samad Sedeeq*[2] *Prof. Dr. Taha Hussein Ali*[3]

[1-2-3] Department of Statistics and Informatics, College of Administration and Economics
Salahaddin University-Erbil, Iraq

Abstract: In this paper, a hybrid method was proposed for multivariate linear regression model and multivariate wavelets. Including the wavelet transform for the dependent variables through the multivariate Daubechies and Fejer-Korovkin wavelets and using the minimax and universal methods with the soft threshold rule to data de-noise when estimating model parameters. Then the comparison between the proposed hybrid and classical method (Ordinary Least Squares), combining simulated and actual data along with a MATLAB program written specifically for this purpose. The best possible multivariate linear regression model for the data may be obtained based on the mean squared error. The research showed that the proposed hybrid method yields more accurate parameter estimates than the traditional approach.

Keywords: *Multivariate Wavelet, multivariate linear regression model, De-noise, and threshold.*

Introduction

The linear regression model is a linear relationship between a single normal distributed response variable with explanatory and more variables. From this relationship, one can explain the changes in response due to changes in explanatory variables (Bartlett, 1937). Linear regression is the most statistical model used in practical applications because these models are linearly dependent on their unknown parameters (Davidson, 1972). This can be fitted much more quickly than the other models in which responses have a non-linear relationship with their unknown parameters, and because the properties of statistical estimators are easier to explain, the estimation of parameters in linear regression is so essential if they are not adequate (they have not BLUE properties) then the model leads to meaningless results (Mallows, 1973).

Multivariate multiple linear regression has the potential to be a very powerful tool in many fields of work and research. This method is used when we have a problem consisting of two or more predictor variables and two or more response variables. Multivariate multiple linear regression is an extremely valuable concept and is applied practically in fields such as business, economics, politics, and medical research. The multivariate multiple regression model is an extension of the standard multiple linear regression model. Multiple linear regression concerns predicting or explaining values of one response variable based on values of a collection of two or more predictor variables. Multivariate multiple linear regression can be used to make predictions about how much a person can lift in a certain powerlifting event depending on his or her body weight and age. (Izenman, 2013)(Quick, 2013)

Wavelet regression is a technique for reducing noise in a sampled function that has been contaminated by noise. The wavelet decomposition coefficients, which mainly represent noise, are thresholded to achieve this. The wavelet theory is a current and important theory that has a wide range of applications in both theoretical and applied disciplines. wavelets are a good tool for the approximation of high dimensional functions, which feature dominant directions of the periodicity. One-dimensional shift invariant spaces and tensor-product wavelets are generalized to multivariate shift invariant spaces on non-tensor-product systems. Wavelet shrinkage estimation, based on a thresholding parameter, has lately become a powerful mathematical technique for de-noising function estimates, and the choice of this threshold dictates, to a large extent, the efficiency and success of de-noising. Signal structure will be lost in general if the threshold setting is set too high. If it is set too low, though, noise will be included in the estimate. In regression modeling, identifying outliers or contaminated observations is a critical step. (Ali, et al, 2023) (Ali and Qadir, 2022).

2. Multivariate Regression Analysis:

When there are several independent variables in a multivariate regression, multivariate multiple regression can be used to estimate a single regression model with multiple dependent variables (Rencher, 2002).

2.1 The Multivariate Linear Model:

Here, we will talk about the multivariate multiple regression model, where the term "multivariate" refers to a group of dependent variables and "multiple" to a group of independent variables. In this situation, several y values are obtained for each x value. All of $x_1, x_2, \ldots, x_q$ should be able to forecast any and all of $y_1, y_2, \ldots, y_p$ (Dawkins, 1989).

There is a matrix in which the n values of the observed y-vector can be listed as rows:

$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{pmatrix} = \begin{pmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_n \end{pmatrix}$$

To put it another way, the values of a subject's p dependent variables are recorded in the columns of Y. Each row in Y represents one of the p independent variables, and hence each column in Y represents the y vector in the (univariate) regression model $y = XB + \Xi$ (Fan, et al. 2017). A matrix can be constructed from the n values of $x_1, x_2, \ldots, x_q$ hat coincides with the X matrix used in the multiple regression formulation and as follows:

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1q} \\ 1 & x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nq} \end{pmatrix}$$

We suppose that X remains constant from one sample to the next (Eyvazian et al. 2011) . Each of the p y's will have its own unique relationship to the x's, necessitating a unique set of for each column in Y (Muller and Peterson, 1984). In this way, we get a matrix $B = (\beta_1, \beta_2, \ldots, \beta_p)$ where each column of Y corresponds to a column of. Therefore, our multivariate model

$$Y = XB + \Xi \qquad \text{............ (1)}$$

where Y is am matrix $n \times P$, X is a matrix $\{n(q+1)\}$ and B is a matrix $\{(q+1)P\}$. The matrix of random error (Ξ) by dimension $n \times P$ (Hamed and Amir 2023).

Using the parameters p = 2, q = 3, we can describe the multivariate model as follows (Noorossana et al. 2010b) :

$$\begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ \vdots & \vdots \\ y_{n1} & y_{n2} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} \\ \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} \\ \varepsilon_{21} & \varepsilon_{22} \\ \vdots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} \end{pmatrix} \qquad \text{............ (2)}$$

The model for the first column of Y is:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} \end{pmatrix} \begin{pmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{21} \\ \beta_{31} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{n1} \end{pmatrix} \quad \text{............ (3)}$$

and for the second column, we have:

$$\begin{pmatrix} y_{12} \\ y_{22} \\ \vdots \\ y_{n2} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} \end{pmatrix} \begin{pmatrix} \beta_{02} \\ \beta_{12} \\ \beta_{22} \\ \beta_{32} \end{pmatrix} + \begin{pmatrix} \varepsilon_{12} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{n2} \end{pmatrix} \quad \text{............ (4)}$$

Additional assumptions that lead to good estimates are as follows :

I. $E(Y) = XB \quad or \quad E(\Xi) = 0$

II. $cov(y_i) = \sum$ for all i = 1, 2, . . . , n, where (y_i') is the i th row of Y.

III. $cov(y_i, y_j) = 0$ for all i ≠ j . (Rencher, 2002) .

First, let's assume that the linear model is accurate and that we don't need any more x-coordinates in order to forecast the y-coordinates. Second, assume that the covariance matrix of Y is uniform over all n observation vectors (rows). The third premise is that the rows of Y that make up the observation vectors are independent of one another. Because of this, we presume that the y values inside a given observation vector (row of Y) are associated with one another but uncorrelated with y values in other observation vectors (Jensen et al.2006).

