IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

ISSN: 3007-1658 (Online)

On the Theory and Pliability of Regressogram Decomposition: Application and Simulation Sensitivity of the Olanrewaju-Olanrewaju Kernel-Based Regressogram

Rasaki Olawale Olanrewaju[1], Sodiq Adejare Olanrewaju[2]

[1]Africa Business School, Business Analytics Value Networks (BAVNs), Mohammed VI Polytechnic University, Rabat 10112, Morocco.
[2] Department of Statistics, Faculty of Science, University of Ibadan, 900001, Ibadan Oyo State, Nigeria.
[1]olanrewaju_rasaq@yahoo.com
[2]sodiqadejare19@gmail.com

ARTICLE INFO

Article history:
Received 01 February 2025
Revised 01 February 2025 ,
Accepted 28 February 2025,
Available online 28 February 2025

Keywords:
Bandwidth
Generalized-Cross Validation
Kernels
Nadaraya-Watson Kernel-Estimator
Regressogram

ABSTRACT

This paper studies the pliability of regressogram decomposition as a technique of modeling bivariate random measurements of $(X_1, Y_1), \dots, (X_n, Y_n)$ either via equal-width bins or via linear smoother with the use of nonparametric, Olanrewaju-Olanrewaju, and machine-learning Boxcar kernels. The mentioned kernels were separately incorporated into the Nadaraya-Watson kernel-estimator as a generalized Mercer kernel. The optimal bandwidth needed as a smoothing parameter for impelling the kernel-based regressogram was derived using Generalized Cross-Validation (GCV). Furthermore, finite and countable sample size bound required for the regressogram modeling was ascertained for reasonable sample sizes needed for effective optimization of coefficients, deductive measurable of some error indexes and strongly universally consistent estimator. In conclusion, the Olanrewaju-Olanrewaju kernel-based regressogram was notably sensitive with miniature scale estimated (σ^2) and GCV estimates in application to real life dataset and simulation study to the nonparametric Doppler regression function.

1. Introduction

Nonparametric regression is a distribution-free regression type modeling that does not presupposes a particular functional error form for linking covariate(s) of interest with associated dependent variable. Instead, it uses division of histogram into either equal or unequal bins; kernels (either as a smoothing or Mercer function), or regression tree to link covariate(s) of interest with the dependent variable with strong assumption of linearity and independence [1, 2]. However, link functions between interested covariate(s) and the associated dependent variable could be in terms of either simple, multiple or multivariate nonparametric regression [3]. Consequently, behavioral histogram via equal or unequal bins driven by a kernel-based function to link covariate(s) of interest to the corresponding response variable is regarded as regressogram [4]. In other words, regressogram is simply the conventional regression modeling plus space of histogram bins been compelled by a kernel or a linear smoother. It is otherwise known as the binning approach of regression analysis. Explicitly, it is the average of the responses corresponding to the associated covariates in the same bin as X (where X is the vector of all covariates). A special version of the

Corresponding author E-mail address: olanrewaju_rasaq@yahoo.com
https://doi.org/10.62933/7tz4w241

regressogram analysis occurs when the bins are chosen recursively by splitting previous bins along axis directions. This could be regarded as *regression tree* [5, 6]. Regressogram is an effective method of examining intricate interactions between variables in a nonparametric setting. One important issue to keep in mind while utilizing regressogram is the bias-variance trade-off. Intuitively, this can be catered for via employing kernel estimators (such as the spline) and linear/non-linear smoothers as drivers for regressogram decomposition [7 - 9].

Harmoniously, bias-variance trade-off describes estimates' accuracy and the cognate ideal of balancing precision in regressogram. Bias-variance trade-off is regarded as trade-off because the biasedness of estimator of interest decreases resulting to higher variability of such estimator and vice versa [10]. The trade-off complexity can be balanced-off in regressogram by appropriately selecting the optimal smoothing parameters (bandwidths), bins, and kernel-based estimator (smoother) for the distribution-free regression-type (regressogram) [11]. On the other hand, kernel smoother is referred to as kernel estimator because it makes use of kernel functions to link covariates of interest to the corresponding dependent variable with the assurance of optimal, efficient, and deductive estimates. Emphatically, regressogram is a useful nonparametric regression technique that makes it possible to choose from a pool of kernels and select the ideal bandwidth that best drive the distribution-free regression analysis [12, 13]. Additionally, smoothen estimate makes it possible via kernel functions to weight observations according to how significant they are to the smoother or covariates. It also makes it possible to ascertain changes in responses due to changes in parameters. Regressogram also makes it possible and easier to adopt and compare different types of kernel estimators to ascertain their variability and error index performances. Among the well-known generalized kernel estimators include Priestley-Chao, Nadaraya-Watson, and Gasser-Müller estimators [14, 15]. The smoother can be linear or non-linear depending on the strength and direction of the relationships between or among covariate(s) and the response variable.

From reviewing point of view, [16] proposed a regressogram-type model estimation for data with equal values. Residual Mean Square Error (RMSE) was adopted as error measurement index with the adoption Mood-Brown, Ordinary Least Square (OLS), Theil, Optimum and Theil Median, Hodges-Lehman and Theil, and Optimum Mean kernels separately. In application to Samsunspor league away goals and scoring goals. That is, the number of away and scoring goals between 1995-2013 seasons. Hodges-Lehman and Theil nonparametric regression analysis yielded the smallest mean and median ranking values, as well as the same miniature RMSE estimate valued at 88,9088 compare to high estimates by the classical regression model. It was concluded that the regressogram-type model should be adopted whenever the classical regression model assumptions are not valid or when the sample number is very low. This advantage makes the former more effective in the presence of outliers. In extension, [17] numerically employed the nonparametric regression model by adopting the Gaussian kernel to predict the average waiting time and time spent moving at 1 ms^{-1} or less by urban bicyclists during rush hours while performing different manoeuvres at intersections in the city of Bologna, Italy. It was reported that predictions made were optimally robust. The robustness was confirmed via boostrapping method that highlighted contributions of different covariates that affected waiting time in the city. Reportedly, it was affirmed that future work should focus more on testing the model transferability to some other case studies.

Furthermore, [18] considered nonparametric regression model with the adoption of k-Nearest Neighbours (k-NN) method and Nadaraya–Watson kernel in a highly flexible compositional response data with inclusion of zeros. In application to simulation studies and real-life data, [18] highlighted that nonparametric regression model via the two adopted methods was robust in complex relationships between compositional response and Euclidean predictor variables. [18] also affirmed that both methods of regression analyzes lead to more accurate predictions compare to the classical regression-type models that usually assume restrictive parametric relationship with the dependent variables. Conclusively, it was summarized that the k-NN regression kernel enjoys higher computational efficiency than Nadaraya–Watson kernel, rendering the former to be highly attractive for use with large sample data sets.

Connectively, regressogram depends on the division of covariates into histogram space bins, say $B_1, \dots, B_n$ with a cognate assumption that covariates $X_i's$ are from a distribution over $[0, 1]$ before exploring the ideal bandwidth, kernel estimators, and other parameters needed to drive it [19 - 21].

Consequently, this article shall workout the practical details of regressogram based on equal-width bins and afterwards adopt the *Nadaraya-Watson estimator* (NW-estimator) to smoothen the defined regressogram function. The NW kernel-estimator will be adopted because of its pliability to accommodate bias-reducing (second and fourth-order kernels: the nonparametric kernels) kernels, built-in kernels, and the newly convoluted machine-learning kernel known as the Boxcar kernel. In other words, the nonparametric kernels of Epanechnikov, Biweight, Triweight, Gaussian, Uniform, and Triangular, as well as the inbuilt Olanrewaju-Olanrewaju kernel and machine-learning kernel of Boxcar shall be incorporated into the generalized NW kernel-estimator differently to smoothen the regressogram function. In addition, linear smoother via collection of nice functions with the use of *Least Square* (LS) method shall be introduced to drive the regressogram function with an optimal bandwidth to be optimized by *Generalized Cross-Validation* (GCV). Lastly, these mentioned functional relations of regressogram shall be subjected to simulation study of a regression function known as the Doppler function and real life application for deductive inference.

The key components of this manuscript includes Introduction, Terminology and symbols, Methodology, Research findings and discussion, Conclusions, Appendix Acknowledgement, Author contributions, Conflicts of interest, and References.

