Libyan Journal of Applied and Contemporary Sciences
ISSN
E: 3136-3709
L: 3136-3695
Vol 1: No 2

ASSESSMENT OF BANDWIDTH SELECTION STRATEGIES TO SMOOTH THE MEAN REGRESSION FUNCTION

Tarek A. Elghazali [1]* ; Fatima H. Alhamroush [2]; Fawzia A. Altobly [3]; Ahmed M. Mami [4]

[1,2,3,4] Department of Statistics, Faculty of Science, University of Benghazi, Benghazi, Libya

Email corresponding author's: tarek.elghazali@uob.edu.ly

Received: 16/8/2026 Accepted: 6/10/2026 Published: 11/10/2026

Copyright © 2026 Author(s) ; Published by Libyan Journal of Applied and Contemporary Sciences
This is an open access article licensed under CC BY: (https://creativecommons.org/licenses/by/4.0)

Abstract

Non-parametric regression is a type of regression analysis in which the functional form of the relationship between the response variable, and the associated predictor variable does not to be specified in order to fit a model to a set of data. There are many different methods for non-parametric regression. In this paper, we have used the local polynomial kernel estimator with the optimal choice of the smoothing parameter(s). Choosing the optimal smoothing parameter(s), which is usually called the bandwidth(s), was considered to be one of the most important issues when using the kernel-based estimators. Moreover, the optimal bandwidth(s) played a great role to uncover the structure of the underlying data. Here, we used simulation approach to make comparisons between four different strategies for selecting the optimal bandwidth(s), namely the Direct Plug In bandwidth selection, the Residual Sum of Square Criterion, the Cross-Validation Criterion, and the Rule of Thumb Criterion. Within the context of these four strategies of selecting the optimal bandwidth(s), there are two different design settings: Fixed Design and Random Design. In addition, four different regression functions have been tested in order to smooth the mean regression function g(x). We have investigated the effect of three important elements: The size of sample (n); the distribution of the errors' (normal or non-normal distribution); and the kind of design (fixed and random). Having conducted 500 simulation studies, interesting results have been achieved.

Keyword : *Local Polynomial Kernel Estimators; Direct Plug-in Selection Strategy; Residual Sum of Square Criterion; the Cross-validation criterion; the Rule of Thumb criterion.*

Doi : https://doi.org/10.64943/ljacs.2026.010213

INTRODUCTION

The primary objective of nonparametric regression methods is to describe the relationship between the response variable Yi and the explanatory variable Xi, typically informed by the structure of a scatter plot. In all cases, some functional relationship between the variables is assumed, and attention is paid to the extent of scatter present in the data (1,2,3). The general form of a nonparametric regression model can thus be expressed as follows:

$$Y_i = g(x_i) + \varepsilon_i, \quad (1)$$

where $g(x_i)$ represents some function that needed to be estimated, and the ε_i is error term with zero mean and constant variance for all i = 1,2,….,n (4) . Of particular interest is the Local polynomial kernel Estimator because it is mathematical tractable and is based on quite simple idea.

Local polynomial kernel Estimator

The Local Polynomial Kernel estimator (5) (Fan, 2018) is defined as

$$\hat{g}(x;p,h) = e_1^T\left(X_{p,x}^T W_x X_{p,x}\right)^{-1}\left(X_{p,x}^T W_x Y\right) \quad (2)$$

Local polynomial kernel regression estimation has been applied across various fields since the late 19th and early 20th centuries. A comprehensive review of its properties and advantages is provided in (5,6) offer a clear introductory exposition of the method (7)

Definition: The fundamental principle of the local polynomial kernel estimator is the local fitting of a p^{th} order polynomial to the data using weighted least squares at a specific point of the regression function. To introduce local polynomial kernel regression, consider a set of multidimensional observations (Y_i, X_i), $i = 1,2,\dots,n$, where Y_i is a scalar response variable and $X_i = (X_{i1}, X_{i2}, \dots, X_{id})$ is a d-dimensional explanatory variable. Depending on the dimensionality d, the analysis may focus on the univariate case ($d = 1$) or the multivariate case ($d > 1$). Additionally, local polynomial kernel regression can be conducted under either fixed design or random design (4,5,6,8).

Univariate Local Polynomial Kernel Regression ($d = 1$)

In the univariate setting with a fixed design, the response variables are assumed to be generated from a model in which the explanatory variables are ordered and non-random, with equally spaced differences between consecutive points, i.e., $(X_{i+1} - X_i)$ is constant

$$Y_i = g(x_i) + \varepsilon_i \quad (3)$$

where the errors ε_i are independent random variables with $E(\varepsilon_i) = 0$, and $E(\varepsilon_i^2) = \sigma^2$ (5).

f is called the mean regression function because $E(y_i) = g(x_i)$ and $var[y_i] = \sigma_i^2$. If $var[y_i] = \sigma^2$ for all X_i, the model is considered to be homoscedastic otherwise, the model is said to be heteroscedastic**.**

Here, we present of an explicit formula for the local polynomial kernel estimator in Fixed design case.

Suppose that the $(p+1)th$ derivative of $g(X)$ exists in a neighborhood of X_0. The unknown regression function $g(X)$ can then be locally approximated by a polynomial of order p using Taylor expansion:

$$g(x) \approx \sum_{i=0}^{p} \frac{g^{(i)}(x_0)}{i!}(x - x_0)^i = \sum_{i=0}^{p} B_i\,(x - x_0)^i, \quad (4)$$

where $B_i = \frac{g^{(i)}(x_o)}{i!}$.

For X in a neighborhood of X_0, we then perform a local polynomial regression fit,

$$\hat{B}(x_0) = \arg\min_{B} \sum_{j=1}^{n} [y_j - \sum\nolimits_{i=0}^{p} B_i (x_j - x_0)^i]^2\, k_h(x_j - x_o), \quad (5)$$

where

$$k_h(u) = \frac{1}{h} k\left(\frac{u}{h}\right).$$

where $B = (B_0, B_1, \dots, B_p)^T$ are the local polynomial coefficients, $k(.)$ is the kernel function, and h is the bandwidth controlling the size of the local neighborhood around x_0. Using the standard weighted least squares theory (in the matrix notation) leads to the following explicit solution for the local polynomial regression fit, which is defined as

$$\hat{g}(x;p,h) = e_1^T\left(X_{p,x}^T W_x X_{p,x}\right)^{-1}\left(X_{p,x}^T W_x Y\right), \quad (6)$$

where $Y = (Y_1, Y_2, \dots, Y_n)^T$, W_x be the diagonal matrix of weights with $W_x = k_h(x_i - x)$ and $e_1$is the $(p+1) \times 1$ vector having 1 in the first entry and zeros elsewhere, and $X_{p,x}$ denotes the design matrix $(n \times (p+1))$ Whose (i,j) the element is $(x_i - x)^{j-1}$. To demonstrate how this estimator works and several useful examples of the local polynomial kernel estimation procedures (5, 7).

The Role of Smoothing Parameter or the Bandwidth h

The LPK estimator is composed of three elements: first, the kernel function K; second, the polynomial order p; and third, the smoothing parameter, also known as the bandwidth h.

The selection of the smoothing parameter, or bandwidth h, is widely recognized as the most critical consideration when applying the Local Polynomial Kernel (LPK) estimator. A very small bandwidth reduces the modeling bias because fewer data points fall within the local neighborhood; however, this results in a higher variance. Conversely, a very large bandwidth increases the modeling bias due to the influence of the underlying function, while reducing variance. Therefore, the bandwidth fundamentally controls the complexity of the model by mediating the trade-off between bias and variance (2,5,9,10).

The Problem Statement

Although the local polynomial kernel (***LPK***) regression model offers considerable flexibility and accurate estimation, its performance critically depends on the choice of the bandwidth parameter (***h***). An inappropriate selection of ***h*** can lead to two major issues: over-smoothing, which occurs when ***h*** is too large and causes the loss of fine-scale details, and under-smoothing, which arises when ***h*** is too small, producing a highly variable model with excessive fluctuations.

The central research problem addressed in this study is the selection of an appropriate bandwidth h using effective data-driven strategies to obtain smooth and reliable estimates of the mean regression function $g(x)$.

The Objectives Study

The objectives of this study are to:

1. Evaluate four bandwidth selection strategies for smoothing the mean regression function $(g(x))$, namely the Direct Plug-in (DPI) Criterion, Cross-Validation (CV) Criterion, Rule of Thumb (RT) Criterion, and Residual Sum of Squares Criterion (RSC).
2. **Compare the performance** of these bandwidth selection strategies under different sample sizes, error distributions, regression function shapes, and design settings, including fixed and random designs.
3. Investigate the effects of sample size, regression-function shape, error distribution, and design setting on the accuracy of the estimated mean regression function.
4. Assess and compare the performance of the competing bandwidth selectors using average bias ($ABias$), average variance ($AVar$), and root average mean squared error ($RAMSE$).

METHODOLOGY

This section presents the theoretical framework for bandwidth selection in local polynomial kernel regression, with particular emphasis on estimating and smoothing the mean regression function $g(x)$ under fixed and random design setting. Four bandwidth selection strategies are considered: Direct Plug-In (DPI), Residual Sum of Square Criterion (RSC), Cross-Validation (CV), and Rule-of-Thumb (RT) (11).

Bandwidth Selection Strategies:

i. **Direct Plug-In** (***DPI***): This method relies on accurate estimation of certain theoretical quantities related to data density.
ii. **Residual Sum of Squares Criterion** (***RSC***): This criterion focuses on minimizing the sum of squared residual errors.

iii. **Cross-Validation** (***CV***): This method divides the data into two sets: a training set and a test set, and selects the bandwidth that minimizes the predictive error on the test set.

iv. **Rule-of-Thumb** (***RT***): This method uses simplified mathematical formulas to estimate the bandwidth based on the characteristics of the data.

1) The Direct Plug-In (*DPI*) Bandwidth Selection Strategy: The bias term in the case of local linear kernel estimator $\hat{g}(x; 1, h)$ when utilizing the fixed design case can be seen as

$$bias\big(\hat{g}(x;1,h)\big) = E[\hat{g}(x,1,h)] - g(x) = (1/2)g''(x)h^2\mu_2(k) + O(n^{-1}) + o(h^2) \tag{7}$$

where $\mu_2(K) = \int z^2\, K(Z)dz$ and the notation $o(n^{-1})$ means that some quantity is less than a constant multiplied of n^{-1}. Once n is sufficiently large (i.e. $\lim_{n\to\infty} \frac{O(n^{-1})}{n^{-1}} \leq C$) and the notation $o(h^2)$ means that as $h \to 0$ the quantity $o(h^2)$ converges to zero with rate of $h2$ (i.e. $\lim_{n\to 0} \frac{o(h^2)}{h^2} \leq 0$) whereas the variance term in the case of the local linear kernel estimator $\hat{g}(x; 1,h)$ when utilizing the fixed design case can be seen as

$$Var[\hat{g}(x;1,h)] = (nh)^{-1}R(K)v(x) + o(nh)^{-1}, \tag{8}$$

where $R(K) = \int K(z)^2\, dx$, and $o(nh)^{-1}$ is defined as above.

