SciencePG
Science Publishing Group

Comparative Study of GCV-MCP Hybrid Smoothing Methods for Predicting Time Series Observations

Samuel Olorunfemi Adams*, Yahaya Haruna Umar**

Department of Statistics, University of Abuja, Abuja, Nigeria

Email address:

samuel.adams@uniabuja.edu.ng (S. O. Adams), yahagumau@gmail.com (Y. H. Umar)

*Corresponding author

Abstract: Generalized Cross Validation (GCV) has been considered a popular model for choosing the complexity of statistical models, it is also well known for its optimal properties. Mallow's CP criterion (MCP) has been considered a powerful tool which is used to select smoothing parameters for spline estimates with non-Gaussian data. Most of the past works applied Generalized Cross Validation (GCV) and Mallow's CP criterion (MCP) smoothing methods to time series data, this methods over fits data in the presence of Autocorrelation error. A new Smoothing method is proposed by taking the hybrid of Generalized Cross Validation (GCV) and Mallow's CP criterion (MCP). The predicting performance of the Hybrid GCV-MCP is compared with Generalized Cross Validation (GCV) and Mallow's CP criterion (MCP) using data generated through a simulation study and real-life data on all SITC export and import price index in Nigeria between the years, 2001-2018, performed by using a program written in R and based on the predictive Mean Score Error (PMSE) criterion. Experimental results obtained show that the predictive mean square error (PMSE) of the three smoothing methods decreases as the sample size and smoothing parameters increases. The study discovered that the Hybrid GCV-MCP smoothing methods performed better than the classical GVV and MCP for both the simulated and real life data.

Keywords: Time Series Predicting, Generalized Cross Validation, Mallow's CP Criterion, Hybrid GCV-MCP

1. Introduction

There are several ways of modeling time-series observations through nonparametric regression techniques to make predictions, one such nonparametric technique is the spline smoothing method [1]. Spline smoothing model assume that observations y_i are taken at times t_i, for $i = 1, ..., $ n and the y_i are generated by a model of the form;

$$y_i = f(x_i) + \varepsilon_i, i = 1, \dots \text{ n} \quad (1)$$

Where; yi is the value of the time series in time t_i, f is an unknown smooth function, and ε_i is normally distributed with the standard deviation σ_i. Smoothing spline is a solution to a nonparametric regression problem with the function $g(x)$ i.e. $\hat{g}$ $\in C^2[a, b]$ in the model that minimizes the penalized residual sum of squares with two continuous derivatives, as given below;

$$S(g) = \sum_{i-1}^{n}(y_i - g(x_i))^2 + \lambda \int_a^b (g''(x))^2 dx \quad (2)$$

The first term in the equation is the residual sum of the square for the goodness of fit of the data, the second term is a roughness penalty, which is large when the integrated second derivative of a regression function $g''(x)$ is also large. λ is a smoothing parameter. If λ approaches 0 then $g(x)$ simply interpolates the observations, if λ is very large, then $g(x)$ will be selected so that $g''(x)$ is everywhere 0, which implies an overall linear least-squares fit the observations.

The smoothing parameter λ plays a key role in controlling the trade-off between the goodness of fit represented by $\sum_{i-1}^{n}(y_i - g(x_i))^2$ and smoothness of the estimate measured by $\int_a^b (g''(x))^2 dx$. The solution based on smoothing spline for a minimum problem in equation (2) is known as a "natural cubic spline" with knots at $x_1, \dots, x_n$. From this point of view, a specially structured spline interpolation which depends on a chosen value λ develops into a suitable approach of function g in a model (2). Let $g = (g(x_1), \dots, g(x_n))$ be the vector of values of function g at the knot points $x1, \dots, xn$. The smoothing spline estimate

$\hat{g}_x$ of this vector for the fitted data $y = (y_1, \ldots, y_n)^T$ are projected by;

$$\hat{g}_\lambda = \begin{bmatrix} \hat{g}_\lambda(x_1) \\ \hat{g}_\lambda(x_2) \\ \cdot \\ \cdot \\ \cdot \\ \hat{g}_\lambda(x_n) \end{bmatrix}_{(n \times 1)} = S_\lambda \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}_{(n \times 1)} \quad \text{or } \hat{g}_\lambda = S_\lambda \qquad (3)$$

where $\hat{g}_\lambda$ is a natural cubic spline with knots at x_1, ..., x_n for a fixed $\lambda > 0$, and S_λ is a well-known positive-definite (symmetrical) smoother matrix which depends on λ and the knot points x_1, ..., x_n, but not on y. The function $\hat{g}_\lambda$, is the estimation of function g, it is obtained by cubic spline interpolation that depends on the condition $\hat{g}_\lambda(xi) = (\hat{g})i$, $i = 1, 2, \ldots, n$.

The smoothing parameter determines the smoothing level. For $\lambda > = 0$, $f(x_i)$ equals the value of the time series at that time, x_i. The greater the smoothing parameter the greater the difference between $f(x_i)$ and x_i. The most popular method used in time series analysis is the classical Autoregressive Moving Average (ARMA) approach; it assumes linear dependence on past observations and discoveries. The classical linear modelling was used with known data samples and discovered that it has its limitations [1]. The increasing knowledge of deviations from the known ARMA model is projected by Spline smoothing researchers in nonlinear time series analysis. Many Spline smoothing researchers have studied modeling of time series observations with Generalized Cross-Validation (GCV), Generalized Maximum Likelihood (GML), and Mallow's C_P criterion (MCP) methods [2]. The GCV method was extended to estimate the smoothing parameter and Autocorrelated error term, [2]. A smoothing spline was represented by a state-space model and extended the CV, GCV, and GML estimation methods to an Autoregressive moving average error term [3]. A Cross-Validation method is employed to estimate some smoothing parameters [4] and more recently, it was extended to GML, GCV, and MCP methods to estimate the smoothing parameter when data are correlated, [5]. Almost all of these methods were developed for time series observations while some others require that the design points are equally spaced. In real-time series, usually, neither the function g nor the standard deviations ai are known, therefore $E('x)$ cannot be calculated according to equation (1). However, in the case when $f(x_i)$ is a linear combination of x_i, the Generalized Cross-Validation (GCV) function is one of the models that is often used in selecting the optimal knot, [6, 7, and 8]. It provides several advantages over other methods, including asymptotic optimal, invariance to transformation and do not needs a known population variance. Mallow's C_P Criterion (MCP) is a smooth parameter selection method that requires an estimated value of the known error variance [9, 10, 11]. The MCP method is used to select the smoothing parameter for spline estimation with non-Gaussian data [12] and [13]. From simulation study, the MCP smoothing method estimated knots which were close with the original knot, so

