Journal of Economics
and Administrative Sciences (JEAS)

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

Comparing Some of Robust the Non-Parametric Methods for Semi-Parametric Regression Models Estimation

Zahraa Kh.Bahez
Al-Mustansiriya University / College of Administration and Economics
Department of Statistics
zahraakhalidbahez@gmail.com

Husam .A. Rasheed
Al-Mustansiriya University / College of Administration and Economics
Department of Statistics
husamstat@gmail.com

Received:30/1/2022 **Accepted:2/3/2022** **Published: June / 2022**

Abstract

In this research, some robust non-parametric methods were used to estimate the semi-parametric regression model, and then these methods were compared using the MSE comparison criterion, different sample sizes, levels of variance, pollution rates, and three different models were used. These methods are S-LLS S-Estimation -local smoothing, (M-LLS)M- Estimation -local smoothing, (S-NW) S-Estimation-NadaryaWatson Smoothing, and (M-NW) M-Estimation-Nadarya-Watson Smoothing.

The results in the first model proved that the (S-LLS) method was the best in the case of large sample sizes, and small sample sizes showed that the (M-LLS) method was the best, while the second model showed in general that the S-LLS method was the best in addition to the method M-LLS was the best in some cases of sample sizes and at different levels of variance. As for the third model, it was shown through the results that in most cases the S-LLS method was the best in addition to the M-LLS method which was better in some cases of sample sizes and at different levels of variance.

Key words: Semi-Parametric Regression;M-Estimation;S-Estimation; Robust Semiparametric Methods; Nonparametric Estimation Method;Kernel Method.

1. Introduction:

Semi-parametric regression (SPR) is considered a partial model that consists of a parameter vehicle and a non-parametric component [1]. The main feature of this model is that it contains all the positive features of the parametric and non-parametric models [2].

But the data often contains outliers, so the use of parametric regression estimation methods will not be appropriate, for example, the method of least squares and the Maximam Likelihood, which calls for searching for alternative methods to traditional estimation methods.

The aim of this study is to estimate a semi-parametric regression model using robust estimation.

Because of its importance, many researchers have written about semi- parametric regression

In the year 2000 (Dylin, Ljwe) [9] studied the semi-parametric regression model for the rate function of events.

In 2005, the researchers (Luisito, Eric) [3] examined the (Kuznets) environmental curve across countries using a semi-parametric regression model.

In 2007, (Anatoly, Demitry, Yuri) [12] studied the semi-parametric regression models to assess the biological productivity.

In 2010, the researchers (Nikolaos , Lefteris) [10] studied the semi-parametric regression model to estimate the cost of programs in the early stages of development.

In 2015, the researchers (Xu , Xiaohua) [13] studied the semi-parametric model to study nutrition improvement and diet change.

In 2017, the researchers (Rui, Chenlei, Jinhong) [8] estimated the semi-parametric regression model for longitudinal data.

In 2019, the researchers (Sebastion, Freek) [11] studied the quasi-parametric model to estimate the dynamic model in robots.

2. Semi-Parametric Regression Model

It is considered that the semi-parametric regression model is one of the statistical methods that achieves the characteristics of both parametric and non-parametric regression [2] and they enjoy more than the two types mentioned above in the first two-part model, the first part contains unknown parameters [11] It is estimated by one of the methods by which parameter regression is estimated such as the Ols method or the Mle method.

As for the second part of the model, the relationship or function that includes the effect is unknown. This part is estimated by one of the nonparametric methods

The Semi-parametric regression model can be represented by the following formula [14].

$$Y = XB + g(z) + \varepsilon_i \quad \dots \quad (1)$$

Y: represents the response variable of degree (n * 1).

X : represents the matrix of explanatory variables of degree (n*p).

B: a feature vector of degree (p*1).

z : a second explanatory variable of degree (n*1).

g(z): represents an unknown parametric function of degree (n*1).

ε_i : vector of errors of degree (n * 1).

3. Nonparametric Estimation Method

To estimate the non-parametric part of the Semi-parametric regression model, the following non-parametric smoothing methods were used.

3.1 Nadaraya-Watson Smoothing

The Nadarya-Watson estimator is the most widely used estimator for estimating non-parametric regression models, but here it will be considered a semi-parametric estimator to estimate a semi-parametric regression model.

