Journal of Economics
and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq
DOI: https://doi.org/10.33095/jeas.v28i133.2352

Robust Estimation OF The Partial Regression Model Using Wavelet Thresholding

Ekhlass Abdulameer Al-Azzawi
University of Karbala, Karbala, Iraq
ekhlass.jaleel1201@coadec.uobaghdad.edu.iq

Lekaa Ali Al-Always
University Of Baghdad, Adhamiya, Baghdad, Iraq
Lekaa.ali.1968@gmail.com

Received: 9/2/2022 Accepted: 13/3/2022 Published: September / 2022

Abstract

Semi-parametric regression models have been studied in a variety of applications and scientific fields due to their high flexibility in dealing with data that has problems, as they are characterized by the ease of interpretation of the parameter part while retaining the flexibility of the non-parametric part. The response variable or explanatory variables can have outliers, and the OLS approach have the sensitivity to outliers. To address this issue, robust (resistance) methods were used, which are less sensitive in the presence of outlier values in the data. This study aims to estimate the partial regression model using the robust estimation method with the wavelet threshold and the PLM estimation method with the Speakman estimation and Nadarya-Watson smoothing, using simulation experiments at different sample sizes and contaminated ratios.

The mean square error criterion was employed to compare the two methods. The robust method is more efficient in obtaining robust estimators than the PLM estimation method.

Keywords: Partial linear regression, Outliers, Robustness , Wavelet thresholding, Spek man , Nadarya-Watson.

1- Introduction

The partial linear regression model technique is theoretically complex and frequently requires extensive software experience, but it is distinguished by the ease of interpretation of the parameter part and the retention of some flexibility of the non-parametric part, as it has greater flexibility in dealing with data that has problems when compared to the restricted parametric regression model, which must fulfill certain assumptions , However, if the number of explanatory variables is large, it may be difficult to interpret and provide inaccurate estimates, resulting in the Dimensional problem. [25;PP.xvii ,1]

The partial linear regression model's goal is to estimate the vector β and function g from the data. The analysis technique is to ignore or remove the non-parametric component (t) and analyze the parameter component βX from the model as if the non-parametric component is not present or vice versa. The partial regression model can be written as follows:

$$Y = X\beta + g(t) + u \qquad (1)$$

Since:

Y: the vector of the response variable at point t_i .

X: the matrix of parametric explanatory variables of dimension (n x m).

t: a continuous variable that represents the data's nonparametric component. It is an indicator, such as the time or distance of the observation, where t ∈ [0,1].

β: a vector of unknown parameter with the dimension (m x 1) of the explanatory variable matrix X.

g(t): unknown nonparametric function .

u: the vector of noise variables (random errors) of dimension (m x 1) which is normally distributed u ~ N(0, σ^2).

Several researchers have developed several methods for estimating partially linear models since Engel et al. [8] introduced it ,and one of the methods for estimating the nonparametric component in these models is based on the Spline Smoothing techniques presented by Hickman, as suggested by Speakman method Kernel Smoothing, Kosick also suggested the Partial Residual method. Profile likelihood method developed by Severini Wong and Carroll . Local linear estimation was also introduced by Hamilton and Trung . Zhang and Zhou were the first to propose the wavelet method. In addition to Härdle . comprehensive introduction to partial linear models theory and applications . Many of the above methods have been found to be closely related to the least-squares (LS) method. The LS method is known to be sensitive to the presence of outliers (outliers) in the data set, as outliers are observations that are numerically far from the rest of the data, which can occur due to artificial errors in data collection or by incorrectly including part of the sample with information. However, If there are outliers in the data set, it is best to replace the Ordinary Least Squares (LS) method with a robust method. As a result, some researchers have developed a set of immune procedures for the partial regression model in order to eliminate the influence of outliers during the data analysis and conclusion-drawing process. [16;PP.1,2] [9;PP.293,294].

In our study, we will address the robust estimation of the partial regression model in equation (1) by using the wavelet transform to estimate the non-parametric component of the model and the robust method to estimate the parameter vector of the parameter component using the M estimator. The method

for linking wavelet estimators to M estimates is based on the use of M estimation algorithms (Half-quardratic algorithms) represented by (ARTUR, LEGEND) algorithms for robust estimation of the parameter vector B.using simulation experiments. The comparison criterion, mean squares error (MSE) , is used to compare the two estimation methods.

2. The concept of outliers

Observations are often defined in terms of spread and concentration through the shape of the normal distribution, however because some observations are located far from the data center, they may follow a different pattern or none at all. Such abnormal observations are called outliers (which are observations with large residual values). Its presence causes severe distortion in the data and takes it away from the shape of the normal distribution, which may occur due to artificial errors when collecting data or by incorrectly including part of the sample with information . When outliers appear in the response variable, they are referred to as (Outliers), and when they appear in the explanatory variables, they are referred to as (Leverage Points), and when they appear in both variables, it makes estimating more difficult. Scientists and researchers have discussed several concepts and definitions of values outliers. It is defined by Broos (1961) as an observation that differs significantly from the other components of the sample. [6;PP.4], As Freeman defined it in 1980, it is a point of view that does not originate in the same manner that most other points of view do. [8;PP.350] .

As a result, outlier values are observations that are distant from the sample's center and have a considerable inaccuracy and bias when compared to the rest of the observations, reducing estimating efficiency. [12;PP.59]

3. Ordinary Estimation Method of Partial Linear Regression Model (PLM)

3.1. Speckman Estimation

Speckman (1988) proposed the following method for estimating the partial regression model: in equation (1), consider the conditional expectation of both sides of the partial regression model as follows: [13;PP.6] [23;PP.417]

$$Y - E(Y|t) = \{X - E(X|t)\}^T \beta + \{ u - E(u|t)\}$$

Let : $\widetilde{Y} = Y\,(I - S)$, $\widetilde{X} = X\,(I - S)$, $\widetilde{U} = u - E(u|t)$

Where S is the smoothing matrix for the estimator (N.W), and it takes the form:

$$S = \frac{K_h(t_i - t_j)}{\sum_{k=1}^{n} K_h(t_i - t_j)} \quad (2)$$

The smoothing matrix is replaced with $\widetilde{X}$ and $\widetilde{Y}$. We'll get the parameter estimator in the format shown below.

$$\widehat{\beta} = \left(\widetilde{X}^T \widetilde{X}\right)^{-1} \widetilde{X}^T \widetilde{Y} \quad (3)$$

