Estimation of adaptive parameters in a nonparametric regression

Noor Salah Hassan — Basra University , College of Administration and Economics, Department of Statistics , Iraq

Sahera Hussein Zain Al Thalabi[2] — Basra University , College of Administration and Economics, Department of Statistics , Iraq

ABSTRACT

The researcher faces several problems when estimating the nonparametric regression functions as they depend heavily on the data and these estimates may be inaccurate, or there may be a problem in finding an efficient method that fits the nonparametric model, so the goal is to find the adaptive capabilities in the nonparametric regression by the "Goldenshluger-lepski" method Modern methods to increase the efficiency and accuracy of estimation through the use of adaptive estimators in non-parametric regression method . In this paper, adaptive estimations were processed in the nonparametric regression method through the use of kernel smoothing and spline. The adaptive "Goldenshluger-Lepski" was included, and to compare the estimation methods three criteria were used, namely (MSE , MAS, RMSE) to choose the best method after applying the procedure to the simulation in the R Package.

Keywords: Adaptive estimation, nonparametric regression, kernel and spline smoothing, Goldenshluger-Lepski bandwidth.

Introduction

There are many problems that the researcher faces when estimating the nonparametric regression functions because they are highly dependent on the data. These estimates may be inaccurate, or there may be a problem in finding an efficient method that fits the nonparametric model, so the aim of this study is to find the adaptive capabilities in the nonparametric regression using modern methods. To increase the efficiency and accuracy of estimation through the use of adaptive estimators in nonparametric regression smoothing . We will discuss some studies that used the adaptive method and its use in nonparametric regression, including:

-The study (Hill and others, 1988) used two nonparametric adaptive procedures to apply multiple comparisons and a test of alternatives required in a one-way ANOVA model, in comparison with the parametric normal theoretical procedure, and the rank-based non-parametric procedure where these procedures are applied to lung cancer data. The results showed the superiority of the adaptive procedures Nonparametric. [6]

-A study (2021, page and Grunewalder) presented an adaptive estimation procedure using the Goldenshluger-Lepski method to choose parameters for the statistical estimator using only the available data without making strong assumptions about the estimation. Nucleus . This method was used to address two

----

[1] Corresponding Autho : Basra University , College of Administration and Economics, Department of Statistics , Iraq , Email noorsalahh1997@gmail.com

[2] Assistant Professor , Basra University , College of Administration and Economics, Department of Statistics , Iraq , Email sahera.zain@uobasrah.edu.iq

regression problems, the kernel regression was fixed in one of them and in the other an adaptation was used. [12]

-The study (Breunig and Chen, 2022) aimed to find an adaptive estimation of the minimum quadratic function in the model of non-parametric automatic variables (NPIV), which is an important problem in the optimal estimation of non-linear functions, this problem is solved through a choice based on data from Lepski type For the smoothing parameter, the results showed that the adaptive estimator of the quadratic function achieves the minimum optimum rate. [3]

Adaptive estimator in nonparametric regression: [10][1][3][4][8][11] An adaptive estimator is defined as an effective estimator for only a partially specified model ("effective" meaning that it is asymptotically equivalent to a non-parametric " likelihood Maximum" local probability estimator Applicable), or a model whose distribution is unknown, so adaptive estimation aims to build estimations entirely based on data without making strong assumptions about the estimation. Nonparametric regression is also a form of regression analysis and a common and flexible tool for data analysis and modeling of the non-linear relationship between dependent and explanatory variables. , that is, it depends mainly on the data, as the goal of nonparametric regression is to estimate the regression function without dependence or having prior knowledge of its functional form , and using adaptive methods, modifications can be made to the traditional methods so that they are as strong as many nonparametric methods. Studies to build a method for selecting data-based smoothing parameters in order to obtain adaptive estimates. The first adaptive procedure was proposed by (lepski 1990) and was developed in (1992) and its goal was to build capabilities from the data in the best possible way and reduce the risk of estimation, the adaptive methods in regression The non-parametric is strong in efficiency as it cannot be outweighed by any non-adaptive method, as the exact adaptive procedure will work well with the data. So the adaptive approach is mainly divided into two types, either it is an adaptive approach to estimate unknown parameters as in nonparametric regression, or the use of Data to determine the appropriate statistical procedure, the adaptive non-parametric approach on the one hand is estimating the parameters from the sample, and on the other hand, data-based methods are the best and most The first to suggest this approach (Randles and Hogg). So the main purpose of adaptive approaches may be to provide a relatively easy alternative to parameterization without much effort on how to choose one from a variety of methods, and to facilitate the decision on the use of the appropriate technique. Adaptive approaches can perform better based on the available information in terms of achieving the desired combination of robustness and efficiency. over the past ten years. The rank-based adaptive tests show that adaptive actions with a potential value can increase the power of certain tests, and the adaptive method can increase the power of the test. If the error distribution is abnormal, the power of classical tests is much lower than adaptive tests.

