World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

On some techniques of selecting spline smoothing parameters for a correlated dataset with autocorrelation structure in the residual

Samuel Olorunfemi Adams * and Mohammed Anono Zubair

Department of Statistics, University of Abuja, Abuja, Nigeria.

World Journal of Advanced Research and Reviews, 2023, 17(02), 068–078

Publication history: Received on 22 December 2022; revised on 31 January 2023; accepted on 02 February 2023

Article DOI: https://doi.org/10.30574/wjarr.2023.17.2.0216

Abstract

Residuals are minimized in a correlated dataset by selecting a smoothing parameter with optimum performance in the smoothing spline. The selection methods utilized in this study include Generalized Maximum Likelihood (GML), Generalized Cross-Validation (GCV), Unbiased Risk (UBR), and the Proposed Smoothing Method (PSM). The aim of this study is to compare the smoothing parameter selection ability of the four parameter selection methods for a correlated dataset with autocorrelation structure in the error term. To achieve this purpose, a Monte-Carlo simulation was conducted by utilizing program written in R-4.2.2. The performance of the parameter selection methods were evaluated using predictive Mean Squared Error (PMSE). Findings from the study indicated that GCV and GML were mostly affected by the presence of auto correlation in the residual and therefore had an asymptotically similar behavioural pattern. The estimators conformed to the asymptotic properties of the smoothing parameter selection methods considered; this is noticed in all the sample sizes and at all the smoothing parameters. The result also showed that; the most consistent and efficient among the four spline smoothing parameter selection methods considered in this study based on sample size and performance in the presence of autocorrelated residual error is the proposed smoothing method (PSM) because it does not undersmooth relative to the other smoothing method especially for small sample and medium sample size of 50 and 100.

Keywords: Autocorrelation; Generalized Maximum Likelihood; Generalized Cross-Validation; Penalized Spline; Splines Smoothing Time series; Spline regression

1. Introduction

In non-parametric regression, smoothing is of great importance because it is used to filter out noise or disturbance in observation; it is commonly used to estimate the mean function in a nonparametric regression model, it is also the most popular method used for prediction in non-parametric regression models. The general spline smoothing model is given as:

$$y_i = f(x_i) + \varepsilon_i \qquad (1)$$

Where; Yi is the observation value of the response variable y, f is an unknown smoothing function, Xi is the observation value of the predictor variable x and ε_i is normally distributed random errors with zero mean and constant variance.

The main objective of this research is to estimate $f(.)$ when $x_i = t_i$ but not necessarily equally spaced, with $t_1 < \ldots < t_n$ (time) and ε_i is assumed to be correlated [1]. Therefore, this research shall consider the spline smoothing for non-parametric estimation of a regression function in a time-series context with the model;

* Corresponding author: Samuel Olorunfemi Adams

$$y_i = f(t_i) + \varepsilon_{it}, \qquad i = 1,2,\ldots,n, t_i \epsilon [0,1] \tag{2}$$

Where; Y_i = observation values of the response variable y, f = an unknown smoothing function, t_i = time for $i = 1 \ldots n, et_i$ = zero mean autocorrelated stationary process.

Smoothing spline arises as the solution to a nonparametric regression problem having the function f(x) with two continuous derivatives that minimize the penalized sum of squares

$$S(g) = \sum_{i-1}^{n} \big(y_i - g(x_i)\big)^2 + \lambda \int_a^b \big(g''(x)\big)^2 dx \tag{3}$$