The variances and covariances of $y_{i1}, y_{i2}, \dots, y_{iP}$ in any y_i are found in the covariance matrix in assumption 2:

$$\text{cov}(y_i) = \sum = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} \quad \text{...... (5)}$$

The covariance matrix $cov(y_i, y_j) = 0$ in assumption 3 contains the covariances of each of $y_{i1}, y_{i2}, \dots, y_{iP}$ with each of $y_{j1}, y_{j2}, \dots, y_{jP}$ (Rao, 1966):

$$cov(y_i, y_j) = \begin{pmatrix} \text{cov}(y_{i1}, y_{i1}) & \text{cov}(y_{i1}, y_{i2}) & \cdots & \text{cov}(y_{i1}, y_{ip}) \\ \text{cov}(y_{i2}, y_{i1}) & \text{cov}(y_{i2}, y_{i2}) & \cdots & \text{cov}(y_{i2}, y_{ip}) \\ \vdots & \vdots & \vdots & \vdots \\ \text{cov}(y_{ip}, y_{i1}) & \text{cov}(y_{ip}, y_{i2}) & \cdots & \text{cov}(y_{ip}, y_{ip}) \end{pmatrix} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

2.2 Least Squares Estimation in MMLR:

By analogy with the univariate case estimate β with:

$$\hat{\beta} = (X'X)^{-1} X'Y \quad \text{............ (6)}$$

Call $\hat{\beta}$ the least squares estimator for β because it "minimizes" $\varepsilon = \hat{\varepsilon}'\hat{\varepsilon}$, a matrix analogous to SSE:

$$\hat{\varepsilon}'\hat{\varepsilon} = (Y - X\hat{\beta})' (Y - X\hat{\beta}) \quad \text{............ (7)}$$

In the following meaning, the matrix $\hat{\beta}$ minimizes $\hat{\varepsilon}'\hat{\varepsilon}$ (Eyvazian et al. 2011) . Adding $X\hat{\beta} - X\beta_0$ to $Y - X$ yields a positive definite matrix, as $\hat{\varepsilon}'\hat{\varepsilon} = (Y - X\hat{\beta})'(Y - X\hat{\beta})$ is an approximation of $Y - X\hat{\beta}$, and β_0 is an estimate that may be better than $\hat{\beta}$ and add $X\hat{\beta} - X\beta_0$. (Rencher, 2002). Thus we cannot improve on $\hat{\beta}$. The least squares estimate $\hat{\beta}$ also minimizes the scalar quantities tr$(Y - X\hat{\beta})'(Y - X\hat{\beta})$ and $(Y - X\hat{\beta})'(Y - X\hat{\beta})$. This means that the:

$\mathrm{tr}(Y - X\hat{\beta})'(Y - X\hat{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{p}\hat{\boldsymbol{\varepsilon}}_{ij}^2$ (Amiri et al. 2014).

Earlier, we mentioned that in the model $Y = X\beta + \Xi$, each column of Y has a corresponding column of β; that is, each $y_{i,j} = 1, 2, \ldots, P$ is predicted differently by $x_1, x_2, \ldots, x_q$. Similarly, we observe a trend in the estimate $\beta = (X'X)^{-1}X'Y$. Each column of Y is multiplied by the matrix product $(X'X)^{-1}X'$ as such, the least squares estimate for the jth dependent variable, y_j, can be found in the jth column of β (Gnanadesikan, 1997). To further describe this, let's use the notation $y_{(1)}, y_{(2)}, \ldots, y_{(p)}$ for the p columns of Y and y'_i for the n rows, where I is an integer between 1 and n. Then :

$$\hat{\beta} = (X'X)^{-1}X'Y$$
$$= (X'X)^{-1}X'\left(y_{(1)}, y_{(2)}, \ldots\ldots\ldots, y_{(p)}\right)$$

$$= [(X'X)^{-1}X'y_{(1)}, (X'X)^{-1}X'y_{(2)}, (X'X)^{-1}X'y_{(p)}]$$

$$= [\hat{\beta}_{(1)}, \hat{\beta}_{(2)}, \ldots, \hat{\beta}_{(p)}] \qquad \ldots\ldots\ldots\ldots (8)$$

1) *3. Wavelet Shrinkage:*

The nonparametric function estimating method known as wavelet shrinkage (WaveShrink) is relatively new, yet it has been demonstrated to possess asymptotic near-optimality properties for a broad class of functions (Ali and Saleh 2022). WaveShrink, as originally conceived by Donoho and Johnstone, presupposes that the data are evenly spaced (Ali and Qadir, 2022).

3.1 Wavelet:

One definition of a wavelet is an oscillation whose amplitude starts at zero, grows or shrinks, and then cycles back to zero again and again. Brief oscillations are what scientists refer to as wavelets (Ali, et al. 2022). Based on the quantity and polarity of its pulses, a taxonomy of wavelets has been developed. Namely, wavelets have characteristics that make them applicable in signal processing. In any discretized wavelet transform, the number of wavelet coefficients for a given upper half-plane rectangular region is finite. Even so, an integral must be evaluated for each coefficient (Ali and Ali, 2019). In exceptional cases, this numerical complexity can be sidestepped by doing a multiresolution analysis using scaled and shifted wavelets. Starting with the scaling function ∅ (also called further wavelet) and the mother wavelet Ψ, we may derive the auxiliary function $f \in L^2R$ (the space of square integrable real functions). Dilations and translations of and produce other wavelets. in L^2R, and that an is a positive integer. The values a = 2 and b = 1 are frequently used as an example. The Daubechies 4-tap wavelet is the most well-known parent-child wavelet pair. One should keep in mind that not all orthonormal discrete wavelet bases can be linked to a multiresolution study (Ali and Saleh 2022). The child wavelets are created by combining the parent wavelets.

$$V_j = \mathrm{span}\left(\emptyset_{j,k}: n \in Z\right), \text{where } \emptyset_{j,k}(y) = 2^{\frac{j}{2}}\emptyset\left(2^j y - n\right) \quad \ldots\ldots (9)$$

$$W_j = \mathrm{span}\left(\Psi_{j,k}: n \in Z\right), \text{where } \Psi_{j,k}(y) = 2^{\frac{j}{2}}\Psi\left(2^j y - n\right) \quad \ldots\ldots (10)$$

The time-domain characteristics of the father wavelet V m are maintained, while the frequency-domain characteristics are maintained by the mother wavelets W_j. In light of them, it's imperative that the sequence $\{0\} \subset \cdots \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \cdots \subset L^2(R)$

Because the discrete wavelet transform (DWT) is such a helpful observation processing technique, it is used in many domains, including science, engineering, mathematics, and computer science. DWT decomposes one observation into many resolutions by using scaled and shifted versions of a compact supported basis function (mother wavelet) (Ali and Ali, 2019). If we have an observation vector y with 2m elements, and we know that the DWT of y is given by (11).

$$W = wy \quad \text{........ (11)}$$

W is a vector of size $(n \times 1)$ that contains the scaling and wavelet coefficients, where w is a $(n \times n)$ wavelet matrix (Alzubaydi and Mustafa, 2016). Each of the wavelet coefficients can be represented by a (j+1) vector. $W=[W_1,W_2,W_3,\ldots,W_j,V_{j0}]^T$. Each Discrete Wavelet Transformation (DWT) employs the same wavelet to band the approximation coefficients, such that information about previous decompositions is added to the next one, as shown in the following formula.

$$y = Ww^T = \sum_{j=1}^{j_0} W_j^T W_j + V_{j_0}^T V_{j_0} \quad \text{........ (12)}$$

The inverse DWT can be used to rebuild the observations from the de-noise data (reduction of the contamination) at each j-level.

