

On Using Multivariate Nonparametric Regression Methods

Ahmed M. Mami*, Maryam M. Mansour, and Abobaker M. Jaber

Department of Statistics, Faculty of Science, University of Benghazi, Benghazi, Libya

DOI: [10.36347/sjpm.2020.v07i02.001](https://doi.org/10.36347/sjpm.2020.v07i02.001)

| Received: 05.02.2020 | Accepted: 14.02.2020 | Published: 23.02.2020

*Corresponding author: Ahmad M.Mami

Abstract

Review Article

Nonparametric regression analysis refers to the absent of the regression estimation of the parameters. There are many drawbacks of using the parametric modeling due to the restrictions imposed on the regression function, as well as the lack of flexibility provided. The answer to the rigidity of parametric regression is to assume that the regression function f does not belong to a parametric family. This will lead us to the nonparametric regression modelling. The scope of nonparametric regression is very broad, ranging from "smoothing" the relationship between two variables in a scatter plot to multivariate regression analysis where we encounter what so called "the curse of dimensionality". This paper aims to propose two methods of the nonparametric regression, namely; the Multivariate Local Polynomial Kernel (MLPK) Regression, and Multivariate Regression Splines (MRS). The theoretical development of these two nonparametric methods, that includes the asymptotic bias, variance, and mean squared errors are presented. Then, data-driven methods of bandwidth selection of the two multivariate nonparametric estimators are derived. Several simulation studies are conducted in order to evaluate and compare the estimated regression curves resulting from utilizing two different classes of the Multivariate Nonparametric regression curve estimation procedures, namely: MLPK estimator, and MRS estimator. Such a study is necessarily to be restrictive, because there are many possibilities for the number of explanatory variables d , the choices of regression functions $f(x)$, sample size (n), and the choice of σ . In this study, we conducted three main simulation studies depending upon $d=1$, $d=2$, and $d=3$. Several interesting results have been achieved. Applications on real dataset also been considered. The dataset used is the Cross Country Growth Panel Data. Lastly, the advantages and disadvantages of these two nonparametric regression estimators have been discussed.

Keywords: Nonparametric Regression, Local polynomial kernel estimator, Smoothing Spline estimator, Multivariate Local Polynomial Kernel (MLPK) Regression, Multivariate Regression Splines (MRS), and GDP growth panel data.

Copyright @ 2020: This is an open-access article distributed under the terms of the Creative Commons Attribution license which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use (NonCommercial, or CC-BY-NC) provided the original author and source are credited.

INTRODUCTION

In statistics, regression analysis examines the causal relation of a dependent variable Y (response variable) with to specified independent variables (explanatory variables). The mathematical model of their relationship is the regression equation. The dependent variable is modeled as a random variable because of uncertainty as to its value, given only the value of each independent variable. A regression equation contains estimates of one or more hypothesized regression parameters. These estimates measure are constructed using data for the variables, such as from a sample. The estimates measure the relationship between the dependent variable and each of the independent variables. They also allow estimating the value of the dependent variable for a given value of each respective independent variable [1].

Uses of regression include curve fitting, prediction, modeling of causal relationships, and testing scientific hypotheses about relationships between variables.

Non-Parametric Regression Models

Nonparametric regression analysis refers to the absent of the regression estimation of the parameters. There are many drawbacks of using the parametric modeling due to the restrictions imposed on the regression function $(\mathbf{x}, \boldsymbol{\theta})$, as well as the lack of flexibility provided. The answer to the rigidity of parametric regression is to assume that the regression function f does not belong to a parametric family. This will lead us to the second kind of the regression modeling, the nonparametric regression models.

The scope of nonparametric regression is very broad, ranging from "smoothing" the relationship between two variables in a scatter plot to multiple-regression analysis and generalized regression models (for example, logistic nonparametric regression for a binary response variable). Unthinkable only a few years ago, methods of nonparametric regression analysis have been used widely because of advances in statistics and computing, and are now a serious alternative to more traditional parametric-regression modeling.

Table-1: Some of the most widely used smoothing methods are listed below:

1-1-Kernel Based Smoothing	2-2-Spline Smoothing	3-3-Wavelet
4-4-K-Nearest Neighbor	5-5-Orthogonal Series Estimators	

Some of these methods are based on quite simple ideas and straightforward applications while others are based on highly sophisticated Mathematics [16]. Of particular interest is the Kernel based smoothing method and the Smoothing Splines because they are mathematically tractable and intuitively simple.

Table-2: Within the context of the Kernel-based smoothing, there are many well-known approaches (estimator) namely:

1-Nadaraya-Watson Estimator	2-Priestley-Chao Estimator	3-Gasser-Muller Estimator
4-LOWESS estimator	5- Local Polynomial Kernel Estimator	

Table-3: On the other hand, within the context of the Cubic Smoothing Splines, there are as well many well-known approaches (estimator) namely:

1- Cubic Smoothing Splines	2- Natural Smoothing Splines	3- B-Smoothing Splines
4- P-Smoothing Splines	5- Thin-plate Smoothing Splines	

This paper aims to provide a broad introduction to two methods of the nonparametric covering the following: introduction to nonparametric regression; Local Polynomial Kernel regression and Smoothing Splines; the statistical inference for nonparametric regression with emphasis on the role of nonparametric regression in data analysis. Then, extend the analysis of these two methods to case of the Multivariate Nonparametric Regression Estimation.

The general nonparametric regression model can be defined as follows

$$y_i = f(x_i) + \varepsilon_i \tag{1}$$

Where $f(x_i)$ represent some function that needed to be estimated, and the ε_i are independent variables with zero mean and constant variance for all $i = 1, 2, \dots, n$.

In the next sub-sections, we will present some important definitions of these two chosen nonparametric methods to smooth the mean regression functions, namely: the Kernel Based Smoothing and the Smoothing Splines.

The Kernel Based Smoothing

The kernel based smoothing method is of particular interest to many statisticians because it is precisely manageable and is based on quite simple idea. The historical development of this method is as follows:

Nadaraya Watson Estimator [2]

The Nadaraya-Watson estimator is independently proposed by (Nadaraya and Watson, 1964) and can be written as

$$\hat{f}(x; h) = \frac{\sum_{i=1}^n k_h(X_i - x) Y_i}{\sum_{i=1}^n k_h(X_i - x)} \tag{2}$$

Priestley-Chao Estimator [3]

The Priestley-Chao estimator (Priestley and Chao, 1972) which has the following form

$$\hat{f}(x; h) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - i/n}{h}\right) Y_i \tag{3}$$

Locally Weighted Scatter Plot Smoother (LOWESS) [4]

The locally Weighted Scatter plot smoother LOWESS (Cleveland [4] is defined as the smoothed point at x_i using locally weighted regression of degree P is (x_i, \hat{y}_i) , where \hat{y}_i is the fitted value of the regression at x_i . Thus

$$\hat{y}_i = \sum_{j=0}^P \hat{B}_j(x_i) x_i^j = \sum_{k=1}^n r_k(x_i) y_k \quad (4)$$

Where $r_k(x_i)$ does not depend on $\hat{y}_i, j = 1, \dots, n$. We have used the notation $r_k(x_i)$ to remind ourselves that these are the coefficients for the y_k that arise from the regression.

Gasser – Muller Estimator (1984)

The Gasser-Muller estimator (Gasser and Muller, 1984) is based on the convolution of a kernel function with a modification of the regress gram, and has the following form

$$\hat{f}(x; h) = \sum_{i=1}^n \int_{s_{i-1}}^{s_i} k_h(u - x) du Y_i \quad (5)$$

Local polynomial kernel Estimator [2]

The Local Polynomial Kernel estimator [5] is defined as

$$\hat{f}(x; p, h) = e_1^T (X_{p,x}^T W_x X)^{-1} (x_{p,x}^T W_x Y) \quad (6)$$

Local polynomial kernel regression estimation was introduced in several different fields in the late nineteenth and early twentieth century's. A comprehensive review of its appealing features and properties are given in [5] in chapter 5, [6] have given a pleasing preliminary exposition [7] have presented an innovative teaching tool which demonstrates local polynomial smoothing via a movie and several new (graphical devices and loader [8] the underlying principle is that a smooth function can be well approximated by a low degree polynomial in the neighborhood of any point Mami [9].

The Smoothing Splines Estimators

A smoother is a tool for summarizing the trend of a response variable y as a function of one or more linear explanatory variables x 's. Since it is an estimate of the trend, it is less variable than y , that is why it is called smoother. There are several types of nonparametric regression, but all of them rely on the data to specify the form of the model. The curve at any point depends only on the observations at that point and some of the specified neighboring points.

