Hacettepe Journal of
Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 52 (5) (2023), 1425 – 1437
DOI : 10.15672/hujms.1167617

Research Article

Adaptive Nadaraya-Watson kernel regression estimators utilizing some non-traditional and robust measures: a numerical application of British food data

Usman Shahzad*[1,2], Ishfaq Ahmad[1], Ibrahim Mufrah Almanjahie[3], Nadia H. Al - Noor[4], Muhammad Hanif[2]

[1] *Department of Mathematics and Statistics, International Islamic University, Islamabad 44000, Pakistan*
[2] *Department of Mathematics and Statistics, PMAS-Arid Agriculture University, Rawalpindi 46300, Pakistan*
[3] *Department of Mathematics, College of Science, King Khalid University, Abha 62223, Saudi Arabia*
[4] *Department of Mathematics, College of Science, Mustansiriyah University, Baghdad 10011, Iraq*

Abstract

In nonparametric regression research, estimation of regression function is a prime concern. Recently, researchers developed some modified Nadaraya-Watson (N-W) regression estimators utilizing robust mean, median and harmonic mean. In this paper, we propose to utilize the additive combination of non-traditional measures i.e. (Hodges-Lehmann, Mid-Range, Tri-Mean, Quartile-Deviation) with the robust minimum covariance determinant (MCD) scale estimator in (N-W) regression estimator. Utilizing these measures, we get some new versions of (N-W) regression estimator. We also attempted to derive the properties of the proposed versions, such as bias, variance, mean square error (MSE), and mean integrated square error (MISE). The proposed estimators are compared with some of the existing estimators available in literature through a simulation study, utilizing two artificial populations. We also incorporated real-life application by taking British food data set denoted as Engel95, and assess the predictive ability of nonparametric regression, based on proposed and existing N-W estimators.

Mathematics Subject Classification (2020). 62G07

Keywords. Nonparametric regression, N-W kernel estimator, robust measures

*Corresponding Author.
Email addresses: usman.stat@yahoo.com (U. Shahzad), ishfaq.ahmad@iiu.edu.pk (I. Ahmad), imalmanjahi@kku.edu.sa (I.M. Almanjahie), nadialnoor@uomustansiriyah.edu.iq (N.H. Al - Noor), mhpuno@hotmail.com (M. Hanif)
Received: 28.08.2022; Accepted: 22.07.2023

1. Introduction

Most research areas, particularly econometrics, focus on methods that address probabilistic or random phenomena involving relevant data. Modelling or Formulating the basic probabilistic structure of data, that is uncertainty in the process, is a critical task, which can be used to describe the adopted mechanism to create the data. Therefore, in such circumstances, parametric and nonparametric methods of explored density estimation have commonly used to define the structures and subsequently create inferences about the unknown but "real or true" models. A parametric model assumes the density is known to a few parameters, whereas a non-parametric model enables a high degree of flexibility in the feasible form, it is generally assumed to belong to some infinite set of curves.

In recent years, several nonparametric models have been created and used for the concerned purposes. There were two reasons behind this, the flexibility in data analysis and advancement in computer technology, which allows us to use the nonparametric techniques easily. In bivariate studies, as an example, exploring the association between the "independent" and "dependent" variables is of common interest, one feasible way is through the mean regression function to define such an association. The flexible estimation method makes no assumption on the form of this function. The form must be determined entirely through the data and this means that non- parametric estimation is desirable [2].

Let $Y_i \epsilon R$ be the response random variable and $X_i \epsilon R$ be the independent random variable such that $Y_i = m(X_i) + \varepsilon_i$, for $i = 1, 2, ..., n$. Where $m(X_i)$ is the regression function due to (X_i, Y_i) with unknown but common probability density function $\mathfrak{f}$. It is also well understood that $m(x)$ is a conditional mean curve, where

$$m(x) = \int \frac{y\mathfrak{f}(x,y)}{\mathfrak{f}(x)} dy. \tag{1.1}$$

It is worth mentioning that $\mathfrak{f}(x, y)$ be the joint and $\mathfrak{f}(x)$ be the marginal density function. The estimate of $m(x)$ is as follows:

$$\hat{m}(x) = \int \frac{y\hat{\mathfrak{f}}(x,y)}{\hat{\mathfrak{f}}(x)} dy. \tag{1.2}$$

As we know that $m(x)$ is unknown, so a nonparametric kernel estimation of the regression function can be obtained [6]. Suppose x_i for $i = 1, 2, ...n$ be the random sample of size n from the random variable with density function $\mathfrak{f}(x)$. The estimate of $\mathfrak{f}(x)$ is as given below:

$$\hat{\mathfrak{f}}(x_i) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1.3}$$

where K denotes kernel probability density function (pdf) $\forall\ x_i$. Further, h represents fixed bandwidth. The kernel i.e. K, based fulfills the following conditions:

- $\int K(u)du = 1$
- $\int uK(u)du = 0$
- $\int u^2K(u)du \neq 0.$

There are many kernel functions available in literature however Gaussian is the most famous kernel function, see [2, 5, 7, 8, 22, 23].