One of the most important features of this estimator is that it is used in both cases, whether the design is fixed or random, and the general formula for this paver is written as follows[5].

$$\hat{g}(z) = \sum_{i=1}^{n} \frac{k(\frac{z - z_i}{h})y_i}{\sum_{i=1}^{n} k(\frac{z - z_i}{h})} \qquad \dots(2)$$

$k(\frac{z-z_i}{h})$;Kernel Function

3.2 Local Linear Smoother

The main idea behind this smoother is a non-parametric estimation of g when Z = z and is used if the design is random or fixed. This leads to least squares solving but with different advantages.

We will illustrate this through Tyler's series

$$\hat{g}(z) = \frac{\sum_{i=1}^{n}(T_i y_i)}{\sum_{i=1}^{n} T_i} \qquad \dots(3)$$

$$T_i = k\left(\frac{Z-Z_i}{h}\right)[S_{n,2} - (Z - Z_i)S_{n,1}]$$

$$S_{n,L} = \sum_{i=1}^{n} k\left(\frac{Z-Z_i}{h}\right)(Z - Z_i)^L \quad \text{L=1,2}$$

4. Robust Semi-Parameter Estimation Methods

4.1 M-Estimation

It is an alternative hippocampal regression estimator to the method of least squares and an extension of the (MLE) method [4]. It is widely used and sometimes called (Huber's estimations) [6] and it is considered a strong estimation against outliers

$\hat{\beta}_{\mu}$= min$P(e_i)$

$$\hat{\beta}_M = \min \sum_{i=1}^{n} \rho\left(Y_i - \sum_{j=0}^{k} x'_{ij}\beta_j\right)$$

$\rho(.)$:It is a symmetric function that has a minimum termination point at zero and gives a contribution to all the values of the remainders in the objective function

The steps of the M method for estimating a Semi-parametric regression model are as follows [4][1].

1. We set the value of the iteration counter to 1=0 we set an initial value of $[B_{10,}B_{20,}B_{30,} \dots . B_{p0}] = B_0$

for we substitute the default value(B_0) in the model(1) and move it to the other side

$y - X\hat{\mathrm{B}} = g(z) + \varepsilon i$

$y^{*} = g(z) + \varepsilon i$ ... (4)

As $y - X\widehat{\mathrm{B}} = y^{*}$

2. Let's assume an initial value of g^(z)

3. We estimate σ^ using the formula

$$\widehat{\sigma} = \frac{MAD}{0.6745}$$

$$MAD = median|e_i - median(e_i)|$$

4. Calculate the weighted value w_i according to a Tukey function that lives on residuals.

$$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{c}\right)^2\right]^2, & |\ u_i\ | \le c \\ 0, & |\ u_i\ | > c \end{cases} \quad \dots(4)$$

$$u_i = \frac{e_i}{\widehat{\sigma}}$$

c=4.685

5. Producing new experimental data Y^{**}

$$Y^{**} = \widehat{g}(z) + \frac{w_i}{2}$$

6. Calculation of $\widehat{g}(z)$ using the nonparametric grader (NW) and (L.L.S)

$$g\hat{}(z)_{NW} = \frac{\sum_{i=1}^{n} k\left(\frac{z - Z_i}{h}\right)(y_i^*)}{\sum_{i=1}^{n} k\left(\frac{z - Z_i}{h}\right)}$$

$$g\hat{}(z)_{LLS} = \frac{\sum_{i=1}^{n}(T_i y_i)}{\sum_{i=1}^{n} T_i}$$

7. We repeat steps (2) through (6) until we obtain a stable estimate

8. We substitute Steps (7) into the model (1)

$$Y = XB + \widehat{g}(z) + \varepsilon_i$$

$$Y - \widehat{g}(z) = XB + \varepsilon_i$$

$$\acute{Y} = XB + \varepsilon_i$$

9. We assign an initial value to $\widehat{B}$

10. The σ^ is calculated in the same formula as before (3)

11. Calculate weights according to the Tukey function in the same formula (4) [6]

12. The calculation of $\widehat{\beta}$ is estimated according to the following formula

$$\widehat{\beta}_M = (X'WX)^{-1}X'WY$$

(W): It is a matrix of diagonal weights with a capacity $(n * n)$ to estimate the parameters of the parameter part