2. Methodology

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a bivariate random sample for ranges of values that can be domain in a continuous or discrete form. In this kind of regression analysis, we are interested in the regression function of the form,

$$m(x) = E(Y|X = x) \tag{1}$$

Equation (1) can be explicitly written as,

$$Y_i = m(X_i) + e_i \tag{2}$$

Such that, $E(Y|X = x) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \cdots + \theta_p X_p = \sum_{i=0}^{p} \theta_i X_i$, with $X_0 = 1$, $e_i \sim (0,1)$ and the assumption that $E(e_i) = 0$, $E(e_i) = Y_i - m(X_i) = 0$.

$m(x)$ could be referred to as the smoothen function, otherwise regarded as the nonparametric regression. Equation (2) may however be problematic in some applications, particularly if the data are tainted by anomalies or needed to be constraint to be non-distributional error form, then the concept of histogram will be required to adjust the regression setting of equation (1). The adjustment of equation (1) with histogram is what is regarded as *regressogram*. The name regressogram was coined by [22] to mean the combination of the normal regression (as a smoother without any error distributional form) plus space of bins of histogram. According to equation [23], regressogram is the averaging of covariates corresponding to $X's$ that falls into disjoint bins spanning of the X-covariates space.

Table 1: Terminology and symbols for the proposed regressogram modeling.

Symbol	Quantity
Y_{j}	Is the average of $Yi's$ in B_j
B_j	Is the bins $B_1,\ldots, B_{\mathrm{j}}$ that represents the entire histogram's regression space.
Y	Is the response variable
$m(x)$	The smoothen function or the nonparametric regression
$\widehat{m}_h(x)_{I(x\in B_j)}$	Is the nonparametric Nadaraya-Watson kernel estimator
$h>0$	Referred to as the bandwidth
$K(\cdot)$	Referred to as the second and fourth order kernel
$Bias(m_{J(x)})$	Biasedness of the regressogram estimator
$Var(m_{J(x)})$	Variance estimator of the regressogram
$CV(h)$	Cross-validation that minimizes the bandwidth
$GCV(h)$	Generalized cross-validation that minimizes the bandwidth
$hopt(x)$	Optimal bandwidth for covariate x to $t-1$
P_n	Empirical measure

2.1 Regressogram via Equal Width-Bins

This can be thought of as an approximation to smoother (x) by a stepwise function evaluated at midpoints of the bins. This makes regressogram a special kind of kernel-based regression function defined for discontinuous stepwise function of bins. Regressogram with equal-width bins is often regarded as the binning approach of regression analysis as represented below as,

Regressogram = regression + histogram (3)

Alternatively, regressogram represents the histogram's regression version that makes bins $B_1,\ldots,B_n$ out of the available space of the entire histogram space. Mathematically,

$$\bar{m}(x) = \bar{Y}_j \quad \text{for} \quad x \in B_j \tag{4}$$

Where $\bar{Y}_j$ is the mean of the responses $(Y's)$ for the data in the bin B_j. In nonparametric regression setting, this implies selecting the bins recursively, by dividing early bins along axis directions. Assuming we are interested in multiple covariates of $x_1, x_2, \ldots, x_5$ for a random variable say X, then the nonparametric regression of equation (2) becomes,

$$Y = m(x_1, x_2, x_3, x_4, x_5) + e_i \tag{5}$$

Additively,

$$Y = m_1(x_1) + m_2(x_2) + m_3(x_3) + m_4(x_4) + m_5(x_5) + e_i \tag{6}$$

such that, $\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{m}_5$ could be estimated by a kernel estimator. This implies that each $\widehat{m}_i$ is an embedded function of (y_i, x_i) for $i = 1, \ldots, n$. For simplicity, if covariates of $x_1, x_2, \ldots, x_5$ for a random variable say X that assumes values between $0 \le X \le 1$, then the simplest nonparametric estimator of m is the regressogram with $k-$integer such that, the $k-$equal-width bins can be divided into $[0,\ 1]$ with

$$B_1 = [0,\ \tfrac{1}{k}),\ B_2 = [\tfrac{1}{k},\ \tfrac{2}{k}),\ B_3 = [\tfrac{2}{k},\ \tfrac{3}{k}),\ B_4 = [\tfrac{3}{k},\ \tfrac{4}{k}), B_5 = [\tfrac{4}{k},\ \tfrac{5}{k}) \tag{7}$$

Assuming n_j denotes observations in the bins, say B_j. This implies that $n_i = \sum_i I(X_i \in B_j)$ where $I(X_i \in B_j) = 1$ if $X_i \in B_j$ and $I(X_i \in B_j) = 0$ provided $X_i \notin B_j$. If $\bar{Y}_j$ is the average of $Y_i's$ in B_j, then, $\bar{Y}_j = \frac{1}{n_j}\sum_{X_i \in B_j} Y_i$. Rewriting in a kernel form gives $\hat{m}(x) = \sum_{j=1}^{k} \bar{Y}_j I(x \in B_j)$. The stated kernel estimator of equation (6) that smoothens Y can be of any form.

2.2 Nadaraya-Watson Kernel-Estimator

The adopted kernel estimator for this write-up that smoothens Y is the nonparametric *Nadaraya-Watson kernel estimator* (NW kernel-estimator) define as,

$$\hat{m}_h(x)_{I(x\in B_j)} = \frac{\sum_{j=1}^{n} Y_i K(\frac{x-X_j}{h})}{\sum_{j=1}^{n} K(\frac{x-X_j}{h})} + e_i \qquad (8)$$

Where $h > 0$ is referred to as the bandwidth and the function $K(\cdot)$ could be any of the second and fourth order kernels, inbuilt kernels or machine-learning kernels, $\hat{m}(x)$ is the local average of $Y_i's$. However, NW-estimator was adopted because it effectively smoothen noisy data, reduces noise interference, and can be conceived as a locally weighted average function (weighting function). This implies that equation (6) could be rewritten as,

$$Y = m_1(x_1) + m_2(x_2) + m_3(x_3) + m_4(x_4) + m_5(x_5) + e_i$$

$$Y_{I(x\in B_j)} = \frac{Y_1 K(\frac{x-X_1}{h})}{K(\frac{x-X_1}{h})} + \frac{Y_2 K(\frac{x-X_2}{h})}{K(\frac{x-X_2}{h})} + \cdots + \frac{Y_n K(\frac{x-X_n}{h})}{K(\frac{x-X_n}{h})} + e_i \qquad (9)$$

In a general term of $i = 1, \dots, n$.

Below are the known properties of $K(x)$

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

It is to be noted that theoretical calculations might be insensitive in terms of deviation of results to the choice of kernel. What matters is the choice of bandwidth(h), which controls the amount of smoothing. Notably, smaller magnitude of bandwidths usually yields rough estimates, while larger bandwidths produce smoother estimates. Among the optimal second and fourth order kernels, inbuilt kernels, and convoluted kernels of $K(\cdot)$ designed such that they can be incorporated into the NW kernel-estimator are the *nonparametric second and fourth kernels of*

1. **Bisquare Kernel:**

$$K(\frac{x-X_j}{h_l})_{I(x\in B_j)} = \frac{\frac{15}{16}\{1-2\|X_i-X_j\|^2+\|X_i-X_j\|^4 I_{[x\in B_j]}\ \|X_i-X_j\|\}}{h_l}$$
$$\forall\ \{i,j\} = 1, \dots, n \qquad (10)$$

2. **Gaussian Kernel:**

$$K(\frac{x-X_j}{h_l})_{I(x\in B_j)} \propto (2\pi)^{-d/2}(-\|X_i - X_j\|_2^2/2h_l^2) \qquad (11)$$

The commonest choice of this kernel is the Gaussian density with $K(x) \propto e^{-x^2/2}$. This does not necessarily mean that it is assumed that the data are normally distributed neither that a distributional random noise is needed. The proportionate kernel is just a way of defining smoothing weights.