The nonparametric estimator can be obtained using the following formula: [20;PP.3]

$$\widehat{g} = \frac{\sum_{i=1}^{n} k_h(t_i - t_j)(y_i - x_i^T \widehat{\beta})}{\sum_{i=1}^{n} k_h(t_i - t_j)} \quad (4)$$

$$\widehat{g} = S\,(y - X\widehat{\beta})$$

For the semi-parametric regression model, the Nadaraya-Watson estimator is used , and it is written as follows:

$$\widehat{g}_h(t) = \frac{n^{-1} \sum_i^n k_h(t - T_i) y_i}{n^{-1} \sum_i^n k_h(t - T_i)} = \sum_{i=1}^{n} w_i\, y_i = Wy \quad (5)$$

$$W_{ni} = \frac{K\left(\frac{t-T_i}{h_n}\right)}{\sum_{j=1}^{n}\left(\frac{t-T_j}{h_n}\right)} = \frac{K(u)}{\sum_{j=1}^{n} K(u)} \qquad (6)$$

Since $\{W_{ni}(t)\}_{j=1}^{n}$ denotes a set of weights that is normal if their sum is one. The size of the weights is represented by the weight function, which also represents the kernel function K(u), which is a real, decreasing function as u changes, and h_n represents the smoothing parameter (bandwidth) by which the size of the weights is defined (h > 0). The value of the observation T is determined by the distance or closeness of data for the observation ti., Furthermore, the (Gaussian) function will be used to estimate (N.W). [4;PP.254] [1;PP. 397] [10;PP.379]

The Cross-validation (CV) method will be used to choose the smoothing parameter, which is one of the most efficient and widely used methods. [14;PP.185] [21;PP.312]

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\{y_i - \hat{g}_h^{-1}(t_i)\}^2 \qquad (7)$$

and the optimal parameter with the smallest value for the criterion (CV) is obtained using the following equation: [2;PP.77,78]

$$h_{cv} = \text{argmin } CV_h \qquad (8)$$

4. The wavelet transformation and robust estimation in the partial regression model

The wavelet transform is a mathematical tool that converts data from the original field to the wavelet field by dividing it into different frequencies. It is used in the fields of mathematics, economics, engineering, social studies, and science for the representation, analysis, and processing of various data. Wavelets are mathematical functions that divide data into a group of compounds of varying frequencies and then study the effect of each compound using a solution consistent with its measurement. The wavelet theory was founded by the French researcher Fourier. [19; PP.xvii] **Thus, any function in the wavelet space can be analyzed into a scale function $\emptyset(t)$ known as the father wavelet and a wavelet function $\psi(t)$ known as the mother wavelet, which generates several functions known as sons wavelets, the function g is expressed in the form**

$$g(t) = \sum_{k=0}^{2^{j_0}-1} c_{j_0k}\emptyset_{j_0k}(t) \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} d_{jk}\psi_{jk}(t) \quad , \qquad t \in [0,1] \qquad (9)$$

Where $\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t-k}{2^j}\right)$ **(10) and**

$$\emptyset_{j_0,k}(t) = \frac{1}{\sqrt{2^{j_0}}}\emptyset\left(\frac{t-k}{2^{j_0}}\right) \qquad (11)$$

The wavelet coefficients are denoted by d_{jk} and the Scaling coefficients by c_{j_0k} .

$$c_{j_0k} = \left(g(t), \emptyset_{j_0k}(t)\right), \qquad (k = 0, 1, \cdots, 2^{j_0} - 1)$$

$$d_{jk} = \left(g, \psi_{jk}\right), (j \geq j_0 , \qquad k = 0, 1, \cdots, 2^j - 1)$$

Mallat proposed an efficient algorithm for calculating wavelet coefficients for a set of noisy data in (1989). He named it the Discrete Wavelet transformation (DWT) process because it divided the input signal into different frequency packets represented by the high-frequency filter G = $\{g_k\}$ and the low-frequency filter H = $\{h_k\}$, yielding the wavelet coefficients d_{jk} and the Scaling coefficients c_{j_0k}. Ingrid

Daubchies, a researcher, proposed a method in 1988 that uses compact supported for orthogonal wavelets and N vanishing moments, To create a family of Scaling functions (Daubchies) that is accompanied by wavelet functions, as this family leads to a family of smoothing Scaling functions. Daubechies wavelets are biorthogonal, sufficiently regular, and asymmetric. [16;PP.14,48,56]

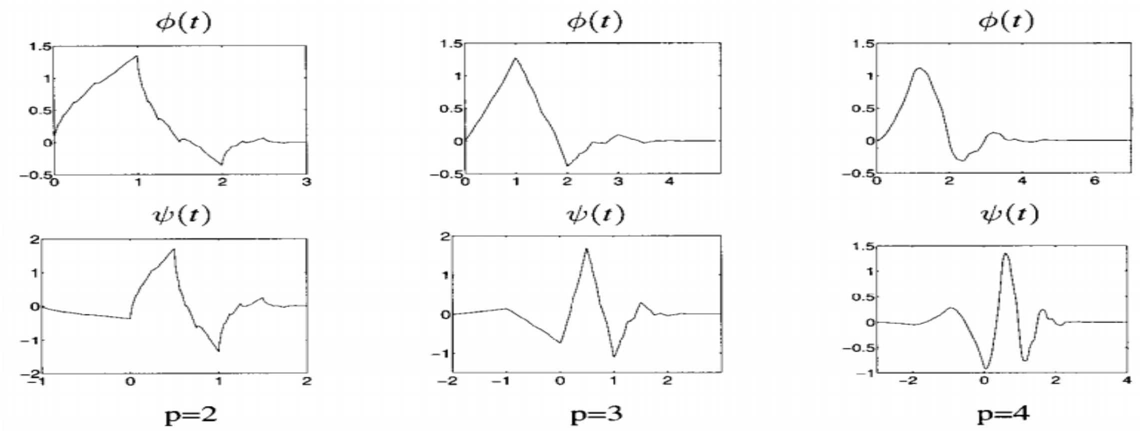


Figure 1. scaling ∅ and wavelet ψ functions of Daubechies with P vanishing moments