13. We repeat steps (9) to (12) until obtaining stability in the estimate of $\widehat{\beta}$

14. We substitute (13) in the Model (1) form

$$Y = X\widehat{B} + \widehat{g}(z) + \varepsilon_i$$

We follow the same steps to employ the M method in the Local Linear Smoother and the iteration process continues to be the linear smoother as follows

$$\left\|\widehat{B}_{LLS}(X)^{I}-\widehat{B}_{LLS}(X)^{I-1}\right\| \leq 0.0001$$

4.2 S-Estimation

This method is called S-estimation because it depends on the estimation of the scale of errors [4]. The method of least squares was generalized by (Rousseeuw & Yahai) [6] to provide this new category of estimation in the framework of (S-estimation) and this method reduces the total errors to the lowest possible and is very resistant to anomalies found in data [1].

The parameters are estimated by minimizing the residual errors and depend on the residuals of the M. Method [7].

$$\widehat{\beta}=min\beta\widehat{\sigma}(e1,e2,\dots,en)$$

By defining the lowest Robust estimator (σ̂S) and it achieves

$$\min\sum_{i=1}^{n}\rho\left(\frac{y_i-\sum_{j=0}^{k}x_{ij}\beta_j}{\widehat{\sigma}_S}\right)$$

The S method is more robust than the M method because the S estimators have smaller asymptotic bias and smaller asymptotic variance in the data with outliers.

The steps of the S method for estimating a Semi-parametric regression model are as follows [4] [1].

1. **We give $\widehat{B}$ an initial value**

The form becomes in the following form

$$y-X\widehat{B}=g(z)+\varepsilon i$$

$$y^{*}=g(z)+\varepsilon i$$

2. **Let's assume an initial value of $\widehat{g}(z)$**
3. **The $\widehat{\sigma}$ of the initial estimate is calculated using the formula**

$$\widehat{\sigma}=\frac{MAD}{0.6745}$$

$$MAD=median|e_i-median(e_i)|$$

And calculate the σ̂ for a frequency greater than one according to the following formula

$$\widehat{\sigma}_s=\sqrt{\frac{\sum_{i=1}^{n}W_i e^2{}_i.}{nK}}$$

4. **Calculate Value $u_i=\frac{e_i}{\widehat{\sigma}}$**

5. **Calculate weights according to the Tukey function (4) but c=1.547**

6. **Calculation of $\widehat{g}(z)$ using the nonparametric grader (NW) and (L.L.S)**

$$\hat{g}(z)_{NW}=\frac{\sum_{i=1}^{n}k\left(\frac{z-Z_i}{h}\right)(y_i^{*})}{\sum_{i=1}^{n}k\left(\frac{z-Z_i}{h}\right)}$$

$$\hat{g}(z)_{LLS}=\frac{\sum_{i=1}^{n}(T_i y_i)}{\sum_{i=1}^{n}T_i}$$

7. **We repeat steps (2) through (6) until we obtain a stable estimate**

8. We substitute (7) into model (1)

$$Y = XB + \hat{g}(z) + \varepsilon_i$$
$$Y - \hat{g}(z) = XB + \varepsilon_i$$
$$\acute{Y} = XB + \varepsilon_i$$

9. We assign an initial value to $\hat{\beta}$

10. The $\hat{\sigma}$ **is calculated in the same form as before (3)**

11. Calculate weights by equation (4) but c=1.547

12. The calculation of $\hat{\beta}$ **is estimated according to the following formula**

$$\hat{\beta}_s = (X'WX)^{-1}X'WY$$

(W) It is a matrix of diagonal weights with a capacity (n*n) to estimate the parameters of the parameter part

13. We repeat steps (9) to (12) until obtaining stability in the estimate of $\hat{\beta}$

14. We substitute (13) in the Model (1) form

$$Y = X\hat{B} + \hat{g}(z) + \varepsilon_i$$

We follow the same steps to employ the M method in the Local Linear Smoother and the iteration process continues to be the linear smoother as follows

$$\left\|\hat{B}_{LLS}(X)^{I} - \hat{B}_{LLS}(X)^{I-1}\right\| \leq 0.0001$$

5. Result And Discussions

The MATLAB program was relied upon to reach these results after assuming three sample sizes, three variance ratios and three pollution ratios, and two explanatory variables (x, z) being generated using the method (Box-Muller). The dependent variable is calculated through the models used in the simulation in terms of the explanatory variable.

