

*Full Length Research Paper***Modification of the adaptive Nadaraya-Watson kernel regression estimator**

Khulood Hamed Aljuhani\* and Lutfiah Ismail Al turk

Statistics Department, King Abdulaziz University, Kingdom of Saudi Arabia, Saudi Arabia.

Received 14 October, 2014; Accepted 14 November, 2014

**Nadaraya-Watson (NW) kernel regression estimator is a widely used and flexible nonparametric estimator of a regression function, which is often obtained by using a fixed bandwidth. Several studies showed that the adaptive kernel estimators with varying bandwidths have better performance results. In this paper, a new improvement of the NW kernel regression estimator is proposed and the bandwidth of this new improvement is obtained depending on the range of the observations. Simulated example is presented, including comparisons with three others NW estimators. The performance of the proposed new estimator is evaluated via the MSE criterion. The results of the simulation study were very promising; it shows that our modified NW estimator performs well in all cases.**

**Key words:** Nonparametric estimation, smoothing parameter, local bandwidth factor, Nadaraya-Watson kernel regression estimator, modified Nadaraya-Watson (NW) estimator.

**INTRODUCTION**

In many statistical problems, nonparametric regression techniques are commonly used for describing the relationship between a response variable and some covariates. Let  $\{(X_i, Y_i)\}_{i=1}^n \in \mathbb{R}$  be a random sample of bivariate data with size  $n$ . The nonparametric regression model is defined as:

$$Y_i = m(X_i) + \varepsilon_i, \quad (1)$$

where

$m(x_i)$ : is unknown regression function, and

$\varepsilon_i$ : are independent random errors with zero mean and

variance  $\sigma^2$ ;  $i = 1, 2, \dots, n$ .

The nonparametric regression techniques are weighted

averages of the response variable, where the weights depend on the technique and the distance between the observations of the explanatory variable scaled by a smoothing parameter. One of the nonparametric regression estimation techniques is the Nadaraya-Watson (NW) kernel estimator. It is more flexible than the other nonparametric methods, and provides an accurate predictor of observations. The NW kernel estimator was first proposed by Nadaraya (1964) and Watson (1964). Nadaraya (1964) introduced the NW estimator as an approximation to the regression curve based on empirical data. He studied the properties of his suggested estimator when the sample size increases infinitely. Watson (1964) presented the NW estimator as a simple computer method for obtaining a "graph" from a large number of observations.

\*Corresponding authors. E-mail: [khaljuhani@kau.edu.sa](mailto:khaljuhani@kau.edu.sa), [lturk@kau.edu.sa](mailto:lturk@kau.edu.sa)

Author(s) agree that this article remain permanently open access under the terms of the [Creative Commons Attribution License 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

The NW kernel estimator depends on one parameter which is called the bandwidth; it controls the amount of curve smoothing where large  $h$  produces a smooth density estimate (Wand and Jones, 1995). The bandwidth of the NW kernel estimator can be fixed or variable; the choice of the optimal bandwidth is a critical issue. The optimal bandwidth is the value that minimized the mean integrated squared error (MISE) which can be obtained by integrating the mean squares of errors (MSE). Several methods of selecting  $h$  can be used, Silverman (1986), Wand and Jones (1995) and Härdle et al. (2004) expanded in the bandwidth selections methods. One of these methods is the least square cross-validation or also called unbiased cross-validation; Scott and Terrell (1987) discussed it and presented a relationship between the biased and unbiased cross-validation. The variable bandwidth should be used rather than the fixed bandwidth in the case of long-tail or multi-modal distributions. Abramson (1982) suggested the inverse-square-root rule for the bandwidth  $h$  of a variable-kernel density, which reduces the bias more than the fixed-bandwidth estimator, even when a nonnegative kernel is used. Silverman (1986) discussed the kernel density estimation exhaustively. He gave details about the assumptions of the kernel weight and the properties of the estimator such as bias and variance. In addition, he proposed an adaptation for the kernel estimator by varying the bandwidth as nonparametric density estimation. Demir and Toktamış (2010) considered the adaptive Nadaraya-Watson (ANW) kernel regression estimators as a way to estimate the regression function. The results of their simulation study showed that the NW kernel estimator has a better performance when evaluating the local bandwidth factor based on the arithmetic mean instead of using the geometric mean. Also, their results did not oppose the previous studies in that the NW kernel estimator with the variable bandwidths is better than the fixed NW kernel estimator.