3.1.1 Daubechies Wavelet (db):

The so-called normal orthogonal wavelets had their origins in 1988, the same year that made possible the use of discrete wavelet analysis, and are named after Ingrid Daubechies, a pioneer in the study of wavelength (Mustafa and Alzubaydi 2013). For instance, 4D and db represent the initials of the researcher (daubechies) and the number of vanishing or ephemeral moments of the wavelet function, respectively; N is the length of the candidate or rank; and (L_1) is the number of ephemeral moments of the wavelet function (db2). In the following relations, (L_1) corresponds to the same person as N, who is the second-ranked person in this family (Ali and Qadir, 2022).

$$L_1 = \frac{N}{2} \quad \text{....... (13)}$$

To generalize, we may say that (dbn) is the family of small waves of order n (recall that the small wave har is a member of this family; it shares the following features with the other members of the family (Ali and Saleh 2022).

- The anchor of the small wave (dbn) is on the period [0, 2n-1].
- A small wave (dbn) has n ephemeral moments, ie:

$$\left|\frac{d^j}{dX^j}\psi(X)\right| < \infty \quad \Rightarrow \quad \int X^j \psi(X) dX = 0\,,\ \ 1 \le j \le n \quad \text{....... (14)}$$

- Anomalies increase with rank, (dbn) have (rn) continuous derivatives (r is about 0.2) (Ali and Ali, 2019). Types of Daubechies wavelet

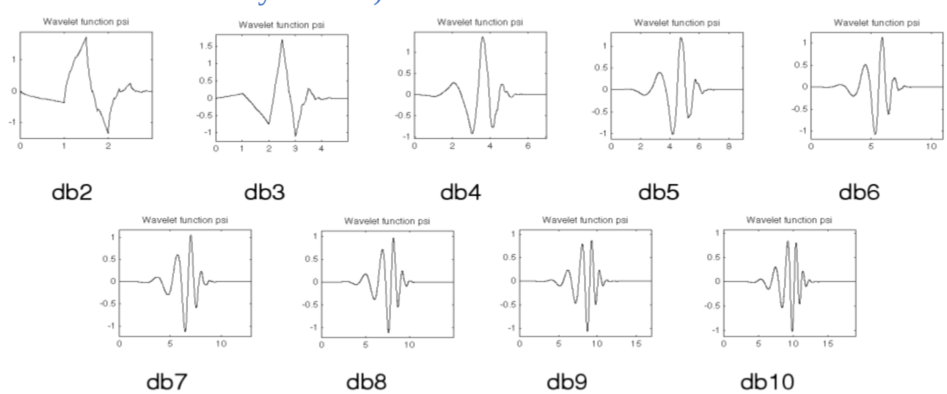

Figure 1: Daubechies Wavelet

II. 3.1.2 Fejer-Korovkin Wavelet (FK):

In discrete (decimated and undecimated) wavelet packet transforms, filters built to minimize the difference between a valid scaling filter and the ideal sinc lowpass filter; filters having N coefficients, are very important. We use the well-known Fejer _Korovkin kernels from approximation theory to build a series of filters with optimal resolution. Name of the Fejér-Korovkin filter to use for scaling L_0= fejerkorovkin(wname). Each name includes a number that corresponds to the number of Fejér-Korovkin filter coefficients. wname specifies the Fejér-Korovkin filter to be returned. fk4, fk6, fk8, fk14, fk18, and fk22 are all acceptable values for wname, where N might be any of 4, 6, 8, 14, 18, or 22. Both the Fejér-Korovkin filters and the L0 scaling filter are returned in a vector format. (Nielson, 2001).

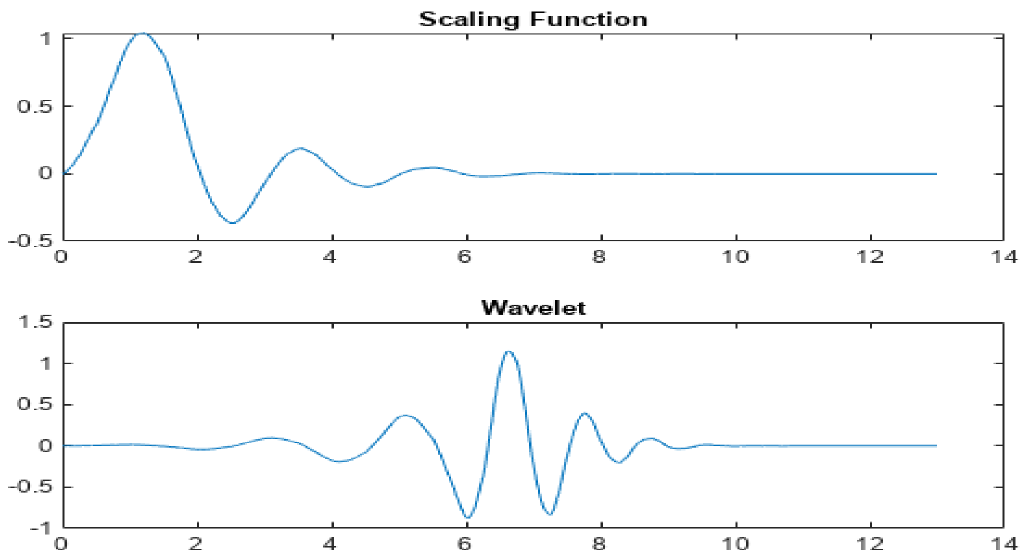

Figure 2: Fejer-Korovkin Wavelet

3.2 Thresholding

To perform thresholding, a signal's value is compared to a predefined threshold and, if it is greater than the threshold, the value is set to zero or equal to the threshold (Ali, et al., 2023). Thresholding is frequently employed in image processing for the purpose of image segmentation. Several thresholding mechanisms are employed for various image types (Ahmadi et al. 2015).The simplest form of non-linear wavelet de-noising is known as thresholding, and it works by splitting the wavelet coefficient into two groups: the signal group and the noise group. Many approaches exist for selecting a threshold value, and different rules govern how the thresholds of the wavelet coefficients are applied (Ali and Ali, 2019).

