

Article

# The Efficiency of the Proposed Smoothing Method over the Classical Cubic Smoothing Spline Regression Model with Autocorrelated Residual

Samuel Olorunfemi Adams<sup>1,\*</sup>, Omorogbe Joseph Asemota<sup>2</sup><sup>1</sup> Department of Statistics, University of Abuja, Abuja, Nigeria<sup>2</sup> Department of Economic and Social Research, National Institute for Legislative and Democratic Studies, Abuja, Nigeria

\*Correspondence: Samuel Olorunfemi Adams (samuel.adams@uniabuja.edu.ng)

**Abstract:** Spline smoothing is a technique used to filter out noise in time series observations when predicting nonparametric regression models. Its performance depends on the choice of the smoothing parameter. Most of the existing smoothing methods applied to time series data tend to over fit in the presence of autocorrelated errors. This study aims to determine the optimum performance value, goodness of fit and model overfitting properties of the proposed Smoothing Method (PSM), Generalized Maximum Likelihood (GML), Generalized Cross-Validation (GCV), and Unbiased Risk (UBR) smoothing parameter selection methods. A Monte Carlo experiment of 1,000 trials was carried out at three different sample sizes (20, 60, and 100) and three levels of autocorrelation (0.2, 0.5, and 0.8). The four smoothing methods' performances were estimated and compared using the Predictive Mean Squared Error (PMSE) criterion. The findings of the study revealed that: for a time series observation with autocorrelated errors,  $Adj. R^2(PSM \lambda = 0.04)$  provides the best-fit smoothing method for the model, the PSM does not over-fit data at all the autocorrelation levels considered ( $\rho = 0.2, 0.5$  and  $0.8$ ); the optimum value of the PSM was at the weighted value of 0.04 when there is autocorrelation in the error term, PSM performed better than the GCV, GML, and UBR smoothing methods were considered at all-time series sizes ( $T = 20, 60$  and  $100$ ). For the real-life data employed in the study, PSM proved to be the most efficient among the GCV, GML, PSM, and UBR smoothing methods compared. The study concluded that the PSM method provides the best fit as a smoothing method, works well at autocorrelation levels ( $\rho = 0.2, 0.5$ , and  $0.8$ ), and does not over fit time-series observations. The study recommended that the proposed smoothing is appropriate for time series observations with autocorrelation in the error term and econometrics real-life data. This study can be applied to; non – parametric regression, non – parametric forecasting, spatial, survival, and econometrics observations.

## How to cite this paper:

Adams, S. O., & Asemota, O. J. (2023). The Efficiency of the Proposed Smoothing Method over the Classical Cubic Smoothing Spline Regression Model with Autocorrelated Residual. *Journal of Mathematics Letters*, 1(1), 19–37. Retrieved from <https://www.scipublications.com/journal/index.php/jml/article/view/618>

## Academic Editor:

Ilyas Haouam

Received: January 20, 2023

Accepted: March 03, 2023

Published: March xx, 2023

**Keywords:** Cubic spline; goodness of fit; Generalized Maximum Likelihood (GML); Generalized Cross-Validation (GCV); and Mallows CP criterion (MCP)



**Copyright:** © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The smoothing spline is a spline consisting of piecewise third-degree polynomials that go through a bunch of control focuses. The second subsidiary of every polynomial is ordinarily set to zero at the endpoints since this gives a limit condition that finishes the framework condition of  $m - 2$  conditions. This creates a purported "normal" cubic spline and prompts a straightforward tri-diagonal framework that can be settled effectively to give the coefficients of the polynomials. The parameters are estimated by minimizing the residual sum of squares (RSS) and a roughness penalty. A general test of "loyalty to

observation" for a curve  $g$  is the residual sum of squares. If  $g$  is allowed to be any curve – unrestricted in functional form, then this distance test can be reduced to zero by any  $g$  that interpolates the observation. The curve would not be admitted because it is not exclusive and because it is a structure-oriented interpretation, [1]. The spline smoothing approach avoids impossible interpolation of the observation by evaluating the contest between the tasks of producing a good fit to the observation and producing a curve without too much rapid local change. The main function of splines is for interpolation, but they can also be used for parametric and non-parametric regression modeling. The most commonly used spline smoothing technique is the cubic splines. Spline smoothing produces another technique for local polynomial regression and it is also a charming component of additive regression models. It is well known that correlation greatly affects the selection of smoothing parameter which is critical to the performance of smoothing spline. The commonly used approach in time series analysis is the classical ARMA method. It assumes linear dependence on past values and past innovations. Generalized Cross Validation (GCV) and Generalized Maximum Likelihood (GML) are the most appropriate spline smoothing method for selecting optimal value for the smoothing parameter and a performance criteria for smoothing parameter selection. So many scholars had carried out research on this area, most of them discovered that time series data assume independence of regressors and error terms which lead to autocorrelation problem. The application of smoothing parameter estimators like GCV and GML do not always solve these problems because they don't occasionally smooth.

Over the last two decades, research on spline smoothing estimation methods has produced a vast amount of information and discoveries from researchers in evaluating the efficiency and performance of the existing estimation techniques when autocorrelation is present in their error terms. In this research work, the proposed smoothing method is compared with three classical smoothing spline parameter selection techniques with the intention of providing a robust smoothing parameter estimation method that will alleviate the problem of over fitting models for time-series data with low, moderate, and high autocorrelation levels and the problem associated with the smoothing methods' performance when different time series sample sizes are utilized.

In Section 2, the cubic smoothing spline was discussed, method of selecting smoothing parameters like Generalized Cross-Validation, Generalized maximum Likelihood, Mallows's CP criterion, and performance evaluation criteria were also discussed in this section. A simulation study and results are given in Sections. Finally, concluding remarks are presented in Section 4.

## 2. Literature Review

A lot of attention has been directed to studies on smoothing smoothing with autocorrelated error. [2] made a comparison between GCV and REML, it was recommended that GCV and REML are good smoothing parameter selection for small and medium-sized samples. [3] & [4] applied the smoothing spline method to fit a curve to a noisy data set, where the selection of the smoothing parameter is essential. An improved  $C_p$  criterion for spline smoothing based on Stein's unbiased risk estimate has been proposed to select the smoothing parameter. The resulting fitted curve is superior and more stable than commonly used selection criteria and possesses the same asymptotic optimality as  $C_p$ . [5] applied most of the data-driven smoothing parameter selection methods and compared them based on large and small sample sizes. The parallel of Akaike's information criterion and Generalized Cross-Validation is recommended as being the best selection criteria. For large samples, the  $GF_{AIC}$  method would seem to be more appropriate while for small samples they proposed the implementation of the GCV criterion. [6] investigates two types of results that support the use of GCV for variable selection under the assumption of sparsity. The first type of result is based on the well-established links between GCV on the one hand and Mallows's  $C_p$  and Stein Unbiased

Risk Estimator on the other hand. The result states that GCV performs as well as Cp or SURE in a regularized or penalized least squares problem as an estimator of the prediction error for the penalty in the neighborhood of its optimal value. [7], [8], and [9] investigated the behavior of the optimal values of gamma and rho to identify simple practical rules to choose their optimal properties. RGCV and modified GCV perform significantly better than GCV. The performance is defined in terms of the Sobolev error, which is shown by example to be more consistent with a visual assessment of the fit than the average squared error. [10-12] discussed UBR and GCV for selecting the optimal knots in spline. The criteria for selecting the best model were based on Mean Squared Error and R-square. The simulation was performed on a spline truncated function with error generated from a Normal distribution for varied sample sizes and error variance. The results of the simulation study showed that GCV estimates the knots more accurately than UBR. [13] considered nonparametric regression problems and developed a model-averaging procedure for smoothing spline regression problems. Model weights were estimated using a delete-one-out cross-validation procedure to minimize the prediction error. A simulation study was performed by using a program written in R. The simulation study provides a comparison of the most well-known CV, generalized GCV, and the proposed method. The model averaging approach is straightforward to implement and gives reliable performances in simulations.

It is clear from the existing literature that the goodness-of-fit of smoothing spline for time series observations has not been investigated so far. The paper aimed at presenting a goodness of fit test for time series observation using three classical cubic spline non-parametric regression functions.

### 3. Methodology

This section discussed the methodology applied in this research work.