The purpose of this paper is to propose a new modification of the NW kernel regression estimator. The bandwidth of our modification is obtained by using the range of the probability density function of  $X_i$ . The idea behind our modification is that by increasing the local bandwidth factor and thus the bandwidth, better performance of the NW kernel estimator will be obtained. In more details, different selected types of the NW estimators, including our modified NW kernel estimator are presented in our study. Also, a brief description of the MSE criterion is given. Finally, a simulation study is conducted with useful concluding remarks given.

**METHODS**

An important factor that has a great impact on the smoothing results is the choice of the bandwidth or the smoothing parameter

$h$ . Here, different Nadaraya-Watson kernel estimators are presented according to the selected type of the bandwidth.

**Fixed Nadaraya-Watson kernel estimator**

The bandwidth can be selected to be a constant over all the range of  $x$ ; this choice is suitable when the unknown regression function behaves the same over all the estimation range. The NW kernel estimator is often obtained with a fixed bandwidth which can be defined as:

$$\hat{m}_{NW}(X_i) = \int y f(y|x) dy = \frac{\int y \hat{f}(x,y) dy}{\hat{f}(x)}$$

$$= \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)} \tag{2}$$

where

$h$ : is the fixed bandwidth,  $h > 0$ , and

$K$ : is the kernel function which satisfying the following conditions

(Silverman, 1986):

- i)  $\int_{-\infty}^{\infty} K(u)du = 1$
- ii)  $\int_{-\infty}^{\infty} u K(u)du = 0$
- iii)  $\int_{-\infty}^{\infty} u^2 K(u)du = \mu_2 \neq 0$
- iv)  $V = \int_{-\infty}^{\infty} K(u)^2 du < \infty$

Several kernel functions are proposed in the literature. Gaussian kernel function is one of the most commonly used in practice (Härdle, 1990; Silverman, 1986); the Gaussian kernel function is defined as

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \tag{3}$$

The fixed bandwidth can be selected depending on various methods such as; Silverman’s rule of thumb and cross-validation. In this paper, the least square cross-validation (LSCV) will be used according to its simple evaluation and its ability to be applied in any regression model. The LSCV minimizes the integrated squared error (ISE) rather than the MISE (Scott and Terrell, 1987), where MISE is the average of the ISE, and ISE is a distance measured between the fitted density  $\hat{f}(x)$  and the true density  $f(x)$  which is defined as

$$ISE = \int [\hat{f}(x) - f(x)]^2 dx \tag{4}$$

and

$$MISE = E(ISE)$$

The  $h_{LSCV}$  is the bandwidth that minimize the LSCV which is defined as

$$LSCV = \int \hat{f}^2(x) dx - \frac{2}{n} \hat{f}_{-i}(x_i) \tag{5}$$

where  $\hat{f}_{-i}(x_i)$ : is the *leave-one-out* kernel density estimator, which is obtained among the remaining  $n - 1$  observations and can be defined by the following equation

$$\hat{f}_{-i}(x_i) = \frac{1}{(n-1)h} \sum_{i \neq j} K\left(\frac{x_i - x_j}{h}\right) \tag{6}$$

**Variable Nadaraya-Watson kernel estimator**

The fixed NW kernel estimator is not a good choice for the cases of multivariate, long-tailed, and the multi-modal distributions. The multivariate problem can be handled by increasing the sample size, but for the cases of the long-tail and the multi-modal distributions, varying the bandwidth is recommended. The estimator which is based on varying the bandwidth is called the variable NW kernel estimator, and has the following form:

$$\hat{m}_{NW}(x_i) = \frac{\sum_{i=1}^n \frac{y_i}{h(X_i)} K\left(\frac{x - X_i}{h(X_i)}\right)}{\sum_{i=1}^n \frac{1}{h(X_i)} K\left(\frac{x - X_i}{h(X_i)}\right)} \tag{7}$$

where  $h(X_i)$ : is the variable bandwidth.

Abramson (1982) gave the following formula to compute  $h(X_i)$ :

$$h(X_i) = \frac{h}{\sqrt{f(X_i)}}$$

where  $f(X_i)$ : is the probability density function of  $X_i$  which can be estimated by the kernel density estimator.

The variable bandwidth  $h(X_i)$  can be obtained by the Silverman algorithm which is presented in Silverman (1986). He presented in his paper, an algorithm for the Abramson style estimator, and referred to it as an adaptive kernel estimator. In the first step, he obtained the prior kernel estimator with fixed bandwidth  $h$  which is denoted by  $\tilde{f}(X_i)$ . Then, he defined the local bandwidth factor  $\lambda_i$ , as:

$$\lambda_i = \left[ \frac{\tilde{f}(x_i)}{g} \right]^{-\alpha} \tag{8}$$

where

$g$ : is the geometric mean of  $\tilde{f}(X_i)$ ,  $g \neq 0$ , and

$\alpha$ : is the sensitivity parameter, which satisfies  $0 \leq \alpha \leq 1$ .