Where; λ is a smoothing constant, the first term in the equation is the residual sum of the square, and the second term is a roughness penalty, which is large when the integrated second derivative of the regression function $f^I(x)$ is large when $f(x)$ is rough (i.e. with a rapidly changing slope). The parameter λ controls the trade-off between goodness-of-fit and the smoothness of the estimate and is often referred to as the smoothing parameter. If λ is 0 then $f(x)$ simply interpolates the data, if λ is very large, then $\hat{f}$ will be selected so that $f^{II}(x)$ is everywhere 0, which implies a globally linear least-squares fit to all data. There is a need to tackle the problem associated with estimating the best spline smoothing methods for time series observation in the presence of correlational error. There is a vast literature on spline smoothing modeling of time series data in the presence of autocorrelated error [2] using the Smoothing Spline model to obtain a Generalized Maximum Likelihood (GML) estimate for the smoothing parameter, then this estimate is compared with the Generalized Cross Validation (GCV) estimate both analytically and by Monte-Carlo method. The comparison was based on a predictive Mean Square Error (PMSE). It was discovered that GCV was somewhat better than GML for n = 64, GCV was decidedly superior for n = 128 while for n= 32, GCV was better for smaller $\sigma^{2,}$ and the comparison was close for larger σ^2. [1] utilized the GCV criterion for choosing the degree of smoothing in spline regression and extended it to accommodate a time-series autocorrelated error sequence. It was demonstrated via simulation that the minimum GCV smoothing spline is an inconsistent estimator in the presence of autocorrelated error. Ignoring the moderate autocorrelation structure can seriously affect the performance of the Cross-Validation smoothing spline. [3] extended the GML, GCV, and UBR to estimate the smoothing parameters and the correlation parameters simultaneously, when the correlation matrix is assumed to depend on a parsimonious set of parameters. The GML method was recommended because it is stable and works well in all simulations. It performs better than other methods, especially when the sample size is small. [4] compared three methods, GML, GCV, and leaving-out-one-pair cross-validation to estimate the smoothing parameters, the weighting parameter, and the correlation parameter simultaneously. Based on simulated data, they concluded that the GML method has smaller Mean Squared Errors for the nonparametric functions and parameters and needs less computational time than the other methods and that it does not overfit data when the sample size is small. [5] reviewed the existing literature in kernel regression, smoothing splines, and wavelet regression under a correlation, both for short-range and long-range dependence. [6] studied smoothing splines with the degree of smoothing selected by Generalized Cross-Validation (GCV-Spline) and provides a method to find an optimal smoother for an fMRI time series, to determine if GCV-Spline of fMRI time series yields unbiased variance estimates of linear regression model parameters. The results from the real data suggest that GCV-Spline determines appropriate amounts of smoothing. The simulations show that the variance estimates are, on average, unbiased. It demonstrates that GCV-Spline is an appropriate method for smoothing fMRI time series. [7] used difference-based methods to construct estimators of error variance and autoregressive parameters in nonparametric regression with time series errors. They proved that the difference-based estimators can be used to produce a simplified version of time series cross-validation. [8] proposed to adjust the GCV criterion for the spatial correlation and showed that it leads to improved smoothing parameter selection results even when the covariance model is misspecified. [9] described the effects of moderate levels of serial correlation on one-sided and ordinary cross-validation in the context of local linear and kernel smoothing investigated. It is shown both theoretically and by simulation that one-sided cross-validation is much less adversely affected by correlation than in ordinary cross-validation. The former method is a reliable means of window width selection in the presence of moderate levels of serial correlation, while the latter is not. It is also shown that ordinary cross-validation is less robust to correlation. [10] investigated the behavior of data-driven smoothing parameters, for penalized spline regression in the presence of correlated data. It was shown for other smoothing methods that mean squared error minimizers, such as Generalized Cross-Validation or the Akaike Information criterion, are extremely sensitive to misspecifications of the correlation structure resulting in over- or (under-)fitting the data. [11] performs an asymptotic analysis of penalized spline estimators and compares P-splines and splines with a penalty of the type used with smoothing splines. It was shown that a P-spline and a smoothing spline are asymptotically equivalent provided that the number of knots of the P-spline is large enough, and the two estimators have the same equivalent kernels for both interior points and boundary points. [12] investigated a bandwidth selector based on the use of a bimodal kernel for nonparametric regression with a fixed design and proved that the proposed selector is quite effective when the errors are severely correlated. [13] 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 (UBR) 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. [14] 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 (GF_{AIC}) and Generalized Cross-Validation (GCV) 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. [15] compared three existing methods used to estimate the degree of smoothness parameter with a proposed smoothing method for time series data under the assumption that the error terms are independent. It was discovered that when the sample size is small (n =20), UBR and GCV were equally preferred and for n = 60 and 100 at smoothing parameters (λ = 1, 2, 3 and 4) UBR method was the best for estimating the degree of smoothness. [16] developed a new spline smoothing estimation method and compare it with three existing methods to eliminate the problem of overfitting associated with the presence of autocorrelation in the error term. The study discovered that the proposed smoothing method is the best for time series observations with autocorrelated error because it doesn't overfit and works well for large sample sizes. [17] proposed an efficient new spline smoothing estimation method and compared it with three classical methods to eliminate the problem of overfitting associated with the presence of Autocorrelation in the error term. The study discovered that the proposed smoothing method is the best for time-series observations with Autocorrelated error because it doesn't overfit and works well for large sample sizes. [18] proposed a smoothing spline technique by taking the hybrid of Generalized Cross Validation (GCV) and Mallow's CP criterion (MCP). The predicting performance of the Hybrid GCVMCP is compared with Generalized Cross Validation (GCV) and Mallow's CP criterion (MCP) using data generated through a simulation study and real-life data. 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.

This study aims to compare the smoothing parameter selection ability of the proposed spline smoothing method (PSM) with three classical estimation methods namely; Generalized Maximum Likelihood (GML), Generalized Cross Validation (GCV), and Unbiased Risk (UBR) for time series observations with autocorrelation structure.

The four spline smoothing estimation methods with autocorrelation structures were presented in section two. Section 3 presents the Monte Carlo simulation study, the equation used for generating values in simulation, experimental design, and data generation, section four compares the four methods through a simulation study, and the discussion of findings was presented in section five while conclusions were presented in the last section.

2. Spline Smoothing Estimation Methods with Autocorrelation Structure

2.1. Generalized Cross-Validation (GCV) Estimate Method

Several methods have been proposed for choosing the smoothing parameter. The most attractive class of such method is the Generalized Cross-Validation (GCV), given as;

$$GCV(\lambda) = \frac{n^{-1}\|(I - S\lambda)y\|^2}{[n^{-1}trace(I - S\lambda)]^2} \quad (4)$$

Where; n is the observations or data set(x_i, y_i), λ = smoothing parameter, S_λ = refers to the ith diagonal member of the smoothing matrix

2.2. Generalized Maximum Likelihood (GML) Estimation Method

A Bayesian model provides a general framework for the GML method and can be used to calculate the posterior confidence intervals of a spline estimate.

The GML estimates of λ is the maximizers of

$$GML(\lambda) = \frac{\lambda^{'}W(I - S\lambda)}{[det^{+}W(I - S\lambda)]^{\frac{1}{n-m}}} \quad (5)$$

$det^{+}(I - S\lambda)$ is the product of the $n - m$ nonzero eigenvalues of (I – Sλ), λ = 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 = number of zero eigenvalues [2].

2.3. Unbiased Risk (UBR) Estimate Method

The UBR method has been successfully used to select smoothing parameters for spline estimates with non-Gaussian data; it can be developed by applying the Weighted Mean Square Errors.

$$UBR(\lambda) = \frac{\frac{1}{n}\left\| W^{\frac{k}{2}}(I - S\lambda)y \right\|^2}{\left[\frac{1}{n}trace\left(W^{k-1}(I - S\lambda)\right)\right]^2} \text{ k = 0, 1, 2} \qquad (6)$$

Where; n is pairs of measurement/observations $\{xi, yi\}$,W is the correlation structure, λ is Smoothing parameters, S_λ is the ith diagonal element of smoother matrix [3].