The term *spline* originally referred to a tool used by draftsmen to draw curves. For our purposes, splines are piecewise regression functions we constrain to join at points called knots. In their simplest form, splines are regression models with a set of explanatory variables on the right hand side of the model that we use to force the regression line to change direction at some point along the range of x . Within the context of splines, one must choose the degree of a polynomial for the piecewise regression functions, the number of knots, and the location of the knots.

Perhaps the most confusing aspect of splines is that there are so many different types. For example, there are regression splines, cubic splines, B-splines, P-splines, Natural splines, Thin-plate splines, and smoothing splines, etc. Moreover, there are often combinations such as Natural cubic with B-splines. The wide variety of splines partially stems from the progress in research on splines. Often a new type of spline either supplants an older type of spline or adds a refinement to existing methods. The splines most often used in statistics is the smoothing splines. See Keele [10] for more comprehensive details on Splines.

LITERATURE REVIEW

Local Polynomial Regression has been shown to curve estimation an excellent adaptation and reasonably simple implementation. Therefore, it has received considerable attention in the statistics communities.

For multivariate Local Polynomial Regression, the bias-variance tradeoff problem is more complicated. It is because the bandwidth H is a matrix in multidimensional curve estimation. Not only is the kernel scale important, but also the kernel shape has significant effects on the estimation results. It explains why there were not many publications concerning the bandwidth matrix selection of multivariate Local polynomial Regression. Although the adaptive bandwidth selection for univariate Local Polynomial Regression has been systematically and intensively studied (Fan and Gijbels [11]).

Traditionally, the bandwidth matrix selection for multivariate Local Polynomial Regression is simplified by using kernels with a specific shape, which is usually symmetric. More precisely, using a symmetric basis kernel $K(u)$ and a bandwidth matrix $H=h*I$ with h the scale parameter and I the identity matrix, the kernel $K_H(u)$ is simplified into d -

dimension of the multivariate Local Polynomial Regression. As a result, the selection of the scale h is much easier than the selection of a d -dimensional bandwidth matrix \mathbf{H} . With this assumption, Ruppert *et al.* [12] have studied the asymptotic bias, variance, and MSE of multivariate Local Polynomial Regression and developed the empirical-bias bandwidths selection algorithm for multivariate Local Polynomial Regression. Based on Ruppert's work, Fan and Gijbels [11], Yang and Tschernig [13] also developed similar scale selection methods for multivariate local linear regression [14].

The so-called regression spline or the least-squares spline approximates the regression function by piecewise polynomials, each defined over a different sub region of the domain of the spline. Many approaches to constructing regression splines have been presented in the literature. Poirier [16] has introduced the cubic spline and the bilinear spline. Buse and Lim [15] have proved that the cubic spline can be computed by the restricted least-squares method. Afterward, Smith [17] has showed that the univariate regression splines of any degree of regression function may be represented in the additional function. This enables us to compute univariate regression splines simply by the least squares method. As a generalization, the bilinear spline presented by Poirier [16] is a model of continuous piecewise multivariate polynomial function of order two. Eubank [18] has pointed out that multivariate regression splines have received only brief attention and further research is required before multivariate regression splines can become a standard method of multivariate surface fitting.

Most recently Chen [19] has proposed a Multivariate regression spline of arbitrary finite polynomial order by the restricted least-squares method.

Theoretical Development

In the study of nonparametric regression, we are given n -pairs of observations $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, then we can write that

$$y_i = f(x_i) + \hat{a}_i \quad i = 1, 2, \dots, n. \quad (7)$$

where y_i represent the response variables, $f(x_i)$ represent some function, and \hat{a}_i are independent variables with zero mean and constant variance for all $i = 1, 2, \dots, n$.

The idea behind smoothing covers many regression problems. Given noisy data, we desire to extract general structure in the observations without following local fluctuations due to noise. Good smoothing methods strike a balance between error and bias.

In regression smoothing, we typically use the variables h (the bandwidth in Kernel smoothing context) or λ (the measure of roughness in the smoothing Splines context) to quantify the amount of "smoothing".

Depending on how one measures error, different ideal h or λ values emerge. Regardless each idea value for h or λ depends upon the sample size n and specific aspects of the problem. For instance, suppose one measures accuracy via the mean integrated squared error, or

$$E[\int_{-\infty}^{\infty} (\hat{m}_h(x) - m(x))^2 dx]$$

Here $m(t) = E[x(t)]$, and $\hat{m}_h(x)$ is an estimate of $m(t)$ with bandwidth h or measure of roughness λ .

Evaluating this integral and performing a detailed asymptotic analysis shows that as $n \sim \infty$, the best bandwidth satisfies

$$h = K_2^{-2/5} \left(\int_{-\infty}^{\infty} K(t)^2 dt \right)^{1/5} \left(\int_{-\infty}^{\infty} m''(x)^2 dx \right)^{-1/5} n^{-1/5} \quad (8)$$

where K_2 is some constant, and $K(\cdot)$ is the kernel function. We usually simplify the optimal value to $cn^{-1/5}$ where c is a constant.

Other optimal values of h or λ are of a similar fashion [20]. Since the optimal value of h or λ depends on the true mean function $m(\cdot)$, we can never, in truth, know the best value of h or λ . It is not hard, however, to determine bad values for h or λ theoretically [11].

Local Polynomial Kernel Regression

The Local Polynomial Kernel (LPK) estimator which is defined as

$$\hat{f}(x; p, h) = e_1^T (X_{p,x}^T W_x X)^{-1} (x_{p,x}^T W_x Y) \tag{9}$$

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 $(x_i - x)^{j-1}$ (Fan and Gijbels [11], and Mami [9]).

The Definition of the Local polynomial Kernel Estimator

The basic idea of this estimator is the local fitting of P^{th} order polynomial to the data via weighted least squares at a particular point of the regression function.

Suppose a set of multidimensional observations: (Y_i, X_i) , $i=1,2,\dots,n$, where for each i , Y_i is a scalar response variable and $X_i = (X_{i,1}, X_{i,2}, \dots, X_{i,d})$ is a d - dimensional explanatory variable, therefore we have two setting depending on $d = 1$ (univariate case) or $d > 1$ (multivariate case).

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

Let the response variables as being generated from the model which ordered and non-random numbers are having the difference between $x_{i+1} - x_i$ being the same (equally-spaced design).

$$y_i = f(x_i) + \hat{a}_i \quad i = 1, 2, \dots, n. \tag{10}$$

where the errors \hat{a}_i are independent random variables with

$$E(\hat{a}_i) = 0, \quad E(\hat{a}_i^2) = \sigma^2$$

f is called the mean regression function because

$$E(y_i) = f(x_i), \quad var[y_i] = \sigma_i^2$$

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.

The Elements of LPK Estimator

There are three elements that form the LPK estimator:

Firstly, the Kernel function K

The term kernel refers to any smooth function such K that

$k(x) \geq 0$ and satisfies the following conditions:

$$\int K(x) dx = 1, \quad \int xK(x) dx = 0, \quad \delta_K^2 = \int x^2 K(x) dx > 0$$

The usual assumptions that needed in the kernel smoothing techniques for the kernel

K are as follows

- the function f''' is continuous on $[0,1]$
- the kernel K is symmetric about zero and is supported on $[-1,1]$ Moreover; it has a bounded first derivative K' .

There are several choices of kernel functions that can be used such as:

- the Boxcar kernel(uniform): $K(x) = \frac{1}{2} I(x)$
- the Gaussian kernel: $K(x) = \frac{1}{\sqrt{2\delta}} e^{-x^2/2}$
- the Epanechnikov kernel: $K(x) = \frac{3}{4} (1 - x^2) I(x)$
- the Tri cube kernel: $K(x) = \frac{70}{81} (1 - |x|^3)^3 I(x)$

In this work, the principle kernel function to be used is the Gaussian kernel.

Secondly, the order of the polynomial p

Recall (9), the explicit expression for the local polynomial kernel estimator of order p (in the matrix notation) can be seen as:

$$\hat{f}(x; p, h) = e_1^T (X_{p,x}^T W_x X_{p,x})^{-1} (X_{p,x}^T W_x Y) \tag{11}$$

The Nadaraya-Watson estimator can be seen as a special case of LPK estimator when ($p=0$) can be written in a simple formula as:

$$\hat{f}(x; h) = \frac{\sum_{i=1}^n k_h(X_i - x) Y_i}{\sum_{i=1}^n k_h(X_i - x)} \tag{12}$$

Thirdly, the Smoothing Parameter or the Bandwidth h

The right choice of the smoothing parameter or bandwidth h is considered to be the most sensitive matter when using the Local Polynomial Kernel estimator. Thus, taking a very small bandwidth, the modeling bias will be small since the number of data points falling in this local neighborhood is also small but the variance will be large. On other side, taking a very large bandwidth creates a large modeling bias depending on the underlying function. In conclusion, the bandwidth governs the complexity of the model through the trade-off between quantities of bias and variance [11].

