Hacettepe Journal of Mathematics and Statistics
Volume 39 (3) (2010), 429 – 437

ON THE ADAPTIVE NADARAYA-WATSON KERNEL REGRESSION ESTIMATORS

S. Demir*† and Ö. Toktamiş‡

Received 11 : 02 : 2009 : Accepted 15 : 03 : 2010

Abstract

Nonparametric kernel estimators are widely used in many research areas of statistics. An important nonparametric kernel estimator of a regression function is the Nadaraya-Watson kernel regression estimator which is often obtained by using a fixed bandwidth. However, the adaptive kernel estimators with varying bandwidths are specially used to estimate density of the long-tailed and multi-mod distributions. In this paper, we consider the adaptive Nadaraya-Watson kernel regression estimators. The results of the simulation study show that the adaptive Nadaraya-Watson kernel estimators have better performance than the kernel estimations with fixed bandwidth.

Keywords: Nonparametric regression, Nadaraya-Watson kernel estimator, Adaptive kernel estimation, Kernel density estimation.

2000 AMS Classification: 62 G 08, 62 G 07.

1. Introduction

For given data points $\{(X_i, Y_i)\}_{i=1}^n \in \mathbb{R}$, let us assume that the regression model is

$$Y_i = m(X_i) + \varepsilon_i,\ i = 1, \ldots, n,$$

with observation errors ε_i and unknown regression function m. Assume that the response variable Y depends on an independent random variable X. Also that ε is a random variable with mean 0 and variance σ^2. As is well known, $m(x)$ is a conditional mean curve

$$m(x) = \mathrm{E}(y/x) = \int \frac{yf(x,y)}{f(x)}\, dy,$$

*Muğla University, Faculty of Arts & Sciences, Department of Statistics, 48000 Muğla, Turkey. E-mail: serdardemir@mu.edu.tr

†Corresponding Author.

‡Department of Statistics, Faculty of Science, Hacettepe University, 06532 Beytepe, Ankara, Turkey. E-mail: oniz@hacettepe.edu.tr

where $f(x,y)$ is the joint density function of (X,Y) and $f(x)$ is the marginal density function of $f(x)$. An estimation of this regression function can be taken as

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

If $\widehat{f}(x) = 0$ then $\widehat{m}(x)$ is defined to be 0. In (1.1), $\widehat{f}(x,y)$ is an estimation of $f(x,y)$, and $\widehat{f}(x)$ is an estimation of $f(x)$. Using kernel estimations instead of estimations of the density functions in the numerator and denominator of (1.1), a nonparametric kernel estimation of the regression function can be obtained.

A kernel estimation

$$\widehat{f}(x) = \frac{1}{nh_1}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_1}\right)$$

can be used instead of the density function estimation $\widehat{f}(x)$ which occurs in the denominator of (1.1). Here h_1 is a fixed smoothing parameter for the kernel density estimation $\widehat{f}(x)$, and K is a symmetric probability density function [4, 8]. This is known as a "kernel function", and satisfies the following general assumptions [7].

A1) $\int K(u)\, du = 1$,
A2) $\int uK(u)\, du = 0$,
A3) $\int u^2 K(u)\, du = \mu_2(K) \neq 0$.

Epanechnikov and Gaussian are the kernel functions which are used most often in practice [7, 2]. The Epanechnikov kernel function is

$$K(u) = 3(1 - u^2)/4,\ |u| \leq 1,$$

and the Gaussian function is

$$K(u) = e^{(-u^2/2)}/\sqrt{2\pi},\ -\infty < u < \infty.$$

In the numerator of (1.1) the multiplicative kernel estimation

$$\hat{f}(x,y) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_1 h_2} K^{[2]}\Big(\frac{x - X_i}{h_1},\ \frac{y - Y_i}{h_2}\Big) \tag{1.2}$$

in $\mathbb{R}\times\mathbb{R}$ can be used instead of the marginal density function estimation $\hat{f}(x,y)$ [3]. Here, $K^{[2]}$ (in 2-dimensional space) is a bivariate kernel function, and h_2 a fixed smoothing parameter for the kernel density estimation $\hat{f}(y)$. Using a single smoothing parameter h, instead of different parameters h_1 and h_2, and substituting the kernel estimators (1.2) and (1.3) in (1.1), the Nadaraya-Watson kernel estimator of the regression function is obtained as

$$\hat{m}_{NW}(x) = \frac{\sum_{i=1}^{n} Y_i K\big(\frac{x - X_i}{h}\big)}{\sum_{i=1}^{n} K\big(\frac{x - X_i}{h}\big)}. \tag{1.3}$$

