Afro-Asian Journal of Scientific Research (AAJSR)
المجلة الأفروآسيوية للبحث العلمي
E-ISSN: 2959-6505
Volume 4, Issue 4, 2026
Page No: 215-231
Website: https://aajsr.com/index.php/aajsr/index
SJIFactor 2024: 5.028 ISI 2025: 0.915 معامل التأثير العربي (AIF) 2025: 0.76

Bandwidth Selection for Local Polynomial Estimation of Regression Derivatives

Tarek A. Elghazali[1*], Fatima H. Alhamroush[2], Ahmed M. Mami[3]
[1,2,3]Department of Statistics, Faculty of Science, University of Benghazi, Benghazi, Libya

اختيار عرض النطاق للتقدير متعدد الحدود الموضعي لمشتقات الانحدار

طارق عبد الجليل الغزالي[1*]، فاطمة الحمروش[2]، أحمد محمد مامي[3]
[1،2،3]قسم الإحصاء، كلية العلوم، جامعة بنغازي، بنغازي، ليبيا

*Corresponding author: tarek.elghazali@uob.edu.ly

Received: September 01, 2026 | Accepted: November 07, 2026 | Published: November 21, 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract:
Bandwidth selection is a fundamental issue in nonparametric regression because it directly influences the accuracy and smoothness of the estimated regression function and its derivatives. An inappropriate bandwidth may result in either excessive smoothing, leading to biased estimates, or insufficient smoothing, producing highly variable estimates. This study presents a comprehensive comparative evaluation of four widely used bandwidth selection strategies for local polynomial kernel estimation, namely the Direct Plug-In (DPI), Cross-Validation (CV), Rule-of-Thumb (RT), and Residual Square Criterion (RSC). The comparison focuses on estimating the first and second derivatives of the mean regression function under both fixed and random design settings. A Monte Carlo simulation experiment consisting of 500 independent replications was conducted using four representative regression functions with different structural characteristics. The performance of each bandwidth selection strategy was investigated under three sample sizes ($n = 50, 100$, and 300) and two error distributions (normal and non-normal). Estimation accuracy was evaluated using Average Bias ($ABias$), Average Variance ($AVar$), and Root Average Mean Squared Error ($RAMSE$). The simulation results demonstrate that the Residual Square Criterion consistently provides superior estimation accuracy across most experimental scenarios, particularly when estimating higher-order derivatives. The Direct Plug-In approach generally ranks second in performance, followed closely by the Rule-of-Thumb method, whereas Cross-Validation frequently produces less stable estimates because of its tendency to select relatively small bandwidths. Furthermore, increasing the sample size substantially improves estimation accuracy regardless of the bandwidth selection strategy or error distribution. These findings indicate that the Residual Square Criterion constitutes a reliable and efficient bandwidth selection method for local polynomial kernel estimation of regression function derivatives in both fixed and random design settings.

Keywords: Local Polynomial Kernel Regression, Bandwidth Selection, Regression Derivatives, Monte Carlo Simulation.

الملخص:
يُعدّ اختيار عرض النطاق مسألةً أساسيةً في الانحدار الغير بارامتري، إذ يؤثر بشكل مباشر على دقة وسلاسة دالة الانحدار المُقدَّرة ومشتقاتها. قد يؤدي اختيار عرض نطاق غير مناسب إما إلى تنعيم مفرط، مما ينتج عنه تقديرات متحيزة، أو إلى تنعيم غير كافٍ، مما ينتج عنه تقديرات شديدة التباين. تُقدّم هذه الدراسة تقييمًا مقارنًا شاملًا لأربع استراتيجيات شائعة الاستخدام لاختيار عرض النطاق لتقدير نواة متعددة الحدود المحلية، وهي: طريقة التوصيل المباشر (DPI)، والتحقق

المتبادل (CV)، وقاعدة التقريب (RT)، ومعيار مربع البواقي (RSC). تركز المقارنة على تقدير المشتقين الأول والثانية لدالة الانحدار المتوسطة في ظل إعدادات لتصاميم ثابتة وعشوائية. أُجريت تجربة محاكاة مونت كارلو تتكون من 500 تكرار مستقل باستخدام أربع دوال انحدار تمثيلية ذات خصائص هيكلية مختلفة. تمّت دراسة أداء كل استراتيجية لاختيار عرض النطاق الترددي في ظل ثلاثة أحجام عينات (n = 50، 100، و300) وتوزيعين للأخطاء (طبيعي وغير طبيعي). وقُيّمت دقة التقدير باستخدام متوسط الانحياز (ABias)، ومتوسط التباين (AVar)، وجذر متوسط مربع الخطأ (RAMSE). أظهرت نتائج المحاكاة أن معيار مربع البواقي يُوفّر باستمرار دقة تقدير فائقة في معظم السيناريوهات التجريبية لدراستنا، لا سيما عند تقدير المشتقات ذات الرتب العليا. ويحتل أسلوب التوصيل المباشر المرتبة الثانية في الأداء عمومًا، يليه مباشرةً أسلوب القاعدة العامة، بينما يُنتج التحقق المتقاطع في كثير من الأحيان تقديرات أقل استقرارًا نظرًا لميله إلى اختيار عروض نطاق ترددي صغيرة نسبيًا. علاوة على ذلك، تُحسّن زيادة حجم العينة دقة التقدير بشكل كبير بغض النظر عن استراتيجية اختيار عرض النطاق الترددي أو توزيع الخطأ. تشير هذه النتائج إلى أن معيار المربع المتبقي يشكل طريقة موثوقة وفعالة لاختيار عرض النطاق لتقدير نواة متعددة الحدود المحلية لمشتقات دالة الانحدار في كل من إعدادات التصميم الثابتة والعشوائية.

الكلمات المفتاحية: الانحدار بالنواة متعددة الحدود المحلية، اختيار عرض النطاق، مشتقات الانحدار، محاكاة مونت كارلو.

Introduction

Nonparametric regression has become one of the most important statistical techniques for modeling complex relationships between response and explanatory variables without assuming a predetermined functional form. Unlike classical parametric regression models, nonparametric approaches provide greater flexibility by allowing the data to determine the shape of the regression function. Consequently, these methods have been widely applied in economics, environmental sciences, engineering, medicine, machine learning, and financial modeling, where nonlinear relationships frequently arise (1–4).

Among the available nonparametric regression techniques, Local Polynomial Kernel (LPK) regression has received considerable attention because of its excellent theoretical properties and practical performance. Local polynomial estimators possess attractive boundary behavior, achieve low estimation bias, and provide reliable estimates of regression derivatives. These advantages make local polynomial regression particularly suitable for estimating the first and higher-order derivatives of regression functions, which play an important role in many scientific and engineering applications (1,2). Despite the flexibility of local polynomial regression, its performance depends critically on the appropriate selection of the smoothing parameter, commonly known as the bandwidth. The bandwidth controls the trade-off between estimation bias and variance. A small bandwidth generally captures local features more accurately but increases estimation variability, whereas a large bandwidth produces smoother estimates at the expense of increased bias. Therefore, selecting an optimal bandwidth remains one of the most challenging and influential aspects of kernel-based regression estimation (1,2,5–7).

Several bandwidth selection methods have been proposed in the statistical literature, including Direct Plug-In (DPI), Cross-Validation (CV), Rule-of-Thumb (RT), and residual-based approaches. Each method follows a different optimization principle and may perform differently depending on sample size, error distribution, data structure, and the complexity of the underlying regression function. Consequently, no single bandwidth selection strategy is universally optimal for all practical situations, motivating further comparative investigations (4,6,8–10).

The primary contribution of this paper is to provide a comprehensive simulation-based comparison of four bandwidth selection strategies for estimating the first and second derivatives of the mean regression function using local polynomial kernel estimation. The comparison considers both fixed and random design settings, multiple regression functions with different levels of complexity, various sample sizes, and both normal and non-normal error distributions. Performance is evaluated using Average Bias ($ABias$), Average Variance ($AVar$), and Root Average Mean Squared Error ($RAMSE$) (8). The findings provide practical guidance for selecting an appropriate bandwidth strategy in derivative estimation problems and contribute to improving the reliability of nonparametric regression analysis.

The effectiveness of local polynomial kernel regression depends primarily on the appropriate selection of the bandwidth parameter, which governs the degree of smoothing applied to the estimated regression function. Although local polynomial estimators possess desirable asymptotic properties, their practical performance may deteriorate considerably when an unsuitable bandwidth is selected. Excessively small bandwidths usually produce highly variable estimates with substantial random fluctuations, whereas excessively large bandwidths oversmooth the data and obscure important local features of the underlying regression function.