3. **Triweight Kernel:**

$$K(\frac{x-X_j}{h_l})_{I(x\in B_j)} = \frac{\frac{35}{36}\{1-3\|X_i-X_j\|^2+3\|X_i-X_j\|^4-\|X_i-X_j\|^6\}\{I_{[x\in B_j]}\ \|X_i-X_j\|\}}{h_l} \qquad (12)$$

4. **Uniform Kernel:**

$$K(\frac{x-X_j}{h_l})_{I(x\in B_j)} = \frac{\frac{1}{2}\{I_{[x\in B_j]}\ \|X_i-X_j\|\}}{h_l} \qquad (13)$$

5. **Epanechnikov Kernel:**

$$K(\frac{x-X_j}{h_l})_{I(x\in B_j)} = \frac{\frac{3}{4}\{1-\|X_i-X_j\|^2\}\{I_{[x\in B_j]}\ \|X_i-X_j\|\}}{h_l} \qquad (14)$$

6. **Triangular Kernel:**

$$K(\frac{x-X_j}{h_l})_{I(x\in B_j)} = \frac{\{1-\|X_i-X_j\|\}\{I_{[x\in B_j]}\ \|X_i-X_j\|\}}{h_l} \quad (15)$$

Where, h_l is the bandwidth of the l individual covariate.

In addition, another notable inbuilt kernel is the

7. **Olanrewaju-Olanrewaju Kernel:**

$$K(\frac{x-X_j}{h_l})_{I(x\in B_j)} = \frac{\lambda(\gamma+\frac{2}{\pi})arctan(\frac{|x-X_j|}{\gamma})tanh(\lambda)}{h_l} = \frac{\lambda(\gamma+\frac{2}{\pi})arctan(\frac{|x-X_j|}{\gamma})(\frac{e^{\lambda}-e^{-\lambda}}{e^{\lambda}+e^{-\lambda}})}{h_l} \quad (16)$$

$\ni \lambda \geq 0$ (λ a non-negative value), usually 0.005, where γ is usually peg at 3.5 according to [24, 25], such that,

$$I(x \in B_j) = B_1 = \left[0,\ \frac{1}{j}\right),\ B_2 = \left[\frac{1}{j},\ \frac{2}{j}\right),\ B_3 = [\frac{2}{j},\ \frac{3}{j}),\ \ldots, B_{j-1} = [\frac{j-2}{j},\ \frac{j-1}{j}), B_j = [\frac{j-1}{j},\ 1) \quad (17)$$

The known convoluted machine-learning kernel is the

8. **Boxcar Kernel:**

$$K(\frac{x-X_j}{h_l})_{I(x\in B_j)} = \begin{cases} 1 & if \quad |\frac{x-X_j}{h_l}| \leq 1 \\ 0 & if \quad |\frac{x-X_j}{h_l}| > 1 \end{cases} \quad (18)$$

For all average of Y_i such that $|X_i - x| \leq h$ where (h) is some small magnitudes of bandwidths. This implies that all the stated kernels from equation (10) to equation (18) shall be individually incorporated into equation (9) of the NW kernel-estimator to give different variants of the regressogram function of equation (6).

The biasedness of the regressogram estimator via each of these stated kernels can be carved-out as:

$$Bias(\widehat{m}_J(x)) = \left(\frac{1}{J}\right) \quad (19)$$

The corresponding variability (variance) estimator of the regressogram is nothing but:

$$Var(\widehat{m}_J(x)) = O\left(\frac{1}{J}\right) \quad (20)$$

such that the Mean Square Error (MSE) and Mean Index Square Error (MISE) indexes are:

$$MSE = O\left(\frac{1}{J}\right) + O\left(\frac{1}{J}\right),\ MISE = O\left(\frac{1}{J}\right) + O\left(\frac{1}{J}\right) \quad (21)$$

2.3 Confidence Interval (C.I) and Variance Estimation of the Regressogram via Equal-Width Bins.

To construct the Confidence Interval (C.I) of $m_h(x)$, the stated assumption of equation (2) that $e_i \sim (0,1)$ shall be adopted. Having ascertained that $e_i \sim (0,1)$, then the z-variate distributed random variable of $m_h(x)$ shall be assumed. That is z-variate of $z = \frac{x-\mu}{\sqrt{\sigma^2}}$. Then,

$$\sqrt{nh}(\widehat{m}_h(x)) - E(\widehat{m}_h(x)) \to \aleph(0, \frac{\sigma^2 \cdot \sigma_K^2}{\pi(x)}) \quad (22)$$

$$\frac{\widehat{m}_h(x) - E(\widehat{m}_h(x))}{Var(\widehat{m}_h(x))} \to \aleph(0,1) \quad (23)$$

No doubt that the variance depends on three quantifications of σ^2, σ_K^2 and $\pi(x)$. Moreover, σ_K^2 is the quantity from the characteristic of the kernel function. $\pi(x)$ is the density of covariates that can be estimated from *Kernel Density Estimation* (KDE). What remains is the unknown estimation of the variance noise σ^2. This can be done via the estimation technique of residuals from the normal regression setting, that is,

$$e_i = Y_i - \widehat{Y}_i = Y_i - \widehat{m}_h(X_i) \quad (24)$$

From the known quantity, that is $\sigma^2 = Var(e_i)$ and the residual square $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} Var(e_i)$ approximation in linear regression setting, then, we have,

$$\sum_{i=1}^{n} e_i^2 \approx \sum_{i=1}^{n} Var(e_i) = \sigma^2(n-p-1) \quad (25)$$

Then,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{(n-p-1)} \quad (26)$$

Where p is the number of parameters estimated and $(n-p-1)$ could be referred to as degree of freedom. The $(1-\alpha)\%$ C.I can be constructed via:

$$\widehat{m}_h(x) \pm z_{1-\alpha/2} \frac{\hat{\sigma} \cdot \sigma_K}{\sqrt{\hat{\pi}(x)}} \quad (27)$$

Where $\hat{\pi}_n(x)$ is the KDE of the covariates.

2.4 Regressogram via Linear Smoother

Apart from embracing the kernel-based style estimator in modeling regressogram function of equation (5), alternatively, the notion of linear smoother might also be introduced. Linear smoother is a collection of *nice functions* (differential functions with starting point(s), say x_0) of any regression estimator. A typical example of a linear smoother is the *Least Squares* (LS) for a simple linear regression of the form,

$$\widehat{m}(x) = \sum_{i=1}^{n} \Psi_i(x) Y_i = \Psi Y \quad (28)$$

Where $\Psi_i(x)$ are nice functions that depend only on $X_i (i = 1, \cdots, n)$ but not on $Y_i (i = 1, \cdots, n)$. The residual for the i^{th} observation is of the form,

$$e_j = Y_j - \widehat{m}(X_j) = Y_j - \sum_{i=1}^{n} \Psi_i(X_j) Y_i \quad (29)$$

Assuming $e = (e_1, e_2, \cdots, e_n)$ is the vector of residuals, then the $n \times n$ matrix of $\Psi_j(X_i) = \Psi_{ij}$ is

$$\Psi = \begin{bmatrix} \Psi_1(X_1) & \Psi_2(X_1) & \cdots & \Psi_n(X_1) \\ \Psi_1(X_2) & \Psi_2(X_2) & \cdots & \Psi_n(X_2) \\ \vdots & \vdots & \vdots & \vdots \\ \Psi_1(X_n) & \Psi_2(X_n) & \cdots & \Psi_n(X_n) \end{bmatrix} = \begin{bmatrix} \Psi(X_1)^T \\ \Psi(X_2)^T \\ \vdots \\ \Psi(X_n)^T \end{bmatrix} \quad (30)$$

This connotes that equation (8) can be rewritten as:

$$\widehat{m}_h(x)_{I(x \in B_j)} = \frac{\sum_{j=1}^{n} Y_i K(\frac{x-X_j}{h})}{\sum_{i=1}^{n} K(\frac{x-X_j}{h})} = \sum_{j=1}^{n} Y_i \Psi_j(X_i) \quad (31)$$

Having said that a typical linear smoother is the LS simple linear regression. Therefore, from the $\hat{\beta}$ of LS estimator, it implies that

$\hat{\beta} = (X^T X)^{-1} X^T Y$ and $\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y$, so,

$$\Psi = X(X^T X)^{-1} X^T \quad (32)$$

Hence, the i^{th}-row of matrix Ψ could be referred to as the *effective kernel for estimating* $m_h(X_i)$, where matrix Ψ could be regarded as the smoothing matrix, such that the alternative degree of freedom as affirmed in equation (26) is the trace of Ψ, that is, $p = Trace(\Psi)$.

