Open Access
Baghdad Science Journal **Vol.16(3) Supplement 2019**

DOI: http://dx.doi.org/10.21123/bsj.2019.16.3(Suppl.).0793

Hazard Rate Estimation Using Varying Kernel Function for Censored Data Type I

Entsar Arebe Al.Doori * *Eqbal Mhomod*

Received 1/10/2018, Accepted 12/3/2019, Published 23/9/2019

Abstract:

In this research, several estimators concerning the estimation are introduced. These estimators are closely related to the hazard function by using one of the nonparametric methods namely the kernel function for censored data type with varying bandwidth and kernel boundary. Two types of bandwidth are used: local bandwidth and global bandwidth. Moreover, four types of boundary kernel are used namely: Rectangle, Epanechnikov, Biquadratic and Triquadratic and the proposed function was employed with all kernel functions. Two different simulation techniques are also used for two experiments to compare these estimators. In most of the cases, the results have proved that the local bandwidth is the best for all the types of the kernel boundary functions and suggested that the 2xRectangle and 2xEpanechnikov methods reflect the best results if compared to the other estimators.

Key words: Bandwidth, Censored Data, Hazard Rat, Kernel Function, Smoothing hazard rate

Introduction:

Man's need to continue his life in the best way is the first motive for the first studies and researches which are related to Survival Time. This takes into consideration the period of his survival when he suffers from certain disease (such as cancer). The nonparametric estimations concerning the hazard rate estimation of life time is a joint means of the statistics to prepare the censored survival data. The scientist **Parzen**(1962)(1) is the first one who has been highly concerned with the estimation by using varying kernel. It has been the weighting function and the kernel estimators have many uses such as (survival studies, **epidemiology, criminology** and **demography**). The kernel estimators are important as far as there are some problems when the stage of the end of the data is reached. This is referred to as the boundary effects. In fact, the boundary effects have been studied by some researchers such as **Breslow and Day(1987**)(2) . The estimates of "the boundary areas" curve did not show an area within the bandwidth of the endpoint. The application of unmodified kernel estimators causes meaningless estimation in the boundary areas near the endpoint. Recently, the researcher **Salha** in (2012)(3) estimated the hazard rate by using the Inverse Gaussian (IG) kernel and studied the nonparametric estimation of hazard rate by using kernel function. And Hind J. Kadhum& Iden H.Alkanani(2014)(4) using Survival estimation for singly type one centered sample based on generalized Rayleigh distribution. The aim of this paper is to compare several estimators of hazard rate estimation and show in the bandwidth. Two bandwidths are used namely global bandwidth and local bandwidth. Each one of them used four types of boundary kernel function: Rectangle, Epanechnikov, Biquadratic and Triquadratic ith a suggested method.

College of Administration and Economic, University of Baghdad, Baghdad, Iraq.
*Corresponding author: entsar_arebe_aldoori@yahoo.com

Materials and Methods (5,6):

Suppose that (t) represents life time variable with the rate of distribution and the hazard function is$\lambda(t)$ and this can be defined as follows (5):

$$\lambda(t) = \frac{f(t)}{1-F(t)} \quad , for\ F(t) < 1 \qquad \ldots(1)$$

Both of them are completed within the positive period of time. And the rate hazard means the specifications of the distribution:

$$\Pr(T > t) = S(t) = exp\left[-\int_0^t \lambda(s)\, ds\right]$$

The integration in the exponent is negative and it is a cumulative kernel function $\Lambda(t)$.

Smoothing hazard rate for continuously observation data (7)

The estimation of hazard rate for continuously observation data is near the concept of density estimation. In order to know this, we replace the equation No (1) by the derivative of the cumulative hazard function:

$$\Lambda(t) = \int_0^t \lambda(x)dx \qquad ...(2)$$

The estimation of hazard rate can be obtained from the increase of the estimation $\boldsymbol{\Lambda(t)}$ .From their part, the two scientists **Watson and Ledbetter** (1964)(5) are the first who suggested and studied the smoothing hazard rate by using the experimental cumulative hazard $\boldsymbol{\Lambda_n(t)}$ according to independent distributions i.i.d and the sample of the periods of hazard$(\delta_{[j]} = 1)$ and they suggested the hazard estimation. The type of the convolution type hazard estimator ([1])

$$\hat{\lambda}_n(t) = \int_0^t \delta_n(t-x)d\Lambda_n(t) \qquad ...(3)$$

Therefore$\{\delta_n\}$ is the sequence of smoothing functions it is near Dirc delta – function when $n \to \infty$ The delta- function sequence is characterized by generality and it contains several types of smoothing and weighting function is one of them which has been used by Parzen (1962)(1) and as follows:

$$\delta_n(x) = \frac{1}{b_n} k\left(\frac{x}{b_n}\right)$$

Where b_n is the bandwidth

The two scientists Watson and Leadbetter (1964)(5) gave another estimation rate:$\hat{\lambda}_n(t) = \frac{\hat{f}_{n\,(t)}}{1-\hat{F}_{n\,(t)}}$

$\hat{f}_n$ is the density estimation of hazard density f and is an $\hat{F}_{n\,)}$ experimental estimation of the kernel hazard time distribution F.Both estimators have the same contrast but with different bias. The amount of convolution $\hat{\lambda}_n(t)$ is predominant because the theoretical measures (mean square error available) outperformed the estimator of the ratio type$\tilde{\lambda}_n(t)$.Under the random control model the current Ti time of the individual can be monitored by another random variable Ci and we will assume that:

1- $T_!, T_2, ... T_n$is lifetime (time to failure) from the observations of size (n) which are random distribution i.i.d identical and independent and it has the same distribution and it is positive with CDF continuous and cumulative hazard and continuous density hazard *f*.

2- $C_1, C_2, ...C_n$ refers to Censoring time. It is a random distribution i.i.d. identical and independent and it has the same distribution. It is also positive with *CDF joint* cumulative hazard and continuous G density hazard.