the MCP method can be taken into consideration in choosing optimal knot, [14]. Numerical experiments showed that the direct MCP methods perform better than existing indirect MCP [15]. Generalized Cross-Validation (GCV), Mallow's CP criterion and Generalized Maximum Likelihood (GML) methods were examined based on a simulation study, it was discovered that for large samples and because of the effect of replication, GCV and Mallow's CP criterion has the same asymptotic result [16].

In this paper, the Hybrid smoothing method is developed by combining the unique optimal properties of GCV and MCP i.e. Hybrid GCV-MCP. The performance and efficiency of the Hybrid GCV-MCP smoothing method was compared with the classical GCV and MCP smoothing methods for predicting time-series observations. After giving a brief introduction of smoothing time series observation using and Generalized Cross-Validation (GCV) and Mallow's C_P Criterion (MCP) the next section presents a literature review on GCV and MCP, section three presents materials and methods. Section four compares the three methods via a simulation study and it application to real life data, and finally, the conclusion was presented in the last section.

2. Literature Review

Literature shall establish the fact that Generalized Cross-validation and Mallow's C_P criterion have been applied to time series observations in past. Generalized Cross-Validation (GCV) was compared with and Mallow's C_P criterion (MCP), it was recommended that Generalized Cross-Validation is a good smoothing parameter selection for small and medium-sized samples [17]. The smoothing spline method was applied to fit a curve to a noisy data set, where the selection of the smoothing parameter is essential. An improved Cp criterion for spline smoothing based on Stein's unbiased risk estimate was proposed to select the smoothing parameter. The resulting fitted curve showed to be superior and more stable than commonly used selection criteria and possesses the same asymptotic optimality as Cp, [18]. Most data-driven smoothing parameter selection methods were compared based on large and small sample sizes. The parallel of Akaike's information criterion (GFAIC) and Generalized Cross-Validation (GCV) is recommended as being the best selection criteria. For large samples, the GFAIC method would seem to be more appropriate while for small samples they proposed the implementation of GCV criterion [19]. Two types of results that support the use of Generalized Cross-Validation (GCV) for variable selection under the assumption of sparsity was investigated. The first type of result is based on the well-established links between GCV on one hand and Mallows's Cp and Stein Unbiased Risk Estimator (SURE) 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 [20]. In the comparison of GCV with GML via a Monte Carlo Method using a program written in R. It was

discovered that GML was better than GCV because it is stable and works well in all simulations and at all sample sizes and it does not over fit data when the sample size is small,[21]. MCP and CV was compared for selecting the optimal knots. The criteria for selecting the best model were based on Mean Squared Error and R-square. A 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 CV estimates the knots more accurately than MCP [14]. Nonparametric regression problems were considered, a model-averaging procedure for smoothing spline regression problem was developed [22].

The main motivation of this paper is to present a new smoothing method to deal with the prediction of time series observations that yields better performance than the classical methods. All of the methods given above have some limitations when modeling time-series observations, which is indicated above. The introduction of the combination of GCV and MCP i.e. Hybrid GCV-MCP was proposed to solve the problem of over smoothing and over fitting of time series observations for the number of knots, smoothing parameters, and time series sizes.

3. Research Methodology

3.1. Data Collection

This research was conducted to evaluate the performances of the three smoothing methods, i.e. GCV, MCP, and hybrid GCV-MCP. Time-series data were simulated by using a program coded in R (version 3.2.3) for time series sizes of; 50, 100, and 150. The number of replications was 1,000 for each of the samples. For each simulated data set, the Predictive Mean Squared-Errors (PMSE) was used to evaluate the quality and performance of the methods.

3.2. Equation Used for Generating Values in the Simulation

The simulation study conducted to evaluate and compare the performance of the four estimation methods is given as;

$$y(t) = Sin\left(\frac{\pi}{t}\right) + \varepsilon_t \qquad t = 1, \ldots, 150 \qquad (4)$$

Where; $\pi = 180°, \varepsilon_t \sim N(0, \sigma)$ and it is independently and identically distributed with zero mean and 0.8 and 1.0 standard deviation.

3.3. Experimental Design

The experimental plan applied in this research work was designed to

1. Three-time series sizes(T) of 50, 100 and 150 were considered for the simulation
2. Four smoothing Parameters were considered, i.e. $\lambda = 1, 2, 3, 4$
3. Two standard deviations were considered i.e. $\sigma_1 \text{ and } \sigma_2 = 0.8 \text{ and } 1.0$

4. Data were generated for 1,000 replications for each of the 3 x 4 x 2 = 24 combinations for cases T's, λ's and σ's.

3.4. Evaluation of the Smoothing Methods

In this study, Predictive Mean squared error (PMSE) is used as quality measurement to evaluate the performance of a smoothing or Curve fitting procedure. PMSE is the expected value of the square difference between the fitted value implied by the predictive function $\hat{f}(x_i)$ and the values of the observed function $f(x_i)$. It is used to measure the performance and quality of a predictor or Smoothing methods like Cross-Validation, Generalized Cross-Validation, Mallow's C_P Criterion, etc. The Predictive Mean Square Error (PMSE) is written mathematically as;

$$PMSE(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \left(f(x_i) - \hat{f}(x_i) \right)^2 \qquad (5)$$