Thus, the MSE as a point-wise measure of consistency of the estimator $\hat{g}(x;1,h)$ is given by

$$\begin{aligned} MSE\{\hat{g}(x;1,h)\} &= E\{\hat{g}(x;1,h) - g(x)\}^2 \\ &= E[\hat{g}(x;1,h) - g(x)]^2 + Var\{\hat{g}(x;1,h)\} \\ &= bias^2\{\hat{g}(x;1,h)\} + Var\{\hat{g}(x;1,h)\} \\ &= (1/2)g''(x)h^2\mu_2(k) + O(n^{-1}) + o(h^2) + nh)^{-1}R(K)v(x) + o(nh)^{-1} \end{aligned}$$

The suitable global loss criterion, since it consists of the two important properties of the estimator namely: Bias and Variance, to obtain asymptotically optimal bandwidth, is the mean integrated squared error ($MISE$) of $\hat{g}(.;1,h)$. It is given by

$$\begin{aligned} MISE[\hat{g}(.;1,h) &= E(\int \{\hat{g}(x;1,h) - g(x)\}^2\, dx) \\ &= n^{-1}h^{-1}R(k)\int v(x)\, dx + h^4\{\mu_2(k)/2\}^2 \int g''(x)^2\, dx + op\{(nh)^{-1} + h^4\} \end{aligned}$$

Similarly, we can obtain the conditional bias term in the case of local linear kernel estimator $\hat{g}(x;1,h)$ in the random design as follows:

$$[E\{\hat{g}(x;1,h)\} - g(x)(X_1^n)^{-1}] = 2^{-1}g''(x)h^2\mu_2(K) + op(h^2),$$

where the notation $op(h^2)$ denotes smaller order in the probability, $g"(x)$ and $\mu_2(K)$ are as defined earlier. However, the variance term in the case of local linear kernel estimator $\hat{g}(x;\ 1,h)$ in the random design is as follows:

$$Var[\hat{g}(x;1,h)(X_1^n)^{-1}] = nh^{-1}R(K)(b(x))^{-1} \tag{9}$$

In both design settings, the bias expression purely depends on $\boldsymbol{x}$ through $g''(x)$ which reflects the error of the linear fitting. Moreover, the higher amount of curvature in g the more bias produced by the local linear fit. On the other hand, Eq. (9) reflects the reality that the variance is punished by larger

conditional variance and sparser data distributed around the point $\boldsymbol{x}$. However, if $b(.)$ is uniform then it will serve as deflection factor in the approximating variance term (7) .

Having obtained the conditional variance term, we can add again it to the squared of the conditional bias term to form the conditional mean squared error (MSE). Thus, the MSE as a pointwise measure of consistency of the estimator $\hat{g}(x;1,h)$ is given by

$$\begin{aligned} MSE\{\hat{g}(x;1,h)(X_1^n)^{-1}\} &= E\{\hat{g}(x;1,h) - g(x)(X_1^n)^{-1}\}^2 \\ &= E\{\hat{g}(x;1,h) - g(x)(X_1^n)^{-1}\}^2 + Var[\hat{g}(x;1,h)(X_1^n)^{-1}] \\ &= bias^2\{\hat{g}(x;1,h)(X_1^n)^{-1}\} + Var[\hat{g}(x;1,h)(X_1^n)^{-1}] \\ &= [2^{-1}g''(x)h^2\mu_2(K) + op(h^2)]^2 + \{(nh)^{-1}R(K)(b(x))^{-1}\}v(x) + op(nh)^{-1}\} \end{aligned}$$

The most suitable global loss criterion, to obtain an asymptotically optimal bandwidth, is the conditional mean integrated squared error ($MISE$) *of* $\hat{g}(.;1,h)$, and is given by

$$\begin{aligned} MISE\{\hat{g}(.;1,h)(X_1^n)^{-1}\} &= E(\int\{\hat{g}(x;1,h) - \hat{g}(x)\}^2 b(x)\,dx\,(X_1^n)^{-1}) \\ &= (nh)^{-1}R(K)\int v(x)\,dx + h^4\{\mu_2(K)2^{-1}\}^2\int g''(x)^2\,b(x)\,dx + op\{(nh)^{-1} + h^4\} \qquad (10) \end{aligned}$$

The Direct Plug-In (*DPI*) Bandwidth Selection to Smooth the Mean Regression Function: The Direct Plug-in (***DPI***) bandwidth selectors are based on the simple idea of "plugging in" estimates of the unknown quantities that appear in formulae for the asymptotically optimal bandwidth (12) . In the case of local linear kernel estimator, the asymptotical bandwidth under ***MISE*** criterion (13) . can be seen as

$$h_{MISE} = C_1(k)[\int v(x)\,dx\ \{\int x^2 b(x)\,dx\}^{-1}]^{1/5}, \qquad (11)$$

where $C_1(k) = [R(K)\{\mu_2(K)^2]^{1/5} = [2\sqrt{\pi}]^{1/5}$ is only a constant depending on the kernel function and the order polynomial p. It is very obvious that Plugin bandwidth strategies rely on replacing the unknown integrals in Eq. (11) by appropriate estimators.

Ruppert et al. (1995) investigated the kernel estimation of regression functional of the form

$$\theta_{rs} = \int g^{(r)}(x)g^{(s)}(x)b(x)\,dx;\ r,s \geq 0;\ r+s \text{ even} \qquad (12)$$

They found that the natural kernel-type estimator of θ_{22} is

$$\hat{\theta}_{22}(g) = n^{-1}\sum_{i=1}^{n}[\hat{g}^{(2)}(X_i;3,g)], \qquad (13)$$

where $\hat{g}^{(2)}(X_i;3,g) = 2!\ e_3^T(X_3^T, W_x\,X_{3,x})^{-1}(X_3^T, W_x\,Y)$.

Clearly, this leads to new bandwidth selection problem of g. The optimal bandwidth for g is obtained by minimizing the squared bias term that appear in the conditional MSE approximation expression (13)

$$g_{AMSE} = C_2(K)[\sigma^2(b-a)/|\theta_{24}|n]^{1/7}, \qquad (14)$$

where, $C_2(K) = [3/8\sqrt{\pi}]^{1/7}$ if $\theta_{24} < 0$, and $C_2(K) = [15/16\sqrt{\pi}]^{1/7}$ if $\theta_{24} > 0$.
The functional θ_{24} and σ^2 that appear in optimal bandwidth g_{AMSE} Eq. (14), are obtained during partitioning the range of the X_1^n data into N blocks. Then, fitting an ordinary least squares quadratic fit

for each block. To partition the range of the data, we let N be the number of sub-samples and letting x_j denote j^{th}sub-sample of the ordered $t's$. If N divides n and $z = n/N$, then $x_j = X_{((J-1)Z+1)}, \dots, X_{(JZ)}\}$. Otherwise, the allocations to sub-samples need some adjustments. Let $\hat{g}_j^Q(x)$ stands for the least squares quartic fit obtained from data having X_I values in x_j. Thus, the blocked quartic estimators of θ_{rs} and $max(r,s) \leq 4$, are given by

$$\hat{\theta}_{rs}^Q(N) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N}(\hat{g}_j^Q)^{(r)}(X_i)(\hat{g}_j^Q)^{(s)}(X_i)1_{\{X_i \in x_j\}}$$

$$\hat{\sigma}_Q^2(N) = (n-5N)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N}\{y_i - \hat{g}_j^Q(X_i)(X_I)\}^2 1_{\{X_i \in x_j\}}.$$

Although there are other possible ways to obtain these initial regression functional estimates, we have found during that the blocked quartic fit along with the use of mallows C_p is a very successful practical approach (14, 15, 16, 17).

The other unknown quantity in the optimal bandwidth Formula (10) is the variance $\int v(x)dx = (b - a)\,\sigma^2$. For local polynomial kernel regression, a natural approach is to extend the variance estimator of (18), which is based on the zero-degree least squares fit, to the first-degree setting. This leads to the first-degree variance estimator

$$\hat{\sigma}_1^2(\lambda) = v^{-1}\sum_{i=1}^{n}\{Y_i - \hat{g}(X_i;1,\lambda)\}^2 \tag{15}$$

Here,

$$v = n - 2\sum_i w_{ii} + \sum_i\sum_j w_{ij}^2, \text{ and } w_{ij} = e_1^T(X_{P,T_i}^T W_{T_i} X_{P,T_i})^{-1}(X_{P,T_i}^T W_{T_i} e_j) \tag{16}$$

Once again, this leads to another new bandwidth selection problem of λ. To obtain a rule for the choice of λ, we appeal to the results for the conditional MSE of $\hat{\sigma}_1^2(\lambda)$ (13). It follows that the asymptotically $AMSE$-optimal choice of λ is

$$\lambda_{AMSE} = C_3(K)[\sigma^4(b-1)/\theta_{22}^2 n^2]^{1/9}, \tag{17}$$

where, $C_3(K) = [4(1/2 + 2\sqrt{2} - 4/3\sqrt{3})/\sqrt{2\pi}]^{1/9}$, and $\theta_{22} = \hat{\theta}_{22}(g)$.

We are now in a position to return to the optimal Plug-in bandwidth Eq. (11), and replacing the unknown quantities by their corresponding estimators

$$\hat{h}_{MISE} = C_1(K)[\hat{\sigma}_1^2(\hat{\lambda}_{AMSE})(b-a)/\hat{\theta}_{22}(\hat{g}_{AMSE})n]^{1/5} \tag{18}$$

THE SIMULATION STUDY

The main purpose of the simulation study is to evaluate and compare the performance of four bandwidth selection strategies--Direct Plug-In (DPI) selection procedure, Cross-validation (CV) Criterion, the Rule of Thumb (RT) Criterion, and Residual Sum of Square Criteria (RSC) –for estimating the mean regression function $g(x)$. The simulation considers two experimental design setting: fixed (equally spaced) design and random design. Four regression functions with different shapes are considered, together with three sample sizes ($n = 50, 100,$ and 300) and two error distributions, normal and non-normal. Each simulation scenario is replicated $L = 500$ time. Performance is assessed using average bias, average variance, and root average mean squared error ($RAMSE$).

Smooth the Mean Regression Functions

This section aims to conduct a simulation study in order to evaluate and compare the estimated mean regression functions. These, estimated regression curves are resulted from utilizing the four bandwidth selection procedures: DPI, CV, RT, RSC selection criteria. Since the regression analysis can be applied to both the fixed (equally-spaced) design and the random design, in the next subsections, we deal with each design setting separately.

Description of the Experiment in the Fixed Design Setting

We have conducted 24 simulation studies, using four different example regression functions, and four selection methods of the bandwidth selection strategies (DPI, CV, RT, and RSC), and three different sample sizes n, and two different types of the errors' distribution. Each simulation study involves ($L = 500$) repetitions. For the error term, two distributions were considered: the standard normal distribution and a non-normal heavy-tailed distribution. Specifically, the non-normal errors were generated from a Student's t-distribution with 3 degrees of freedom ($\nu = 3$). This distribution has mean $E(\varepsilon) = 0$ and variance $Var(\varepsilon) = 3$ and was selected to represent errors with heavier tails and a higher probability of extreme observations than the normal distribution. In the simulation, the Student's t errors were generated in R using the command $err <- rt(n, 3)$.

We consider four different example regression function, and for each regression function, the error variance was set according to the corresponding value of σ as reported in Table 1. The shape of these functions exhibited several features such as concave, convex, and flat parts. We have also chosen three different sample sizes of $n = 50$, $n = 100$ and $n = 300$.

Table 1: *The Example Regression Functions used in the simulation study.*

Example	$g_i(x)$	Interval	Sigma
1	$g_1(x) = 0.3\, exp\,(-16\,(x - 0.25)^2) + 0.7\, exp\,(-64\,(x - 0.75)^2)$	(0,1)	0.15
2	$g_2(x) = 0.75\, sin\,(3\,\pi\, x/2 + 1.25)$	(0,1)	0.6
3	$g_3(x) = 1 + sin(x) + 2\, cos(x) + 3\, sin\,(5\, x)$	(-1,1)	1.2
4	$g_4\,(x) = sin\,(2\,\pi\, x3)^3$	(0,1)	0.5

Description of the Experiment in the Random Design Setting

We have conducted another 24 simulation studies, using the same four different example regression functions (as present in Table 1), and four selection methods of the bandwidth selection strategies (DPI, CV, RT, and RSC), and three different sample sizes n (50, 100, and 300), and two different types of the errors' distribution. Each simulation study involves ($L = 500$) repetitions. The same two error distributions described in Section 3.2, namely the standard normal distribution and the Student's t-distribution with 3 degrees of freedom ($\nu = 3$), were used in the random-design simulations.