The nonparametric component g is represented as a function in infinite dimensional space in nonparametric analysis. The idea behind wavelet approaches is that the unknown function has an economical wavelet expression, i.e. g is, or is well approximated by, a function with a small proportion of nonzero wavelet coefficients. [9;PP. 294]

When there is incomplete or garbled data, signal noise occurs, and it is necessary to remove this noise from the original signal to receive correct info. As a result, comparing the set of wavelet coefficients with a value or set of values from the (Threshold) is an acceptable approach for reducing noise, as the idea of the threshold comprises setting zero for all wavelet coefficients whose values are less than the value of the threshold limit. As a result, in the kernel functions, the wavelet threshold will replace the smoothing parameter, and its rise or decrease will affect the amount of smoothing tainted data. The hard threshold function and the soft threshold function are two types of threshold law. [18;PP. 347] [15;PP.431]

$$\varphi_\lambda^H(t) = \begin{Bmatrix} t, & if \quad |t| > \lambda \\ 0, & if \quad |t| \le \lambda \end{Bmatrix} \qquad (12)$$

$$\varphi_\lambda^s(t) = \begin{Bmatrix} t-\lambda, & if \quad t > \lambda \\ t+\lambda, & if \quad t < -\lambda \\ 0, & if \quad |t| \le \lambda \end{Bmatrix} \qquad (13)$$

Whereas $\varphi_\lambda^H(t)$ denotes the hard threshold function , and $\varphi_\lambda^s(t)$ denotes to the soft threshold function.

It is assumed that the t points have the same dimensions $t_i = i/n$ and that the sample size is $n = 2^j$, where J represents positive integers. In the following steps, the data is converted from its original field to the wavelet (time-frequency) domain by applying the Daubechies discrete wavelet transform (DWT) to X and Y to obtain their corresponding wavelet representations A and Z. Where this transformation is based on both the scale function $\emptyset_{jk}(t)$ and the wavelet function

ψ_{jk}(t), both of which represent orthogonal wavelets of the space L^2 [0,1] to any nonparametric function $g \in L^2$ [0,1] .

We suppose that e = (e_1 , ….. , e_n $)^T$ is a vector of real numbers . The following equation gives the discrete wavelet transform of the vector e:

$d = W_{n \times n} e$ (14)

were d is the discrete scaling coefficients S_{j0k} and discontinuous wavelet coefficients w_{jk} of the vector e are both included in this vector of dimension (n x 1). And $W_{n \times n}$ is an orthogonal matrix connected with the base of the mother wavelet's orthogonal recurring wavelet. To create a PLM model in the form (1) based on wavelets, multiply both sides by the orthogonal matrix W to obtain the transformed model using the formula: [9;PP.295]

$K = V\beta + \theta + \epsilon$ (15)

Where $K = W_{n\times n}Y$, $V = W_{n\times n}X$, $\theta = W_{n\times n}g$, $\epsilon = W_{n\times n}U$

The Penalized Least Square PLS is used to estimate the parameters and in the model (6) by reducing the Object Function, which is divided into two parts, the Loss Function and the Penalty Function. in the following format:

$$(\widehat{\beta}_n, \widehat{\theta}_n) = arg \min_{(\beta,\theta)} \left\{ j_n(\beta,\theta) = \sum_{i=1}^{n} \frac{1}{2}\left(k_i - V_i^T\beta - \theta_i\right)^2 + \lambda \sum_{i=i_0}^{n} |\theta_i| \right\} \qquad (16)$$

Were λ is the smoothing parameter represented by the threshold value, which is chosen using the universal threshold approach. Only the empirical wavelet coefficients of the nonparametric section of the model, not the scaling coefficients, are penalized by the penalty term in the preceding equation. Were $i_0 = 2^{j0} + 1$. Both the estimators $\widehat{\beta}_n$ and $\widehat{\theta}_n$ represent two solutions to the optimization problem, employing the M-Huber method to estimate β and half-squared algorithms (ARTUR and LEGEND) to estimate θ, as shown in the following formulas

$$\widehat{\beta}_n = arg \min_{\beta} \sum_{i=i_0}^{n} \rho_\lambda(k_i - V_i^T\beta) \qquad (17)$$

$$\widehat{\theta}_{i,n} = \begin{cases} k_i - V_i^T\widehat{\beta}_n & if\ i < i_0, \\ \gamma_{soft,\lambda}\left(k_i - V_i^T\widehat{\beta}_n\right) & if\ i \geq i_0, \quad i = 1, \cdots, n \end{cases} \qquad (18)$$

where is Huber's cost functional ρ_λ defined by

$$\rho_\lambda(u) = \begin{cases} u^2/2 & if\ |u| \leq \lambda \\ \lambda|u| - \lambda^2/2 & if\ |u| > \lambda \end{cases} \qquad (19)$$

The soft threshold function $\gamma_{soft,\lambda}$ is defined by :

$$\gamma_{soft,\lambda}(u) = sign(u)(|u| - \lambda)_+ \qquad (20)$$

The inverse discrete wavelet transform IDWT is used to estimate the robust nonparametric component g by [3;PP.7,8] [7;PP.2,3]

$g = W^T_{n \times n} D$ (21)

4.1 Half-quaadratic algorithms

To obtain the estimator of the parameters vector β , we must solve the minimization problem by Standard optimization tools, such as the convex Huber (ρ_λ) loss function, whose second derivative is large near zero, which may result in slow optimization. As a result, for model (13), the half-squared optimization method for cost functions is used, which involves associating an auxiliary variable c with every β and creating an augmented criterion F so that the function F(β,c) is quadratic for each β (therefore, quadratic programming can be used) where each c can be calculated independently for each constant β using an appropriate formula. [9;PP.299]

$$\hat{\beta}_n = arg\min_{\beta} J(\beta) \quad , \quad J(\beta) = \sum_{i=1}^{n} \rho_\lambda\left(k_i - V_i^T\beta\right) \qquad (22)$$

Each c can be determined independently using an explicit formula for each β fixed. The augmented criterion F is chosen because it achieves the same minimum as J for the same value of β. The augmented energy optimization problem can be solved iteratively. At each iteration, one realized an optimization for c with respect to β for c fixed and a second optimization for β fixed with respect to c . If $\beta^{(m)}$ and $c^{(m)}$ are the values obtained after m iterations, the algorithm's (m + 1)th step actualizes these values by

$$\left.\begin{aligned} \beta^{(m+1)} &= arg\min_{\beta} F(\beta, c^{(m)}) \\ c^{(m+1)} &= arg\min_{c} F(\beta^{(m+1)}, c) \end{aligned}\right\} \qquad (23)$$

This method yields two algorithms, ARTUR and LEGEND, which are also known as IRLS and IMR in the literature.