The estimation of regression function derivatives presents an even greater challenge because derivative estimators are inherently more sensitive to bandwidth selection than the estimation of the

regression function itself. Consequently, selecting an optimal bandwidth becomes increasingly important when estimating first- and second-order derivatives (5,7).

Although numerous bandwidth selection procedures have been proposed in the statistical literature, including Direct Plug-In (DPI), Cross-Validation (CV), Rule-of-Thumb (RT), and residual-based approaches, their relative performance varies according to sample size, error distribution, regression complexity, and experimental design. Therefore, there remains a need for a systematic comparison of these methods under a wide range of practical conditions.

Accordingly, this study investigates the performance of four bandwidth selection strategies through an extensive Monte Carlo simulation study in order to identify the most reliable approach for estimating the first and second derivatives of the mean regression function under both fixed and random design settings.

Problem Statement:

The effectiveness of local polynomial kernel regression depends primarily on the appropriate selection of the bandwidth parameter, which governs the degree of smoothing applied to the estimated regression function. Although local polynomial estimators possess desirable asymptotic properties, their practical performance may deteriorate considerably when an unsuitable bandwidth is selected. Excessively small bandwidths usually produce highly variable estimates with substantial random fluctuations, whereas excessively large bandwidths oversmooth the data and obscure important local features of the underlying regression function.

The estimation of regression function derivatives presents an even greater challenge because derivative estimators are inherently more sensitive to bandwidth selection than the estimation of the regression function itself. Consequently, selecting an optimal bandwidth becomes increasingly important when estimating first and second-order derivatives (11).

Although numerous bandwidth selection procedures have been proposed in the statistical literature, including Direct Plug-In (DPI), Cross-Validation (CV), Rule-of-Thumb (RT), and residual-based approaches, their relative performance varies according to sample size, error distribution, regression complexity, and experimental design. Therefore, there remains a need for a systematic comparison of these methods under a wide range of practical conditions (12).

Accordingly, this study investigates the performance of four bandwidth selection strategies through an extensive Monte Carlo simulation study in order to identify the most reliable approach for estimating the first and second derivatives of the mean regression function under both fixed and random design settings.

Research Objectives:

The main objective of this study is to evaluate and compare the performance of several bandwidth selection strategies for local polynomial kernel estimation of regression function derivatives. The specific objectives are (a) To investigate the performance of four bandwidth selection methods, namely Direct Plug-In (DPI), Cross-Validation (CV), Rule-of-Thumb (RT), and Residual Square Criterion (RSC); (b) To compare the estimation accuracy of these methods for the first and second derivatives of the mean regression function; (c) To evaluate the influence of sample size on bandwidth selection performance; (d) To investigate the effect of normal and non-normal error distributions on derivative estimation; (e) To compare the performance of bandwidth selection methods under both fixed and random design settings; and (f) To identify the most efficient bandwidth selection strategy based on Average Bias ($ABias$), Average Variance ($AVar$), and Root Average Mean Squared Error ($RAMSE$).

Literature Review:

Elghazali et al. (2026) have investigated a simulation study that assessed four bandwidth selection strategies—Direct Plug-In (*DPI*), Cross-Validation (CV), Rule of Thumb (RT), and Residual Squares Criterion (RSC)—for smoothing the mean regression function using local polynomial kernel regression. The study considered fixed and random designs, different regression functions, sample sizes of 50, 100, and 300, and both normally and non-normally distributed errors. Performance was evaluated using average bias, average variance, and $RAMSE$ over 500 simulation replications. The results showed that increasing the sample size generally improved estimation accuracy, while non-normal errors increased variability. Overall, DPI demonstrated the most consistent performance, producing relatively lower bias and $RAMSE$ across the investigated scenarios, followed by RSC in several settings. These findings provide an important methodological basis for the current study in evaluating bandwidth selection strategies and identifying factors affecting the accuracy of nonparametric regression estimation. The study has primarily focused on bandwidth selection for estimating the regression function itself, whereas relatively limited attention has been given to the comparative performance of bandwidth selectors specifically for higher-order regression derivatives under varying design structures and error distributions. The present study addresses this gap by systematically comparing DPI, CV, RT, and RSC for first and second-order derivative estimation under both fixed and random designs.

Recent developments have also addressed bandwidth selection for estimating both regression functions and their derivatives under more complex error structures, including unknown error correlations (13).

Material and methods:

This study employs a simulation-based methodology to compare the performance of four bandwidth selection strategies within the framework of local polynomial kernel regression. The methodological framework consists of constructing local polynomial estimators, selecting the optimal bandwidth using different selection procedures, and evaluating the resulting derivative estimates through repeated Monte Carlo simulations.

The comparison is carried out under both fixed and random design settings in order to assess the robustness of each bandwidth selection strategy across different experimental conditions. The investigation considers the estimation of both the first and second derivatives of the mean regression function, since higher-order derivative estimation is considerably more sensitive to bandwidth choice than estimation of the regression function itself.

To ensure a comprehensive evaluation, four representative regression functions with different levels of smoothness and curvature are employed. These functions include multimodal, oscillatory, nonlinear, and highly varying patterns that reflect a broad range of practical regression problems encountered in statistical applications.

The simulation experiments are repeated 500 times for each experimental configuration. Three sample sizes (50, 100, and 300 observations) are considered to examine the effect of increasing sample information on estimation performance. In addition, two different error distributions are investigated, namely normally distributed errors and non-normal error distributions, allowing the robustness of each bandwidth selection procedure to be assessed under different stochastic environments.

The performance of the competing bandwidth selection strategies is evaluated using three widely accepted statistical measures: Average Bias ($ABias$), Average Variance ($AVar$), and Root Average Mean Squared Error ($RAMSE$). These performance criteria jointly measure estimation accuracy, stability, and overall efficiency, providing a comprehensive basis for comparing the competing bandwidth selection procedures.

Bandwidth Selection Strategies:

There are four main strategies for selecting the bandwidth as follows:

1. **Direct Plug-In** ($\boldsymbol{DPI}$): This method relies on accurate estimation of certain theoretical quantities related to data density.
2. **Residual Squares Criterion** ($\boldsymbol{RSC}$): This criterion focuses on minimizing the sum of squared residual errors.
3. **Cross-Validation** ($\boldsymbol{CV}$): 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.
4. **Rule-of-Thumb** ($\boldsymbol{RT}$): This method uses simplified mathematical formulas to estimate the bandwidth based on the characteristics of the data.
5. **The Direct Plug-In ($\boldsymbol{DPI}$) 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:**

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

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:

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

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:

$$MISE[\hat{g}(.;1,h) = E(\int \{\hat{g}(x;1,h) - g(x)\}^2\, dx)$$

$$= n^{-1}h^{-1}R(k)\int v(x)\, dx + h^4\{\mu_2(k)/2\}^2 \int g''(x)^2\, dx + op\{(nh)^{-1} + h^4\}$$

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:

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

In both design settings, the bias expression purely depends on ***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 ***x***. However, if $b(.)$ is uniform then it will serve as deflection factor in the approximating variance term (9).

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:

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

The Direct Plug-In (DPI) Bandwidth Selection to Smooth the Mean Regression Function:

The Direct Plug-in ($\boldsymbol{DPI}$) 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. In the case of local linear kernel estimator, the asymptotical bandwidth under $MISE$ criterion (6) can be seen as:

$$h_{MISE} = C_1(k)[\int v(x)\, dx\ \{\int x^2 b(x)\, dx\}^{-1}]^{1/5}, \quad (10)$$

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:

$$\theta_{rs} = \int g^{(r)}(x)g^{(s)}(x)b(x)\, dx;\ r,s \geq 0;\ r+s \text{ even} \quad (11)$$

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

$$\hat{\theta}_{22}(g) = n^{-1}\textstyle\sum_{i=1}^{n}[\hat{g}^{(2)}\,(X_i;3,g)], \quad (12)$$

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 (6)

$$g_{AMSE} = C_2(K)[\sigma^2(b-a)/|\theta_{24}|n]^{1/7}, \quad (13)$$

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. (13), 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).