The rest of the article is structured as follows. Section 2 includes a brief review of some existing estimators of the N-W kernel function. In Section 3, Non-traditional measures besides MCD robust estimator are discussed. Section 3 also includes the proposed class of the N-W kernel regression estimator. Numerical illustration, includes simulation study and real-life application, besides its computational results, are presented in Section 4. Finally, the essential conclusions that drown from the results are covered in Section 5.

2. Review of some N-W kernel regression estimators

The classical Nadaraya-Watson (N-W) kernel estimator of the regression function is obtained as

$$\hat{m}_1(x) = \frac{\sum_{i=1}^{n} y_i K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)}, \tag{2.1}$$

where h is a fixed bandwidth based on Silverman's [19] rule of thumb.

According to [10] and [2], for long-tailed and multi-modal distributions varying bandwidth is preferable rather than fixed bandwidth. So (N-W) kernel estimator with varying bandwidth can be used as given below:

$$\hat{m}_2(x) = \frac{\sum_{i=1}^{n} \frac{y_i}{b(x_i)} K\left(\frac{x - X_i}{b(x_i)}\right)}{\sum_{i=1}^{n} \frac{1}{b(x_i)} K\left(\frac{x - X_i}{b(x_i)}\right)}, \tag{2.2}$$

where

$$b(x_i) = \frac{h}{\sqrt{\hat{\mathfrak{f}}(x_i)}}$$

due to [1].

Silverman [19] extended the idea of [1] and introduced the following bandwidth factor

$$b(x_i) = h\Psi_{G.M},$$

where

$$\Psi_{G.M} = \left[\frac{\bar{\mathfrak{f}}(x_i)}{G.M}\right]^{-\alpha},$$

where G.M denotes geometric mean and $\alpha = 0.5$ due to [1]. Hence, (N-W) kernel estimator due to [19], as given below:

$$\hat{m}_3(x) = \frac{\sum_{i=1}^{n} \frac{y_i}{h\Psi_{G.M}} K\left(\frac{x - X_i}{h\Psi_{G.M}}\right)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h\Psi_{G.M}}\right)}. \tag{2.3}$$

Demir and Toktamis [6] utilized A.M rather than G.M in bandwidth factor as follows:

$$b(x_i) = h\Psi_{A.M},$$

where

$$\Psi_{A.M} = \left[\frac{\bar{\mathfrak{f}}(x_i)}{A.M}\right]^{-\alpha},$$

where A.M denotes arithematic mean and $\alpha = 0.5$. Hence, (N-W) kernel estimator due to [6], as given below:

$$\hat{m}_4(x) = \frac{\sum_{i=1}^{n} \frac{y_i}{h\Psi_{A.M}} K\left(\frac{x - X_i}{h\Psi_{A.M}}\right)}{\sum_{i=1}^{n} \frac{1}{h\Psi_{A.M}} K\left(\frac{x - X_i}{h\Psi_{A.M}}\right)}. \tag{2.4}$$

Khulood and Lutfiah [13] utilized range in bandwidth factor as follows:

$$b(x_i) = h\Psi_R,$$

where

$$\Psi_R = \left[\frac{\bar{\mathfrak{f}}(x_i)}{R}\right]^{-\alpha},$$

where R denotes range and $\alpha = 0.5$. Hence, (N-W) kernel estimator due to [13], as given below:

$$\hat{m}_5(x) = \frac{\sum_{i=1}^{n} \frac{y_i}{h\Psi_R} K\left(\frac{x - X_i}{h\Psi_R}\right)}{\sum_{i=1}^{n} \frac{1}{h\Psi_R} K\left(\frac{x - X_i}{h\Psi_R}\right)}. \tag{2.5}$$

Ali [2] utilized Trimmed Mean (Tr), Median (M.D) and Harmonic Mean (H.M) in bandwidth factor as follows:

$$b(x_i) = h\Psi_{Tr}, \quad b(x_i) = h\Psi_{M.D}, \quad b(x_i) = h\Psi_{H.M},$$

where

$$\Psi_{Tr} = \left[\frac{\bar{\mathfrak{f}}(x_i)}{Tr}\right]^{-\alpha}, \quad \Psi_{M.D} = \left[\frac{\bar{\mathfrak{f}}(x_i)}{M.D}\right]^{-\alpha}, \quad \Psi_{H.M} = \left[\frac{\bar{\mathfrak{f}}(x_i)}{H.M}\right]^{-\alpha}.$$