#### 3.1. Cubic Smoothing Spline Regression Model

The most common example of the smoothing spline is the cubic spline; it is the smoothing spline's functional form and a piecewise cubic function that interpolates the dataset and ensures the smoothness of the observation. It is piecewise third-degree polynomials that go through a bunch of focuses. It has a nonstop first and second subordinate with the request for the coherence of  $(d-1)$ , where  $d$  is the polynomial degree. The Model with shortened force premise work  $b(x)$  changes the factors  $X_i$  by applying a premise work  $b(x)$  and fits a model utilizing these changed factors, which adds non-linearity to the model and empowers the splines to fit smoother and adaptable Non-straight capacities. The spline smoothing model is written as;

$$y_i = f(t_i) + \varepsilon_i \quad (1)$$

Where;  $y_i$  is the response variable,  $f$  is an unknown smoothing function,  $t_i$  is the independent/predictor variable and  $\varepsilon_i$  is zero means autocorrelated stationary process.

The general cubic spline function is given as;

$$f(t) = at^3 + bt^2 + ct + d + \varepsilon \quad (2)$$

Where;  $a, b, c,$  and  $d =$  real number coefficients and  $a \neq 0, t =$  independent variable,  $\varepsilon =$  error term, and d.f. =  $k-d-1$  ( $k =$  number of knots and  $d =$  degree of cubic spline)

The cubic spline smoothing estimate function is  $\hat{f}$  while;  $f$  refers to the minimizer of a twice differentiable function of;

$$S(f) = \sum_{i=1}^n (y_i - \hat{f}(t_i))^2 + \lambda \int_a^b (\hat{f}''(t))^2 dt \quad (3)$$

Where;

- $\lambda > 0$  is a smoothing parameter,
- The initial part in equation (3) refers to the residual sum of the square for the integrity of the information's attack.
- The roughness penalty in the subsequent term of equation (3), is enormous when the incorporated second subsidiary of a regression function  $f''(t)$  is likewise huge
- If  $\lambda$  moves toward 0, then  $f(t)$  only interpolates the data set.
- If  $\lambda$  is very big, then  $f(t)$  would be chosen wherefore  $f''(t)$  is wherever 0, which will suggest a by and large direct least-squares fit the perceptions.

If  $f(t)$  values are fixed at  $f(t_1), \dots, f(t_2)$  the roughness  $\int_a^b (\hat{f}''(t))^2 dt$  is minimized by a natural cubic spline, this solution is written as a basic function as;

$$f(t) = \beta_0 + \beta_1 f_1(t) + \dots + \beta_{n+3} f_{n+3}(t)$$

### 3.2. Generalized Cross-Validation (GCV) Estimation Method with an Autocorrelation Structure

The term Generalized Cross-Validation (GCV) was proposed by [14] and [17] as a replacement for Cross-Validation (CV), it is the most popular method for choosing the complexity of statistical models. The basic principle of cross-validation is to leave the data points out one at a time and to choose the value of  $\lambda$  under which the missing data points are best predicted by the remainder of the data. To be precise, let  $g_\lambda^{-1}$  be the smoothing spline determined from all the information sets aside from  $(t_i, y_i)$ , utilizing the worth  $\lambda$  for the smoothing boundary. The cross-validation decision regarding  $\lambda$  can then be the estimation of  $\lambda$  which can be written as;

$$CV(\lambda) = \frac{1}{n} \sum \{y_i - \hat{g}(t_i)\}^2 \tag{4}$$

Equation (4) is similar to the test for regression model estimation [16]. Define a matrix  $A(\lambda)$  by;

$$A_{ij}(\lambda) = n^{-1} g(t_i, t_j) \tag{5}$$

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \frac{\{y_i - \hat{g}(t_i)\}^2}{\{1 - A_{ii}(\lambda)\}^2} \tag{6}$$

Wang, Meyer & Opsomer (2013) [15] also proposed the application of a related test referred to as the Generalized Cross-validation, acquired from equation (6) by substituting  $A_{ii}(\lambda)$  with its mean value,  $n^{-1} trA(\lambda)$ , this gives the score.

$$GCV(\lambda) = \frac{n^{-1} RSS(\lambda)}{(1 - n^{-1} trA(\lambda))^2} \tag{7}$$

Where;  $RSS(\lambda)$  refers to the residual sum of squares. [17] also gave hypothetical contentions to prove that GCV ought to pick an ideal estimation of  $\lambda$  in the function of minimizing the mean squared error (MSE) at the design points. The forecast published practical examples bear out a good performance in [18]. The summed-up Cross-validation technique is notable for its optimal qualities [19]. For any given  $n \times n$ , the impact matrix is given as;

$$\begin{bmatrix} \hat{f}_n, \lambda(t_1) \\ \hat{f}_n, \lambda(t_2) \\ \vdots \\ \hat{f}_n, \lambda(t_n) \end{bmatrix} = S(\lambda)y, \tag{8}$$

therefore  $W_0(\lambda)$  can be revised as;

$$W_0(\lambda) = \frac{\sum_{k=1}^n (a_{kj}y_j - y_k)^2}{(1 - a_{kk})^2} \tag{9}$$

Where; Generalized Cross-Validation is the changed type of Cross-Validation, a customary method for assessing the smoothing boundary. The GCV score is built by correlation with the CV score which is gotten from the normal residuals by dividing them by  $1 - (S_\lambda)_{ii}$ . The acknowledged arrangement of GCV is to replace the documentation  $1 - (S_\lambda)_{ii}$  in Cross-Validation with a mean score of  $1 - n^{-1} \text{trace}(S_\lambda)$ . Consequently, by adding the residual squared and notation  $\{1 - n^{-1} \text{trace}(S_\lambda)\}^2$ , by the known conventional cross-approval, the GCV smoothing technique is composed numerically as;

$$GCV(\lambda) = \frac{1}{n} \frac{\sum_{k=1}^n \{y - f_k(x_k)\}^2}{\{1 - n^{-1} \text{trace}(S_\lambda)\}^2} \tag{10}$$

$$GCV(\lambda) = \frac{n^{-1} \|(I - S_\lambda)y\|^2}{[n^{-1} \text{trace}(I - S_\lambda)]^2} \tag{11}$$

Where;  $n$  is observations or data set,  $\lambda$  is the smoothing parameter,  $S_\lambda$  refers to the  $i$ th diagonal member of the smoothing matrix

The first research on cross-validation was conducted by [20], which was subsequently augmented to the log periodogram's smoothing [21]. The term Generalized Cross-Validation (GCV) was determined by [20]. The GCV score figured by similarity to the CV score can be obtained from the normal residuals by isolating them by  $1 - (S_\lambda)_{ii}$ . The essential plan of GCV is to supplant the components  $1 - (S_\lambda)_{ii}$  with the mean score  $1 - n^{-1} \text{tr}(S_\lambda)$ . Consequently, adding the squared revised remaining and factor  $\{1 - n^{-1} \text{tr}(S_\lambda)\}$ . Given the spline smoothing for non-parametric assessment of a relapse work in a period series setting and accepting that the reaction variable  $y_i$  are taken on the occasion  $t_i$ , for  $i = 1, \dots, n$  and that a model of the structure creates the  $y_i$

$$y_i = f(t_i) + Z(t_i) \tag{12}$$

Where  $f(\cdot)$  refers to the smoothing function and  $Z(t_i)$  refers to the zero-mean, Autocorrelated stationary process. It can be said that even though  $t_i$  is specific, it is not uniformly spaced, with  $t_1 < \dots < t_n$

If the  $Z(t_i)$  in (12) has a known correlation function, with  $CovZ(t_1, \dots, t_n) = \sigma^2 v_{ij}$ , a normal addition of the usual smoothing spline approach amongst is to estimate  $f$  by the  $\hat{f}$  which minimizes;

$$(y - \hat{f})^T W (y - \hat{f}) + \lambda \int_a^b \{f''(t)\}^2 dt \tag{13}$$

Amid every properly smoothed function  $f$ , it is confirmed that  $W = V - 1 = [v_{ij}]$ ,  $y = (y_1, \dots, y_n)^T$  and  $f = (f(t_1), \dots, f(t_n))^T$ . It has been proven that the function  $\hat{f}$  remains a natural cubic spline that has knots at the  $t_i$ . Also, if  $\hat{f}$  denotes the vector with the  $i$ th element  $\hat{f}(t_i)$  then there is a matrix  $S_\lambda$  such that  $\hat{f} = S(\lambda)y$ , i.e. for fixed  $\lambda$ , the

estimate is a direct capacity of  $y$ . This linearity proposes a nearby association between spline smoothing and bit smoothing, as shown unequivocally in [22]. One approach to picking the parameter denoted by  $\lambda$  is for the generalized cross-correlation to be minimized [17]. In the current setting, the common extension of this model is to limit equation (13); this gives a technique for assessing  $g$  within the sight of a realized autocorrelation structure. Concerning span assessment of  $g$ , the Bayesian assumption presented by [22] extends with the connection network and  $V$  is replaced by Silverman's inverse inclining weighting lattice, which presents the posterior difference matrix, written as;

$$Var(\hat{f}) = \sigma^2 A(\lambda) V \tag{14}$$

The minimization of GCV ( $\lambda$ ) as proposed by [23] and [24] is written as;

$$GCV(\lambda) = \frac{(y - \hat{f})^T W (y - \hat{f})}{[\text{trace}(I - S\lambda)]^2} \tag{15}$$