The smoothing parameter h of the Nadaraya-Watson kernel estimator controls the smoothing level of the estimation, and is called the "bandwidth". The bandwidth h plays a very important role in the performance of the kernel estimators. Various methods of choosing h are available. The methods used mostly are cross-validation, penalized functions, plug-in and bootstrap [5]. The question as to which is the best is controversial. The cross-validation method is often preferred because it is easily computable and applicable for any regression model. In the cross-validation method, the bandwidth which minimizes

the cross-validation (CV) function with a nonnegative weight function $w(X_i)$ is obtained as

$$\text{CV}(h) = n^{-1} \sum_{i=1}^{n} \{Y_i - \hat{m}(X_i)\}^2 w(X_i), \tag{1.4}$$

see [2]. The CV function in (1.5) contains the leave-one-out kernel estimator

$$\hat{m}_i(X_i) = \frac{\sum_{j \neq i}^{n} Y_j K(\frac{X_i - X_j}{h})}{\sum_{j \neq i}^{n} K(\frac{X_i - X_j}{h})}. \tag{1.5}$$

The leave-one-out estimator $\hat{m}_i(X_i)$ is obtained over the remaining $n-1$ data after leaving out X_i and Y_i. The bandwidth that minimizes the cross-validation function also minimizes the integrated mean square error, which is a performance criterion of the estimator.

2. Adaptive kernel estimators of the density function

The kernel estimator of the probability density function with fixed bandwidth given by (1.2) is not sufficient for long tail distributions, multi-mode distributions and the multivariate case. Silverman [7] showed this is the case in a study using right-long tailed data. Silverman's procedures are based on a varying bandwidth. One estimator which is used with varying bandwidth is the adaptive kernel (or sample point) estimator.

In the case of one variable, the adaptive kernel estimator which uses different bandwidths for the data point X_i is,

$$\hat{f}_U(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h(X_i)} K \left\{ \frac{x - X_i}{h(X_i)} \right\}. \tag{2.1}$$

Here the varying bandwidth $h(X_i)$ is an adaptive bandwidth which depends on X_i. For any dimension, Abramson [1] proposed a method (the square-root rule) which uses a value of $h(X_i)$ proportional to $f(X_i)^{-1/2}$.

Silverman [7] gave an algorithm with three steps for the Abramson-type estimator. At the first step, a prior kernel estimator $\tilde{f}(X_i)$ with a fixed bandwidth is obtained. At the second step, the local bandwidth factor λ_i is defined as

$$\lambda_i = \{\tilde{f}(X_i)/g\}^{-\alpha},$$

where g (assuming $g \neq 0$) is the geometric mean of $\tilde{f}(X_i)$, and α is called the *sensitivity parameter*, which satisfies $0 \leq \alpha \leq 1$. At the last step, for one variable the kernel estimator is obtained as

$$\hat{f}_U(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h\lambda_i} K\Big(\frac{x - X_i}{h\lambda_i}\Big). \tag{2.2}$$

As seen from (2.2), the adaptive bandwidth h is taken as $h(X_i) = h\lambda_i$. The adaptive kernel estimation is equivalent to the kernel estimation with fixed bandwidth when the sensitivity parameter α is equal to 0. When $\alpha = 1$, then the adaptive kernel estimation is equivalent to the nearest neighbor estimation.

Abramson [1] and Silverman [7] emphasized that taking $\alpha = 0.5$ leads to good results.

