

Empirical Performance of Nonparametric Kernel Estimators with Data-Driven Bandwidths for Lifetime Distributions

 Javaria Ahmad Khan^{1*}

 Muhammad Ejaz²

 Dursun Aydin³

How to cite this article:

Khan, J., Ejaz, M., & Aydin, D. (2025). Empirical performance of nonparametric kernel estimators with Data-Driven bandwidths for lifetime distributions. *Applications of Mathematical Sciences*, 2(1) 1–20.

Received: 2 January 2025 / Accepted: 7 May 2025 / Published online: 15 June 2025

Abstract

During the past several years, many methods for selecting the smoothing parameter have been added in nonparametric kernel regression. The performance of newly introduced selection methods differs quite a bit, but it is better to consider the recently developed methods for the comparison. Due to the demand for automatic data-driven bandwidth selectors for practical statistics, specifically with lifetime distributions, this article provides a review to explain and compare these methods. We conducted an extensive simulation study along with varied estimators and sample sizes to assess the relative performance of the 20 selection methods. From MSEs of simulation, the suitable selection criteria are provided for a suitable bandwidth selection for both classical and plug-in categories. Also, we illustrated these methods with a real example, where we concluded that GCV and Mallows's C_p perform better with LOWESS estimator.

Keywords: Nonparametric Regression, Kernels, Bandwidth, Lifetime distributions, Comparison.

1 Introduction

In nonparametric statistics, especially in lifetime distributions, kernel estimators are

¹Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan

Corresponding Author: jakhan0@yahoo.com

²Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan

³Department of Statistics, Muğla Sıtkı Koçman University, Muğla, Turkey



important for smoothing and data analysis. The choice of the bandwidth parameter is important because it affects the bias and variance of the estimator. Recent developments have introduced a variety of bandwidth selection methods, including classical methods and new data-driven methods that meet the requirements of automated and reliable methods for useful calculations role.

Recent research emphasizes the importance of robust non-parametric regression and efficient bandwidth selection. Čížek and Sadıkoğlu (2020) review robust non-parametric regression, and emphasize robustness in the face of outliers. Ali, Hayawi, and Botani (2023) propose methods for estimating bandwidth in Nadrai-Watson kernel regression using universal threshold levels. Linton and Xiao (2019) focus on computational efficiency under dynamic heterogeneous skedasticity, focusing on complex real-world data.

New bandwidth selection criteria have also emerged. Smith, Li, and Zhang (2019) propose new formulas for kernel density estimation, while Jones-Ma (2020) explores machine-learning techniques to increase selection efficiency. Chen and Fan (2021) provide methods using data for local polynomial regression do, thereby improving performance on a variety of data types.

Other developments include hybrid methods for bandwidth selection, demonstrated by Zhao and Li (2022) to provide robust performance in kernel density calculations. Garcia and Fernandez (2023) compare different methods for lifetime data analysis, delivering actionable insights. Putra Fadhalur Rahman (2023) discusses knots and bandwidth optimization in territory weighted truncated spline regression using generalized cross-validation. Lawrence, (1995) presents a good overview of many of these distributions and include the exponential, logistic, log-logistic, gamma, lognormal, Weibull and Gumbel distributions.

In nonparametric kernel regression selection of bandwidth is a critical issue that becomes more serious due to lack of devices or guidelines for practitioners. Generally, in bandwidth studies, the aim was typically to introduce a new method and compare its performance with others, see e. g. Stute, (1992), Kokonendji (2007). There are hardly any reviews available for nonparametric regression that compares the different existing bandwidth selection procedures for continuous non-normal responses which belong to exponential family of distributions. General reviews or comparisons can be found mostly about bandwidth selection in nonparametric regression for normal responses; see Rice(1984), Aydına (2016) and others. Some considered the non-normal responses, but they focused only on discrete responses that belong to exponential family. For non-normal responses, a vast literature is available for discrete data. In case of life time distributions, particularly, Brown (2010), Cai and Zhou (2010), and Chen (2015) among others.

As it can be examined, almost not any publication is available, where author considered a continuous type of distributions as a response. Here we focused on lifetime distribution responses, and select optimum bandwidth among 20 bandwidths for different sample sizes. This study focuses on the kernel smoothing methods used in nonparametric kernel regression.

This paper is organized in five sections. The second section presents the kernel smoothing methods for nonparametric regression. The section 3 discusses the classical and plug-in type of methods for selecting the penalty parameter. Section 4 indicates the Monte Carlo simulation experiments and real data example. Section 5 gives the conclusions and recommendations.

2 Kernel Estimators

For nonparametric kernel regression model,

$$Y_i = m(X_i) + \varepsilon_i, \quad i=1, 2, \dots, n \quad (1)$$

where, X_i is predictor variable, ε_i is independent error terms with zero mean and common variance σ^2 , and m is an unknown smooth function, Y_i is response variable and follows exponential, Gamma and Weibull distributions (which are member of exponential family of distribution).

To estimate $m(X_i)$ given in (1), we focus only on Kernel estimators. Different Kernel methods are available in literature. For this purpose, we used following weights to estimate “ m ” of (1). We have

$$\hat{m}(x) = \sum_{i=1}^n w_{ni}(x) y_i.$$

The function $\hat{m}(x)$ is called “linear smoother” for appropriately defined weights $w_{ni}(x)$, usually with requirement that weights has a property that $|w_{ni}(x)| = 0$ if $|X_i - x| \geq a_n$ (Bhattacharya & Müller, 1993).

Boente and Fraiman (1989) studied the assumptions related to weights which are; **Assumption.** The weight functions w_n satisfy these conditions

- (i.) $\max_{1 \leq i \leq n} \sum_{l=1}^n w_{ni}(x) = O(1)$
- (ii.) $\max_{1 \leq i, l \leq n} \sum_{l=1}^n w_{ni}(x) = O(n^{-2/3})$
- (iii.) $\max_{1 \leq i \leq n} \sum_{l=1}^n w_{ni}(x) I(|x_i - x_l| > a_n) = O(h_n)$, where $I(\cdot)$ is an indicator function, a_n satisfies $\limsup_{n \rightarrow \infty} n a_n^3 < \infty$, and h_n satisfies $\limsup_{n \rightarrow \infty} n h_n^3 < \infty$.
- (iv.) $tr(w'_n w_n) = tr(w_n) = O(h^{-1})$, where $tr(A)$ denotes the trace of a square matrix A .

Assumption. The bandwidth parameter h satisfies $nh^8/\sqrt{(\log(n))} \rightarrow 0$ and $nh^2/(\log(n))^2 \rightarrow \infty$ as $n \rightarrow \infty$.

Stone (1977) studied the conditions for estimation consistency. He showed that $\hat{m}$ is consistent (actually he showed that $E|Y|^r < \infty$ implies $E|\hat{m}(X) - m(X)|^r \rightarrow 0$) if

1. $\sum_{i=1}^n w_{ni}(x) 1_{\{|X_i - x| > \alpha\}} \rightarrow 0$ in probability for all $\alpha > 0$;
2. $\sum_{i=1}^n w_{ni}(x) \rightarrow 1$ in probability
3. $\max_{1 \leq i \leq n} w_{ni}(x) \rightarrow 0$ in probability; here exist constants $C, D > 0$ such that
4. $E \sum_{i=1}^n |w_{ni}(x)| f(X_i) \leq C E f(X)$ for any non-negative Borel function f ; and
5. $P(\sum_{i=1}^n |w_{ni}(x)| \leq D) = 1$ for all $n \geq 1$.