3- The life time Xi and the censoring times Ci are independent and: $X_i = \min(T_i, C_i)$

The $\delta_i = \mathrm{I}_{\{x_i = T_i\}}, i = 1, ...n$ progressively arranged x $(x_{(i)}, \delta_{(i)})$ the sample is arranged according to $\delta_{(i)}$ $X_{(1)} \le X_{(2)} \le \cdots . \le X_{(n)}$ X_i s the indicator of censoring to X_i and hazard estimators can be obtained by smoothing estimators $\boldsymbol{\Lambda_n(t)}$ Nelson-Aalen for the cumulative hazard function and we suppose that:

$$N_n(t) = \sum_{i=1}^{n} \mathrm{I}_{\{X \le t, \delta_i = 1\}}$$

$$Y_n(t) = \sum_{i=1}^{n} \mathrm{I}_{\{X \ge t\}}$$

The estimator $\boldsymbol{\Lambda}(.)$ Nelson-Aalen in the analysis of observation data is defined by the following form:

$$\boldsymbol{\Lambda_n(t)} = \int_0^t \frac{\mathrm{I}_{\{y_n\,(s)>0\}}}{(Y)_n(s)} \, dN_n(s)$$

$$= \sum_{i=1}^{n} \frac{\delta_{[i]} I_{\{x_{(i)} \le t\}}}{n-i+1} \; ...(4)$$

Provided that there is no link between the observations

Kernel Hazard Estimators (8):

To obtain the Kernel Hazard Estimators we choose the value:

$$\delta_n(x) = \frac{1}{b_n} k\left(\frac{x - X_{(i)}}{b_n}\right)$$

Where $b_n = b \; bandwidth$ . . . (5)

The equations No. (4) And (5) are replaced by the equation no. (3) By choosing kernel k and bandwidth $b = b_n$ therefore we obtain the kernel hazard estimator as follows:

$$\tilde{\lambda}(x) = \int \frac{1}{b} k\left(\frac{x - X}{b_n}\right) \Lambda_n(x)$$

$$= \frac{1}{b} \sum_{i=1}^{n} k\left(\frac{x - X_{(i)}}{b_n}\right) \frac{\delta_{(i)}}{n-i+1} \; ...(6)$$

K is fixed kernel function. The bandwidth is of extreme importance and it organizes the differentiation between bias and contrast of the estimator in the equation (6). The small bandwidth

([1]) The convolution of the two kernels g , h can be defined in the following form:

$$(h * g)(t) = \int_{-\infty}^{\infty} g\,(t-\tau)dt$$

causes little smoothing curve with small bias and great contrast if it is compared to big bandwidth. The characteristics of the hazard rate estimator function no.(5) has made use of by many researchers and we mention some of them Ramlau-Hansen (1983)(7), Taner and Wang (1983)(8) and we are able to get the characteristics of the adjacent consistency under certain assumptions are. Suppose that k is a round figure $k \geq 0$ and that λ is derivable and continuous for k of the times in the (O. R) period where R is the right endpoint so that L(R) < 1. The rate of the rounding of equation (6) is dependent on the degree of the core and the beam width. Application of the k-degree in the core selects an even number k = 2 the bias and the variation are respectively.

$$bias\left(\hat{\lambda}_{(t)}\right) = b^k\left[\lambda^{(k)}(t)B_k + o(1)\right] \text{ ... (7)}$$

$$var\left(\hat{\lambda}(t)\right) = \frac{1}{nb}\left\{\frac{\lambda(t)}{[1-F(t)][1-G(t)]}v + o(1)\right\} \text{ ... (8)}$$

Where as

$$B_k = \frac{(-1)^k}{k!\int k^2K(x)dx} \text{ ... (9)}$$

$$v = \int k^2(x)dx < \infty \quad \text{ ... (10)}$$

By choosing Epanechnikov kernel as it is shown in the following equation:

$$k(x) = 0.075(1-x^2), -1 < x < 1 \text{ ... (11)}$$

The effect of bandwidth b and the differentiation between bias and variance is evident in the equations (8) (7) and for the parallel distribution, we assume that: $d = \lim_{n\to\infty} n\, b^{2k+1}$ Exist for some $0 \leq d \leq \infty$

$$(nb)^{\frac{1}{2}}\left(\hat{\lambda}(t) - \lambda(t)\right) \xrightarrow{D} N\left\{d^{\frac{1}{2}}\lambda^{(k)}(x)B_{k,} \frac{\lambda_{(t)}}{\left[[1-F(t)][1-G(t)]\right]^v}\right\} \text{ ... (12)}$$

where D converge in distribution

Boundary Effects (9):

The unmodified kernel estimation is unreliable in the border area, whereas the bandwidth area is within larger or smaller observations to address the boundary effects of different data we refer to by boundary kernel which can be used with he boundary area that is to say hazard function estimator with the different kernel functions and different bandwidth. The increase of bandwidth is different according to what is stated in order to realize the balance between the bias and the contrast thus reducing the error. e. The hazard estimator rate is defined with different degrees as it is shown in the following formula (3):

$$\hat{\lambda}(x) = \hat{\lambda}\left(x, b(x)\right) = \frac{1}{b(x)}\sum_{i=1}^{n} k_x\left(\frac{x - X_{(i)}}{b(x)}\right)\frac{\delta_{(i)}}{n-i+1} \text{ ... (13)}$$

The $\boldsymbol{b} = \boldsymbol{b}(\boldsymbol{x})$ represents the bandwidth and $k = k_x$ which is kernel and it depends upon the x point where the estimation has been calculated. We will discuss the choice of kernel k_x and bandwidth b(x). The moment conditions in kernel boundary means that it carried the negative values and this will lead to negative hazard rate estimates as it is shown in the equation (No. 13) near endpoints. This might happen in the interior. In this case, we have to assume that: $\hat{\lambda}(x) = \max(\hat{\lambda}(x), 0)$

Bandwidth b(x) kernel k_x chorice (9)(10):