The first form:

$$g_1(z_{1i}) = 0.5\sin(2\pi z_i)$$

The second form:

$$g_2(z_{2i}) = \sin 2z + exp(-16z^2)$$

The third form:

$$g_3(z_{3i}) = exp(-(z_i - 0.5)^2$$

Three sample sizes (10,50,100) and three levels of variance (0.01,0.5,1.5) and three pollution rates (15,20,25) were assumed, with repeated values as 500 times. Note that an explanation of the symbols in the tables is as follows.

1- M-NW: M-Estimation Nadarya-Watson Smoothing.

2- M-LLS: M-Estimation-Local Smoothing.

3- S-NW: S-Estimation Nadarya-Watson Smoothing.

4- S-LLS: S-Estimation-Local Smoothing.

The results of the first form

Table (1) shows MSE for each form (pollution rate 15%)

n	σ	M-NW	S-NW	M-LLS	S-LLS	Best
10	0.01	0.1067	0.1438	0.1066	0.5119	M-LLS
	0.5	0.660150	0.80169	0.660143	1.54736	M-LLS
	1.5	4.9788	5.9903	4.9773	6.7018	M-LLS
50	0.01	0.02372	0.0319	0.02370	0.0137	S-LLS
	0.5	0.146700	0.1782	0.146701	0.1039	S-LLS
	1.5	1.1064	1.3312	1.1065	1.0893	S-LLS
100	0.01	0.01227	0.0165	0.01226	0.0118	S-LLS
	0.5	0.075879	0.0921	0.075880	0.0739	S-LLS
	1.5	0.57231	0.6885	0.57233	0.4703	S-LLS

Table (2) shows MSE for each form (pollution rate 20%)

n	σ	M-NW	S-NW	M-LLS	S-LLS	Best
10	0.01	0.1078	0.1452	0.1077	0.5170	M-LLS
	0.5	0.666818	0.8098	0.666802	1.5630	M-LLS
	1.5	5.0291	6.0508	5.0290	6.7694	M-LLS
50	0.01	0.0515	0.0694	0.0515	0.0470	S-LLS
	0.5	0.318561	0.3869	0.318562	0.2967	S-LLS
	1.5	2.4026	2.8907	2.4028	1.2340	S-LLS
100	0.01	0.0384	0.0517	0.0384	0.0242	S-LLS
	0.5	0.237582	0.2885	0.237581	0.2469	M-LLS
	1.5	1.7918	2.1559	1.7920	1.0119	S-LLS

Table (3) shows MSE for each form (pollution rate 25%)

n	σ	M-NW	S-NW	M-LLS	S-LLS	Best
10	0.01	0.1334	0.1797	0.1333	0.6398	M-LLS
	0.5	0.825188	1.0021	0.825182	1.9342	M-LLS
	1.5	6.2235	7.4879	6.2222	8.3772	M-LLS
50	0.01	0.0654	0.0881	0.0653	0.3136	M-LLS
	0.5	0.404518	0.4912	0.404519	0.3482	S-LLS
	1.5	3.0509	3.6707	3.0512	3.0066	S-LLS
100	0.01	0.0542	0.0730	0.0541	0.0518	S-LLS
	0.5	0.335044	0.4069	0.335046	0.2853	S-LLS
	1.5	2.5269	3.0402	2.5271	1.4013	S-LLS

The results of the second form:

Table (4) shows MSE for each form (pollution rate 15%)

n	σ	M-NW	S-NW	M-LLS	S-LLS	Best
10	0.01	0.7117	0.9801	0.7229	0.7059	S-LLS
	0.5	1.2798	1.6243	1.2762	1.2851	M-LLS
	1.5	5.2160	6.2325	5.2154	5.2285	M-LLS
50	0.01	0.1655	0.2279	0.1681	0.1351	S-LLS
	0.5	0.2976	0.3777	0.2968	0.2614	S-LLS
	1.5	1.2130	1.4494	1.2129	1.0810	S-LLS
100	0.01	0.0782	0.1077	0.0794	0.0920	M-LLS
	0.5	0.1406	0.1785	0.1402	0.2011	M-LLS
	1.5	0.5732	0.6849	0.5731	0.7943	M-LLS