There defined (N-W) kernel estimators are as given below:

$$\hat{m}_6(x) = \frac{\sum_{i=1}^{n} \frac{y_i}{h\Psi_{Tr}} K\left(\frac{x - X_i}{h\Psi_{Tr}}\right)}{\sum_{i=1}^{n} \frac{1}{h\Psi_{Tr}} K\left(\frac{x - X_i}{h\Psi_{Tr}}\right)}, \tag{2.6}$$

$$\hat{m}_7(x) = \frac{\sum_{i=1}^{n} \frac{y_i}{h\Psi_{M.D}} K\left(\frac{x - X_i}{h\Psi_{M.D}}\right)}{\sum_{i=1}^{n} \frac{1}{h\Psi_{M.D}} K\left(\frac{x - X_i}{h\Psi_{M.D}}\right)}, \tag{2.7}$$

$$\hat{m}_8(x) = \frac{\sum_{i=1}^{n} \frac{y_i}{h\Psi_{H.M}} K\left(\frac{x - X_i}{h\Psi_{H.M}}\right)}{\sum_{i=1}^{n} \frac{1}{h\Psi_{H.M}} K\left(\frac{x - X_i}{h\Psi_{H.M}}\right)}. \tag{2.8}$$

The aim of the present paper is to propose some new improvements of the AN-W kernel regression utilizing some non-traditional and robust measures. Note that in upcoming lines, we will denote $\hat{m}_1(x) - \hat{m}_8(x)$ as given below:

- N-W
- V.N-W
- G.M.N-W
- A.M.N-W
- R.N-W
- Tr.N-W
- M.D.N-W
- H.M.N-W.

3. Proposed class of N-W kernel regression estimators

In current section, initially we describe some non-traditional and robust measures. After that, we define novel class of adaptive Nadaraya-Watson kernel regression estimators.

3.1. Hodges-Lehmann (H-L) estimator

The $H - L$ estimator was proposed by [12]. Initially, it was suggested to predict the location parameter of one-dimensional populations, and then it was used for many more reasons, such as evaluating the differences between two populations' members based on their generalized vector sampling population from univariate to multivariate. The set of all feasible subsets of a dataset with n observations it has $n(n+1)/2$ elements. The mean is calculated for each such subset; then, the average of these $n(n+1)/2$ averages is defined as the location H-L estimator. The construction of (H-L) based on median of the pairwise Walsh averages [11].

3.2. Mid-range (M-R) estimator

The mid-range (M-R) estimator is a measure of central tendency defined as the average of the two extreme values in the data set. Rider [15] conducted a research on the distribution of M-R samples from five distinct symmetric populations of restricted range, and on the relative efficiency of M-R and mean in estimating the M-R/mean/median population. He stated that the M-R is more effective than the mean for all the populations studied (for more details see [4]). The (M-R) calculates as: $\dfrac{X_L + X_M}{2}$, where X_L and X_M denoting the minimum/lowest and maximum/highest order statistics of X. The construction of (M-R) based upon extreme values [9].

3.3. Tri-mean (T-M) estimator

The Tri-mean (T-M) estimator is a non-traditional measure of a probability distribution's location parameter. The T-M estimator does not only take the central tendency into consideration but it also provides due significance to data distribution, so, it is described as a weighted average of the distribution's mid quartile or median (Q_2) and its lower and upper quartiles (Q_1 and Q_3). For the lower quartile, 25% of the data set are less than or equal to its value, for the median 50% of the data set are less than or equal to its value and for the upper quartile 75% of the data set are less than or equal to its value. For more details see [3]. (T-M) is the weighted mean of quartiles and calculated as: $\dfrac{Q_1 + 2Q_2 + Q_3}{4}$, see [21].

3.4. Quartile-deviation (Q-D) estimator

The Quartile deviation (Q-D) estimator is a positional and absolute measure of data dispersion in any series which tries to minimize the error of range as a measure of dispersion. Unlike range, it avoids the use of extreme values and in its place uses the difference of and as a measure of dispersion. Since the Q-D estimator is equal to half of the IQR "Inter-Quartile Range" so sometimes it named the Semi IQR. It is calculated as: $\dfrac{Q_3 - Q_1}{2}$. For more details see [18].

3.5. MCD estimator

Rousseeuw [16] presented a minimum covariance determinant (MCD) robust estimator with a high breakdown point. Since Rousseeuw and Van Driessen [17] designed the computationally efficient fast algorithm, the MCD estimator appears to be used widely in many fields also used to supports the development of many robust multivariate techniques.

In the MCD method, all k-subsets $(n/2 < k < n)$ out of the dataset with n observations are considered. The MCD estimator is given by a subset of k observations with the lowest covariance determinant. The MCD scale estimate is the covariance matrix of those k observations.