The formula for nonparametric regression is as follows:

$$, \quad i=1,2,\ldots,n \quad y_i = m(xi) + \varepsilon i \quad \ldots. \quad (1)$$
$$, \quad \varepsilon \sim N(0, \sigma^2)$$

Yi: the response variable, m (xi): the unknown function to be estimated,

xi: the explanatory variable , εi: the values of the random variable, which is white noise that is normally distributed.

The adaptive estimator for the parameter vector is as follows [15] :

$$\hat{\theta}(x) = \hat{\theta}_k(x) = \left(\theta_k^1(x), \ldots\ldots, \theta_k^p(x)\right)^T \quad \ldots.(2) \qquad k = 1, \ldots., p$$

$\theta_k^1, \ldots\ldots, \theta_k^p$:Unknown parameter, θ is estimated based on sample observations (xi , yi)

KERNEL SMOOTHERS

The positional polynomial regression smoother (LLS) is one of the best smoothing methods because it deals with static and random models, and it is sometimes called the weight or window function, as this function is continuous and symmetric, its integral is equal to the integer one, when (the bandwidth) is small very . [10]

The formula for smoothers is as follows

$$\widehat{m_h}(x) = \frac{\sum_{i=1}^{n} yi\, k\,(x - Xi)/h}{\sum_{i=1}^{n} k\,(x - Xi)/h} \quad \ldots.(3)$$

$$w_i(x) = \frac{\frac{k(x - Xi)}{h}}{\sum_{i=1}^{n} k\,(x - Xi)/h} \quad \ldots..(4)$$

$\frac{\sum_{i=1}^{n} k(x-xi)}{h}$: represents the endodontic function , $w_i(x)$: represents the weight function and one of its conditions is positive, h: represents the smoothing parameter (the bandwidth) in the estimator $(m(x))$. If its value is large, the function is smooth, and if its value is small, the function is not smooth. [1]

-The Gasser-Müller (G-M) smoother is one of the most widely used gradient smoothing tools. The Gasser-Müller estimator is used to build a nonparametric estimation of the regression function, and the Gasser-Müller estimator is a modification of an earlier version of priestley-chao), a new type of kernel. [4]

Its general form is as follows. [4]

$$\widehat{m}_h(x) = \frac{1}{n}\sum_{i=1}^{n} Y_i \int_{s_{i-1}}^{s_i} K\left(\frac{x-u}{h}\right) du \quad \ldots \quad (5) \quad s_0 = 0 \quad , \quad s_n = 1$$

$$s_i = \frac{x_i + x_{i+1}}{2} \qquad , x_i \le s_i \le x_{i+1}$$

- Also, the nearest neighbor smoother (K-NN) depends on calculating the Euclidean distance between each point and the point closest to it. If the data are close to each other, the distance will be small and vice versa. [9] So, its general form is as follows:

$$\widehat{m}_k(x) = \frac{\frac{k(xi - x)}{k_l}}{\sum_{i=1}^{n} \frac{k(xi - x)}{k_l}} \qquad k_l \to \infty \ldots (6)$$

$$, k_l = d(i,j) = \sqrt[2]{\sum_{i=1}^{n} (xi - xj)^2} \ \ldots \ (7)$$

:represents the Euclidean distance between k_l x, k and xi ,xj : data points

Spline Smoother

depend on the sum of the squares of the residuals as used when the regression line is divided into pieces, as the explanatory variable x with period (a,b) is divided and the lines cut are called slide nodes so that smoothing the slides overcomes the problem of choosing a node and from During the identification of new nodes or changing the existing nodes, they are divided into linear spline (SPL) and cubic spline (SPC). [2][9]

$$S(m) = \sum_{i=1}^{n} \left(yi - \widehat{m}(xi)\right)^2 + \lambda \int_a^b [\widehat{m}''(x)]^2\, dx \ \ldots \ (8) \ , \quad \lambda > 0$$