Some of the well-known bandwidth selection strategies are:

- Rule of thumb
- Rice’s criterion
- Solve the equation
- Cross-validation criterion (CV)
- Direct plug-in selection strategy (DPI)
- Residual squares criterion (RSC)

The Exact Bias and Variance of LPK Estimator

The bias expression in the case of local linear kernel estimator $\hat{f}(x; 1, h)$ when utilizing the equally-spaced design case can be seen as:

$$E[\hat{f}(x; 1, h)] - f(x) = \frac{1}{2} f''(x) h^2 M_2(x) + O(n^{-1})o(h^2), \tag{13}$$

Where: $M_2(k) = \int_{-1}^1 z^2 k(z) dz$ 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. $\lim_{n \rightarrow \infty} \frac{O(n^{-1})}{n^{-1}} \leq C$); and the notation $o(h^2)$ means that $h \rightarrow 0$ the quantity $O(h^2)$ converges to zero with rate of h^2 (i.e. $\lim_{h \rightarrow 0} \frac{o(h^2)}{h^2} = 0$) for complete details of the derivations see Ruppert, et al. [12].

While, the variance expression in the case of local linear kernel estimator $\hat{f}(x; 1, h)$ when utilizing the equally-spaced design case can be seen as:

$$var[\hat{f}(x; 1, h)] = (nh)^{-1} R(k) v(x) o(n^{-1} h^{-1}) \tag{14}$$

is defined as above moreover, for complete details of the derivation. The appropriate global loss criterion is the mean integrated squared error (MISE) of $\hat{f}(\cdot; 1, h)$, and it is given by :

$MISE[\hat{f}(x; 1, h)]$	$= E \{ \int E \{ \hat{f}(x; 1, h) - f(x) \}^2 dx \}$
	$= \int [n^{-1} h^{-1} R(k) + v(x)] dx +$
	$h^4 \{ m_2(k)/(2!) \}^2 \int m^2 x^2 f(x) dx + op(nh^{-1} + h^4) \}$ (15)

In this work, the selection of h is based on Direct Plug-In strategy. The formula to obtain h_{opt} is simply minimizing MISE [$\hat{f}(x; 1, h)$] with respect with h .

Multivariate Local Polynomial Kernel Regression ($d > 1$)

In Multivariate Local Polynomial Kernel (MLPK) regression context, we have a set of multidimensional observations: (Y_i, X_i) , $i=1,2,\dots,n$, where for each i, Y_i is a scalar response variable and $X_i=(X_{i,1}, X_{i,2}, \dots, X_{i,d})$ is a d -dimensional explanatory variable. We assume the homoscedastic model

$$Y_i = f(X_{i,j}) + \epsilon_{i,j}, \quad \text{such that } i=1,2,\dots,n, j=1,2,\dots,d \quad (16)$$

Where $f(X_{i,j})$ is a smooth function specifying the conditional mean of Y_i given $(X_{i,j}), \epsilon_{i,j}$ is an independent identically normally distributed with zero mean and unit variance.

The goal is to estimate $f(X_{i,j})$ and its partial derivatives from the noisy samples Y_i . Based on the work of Zhang and Chan [14], the smooth function $f(X_{i,j})$ can be approximated locally, say at a given point $X=(X_1, X_2, \dots, X_d)^T$, by a multivariate polynomial of a certain degree

$f(X: x)$	$=f(x) + \{\nabla_f(x)\}^T (X-x) + \frac{1}{2} (X-x)\{H_f(x)\}^T (X-x) + \dots$
	$=f(x) + \{\nabla_f(x)\}^T (X-x) + \frac{1}{2} \text{vec}^T\{H_f(x)\} \cdot \text{vec}\{(X-x)(X-x)^T\} + \dots$
	$=\beta_0 + \beta_V^T \cdot (X-x) + \beta_H^T \cdot \text{vech}\{(X-x)(X-x)^T\} + \dots$ (17)

Where $\beta_0 = m(x)$, $\beta_V = \nabla_m(x) = [\frac{\partial m(x)}{\partial x_1}, \dots, \frac{\partial m(x)}{\partial x_d}]^T$,
 $\beta_H = \text{vech}\{H_f(x)\} = \frac{1}{2} [\frac{\partial^2 m(x)}{\partial x_1^2}, 2 \frac{\partial^2 m(x)}{\partial x_1 \partial x_2}, \dots, \frac{\partial^2 m(x)}{\partial x_d^2}]^T$,

$X=(X_1, X_2, \dots, X_d)^T$ is neighborhood of x , ∇^f is the $d \times 1$ gradient operator, H_f is the $d \times d$ Hessian matrix of $f(x)$, $\text{vec}(\cdot)$ is the vectorization operator which converts a matrix into a column vector lexicographically, and $\text{vech}(\cdot)$ is the half-vectorization operator which converts, for example

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ into } \text{vech}(A) = [a_{11}, a_{21}, a_{12}, a_{22}]^T.$$

Since $\epsilon_{i,j}$ is *i.i.d* and normally distributed, the maximum likelihood (ML) estimation of the coefficient vector β at location x can be obtained by solving the following weighted least-squares (WLS) regression problem

$$\hat{\alpha}(x; H) = \arg \min_{\beta} \{J(x; H)\} \quad (18)$$

Where

$$J(x; H) = \sum_{i=1}^n k_H(X_i - x) \left[Y_i - \sum_{k=0}^p \sum_{k_1+\dots+k_d=k} \beta_{k_1 \dots k_d} \prod_{j=1}^d (X_{i,j} - x_j)^{k_j} \right]^2 \quad (19)$$

is a least-squares cost function and $K_H(X_i - x) = \frac{1}{H} k(H^{-1}(X_i - x))$, is a d -variate non-negative kernel function which gives emphasis to neighboring observations around x in estimating $\beta(x)$. The bandwidth matrix H determines the weights of neighboring samples around x to be used in estimating $\beta(x)$. For separable windows, we have

$$K_h(X_i - x) = \prod_{j=1}^d K_{h_j}(X_i - x) = \frac{1}{h_1 \dots h_d} \prod_{j=1}^d K_{h_j} \left(\frac{1}{h_j} (X_{i,j} - x_j) \right) \quad (20)$$

To facilitate the development of the MLPK estimator, we assume that the basis-kernel $K(u)$ is known a priori. The WLS solution to (18) is:

$$\hat{\alpha}(x; h) = (X^T W X)^{-1} X^T W Y, \quad (21)$$

Where $\mathbb{X} = \begin{bmatrix} 1 & (x_1 - x)^t \text{vech}\{(x_1 - x)(x_1 - x)^t\} & \dots \\ 1 & (x_2 - x)^t \text{vech}\{(x_2 - x)(x_2 - x)^t\} & \dots \\ 1 & (x_n - x)^t \text{vech}\{(x_n - x)(x_2 - x)^t\} & \dots \end{bmatrix}$ and $Y = [y_1, y_2, \dots, y_n]^T$,

$W = \text{diag}\{k_h(X_1 - x), \dots, k_h(X_n - x)\}$, $k_h(x_n - x)$ is the weighting matrix.

Recall the key problem in **LPK** regression, is the appropriate selection of the scale parameter h . Therefore, you do need to achieve the best bias-variance tradeoff in estimating $m^{(k)}(x)$. Now, we need to examine the analytical expressions for the conditional bias, variance and MSE of the polynomial coefficient estimate \hat{a} .

The asymptotic bias and variance of the k -th partial derivative $\hat{m}^{(k)}(x;h)$ are

$$\text{Bias}\{\hat{m}^{(k)}(x;h)\} = \sqrt{B}h^{p+1-k} \tag{22}$$

$$\text{Var}\{\hat{m}^{(k)}(x;h)\} = Vh^{-(d+2k)} \tag{23}$$

Where $B = \left(k! \left(M^{-1} B m_{p+1}(X)\right)_k\right)^2$, $V = (K!)^2 \phi^2(x) (M^{-1} \Gamma M^{-1})_{k,k} / n f(x)$,

$B = [M_{0,p+1}, M_{1,p+1}, \dots, M_{p,p+1}]^T$, $M_{p+1}(x) = \text{vech}(\hat{m}^{(k)}(x))$ is the regression result obtained with a higher order $(p+1)$, $(M^{-1} \Gamma M^{-1})_{k,k}$ is the (K, K) diagonal element of the constant matrix $M^{-1} \Gamma M^{-1}$, $(M^{-1} B m_{p+1}(x))_k$ is the K -th element of the vector $M^{-1} B m_{p+1}(x)$, and the (i, j) th element of the matrices M and Γ are $M_{i,j} = \int u^{(i+j)} K(u) du$ and $\Gamma_{(i,j)} = \int u^{(i+j)} K^2(u) du$ Respectively.