If a weighting scheme is consistent, (1) and (3) are also necessary, and if it is nonnegative and consistent, then (1), (2), (3) are necessary.

Different approaches which we will consider in this paper, to estimate $w_{ni}(x)$ are;

Nadaraya-Watson (NW) Estimator (Nadaraya, [1964](#); Watson, [1964](#)), K-Nearest Neighbor (k-NN) Estimator Royall ([1966](#)), Priestley-Chao (PC) Estimator (Priestley and Chao, [1972](#)), Locally Weighted Regression and Smoothing Scatter Plots (LOESS) Cleveland, ([1979](#)) and Gasser and Muller (GM) Estimator Gasser and Muller, ([1984](#)).

There is vast literature in which researchers focus on their performance, behavior and properties. In which they enhanced the performance of these estimators. The asymptotic properties of the estimators give us valuable insights into its performance. Chu and Marron ([1991](#)) provided an extensive comparison and also discussed their merits. Some others are; Cheng ([1984](#)) proved that k- nearest neighbor is uniformly consistent with probability one. He also showed that its rate of convergence is also near-optimal.

The bias of local polynomial estimator decreases with h quadratically, that means that small bandwidths h gives estimators with low bias, whereas large bandwidths provide largely biased estimators. Similarly, the variance decreases as a factor of $(nh)^{-1}$ this is related to the so-called effective sample size nh . Cui and Wei ([2013](#)), Fan ([2018](#)) established strong consistency (point-wise) under weaker conditions of the well-known Nadaraya-Watson estimator. Such consistency provided a general approach for constructing strongly consistent kernel estimates of regression functions and also permit to apply kernels with unbounded support and even not integrable ones.

He ([2019](#)) showed the strong and complete consistency of Priestly- Chao estimator in nonparametric regression model with their rates by using widely orthant dependent errors under some general conditions. They showed that these rates can be approximate at $O(n^{-1/3})$. Zhang ([2022](#)) studied the weak consistency for the Gasser and Müller (G-M) kernel estimator (a weighted kernel estimator with integral form) in a nonparametric regression model based on negatively associated random errors under some suitable conditions. To establish the weak consistency of the estimator GM estimator they used the following assumptions

- $K \in \text{Lip}(r_1)$, for some $r_1 > 0$, which means that K is a Lipschitz function of order r_1 ;
- $\int_{-\infty}^{\infty} K(u)du = 1$, $\int_{-\infty}^{\infty} |K(u)|du < \infty$, and $K(\cdot)$ is bounded in $\mathbb{R}^1$;
- $g \in \text{Lip}(r_2)$, for some $r_2 > 0$;
- $h_n \rightarrow 0, \delta_n \rightarrow 0$, as $n \rightarrow \infty$
- $h_n^{-1} \left[\left(\delta_n / h_n \right)^{r_1} + \delta_n^{r_2} \right] \rightarrow 0$ as $n \rightarrow \infty$
- $\sum_{i=1}^n w_{ni}(x) \rightarrow 1$ as $n \rightarrow \infty$
- $\sum_{i=1}^n |w_{ni}(x)| \leq C$ for all $n \geq 1$
- $\sum_{i=1}^n |w_{ni}(x)|^q \rightarrow 0$ as $n \rightarrow \infty$ for some $1 < q \leq 2$.

2.1 Bandwidth Selection

In literature, various methods of bandwidth are featured. Basically bandwidth selection methods can be categorized in two different categories; Classical and Plug-in Hardle, ([1990](#)), Jones ([1995](#)), Sheather and Jones ([1991](#)). The ideal value of h , minimized the mean integrated square error is;

$$h_{AMISE} = \left[\frac{R(k)}{nR(f'')(f x^2 k)^2} \right]^{\frac{1}{5}}, \quad (2)$$

Where $R(K) = \int_R K(x)^2 dx$, the function $R(f'')$ measures the total curvature of f . We have implemented 21 bandwidth selection methods in our study and seek for the optimal method that minimizes averaged mean integrated square error (AMISE). The details regarding classical and plug-in bandwidth methods are given in following sections;

2.2 Classical Bandwidths

In this category, different methods are included, which are mentioned in Table 1 with their notations and references. These are more or less natural extensions of methods used in parametric modeling.

Table 1: Classical Bandwidths

#	h	Description	References
1	Normal Scale Rule (NSR)	$h = 1.06\sigma(n)^{-\frac{1}{5}}$	Silverman (1986)
2	Unbiased Cross Validation (UCV)	$UCV(h; r) = \frac{R(K^{(r)})}{nh^{2r+1}} + \frac{(-1)^r}{n(n-1)h^{2r+1}} \sum_{i=1}^n \sum_{j=1, j \neq i}^n (K^{(r)} * K^{(r)} - 2K^{(2r)}(\frac{X_j - X_i}{h}))$	Bowman, (1984), Rudemo, (1982)
3	Biased Cross Validation (BCV)	$BCV1(h; r) = \frac{R(K^{(r)})}{nh^{2r+1}} + \frac{1}{4} \mu_2^2(K) h^4 \hat{\theta}_{r+2}(h)$ $BCV2(h; r) = \frac{R(K^{(r)})}{nh^{2r+1}} + \frac{1}{4} \mu_2^2(K) h^4 \bar{\theta}_{r+2}(h)$	Scott and Terrell, (1987) Jones and Kappenman, (1992)
4	Maximum Likelihood Cross Validation (MLCV)	$MLCV(h) = n^{-1} \sum_{i=1}^n \log[\hat{f}_{h,i}(x)]$	Habbema (1974)
5	Modified Cross Validation (MCV)	$MCV(h; r) = \frac{R(K^{(r)})}{nh^{2r+1}} + \frac{(-1)^r}{n(n-1)h^{2r+1}} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \varphi^{(r)}(\frac{X_j - X_i}{h})$	Stute, (1992)
6	One-sided Cross Validation (OSCV)	$OSCV(b) = \frac{1}{n-m} \sum_{i=m+1}^n (r_b^{-i}(x_i) - Y_i)^2$	Savchuk (2013)
7	Indirect Cross Validation (ICV)	$ICV(x, b) = \frac{1}{n} \int_{-\infty}^{\infty} \phi\left(\frac{x-\mu}{\eta}\right) \hat{f}_b^2(u) du - \frac{2}{n\eta} \sum_{i=1}^n \phi\left(\frac{x-X_i}{\eta}\right) \hat{f}_{b,-i}^2(X_i)$	Savchuk (2013)
8	Smooth Cross Validation (SCV)	$SCV(h) = (nh)^{-1} R(K) + \mathbb{I} \mathbb{S} B(h)$	Hall (1992)
9	Generalized Cross Validation (GCV)	$GCV(h) = \frac{RSS_h}{n(1 - \frac{w(0)}{nh})^2}$	Craven and Wahba, (1978)
10	Complete Cross Validation (CCV)	$CCV(h; r) = R(\hat{f}_h^{(r)}) - \bar{\theta}_r(h) + \frac{1}{2} \mu_2(K) h^2 \bar{\theta}_{r+1}(h) + \frac{1}{24} (6\mu_2^2(K) - \delta(K)) h^4 \bar{\theta}_{r+2}(h)$	Jones and Kappenman, (1992)