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 Hall & Marron (1990), which is based on the zero-degree least squares fit, to the first-degree setting. This leads to the first-degree variance estimator:

$$\hat{\sigma}_1^2(\lambda) = v^{-1}\sum_{i=1}^{n}\{Y_i - \hat{g}(X_i;1,\lambda)\}^2 \qquad (14)$$

Here,

$$v = n - 2\sum_i w_{ii} + \sum_i\sum_j w_{ij}^2, \text{ and } w_{ij} = e_1^T(X_{P,T_i}^T W_{T_i} X_{P,T_i})^{-1}(X_{P,T_i}^T W_{T_i} e_j) \qquad (15)$$

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 $\hat{\sigma}_1^2(\lambda)$ (6). It follows that the asymptotically $AMSE$-optimal choice of λ **is:**

$$\lambda_{AMSE} = C_3(K)[\sigma^4(b-1)/\theta_{22}^2 n^2]^{1/9}, \qquad (16)$$

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:

$$\hat{h}_{MISE} = C_1(K)[\hat{\sigma}_1^2(\hat{\lambda}_{AMSE})(b-a)/\hat{\theta}_{22}(\hat{g}_{AMSE})n]^{1/5} \qquad (17)$$

The Simulation Study:

A simulation is an imitation of some real things, or process. The act of simulating something generally entails representing certain key characteristics or behaviors of a selected statistical models. In this chapter the main purpose is to conduct a simulation study in order to evaluate and compare the estimated mean regression functions and their derivatives. These estimated regression curves are resulted from utilizing the four bandwidth selection procedures: namely Direct Plug-In (DPI) selection procedure, Cross-validation (CV) Criterion, the Rule of Thumb (RT) Criterion, and Residual Square Criteria (RSC) selection procedure in two different design settings namely: The Fixed (equally-spaced) design and the Random design. Such a study is necessarily to be restrictive, because there are many possibilities for the choices of example regression functions $g(x)$**,** sample sizes (n), bandwidth selection strategies, and the distribution of the errors term.

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. We consider four different example regression function as present 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

Results and discussion:

Description of the Experiment in Fixed Design Setting:

We have conducted 48 simulation studies, using the first and second derivatives of four example derivative regression functions, and three different sample sizes (n), and the errors' distribution were generated using both the normal distribution and non-normal distribution. Each simulation study involves ($L = 500$) repetitions. More specifically, the first and second derivatives of example regression functions $f(x)$ that have been used in Table 1. We also have chosen three different choices of sample sizes of $n = 50, 100,$ and 300 respectively.

Table (2): The First and Second Derivatives of the Example Regression Functions.

Example	$f_i^j(x);\ j = 1st, 2nd\ derivates$	Interval	Sigma
1	$f_1'(x) = -(0.7\ (exp(-64\ (x - 0.75)^2)\ (64\ (2\ (x - 0.75)))) + 0.3\ (exp(-16\ (x - 0.25)^2)\ (16\ (2\ (x - 0.25)))))$ $f_1''(x) = -(0.7\ (exp(-64\ (x - 0.75)^2)\ (64 * 2) - exp(-64\ (x - 0.75)^2)$ $(64\ (2\ (x - 0.75)))\ (64\ (2\ (x - 0.75)))) + 0.3\ (exp(-16\ (x - 0.25)^2)\ (16 * 2) - exp(-16(x - 0.25)^2)(16\ (2\ (x - 0.25)))\ (16\ (2\ (x - 0.25)))))$	(0,1)	0.15
2	$f_2'(x) = 0.75\ (cos((3\pi\ x)/2 + 1.25)\ (3\pi/2))$ $f_2''(x) = -(0.75\ (\sin((3\pi\ x)/2 + 1.25)\ (3\ \pi/2)\ (3\pi/2)))$	(0,1)	0.6
3	$f_3'(x) = \cos(x) - 2\sin(x) + 3(\cos(5x)5)$ $f_3''(x) = -(3\ (\sin(5x) * 5 * 5) + (\sin(x) + 2\cos(x)))$	(-1,1)	1.2
4	$f_4'(x) = 3(\cos(2\ \pi\ x^3)\ (2\ \pi\ (3\ x^2))\ (\sin(2\ \pi\ x^3))^2)$ $f_4''(x) = 3((cos(2\pi x^3)\ (2\pi(3(2\ x))$ $- sin(2\ \pi\ x^3)\ (2\ \pi\ (3\ x^2))\ (2\ \pi\ (3\ x^2)))\ (sin(2\ \pi\ x^3))^2$ $+ cos(2\pi x^3)(2\pi(3\ x^2))\ (2\ (cos(2\pi\ x^3)(2\ \pi\ (3x^2))(sin(2\pi x^3)))))$	(0,1)	0.5

Numerical Summary for Fixed Design Setting:

Having conducted the simulation runs (L=500), the numerical results are tabulated in Table 2. Thus, Table 2 displays the numerical summary of results obtained using four different bandwidth selection procedures $(DPI, CV, RT\ or\ RSC)$ when the bandwidth selection setting is either Fixed Design or Random Design, with three different choices of sample sizes $(50,\ 100,$ and $300)$. The numerical summary again includes three statistical properties namely the Average Bias $(ABias)$, the Average Variance $(AVar)$ and the Square Root of Average Mean Square Errors $(RAMSE)$ for each example derivative regression function $f_1'(x),\ f_1''(x),\ f_2'(x), f_2''(x)\ , f_3'(x),\ f_3''(x)\ f_4'(x),$ and $f_4''(x)$ respectively.The formulae that we have been used to calculate $ABias$, $AVar$, and $RAMSE$ are as follows:

$$ABias(\hat{f}') = (1/n)\sum_{i=1}^{n}[\frac{1}{L}\sum_{J=1}^{L}\hat{f}'(x_{ij}) - f'(x_i)] \quad (1)$$

$$AVar(\hat{f}') = (1/n)\sum_{i=1}^{n}[\frac{1}{L}\sum_{J=1}^{L}\hat{f}'^2(x_{ij}) - (\frac{1}{L}\sum_{j=1}^{L}\hat{f}'(x_{ij}))^2] \quad (2)$$

$$RAMES(\hat{f}') = \sqrt[2]{ABias(\hat{f}')^2 + AVar(\hat{f}')} \quad (3)$$

$$ABias(\hat{f}'') = (1/n)\sum_{i=1}^{n}[\frac{1}{L}\sum_{J=1}^{L}\hat{f}''(x_{ij}) - f''(x_i)] \quad (4)$$

$$AVar(\hat{f}'') = (1/n)\sum_{i=1}^{n}[\frac{1}{L}\sum_{J=1}^{L}\hat{f}''^2(x_{ij}) - (\frac{1}{L}\sum_{j=1}^{L}\hat{f}''(x_{ij}))^2] \quad (5)$$

$$RAMES(\hat{f}'') = \sqrt[2]{ABias(\hat{f}'')^2 + AVar(\hat{f}'')} \quad (6)$$

The simulation programs, and auxiliary programs (in Rgui code) are used to calculate the bandwidths as well as obtaining the corresponding fitted values of the derivative regression curves through utilizing four different criteria, namely: (DPI) procedure, (CV) procedure, (RT) procedure, and (RSC) procedure when the bandwidth is in fixed design setting.

Table (3): The numerical summary of results obtained using four different bandwidth selection (DPI, RT, CV, and RSC) in two settings (normally and non-normally distributed errors) from the simulation study in fixed design. That includes ABias, Avar, and RAMSE for 500 simulation runs using $f_1'(x)$, and $f_1''(x)$ of first example regression function.

		***DPI* Criterion**		***CV* Criterion**		***RT* Criterion**		***RSC* Criterion**	
		Normal	**Non-Nor.**	**Normal**	**Non-Nor.**	**Normal**	**Non-Nor.**	**Normal**	**Non-Nor.**
	n	**The First Derivative**							
		The Example Derivative Regression Function $f_1'(x)$							
Abias	50	0.36564	0.36295	0.46500	0.49138	0.37305	0.38227	0.24463	0.21811
	100	0.22757	0.19721	0.32929	0.31284	0.25951	0.22670	0.20581	0.17957
	300	0.11024	0.10922	0.15733	0.15286	0.10831	0.10848	0.11569	0.12133
Avar	50	1.96779	1.73164	2.13321	1.99822	1.97567	1.75997	1.71585	1.6004
	100	1.28425	1.53041	1.56213	1.85882	1.37106	1.60586	0.75464	0.95030
	300	0.75003	0.65880	0.95105	0.85516	0.73285	0.66167	0.34802	0.31867
ARMSE	50	1.52756	1.44727	1.61190	1.57785	1.53255	1.46364	1.40017	1.34852
	100	1.22956	1.31439	1.37529	1.47365	1.27694	1.35372	0.98003	1.06295
	300	0.92751	0.87637	1.05280	1.00400	0.91715	0.87758	0.68097	0.66332
		The Second Derivative							
		The Example Derivative Regression Function $f_1''(x)$							
Abias	50	318.6828	866.5006	381.5719	899.8487	380.1832	1047.507	276.3581	480.2801
	100	139.4292	680.1352	226.8741	738.4478	174.232	954.6077	210.8734	560.4803
	300	81.98444	490.1875	111.9861	658.3791	81.49999	720.5475	155.1699	547.1426
Avar	50	2140.261	49011.35	2170.216	49092.31	2083.797	49152.58	1299.322	120609.5
	100	2668.523	53474.04	2675.319	53779.78	2642.983	53676.35	803.1587	80016.39
	300	1572.054	23941.19	1607.27	23976.41	1560.644	24194.41	301.6717	27101.19
ARMSE	50	49.58774	223.3335	50.51522	223.5893	49.63849	224.0538	39.69484	347.9796
	100	52.99011	232.7105	53.87201	233.4914	53.07744	233.7327	31.84387	283.8606
	300	40.66988	156.3054	41.46391	156.9547	40.52338	157.8447	21.37385	166.2779

