African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 3, 2020 (pp. 51-59)
www.abjournals.org

KERNEL FUNCTION AND NONPARAMETRIC REGRESSION ESTIMATION: WHICH FUNCTION IS APPROPRIATE?

Langat Reuben Cheruiyot[1], George O. Orwa[2] and Odhiambo Romanus Otieno[2]

[1]Department of Mathematics and Computer Science, University of Kabianga, P. O. Box 2030-20200, Kericho, Kenya.
[2]Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, P. O. BOX 62000-00200, Nairobi, Kenya.

ABSTRACT: *In regression estimation, researchers have the option of using parametric or nonparametric regression estimation. Because of the challenges that one can encounter as a result of model misspecification in the parametric type of regression, the nonparametric type of regression has become popular. This paper explores this type of regression estimation. Kernel estimation usually forms an integral part in this type of regression. There are a number of functions available for such a use. The goal of this study is to find out an appropriate function that can be used for weighting in regression estimation. Though from the theoretical results epanechnikov function is the optimal one, there are situations where Gaussian function may be advantageous. Simulations show that the estimates inherit the smoothness of the kernel functions used.*

KEYWORDS: Kernel Functions, Smooth Estimates, Density Estimation, Nonparametric Regression Estimation

INTRODUCTION

Many non-parametric techniques that can be used in regression estimation are in existence in the current researches. They include techniques such as the local polynomial regression, spline regression, and orthogonal series. In the framework of the model-based approach, regression estimation is paramount in obtaining estimates of the non-sample population. The flexible nature of the non-parametric technique has made it an attractive option in statistical researches. The technique entails use of kernel functions in assignment of weights to observations used in estimation.

This paper has been organized as follows: in section 2, we give a brief review of the literature regarding non-parametric regression and kernel functions, in section 3; density and regression estimation and use of various functions is explored. Empirical analysis has been done in section 4 using some artificially simulated datasets. Discussion of results and conclusion is given in section 5.

Kernel Function in Nonparametric Regression

As earlier stated, there are many kernel functions in literature that a researcher can use in nonparametric regression estimation. A kernel is simply a smoothing function or weight-assigning function. Different researchers require different assumptions to be made about the functions but most are common. The common assumptions include those that require it to be

symmetric and unimodal. The existence of some moments is also another common assumption though Fan and Gijbels (1992) require the existence of all moments. Though not universal, it is common to assume a bounded support for the kernel and that kernel is smooth, Avery (2010). The functions commonly used are the Gaussian function and those kernels derived from the Beta function with the parameter changing from 0, 1, 2 and 3 which respectively yields the uniform, Epanechnikov, biweight, and triweight kernels. Another kernel is the triangular kernel rarely used because it lacks the smoothness property, Wand and Jones (1995).

Within the kernel window observations may receive the same weights as in histograms or weights that reduce gradually as one moves away from the target observation where the kernel is centred. The kernel used in histogram in density estimation is typical of the uniform or the rectangular kernel- so called because it treats the points in a bin the same, in fact such a particular choice of a kernel is termed "naive" since weight of $\frac{1}{2}$ is assigned to all points regardless of how far or close the point is from the central point of the bin or window of the kernel, Di Nardo and Tobias (2001). A variety of kernel functions are possible in general, but both practical and theoretical considerations limit the choice, Härdle (1994). It is because of the knowledge that points that are closer to the centre of the window of a chosen kernel have closer association and thus can give more details or contribution on the target observation. These points undoubtedly deserve to be given more weight than the ones further away, Irizarry and Bravo (2010). This obviously rids out the use of kernels that assign equal weights across the window such as the uniform kernel function and leaves out those that assign reducing weights further away from the centre, see table 3.1 for the common kernel functions. It would also be desirable to use a kernel that optimizes the error criterion measurement such as AMISE. If this is the goal of the researcher then Epanechnikov would be the right choice. It is worth noting that the difference between this kernel and the others discussed and tabulated is not significant. In fact a slight increase in the sample size brings the corresponding efficiency at par with the optimal kernel. A disadvantage noted with Epanechnikov is that it has a discontinuous first derivative which may be undesirable, resulting to the choice of Gaussian function instead, Wand and Jones (1995). This is the reason why Gaussian kernel has taken for preference in this study. It can also be noted that this function has an optimal bandwidth choice in the event that the underlying distribution is normal.