11	Trimmed Cross Validation (TCV)	$TCV(h; r) = \frac{R(K^{(r)})}{nh^{2r+1}} + \frac{(-1)^r}{n(n-1)h^{2r+1}} \sum_{i=1}^n \sum_{j=1; j \neq i}^n \varphi^{(r)}\left(\frac{X_{[j]} - X_{[i]}}{h}\right)$	Feluch and Koronacki (1992)
12	Akaike Information Criterion (AIC)	$AIC(\hat{\theta}) = \sum_{i=1}^n D\left(Y_i, \hat{\theta}(x_i)\right) + 2v_1$	Akaike (1974)
13	Akaike Information Criterion Corrected (AICC)	$AIC_C(h) = \log \frac{\ (S_h - I)y\ ^2}{n} + 1 + \frac{2\{\text{tr}(S_h) + 1\}}{n - \text{tr}(S_h) - 2}$	Hurvich (1998)

2.3 Plug in Bandwidths

In this category, estimate of f which is mostly approximated through Taylor series expansions is written as a function of unknown f . this pilot estimate of f is then "plugged in" to estimate mean integrated squared error.

Table 2: Plug-in Bandwidths

#	h	Description	References
1	Bandwidth of Gasser, et al. (GKK)	$CV(b) = \sum_{i=1}^n v(t_i)(\hat{f}_i(t_i; b) - Y_i)^2 (t_i - t_{i-1})$	Gasser (1991)
2	Mallow C_p (Mallow)	$RR(h) = RSS_h - n\sigma^2 + 2\sigma^2 \frac{w(0)}{h}$	Mallows (1973)
3	Bootstrap without Sampling (bWOs)	$\begin{aligned} \overline{MISE}_{boot}^*(h) &= \frac{1}{2\pi nh} \int \varphi_k(t) ^2 \varphi_u(t/h) ^{-2} dt \\ &- \frac{1}{\pi} \int \hat{\varphi}_{X,g}(t) ^2 \varphi_k(ht) dt \\ &+ \frac{n-1}{2\pi n} \int \hat{\varphi}_{X,g}(t) ^2 \varphi_k(ht) dt \end{aligned}$	(Wan & Wang, 2011)
4	Bootstrap with Sampling (bWs)	$\widehat{Var}(f_n(x; h)) = B^{-1} \sum_{j=1}^B \int (f_{nj}^*(x; h) - \bar{f}_{nj}^*(x; h))^2 dx$	(Wang & Wang, 2011)
5	Bandwidth of Altman and Leger (AL)	$\hat{h}_{opt} = (0.25\hat{V}_2/\hat{B}_3)^{1/3} n^{-1/3}$	(Altman & Leger, 1995)
6	Rice' T Criterion (Rice)	$RT(h) = \frac{RSS(h)}{n} \bigg/ \left(\frac{1 - tr(S_h)}{n}\right)$	Rice (1984)
7	Direct Plug-In (DPI)	$\widehat{h_{DPI,2}} = \left[\frac{R(K)}{\mu_2(K)^2 \widehat{\Psi}_4(g_n) n} \right]^{1/5}$	(Sheather & Jones, 1991)

2.4 Finite Sample Performance and Comparison

A fair comparison requires a reasonable number of bandwidths, sample sizes and different estimators. This study established the performance of above-mentioned bandwidths with

most frequent used kernel estimators based on Gaussian kernel, which is

$$\text{Gaussian Kernel} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad -\infty < x < \infty$$

We have used three life time distributions one after another in (1), i.e. Exponential distributions, Gamma distribution and Weibull distribution as responses. We consider $x_i \sim U(-1, 1)$ and $\varepsilon_i \sim N(0, 1)$. For sample sizes $n = 25, 50, 100, 200, 300, 500$, with 500 replications.

Specifically, we set $Y_i \sim (\text{Exp}(1), \text{Gamma}(1.5502, 91.1034), \text{Weibull}(76.318, 1.4))$.

2.5 Empirical Evaluations for Classical

For each simulated data set used in the experiments, the mean square error (MSE) values were used in order to evaluate the performance of bandwidth with estimators.

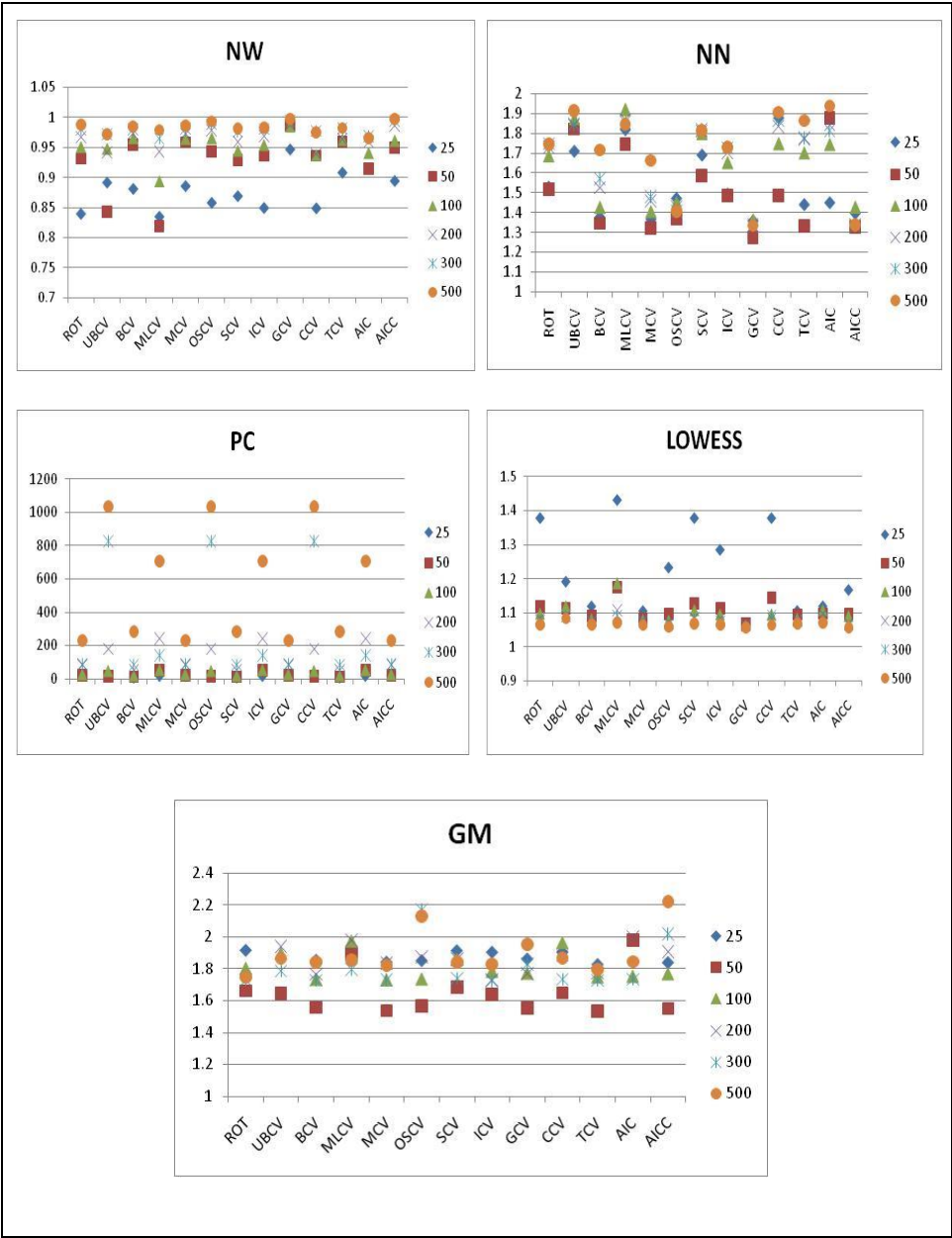


Figure 1: Classical Bandwidth Performance for Different Estimators (Exp)