The bandwidth in the kernel hazard function estimation can be fixed at all points (the global bandwidth b) or can be different in various points (local bandwidth). Normally, the global bandwidth is used to measure the density or to estimate the slope which is prevalent for its simplicity. The K_x kernels are:

$$k_x(z) = \begin{cases} k_+(1, z) & x \epsilon I \\ k_+\left(\frac{x}{b(x)}, z\right) & x \in B_l \\ k_-\left(\frac{R-x}{b(x)}, z\right) & x \in B_R \end{cases} \text{ ... (14)}$$

$I = \{x: b(x) \leq x \leq R - b(x)\}$= Is **interior** (inside) and the effect of boundary is not calculated and $B_l = \{x: 0 \leq x \leq b(x)\}$ left boundary region $B_R = \{x: R - b(x) < x \leq R\}$ right boundary region $K_+, K_- : [0,1] \times [-1,1] \to \mathbb{R}$ we notice that $K_+(q, .)$ provide us in the time **[-1, q]** , $0 \leq q \leq 1$ and $K_-(q, .) = K_+(q, .)$ $0 \leq q < 1$. The kernels $K_+(q, .)$ are called **boundary** kernels.

Table 1. The best forms of boundary kernels

	boundary kernel on $[-1,q]$	
0	$\frac{2}{(1+q)^3}[3((1-q)x + 2((1-q+q^2]$	Rectangle
1	$\frac{12}{(1+q)^4}(x+1)[x(1-q)(3q^2-2q+1/2]$	Epanechnikov
2	$\frac{15}{(1+q)^5}(x+1)^2(q-x\left[(2x)\left(5\frac{1-q}{1+q}-1\right)+(3q-1)+\frac{5(1-q)^2}{1+q}\right]$	Biquadratic
3	$\frac{70}{(1+q)^7}(x+1)^3(q-x\left[(2x)\left(7\frac{1-q}{1+q}-1\right)+(3q-1)+\frac{7(1-q)^2}{1+q}\right]$	Triquadratic

There are two familiar methods of showing the bandwidth which will be tackled later.

Local bandwidth (10) (11):

The display of the optimal local bandwidth at each point in the grid is obtained from the minimization of local MSE.

$MSE_{est}\big(x, b(x)\big) = \hat{V}\big(x, b(x)\big) + \hat{B}^2(x, b(x))$ . . . (15)

And the $\hat{v}$, $\hat{\beta}$ they are the variance bias respectively as follows:

$\hat{V}\big(x, b(x)\big) = \frac{1}{nb(x)} \int K_x^2\,(y)\, \frac{\hat{\lambda}(x-b(x)y)}{l_n(x-b(x)y)} dy$ ...(16)

$\hat{B}^2\big(x, b(x)\big) = \int \hat{\lambda}(x - b(x)y) K_x(y) dy - \hat{\lambda}(x)$ . . . (17)

$L_n(x) = 1 - \frac{1}{n+1} \sum_{i=1}^{n} I_{\{X \le x, \delta_i = 1\}}$ …. (18)

The equation no. (18) is the empirical survival function of the uncensored observations and the equation Ln(x)=0 ,$\hat{\lambda}(.)$ is pilot estimate to λ which has been formed by boundary correction and the fixed initial bandwidth bo is limited by the researcher. These estimations depend on the assumption of finite sample of bias and contrast which is from the convolution type derived by Muller and Wang (1990) (9). The minimizing bandwidth is:

$\hat{b}(x) = \arg minMSE_{est}(x, b)$ … (19)

Which has been referred as $\hat{\lambda}$(x,$\hat{b}$ (x)) hazard rate estimator in the equation (13) with the bandwidth $\widehat{b}(x)$.

Global bandwidth (9):

The global bandwidth is reflected itself in all the points of the grid. The optimal global bandwidth can be got from minimizing the integrated mean square error (IMSE)

$$IMSE_{est}(b) = \int_0^R \big\{\hat{V}(x, b) + \hat{B}^2(x, b)\big\} \mathrm{dx} \quad ...(20)$$

$\boldsymbol{\hat{b}} = arg_b IMSE_{est}(b)$ … (21)

B does not depend on x and the global bandwidth estimation

The following is the summary of hazard function estimation in algorithm:

Function estimation algorithm by using the local bandwidth (9):

__The first step__

We choose the kernel k+(q ,.) From the timetable No. (1) About the methods of determining $\mu\epsilon\{0,1,2,3\}$ and we recommend that we choose the initial bandwidth b_o depending on the situation and it can be got in the following equation:

$$bndwidth\, pilot = \frac{R}{8nz^{0.2}}$$

$$= (max.time - min.time)/(8nz^{0.2})$$

nz is the number of uncensored observations. We suppose that the data is available during the period. [0, R] then we get the pilot estimators $\widehat{\lambda(.)}$ according to the equation (13) and use $b(x) = b_o$ it does not depend on x and we get Kx according to the equation No. (14).

Instead of that we choose a parametric model and we fit it to the data by maximum likelihood method and the $\hat{\lambda}_{par}$ refers to the fitting model.

__The second step__ We choose an equidistant grid and minimizing grid m_1 from the points $\tilde{x}_i, i = 1, \dots, m_i$ between o and R the number m_1 is the grid point determine the calculated time to the largest expansion and it must not be very large. Then we choose the grid L bandwidth $\tilde{b}_j, j = 1, \dots, L_i$ equidistant between b_1 , b_2 and we recommend choosing:$b_1 = 2b_0/3$,$b_2 = 4b_0$: for the points of the grid $\tilde{x}_i$ and all the bandwidth $\tilde{b}_j$ we calculate the variance and the bias $\hat{v}$ and $\hat{\beta}$ according to the equations (16) and (17) and we get the minimizers $\tilde{b}(\tilde{x}_i)$ for $MSE_{est}(\tilde{x}_i)$ according to the equations No.(19) and the minimizing$(\tilde{x}_i)$ for$\tilde{b}_1 \dots . . \tilde{b}_L$