However, after rigorous derivations and approximations, it was concluded that regressogram is also a linear smoother. If the bins of the covariates are $B_1, B_2, \cdots, B_m$ with $B(x)$ being the bin that belongs to $x's$. Then,

$$\Psi_j(x) = \frac{I(X_j \in B(x))}{\sum_{i=1}^{n} I(X_j \in B(x))} \quad (33)$$

Additionally, kernel driven regressogram with equal-width bins is also a linear smoother from a starting point value of $x = x_0$ as a differential equation with,

$$\widehat{m}_h(x_0)_{I(x\in B_j)} = \frac{\sum_{j=1}^{n} Y_i K(\frac{x_0-X_j}{h})}{\sum_{i=1}^{n} K(\frac{x_0-X_j}{h})} = \sum_{j=1}^{n} Y_i \Im_i^K(x_0),$$ such that,

$$\Im_i^K(x_0) = \frac{K(\frac{x_0-X_j}{h})}{\sum_{\Psi=1}^{n} (\frac{x_0-X_j}{h})}$$

So, invariably, $\Psi_j(x) = \frac{K(\frac{x_0-X_j}{h})}{\sum_{\Psi=1}^{n} K(\frac{x_0-X_j}{h})}$

In addition, the variance of any kernel driven regressogram coincides with that of the regressogram driven by a linear smoother with

$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{(n-p-1)}$ (See Appendix).

2.5 Selection of Bandwidth h by a Cross-Validation (CV) Technique

Bandwidth (h) could be chosen via Cross-Validation (CV), or via a more redefined Leave-One-Out Cross-Validation (LOOCV). Assuming $\widehat{m}_h^{(-j)}$ is the kernel estimator with bandwidth (h) obtained after leaving-out (X_i, Y_i). The CV (similar to that of the AIC of continuous parametric index) index score is,

,

$$CV(h) = \frac{1}{n}\sum_{j} (Y_j - \widehat{m}_h^{(-j)}(X_i))^2 \quad (34)$$

For

$$\Psi_{i(-j)}(x) = \begin{cases} 0 & if\ i = j \\ \dfrac{\Psi_j(x)}{\sum_{k\neq i}\Psi_k(x)} & if\ i \neq j \end{cases}$$

Such that, $\widehat{m}_{(-j)}(x) = \sum_{i=1}^{n} Y_i \Psi_{i,}(-j)$

However, there is still a short cut to it via,

$$CV(h) = \frac{1}{n}\sum_{i} (\frac{(Y_i - \widehat{m}_h(X_i))}{1-\Psi_{jj}}) \quad (35)$$

The idea here is to fit $\widehat{m}_h$ for some values of (h). We then compute $CV(h)$ by trying to find $\hat{h}$ that minimizes $CV(h)$. By doing so, Ψ_{jj} could be replaced by the averaging value of $\frac{1}{n}\sum_{j}\Psi_{jj} = \frac{1}{n}Tr(\Psi) = \frac{p}{n}$. By this replacement of Ψ_{jj} with $\frac{p}{n}$, we have *Generalized Cross-Validation* (GCV) as,

$$GCV(h) = \frac{1}{(1-\frac{p}{n})^2}\frac{1}{n}\sum_i (Y_i - \widehat{m}_h(X_i))^2 \quad (36)$$

Alternatively, local bandwidth can also be selected via a plug-in principle. The formula depends on asymptotic best optimal local bandwidth h_{opt} of the variate x, that is, $h_{opt}(x)$;

$$h_{opt}(x) = \frac{1}{n^{1/5}}\left(\frac{\sigma_\varepsilon^2 \int K^2(z)dz}{\{m''(x)\int z^2 K(z)dz\}^2}\right)^{1/5} \quad (37)$$

Iteratively by starting with an initial bandwidth h_0 to estimate $m'(\cdot)$ by an inflated version of $n^{1/10}h_0$ and σ_ε^2. In case there is need for smoothing of two continuous derivative functions of a regression problem that need to minimize the sum of squares, we then have,

$$SS\left(h_{opt}(x)\right) = \sum_{i=1}^{n} Y[i^{-}m_h(X_i)]^2 + \int_{xmin}^{xmax} [m_h''(X_i)]^2 dx \quad (38)$$

Where $h_{opt}(x)$ is the smoothing parameter.

2.6 Finite Sample Bounds of Regressogram with Appropriate Linear Smoother

Theorem 1: If $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty}\frac{nh_n^d}{logn} = \infty$, then the estimate m_n is strongly universally consistent (See [26]), that is,

$$\lim_{n\to\infty}\int (m(x) - m_n(x))^2\mu(dx) = 0 \text{ (a.s.)} \quad (39)$$

for any bivariate random sample of (X, Y) with $E(Y^2) < \infty$.

$$\lim_{n\to\infty} \int E[\{m(x) - m_n(x)\}^2]\mu(dx) = 0 \quad (40)$$

$$E\int [\{m(x) - \widetilde{m}_n(x)\}^2]\mu(dx) = O(\tfrac{1}{nc_n^2 h_n^{2d}}) + O(h_n^2) \quad (41)$$

for any bivariate random sample of (X, Y) with $E(Y^2) < \infty$, such that, $h_n = cn^{-1/(2(d+1))}$, and $c_n = \frac{1}{\sqrt{\log n}}$.

Proof.

Assuming the smoothing estimator is universally consistent, such that, $[E\|\widehat{m}_h(x) - m_{h_0}(x)\|_2^2 \to 0$ as $n \to \infty$ with assumption that $E(Y^2 < \infty)]$ for any compacted kernel $K(\cdot)$ and bandwidth $h \to h_n$ satisfying $h_n \to 0$ and $nh_n^d \to \infty$ as $n \to \infty$.

To solve the kernel smoothing estimator consistency of $E\|\widehat{m}_h(x) - m_{h_0}(x)\|_2^2 \to 0$ depends on another emphatic probabilistic assumption that P_X is known such that $E\|\widehat{m}_h(x) - m_{h_0}(x)\|_2^2 = \int (\widehat{m}_h(x) - m_{h_0}(x))P_X(x)$ or equivalently $\sum(\widehat{m}_h(x) - m_{h_0}(x))P_X(x)$ (In a discrete setting). By dropping this assumption in order to move away from the parametric probabilistic assumption of known P_X, we shall be employing a spherical kernel $K(\|x\|) = I(\|x\| \le 1)$ to nullify the constraint of known P_X as a result of extension to incorporated kernels.

$$\widehat{m}_h(x) = \frac{\sum_{i=1}^n Y_i I(\|X_i - x\| \le h)}{\sum_{i=1}^n I(\|X_i - x\| \le h)} = \frac{\sum_{i=1}^n Y_i I(\|X_i - x\| \le h)}{nP_n(B(x,h))} \quad (42)$$

Where P_n is an empirical measure, where $B(x, h) = \{u: \|x - u\| \le h\}$. Obviously, if $B(x, h) = \{u: \|x - u\| \le h\} = 0$, then $\widehat{m}_h(x) = 0$.

Assuming $\Omega = \{m: |\widehat{m}_h(x) - m_h(x)| \le L\|x - u\|^1,\ x, u, \in \mathbb{R}^d\}$ such that $L\|\ \cdot\ \|^p$ is an L^p −space with $p = 1$ for $L\|x - u\|^1$.