The Predictive Mean Square Error can be divided into two terms, the first term is the sum of square biases of the fitted values while the second is the sum of variances of the fitted values. Where; $f(x_i)$ is observed value and $\hat{f}(x_i) =$ fitted/predicted/estimated value

3.5. Generalized Cross-Validation (GCV)

The term Generalized Cross-Validation (GCV) was proposed by [23] and [24] as a replacement of 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 λ under which the missing data points are best predicted by the remainder of the data. To be precise, let g_λ^{-1} be the smoothing spline calculated from all the data pairs except (ti, yi), using the value λ for the smoothing parameter. The cross-validation choice of λ is then the value of λ which minimizes the Cross-Validation score;

$$CV(\lambda) = \frac{1}{n} \sum \{y_i - \hat{g}(t_i)\}^2 \qquad (6)$$

This is identical to the criterion for model selection in regression generally defined by matrix A (λ) [25]

$$A_{ij}(\lambda) = n^{-1} g(t_i, t_j) \qquad (7)$$

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \frac{\{y_i - \hat{g}(t_i)\}^2}{\{1 - A_{ii}(\lambda)\}^2} \qquad (8)$$

The use of related criterion was suggested, it was called Generalized Cross-validation, obtained from (8) by replacing $A_{ii}(\lambda)$ by its average value, $n^{-1} tr A(\lambda)$, this gives the score [24].

$$GCV(\lambda) = \frac{n^{-1} RSS(\lambda)}{(1 - n^{-1} tr A(\lambda))^2} \qquad (9)$$

Where; RSS (λ) is the residual sum of squares, $\sum \{y_i - \hat{g}(t_i)\}^2$, in their study [24] also give theoretical arguments to show that Generalized Cross-Validation should, asymptotically choose the best possible value of λ in the

sense of minimizing the average squared error at the design points. This predicted good performance is borne out by published practical examples in [27]. The generalized Cross-validation method is well known for its optimal properties [14]. If there exists an n x n, the influence matrix, with the property;

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

It is such that $W_0 (\lambda)$ can be rewritten as;

$$W_0(\lambda) = \frac{\sum_{k=1}^{n}(a_{kj}y_j - y_k)^2}{(1 - a_{kk})^2} \qquad (10)$$

Where; $a_{ky}y_j \in 1, 2, \ldots, n$

GCV is a modified form of the CV which is a conventional method for choosing the smoothing parameter. The GCV score which is constructed by analogy to CV score can be obtained from the ordinary residuals by dividing by the factors $1 - (S\lambda)_{ii}$. The approved design of

GCV is to replace the factors $1 - (S_\lambda)_{ii}$ in Cross Validation with the average score $1 - n^{-1} trace\ (S\lambda)$. Thus, by summing the squared corrected residual and factor $\{1 - n^{-1} trace\ (S_\lambda)\}^2$, by the example ordinary cross-validation, the GCV score function can be written as;

$$GCV(\lambda) = \frac{1}{n} \frac{\sum_{k=1}^{n}\{y - f_k(x_1)\}^2}{\{1 - n^{-1} trace(S\lambda)\}^2}$$

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

Where; n is the measurement/observations $\{xi, yi\}$, λ is smoothing parameters and $S\lambda$ is the ith diagonal element of a smoother matrix.

3.6. Mallow's CP Criterion

Mallow's CP criterion (MCP) was developed by [26] to estimate the fit of a regression model based on Ordinary Least Square. It is applied to a model selection case where predictor variables are present for forecasting some outcomes and for finding the best model involved in subset predictors. The smaller the value of the Cp, the relatively precise it is, the Cp is written mathematically as;

$$MCP\ (\lambda) = \frac{1}{n}\{\|(S_\lambda - 1)y\|^2 + 2\sigma^2 tr(S_\lambda) - \sigma^2\}$$

$$MCP\ (\lambda) = \frac{1}{n}\{\|(y - \hat{f}_\lambda)y\|^2 + 2\sigma^2 tr(S_\lambda) - \sigma^2\}$$

$$MCP\ (\lambda) = \frac{1}{n}\frac{\sum_{i=1}^{n}\{y_i - \hat{f}_{\hat{\lambda}}(x_i)\}^2}{tr\{I - (s_{\hat{\lambda}})\}^2} = \frac{\|(s_{\hat{\lambda}} - I)y\|^2}{tr(I - s_{\hat{\lambda}})} \qquad (12)$$

3.7. Hybrid GCV-MCP Method

Hybrid is the combination of two models or methods. The combination of different methods have been frequently used in research for better performance by manipulating the unique strength of the two methods. A combination of GCV and MCP will provide more accurate/precise predicting model for forecasting as compared to an individual smoothing methods. The GCV method is well known for its optimal properties in smoothing estimation method [6] while MCP method has been successfully applied to estimate smoothing parameters for spline estimates with non-Gaussian data and to fit data appropriately [8, 24, 25] and [26]. The proposed Smoothing method combines the optimal properties of GCV and MCP.