The asymptotic analytical results on bias and variance of MLPK estimator can be found in Zhang and Chan [14], Rupert and Ward [12], and Fan and Gijbels [11].

$\text{MSE}\{\hat{m}^{(k)}(x;h)\}$	$= \text{Bias}^2\{\hat{m}^{(k)}(x;h)\} + \text{var}\{\hat{m}^{(k)}(x;h)\}$	
	$= B h^{2(p+1-k)} + V h^{-(d+2k)}$	(24)

By minimizing (24), an analytical formula for $h^{opt}(x)$ can be determined as follows:

$$h^{opt}(x) = \left(\frac{(2k+d)v}{2(p-k+1)B}\right)^{\frac{1}{d+2p+1}}$$

The Smoothing Splines

Traditionally, the term Spline referred to a tool used by draftsmen to draw curves. In this framework, Splines are piecewise regression functions that are constrain to joint at points called “knots”. In simpler form, Splines are regression models with a set of explanatory variables. Within the context of splines, one must choose:

- The degree of a polynomial for the piecewise regression functions
- The number of knots
- The location of the knots

Within the framework of Splines, there are too many different types. For example, there are Regression Splines, Cubic Splines, B-splines, P-splines, Natural Splines, Thin-plate Splines, and Smoothing Splines,.....etc (Keele [10] has provided comprehensive reviews and extensive details on the various Smoothing Splines types).

The Definition of the Cubic Smoothing Spline

The basic idea of the Cubic Smoothing Spline was developed using the following process: *a series of unique cubic polynomials are fitted between each of the data points, with the stipulation that the curve obtained be continuous and appear smooth*. These Cubic Smoothing Splines can then be used to determine rates of change and cumulative change over the interval $[a, b]$.

Cubic Smoothing Spline Estimator ($d=1$)

Recall the general nonparametric regression model, that is

$$y_i = f(x_i) + \epsilon_i \quad i = 1, 2, \dots, n. \tag{25}$$

If f is allowed to be any curve then this distance measure can be reduced to zero by any f that interpolates the data. Such a curve would not be satisfactory on the grounds that it is not unique and that it is too wiggly for a structure-oriented interpretation. The Cubic Smoothing Spline approach stays away from this doubtful interpolation of the data by computing the competition between the goal to produce a good fit to the data and the aim to produce a curve without too much rapid local variation [2].

There are severing always to compute local variation. One is to define measures of roughness based, for instance, on the first, second, and so forth derivative. In order to explain the main idea, the integrated squared second derivative is most convenient, that is, the penalty of roughness.

$$\int (f''(x))^2 dx$$

is used here to compute local variation. Using this measure, the weighted sum defined as follows:

$$ss(f, \lambda) = \sum_{i=1}^n \{y_i - f(x_i)\}^2 + \lambda \int_a^b f''(x)^2 dx \tag{26}$$

Where λ denotes a smoothing parameter. The smoothing parameter λ represents the rate of exchange between residual error and roughness of the curve f .

The parameter λ in (26) governs the tradeoff between smoothness and goodness of fit. When λ is large a premium is being placed on smoothness and potential estimators with large derivatives are penalized. The limiting case of $\lambda = \infty$ produces an order polynomial regression fit to the data. Conversely, a small value of corresponds to more emphasis on goodness of it with $\lambda = 0$.

In general, as λ increases, we get a smoother fit to the data, but also again perhaps a biased fit. As the parameter decreases, we get a very good fit that has increased of variance.

For Cubic Smoothing Splines estimator, knot location is not a big matter anymore. Once the penalty is introduced, the number of knots and their location no longer has much influence on the smoothness of the fit. For the Cubic Smoothing Splines estimator, knots are placed at all unique values of x_i .

One might think that having that many knots would mean that we have too many parameters, but the penalty term ensures that the spline coefficients are shrunk towards linearity thus limiting the approximate degrees of freedom [12].

Selecting the Smoothing Parameter

Several data-driven methods have been successfully used in practice. The following two methods, namely: the Cross Validation (CV), and the Generalized Cross Validation (GCV) are frequently used. First, we rewrite the penalized spline model in (26) in matrix form. Given the correspondence between splines models and linear regression models, we can write the first term in (26) as linear regression model in matrix form. It can be shown that the penalty term from (26) can be written as a quadratic form in β [12].

It allows writing the penalty term in matrix form as

$$\int_a^b f''(x)^2 dx = \beta'D\beta \tag{27}$$

where D is a matrix of the following form

$$D = \begin{bmatrix} 0_{2 \times 2} & 0_{2 \times k} \\ 0_{k \times 2} & I_{k \times k} \end{bmatrix}_{(k+2) \times (k+2)} \tag{28}$$

$$\beta = [\beta_0, \beta_1, \beta_{11}, \dots, \beta_{1k}] \tag{29}$$

In the above matrix k denotes the number of knots. The penalized spline regression model in (29) can be expressed in matrix notation as follows:

$$ss(f, \lambda) = \|y - X\beta\|^2 + \lambda\beta'D\beta \tag{30}$$

As stated above, there are two procedures to select the optimal value of lambda λ , namely: the Cross-Validation (CV) and Generalized Cross-Validation (GCV).

In this study, we deal with the cross validation method (CV), as a performance criterion for selecting the smoothing parameter. Using such a criterion, we implement a CV-based algorithm, in *Rgui* (Open Source Statistical Package) by using *ssanova* function in *gss* package. Then, we apply it on the selected example regression function. In order to emphasize the quality of the fitting by the CV-smoothing spline functions. Then, we fit some real data with this kind of function. Such estimator will minimize (30). First part of this expression represents the goodness-of-fit to the data and the second part represents the smoothness of the estimator.

Multivariate regression splines ($d > 1$)

Let $x_1, \dots, x_p, p > 1$, be explanatory variables that are related to the response variable y . Consider the joint-points $\{dig : s_i = 0, 1, \dots, a_i\}$, with $a_i \geq 1, i = 1, \dots, p$, the elements of which are known as knots. These joint-points define a p -dimensional rectangular grid in the space of variables x_1, \dots, x_p consisting of a_1, \dots, a_p multivariate rectangles as the region set:

$$\{(x_1, \dots, x_p) : \delta_{s_i}^i < x_i \leq \delta_{s_{i+1}}^i, i = 1, \dots, p\}, S_i = 0, 1, \dots, a_i, i = 1, \dots, p \tag{31}$$

Here $\delta_0^i, \delta_{a_i+1}^i$ are set such that the region $\{(x_1, \dots, x_p) : \delta_0^i < x_i < \delta_{a_i+1}^i, i = 1, \dots, p\}$ is the bounded domain of the regression function. The regression function Has the following form:

$$\sum_{c_1 + \dots + c_p = 0}^k \beta_{c_1 \dots c_p} \prod_{i=1}^p x_i^{c_i} \tag{32}$$

Where c_1, \dots, c_p are nonnegative integers and \mathcal{D} denotes for product of polynomial functions. In the following, we state the setting of the piecewise multivariate polynomial regression model.

Definition 2.1. It is based on the work of Chen (1997), and stated as:

(a) An order k piecewise multivariate polynomial function is a function defined on the region of (31) such that each multivariate rectangle $\{(x_1, \dots, x_p) : \delta_{s_{i-1}}^i < x_i \leq \delta_{s_i}^i, i = 1, \dots, p\}$ has an order k multivariate polynomial defined on it.

(b) An order k piecewise multivariate polynomial regression model is the following:

$$y = f(x_1, \dots, x_p) + \varepsilon \tag{33}$$

for which f is an order k piecewise multivariate polynomial function and ε is the error term.

The class of order k piecewise multivariate polynomial function is a vector space with dimension $\binom{p+k}{p} \prod_{i=1}^p (a_i + 1)$

.The e multivariate regression spline is then a model (33) for which the regression function f satisfies some continuity condition of partial derivatives.

Definition 2.2. It is based on the work of Chen [19], and stated as:

(a) An order k piecewise multivariate polynomial function of model (25) is said to be a multivariate regression spline if it has some continuous partial derivatives.