3.2.1 Threshold Estimation:

Wavelet de-noising relies on measuring the noise level of a signal in order to recover the optimum components of the signal. The amount of background noise can be estimated in several different ways. In this analysis, we compare the performance of four distinct threshold values. Using Stein's unbiased risk estimation criterion, the process of choosing an adoption threshold is called rigresure (Ali and Qadir, 2022).

$$\text{Value} = \sigma\sqrt{2\log(N\log_2 N)} \qquad \ldots\ldots\ldots (15)$$

Where N is the signal length and is the noise standard deviation.

We look at a unique type of nonlinear estimator called a truncated threshold wavelet estimator. As in (2), define the empirical coefficients $\widehat{\alpha}_{jk}, \widehat{\beta}_{jk}$ and use hard thresholding:

$$\hat{\beta}_{jk} = \begin{cases} \hat{\beta}_{jk} & \text{if } |\hat{\beta}_{jk}| > KC(J)n^{-1/2} \\ 0 & \text{if } |\hat{\beta}_{jk}| \le KC(J)n^{-1/2} \end{cases}$$

……… (16)

The estimator TW is then applied to the functions $j_0(n), j_1(n), C(j)$, and K.

$$TW(x) = \hat{f}_{n,j_1} + \hat{D}_{j_1,j_0} = \sum_{k\in Z} \hat{\alpha}_{j_1 k}\ \phi_{j_1 k}(x) + \sum_{j_1}^{j_0} \sum_{k\in Z} \hat{\beta}_{jk}\ \psi_{jk}(x) \quad ……… \ (17)$$

Analysis of the error's bias and variance could shed light on this. When using level j(n), the bias of LW is of order $2^{-j(n)\acute{\sigma}\acute{p}}$, while the stochastic term is of order $(2^{j(n)}/n)^{\acute{p}/2}$) (Mustafa and Alzubaydi 2013) As a result, we can reason that it is best to begin with a low-frequency estimator $LW(j_1(n))$, with $j_1(n)$set low enough that the stochastic term has the correct rate, and then to add in specific "details" up to the higher order $j_0(n)$in such a manner that the bias term also has the correct order. (It is evident that choosing $j_0 < j_1$is sufficient if p' = p, but choosing $j_0 > j_1$is necessary if p' > p.

3.3 Shrinkage:

is the lessening of the influence of sampling error. In the last few decades, the practice of reducing sample sizes has gained favor. These days, it's not uncommon to hear about or even practice some form of statistical shrinking (Ali, et al., 2023). Shrinkage is the most misunderstood statistical catchall term. Since it goes by a variety of aliases, this is why. Consider the widely-used RiskMetrics volatility estimator (Ali, et al. 2022).

3.3.1 Universal Method:

Donoho and Johnstone (1994) submitted the universal threshold method, which is given by formula (17).

$$\delta^{U} = \hat{\sigma}_{MAD} \sqrt{2\ \log(n)} \quad ….. \ (18)$$

Standard error of the estimate (σ^_MAD) of the coefficients of interest is equal to MAD 0.6745. The wavelet coefficients at the smallest scale can be characterized by their median absolute deviation (MAD).

3.3.2 Minimax Method:

As a refinement of the universal threshold technique, Donoho and Johnstone (1994) proposed the optimal minimax threshold approach. An estimator that satisfies the minimax risk assumption underpins the minimax approach, as in:

$$\tilde{R}(F) = \inf_{\tilde{f}} \sup_{f\in\tilde{R}(F)} R(\tilde{f}, f) \quad ….. \ (19)$$

Where

$$R(\tilde{f}, f) = \frac{1}{n}\sum_{i=1}^{n} E[\tilde{f}, f]^2 \quad ….. \ (20)$$

Where $f = f(x_i)$ and $\tilde{f} = \tilde{f}(x_i)$, Indicate the true and estimated sample values as vectors. When compared to its universal counterpart, the minimax threshold estimator emphasizes overall mean square error (MSE) reduction while avoiding over-smoothing of estimates.

4. Application part:

The simulation study was conducted by first simulating the multivariate regression model, and then applying it to real data based on mean square error to compare the suggested method with the conventional approach in terms of efficiency and accuracy of the estimated multivariate regression model (MSE). Also, we achieve this by developing a specialized MATLAB (version 2020a) software for this work (Appendix).

4.1 Simulation study:

To compare between the proposed and classical method in estimating the parameters of the multivariate regression model, a random error was generated for the multivariate normal distribution (see Appendix), with a dimension (*n*, *p*), where *n* = 30, 50, 100, and 200 (sample size) and *p* = 3, 4, 5 (number of dependent variables) by mean vector equal to **0** and covariance matrix Σ, such that (for *p* = 3):

$$\Sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix}$$

The matrix of independent variables **X** is by dimension (n, q), q is number of independent variables (q = 4) and the multivariate regression coefficients (intercept and slope) has $(q+1\times p)$ dimension are:

$$\beta = \begin{pmatrix} 2 & 4 & 6 \\ 4 & 3 & 5 \\ 3 & 6 & 4 \\ 3 & 5 & 2 \\ 6 & 4 & 2 \end{pmatrix}$$

The noise matrix that was added to the multivariate regression model is randomly generated with a normal distribution for zero means and a constant variance-covariance matrix equal to (3).The matrix of dependent variables Y is by dimension (n, p), and it has a correlation matrix for the first experiment (n = 30, p = 3) are:

$$R = \begin{pmatrix} 1 & 0.9050 & 0.8027 \\ 0.9050 & 1 & 0.9477 \\ 0.8027 & 0.9477 & 1 \end{pmatrix}$$

The correlation matrix indicates the presence of a correlation between the dependent variables, which fits a multivariate regression model. Figure (1) shows the scattered plot for the values of the dependent variables and their corresponding values after using the proposed method, including the Wavelet transform for observations of the dependent variables through the (Db) wavelet of the third order and de-noising of data by using the minimax method with the Soft threshold rule.