Where;  $(S\lambda)$  is the  $i$ th diagonal element of the smoother matrix,  $W = V - 1 = [v_{ij}]$ , the correlation function,  $y$  is  $(y_1, \dots, y_n)^T$ ,  $f = (f(t_1), \dots, f(t_n))^T$

### 3.3. Generalized Maximum Likelihood (GML) Estimation Method with Autocorrelation Structure

[25] proposed the GML technique for correlated data that possesses a single parameter for smoothing observations. However, there exist two parameters for smoothing in the case of a bivariate model that should be assessed along with the covariance boundaries. Following a comparative determination, GML is given as;

$$GML(\lambda) = \frac{y^i(I - S\lambda)}{[\det^+(I - S\lambda)]^{\frac{1}{n-m}}} \tag{16}$$

Where;  $\det^+(I - S\lambda)$  refers to the product of  $(n - m)$  non-zero eigenvalues of  $(I - S\lambda)$ . [25] provided a Bayesian model for the GML method's general framework and can calculate a spline estimate's posterior confidence intervals. Suppose that the data are simulated via the;

$$y_i = f(t_i) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad t_i \in [0, 1] \tag{17}$$

Where;  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \sim N(0, \sigma^2, W^{-1})$  which do not dependent on  $f$ , Model (17) is usually referred to as a Bayesian model, it can also be known as a hierarchical model or a mixed-effects model. This Bayesian model is similar to the model illustrated by [19], though the residuals are correlated. Based on the justification of [19], it can be shown that;

$$\lim_{n \rightarrow \infty} E(f(t)/y) = \hat{f}t \text{ and } \lim_{n \rightarrow \infty} cov(f/y) = \sigma^2, W^{-1} \tag{18}$$

Where;  $F = (F(t_1), \dots, F(t_n))'$  and  $\sigma \rightarrow \infty$  expanded prior are estimated for polynomial coefficients with degrees smaller than  $m$ .

According to [25], the covariance matrix  $W^{-1}$  relies on several correlations with parameter vector of  $\tau$ . Interestingly, covariance structures refer to first-order autoregressive for time-series observation, structured symmetry or unstructured for repeated measurements, and spatial data. GML with Autocorrelation structure is therefore given by;

$$GML(\lambda) = \frac{\lambda^i W(I - S\lambda)}{[\det^+ W(I - S\lambda)]^{\frac{1}{n-m}}} \tag{19}$$

Where;  $\det^+(I - S\lambda)$  is the product of the  $n - m$  nonzero eigenvalues of  $(I - S\lambda)$ ,  $\lambda$  is the Smoothing parameter,  $W$  is the structure of the correlation,  $S\lambda$  is the smoother matrix diagonal elements,  $n = n_1 + n_2$  are the pair of observations and  $m =$  the number of zero eigenvalues.

**3.4. Unbiased Risk (UBR) Estimation Method with Autocorrelation Structure**

Unbiased Risk is also known as Mallows’ CP criterion; it was developed by [26] to evaluate the regression model fit dependency on Ordinary Least Square (OLS). It is used to estimate choice situations where explanatory variables can predict a few results and locate the best model associated with the subset of independent variables. The more modest the estimation of the Cp, the generally exact it is, the Cp is written numerically as;

$$UBR(\lambda) = \frac{\|(S\lambda - I)y\|^2}{tr(I - S\lambda)} \tag{20}$$

[25] provides the UBR technique that can be used effectively to choose a smoothing parameter for cubic spline smoothing that possesses non-Gaussian information. It was developed by using Predictive Mean Square Errors (PMSE).

The Unbiased Risk with Autocorrelation structure can be written mathematically as;

$$UBR(\lambda) = \frac{\frac{1}{n} \|W^{\frac{k}{2}}(I - S\lambda)y\|^2}{\left[\frac{1}{n} trace(W^{k-1}(I - S\lambda))\right]^2} \quad k = 0, 1, 2 \tag{21}$$

Where;  $n$  is the measurement/observations  $\{x_i, y_i\}$ ,  $W$  is the Autocorrelation structure,  $\lambda$  is the parameter used for smoothing and  $S\lambda$  is the matrix smoother of the  $i$ th diagonal member.

**3.5. Proposed Smoothing Method (PSM) with Autocorrelation Structure**

A smoothing spline model is usually written as:

$$y_i = f(x_i) + \varepsilon_i \tag{22}$$

Where;  $y$  refers to the response variable,  $x$  refers to a predictor variable,  $f$  is the Regression function and  $\varepsilon_i \sim N(N(0, \sigma W^{-1}))$ .

There are several options to examine whenever model (22) is used for non-linearity, it incorporates, observation change, and adds substance items, for example, cubic spline and Spline smoothing. This research work is keen on spline smoothing because it examines non-linearity dependent on regression bend by presenting a wrinkle or twists in these crimps created by pivot work, and the place of the turn on the fit is called hitches.

The traditional regression analysis’s primary purpose is to minimize the residual sum of squares (RSS); the model with the minimum RSS is the preferred model. It is important to note that [27] proposed Cross-Validation (CV) as a technique for estimating Spline Smoothing. Instead of RSS in the customary straightforward simple regression, the residual is characterized.

In this manner, an improved spline smoothing technique is proposed by adding the weighted parameters  $k$  and  $k - 1$  with the other properties and qualities of the UBR and GCV [28] and [29]. The combination of the two smoothing methods’ quantities will result in the optimal performance of smoothing methods whose model does not overfit time-series observations. The minimizer is the Proposed Smoothing Method (PSM) with autocorrelation structure given as;

PSM = ( $k$ ) overfitting and optimal knot detector +  $(1 - k)$  best for forecasting non-Gaussian data

$$PSM(\lambda) = k \frac{(y - \hat{f})^T W (y - \hat{f})}{[\text{trace}(I - S\lambda)]^2} + (1 - k) \frac{\frac{1}{n} \left\| W^{\frac{g}{2}} (I - S\lambda) \right\|^2}{\left[ \frac{1}{n} \text{trace}\{W^{g-1}(I - S\lambda)\} \right]^2} \quad (23)$$

The behavior of the minimized  $\lambda$  in UBR and GCV techniques under the alternate value of  $g = 1$  as the optimum value of PSM yields;

$$PSM(\lambda) = k \frac{(y - \hat{f})^T W (y - \hat{f})}{[\text{trace}(I - S\lambda)]^2} + (1 - k) \frac{\frac{1}{n} \left\| W^{\frac{1}{2}} (I - S\lambda) \right\|^2}{\left[ \frac{1}{n} \text{trace}\{W(I - S\lambda)\} \right]^2} \quad (24)$$

The proposed method for estimating  $f$  is given in (27) subject to the condition that  $0 < g < 1$  is chosen, using the algorithm in section 3.6. [30-32].

Where;  $n$  is the number of the dataset,  $k$  is the weighted value,  $0 < k < 1$ ,  $W = V^{-1} = \text{Correlation Matrix for the error term}$ ,  $y = (y_1, \dots, y_n)^T$  is the Smoothing function,  $\hat{f} = (f(t_1), \dots, f(t_n))^T = S\lambda y$ ,  $S\lambda$  is the diagonal member of the smoothing matrix,  $\left\| W^{\frac{1}{2}} (I - S\lambda) y \right\| = \text{norm of the Euclidean vector } W^{\frac{1}{2}}(y - \hat{f})$ .

### 3.6. Proposed Smoothing Method (PSM) Algorithm

Step 1: Read the simulated sample data  $(x_i, y_i)$  for  $i = 1 - T$  and for each of the  $i$ 's determine the Pre-selected smoothing parameters  $\lambda_1, \dots, \lambda_t$ , calculate the respective set of smoothing Spline estimates  $f(\lambda) = \{\hat{f}_{\lambda_1}, \dots, \hat{f}_{\lambda_t}\}$

Step 2: For the given  $\lambda$ ,  $\sigma$  and  $T$  use the data in 1 above to fit a curve and the estimate ahead by linear extension  $f(x_i)$  and  $\hat{f}(x_i)$

Step 3: Insert the weighted value ( $k$ ) of the coefficients of GCV and UBR

Step 4: Obtain the predictive mean square error  $PMSE(\hat{f}\lambda) = \sum_{i=1}^t \left[ (f(x) - \hat{f}(x_i))^2 \right]$  for these points

Step 5: add all values of PMSEs to get the resulting PSM value for the given  $\lambda$  and  $q$ .