4.1.1 ARTUR Algorithms

The Iterative Reweighted Least Squares Algorithm is also known as (ARTUR) in the robustness literature (IRLS). The Gaiman and Reynolds theorem leads to an improved criterion for the following form:

$$K(\beta, c) = \sum_{i=1}^{n} c_i\left(k_i - V_i^T\beta\right)^2 + \psi(c) \qquad (24)$$

The auxiliary variable c corresponds to a weight on the residuals of the least squares fit, which explains the IRLS terminology. Weights on large residuals, intuitively, have a tendency to eliminate the corresponding responses from the fit. As a result, the ARTUR algorithm's m+1 step can be described as follows

$$\begin{cases} r_i^{(m)} = k_i - V_i^T\beta^{(m)} \\ c_i^{(m+1)} = \dfrac{\rho_\lambda'\left(2r_i^{(m)}\right)}{2r_i^{(m)}}, \qquad \forall i \in \{1, \cdots, n\} \\ \beta^{(m+1)} = \left(V^T c^{(m+1)} V\right)^{-1} V^T c^{(m+1)} K \end{cases}$$

4.1.2 LEGEND Algorithms

Is an algorithm that is a little different. Instead of weighing the residuals, the auxiliary variable subtracts the larger residual values. The existence of the related augmented energy functional is proven by Geman and Reynolds' second theorem (1992). The criterion that should be minimized is as follows

$$F(\beta, c) = \sum_{i=1}^{n} \left(k_i - V_i^T \beta - c_i\right)^2 + \xi(c) \qquad (25)$$

The LEGEND algorithm's m + 1 step can be expressed as follows, using similar notation as the ARTUR algorithm.

$$\begin{cases} r^{(m)} = K - V\beta^{(m)} \\ c_i^{(m+1)} = r_i^{(m)} \left(1 - \dfrac{\rho'_\lambda\left(2r_i^{(m)}\right)}{2r_i^{(m)}}\right), \qquad \forall i \in \{1, \cdots, n\} \\ \beta^{(m+1)} = (V^T V)^{-1} V^T \left(K - c^{(m+1)}\right) \end{cases}$$

5. Simulation Study

A simulation is an imitation or portrayal of the operation of a genuine system over a period of time. Simulation, whether done manually or with a computer, is based on the creation of a model of the real system. This model is made up of a set of assumptions about how the system operates, which can be expressed as mathematical, logical, or symbolic relationships between system elements. Following the development and activation of the model, it is used to conduct some experiments that cannot be conducted on the real system in order to observe and deduce the many changes and interactions that would occur in the system if they were conducted on it. [5;PP.14]

The following steps can be used to describe the stages of simulation experiments:

1- Explanatory variables (x_k) with a normal distribution are generated using the Box-Muller method by first generating two random variables u_1 and u_2 that follow the uniform distribution u (0,1), and then converting these two variables into independent random variables x_1 and x_2 that follow the standard normal distribution. Random errors e_i with a mean of zero and a variance of σ_e^2 also generated . then nonparametric explanatory variables (t) are generated that follow the uniform distribution.

2- The following formulas are used to select the nonparametric smoothing functions g(z i) (quadratic, sinusoidal, and cubic):

$g_1(t_i) = 3.2t^2 - 1$ [24;PP.16]

$g_2(t_i) = \sin(2t) + 2\exp(-16t^2)$ [11;PP.12]

$g_3(t_i) = t - 3t^2 + 3t^3$ [22;PP.534]

3- According to the partial regression model, the dependent variable is generated directly using the explanatory variables, random errors, and smoothing functions that were generated as follows:

$$y_i = \sum_{j=1}^{m} \beta_j x_{ij} + g_j\left(t_{ij}\right) + e_{ij}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, m \qquad (26)$$

4- The values of the semi-parametric model's parameters are determined by using the least squares method to estimate them as follows:

$(\beta_1 = 0.5, \beta_2 = 1.5, \beta_3 = -1.25)$

5- The following formula is used to contaminated the variables with contaminated rates (C = 10%,20%,40%) :

$$Y = (1 - C)M_1 + CM_2 \quad , \quad C \neq 0 \qquad (27)$$

whereas:

C: the contamination percentage to be used to contaminate the data.

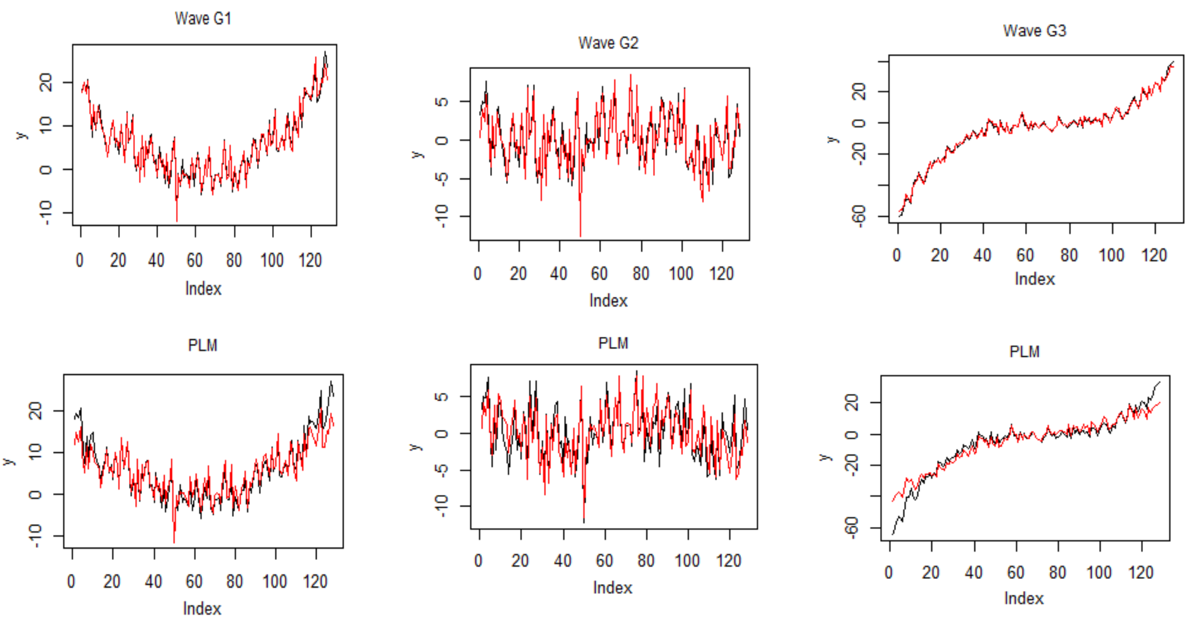
$(1-C)$: Percentage of uncontaminated data.
Y, M_1, M_2 : It shows the final contaminating variable, the uncontaminated variable, and the contaminating variable in that order.