At the last step, his suggested adaptive bandwidth is defined as:

$$h(X_i) = h \lambda_i \tag{9}$$

In (1982), Abramson gave the sensitivity parameter the value 0.5 since this value leads to good prediction results. Then, the variable NW kernel estimator can be written as follow:

$$\hat{m}_{NW}(x_i) = \frac{\sum_{i=1}^n \frac{y_i}{h \lambda_i} K\left(\frac{x - X_i}{h \lambda_i}\right)}{\sum_{i=1}^n \frac{1}{h \lambda_i} K\left(\frac{x - X_i}{h \lambda_i}\right)} \tag{10}$$

**Adaptive Nadaraya-Watson kernel estimator**

Demir and Toktamış (2010) modified the NW kernel estimator, their modification based on using the arithmetic mean instead of the geometric mean when computing the local bandwidth factor which is given as

$$\lambda_i^* = \left[ \frac{\tilde{f}(X_i)}{\bar{X}} \right]^{-0.5} \tag{11}$$

where  $\bar{X}$ : is the arithmetic mean of  $\tilde{f}(X_i)$ .

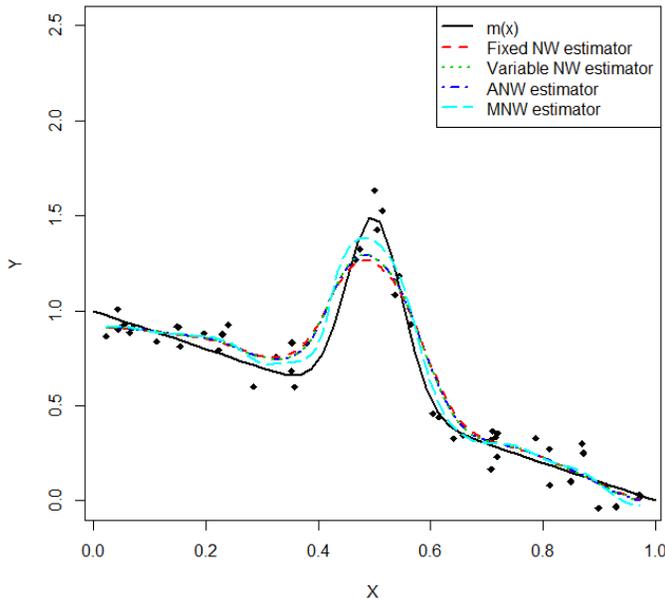
Then, the ANW kernel estimator is defined by Equation 10 with replacing  $\lambda_i$  by  $\lambda_i^*$ . The authors used the arithmetic mean since its value is greater than or equal to the geometric mean (Lidstone, 1932), and that has made the value of the  $\lambda_i^*$  greater than  $\lambda_i$ . By

maximizing the value of the local bandwidth factor, the value of the bandwidth will increase too, and this will enhance the performance of the NW estimator. According to their simulation study and real data application, they showed that the performance of the ANW kernel estimator is better than the performance of the NW kernel estimator.

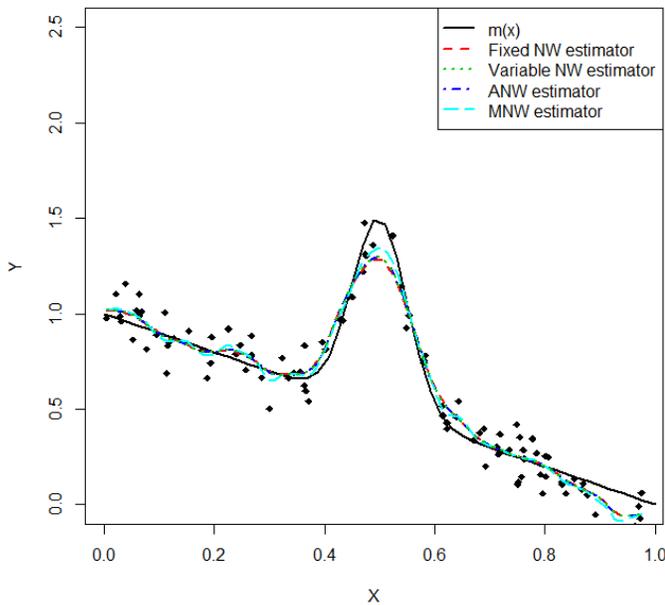
**Modified Nadaraya-Watson kernel estimator**