3.6. Proposed versions of N-W

In current sub-section, we propose some new versions of N-W estimator using Hodges-Lehmann (H-L), Mid-Range (M-R), Tri-Mean (T-M) and Quartile-Deviation (Q-D), taking inspiration from [6], [13] and [2] work. The new versions of N-W estimators based on additive combination of non-traditional measures i.e. (H-L, M-R, T-M, Q-D) with the robust MCD scale estimator, instead of (G.M, A.M, H.M, Tr), during the computation of bandwidth factor in N-W function. The proposed versions of N-W are as given below:

$$b(x_i) = h\Psi_{P1} = h(\Psi_{H-L} + \Psi_{MCD}), \quad b(x_i) = h\Psi_{P2} = h(\Psi_{M-R} + \Psi_{MCD}),$$

$$b(x_i) = h\Psi_{P3} = h(\Psi_{T-M} + \Psi_{MCD}), \quad b(x_i) = h\Psi_{P4} = h(\Psi_{Q-D} + \Psi_{MCD}).$$

where

$$\Psi_{H-L} = \left[\frac{\bar{\mathfrak{f}}(x_i)}{H-L}\right]^{-\alpha}, \quad \Psi_{M-R} = \left[\frac{\bar{\mathfrak{f}}(x_i)}{M-R}\right]^{-\alpha}, \quad \Psi_{T-M} = \left[\frac{\bar{\mathfrak{f}}(x_i)}{T-M}\right]^{-\alpha},$$

$$\Psi_{Q-D} = \left[\frac{\bar{\mathfrak{f}}(x_i)}{Q-D}\right]^{-\alpha}, \quad \Psi_{MCD} = \left[\frac{\bar{\mathfrak{f}}(x_i)}{MCD}\right]^{-\alpha}.$$

Hence, our defined (N-W) kernel estimators are as given below:

$$\hat{m}_{P1}(x) = \frac{\sum_{i=1}^{n} \frac{y_i}{h\Psi_{P1}} K\left(\frac{x - X_i}{h\Psi_{P1}}\right)}{\sum_{i=1}^{n} \frac{1}{h\Psi_{P1}} K\left(\frac{x - X_i}{h\Psi_{P1}}\right)}, \tag{3.1}$$

$$\hat{m}_{P2}(x) = \frac{\sum_{i=1}^{n} \frac{y_i}{h\Psi_{P2}} K\left(\frac{x - X_i}{h\Psi_{P2}}\right)}{\sum_{i=1}^{n} \frac{1}{h\Psi_{P2}} K\left(\frac{x - X_i}{h\Psi_{P2}}\right)}, \tag{3.2}$$

$$\hat{m}_{P3}(x) = \frac{\sum_{i=1}^{n} \frac{y_i}{h\Psi_{P3}} K\left(\frac{x - X_i}{h\Psi_{P3}}\right)}{\sum_{i=1}^{n} \frac{1}{h\Psi_{P3}} K\left(\frac{x - X_i}{h\Psi_{P3}}\right)}, \tag{3.3}$$

$$\hat{m}_{P4}(x) = \frac{\sum_{i=1}^{n} \frac{y_i}{h\Psi_{P4}} K\left(\frac{x - X_i}{h\Psi_{P4}}\right)}{\sum_{i=1}^{n} \frac{1}{h\Psi_{P4}} K\left(\frac{x - X_i}{h\Psi_{P4}}\right)}. \tag{3.4}$$

Note that in upcoming lines, we will denote $\hat{m}_{P1}(x) - \hat{m}_{P4}(x)$ as given below:

- P1.N-W
- P2.N-W
- P3.N-W
- P4.N-W.

It is worth mentioning that the authors, Demir and Toktamis [6], Khulood and Lutfiah [13], and, Ali [2] haven't showed the properties of estimators analytically. So, in this article we have made an attempt to give the properties of estimators analytically. In general, let we represent estimators $\hat{m}_{P1}(x) - \hat{m}_{P4}(x)$ as $\hat{m}_P(x)$ and $\Psi_{P1} - \Psi_{P4}$ denoted as Ψ_P, respectively. Now, taking motivation from [20]. The bias, variance, mean square error (MSE) and mean integrated square error (MISE) of $\hat{m}_P(x)$ is

$$B(\hat{m}_P(x)) = \frac{(h\Psi_P)^2}{2}\mu_K\left(m''_P(x) + \frac{2m'_P(x)g'_P(x)}{g_P(x)}\right) + O(h\Psi_P)^2, \tag{3.5}$$

$$V(\hat{m}_P(x)) = \left(\frac{\sigma^2\sigma_K^2}{g_P(x)}\right)\left(\frac{1}{nh\Psi_P}\right) + O\left(\frac{1}{nh\Psi_P}\right), \tag{3.6}$$

where $g_P(x)$ is the probability density function of the covariates $X_1, X_2, \ldots, X_n$ and $\mu_K = \int x^2K(x)dx$. Further, σ^2 is the error of regression model and $\sigma_K^2 = \int K^2(x)dx$ is a constant of kernel function. As we know that

$$MSE(\hat{m}_P(x)) = [B(\hat{m}_P(x))]^2 + V(\hat{m}_P(x)), \tag{3.7}$$

$$\begin{aligned} MSE(\hat{m}_P(x)) &= \frac{(h\Psi_P)^4}{4}\mu_K^2\left(m''_P(x) + \frac{2m'_P(x)g'_P(x)}{g_P(x)}\right)^2 \\ &+ \left(\frac{\sigma^2\sigma_K^2}{g_P(x)}\right)\left(\frac{1}{nh\Psi_P}\right) + O(h\Psi_P)^4 + O\left(\frac{1}{nh\Psi_P}\right), \end{aligned}$$