6- The simulation experiments next apply the two studied estimation methods. For the wavelet estimation method, filters with level (4) , soft Universal threshold and Daubechies function were used. [17;PP.29]

$$\lambda_{UV} = \sigma\sqrt{2\ln(n)} \qquad (28)$$

7- Three different sizes (n = 64,128,256) and three contamination rates (= 10, = 20, = 40) were used, and the steps were repeated (R = 1000) . and the average was taken to find the estimators and get the final values of the estimators and the average error squares .

The following tables and figures show the results:



M_1 **M_2** **M_3**

Figure 2.Curves of true values and estimated values of response variable Y for the three models and for the two estimation methods WAVE and PLM at (β1=0.5,2=0.75,β3=-1.25), θ=10%, n=128

Table 1: Estimated values of the parameters for the two estimation methods WAVE and PLM at C = 10% when (β_1 = 0.5, β_2 = 1.5, β_3 = -1.25)

Models	Method $g(t)$	n	$\hat{\beta}$	WAVE	Baise	PLM	Baise
M_1	g_1	64	$\hat{\beta}_1$	0.3926	0.1074	0.5641	-0.0641
			$\hat{\beta}_2$	1.4679	0.0321	1.2108	0.2892
			$\hat{\beta}_3$	-1.1799	-0.0701	-1.3851	0.1351
		128	$\hat{\beta}_1$	0.4605	0.0395	0.633	-0.133
			$\hat{\beta}_2$	1.4623	0.0377	1.414	0.086
			$\hat{\beta}_3$	-1.2570	0.007	-1.322	0.072
		256	$\hat{\beta}_1$	0.4825	0.0175	0.5525	-0.0525
			$\hat{\beta}_2$	1.4508	0.0492	1.4616	0.0384
			$\hat{\beta}_3$	-1.2701	0.0201	-1.3071	0.0571
M_2	g_2	64	$\hat{\beta}_1$	0.518	-0.018	0.5669	-0.0669
			$\hat{\beta}_2$	1.451	0.049	1.5355	-0.0355
			$\hat{\beta}_3$	-1.313	0.063	-1.0660	-0.184
		128	$\hat{\beta}_1$	0.3938	0.1062	0.5585	-0.0585
			$\hat{\beta}_2$	1.5171	-0.0171	1.4507	0.0493
			$\hat{\beta}_3$	-1.2745	0.0245	-1.1763	-0.0737
		256	$\hat{\beta}_1$	0.469	0.031	0.5611	-0.0611
			$\hat{\beta}_2$	1.410	0.09	1.4357	0.0643
			$\hat{\beta}_3$	-1.229	-0.021	-1.2329	-0.0171
M_3	g_3	64	$\hat{\beta}_1$	0.2494	0.2506	0.8418	-0.3418
			$\hat{\beta}_2$	1.2424	0.2576	1.9908	-0.4908
			$\hat{\beta}_3$	-0.6130	-0.637	-0.7825	-0.4675
		128	$\hat{\beta}_1$	0.3725	0.1275	0.5145	-0.0145
			$\hat{\beta}_2$	1.5759	-0.0759	1.6945	-0.1945
			$\hat{\beta}_3$	-1.4934	0.2434	-0.8641	-0.3859
		256	$\hat{\beta}_1$	0.411	0.089	0.4732	0.0268
			$\hat{\beta}_2$	1.316	0.184	1.4857	0.0143
			$\hat{\beta}_3$	-1.288	0.038	-1.2881	0.0381

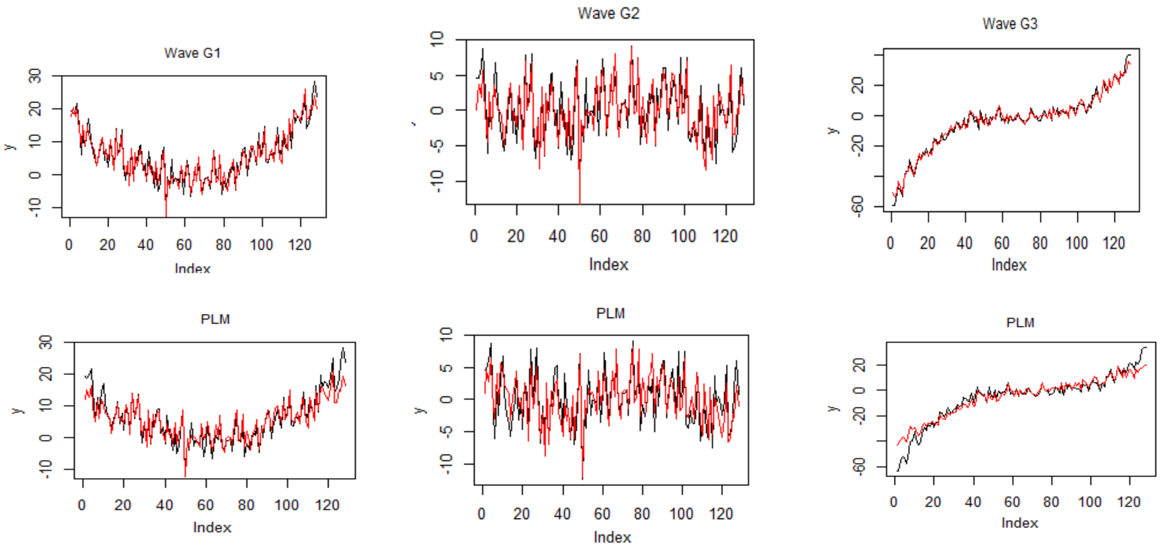
5.1 Results and Discussion

We note from table1 and figure2 that the bias values show that they decrease as the estimated values of the parameters approach the default values. The robust estimation technique (WAVE) was chosen over the (PLM) approach in the M_1 and, M_2 models since the estimated values of its parameters converged to the default values at sample sizes (256,64,128) .