__The third step__ we choose the equidistant grid and estimation grid m_2 from the points $x_r, r = 1, \dots . , m_2$ between 0 and R which we desire the final hazard rate. We get the similar bandwidth $\hat{b}(x)$ by smoothing the bandwidth $\hat{b}(\tilde{x}_i)$ with boundary - modified smoother by using the bandwidth$\tilde{b}_0$ and $\tilde{b}_0 = b_0\; (or\; \tilde{b}_0 = \frac{3}{2} b_0)$

$$\hat{b}(x_r) = \frac{\sum_{i=1}^{n} k_{x,} \left(\frac{x_r - \tilde{x}i}{\tilde{b}_o}\right) \tilde{b}(\tilde{x}_i.)}{\sum_{i=1}^{n} k_x \left(\frac{x_r - \tilde{x}i}{\tilde{b}_o}\right)} \dots (22)$$

__The fourth step__ To get the final hazard function estimation$\widehat{h}\,(x_r)$ from equation No.(13) by using: $b(x_r) = \hat{b}(x_r),\ 1 \leqslant r \leqslant m_2$

Hazard function estimation algorithm by using global bandwidth (10):

__The first step__: Hazard function estimation algorithm by using local bandwidth

The second step we get the integrated mean square error (IMSE) according to the following

$$IMSE^{*}(\tilde{b}_j) = \sum_{i=1}^{m_i} MSE_{est}(\tilde{x}_i \tilde{b}_j) \ldots (23)$$

equation:

$$\hat{b} = min_{\tilde{b}_j} IMSE * (\tilde{b}_j) \ldots\ldots (24)$$

The third step: We choose the equidistant estimation grid m_2 from the points $x_r, r = 1, \ldots., m_2$ as it is the case in the third step from the hazard function estimation algorithm by using local.

The fourth step: We calculate $\hat{h}(x_r)$ according to the equation No.(13) by using $b(x_r) = \hat{b}$

The proposed function the researcher suggested the following density function

$f(x) = 2x \ , 0 \leq x \leq 1$ …. (25)

The empirical part(12):

The implementation of all the empirical simulation by using the programing language R (version 2.15.1 R 2012). This language is a large number of packages used in many different statistical fields. To carry out the experimentation of simulation different levels of factors are used and as follows:1-Size of different big medium and small samples :n = 30.60.80.100.2-The proposed density function in equation No. (25).3-The distributions used in this research Exponential distribution, Gamma distribution, Normal distribution, Log-normal distribution, Bimodal distribution. In order to determine the best method of distribution two criteria are used Average Mean Square Error (AMSE), Average Square Error (ASE), Average Maximum Deviation (AMXDV)

$$AMSE = \frac{1}{1000} \sum_{I=1}^{1000} ASE_i$$

$$ASE(\lambda) = m_2^{-1} \sum_{r=1}^{m_2} (\lambda(x_r) - \hat{\lambda}(x_r))^2$$

$$AMXDV = \frac{1}{1000} \sum_{I=1}^{1000} MXDV_i$$

Plan of the performance of simulation experimentation

The programming language is a large increasing number of packages such as ‘dist’ package by which many distributions can be generated. From these packages we mention (muhaz package) by which it is possible to get the hazard function estimation of the first type and the kernel with observation data m1=101 and m2=50 this experiment has been repeated (1000) times.

Perform the first simulation experiment

1- A. The sample x_i ,$1 \leq i \leq n$ has been generated with gamma distribution G (5.1) with the parameter Fig. No 5 and the parameter 1.B. The variable c_i ,$1 \leq i \leq n$ has been generated from the exponential distribution Exp(1/2.5) with the arithmetic mean 2.5.C. in A and B we find that the real hazard function takes the symbol hreal according to the equation No. (1).

2- Perform the global bandwidth algorithm by using muhaz function and we get the hazard function estimator and give it the symbol hest as it is in the equation No.(13) for four types of boundary kernel function which are:Rectangle, Epanechnikov, Biquadratic and Triquadratic which are explained in the table No. (1).We get AMSE for the four boundary kernels to find the difference with the square hest and hreal.

3-Perform the local bandwidth algorithm by using muhaz function and we get the hazard function estimator and give it the symbol hest as it is in the equation No.(13) for four types of boundary kernel function which are : Rectangle, Epanechnikov, Biquadratic and Triquadratic which are explained in the table No. (1).We get AMSE for the four boundary kernels to find the difference with the square hest and hreal.

Performing the proposed function experiment

1- A. The samplex_i ,$1 \leq i \leq n$ has been generated with the proposed density function x=2G which is the density function according to gamma distribution.

2- B. The variable C has been generated from the exponential distribution Exp(1/2.5) with the arithmetic mean 2.5

C. In A and B we find that the real hazard function takes the symbol hreal according to the equation No. (1).

3- Performing the global bandwidth algorithm by using muhaz function and we get the hazard function estimator and give it the symbol hest as it is in the equation No.(13) for four types of boundary kernel function which are : Rectangle, Epanechnikov, Biquadratic and Triquadratic which are explained in the table No. (1).We get AMSE for the four boundary kernels to find the difference with the square hest and hreal.

4- Performing the local bandwidth algorithm by using muhaz function and we get the hazard function estimator and give it the symbol hest as it is in the equation No.(13) for four types of boundary kernel function which are : Rectangle, Epanechnikov, Biquadratic and Triquadratic which are explained in the table No. (1).We get AMSE for the four boundary kernels to find the difference with the square hest and hreal.

View and discussion of the results the first simulation experiment

1- To show the local and global bandwidth and the size of the sample n=100,80 the lowest value of the standard AMSE in the proposed method 2xBiquadratic and the lowest value of the standard AMXDV in the proposed method are 2xEpanechnikov, 2xBiquadratic respectively, as it shown in Tables (1)and (2) and Fig. (1).

2- To show the local and global bandwidth and the size of the sample n=60 the lowest value of the standard AMSE in the proposed method 2xBiquadratic and the standard AMXDV in the proposed method is 2xRectangle, 2xBiquadratic respectively, as shown in Tables (1, 2) and Fig. (1).