This part of the paper is dedicated to our modification which aims to enhance the predictive ability of the NW kernel estimator through increasing the value of the bandwidth. In our proposed NW kernel estimator we suggest to evaluate the local bandwidth factor based on the range of the observations instead of using the geometric or the arithmetic mean. Most of the times, the range will have a larger value, particularly if the phenomenon being studied has outliers. Thus, the local bandwidth factor and the value of the bandwidth  $h$  will be increased. The modified local bandwidth factor is given as:

$$\lambda_i' = \left[ \frac{\tilde{f}(x)}{R} \right]^{-0.5} \tag{12}$$



**Figure 1.** The real regression function and the NW kernel estimators of the regression function using sample size 50 and  $h=0.16$ .



**Figure 1.** The real regression function and the NW kernel estimators of the regression function using sample size 100 and  $h=0.08$ .

where  $R$ : is the range of  $\tilde{f}(X_i)$ , which is the difference between the largest and smallest values.

Therefore, the modified Nadaraya-Watson (MNW) kernel estimator can be obtained by substituting  $\lambda'_i$  instead of  $\lambda_i$  in Equation (10).

**Evaluation criteria**

For the selection of the best performance NW kernel estimators, several criteria can be used. In general, the evaluation criteria are based on computing the distance between the observed values  $y_i$  and the predicted values  $\hat{y}_i$  which are obtained by using the estimated models. Here, the mean squared error (MSE) will be used. The best estimator is the one with the smallest MSE value. The MSE can be computed mathematically by using the following formula:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{13}$$

$n$ : is the number of observations.

**Simulation study**

Here, the performance of the new proposed MNW kernel estimator is examined over three different selected NW kernel estimators; the fixed NW, the variable NW, and the ANW through a simulation study. The explanatory variables are generated from the uniform distribution based on the interval  $[0,1]$  with six different sample sizes 25, 50, 100, 250, 300 and 600. The regression function is given by Hardle (1990) as:

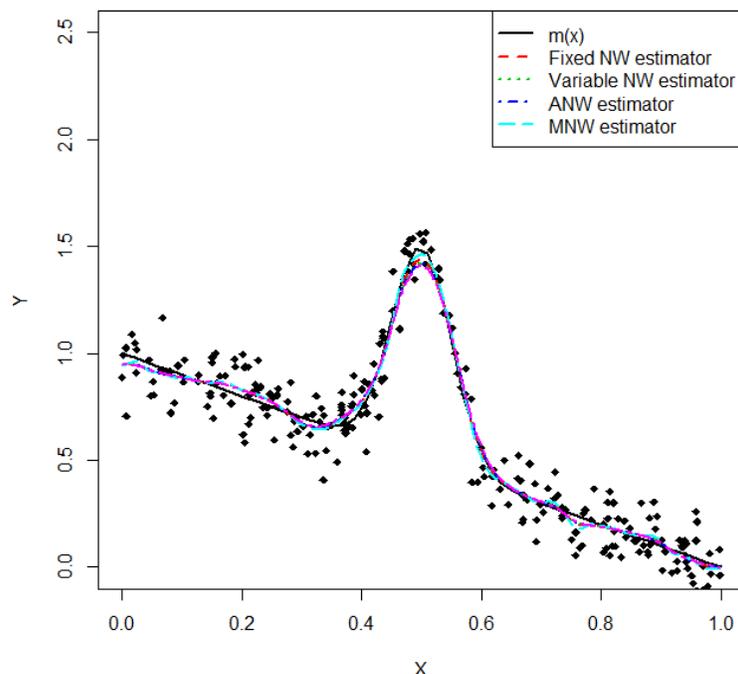
$$Y_i = 1 - X_i + e^{[-200(x_i-0.5)^2]} + \varepsilon_i \tag{14}$$

Where the random errors  $\varepsilon_i$  have normal distribution with 0 mean and 0.1 variance. The fixed bandwidth  $h$  is obtained by using the unbiased cross-validation method. To evaluate the NW kernel estimators, the Gaussian kernel function is used. A 1000 simulation repetition for each sample size is used to compute the MSE criterion.

The graphs of the real regression function and the estimated regression functions which are computed based on the sample of sizes 50, 100, 250 are presented in Figures 1, 2, and 3. While the performance of the MNW kernel estimator comparable with the three selected NW kernel estimators is considered objectively, our comparable study is based on the MSE criterion. For each sample size, the MSE value of the fixed NW, variable NW, ANW and MNW kernel estimators which is based on Gaussian kernel function are computed. The results of the MSE criterion is presented in Table 1.

