Journal of Economics and Administrative Sciences
2024; 30(141), pp. 479-496

P-ISSN 2518-5764
E-ISSN 2227-703X

Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq

Comparison of Some Robust Methods of Difference Based Ridge Estimator in Semiparametric Regression Model

Wissam Khalid Saed *
College of Administration and Economics,
University of Baghdad
Baghdad, Iraq
wissam.k.saed.1988@gmail.com
***Corresponding author**

Rabab Abdulrida Saleh
College of Administration and Economics,
University of Baghdad
Baghdad, Iraq
rabab.saleh@coadec.uobaghdad.edu.iq

Received:27/7/2023 Accepted:13/9/2023 Published Online First: 30 /6/ 2024

Abstract:

The semiparametric regression models have received a lot of attention from researchers recently because it combines parametric and nonparametric methods, it is one of the advanced topics in data analysis for various studies, which aims to find the best capabilities and a high level of efficiency.

One of the most important semiparametric regression models is the partial linear regression model (PLM), which consists of a parametric component and a nonparametric component, for the purpose of estimating the parametric component, the difference method will be used to remove the nonparametric component.

When the analysis hypotheses of the parametric component are not fulfilled, it will suffer from several problems, the most important of which is the problem of complete multicollinearity, besides the multicollinearity, there are also outliers in the data.

In this research, the problems of multicollinearity and outliers of the semiparametric regression model were addressed, where simulation was used to generate data with different sample sizes and for different correlations and outlier ratios and for different methods such as [Difference Ridge based M robust with Nadaraya – Watson (DRMNW), Difference Ridge based S robust with Nadaraya – Watson (DRSNW), Difference Ridge based M robust with Smoothing spline (DRMSP), Difference Ridge based S robust with Smoothing spline (DRSSP)], the results showed that method Difference Ridge based M robust with Smoothing spline (DRMSP) is the best estimator.

Paper type: *Research paper*.
Keywords: Semiparametric Regression Model, Multicollinearity, Outliers, Robust estimates, Difference-Based Ridge Estimator, M, S, Kernel function, Nadaraya – Watson estimator, Cubic Smoothing Spline estimator.

1.Introduction:

Semiparametric models as having the parametric component and the nonparametric component). (Powell, 1994)

One of the most important semiparametric models is the partial linear regression model (PLM). (Speckman, 1988)

The problem of multicollinearity only when there is a linear relationship between some or all of the explanatory variables, and that the correlations between these variables are known as multicollinearity, and one of the most important conditions that must be met is the rank condition. (Abboud & Khorshid, 2018)

When there is a multicollinearity in the semiparametric regression model, it results of the estimates are inaccurate and inefficient. Besides the multicollinearity, there are also outliers. Robust ridge regression methods are used to overcome the effects of multilinearity and outliers. (Roozbeh, 2016)

For instance, the principle difference based ridge regression estimator in (2010), the researchers (Tabakan & Akdeniz, 2010) used difference based ridge regression estimator $\widehat{\boldsymbol{\beta}}_{(\mathbf{K})}$ in (PLM). It is (MSE) compared analytically with the nonridge $\widehat{\boldsymbol{\beta}}_{(\mathbf{0})}$, the $\widehat{\boldsymbol{\beta}}_{(\mathbf{K})}$ superiority over $\widehat{\boldsymbol{\beta}}_{(\mathbf{0})}$.

In (2016), the researcher (Roozbeh, 2016) in the case of a multicollinearity in addition to the presence of outliers in the semiparametric regression model, in this regard, the researcher proposed a new form of ridge for robust and generalized restricted estimators based on Least trimmed squares (LTS) method in a semiparametric. The proposed estimator proved to be the best estimator for the parametric part of the model.

In (2021), the researcher (Shih et al., 2021), in case the data contains outliers and multicollinearity, in this case ridge M estimator is the best estimator, better than the LS estimator. A lot of estimators, such as the M estimator before the ridge and the ridge M estimator for Stein-rule shrinkage, while it was developed on the basis of the Ridge M. estimator. Three robust estimations are proposed, which are used for resistance to outliers and multicollinearity. The simulation confirmed that all the proposed estimators and the Ridge M. estimators are better than the Least squares estimator (LSE).

In (2022), the researcher (Herawati et al., 2022), if there are two problems in the data of the multilinearity problem and the outliers problem, to overcome them, robust regression is used. So, it was using ridge least absolute deviation method. A comparison by MSE of the two methods ridge least absolute deviation (RLAD) and Least squares (LS), all results showed that (RLAD) had a lower MSE than (LS).

2.Robust estimation methods:

2.1 M estimation:

It is symbolized by the symbol (M), this method is considered one of the most robust methods. As the work of this method works with the problem of outliers, through which the squares of the residuals are replaced by the loss function ρ. (Irshayyid & Saleh, 2022)

This method was introduced by Hooper in 1964. This method is almost as efficient as the ordinary least squares method. Instead of reducing the sum of the squares of errors as the objective, the M estimate reduces the errors function (p). (Alma, 2011) (Rashid & Hafez, 2013) (A. F. Lukman et al., 2015)

$$\min \sum_{i=1}^{n} p\left(\frac{e_i}{S}\right) = \min \sum_{i=1}^{n} p\left(\frac{y_i - x_i'\beta}{S}\right) \qquad (1)$$

The properties for the function ρ can be written as:
$p(e) \geq 0$, $p(0) = 0$, $p(e) = p(-e)$ and $p(e_i) \geq p(e_i')$ for $|e_i| \geq |e_i'|$ (A. F. Lukman et al., 2015)
The algorithm for the M estimation can be written as follows: (Susanti et al., 2014)

1) Calculate the parameter $\hat{\beta}_{OLS}(0)$ by assuming initial values such as OLS.
2) Calculate the residual: $e_i = y_i - \hat{y}_i$
3) Calculate:

$$\hat{\sigma}_i = \frac{MAD}{0.6745} = \frac{median|e_i - median(e_i)|}{0.6745}$$

4) Calculate: $u_i = \frac{e_i}{\hat{\sigma}_i}$
5) Calculating the weighted value of the Tukey function:

$$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{4.685}\right)^2\right]^2 & |u_i| \leq 4.685 \\ 0 & |u_i| > 4.685 \end{cases}$$

6) Calculating the $\hat{\beta}_M$ using the weighted least squares method with w_i:

$$\hat{\beta}_M = (X'W_iX)^{-1}(X'W_iY)$$

7) Repeat steps 2 to 5 to obtain a convergent value of $\boldsymbol{\hat{\beta}_M}$.