3-To show the local and global bandwidth and the size of the sample n=30 the lowest value of the standard AMSE in the proposed method 2xEpanechnikov for two bandwidth and the lowest value of the standard AMXDV in the proposed method is 2x Epanechnikov,x Biquadratic respectively as shown in Table (1, 2)and Fig. (1).

4-The lowest value of the standard AMSE for the local and global bandwidth is in the proposed method 2xEpanechnikov and the size of the sample is n=30 as it is shown in Table (2).

5-The lowest value of the standard AMXDV for the local and global bandwidth is in the proposed method 2xEpanechnikov ,xBiquadratic respectively and the size of the sample is n=30 as it is shown in Table (2).

Table 2. Results of the first simulation experiment

	Band width	**Local**		**global**	
Carnal function		**AMXDV**	**AMSE**	**AMXDV**	**AMSE**
n=100	**xEpanechnikov**	0.199516	0.007004	0.947509	0.037113
	x Biquadratic	0.214452	0.006322	0.530998	0.022392
	x Rectangle	0.191194	0.006242	0.678972	0.025083
	x Triquadratic	0.244015	0.00721	0.610106	0.027008
	2xEpanechnikov	0.104423	0.001947	0.469996	0.009662
	2x Biquadratic	0.112896	0.001741	0.266686	0.00564
	2x Rectangle	0.109332	0.001827	0.353497	0.006659
	2x Triquadratic	0.126241	0.002033	0.308309	0.006827
	xEpanechnikov	0.196201	0.007455	0.916713	0.036786
	x Biquadratic	0.216402	0.006853	0.534675	0.023486
	x Rectangle	0.198119	0.006927	0.647805	0.024855
n=80	**x Triquadratic**	0.242221	0.007549	0.621194	0.028588
	2xEpanechnikov	0.110097	0.002215	0.457324	0.009965
	2x Biquadratic	0.119016	0.001825	0.266396	0.005795
	2x Rectangle	0.107435	0.001859	0.334903	0.006777
	2x Triquadratic	0.128822	0.001967	0.311707	0.007099
	xEpanechnikov	0.194696	0.007405	0.374316	0.044396
	x Biquadratic	0.206463	0.006461	0.520459	0.032896
	x Rectangle	0.196543	0.006603	0.78115	0.021617
n=60	**x Triquadratic**	0.239567	0.007458	0.622154	0.037184
	2xEpanechnikov	0.101373	0.002042	0.806763	0.014478
	2x Biquadratic	0.112665	0.001674	0.160962	0.003041
	2x Rectangle	0.100083	0.001745	0.63192	0.01023
	2x Triquadratic	0.137244	0.002372	0.229742	0.004687
	xEpanechnikov	0.139827	0.001817	0.93525	0.023141
n=30	**x Biquadratic**	0.061881	0.00071	0.068073	0.001379
	x Rectangle	0.093536	0.000499	0.473278	0.01021
	x Triquadratic	0.075463	0.000574	0.093768	0.001903
	2xEpanechnikov	0.030457	0.000155	0.076754	0.0005
	2x Biquadratic	0.076754	0.000343	0.076754	0.000576
	2x Rectangle	0.076754	0.000303	0.076754	0.000606
	2x Triquadratic	0.076754	0.000441	0.091643	0.001112