Step 6: Repeat steps 1–5 for 1000 times.

### 3.7. Monte Carlo Simulation study

This part is concerned with the outcome of a Monte Carlo simulation study. This study was led to assess the achievement of the four smoothing techniques depicted in this research, for example, GML, GCV, UBR, and PSM. The dataset was generated by applying a program written in R (version 3.2.3) for time-series sample sizes of; 20, 60, and 100. The experiment was replicated at 1,000 for every one of the examples. The Predictive Mean Squared-Errors (PMSE), adjusted R-Square and predicted R-square was utilized to assess the smoothing techniques' quality and performance for each simulated data.

### 3.8. Equation used to generate the value in the data

The data generation study performed to assess and measure the performance of the four spline smoothing methods is given as;

$$y(t) = 2\text{Sin}\left(\frac{\pi}{t}\right) + \varepsilon_t \quad t = 20, 60, \text{ and } 100 \quad (25)$$

Where;  $\pi = 180^\circ$ ,  $\varepsilon_t \sim N(0, \sigma W^{-1})$ , a first-order autoregressive process AR (1) with a mean of 0, a standard deviation of 0.8, and autocorrelation levels ( $\rho$ ) of 0.2, 0.5, and 0.8 with a 95% confidence limit, Note that;  $e_t = \rho\varepsilon_{t-1} + v_t$  and  $v_t \sim N(0, \sigma^2)$

### 3.9. Experimental design and data generation

The experimental design adopted in this study is;

- Three-time-series samples (T) of 20, 60, and 100 were considered in the data generation
- Three autocorrelation levels were considered, i.e.  $\rho = 0.2, 0.5$  and  $0.8$
- One standard deviation value was considered, i.e.  $\sigma = 0.8$
- The dataset was simulated for 1,000 replications in each of the  $3 \times 3 \times 4 \times 1 = 36$  combinations for cases T's,  $\rho$ 's,  $\lambda$ 's, and  $\sigma$ 's.

All the selected parameters the in experimental design are similar to the ones used in [33].

### 3.10. Smoothing Spline Assessment methods used in this Study

Efforts were made in this study to examine and compare the strength of the four spline smoothing estimators, namely; Generalized Cross-Validation (GCV), Unbiased risk (UBR), Generalized Maximum Likelihood (GML), and the Proposed Smoothing Method (PSM) developed by taking the weighted hybrid of GCV and UBR.

#### (i) Predictive Mean Square Error

A comparison was made to test the four estimation methods' effect and performance in the presence of autocorrelation error. An estimate of the four smoothing methods, the criterion, effect, and performance of different autocorrelation errors of the four estimation methods (i.e. Generalized Cross Validation (GCV), Generalized Maximum Likelihood (GML), Proposed Smoothing Methods (PSM) ( $0 < k < 1$ ) and Unbiased Risk (UBR)) were performed using codes written in R-console. Four different estimation methods were used i.e. GCV (V), GML (M), PSM ( $0 < k < 1$ ), and UBR (U). This data generation was carried out for V, M, P, and U. At the same time, the Evaluation and comparison of the Four (4) Spline Smoothing estimation methods were investigated by applying the asymptotic sampling qualities of the criterion given as; Mean Square Prediction Error (MSPE).

The Predictive Mean squared error (PMSE) of a smoothing curve or model fitting process, according to [26] and [4], is the difference between the expected value of the square difference of the fitted value, that is; function  $\hat{f}(x_i)$  and the observed value estimate is given as the function  $f(x_i)$ . It is utilized to estimate the performance and attributes of smoothing methods like Cross-Validation, Generalized Cross-Validation, and Generalized Maximum Likelihood, etc. The Predictive Mean Square Error (PMSE) is written numerically as;

$$PMSE(\lambda) = E \left[ \sum_{i=1}^n (f(x_i) - \hat{f}(x_i))^2 \right] \quad (26)$$

The Predictive Mean Square Error is usually separated into two parts; the initial part is the sum of square biases of the fitted qualities, and the other part is the number of changes in the fitted observations.

Where;

$f(x_i)$  = observed value

$\hat{f}(x_i)$  = predicted or estimated value

At each scenario of specification, for instance, say, time-series size (T) = 20, autocorrelation level ( $\rho$ ) = 0.2, d.f = 1, and standard deviation ( $\sigma$ ) = 0.8, the smoothing

methods were tested and compared using the asymptotic properties of the estimators based on the PMSE criterion.

### (ii) Test for Over-fitting in Spline Smoothing

In statistics, overfitting occurs when a model fails to fit extra information or neglects to anticipate future perceptions reliably. PRESS and Predicted R-square are the best and easiest ways to discover overfitted in smoothing methods and models. The result may be interpreted by simply comparing the predicted R-Square to the normal R-Square and observing if there exist a great difference between the two test techniques. If there is a large difference between the two values, the model doesn't predict new observations and fits the true data, and there is the possibility of overfitting the model. Overfit model has too many numbers and terms and begins to fit the random noise in the sample; it is not possible to predict random noise. The Predictive R-square is a statistical technique that determines how well a model predicts a response for new observation. It is something of an in-house cooked measure, which is computed by effectively eradicating each variable from the data set, estimating the regression model, and deciding how suitable the model forecasts the removed variables. Predictive R-square is usually written mathematically as;

$$\text{Pred. R - Square} = \left(1 - \frac{\text{Predicted residual sum of squares (PRESS)}}{\text{Sum of square total}}\right) \times 100 \quad (27)$$

While R-square, also known as the coefficient of determination, can be derived through;

$$R - \text{Square} = \left(1 - \frac{\text{Sum of squares error}}{\text{Sum of square total}}\right) \quad (28)$$

### (iii) Test for Goodness-of-fit for the Smoothing Methods

The goodness-of-fit of the smoothing methods explains how well the methods fit the simulated and real-life data. It also summarizes the differences between the observed value and predicted or estimated values. The Adjusted R-square was used to determine the best-fit smoothing methods. It is written mathematically as;

$$\text{Adjusted R - Square} = \left(1 - \frac{(1-R\text{square}) \times (n-1)}{n-p}\right) \quad (29)$$

Where; n = several observations and p = many parameters.

## 4. Results and Discussion

Tables 1 to 3 present the summary fit result of the smoothing spline regression model and the model performance criteria, i.e. the PMSE, multiple R-square, adjusted R-Square, and predicted R-square based on time-series periods ( $T=60$ ), four degrees of smoothing ( $D.S.=1, 2, 3$  and  $4$ ) and autocorrelation level ( $\rho = 0.5$ ). It was revealed from the result that all the coefficients of the smoothing methods' parameters were significant at ( $P$ -value  $<0.001, <0.01, \text{ and } <0.05$ ).

The PMSE of the four smoothing techniques indicated that; the Proposed Smoothing Method ( $PSM = 0.18$ ) had the smallest PMSE of  $0.757980$  at  $T = 60, D.S. = 2, \text{ and } \rho=0.5$ . This was closely followed by,  $UBR$  with PSME of  $1.017353$  at  $T = 60, D.S. = 2, \text{ and } \rho = 0.5$  then,  $GML$  with PSME of  $1.300494$  at  $T = 60, D.S. = 2 \text{ and } \rho = 0.5$ . The result implies that; the Proposed Smoothing Method ( $PSM = 0.18$ ) performs better than the other smoothing methods at a time series size ( $T=60$ ) and  $\rho = 0.5$ .

The adjusted R-Square result showed that the Proposed Smoothing Method ( $PSM = 0.18$ ) had the largest values of  $0.8095$  at,  $T = 60, D.S. = 2, \text{ and } \rho=0.5$ , which is closely followed by the Proposed Smoothing Method ( $PSM = 0.20$ ) with the value of  $0.7879$  when  $T = 20$ ,

$D.S. = 2$  and  $\rho=0.5$ , then GCV smoothing method with  $0.7828$  at  $T = 60$ ,  $D.S. = 2$  and  $\rho = 0.5$ . It can be inferred from the result above that; the Proposed Smoothing Method (PSM =  $0.18$ ), provides the best fit to the time-series observations at a time-series size ( $T = 60$ ) and  $\rho = 0.5$ .

It can be seen from the result presented in Table 1-3 that the difference between the Multiple R-square and predictive R-square of the Proposed Smoothing Method was the least when compared to the other smoothing methods. At  $T = 60$ ,  $D.S. = 1, 2, 3,$  and  $4, \rho = 0.5$ , the differences between the Multiple R-Square and predictive R-square were  $0.3669, 0.0364, 0.4599,$  and  $0.1759$  respectively. This result shows that the Proposed Smoothing Method does not overfit the time series observations when the time-series size of  $60$  and  $\rho = 0.5$ .