RESULTS AND DISCUSSION

Numerical Summary for Fixed Design Setting

Having conducted the simulation runs using the (L=500) simulated data sets, we obtained the results that are tabulated in Tables 2-5. These tables display the numerical summary of results obtained using four different bandwidth selection procedures (DPI, CV, RT, and RSC). Moreover, two error distributions were considered: the standard normal distribution and the Student's t-distribution with 3 degrees of freedom ($\nu = 3$), representing non-normal heavy-tailed errors. The latter has mean zero and variance three with three different sample sizes of n=50, n=100, and n=300. The numerical summary includes three statistical properties, namely the Average Bias ($ABias$), the Average Variance ($AVar$), and the Square Root of Average Mean Square Errors ($RAMSE$). The formulae that we have been used to calculate $ABias, AVar$, and $RAMSE$ are as follows:

$$ABias(\hat{g}) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{L}\sum_{j=1}^{L}\hat{g}\left(x_{ij}\right) - g(x_i)\right] \qquad (19)$$

$$AVAR(\hat{g}) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{L}\sum_{j=1}^{L}\hat{g}^2\left(x_{ij}\right) - \left(\frac{1}{L}\sum_{j=1}^{L}\hat{g}\left(x_{ij}\right)\right)^2\right] \qquad (20)$$

$$RAMSE(\hat{g}) = \sqrt[2]{ABias(\hat{g})^2 + AVar(\hat{g})} \qquad (21)$$

The numerical summary of results obtained using four different bandwidth selection strategies, $DPI, CV, RT,$ and RSC criteria, from the simulation study using the first, the second, third, and fourth examples regression functions (fixed design setting) are given in Tables 2, 3, 4, and 5**.** In addition, the plots of true function and four other estimated functions using four different bandwidth selection strategies DPI, CV, RT, and RSC criteria from the simulation study using the four examples regression functions (fixed design setting) are showing by Figures. 1, 2, 3, and 4. **Note that,** Normal = standard normal errors; and Student's t (ν=3) = Student's t-distributed errors with 3 degrees of freedom.

Table 2. Summary of results from the simulation study using the first example regression function (Fixed Design Setting).

Metric	n	DPI Normal	DPI Non-N	CV Normal	CV Non-N	RT Normal	RT Non-N	RSC Normal	RSC Non-N
$ABias$	50	0.00095	0.00478	0.00126	0.00527	0.00132	0.01241	0.00071	0.00368
$ABias$	100	0.00059	0.00363	0.00087	0.00481	0.00076	0.01079	0.00049	0.00339
$ABias$	300	0.00026	0.00241	0.00036	0.00316	0.00027	0.00668	0.00022	0.00245
$Avar$	50	0.00602	0.17358	0.00708	0.17361	0.00806	0.18133	0.00649	0.19915
$Avar$	100	0.00361	0.09560	0.00509	0.09712	0.00481	0.10628	0.00372	0.10082
$Avar$	300	0.00137	0.03382	0.00220	0.03517	0.00150	0.04316	0.00132	0.03235
$ARMSE$	50	0.08351	0.42233	0.09136	0.42294	0.09687	0.44016	0.08483	0.45036
$ARMSE$	100	0.06478	0.31501	0.07721	0.31927	0.07462	0.34216	0.06494	0.32282
$ARMSE$	300	0.04034	0.19034	0.05062	0.19579	0.04209	0.22323	0.03928	0.18653

Table 3. *Summary of results from the simulation study using the second example regression function (Fixed Design Setting).*

Metric	n	DPI Normal	DPI Non-N	CV Normal	CV Non-N	RT Normal	RT Non-N	RSC Normal	RSC Non-N
$ABias$	50	0.00157	0.00130	0.00136	0.00115	0.00659	0.01012	0.00128	0.00144
$ABias$	100	0.00079	0.00156	0.00078	0.00172	0.00352	0.00827	0.00071	0.00145
$ABias$	300	0.00074	0.00103	0.00064	0.00102	0.00159	0.00400	0.00068	0.00094
$Avar$	50	0.05080	0.16893	0.04895	0.16692	0.07207	0.19084	0.05886	0.19815
$Avar$	100	0.02828	0.08992	0.02834	0.09210	0.04680	0.11296	0.03116	0.09760
$Avar$	300	0.01042	0.03198	0.00884	0.03179	0.02025	0.05103	0.01046	0.03145
$ARMSE$	50	0.22884	0.41259	0.22430	0.40996	0.28045	0.44828	0.24525	0.44675
$ARMSE$	100	0.17052	0.30246	0.17065	0.30630	0.22433	0.34818	0.17851	0.31471
$ARMSE$	300	0.10564	0.18169	0.09738	0.18115	0.14777	0.23459	0.10552	0.17997

Table 4. *Summary of results from the simulation study using the third example regression function (Fixed Design Setting).*

Metric	n	DPI Normal	DPI Non-N	CV Normal	CV Non-N	RT Normal	RT Non-N	RSC Normal	RSC Non-N
$ABias$	50	0.04678	0.04134	0.04921	0.04490	0.05090	0.04615	0.03394	0.03104
$ABias$	100	0.02276	0.02042	0.02586	0.02363	0.02575	0.02434	0.01970	0.01820
$ABias$	300	0.01142	0.01393	0.01224	0.01409	0.01075	0.01339	0.00972	0.01223
$Avar$	50	0.21474	0.24393	0.22630	0.26559	0.24600	0.28127	0.23942	0.26454
$Avar$	100	0.12619	0.13967	0.16057	0.17688	0.16223	0.18586	0.12859	0.14226
$Avar$	300	0.04915	0.05467	0.06371	0.05746	0.03802	0.04667	0.04935	0.05436
$ARMSE$	50	0.51139	0.53411	0.52489	0.55721	0.54489	0.57221	0.52283	0.54368
$ARMSE$	100	0.38594	0.40011	0.43177	0.44777	0.43357	0.45847	0.38508	0.40058
$ARMSE$	300	0.24610	0.26190	0.27559	0.26750	0.22083	0.24507	0.24304	0.25804

Table 5. *Summary of results from the simulation study using the fourth example regression function (Fixed Design Setting).*

Metric	n	DPI Normal	DPI Non-N	CV Normal	CV Non-N	RT Normal	RT Non-N	RSC Normal	RSC Non-N
$ABias$	50	0.01258	0.02394	0.02275	0.03172	0.01560	0.04600	0.01435	0.02349
$ABias$	100	0.00651	0.01625	0.01730	0.02814	0.00791	0.03027	0.00989	0.02172
$ABias$	300	0.00333	0.00915	0.01195	0.02170	0.00298	0.01402	0.00512	0.01744
$Avar$	50	0.03747	0.18232	0.05805	0.18982	0.04794	0.22076	0.04139	0.20809
$Avar$	100	0.02381	0.11754	0.05216	0.13476	0.03104	0.14815	0.02332	0.11324
$Avar$	300	0.00979	0.04746	0.03710	0.07437	0.00875	0.06295	0.00880	0.04022
$ARMSE$	50	0.22371	0.45416	0.28427	0.47068	0.25207	0.51648	0.23611	0.48123
$ARMSE$	100	0.17412	0.36576	0.26356	0.40361	0.19736	0.42240	0.18222	0.36737
$ARMSE$	300	0.11452	0.23794	0.22147	0.30995	0.10832	0.27744	0.11798	0.24013

Having a close look at Table 2-5, we have made several comparisons as following:

Regarding the Bias measure ($ABias$): All bandwidth selection criteria have shown a significant improvement in performance with increasing sample sizes. As the sample size increases, the numerical values of $ABias$ decrease, reflecting the increased consistency of the estimates, with an exception in $g_2(x)$ where there is an irregular decrease (2,3) . This emphasizes the importance of using sufficient sample sizes to achieve accurate and reliable results regardless the choice of the example regression functions (14,15).

The $g_1(x)$, $g_3(x)$, and $g_4(x)$ have shown a considerable reduction in $ABias$ estimated values as the error terms changes from non-normally to normally distributed regardless the choice of the bandwidth selection criteria. This emphasizes the importance of the normality assumption in order to achieve fewer bias results. There are two exceptions when $n = 50$ with CV and DPI bandwidth selection criteria for all example regression functions used.
It has been noticed that the RT and RSC bandwidth selection criteria have attained higher ABias terms, suggesting that they may be less accurate in estimating parameters under certain conditions. While the DPI and CV bandwidth selection criteria have provided smaller bias terms (9,17) . This have been seen in all example regression functions as well as in both normally distributed errors and non-normally distributed errors.

Regarding the Variance measure ($AVar$): Once again, all bandwidth selection criteria have shown a substantial decrease in $AVar$ estimated values as the sample sizes n increase regardless the choice of the example regression function used (2).

Generally, all bandwidth selection criteria have shown a considerable increase in $AVar$ estimated values as the error terms changes from normally to non-normally distributed the regardless the choice of the example regression function. This emphasizes again the importance of the normality assumption in order to achieve less variance results (15) .

Regarding the Root Means Squared Error measure ($ARMSE$): Regarding the Root Average Mean Squared Error measure ($RAMSE$), the results show that the magnitude of the error measure varied across the bandwidth selection criteria, regression functions, sample sizes, and error distributions. The DPI procedure generally produced comparatively small $RAMSE$ values across many of the investigated scenarios; however, it did not consistently achieve the smallest value in every case. In some scenarios, CV or RSC produced comparable or smaller $RAMSE$ values. Therefore, the relative performance of the bandwidth selection procedures should be interpreted according to the specific simulation conditions rather than through a universal ranking (9,13).

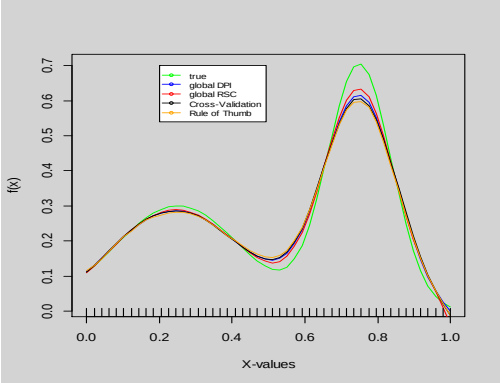
When comparing the normal and non-normal error distributions under the DPI bandwidth selection procedure, the results generally show smaller error measures under normally distributed errors across the investigated sample sizes and regression functions. This pattern indicates that the presence of heavy-tailed errors can increase the difficulty of accurately estimating the mean regression function. However, this observation should be interpreted within the simulation settings considered in the present study.

The non-normal distributed errors represent a greater challenge in accurately estimating parameters (14). This emphasizes the need to carefully choose the bandwidth selection criterion based on the characteristics of the data used.

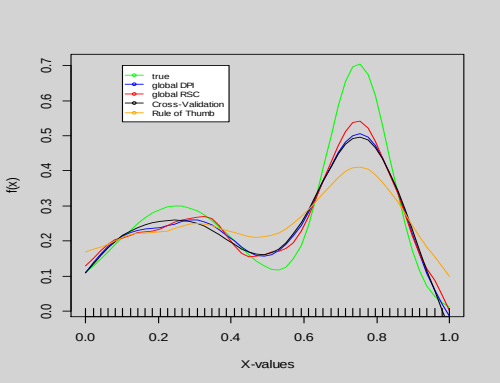
To sum up all of the above: To summarize the fixed-design results, the RT and RSC bandwidth selection procedures produced relatively larger bias values in several simulation scenarios, whereas DPI and CV generally produced comparatively smaller bias values. Nevertheless, the differences among the procedures varied according to the regression function, sample size, and error distribution. The variance measure generally increased under Student's t errors compared with normal errors, indicating the greater variability associated with heavy-tailed error distributions. Overall, the results suggest that bandwidth selection should be considered in relation to the characteristics of the particular estimation scenario.

In summary, the fixed-design results indicate that DPI performed competitively across a substantial proportion of the investigated scenarios and frequently produced relatively small bias and $RAMSE$ values. However, the results do not support a universal superiority of DPI over the other bandwidth selection procedures. The relative performance of DPI, CV, RT, and RSC depended on the regression-function shape, sample size, and error distribution. These findings emphasize the importance of evaluating bandwidth selection procedures under the specific characteristics of the data and estimation problem (9,13,17).