Having a close look at Table 2, that summarize the results of using several smoothing parameter selection criteria (DPI, CV, RT, and RSC) using the first example regression functions, to smooth $f_1'(x)$, and $f_1''(x)$ in fixed design setting: we have made several comparisons as following: For the first derivative, the RSC criterion showed the lowest bias values in most scenarios, especially with a small sample size ($n = 50$) and under both error distributions (normal and non-normal), indicating its accuracy in estimating the true value without systematic deviation. Both the DPI and RT criteria ranked close to and good after the RSC criterion, while the CV criterion recorded the highest bias values overall .For the second derivative, bias values increased significantly for all criteria, which is expected given the difficulty of estimating higher-order derivatives. In this context, the RSC criterion maintained a relatively satisfactory performance compared to the other criterions, whose bias values were largely similar.

Regarding the Variance measure ($AVar$):

For the first derivative, the RSC criterion demonstrated exceptional performance, recording the lowest variance values across all sample sizes and error types. The difference was particularly pronounced at large sample sizes (100, and 500). This indicates that the estimates produced by this criterion are the most stable and reliable. The performance of the DPI and RT criteria was similar and good, while the CV criterion continued to show the highest variance values. For the second derivative, a dramatic increase in variance values was observed across all criteria. However, the RSC criterion maintained its clear superiority, recording the lowest variance values by a significant margin compared to the other three criteria that showed similar performance.

Regarding the Root Means Squared Error measure ($ARMSE$):

For the first derivative, the superiority of the RSC criterion in reducing both bias and variance resulted in it achieving the lowest $ARMSE$ value, making it the optimal criterion for overall estimation accuracy. DPI and RT ranked close behind, while CV ranked last. For the second derivative, the overwhelming superiority of the RSC criterion in reducing variance compensated for its unexceptional performance in

the bias measure, resulting in it achieving the lowest $ARMSE$ value across all sample sizes, confirming its superior overall performance for the second derivative. DPI and RT ranked close behind, while CV ranked last.

It can be concluded that the RSC criterion is the most efficient and reliable criterion tested for selecting the optimal bandwidth in estimating both the first and second derivatives of a function, under fixed design conditions and for various sample sizes and error distributions. It is recommended for use in similar practical applications. It is followed in performance by both DPI and RT criteria, while CV criterion shows less reliable performance across different measures.

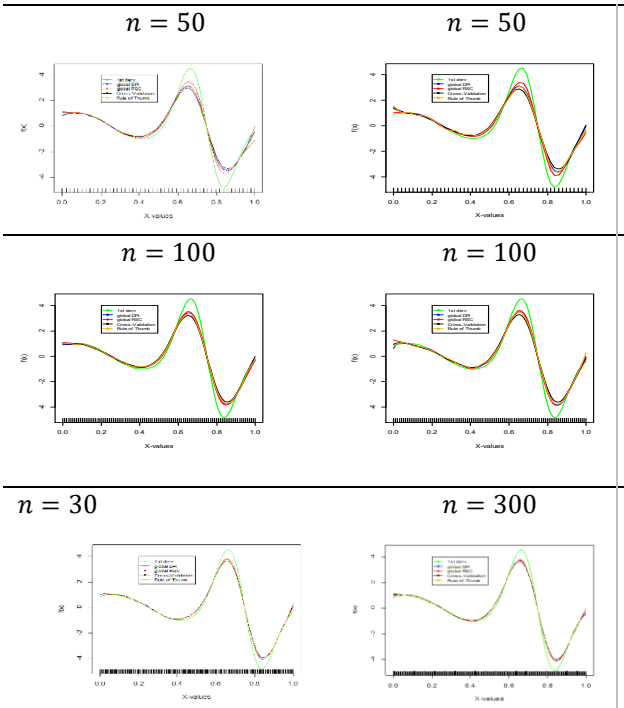


Figure (1): The plot of true $f_1'(x)$ (green line) and four others estimated 1st derivatives using four different procedures of selection optimal bandwidth (DPI, CV, RT, and RSC) in normally and non-normally distributed errors setting, DPI (blue line).

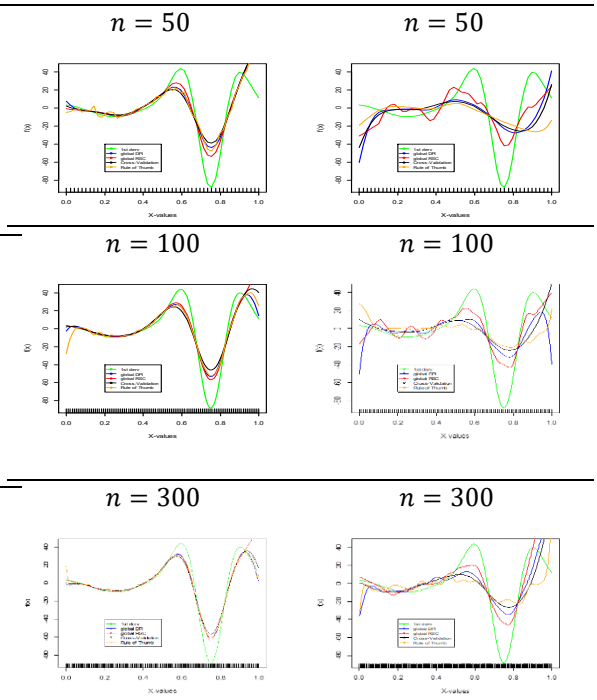


Figure (2): The plot of true $f_1''(x)$ (green line) and four others estimated second derivatives using four different procedures of selection optimal bandwidth (DPI, CV, RT, and RSC) in normally and non-normally distributed errors setting, DPI (blue line).

Figs. 1 and 2 plot the true derivative curve (green line) and the four other estimated curves (the other colored lines) 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 (normally and non-normally distributed errors). The errors' distributed once as normal (left hand side panels in Figs. 1 and 2) and the errors' non-normally distributed (right hand side panels in Figs. 1 and 2). Three different sample sizes ($n = 50, 100,$ and 300) and the first example regression functions $f_1'(x)$, *and* $f_1''(x)$.

For the first derivative $f_1'(x)$ in Fig. 1: It is clear that the four estimated curves closely approximate the true curve, especially as the sample size (n) increases. This indicates that all bandwidth selection methods performed well in estimating the first derivative of the function.

For the second derivative $f_1''(x)$ in Fig. 2: It can be seen that estimating the second derivative is more difficult, as the estimated curves exhibit some inconsistency from the true curve, and this inconsistency is greater in the case of non-normal errors. However, increasing the sample size (n) from 50 to 300 significantly improves the estimation accuracy for all methods.

In general, the figures show that the performance of the four methods improves as the sample size increases, making the estimates closer to the true curves.

Table (4): The numerical summary of results obtained using four different bandwidth selection (DPI , RT, CV, and RSC) in two settings (normally and non-normally distributed errors) from the simulation study in fixed design. That includes Abias, Avar, and RAMSE for 500 simulation runs using $f_2'(x)$, and $f_2''(x)$ of second example regression function.