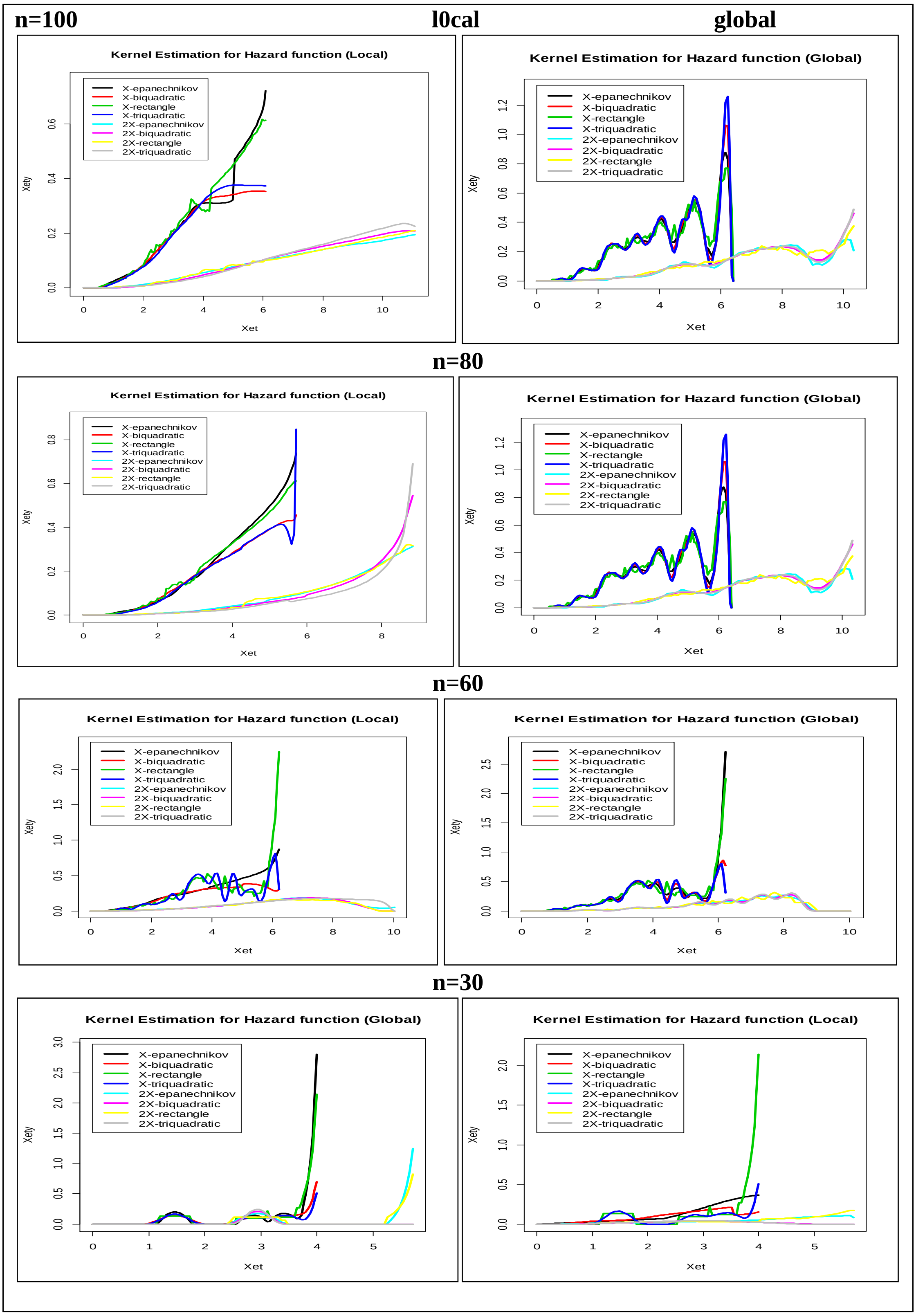


Figure 1. The estimation of the failure function represents the results of the first simulation experiment for two types of beam widths and different sample

Performing the second simulation experiment

1- A. The variable x_i has been generated from the density function of bimodal distribution which is used in (1992) by Kooperberg and Stone:

g is density function of normal algorithm $f = 0.8g + 0.2h$ distribution Lognormal, (O. ½) and h is the density function of normal distribution Normal (2,0.17) with arithmetic mean 2 with standard deviation 0.17.B. The variable C_i has been generated from the exponential distribution

Exp(1/2.5) with the arithmetic mean 2.5.C. In A and B we find that the real hazard function takes the symbol hreal according to the equation No. (1).

2- Performing the global bandwidth algorithm by using muhaz function and we get the hazard function estimator and give it the symbol hest as it is in the equation No.(13) for four types of boundary kernel function which are : Rectangle, Epanechnikov, Biquadratic and Triquadratic which are explained in the table No. (1).We get AMSE for the four boundary kernels to find the difference with the square hest and hreal.

3- Performing the local bandwidth algorithm by using muhaz function and we get the hazard function estimator and give it the symbol hest as it is in the equation No.(13) for four types of boundary kernel function which are : Rectangle, Epanechnikov, Biquadratic and Triquadratic which are explained in the table No. (1).We get AMSE for the four boundary kernels to find the difference with the square hest and hreal.

Performing the proposed function experiment

1- A. Generating the variable x_i from the proposed density function $x = 2f$ and f is the bimodal distribution density function $f = 0.8g + 0.2h$

g is density function of normal algorithm distribution Lognormal ,

Lnorm (0.1/2) and h density function of normal distribution (2.0.17) with arithmetic mean 2 and standard deviation 0.17. B. The variable C_i has been generated from the exponential distribution Exp(1/2.5) with the arithmetic mean 2.5.

C. The paragraphs a and b we find the real hazard function with the symbol hreal according to the equation No. (1)

2- Performing the global bandwidth algorithm by using muhaz function and we get the hazard function estimator and give it the symbol hest as it is in the equation No.(13) for four types of boundary kernel function which are : Rectangle, Epanechnikov, Biquadratic and Triquadratic which are explained in the table No. (1).We get AMSE for the four boundary kernels to find the difference with the square hest and hreal.

3. Performing the local bandwidth algorithm by using muhaz function and we get the hazard function estimator and give it the symbol hest as it is in the equation No.(13) for four types of boundary kernel function which are : Rectangle, Epanechnikov, Biquadratic and Triquadratic which are explained in the table No. (1).We get AMSE for the four boundary kernels to find the difference with the square hest and hreal.

View and discussion of the results of the second simulation experiment

1- To show the local and global bandwidth and the size of the sample n=100 the lowest value of the standard AMSE in the proposed method 2xRectangle and of the standard AMXDV in the proposed Rectangle, 2xBiquadratic respectively as shown in Table (3) and Fig. (2).

2- To show the local and global bandwidth and the size of the sample n=80 the lowest value of the standard AMSE is in the proposed method **2xBiquadratic** for two bandwidth and the lowest value of the standard AMXDV is in the proposed method 2xBiquadratic , xTriquadratic respectively as it is shown in Table (3) and Fig. (2).

3- To show the local and global bandwidth and the size of the sample n=60 the lowest value of the standard AMSE is in proposed xEpanechnikov, 2xBiquadratic respectivly and of the standard AMXDV is in the proposed method 2xEpanechnikov **Rectangle**g_ as it is shown in Table (3) and Fig. (2).

4- To show the local and global bandwidth and the size of the sample n=30 the lowest value of the standard AMSE is in the proposed method respectively and the lowest value of the standard AMXDV is in the proposed method x Rectangle 2x Triquadratic respectively as it is shown in Table (3) and Fig. (2).

5- The lowest value of the standard AMSE for the local and global bandwidth is in the proposed method 2x Rectangle for two bandwidth and the size of the sample is n=100 as it is shown in Table (3).

6- The lowest value of the standard AMXDV for the local bandwidth is in the proposed method 2x Rectangle and the size of the sample is n=100 .and the lowest value of the standard AMXDV for the global bandwidth is in the proposed method x Rectangle and the size of the sample is n=30 as it is shown in Table (3).

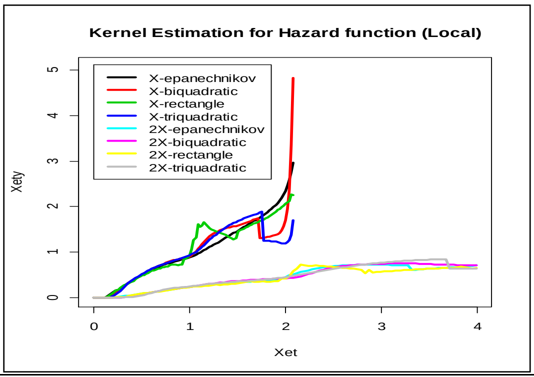
7- The value of the standard AMSE**,** AMXDV of the local bandwidth is less than the global bandwidth for the boundary of kernel function types and all the sizes of the samples.