The risk bound of $E\|\widehat{m}_h(x) - m_{h_0}(x)\|_2^2$ without density P_X for compact support of X possesses $(Y|X = x) \le \sigma^2 < \infty\ \forall x$, then,

$$\sup_{P\in H_d(1,L)} E\|\widehat{m}_h(x) - m_h(x)\|_P^2 \le c_1 h^2 + \frac{c_2}{nh^d} \quad (43)$$

$$\approx \frac{c}{n^{\frac{2}{(d+2)}}} \text{ if } h \approx n^{\frac{1}{(d+2)}}$$

Such that the smoothen manifold of dimension $r < d$ is

$$\int \frac{dP(x)}{nP(B(x,h))} \approx O(\tfrac{1}{nh^r}) \text{ Instead of } O(\tfrac{1}{nh^d}) \quad (44)$$

Let,

$$m_h(x) = \frac{\sum_{i=1}^n I(\|X_i - x\| \le h)m(X_i)}{nP_n(B(x,h))} \quad (45)$$

If D_n exist and is true such that $D_n = \{P_n(B(x,h)) > 0\}$, it implies that,

$$E((\widehat{m}_h(x) - m_h(x))^2|X = X_1, X_2, \dots, X_n) = \frac{\sum_{i=1}^n (\|X_i - x\| \le h)Var(Y_i|X)}{n^2 P_n^2(B(x,h))} \le \frac{\sigma 2}{nP_n(B(x,h))} \quad (46)$$

For $m \in \Omega$, we have $|m(X_i) - m(x)| \le L\|X_i - x\| < Lh$ for $X_i \in B(x, h)$, hence,

$$|m_h(x) - m(x)|^2 \le L^2h^2 + m^2(x)I_{A_n(x)^c} \quad (47)$$

Therefore,

$$E\int (\widehat{m}_h(x) - m(x))^2 dP(x) = E\int (\widehat{m}_h(x) - m_h(x))^2 dP(x) + E\int (m_h(x) - m(x))^2 dP(x)$$

$$\le E\int \frac{\sigma^2}{nP_n(B(x,h))} I_{A_n(x)} dP(x) + L^2h^2 + \int m^2(x)E(I_{A_n(x)^c})dP(x) \quad (48)$$

Bounding the first term, assuming $G = nP_n(B(x,h))$. It is to be noted that $G \sim Bin(n,q)$ where $q = P(X \in B(x,h))$. Now,

$$E(\frac{I(G>0)}{G}) \leq E(\frac{2}{1+G}) = \sum_{k=0}^{n} \frac{2}{k+1} \binom{n}{k} q^k (1-q)^{n-k} \quad (49)$$

$$= \frac{2}{(n+1)q} \sum_{k=0}^{n} \binom{n+1}{k+1} q^{k+1} (1-q)^{n-k} \quad (50)$$

$$\leq \frac{2}{(n+1)q} \sum_{k=0}^{n} \binom{n+1}{k} q^{k+1} (1-q)^{n-k+1} \quad (51)$$

$$= (q + (1-q))^{n+1} \frac{2}{(n+1)q} = \frac{2}{q(n+1)} \leq \frac{2}{qn} \quad (52)$$

Therefore,

$$E\int \frac{\sigma^2 I_{A_n(x)}}{nP_n(B(x,h))} dP(x) \leq 2\sigma^2 \int \frac{dP(x)}{nP(B(x,h))} \quad (53)$$

Our interest is to choose points $(t_1, \ldots, t_M)$ such that the support of P_X is covered by $\cup_{j=1}^{\Omega} B(t_j, \frac{h}{2})$ where $\Omega \leq \frac{c_2}{(nh^d)}$. So,

$$\int \frac{dP(x)}{nP(B(x,h))} \leq \sum_{j=1}^{\Omega} \int \frac{I(k \in B(k_j, \frac{h}{2}))}{nP(B(x,h))} dP(x) \leq \sum_{j=1}^{\Omega} \int \frac{I(k \in B(k_j, \frac{h}{2}))}{nP(B(x, \frac{h}{2}))} dP(x) \leq \frac{M}{n} \leq \frac{c_1}{nh^d} \quad (54)$$

From the third entity of equation (48) we have,

$$\int m^2(x) E(I_{A_n(x)^c}) dP(x) \leq \sup_x m^2(x) \int (1 - P(B(x,h)))^n dP(x) \quad (55)$$

$$\leq \sup_x m^2(x) \int e^{-nP(B(x,h))} dP(x)$$

$$= \sup_x m^2(x) \int e^{-nP(B(x,h))} \frac{nP(B(x,h))}{nP(B(x,h))} dP(x)$$

$$\leq \sup_x m^2(x) \sup_u (ue^{-u}) \int \frac{1}{nP(B(x,h))} dP(x)$$

$$\leq \sup_x m^2(x) \sup_u (ue^{-u}) \frac{c_1}{nh^d} = \frac{c_2}{nh^d} \quad (56)$$

Remark:

Equation (56) ascertain the finite and countable sample size bound require for kernel-based regressogram modeling. It implies that estimates from linear smoother or kernel-based regressogram are error and estimates bounded, coupled with the un-deviated linearity assumption without making strong assumptions. Furthermore, what this connotes is that for strongly universally consistent estimates of $m_h(x)$ to be ascertained, sample size (n) of bivariate data points ideal for optimal regressogram estimates and bounded error must be less than $nh^d c_1 = nh^d c_2$ for some c_1 and c_2 in $\mathbb{R}$. Additionally, futuristically, rates of convergence for estimates could be looked into.

2.7 Algorithm

Input: Observations/samples/datapoints stored in a trained bivariate dataset $\{(X_i, Y_i) = (x_1, y_2), \ldots, (x_n, y_n)\}$ for $i = 1, \ldots, n$ with T, smooth function from Hilbert space H, bins

$$I(x \in B_j) = [[0, \frac{1}{k}), [\frac{1}{k}, \frac{2}{k}), [\frac{2}{k}, \frac{3}{k}), \ldots,$$

$[\frac{(n-2)}{k}, \frac{(n-1)}{k}), [\frac{(n-1)}{k}, \frac{n}{k})]$ and regularization parameter γ that is usually 3.5 in magnitude and penalization λ included in the Olanrewaju-Olanrewaju kernel.

1. **Initialize** $\widehat{m}_h(x)_{I(x \in B_1)}, \widehat{m}_h(x)_{I(x \in B_2)}, \cdots, \widehat{m}_h(x)_{I(x \in B_j)}$ for $n_i = \sum_i I(X_i \in B_j)$
2. If $\bar{Y}_j$ of $Y_i's$ are in B_j, then, estimate $\bar{Y}_j = \frac{1}{n_j} \sum_{X_i \in B_j} Y_i$

3. If $n_i = \sum_i I(X_i \in B_j)$, then $\widehat{m}_h(x) = \sum_{j=1}^{k} \bar{Y}_j I(x \in B_j)$

4. Compute the initial estimator based on D:

$\hat{Y}_{I(x \in B_j)} = Y_i K(\cdot)$ such that $K(\cdot)$ is any of the second, fourth, inbuilt, machine-learning kernels to be incorporated into NW

5. Apply a smoother to $\hat{Y}_{I(x \in B_j)}$ on X_j to obtain $\widehat{m}_h(x)$

6. Compute $\hat{Y}_{I(x \in B_j)} = Y_i K(\cdot)$ and transfer it to the machine

7. Compute

$$h_{opt}(x) = \frac{1}{n^{1/5}} \left(\frac{\sigma_\varepsilon^2 \int K^2(z)dz}{\{m''(x) \int z^2 K(z)dz\}^2} \right)^{1/5}$$

8. $GCV(h) = \frac{1}{\left(1-\frac{p}{n}\right)^2} \frac{1}{n} \sum_i \left(Y_i - \widehat{m}_h(X_i)\right)^2$ for $t = 1, \ldots, T$

9. Estimates of λ, σ^2, Ψ, and $h_{opt}(x)$ are extracted from $\widehat{m}_h(x)$ with the involved NW kernel-based estimator.

3. Research Findings and Discussions

Two set of analyzes shall be carried-out in this section: the regressogram analyzes shall be applied to simulation study of nonparametric Doppler regression function and a real life dataset. Two set of analyzes will be used to validate the embedded nonparametric, Olanrewaju-Olanrewaju, and Box kernels into the NW generalized-estimator as a regressogram function. In each of the set of these analyzes, the kernel-based regressogram shall be juxtaposed performance and optimality wise via some error indexes.

3.1 Simulation Study

The well-known nonparametric regression function regarded as the Doppler function,

$$f(x) = 3sin(2x) + 10(x-5)I(x > 5) \quad (57)$$

shall be subjected to the above stated kernel-based regressogram functions, where $I(\cdot)$ is the indicator function of the constraint $x > 5$. The sine function shall be kinked at $x = 5$ with a slope linear component. The error distribution of interest is the $(Ga-1)((X-5)^2+3) + T_t$. Where $Ga \sim Gamma(2,2)$, while T_t is the student-*t* distribution with three (3) degrees of freedom. Five thousand (5000) paired samples of $(Y = f(x), X)$ of $\frac{X}{10} \sim Beta(2,2)$ with distribution of $Y = m_h(x)|X$ shall be used to run $f(x) = 3sin(2x) + 10(x-5)I(x > 5)$.