2.4. Proposed Smoothing Method (PSM)

The proposed smoothing method (PSM) derived as the minimizer of equation 4 and 6 given by;

$$PSM\ (\lambda) = k\frac{(y - \hat{f})^T W(y - \hat{f})}{[trace(I - S\lambda)]^2} + (1 - k)\frac{\frac{1}{n}\left\| W^{\frac{1}{2}}(I - S\lambda) \right\|^2}{\left[\frac{1}{n}trace\{W(I - S\lambda)\}\right]^2} \qquad (7)$$

The proposed method for estimating $\boldsymbol{f}$ is given in equation (7) subject to the condition that $0 < k < 1$

Where; n is the number of dataset, k is the weighted value, $0 < k < 1$, W = V^{-1} = Correlation Matrix for the error term, $y = (y_1, \ldots, y_n)^T$ = Smoothing function, $\hat{f} = (f(t_1) \ldots f(t_n)).y_n)^T$ = $S_\lambda y$, Sλ= the diagonal member of the smoothing matrix, $\left\| W^{\frac{1}{2}}(I - S\lambda)y \right\|$ is the norm of the Euclidean vector $W^{\frac{1}{2}}(y - \hat{f})$, [16]-[19]

3. Material and method

3.1. Equation used for generating values in simulation

A simulation study is conducted to evaluate and compare the performance of the four estimation methods presented in previous sections. The model considered is;

$$y(t) = 2Sin\left(\frac{\pi}{t}\right) + \varepsilon_t \quad \text{t = 50, 100 and 150} \qquad (8)$$

Where; ε's are generated by a first-order autoregressive process AR (1) with mean 0, standard deviations 0.3 and 0.7 and first-order correlations (i.e. ρ = 0.1, 0.5 and 0.9) and its 95% Bayesian confidence interval, [1] and [20].

3.2. Experimental design and data generation

The experimental plan applied in this research work was designed to have three sample Sizes (n) of 50, 100 and 150, three autocorrelation levels, i.e. ρ = 0.1, 0.5 and 0.9, four smoothing functions were considered i.e. λ = 1, 2, 3 and 4, two standard deviation were considered, i.e. σ = 0.3 and 0.7. The data were generated for 1000 replications for each of the $3\times3\times4\times2=72$ combinations of cases n, ρ , λ, and σ. The criterion used is the PMSE values to evaluate $\hat{f}_\lambda$ computed according to each of the estimation given as;

$$PMSE(\lambda) = \sum_{i=1}^{n}\left(E\left[\hat{f}(x_i)\right] - f(x_i)\right)^2 \qquad (9)$$

Where; $f(x_i)$ is the value at knots x_i of the appropriate function given as $x_i = \frac{i-0.05}{n}$ [14]. A Simulation study was performed by using a program written in R, it was used to estimate all the model parameters, the criterion, the effect of autocorrelation on the estimated parameters and the performances of the four estimation methods i.e. Generalized Maximum Likelihood (GML), Generalized Crossed Validation (GCV), Unbiased Risk (UBR) and the Proposed Smoothing Method (PSM).

4. Result and discussion

In this study, we presented a modified Spline smoothing estimation method and compared its efficiency with three existing estimation methods namely; the Generalized Cross-Validation, Generalized Maximum Likelihood, and Unbiased Risks, we computed Predictive mean square errors criterion to measure their efficiency

4.1. Performance of the four smoothing methods based on predictive mean square error criterion when σ = 0.3

Table 1 presents the predictive mean square error for the four estimators, three sample sizes, four spline smoothing levels, and three correlation error levels at 0.3 sigma level. It was discovered that for GCV and sample size 50 the predictive mean square error of 4.938284 at λ = 1, decreases to 2.789043 at λ = 2 and further decreases to 2.018062 when λ = 4. The predictive mean square error increases as the level of autocorrelation increases from 4.938284 when ρ = 0.1 to 5.735483 when ρ = 0.5 and to 5.70041 when ρ = 0.9 for smoothing function (λ) = 1 and sample size = 50. It was also discovered that the predictive mean square error decreases as the sample size increases; at n = 50, the PMSE decreased from 4.938284 to 1.353605 at n = 100 and further decreased from 1.353605 to 0.394855 at n = 150 and smoothing function (λ) = 1.

The predictive mean square error (PMSE) of GML decreases from 3.788134 at λ = 1, to 3.624478 at λ = 3 and then decreased to 3.615046 at λ = 4. At sample size 50 the predictive mean square error is 3.902353, it decreased to 2.328352 as the sample size increased to 100 and further decreased to 2.314015 as the sample size increased to 150. It is noticed that the PMSE of GML increases from 2.638143 to 2.804273 as the autocorrelation error level increases from 0.1 to 0.5 but decreases from 2.804273 to 2.625861 as the autocorrelation level increases from 0.5 to 0.9. For all the other increases in autocorrelation error levels, the PMSE increased correspondingly, there is efficiency in GML.

For the Proposed Smoothing Method (PSM), it was discovered that the predictive mean square error increases as the autocorrelation level increases and decreases as the sample size increases. At sample size 50 the predictive mean square error of 4.208490 at λ = 2 decreases to 4.202272 at λ = 3 and further decreases to 3.615946 when λ = 4. The predictive mean square error of PSM decreases as the sample size increases, for λ = 1 and autocorrelation level of 0.1. PSM decreased from 4.188747 at sample size = 50 to 2.853925 at sample size 100 and further decreased to 2.287803 at sample size 150. The predictive mean square error of PSM increases from 2.853925 to 1.822216 as the autocorrelation error level increases from 0.1 to 0.5 for a sample size is 150 and increases from 1.822216 and 1.812007 as the autocorrelation error level increases of 0.5 to 0.9 for sample size is 150.