2.2 S estimation:

S estimation by Rousseeuw and Yohai (1984), it is a method that has a high breakdown. As this method has a higher statistical efficiency than trimmed least squares estimation. (Chen, 2002) (ŞAMKAR & ALPU, 2010) (Lukman et al., 2014)

The S method is more robust than the M method because the S estimators have smaller asymptotic variance and smaller asymptotic bias in the data with outliers. (Bahez & Rasheed, 2022)

where the S estimator is derived by scale statistics in an implicit manner corresponding to S(θ). (Lukman et al., 2015)

According to Salibian and Yohai, the S-estimator is defined by: (Susanti et al., 2014)

$$\boldsymbol{\hat{\beta}_S = \min \beta\, \hat{\sigma}_S(e_1, e_2, \dots, e_n)} \qquad (2)$$

With determining minimum robust scale estimator $\boldsymbol{\hat{\sigma}_S}$:

$$\boldsymbol{\min \sum_{i=1}^{n} p\left(\frac{y_i - \sum_{i=0}^{n} x_{ij}\beta}{\hat{\sigma}_S}\right)} \qquad (3)$$

The algorithm for the S estimation can be written as follows: (Susanti et al., 2014) (Abbas & Abood, 2022)

1) Calculate the parameter $\boldsymbol{\hat{\beta}_{OLS}(0)}$ by assuming initial values such as OLS.
2) Calculate the residual: $\boldsymbol{e_i = y_i - \hat{y}_i}$
3) Calculate:

$$\boldsymbol{\hat{\sigma}_i = \begin{cases} \frac{median|e_i - median(e_i)|}{0.6745} & iteration = 1 \\ \sqrt{\frac{1}{nk}\sum_{i=1}^{n} w_i e_i^2} & iteration > 1 \end{cases}}$$

4) Calculate: $\boldsymbol{u_i = \frac{e_i}{\hat{\sigma}_i}}$
5) Calculating the weighted value of the Tukey function:

 P-ISSN 2518-5764
E-ISSN 2227-703X

$$w_i = \begin{cases} \begin{cases} \left[1-\left(\frac{u_i}{1.547}\right)^2\right]^2 & |u_i| \le 1.547 \\ 0 & |u_i| > 1.547 \end{cases} & \text{iteration} = 1 \\ \frac{p(u_i)}{u^2} & \text{iteration} > 1 \end{cases}$$

6) Calculating the $\hat{\beta}_S$ using the weighted least squares method with w_i:

$$\hat{\beta}_S = (X'W_iX)^{-1}(X'W_iY)$$

7) Repeat steps 2 to 5 to obtain a convergent value of $\hat{\beta}_S$.

3. Semiparametric Regression Model:

Semiparametric models were introduced by Begun et al in 1983, the term being attributed to Oakes in 1981. (Powell, 1994)

Semiparametric linear regression models believe that the dependent variable (y) depends on the independent variable (X) in a linear way, while it is not linearly related to the other independent variable (z). (Duran & Akdeniz, 2013)

The partial linear regression model (PLM) proposed by researchers (Robinson & Speckman) in (1988). (Speckman, 1988) (Al-Azzawi & Al-Always, 2022)

The (PLM) consists of a linear part represented by Parametric regression and a part Nonlinear represented by nonparametric regression. (AL-Adilee & Aboudi, 2021)

Can be written in the following form:

$$Y_i = \sum_{i=1}^{p} \beta_j X_{ij}\, g(Z_i) + \varepsilon_i \quad , \quad i = 1, 2, \dots, n \qquad (4)$$

In matrix form, the model can be rewritten as follows:

$$Y = X\beta + g(Z) + \varepsilon \qquad (5)$$

Where: $\mathbf{Y}$: Vector of response variable of degree (n×1).

$\mathbf{X\beta}$: Parametric part which contains:

$\mathbf{X}$: Explanatory variable of degree (n×p). $\boldsymbol{\beta}$: Vector of parameter of degree (p×1).

$\mathbf{g(Z)}$: Nonparametric part (smooth function unknown) of the degree (n×1).

$\mathbf{Z}$: Nonparametric variable (continuous variable) of degree (n×1).

$\boldsymbol{\varepsilon}$: Vector random errors (independently and identically distributed) of degree (n×1), with mean $E(\varepsilon) = 0$, and Fixed variance $Var(\varepsilon) = \sigma^2$. (Aydın, 2014)

3.1 Difference Method:

This method was proposed by the researcher (Yatchew) in (2003), which is used to estimate the parametric component of a (PLM) by removing the influence of the nonparametric component. (Yatchew, 2003)

The difference matrix of [(n-m)×n] rank. Difference matrix be written as: (Duran et al., 2012) (Duran & Akdeniz, 2013) (Wu, 2016)

$$D = \begin{bmatrix} d_0 & d_1 & d_2 & \cdots & d_m & 0 & \cdots & \cdots & 0 \\ 0 & d_0 & d_1 & d_2 & \cdots & d_m & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & d_0 & d_1 & d_2 & \cdots & d_m & 0 \\ 0 & 0 & \cdots & \cdots & d_0 & d_1 & d_2 & \cdots & d_m \end{bmatrix}$$

Where: $\mathbf{m}$: difference rank.

$(d_0, d_1, \dots, d_m)$: Weights that satisfy the following two conditions:

$\sum_{j=0}^{m} d_j = 0 \quad , \quad \sum_{j=0}^{m} d_j^2 = 1$ (Hussein, 2019)

The formula can be written as follows:

$$\tilde{Y} \approx \tilde{X}\beta + \tilde{\varepsilon} \qquad (6)$$

Where: $\widetilde{\mathbf{Y}} = \mathbf{DY}$, $\widetilde{\mathbf{X}} = \mathbf{DX}$, $\widetilde{\boldsymbol{\varepsilon}} = \mathbf{D}\boldsymbol{\varepsilon}$
$\widetilde{\mathbf{Y}}$: Vector of observations of the response variable of degree $[(\mathbf{n}-\mathbf{m})\times \mathbf{1}]$.
$\widetilde{\mathbf{X}}$: Matrix of observations of explanatory variables of degree $[(\mathbf{n}-\mathbf{m})\times \mathbf{p}]$.
$\boldsymbol{\beta}$: Vector of unknown parameters of degree $[\mathbf{p}\times \mathbf{1}]$.
$\mathbf{D}$: difference matrix of degree $[(\mathbf{n}-\mathbf{m})\times \mathbf{n}]$.
$\mathbf{n}$: Number of observations.
$\widetilde{\boldsymbol{\varepsilon}}$: vector of random errors of degree $[(\mathbf{n}-\mathbf{m})\times \mathbf{1}]$. (Akdeniz et al., 2015) (Tabakan & Akdeniz, 2010)
The first condition of the difference estimator guarantees the removal of the nonparametric effect in the regression equation, the second condition is that the variance of the residuals is not affected by removing the nonparametric effect. (Yatchew, 1997)

After reducing the nonparametric part, the (Yatchew) method was proposed to estimate the parameters by using the method of least squares depending on the difference estimator by increasing their energies: (Roozbeh et al., 2011)

$$\widehat{\boldsymbol{\beta}}_{\mathbf{diff}} = \left(\widetilde{\mathbf{X}}^{\mathbf{t}}\widetilde{\mathbf{X}}\right)^{-\mathbf{1}}\widetilde{\mathbf{X}}^{\mathbf{t}}\widetilde{\mathbf{Y}} \qquad (\mathbf{7})$$

As for the error variance depending on the difference estimator, it is written in the following form: (Duran & Akdeniz, 2013)

$$\boldsymbol{\sigma}^{\mathbf{2}}_{\mathbf{diff}} = \frac{\mathbf{1}}{\mathbf{n}}\left(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathbf{diff}}\right)^{\mathbf{t}}\mathbf{D}^{\mathbf{t}}\mathbf{D}\left(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathbf{diff}}\right) \qquad (\mathbf{8})$$

The researchers (Turkmen & Tabakan) in (2015) in the presence of outliers in the data of the semiparametric regression model employed the method of differences presented by (Yatchew) in (1997) with the robust methods such as the robust (MM) method in addition the smoothing spline estimator. Create a Proposed algorithm. (Turkmen & Tabakan, 2015)

In the study, the difference method will be employed with the robust methods (M, S) to model the semiparametric regression in the presence of outliers in addition to the existence of the problem of multicollinearity.