Whereas

$\sum_{i=1}^{n}\left(yi - \widehat{m}(xi)\right)^2$: It represents the sum of the squares of the residuals , $\widehat{m}''(x)$: Represents the second derivative of the bootstra , λ : Represents the penalty factor indicating the width of the appropriateness quality package represented by $\sum_{i=1}^{n}\left(yi - \widehat{m}(xi)\right)^2$ And the smoothing of appreciation represented by $\int_a^b [\widehat{m}''(x)]^2\, dx$

The adaptive Goldenshluger-Lepski method extends the Lepski method in order to perform adaptation across multiple parameters. This method has been used in different contexts as it was used for the first time in a multidimensional white noise model. As it has been widely used in recent studies of non-parametric estimation, the idea of this method for adaptive non-parametric estimation is to choose an estimator that reduces the sum of the unknown bias factor of variance. [8][12]

The Goldenshluger-lepski formula is as follows [5] :

$$\widehat{h}(x_i) = \arg\min_{h \in Ӈ_n}\{\widehat{A}(h, x_i) + \widehat{V}(h, x_i)\} \ \ldots (9)$$

$$\widehat{A}(h, x_i) = \max_{h' \in Ӈ_n}([\widehat{m}_{h'}(x_i) - \widehat{m}_{h \vee h'}(x_i)]^2 - V(h', x_i)) .. (10)$$

$$\widehat{V}(h, x_i) = k\sigma^2 \frac{\ln n}{n\widehat{\varphi}(h)} \ , h \neq 0 \ .. (11)$$

K : represent a constant that does not depend on h , $\widehat{m}_h(x_i)$: function estimator , $Ӈ_n$: Represents a set of smoothing parameter (bandwidth) .

$\widehat{V}(h, x_i)$: Represents an empirical analogue of variance , $\widehat{A}(h, x_i)$: Represents an approximation of the term bias .

In order to estimate the regression curve, there are several criteria that are relied upon in the differentiation, and among these criteria are the mean absolute error squares (MAS), the roots mean squares error (RMSE), and the mean squared error (MSE) standard [9][14]. The function was used Endodontic (Epanchnickov) and adaptive bandwidth (Goldenshluger-lepski) on the experimental side.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{m}(x)| \quad \ldots (12)$$

$$\text{RMSE} = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{m}(x)\right)^2} \quad \ldots (13)$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{m}(x))^2 \ldots (14)$$

The analysis of this study is carried out using simulation, as it is known as a method that includes the use of a theoretical mathematical model and similar to the real model that represents the studied problem. Simulation experiments were carried out using three sample sizes (n = 30, 60, 100) and with a frequency of 500 for each experiment. The nonparametric methods will be compared, and two models were used in the simulation.

first model $m(x_i) = \sqrt{2}\text{Cos}(2\pi x)$

second model $m(x_i) = 0.5 + 0.2\text{Cos}[4\pi x] + 0.1\text{Cos}[4\pi x]$

The variables (independent and random error) were generated. The random errors are normally distributed with a mean of zero and variance σ^2, The nonparametric explanatory variable Xi is generated according to the standard normal distribution .

Statistical Analysis

TABLE 1. The first model, (RMSE, MSE, MAE) criteria for the first model according to the different sample sizes and levels of variation.