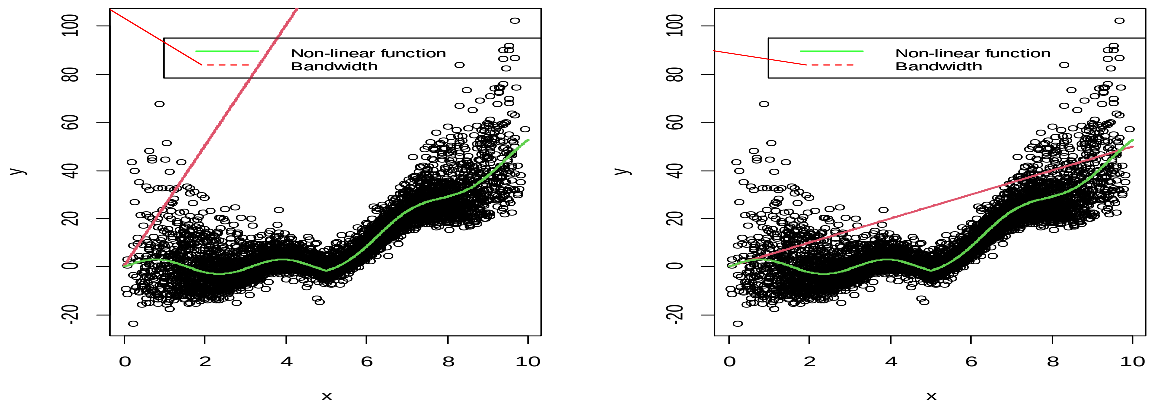

Figure 1. Nadaraya-Watson regression estimates of the nonlinear function of equation (39)

Source: Authors' Computation (2025)

From Figure 1. above, the true function of the non-linear regression function is the green color, while estimates are in red color with estimated bandwidth close to 0.25 (left) and 0.27 (right) respectively. Bandwidth values lead to estimates, which capture the function well with considerably lesser variation than the smaller bandwidth from the first set of estimates. It is to be noted that the first plot was based on approximation of the *Mean Squared Error* (MSE) minimizing bandwidth, while the other is based on *Leave-One-Out Cross-Validation* (LOOCV). The NW-estimator kernel was adopted as default because it was also perceived as a local-linear estimator. Smaller bandwidths are accurate over much of the range, but exhibits increased variation in the tails. The selection of the bandwidth was from the histogram of each of the component of the bin of the generated 5000 samples. Significantly, it connotes that bandwidths between zero (0) and ten (10) were used to drive the optimal bandwidths to be estimated Generalized Cross-Validation (GCV), that is, they were used as guessing bandwidths.

From Table 2, five thousand (5000) pair sample were sampled from $\frac{1}{10}X \sim Beta(2,2)$ and the distribution of $Y = \widehat{m}_h(x)|X$ was described for each of the kernel-based regressogram. Notably, all the embedded kernels: Triangular, Uniform, Epanechnikov, Triweight, Bisquare, Gaussian, Olanrewaju-Olanrewaju, and Boxcar into NW-estimator produced significant respondents for the regressogram $\widehat{m}_h(x)$ with corresponding covariate $3\sin(2x) + 10(x-5)I(x>5)$ with distributional random noise of Student-t (of three (3) as the degree of freedom). A Durbin-Watson estimate of 1.8 indicates a positive serial correlation (Ascertaining that the independence assumption was met).

The adjusted R-squared (R-Sq. (adj)) is a modified version of the R-squared that adjusted number of predictors in the model. The adjusted R-squared increases when terms improve the model more than expected by chance. It decreases when a predictor improves the model by less expected. Alternatively, adjusted R-squared increases when an added explanatory variable has contributed significantly to the model. However, all the embedded kernels significantly contributed to the covariate $3\sin(2x) + 10(x-5)I(x>5)$, that is, $3\sin(2x) + 10(x-5)I(x>5)$ contributed immensely to $\widehat{m}_h(x)$. With this, only uniform, Olanrewaju-Olanrewaju, and Boxcar kernels yielded higher adjusted R-squared of 26.1% for the three kernels compare to 9.81%, 4.2%, 0.873%, 3.48%, and 0.257% by triangular, Epanechnikov, Triweight, Bisquare, and Gaussian kernels respectively. However, the 26.1% adjusted R-squared yielded by uniform, Olanrewaju-Olanrewaju, and Boxcar kernels is still somewhat low in explaining the proportion of variance in the dependent variable that could explain the independent variable. Explicitly, this literally connotes that R-Sq.(adj) must lie between zero(0) and one (1), such that any value of R-Sq.(adj) close or approximately equal to one(1) is a good fit. Affirmatively, only the uniform, Olanrewaju-Olanrewaju, and Boxcar kernel R-Sq. (adj) for the regressogram are closer to 1 (but still somehow low), this implies that these kernels are somewhat of indication of interest as far as the simulation study is concern. To cement the final deduction from the pinpointed three kernels by the adjusted R-squared inference, it could be inferred that only the Olanrewaju-Olanrewaju kernel yielded the smallest magnitude of GCV, scale estimated(σ^2), and residual estimates valued at 0.0002081, 0.00020802, and 1.039677 respectively compare to that of higher magnitudes yielded by uniform and Boxcar kernels valued at 59592, 59568, 297721710; and 353.7724, 353.624, 1190886841 respectively.

Table 2: Simulation study of the kernel-based regressogram of the Doppler function $Y = m_h(x) = (Ga - 1)((X - 5)^2 + 3) + T_t$

NW ⇓ Kernel	Intercept via $Y = \hat{m}_h(x)$	X	R-Sq.(adj)	Deviance Explained	GCV	Scale Estimated	Residual
Triangular	15609.5 (1071.0) <2e-16 ***	-4546.7 (194.8) <2e-16 ***	0.0981	9.83%	28863.42	28851.8	787312.3
Uniform	-213.186 (8.402) <2e-16 ***	64.281 (1.528) <2e-16 ***	0.261	26.1%	59592	59568	297721710
Epanechnikov	556421 (58204) <2e-16 ***	-157057 (10587) <2e-16 ***	0.042	4.22%	465397.3	465218.3	12694777
Triweight	103522.4 (83370.17) 5.07e-06 ***	-101762.3 (55789.73) 2.16e-11 ***	0.00873	0.893%	32033295392	32020050831	523148.6
Bisquare	-577488 (67335) <2e-16 ***	165004 (12248) <2e-16 ***	0.0348	3.5%	622878.9	622634.7	16990545
Gaussian	0.386903 (0.034059) <2e-16 ***	-0.023090 (0.006195) 0.000196 ***	0.00257	0.277%	0.97916	0.97877	4891.868
Olanrewaju-Olanrewaju	-0.1705225 (0.0909371) <2e-16 ***	0.1891411 (0.06085714) <2e-16 ***	0.261	26.1%	0.0002081	0.00020802	1.039677
Boxcar	-426.371 (16.805) <2e-16 ***	128.562 (3.057) <2e-16 ***	0.261	26.1%	353.7724	353.624	1190886841

Keys: Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Source: Authors' computation (2025).

3.2 Real Life Application

The *English Premier League* (EPL) is one of the famous and highest levels of the English professional football league, founded in 1992 and is usually contested by twenty (20) clubs year-in year-out. It is also known as the Premiership. The football league is usually played per season. Season typically runs from August to May of another year, such that each team plays third-eight (38) matches against all other teams, both home and away. Unarguably, the accumulated points by each of the involved clubs in the EPL is a cognate function of Goals Scored (GS) and Goals Conceded (GC), that is, $(x)_{Points} = f(GS, GC)$. However, there are still some un-neglected and less determinant exogenous variables like number of wins, draws, and losses that contribute to the aggregate points $(i = 1, \dots, 17)$ by each of the twenty (20) participated clubs year-in year-out. Aggregated points year-in and year-out from 2008 to 2023 shall be subjected to regressogram of embedded kernels of the nonparametric, Olanrewaju-Olanrewaju, and Boxcar kernels for deductive inference. However, seventeen (17) yearly data points of Goals Scored (GS) and Goals Conceded (GC) that yielded number of wins, draws, and losses shall be considered.