In the second model M_2 The ordinary estimating technique (PLM) has advanced above the robust estimation method (WAVE) through the convergence of the values of its estimated parameters at sample sizes (256,128,64), followed by (WAVE) method at sample sizes (64, 128).

Table 2. Estimated values of the parameters for the two estimation methods WAVE and PLM at C = 20% when (β_1 = 0.5, β_2 = 1.5, β_3 = -1.25)

Models	Method / g(t)	n	$\hat{\beta}$	WAVE	Baise	PLM	Baise
M_1	g_1	64	$\hat{\beta}_1$	0.3804	0.1196	0.5836	-0.0836
			$\hat{\beta}_2$	1.4024	0.0976	1.0712	0.4288
			$\hat{\beta}_3$	-1.1529	-0.0971	-1.4519	0.2019
		128	$\hat{\beta}_1$	0.4127	0.0873	0.7093	-0.2093
			$\hat{\beta}_2$	1.4691	0.0309	1.3885	0.1115
			$\hat{\beta}_3$	-1.2938	0.0438	-1.3520	0.102
		256	$\hat{\beta}_1$	0.4688	0.0312	0.5823	-0.0823
			$\hat{\beta}_2$	1.4034	0.0966	1.4414	0.0586
			$\hat{\beta}_3$	-1.2695	0.0195	-1.3375	0.0875
M_2	g_2	64	$\hat{\beta}_1$	0.5052	-0.0052	0.5433	-0.0433
			$\hat{\beta}_2$	1.3956	0.1044	1.5586	-0.0586
			$\hat{\beta}_3$	-1.2911	0.0411	-1.0951	-0.1549
		128	$\hat{\beta}_1$	0.3458	0.1542	0.5833	-0.0833
			$\hat{\beta}_2$	1.5242	-0.0242	1.4463	0.0537
			$\hat{\beta}_3$	-1.3082	0.0582	-1.1453	-0.1047
		256	$\hat{\beta}_1$	0.4564	0.0436	0.5815	-0.0815
			$\hat{\beta}_2$	1.3642	0.1358	1.4147	0.0853
			$\hat{\beta}_3$	-1.2282	-0.0218	-1.2369	-0.0131
M_3	g_3	64	$\hat{\beta}_1$	0.2319	0.2681	0.9371	-0.4371
			$\hat{\beta}_2$	1.1696	0.3304	2.0784	-0.5784
			$\hat{\beta}_3$	-0.5780	-0.672	-0.6797	-0.5703
		128	$\hat{\beta}_1$	0.3306	0.1694	0.5825	-0.0825
			$\hat{\beta}_2$	1.5693	-0.0693	1.7834	-0.2834
			$\hat{\beta}_3$	-1.5355	0.2855	-0.7964	-0.4536
		256	$\hat{\beta}_1$	0.3926	0.1074	0.5105	-0.0105
			$\hat{\beta}_2$	1.2755	0.2245	1.5101	-0.0101
			$\hat{\beta}_3$	-1.2845	0.0345	-1.2721	0.0221



M_1 **M_2** **M_3**

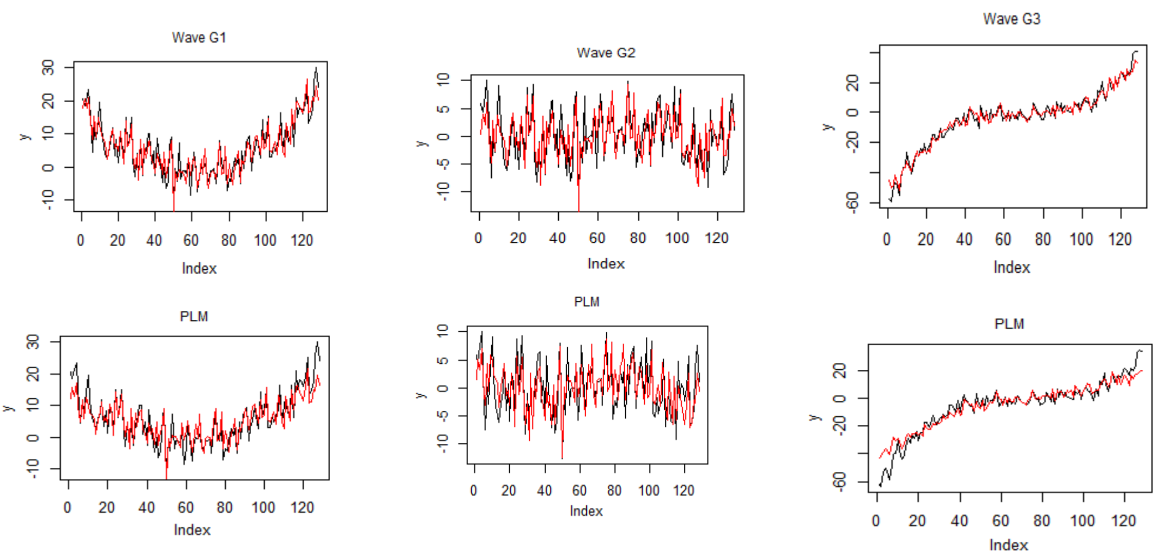
Figure 3.Curves of true values and estimated values of response variable Y for the three models and for the two estimation methods WAVE and PLM at (β1=0.5,2=0.75,β3=-1.25), θ=20%, n=128

Table2 and figure3 reveal that the (WAVE) technique outperforms the estimate method (PLM) in the first and third models M_1 and M_3 at sample sizes of 128, 256, and 64, respectively, with a 20% contaminated percent. Unlike the first and third models, the second model demonstrates that the (PLM) approach outperforms the (WAVE) estimate method when the sample size is (64,128,256).