(b) A multivariate regression spline is said to have continuity-degree $j, 0 \leq j \leq k - 1$ if its (j_1, \dots, j_p) -th partial derivative is continuous for $0 \leq j_1 + \dots + j_p \leq j$

We now state an equivalent form of the order k piecewise multivariate polynomial function, which generalizes the formulation of the piecewise univariate polynomial function to the multivariate case. The order k piecewise multivariate polynomial function can be formulated in the form,

$$\begin{aligned}
 & P + \sum_{s=1}^p \sum_{t_s}^{a_s} p^{t_s} I(x_s > \delta_{t_s}^s) \\
 & \sum_{1 \leq s_1 < s_2 \leq p} \sum_{t_{s_2}=1}^{a_{s_2}} \sum_{t_{s_1}=1}^{a_{s_1}} p^{t_{s_1} t_{s_2}} I(x_{s_1} > \delta_{t_{s_1}}^{s_1}, x_{s_2} > \delta_{t_{s_2}}^{s_2}) \\
 & + \sum_{t_p=1}^{a_p} \sum_{t_1=1}^{a_s} p^{t_1 \dots t_p} I(x_1 > \delta_{t_1}^1, \dots, x_p > \delta_{t_p}^p), \tag{34}
 \end{aligned}$$

Where P are multivariate polynomials with the form of (34).

Denote the additional function “+” by $x_+ = \max\{x, 0\}$. The next result provides a basis based on additional function for the space of order k piecewise multivariate polynomial functions that guides an easy way to construct multivariate regression splines. The following two theorems are based on the work of Chen [19].

Theorem 2.1. The following set is a basis for the space of order k piecewise multivariate polynomial functions,

$$\prod_{i=1}^p x_i^{c_i}, \prod_{i \neq s} x_i^{c_i} (x_s - \delta_{t_s}^s)^{c_s} +, \prod_{i \neq s, s_2} x_i^{c_i} (x_{s_1} - \delta_{t_{s_1}}^{s_1})^{c_{s_1}} + (x_{s_2} - \delta_{t_{s_2}}^{s_2})^{c_{s_2}} +, \dots, \prod_{i=1}^p (x_i - \delta_{t_i}^i)^{c_i} +, \tag{35}$$

where nonnegative integers c_1, \dots, c_p satisfy that $0 \leq \sum_{i=1}^p c_i \leq k$, and domain of t_i is $1, \dots, a_i$ for $i=1, \dots, p, s, s_1$ and s_2, \dots .

We can easily check that the number of basic functions above is exactly the necessary integer $\binom{p+k}{p} \prod_{i=1}^p (a_i + 1)$. With

Theorem 2.1 emphasizes that any linear combination of the functions in the basis of (34) provides some kind of multivariate regression spline. However, the interesting regression spline might be the one with continuity degree $j \in \{0, 1, k - 1\}$. We now state it in the following theorem.

Theorem 2.2. Let $b = \min\{b^*, p\}$, where $b^* = \max\{b_0: b_0(j+1) \leq k, b_0 \text{ is integer}\}$.

The set of the following functions is a basis of the space of order k multivariate regression splines with continuity-degree j

$$\prod_{i=1}^p x_i^{c_i}, (x_s - \delta_{t_s}^s)^{c_s} + \prod_{i \neq s} x_i^{c_i}, (x_{s_1} - \delta_{t_{s_1}}^{s_1})^{c_{s_1}} + (x_{s_2} - \delta_{t_{s_2}}^{s_2})^{c_{s_2}} + \prod_{i \neq s_1, s_2} x_i^{c_i}, \dots, \prod_{i=1}^b (x_{s_i} - \delta_{t_{s_i}}^{s_i})^{c_{s_i}} + \prod_{i \neq s, \dots, s_b} x_i^{c_i}, \tag{36}$$

where c_s, c_{s_i} are restricted on the set $\{j + 1, j + 2, \dots, k\}$ and c_1, \dots, c_p satisfy $\sum_{i=1}^p c_i \leq k$. also, domain of t_i is $1, \dots, a_i$.

In the particular case where $j = k - 1$, the smoothest multivariate regression spline might be of great interest.

The smoothest multivariate regression spline can be formulated as

$$y = \sum_{c_1 + \dots + c_p = 0}^k \beta_{c_1 \dots c_p} \prod_{i=1}^p x_i^{c_i} + \sum_{s=1}^p \sum_{t_s=1}^{a_s} \beta_k^{t_s} (x_s - \delta_{t_s}^s)^k + \varepsilon \tag{37}$$

The Simulation Study

The simulation study is conducted in order to evaluate and compare the estimated regression curves resulting from utilizing two different classes of the nonparametric regression curve estimation procedures, namely: the Local polynomial kernel estimators and the Splines estimators. Such a study is necessarily to be restrictive, because there are many possibilities for the number of explanatory variables d , the choices of regression function $f(x)$, sample size (n) , and the choice of σ . In this study, we conducted two main simulation studies depending upon $d=2$, and $d=3$.

Description of the Experiment in the case of $d = 2$

We have conducted another 18 simulation studies, using the same three different example regression functions as shown in Table 1. We utilized three selected methods of estimation (Nadaraya-Watson estimator (**NW**), Local Linear Kernel estimator (**LLK**) and Cubic Smoothing Spline estimator (**CSS**), three different sample sizes n , and two different values of σ 's. Each simulation study involves $L=500$ repetitions. The shape of the used three example regression functions graphed in Figure 1. It is worth noting that $f_5(x_1, x_2)$ and $f_6(x_1, x_2)$ are unimodal shaped, whereas $f_7(x_1, x_2)$ is bimodal shaped. We have also chosen three sample sizes of $n=50$, $n=100$ and $n=300$, two levels of σ 's, and the x 's are being generated using the uniform distribution on the interval $[a, b]$.

Table-1: The Example Regression Functions used in the 1st simulation study

Ex	$f_i(x_1, x_2)$	Sigma	Interval
1)	$f_5(x_1, x_2) = \pi^{(0.2)(0.4)} * \left[(1.2)e^{\frac{-(x_1-0.2)^2}{(0.2)^2} - \frac{(x_2-0.3)^2}{(0.4)^2}} \right]$	$\sigma_1 = 0.05$ $\sigma_2 = 0.10$	$x \in U[0,1]$
2)	$f_6(x_1, x_2) = \pi^{(0.2)(0.4)} * \left[(0.8)e^{\frac{-(x_1-0.7)^2}{(0.2)^2} - \frac{(x_2-0.8)^2}{(0.4)^2}} \right]$	$\sigma_1 = 0.05$ $\sigma_2 = 0.10$	$x \in U[0,1]$
3)	$f_7(x_1, x_2) = \pi^{(0.2)(0.4)} * \left[(1.2)e^{\frac{-(x_1-0.2)^2}{(0.2)^2} - \frac{(x_2-0.3)^2}{(0.4)^2}} + (0.8)e^{\frac{-(x_1-0.7)^2}{(0.2)^2} - \frac{(x_2-0.8)^2}{(0.4)^2}} \right]$	$\sigma_1 = 0.05$ $\sigma_2 = 0.10$	$x \in U[0,1]$

Numerical Summary for the 1st simulation study

Having conducted the simulation runs using the $L = 500$ simulated data sets, we obtained the results that are tabulated in Tables (2), (3), and (4). Once more, these tables display the numerical summary of results obtained using three curve estimation procedures (Nadaraya-Watson estimator (**NW**), Local Linear Kernel estimator (**LLK**) and Cubic Smoothing Spline estimator (**CSS**)). Moreover, two different values of σ 's, with three different sample sizes of $n=50$, $n=100$, and $n=300$. Once more, the numerical summary includes three statistical properties, namely the Average Square Bias (ASB), the Average Variance (AVAR) and the Average Mean Square Errors (AMSE), as well as the mean of the coefficient of determination R^2 with its standard deviation.