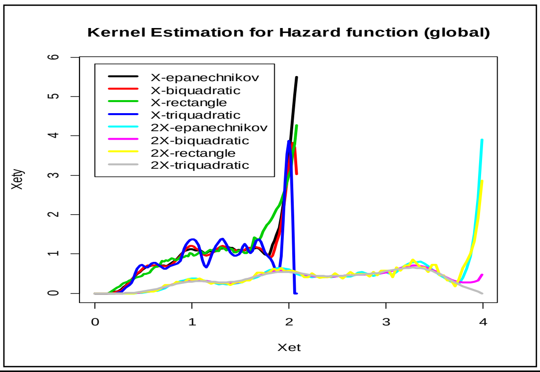
Table 3. Results of the second simulation

Carnal function	Band width	Local		global	
		AMXDV	**AMSE**	**AMXDV**	**AMSE**
n=100	**xEpanechnikov**	0.205421	0.006982	0.941414	0.041447
	x Biquadratic	0.230874	0.006193	0.569635	0.024742
	x Rectangle	0.198622	0.005784	0.696778	0.028549
	x Triquadratic	0.240959	0.007261	0.695874	0.031059
	2xEpanechnikov	0.092793	0.001592	0.4013	0.007687
	2x Biquadratic	0.107932	0.001716	0.267224	0.005843
	2x Rectangle	0.101838	0.001558	0.302092	0.00539
	2x Triquadratic	0.126945	0.002146	0.301196	0.00694
	xEpanechnikov	0.565407	0.075322	0.682262	0.184819
	x Biquadratic	0.664094	0.091413	0.37705	0.335125
	x Rectangle	0.615393	0.076498	0.344032	0.174117
n=80	**x Triquadratic**	0.646536	0.106088	0.129477	0.292992
	2xEpanechnikov	0.455608	0.028483	0.882476	0.131723
	2x Biquadratic	0.265638	0.017664	0.599564	0.052142
	2x Rectangle	0.518503	0.04034	0.387802	0.088067
	2x Triquadratic	0.509506	0.032869	0.651697	0.059228
	xEpanechnikov	0.238575	0.153299	0.598235	0.358896
	x Biquadratic	0.49294	0.178485	0.7918	0.240096
	x Rectangle	0.35474	0.165332	0.054603	0.260144
	x Triquadratic	0.634161	0.209833	0.98441	0.277859
	2xEpanechnikov	0.478388	0.028633	0.372719	0.092686
	2x Biquadratic	0.569475	0.033705	0.777055	0.057392
n=60	**2x Rectangle**	0.498595	0.029503	0.965941	0.062014
	2x Triquadratic	0.630132	0.038683	0.870681	0.066934
	xEpanechnikov	0.694178	0.14493	0.383564	0.984759
	x Biquadratic	0.602721	0.095829	0.871871	0.289783
	x Rectangle	0.664538	0.108827	0.005147	0.225256
	x Triquadratic	0.760746	0.122607	0.871871	0.301658
	2xEpanechnikov	0.302178	0.033147	0.383368	0.038804
	2x Biquadratic	0.295193	0.024691	0.460215	0.04358
n=30	**2x Rectangle**	0.31369	0.025521	0.308012	0.030213
	2x Triquadratic	0.289105	0.028277	0.354832	0.035017