After the difference matrix was used to remove the nonparametric component from the semiparametric model, the new model will be relied upon in order to estimate the robust methods (M, S), depending on equation (6) of the difference method. So, the robust method (M, S) algorithms will be modified by multiplying the difference matrix by the explanatory variable's matrix X, and multiplying the difference matrix by the response variable vector Y.

3.1.1 Difference-Based Ridge Estimator:

This Ridge regression method has been discussed by the researchers (Hoerl and Kennard) in 1970, this is the most efficient way to deal with the problem of multilinearity, where this method gets rid of adding a small positive amount to the elements of the diameter of the information matrix. (Rashid & Hafez, 2013) (Husein, 2016) (Abboud & Khorshid, 2018) (Kamal & Khazal, 2019)

It was proposed by the two researchers (Tabakan & Akdeniz) in (2010), where they employed the difference method of the researcher (Yatchew) in the ridge regression, where the difference-based ridge estimator can be obtained, by relying on equation (6) of the difference method. (Tabakan & Akdeniz, 2010)

The formula for the difference-based ridge estimator can be written as follows:

$$\widehat{\boldsymbol{\beta}}^{\mathbf{diff}}_{\mathbf{Ridge}} = \left(\widetilde{\mathbf{X}}^{\mathbf{t}}\widetilde{\mathbf{X}} + \mathbf{KI}\right)^{-\mathbf{1}}\widetilde{\mathbf{X}}^{\mathbf{t}}\widetilde{\mathbf{Y}} \qquad (\mathbf{9})$$

Where: $\widetilde{\mathbf{Y}}$: Vector of observations of the response variable of degree $[(\mathbf{n}-\mathbf{m})\times \mathbf{1}]$.
$\widetilde{\mathbf{X}}$: Matrix of observations of explanatory variables of degree $[(\mathbf{n}-\mathbf{m})\times \mathbf{p}]$.
$\boldsymbol{\beta}$: Vector of unknown parameters of degree $[\mathbf{p}\times \mathbf{1}]$.
$\mathbf{D}$: difference matrix of degree $[(\mathbf{n}-\mathbf{m})\times \mathbf{n}]$.
$\mathbf{I}$: Unit matrix of degree (p×p).
$\mathbf{K}$: Bais parameter It is a constant value $\mathbf{K} > \mathbf{0}$.
$\mathbf{n}$: Number of observations.
$\mathbf{m}$: difference rank.
$\widetilde{\boldsymbol{\varepsilon}}$: vector of random errors of degree $[(\mathbf{n}-\mathbf{m})\times \mathbf{1}]$.

P-ISSN 2518-5764
E-ISSN 2227-703X

3.1.2 Difference-Based Ridge Estimator based on the M robust estimator:

An estimator that combines the ridge estimator and the M estimator was proposed by Silvapulle in (1991), In the case of the outliers and multicollinearity problems. Huber in (1981) is preferred the robust M-estimator to the LSE. In the event that the data suffers from the presence of outliers. The main formula for the difference ridge estimator based on M robust is written in the following: (Shih et al., 2021)

$$\hat{\beta}^{diff}_{Ridge}(M) = \left(\tilde{X}^t\tilde{X} + \hat{K}^{diff}_M I\right)^{-1}\tilde{X}^t\tilde{Y} \qquad (10)$$

Where the value of the robust bias parameter $\hat{K}^{diff}_M$ is calculated as follows:

$$\hat{K}^{diff}_M = \frac{p{\sigma^2}^{diff}_M}{\hat{\beta}^{diff^t}_M \hat{\beta}^{diff}_M} \quad , \quad \hat{\beta}^{diff}_M \neq 0 \qquad (11)$$

To calculate the variance of difference ridge estimator based on the M robust estimator: (Duran & Akdeniz, 2013)

$$ {\sigma^2}^{diff}_M = \frac{1}{n}\left(Y - X\hat{\beta}^{diff}_M\right)^t D^t D\left(Y - X\hat{\beta}^{diff}_M\right) \qquad (12)$$

3.1.3 Difference-Based Ridge Estimator based on the S robust estimator:

It is assumed that the robust estimator parameter obtained using the S estimator is the Ridge Estimator based on the S estimator. (Jeremia et al., 2020)

The main formula for the difference ridge estimator based on the S robust it is written in the following form: (Duran & Akdeniz, 2013)

$$\hat{\beta}^{diff}_{Ridge}(S) = \left(\tilde{X}^t\tilde{X} + \hat{K}^{diff}_S I\right)^{-1}\tilde{X}^t\tilde{Y} \qquad (13)$$

Where the value of the robust bias parameter $\hat{K}^{diff}_S$ is calculated as follows:

$$\hat{K}^{diff}_S = \frac{p{\sigma^2}^{diff}_S}{\hat{\beta}^{diff^t}_S \hat{\beta}^{diff}_S} \quad , \quad \hat{\beta}^{diff}_S \neq 0 \qquad (14)$$

To calculate the variance of difference ridge estimator based on S robust estimator:

$$ {\sigma^2}^{diff}_S = \frac{1}{n}\left(Y - X\hat{\beta}^{diff}_S\right)^t D^t D\left(Y - X\hat{\beta}^{diff}_S\right) \qquad (15)$$

4. Nonparametric Estimation Method:

Nonparametric regression was proposed by (Jacob Wolfowitz) in 1942. (Kvam et al., 2022)

In nonparametric models, knowledge of the data distribution is not required, and these models never contain parameters. Where the relationship between the explanatory variables and the response variable is not known. (Mahmoud, 2019)

The general formula for nonparametric regression can be written as: (Ali et al., 2020) (Hameed & Khalaf, 2021)

$$Y_i = g(Z_i) + \varepsilon_i \quad , \quad i = 1, 2, \dots, n \qquad (16)$$

Where: $g(Z_i)$: The unknown smoothing function.

4.1 Kernel function (Nadaraya – Watson estimator) (N.W):

The estimator (Nadaraya-Watson) was used extensively in many areas of statistical research, and is the simplest type of the smoothers. It was proposed by researchers (Nadaraya) and (Watson) in (1964), is an important estimator for estimating a function (Kernel). (Hameed & Khalaf, 2021)

The properties of the kernel function that are used with the (NW) estimator can be written:

$$(1) \int k(z)dz = 1 \;,\; (2) \int zk(z)dz = 0 \;,\; (3) \int z^2k(z)dz = 0 \;,\; \forall z = 1, 3, \dots, k-1$$

(Demir & Toktamiş, 2010) (Al-Tai & Al-Kazaz, 2022)

The general formula for the (N.W) estimator can be written as: (Härdle et al., 2004)

P-ISSN 2518-5764
E-ISSN 2227-703X

$$\hat{g}(Zi) = \frac{n^{-1}\sum_{i=1}^{n} K_h(z - Z_i)Y_i^*}{n^{-1}\sum_{i=1}^{n} K_h(z - Z_i)} \qquad (17)$$

Equation (17) can be rewritten in the form of matrices in the following form: (Khalaf & Mohammed, 2023)

$$\hat{g}(Zi) = W_h(z)Y_i^* \qquad (18)$$

Where: $Y_i^* = Y_i - X_i\hat{\beta}^{diff}_{Ridge(M\ or\ S)}$

h : bandwidth or smoothing parameter. $W_h(z)$: the weight function of (N.W) estimator.