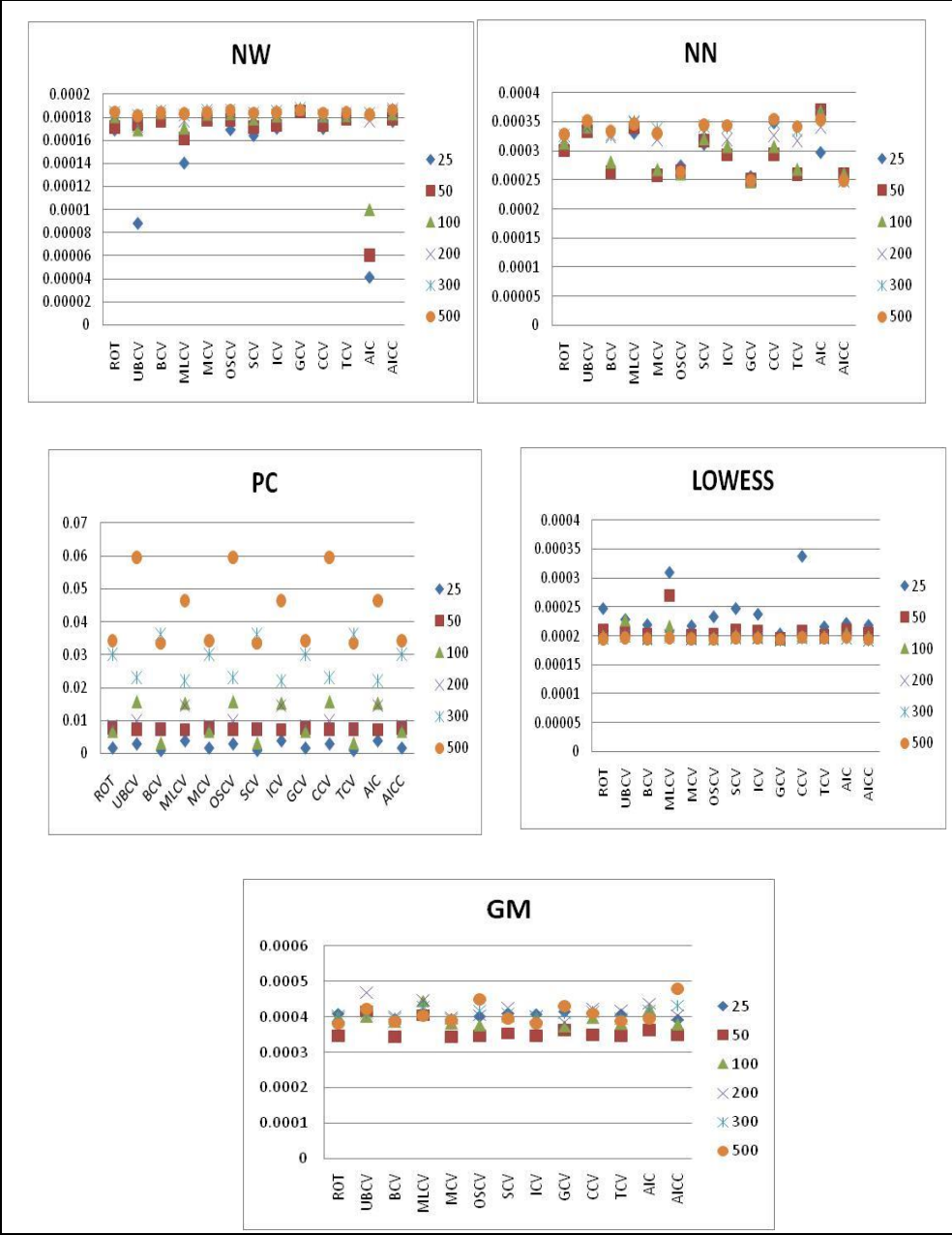


Figure 2: Classical Bandwidth Performance for Different Estimators (Gamma)