Table 3. Estimated values of the parameters for the two estimation methods WAVE and PLM at C = 40% when (β_1 = 0.5, β_2 = 1.5, β_3 = -1.25)

Models	Method / g(t)	n	$\hat{\beta}$	WAVE	Baise	PLM	Baise
M_1	g_1	64	$\hat{\beta}_1$	0.3693	0.1307	0.6031	-0.1031
			$\hat{\beta}_2$	1.3458	0.1542	0.9316	0.5684
			$\hat{\beta}_3$	-1.1281	-0.1219	-1.5187	0.2687
		128	$\hat{\beta}_1$	0.3718	0.1282	0.7854	-0.2854
			$\hat{\beta}_2$	1.4800	0.02	1.3634	0.1366
			$\hat{\beta}_3$	-1.3340	0.084	-1.3822	0.1322
		256	$\hat{\beta}_1$	0.4595	0.0405	0.6121	-0.1121
			$\hat{\beta}_2$	1.3622	0.1378	1.4212	0.0788
			$\hat{\beta}_3$	-1.2751	0.0251	-1.3680	0.118
M_2	g_2	64	$\hat{\beta}_1$	0.4923	0.0077	0.5198	-0.0198
			$\hat{\beta}_2$	1.3403	0.1597	1.5818	-0.0818
			$\hat{\beta}_3$	-1.2690	0.019	-1.1243	-0.1257
		128	$\hat{\beta}_1$	0.311	0.189	0.608	-0.108
			$\hat{\beta}_2$	1.538	-0.038	1.442	0.058

			$\hat{\beta}_3$	**-1.362**	**0.112**	**-1.114**	**-0.136**
		256	$\hat{\beta}_1$	**0.4439**	**0.0561**	**0.6019**	**-0.1019**
			$\hat{\beta}_2$	**1.3181**	**0.1819**	**1.3937**	**0.1063**
			$\hat{\beta}_3$	**-1.2271**	**-0.0229**	**-1.2409**	**-0.0091**
M_3	g_3	64	$\hat{\beta}_1$	**0.2031**	**0.2969**	**1.0325**	**-0.5325**
			$\hat{\beta}_2$	**1.1127**	**0.3873**	**2.1660**	**-0.666**
			$\hat{\beta}_3$	**-0.5294**	**-0.7206**	**-0.5769**	**-0.6731**
		128	$\hat{\beta}_1$	**0.2816**	**0.2184**	**0.6505**	**-0.1505**
			$\hat{\beta}_2$	**1.5801**	**-0.0801**	**1.8722**	**-0.3722**
			$\hat{\beta}_3$	**-1.6027**	**0.3527**	**-0.7287**	**-0.5213**
		256	$\hat{\beta}_1$	**0.377**	**0.123**	**0.5478**	**-0.0478**
			$\hat{\beta}_2$	**1.236**	**0.264**	**1.5344**	**-0.0344**
			$\hat{\beta}_3$	**-1.287**	**0.037**	**-1.2560**	**0.006**



M_1 M_2 M_3

Figure 4.Curves of true values and estimated values of response variable Y for the three models and for the two estimation methods WAVE and PLM at $(\beta1=0.5,2=0.75,\beta3=-1.25)$, $\theta=40\%$, n=128

as shown in Table3 and figure4 , and through small bias values and At a 40 percent contaminated ratios, note that the (WAVE) method is also better than the (PLM) method in the first model M_1 at the samples sizes (64,128,256) respectively, while we note the progress of the estimation method (PLM) on the method (WAVE) in the second M_2 and third M_3 models with sample sizes (256,128,64).

Table 4: The values of the MSE criterion to compare between the two estimation methods WAVE and PLM for the partial linear regression model when ($\beta_1 = 0.5, \beta_2 = 1.5, \beta_3 = -1.25$) and at contaminated rates θ = 10%, θ = 20%, θ = 40%

C	Models	Method g(t)	n	WAVE	PLM
10%	M_1	g_1	64	0.07962	0.2552
			128	0.05699	0.2227
			256	0.05121	0.1358
	M_2	g_2	64	0.07349	0.2406
			128	0.05622	0.1378
			256	0.03829	0.08044
	M_3	g_3	64	0.2026	0.8033
			128	0.08761	1.287
			256	0.06739	0.9733
20%	M_1	g_1	64	0.1615	0.368
			128	0.1253	0.3028
			256	0.1089	0.2058
	M_2	g_2	64	0.1582	0.3843
			128	0.1147	0.2207
			256	0.08284	0.1507
	M_3	g_3	64	0.3175	0.8749
			128	0.1795	1.349
			256	0.1499	1.042
40%	M_1	g_1	64	0.2749	0.5255
			128	0.2162	0.4149
			256	0.1867	0.3031
	M_2	g_2	64	0.2747	0.5807
			128	0.1948	0.3381
			256	0.1454	0.2495
	M_3	g_3	64	0.5438	0.9906
			128	0.2957	1.442
			256	0.259	1.138

The values of the comparison standard (MSE) for the two estimating methods (WAVE) and (PLM) are shown in Table (4), where we see that the (WAVE) approach recorded the lowest value for the (MSE) standard in all three models and at all sample sizes and pollution rates.

We also notice that as the sample size is increased, the (MSE) value of the two estimation methods decreases, with some fluctuation in the MSE values when the sample size is increased in the (PLM) method in the third model M_3 and at the three ratios of contamintaed . The results also revealed that the second model had the lowest MSE criterion value for the two estimation methods and the three contamintaed ratios than the first and third models.

6- Conclusions

1- For all models, sample sizes, and contamination rates allowed in experiments, the (WAVE) estimating method outperforms the ordinary estimation method (PLM) in terms of obtaining more efficient estimations because it is the most effective in data trimming and smoothing.

2- For all sample sizes in the first and third models at the 10% and 20% contaminated ratios, and in the second model at the 10% contaminated ratio, the estimated values of the parameters using the robust estimation method (WAVE) are the closest to the default parameter values.

3- For most models and sample sizes, the value of the mean square error (MSE) is inversely proportional to the sample size, with the larger the sample size, the smaller the (MSE).

4- By recording the minimum value of the MSE criterion at the three sample sizes and the recorded contaminated ratio, the second model is more efficient than the first and third models.

5- Because it records the highest values of the MSE criterion at all three sample sizes and contaminated ratios, the third model is regarded as the least efficient of the first and second models.

References

1- Abdul Hussein,R.H and Kadhem,S.H,(2020)." Comparison of weighted estimated method and proposed method (BEMW) for estimation of semi-parametric model under incomplete data". Journal of Economics and Administrative Sciences, Vol.26 (NO. 120), pp. 395-414 .

2- Adrian W. Bowman & Azzalini A., 1997. "Applied Smoothing Techniques for Data Analysis". CLARENDON PRESS • OXFORD.