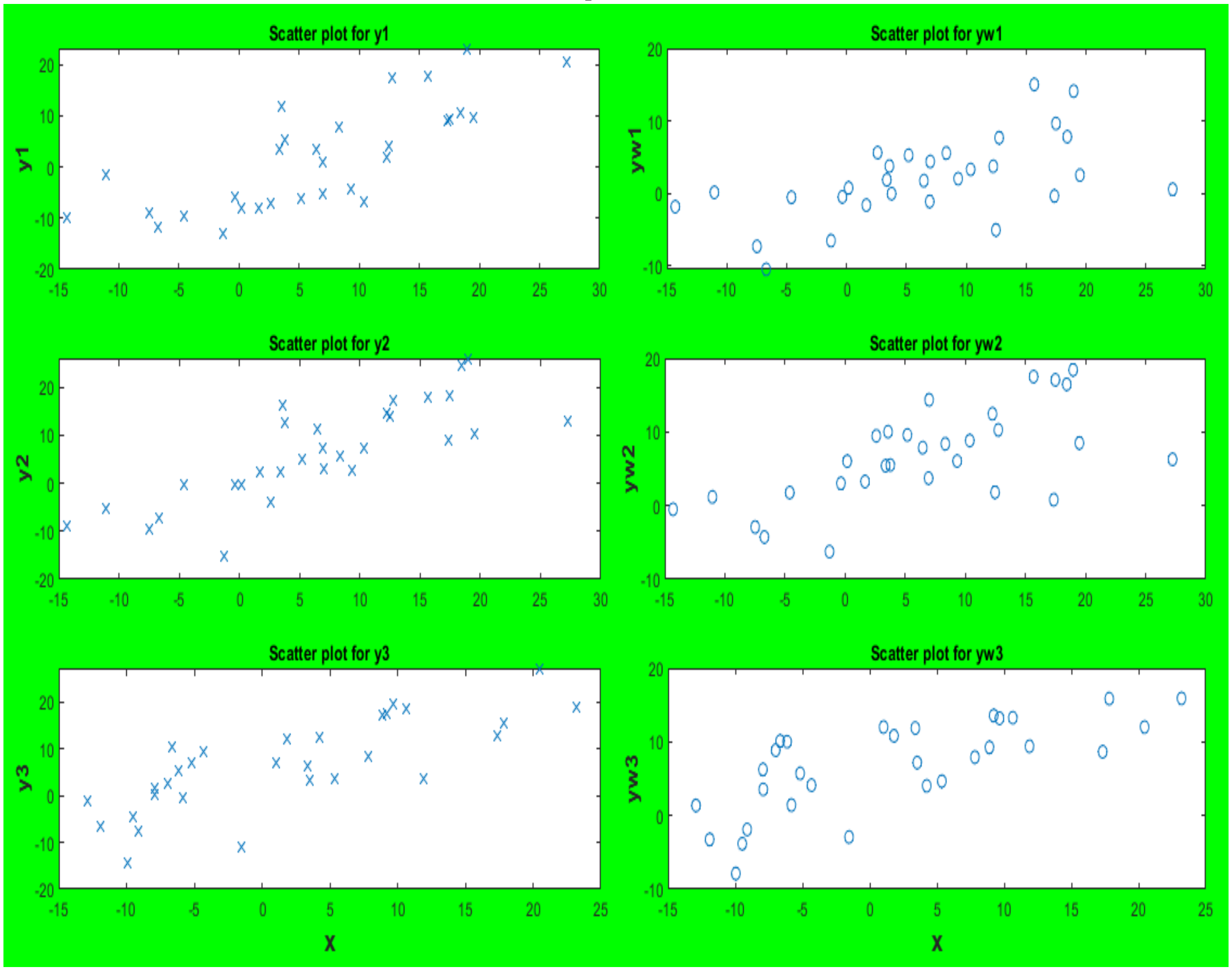

Figure 3: Scatter plot for three dependent variables and with de-noise Wavelet

Table (1) summarizes the results of the proposed and classical method for the first simulation experiment (n = 30, p = 3, and q = 4) in the analysis of the multivariate regression model.

Table (1) : Classical and Proposed MR analysis for first simulation experiment

Classical Method					Proposed Method				
β			MSE	LogL	βw			MSEW	LogLW
1.56	4.94	5.67	64.15	-246.68	1.84	5.98	6.21	49.44	-214.47
5.02	4.33	6.10			2.05	1.91	3.06		
4.89	6.70	6.35			2.57	3.87	3.89		
3.41	4.28	1.33			1.28	1.85	0.92		
7.89	5.88	4.04			3.51	3.62	3.03		

Estimated regression coefficients, returned as a column vector or matrix for the classical and proposed method and it's β and βw respectively. Residuals for the fitted regression model returned as an n-by-p matrix E. And the trace of $(E'E)/(n-q-1)$ is equal to (64.15) and (49.44) for classical and proposed methods, respectively, as mean square error (MSE), to measure the efficiency of the estimated multivariate regression model. Log-likelihood objective function value after the last iteration for classical and proposed method (Log L and LogLW), returned as a scalar value equal to (-246.68 and -214.47), respectively.

Table (2) shows the averaged MSE when the number of repeated simulations for classical and proposed methods is 1000 after estimating the parameters of the multivariate regression model. For the proposed method, two types of Wavelet were used, namely (db) of the third order and (fk) of the sixth order, with two types of threshold methods, namely (Minimax) and (Universal), and the application of the threshold rule (Soft). The proposed methods were better than the classical method for all simulation cases because the average of MSE is less than that of the classical approach. (fk6) wavelet was better than (db3) wavelet for all simulation cases in data de-noising and obtaining estimated parameters with higher accuracy. The Minimax threshold method was better than the Universal for n = 30 and 50 observations (small sample sizes), while the preference was for the Universal threshold if the sample sizes were equal to 100 and 200 observations (large sample sizes), for all cases, the number of dependent variables.

Table (2):The averaged MSE for classical and proposed methods

N	p	Classical	Proposed			
			Multivariate Wavelet			
			db3-minimax	db3-universal	fk6-minimax	fk6-universal
30	3	64.6064	47.0549	48.9953	**46.2326**	48.1947
	4	131.369	97.8040	103.273	**95.7366**	100.959
	5	251.960	153.988	155.950	**151.078**	152.734
50	3	63.9534	48.8298	51.4854	**48.4391**	51.0052
	4	131.237	101.471	106.985	**100.959**	106.447
	5	247.986	156.543	160.449	**155.039**	159.055
100	3	63.7563	34.1261	32.9092	33.1901	**32.0045**

	4	132.392	70.5684	68.0618	68.7419	**66.1460**
	5	249.939	107.937	102.333	105.161	**99.5145**
200	3	63.9568	18.4450	15.6740	18.1688	**15.5037**
	4	131.836	38.6104	32.9356	38.0655	**32.5656**
	5	249.634	57.1962	48.0793	56.6969	**47.8257**

4.2 The real data:

Assuming unique values for the intercepts and slopes in a single n by p design matrix for all dependent variables, the actual data fit an MRM from panel data. Load("flu") is being used to load the sample data. Nine different regional flu estimates based on Google® query data are included with national CDC estimates in the flu dataset array. The nine seasonal flu forecasts serve as the dependent variables in this analysis. N = 52 since weekly observations have been collected for over a year. It follows that p = 9 because the regions have 9 dimensions and the dependents have 9 dimensions. Weekly nationwide flu estimates serve as the independent variables in Figure (2).