and MISE is

$$\begin{aligned} MISE(\hat{m}_P(x)) &= \frac{(h\Psi_P)^4}{4}\mu_K^2\int\left(m''_P(x) + \frac{2m'_P(x)g'_P(x)}{g_P(x)}\right)^2 dx \\ &+ \left(\frac{\sigma^2\sigma_K^2}{nh\Psi_P}\right)\int\left(\frac{1}{g_P(x)}\right)dx + O(h\Psi_P)^4 + O\left(\frac{1}{nh\Psi_P}\right). \end{aligned}$$

Similarly, properties of [6], [13], and [2] estimators can also be obtained.

4. Numerical illustration

Two artificial populations (Pop-1 & Pop-2) and one real-life data set (Pop-3) considered for the purposes of the article in upcoming sub-sections.

4.1. Simulation Study (Pop-1 & Pop-2)

For simulation study, let us consider two simulated data sets. Simulated populations generated from the following regression model:

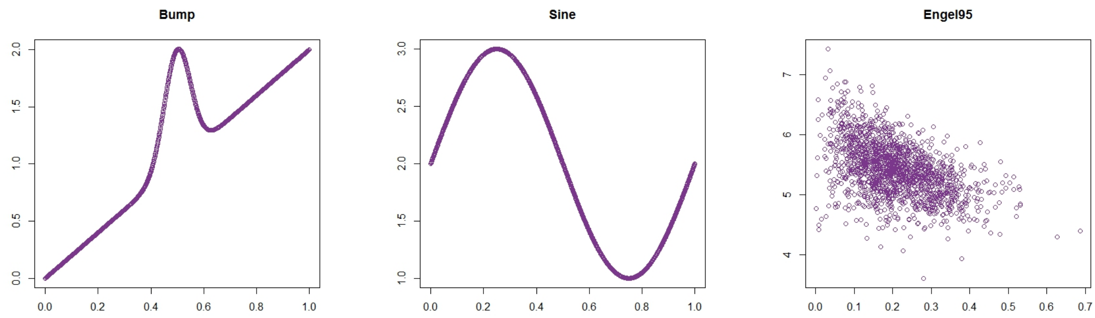
$$y_i = m(X_i) + \in_i, \quad 1 \leqslant i \leqslant 1500, \tag{4.1}$$

with the functions

$$m(X_i) = 2\left(x - \frac{1}{2}\right) + \exp\left(-200\left(x - \frac{1}{2}\right)^2\right) + 1, \quad (Bump) \tag{4.2}$$

$$m(X_i) = sin(2x\pi) + 2. \quad (Sine) \tag{4.3}$$

The errors ($\in_i$) are normally distributed with zero mean and variance 0.1 [6]. Graphical representation of simulated populations are also available in Figure 1. We draw four different sample sizes i.e. 10, 20, 100 and 150 for both data sets. Both populations in Figure 1 shows non-linearity, hence suitable for non-parametric techniques. Further, Gaussian kernel function is used for the purposes of the article with $h = (0.5, 0.75, 1)$.



(a) Pop-1 Plot. **(b)** Pop-2 Plot. **(c)** Pop-3 Plot.

Figure 1. Simulated and real populations.

The steps of simulation can be summarized (in points), as follows:

- A sample containing n units is selected from Pop-1.
- All the versions of N-W i.e. (N-W, V.N-W, G.M.N-W, A.M.N-W, R.N-W, Tr.N-W, M.D.N-W, H.M.N-W, P1.N-W, P2.N-W, P3.N-W, P4.N-W) are estimated with $h = 0.5, 0.75, 1$, using the sample data of above step, at all values of x (Ali 2019).
- Obtain MSE through each version of N-W i.e. $MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$.
- Steps (1),(2) and (3) are repeated $Z' = 7000$ (say) times.
- The overall MSE is computed for each up-to Z' and than averaged as: $AMSE = \frac{\sum_{j=1}^{Z'}(MSE_j)}{Z'}$.
- Through overall MSE computations, we find the percentage relative efficiency (PRE) of each N-W w.r.t classical N-W.

Where $\hat{y}_i$ denotes estimated values through all the versions of N-W. Note that for Pop-2 we adapt same steps and calculate overall MSEs' and hence their PREs, respectively. The PREs' are available in Tables 1–4 for both populations.

Table 1. PREs of Pop-1 for small sample sizes (n=10,20).