**Table 1. Simulation Result for Smoothing Spline Regression Model of GML, GCV, PSM, and UBR for T=20, and  $\rho = 0.2, 0.5$**

Conditions	Smoothing Methods	$\beta_i$	Smoothing Parameter	Selection criteria			
				PMSE	R Square	Adj. R Square	Pred. R Square
T = 20, $\rho = 0.2$	GML	$\beta_0 = 0.5596$ $\beta_1 = 5.9713^{**}$ $\beta_2 = -23.5218^*$ $\beta_3 = 17.3125^{**}$	0.04384406	1.34165	0.6663	0.6037	0.000
T = 20 $\rho = 0.2$	GCV	$\beta_0 = -0.2034$ $\beta_1 = 14.7471^{**}$ $\beta_2 = -42.272^{***}$ $\beta_3 = 27.8714^{***}$	0.0001058604	0.956241	0.6884	0.6299	0.0009
T = 20 $\rho = 0.2$	UBR	$\beta_0 = -0.1656$ $\beta_1 = 9.2593$ $\beta_2 = -25.7615^*$ $\beta_3 = 16.5350^*$	0.06503716	2.59853	0.5401	0.4538	0.000
T = 20, $\rho = 0.2$	PSM (K =0.04)	$\beta_0 = -0.1376$ $\beta_1 = 12.0791^{***}$ $\beta_2 = -35.2248^{***}$ $\beta_3 = 234056^{***}$	0.06243990	0.046857	0.9678	0.9618	0.6428
T = 20, D.S. = 4, $\rho = 0.5$	GML	$\beta_0 = 0.1047$ $\beta_1 = 11.7009^{***}$ $\beta_2 = -36.7510^{***}$ $\beta_3 = 25.1998^{***}$	0.02464804	1.47087	0.7143	0.6608	0.0009
T = 20, D.S. = 4, $\rho = 0.5$	GCV	$\beta_0 = 0.1513$ $\beta_1 = 9.9033^{**}$ $\beta_2 = -30.5509^{***}$ $\beta_3 = 20.5665^{***}$	0.00012548	1.437364	0.7653	0.7213	0.020
T = 20, D.S. = 4, $\rho = 0.5$	UBR	$\beta_0 = -0.1341$ $\beta_1 = 10.1868^{***}$ $\beta_2 = -31.8230^{***}$ $\beta_3 = 21.9838^{***}$	0.2490225	1.14616	0.7363	0.6868	0.0013
T = 20, D.S. = 4, $\rho = 0.5$	PSM (K =0.16)	$\beta_0 = -0.3008$ $\beta_1 = 10.4908^{***}$ $\beta_2 = -28.9898^{***}$ $\beta_3 = 18.8467^{***}$	0.20919897	1.794844	0.8802	0.8577	0.742
T = 20, D.S. = 4, $\rho = 0.8$	GML	$\beta_0 = 0.3308$ $\beta_1 = 9.6598$ $\beta_2 = -32.6249^*$ $\beta_3 = 23.2359^{**}$	0.0150933	2.16225	0.5406	0.4545	0.000

T = 20, D.S. = 4, $\rho = 0.8$	GCV	$\beta_0 = 0.5529$ $\beta_1 = 8.6756$ $\beta_2 = -30.8329^*$ $\beta_3 = 21.8131^{**}$	0.0006287	2.188967	0.6508	0.5854	0.122
T = 20, D.S. = 4, $\rho = 0.8$	UBR	$\beta_0 = 0.02747$ $\beta_1 = 13.26481^{**}$ $\beta_2 = -42.53308^{***}$ $\beta_3 = 30.18391^{***}$	0.6178076	2.49091	0.7091	0.6545	0.000
T = 20, D.S. = 4, $\rho = 0.8$	PSM (K =0.28)	$\beta_0 = -0.04164$ $\beta_1 = 10.37883^*$ $\beta_2 = -32.14938^{**}$ $\beta_3 = 21.52423^{**}$	0.44499741	1.483121	0.7442	0.6962	0.3567

**Table 2. Simulation Result for Cubic Spline Regression Model of GML, GCV, PSM, and UBR for T=60, and  $\rho = 0.2$**

Parameters	Smoothing Methods	$\beta_i$	Smoothing Parameters	Selection criteria			
				PMSE	R Square	Adjusted R-Square	Pred. R -Square
T = 60, D.S. = 4, $\rho = 0.2$	GML	$\beta_0 = -0.02920$ $\beta_1 = 10.77812^{***}$ $\beta_2 = -31.19231^{***}$ $\beta_3 = 21.57675^{***}$	0.001546984	3.615946	0.9478	0.945	0.000
T = 60, D.S. = 4, $\rho = 0.2$	GCV	$\beta_0 = -0.14492$ $\beta_1 = 10.83452^{***}$ $\beta_2 = -31.70387^{***}$ $\beta_3 = 21.11344^{***}$	6.08904e-05	2.018062	0.9454	0.9425	0.022
T = 60, D.S. = 4, $\rho = 0.2$	UBR	$\beta_0 = -0.11607$ $\beta_1 = 11.22499^{***}$ $\beta_2 = -33.07054^{***}$ $\beta_3 = 22.13243^{***}$	0.04321843	3.398581	0.9553	0.9529	0.000
T = 60, D.S. = 4, $\rho = 0.2$	PSM (K =0.08)	$\beta_0 = -0.25494^{**}$ $\beta_1 = 12.44873^{***}$ $\beta_2 = -35.69443^{***}$ $\beta_3 = 23.63443^{***}$	0.039765827	2.416724	0.9459	0.9422	0.7752
T = 60, D.S. = 4, $\rho = 0.5$	GML	$\beta_0 = -0.2659$ $\beta_1 = 12.3001^{***}$ $\beta_2 = -35.1093^{***}$ $\beta_3 = 23.1599^{***}$	0.001259269	2.16225	0.7355	0.7213	0.000
T = 60, D.S. = 4, $\rho = 0.5$	GCV	$\beta_0 = -0.2518$ $\beta_1 = 12.452^{***}$ $\beta_2 = -36.210^{***}$ $\beta_3 = 25.1076^{***}$	9.37482e-05	2.188967	0.7384	0.7453	0.021
T = 60, D.S. = 4, $\rho = 0.5$	UBR	$\beta_0 = -0.3808$ $\beta_1 = 13.5769^{***}$ $\beta_2 = -38.1391^{***}$ $\beta_3 = 25.0224^{***}$	0.02952382	2.49091	0.7791	0.7672	0.000
T = 60, D.S. = 4, $\rho = 0.5$	PSM (K =0.20)	$\beta_0 = -0.2591$ $\beta_1 = 11.4306^{***}$ $\beta_2 = -32.4205^{***}$ $\beta_3 = 21.5264^{***}$	0.023637806	1.733162	0.7879	0.7766	0.612

T = 60, D.S. = 4, $\rho = 0.8$	GML	$\beta_0 = 0.1528$ $\beta_1 = 10.3817^{***}$ $\beta_2 = -32.6398^{***}$ $\beta_3 = 22.4550^{***}$	0.001038157	2.996486	0.5791	0.5565	0.000
T = 60, D.S. = 4, $\rho = 0.8$	GCV	$\beta_0 = 0.1035$ $\beta_1 = 8.5890^{***}$ $\beta_2 = -25.6512^{***}$ $\beta_3 = 16.6920^{***}$	9.757026e-05	1.456264	0.5882	0.5661	0.010
T = 60, D.S. = 4, $\rho = 0.8$	UBR	$\beta_0 = -0.5286$ $\beta_1 = 15.0744^{***}$ $\beta_2 = -41.1956^{***}$ $\beta_3 = 22.2068^{***}$	0.05817642	2.564013	0.5723	0.5494	0.000
T = 60, D.S. = 4, $\rho = 0.8$	PSM (K =0.32)	$\beta_0 = 0.4314$ $\beta_1 = 8.4836^{**}$ $\beta_2 = -30.6451^{***}$ $\beta_3 = 22.0310^{***}$	0.039591188	1.0654	0.6401	0.6208	0.433