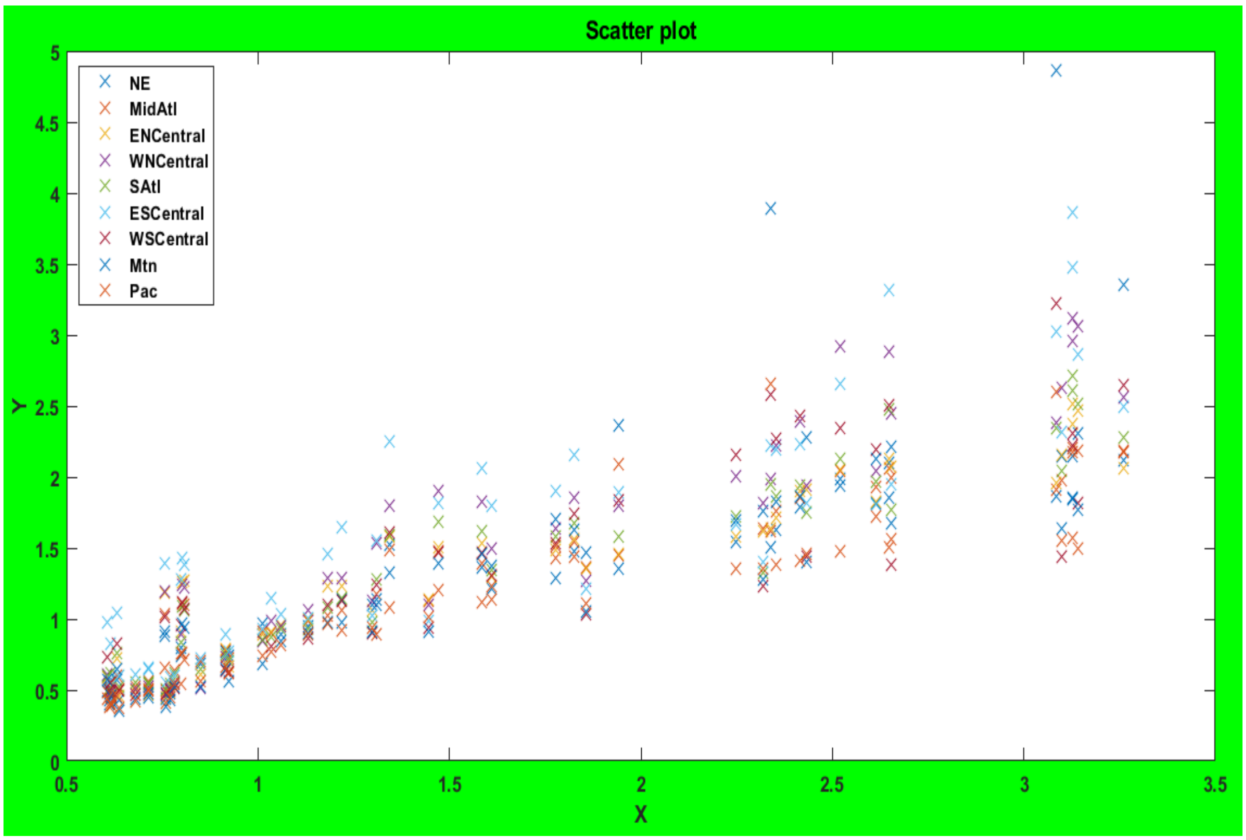

Figure 4: Scatter plot for data

Make a design matrix X of size n by p, with q equal to 1. In order to incorporate a constant term into the regression, an additional column of ones must be added. The standard MRM formula is where and, with contemporaneous correlation between regions.

Estimation of the 18 regression coefficients (29) involves determining the values of 9 intercept terms and 9 slope terms (3).

Table (3) :Classical Multivariate Regression coefficients

β_{0j}	0.1857	0.2425	0.2452	0.0681	0.1813	0.2542	0.1539	-0.0161	0.1599
β_{1j}	0.6622	0.6325	0.6574	0.8884	0.7132	0.8350	0.7265	0.8208	05765

Plotting a fit multivariate regression model with a scattering in Figure (3).

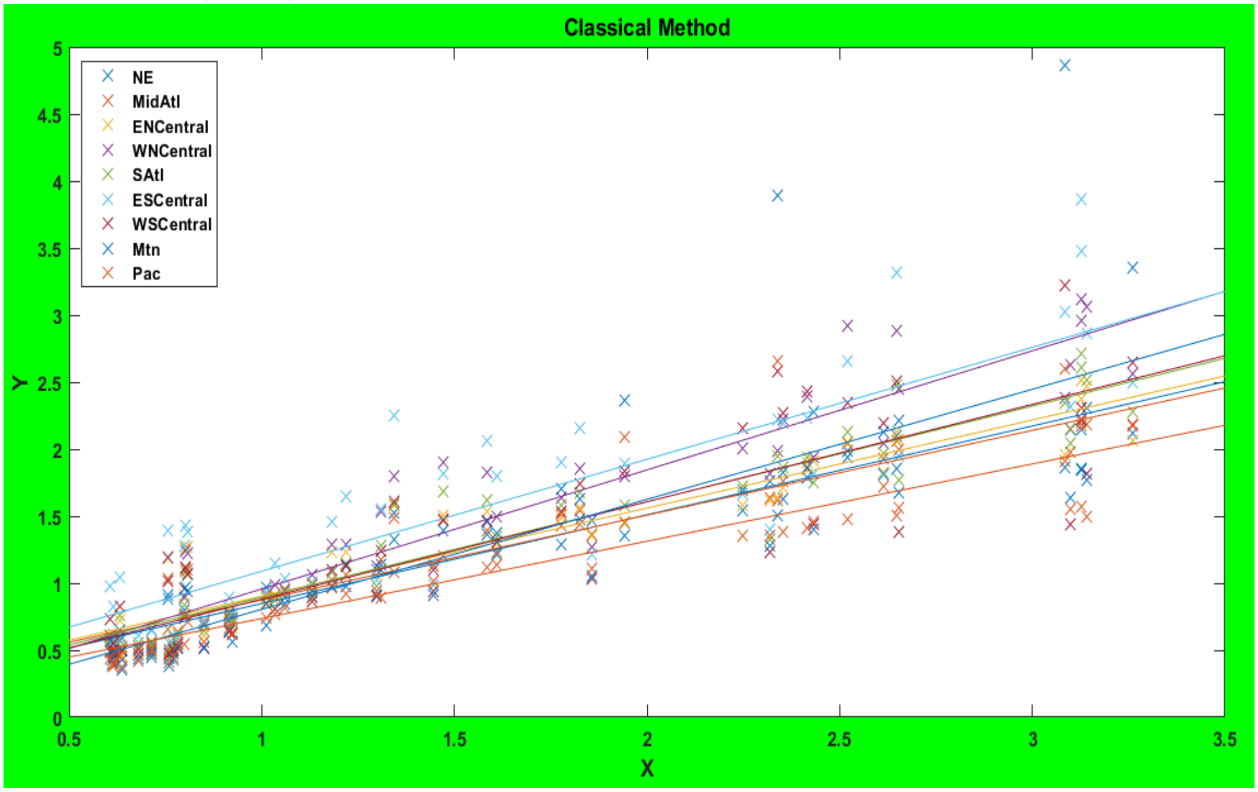

Figure 5: Classical multivariate regression model with scatter

Each regression line in Figure 3 has a unique intercept and slope. Since the last iteration, the log-likelihood objective function value has been minimized to be equal to (0.8935), and the scalar values have been minimized to be equal to (0.8935) for the multivariate regression coefficients (291.1413). The proposed method included using the Wavelet transform for observations the dependent variables through the (db) Wavelet of the second order and de-noising of data by using the universal method with the Soft threshold rule (see table 4).