		***DPI* Criterion**		***CV* Criterion**		***RT* Criterion**		***RSC* Criterion**	
		Normal	**Non-Nor.**	**Normal**	**Non-Nor.**	**Normal**	**Non-Nor.**	**Normal**	**Non-Nor.**
	n	**The First Derivative**							
		The Example Derivative Regression Function $f_2'(x)$							
$Abias$	50	0.11661	0.90863	0.08744	0.68319	0.31264	0.29645	0.13896	1.05888
	100	0.07447	0.17068	0.09825	0.14903	0.13238	0.53891	0.10312	0.54859
	300	0.02975	0.16346	0.02284	0.06474	0.05668	0.32277	0.01722	0.10675
$Avar$	50	10.0352	67.75846	10.07289	68.02855	10.50595	69.0935	17.6897	122.9138
	100	6.84346	50.06920	6.65461	50.11321	7.42263	51.04493	7.67285	75.14463
	300	2.53837	20.06129	2.52689	20.2963	2.73193	20.98825	2.22333	19.86354
$ARMSE$	50	3.18619	8.28656	3.18753	8.28925	3.28916	8.33006	4.22240	11.13430
	100	2.63020	7.08801	2.59863	7.08959	2.74864	7.18219	2.78854	8.700185
	300	1.60254	4.49719	1.59679	4.51232	1.66991	4.61639	1.49685	4.46881
		The Second derivative							
		The Example Derivative Regression Function $f_2''(x)$							
$Abias$	50	21.11182	187.7950	14.15098	77.07283	48.89424	7.744034	118.0269	2035.371
	100	6.18165	51.58501	12.77312	10.95858	29.78842	87.26171	10.5282	943.7741
	300	13.65544	31.97379	5.149931	27.56135	29.21344	133.7293	4.04465	46.90091
$Avar$	50	5389.345	56936.87	5412.301	57058.15	5337.762	57182.57	19265.93	286666.1
	100	6174.032	39306.19	6143.185	39334.77	6096.177	39171.86	9386.008	266159
	300	1882.022	16318.08	1876.271	16305.82	1871.606	16166.99	2428.329	20397.45
$ARMSE$	50	73.55581	239.0077	73.66446	239.0297	73.39384	239.145	139.2263	537.3095
	100	78.61433	198.3879	78.45991	198.3576	78.26854	198.1391	96.93573	516.8198
	300	43.53937	127.8673	43.37535	127.8021	43.59839	127.6743	49.3191	142.9837

Having a close look at Table 4, that summarize the results of using several smoothing parameter selection criteria (DPI, CV, RT, and RSC) using the second example regression functions, $f_2'(x)$, and $f_2''(x)$ in Fixed design setting: **we have made several comparisons as following:**

Regarding the Bias measure ($ABias$):

It is noted that the RSC criterion achieved the lowest bias values in most cases, especially with a small sample size ($n = 50$) and across both error distribution types (normal and non-normal), indicating its accuracy in estimating the true value. Meanwhile, the CV criterion showed the highest bias values overall, making it the least accurate in this regard. The performance of the DPI and RT criteria was comparable and acceptable.

For the second derivative: Bias values increased significantly for all criteria, which is expected due to the increased difficulty of estimating higher-order derivatives. In this context, the RSC criterion demonstrated relatively satisfactory performance compared to the other criteria.

Regarding the Variance measure ($AVar$):

For the first derivative: The RSC criterion stood out significantly, recording the lowest variance values across all sample sizes and error types, demonstrating the stability and reliability of its resulting estimates. Meanwhile, the CV criterion recorded the highest variance values in most comparisons. The performance of the DPI and RT criteria was close in the middle.

For the second derivative: Despite the significant increase in variance values across all criteria, the RSC criterion maintained its clear superiority, recording the lowest variance values by a significant margin.

Regarding the Root Means Squared Error measure ($ARMSE$):

For the first derivative: The RSC criterion's superiority in both bias and variance measures resulted in achieving the lowest $ARMSE$ values, making it the optimal criterion in terms of overall accuracy. The CV criterion ranked last in terms of performance. For the second derivative: Despite the complexity of estimating the second derivative, the RSC criterion achieved the lowest $ARMSE$ values in most cases, due to its significant advantage in reducing variance.

Overall, the RSC criterion demonstrated its superiority over the other criteria (DPI, CV, RT) across all measures (bias, variance, and root mean square error) and across various sample sizes and error

distributions. Therefore, it is recommended as a selection criterion for determining the optimal bandwidth for estimating derivatives of regression functions under fixed design settings.

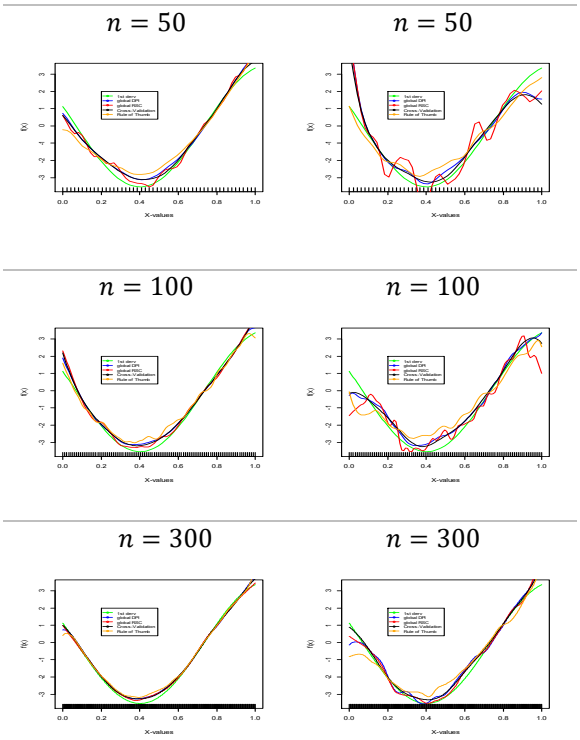


Figure (3): The plot of true $\mathrm{f}_2'(\mathrm{x})$ (green line) and four other estimated first derivatives using four different procedures of selection optimal bandwidth (DPI, CV, RT, and RSC) in normally and non-normally distributed errors setting, DPI (blue line), CV (red line), RT (black line), RSC (yellow line) from the simulation study in fixed design.

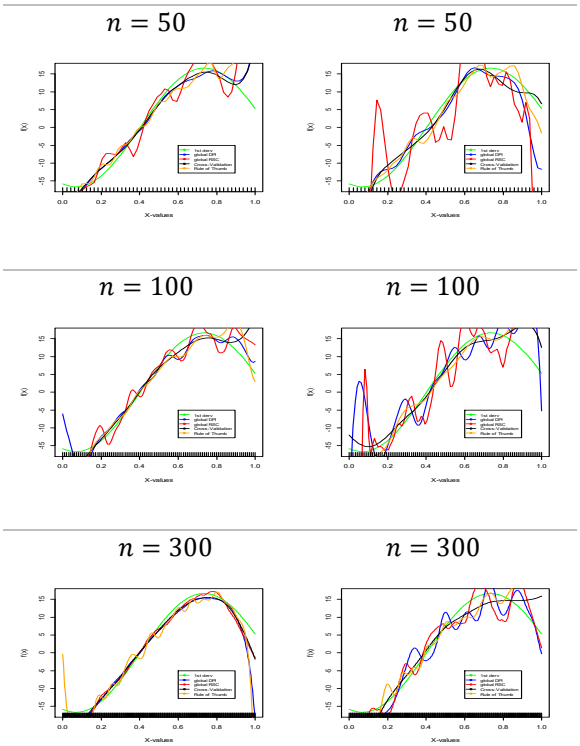


Figure (4): The plot of true $\mathrm{f}_2''(\mathrm{x})$ (green line) and four other estimated second derivatives using four different procedures of selection optimal bandwidth (DPI, CV, RT, and RSC) in normally and non-normally distributed errors setting, DPI (blue line), CV (red line), RT (black line), RSC (yellow line) from the simulation study in fixed design.

Figs. 3 and 4 plot the true derivative curve (green line) and the four other estimated curves (the other colored lines) 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 (normally and non-Normally distributed errors). The errors' distributed once as normal (left hand side panels in Figs. 3 and 4) and the errors' non-normally distributed (right hand side panels in Figs. 3 and 4). Three different sample sizes ($n = 50, 100$, and 300) and the second example regression functions $f_2'(x)$, *and* $f_2''(x)$.

In Fig. 3, which represents the first derivative of the second function $f_2'(x)$, we find that all four methods ($DPI, CV, RT,$ and RSC) provide estimates that are very close to the true function $f_2'(x)$, represented by the green line. This is true for both normally and non-normally distributed errors, and for all sample sizes ($50, 100,$ and 300). When the sample size increases to $n = 300$, all estimated lines converge better to the true line, indicating that accuracy increases with the number of samples.

In Fig. 4, which represents the second derivative of the second function $f_2''(x)$, the situation is different. The discrepancy in the estimates is more noticeable, especially at the smaller sample size of $n = 50$, where we see that the four estimates do not closely follow the true line $f_2''(x)$. However, as the sample size increases to $n = 100$ and then to $n = 300$, the quality of the estimates improves significantly. At $n = 300$, we observe that the estimated lines begin to better approximate the true line, confirming that increasing the sample size has a positive effect on the estimation accuracy, especially for higher derivatives.

In summary, it can be observed that the performance of all bandwidth selection methods improves significantly with increasing sample size, and that the first derivative estimate $f_2'(x)$, is more stable and accurate than the second derivative estimate $f_2''(x)$. This is consistent with theoretical expectations that higher derivative estimates are more sensitive to changes in data and sample size.