Researchers have found out that even with this advantage the choice of the kernel function is not as important as that of the bandwidth itself, Faraway (2006). If one misses the optimal bandwidth that minimizes AMISE/MISE or other measure of accuracy by ten percent there is more drastic effect on the smoothing parameter than if one selected one of the "suboptimal" kernels, Härdle (1994).

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 3, 2020 (pp. 51-59)

www.abjournals.org

Kernel Functions Commonly used in Nonparametric Density and Regression Estimation

Some of the common kernel functions and their efficiencies are given in the table 3.1.

Table 3.1: Common Kernel Functions

Kernel	Equation	R(K)	$K_2(K)$	Eff(K)
Uniform	$K(z)=\frac{1}{2}I[\lvert z\rvert\leq 1]$	$\frac{1}{2}$	$\frac{1}{3}$	0.9295
Epanechnikov	$K(z)=\frac{3}{4}(1-z^2)I[\lvert z\rvert\leq 1]$	$\frac{3}{5}$	$\frac{1}{5}$	1.0000
Biweight	$K(z)=\frac{15}{16}(1-z^2)^2 I[\lvert z\rvert\leq 1]$	$\frac{5}{7}$	$\frac{1}{7}$	0.9939
Triweight	$K(z)=\frac{35}{32}(1-z^2)^3 I[\lvert z\rvert\leq 1]$	$\frac{350}{429}$	$\frac{1}{9}$	0.9867
Gaussian	$K(z)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}z^2)$	$\frac{1}{2\sqrt{\pi}}$	1	0.9512
Triangular	$K(z)=(1-\lvert z\rvert)I[\lvert z\rvert\leq 1]$	$\frac{2}{3}$	$\frac{1}{6}$	0.9859

At this point one may ask whether the choice of the kernel functions matter. It has been proven that the kernel that optimizes the AMISE, the measure of accuracy, is the Epanechnikov kernel Wand and Jones (1995). In table 3.1 it is important to note that Eff(k) which represents the efficiency of the kernel has been given relative to Epanechnikov kernel - the minimizer of AMISE and $R(K)=\int_{-\infty}^{\infty}K(z)^2dz$ is the roughness of the kernel. $K_2(K)$= $\int z^2K(z)dz$ is the second moment of the kernel- actually the spread ("variance") of the kernel density. From this information one can see that there are no much differences in these efficiencies, an indication that kernel selection has rather limited impact on them.

It is clear in the table that the uniform (also called the rectangular kernel) has an approximate efficiency value of 93%. The interpretation for this is that for n = 93, one can obtain an optimal AMISE with Epanechnikov kernel, while n = 100 would be required to obtain approximately the same value using the uniform kernel. Since both eventually lead to roughly the same estimate, it implies that identification of the appropriate kernel function should not be a big deal. This is the reason why in many researches the choice of kernel is based on other considerations like the desired smoothness, Alberts and Karunamuni (2007) and Zucchini (2003).

Frequently researchers may not opt for the uniform kernel function because it assigns constant weights across the observations in its window. In particular weights of ½ are assigned to points within the distance of h (the bandwidth) away from x- the point at the centre of the window. Points farther away are all assigned zero weights because the indicator function, $I[\lvert z\rvert\leq 1]$, is by definition equal to 0 for all values of the scaled distance, $z=(x-X_i)/h$, that are bigger than 1.

The others- the Epanechnikov, bi-weight, tri-weight, Gaussian and triangular kernels assign relatively smaller weights progressively away from the point at the centre. They assign more weight to the points closer to the centre. It should be noted that for all the kernels shown in

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 3, 2020 (pp. 51-59)

the table 3.1 their indicator functions, $I[|z| \leq 1]$, are by definition also equal to 0 for all values of the scaled distance, $z = (x - X_i)/h$, that are bigger than 1, except for Gaussian which is unbounded. The individual graphs of these functions are presented in Fig. 3.1 parts (a)-(f).