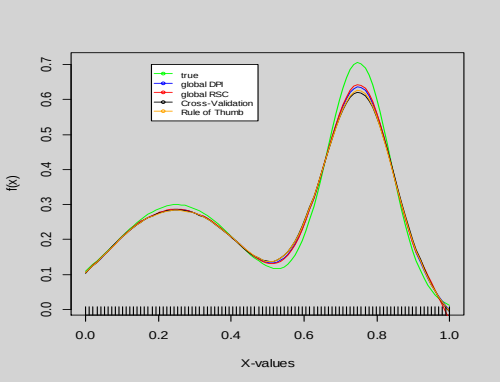
(a) Normal errors, $\boldsymbol{n = 50}$



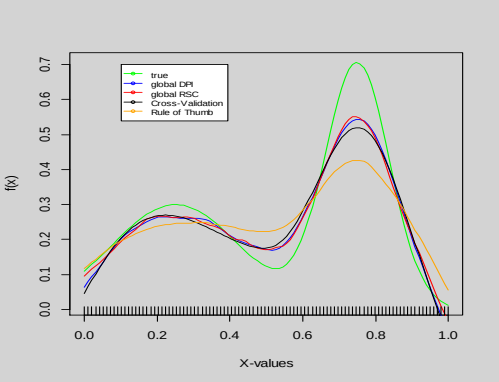
(b) Non-normal errors, $\boldsymbol{n = 50}$



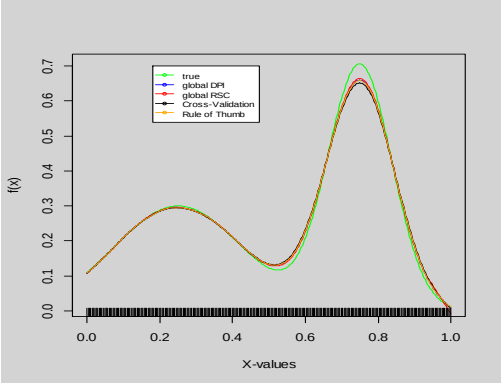
(c) Normal errors, $\boldsymbol{n = 100}$



(d Non-normal errors, $\boldsymbol{n = 100}$



(e) Normal errors, $\boldsymbol{n = 300}$



(f) Non-normal errors, $\boldsymbol{n = 300}$

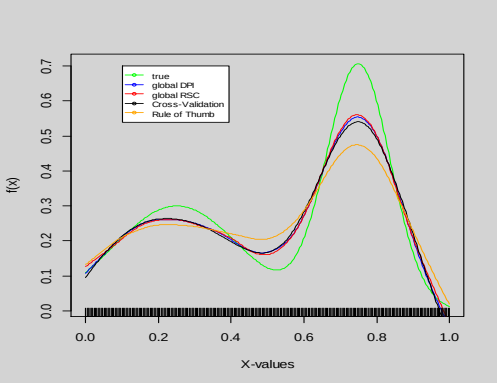
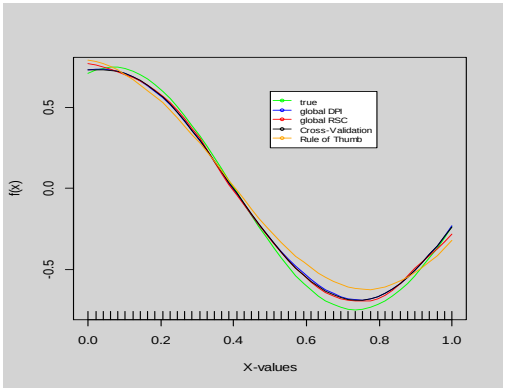
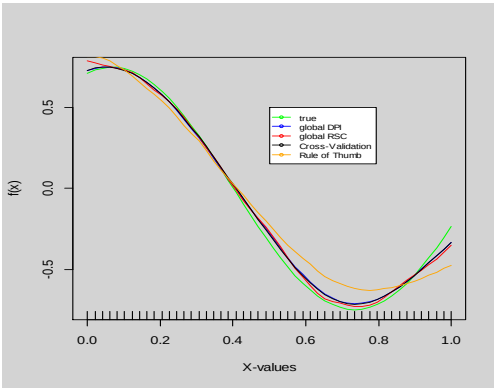


Figure 1. Estimated mean regression function (g(x)) under the fixed-design setting for the first example regression function. Panels (a)–(f) correspond to normal and Student's t (v = 3) error distributions for sample sizes (n=50), (100), and (300). The curves represent the true regression function and the estimates obtained using the DPI, RSC, CV, and RT bandwidth selection strategies.

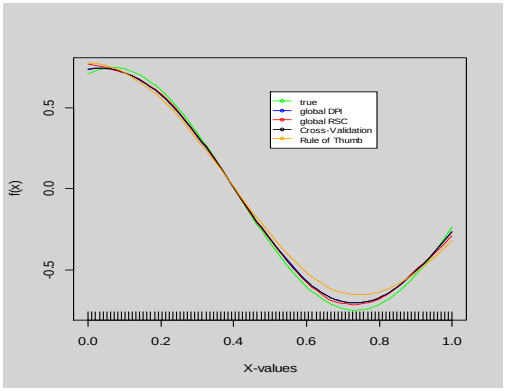
(a) Normal errors, $\boldsymbol{n = 50}$



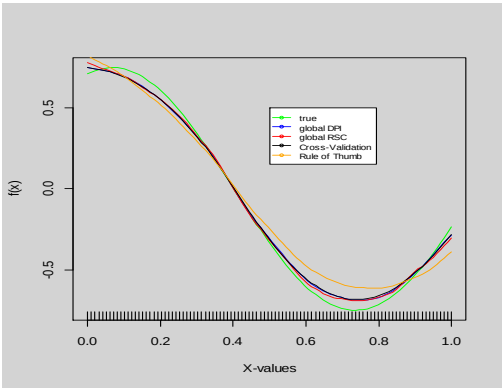
(b) Non-normal errors, $\boldsymbol{n = 50}$



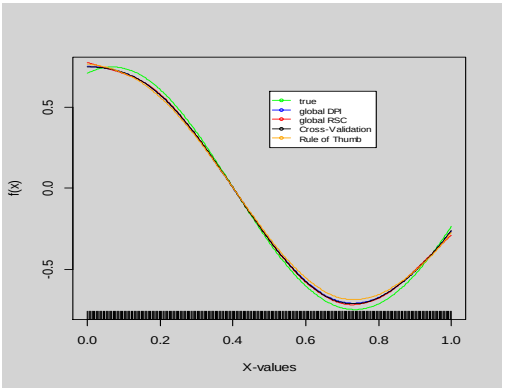
(c) Normal errors, $\boldsymbol{n = 100}$



(d) Non-normal errors, $\boldsymbol{n = 100}$



(e) Normal errors, $\boldsymbol{n = 300}$



(f) Non-normal errors, $\boldsymbol{n = 300}$

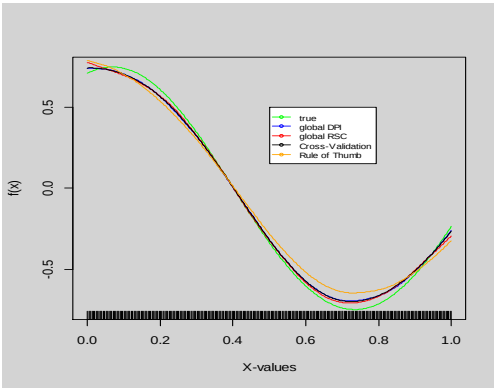
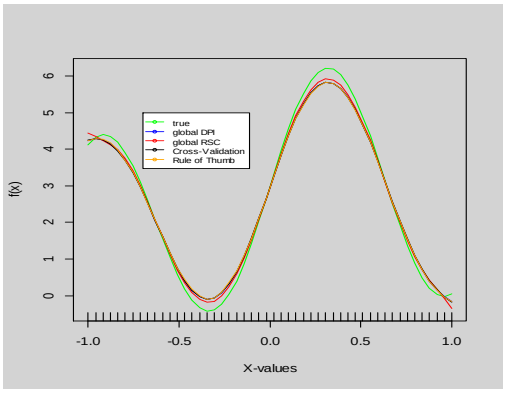
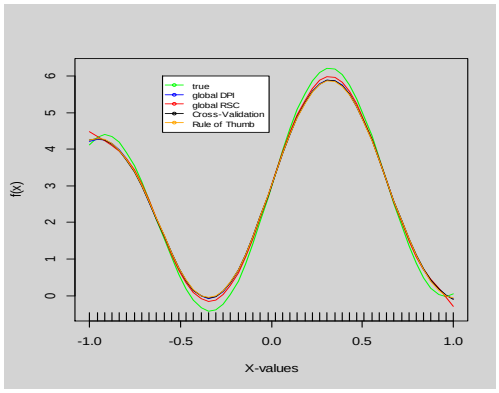


Figure 2. Estimated mean regression function (g(x)) under the second-design setting for the first example regression function. Panels (a)–(f) correspond to normal and Student's t (v = 3) error distributions for sample sizes (n=50), (100), and (300). The curves represent the true regression function and the estimates obtained using the DPI, RSC, CV, and RT bandwidth selection strategies.

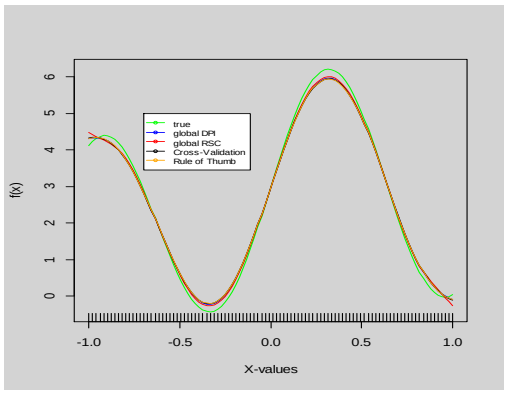
(a) Normal errors, $\boldsymbol{n = 50}$



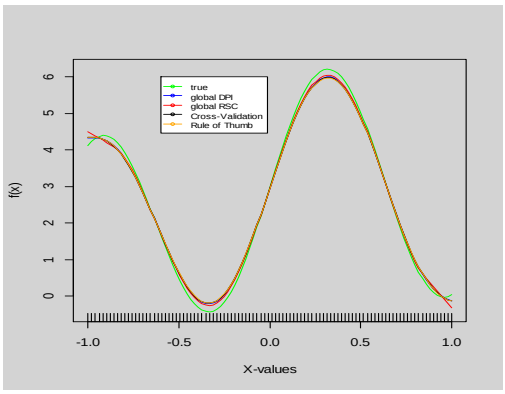
(b) Non-normal errors, $\boldsymbol{n = 50}$



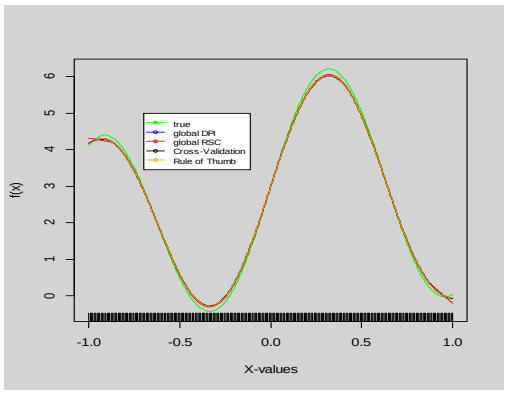
(c) Normal errors, $\boldsymbol{n = 100}$