Table (4) :Proposed Multivariate Regression coefficients

β_{0j}	0.1960	0.2517	0.2557	0.0795	0.1913	0.2765	0.1597	-0.0103	0.1658
β_{1j}	0.6548	0.6257	0.6498	0.8802	0.7062	0.8204	0.7224	0.8166	0.5727

Multivariate Regression coefficients have minimized the scalar quantities (MSE) equal to **(0.8234)**and its less than that of the classical method. The value of the log-likelihood objective function after the last iteration is equal to **(383.2094)** and its greater than that of the classical method, therefore the proposed method was better than the classical method. Plotting a fit multivariate regression model with scatter for proposed method in Figure (4).

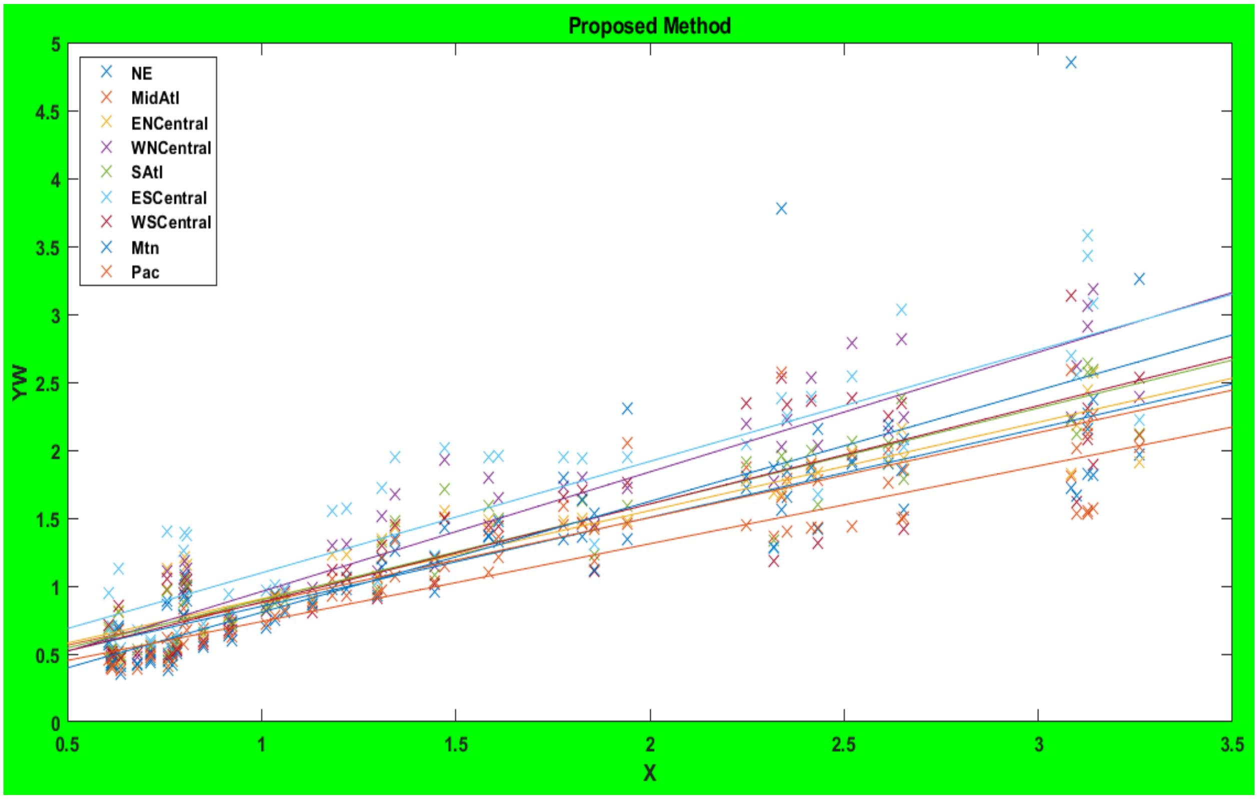

Figure 6: Proposed multivariate regression model with scatter

5. Conclusion & Recommendations:

Through the study of simulation and real data, the following main conclusions and recommendations were summarized:

5.1 Conclusions:

1. The suggested hybrid technique is more effective than OLS for estimating models in all simulated and real-world data.
2. Wavelet fk6 was the best compared to Db3 for all sample sizes.
3. Minimax method was the best compared to Universal method for small sample sizes (30 and 50) and vice versa for large sample sizes (100 and 200).
4. The proposed hybrid method addressed the problem of noise in the data.

5.2 Recommendations:

1. Using the proposed method in estimating the parameters of a multivariate linear regression model.
2. Conducting a prospective study on the use of multivariate wavelet in estimating the parameters of multivariate quantitative regression model.
3. Apply a statistical method with the help of MATLAB or R. The goal of developing this software is to have the multivariate wavelet linear regression model pick the optimal wavelet for the data entered.4. An increase in the number of dependent variables in the model leads to an increase in the MSE criterion values.

Appendix

```
clc
clear all
```

```
n=30;q=4; p =3;beta=[2 4 6;4 3 5;3 6 4;3 5 2;6 4 2];
randn('seed',1234); x=randn(n,q);
E=randn(n,p)*[1 2 3;2 1 2;3 2 1]; X=[ones(n,1) x];
noise=randn(n,p)*3;
y= X*beta+E+noise; corr(E) , corr(y); [betah sigma Eh c logl]=mvregress(X,y)
logl, Eh=y-X*betah; EhpEh=Eh'*Eh,
MSE=trace(EhpEh)/(n-q-1)
yn =
wdenoise(y,'Wavelet','db3','DenoisingMethod','minimax','ThresholdRule','soft');
[betaw sigmaw Ehw c loglw]=mvregress(X,yn),
Ehw=yn-X*betaw;
EhwpEhwd=Ehw'*Ehw;
MSED=trace(EhwpEhwd)/(n-q-1), loglw, x=y(:,3);
figure;
subplot(3,2,1),plot(x,y(:,1),'x'), ylabel('y1'),
title('Scatter plot for y1')
subplot(3,2,2),plot(x,yn(:,1),'.'), ylabel('yw1'),
title('Scatter plot for yw1')
subplot(3,2,3),plot(x,y(:,2),'x'), ylabel('y2'),
title('Scatter plot for y2')
subplot(3,2,4),plot(x,yn(:,2),'.'),
x=y(:,1);ylabel('yw2'), title('Scatter plot for yw2')
subplot(3,2,5),plot(x,y(:,3),'x'),
xlabel('x'),ylabel('y3'), title('Scatter plot for y3')
subplot(3,2,6),plot(x,yn(:,3),'.'),
ylabel('yw3'),xlabel('x'), title('Scatter plot for yw3')
```