Table (5): The numerical summary of results obtained using four different bandwidth selection (DPI, RT, CV, and RSC**)** in two settings (normally and non-normally distributed errors) from the simulation study in fixed design. That includes ABias, Avar, and RAMSE for 500 simulation runs using $f_3'(x)$, and $f_3''(x)$ of third example regression function.

		***DPI* Criterion**		***CV* Criterion**		***RT* Criterion**		***RSC* Criterion**	
		Normal	**Non-Nor.**	**Normal**	**Non-Nor.**	**Normal**	**Non-Nor.**	**Normal**	**Non-Nor.**
	n	**The First derivative**							
		The Example Derivative Regression Function $f_3'(x)$							
$Abias$	50	2.77683	3.08726	2.91469	3.09547	3.49754	3.66545	2.58115	2.39712
	100	2.01887	2.09981	2.01026	2.13648	2.39227	2.56134	2.44049	3.14933
	300	0.96775	1.04263	0.97980	0.96420	1.07746	1.11604	1.39197	1.32893
$Avar$	50	49.55809	49.71956	50.38173	50.31075	52.55056	49.98182	32.71416	41.58796
	100	33.84347	47.62709	34.74946	48.65223	35.83436	49.89147	24.78997	56.67389
	300	15.70948	15.83878	15.56116	15.41303	16.45377	15.9224	8.00723	6.51575
$ARMSE$	50	7.23429	7.26683	7.30044	7.30796	7.48653	7.32443	5.94099	6.63213
	100	5.98852	7.05173	6.06298	7.12662	6.18277	7.24243	5.21828	7.73455
	300	4.08378	4.10870	4.06706	4.04688	4.18703	4.12776	3.06581	2.80084
		The Second derivative							
		The Example Derivative Regression Function $f_3''(x)$							
$Abias$	50	122.3843	187.4987	130.3298	190.9359	267.5346	200.4807	116.466	235.3494
	100	183.3202	130.5295	177.849	125.2402	240.5113	436.6262	150.0061	146.1766
	300	65.93659	59.79989	48.89262	59.87224	95.52943	107.4364	64.07839	107.949
$Avar$	50	5899.785	13005.87	5922.582	12970.61	5247.487	13294.81	4330.336	11663.75
	100	6022.901	14475.83	6004.116	14393.34	5965.445	13645.11	2110.405	6112.64
	300	5676.304	5988.629	5627.242	5827.672	5766.260	5669.255	910.3976	1427.92
$ARMSE$	50	77.60264	114.8624	77.80046	114.7238	74.26319	116.1692	66.68434	109.083
	100	78.77958	120.8568	78.62547	120.4931	78.77789	118.6665	47.54378	79.11269
	300	75.77757	77.77165	75.34012	76.73033	76.56233	76.00455	31.2166	39.19016

Having a close look at Table 5, that summarize the results of using several smoothing parameter selection criteria (DPI, CV, RT, and RSC) using the third example regression functions, $f_3'(x)$, and $f_3''(x)$ in Fixed design setting: **we have made several comparisons as following:**

Regarding the Bias measure ($ABias$):

The DPI and CV quantities showed similar and stable performance in reducing bias under both types of errors (normal and non-normal) and across different sample sizes (50, 100, and 300).

The RT quantity's performance was variable, recording a very low bias (3.49754) at $n = 50$ with normal distributed errors, but high bias values in other cases (e.g., 3.66545 with non-normal distributed errors for the same sample size), indicating its instability.

The RSC quantity showed a relatively low bias at $n = 50$, but recorded the highest bias value at $n = 100$ with non-normal distributed errors (3.14933), which may indicate its sensitivity to the error distribution.

Regarding the Variance measure ($AVar$):

In general, the variance ($AVar$) decreases with increasing sample size (from 50 to 300) for all criteria, consistent with the asymptotic properties of kernel estimators.

The RSC criterion tends to produce the lowest variance value, especially at large sample sizes ($n = 300$), recording 8.007231 and 6.515752 under normal and non-normal distributed errors, respectively. This suggests the potential ability of this criterion to achieve greater estimation stability.

The DPI and CV criteria exhibit similar and consistent performance in terms of variance across all sample sizes. RT Criterion typically records slightly higher contrast values than DPI and CV criteria, indicating lower efficiency.

Regarding the Root Means Squared Error measure ($ARMSE$):

The $ARMSE$ measure, which combines bias and variance, confirms that the RSC criterion performs best overall, especially with a large sample size ($n = 300$), where it recorded the lowest $ARMSE$ values (3.06581 and 3.00084).

The DPI and CV criteria perform similarly well, with DPI having a slight advantage in most scenarios. The RT criterion typically records higher $ARMSE$ values than its competitors, making it the least efficient of the four criteria tested.

To sum up: Under the different scenarios of this study (fixed design, third example regression function), the RSC criterion appears to be the most efficient and reliable, especially for large sample sizes, as it achieved the best balance between bias and variance, resulting in the lowest mean squared error. While DPI and CV are robust and reliable options across different scenarios, the RT criterion exhibits instability and is not recommended based on these results.

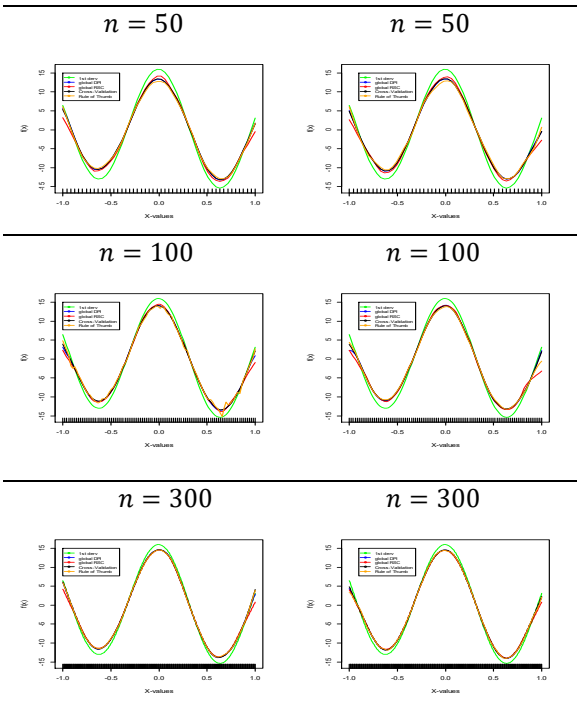


Figure (5): The plot of true $\mathrm{f}_3'(\mathrm{x})$ (green line) and four other estimated first derivatives using four different procedures of selection optimal bandwidth (DPI, CV, RT, and RSC) in normally and non-normally distributed errors setting, DPI (blue line), CV(red line), RT(black line), RSC(yellow line) from the simulation study in fixed design.

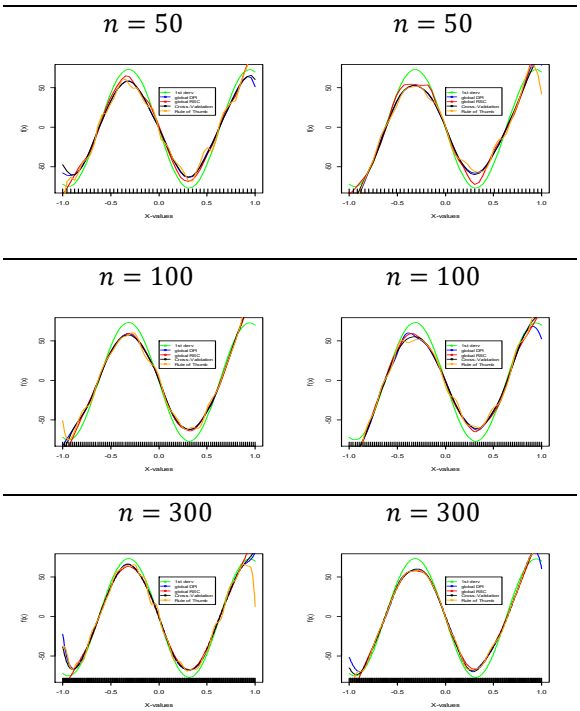


Figure (6): The plot of true $\mathrm{f}_3''(\mathrm{x})$ (green line) and four others estimated second derivatives using four different procedures of selection optimal bandwidth (DPI, CV, RT, and RSC) in normally and non-normally distributed errors setting, DPI (blue line), CV (red line), RT (black line), RSC (yellow line) from the simulation study in fixed design.

Figs. 5 and 6 plot the true derivative curve (green line) and the four other estimated curves (the other colored lines) 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 (normally and non-normally distributed errors). The errors' distributed once as normal (left hand side panels in Figs. 5 and 6) and the errors' non-normally distributed (right hand side panels in Figs. 5 and 6). Three different sample sizes (50, 100, and 300) and the third example regression functions $f_3'(x)$, *and* $f_3''(x)$.