The predictive mean square error for UBR increases as the autocorrelation level increases and decreases as the smoothing levels and sample sizes increase. At sample size 50 the predictive mean square error of 3.777261 at λ = 1, decreases to 3.469432 at λ = 2, decreases to 3.416732 at λ = 3 but increased slightly to 3.98581 when λ = 4. The predictive mean square error of UBR decreases as the sample size increases, for λ = 2 and autocorrelation level of 0.5, UBR decreases from 3.469432 at sample size = 50 to 1.88788 at sample size 100 and further decreases to 1.431244 at sample size 150. The predictive mean square error of UBR increases from 3.416732 to 3.526772 as the autocorrelation error level increases from 0.1 to 0.5 for sample size is 50 and increases from 3.526772 and 3.611808 as the autocorrelation error level increases of 0.5 to 0.9 for sample size the same sample size.

Table 1 The PMSE result for GML, GCV, PSM and UBR with Autocorrelation Structure ρ = 0.1, 0.5 and 0.9 for n = 50, 100 and 150 when standard deviation (σ) = 0.3

		PMSE								
		n = 50			**n = 100**			**n = 150**		
Lamda	Smoothing Methods	**ρ= 0.1**	**ρ= 0.5**	**ρ= 0.9**	**ρ= 0.1**	**ρ= 0.5**	**ρ= 0.9**	**ρ= 0.1**	**ρ= 0.5**	**ρ= 0.9**
λ = 1	GCV	4.93828	5.73548	5.70041	1.35360	3.17988	5.81730	0.39485	4.19007	4.75306
	GML	3.78813	3.90233	4.55785	2.32835	2.42954	2.62586	2.31401	2.83604	2.43808
	PSM(k=1)	4.18874	1.97744	2.05909	2.85392	1.82221	1.81200	2.28780	1.57344	1.60574

	UBR	3.77726	2.81087	1.44908	2.10140	2.31704	1.11851	1.91307	2.07978	0.84175
λ = 2	GCV	2.78904	3.75568	5.36890	1.12314	1.37403	4.40631	0.34156	2.96876	3.18899
	GML	2.63814	2.80423	1.30049	2.19448	2.01800	1.02794	2.04044	1.33480	0.17112
	PSM(k=1)	4.20849	2.01893	2.10515	2.82329	1.87953	1.77842	2.28780	1.57340	1.20083
	UBR	3.46943	2.50677	1.01735	1.88788	1.61657	1.23034	1.43124	0.22050	1.53258
λ = 3	GCV	3.17514	3.50762	4.21841	2.47222	1.73035	1.45626	0.33490	0.81536	1.99245
	GML	3.62447	3.80280	4.26333	2.09433	2.95858	2.99648	1.99026	2.22264	0.80309
	PSM(k=1)	4.20227	2.02576	2.11214	1.81691	0.17547	1.76522	1.53195	0.46713	0.12489
	UBR	3.41673	3.52677	3.61180	1.85792	2.52561	2.56401	1.36111	1.86693	3.32113
λ = 4	GCV	2.01806	3.42688	2.16943	1.094332	0.173144	2.74644	0.332736	2.76541	2.928445
	GML	3.61594	2.80051	1.25093	2.175146	1.938749	5.985579	1.973208	1.98451	5.983278
	PSM(k=1)	4.11762	2.02809	2.11447	1.814626	1.701375	1.760514	1.500005	1.43017	1.098286
	UBR	3.39881	3.51261	4.92771	1.857928	1.94582	3.615934	1.337717	1.81572	3.257353

Table 2 presents the predictive mean square error for the four estimators, three sample sizes, four spline smoothing levels, three correlation error levels, and at 0.7 sigma level. It was discovered that for GCV, at ρ = 0.5 and sample size 50 the predictive mean square error of 2.217985 at λ = 1, decreases to 2.038837 at λ = 2, decreases to 1.975886 at λ = 3 and further decreases to 0.873763 when λ = 4. The predictive mean square error increases as the level of autocorrelation increases from 2.217985 when ρ = 0.1 to 4.652218 when ρ = 0.5 and to 5.219997 when ρ = 0.9 for smoothing function (λ) = 1 and sample size = 50. It was also discovered that for smoothing function (λ) = 2, the predictive mean square error decreases as the sample size increases; at n = 50 the PMSE decreased from 2.038837 to 1.036064 at n = 100 and further decreased to 0.106917 at n = 150.

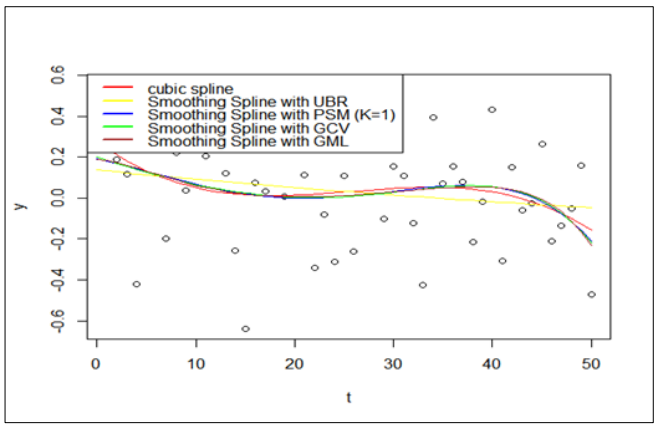
The predictive mean square error (PMSE) of GML decreases as the smoothing parameter increases. For small sample size and at ρ = 0.9, the predictive mean square error decreased from 1.460676 at λ = 1 to 1.191663 at λ = 2 then decreased to 1.152826 at λ = 3 and further decreased to 1.139958 at λ = 4. The predictive mean square error of GML decreases as the sample size increases. At sample size 50 the predictive mean square error is 1.402249, it decreased to 1.285324 as the sample size increased to 100 and further decreased to 0.917754 as the sample size increased to 150. It is noticed that the predictive mean square error of GML increases from 1.344602 to 2.150393 as the autocorrelation error level increases from 0.1 to 0.5, and increases from 2.150393 to 2.723054 as the autocorrelation level increases from 0.5 to 0.9. Thus there is efficiency in GML, but it was observed that predictive mean square error decreased as the autocorrelation error level increased.