References

1. Ahmadi, K., Javaid, A. Y. & Salari, E. (2015). An efficient compression scheme based on adaptive thresholding in wavelet domain using particle swarm optimization. *signal processing: image communication*, 32**,** 33.39-
2. Ali, Taha Hussein & Ali, Mardin Samir (2019), *Analysis of Some Linear Dynamic Systems with Bivariate Wavelets*, Iraqi Journal of Statistical Science (30), pp[31- 54]
3. Ali, Taha Hussein and Qadir, Jwana Rostam (2022), *Using Wavelet Shrinkage in the Cox Proportional Hazards Regression model (simulation study)*, Iraqi Journal of Statistical Science (35), pp[32- 54],
4. Ali, Taha Hussein, Avan Al-Saffar, and Sarbast Saeed Ismael. (2023) "Using Bayes weights to estimate parameters of a Gamma Regression model." Iraqi Journal of Statistical Sciences 20.20 (2023): 43-54.
5. Ali, Taha Hussein, Hayawi, Heyam Abd Al-Majeed, and Botani,.Delshad Shaker Ismael "Estimation of the bandwidth parameter in Nadaraya-Watson kernel non-parametric regression based on universal threshold level." Communications in Statistics-Simulation and Computation (2023): 1-34. https://doi.org/10.1080/03610918.2021.1884719
6. Ali, Taha Hussein, Mahmood, Saman Hussein and Wahdi, Awat Sirdar (2022), *Using Proposed Hybrid method for neural networks and wavelet to estimate time series model*, Tikrit Journal of Administration and Economics Sciences, Vol. 18, No. 57, Part (3): 432-448 Doi: www.doi.org/10.25130/tjaes.18.57.3.26 http://jaes.tu.edu.iq
7. Ali, Taha Hussein, Nasradeen Haj Salih Albarwari, and Diyar Lazgeen Ramadhan. (2023) "Using the hybrid proposed method for Quantile Regression and Multivariate Wavelet in estimating the linear model parameters." IRAQI JOURNAL OF STATISTICAL SCIENCES 20.20 (2023): 9-24.
8. Ali, Taha Hussien & Saleh, Dlshad Mahmood (2022), *Proposed Hybrid Method for Wavelet Shrinkage with Robust Multiple Linear Regression Model With Simulation Study*, QALAAI ZANISTSCIENTIFIC JOURNAL A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq, Vol. (7), No (1), Winter 2022 ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print), Doi:10.25212/lfu.qzj.7.1.36
9. Alzubaydi, Taha Hussein Ali and Mustafa, Qais (2016), *Reducing the orders of mixed model (ARMA) before and after the wavelet de-noising with application*, University

Journal of Human Science, Vol.20, No.6, (PP 433-442)

10. Amiri, A., A. Saghaei, M. Mohseni, and Zerehsaz, Y..(2014). *Diagnosis aids in multivariate multiple linear regression profiles monitoring.* Communications in Statistics –Theory and Methods 43 (14):3057–79. doi:10.1080/03610926.2012.694543.
11. Bartlett, M. S. (1937), "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Society of London, Ser. A*, 160, 268–282.
12. Davidson, M. L. (1983), *Multidimensional Scaling*, New York: Wiley.
13. Dawkins, B. (1989), "Multivariate Analysis of National Track Records," *American Statistician*,43, 110–115.
14. Donoho, D. L. & Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *biometrika*, 81**,** 425-455.
15. Eyvazian, M., R. Noorossana, A. Saghaei, and A. Amiri. 2011. *Phase II monitoring of multivariate multiple linear regression profiles*. Quality and Reliability Engineering International 27 (3):281–96. doi:10.1002/qre.1119.
16. Fan, S.-K S., Jen, C.-H., and Lee, T.-Y.. (2017). *Modeling and monitoring the nonlinear profile of heat treatment process data by using an approach based on a hyperbolic tangent function. Quality Engineering* 29 (2):226–43. doi:10.1080/08982112.2016.1193614.
17. Gnanadesikan, R. (1997), *Methods for Statistical Data Analysis of Multivariate Observations*,2nd ed., New York: Wiley.
18. Hamed, Sabahno & Amiri, Amirhossein (2023): Simultaneous monitoring of the mean vector and covariance matrix of multivariate multiple linear profiles with a new adaptive
19. Izenman, A. J.(2013)Multivariate Regression. In: Modern Multivariate Statistical Techniques. Springer Texts in Statistics. Springer, New York, NY.
20. Jensen, W. A., Jones-Farmer,L. A. Champ, C. W. and Woodall, W. H.. 2006. Effects of parameter estimation on control chart properties: A literature review. Journal of Quality Technology 38 (4):349–64. doi:10.1080/00224065.2006.11918623.
21. Mallows, C. L. (1973), "Some Comments on *Cp*," *Technometrics*, 15, 661675.
22. Muller, K. E., and Peterson, B. L. (1984), "Practical Methods for Computing Power in Testing the Multivariate General Linear Hypothesis," *Computational Statistics and Data Analysis*, 2, 143–158.
23. Mustafa, Qais and Alzubaydi Taha Hussein Ali (2013), *Comparing The Box-Jenkins Models Before And After The Wavelet Filtering In Terms Of Reducing The Orders With Application*, J. CONCRETE AND APPLICABLE MATHEMATICS, VOL. 11, NO. 2, 190-198, 2013, COPYRIGHT 2013 EUDOXUS PRESS LLC
24. Myers, R. H. (1990), *Classical and Modern Regression with Applications*, 2nd ed, Boston: PWS-Kent.
25. Nielsen, M., (2001), *On The Construction And Frequency Localization Of Finite Orthogonal Quadrature Filters*, Journal of Approximation Theory. Vol. 108, Number 1, pp(36-52)
26. Noorossana, R., M. Eyvazian, A. Amiri, and Mahmoud, M. A.. 2010b. *Statistical monitoring of multivariate multiple linear regression profiles in Phase I with calibration application. Quality and Reliability Engineering International* 26 (3):291–303. doi:10.1002/qre.1066.
27. Quick, C. (2013). Multivariate Multiple Regression with Applications to Powerlifting Data. Advisor: Kang James. University of Minnesota Duluth Department of Mathematics and Statistics.
28. Rao, C. R. (1966), "*Covariance Adjustment and Related Problems in Multivariate Analysis*," in Multivariate Analysis, P. Krishnaiah (ed.), New York: Academic Press, pp. 87–103.
29. Rencher, A. C. (2002). *Methods of multivariate analysis*. Wiley: New York.

Shewhart-type control chart, Quality Engineering, DOI: 10.1080/08982112.2022.2164725