**RESULTS**

From the figures, it is clear that the performance of the MNW is superior to the fixed NW, the variable NW and the ANW kernel estimators. Also, the figures show that generally the performance of all the studied NW kernel estimators becomes better by increasing the sample size.



**Figure 3.** The real regression function and the NW kernel estimators of the regression function using sample size 250 and  $h=0.06$ .

**Table 1.** The MSE values of the NW kernel estimators.

| Kernel functions | Sample size | Fixed NW kernel estimator | Variable NW kernel estimator | ANW kernel estimator | MNW kernel estimator |
|------------------|-------------|---------------------------|------------------------------|----------------------|----------------------|
| Gaussian         | 25          | 6.6588                    | 4.1133                       | 3.9618               | 3.8548               |
|                  | 50          | 3.5671                    | 2.1655                       | 2.0818               | 2.0120               |
|                  | 100         | 1.8340                    | 1.1040                       | 1.0570               | 1.0221               |
|                  | 250         | 0.7573                    | 0.4685                       | 0.4457               | 0.4265               |
|                  | 300         | 0.6325                    | 0.3959                       | 0.3764               | 0.3593               |
|                  | 600         | 0.3205                    | 0.2120                       | 0.2021               | 0.1920               |

And according to the simulation results in Table 1 we can conclude that:

- 1) The variable NW estimator gives noticeably better prediction results than the fixed NW estimator.
- 2) The ANW estimator has better performance than the variable NW estimator; same results were obtained by Demir and Toktamış in (2010).
- 3) Our suggested estimator gives better predictive capability in all cases.
- 4) All the estimators are enhanced by increasing the sample size.

Generally and according to our study and the study of Demir and Toktamış in (2010), we can conclude that any modification that aims to increase the local bandwidth factor will give an improved prediction results.

## DISCUSSION

The Nadaraya-Watson (NW) kernel estimator is a non-parametric method that can be used for regression estimation, it is an easy and flexible method and has previously been shown to provide an accurate prediction results. In this paper we have proposed a new NW kernel regression estimator as a modification to the adaptive Nadaraya-Watson kernel estimator. Our suggestion based on enhancing the predictive ability of the ANW kernel estimator through increasing the value of the Local bandwidth factor by using the range instead of the arithmetic mean when calculating the bandwidths. By conducting a simulation study with different sample sizes, various NW kernel regression estimators have been compared with our new suggested kernel estimator. The

MNW kernel estimator seems to be superior to the other three NW kernel estimators in all cases. This estimator was more stable in comparison with the other kernel estimators; so according to our study when aiming to estimate the regression function the MNW kernel estimator is recommended.

### Conflict of Interest

The authors have not declared any conflict of interest.

### REFERENCES

- Abramson IS (1982). On bandwidth variation in kernel estimates-a square root law. *The annals of Statistics*. 10(4):1217-1223. <http://dx.doi.org/10.1214/aos/1176345986>
- Demir S, Toktamiş Ö (2010). On the Adaptive Nadaraya-Watson Kernel Regression Estimators. *Hacettepe J. Math. Stat.* 39(3):429-437.
- Härdle W (1990). *Applied nonparametric regression*. Cambridge University Press, Cambridge. <http://dx.doi.org/10.1017/CCOL0521382483>
- Härdle W, Werwatz A, Müller M, Sperlich S (2004). *Nonparametric and semiparametric models: An Introduction*. Springer, New York. <http://dx.doi.org/10.1007/978-3-642-17146-8>
- Lidstone GJ (1932). 1023: The Arithmetic and Geometric Mean. *Math. Gazette*. 16(218):127-127.
- Nadaraya EA (1964). On Estimating Regression, *Theory of Probability & Its Applications*. *Soc. Ind. Appl. Math.* 9(1):141-142.
- Scott DW, Terrell GR (1987). Biased and Unbiased Cross-Validation in Density Estimation. *J. Am. Stat. Assoc.* 82(400):1131-1146. <http://dx.doi.org/10.1080/01621459.1987.10478550>
- Silverman BW (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, New York. <http://dx.doi.org/10.1007/978-1-4899-3324-9>
- Wand MP, Jones MC (1995). *Kernel Smoothing*. Chapman & Hall, London. <http://dx.doi.org/10.1007/978-1-4899-4493-1>
- Watson GS (1964). Smooth Regression Analysis. *Indian J. Stat. Series A*:359-372.