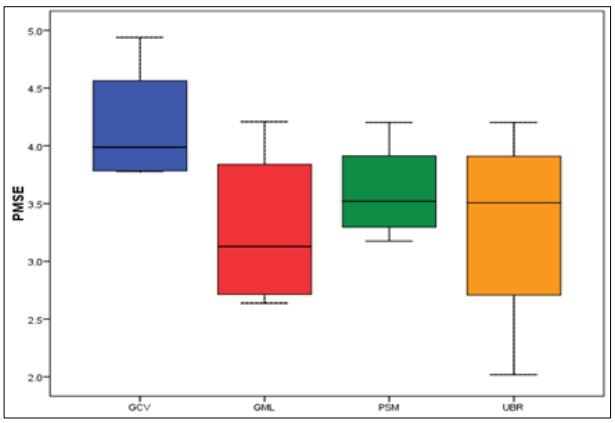
For the Proposed Smoothing Method (PSM), it was discovered that the predictive mean square error decreases as the autocorrelation level, smoothing parameter and sample size increases. At sample size 50 the predictive mean square error of 4.188747 at λ = 1 increased to 4.208498 at λ = 2 but decreases to 4.02272 when λ = 3 and further decreased to 4.117621 when λ = 4. The predictive mean square error of PSM decreases as the sample size increases, for λ = 2 and autocorrelation level of 0.1. PSM decreased from 1.706005 at sample size = 50 to 1.337262 at sample size 100 and further decreased to 1.111343 at sample size 150. The predictive mean square error of PSM decreases from 1.9762941 to 1.878994 as the autocorrelation error level increases from 0.1 to 0.5 for sample size 50 and further decreases from 1.878994 to 1.62727 as the autocorrelation error level increases of 0.5 to 0.9 for sample size is 50.

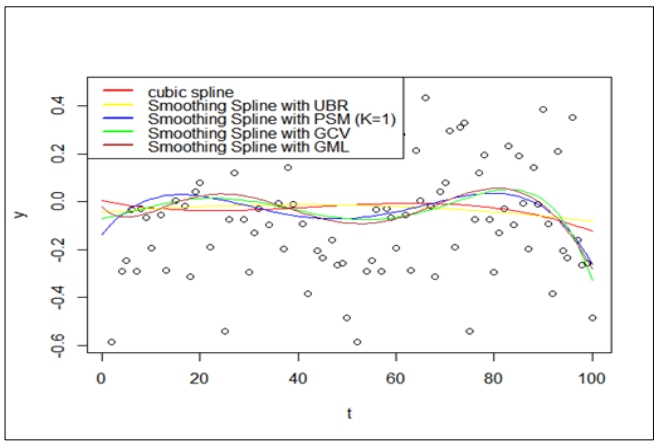
The predictive mean square error for UBR increases as the autocorrelation level decreases as the smoothing level and sample size increases. At sample size 50 the predictive mean square error of 3.946115 at λ = 1, decreases to 2.285086 at λ = 2 to 2.166318 at λ = 3 and further decreases to 1.259853 when λ = 4. The predictive mean square error of UBR decreases as the sample size increases, for λ = 4 and autocorrelation level of 0.9, UBR decreases from 2.549091 at sample size = 50 to 2.412688 at sample size 100 and further decreases to 1.540203 at sample size 150. The predictive mean square error of UBR increases from 2.166318 to 2.202126 as the autocorrelation error level increases from 0.1 to 0.5 for sample size is 50 and increases from 2.202126 to 2.563679 as the autocorrelation error level increases of 0.5 to 0.9 for sample size the same sample size, but it was observed that predictive mean square error decreased as the autocorrelation error level increases.

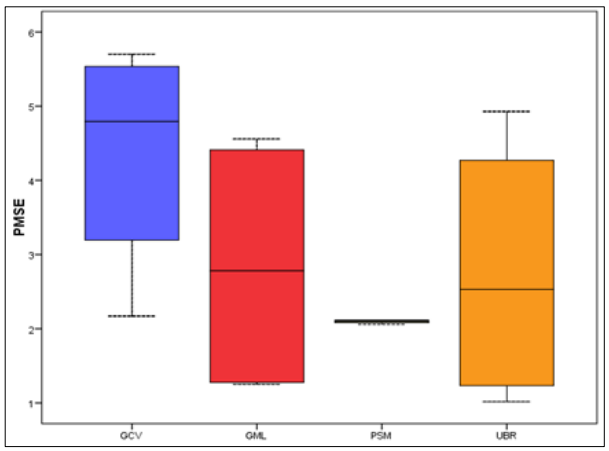
Table 2 The PMSE result for GML, GCV, PSM and UBR with Autocorrelation Structure ρ = 0.1, 0.5 and 0.9 for n = 50, 100 and 150 when standard deviation (σ) = 0.7