The minimizer of GCV is given as;

$$GCV(\lambda) = \frac{1}{n} \frac{\sum_{k=1}^{n}\{y - f_k(x_1)\}^2}{\{1 - n^{-1} trace(S\lambda)\}^2}$$

$$GCV(\lambda) = \frac{n^{-1}\|(y - S\lambda)y\|^2}{[n^{-1}trace(I - S\lambda)]^2} = \frac{n^{-1}\|(I - S\lambda)y\|^2}{[n^{-1}trace(I - S\lambda)]^2} \quad (13)$$

While the Mallow's CP criterion method of is given as;

$$MCP(\lambda) = \frac{\|(S_{\hat{\lambda}} - I)y\|^2}{tr(I - S_{\hat{\lambda}})} \qquad (14)$$

Therefore, a new smoothing method is proposed by introducing an additional weighted parameter g and combining properties of the Generalized Cross-Validation (GCV) and Mallow C_P criterion (MCP). The combination, measurement, and expression of the quantities of the two methods will yield an optimal performance and smoothing model that does not over fit data. The minimizer of the hybrid methods of (13) and (14) is given as;

$$Hybrid\ (\lambda) = gf_1(\lambda) + (1 - g)f_2(\lambda) \qquad (15)$$

$$0 < g < 1, f_1(\lambda) = \text{GCV and } f_2(\lambda) = \text{MCP}$$

$$Hybrid\ (\lambda) = g\frac{n^{-1}\|(I - S\lambda)y\|^2}{[n^{-1}trace(I - S\lambda)]^2} + (1 - g)\frac{\|(S\lambda - I)y\|^2}{tr(I - S\lambda)} \quad (16)$$

Where; n is number of observations, $0 < g < 1$, g is weighted values, $y = (y_1, \ldots, y_n)^T$ is the smoothing function, $\hat{f} = (f(t_1).f(t_n)).y_n)^T$ is $S_\lambda y$, $S\lambda =$ is the diagonal element of the smoother matrix.

4. Simulation Study

In this section, a simulation study is carried out to compare the behaviors of the Hybrid GCV-MCP smoothing methods with two classical smoothing methods namely; Generalized Cross-Validation (GCV) and Mallow's C_P (MCP) criterion when estimating time-series observation. Before the results of simulation experiments, datasets for the different simulation combinations are generated by using codes written in the R 3.2.3 software. Our data generation procedure, with accompanying descriptions, is given in Table 1.

Table 1. *Data generation steps with explanation.*

	Steps	Explanation
Step 1.	Decide on T, n_{sm} and λ	The sample size of the simulated dataset, number of replication and smoothing parameter
Step 2.	Produce on $x_i's, f(.)$ and ε_i	Nonparametric Covariate, smoothing function, random error terms
Step 3.	Generate $y_i's$	Generate the observed or actual data point for the smoothing model
Step 4.	Generate $\hat{y}_i's$	Generate the estimated or predicted observation for the smoothing model
Step 5.	Obtain the PMSE (λ)	Compute the Predictive Mean Square error

For this simulation study, within the scope of Step 1, $T = 50, 100$ and $150, n_{sm} = 1000$ and $\lambda = 1, 2, 3$ 4 . The Nonparametric Covariate, smoothing function, random error terms in step 2 are generated as; $x_i = \frac{i-0.5}{n}, \sigma = 0.8, 1.0, \varepsilon_i \sim iid\ N(0,1)$ and $f(x) = 1.5\theta\left(\frac{x-0.35}{0.15}\right) - \theta\left(\frac{x-0.6}{0.09}\right), \theta(\mu) = \frac{1}{\sqrt{2\pi}}\ exp\left[\frac{-\mu^2}{2}\right]$, where θ is a constant that determines the shape of the curve. In step 3, $y(i) = Sin\left(\frac{\pi}{1}\right) + \varepsilon_i$ the function in step 5 used to test the hybrid GCV-MCP smoothing method under various conditions is given as; $PMSE(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(f(x_i) - \hat{f}(x_i)\right)^2$. Where $\hat{f}(x_i) = \left(\hat{f}\right)_i$ and $f(x_i)$ is the value at the knot x_i of the appropriate function f.

4.1. Smoothing Results Methods for Smoothing Parameters ($\lambda = 1, 2, 3$ and 4), Time (T = 50, 100 and 150) and Std. Deviation ($\sigma = 0.8$)

Table 2 presents the predictive mean square error of the three smoothing methods, three sample sizes at 0.8 sigma level. It was discovered that for the GCV smoothing method, the predictive mean square error decreases as the time increases; when T = 50 the PMSE decreased from 0.053273 to 0.027264 when T = 100 and further decreases from 0.027264 to 0.025485 when T = 150 for smoothing function $\lambda = 1$. When T = 50, the predictive mean square error (PMSE) of the hybrid GCV-MCP smoothing method,

decreased from 0.042392 to 0.040716 when T = 100 and further decreased to 0.003851 when T = 150. Furthermore, for the MCP smoothing method, the predictive mean square error decreases as the time increases; when T = 50 the PMSE decreased from 0.048021 to 0.034561 when T = 100 and further decreases from 0.034561 to 0.025485 when T = 150 for smoothing function $\lambda = 1$.

4.2. Smoothing Results Methods for Smoothing Parameters ($\lambda = 1, 2, 3$ and 4), Time (T = 50, 100 and 150) and Std. Deviation ($\sigma = 1.0$)

Table 3 presents the predictive mean square error of the three smoothing methods and the three sample sizes at 1.0 sigma level. It was discovered that for the GCV smoothing method, the predictive mean square error decreases as the time increases; when T = 50 the PMSE decreased from 0.126826 to 0.0424942 when T = 100 and further decreases from 0.0424942 to 0.0242396 when T = 150 for smoothing function $\lambda = 1$. When T = 50, the predictive mean square error (PMSE) of the hybrid GCV-MCP smoothing method, decreased from 0.087162 to 0.016832 when T = 100 and further decreased to 0.004103 when T = 150. Furthermore, for the MCP smoothing method, the predictive mean square error decreases as the time increases; when T = 50 the PMSE decreased from 0.056372 to 0.044455 when T = 100 and further decreases from 0.044455 to 0.037378 when T = 150 for smoothing function $\lambda = 1$.