For the First derivative (Fig. 5):

Generally, performance is good and stable. As the sample size increases, accuracy increases, and the estimated curves become closer to the actual value. The error distribution (normal or non-normal) has some effect, but its impact is limited to the first derivative.

For the Second derivative (Fig. 6):

Estimation here was much more difficult than for the first derivative. The curves follow unsteady pattern and are far from the actual value. The non-normal distributed errors aggravated the problem significantly. The curves became very inconsistent and unstable, especially in small samples. Even when the sample size increased ($n = 300$), accuracy improved, but not sufficiently; there was still some fluctuation.

To sum up: Sample size is very important for improving accuracy. Estimating the second derivative is more complex and very sensitive to noise (error) than the first derivative. If the data have non-normally distributed errors, this will have a very negative impact on the accuracy of the second derivative estimation and make the results volatile.

Table (6): The numerical summary of results obtained using four different bandwidth selection (DPI, RT, CV, and RSC) in two settings (normally and non-normally distributed errors) from the simulation

study in fixed design. That includes ABias, Avar, and RAMSE for 500 simulation runs using $f_4'(x)$, and $f_4''(x)$ of fourth example regression function.

		***DPI* Criterion**		***CV* Criterion**		***RT* Criterion**		***RSC* Criterion**	
		Normal	**Non-Nor.**	**Normal**	**Non-Nor.**	**Normal**	**Non-Nor.**	**Normal**	**Non-Nor.**
	n	**The First Derivative**							
		The Example Derivative Regression Function $f_4'(x)$							
$Abias$	50	19.77259	22.13356	23.4799	24.93962	20.85415	25.74188	18.71649	20.1095
	100	11.63057	19.82428	22.18073	23.73527	12.23251	23.5545	16.06915	18.85178
	300	6.543963	15.44432	19.60836	22.08329	5.806251	18.04101	15.68641	18.88247
$Avar$	50	17.42997	81.64828	24.29166	84.94796	20.70198	89.11598	16.48522	128.0421
	100	15.42508	73.99334	18.75585	78.81096	16.1024	83.19837	13.75041	81.83464
	300	9.67183	34.59179	8.806891	39.15394	9.628075	40.28698	4.492727	17.24064
$ARMSE$	50	6.09939	10.18734	6.911697	10.48273	6.446405	10.71718	5.933103	12.17175
	100	5.201505	9.685949	6.39817	10.12651	5.323054	10.33213	5.460729	10.03426
	300	4.026884	7.073621	5.330596	7.825422	3.928655	7.637276	4.49212	6.01025
		The Second Derivative							
		The Example Derivative Regression Function $f_4''(x)$							
$Abias$	50	38508.76	42498.09	44735.27	43452.95	35470.51	39671.28	44149.68	43370.43
	100	27228.4	38229.11	46791	41970.27	27060.53	38798.74	41381.7	43267.59
	300	19191.7	34227.76	44899.42	43929.92	17441.54	40183.6	39717.42	43014.56
$Avar$	50	17743.14	116881.1	17075.12	118489.9	19698.12	120346.5	14157.54	213475.8
	100	40471.59	80428.66	39416.46	83284.34	42205.91	84991.02	13045.25	129541.7
	300	28013.94	44537.9	32721.54	43452.88	28222.69	44635.54	5926.136	19133.62
$ARMSE$	50	237.1748	399.2233	248.6169	402.4212	234.88	400.0222	241.4689	506.8
	100	260.1922	344.4674	293.6111	353.9133	263.1852	351.8377	233.2958	415.7033
	300	217.2686	280.6522	278.6054	295.6058	213.6919	291.2373	213.6435	249.2954

Having a close look at Table 6, that summarize the results of using several smoothing parameter selection criteria (DPI, CV, RT, and RSC) using the fourth example regression functions, $f_4'(x)$, and $f_4''(x)$ in Fixed design setting: we have made several comparisons **as following:**

Regarding the Bias measure ($ABias$):

When comparing the bandwidth selection criteria (DPI, CV, RT, and RSC) for the first derivative $f_4'(x)$, we find that DPI and RSC criteria often achieve lower values for average bias ($Abias$) than CV and RT criteria, whether the errors are normally or non-normally distributed. This suggests that these two criteria provide more accurate and reliable estimates of the first derivative. For the second derivative $f_4''(x)$, Table 5 shows that $Abias$ estimates are significantly higher than for the first derivative. This is expected because estimating the second derivative is considered more difficult and data-sensitive. However, the RSC criterion tends to achieve better results in various situations, especially when compared to criteria such as CV and RT.

Regarding the Variance measure ($AVar$):

In general, the results show a significant improvement in performance as the sample size (n) increases from 50 to 300.

For the first derivative, increasing sample size (n) leads to a gradual decrease in the values of $AVar$ for all criteria. This confirms that providing more data points helps reduce bias and variance, thus improving overall estimation accuracy. For the second derivative, although the values are still high, there is a decrease in $AVar$ as the sample size increases, highlighting the great importance of increasing the sample size to obtain a more stable and accurate estimate, even for higher derivatives.

Regarding the Root Means Squared Error measure ($ARMSE$):

Table 5 clearly shows that the values of $ABias$, $AVar$, and $ARMSE$ are significantly higher when the errors are non-normally distributed compared to their normal distribution counterparts.

This significant difference in performance confirms that the criteria used are more efficient and effective under the assumption of a normal distribution of errors. When the error distribution deviates from the normal distribution, the difficulty of estimating the derivatives increases, leading to a significant increase in bias and variance, especially for estimating the second derivative $f_4''(x)$, where the $ARMSE$ values reach very high levels.

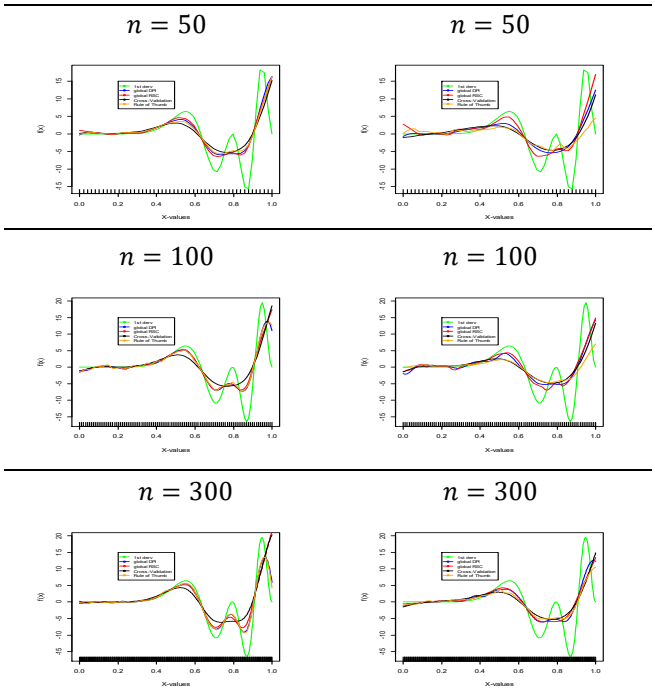


Figure (7): The plot of true $\mathrm{f}_4'(\mathrm{x})$ (green line) and four other estimated first derivatives using four different procedures of selection optimal bandwidth (DPI, CV, RT, and RSC) in normally and non-normally distributed errors setting, DPI (blue line), CV (red line), RT (black line), RSC (yellow line) from the simulation study in fixed design.

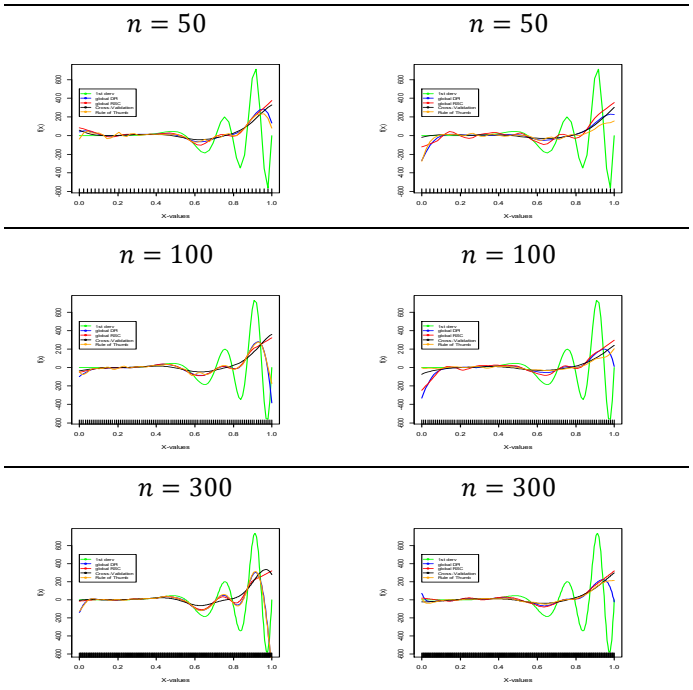


Figure (8): The plot of true $\mathrm{f}_4''(\mathrm{x})$ (green line) and four other estimated second derivatives using four different procedures of selection optimal bandwidth (DPI, CV, RT, and RSC) in normally and non-normally distributed errors setting, DPI (blue line), CV (red line), RT (black line), RSC (yellow line) from the simulation study in fixed design.