4.2 Cubic Smoothing Spline estimator:

The smoothing spline is attributed to the researcher (Whittaker) in (1923). (Härdle, 1990)
Smoothing Spline method is used to estimate the regression model, penalized least squares method can be used to estimate the smoothing function.

$$S(g) = \arg\min_{f(.)}\left[\sum_{i=1}^{n}\left(y_i - x'_i\beta - g(z_i)\right)^2 + \lambda\int_a^b \left(g''(z)\right)^2 dz\right] \qquad (19)$$

Where: λ: Smoothing Parameter.
$g(z_i)$: represents the smoothing spline estimator for a roughness penalty.
$g''(z)$: The second derivative of the smoothing function. (Aydin, 2007b) (Hmood & Katea, 2014)

Idea of estimating method depends on minimization two main parts to obtain the curve best. First part is sum of squares of residuals, the second part is roughness penalty or the penalty term. (Habeeb et al., 2021)

As $\lambda \to 0$ the roughness penalty dominates in (1) and the spline estimate interpolates the data.
As $\lambda \to \infty$ the roughness penalty dominates in (1) and the spline estimate is forced to be a constant.
(Aydin, 2007a) (Aydin et al., 2013)

The difference between smoothing splines (knots = n), regression splines (knots < n) and penalized regression splines (regression splines with penalization for the number of knots) lies in the number of knots chosen. (Hens, 2005) (Hmood & Burhan, 2018)
Suppose given n real numbers $(z_1, z_2, \dots, z_n)$ in the interval [a,b].

A function (g) in the interval [a,b] as cubic spline if two conditions are satisfied:
1) In each interval $(a, z_1), (z_1, z_2) \dots\dots\dots\dots\dots (z_n, b)$, the function g is polynomial cubic spline.
2) The polynomial pieces fit together at point z_i in such a way g, g' and g'' are contiguous at each z_i, The function g is contiguous in [a,b]. (Ibrahim & Suliadi, 2010)

Depending on the matrix formula, the estimator can be represented as in the following formula:

$$\hat{g}_\lambda = S_\lambda \underline{y} \qquad (20)$$

Where: λ : Parameter of smoothing Spline.
We get an estimate of $\hat{g}$ using the Cubic Smoothing Spline method for the value Y^* of the nonparametric part, so the estimate is as follows:

$$\hat{g} = S_\lambda\left(y_i - x'_i\hat{\beta}^{diff}\right) \qquad (21)$$

$$\hat{g} = S_\lambda y_i^* \qquad (22)$$

Where: $y_i^* = y_i - x'_i\hat{\beta}^{diff}_{ridge(M\ or\ S)}$
S_λ : Smoothing matrix (definite and nonnegative and symmetric) (n×n) that depends on the value of λ and the values of z_i and does not depend on the values of y_i. (Bickel et al., 2009)
The smoothing parameter (λ) will be selected by using the cross-validation (CV) method, as this method is one of the most used and highly efficient methods. Which is used in the (NW) method and the smoothing spline method. (Al-Azzawi & Al-Always, 2022)

5. Simulation:

Simulations for this study were tested using (MATLAB) language in order to generate simulation data to compare methods (DRMNW, DRSNW, DRMSP, DRSSP) with different sample sizes ($\mathbf{n_1 = 50, n_2 = 100, n_3 = 150}$) after assuming three outliers in the response variables ($\boldsymbol{\tau_1 = 10\%, \tau_2 = 20\%, \tau_3 = 30\%}$), five correlation ratios ($\boldsymbol{\rho_1 = 0.50, \rho_2 = 0.60, \rho_3 = 0.70, \rho_4 = 0.80, \rho_5 = 0.90}$) and four parameters ($\boldsymbol{\beta_1 = 1.5, \beta_2 = -1.5, \beta_3 = 1, \beta_4 = 2}$) and four explanatory variables (X1, X2, X3, X4, Z) being generated using the method (Box-Muller), each experiment was repeated 500 time.

The mean absolute percentage error (MAPE) scale was used, which is the most widely used measure for error prediction. It measures accuracy as a percentage. It can be calculated through the following equation:

$$\mathbf{MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right|}$$

Models used to generate the nonparametric component (Z):

1) $\hat{\mathbf{g}}(\mathbf{Zi}) = \mathbf{0.5\,\sin(2\,\pi\,Z)}$
2) $\hat{\mathbf{g}}(\mathbf{Zi}) = \mathbf{\sin(2Z) + 2e^{(-16Z^2)}}$
3) $\hat{\mathbf{g}}(\mathbf{Zi}) = \mathbf{e^{(-(Z-0.5)^2)}}$

6. Analyzing the results:

The results of the first semiparametric partial linear regression model:

Table 1: The mean absolute percentage error (MAPE), when pollution percentage with outlier 10%

n	ρ	DRMNW	DRMSP	DRSNW	DRSSP	Best
50	0.5	0.552919	1.114056	0.938103	4.232229	DRMNW
	0.6	0.556217	1.107696	0.554606	9.31E-05	DRSSP
	0.7	0.551099	1.108937	0.443173	3.78E-06	DRSSP
	0.8	0.553149	1.1104	0.350917	2.93E-07	DRSSP
	0.9	0.559684	1.107422	0.320467	2.89E-07	DRSSP
100	0.5	0.536971	1.073228	0.921005	0.38337	DRSSP
	0.6	0.538642	1.072889	0.55377	5.37E-07	DRSSP
	0.7	0.533202	1.073854	0.445453	6.92E-09	DRSSP
	0.8	0.535658	1.068288	0.351682	5.57E-11	DRSSP
	0.9	0.537732	1.071636	0.326029	1.32E-11	DRSSP
150	0.5	0.526016	1.054253	0.914008	0.075336	DRSSP
	0.6	0.528341	1.053823	0.554046	6.1E-09	DRSSP
	0.7	0.525412	1.050166	0.444231	8.95E-12	DRSSP
	0.8	0.527886	1.052417	0.352265	1.95E-14	DRSSP
	0.9	0.524936	1.051357	0.329051	3.61E-14	DRSSP

The results of table (1) showed outlier of (10%) in relation to a sample size (n=50) and at a correlation coefficient level (ρ=0.50), the best estimator is (DRMNW) because it has the lower of (MAPE), and at the rest of the levels of the correlation coefficient, the best estimator is (DRSSP), but at the sample size (n=100,150) and for all levels of the correlation coefficient, the best estimator is (DRSSP), which is better than the rest of the estimators.