		PMSE								
		n = 50			n = 100			n = 150		
Lamda	Smoothing Methods	ρ = 0.1	ρ = 0.5	ρ = 0.9	ρ = 0.1	ρ = 0.5	ρ = 0.9	ρ = 0.1	ρ = 0.5	ρ = 0.9
λ = 1	GCV	2.217985	4.652218	5.219991	1.5079261	3.032906	3.355379	0.109678	0.205153	4.068174
	GML	1.402249	2.213838	2.854191	1.285324	2.424851	2.860878	0.917754	1.498209	1.460676
	PSM(k=1)	1.9762941	1.878994	1.62727	1.681525	1.655205	2.622758	1.625184	1.060796	1.814121
	UBR	3.946115	2.170123	2.854018	3.477279	1.895938	1.904192	0.715411	1.410622	1.391461
λ = 2	GCV	2.038837	1.550266	2.357644	1.036064	3.064901	3.686213	0.106917	0.204841	2.641265
	GML	2.353263	2.159928	2.742754	1.61744	1.745815	1.801702	0.916592	1.484834	1.191663
	PSM(k=1)	1.706005	1.883573	1.512748	1.337262	1.815278	1.258637	1.111343	1.555058	0.824054
	UBR	2.285086	2.043898	2.606053	1.686028	1.615925	1.94976	0.715436	0.391479	1.213843
λ = 3	GCV	1.975886	2.465147	2.230474	1.106586	1.865407	1.493562	0.914299	1.204822	1.462472
	GML	1.344602	2.150393	2.723054	2.376657	1.703152	1.747526	0.916174	0.482901	1.152826
	PSM(k=1)	1.691873	1.799777	1.490825	1.289702	1.65212	1.185653	1.188291	1.786081	1.525496
	UBR	2.166318	2.202126	2.563679	1.335866	2.149228	2.283664	0.715459	0.388746	1.832608
λ = 4	GCV	0.873763	1.437364	2.188967	0.106479	2.800442	1.430831	0.956241	0.204817	1.404276
	GML	1.341634	2.147087	2.716225	1.296255	2.050446	1.895078	0.916018	0.482256	1.139858
	PSM(k=1)	1.686857	1.794844	1.483121	1.27395701.	1.659382	1.159813	1.104291	1.454671	1.259721
	UBR	1.259853	2.014616	2.549091	221922	1.578077	2.412688	0.715468	0.387835	1.540203









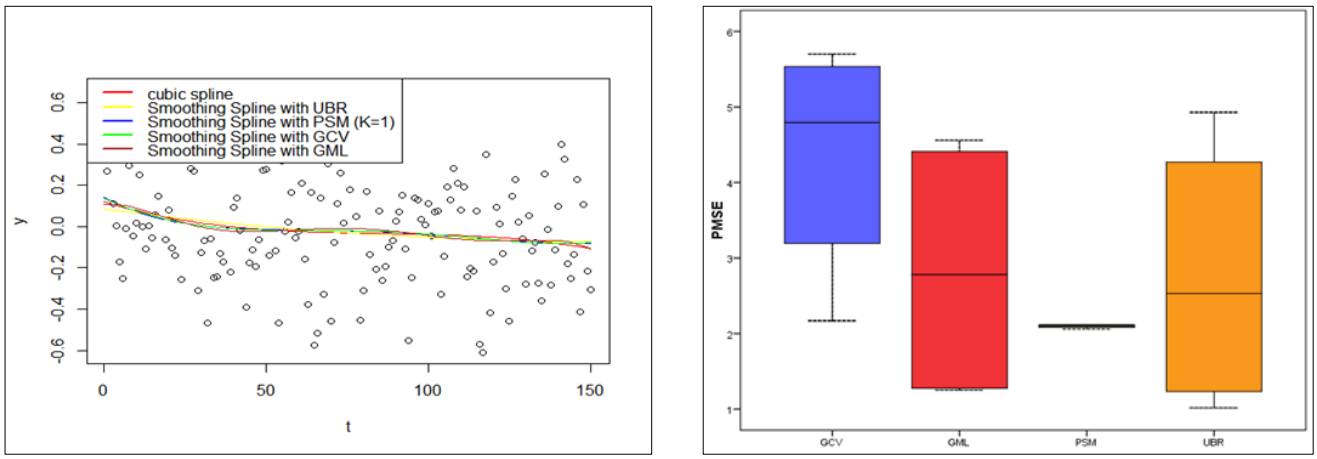
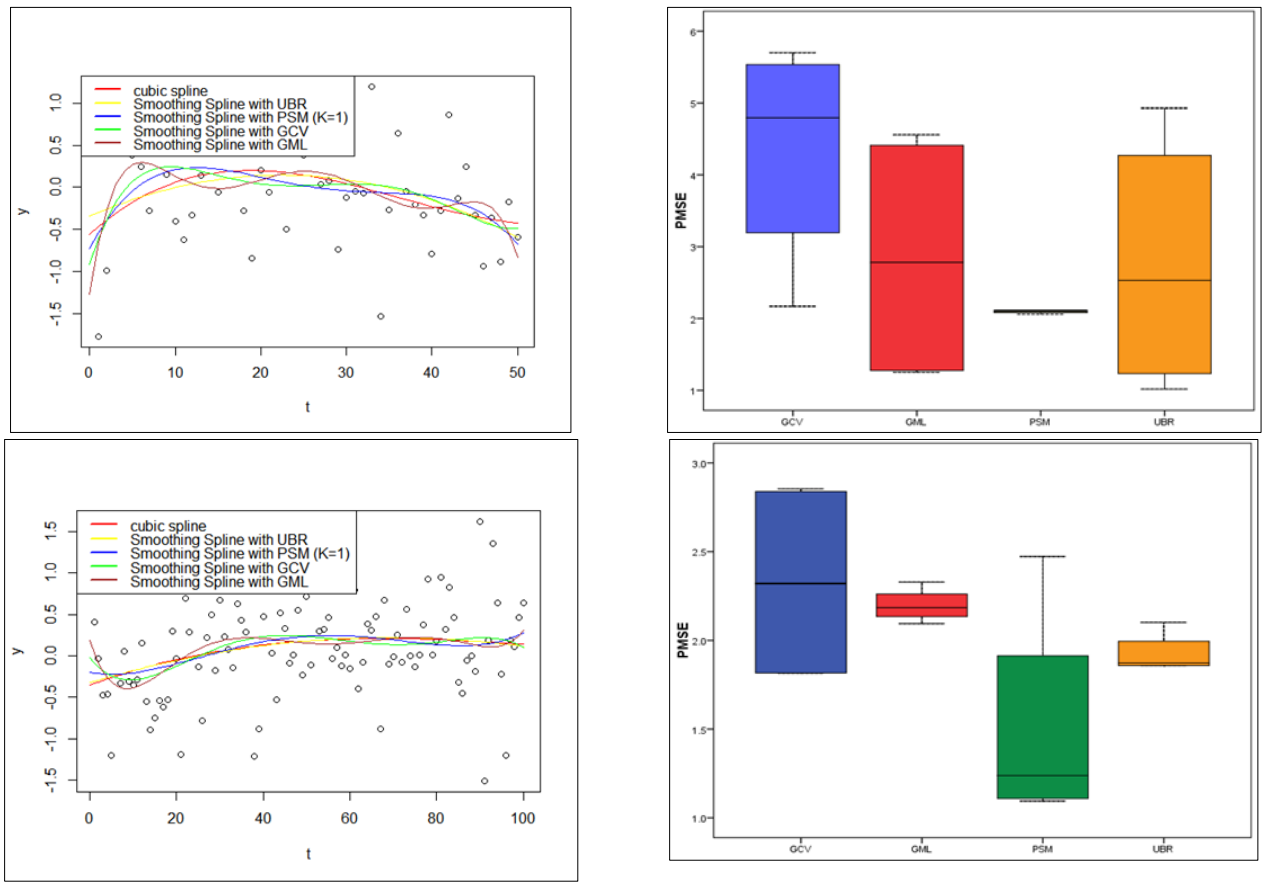