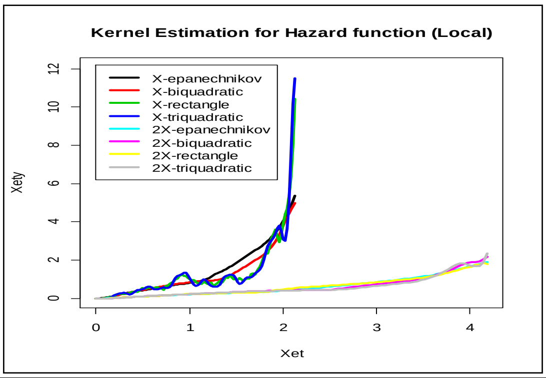
Global local

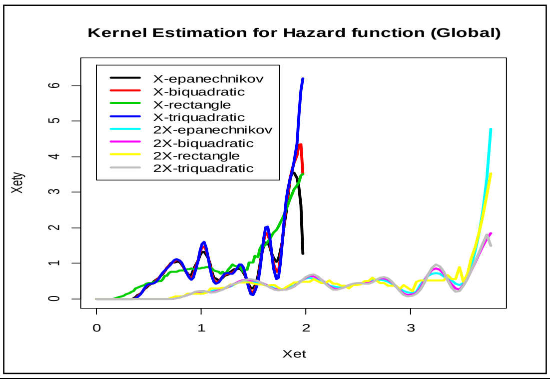
n=100





n=80





n=60

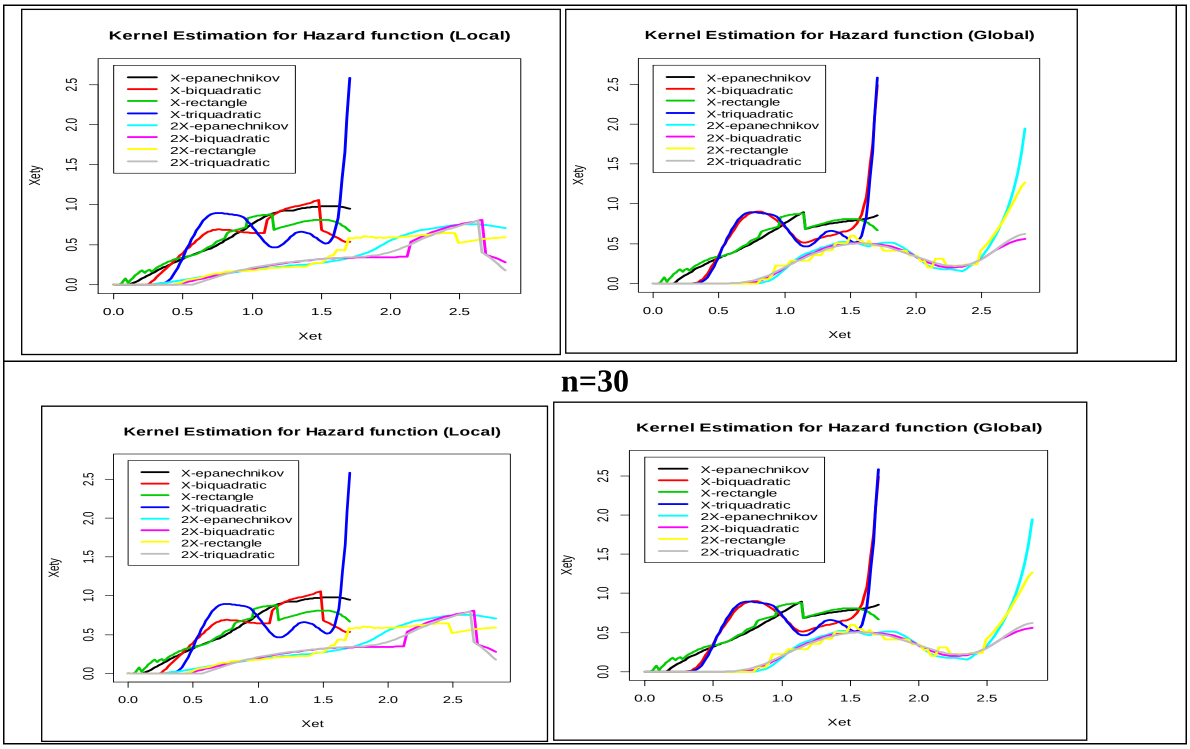


Figure 2. Estimation of the failure function represents the results of the second simulation experiment for two types of beam widths and different sample sizes

Conclusions:

1- The advantage of using the standard AMSE and for the two types of bandwidth for all the boundary of kernel function types and all the sizes of the samples except showing the global bandwidth in the x Biquadratic with the size of n=60.

2- The advantage of using the kernel function determines the proposed function 2x Epanechnikov with the local bandwidth and the size of the sample is n=30.

3- The advantage of using the kernel function determes the proposed function x Biquadratic with the local bandwidth and the size of the sample is n=30.

4- The advantage of using the kernel function determines the proposed function 2x Rectangle with the global bandwidth and the size of the sample is n=100.

5- The advantage of using the local bandwidth and the global bandwidth in reliability.

6- The increase of the value of AMSE with the increase of the sizes of the samples.

Conflicts of Interest: None.

References:

1. Parazen E . On Estimation of a Probability density Function and Mode, JMSA.1962;l(33):1065-1076.

2. Breslow N E, Day N E. The Design and Analysis of Studies. Statistical Methods in Cancer Research 1987;2;Oxford University Press.

3. Salha R.Hazard Rate function Estimation Using Inverse Gaussian Kernel .IUGNS. 2012;20;(1): 73-84.

4. Hind J K, Iden H Alkanani .Survival estimation for singly type one centered sample based on generalized Rayleigh distribution. Baghdad Sci.j.2014; 11 (2): 193-201. Publisher: Baghdad University.

5. Watson G S, Leadbetter, M R. Hazard analysis1. Biometrika.1964; 51:175-184.

6. Wang J L. Smoothing hazard rate .Encyclopedia of Biostatistics.2003; 2nd Edition; 7: 4986-4997.

7. Ramlau Hansen H. Smoothing Counting process Intensities by means of kernel function. Ann. statist.1983; 11(2): 453-466.

8. Tanner M A, Wong W H. The estimation of the hazard function from randomly censored data by the kernel method. Ann.Statit.1983; 1: 989-993.

9. Müller HG Wang J L .Hazard rate estimation under random censoring with varying kernels and bandwidths.Biometrics.1994; 50 (1): 61-76.

10. Müller HG, Wang JL. Locally adaptive hazard smoothing. Probab. Theory Related Fields.1990; 85; 523-538.

11. Philip Sieger, Eirini. Tatsi. nonparametric estimation of the hazard function .Goeth UniversityFrankfurt Fculty of Economicsand business administration. 2010, September 15.

12. Kooperberg C, Stone C J. Logspline density estimation for censored data. J. comput. Graph. Statist.1992; 1: 301-328.

تقدير دالة الفشل باستعمال دوال لبية مختلفة لبيانات مراقبة من النوع الاول

انتصار عريبي فدعم **اقبال محمود**

قسم الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد، بغداد، العراق

الخلاصة:

في هذا البحث تم تقديم عدد من المقدرات الخاصة بتقدير دالة الفشل باستعمال إحدى الطرق اللامعلمية وهي الدوال اللبية لبيانات مراقبة من النوع الأول لأنواع مختلفة من عرض الحزم والدوال اللبية الحدود ، حيث استعمل نوعين من عرض الحزم عرض الحزمة الشامل ولاربع دوال ولاربع دوال لبية gobal bandwidth وعرض الحزمة الموضعي local bandwidth Rectangle, Epanechnikov, Biquadratic and Triquadratic وكذلك توضيف دالة مقترحة في الدوال اللبية كافة ولتجربتين محاكاة مختلفتين لمقارنة تلك المقدرات وقد أثبتت النتائج أفضلية أسلوب عرض الحزمة الموضعي local bandwidth لجميع أنواع الدوال اللبية الحدود وافضلية إستعمال الدوال اللبية الحدود للدالة المقترحة .2x Rectangle والدالة 2xEpanechnikov لاكثر التجارب المقامة .

الكلمات المفتاحية: عرض الحزمة، بيانات مراقبة، معدل الفشل، الدالة اللبية، تمهيد دالة الفشل