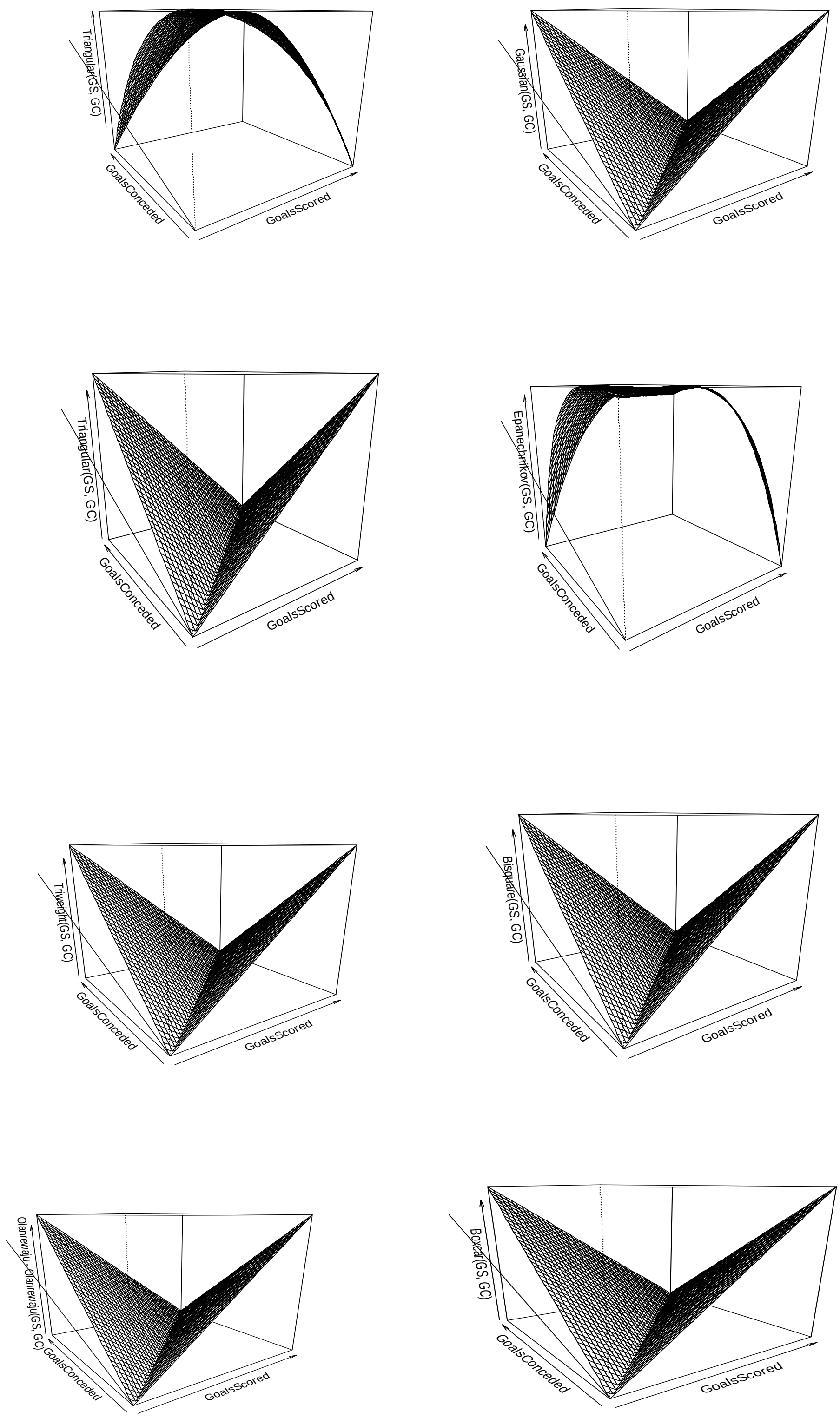

Figure 2. Nonparametric, Olanrewaju-Olanrewaju, and boxcar kernels grids of (975:1087) goals scored and goal conceded.

Source: Authors' computation (2025).

It is to be noted that kernels are estimators of sum of 'bumps' placed at the observations. Kernel function determines the shape of the bumps, while window width (h) determines their width. In other words, kernels are estimators of sum of 'bumps' centred at the observations of two bivariate dimension of (x_1, y_1), $(x_2, y_2), \cdots$, (x_{16}, y_{16}). The EPL scaled data are equal for each kernel for both dimensions and uses a single smoothing parameter each. Figure 2 above made-up the Gaussian, Epanechnikov, Triangular, Triweight, Bisquare (nonparametric); Olanrewaju-Olanrewaju, and Boxcar kernels in descending the order with positive bandwidths of 2.3, 2.5, 2.1, 2.4, 2.12, 2.6, 2.05 respectively. Because the grids of scores considered for the mentioned kernels are non-negative, that is symmetric nonnegative kernels, they could be referred to as second-order kernels (bias-reducing kernels). Notably, only Epanechnikov and Triweight possessed a relatively rhombus like (Bell like: See Figure 2 above) shape for the seventeen (17) year data points with grids of (975:1087) of goals scored and goals conceded respectively. Other remaining kernels: Gaussian, Triangular, Triweight, Bisquare, Olanrewaju-Olanrewaju, and Boxcar yielded a skewed-skewed (Cup like: See Figure 2 above) shape for the seventeen (17) year data points with grids of (975:1087) of goals scored and goals conceded respectively for the years of studied. Durbin-Watson estimate of 1.6 indicates a positive serial correlation to ascertain and confirm the independence assumption.

Table 3. English premier league (EPL) aggregated points from (2008 to 2023) in application to kernels' embedded NW estimator of the regressogram function.

NW ⇓ Kernel	Gaussian	Triangular	Uniform	Epanechnikov	Triweight	Bisquare	Olanrewaju-Olanrewaju
Intercept	1364.96092 (59.76913) 0.002623***	3605258094 (59.76913) 0.00262***	363880.13 (69395.03) 0.00119**	659638.4 (504798.9) 0.004158	-56802.6 (19713.08) 0.0098***	-63269107 (14027354) 0.00642***	21.503160 (4.100835) 0.00119 **
Wins	-0.88266 (0.13171) 0.00027***	9969461 (749470) 0.00788***	-1557.17 (152.92) 0.01280***	49735.76 (15055.53) 0.01447	-2126.372 (587.7958) 0.0078***	-3865808 (1538847) 0.024189***	-0.092020 (0.009037) 0.0128***
Draws	-0.73675 (0.07046) 0.010713 ***	-969983 (400922) 0.04613*	-316.25 (81.81) 0.00616**	-47362.53 (29627.74) 0.0006***	760.8667 (1157.029) 0.04613*	1359979 (823213.7) 0.000502 ***	-0.018689 (0.004834) 0.00616**
Losses	0.30853 (0.02751) 0.024762***	-7073183 (156528) 0.000308***	912.41 (31.94) 0.0006***	-39030.9 (11566.12) 0.0003***	1507.871 (451.7697) 0.0003***	3232272 (321440.7) 0.0003***	0.053918 (0.001887) 0.00056***
Goals Scored	-1.91795 (0.15901) 0.01521***	5810170 (904808) 0.00036***	-1699.58 (184.62) 0.02479***	57410.35 (18175.22) 0.01595 *	-1239.195 (709.6312) 0.0004 ***	-631323.9 (504793.7) 0.251466	-0.100435 (0.010910) 0.02479***
Goals Conceded	1.90162 (0.15506) 0.01360***	-15070507 (882307) 0.00527***	2723.83 (180.03) 0.00327***	-67202.75 (17721.44) 0.00570***	3213.135 (692.2838) 0.0053	1296407 (492406.1) 0.017788***	0.16096 (0.010639) 0.00327***
R-sq.(adj)	0.999	0.999	0.998	0.999	0.999	0.999	0.993
Deviance Explained	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%
GCV	8.1032	3155273	1197.852	2126326609	11929714	982945.4	0.038146
Scale Estimated	4.3633	1699039	2372.887	4212058677	6423711	529287.4	0.02054

NW ⇓ Kernel	Intercept	Wins	Draws	Losses	Goals Scored	Goals Conceded	R-sq.(adj)	Deviance Explained	GCV
Boxcar	727760.27 (138790.06) 0.00119**	-3114.34 (305.85) 0.0128***	-632.50 (163.61) 0.00616**	1824.81 (63.88) 0.0006 ***	-3399.16 (369.24) 0.0248 ***	5447.66 (360.06) 0.00327***	0.999	99.9% Scale Estimated 2580.049	4791.629

Keys: Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Source: Authors' computation (2025).