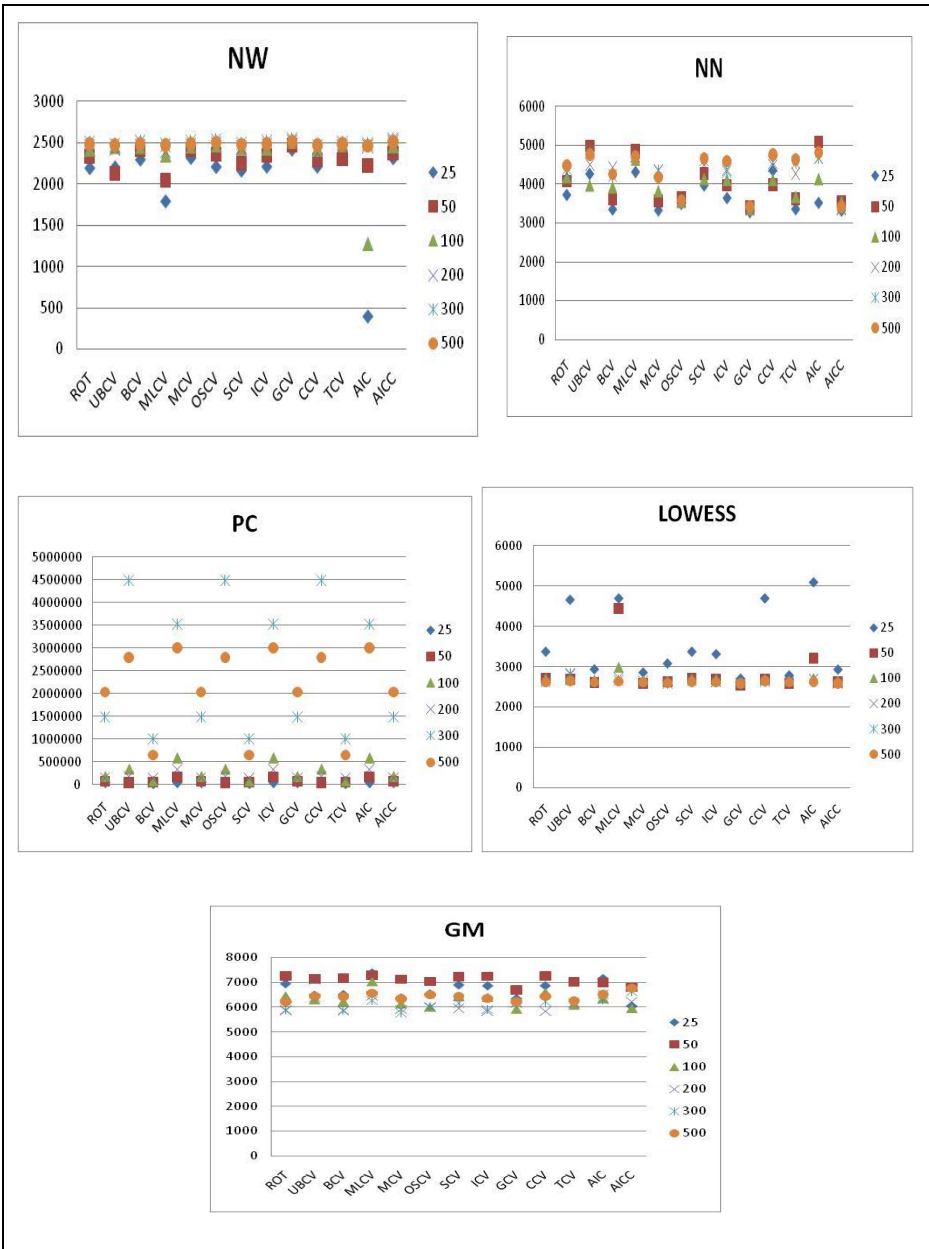


Figure 3: Classical Bandwidth Performance for Different Estimators (Weibull)

Numerical findings of our work have been summarized in the graphs. For all life time distributions (under consideration) MSEs of NW, are minimum comparatively others and they over smooth in all cases. For LOESS, the MSE indicate for increasing sample size, i.e., the bigger the n , the more they tend to under smooth. PC and GM performs very poor in MSE context. By use of PC all bandwidths over smooth for large sample sizes.

In Figure 1-3, for all responses, NW provided the smallest MSE but empirical performance of NW is not consistent; their MSEs increases as sample size increases.

Performance of no single bandwidth is better in all sample sizes. According to the results, in Figure 1-3, for LOESS and NN, on the whole GCV had the best empirical performance for all sample sizes. Particularly, performance of AICC is same as GCV for large sample size when LOESS is used. It is indicated that PC and GM provided better MSE with different bandwidths but with life time distributions their performance is underprivileged.

2.6 Empirical Evaluations for Plug-In

We will now summarize the performance for Plug-in bandwidths according to MSEs. Overall, we are providing here the comparison of the methods along different estimators over lifetime distributions. For each of above mentioned four estimators, we show the MSEs for all sample sizes, i.e., for $n = 25, 50, 100, 200, 300$ and 500 . In order to examine the most important results, NW estimator provides smallest MSEs, but here too; results are increasing with increase of sample size.

Empirical performance of NW estimator is better with Altman and Leger plug-in bandwidth then the rest of bandwidth selection methods. From Figure 4-6 it is evident that Mallow's C_p performs very well for all type of sample sizes with LOESS and NN estimator. It is important to note that for small sample size ($n = 25, 50$) performance of GKK is quite similar to Mallow's C_p for both cases (LOESS and NN). In case of PC, with all responses Rice' T bandwidth performs well, compared with all other candidates. By using GM estimator, with lifetime distributions, we showed that GKK performs better for small sample sizes but in case of large sample size, MSE's derived by Bootstrap bandwidth (without sampling) are smallest among other bandwidths.

2.7 Real Data Example

In this subsection, we present the behavior of different bandwidths for NW, LOESS and NN. PC and GM are omitted due to poor performance. Data is adopted from ReliaSoft (2015), 14 units are used. Here, y_i are responses (time) and strongly follow exponential distribution then the other lifetime distributions, which can be examined by

Table 3: Identification of Response Distribution

	Exponential	Gamma	Weibull
Kolmogorov-Smirnov statistic	0.14453390	0.10034541	0.09755905
Anderson-Darling statistic	0.54047811	0.16375630	0.16096807
Akaike's Information Criterion	136.5865	135.8744	135.5197
Bayesian Information Criterion	137.2256	137.1526	136.7978

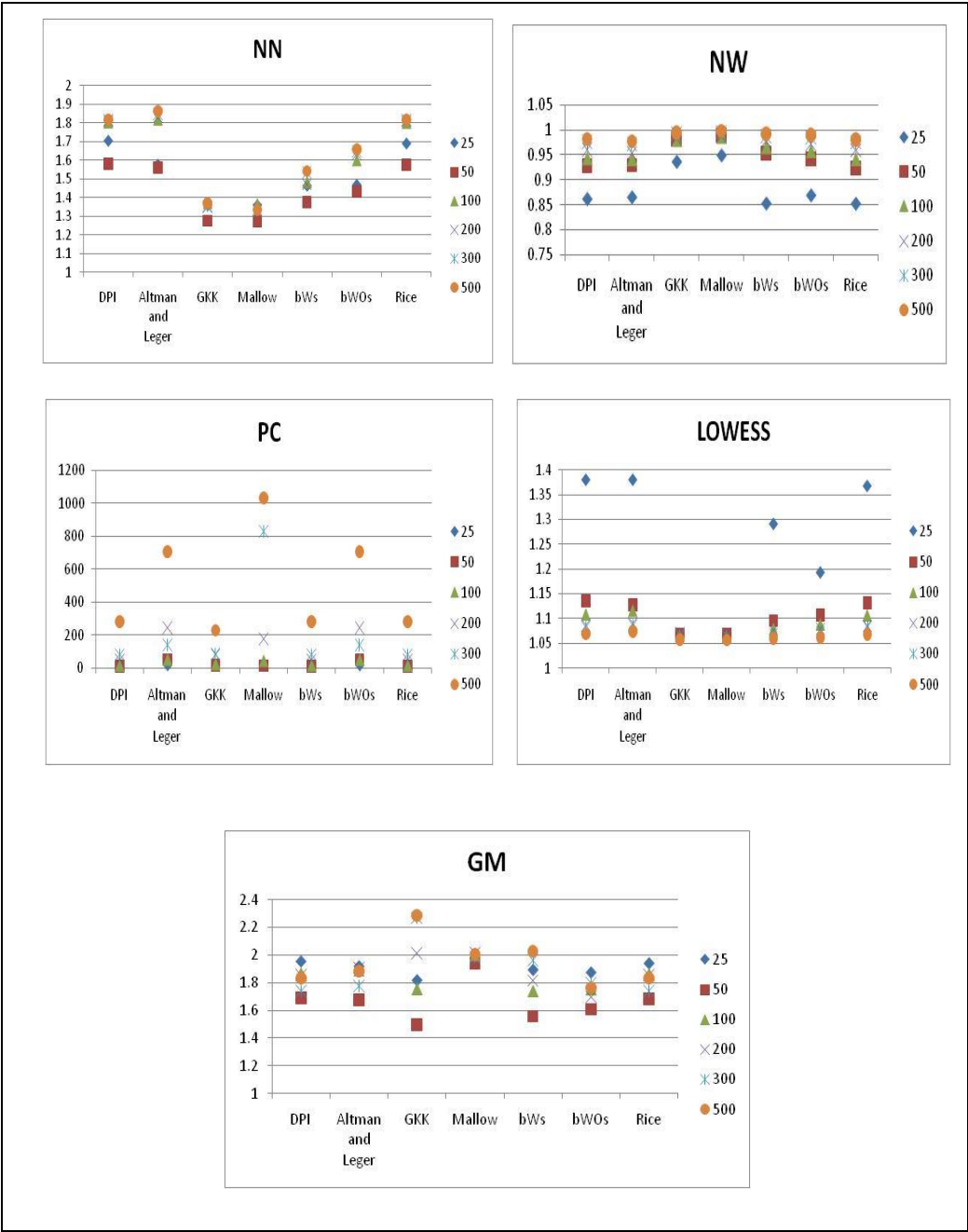


Figure 4: Plug-in Bandwidth Performance for Different Estimators (Exp)