The formulae that have been used to calculate ASB, AVAR, and AMSE are as following:

$$ABS(\hat{f}) = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{L} \sum_{j=1}^L \hat{f}(x_{ij}) - f(x_i) \right]^2 \tag{3.1}$$

$$AVAR(\hat{f}) = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{L} \sum_{j=1}^L \hat{f}^2(x_{ij}) - \left(\frac{1}{L} \sum_{j=1}^L \hat{f}(x_{ij}) \right)^2 \right] \tag{3.2}$$

$$AMSE(\hat{f}) = ABS(\hat{f}) + AVAR(\hat{f}) \tag{3.3}$$

$$mean(R^2) = \left[\frac{1}{L} \sum_{j=1}^L \left(1 - \frac{sse_j}{sst_j} \right) \right] \tag{3.4}$$

where $sse_j = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ $sst_j = \sum_{i=1}^n (y_i - \bar{y})^2$

The main program (in **Rgui** code) used to calculate these are available upon request from auth

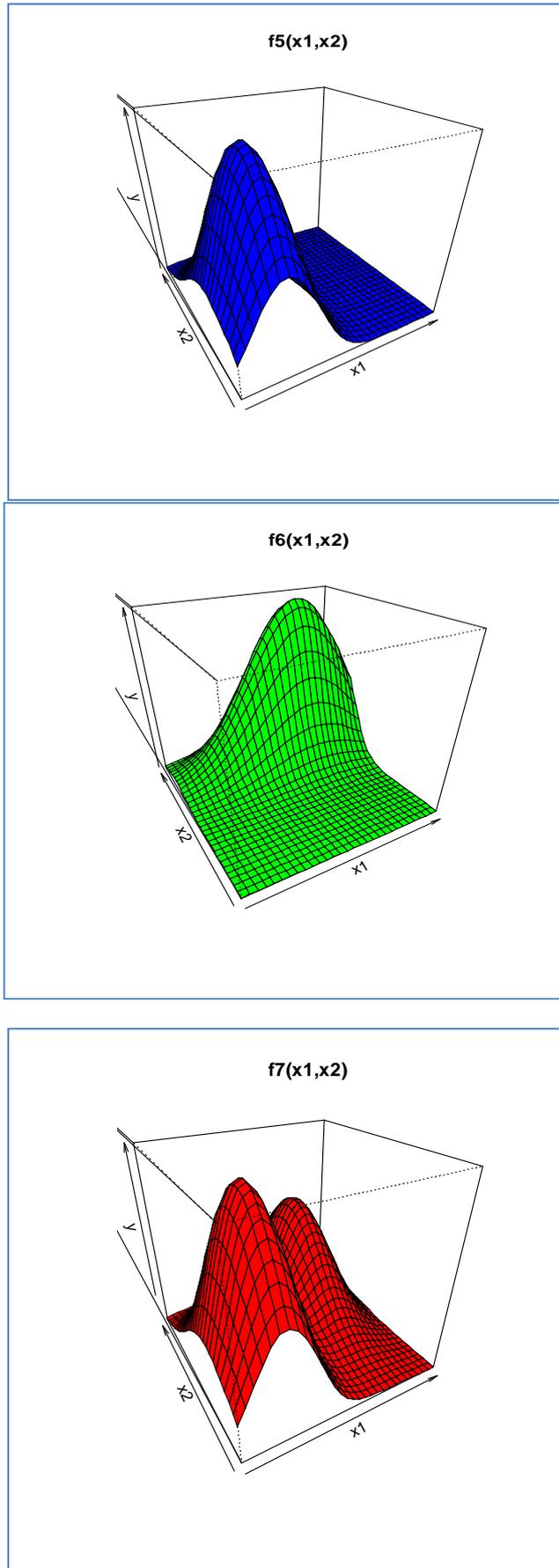


Fig-1: Represents the true surfaces used in the 1st simulation study

Table-2: The numerical summary of results obtained using three different estimation methods, (NW), (LLK) and (CSS) estimators from the 1st simulation study using $f_5(x_1, x_2)$.

		NW estimator		LLK estimator		CSS estimator	
		$\sigma = 0.1$	$\sigma = 0.05$	$\sigma = 0.1$	$\sigma = 0.05$	$\sigma = 0.1$	$\sigma = 0.05$
ASB	$n = 50$	0.0034327	0.0019683	0.0042990	0.0011476	0.0035546	0.0026029
	$n = 100$	0.0013505	0.0013356	0.0009715	0.0008091	0.0012545	0.0007123
	$n = 300$	0.0008206	0.0004848	0.0005508	0.0002484	0.0006540	0.0002599
AVAR	$n = 50$	0.0016682	0.0007038	0.0028042	0.0017106	0.0018184	0.0006922
	$n = 100$	0.0013483	0.0004790	0.0027283	0.0010386	0.0014717	0.0006079
	$n = 300$	0.0006398	0.0002559	0.0009874	0.0004314	0.0007157	0.0003354
AMSE	$n = 50$	0.0051009	0.0026721	0.0071033	0.0028582	0.0053730	0.0032951
	$n = 100$	0.0026989	0.0018147	0.0036999	0.0018478	0.0027262	0.0013202
	$n = 300$	0.0014602	0.0007408	0.0015382	0.0006798	0.0013697	0.0005953
Mean(R^2) (Sd)	$n = 50$	0.9060908 (0.03132)	0.9538316 (0.0095932)	0.8735855 (0.05027)	0.9478062 (0.01857)	0.9036228 (0.03915)	0.945385 (0.01310)
	$n = 100$	0.9599429 (0.01226)	0.9660516 (0.0066339)	0.9446779 (0.01859)	0.9644602 (0.01031)	0.9596759 (0.01189)	0.9766236 (0.00869)
	$n = 300$	0.9748263 (0.0057079)	0.9861806 (0.0024882)	0.9728196 (0.00660)	0.9869721 (0.00314)	0.976949 (0.00783)	0.9891441 (0.00428)

Table-3: The numerical summary of results obtained using three different estimation methods, (NW), (LLK) and (CSS) estimators from the 1st simulation study using $f_6(x_1, x_2)$.

		NW estimator		LL estimator		CSS estimator	
		$\sigma = 0.1$	$\sigma = 0.05$	$\sigma = 0.1$	$\sigma = 0.05$	$\sigma = 0.1$	$\sigma = 0.05$
ASB	$n = 50$	0.0011293	0.0006789	0.0010544	0.0008776	0.0015724	0.0010294
	$n = 100$	0.0007791	0.0005187	0.0006924	0.0003783	0.0010779	0.0006509
	$n = 300$	0.0004112	0.0001357	0.0003126	0.0001124	0.0004593	0.0001136
AVAR	$n = 50$	0.0011627	0.0004510	0.0023501	0.0006464	0.0010049	0.0004392
	$n = 100$	0.0008580	0.0003368	0.0010442	0.0005173	0.0006329	0.0003250
	$n = 300$	0.0004041	0.0001502	0.0006082	0.0002077	0.0003707	0.0001478
AMSE	$n = 50$	0.0022919	0.0011299	0.0034047	0.0015239	0.0025773	0.0014687
	$n = 100$	0.0016372	0.0008553	0.0017367	0.0008957	0.0017109	0.0009760
	$n = 300$	0.0008154	0.0002859	0.0009208	0.0003202	0.0008300	0.0002613
Mean(R^2) (Sd)	$n = 50$	0.8672289 (0.06552)	0.8964127 (0.03720)	0.830818 (0.07776)	0.8638513 (0.04469)	0.8565963 (0.05303)	0.8710634 (0.04541)
	$n = 100$	0.8999951 (0.04289)	0.9264813 (0.02710)	0.8975129 (0.03892)	0.9183678 (0.02671)	0.8973814 (0.03817)	0.914787 (0.03141)
	$n = 300$	0.9571538 (0.01252)	0.9747182 (0.00631)	0.9511708 (0.01483)	0.9708111 (0.00821)	0.9566303 (0.01470)	0.9773587 (0.00709)

Table-4: The numerical summary of results obtained using three different estimation methods, (NW), (LLK) and (CSS) estimators from the 1st simulation study using $f_7(x_1, x_2)$.

		NW estimator		LL estimator		CSS estimator	
		$\sigma = 0.1$	$\sigma = 0.05$	$\sigma = 0.1$	$\sigma = 0.05$	$\sigma = 0.1$	$\sigma = 0.05$
ASB	$n = 50$	0.004624761	0.00250646	0.0052698	0.00231440	0.00580817	0.00202148
	$n = 100$	0.002333852	0.00130354	0.0020010	0.00099558	0.00246119	0.00098441
	$n = 300$	0.000845883	0.00059507	0.0006078	0.00027176	0.00068653	0.00026869
AVAR	$n = 50$	0.0018892	0.0006123	0.0026263	0.0024727	0.0020046	0.0007910
	$n = 100$	0.0013124	0.0004498	0.0018887	0.0006766	0.0015039	0.0006679
	$n = 300$	0.0007670	0.0002914	0.0012911	0.0004659	0.0009010	0.0004093
AMSE	$n = 50$	0.006514022	0.00311879	0.0078960	0.00478705	0.00781273	0.00281258
	$n = 100$	0.003646274	0.00175327	0.0038898	0.00167236	0.00396516	0.00165239
	$n = 300$	0.001612944	0.00088649	0.0018989	0.00073758	0.00158756	0.00067799
Mean(R^2) (Sd)	$n = 50$	0.8742632 (0.03807)	0.9201824 (0.0192951)	0.868596 (0.03688)	0.8864845 (0.075616)	0.858796 (0.041657)	0.9336696 (0.0208229)
	$n = 100$	0.9342583 (0.01714565)	0.964989 (0.0056040)	0.9275807 (0.01952)	0.9661041 (0.007302)	0.9290936 (0.021679)	0.9665886 (0.0108449)
	$n = 300$	0.968159 (0.00762577)	0.9825549 (0.0028068)	0.962918 (0.00996)	0.9853254 (0.002699)	0.9694617 (0.009333)	0.9869623 (0.0044529)

The Discussion of the 1stSimulation Study

The numerical results of the first experiment have three common points:

- All the estimated values of the statistical aspects such as ASB, AVAR, AMSE, and R^2 increase as a choice of σ increments from 0.05 to 0.1 regardless of the estimator employed, the **NW**, **LLK**, or **CSS** estimators.
- By increasing the sample size from 50, 100 to 300, all estimates ASB, AVAR, AMSE, R^2 are steadily decrease regardless of the type of estimation method.
- The **CSS** estimator has provided smaller estimated values than the other estimated values obtained through utilizing the **NW**, and **LLK** estimators regardless of the choice of the value σ or sample size n .