RMSE						
σ^2	N	ALLS	AGM	KNN	ASPL	ASPC
$\sigma^2 = 0.5$	30	1.076737	1.469676	0.83454	1.118572	1.228584
	60	1.189452	1.443043	0.819202	1.176911	1.180858
	100	1.143882	1.399573	0.661001	1.116827	1.131377
$\sigma^2 = 1$	30	1.370108	1.698854	1.248197	1.405926	1.517653
	60	1.519556	1.603637	1.387996	1.498323	1.494812
	100	1.390022	1.555937	1.178932	1.371235	1.387317
$\sigma^2 = 1.5$	30	1.733745	2.021046	1.736418	1.7713	1.889049
	60	1.937968	1.917974	1.998888	1.910516	1.901915
	100	1.729486	1.721086	1.721623	1.716286	1.734665

						MSE
σ^2	N	ALLS	AGM	KNN	ASPL	ASPC
$\sigma^2 = 0.5$	30	1.159363	2.159948	0.696456	1.251202	1.509419
	60	1.414797	2.082373	0.671092	1.385119	1.394425
	100	1.308466	1.958805	0.436922	1.247304	1.280015
$\sigma^2 = 1$	30	1.877197	2.886106	1.557996	1.976629	2.30327
	60	2.309049	2.571653	1.926532	2.244971	2.234463
	100	1.93216	2.42094	1.389879	1.880285	1.924647
$\sigma^2 = 1.5$	30	3.005872	4.084626	3.015147	3.137504	3.568506
	60	3.755721	3.678624	3.995551	3.650071	3.617279
	100	2.991121	2.962136	2.963985	2.945637	3.009063

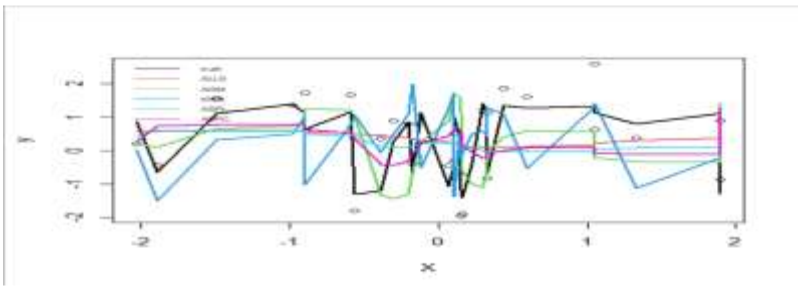
						MAE
σ^2	N	ALLS	AGM	KNN	ASPL	ASPC
$\sigma^2 = 0.5$	30	0.911211	1.11404	0.659559	0.90587	0.97893
	60	0.970488	1.162448	0.59967	0.984768	0.986437
	100	1.002256	1.109086	0.515384	0.964442	0.959727
$\sigma^2 = 1$	30	1.099767	1.319684	1.041636	1.122617	1.211347
	60	1.214723	1.341412	1.054362	1.222712	1.231411
	100	1.114207	1.203239	0.937794	1.079422	1.078488
$\sigma^2 = 1.5$	30	1.342265	1.541023	1.443969	1.375983	1.474218
	60	1.536303	1.581209	1.525517	1.541324	1.542769
	100	1.307936	1.294981	1.368802	1.288311	1.291902

Source/ From the (R.4.1.2) Package using simulation method

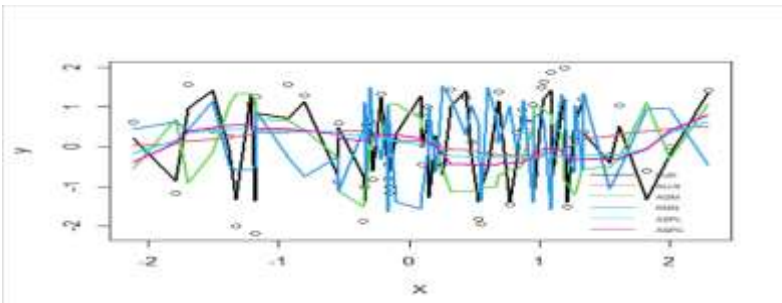
Explanation of Table 1. for the first model

1-The results showed, depending on the comparison criteria (RMSE) and (MSE) when the sample size is (n = 30, 60, 100) and with a level of variance ($\sigma^2 = 0.5$) that the best adaptive estimator is (KNN), followed by the adaptive estimator(ALLS), but when the level of variance is ($\sigma^2 = 1$) and sample size (n = 30), then the best adaptive estimator is (KNN). As for (n=30 , $\sigma^2 = 1.5$) the best adaptive estimator is (ALLS), then the adaptive estimator (KNN) .