Let $(X_i,\ Y_i)$ be a random sample from a population which has the density function $f(x, y)$, $(i = 1, \ldots, n)$. The kernel estimation of the bivariate joint density function is given by Epanechnikov as follows:

$$\hat{f}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_X h_Y} K^{[2]}\Big(\frac{x - X_i}{h_X},\ \frac{y - Y_i}{h_Y}\Big).$$

A bivariate kernel function can be obtained as

$$K^{[2]}\left(\frac{x-X_i}{h_X},\ \frac{y-Y_i}{h_Y}\right) = K_1\left(\frac{x-X_i}{h_X}\right) K_2\left(\frac{y-Y_i}{h_Y}\right)$$

by using multiplicative kernel functions [2]. Using the same kernel functions $K_1 = K_2 = K$, the kernel estimator of the bivariate probability density function for X and Y is

$$\hat{f}(x,y) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_X h_Y} K\left(\frac{x-X_i}{h_X}\right) K\left(\frac{y-Y_i}{h_Y}\right).$$

This estimator was used to obtain the Nadaraya-Watson kernel estimator with fixed bandwidth in Equation (1.4).

Using varying bandwidths instead of fixed bandwidths, Sain [6] gives the adaptive multiplicative kernel estimator (in a d-dimensional space) of the multivariate density function as

$$\hat{f}_U(x_1,\ldots,x_d) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_{1i}\cdots h_{di}} \left[\prod_{j=1}^{d} K\left\{\frac{x_j - X_{ij}}{h(X_{ij})}\right\}\right]$$

for variables $x_1,\ldots,x_d$ with n observations. Thus, the adaptive multiplicative kernel estimator of the bivariate density function is obtained as

$$\hat{f}_U(x,y) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h(X_i)\,h(Y_i)} K\left\{\frac{x-X_i}{h(X_i)}\right\} K\left\{\frac{y-Y_i}{h(Y_i)}\right\}. \tag{2.3}$$

3. Adaptive Nadaraya-Watson kernel estimators

Here, we use the estimators $\hat{f}_U(x)$ and $\hat{f}_U(x,y)$ of the density function to estimate the regression function in Equation (1.1). Plugging $\hat{f}_U(x)$ and $\hat{f}_U(x,y)$ into the numerator and denominator of Equation (1.1), we obtain the adaptive Nadaraya-Watson (NWU) kernel estimator with varying bandwidths as follows (for the proof see the appendix):

$$\begin{aligned} \hat{m}_{NWU}(x) &= \int \frac{y\hat{f}_U(x,y)}{\hat{f}_U(x)}\,dy \\ &= \frac{\sum_{i=1}^{n} \frac{Y_i}{\lambda_i} K\left(\frac{x-X_i}{\lambda_i h}\right)}{\sum_{i=1}^{n} \frac{1}{\lambda_i} K\left(\frac{x-X_i}{\lambda_i h}\right)}. \end{aligned} \tag{3.1}$$

The local bandwidth factors λ_i in Equation (3.1) can be determined by using the same three-stage algorithm given by Silverman to obtain the adaptive estimation of the density function. In practice, Abramson [1] and Silverman [7] propose that taking α equal to 0.5 leads to good results.

In addition, we want to see how using arithmetic mean instead of the geometric mean when computing the local bandwidths λ_i affects the performance of the adaptive Nadaraya-Watson kernel estimations. This is only for intuitive reasons. Thus, the modified local bandwidth factor λ_i^* is obtained as

$$\lambda_i^* = \left\{\tilde{f}(X_i)/a\right\}^{-\alpha}, \tag{3.2}$$

where $a = \sum_{i=1} \tilde{f}(X_i)/n$. Using the λ_i^* in Equation (3.2), the modified adaptive Nadaraya-Watson (NWUA) kernel estimator can be written as

$$\hat{m}_{NWUA}(x) = \frac{\sum_{i=1}^{n} \frac{Y_i}{\lambda_i^*} K\left(\frac{x-X_i}{\lambda_i^* h}\right)}{\sum_{i=1}^{n} \frac{1}{\lambda_i^*} K\left(\frac{x-X_i}{\lambda_i^* h}\right)}.$$

For comparing the performances of the Nadaraya-Watson and adaptive Nadaraya-Watson estimators, firstly we have tried to find the theoretical mean square errors of the estimators. But we could not obtain these due to mathematical difficulties. Therefore we focused on a simulation study. The results of the simulation study, whose aim is to compare the adaptive kernel Nadaraya-Watson NWU and the modified adaptive kernel Nadaraya-Watson NWUA will be given in next section.