Therefore, the CSS estimator is superior because of the following main reasons:

- Estimated **NW** and **LLK** estimators take into account the two explanatory variables in the local parts. So we gain further reduction in the term bias only.
- Whereas, in the **CSS** estimator, we consider the two explanatory variables linear in some parts of the regression curve and non-linear in other parts of the curve. Thus, we get further reduction in the variance term from the linear part and get further reduction in the bias term from the nonlinear part.
- The Smoothing spline method gives the best results in all example regression functions used when the number of explanatory variable $p = 2$.

Description of the Experiment in the case of $d = 3$

Once More, we have conducted another 18 simulation studies, using modified versions from the three different example regression functions used earlier as shown in Table 5. We utilized three selected methods of estimation (Nadaraya-Watson estimator (**NW**), Local Linear Kernel estimator (**LLK**) and Cubic Smoothing Spline estimator (**CSS**), three different sample sizes n , and two different values of σ 's. Each simulation study involves $L=500$ repetitions. We have also chosen three sample sizes of $n=50$, $n=100$ and $n=300$, two levels of σ 's, and the x 's are being generated using the uniform distribution on the interval $[a, b]$.

Table-5: The Example Regression Functions used in the 2nd simulation study

Ex	$f_i(x_1, x_2, x_3)$	Sigma	Interval
4)	$f_8(x_1, x_2, x_3) = \pi^{(0.2)(0.4)(0.6)} * \left[(1.2)e^{-\frac{(x_1-0.2)^2}{(0.2)^2} - \frac{(x_2-0.3)^2}{(0.4)^2} - \frac{(x_3-0.4)^2}{(0.6)^2}} \right]$	$\sigma_1 = 0.10$ $\sigma_2 = 0.20$	$x \in U[0,1]$
5)	$f_9(x_1, x_2, x_3) = \pi^{(0.2)(0.4)(0.6)} * \left[(0.8)e^{-\frac{(x_1-0.2)^2}{(0.2)^2} - \frac{(x_2-0.3)^2}{(0.4)^2} - \frac{(x_3-0.4)^2}{(0.6)^2}} \right]$	$\sigma_1 = 0.10$ $\sigma_2 = 0.20$	$x \in U[0,1]$
6)	$f_{10}(x_1, x_2, x_3) = \pi^{(0.2)(0.4)(0.6)} * \left[(1.2)e^{-\frac{(x_1-0.2)^2}{(0.2)^2} - \frac{(x_2-0.3)^2}{(0.4)^2} - \frac{(x_3-0.4)^2}{(0.6)^2}} \right] + \left[(0.8)e^{-\frac{(x_1-0.2)^2}{(0.2)^2} - \frac{(x_2-0.3)^2}{(0.4)^2} - \frac{(x_3-0.4)^2}{(0.6)^2}} \right]$	$\sigma_1 = 0.10$ $\sigma_2 = 0.20$	$x \in U[0,1]$

Numerical Summary for the 2nd simulation study

Again, having conducted the simulation runs using the $L = 500$ simulated data sets, we obtained the results that are tabulated in Tables (5), (6), and (7). Once more, these tables display the numerical summary of results obtained using three curve estimation procedures (Nadaraya-Watson estimator (NW), Local Linear Kernel estimator (LLK) and Cubic Smoothing Spline estimator (CSS)). Moreover, two different values of σ 's, with three different sample sizes of $n=50$, $n=100$, and $n=300$. Once more, the numerical summary includes three statistical properties, namely the Average Square Bias (ASB), the Average Variance (AVAR) and the Average Mean Square Errors (AMSE), as well as the mean of the coefficient of determination R^2 with its standard deviation.

Table-6: The numerical summary of results obtained using three different estimation methods, (NW), (LLK) and (CSS) estimators from the 2nd simulation study using $f_8(x_1, x_2, x_3)$.

		NW estimator		LLK estimator		CSS estimator	
		$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.1$	$\sigma = 0.2$
ASB	$n = 50$	0.0016288	0.0026226	0.0017298	0.0023301	0.0020659	0.0028168
	$n = 100$	0.0008892	0.0016231	0.0014483	0.0020050	0.0017363	0.0025630
	$n = 300$	0.0006213	0.0009789	0.0009026	0.0014199	0.0010229	0.0019271
AVAR	$n = 50$	0.0012809	0.0034548	0.0011598	0.0040723	0.0008489	0.0032392
	$n = 100$	0.0007977	0.0027076	0.0007781	0.0026064	0.0005127	0.0017163
	$n = 300$	0.0004771	0.0013664	0.0006675	0.0014769	0.0003966	0.0007769
AMSE	$n = 50$	0.0029099	0.0060772	0.0028896	0.0064024	0.0029148	0.0060560
	$n = 100$	0.0016869	0.0043308	0.0022264	0.0046115	0.0022490	0.0042793
	$n = 300$	0.0010986	0.0023452	0.0015702	0.0028969	0.0014194	0.0027040
Mean(R^2) (Sd)	$n = 50$	0.6886992 (0.12609)	0.8229871 (0.10011)	0.6954374 (0.11239)	0.8160735 (0.09977)	0.7177756 (0.08575)	0.8401811 (0.07108)
	$n = 100$	0.8172462 (0.06652)	0.8780366 (0.05873)	0.7821664 (0.05268)	0.8743705 (0.05454)	0.7920935 (0.04267)	0.8908206 (0.03472)
	$n = 300$	0.8815194 (0.02359)	0.937047 (0.02166)	0.8431666 (0.03188)	0.9248905 (0.02068)	0.8618307 (0.02616)	0.9326753 (0.01406)

Table-7: The numerical summary of results obtained using three different estimation methods, (NW), (LLK) and (CSS) estimators from the 2nd simulation study using $f_9(x_1, x_2, x_3)$.

		NW estimator		LLK estimator		CSS estimator	
		$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.1$	$\sigma = 0.2$
ASB	$n = 50$	0.0010729	0.0013059	0.0012817	0.0010347	0.0014929	0.0011749
	$n = 100$	0.0006250	0.0014087	0.0007746	0.0015939	0.0010692	0.0018170
	$n = 300$	0.0003194	0.0006786	0.0004401	0.0009867	0.0007719	0.0012963
AVAR	$n = 50$	0.0011757	0.0025414	0.0010891	0.0038460	0.0008100	0.0030809
	$n = 100$	0.0007904	0.0022236	0.0008324	0.0024173	0.0004594	0.0017459
	$n = 300$	0.0003690	0.0013154	0.0005458	0.0012443	0.0003142	0.0006534
AMSE	$n = 50$	0.0022487	0.0038473	0.0023709	0.0048808	0.0023029	0.0042557
	$n = 100$	0.0014154	0.0036321	0.0016070	0.0040111	0.0015288	0.0035630
	$n = 300$	0.0006886	0.0019940	0.0009859	0.0022309	0.0010861	0.0019498
Mean(R^2) (Sd)	$n = 50$	0.7441337 (0.12758)	0.8798543 (0.09894)	0.1275828 (0.10389)	0.8456454 (0.15237)	0.7713977 (0.07714)	0.8803781 (0.06739)
	$n = 100$	0.8444937 (0.06781)	0.8999332 (0.04664)	0.8280891 (0.06242)	0.8909915 (0.05733)	0.8500793 (0.03701)	0.9084415 (0.03237)
	$n = 300$	0.9258467 (0.02464)	0.947289 (0.02096)	0.8995048 (0.03879)	0.9425813 (0.01753)	0.8926859 (0.02176)	0.9514676 (0.01069)

Table-8: The numerical summary of results obtained using three different estimation methods, (NW), (LLK) and (CSS) estimators from the 2nd simulation study using $f_{10}(x_1, x_2, x_3)$.