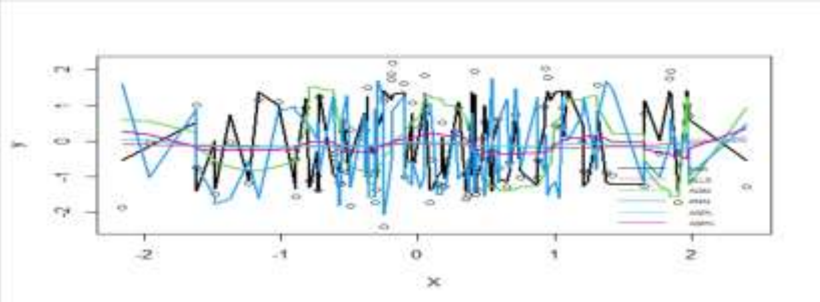
2-The results showed that, depending on the comparison standard (MAE), when the sample size is (n=60) and with the level of variance $\sigma^2 = 0.5$, the estimator is (KNN), then the estimator is (ASPL), but when the variance level is ($\sigma^2 = 1$) At the sample size (n=30), the best adaptive estimator is (KNN), followed by the adaptive estimator (ALLS), and when the level of variance is ($\sigma^2 = 1.5$) at the sample size (n=30), the best estimator It is an adaptive estimator (ALLS), followed by an adaptive estimator (ASPL)



n=30



n=60



n=100

FIGURE1. The adaptive nonparametric capabilities of the first model when the sample size is (30, 60, 100)

TABLE 2. For the second model, (RMSE, MSE, MAE) criteria for the second model according to the different sample sizes and levels of variance

RMSE						
σ^2	**N**	**ALLS**	**AGM**	**KNN**	**ASPL**	**ASPC**
$\sigma^2 = 0.5$	30	1.076737	1.469676	0.83454	1.118572	1.228584
	60	1.189452	1.443043	0.819202	1.176911	1.180858
	100	1.143882	1.399573	0.661001	1.116827	1.131377
$\sigma^2 = 1$	30	1.370108	1.698854	1.248197	1.405926	1.517653
	60	1.519556	1.603637	1.387996	1.498323	1.494812
	100	1.390022	1.555937	1.178932	1.371235	1.387317
$\sigma^2 = 1.5$	30	1.733745	2.021046	1.736418	1.7713	1.889049
	60	1.937968	1.917974	1.998888	1.910516	1.901915
	100	1.729486	1.721086	1.721623	1.716286	1.734665
						MSE
σ^2	**n**	**ALLS**	**AGM**	**KNN**	**ASPL**	**ASPC**
$\sigma^2 = 0.5$	30	1.159363	2.159948	0.696456	1.251202	1.509419
	60	1.414797	2.082373	0.671092	1.385119	1.394425
	100	1.308466	1.958805	0.436922	1.247304	1.280015
$\sigma^2 = 1$	30	1.877197	2.886106	1.557996	1.976629	2.30327
	60	2.309049	2.571653	1.926532	2.244971	2.234463
	100	1.93216	2.42094	1.389879	1.880285	1.924647
$\sigma^2 = 1.5$	30	3.005872	4.084626	3.015147	3.137504	3.568506
	60	3.755721	3.678624	3.995551	3.650071	3.617279
	100	2.991121	2.962136	2.963985	2.945637	3.009063
						MAE
σ^2	**n**	**ALLS**	**AGM**	**KNN**	**ASPL**	**ASPC**
$\sigma^2 = 0.5$	30	0.97893	0.90587	0.659559	1.11404	0.911211
	60	0.986437	0.984768	0.59967	1.162448	0.970488
	100	0.959727	0.964442	0.515384	1.109086	1.002256
$\sigma^2 = 1$	30	1.211347	1.122617	1.041636	1.319684	1.099767
	60	1.231411	1.222712	1.054362	1.341412	1.214723