Table 2: The mean absolute percentage error (MAPE), when pollution percentage with outlier 20%

n	ρ	DRMNW	DRMSP	DRSNW	DRSSP	Best
50	0.5	0.323217	1.257146	0.8839	4047.02	DRMNW
	0.6	0.328181	1.240443	0.314351	1.35E-07	DRSSP
	0.7	0.319752	1.241572	0.203029	3.76E-10	DRSSP
	0.8	0.324601	1.250197	0.130638	7.86E-12	DRSSP
	0.9	0.330937	1.240854	0.109675	4.52E-11	DRSSP
100	0.5	0.298528	1.157944	0.849626	11.8779	DRMNW
	0.6	0.30121	1.156681	0.310819	7.08E-12	DRSSP
	0.7	0.295009	1.159411	0.203149	2.16E-15	DRSSP
	0.8	0.297925	1.146402	0.128393	3.89E-19	DRSSP
	0.9	0.300562	1.154443	0.111073	1.37E-20	DRSSP
150	0.5	0.284705	1.114212	0.836197	0.139037	DRSSP
	0.6	0.286998	1.113642	0.310066	2.47E-15	DRSSP
	0.7	0.283468	1.105568	0.200819	8.75E-21	DRSSP
	0.8	0.286029	1.110462	0.127499	7.38E-26	DRSSP
	0.9	0.282694	1.108225	0.111612	1.05E-24	DRSSP

The results of table (2) showed outlier of (20%) with respect to a sample size (n = 50) and at a correlation coefficient level (ρ=0.50), the best estimator is (DRMNW) because it has the lower of (MAPE), and at the rest of the correlation coefficient levels the best estimator is (DRSSP), but at the sample size (n=100) and at the level of correlation coefficient (ρ=0.50) the best estimator is (DRMNW), and at the rest of the correlation coefficient levels the best estimator is (DRSSP), and at the sample size (n=150) and for all levels of the correlation coefficient, the best estimator is (DRSSP), which is better than the rest of the estimators.

Table 3: The mean absolute percentage error (MAPE), when pollution percentage with outlier values is 30%

n	ρ	DRMNW	DRMSP	DRSNW	DRSSP	Best
50	0.5	0.552919	1.114056	0.938103	4.232229	DRMNW
	0.6	0.556217	1.107696	0.554606	9.31E-05	DRSSP
	0.7	0.551099	1.108937	0.443173	3.78E-06	DRSSP
	0.8	0.553149	1.1104	0.350917	2.93E-07	DRSSP
	0.9	0.559684	1.107422	0.320467	2.89E-07	DRSSP
100	0.5	0.536971	1.073228	0.921005	0.38337	DRSSP
	0.6	0.538642	1.072889	0.55377	5.37E-07	DRSSP
	0.7	0.533202	1.073854	0.445453	6.92E-09	DRSSP
	0.8	0.535658	1.068288	0.351682	5.57E-11	DRSSP
	0.9	0.537732	1.071636	0.326029	1.32E-11	DRSSP
150	0.5	0.526016	1.054253	0.914008	0.075336	DRSSP
	0.6	0.528341	1.053823	0.554046	6.1E-09	DRSSP
	0.7	0.525412	1.050166	0.444231	8.95E-12	DRSSP
	0.8	0.527886	1.052417	0.352265	1.95E-14	DRSSP
	0.9	0.524936	1.051357	0.329051	3.61E-14	DRSSP

The results of table (3) showed outlier of (30%) for a sample size (n=50) and at a correlation coefficient level (ρ=0.50), the best estimator is (DRMNW) because it has the lower of (MAPE), and at the rest of the levels of the correlation coefficient, the best estimator is (DRSSP), but at the sample size (n=100,150) and for all levels of the correlation coefficient, the best estimator is (DRSSP), which is better than the rest of the estimators.

The results of the second semiparametric partial linear regression model:

Table 4: The mean absolute percentage error (MAPE), pollution percentage with outlier 10%

n	ρ	DRMNW	DRMSP	DRSNW	DRSSP	Best
50	0.5	1.112185	1.05045	0.825251	1.085766	DRSNW
	0.6	1.103272	1.051371	0.300005	4.76E-08	DRSSP
	0.7	1.105412	1.056769	0.194536	8.8E-10	DRSSP
	0.8	1.119472	1.0519	0.12019	4.92E-13	DRSSP
	0.9	1.108578	1.051376	0.105819	1.1E-13	DRSSP
100	0.5	1.074086	1.033507	0.815889	0.007441	DRSSP
	0.6	1.071341	1.034178	0.300725	1.66E-12	DRSSP
	0.7	1.063751	1.032332	0.195815	1.26E-15	DRSSP
	0.8	1.081894	1.036636	0.122132	1.38E-19	DRSSP
	0.9	1.068827	1.03259	0.107872	2.78E-20	DRSSP
150	0.5	1.040681	1.025228	0.813895	0.004114	DRSSP
	0.6	1.048629	1.025725	0.304076	2.24E-16	DRSSP
	0.7	1.0471	1.024753	0.196779	3.1E-21	DRSSP
	0.8	1.056334	1.025245	0.123691	5.09E-26	DRSSP
	0.9	1.048074	1.026832	0.109693	1.45E-27	DRSSP

The results of table (4) showed outlier of (10%) in relation to a sample size (n=50) and at a correlation coefficient level (ρ=0.50), the best estimator is (DRSNW) because it has the lower of (MAPE), and at the rest of the levels of the correlation coefficient, the best estimator is (DRSSP), but at the sample size (n=100,150) and for all levels of the correlation coefficient, the best estimator is (DRSSP), which is better than the rest of the estimators.

Table 5: The mean absolute percentage error (MAPE), when pollution percentage with outlier 20%

n	ρ	DRMNW	DRMSP	DRSNW	DRSSP	Best
50	0.5	1.308956	1.106296	0.68479	1046.064	DRSNW
	0.6	1.289023	1.108279	0.096645	4.18E-13	DRSSP
	0.7	1.296947	1.120491	0.043292	4.05E-16	DRSSP
	0.8	1.324507	1.109389	0.017543	1.35E-22	DRSSP
	0.9	1.297891	1.108456	0.013743	8.67E-24	DRSSP
100	0.5	1.197001	1.069236	0.667504	0.000755	DRSSP
	0.6	1.188673	1.070744	0.09466	1.92E-22	DRSSP
	0.7	1.172933	1.066774	0.041697	3.94E-28	DRSSP
	0.8	1.213785	1.075988	0.017128	7.97E-36	DRSSP
	0.9	1.183279	1.067335	0.013503	6.38E-37	DRSSP
150	0.5	1.107317	1.05187	0.663614	0.005586	DRSSP
	0.6	1.129077	1.05274	0.0958	4.46E-30	DRSSP
	0.7	1.125009	1.050788	0.041195	1.87E-39	DRSSP
	0.8	1.143236	1.051808	0.016856	1.92E-48	DRSSP
	0.9	1.12816	1.055038	0.01351	1.46E-51	DRSSP

The results of table (5) showed outlier of (20%) for a sample size (n=50) and at a correlation coefficient level (ρ=0.50), the best estimator is (DRSNW) because it has the lower of (MAPE), and at the rest of the levels of the correlation coefficient, the best estimator is (DRSSP), but at the sample size (n=100,150) and for all levels of the correlation coefficient, the best estimator is (DRSSP), which is better than the rest of the estimators.