		NW estimator		LLK estimator		CSS estimator	
		$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.1$	$\sigma = 0.2$
ASB	$n = 50$	0.0016107	0.0031957	0.0016824	0.0028936	0.0026317	0.0034519
	$n = 100$	0.0013037	0.0026870	0.0019808	0.0025529	0.0023889	0.0034500
	$n = 300$	0.0006937	0.0012802	0.0010369	0.0018310	0.0013529	0.0022224
AVAR	$n = 50$	0.0011958	0.0030312	0.0014141	0.0042713	0.0009592	0.0033399
	$n = 100$	0.0012189	0.0026250	0.0013339	0.0029247	0.0008111	0.0018826
	$n = 300$	0.0006589	0.0016277	0.0009829	0.0016224	0.0004807	0.0009021
AMSE	$n = 50$	0.0028064	0.0062269	0.0030967	0.0071649	0.0035909	0.0067918
	$n = 100$	0.0025226	0.0053120	0.0033146	0.0054777	0.0031999	0.0053327
	$n = 300$	0.0013527	0.0029079	0.0020199	0.0034536	0.0018336	0.0031247
Mean(R^2) (Sd)	$n = 50$	0.7050192 (0.13687)	0.8205342 (0.10319)	0.6686886 (0.13549)	0.7906079 (0.13237)	0.668856 (0.10529)	0.8208825 (0.07154)
	$n = 100$	0.7268483 (0.08416)	0.8550383 (0.05535)	0.6770546 (0.07729)	0.852648 (0.05191)	0.7041907 (0.06738)	0.8660438 (0.03602)
	$n = 300$	0.8473764 (0.03079)	0.9223196 (0.02673)	0.7979329 (0.04349)	0.9116854 (0.02190)	0.825884 (0.02763)	0.9228253 (0.01767)

The Discussion of the 2nd Simulation Study

The numerical results of the second experiment have three common points:

- The statistical aspects ASB, AVAR, AMSE, and R^2 estimated values increase as the choice of σ increments from 0.05 to 0.1 regardless of the estimation method used **NW**, **LLK**, or **CSS** estimators.
- The statistical aspects ASB, AVAR, AMSE, and R^2 estimated values decrease as the choice of sample size n increases from 50, 100 to 300, regardless of the estimation method used **NW**, **LLK**, or **CSS** estimators.
- The **NW** estimator has provided smaller estimated values than the other estimated values obtained through utilizing the **CSS**, and **LLK** estimators regardless of the choice of the value σ or sample size n .

However, we have got a further reduction in AMSE values when using the estimated **NW** for the following main reasons:

- When the number of explanatory variables increases, it becomes difficult to obtain a good estimate due to the increased curvature of the basic curve. Therefore, the estimated **NW** estimates behave well because it is based on the local zero estimate (constant).

- By increasing the number of explanatory variables to $p = 3$, the kernel based estimators gave the much better results.
- Increasing the sample size leads to more accurate estimates indicating that **CSS**, **LLK**, and **NW** provide properties with decent appreciation.

Applications on Real Data

Cross country Goods Domestic Products (**GDP**) growth panel covering the period 1960-1995 used by Liu and Stengos [8] and Maasoumi, Racine, and Stengos [21]. There are 616 observations in total, and it can be found in Rgui in *np* package under the name (oecdpanel data). The data frame with 616 observations on 4 variables.

The Response variable

Growth: rate of real **GDP** per capita for each 5-year period

The Explanatory variables

Initgdp: per capita real **GDP** at the beginning of each 5-year period

Popgro: average annual population growth rate for each 5-year period

Humancap: average secondary school enrolment rate for each 5-year period.

Firstly, from the descriptive statistics for the four selected variables used in the **CCGP** data, we notice that there are big differences in the standard deviations values among the variables. For example, the standard deviation value of the *Growth* is 0.03035 whereas the standard deviation value of the *Initgdp* is 1.013756. The same remarks can be seen for the characteristics of the two variables.

Secondly, we check the normality of the individual variables by plotting the Q-Q plots of the four variables that are formed the **CCGP** data.

Having examined the Q-Q plots of the four variables that are formed the **CCGP** data, we notice the following remarks: First, the variable *Growth* behaves to some extent close to normal. Second, the variables *Intigdp*, *popgro*, and *humancap* seem to have a severe deviation from normal. Overall, we can conclude that the **CCGP** data is not normally distributed data since the variables are clearly not following the normal behavior. This means the set of data a mixture data.

Thirdly, we utilize the three different estimators to handle the regression curve estimation (**LLK**, **LQK**, and **CSS**) for the **CCGP** data. Having obtained the four resulted estimation curves, we have obtained the results that have been tabulated in following table:

	NW	LLK	CSS
MSE	0.0006358	0.0006844	0.000842
			2

From the table we notice that the **NW** estimator work well than other two estimators in the light of the results obtained. We can see that it gives the smallest values of the MSE. It can also be noticed that the **LLK** estimator is 2nd best estimator. Also, we noticed that the **CSS** estimator failed to give a good result because the data was not normally behaved. This confirms the results obtained in section 3, which means that when the data is distributed as mixture then the Kernel based estimator’s works much better than corresponding counterparts of the cubic smoothing Splines estimator. There are two reasons for superiority of the Kernel based smoothers. One is that the **NW** and **LKK** estimators have provide well balance trading between the bias and variance in the dense and sparse regions of the data comparing with the **CSS** estimator. The second is that the Local Polynomial Kernel estimators generally do not have boundary effect problem whereas the **CSS** estimator does.

REFERENCES

1. Weisberg, S. (2005). Applied Linear Regression. 3rd edition. Wiley and Sons, Inc
2. Elysear, F. A. (2010). Comparisons between the Cubic Smoothing Spline and the Local Polynomial Kernel Estimators. Unpublished M.Sc. thesis, Garyounis University, Benghazi, Libya.
3. Priestly, M.B., & Chao, M.T. (1972). Nonparametric function fitting. Journal of Royal Statistical Society (B), 34, 384-392.
4. Cleveland, W.S. (1979). Robust Locally Weighted Regression and Smoothing Scatter Plots. Journal of the American Statistical Association. 87, 998-1004.
5. Fan, J., & Gijbels, I. (1995). Data-driven bandwidth seletion in local polynomial fitting: variable bandwidth and spatial adoption. Journal of Royal Statistical Society (B), 57(2), 371-394.

6. Wand, M.P., & Jones, M.C. (1995). Kernel Smoothing, Chapman & Hall.
7. Marron, J.S., Ruppert, D., Smith, E.K., & Conley, G. (2002). Motion picture analysis of smoothing. North Carolina Institute of Statistics Mimeo Series Number 2367.
8. Liu, Z., & T. Stengos. (1999). "Non-linearities in cross country growth regressions: a semiparametric approach," *Journal of Applied Econometrics*, 14, 527-538.
9. Mami, A. M. (2002). Local Polynomial Regression with Applications to both Independent and Longitudinal Data. Ph.D unpublished Thesis, Dept. of Statistics, University of Manchester, and Manchester, UK.
10. Keele, L. (2006). How To Be Smooth: Automated Smoothing in Political Science. Columbus, OH 43210.
11. Fan, J., & Gijbels, I. (1996). Local Polynomial Modelling and Its Applications. London: Chapman & Hall.
12. 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, 1257-1270.
13. Yang, L., & Tschernig, R. (1999). Multivariate bandwidth selection for local linear regression. *Journal of the Royal Statistical Society: Series B*, 61(4), 793-815.
14. Zhang, Z. G., & Chan, S. C. (2011). On Kernel Selection of Multivariate Local Polynomial Modelling and its Applications to Image Smoothing and Reconstruction. *J Sign Process System*, 64, 361-374.
15. Buse, A., Lim, L. (1977). Cubic splines as a special case of restricted least squares. *J. Amer. Statist. Asso*, 72, 64-68.
16. Poirier, D.J. (1975). On the use of bilinear splines in economics, *J. Econometrics*, 3, 23-34.
17. Smith, P.L. (1979). Splines as a useful and convenient statistical tool. *The American Statistician*, 33, 57-62.
18. Eubank, R.L. (1988). *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York.
19. Chen, L. A. (1997). Multivariate regression splines. *Computational Statistics & Data Analysis*, 26, 71-82
20. Green, P. J., & Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. Boca Raton, FL. Chapman & Hall.
21. Maasoumi, E., & J.S. Racine., & T. Stengos. (2007). "Growth and convergence: a profile of distribution dynamics and mobility," *Journal of Econometrics*, 136, 483-508