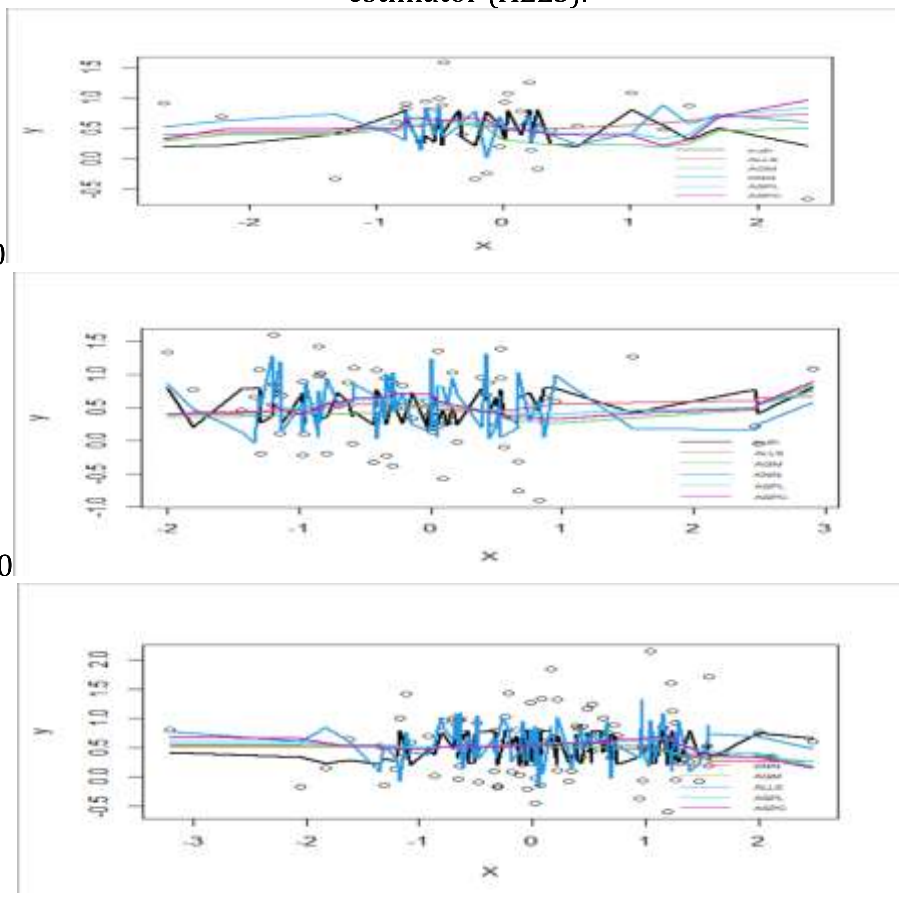
	100	1.078488	1.079422	0.937794	1.203239	1.114207
$\sigma^2 = 1.5$	30	1.474218	1.375983	1.443969	1.541023	1.342265
	60	1.542769	1.541324	1.525517	1.581209	1.536303
	100	1.291902	1.288311	1.368802	1.294981	1.307936

Source/ From the (R.4.1.2) Package using simulation method

Explanation of TABLE 2. for the second model

1-The results showed, depending on the comparison criteria (RMSE) and (MSE) when the sample size is (n = 30, 60, 100) and with a level of variance ($\sigma^2 = 0.5$) that the best adaptive estimator is (ASPL), followed by the adaptive estimator(ALLS), but when the level of variance is ($\sigma^2 = 1$) and sample size (n = 30), then the best adaptive estimator is (ALLS). As for (n=30 , $\sigma^2 = 1.5$) the best adaptive estimator is (ALLS), then the adaptive estimator (ASPL) .

2-The results showed that, depending on the comparison standard (MAE), when the sample size is (n=60) and with the level of variance $\sigma^2 = 0.5$, the estimator is (ALLS), then the estimator is (ASPL), but when the variance level is ($\sigma^2 = 1$) At the sample size (n=30), the best adaptive estimator is (AGM), followed by the adaptive estimator (ALLS), and when the level of variance is ($\sigma^2 = 1.5$) at the sample size (n=30), the best estimator It is an adaptive estimator (AGM), followed by an adaptive estimator (ALLS).



n=30

n=60

n=100

Figures 2. The adaptive nonparametric capabilities of the second model when the sample size is (30, 60, 100)

Conclusions

1-When implementing simulation experiments using three sample sizes (n = 30, 60, 100) and with a frequency of 500 for each experiment and depending on the comparison criteria at a level of variance (σ^2= 0.5), it was found that the best estimator of the first nonparametric model is that the estimator (ALLS) is the best estimator, then It is followed by the estimator (KNN).