Standard Kernel Density Estimator

This section briefly highlights how the kernel density function works. This idea has been illustrated using Fig. 3.2 constructed from an artificial data set. It should be noted that while the area under the density estimate is equal to one, each of the rescaled kernel function has an area equal to $\frac{1}{n}$. This can be obtained by integration as follows.

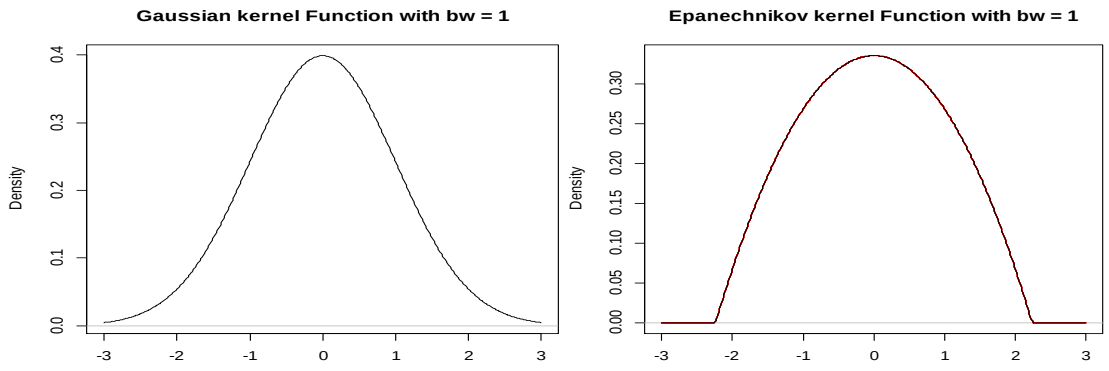
$$\int \frac{1}{nh} K\left(\frac{x - X_i}{h}\right) dx = \frac{1}{nh} \int K(z) h dz = \frac{1}{nh} h \int K(z) dz = \frac{1}{n} \quad (1)$$

This follows from condition $K(z) = K(-z)$. $K(.)$ denote a kernel function which is also twice continuously differentiable.

This is the sum over all the rescaled kernels (the little bumps). The estimated function is a result of the vertical sum of the bumps centred on the observed values of *x*.

Unlike histograms, kernel density estimates do not depend on the choice of the origin. They are smoother than histogram estimators since they inherit the smoothness of the kernel chosen and have a faster rate of convergence. As will be noticed later, increasing the bandwidth increases the amount of smoothing in the estimate (i.e. large $h(\to \infty)$ would give a flat estimate) while a small $h(\to 0)$ reveals the finer details of the distribution.

As stated earlier the choice of the kernel function is not very crucial as the bandwidth itself. For this reason and that of its ease in obtaining the optimal bandwidth during the bandwidth selection, the Gaussian kernel function has been used in this study. It also has an advantage in that the weights are always positive, Todd (2013)



(a) (b)

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 3, 2020 (pp. 51-59)
www.abjournals.org