(d) Non-normal errors, $\boldsymbol{n = 100}$



(e) Normal errors, $\boldsymbol{n = 300}$



(f) Non-normal errors, $\boldsymbol{n = 300}$

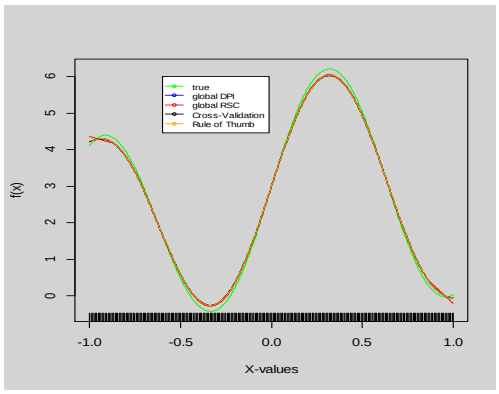
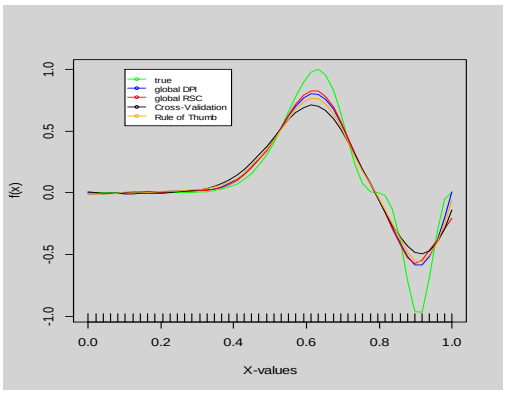
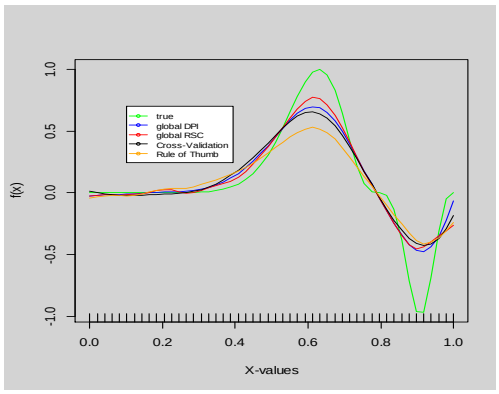


Figure 3. Estimated mean regression function (g(x)) under the third-design setting for the first example regression function. Panels (a)–(f) correspond to normal and Student's t (v = 3) error distributions for sample sizes (n=50), (100), and (300). The curves represent the true regression function and the estimates obtained using the DPI, RSC, CV, and RT bandwidth selection strategies.

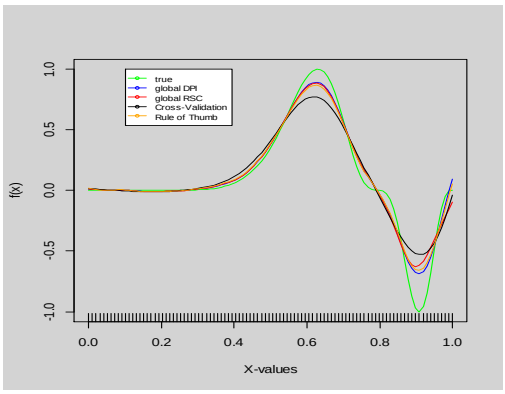
(a) Normal errors, $\boldsymbol{n = 50}$



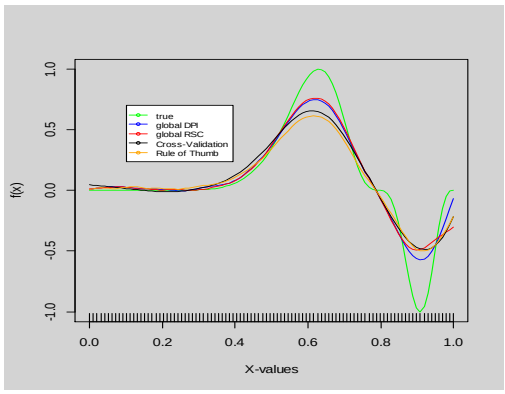
(b) Non-normal errors, $\boldsymbol{n = 50}$



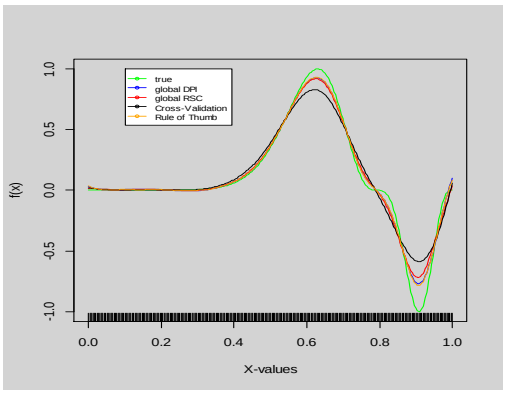
(c) Normal errors, $\boldsymbol{n = 100}$



(d) Non-normal errors, $\boldsymbol{n = 100}$



(e) Normal errors, $\boldsymbol{n = 300}$



(f) Non-normal errors, $\boldsymbol{n = 300}$

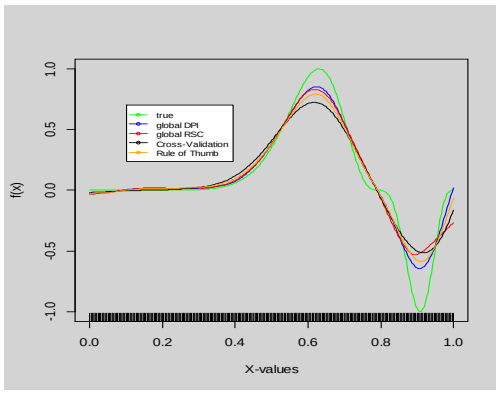


Figure 4. Estimated mean regression function (g(x)) under the fixed-design setting for the fourth example regression function. Panels (a)–(f) correspond to normal and Student's t (v = 3) error distributions for sample sizes (n=50), (100), and (300). The curves represent the true regression function and the estimates obtained using the DPI, RSC, CV, and RT bandwidth selection strategies.

Figures. 1- 4 plot the true curve and the four other estimated curves; each describes the average curve results from 500 runs. The estimated curves are obtained using four different bandwidth selection procedures (DPI, Rule of Thumb, Cross-validation, and RSC) in two different error-type distributions (normally and non-normally distributed errors). The errors are normally distributed (left-hand side panels in Figures. 1-4) and non-normally distributed (right-hand side panels in Figures. 1-4). Three different sample sizes ($n = 50,\ 100,$ and 300) and four example regression functions $g_1(x), g_2(x), g_3(x)$, and $g_4(x)$, respectively.

By examining Figures. 1-4 closely, we have noticed the following points: First, the estimated curves obtained using the different bandwidth selection procedures showed noticeable differences across sample sizes. In several scenarios with $(n = 50)$, the curves obtained using DPI followed the true regression functions relatively closely, whereas at $(n = 300)$, the RSC procedure also produced curves that closely followed the true functions in several scenarios. These observations indicate that the graphical performance of the bandwidth selection procedures varies with sample size and regression-function shape. Second, the estimated curves obtained using the normally distributed errors result in optimum curves that succeed in capturing almost all the features of the true curves compared with those obtained using the non-normally distributed errors. Third, the CV and RT procedures also produced reasonable estimates in several scenarios, although greater variation was observed across the regression functions and error distributions. The differences among DPI, RSC, CV, and RT were more noticeable under Student's t errors, indicating that the effect of the bandwidth selection procedure depends on the underlying simulation conditions.

Numerical Summary for Random Design Setting

Again, the numerical summary of results obtained using four different bandwidth selection strategies, DPI, CV, RT, and RSC criteria, from the simulation study using the first, second, third, and fourth examples regression functions (random design setting) are given in Tables 6, 7, 8, and 9**.** Moreover, the plots of true function and four other estimated functions using four different bandwidth selection strategies DPI, CV, RT, and RSC criteria from the simulation study using the four examples regression functions (random design setting) are showing by Figures. 5, 6, 7, and 8. **Note that,** Normal = standard normal errors; and Student's t (ν=3) = Student's t-distributed errors with 3 degrees of freedom.

Table 6. *Summary of results from the simulation study using the first example regression function (Random Design Setting).*

Metric	n	DPI Normal	DPI Non-N	CV Normal	CV Non-N	RT Normal	RT Non-N	RSC Normal	RSC Non-N
$ABias$	50	0.00230	0.00435	0.00238	0.00463	0.00229	0.01118	0.01098	0.01348
$ABias$	100	0.00135	0.00521	0.00138	0.00585	0.00135	0.01407	0.00092	0.00654
$ABias$	300	$5.411e^{-0.5}$	0.00319	0.00017	0.00438	$4.673e^{-0.5}$	0.00923	0.00104	0.00445
$Avar$	50	0.00052	0.30916	0.00174	0.30899	0.00049	0.31705	0.00066	0.47039
$Avar$	100	0.00037	0.16771	0.00182	0.16783	0.00069	0.18367	0.00038	0.21662
$Avar$	300	0.00013	0.05882	0.00204	0.06076	$3.720e^{-0.5}$	0.07039	0.00019	0.06560
$ARMSE$	50	0.05312	0.55992	0.06418	0.56001	0.05270	0.57292	0.10786	0.6956
$ARMSE$	100	0.04139	0.41584	0.05661	0.41675	0.04513	0.44468	0.03605	0.47240
$ARMSE$	300	0.01361	0.24902	0.04696	0.25523	0.00309	0.28218	0.03510	0.26468

Table 7. *Summary of results from the simulation study using the second example regression function (Random Design Setting).*

Metric	n	DPI Normal	DPI Non-N	CV Normal	CV Non-N	RT Normal	RT Non-N	RSC Normal	RSC Non-N
$ABias$	50	0.00293	0.00263	0.00281	0.00258	0.00859	0.01260	0.02917	0.02006
$ABias$	100	0.00174	0.00226	0.00160	0.00263	0.00456	0.01118	0.02191	0.01410
$ABias$	300	0.00280	0.00153	0.00050	0.00168	0.00109	0.00656	0.00280	0.02251
$Avar$	50	0.04255	0.27837	0.03997	0.27879	0.06866	0.30718	0.04712	0.42518
$Avar$	100	0.02433	0.18782	0.02350	0.19120	0.03530	0.19763	0.02782	0.22123
$Avar$	300	0.00767	0.06452	0.00634	0.06601	0.01496	0.81315	0.00901	0.77110
$ARMSE$	50	0.21328	0.53009	0.20684	0.53044	0.27794	0.56548	0.27619	0.66727
$ARMSE$	100	0.16147	0.43599	0.15844	0.44027	0.19966	0.45696	0.22302	0.48511
$ARMSE$	300	0.09081	0.25700	0.08275	0.26017	0.12670	0.29644	0.10868	0.31563

Table 8. *Summary of results from the simulation study using the third example regression function (Random Design Setting).*

Metric	n	DPI Normal	DPI Non-N	CV Normal	CV Non-N	RT Normal	RT Non-N	RSC Normal	RSC Non-N
$ABias$	50	0.47093	1.15522	0.468	1.17104	0.48748	1.15584	1.11991	1.12075
$ABias$	100	0.03659	0.09493	0.03690	0.09654	0.04324	0.10349	0.28929	0.59055
$ABias$	300	0.01427	0.02674	0.01369	0.02448	0.01389	0.02535	0.64055	0.25292
$Avar$	50	0.25215	0.36940	0.08850	0.10527	0.53943	0.49240	0.27008	0.49849
$Avar$	100	0.13772	0.23407	0.13327	0.23418	0.23272	0.33402	0.15341	0.27185
$Avar$	300	0.04989	0.09778	0.04016	0.06738	0.04587	0.08341	0.05544	0.10876
$ARMSE$	50	0.85034	1.23476	0.74599	1.12974	1.01337	1.28384	1.17898	1.27249
$ARMSE$	100	0.41751	0.57359	0.41252	0.57509	0.52532	0.66145	0.66536	0.92865
$ARMSE$	300	0.25331	0.35287	0.23204	0.30309	0.24447	0.32978	0.93426	0.60140

Table 9. *Summary of results from the simulation study using the fourth example regression function (Random Design Setting).*

Metric	n	DPI Normal	DPI Non-N	CV Normal	CV Non-N	RT Normal	RT Non-N	RSC Normal	RSC Non-N
$ABias$	50	0.01430	0.01940	0.02324	0.02960	0.01331	0.05418	0.13371	0.05104
$ABias$	100	0.00764	0.02897	0.01578	0.03766	0.00664	0.04852	0.02323	0.03672
$ABias$	300	0.00251	0.01630	0.01057	0.02829	0.00233	0.02502	0.00513	0.02608
$Avar$	50	0.01518	0.36562	0.03342	0.38072	0.01738	0.42460	0.01930	0.43164
$Avar$	100	0.00942	0.18612	0.03065	0.19376	0.00759	0.21811	0.01130	0.24864
$Avar$	300	0.00433	0.07764	0.03481	0.09416	0.00328	0.09830	0.00462	0.07597
$ARMSE$	50	0.17169	0.62050	0.23804	0.64056	0.17518	0.69194	0.39117	0.69475
$ARMSE$	100	0.13063	0.46377	0.21549	0.48106	0.11931	0.51637	0.18582	0.53419
$ARMSE$	300	0.08272	0.30651	0.21303	0.34993	0.07493	0.35118	0.09870	0.31945

Having a close look at Tables 6-9, that summarize the results of using several smoothing parameter selection criteria (DPI, CV, RT, and RSC) across different example regression functions in Random Design setting: we have made several comparisons as following:

Regarding the Bias measure ($ABias$)

Sample size: As expected, all bandwidth selection criteria showed significant performance improvement as sample sizes increased. As the sample size increases, the numerical values of $ABias$ decrease, reflecting increased consistency of the estimates. This emphasizes the importance of using sufficient sample sizes to achieve accurate and reliable results regardless of the choice of regression function.