Figure 1 Right: Spline smoothing curve of the PMSE of the smoothing splines curve, with λ selected by UBR (yellow), PSM (blue), GCV (green), and GML (brown), by using different time series sample sizes of (50, 100 and 150). Left: Box plot of GCV, GML, PSM, and UBR for one of the simulated sample curves ρ = 0.9 and σ=0.3



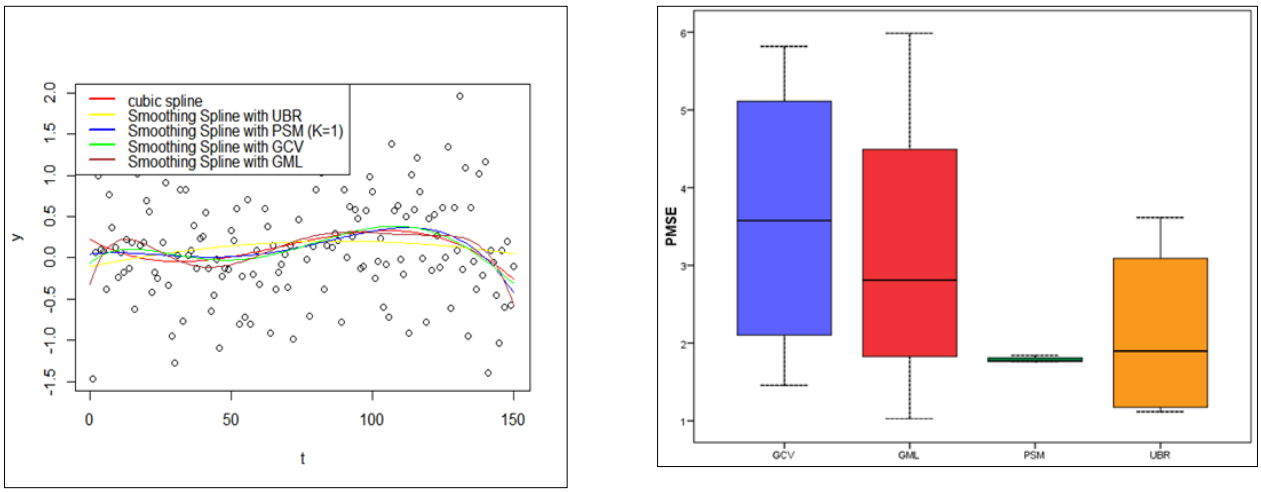


Figure 2 Right: Spline smoothing curve of the PMSE of the smoothing splines curve, with λ selected by UBR (yellow), PSM (blue), GCV (green), and GML (brown), by using different time series sample sizes of (50, 100 and 150). Left: Box plot of GCV, GML, PSM, and UBR for one of the simulated sample curves ρ = 0.9 andσ=0.7

Figure 1 and 2 presents the predictive mean square error estimates of GCV, GML, PSM, and in 1000 replications. From these plots, we can see that the PSM and UBR estimates have small PSMEs compared with GCV and GML, an indication that the four smoothing methods select the smoothing parameters very well but the PSM and UBR provide better estimates than GCV and GML through a simulation study. The PSM method is more stable when the sample size is small, such as when n = 50 while the UBR method performs slightly better when n = 100. In this case, there were several replications where GCV and GML provided more estimates of smoothing parameters which lead to the under-smoothing of the data. This behavior of the GCV method was investigated in [3] and [21]

Table 3 Summary of the predictive mean square error and ranks of the smoothing methods in the presence of autocorrelation error