	$n = 10$			$n = 20$		
$\hat{\theta}$	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$
N-W	100	100	100	100	100	100
V.N-W	117.43	165.42	220.05	130.17	175.74	228.04
G.M.N-W	280.29	147.04	86.24	281.24	147.76	86.79
A.M.N-W	280.30	147.05	86.25	281.40	147.78	86.84
R.N-W	31.58	13.33	6.21	63.08	24.13	11.04
Tr.N-W	292.43	159.08	98.26	293.51	159.82	98.81
M.D.N-W	293.00	159.25	98.31	294.03	159.98	98.87
H.M.N-W	292.28	159.04	98.24	293.08	159.73	98.78
P1.N-W	508.18	544.67	333.50	504.24	542.86	332.64
P2.N-W	505.41	543.85	333.20	501.80	542.08	332.33
P3.N-W	508.70	544.83	333.56	487.91	543.16	332.32
P4.N-W	332.48	173.37	103.55	385.24	194.23	113.68

Table 2. PREs of Pop-1 for large sample sizes (n=100,150).

$\hat{\theta}$	$n = 100$			$n = 150$		
	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$
N-W	100	100	100	100	100	100
V.N-W	131.36	179.35	233.98	144.10	189.67	241.97
G.M.N-W	294.22	160.97	100.17	295.17	161.69	100.72
A.M.N-W	294.23	160.98	100.18	295.33	161.71	100.77
R.N-W	34.51	16.26	9.14	66.01	27.06	13.97
Tr.N-W	294.36	161.01	100.19	295.44	161.75	100.74
M.D.N-W	294.93	161.18	100.24	295.96	161.91	100.80
H.M.N-W	294.21	160.97	100.17	295.01	161.66	100.71
P1.N-W	522.11	558.60	347.43	518.17	556.79	346.57
P2.N-W	519.34	557.78	347.13	515.73	556.01	346.26
P3.N-W	522.63	558.76	347.49	501.84	557.09	346.25
P4.N-W	337.41	178.30	108.48	390.17	199.16	118.61

Table 3. PREs of Pop-2 for small sample sizes (n=10,20).

$\hat{\theta}$	$n = 100$			$n = 150$		
	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$
N-W	100	100	100	100	100	100
V.N-W	136.84	183.25	236.65	150.80	193.71	244.75
G.M.N-W	299.68	159.45	97.05	295.40	158.41	96.80
A.M.N-W	299.93	159.49	97.06	295.52	158.43	96.81
R.N-W	68.16	23.57	9.81	56.55	20.27	8.60
Tr.N-W	300.20	159.57	97.09	295.55	158.45	96.77
M.D.N-W	301.35	159.92	97.26	295.69	158.49	96.83
H.M.N-W	299.44	159.42	97.04	295.27	158.39	96.74
P1.N-W	658.45	565.83	347.59	535.62	558.96	345.02
P2.N-W	652.90	564.15	346.95	534.98	558.71	344.92
P3.N-W	659.89	566.26	347.75	539.83	559.03	345.05
P4.N-W	409.88	200.93	116.45	385.30	192.73	113.01

Table 4. PREs of Pop-2 for large sample sizes (n=100,150).

$\hat{\theta}$	$n = 100$			$n = 150$		
	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$
N-W	100	100	100	100	100	100
V.N-W	140.62	187.03	240.43	154.58	197.49	248.53
G.M.N-W	303.46	163.23	100.83	299.18	162.19	100.58
A.M.N-W	303.71	163.27	100.84	299.30	162.21	100.59
R.N-W	71.94	27.35	13.59	60.33	24.05	12.38
Tr.N-W	303.98	163.35	100.87	299.33	162.23	100.55
M.D.N-W	305.13	163.70	101.04	299.47	162.27	100.61
H.M.N-W	303.22	163.20	100.82	299.05	162.17	100.52
P1.N-W	662.23	569.61	351.37	539.40	562.74	348.80
P2.N-W	656.68	567.93	350.73	538.76	562.49	348.70
P3.N-W	663.67	570.04	351.53	543.61	562.81	348.83
P4.N-W	413.66	204.71	120.23	389.08	196.51	116.79

4.2. Real-life application (Pop-3)

British cross-section data (Engel95), consisting of a random sample taken from the British Family Expenditure Survey for 1995. The households consist of married couples with an employed head-of-household between the ages of 25 and 55 years. There are 1655 household-level observations in total. Data set contained ten variables. However, we consider two variables from the available ten variables namely, "Expenditure on food" as dependent variable and "Logarithm of total expenditure i.e. (logexp)" as independent variable. For more details see e.g. [14].

For diagnostic checking, either we should apply traditional linear regression on the referenced data set, let us check the presence of two major issues which may occur in a data set and may disturb the classical assumptions of OLS, namely, heteroskedasticity and autocorrelation. For assessment of heteroskedasticity issue, we apply Breusch-Pagan test (bptest) and Ramseys RESET test (resettest). For assessment of autocorrelation issue, we apply Durbin-Watson test (dwtest) and Breusch-Godfrey test (bgtest). The numerical descriptives' of these tests are also provided in Table 5. As we see in Table 5, all the P-values < 0.05, of referenced tests. Hence, providing clear indication of the presence of heteroskedasticity and autocorrelation issues. Also from scatter plot in figure 3, we can see the non-linear relationship between dependent and independent variables. In light of diagnostic tests and scatter plot, we can say that traditional OLS is not suitable for our data set. So there is a need to incorporate a some sort of technique, which can provide us better results even under the presence of such issues i.e. heteroskedasticity and autocorrelation. One of the solution is to incorporate nonparametric (kernel) regression. So we apply nonparametric (kernel) regression technique utilizing existing and proposed versions of N-W kernel estimator. The objective of applying different versions of N-W is to compare the performance between them on the premise of predictive ability and highlight the most enhanced N-W kernel estimator. It is worth mentioning that we are using Gaussian kernel and the same values of h as we utilized for artificial populations in Section 4.1. The results of PREs' are available in Table 6 and 7.