Assumption of normality: A significant reduction in the estimated $ABias$ values is observed when the error terms are transformed from non-normal distribution to normal distribution in the example regression functions $g_1(x), g_3(x)$, and $g_4(x)$. This reinforces the importance of assuming normality to achieve less biased results. There are exceptions when $n = 50$ with the CV and DPI bandwidth selection criteria for all example regression functions used.

Criteria Performance: The RT and RSC bandwidth selection criteria showed higher $ABias$ values, indicating that they may be less accurate in estimating parameters under certain conditions. The DPI and CV bandwidth selection procedures generally produced comparatively smaller $ABias$ values in a number of the investigated scenarios. However, the magnitude of the differences varied across regression functions, sample sizes, and error distributions. Therefore, the observed bias results do not indicate that one procedure consistently provides the smallest bias under all random-design conditions.

Overall, the random-design results indicate that DPI and CV performed competitively with respect to bias in several simulation scenarios. At the same time, differences among the procedures were observed across regression functions, sample sizes, and error distributions. These findings support a scenario-specific comparison of bandwidth selection procedures rather than a universal preference for a particular criterion.

Regarding the Bias measure ($AVar$)

Sample size: As expected, all bandwidth selection criteria showed significant decreases in $AVar$ as sample sizes increased. This suggests that the precision of the estimates increases with increasing sample size regardless of the example regression function used. Using larger sample sizes helps reduce variability and increase the stability of results.

Assumption of normality: In general, all bandwidth selection criteria showed a significant increase in the variance values ($AVar$) when the error terms were transformed from a normal distribution to a non-normal distribution. This again reinforces the importance of assuming normality to achieve results with lower variance. A normal error distribution facilitates fewer variable estimates, while a non-normal distribution increases the challenges.

Criteria Performance: Different criteria for bandwidth selection showed differences in performance. The DPI and CV procedures showed comparatively stable variance estimates in several of the investigated scenarios, although their performance varied across regression functions and sample sizes. In contrast, RT and RSC showed larger variance values in some scenarios, particularly under Student's t errors. These differences suggest that the stability of a bandwidth selection procedure can depend on the characteristics of the regression function and the error distribution (19).

RT and RSC parameters showed greater variability, suggesting that they may be less stable in some circumstances.

In summary, Increasing the sample size reduces the variance ($AVar$), which increases the accuracy of the estimates; the normal distribution of error terms reduces the variance, which enhances the accuracy of the results; the DPI and CV benchmarks showed better performance in terms of contrast, which indicates that they are more stable than other bandwidth selection criteria such as RT and RSC.

Regarding the Bias measure ($ARMSE$)

Sample size: All bandwidth selection criteria showed clear improvement in $ARMSE$ values with increasing sample size. This is reflected in the small values of the $ARMSE$ indicating that the precision of the estimates increases as the sample size increases. This pattern is observed across all example regression functions and across both normal and non-normal distributed error terms.

Assumption of normality: The $RAMSE$ results show differences among the bandwidth selection procedures across the investigated simulation scenarios. DPI produced comparatively small $RAMSE$ values in several cases, but it did not consistently yield the smallest value across all regression functions, sample sizes, and error distributions. CV, RT, or RSC occasionally produced comparable or smaller $RAMSE$ values. Consequently, the performance of the bandwidth selection procedures should be assessed with respect to the specific simulation scenario rather than assuming a consistent ordering among the methods.

Criteria Performance: The bandwidth selection procedures showed different $RAMSE$ patterns across the simulation scenarios. DPI generally produced relatively small $RAMSE$ values in a number of cases, while CV, RT, and RSC also performed competitively under particular regression functions or sample sizes. The results therefore indicate that the relative performance of the procedures is scenario-dependent, particularly when the error distribution changes from normal to Student's t.

In summary, the RAMSE results indicate that DPI performed competitively across many of the investigated scenarios and frequently produced relatively small error values. However, the smallest RAMSE was not obtained by DPI in every scenario. The performance of all four procedures was affected by sample size, regression-function shape, and error distribution, with Student's t errors generally associated with larger error measures. These findings emphasize the importance of evaluating bandwidth selection procedures under the specific characteristics of the estimation problem. To summarize all of the above.

The Bias measure ($ABias$)

Sample size: As sample sizes increase, the bias values ($ABias$) for all bandwidth selection criteria decrease, reflecting increased consistency of the estimates.

Assuming normality: Switching to normal distributed error terms reduces the bias values in the example regression functions $g_1(x), g_3(x)$, and $g_4(x)$. There are exceptions at $n = 50$ with the CV and DPI criteria.

Criteria performance: The DPI and CV bandwidth selection procedures generally showed comparatively lower bias values in several of the investigated scenarios, whereas RT and RSC produced relatively higher bias values in some cases. Nevertheless, the magnitude and direction of these differences varied across the regression functions, sample sizes, and error distributions. Thus, the results do not imply a uniform performance ordering among the four procedures.

The Variance measure ($AVar$)

Sample size: All bandwidth selection criteria showed a significant decrease in variance as sample size increased.

Assuming normality: The normal distributed error terms reduces the variance values ($AVar$), which reinforces the importance of assuming normality to achieve fewer variable results.

The Root Mean Square Error measure ($ARMSE$)

DPI Performance: The DPI procedure produced relatively small RAMSE values in several of the investigated scenarios. However, comparisons across the simulation settings show that CV, RT, or RSC could also produce smaller RAMSE values in particular cases. Therefore, the results indicate competitive performance of DPI rather than consistent superiority over the other bandwidth selection procedures.

Normality assumption: The DPI criterion prefers normally distributed error terms to achieve the lowest $ARMSE$ values.

Challenges with non-normal distributed errors: non-normal distributed errors make it more difficult to achieve accurate estimates, which reinforces the importance of choosing criteria based on the characteristics of the data.

Overall, the results indicate that DPI is a competitive bandwidth selection procedure for estimating the mean regression function. Its performance, however, varies according to the regression-function shape, sample size, error distribution, and design setting. The findings therefore support comparing the competing procedures within the specific characteristics of each estimation scenario.

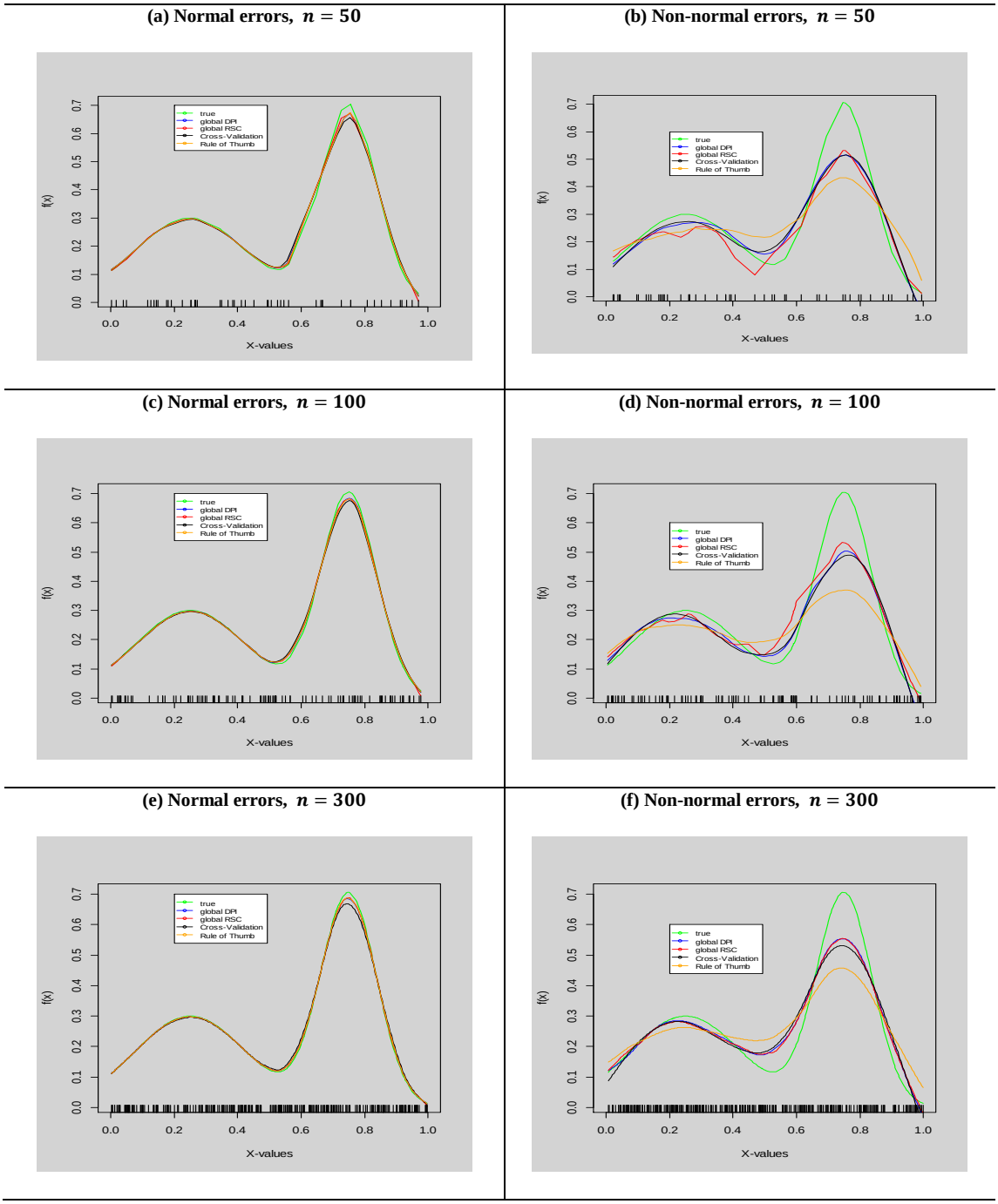


Figure 5. Estimated mean regression function (g(x)) under the random-design setting for the first example regression function. Panels (a)–(f) correspond to normal and Student's t (v = 3) error distributions for sample sizes (n=50), (100), and (300). The curves represent the true regression function and the estimates obtained using the DPI, RSC, CV, and RT bandwidth selection strategies.