Autocorrelation Levels	**Smoothing method**			
	GCV	GML	PSM (k=1)	UBR
α = 0.1	1.08	1.39	1.47	1.63
α = 0.5	1.89	1.71	1.66	1.48
α = 0.9	2.63	1.99	1.27	2.09
Grand mean	1.87	1.70	1.47	1.73
Rank	4	2	1	3

Table 4 Summary of the predictive mean square error and ranks of the smoothing methods based on sample size

Sample Size	**Smoothing method**			
	GCV	GML	PSM (k=1)	UBR
n = 50	2.434	2.179	1.711	2.326
n = 100	2.041	1.900	1.549	1.921
n = 150	1.124	1.047	1.145	0.951
Grand mean	1.867	1.709	1.468	1.732
Ranks	4	2	1	3

5. Conclusion

In this study, we presented Spline smoothing estimation method for time series observations in the presence of autocorrelated errors and based on sample size. The result presented in tables 3 and 4 showed that all the smoothing methods compared and compete favorably in the presence of autocorrelation error and an increase in sample size. The simulation result under the finite sampling properties of the PMSE criterion shows that all estimators are consistent and adversely affected by autocorrelated error the estimators' ranks are as follows, PSM, GML, UBR, and GCV. The result suggested that PSM should be preferred when the autocorrelation level is mild and high ($\rho = 0.5 - 0.9$). This finding is corroborated by those of [1], [19], [22], and [23].

If there is low autocorrelation in the observations, (i.e. $\rho = 0.1$) the unbiased Risk (UBR) should be considered. It was observed that GCV and GML were mostly affected by the presence of autocorrelation and therefore had an asymptotically similar behavioral pattern. It was also discovered that the estimators conformed to the asymptotic properties of the smoothing methods considered; this is noticed in all the sample sizes and all the smoothing parameters.

The most consistent and efficient among the four spline smoothing methods considered in this study based on sample size and performance in the presence of autocorrelation error is the proposed smoothing method (PSM) because it does not under smooth relative to the other smoothing method, especially for small sample size i.e. n = 50 and 100. (See Figures 1 and 2). This discovery is in agreement with the Monte-Carlo experiments' results from [2], [18], [19], [24], [25], [26], and [27]. It is also noticed that the predictive mean square error of the proposed smoothing method (PSM) goes to zero at a faster rate in the presence of autocorrelation error than the PMSE of the other smoothing methods considered in this study (see Tables 3 and 4). The next in terms of performance, consistency, and efficiency in the presence of autocorrelation is Generalized Maximum Likelihood (GML), Unbiased Risk (UBR) and the least in is Generalized Cross-Validation (GCV).

Compliance with ethical standards

Acknowledgments

The authors appreciate the effort of everyone that contributed to this study, the anonymous reviewers and editors for their constructive input in this manuscript.

Disclosure of conflict of interest

The author declared that there was no conflict of interest during the cause of this study and producing and submitting this manuscript for publication.

References

[1] Diggle, P.J. and Hutchinson, M.F. (1989). On spline smoothing with autocorrelated errors. Australian Journal of Statistics, 31: 166 –182.

[2] 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.

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

[4] Yuedong, W., Wensheng G. and Brown M.B. (2000). Spline Smoothing for Bivariate data with application to association between hormones, Statistica Sinica, 10: 377 – 397.

[5] Opsomer J., Yuedong W. and Yang Y. (2001). Nonparametric Regression with correlated Error Statistical Sciences, 6: (2) 134 – 153.

[6] Carew, J. D.,Wahba, G., Xie X, Nordheim, E.V. and Meyerand M. E. (2003), Optimal Spline Smoothing of FMRI Time Series by Generalized Cross-Validation, NeuroImage, 18(4): 950 – 961.

[7] Hall, P. and Keilegom, I. (2003). Using Difference-Based Methods for Inference in Nonparametric Regression with Time Series Errors. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65 (2): 443 – 456.

[8] Francisco-Fernandez, M. and Opsomer, J.D. (2005). Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. Canadian Journal of Statistics, 33(2): 279–295.

[9] Hart, J. D. and Lee, C. (2005). Robustness of one-sided cross-validation to autocorrelation. Journal of Multivariate Analysis, 92:77 – 96.

[10] Krivobokova T. and Kauermann G. (2007). A note on Penalized Spline Smoothing with Correlated Errors. Journal of the American Statistical Association, 102: 1328 – 1337.

[11] Wang Xiao, Shen Jinglai, Ruppert David (2011). On the Asymptotics of Penalized Spline Smoothing, Electronic Journal of Statistics, 5, 1-17 https://doi.org/10.101214/10-EJS593

[12] Kim, T., Park, B., Moon, M. and Kim C. (2009). Using bimodal kernel for inference in Nonparametric regression with correlated errors. Journal of Multivariate Analysis. 100 (7), 1487 – 1497.

[13] 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.

[14] 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.

[15] 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.

[16] Adams, S.O., Ipinyomi, 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

[17] Adams, S.O., Ipinyomi, 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.

[18] 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

[19] Adams, S.O. (2021). An Improved Spline Smoothing Method for Estimation in the Presence of Autocorrelation Errors. University of Ilorin.

[20] Wahba, G. (1983), Bayesian Confidence intervals for the cross-validated smoothing Spline, Journal of Royal Statistical Society Service. B. 45:133 – 150.

[21] 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.

[22] 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

[23] 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

[24] Barry, D. (1983). Nonparametric Bayesian regression, Ph.D. thesis, Yale University, New Haven, Connecticut.

[25] Adams, S.O., Obaromi, A.D, Alumbugu, A.I. (2021). 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

[26] Adams, S.O., Ipinyomi R.A. and Yahaya H.U. (2017). Smoothing Spline of ARMA Observations in the Presence Of Autocorrelation Error. European Journal of Statistics and Probability, 5(1): 1 – 8.

[27] 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