Table 5. Results of diagnostic tests for (Engel95).

bptest	BP = 108.36	df = 1	p-value = 2.2e-16
resettest	RESET = 4.0872	df = 1	p-value = 0.01696
dwtest	DW = 1.8059	-	p-value = 3.868e-05
bgtest	BPLM = 15.315	df = 1	p-value = 9.097e-05

Table 6. PREs of (Engel95) for small sample sizes (n=10,20).

	$n=10$			$n=20$		
$\hat{\theta}$	$h=0.5$	$h=0.75$	$h=1$	$h=0.5$	$h=0.75$	$h=1$
N-W	100	100	100	100	100	100
V.N-W	132.64	179.05	232.45	146.60	189.51	240.55
G.M.N-W	295.48	155.25	92.85	291.20	154.21	92.60
A.M.N-W	295.73	155.29	92.86	291.32	154.23	92.61
R.N-W	63.96	19.37	5.61	52.35	16.07	4.40
Tr.N-W	296.00	155.37	92.89	291.35	154.25	92.57
M.D.N-W	297.15	155.72	93.06	291.49	154.29	92.63
H.M.N-W	295.24	155.22	92.84	291.07	154.19	92.54
P1.N-W	654.25	561.63	343.39	531.42	554.76	340.82
P2.N-W	648.70	559.95	342.75	530.78	554.51	340.72
P3.N-W	655.69	562.06	343.55	535.63	554.83	340.85
P4.N-W	405.68	196.73	112.25	381.10	188.53	108.81

Table 7. PREs of (Engel95) for large sample sizes (n=100,150).

	$n = 100$			$n = 150$		
$\hat{\theta}$	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$
N-W	100	100	100	100	100	100
V.N-W	133.76	167.98	208.40	145.45	179.67	220.09
G.M.N-W	211.19	141.53	100.21	222.88	153.22	111.90
A.M.N-W	211.70	141.62	100.25	223.39	153.31	111.94
R.N-W	108.23	58.98	36.16	119.92	70.67	47.85
Tr.N-W	212.28	141.91	100.37	223.97	153.60	112.06
M.D.N-W	214.55	142.85	100.81	226.24	154.54	112.50
H.M.N-W	210.66	141.42	100.81	222.35	153.11	112.50
P1.N-W	583.55	387.47	272.95	595.24	399.16	284.64
P2.N-W	571.20	381.92	270.31	582.89	393.61	282.00
P3.N-W	585.83	388.31	273.36	597.52	400.00	285.05
P4.N-W	296.67	183.27	123.07	308.36	194.96	134.76

4.3. Findings of numerical illustration

The computational results of numerical study are available in Tables 1, 3 and 6. These tables report the predictive ability performance of different versions of N-W, based on their PRE. Some major highlights are as follows:

- Tables 1 and 2 show the computational results of first simulated population. In Table 1, P3.N-W has the maximum PRE with sample size n=100, and the maximum PRE is associated with P1.N-W for sample size n=150 when h=0.5 or 1 while it returns to associated with P3.N-W when h=0.75.
- Tables 3 and 4 show the computational results of second simulated population. In Table 3, P3.N-W has the maximum PRE with both sample sizes i.e. n=100 and n=150.
- Tables 6 and 7 show the computational results of real-life data set. In Tables 6 and 7, P3.N-W has also the maximum PRE.
- In all the referenced populations, R.N-W has the least PRE as compare to all the proposed and existing versions of N-W.
- In all the referenced populations, P4.N-W is not providing improved results w.r.t. proposed versions of N-W. However, P4.N-W providing an improvement in results w.r.t. existing ones except for V.N-W.
- First three proposed versions (P1.N-W, P2.N-W, P3.N-W) are better than existing ones, in all the referenced populations. However, P3.N-W is dominating all the proposed and existing versions of N-W.
- By increasing sample size, MSE is decreasing. Hence, PRE is increasing by increasing sample size.

According to numerical illustration, we can say, proposed versions of N-W are the best ones except P4.N-W, because P4.N-W is only performing better with the minimum value of bandwidth i.e. h=0.5, at-least for the numerical experiments considered in current article. Which emphatically showed the improved performance in the sense of predictive ability of N-W versions. So the utilization of proposed versions is highly recommended, except P4.N-W.