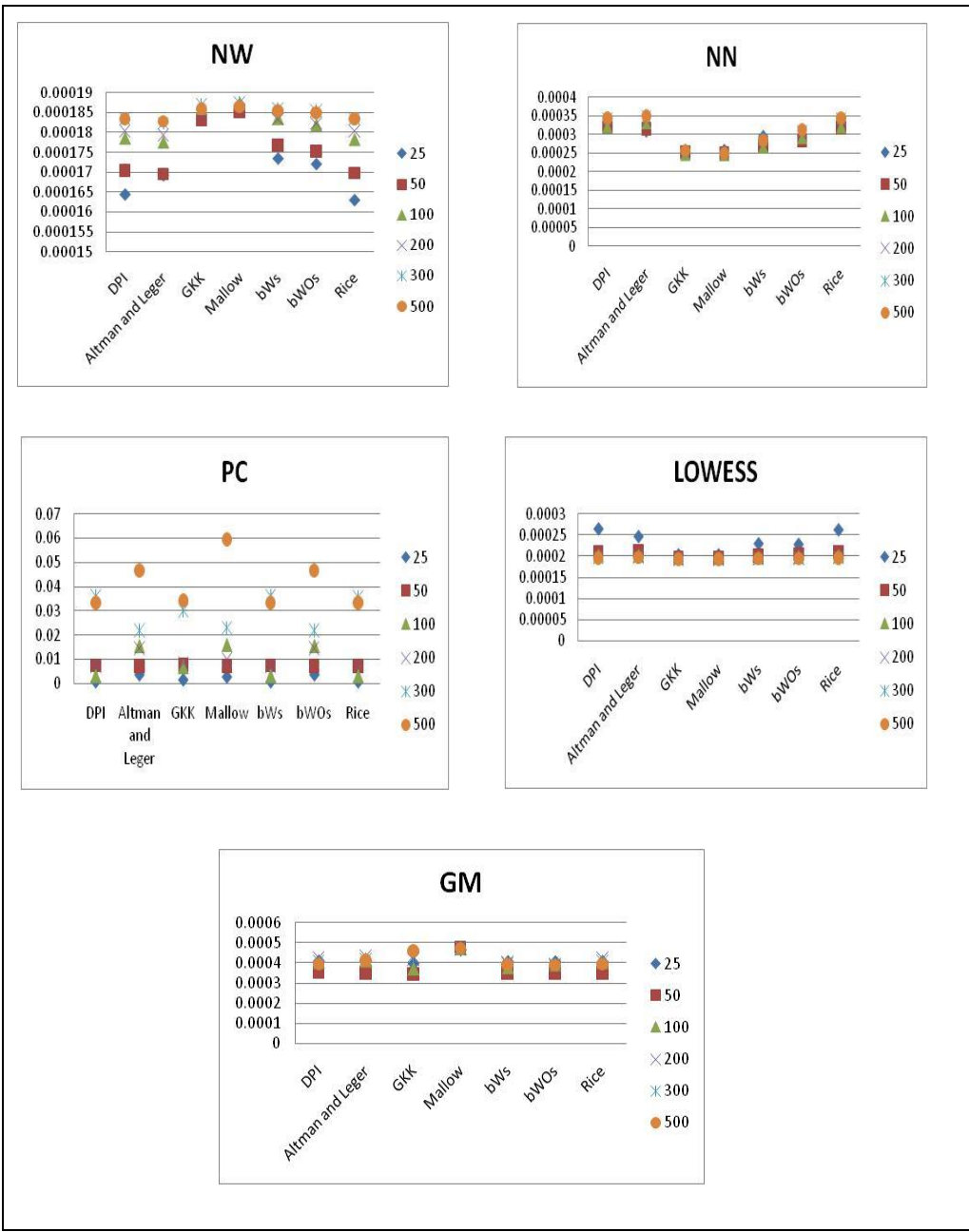


Figure 5: Plug-in Bandwidth Performance for Different Estimators (Gamma)

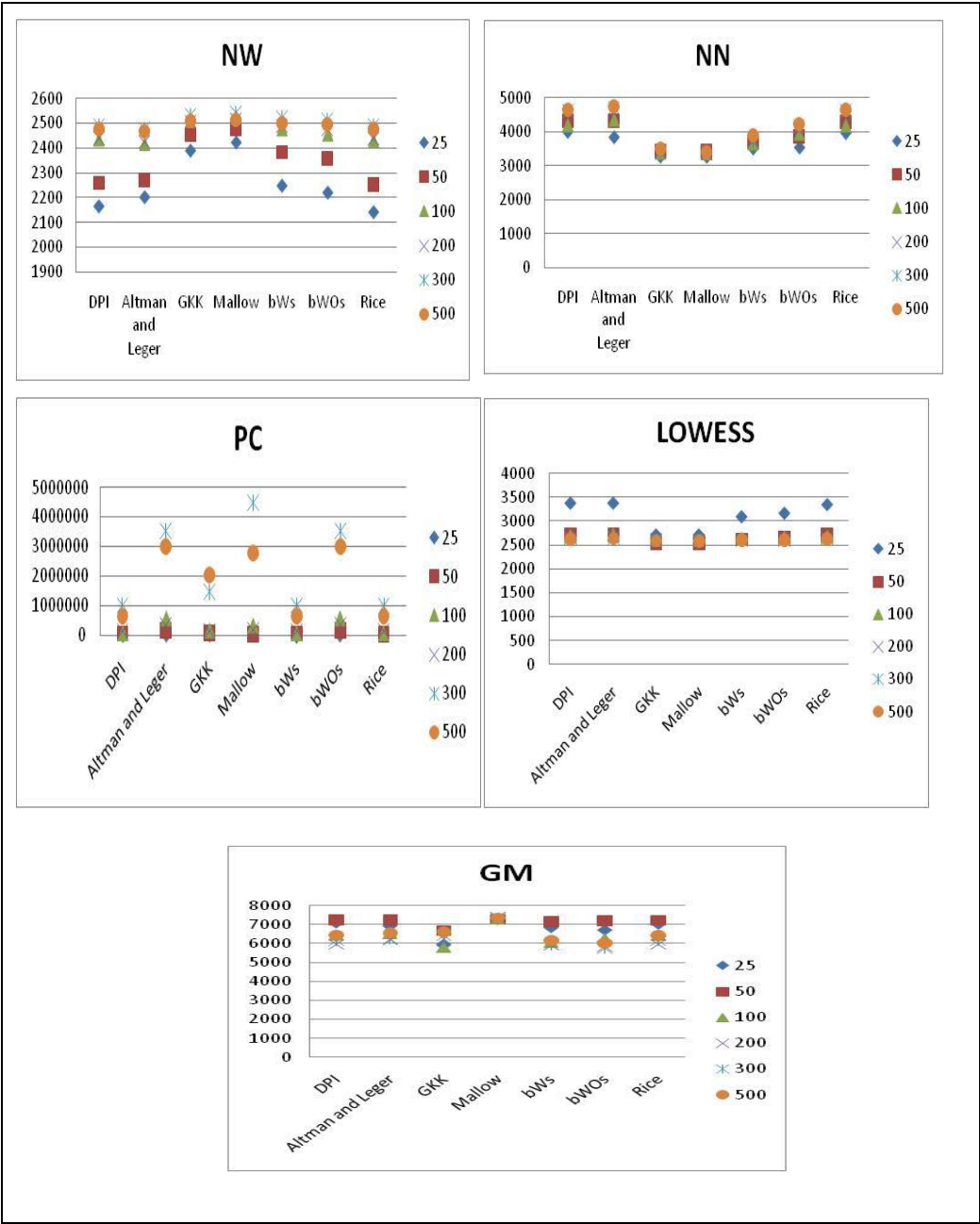


Figure 6: Plug-in Bandwidth Performance for Different Estimators (Weibull)