4. Simulation results

A simulation study was conducted to compare the performances of the estimators with the classical Nadaraya-Watson estimators. For the simulation, we used the regression function given by Hardle [2] as

$$Y_i = 1 - X_i + e^{\left\{-200(X_i-0.5)^2\right\}} + \varepsilon_i, \tag{4.1}$$

where the X_i were drawn from a uniform distribution based on the interval $[0, 1]$. The ε_i have a normal distribution with 0 mean and 0.1 variance. In this way, we generated samples of size 25, 100, 250 and 500. The fixed bandwidth h was computed using the cross-validation method with $w(X_i) = 1$. The NW, NWU and NWUA kernel estimations were computed using the Epanechnikov and Gaussian kernel functions. The number of simulation repetitions for each estimation was 1000. The graphs of the real regression function and the estimations of the regression functions computed over a sample of 100 are illustrated in Figure 1 and Figure 2.

Figure 1. The regression curves of the kernel estimations using the Epanechnikov kernel for $h = 0.171$

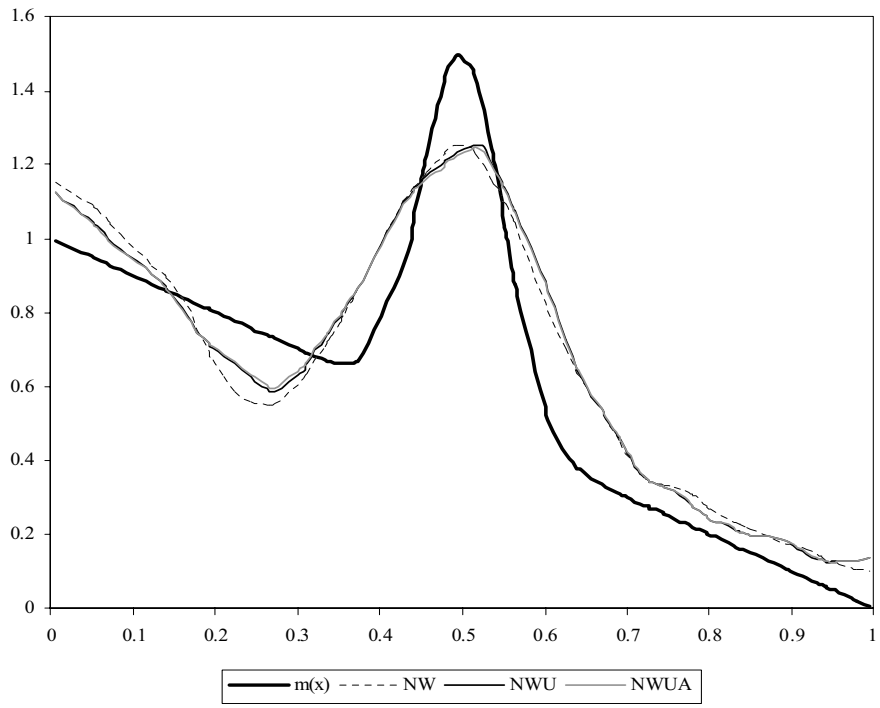
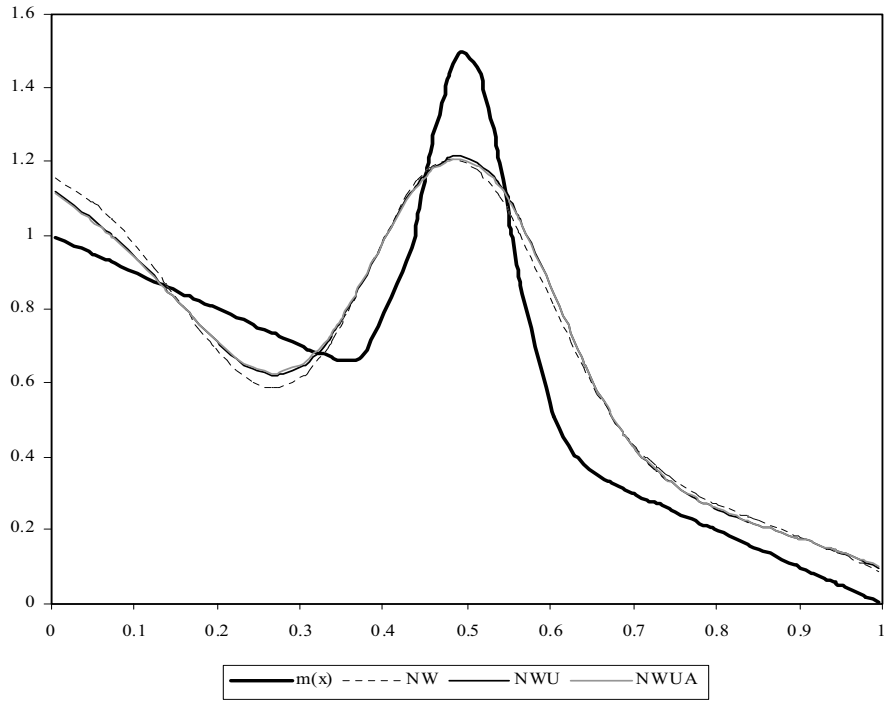