**Table 3. Simulation Result for Cubic Spline Regression Model of GML, GCV, PSM, and UBR for T=100, and  $\rho = 0.2$**

Parameters	Smoothing Methods	$\beta_i$	Smoothing Parameters	Selection criteria			
				PMSE	R Square	Adjusted R-Square	Pred. R -Square
T = 100, D.S. = 4, $\rho = 0.2$	GML	$\beta_0 = -0.12003^*$ $\beta_1 = 10.88492^{***}$ $\beta_2 = -31.70714^{***}$ $\beta_3 = 21.05287^{***}$	0.000955618	1.973208	0.9300	0.9279	0.000
T = 100, D.S. = 4, $\rho = 0.2$	GCV	$\beta_0 = -0.11858$ $\beta_1 = 11.47119^{***}$ $\beta_2 = -33.74040^{***}$ $\beta_3 = 22.58175^{***}$	0.000054114	0.332736	0.9474	0.9458	0.000
T = 100, D.S. = 4, $\rho = 0.2$	UBR	$\beta_0 = -0.19474^{**}$ $\beta_1 = 11.64251^{***}$ $\beta_2 = -33.84756^{***}$ $\beta_3 = 22.51969^{***}$	0.02506713	1.337717	0.9535	0.9521	0.011
T = 100, D.S. = 4, $\rho = 0.2$	PSM (K =0.12)	$\beta_0 = -0.20195^{**}$ $\beta_1 = -12.07546^{***}$ $\beta_2 = -35.23636^{***}$ $\beta_3 = 23.62322^{***}$	0.022065568	1.207069	0.9353	0.9333	0.5438
T = 100, D.S. = 4, $\rho = 0.5$	GML	$\beta_0 = -0.1604$ $\beta_1 = 11.3986^{***}$ $\beta_2 = -32.6706^{***}$ $\beta_3 = 21.5384^{***}$	0.000842476	1.984518	0.7532	0.7455	0.009
T = 100, D.S. = 4, $\rho = 0.5$	GCV	$\beta_0 = -0.2635$ $\beta_1 = 12.6782^{***}$ $\beta_2 = -36.9091^{***}$ $\beta_3 = 24.7118^{***}$	0.0001544047	2.765412	0.7876	0.7803	0.0112
T = 100, D.S. = 4, $\rho = 0.5$	UBR	$\beta_0 = -0.4740^{**}$ $\beta_1 = 14.5119^{***}$ $\beta_2 = -41.2996^{***}$ $\beta_3 = 27.7351^{***}$	0.02684178	1.815722	0.7768	0.7698	0.000

T = 100, D.S. = 4, $\rho = 0.5$	PSM (K =0.24)	$\beta_0 = -0.08351$ $\beta_1 = 12.12945^{***}$ $\beta_2 = -35.37438^{***}$ $\beta_3 = 23.59913^{***}$	0.02043681	1.739182	0.8046	0.7984	0.2133
T = 100, $\lambda = 4$ , $\rho = 0.8$	GML	$\beta_0 = 0.3072$ $\beta_1 = 6.7774^{***}$ $\beta_2 = -24.1129^{***}$ $\beta_3 = 17.3734^{***}$	0.0007463417	0.19754	0.5267	0.5119	0.000
T = 100, $\lambda = 4$ , $\rho = 0.8$	GCV	$\beta_0 = -0.09507$ $\beta_1 = 13.05444^{***}$ $\beta_2 = -38.07188^{***}$ $\beta_3 = 25.36748^{***}$	0.0001467109	2.6543	0.6445	0.6333	0.013
T = 100, $\lambda = 4$ , $\rho = 0.8$	UBR	$\beta_0 = -0.2412$ $\beta_1 = 14.2442^{***}$ $\beta_2 = -38.4952^{***}$ $\beta_3 = 24.5294^{***}$	0.01991017	1.532589	0.5904	0.5776	0.071
T = 100, $\lambda = 4$ , $\rho = 0.8$	PSM (K =0.36)	$\beta_0 = -0.8702$ $\beta_1 = 10.95087^{***}$ $\beta_2 = -33.32863^{***}$ $\beta_3 = 22.61821^{***}$	0.012795325	0.056745	0.849	0.8207	0.4532

Figures 1 to 3 below clearly show the comparisons of the behaviors of the cubic smoothing spline selected by GCV, GML, and MCP for sample sizes 20, 60, and 100 respectively. It was observed that the observed value of Generalized Cross-Validation was closer to the fitted/estimated value when compared to the

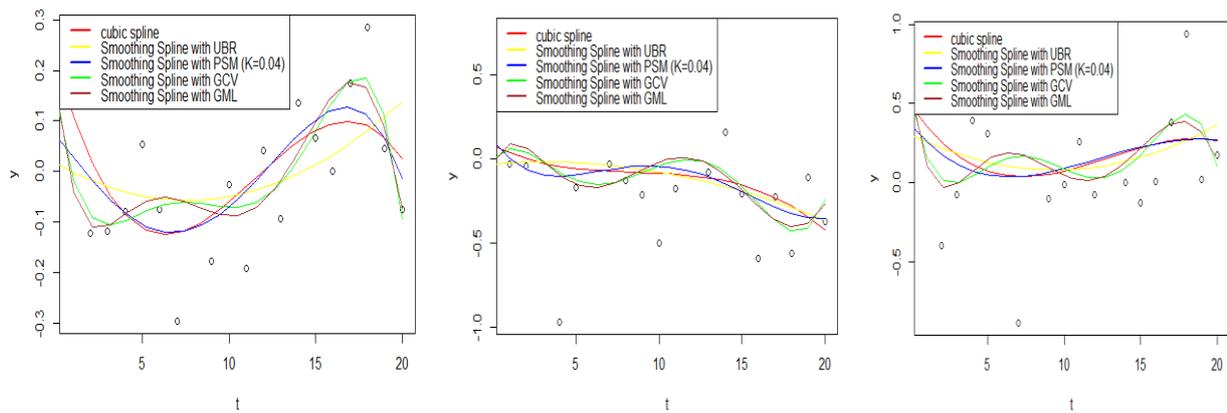


Figure 1. Cubic smoothing spline and fitted curve using GCV, GML, UBR, and PSM, (K=0.04) for T = 20, rho = 0.2, 0.5, and 0.8, sigma = 0.8

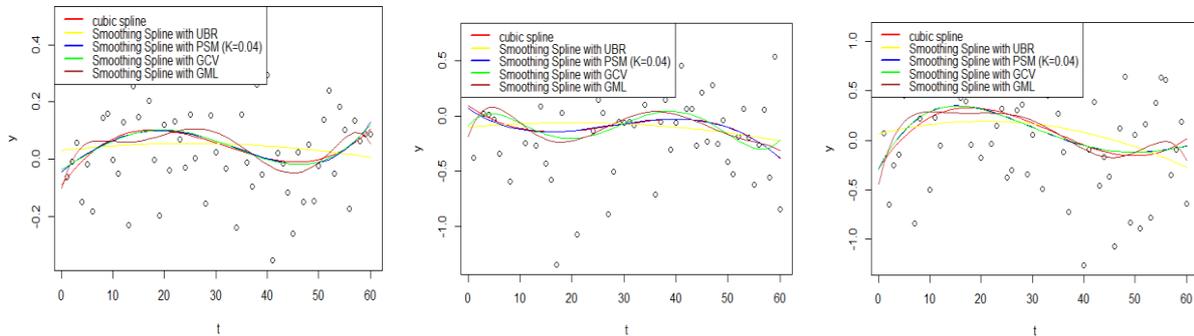


Figure 2. Cubic smoothing spline and fitted curve using GCV, GML, UBR, and PSM, ( $K=0.04$ ) for  $T = 60$ ,  $\rho = 0.2, 0.5$ , and  $0.8$ ,  $\sigma = 0.8$

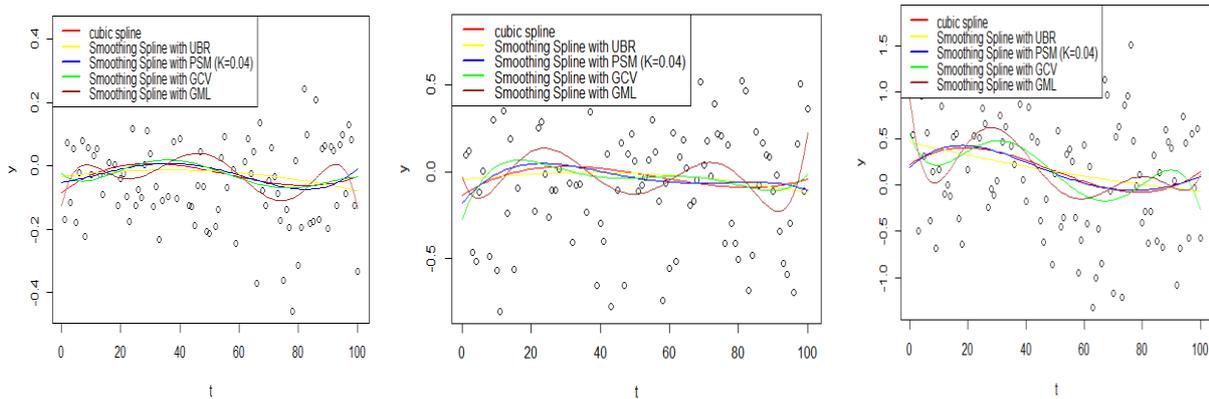


Figure 3. Cubic smoothing spline and fitted curve using GCV, GML, UBR, and PSM, ( $K=0.04$ ) for  $T = 100$ ,  $\rho = 0.2, 0.5$ , and  $0.8$ ,  $\sigma = 0.8$