Table 2. *PMSE result for the smoothing methods for smoothing parameters (λ = 1, 2, 3 and 4), time (T = 50, 100 and 150) and std. deviation (σ = 0.8).*

Time size	Smoothing Methods	Smoothing Parameter levels			
		$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$
T =50	GCV	0.053273	0.044131	0.042976	0.042599
	Hybrid	0.042390	0.071148	0.070962	0.068089
	MCP	0.048020	0.032269	0.030172	0.017179
T = 100	GCV	0.027264	0.021027	0.020245	0.029484
	Hybrid	0.044716	0.091883	0.081778	0.044087
	MCP	0.034561	0.024042	0.022625	0.022158
T = 150	GCV	0.025094	0.018843	0.018061	0.017806
	Hybrid	0.003851	0.098199	0.069462	0.002584
	MCP	0.025485	0.018456	0.017496	0.017179

Table 3. *PMSE result for the smoothing methods for smoothing parameters (λ = 1, 2, 3 and 4), time (T = 50, 100 and 150) and std. deviation (σ = 1.0).*

Time size	Smoothing Methods	Smoothing Parameter levels			
		$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$
T =50	GCV	0.126826	0.104707	0.101912	0.101001
	Hybrid	0.087162	0.052266	0.062848	0.065031
	MCP	0.056372	0.039990	0.037767	0.037034
T = 100	GCV	0.042494	0.033899	0.032822	0.032471
	Hybrid	0.068320	0.093528	0.085243	0.048527
	MCP	0.044455	0.030348	0.028460	0.027839
T = 150	GCV	0.024239	0.019535	0.018947	0.018755
	Hybrid	3.704103	2.900917	2.773740	2.730723
	MCP	0.0373787	0.025315	0.023706	0.023177

4.3. Smoothing Curves of the Time Series Observation

Figures 1 to 6 presents the observed and estimated values of GCV, Hybrid and MCP after 1000 replications. From these plots it is observed that Hybrid and GCV have smaller PMSEs when compared to MCP. The plot also indicated that Genralized Cross-Validation (GCV), Mallow's C_p Criterion (MCP) and Hybrid GCV-MCP smoothing methods provides

a good fits for time series observations, but Hybrid GCV-MCP and GCV smoothing methods provide a better estimates than MCP based on Predictive Mean Square Error (PMSE). And at all sample sizes. The Hbybrid method is more stable when the sample size is small and medium (T = 50 and 100) This behavior of the Hybrid smoothing method is somehow similar to the finding of [27] and [13].

***Figure 1.** Plots of the observations (...) and Estimates (---) With Smoothing Parameters Chosen by GCV, Hybrid and MCP for λ=1, 2, 3, 4 σ = 0.8 and T = 50.*