Figure 2. The regression curves of the kernel estimations using the Gaussian kernel for $h = 0.084$



For each sample, we computed the values of the mean square errors (MSE) related to the kernel estimations NW, NWU and NWUA. Finally we obtained an integrated MSE over the 1000 sample. The IMSE values of the kernel estimators which are obtained using the Epanechnikov and Gaussian kernel functions are given in Table 1.

Table 1. IMSE values of the estimations for the Epanechnikov kernel

n	NW	NWU	NWUA
25	179.16	175.60	173.39∗
100	76.14	74.57	74.12∗
250	39.98	39.29	39.19∗
500	21.20	20.92	20.88∗

∗Minimum IMSE in each row.

Table 2. IMSE values of the estimations for the Gaussian kernel

n	NW	NWU	NWUA
25	187.59	186.02	184.00∗
100	74.91	73.54	73.10∗
250	38.30	37.64	37.57∗
500	20.87	20.61	20.58∗

∗Minimum IMSE in each row.

As seen from Table 1 and Table 2, for all sample sizes, the kernel estimators NWU and NWUA using varying bandwidths for the Epanechnikov and Gaussian kernels have smaller IMSE values than the NW kernel estimator with fixed bandwidth. In each case, it is seen that NWUA has the best performance.

In addition, comparing Table 1 and Table 2, in the case of a small sample size ($n = 25$), we see that the kernel estimations NW, NWU and NWUA computed using the Epanechnikov kernel function show a better performance than the estimations computed using the Gaussian kernel function.

5. A real data example

We apply the classical and adaptive Nadaraya-Watson kernel regression estimators described above to economics data coming from the Central Bank of the Republic of Turkey (http://tcmbf40.tcmb.gov.tr/cbt.html). We have 215 observation pairs (the monthly data between January 1989 and November 2006). The independent variable X is the effective exchange rate index (real; 1995 = 100). The dependent variable Y is total exports ($ Millions; Foreign Trade International Standard Industry Categorization-ISIC REVISE 3).

Figure 3 and Figure 4 show the regression curves of the computed Nadaraya-Watson kernel estimations with the Epanechnikov and Gaussian kernel functions respectively.

Figure 3. The regression curves of the kernel estimations with the Epanechnikov kernel function for the real dataset and $h = 2.82$