5. Application

Table 4. Test for autocorrelation for the real-life data in the presence of Autocorrelation

<b>Box-Ljung test</b>
data: Residuals
X-squared = 96.7395, df = 1, p-value < 2.2e-16

$H_0$ : The data are independently distributed, or the correlations in the population from which the samples are drawn are zero

$H_1$ : The data are not independently distributed; they have serial correlation

**Decision:** autocorrelation exists in the model

Table 5. Stationarity test for the real-life data with smoothing parameters

<b>Augmented Dickey-Fuller Test</b>
data: Residuals
Dickey-Fuller = -3.6471, Lag order = 5, p-value = 0.03021
alternative hypothesis: stationary

H<sub>0</sub>: The observation is not stationary, there exists a unit root

H<sub>1</sub>: The observation is stationary, there is no unit root

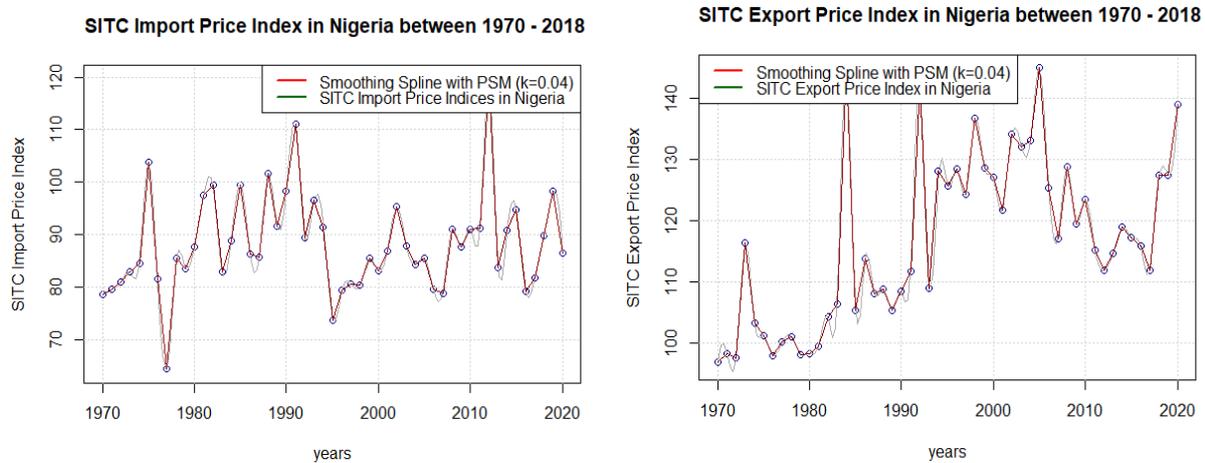
**Decision:** The data is stationary, there is no unit root

**Table 6. Cubic Smoothing Spline regression and Predictive Mean Square Error Result for Real-life data**

Smoothing Methods	$\beta_i$	Df	Selection criteria			
			PMSE	Multiple R-Square	Adjusted R Square	Predicted R Square
GML	$\beta_0 = 9.046$	209	42.77	0.5997	0.5940	0.102
	$\beta_1 = -1.275$	209				
	$\beta_2 = 5.687$	209				
	$\beta_3 = -9.716$	209				
GCV	$\beta_0 = 9006$	209	45.78	0.6004	0.5947	0.118
	$\beta_1 = -1.230$	209				
	$\beta_2 = 5.677$	209				
	$\beta_3 = -9728$	209				
UBR	$\beta_0 = 9.162$	209	92.80	0.6014	0.5957	0.099
	$\beta_1 = -1.524$	209				
	$\beta_2 = 5.840$	209				
	$\beta_3 = -9.987$	209				
PSM (K = 0.04)	$\beta_0 = 9.187$	209	32.934	0.6021	0.5964	0.4550
	$\beta_1 = -1.568$	209				
	$\beta_2 = 5.863$	209				
	$\beta_3 = -1002$	209				

Table 6 above presents the predictive mean square error of the real-life data on the standard international trade classification (SITC) export and import price indices in Nigeria between 2001 and 2020. It was discovered that the proposed smoothing Method (PSM) had the least predictive mean square error (PMSE) a confirmation that it is the preferred smoothing method for simulated and real-life data. The result also presented the multiple, adjusted, and predictive R-Square. It can be inferred from the adjusted R-square of the proposed smoothing method of 59.6% that it has the best fit among the four smoothing methods.

The plot above presents the smoothing curve of the annual standard international trade classification import price index dataset in Nigeria from 1970-2020. The data used for analysis were earlier tested for stationarity and autocorrelation. As can be seen from Figure 4 the proposed smoothing method with optimal smoothing parameter  $\lambda = 0.062439908$  and for weighted value ( $k = 0.04$ ) was used to carefully analyze the residuals to try to detect disturbances or errors in the stationary part of the series. It was observed that the PSM curve is very close to the real-life data and also provides a good fit.



**Figure 4.** Smoothing curve of SITC import and Export price index dataset (dark blue) and fitted value (red) with Smoothing Parameters Chosen by PSM (0.04).

The plot above presents the smoothing curve of the annual standard international trade classification export price index dataset in Nigeria from 1970-2018. The curve presented in Figure 4 indicated that the proposed smoothing method with optimal smoothing parameter  $\lambda = 0.062439908$  and for weighted value ( $k = 0.04$ ) was used to smooth the residuals for disturbances or errors in the stationary part of the series. It was observed that the PSM curve is very close to the real-life data and also provides a good fit.

## 6. Conclusion

The results generated from the simulation and real-life data conducted in this study have provided great insight into the smoothing method whose model produces the best fit for the time-series observations, the model whose smoothing method does not over fit data, the optimum value of the proposed smoothing method and the performance of the smoothing methods when autocorrelation is present in the error term.

The result of the goodness-of-fit test revealed that the proposed smoothing method had the best-fit model among the competing smoothing methods on the simulated and real-life data. The proposed smoothing method's model fitted without any deflection and shortcoming under cubic spline functional form with the highest adjusted R-Square of 0.9618, at  $T = 20$ ,  $D.S. = 4$ ,  $\rho = 0.2$ , and the weight value of  $k = 0.04$ .

The finding on the effect of autocorrelation in the error terms of the four smoothing methods considered in this study showed that the proposed smoothing method (PSM) works well for all levels of autocorrelation ( $\rho = 0.2, 0.5, \text{ and } 0.8$ ). It also provided a better estimate, proved to be the most preferred smoothing method than the GML, GCV, and UBR, and does not overfit a time series observation with autocorrelation in the error term at a predicted R-square of 0.6218. This result is slightly similar for GML with; [25] and [35] but differs from [24], [36-39].

The study on the optimum value of the Proposed Smoothing Method (PSM) indicated that the smoothing method performs at an optimal level when ( $k = 0.04$ ) with a predictive mean square error value of 0.046857, multiple R-Square of 0.9678, Adjusted R-Square of 0.9618 and predictive R-square of 0.6428.

The result on the effect of sample size on the performance of the four smoothing methods shows that the proposed smoothing method is computationally more efficient and consistent, and works well at all sample sizes ( $T = 20, 60, \text{ and } 100$ ) Monte-Carlo experiment. The plots and results presented in Tables and Figure 1 – 4 indicated that;

GML, GCV, and UBR showed signs of inefficiency for all the time series sizes ( $T = 20, 60,$  and  $100$ ). This finding is quite different from; [25], [35] and [40].

The findings from the result also proved the Proposed Smoothing Method (PSM) to be more efficient among the four competing smoothing methods for real-life data. This result disagrees with the finding by [41].