Table 6: The mean absolute percentage error (MAPE), when pollution percentage with outlier 30%

n	ρ	DRMNW	DRMSP	DRSNW	DRSSP	Best
50	0.5	1.112185	1.05045	0.825251	1.085766	DRSNW
	0.6	1.103272	1.051371	0.300005	4.76E-08	DRSSP
	0.7	1.105412	1.056769	0.194536	8.8E-10	DRSSP
	0.8	1.119472	1.0519	0.12019	4.92E-13	DRSSP
	0.9	1.108578	1.051376	0.105819	1.1E-13	DRSSP
100	0.5	1.074086	1.033507	0.815889	0.007441	DRSSP
	0.6	1.071341	1.034178	0.300725	1.66E-12	DRSSP
	0.7	1.063751	1.032332	0.195815	1.26E-15	DRSSP
	0.8	1.081894	1.036636	0.122132	1.38E-19	DRSSP
	0.9	1.068827	1.03259	0.107872	2.78E-20	DRSSP
150	0.5	1.040681	1.025228	0.813895	0.004114	DRSSP
	0.6	1.048629	1.025725	0.304076	2.24E-16	DRSSP
	0.7	1.0471	1.024753	0.196779	3.1E-21	DRSSP
	0.8	1.056334	1.025245	0.123691	5.09E-26	DRSSP
	0.9	1.048074	1.026832	0.109693	1.45E-27	DRSSP

The results of table (6) showed outlier of (30%) in relation to a sample size (n=50) and at a correlation coefficient level (ρ=0.50), the best estimator is (DRSNW) because it has the lower of (MAPE), and at the rest of the levels of the correlation coefficient, the best estimator is (DRSSP), but at the sample size (n=100,150) and for all levels of the correlation coefficient, the best estimator is (DRSSP), which is better than the rest of the estimators.

The results of the third semiparametric partial linear regression model:

Table 7: The mean absolute percentage error (MAPE), when pollution percentage with outlier 10%

n	ρ	DRMNW	DRMSP	DRSNW	DRSSP	Best
50	0.5	1.33765	1.04237	0.785873	0.09263	DRSSP
	0.6	1.338779	1.043765	0.23144	2.33E-09	DRSSP
	0.7	1.33792	1.044547	0.136992	3.34E-12	DRSSP
	0.8	1.343945	1.041797	0.080227	2.41E-15	DRSSP
	0.9	1.331647	1.046576	0.070064	6.97E-14	DRSSP
100	0.5	1.285518	1.027823	0.778524	0.002175	DRSSP
	0.6	1.294201	1.027889	0.234223	3.88E-14	DRSSP
	0.7	1.281832	1.028904	0.141192	3.83E-18	DRSSP
	0.8	1.280866	1.030257	0.083846	2.28E-23	DRSSP
	0.9	1.291006	1.026995	0.068378	9.69E-25	DRSSP
150	0.5	1.253946	1.021178	0.77695	0.00028	DRSSP
	0.6	1.258299	1.021688	0.239028	9.86E-19	DRSSP
	0.7	1.258365	1.020909	0.142863	2.89E-25	DRSSP
	0.8	1.264872	1.021876	0.083208	1.61E-31	DRSSP
	0.9	1.273791	1.022553	0.069507	2.99E-34	DRSSP

The results of table (7) showed outlier of (10%) and for all sample sizes (n=50,100,150) and for all levels of the correlation coefficient, the best estimator is (DRSSP) because it has the lower of (MAPE) of the rest of the estimators.

Table 8: The mean absolute percentage error (MAPE), when pollution percentage with outlier 20%

n	ρ	DRMNW	DRMSP	DRSNW	DRSSP	Best
50	0.5	1.894762	1.088312	0.621494	4.610488	DRSNW
	0.6	1.887837	1.091418	0.059188	1.17E-15	DRSSP
	0.7	1.88786	1.093134	0.022362	2.12E-21	DRSSP
	0.8	1.918725	1.087207	0.008324	2.96E-27	DRSSP
	0.9	1.876202	1.09785	0.006745	4.74E-24	DRSSP
100	0.5	1.715983	1.057177	0.608271	6.08E-05	DRSSP
	0.6	1.737842	1.057431	0.05865	3.06E-25	DRSSP
	0.7	1.702814	1.059571	0.022441	7.77E-33	DRSSP
	0.8	1.702536	1.062404	0.008486	1.02E-43	DRSSP
	0.9	1.723562	1.055451	0.005694	2.65E-46	DRSSP
150	0.5	1.620291	1.043278	0.605206	5.83E-06	DRSSP
	0.6	1.622051	1.044327	0.059843	7.28E-34	DRSSP
	0.7	1.629796	1.042683	0.022352	2.11E-47	DRSSP
	0.8	1.644565	1.044704	0.008004	9.7E-60	DRSSP
	0.9	1.6719	1.046153	0.00572	5.76E-65	DRSSP

The results of table (8) showed outlier of (20%) for a sample size (n=50) and at a correlation coefficient level (ρ=0.50), the best estimator is (DRSNW) because it has the lower of (MAPE), and at the rest of the levels of the correlation coefficient, the best estimator is (DRSSP), but at sample sizes (n=100,150) and for all levels of the correlation coefficient, the best estimator is (DRSSP), which is better than the rest of the estimators.

Table 9: The mean absolute percentage error (MAPE), when pollution percentage with outlier 30%

n	***ρ***	**DRMNW**	**DRMSP**	**DRSNW**	**DRSSP**	**Best**
50	0.5	1.33765	1.04237	0.785873	0.09263	DRSSP
	0.6	1.338779	1.043765	0.23144	2.33E-09	DRSSP
	0.7	1.33792	1.044547	0.136992	3.34E-12	DRSSP
	0.8	1.343945	1.041797	0.080227	2.41E-15	DRSSP
	0.9	1.331647	1.046576	0.070064	6.97E-14	DRSSP
100	0.5	1.285518	1.027823	0.778524	0.002175	DRSSP
	0.6	1.294201	1.027889	0.234223	3.88E-14	DRSSP
	0.7	1.281832	1.028904	0.141192	3.83E-18	DRSSP
	0.8	1.280866	1.030257	0.083846	2.28E-23	DRSSP
	0.9	1.291006	1.026995	0.068378	9.69E-25	DRSSP
150	0.5	1.253946	1.021178	0.77695	0.00028	DRSSP
	0.6	1.258299	1.021688	0.239028	9.86E-19	DRSSP
	0.7	1.258365	1.020909	0.142863	2.89E-25	DRSSP
	0.8	1.264872	1.021876	0.083208	1.61E-31	DRSSP
	0.9	1.273791	1.022553	0.069507	2.99E-34	DRSSP

The results of table (9) showed outlier of (30%) and for all sample sizes (n=50,100,150) and for all levels of the correlation coefficient, the best estimator is (DRSSP) because it has the lower of (MAPE) of the rest of the estimators.