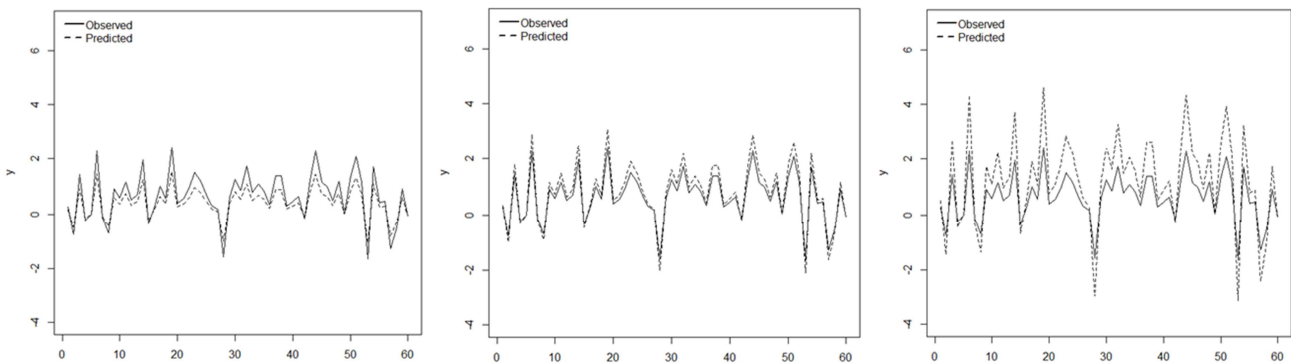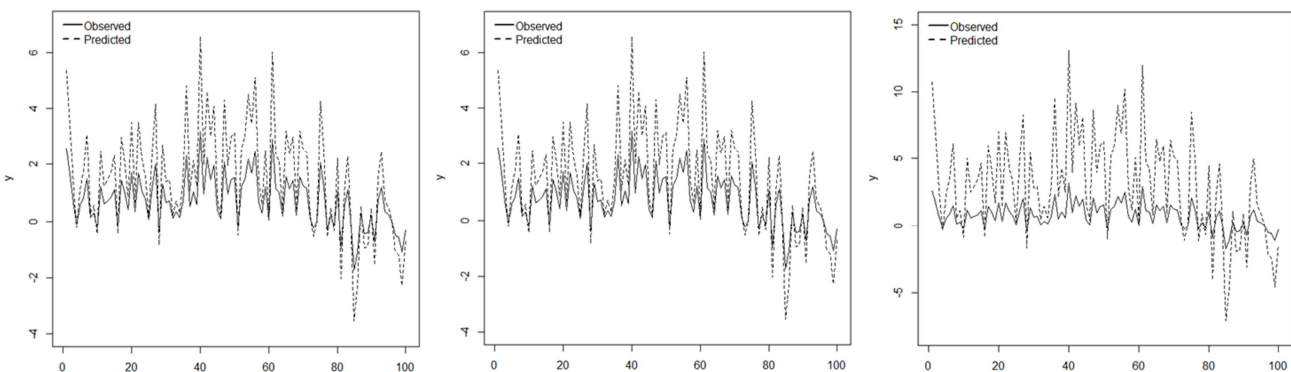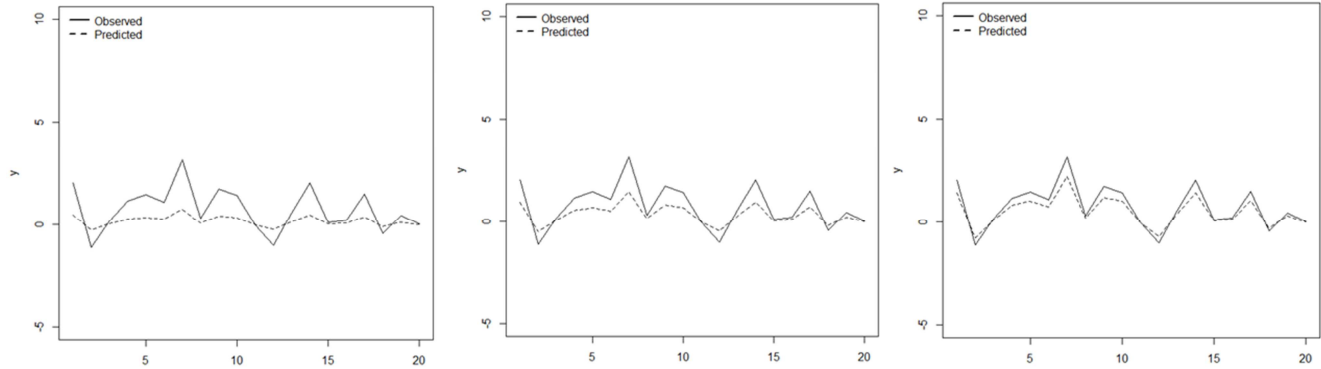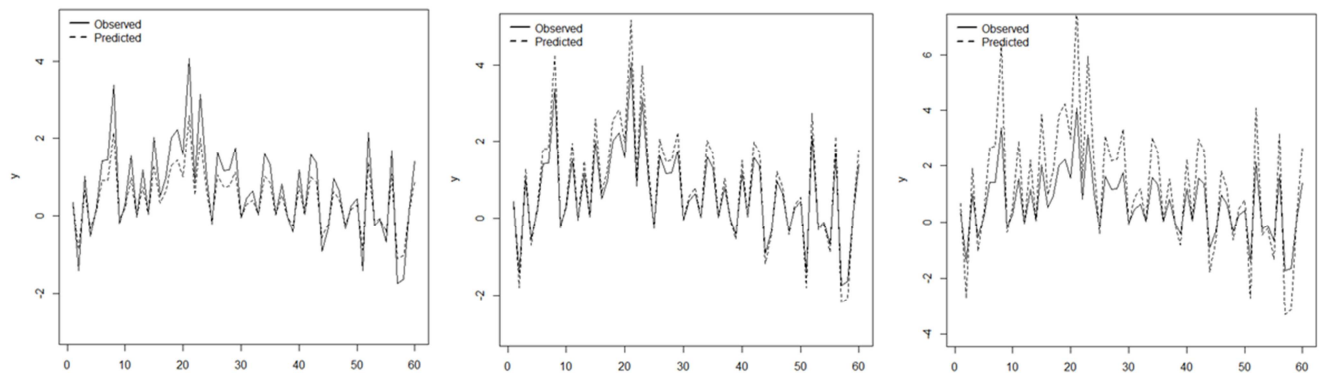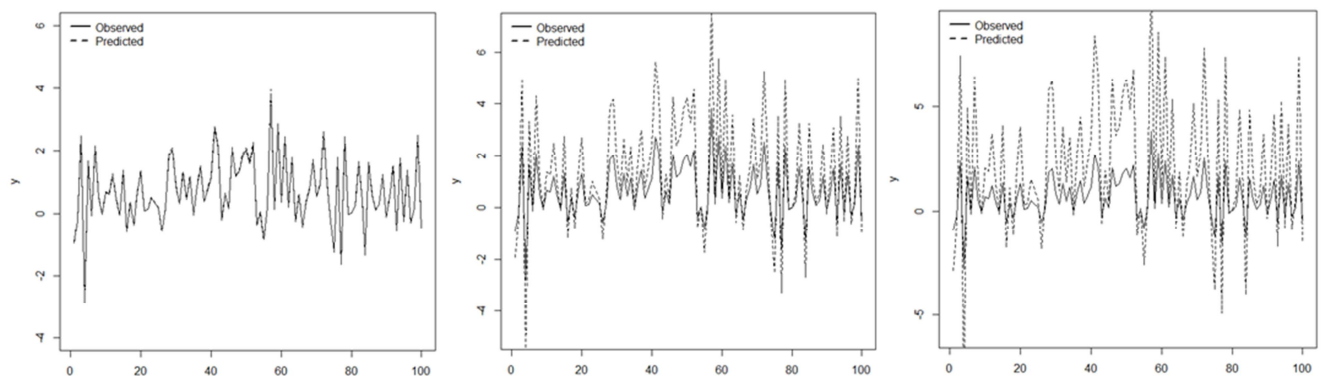***Figure 2.** Plots of the observations (...) and Estimates (---) With Smoothing Parameters Chosen by GCV, Hybrid and MCP for λ=1, 2, 3, 4 σ = 0.8 and T = 100.*

***Figure 3.** Plots of the observations (...) and Estimates (---) With Smoothing Parameters Chosen by GCV, Hybrid and MCP for λ=1, 2, 3, 4 σ = 0.8 and T = 150.*

***Table 4.** Rank of the performance of smoothing methods for time periods = 50, 100 and 150, smoothing function = 1, 2, 3 and 4 and for std. dev. = 0.8.*

Time size	Smoothing Methods			
	GCV	Hybrid GCV-MCP	MCP	Preferred Method
T =50	3	1	2	Hybrid
T = 100	2	1	3	Hybrid
T = 150	3	2	1	MCP