3- Antoniadis, A., Fan, J., (2001). "Regularization of wavelet approximations".J. Am. Stat. Assoc. 96, 939–967 .

4- Aydin,D.(2007). "Acomparison of the nonparametric regrestion model using smoothing spline and kernel regression" , World Academy of Science, Engineering and Technology 36,253-257

5- Beri,A.,(2002). "Modeling and simulation using Excel, SIMAN, Arena and General purpose simulation system (GBS WORLD)". King Saud University / Department of Statistics and Operations Research.

6- Bross, I. D. (2012). "Outliers in patterned experiments a strategic appraisal". Technometrics, 3(1), 91-102.

7- Chang, X., Qu, L., (2004)".Wavelet estimation of partially linear models". Comput. Stat. Data Anal. 47, 31–48 .

8- Freeman, P.R. , (1980). "On the Number of Outliers in Data from a Linear Model". In Bayesian Statistics, Ed. J.M.

9- Gannaz,I.,(2007)."Robust Estimation and Wavelet Thresholding in Partially Linear Models".Stat Comput 17:293-310.

10- Gatea,M.M, Hamoud,M.Y,(2014)." A comparison of the Semiparametric Estimators model using different smoothing methods". Journal of Economics and Administrative Sciences, Vol.20 (NO. 75), pp. 376-394 .

11- Gerolimetto, M., & Magrini, S. (2009). "Nonparametric regression with spatially dependent data". Department of Economics WP, 20- 2009.

12- Hampel F., Ronchetti E., Rousseeuw P., Stahel W., (1986)".Robust Statistics: The Approach Based on Influence Functions", Johan Wiley & Sons, Inc.

13- Hardle ,W., Liang,H., and Gao,J., (2000)."Partially linear models" .

14- Hardle,W., (1990). "Applied nonparametric regression" Gambridg-MA:cambirg university press.

15- Hassan,Y.A. and Hmood,M.Y. ,(2020)," Estimate the Partial Linear Model Using Wavelet and Kernel Smoothers", Journal of Economics and Administrative Sciences, Vol.26 (NO. 119), pp. 428-443 .

16- Jiang,Y.,(2015). "Robust estimation in partially linear regression models", J. Appli. Statist. pp. 1–12.

17- Kadhem,S.and Hussain,Dh.,(2015)," Wavelet Analysis For Regression Model With Missing Data". A dissertation of PH.D of Science in Statistics, College of Economics and Administration, Baghdad University.

18- Kozlowski, J., Serafinczuk, J and Kozik, A. (2003). "Wavelet shrinkage-based noise reduction from the high resolution X-ray images of epitaxial layers", Materials Science, Vol. 21,No. 3.

19- Misiti, M., Misiti, Y., Oppenheim, G., Michel, J., (2007). "Wavelets and their Applications", ISTE Ltd.

20- M*ü*ller,M.(2014)." An Introduction to the Estimation of GPLMs and Data Examples for the R gplm Package", https://cran.r-project.org/web/packages/gplm/vignettes/gplm-examples.pdf.

21- Robert, F.E., Garanger, C.W.J., Rice,J. and Weiss, A., (1986). "Semiparametric estimates of the relation between weather and electricity sales", JASA, Vol.81, No.394, PP 310-320.

22- Schimek,M.G.,(2000)," Estimation and inference in partially linear models with smoothing splines" Journal of Statistical Planning and Inference(91) ,PP 525_540.

23- Speckman,P.,(1988). "Kernel smoothing in partial linear models", J. R. Statist. Soc. Ser. B (Methodol.) 50, pp. 413–436.

24- Wang,Q.,Linton,O.,and Härdle,W.,(2004),*"* Semiparametric Regression Analysis with Missing Response at Random " the institute for fiscal studies department of economics, UCL ,*ce*mmap working paper CWP11/03.

25- Yatchew,A.,(2003)".semiparametric regression for the applied conometrician",Cambridge university press.

التقدير الحصين لانموذج الانحدار الجزئي باستعمال عتبة مويجية

الباحث/ اخلاص عبد الامير حميد العزاوي
جامعة كربلاء/كلية الصيدلة
كربلاء / حي شهداء الموظفين
ekhlass.jaleel1201@coadec.uobaghdad.edu.iq

أ.د. لقاء علي محمد العلوي
جامعة بغداد/كلية الادارة والاقتصاد
بغداد/ الاعظمية
Lekaa.ali.1968@gmail.com

Received: 9/2/2022 Accepted: 13/3/2022 Published: September / 2022

مستخلص البحث

توالت الدراسات حول نماذج الانحدار الشبه معلمية في العديد من التطبيقات والمجالات العلمية المتنوعة وذلك لما تتمتع به من مرونة عالية في التعامل مع البيانات التي تعاني من المشاكل حيث انها تتميز بسهولة تفسير الجزء المعلمي مع الاحتفاظ بمرونة الجزء اللامعلمي ، و يمكن لمتغير الاستجابة أو المتغيرات التوضيحية أن تحتوي على قيم شاذة (Outliers)،حيث تفشل طريقة المربعات الصغرى OLS في تقديرها وذلك لحساسيتها للقيم الشاذة . ولعلاج هذه المشكلة تم استخدام الطرائق الحصينة او (المقاومة) التي تكون أقلّ حساسية في حالة وجود القيم الشاذة في البيانات. يهدف هذا البحث الى تقدير انموذج الانحدار الجزئي باستعمال طريقة التقدير الحصين باستعمال الاسلوب المويجي ومقارنته مع طريقة التقدير الاعتيادية (PLM) من خلال تقدير (Speakman) وممهد Nadarya-Watson وذلك عن طريق تنفيذ تجارب المحاكاة عند حجوم عينات ونسب تلوث مختلفة. تمت المقارنة بين الطريقتين باستخدام معيار متوسط مربعات الخطأ (Mean square error) . حيث وجد ان طريقة التقدير الحصين من خلال الاسلوب المويجي هي الاكثر كفاءة في الحصول على مقدرات حصينة من طريقة التقدير العادية للانموذج .

المصطلحات الرئيسة للبحث: انموذج الانحدار الجزئي ، القيم الشاذة ، الانحدار الحصين ، العتبة المويجية ، Speakman ، Nadarya-Watson .

*البحث مستل من رسالة ماجستير