7.Conclusion:

1) In the first, second and third semiparametric partial linear regression models (when the level of the correlation coefficient is low and the sample size is small and for all outliers), the (DRMNW) estimator is the best estimator, but for the rest of the correlation levels, for all sample sizes, for all pollution rates and for all models, the (DRSSP) estimator is the best estimator.

2) The values of the mean absolute percentage error (MAPE) decrease when the sample size is increased (an inverse relationship).

3) We conclude that in all models, for all sample sizes, and for all levels of the correlation coefficient, the best estimator is (DRSSP).

Authors Declaration:

Conflicts of Interest: None

-We Hereby Confirm That All The Figures and Tables In The Manuscript Are Mine and Ours. Besides, The Figures and Images, Which are Not Mine, Have Been Permitted Republication and Attached to The Manuscript.

- Ethical Clearance: The Research Was Approved By The Local Ethical Committee in The University.

References:

1. Abbas, H. H., & Abood, S. N. (2022). Comparison of Robust Circular S and Circular Least Squares Estimators for Circular Regression Model using Simulation. Journal of Economics and Administrative Sciences, 28(134), 168–185.

2. Abboud, S. N., & Khorshid, E. S. (2018). Comparison between the Methods of Ridge Regression and Liu Type to Estimate the Parameters of the Negative Binomial Regression Model Under Multicollinearity Problem by Using Simulation. Journal of Economics And Administrative Sciences, 24(109).

3. Akdeniz, F., Duran, E. A., Roozbeh, M., & Arashi, M. (2015). Efficiency of the generalized difference-based Liu estimators in semiparametric regression models with correlated errors. Journal of Statistical Computation and Simulation, 85(1), 147–165.

4. AL-Adilee, R. T. K., & Aboudi, E. H. (2021). ESTIMATED NON-PARAMETRIC AND SEMI-PARAMETRIC MODEL FOR LONGITUDINAL DATA. International Journal of Agricultural & Statistical Sciences, 17.

5. Al-Azzawi, E. A., & Al-Always, L. A. (2022). Robust Estimation OF The Partial Regression Model Using Wavelet Thresholding. Journal of Economics and Administrative Sciences, 28(133), 97–113.

6. Ali, O. A., Naji, M. Q., & Ismaeel, M. M. (2020). Kernel estimation of returns of retirement funds of employers based on monetary earnings (subscriptions and compensation) via regression discontinuity in Iraq. Periodicals of Engineering and Natural Sciences, 8(3), 1752–1766.

7. Alma, Ö. G. (2011). Comparison of robust regression methods in linear regression. Int. J. Contemp. Math. Sciences, 6(9), 409–421.

8. Al-Tai, A. A., & Al-Kazaz, Q. N. N. (2022). Semi parametric Estimators for Quantile Model via LASSO and SCAD with Missing Data. Journal of Economics and Administrative Sciences, 28(133), 82–96.

9. Aydin, D. (2007a). A comparison of the nonparametric regression models using smoothing spline and kernel regression. World Academy of Science, Engineering and Technology, 36, 253–257.

10. Aydin, D. (2007b). Estimation of GDP in Turkey by nonparametric regression models. Proceedings of the 6th WSEAS International Conference on Instrumentation, Measurement, Circuits and Systems, 221–225.

11. Aydin, D., Memmedli, M., & Omay, R. E. (2013). Smoothing parameter selection for nonparametric regression using smoothing spline. European Journal of Pure and Applied Mathematics, 6(2), 222–238.

12. Aydın, D. (2014). Partially linear models based on smoothing spline estimated by different selection methods: a simulation study. Pakistan Journal of Statistics, 30(1), 35–56.

13. Bahez, Z. K., & Rasheed, H. A. (2022). Comparing Some of Robust the Non-Parametric Methods for Semi-Parametric Regression Models Estimation. Journal of Economics and Administrative Sciences, 28(132), 105–117.

14. Bickel, P., Diggle, P., Fienberg, S., Gather, U., Olkin, I., & Zeger, S. (2009). Springer series in statistics. Springer.

15. Chen, C. (2002). Paper 265-27 Robust regression and outlier detection with the ROBUSTREG procedure. Proceedings of the Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference.

16. Demir, S., & Toktamiş, Ö. (2010). On the adaptive Nadaraya-Watson kernel regression estimators. Hacettepe Journal of Mathematics and Statistics, 39(3), 429–437.

17. Duran, E. A., & Akdeniz, F. (2013). New difference-based estimator of parameters in semiparametric regression models. Journal of Statistical Computation and Simulation, 83(5), 810–824.

P-ISSN 2518-5764
E-ISSN 2227-703X

18. Duran, E. A., Härdle, W. K., & Osipenko, M. (2012). Difference based ridge and Liu type estimators in semiparametric regression models. Journal of Multivariate Analysis, 105(1), 164–175.

19. Habeeb, A. S., Hmood, M. Y., & Mohammed, M. J. (2021). ESTIMATING NONPARAMETRIC AUTOREGRESSIVE CURVE BY SMOOTHING SPLINES METHOD. International Journal of Agricultural & Statistical Sciences, 17(2).

20. Hameed, L. M. A., & Khalaf, K. I. (2021). COMPARISON OF SOME NONPARAMETRIC METHODS TO DETERMINE THE NUMBER OF RADIATION DOSES FOR BREAST CANCER PATIENTS. International Journal of Agricultural & Statistical Sciences, 17.

21. Härdle, W. (1990). Applied nonparametric regression (Issue 19). Cambridge university press.

22. Härdle, W., Müller, M., Sperlich, S., & Werwatz, A. (2004). Nonparametric and semiparametric models (Vol. 1). Springer.

23. Hens, N. (2005). Non-and semi-parametric techniques for handling missing data. Limburgs Universitair Centrum (Belgium).

24. Herawati, N., Saidi, S., & Azis, D. (2022). RIDGE LEAST ABSOLUTE DEVIATION PERFORMANCE IN ADDRESSING MULTICOLLINEARITY AND DIFFERENT LEVELS OF OUTLIER SIMULTANEOUSLY. Barekeng: Jurnal Ilmu Matematika Dan Terapan, 16(3), 779–786.

25. Hmood, M. Y., & Burhan, Y. K. (2018). comparison between the methods estimate nonparametric and semiparametric transfer function model in time series the Using simulation. Journal of Economics And Administrative Sciences, 24(106).

26. Hmood, M. Y., & Katea, M. M. (2014). A comparison of the Semiparametric Estimators model using smoothing methods different. Journal of Economics And Administrative Sciences, 20(75).

27. Husein, S. M. (2016). A Simulation study to comparison some biased estimators. International Journal of Advanced Research in Engineering and Technology, 7(4), 57–64.

28. Hussein, S. M. (2019). Comparison of Some Suggested Estimators Based on Differencing Technique in the Partial Linear Model Using Simulation. Baghdad Science Journal, 16(4), 918.

29. Ibrahim, N. A., & Suliadi, S. (2010). GEE-smoothing spline in semiparametric model with correlated nominal data: estimation and simulation study. Proceedings of the 4th International Conference on Applied Mathematics, Simulation, Modelling, 19–26.