Table 5. *Rank of the performance of smoothing methods for time periods = 50, 100 and 150, smoothing function = 1, 2, 3 and 4 and for std. dev. = 1.0.*

Time size	Smoothing Methods			
	GCV	Hybrid GCV-MCP	MCP	Preferred Method
T =50	2	3	1	MCP
T = 100	2	1	3	Hybrid
T = 150	2	1	3	Hybrid

Figure 4. *Plots of the observations (...) and Estimates (---) With Smoothing Parameters Chosen by GCV, Hybrid and MCP for λ=1, 2, 3, 4 σ = 1.0 and T = 50.*

Figure 5. *Plots of the observations (...) and Estimates (---) With Smoothing Parameters Chosen by GCV, Hybrid and MCP for λ=1, 2, 3, 4 σ = 1.0 and T = 100.*

Figure 6. *Plots of the observations (...) and Estimates (---) With Smoothing Parameters Chosen by GCV, Hybrid and MCP for λ=1, 2, 3, 4 σ = 1.0 and T = 150.*

Tables 4 and 5 above present ranks and preferred smoothing methods of the three smoothing methods (GCV, Hybrid GCV-MCP, and MCP) at three time periods (i.e. T = 50, 100 and 150) when the standard deviation is 0.8 and.

From the results present in table 4, it can be seen that Hybrid GCV-MCP methods had the least predictive mean square error when the time series size is small and moderate (T = 50 and 100) while MCP smoothing method had the

smallest predictive mean square error when the time series size is large (T = 150).

From the results present in table 5, it can be seen that MCP smoothing methods had the least predictive mean square error when the time series size is small (T = 50) while Hybrid GCV-MCP smoothing methods had the smallest predictive mean square error when the time series size is small (T = 100 and 150).

In summary, Hybrid GCV-MCP is the best smoothing method for predicting time-series observations at three time periods (i.e. T = 50, 100, and 150), for standard deviation (σ = 0.8 and 1) based on the predictive mean square error (PSME) criterion. This finding is quite different from, [15] but somehow similar to those from; [27] and [13].

4.3. Application of Smoothing Methods to Real Life Data

This section is prepared to show the performance of the hybrid GCV-MCP smoothing method for time series observation. The dataset represents data from all Standard International Trade Classification (SITC) Export and Import price index in Nigeria between 2001-2018 collected from CBN 2018. To provide continuity, the logarithms of the All SITC product export price index are considered as a response variable (*export*), and *import* is taken as a nonparametric covariate (*import*). The right smoothing spline time series model is thus given by;

$$y_t = log(import_t) = f(export) + \varepsilon_i \ 1 \le t \le 216 \ (17)$$

The dataset contains information for 216 all Standard International Trade Classification (SITC) Export and Import price index in Nigeria to be used for this analysis. The outcomes calculated for the model in Equation (17) are given the following table and figure.

Table 8 present the result for real-life data on all SITC export and import price index in Nigeria between 2001-2018. The Autocorrelation result showed Autocorrelation existed in the time series observation (X-squared = 96.7395, df = 1, p-value < 2.2e-16). Table 6 showed the time series observation is stationary (Dickey-Fuller = -3.6471, Lag order = 5, p-value < 0.03021) while table 7 indicated that hybrid GCV-MCP smoothing method performed better than the GCV and MCP.

Table 6. *Test for autocorrelation for the real-life data for smoothing parameters (λ = 1, 2, 3 and 4).*

Box-Ljung test
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_I: The data are not independently distributed; they exhibit serial correlation

Decision: Autocorrelation exist in the data

Table 7. *Stationarity test for the real-life data for smoothing parameters (λ = 1, 2, 3 and 4).*

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

H_0: The data is not stationary, there is a unit root
H_I: The data is stationary, there is no unit root
Decision: The data is stationary, there is no unit root

Table 8. *PMSE result for the real-life data of smoothing parameters (λ = 1, 2, 3 and 4).*

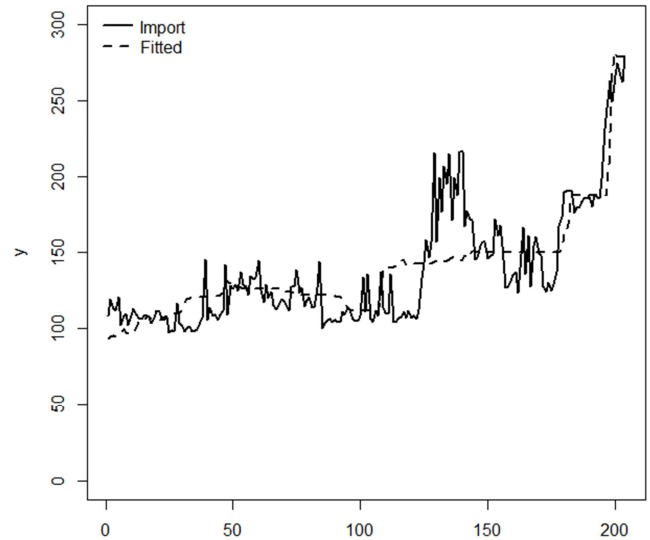| Method	1	2	3	4
GCV	13.70312	13.59623	13.53751	13.49974
MCP	7.796751	7.687323	7.637122	7.611105
Hybrid	36.92479	36.69295	36.0000	36.5906

Figure 7. *Plots of the SITC import price index in Nigeria dataset (. . .) and Estimates (---) with Smoothing Parameters Chosen by Hybrid GCV-MCP for λ=1,2,3,4, σ = 0.8 and T = 100.*

5. Conclusions

In this paper, the experimental results are being obtained by using GCV, MCP, and Hybrid GCV-MCP for predicting time-series observations. The performance of the Hybrid model is compared with Generalized Cross-Validation (GCV) and Mallow's C$_P$ Criterion (MCP). GCV is used to choose the complexity of statistical models while Mallow's CP criterion (MCP) is a powerful tool used to select smoothing parameters for spline estimates with non-Gaussian data. The result showed that MCP provided better results when compared to GCV for small sample size. In all, the Hybrid GCV-MCP method achieved good forecasted results and performed better than the two classical smoothing methods (i.e. GCV and MCP) based on the least predictive mean square error (PMSE) criterion. The proposed smoothing method was also applied to the real-life data on all SITC export and import price index in Nigeria between the years, 2001-2018, result from PMSE proved that hybrid GCV-MCP performed better than the GCV and MCP.