5. Conclusion

In this study, taking inspiration from recent utilization's of N-W kernel regression, we proposed some new versions of N-W based on additive combinations of non-traditional measures with the robust MCD scale estimator. The merits of the existing and proposed versions have been evaluated by simulation study with different values of smoothing parameter i.e. h and Gaussian kernel function. A real-life application data (Engel95) is also considered for the purposes of the article. For incorporating (Engel95) data set according to our existing and proposed setting, we examined the issues of heteroskedasticity and autocorrelation through Breusch-Pagan test, Ramseys RESET test, Durbin-Watson test and Breusch-Godfrey test. As these issues were present, hence the considered data set is suitable for our existing and proposed tools. The PRE computations are shown numerically in Tables 6 and 7. These computations highlight the predominance of the proposed N-W versions, at-least for the experimental circumstances considered in the article. Hence, the proposed versions of N-W are recommended for the enhancement of the predictive ability of kernel regression technique.

Acknowledgment. The authors thank and extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups Project under grant number R.G.P. 2/406/44.

References

[1] I.S. Abramson, *On bandwidth variation in kernel estimates-a square root law*, Ann. Statist. **10** (4), 1217-1223, 1982.

[2] T.H. Ali, *Modification of the adaptive Nadaraya-Watson kernel method for nonparametric regression (simulation study)*, Comm. Statist. Simulation Comput. **51** (2), 391-403, 2022.

[3] D.F Andrews, P.J. Bickel, F.R. Hampel, P.J. Huber, W.H. Rogers and J.W. Tukey, *Robust Estimates of Location: Survey and Advances*, Princeton University Press, Princeton, New Jersey, 1972.

[4] G.R. Arce and S.A. Fontama, *On the midrange estimator*, IEEE Trans. Acoust., Speech, Signal Process. **36** (6), 920-922, 1988.

[5] S. Demir, *Adaptive kernel density estimation with generalized least square cross-validation*, Hacet. J. Math. Stat. **48** (2), 616-625, 2019.

[6] S. Demir and O. Toktamis, *On the adaptive nadaraya-watson kernel regression estimators*, Hacet. J. Math. Stat. **39** (3), 429-437, 2010.

[7] A. Eftekharian and M. Razmkhah, *On estimating the distribution function and odds using ranked set sampling*, Statist. Probab. Lett. **122**, 1-10, 2017.

[8] A. Eftekharian and H. Samawi, *On kernel-based quantile estimation using different stratified sampling schemes with optimal allocation*, J. Stat. Comput. Simul. **91** (5), 1040-1056, 2021.

[9] E.B. Ferrell, *Control charts using midranges and medians*, Industrial Quality Control **9** (5), 30-34, 1953.

[10] M. Hanif, S. Shahzadi, U. Shahzad and N. Koyuncu, *On the adaptive Nadaraya-Watson kernel estimator for the discontinuity in the presence of jump size*, Suleyman Demirel Univ. Fen Bilim. Enst. Derg. **22** (2), 511-520, 2018.

[11] T.P. Hettmansperger and J.W. McKean, *Robust Nonparametric Statistical Methods*, Arnold, London, 1998.

[12] J.L. Hodges and E.L. Lehmann, *Estimates of location based on rank tests*, Ann. Math. Statist. **34** (2), 598-611, 1963.

[13] H.A. Khulood and I.A. Lutfiah, *Modification of the adaptive Nadaraya-Watson kernel regression estimator*, Sci. Res. Essays **9** (22), 966-971, 2014.

[14] Q. Li and J.S. Racine, *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, 2007.

[15] P.R. Rider, *The midrange of a sample as an estimator of the population midrange*, J. Amer. Statist. Assoc. **52** (280), 537-542, 1957.

[16] P.J. Rousseeuw, *Multivariate estimation with high breakdown point*, in: W. Grossmann, G. Pflug, I. Vincze and W. Wertz (ed.) Mathematical Statistics and Applications, Reidel Publishing Company, Dordrecht, 283297, 1985.
[17] P.J. Rousseeuw and K. van Driessen, *A fast algorithm for the minimum covariance determinant estimator*, Technometrics **41** (3), 212-223, 1999.
[18] R.S. Saksena, *A Hand Book of Statistics*, Motilal Banarsidass, Delhi, 1981.
[19] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.
[20] M.P. Wand and M.C. Jones, *Kernel Smoothing*, CRC Press, 1994.
[21] T. Wang, Y. Li and H. Cui, *On weighted randomly trimmed means*, J. Syst. Sci. Complex. **20** (1), 47-65, 2007.
[22] E. Zamanzade and M. Vock, *Variance estimation in ranked set sampling using a concomitant variable*, Statist. Probab. Lett. **105**, 1-5, 2015.
[23] E. Zamanzade and X. Wang, *Improved nonparametric estimation using partially ordered sets*, in: Statistical Methods and Applications in Forestry and Environmental Sciences, Springer, Singapore, 57-77, 2020.