2-But when the level of variance is (σ^2=1.5) at the sample size (n=30), the best estimator for the first model is (ALLS), followed by (KNN) estimator, as the values of the criteria (RMSE), (MSE) and (MAE) are less. With increasing sample sizes and for all estimators used, and increasing the values of (RMSE), (MSE) and (MAE) for all estimators with increasing values of residual variance.

3-As for the second model, the results showed, depending on the comparison criteria (RMSE) and (MSE), when the sample size (n = 30,60) at a level of variation (σ^2= 0.5, 1) that the estimator (ALLS) is the best estimator, then follows 2nd place estimator (KNN).

4-But when the level of variance is (σ^2=1) at the sample size (n=30), the best estimator is (KNN), followed by the estimator (ALLS), and in general the best adaptive estimator for the second nonparametric model is (ALLS), followed by (KNN) estimator adaptive ,

5-Finally, we can say that the best estimation adaptive method for the three criteria and by increasing the sample sizes at three different levels of variance was the ALLS method, which represents the smoothing of the adaptive local polynomial regression.

Acknowledgments

The authors would like to thank Basra University (www.uo Basra.edu.iq) Basra - Iraq for its support in the present work.

References

1-Ali, Noor Abdul-Karim (2021), Comparing some of the traditional and immunized nonparametric capabilities of the nonparametric regression model with application. Master's thesis, College of Administration and Economics, University of Basra

2-Aydin , D. & Memmedli2 , M. & Omay2, R. (2013), Smoothing Parameter Selection for Nonparametric Regression Using Smoothing Spline. EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS Vol. 6, No. 2, 222-238 ISSN 1307-5543

3-Breunig, C. & Chen, X. (2022), simple adaptive estimation of quadratic functionals in nonparametric iv models. mathematics statistics theory , version, v2

4-Fan, J. & Gijbels, L. (2017), Local Polynomial Modelling and Its Applications. MONOGRAPHS ON STATISTICS AND APPLIED PROHABILITY , First edition 1996, Taylor & Francis Group journal

5-Chagnya, G. & Roche, A. (2016), Adaptive estimation in the functional nonparametric regression model. Journal of Multivariate Analysis ,Volume 146, April

6-Hill ,N. J. , Padmanabhan, A. R. , Puri, M. L. (1988) Adaptive Nonparametric Procedures and Applications, Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 37, No. 2,1988

7-Lacour, C.& Massart, P. (2016), Minimal penalty for Goldenshluger–Lepski method. Stochastic Processes and their ApplicationsVolume 126, Issue 12, December

8-Lederer, J.& YU , L.& GAYNANOVA , I. (2019), Oracle Inequalities for High-dimensional Prediction. Vol.25 • No. 2 Bernoulli Society for Mathematical Statistics and Probability.

9-Matta, Nour Sabah and Al-Safawi, Safa Younes (2011), Estimation of nonparametric regression functions using some improvement . Iraqi Journal of Statistical Sciences, Vol. 2011, No. 20

10-Muhammad, Abdul-Hussain Muhammad (2011), Using the estimator ((Nadaraya-Watson)) kernel in estimating the nonparametric regression function. Al-Qadisiyah Journal for Administrative and Economic Sciences, Volume 13, Issue 1

11-O'Gorman, T. W. (2004), Applied Adaptive Statistical Methods Tests of Significance and Confidence Intervals. by the American Statistical Association and the Society for Industrial and Applied Mathematics.

12-page, S.& grunewalder , S. (2021), The Goldenshluger–Lepski Method for Constrained

Least-Squares Estimators over RKHSs. Journals Project Euclid , Bernoulli 27(4): 2241-2266 ,November Volume 27 , Issue 4

13-Serdyukova , N. (2012), Adaptive estimation in regression and complexity of approximation of random fields. Statistics Theory (math.ST); Probability (math.PR)

14-Tali, Nada Hussein & Taher, Ahmed Shaker Mohammed Taher (2022), Using kernel regression in estimating the coefficients of the regression model with random error limits that are self-correlated with practical application. Journal of Administration and Economics - Al-Mustansiriya University, issue 132, pages 234-247

15-Zhang , Z. (2014), Adaptive Robust Regression Approaches in data analysis and their Applications. PhD thesis , University of Cincinnati ,ProQuest Publishing,