Finally, the Proposed Hybrid GCV-MCP is recommended as the best smoothing method for time series observations for mostly medium and large sample sizes.

Acknowledgements

References

[1] Stheim, D. (1994). Nonlinear Time Series, A Selective Review, Scandinavian Journal of Statistics, 21: 97-130.

[2] Diggle, P. J. and Hutchinson, M. F. (1989). On Spline Smoothing with Autocorrelated Errors, Australian Journal of Statistics, 31: 166-182.

[3] Kohn, R., Ansley, C. F., and Wong, C. (1992), 'Nonparametric Spline Regression with Autoregressive Moving Average Errors,' Biometrika, 79: 44-50.

[4] Hurvich, C. M., and Zeger, S. L. (1990). A Frequency Domain Selection Criterion for Regression with Autocorrelated Errors. Journal of the American Statistical Association. 85: 705-714.

[5] Yuedong, W. (2011). Smoothing Spline: Methods and Applications. Technometrics. 54 (3): 324-326. Chen, C. S. and Huang H. C. (2011): An improved Cp criterion for spline smoothing. *Journal of Statistical Planning and Inference, 144 (1): 445-471.*

[6] Reinsch, C. H. (1967). Smoothing by Spline Functions. *Numer. Math*. 10: 177–183.

[7] Craven, P., Wahba, G. (1979). Smoothing Noisy Data with Spline Functions. *Numer. Math*. 31: 377–403.

[8] Hutchinson, M. F., de Hoog, F. R. (1985): Smoothing Noisy Data with Spline Functions. Numer. Math., 47: 99–106.

[9] Golub, G. H., Heath, M., Wahba, G (1979). Generalized Cross-Validation as a Method for Choosing a Good Ridge, 21 (2): 215-223.

[10] Wahba, G. (1990). Spline Models for Observational Data. *CBMS-NSF Regional Conference Series in Applied Mathematics.* 59, Philadelphia: SIAM. 1-180.

[11] Hurvich, C. M., Simonoff, J. S. and Tsai, C. (1998). Smoothing Parameter Selection in Non-Parametric Regression using an Improved Akaike Information Criterion. Journal of Royal Statistical Society, B, 60 (2): 271-293.

[12] Gu, C. (1992). Diagnostic for Nonparametric Regression Models with Additive Term. Journal of American Statistical Association. 87 (420): 1051-1058.

[13] Yuedong, W. (1998). Smoothing Spline Models with Correlated Random Errors. Journal of American Statistical Association. (93) 441: 341-348.

[14] Devi, A. R., Budiantara, I. N. and 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.

[15] Anna L., Tong, T. and Yuedong W. (2006). Smoothing Spline Estimation of Variance Function. Journal of Computational and Graphical Statistics 16 (2): 1-10.

[16] Aydin, D and Tuzemen, S. (2012). Smoothing Parameter Selection Problem in Nonparametric Regression Based on Smoothing Spline: A Simulation Study. Journal of Applied Sciences. 12 (7): 636-644.

[17] Aydin, D. and Memmedli, M. (2011). Optimum Smoothing Parameter Selection for Penalized Least squares inform of Linear Mixed Effect Models. Optimization, iFirst: 1–18.

[18] Chen, C. S. and Huang H. C. (2011): An improved Cp criterion for spline smoothing. Journal of Statistical Planning and Inference, 144 (1): 445-471.

[19] Aydin, D., M. Memmedli, and R. E. Omay. (2013). Smoothing Parameter Selection for Nonparametric Regression using Smoothing Spline. European Journal of Pure and Applied Mathematics. 6: 222–38.

[20] Jansen J. P. and Vieira, M. C. and Cope S. (2015). Network Meta-Analysis of Longitudinal Data Using Fractional Polynomials. Statistics in Medicine, 34 (15).

[21] Adams S. O. and Ipinyomi R. A. (2017). A Proposed Spline Smoothing Estimation Method for Time Series Observations. International Journal of Mathematics and Statistics Invention (IJMSI). 7 (2): 18-25.

[22] Xu, L and Zhou J. (2019). A Model-Averaging Approach for Smoothing Spline Regression. Communications in Statistics - Simulation and Computation. 48 (8): 2438-2451.

[23] Wahba, G. (1977). Applications of Statistics, in P. Krishnaiah (ed.), A Survey of Some Smoothing Problems and the Method of Generalized Cross-Validation for solving them, Northern Holland, Amsterdam. 14 (4): 651-667.

[24] Wahba, G. (1980). Automatic Smoothing of the Log Periodogram. Journal of the American Statistical Association 75: 122-132.

[25] Wahba, G., and Wang, Y. (1993). Behavior near Zero of the Distribution of GCV Smoothing Parameter Estimates for Splines. Statistics and Probability Letters. 25: 105-111.

[26] Mallows, C. L., (1973). Some Comments on Cp, Technometrics, 15 (4), 661-675.

[27] Wahba, G., Wang, Y., Gu, C., Klein, R., and KIein, B. (1995). Smoothing Spline ANOVA for Exponential Families, With Application to the Wisconsin Epidemiological Study of Diabetic Retinopathy. The Annals of Statistics, 23, 1865-1895.