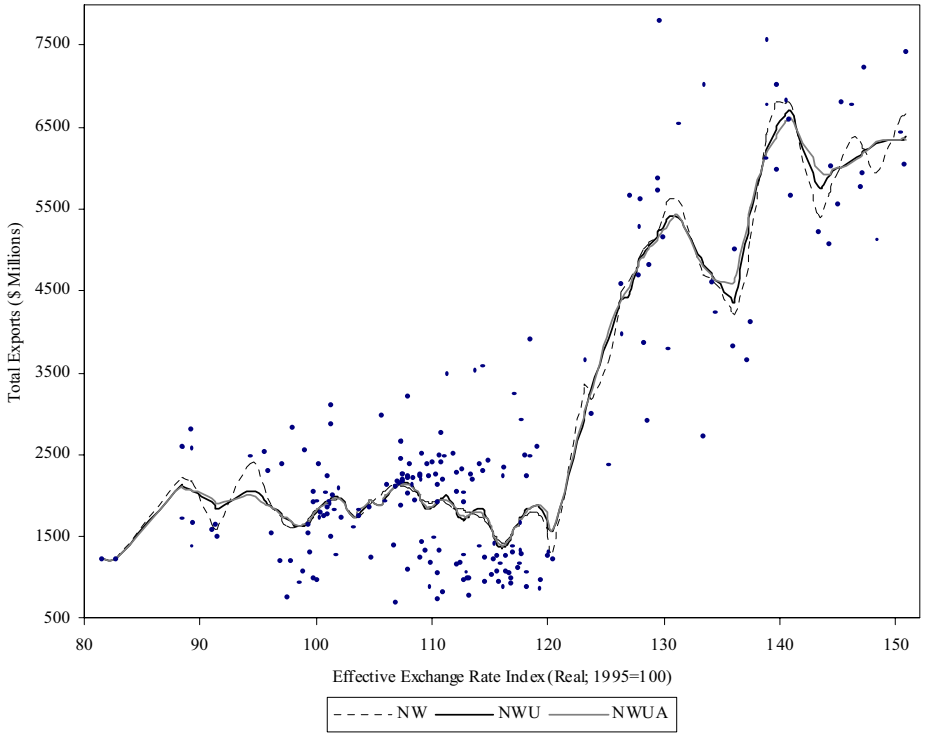
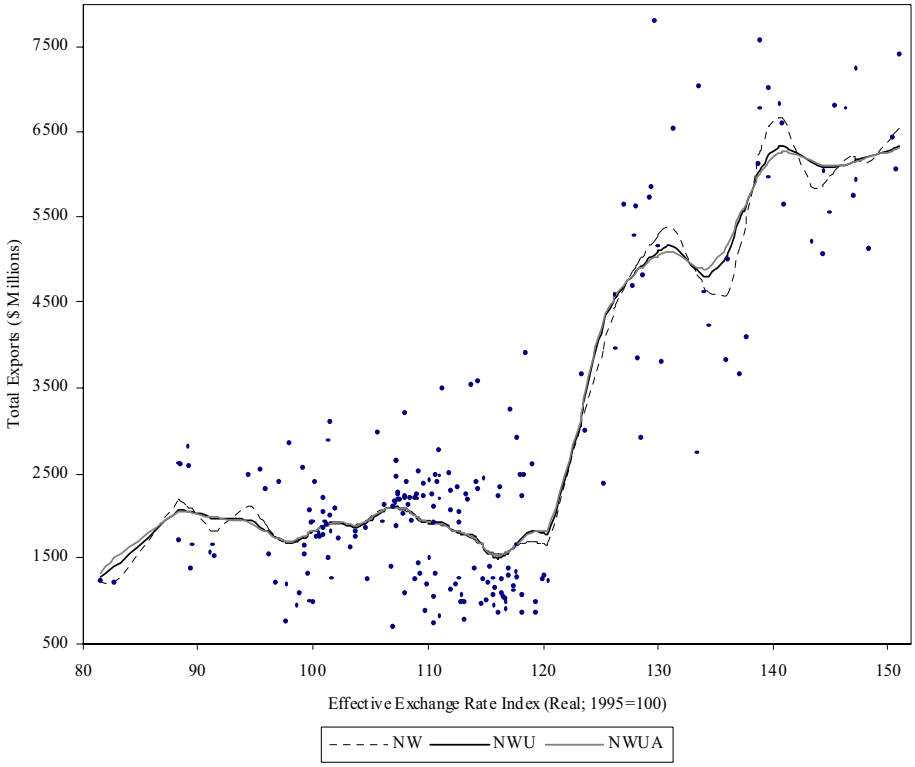


Figure 4. The regression curves of the kernel estimations with the Gaussian kernel function for the real dataset and $h = 1.47$



As seen from Figure 3 and Figure 4, the adaptive kernel estimations differ from the kernel estimations with fixed bandwidths especially in regions where the data points are sparsely located.

6. Conclusion

In this paper, we have studied the adaptive Nadaraya-Watson kernel estimators when used to estimate a regression function.