Table (5) shows MSE for each form (pollution rate 20%)

n	σ	M-NW	S-NW	M-LLS	S-LLS	Best
10	0.01	0.7189	0.9900	0.7102	0.7226	M-LLS
	0.5	1.2927	1.6407	1.2891	2.3082	M-LLS
	1.5	5.2687	6.2955	5.2681	5.3015	M-LLS
50	0.01	0.3434	0.4730	0.3388	0.3991	M-LLS
	0.5	0.6176	0.7838	0.6158	1.1027	M-LLS
	1.5	2.5170	3.0075	2.5167	1.4882	S-LLS
100	0.01	0.2553	0.3516	0.2593	0.0940	S-LLS
	0.5	0.4591	0.5826	0.4578	0.3197	S-LLS
	1.5	1.8710	2.2356	1.8707	0.9928	S-LLS

Table (6) shows MSE for each form (pollution rate 25%)

n	σ	M-NW	S-NW	M-LLS	S-LLS	Best
10	0.01	0.8968	1.2350	0.9109	0.8661	S-LLS
	0.5	1.6127	2.0468	1.6081	2.8795	M-LLS
	1.5	6.5727	7.8536	6.5719	6.1086	S-LLS
50	0.01	0.4502	0.6200	0.4573	0.4475	S-LLS
	0.5	0.8095	1.0275	0.8073	1.4455	M-LLS
	1.5	3.2994	3.9424	3.2990	2.9724	S-LLS
100	0.01	0.3620	0.4986	0.3677	0.2423	S-LLS
	0.5	0.6510	0.8263	0.6492	1.1624	M-LLS
	1.5	2.6534	3.1704	2.6531	2.6771	M-LLS

The results of the third form:

Table (7) shows MSE for each form (pollution rate 15%)

n	σ	M-NW	S-NW	M-LLS	S-LLS	Best
10	0.01	0.1016	0.1531	0.1019	0.0520	S-LLS
	0.5	0.5840	0.8100	0.5846	0.4335	S-LLS
	1.5	5.2823	6.1373	5.2729	6.6716	M-LLS
50	0.01	0.0236	0.0356	0.0237	0.0151	S-LLS
	0.5	0.1358	0.1884	0.1360	0.1334	S-LLS
	1.5	1.2284	1.4273	1.2263	1.1515	S-LLS
100	0.01	0.0109	0.0165	0.0110	0.0086	S-LLS
	0.5	0.0628	0.0871	0.0629	0.0541	S-LLS
	1.5	0.5680	0.6599	0.5670	0.3174	S-LLS

Table (8) shows MSE for each form (pollution rate 20%)

n	σ	M-NW	S-NW	M-LLS	S-LLS	Best
10	0.01	0.1026	0.1546	0.1029	0.0566	S-LLS
	0.5	0.5899	0.8182	0.5805	1.4480	M-LLS
	1.5	5.3357	6.1993	5.3262	6.7390	M-LLS
50	0.01	0.0490	0.0739	0.0492	0.0381	S-LLS
	0.5	0.2818	0.3909	0.2811	0.2917	M-LLS
	1.5	2.5490	2.9616	2.5445	2.2194	S-LLS
100	0.01	0.0363	0.0547	0.0362	0.0366	M-LLS
	0.5	0.2088	0.2896	0.2090	0.1125	S-LLS
	1.5	1.8886	2.1942	1.8852	0.3853	S-LLS

Table (9) shows MSE for each form (pollution rate 25%)

n	σ	M-NW	S-NW	M-LLS	S-LLS	Best
10	0.01	0.1270	0.1914	0.1264	0.1650	M-LLS
	0.5	0.7300	1.0125	0.7208	1.7919	M-LLS
	1.5	6.6029	7.6716	6.5911	8.3395	M-LLS
50	0.01	0.0643	0.0968	0.0642	0.2859	M-LLS
	0.5	0.3694	0.5124	0.3698	0.3068	S-LLS
	1.5	3.3414	3.8822	3.3354	2.2202	S-LLS
100	0.01	0.0517	0.0779	0.0518	0.0499	S-LLS
	0.5	0.2971	0.4120	0.2974	0.1292	S-LLS
	1.5	2.6871	3.1220	2.6823	1.3938	S-LLS