Figs. 7 and 8 plot the true derivative curve (green line) and the four other estimated curves (the other colored lines) 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 (normally and non-normally distributed errors). The errors' distributed once as normal (left hand side panels in Figs. 7 and 8) and the errors' non-normally distributed (right hand side panels in Figs. 7 and 8). Three different sample sizes (50, 100, and 300) and the fourth example regression functions $f_4'(x)$, *and* $f_4''(x)$.

For smoothing the first derivative $f_4'(x)$ (Fig. 7):

For the most part, all four criteria (DPI, CV, RT, and RSC) were able to capture the general pattern of the true first derivative (green line), especially for normal distributed errors (left panels). This means that the estimated curves follow roughly the same path.

The improvement with increasing sample size from 50 to 300 is noticeable, but uncertain compared to other regression examples functions. The estimated curves are closer to the true line and less jumpy, but some minor deviations are still present.

The effect of non-normal distributed errors (right panels) is more obvious. The estimated curves are rough, less smooth, and exhibit higher variance, especially at small sample sizes (50 and 100). This shows as a "jitter" in the lines, away from the smooth path of the true line.

Comparison between criteria: It is difficult to distinguish one criterion as clearly superior to the other in all cases. However, in general, RSC criterion (yellow line) and DPI criterion (blue line) produced curves that appeared to be slightly smoother and closer to the actual value in some panels, especially as sample size (n) increased.

For smoothing the second derivative $f_4''(x)$ (Fig. 8):

The situation is much more difficult here. Estimating the second derivative for fourth Example regression function appears very difficult, even with large samples. The estimated curves are all very far from the actual value (green line) and suffer from very high variance. This demonstrates itself in the form of jagged, scattered "paths" not centered around the actual values .The right-hand panels (non-normal distributed error) demonstrate the almost complete failure of the criteria to reliably estimate the second derivative. The curves are very sparse and inconsistent, and provide no indication of the shape of the true derivative. This confirms the extreme sensitivity of the second derivative to the nature of the distribution.

The role of sample size is negligible almost. Increasing the sample size to $n = 300$ did not significantly improve the accuracy of the second derivative in this example regression function. The problem lies in the difficulty of estimating the derivative, not in the size of the data alone. In such challenging scenarios, no criterion was clearly superior to the other. All performed very poorly in retrieving the shape of the true second derivative function.

To Sum up:

For the first derivative estimation, it can be reasonably estimated, especially with normal distributed errors and larger samples. The performance of the criteria is fairly close. For the second derivative estimation, it is very difficult and inaccurate for this particular example. All methods failed to provide reliable estimation, especially in the presence of non-normal distributed errors.

In general, these results confirm that the difficulty of estimation depends on the form of the original function itself. The fourth example regression function (of the form $(3\sin(2\pi\, x^3))^2$) seems to have derivatives, especially the second derivative, that are particularly challenging for the Suavizado technique in the kernel under study, which limits its effectiveness in this case.

Conclusion:

This section is devoted to highlight the major findings of our study in the light of results obtained through using four bandwidth selection procedures (Direct Plug-In (DPI), Residual Squares Criterion (RSC), Rule-of-Thumb (RT), and Cross-validation (CV) Criterion) in Fixed design setting and in the Random design setting in order to smooth the derivatives of mean regression functions.

The discussion is based on the numerical statistical aspect obtained from the simulation study. Finally, we provide the limitations, and several merits that warrant further investigation that could be used to extend scope of the present work.

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 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**: $\boldsymbol{DPI}$ Criterion consistently outperforms other methods, producing the smallest bias, variance, and $RMSE$ across different sample sizes and error distributions. It is the most efficient and reliable strategy overall. $\boldsymbol{RSC}$ performs second-best, especially with larger sample sizes, but shows higher bias and variance under more restrictive conditions. $\boldsymbol{CV}$ Criterion performs moderately well with lower bias than RT and RSC Criteria, but results are more variable, particularly under non-normal errors. $\boldsymbol{RT}$ Criterion shows the weakest performance, with consistently higher bias and variance, making it less suitable for accurate estimation.

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. Among the bandwidth selection methods, the DPI criterion consistently achieved the lowest bias, variance, and $RMSE$, making it the most reliable and efficient bandwidth selection strategy in both fixed and random designs. RSC performed second-best with larger samples, while CV Criterion showed moderate accuracy but greater variability. RT Criterion was the least effective, with higher bias and variance values. RT and RSC Criteria should be applied with caution, especially under non-normal error conditions, due to their reduced accuracy and stability.

Overall, DPI Criterion emerges as the preferred method for achieving accurate and robust regression function estimation across both fixed and random design settings. Moreover, *Recent studies have further demonstrated that both the error structure and bandwidth selection can substantially affect the accuracy and robustness of nonparametric derivative estimation (16).*

Limitations:

Although the simulation study considered several regression functions, sample sizes, and error distributions, certain limitations remain. The investigation was restricted to univariate local polynomial regression and focused exclusively on four global bandwidth selection procedures. Furthermore, only Gaussian kernel smoothing was considered throughout the analysis. Additional study is therefore required to examine more complex data structures and alternative smoothing techniques.

Recommendations:

Based on this study, several research questions may be considered for future investigation:

1. The comparative study may be extended to multivariate local polynomial regression models involving multiple explanatory variables.
2. Adaptive and variable bandwidth selection methods may be investigated and compared with the global bandwidth procedures considered in this study.
3. The proposed methodology may be applied to real-world datasets from medicine, economics, environmental sciences, and artificial intelligence to evaluate its practical effectiveness under realistic conditions.

References:

1. Fan J. Local polynomial modelling and its applications: monographs on statistics and applied probability 66. Routledge; 2018.
2. Wand MP, Jones MC. Kernel smoothing. CRC press; 1994.
3. Härdle W. Applied nonparametric regression. Cambridge university press; 1990.
4. Li Q, Racine JS. Nonparametric econometrics: theory and practice. Princeton University Press; 2007.
5. Liu, S., & Charnigo R. Tuning parameter selection for nonparametric derivative estimation in random design. Statistics (Ber). 2023;57(6):1402–25.
6. Ruppert D, Sheather SJ, Wand MP. An effective bandwidth selector for local least squares regression. J Am Stat Assoc. 1995;90(432):1257–70.
7. Zhou, J., & Parmeter CF. Data-driven resistant kernel regression. J Nonparametr Stat. 2025;37(1):33–59.
8. Tarek A. Elghazali, Fatima H. Alhamroush, Fawzia A. Altobly AMM. ISSN E : 3136-3709 L : 3136-3695 Vol 1 : No 2 ASSESSMENT OF BANDWIDTH SELECTION STRATEGIES TO Local polynomial kernel Estimator Univariate Local Polynomial Kernel Regression (? Libyan J Appl Contemp Sci. 2026;1(2):157–81.
9. Mami AM. Local polynomial regression with applications to both independent and longitudinal data. University of Manchester; 2002.
10. Loader CR. BANDWIDTH SELECTION: CLASSICAL OR PLUG-IN? Ann Stat. 1999;27(2):415–38.
11. Dong, H., Otsu, T., & Taylor L. Bandwidth selection for nonparametric regression with errors-in-variables. Econom Rev. 2023;42(4):393–419.
12. Wu, C. Y., Huang, L. S., & Jin Z. Decomposition and reproducing property of local polynomial equivalent kernels in varying coefficient models. J Nonparametr Stat. 2024;36(2):357–72.
13. Sisheng, L., & Jing Y. Kernel regression for estimating regression function and its derivatives with unknown error correlations: L. Sisheng, Y. Jin. Metrika. 2024;87(1):1–20.
14. Mallows CL. More comments on Cp. Technometrics. 1995;37(4):362–72.
15. Hall P, Marron JS. On variance estimation in nonparametric regression. Biometrika. 1990;415–9.
16. Kong, D., Zhao, S., & Wang W. Nonparametric derivative estimation with bimodal kernels under correlated errors: D. Kong et al. Comput Stat. 2024;39(4):1847–65.