In a similar vein, from Table 3 above, all the involved regressogram covariates of $m_h(X_i)$ Wins, Draws, Losses, Goals scored, and Goals conceded of the involved kernels yielded significant contributions (that is, they contributed immensely to the dependent variable Y_i: Points accumulated with their p-values $<$ 0.05). Unarguably, the stated covariates immensely contributed to the accumulated points by the involved participated clubs within the seventeen (17) years of study in line with the ascertained finite and countable sample size bound of equation (56). Interestingly, all the kernels produced approximately the same magnitude of adjusted R-Squared and deviance estimates of 0.999 and 99.9%. Notably, differences among the kernel-based regressogram analyzes were in the scaled estimated (σ^2) and *Generalized Cross-Validation* (GCV). Among the kernel-based regressogram modeling to the EPL datasets is the Olanrewaju-Olanrewaju kernel-based regressogram analysis that yielded the smallest magnitude of scale estimated (σ^2) and GCV valued at (0.038146 and 0.02054) compare to the ones valued at (8.1032 and 4.3633), (3155273 and 1699039), (1197.852 and 2372.887), (2126326609 and 4212058677), (11929714 and 6423711), (982945.4 and 529287.4), and (2580.049 and 4791.629) by Gaussian, Triangular, Uniform, Epanechnikov, Triweight, Bisquare, and Boxcar respectively.

Deductively, from the simulation study of all the kernel-based regressogram for the Doppler function of $f(x) = 3\sin(2x) + 10(x-5)I(x > 5)$, the sensitivity of the Olanrewaju-Olanrewaju kernel-based regressogram with the smallest magnitudes of GCV, scale estimated (σ^2) and residual estimates valued at 0.0002081, 0.00020802, and 1.039677 compare to some higher valued magnitude GCV, scale estimated (σ^2) and residual estimates by other kernel-based regressogram analyzes was optimally noted. Similarly, in application to the nonparametric kernels; Olanrewaju-Olanrewaju and Boxcar kernels to the 2008 to 2023 English Premier League yearly-accumulated points, the sensitivity of the Olanrewaju-Olanrewaju kernel-based regressogram modeling was noted with the smallest magnitude of scale estimated (σ^2) and GCV estimates.

4. Conclusions and Future Recommendations

We introduced regressogram as a technique for modeling bivariate random samples of $(X_1, Y_1), \ldots, (X_n, Y_n)$ via equal-width bins and with the use of nonparametric, Olanrewaju-Olanrewaju, and Boxcar kernels. Also expounded is the introduction of linear smoother as a technique for driven regressogram analysis with the Nadaraya-Watson estimator as the basis. A derivation for the optimal bandwidth $(h_{opt}(x))$ was derived for driving the mentioned kernels as a smoothing function needed for minimizing the sum of squares, fitting efficient regressogram coefficients, and other measureable indexes like *Generalized Cross-Validation* (GCV). Finite and countable sample size bound required for the kernel-based regressogram modeling was ascertained for reasonable sample sizes needed for the nonparametric modeling coefficients and measurable to be efficient. In conclusion, the Olanrewaju-Olanrewaju kernel-based regressogram was noted to be sensitive with the smallest magnitude of scale estimated (σ^2) and GCV estimates in application to real life datasets and simulation study of the Doppler nonparametric function. In recommendation, futurist expansion could be tailored through a naive estimator, curse of dimensionality, as well as penalized regressogram type-estimator (situation where the sample size (n) is less than the covariates needed to drive the regressogram function via linear smoother, equal-width bins, or unequal-width bins.

References

[1] M. Mächler, "Nonparametric regression," Seminar Für Statistik, ETH Zürich, Switzerland, 2018, pp. 1-35.

[2] Berrett, T. B. (2021). Strongly universal consistent nonparametric regression and classification with privatized data. *Electronic on Statistics*, 18(15), 2430-2453.

[3] Hofymer, D. P. (2021). Fast exact evaluation of univariate kernel sums. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43, 447-458.

[4] Correa, J. P., Kneib, T., Ospina, R., Tejada, J., and Marmolejo-Ramos, F. (2020). Assessing the potential heteroscedasticity in psychological data: A GAMLSS approach deep learning approach. *The Quantitative Methods for Psychology*, 19(4), 333-346.

[5] Farokhi, F. (2025). Deconvoluting kernel density estimation and regression for locally differentially private data. *Scientific Report*, 10(1), 8-30.

[6] Hofymer, D. P. (2022). Fast kernel smoothing in R with applications to projection pursuit. *Journal of Statistical Software*, 101(3), 1-33.

[7] T. Hastie, R. Tibshirani, and J. Friedman, "Elements of statistical learning:data mining inference and prediction. Springer, Second Edition, 2009.

[8] L. Györfi, M. Kohler, A. Krzyzak, and H.A. Walk, "A Distribution-free theory of nonparametric regression. Springer, 2002.

[9] A. Bowman, and A. Azzalini, " Sm: Smoothing methods of nonparametric regression and density estimation. R Package version 2.2-5.7, 2021.

[10] Langrené, N., & Warin, X. Fast and stable multivariate kernel density estimation by fast sum updating. *Journal of Computational Graphical Statistics*, 28(3), 596-608.

[11] Yen-Chin, C. "STAT 415 introduction to nonparametric statistics, lecture 9: regression: regressogram and kernel regression, winter, 2018.

[12] Li, Q, Lin, J., & Racine, J. S. (2013). Optimal bandwidth selection for nonparametric distribution and quantile functions. *Journal of Business and Economic Statistics*, 31, 57-65.

[13] Noh, H. (2014). Estimation using kernel smoothing with data transformation. *Journal of the Korean Statistical Society*, 43, 503-514.

[14] Li, Q., Maasounmi, E., & Racine, J.S. (2009).A nonparametric test for equality of distributions with mixed categorical and continuous data. *Journal of Econometrics*, 148,186-200.

[15] Sparapani, R., Spanbauer, C., & McCulloch,R. (2021). Nonparametric machine learning and efficient computation with Bayesian additive regression trees: The BART R Package. *Journal of Statistical Software*, 97(1), 457-472.

[16] Erillis, N.A., & Alakus, K. (2014). Non-parametric regression estimation for data with equal values. *European Scientific Journal*, 10(4), 72-82.

[17] Walker, J., Poliziani, C., Tortora, C., Schweizer, J., & Rupi, F. (2022). Nonparametric regression analysis of cyclist waiting times across three behavioural typologies. *ISPRS International Journal of Geo-information*, 11(3), 3-14.

[18] Tsagris, M. Alenazi, A., & Stewart, C. (2023). Flexible non-parametric regression models for compositional response data with zeros. *Statistics and Computing*, 33(106), 1-17.

[19] B. E. Hansen. Lecture notes on nonparametrics. University of Wisconsin, Spring, 2009.

[20] J. Fox, and S. Weisberg, An R companion to applied regression. Sage, Thousand Oaks, CA, Third Edition, 2019.

[21] R. Tibshirani, and L. Wasserman, "Nonparametric regression: statistical machine learning," Spring, 2012.

[22] J.W. Tukey, "Curves as parameters and toch estimation," Proc. 4th Berkeley Symposium, 681-694, 1961.

[23] Tukey, J.W. (1947). Nonparametric estimation II. statistically equivalent blocks and tolerance Regions. *The continuous case, Annals of Mathematical Statistics*, 18, 529-539.

[24] Olanrewaju, R.O. (2020). Penalized likelihood estimation of gamma distributed response variable via corrected solution of regression coefficients. *Journal of Modern Applied Statistical Methods*, 19(1), 1-18.

[25] Olanrewaju, R.O., & Olanrewaju S. A. (2024). On the Efficiency of the newly proposed Olanrewaju-Olanrewaju $(L_{(O-O)\lambda_{Y(|\theta|)}})$ penalized regression-type estimator via GLMs technique. *Studies of Applied Economics*, 42(1), 1-23.

[26] L. Györfi, M. Kohler A. Krzyzak, and H.A. Walk "Distribution-free theory of nonparametric regression," Springer Science and Business Media, 2006.