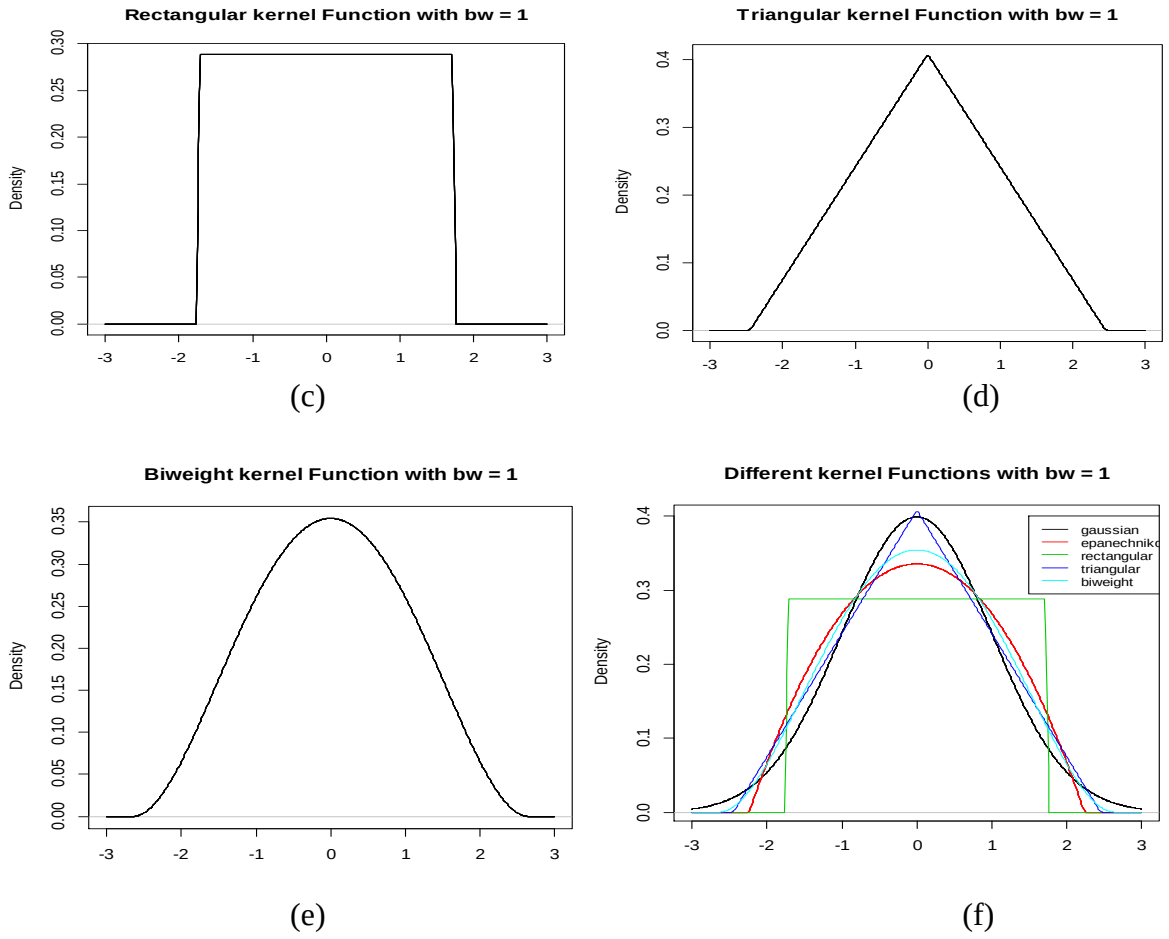


Fig. 3.1 Graphs of Some Selected Kernel Functions

The last graph of fig. 3.1 shows all the five selected graphs combined. Notice the shapes of each of the kernel function and how it will have an effect on the assignment of the respective weights.