30. Irshayyid, A. J., & Saleh, R. A. (2022). Robust Estimates for One-Parameter Exponential Regression Model. Journal of Economics and Administrative Sciences, 28(134), 147–159.

31. Jeremia, N. E., Nurrohmah, S., & Fithriani, I. (2020). Robust Ridge regression to solve a multicollinearity and outlier. Journal of Physics: Conference Series, 1442(1), 12030.

32. Kamal, G. I., & al-Imam Saadi Khazal, S. (2019). Using Some Robust Methods For Handling the Problem of Multicollinearity. Journal of Economics and Administrative Sciences, 25(112), 500–514.

33. Khalaf, N. B., & Mohammed, L. A. (2023). Comparison of Some Methods for Estimating Nonparametric Binary Logistic Regression. Journal of Economics and Administrative Sciences, 29(135), 56–67.

34. Kvam, P., Vidakovic, B., & Kim, S. (2022). Nonparametric statistics with applications to science and engineering with R (Vol. 1). John Wiley & Sons.

35. Lukman, A., Arowolo, O., & Ayinde, K. (2014). Some robust ridge regression for handling multicollinearity and outlier. International Journal of Sciences: Basic and Applied Research, 16(2), 192–202.

36. Lukman, A. F., Osowole, O. I., & Ayinde, K. (2015). Two stage robust ridge method in a linear regression model. Journal of Modern Applied Statistical Methods, 14(2), 8.

P-ISSN 2518-5764
E-ISSN 2227-703X

37. Mahmoud, H. F. F. (2019). Parametric versus semi and nonparametric regression models. ArXiv Preprint ArXiv:1906.10221.
38. Powell, J. L. (1994). Estimation of semiparametric models. Handbook of Econometrics, 4, 2443–2521.
39. Rashid, D. H., & Hafez, A. S. A. (2013). Robust Two-Step Estimation and Approximation Local Polynomial Kernel For Time-Varying Coefficient Model With Balance Longitudinal Data. Journal of Economics And Administrative Sciences, 19(70).
40. Roozbeh, M. (2016). Robust ridge estimator in restricted semiparametric regression models. Journal of Multivariate Analysis, 147, 127–144.
41. Roozbeh, M., Arashi, M., & Niroumand, H. A. (2011). Ridge regression methodology in partial linear models with correlated errors. Journal of Statistical Computation and Simulation, 81(4), 517–528.
42. ŞAMKAR, H., & ALPU, Ö. (2010). Ridge regression based on some robust estimators. Journal of Modern Applied Statistical Methods, 9(2).
43. Shih, J.-H., Lin, T.-Y., Jimichi, M., & Emura, T. (2021). Robust ridge M-estimators with pretest and Stein-rule shrinkage for an intercept term. Japanese Journal of Statistics and Data Science, 4, 107–150.
44. Speckman, P. (1988). Kernel smoothing in partial linear models. Journal of the Royal Statistical Society Series B: Statistical Methodology, 50(3), 413–436.
45. Susanti, Y., Pratiwi, H., Sulistijowati, S., & Liana, T. (2014). M estimation, S estimation, and MM estimation in robust regression. International Journal of Pure and Applied Mathematics, 91(3), 349–360.
46. Tabakan, G., & Akdeniz, F. (2010). Difference-based ridge estimator of parameters in partial linear model. Statistical Papers, 51(2), 357–368.
47. Turkmen, A. S., & Tabakan, G. (2015). Outlier resistant estimation in difference-based semiparametric partially linear models. Communications in Statistics-Simulation and Computation, 44(2), 417–432.
48. Wu, J. (2016). Difference-based ridge-type estimator of parameters in restricted partial linear model with correlated errors. SpringerPlus, 5(1), 1–10.
49. Yatchew, A. (1997). An elementary estimator of the partial linear model. Economics Letters, 57(2), 135–143.
50. Yatchew, A. (2003). Semiparametric regression for the applied econometrician. Cambridge University Press.

Journal of Economics and Administrative Sciences
2024; 30(141), pp. 479-496
P-ISSN 2518-5764
E-ISSN 2227-703X

مقارنة بين بعض الطرائق الحصينة لمقدر الحرف المبني على الفروق في نماذج الانحدار شبه المعلمية

وسام خالد سعيد [1]
جامعة بغداد / كلية الادارة والاقتصاد/ قسم الاحصاء
wissam.k.saed.1988@gmail.com

رباب عبد الرضا صالح[2]
جامعة بغداد / كلية الادارة والاقتصاد/ قسم الاحصاء
rabab.saleh@coadec.uobaghdad.edu.iq

Received:27/7/2023 Accepted:13/9/2023 Published Online First: 30 /6/ 2024

مستخلص البحث:

يحظى موضوع نماذج الانحدار شبه المعلمية بأهتمام كبير من قبل الباحثين في الآونة الأخيرة كونه يدمج بين الطرائق المعلمية والطرائق اللامعلمية، وهو من المواضيع المتقدمة في تحليل البيانات لمختلف الدراسات والذي يهدف الى إيجاد افضل مقدرات وذات مستوى كفاءة عالي.

ومن أهم نماذج الانحدار شبه المعلمية أنموذج الانحدار الخطي الجزئي (PLM) والذي يتكون من مركبة معلمية ومركبة لامعلمية ، ولغرض تقدير المركبة المعلمية سيتم أستخدام تقنية الفروق من أجل إزالة المركبة اللامعلمية . وعند عدم تحقق فروض التحليل الخاصة بالمركبة المعلمية سوف تعاني مشاكل عدة ومن أهمها مشكلة التعدد الخطي التام بالإضافة الى وجود القيم الشاذة في البيانات .

تم في هذا البحث معالجة مشكلتي التعدد الخطي والقيم الشاذة لأنموذج الأنحدار شبه المعلمي ، حيث تم أستخدام المحاكاة لتوليد البيانات وبأحجام عينات مختلفة ولنسب أرتباطات وتلويث مختلفة ولطرائق مختلفة مثل (مقدر الحرف المبني على الفروق بالأعتماد على مقدر M الحصين مع مقدر ندارايا واتسن (DRMNW)، مقدر الحرف المبني على الفروق بالأعتماد على مقدر M الحصين مع مقدر الشرائح التمهيدية (DRMSP) ، مقدر الحرف المبني على الفروق بالأعتماد على مقدر S الحصين مع مقدر ندارايا واتسن (DRSNW) ، مقدر الحرف المبني على الفروق بالأعتماد على مقدر S الحصين مع مقدر الشرائح التمهيدية (DRMSP)) وبأستخدام معيار متوسط نسبة الخطأ المطلق (MAPE) ، وأظهرت النتائج أن طريقة مقدر الحرف المبني على الفروق بالأعتماد على مقدر S الحصين مع مقدر الشرائح التمهيدية (DRMSP) هي الأفضل .

نوع البحث: ورقة بحثية .

المصطلحات الرئيسة للبحث: نماذج الانحدار شبه المعلمي، التعدد الخطي، القيم الشاذة، التقديرات الحصينة، مقدر الحرف المبني على الفروق، مقدر M، مقدر S، دالة النواة، مقدر ندارايا واتسن، مقدر الشرائح التمهيدية التكعيبية

***البحث مستل من رسالة ماجستير**

(2)
(3)