Analysis of the simulation results:

Through the results, it was found that the values of the mean squares of error decreased as the sample sizes increased and increased with the increase in the variance values for all the models studied. As for the comparison of methods, the following shows:

<u>First: The results of the first model</u>

1- In the case of 15% pollution, it was shown in Table (1) that the S-LLS method was the best when sample sizes 50 and 100, but when sample size 10, the M-LLS method was the best.

2- In the case of 20% pollution, it was shown in Table (2) that the S-LLS method is the best except for the sample size 10. The M-LLS method was the best, and at the size 100 and the level of variation 0.5 also M-LLS was the best.

3- In the case of 25% pollution, we note from Table (3) that the S-LLS method is the best for the sample size 100 and 50, except for sample size 10 and 50 at the level of variation 0.01 the M-LLS method was the best.

<u>Second: The results of the second Model:</u>

1- In the case of 15% pollution, it was shown from Table (4) that the M-LLS method was the best at sample size 10 and 100, except for sample size 50 and 10 at the level of variation 0.01 the S-LLS method was the best.

2- In the case of 20% pollution, it was shown in Table (5) that the M-LLS method was the best at sample size 50 and 10, except for sample size 100 and 50 at a level of variation of 1.5 the S-LLS method was the best.

3- In the case of 25% pollution, it was shown in Table (6) that the S-LLS method was the best when the sample size was 10 at a level of variance 0.01 and 1.5, the sample size was 50 at a level of variance 0.01 and 1.5 and the sample size was 100 at a level of variance 0.01 and 1.5, as for the rest of the cases the M-LLS method was the best .

<u>Third: The results of the third Model:</u>

1- In the case of 15% pollution, it was shown in Table (7) that the S-LLS method was the best, except for the sample size 10 at the level of variation of 1.5. It was found that the M-LLS method was the best.

2- In the case of 20% pollution, it was shown in Table (8) that the S-LLS method is the best except for the sample size 10 at the level of variation 0.01, the level of variation 1.5, the sample size of 100 and the level of variation 1.5. It turned out that the M-LLS method was the best.

3- In the case of 25% contamination, it was shown in Table (9) that the M-LLS method was the best at a sample size of 10 and a sample size of 50 at a variance level of 0.01. As for the rest of the cases, the S-LLS method was the best.

6. Conclusions and Recommendations

<u>6.1 Conclusions</u>

Depending on the simulation results for each of the semi-parametric regression models and for all sample sizes and levels of variance, the following conclusions are reached:

1- It is found that the MSE values for all models are inversely proportional to the sample sizes, that is, they increase as the sample sizes decrease.

2- It is found that the MSE values decrease as the variance values decrease, and this means that it is directly proportional to the variance values.

As for the conclusions for each of the semi-parametric regression models, it reaches the following:

The first Model:

3- In general, it turns out that the S-LLS method is the best in the case of large sample sizes, and when small sample sizes show that the M-LLS method is the best.

The second Model:

4- In general, it is found that the S-LLS method is the best in addition to the M-LLS method is better in some cases of sample sizes and at different levels of Variance.

The third Model:

5- It is found in general in most cases that the S-LLS method is the best in addition to the M-LLS method is better in some cases of sample sizes and at different levels of Variance.

6.2 Recommendations

1- In the first model, the S-LLS method is used in cases of large sample sizes because it gives less MSE.

2- Using the S-LLS method in the second model, it gives the best capabilities and the lowest MSE.

3- Using the S-LLS method in the third model because it gives the best estimates with the least error.

4- Using an unused kernel smoother such as the Gasser-Muller Kernel and other non-parametric functions such as the Spline function.

5- Using unused robust methods such as (LMS) and (LTS) in estimating the semi-parametric regression model.

References

1- Alves, M. F., Gomes, M. I., & de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. Portugaliae Mathematica, 60(2), 193-214.

2 - Amini, M., & Roozbeh, M. (2015). Optimal partial ridge estimation in restricted semiparametric regression models. Journal of Multivariate Analysis, 136, 26-40.