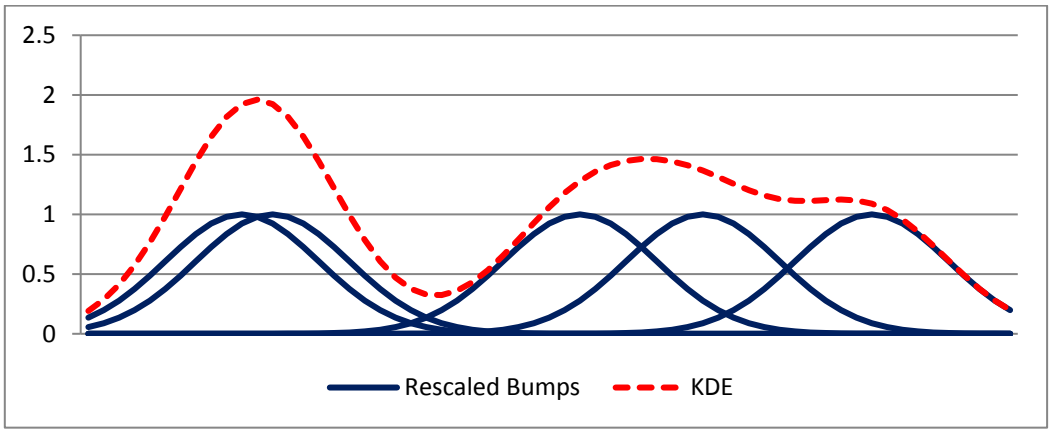


Fig. 3.2: Kernel Density Estimate Viewed as Sum of Bumps

Kernel Functions in Regression Estimation

The goal of this paper from the onset is to identify an appropriate kernel. This require practical comparison of the graphs as well as the theoretical considerations. With a simple simulation using the R software, the Gaussian kernel function was compared with the rest. The influence of the different selected kernel functions was checked, with the bandwidths and data kept constant. The resulting graphs of this study are given in fig. 4.1 for the density estimation and regression estimation in the subsequent graphs in fig. 5.1 given in the next section. The graphs in fig. 4.1 are from an artificial data set simulated using a random normal distribution with mean 20 and standard deviation of 1. Note, however, that every kernel function has its own optimal bandwidth. These bandwidths may not necessarily be the same. The histograms only serve to give a rough picture on the data used in the density plot. The dataset of Faithful (waitings and eruption) in R was used for fig. 5.1.

RESULTS, DISCUSSION AND CONCLUSION

The study revealed that quite a number of kernel functions have the suitable features such as that of assignment of weights so that the farthest point is given the least while that closest to the central point of the window of the kernel function receives the most. These kernel functions included the Gaussian, Epanechnikov, bi-weight, and the tri-weight among others. There is the so-called "naive" kernel function also variously referred to as the uniform or rectangular kernel function which assigns weights of ½ uniformly within the window. The constant assignment of the weights may not be a pleasant thing as it is believed that the points closer to a given observation have more information than those that are farther away. Also, its density function is not smooth because of the jumps over the short intervals of the window while being constant elsewhere see fig. 4.1 part (c). These jumps often end up affecting the smoothness of the estimate. Also noted in previous study by Wand and Jones (1995) is the triangular function which Avery (2010) reports to be lacking the smoothness property. On such grounds one may not opt for such a kernel function.

As for the others stated they have almost similar characteristics but Epanechnikov is the optimal kernel. This function posses' smooth properties but has discontinuous first derivative thus making the Gaussian function to be the best substitute.

The determination of the bandwidth which is known to play the key role in the smoothness of the curve is usually a point of concern in any study. Though there is no known universal way of obtaining a bandwidth, some researchers often narrowed down on the Gaussian kernel function that has optimal criteria of choosing the bandwidth for data that is or near normally distributed. This often allows a fine and precise trade-off between fitting the data and smoothing.

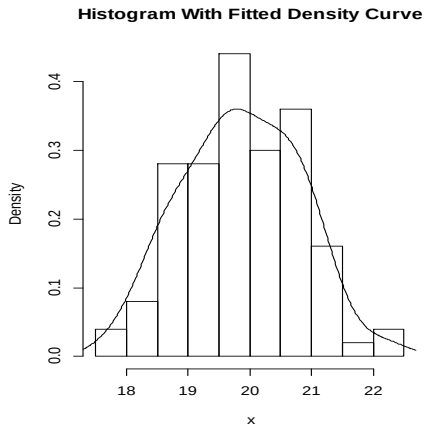
Generally, with properly chosen bandwidth most of the functions give satisfactory estimates. Thus, it can be concluded that of the two- the function and the bandwidth, it is the later that has more impact. The function has negligible effect.