Generally, results of NW, LOESS and NN verified our simulation results. In which, MSEs calculated by NW estimator are smallest than others but not consistent

Table 4: MSE of Real Data Example

Bandwidths	Kernel Estimators		
	NW	NN	LOESS
ROT	92.57	38.67	3350.84
UBCV	88.48	38.26	3350.84
BCV	380.58	50.41	3999.03
MLCV	121.61	41.19	3347.15
MCV	380.57	50.41	3999.03
OSCV	0.04	52.94	3400.00
SCV	60.25	34.85	3376.34
ICV	109.62	40.23	3347.15
GCV	842.97	54.96	3999.03
CCV	18.75	27.22	3292.86
TCV	88.48	38.26	3350.84
AIC	192.97	45.32	3670.84
AICC	0.001	26.87	3400.00
DPI	46.20	32.70	3309.76
Altman and Leger	78.74	37.20	3350.84
GKK	610.20	53.19	3999.03
Mallow	873.35	55.15	3999.03
bWs	69.67	36.11	3376.34
bWOs	142.41	42.63	3346.39
Rice	45.14	31.95	3233.78

3 Conclusion

The problem of choosing an optimum smoothing parameter is as old as the nonparametric estimation. It is a modern data analytic method with its expertise in capturing the local humps and valley in the distributions or regression functions. Kernel regression is becoming a famous implementation procedure for explorative empirical research and can be found in statistical, mathematical and econometric software packages, but selecting the bandwidth is still the main difficulty for researchers and no bandwidth comes with an undertaking that it will work satisfactorily in all situations.

A sensible data analytic strategy is that of obtaining estimates for a variety of bandwidths and this article is trying to present an inspection, comparison and guiding principle for kernel regression with lifetime distributions. Although it is true that, especially for large and huge data sets, Spline estimation is much more common, but this method has some disadvantages, that it can be used only for continuous predictors and is limited to only one predictor and locally weighted smoothing etc. Presumably, we are the earliest to offer such a widespread review and comparison study for selection of bandwidth methods in the context of kernel regression with non-normal responses. We have discussed, implemented and compared 20 bandwidth methods, with 5 kernel estimators.

The probably appealing results are summarized in the section of simulation. Where we could examine which methods behave quite better, although no selection method performed uniformly best. For Classical bandwidths category, GCV bandwidth performs well for different sample sizes and in category of plug-in, Altman and Leger bandwidth is the best selection criterion. While using NN estimator, GCV and Mallow's C_p are also best criteria for all sample sizes. In case of GM estimator, MSEs are largest as compared to other estimator results. No common bandwidth performs best for all sample sizes.

The performance of LOWESS is better than NN, NW, PC and GM on the basis of MSE with all responses. LOWESS obtained the consistent regression errors. For all sample sizes, GCV and Mallow's C_p are best methods from classical and Plug-In categories, respectively. NW estimator provides smallest MSE for all type of bandwidths but their results are not consistent. In this case, according to the above results, we can make the following suggestions: In case when life time distribution(s) is presented as a response, it is resultant that LOESS estimator with GCV (in case, when Classical type of bandwidths has to use) and Mallow's C_p (in case, when Plug-In type of bandwidths have to use) criterion, because of its superior empirical performance; otherwise the close competitor is AICC, because it perform similar to GCV for large sample sizes.

Author Contributions

Javaria Ahmad Khan, Muhammad Ejaz, and Dursun Aydin contributed to the conception and development of the study, the analytical work, interpretation of the findings, and preparation of the manuscript. All authors contributed to writing-original draft and writing-review and editing. All authors read and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Conflict of Interest

The authors declare that they have no financial, professional, institutional, commercial, or personal conflicts of interest that could have influenced the conduct, interpretation, or presentation of this study.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

This study uses simulated data and an illustrative lifetime-data example. The simulation design, analytical procedures, and sources supporting the findings are described in the manuscript and its cited references. No new human-participant data were collected.

Ethical Approval

Ethical approval was not required because this study did not involve human participants, animals, biological specimens, or identifiable personal information.

Informed Consent

Not applicable. The study did not involve human participants or primary participant-data collection.

Use of Generative AI

The authors declare that no generative artificial intelligence tools were used to conduct the literature search or synthesis, generate the scholarly content, interpret the evidence, or formulate the conclusions.

4 References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723. [10.1109/TAC.1974.1100705](https://doi.org/10.1109/TAC.1974.1100705)
- Ali, T. H., Hayawi, H. A. A. M., & Botani, D. S. I. (2023). Estimation of the bandwidth parameter in Nadaraya-Watson kernel non-parametric regression based on universal threshold level. *Communications in Statistics-Simulation and Computation*, 52(4), 1476-1489.
- Altman, N., & Leger, C. (1995). Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference*, 46(2), 195-214. [https://doi.org/10.1016/0378-3758\(94\)00102-2](https://doi.org/10.1016/0378-3758(94)00102-2)
- Aydina, D., Gunerib, O. I., & Fitc, A. (2016). Choice of bandwidth for nonparametric regression models using kernel smoothing: A simulation study.
- Bhattacharya, P., & Müller, H.-G. (1993). Asymptotics for nonparametric regression. *Sankhyā: The Indian Journal of Statistics, Series A*, 420-441. <https://www.jstor.org/stable/25050952>
- Boente, G., & Fraiman, R. (1989). Robust nonparametric regression estimation for dependent observations. *The Annals of Statistics*, 17(3), 1242-1256 [2010.1214/aos/1176347266](https://doi.org/10.1214/aos/1176347266)
- Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2), 353-360. <https://doi.org/10.1093/biomet/71.2.353>
- Brown, L. D., Cai, T. T., & Zhou, H. H. (2010). Nonparametric regression in exponential families. [2010.1214/09-AOS762](https://doi.org/10.1214/09-AOS762)
- Cai, T. T., & Zhou, H. H. (2010). Nonparametric regression in natural exponential families. In *Borrowing Strength: Theory Powering Applications—A Festschrift for Lawrence D. Brown* (Vol. 6, pp. 199-216). Institute of Mathematical Statistics. <https://projecteuclid.org/ebook/Download?urlId=10.1214/10-IMSCOLL614&isFullBook=false>
- Chen, Y. (2015). Nonparametric Regression In Natural Exponential Families: A Simulation Study. https://tigerprints.clemson.edu/all_theses/2204
- Chen, Y., & Fan, J. (2021). Data-driven bandwidth selection for local polynomial regression. *Journal of Computational and Graphical Statistics*, 30(4), 1089-1103. <https://doi.org/10.1080/10618600.2021.1910102>
- Cheng, P. E. (1984). Strong consistency of nearest neighbor regression function estimators. *Journal of Multivariate Analysis*, 15(1), 63-72. [https://doi.org/10.1016/0047-259X\(84\)90067-8](https://doi.org/10.1016/0047-259X(84)90067-8)
- Čížek, P., & Sadıkoğlu, S. (2020). Robust nonparametric regression: A review. Wiley