3- Bertinelli, L., & Strobl, E. (2005). The environmental Kuznets curve semi-parametrically revisited. Economics Letters, 88(3), 350-357.

4- Boente, G., Salibian-Barrera, M., & Vena, P. (2020). Robust estimation for semi-functional linear regression models. Computational Statistics & Data Analysis, 152, 107041.

5- Gao, J., & Tong, H. (2002). Nonparametric and semiparametric regression model selection Journal Australia ,Crawely WA 6009.

6- Huber, P. J. (2002). John W. Tukey's contributions to robust statistics. Annals of statistics, 1640-1648.).

7- Li, G. (1985). Robust regression. Exploring data tables, trends, and shapes, 281, U340.

8- Li, R., Leng, C., & You, J. (2017). A Semiparametric Regression Model for Longitudinal Data with Non-stationary Errors. Scandinavian Journal of Statistics, 44(4), 932-950.
9- Lin, D. Y., Wei, L. J., Yang, I., & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62(4), 711-730
10- Mittas, N., & Angelis, L. (2010). LSEbA: least squares regression and estimation by analogy in a semi-parametric model for software cost estimation. Empirical Software Engineering, 15(5), 523-555.
11- Riedel, S., & Stulp, F. (2019). Comparing semi-parametric model learning algorithms for dynamic model estimation in robotics. arXiv preprint arXiv:1906.11909.
12- Shvidenko, A., Schepaschenko, D., Nilsson, S., & Bouloui, Y. (2007). Semi-empirical models for assessing biological productivity of Northern Eurasian forests. Ecological Modelling, 204(1-2), 163-179.
13- Tian, X., & Yu, X. (2015). Using semiparametric models to study nutrition improvement and dietary change with different indices: The case of China. Food Policy, 53, 67-81.
14- Zhao, Y., Gijbels, I., & Van Keilegom, I. (2021). Parametric copula adjusted for non-and semi-parametric regression. Annals of Statistics.

مقارنة بعض الطرائق اللامعلمية الحصينة لتقدير أنموذج الانحدار شبه المعلمي

الباحث/ زهراء خالد باهض
الجامعة المستنصرية / كلية الادارة والاقتصاد
قسم الاحصاء
zahraakhalidbahez@gmail.com

أ.م.د حسام عبد الرزاق رشيد
الجامعة المستنصرية / كلية الادارة والاقتصاد
قسم الاحصاء
husamstat@gmail.com

Received:30/1/2022 Accepted:2/3/2022 Published: June / 2022

مستخلص البحث:

في هذا البحث تم استعمال بعض الطرائق اللامعلمية الحصينة لتقدير أنموذج الانحدار شبه المعلمي ومن ثم مقارنة هذه الطرائق بالاعتماد على معيار المقارنة الـ MSE اذ تم استعمال احجام عينات ومستويات تباين ونسب تلوث مختلفة وثلاثة نماذج مختلفة وهذه الطرائق تمثلت بطريقة S-Estimation-Local Smoothing (S-LLS) وM-Estimation-Local Smoothing (M-LLS) و S-Estimation-Nadarya_Watson Smoothing (S-NW) وM-Estimation-Nadarya-Watson Smoothing (M-NW).

واثبتت النتائج في الانموذج الاول ان طريقة (S-LLS) كانت هي الافضل في حالة احجام العينات الكبيرة وعند احجام العينات الصغيرة تبين ان طريقة (M-LLS) هي الافضل اما الانموذج الثاني تبين بشكل عام ان طريقة S-LLS هي الافضل بالاضافة الى طريقة M-LLS هي الافضل في بعض حالات احجام العينات وعند مستويات تباين مختلفة اما الانموذج الثالث تبين من خلال النتائج ان اغلب الحالات طريقة S-LLS هي الافضل بالاضافة طريقة M-LLS افضل في بعض حالات احجام العينات وعند مستويات تباين مختلفة .

المصطلحات الرئيسة للبحث: طريقة M-estimation, طريقة S-estimation , انموذج الانحدار شبه المعلمي , طرائق التقدير الحصينة , طرائق التقدير اللامعلمية الحصينة , طريقة Kernel

*البحث مستل من رسالة ماجستير