REFERENCES

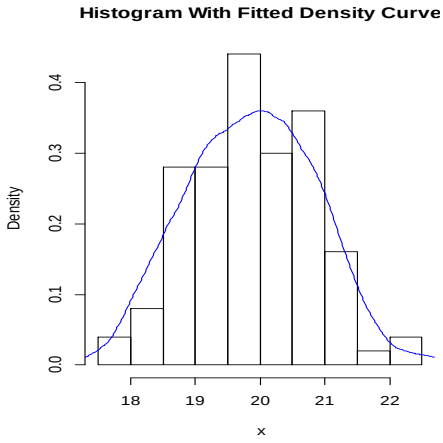
[1]. Alberts T and Karunamuni R. J. (2007). Boundary correction methods in Kernel density estimation. Presentation slides downloaded from internet.
[2]. Avery M. (2010). Literature Review for Local Polynomial Regression. Unpublished manuscript.
[3]. DiNardo J and Tobias J. L.(2001) Nonparametric Density and Regression Estimation. *Journal of Economic Perspectives* Vol **15** No. 4 pp 11-28
[4]. Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Annals of Statistics* **20**, 2008-2036.
[5]. Faraway J. (2006). Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. Chapman & Hall/CRC Taylor & Francis Group, LLC. New York.
[6]. Härdle, W. (1994), *Applied Nonparametric Regression Analysis,* Cambridge: Cambridge University Press
[7]. Irizarry R. A. and Bravo H.C. (2010). Smoothing. Lecture notes
[8]. Karunamuni and Alberts (2004). On the boundary correction in Kernel density estimation. A paper presented in the Fifth Biennial IISA International Conference on Statistics, Probability and Related Areas held at the University of Georgia, Athens, Georgia, from May 14-16, 2004.
[9]. Todd P. E. (2013). Nonparametric Density and Regression Estimation. Lecture Notes.
[10]. Wand and Jones, (1995) *Kernel Smoothing,* Chapman and Hall. New York.
[11]. Zucchini W., (2003). Applied Smoothing Techniques Part 1: Kernel Density Estimation

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 3, 2020 (pp. 51-59)
www.abjournals.org

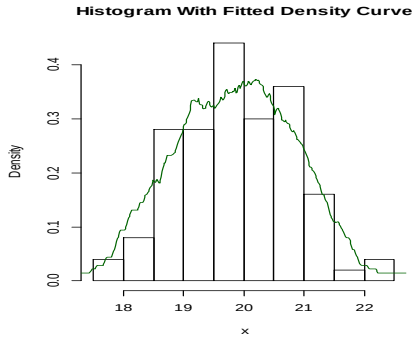
APPENDIX



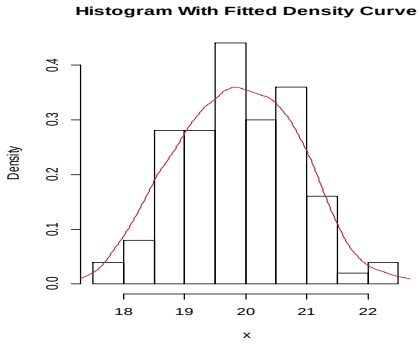
(a) Gaussian



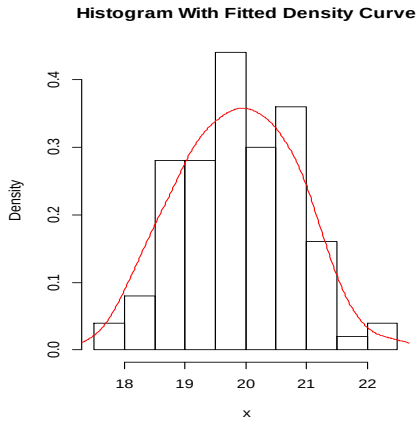
(b) Epanechnikov



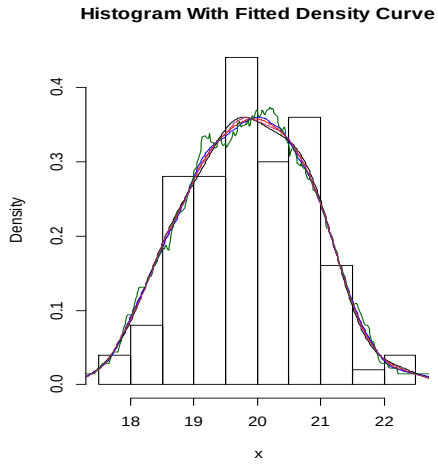
(c) Rectangular



(d) Triangular



(e) Biweight



(f) Combined Graphs

Fig. 4.1 Density Estimates Obtained using Various Kernel Functions with bw=0.39

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 3, 2020 (pp. 51-59)
www.abjournals.org