- Interdisciplinary Reviews: Computational Statistics*, 12(3), e1492.
- Chu, C.-K., & Marron, J. S. (1991). Choosing a kernel regression estimator. *Statistical science*, 404-419.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368), 829-836.
- Craven, P., & Wahba, G. (1978). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische mathematik*, 31(4), 377-403. <https://doi.org/10.1007/BF01404567>
- Cui, W., & Wei, M. (2013). Strong consistency of kernel regression estimate. *Open J. Stats*, 3, 179-182. <http://dx.doi.org/10.4236/ojs.2013.33020>
- Fan, J. (2018). *Local polynomial modelling and its applications: monographs on statistics and applied probability* 66. Routledge. <https://doi.org/10.1201/9780203748725>
- Feluch, W., & Koronacki, J. (1992). A note on modified cross-validation in density estimation. *Computational statistics & data analysis*, 13(2), 143-151. [https://doi.org/10.1016/0167-9473\(92\)90002-W](https://doi.org/10.1016/0167-9473(92)90002-W)
- Garcia, M., & Fernandez, D. (2023). A comparative study of bandwidth selection methods for lifetime data analysis. *Lifetime Data Analysis*, 29(2), 311-334. <https://doi.org/10.1007/s10985-023-09517-0>
- Gasser, T., Kneip, A., & Köhler, W. (1991). A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association*, 86(415), 643-652. <https://doi.org/10.1080/01621459.1991.10475090>
- Gasser, T., & Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, 171-185. <https://www.jstor.org/stable/4615954>
- Habbema, J., Hermans, J., & Van den Broek, K. (1974). A stepwise discriminant analysis program using density estimation.
- Hall, P., Marron, J., & Park, B. U. (1992). Smoothed cross-validation. *Probability theory and related fields*, 92(1), 1-20. <https://doi.org/10.1007/BF01205233>
- Hardle, W. (1990). *Applied nonparametric regression*. Cambridge university press. <https://doi.org/10.1017/CCOL0521382483>
- He, Q. (2019). Consistency of the Priestley–Chao estimator in nonparametric regression model with widely orthant dependent errors. *Journal of Inequalities and Applications*, 2019(1), 1-13. <https://doi.org/10.1186/s13660-019-2016-8>
- Hurvich, C. M., Simonoff, J. S., & Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 60(2), 271-293. <https://doi.org/10.1111/1467-9868.00125>
- Jones, M., & Kappenman, R. (1992). On a class of kernel density estimate bandwidth selectors. *Scandinavian Journal of Statistics*, 337-349. <https://www.jstor.org/stable/4616251>
- Jones, A., & Ma, Y. (2020). Advances in bandwidth selection for kernel density estimation using machine learning. *Journal of Statistical Computation and Simulation*, 90(7), 1231-1245. <https://doi.org/10.1080/00949655.2020.1716439>
- Jones, W. M. (1995). Kernel Smoothing, Chapman & Hall. In: London.
- Kokonendji, C., Senga Kiesse, T., & Zocchi, S. S. (2007). Discrete triangular distributions

- and non-parametric estimation for probability mass function. *Journal of Nonparametric Statistics*, 19(6-8), 241-254.
<https://doi.org/10.1080/10485250701733747>
- Lawrence, L. M. (1995). Reliability: Probabilistic Models and Statistical Methods. In: New Jersey, Englewood Cliffs, Prentice Hall, Inc.
- Linton, O., & Xiao, Z. (2019). Efficient estimation of nonparametric regression in the presence of dynamic heteroskedasticity. *Journal of Econometrics*, 213(2), 608-631.
- Mallows, C. (1973). Some remarks of Cp. *Technometrics*, 15, 661-675.
- Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1), 141-142. <https://doi.org/10.1137/1109020>
- Priestley, M. B., & Chao, M. (1972). Non-parametric function fitting. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 34(3), 385-392.
<https://doi.org/10.1111/j.2517-6161.1972.tb00916.x>
- Putra, R., & Fadhlurrahman, M. G. (2023). Determination of the best knot and bandwidth in geographically weighted truncated spline nonparametric regression using generalized cross validation. *MethodsX*, 10, 101994.
- ReliaSoft. (2015). *Life Data Analysis Reference*. USA: ReliaSoft Corporation.
<http://www.ReliaSoft.com>.
- Rice, J. (1984). Bandwidth choice for nonparametric regression. *The Annals of Statistics*, 1215-1230. <https://www.jstor.org/stable/2240998>
- Royall, R. M. (1966). *A class of non-parametric estimates of a smooth regression function*. Stanford University.
- Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 65-78. <https://www.jstor.org/stable/4615859>
- Savchuk, O. Y., Hart, J. D., & Sheather, S. P. (2013). One-sided cross-validation for nonsmooth regression functions. *Journal of Nonparametric Statistics*, 25(4), 889-904. <https://doi.org/10.1080/10485252.2013.817575>
- Scott, D. W., & Terrell, G. R. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82(400), 1131-1146.
[10.1080/01621459.1987.10478550](https://doi.org/10.1080/01621459.1987.10478550)
- Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3), 683-690.
<https://doi.org/10.1111/j.2517-6161.1991.tb01857.x>
- Silverman, B. (1986). *Density estimation* (Chapman Hall, London).
<https://www.taylorfrancis.com/books/mono/10.1201/9781315140919/density-estimation-statistics-data-analysis-bernard-silverman>
- Smith, B., Lee, C., & Zhang, X. (2019). New criteria for bandwidth selection in kernel density estimation. *Computational Statistics & Data Analysis*, 135, 1-15.
<https://doi.org/10.1016/j.csda.2019.07.005>
- Stute, W. (1992). Modified cross-validation in density estimation. *Journal of Statistical Planning and Inference*, 30(3), 293-305.
[https://doi.org/10.1016/0378-3758\(92\)90157-N](https://doi.org/10.1016/0378-3758(92)90157-N)
- Wang, X. F., & Wang, B. (2011). Deconvolution estimation in measurement error models: the R package decon. *Journal of statistical software*, 39(10).
<https://doi.org/10.18637/jss.v039.i10>

- Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 359-372. <https://www.jstor.org/stable/25049340>
- Zhang, L., Wang, R., Wang, M., & Wang, X. (2022). Weak consistency for the nonparametric kernel regression estimator based on negatively associated random errors. *Communications in Statistics-Theory and Methods*, 1-18. <https://doi.org/10.1080/03610926.2022.2158342>
- Zhao, H., & Li, W. (2022). Improving bandwidth selection for kernel density estimation with hybrid methods. *Statistics and Computing*, 32(3), 467-485. <https://doi.org/10.1007/s11222-021-10020-6>