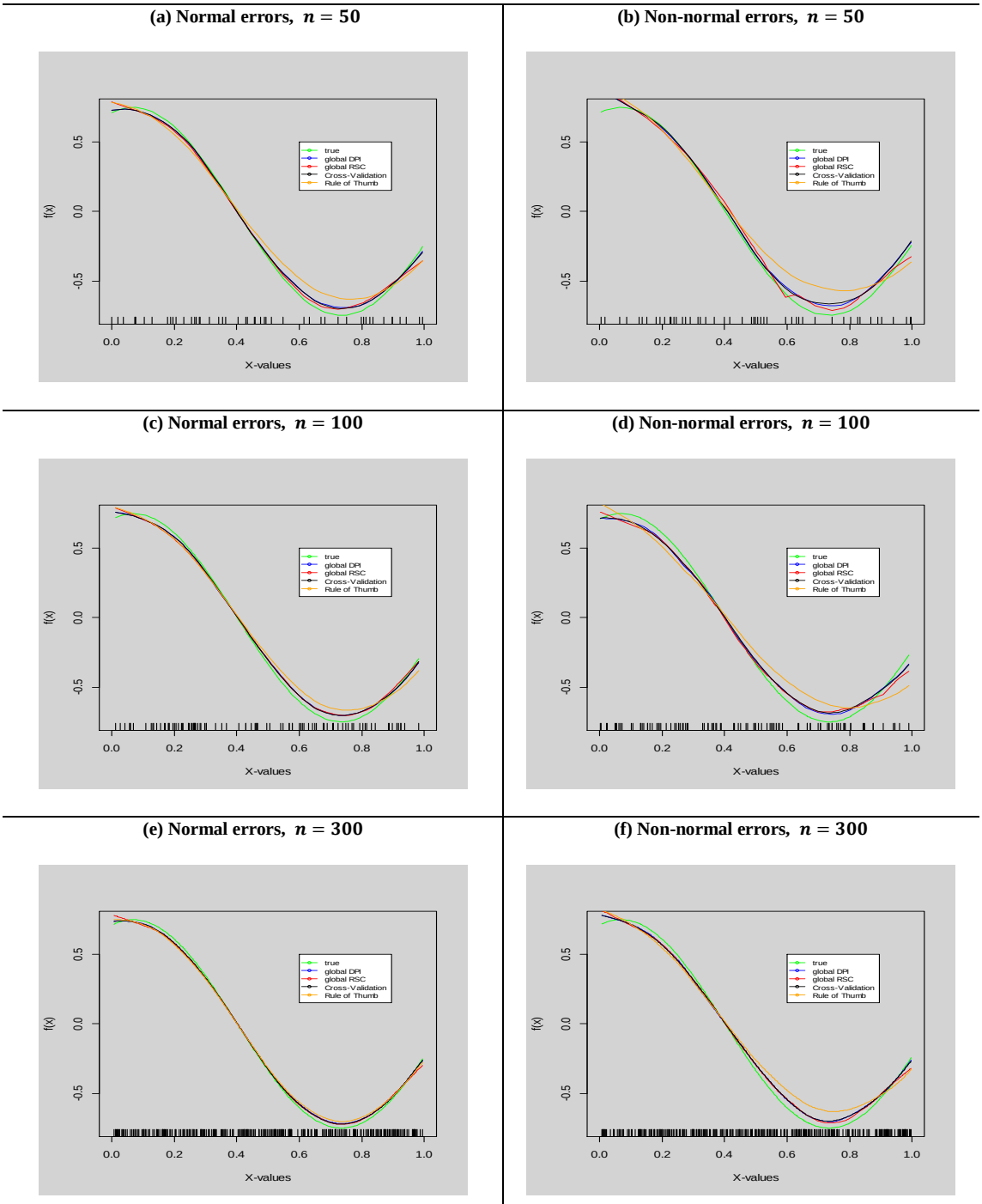


Figure 6. Estimated mean regression function (g(x)) under the random-design setting for the second example regression function. Panels (a)–(f) correspond to normal and Student's t (v = 3) error distributions for sample sizes (n=50), (100), and (300). The curves represent the true regression function and the estimates obtained using the DPI, RSC, CV, and RT bandwidth selection strategies.

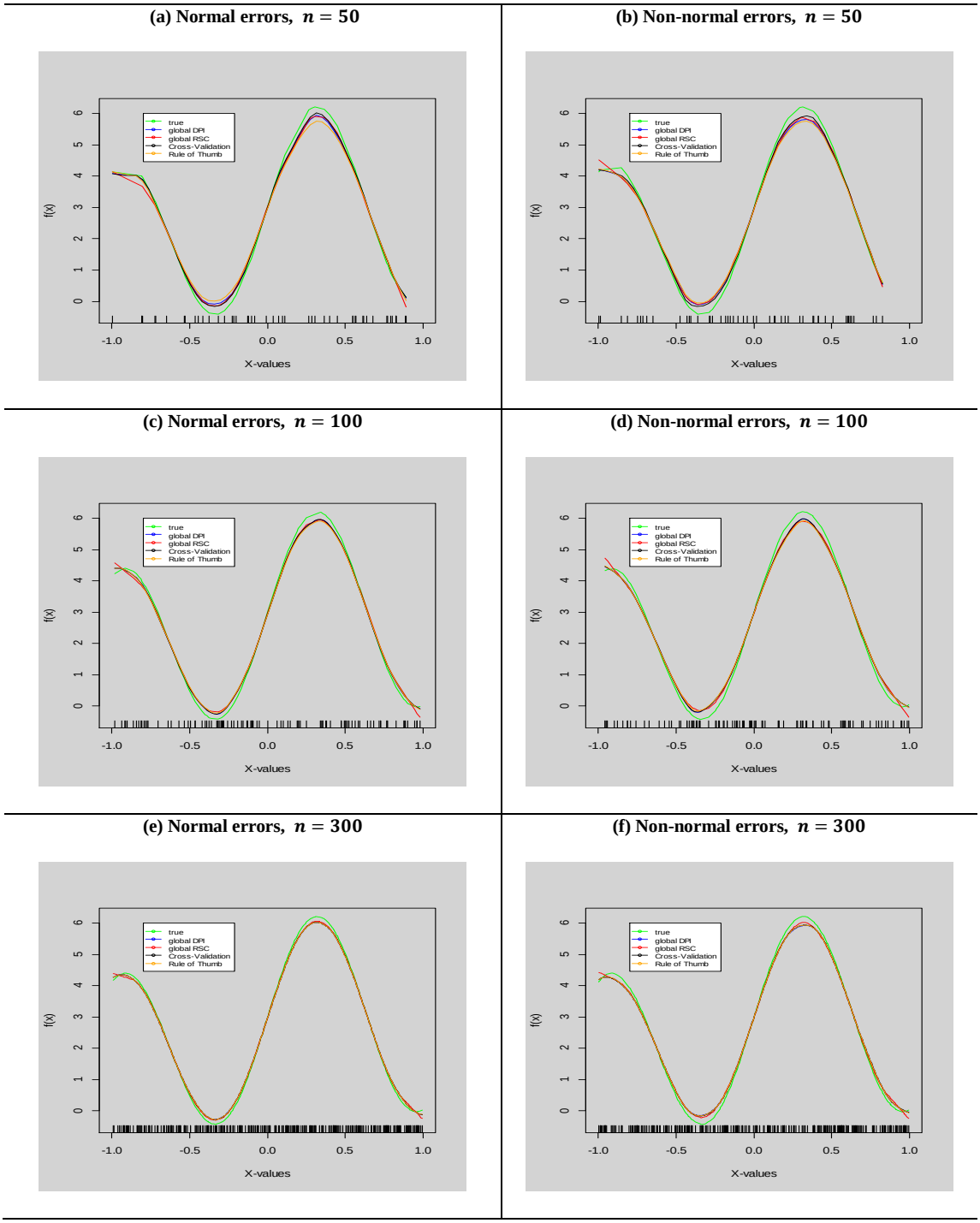


Figure 7. Estimated mean regression function (g(x)) under the random-design setting for the fithirdrst example regression function. Panels (a)–(f) correspond to normal and Student's t (v = 3) error distributions for sample sizes (n=50), (100), and (300). The curves represent the true regression function and the estimates obtained using the DPI, RSC, CV, and RT bandwidth selection strategies.

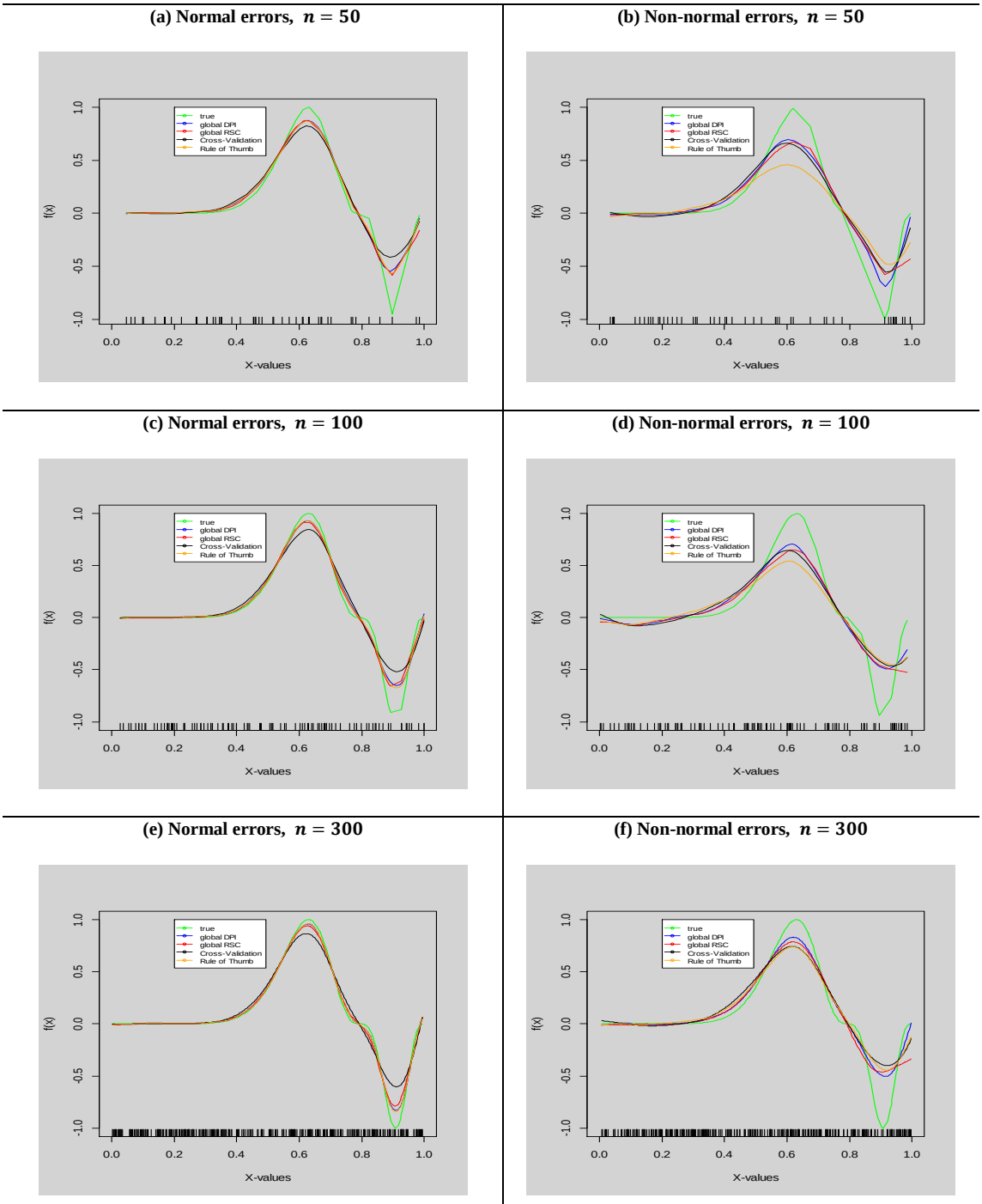


Figure 8. Estimated mean regression function (g(x)) under the random-design setting for the fourth example regression function. Panels (a)–(f) correspond to normal and Student's t (v = 3) error distributions for sample sizes (n=50), (100), and (300). The curves represent the true regression function and the estimates obtained using the DPI, RSC, CV, and RT bandwidth selection strategies.