The results of the simulation study, which was performed to evaluate the performances of the kernel estimators considered, showed that the adaptive Nadaraya-Watson kernel regression estimators with varying bandwidths provide better estimates than the Nadaraya-Watson estimator with fixed bandwidth. In particular, the adaptive Nadaraya-Watson kernel regression estimator in which the bandwidths are obtained using the arithmetic mean instead of the geometric mean, leads to a better performance. Finally, the adaptive Nadaraya-Watson kernel regression estimators are preferable for estimating a regression function non-parametrically.

7. Appendix

The formula for the adaptive Nadaraya-Watson kernel regression estimator is obtained as follows:

$$\begin{aligned}\hat{m}_{NWU}(x) &= \int \frac{y\hat{f}_U(x,y)}{\hat{f}_U(x)}\,dy \\ &= \frac{1}{\hat{f}_U(x)}\int y\,\hat{f}_U(x,y)\,dy \\ &= \frac{1}{\hat{f}_U(x)}\int y\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h(X_i)h(Y_i)}K\left(\frac{x-X_i}{h(X_i)}\right)K\left(\frac{y-Y_i}{h(Y_i)}\right)dy \\ &= \frac{1}{\hat{f}_U(x)}\sum_{i=1}^{n}\frac{1}{nh(X_i)}K\left(\frac{x-X_i}{h(X_i)}\right)\int \frac{y}{h(Y_i)}\,K\left(\frac{y-Y_i}{h(Y_i)}\right)dy.\end{aligned}$$

Using the variable transformation $\big(y-Y_i\big)/h(Y_i\big)=t$, we get

$$\begin{aligned}\hat{m}_{NWU}(x) &= \frac{1}{\hat{f}_U(x)}\sum_{i=1}^{n}\frac{1}{nh(X_i)}K\left(\frac{x-X_i}{h(X_i)}\right)\int \big[h(Y_i)\,t+Y_i\big]\,K(t)\,dt \\ &= \frac{1}{\hat{f}_U(x)}\sum_{i=1}^{n}\frac{1}{nh(X_i)}K\left(\frac{x-X_i}{h(X_i)}\right)\left[h(Y_i)\int t\,K(t)\,dt+Y_i\int K(t)\,dt\right].\end{aligned}$$

Using Equation (2.1) and Assumptions A1, A2, we get the next formula as

$$\begin{aligned}\hat{m}_{NWU}(x) &= \frac{\sum_{i=1}^{n}\frac{Y_i}{nh(X_i)}K\big(\frac{x-X_i}{h(X_i)}\big)}{\hat{f}_U(x)} \\ &= \frac{\sum_{i=1}^{n}\frac{Y_i}{nh(X_i)}K\big(\frac{x-X_i}{h(X_i)}\big)}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h(X_i)}K\big(\frac{x-X_i}{h(X_i)}\big)}\end{aligned}$$

Taking $h(X_i)=\lambda_i h$, we obtain the adaptive Nadaraya-Watson kernel regression estimator as

$$\hat{m}_{NWU}(x)=\frac{\sum_{i=1}^{n}\frac{Y_i}{\lambda_i}K\big(\frac{x-X_i}{\lambda_i h}\big)}{\sum_{i=1}^{n}\frac{1}{\lambda_i}K\big(\frac{x-X_i}{\lambda_i h}\big)}.$$

Acknowledgment The authors are very grateful to the referee for helpful comments and suggestions to improve this paper.

References

[1] Abramson, I. *On bandwidth variation in kernel estimates - a square-root law*, Ann. Statist. **10**, 1217–1223, 1982.
[2] Hardle, W. *Applied Nonparametric Regression* (Cambridge, New Rochelle, 1990).
[3] Hardle, W. *Smoothing Techniques. With Implementation in S.* (Springer-Verlag, New York, 1991).
[4] Nadaraya, E. A. *On nonparametric estimates of density functions and regression curves*, Theory Appl. Probability **10**, 186–190, 1965.
[5] Pagan, A. and Ullah, A. *Nonparametric Econometrics* (Cambridge University Press, Cambridge, 1999).
[6] Sain, S. R. *Adaptive Kernel Density Estimation* (Unpublished Ph.D. Thesis, Department of Statistics, Rice University, 1994).
[7] Silverman, B. W. *Density Estimation for Statistics and Data Analysis* (Chapman & Hall, New York, 1986).
[8] Watson, G. S. *Smooth regression analysis*, Sankhya **26** (15), 175–184, 1964.