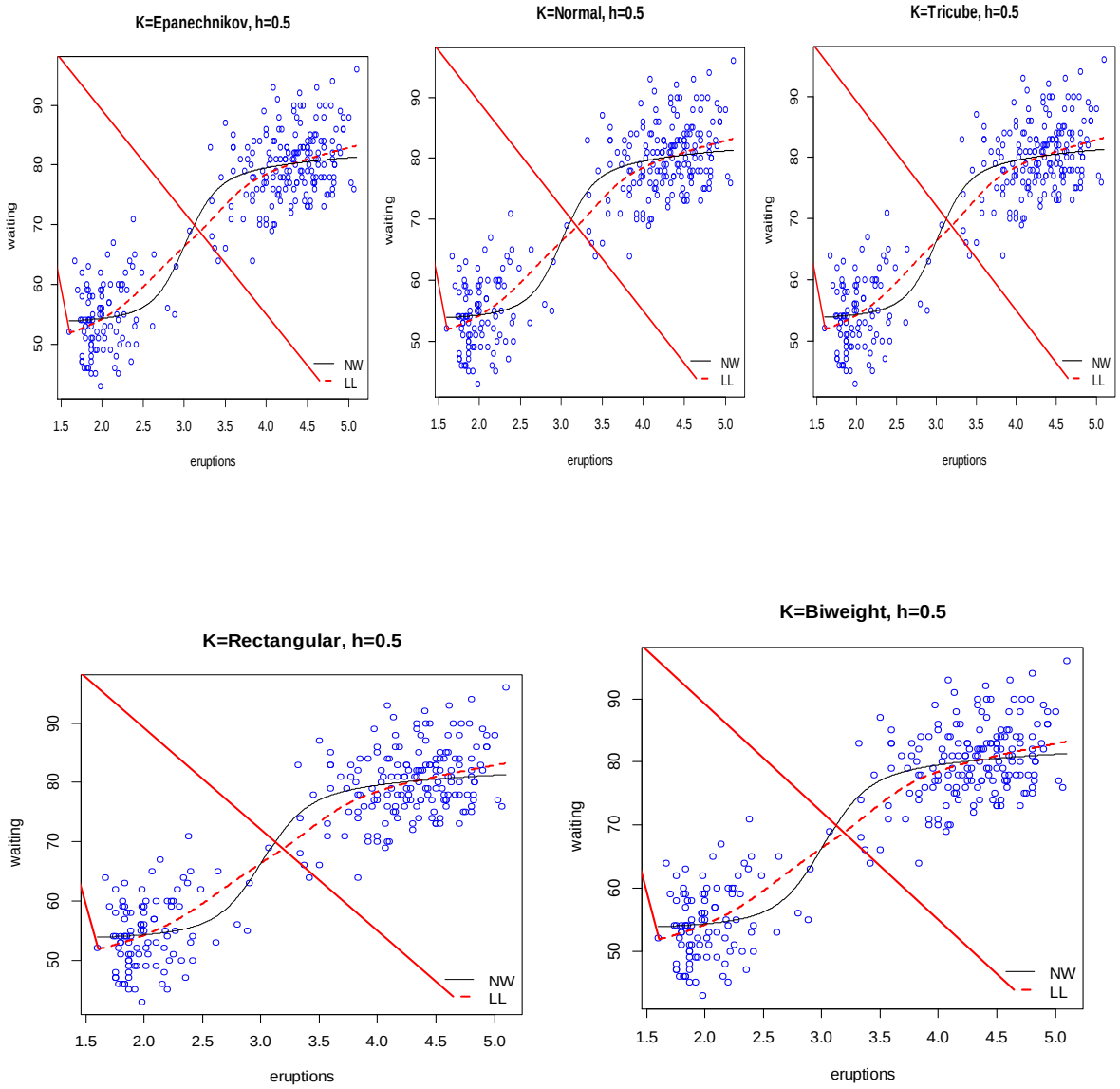


Fig. 5.1 Regression estimates with different kernel functions