Figures. 5-8 plot the true curve and the four other estimated curves; each describes the average curve results from 500 runs. The estimated curves are obtained using four different bandwidth selection procedures (DPI, Rule of Thumb, Cross validation, and RSC) in two different error types distributions (the normally and non-normally distributed errors). The errors' distributed once as normal (left hand side panels in Figures. 5-8) and the errors' non-normally distributed (right hand side panels in Figures, 5-8). Three different sample sizes (50, 100, and 300) and four example regression functions $g_1(x)$, $g_2(x)$, $g_3(x)$, and $g_4(x)$ respectively.

By examining Figures. 5-8 closely, we have noticed the following points: First, all estimated curves have shown significant performance improvement as sample sizes increased. Second, the estimated curves obtained using the normally distributed errors result in optimum curves that succeed in capturing almost all the features of the true curves comparing with those obtain using the non-normally distributed errors. Third, the estimated curves obtained using (DPI) selection strategy result in optimum curves almost in all examples regression functions whether the errors' distribution are either normal or non-normal, followed by the estimated curves result using (RSC) bandwidth selection criterion in almost all examples regression functions under study whether the errors' distributed as normal or non-normal. Lastly, other bandwidth selection criteria (CV, and RT) have performed quite well, the graphical panels have shown greater performance variation compared to DPI, and RSC especially in cases of non-normal distributed error terms.

CONCLUSION

This study evaluated four bandwidth selection procedures (Direct Plug-In (DPI), Residual Sum of Squares Criterion (RSC), Rule-of-Thumb (RT), and Cross-validation (CV) Criterion) for smoothing the mean regression function under fixed and Random design setting. The simulation study considered four regression function shapes, three sample sizes (50, 100, and 300), and two error distributions. The discussion is based on the numerical statistical aspect obtained from the simulation study.

Discussion of the Simulation Results

The purpose of this section is to summarize the simulation and the differences between the main results obtained from the study, also to make a statement the effect of the sample size (n), errors' distribution either as normal or non-normal, and bandwidth selection setting of either fixed or random design using four well known bandwidth selection strategies namely (Direct Plug-In (DPI), Residual Sum of Squares Criterion (RSC), Rule-of-Thumb (RT), and Cross-validation (CV) Criterion).

Smoothing the Mean Regression Functions: The first simulation study, conducted under both fixed and random design settings, provides clear insights into the performance of different bandwidth selection criteria (DPI, CV, RT, and RSC) in estimating mean regression functions. The Key finding are as follows:

Effect of Sample Size: Increasing sample size consistently improves accuracy by reducing bias ($ABias$), variance ($AVar$), and root mean squared error ($ARMSE$) across all methods. Larger samples produce more stable and reliable estimates.

Effect of Error Distribution: Assuming normality in error terms significantly reduces bias and variance, resulting in more accurate estimates. Non-normal error distributions create greater challenges for all criteria, highlighting the importance of distributional assumptions.

Performance of Selection Criteria: The simulation results show that the relative performance of the four bandwidth selection procedures varies across the investigated conditions. DPI produced comparatively small bias and $RAMSE$ values in many scenarios and performed competitively overall. However, it did not consistently produce the smallest bias, variance, or $RAMSE$ in every scenario. CV, RT, and RSC also produced competitive results under particular combinations of regression-function shape, sample size, error distribution, and design setting. Therefore, the results do not support a universal ranking of the four bandwidth selection procedures (20-23).

In summary, the simulation study demonstrated that sample size and error distribution play a crucial role in the accuracy of smoothing mean regression functions. Larger sample sizes consistently reduced bias, variance, and error, while assuming normality in error terms improved estimation stability. Across the simulation scenarios considered in this study, the DPI procedure generally produced comparatively small bias and $RAMSE$ values. However, its performance varied according to the regression-function shape, sample size, error distribution, and design setting, and it did not achieve the smallest error measure in every scenario. CV and RSC also performed competitively in several settings, while the performance of RT varied across the investigated conditions. These findings indicate that bandwidth

selection should be evaluated according to the characteristics of the data and the specific estimation problem rather than through a universal ranking of the competing procedures (24).

Recommendations

Based on this study, future research may extend the present simulation framework to bandwidth selection for estimating derivatives of the mean regression function, including settings involving correlated errors (25) . Moreover, Future studies could also consider bandwidth selection for estimating the regression function when the correlation structure of the errors is unknown (26-29) .

REFERENCES

1. Gasser, T., & Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. Scandinavian Journal of Statistics, 171–185.
2. Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2017). The Elements of Statistical Learning Data Mining, Inference, and Prediction (2 (ed.)). Springer.
3. James, G., Witten, D., Hastie, T., & T. (2023). An introduction to statistical learning with applications in R (2nd ed.). Springer.
4. Härdle, W. (1990). Applied nonparametric regression (Issue 19). Cambridge university press.
5. Fan, J. (2018). Local polynomial modelling and its applications: monographs on statistics and applied probability 66. Routledge. https://doi.org/10.1201/9780203748725
6. Yousuf, A., Jamal Saadoon , A., & Said Khan , F. (2026). Phytochemical Composition and Heavy Metals Analysis of Crude Ethanolic Extract of Geranium wallichianum, Elaeagnus parvifolia, and Taraxacum officinale. Libyan Journal of Applied and Contemporary Sciences, 1(1), 30–37. https://doi.org/10.64943/ljacs.2026.010104
7. Mami, A. M. (2002). Local polynomial regression with applications to both independent and longitudinal data. University of Manchester.
8. Chacón, J. E., & Duong, T. (2018). Multivariate kernel smoothing and its applications. Chapman and Hall/CRC.
9. Guidoum, A. C. (2015). Kernel estimator and bandwidth selection for density and its derivatives. Department of Probabilities and Statistics, University of Science and Technology, Houari Boumediene, Algeria.
10. 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.
11. Liu, S., & Charnigo, R. (2023). Tuning parameter selection for nonparametric derivative estimation in random design. Statistics, 57(6), 1402–1425. https://doi.org/10.1080/02331888.2023.2278042
12. Loader, C. R. (1999). BANDWIDTH SELECTION: CLASSICAL OR PLUG-IN? The Annals of Statistics, 27(2), 415–438.
13. Ruppert, D., Sheather, S. J., & Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. Journal of the American Statistical Association, 90(432), 1257–1270.
14. Henderson, D. J., & Parmeter, C. F. (2015). Applied nonparametric econometrics. Cambridge University Press.
15. Li, Q., & Racine, J. S. (2007). Nonparametric econometrics: theory and practice. Princeton University Press.
16. 16. Mallows, C. L. (1995). More comments on Cp. Technometrics, 37(4), 362–372.
17. Park, B. U., & Marron, J. S. (1990). Comparison of data-driven bandwidth selectors. Journal of the American Statistical Association, 85(409), 66–72.
18. Hall, P., & Marron, J. S. (1990). On variance estimation in nonparametric regression. Biometrika, 415–419.
19. Zhou, J., & Parmeter, C. F. (2025). Data-driven resistant kernel regression. Journal of Nonparametric Statistics, 37(1), 33–59./10.1080 https://doi.org/10485252.2024.2335494

20. Hossen Mohammed Al-Mahdi Al-Sharif. (2025). Analysis and Evaluation of ARIMA and SARIMA Models Performance in Time Series Forecasting: An Applied Study. Libyan Journal of Medical and Applied Sciences, 3(2), 146–157. https://doi.org/10.64943/ljmas.v3i2.89
21. Fowzi S. Alarabi. (2025). Investigation of Scattered Power and Cross-Polarization Effects on Linearly Polarized Waves under Sandstorm Conditions. Libyan Journal of Medical and Applied Sciences, 3(3), 81–90. https://doi.org/10.64943/ljmas.v3i3.140
22. García-Portugués, E., & Prieto-Alaiz, M. (2022).A review and comparison of bandwidth selection criteria for kernel regression and local polynomial estimators. Statistical Science, 37(4), 582–601. https://doi.org/10.1214/22-STS856
23. Abdelsalam Milad, F. (2026). Optimization of Laser-Induced Optical Response in Advanced Materials for High-Performance Photonic Applications. Libyan Journal of Applied and Contemporary Sciences, 1(2), 70–89. https://doi.org/10.64943/ljacs.2026.010207
24. Zhang, Y., & Chen, X. (2022). A comparative study of bandwidth selection criteria in kernel regression under fixed and random designs. Communications in Statistics - Simulation and Computation, 51(8), 4520–4539. https://doi.org/10.1080/03610918.2020.1741623
25. Kong, D., Zhao, S., & Wang, W. (2024). Nonparametric derivative estimation with bimodal kernels under correlated errors: D. Kong et al. Computational Statistics, 39(4), 1847–1865. https://doi.org/10.1007/s00180-023-01419-4
26. Sisheng, L., & Jing, Y. (2024). Kernel regression for estimating regression function and its derivatives with unknown error correlations: L. Sisheng, Y. Jin. Metrika, 87(1), 1–20. https://doi.org/10.1007/s00180-023-01419-4
27. Najeeb Ahmed , O., & Amer Mohammed, A. (2026). Intelligent Deep Learning-based Systems to Detect Early Cyber Attacks in IoT Networks. Libyan Journal of Applied and Contemporary Sciences, 1(1), 38–52. https://doi.org/10.64943/ljacs.2026.010105
28. Barreiro-Ures, D., Cao, R., & Francisco-Fernández, M. (2023).Subsampling-based cross-validation bandwidth selection in non-parametric kernel regression under fixed and random designs. Computational Statistics & Data Analysis, 181, 107689. https://doi.org/10.1016/j.csda.2022.107689
29. Kassi, O., & Wang, S. G. W. (2024).Structural adaptation and rate-accelerated bandwidth selection in non-parametric local polynomial smoothing. Journal of Nonparametric Statistics, 36(3), 512–535. https://doi.org/10.1080/10485252.2024.2315891

تقييم استراتيجيات اختيار عرض النطاق لتنعيم دالة الانحدار المتوسطة

الملخص

يُعد الانحدار غير المعلمي (Non-parametric regression) نوعاً من أنواع تحليل الانحدار الذي لا يتطلب تحديد الشكل الدالي للعلاقة بين متغير الاستجابة والمتغير التنبؤي المرتبط به لملاءمة نموذج ما لمجموعة من البيانات. وتوجد طرق عديدة ومختلفة للانحدار غير المعلمي؛ ففي هذه الورقة البحثية، استخدمنا مُقدِّر النواة متعدد الحدود المحلي (local polynomial kernel estimator) مع الاختيار الأمثل لمعامل (أو معاملات) التنعيم. ويُعد اختيار معامل التنعيم الأمثل -الذي يُعرف عادةً باسم "عرض النطاق" (bandwidth)- من أهم المسائل عند استخدام المُقدِّرات القائمة على النواة؛ إذ يلعب عرض النطاق الأمثل دوراً كبيراً في الكشف عن بنية البيانات الأساسية. وقد استخدمنا هنا أسلوب المحاكاة للمقارنة بين أربع استراتيجيات مختلفة لاختيار عرض النطاق الأمثل، وهي: استراتيجية التعويض المباشر (Direct Plug-in)، ومعيار مجموع مربعات البواقي (Residual Sum of Squares)، ومعيار التحقق المتقاطع (Cross-Validation)، ومعيار القاعدة التجريبية (Rule of Thumb). وفي سياق هذه الاستراتيجيات الأربع لاختيار عرض النطاق الأمثل، تم النظر في إعدادين مختلفين للتصميم: التصميم الثابت (Fixed Design) والتصميم العشوائي (Random Design). وبالإضافة إلى ذلك، جرى اختبار أربع دوال انحدار مختلفة لتنعيم دالة متوسط الانحدار g(x)، كما بحثنا في تأثير ثلاثة عناصر مهمة: حجم العينة (n)، وتوزيع الأخطاء (توزيع طبيعي أو غير طبيعي)، ونوع التصميم (ثابت أو عشوائي). وقد أفضت دراسات المحاكاة -التي أُجريت بواقع 500 تكرار- إلى نتائج مثيرة للاهتمام.

الكلمات المفتاحية: مُقدِّرات النواة متعددة الحدود المحلية، استراتيجية الاختيار بالتعويض المباشر (Direct Plug-in)، معيار مجموع مربعات البواقي، معيار التحقق المتقاطع، معيار القاعدة التجريبية.