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- [1] Wahba, G. & Wang, Y. (1993). Behavior near Zero of the Distribution of GCV Smoothing Parameter Estimates for Splines. *Statistics and Probability Letters*, (25), 105 – 111.
- [2] Aydin, D. & Memmedli, M. (2011). Optimum Smoothing Parameter Selection for Penalized Least-squares in the Form of Linear Mixed Effect Models. *Optimization, iFirst*: 1–18.
- [3] Chen, C.S. & Huang, H.C. (2011). An improved Cp Criterion for Spline Smoothing. *Journal of Statistical Planning and Inference*, 144(1), 445 – 471.
- [4] Adams, S.O., Ipinyomi, R.A. (2019). A Proposed Spline Smoothing Estimation Method for Time Series Observations. *International Journal of Mathematics and Statistics Invention (IJMSI)*, 07(02), 18-25.
- [5] Aydin, D., M. Memmedli, and R. E. Omay. (2013). Smoothing Parameter Selection for Nonparametric Regression using a Smoothing Spline. *European Journal of Pure and Applied Mathematics*, 6, 222–38.
- [6] Jansen, Maarten (2015). Generalized Cross Validation in Variable Selection with and Without Shrinkage. *Journal of Statistical Planning and Inference*, 159, 90-104. <https://doi.org/10.1016/j.jsp.2014.10.007>
- [7] Lukas, M.A., De Hoog, F.R. & Anderssen R.S. (2016). Practical Use of Robust GCV and Modified GCV for Spline Smoothing. *Computational Statistics*. 31(1), 269-289.
- [8] Adams, S.O. (2021). An Improved Spline Smoothing Method for Estimation in the Presence of Autocorrelation Errors. *The University of Ilorin*.
- [9] Adams, S.O., Balogun, P.O. (2020). Panel Data Analysis on Corporate Effective Tax Rates of Some Listed Large Firms in Nigeria. *Dutch Journal of Finance and Management*, 4(2), 1-9, 2542–4750. <https://doi.org/10.21601/djfm/9345>
- [10] Devi, A.R., Budiantara, I.N. & Vita-Ratnasari, V. (2018). Unbiased Risk and Cross-Validation Method for Selecting Optimal Knots in Multivariable Nonparametric Regression Spline Truncated (Case Study: The Unemployment Rate in Central Java, Indonesia, 2015). *AIP Conference Proceedings 2021*.
- [11] Adams, S.O., Ipinyomi, R.O., Yahaya, H.U. (2017). Smoothing Spline of ARMA Observations in the Presence Autocorrelation Error. *European Journal of Statistics and Probability*, 5(1), 1-8.
- [12] Adams, S.O., Yahaya, H.U., Nasiru, M.O. (2017). Smoothing Parameter Estimation of the Generalized Cross Validation and Generalized Maximum Likelihood. *IOSR Journal of Mathematics*, 13(1), 41 – 44. <https://doi:10.9790/5728-1301054144>
- [13] Xu, L. & Zhou, J. (2019). A Model-Averaging Approach for Smoothing Spline Regression. *Communications in Statistics - Simulation and Computation*. 48(8), 2438 – 2451.
- [14] Wahba, G. (1977). Applications of Statistics in P. Krishnaiah Edition, A Survey of Some Smoothing Problems and the Method of Generalized Cross-Validation for solving them, *Northern Holland, Amsterdam*.
- [15] Wang, H., Meyer M. C. & Opsomer J.D. (2013). Constrained Spline Regression in the Presence of AR(p) Errors. *Journal of Nonparametric Statistics*, 25, 809 – 827.
- [16] Cook R.D. & S. Weisberg, S. (1982). Residuals and Influence in Regression, *Journal of the American Statistical Association*, 86, 328–332.
- [17] Craven P. & Wahba, G. (1979). Smoothing Noisy Data with Spline Functions, *Numerical Mathematics*, 31, 377 – 403.
- [18] Xiang, D. & Wahba, G. (1998). Approximate Smoothing Spline Methods for Large Data Sets in the Binary Case. *Proceedings of the 1997 ASA Joint Statistical Meetings, Biometrics Section* 94-98.
- [19] Wahba, G. (1990). Spline Models for Observational Data. *CBMS-NSF Regional Conference Series in Applied Mathematics*, Philadelphia: SIAM 59.
- [20] Wahba, G. (1975). Optimal Convergence Properties of Variable Knot Kernel and Orthogonal Series Methods for Density Estimation, *Annals of Statistics*, 3, 15 – 29.
- [21] Wahba, G. (1980). Automatic Smoothing of the Log Periodogram. *Journal of the American Statistical Association*, 75, 122-132.
- [22] Silverman, B.W. (1984). Spline Smoothing: The Equivalent Variable Kernel Method. *Annals of Statistics*, 12(3), 898 – 916.
- [23] Wahba, G. (1983). Bayesian Confidence Intervals for the Cross-Validated Smoothing Spline. *Journal of Royal Statistical Society (Series B)*, 45, 133-150.
- [24] Diggle, P.J. & Hutchinson, M.F (1998). On Spline Smoothing with Autocorrelated Errors. *Australian Journal of Statistics*, 31, 166 –182.

- 
- [26] Yuedong, W. (1998). Smoothing Spline Models with Correlated Random Errors. *Journal of American Statistical Association*, 93(441), 341 – 348.
- [27] Mallows, C.L. (1973). Some Comments on Cp, *Technometrics*, 15(4), 661 – 675.
- [28] Wahba, G. (1979). Convergence Rates of Thin Plate Smoothing Splines when the Data are Noisy in T. Gasser and M. Rosenblatt. *Smoothing Techniques for Curve Estimation*, Springer-Verlag, New York.
- [29] Adams, S.O., Gayawan, E., Garba, M.K. (2009). Empirical Comparison of the Kruskal - Wallis Statistics and its Parametric Counterpart. *Journal of Modern Mathematics and Statistics*, 3(2), 38 – 42. *Medwell Journal*. <https://doi:jmmstat.2009.38.42>.
- [30] Adams, S.O., Ipinoyomi, R.A. (2019). A New Smoothing Method for Time Series Data in the Presence of Autocorrelated Error. *Asian Journal of Probability and Statistics (AJPAS)*, 04(04), 1-19. <https://doi.org/10.9734/ajpas/2019/v4i430121>
- [31] Adams, S.O., Ipinoyomi, R.A. (2020). On the Efficiency of the Weighted Generalized Cross Validation and Unbiased Risk Smoothing Method for Time Series Observations with Autocorrelated Error. *International Journal of Academic and Applied Research*, 04(07), 70-81.
- [32] Adams, S.O., Yahaya, H.U. (2020). Comparative Study of GCV-MCP Hybrid Smoothing Methods for Predicting Time Series Observations. *American Journal of Theoretical and Applied Statistics*, 9(5), 219-227. <https://doi:10.11648/j.ajtas.20200905.15>
- [33] Adams, S.O., Obaromi, A.D, Alumbugu, A.I. (2021). The goodness of Fit test of an Autocorrelated Time Series Cubic Smoothing Spline Model. *Journal of the Nigerian Society of Physical Sciences*. 3(3), 191-200. <https://doi.org/10.46481/jnsps.2021.265>.
- [34] Wahba, G. (1985). A Comparison of GCV and GML for Choosing the Smoothing Parameters in the Generalized Spline Smoothing Problem. *The Annals of Statistics* 4, 1378 – 1402.
- [35] Daniel, C. (1973). One at a time Plans. *Journal of American Statistical Association*, 68 (342), 353 – 360.
- [36] Yuedong, W., Wensheng G. & Brown, M.B. (2000). Spline Smoothing for Bivariate Data with Application to Association between Hormones, *Statistica Sinica*, 10, 377 – 397.
- [37] Hart, J. D. & Wehrly, T. E. (1986). Kernel Regression Estimation using Repeated Measurements Data. *Journal of the American Statistical Association*, 81, 1080 – 1088.
- [38] Altman, N.S. (1990). Kernel Smoothing of Data with Correlated Errors. *Journal of the American Statistical Association*. 85, 749–759.
- [39] Herrmann, E., Gasser, T. & Kneip, A. (1992). Choice of Bandwidth for Kernel Regression When Residuals are Correlated. *Biometrika*, 79(4), 783–795.
- [40] Krivobokova T. & Kauermann, G. (2007). A Note on Penalized Spline Smoothing with Correlated Errors. *Journal of the American Statistical Association*, 102, 1328 – 1337.
- [41] Kim, T., Park, B., Moon, M. & Kim, C. (2009). Using Bimodal Kernel for Inference in Nonparametric Regression with Correlated Errors. *Journal of Multivariate Analysis*, 100 (7), 1487 – 1497.
- [42] Carew, J. D., Wahba, G., Xie, X., Nordheim, E.V. & Meyerand M. E. (2003). Optimal Spline Smoothing of fMRI Time Series by Generalized Cross-Validation. *NeuroImage*, 18(4), 950 – 961.