A Modified Smoothing Estimation Method for Time Series Data in The Presence of Autocorrelated Error

[1*]**Adams Samuel Olorunfemi; [2]Rueben Adeyemi Ipinyomi**

[*1]Department of Statistics,
University of Abuja, Nigeria

[2]Department of Statistics,
University of Ilorin, Nigeria.
E-mail: samuel.adams@uniabuja.edu.ng[*]

***Abstract*- Spline Smoothing is used to filter out noise or disturbance in an observation, its performance depends on the choice of smoothing parameters. There are many methods of estimating smoothing parameters; most popular among them are; Generalized Maximum Likelihood (GML), Generalized Cross-Validation (GCV), and Unbiased Risk (UBR), this methods tend to underestimate smoothing parameters in the presence of autocorrelation error. A new Spline Smoothing Estimation method is proposed by modifying the Generalized Cross-Validation and Unbiased Risk methods. It is demonstrated through a simulation study performed by using a program written in R to compare the new Spline Smoothing Estimation method and the three existing methods, the comparison was based on the predictive Mean Score Error criteria. The Proposed method is recommended; because it performed better than other methods, especially for a small sample size.**

Keywords - *Autocorrelation, Generalized Maximum Likelihood, Generalized Cross-Validation, Penalized Spline, Splines Smoothing, Time series and Spline regression.*

I. Introduction

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

$$y_i = f\left(X_i\right) + \varepsilon_i \tag{1.1}$$

Where; Y_i is the observation values of the response variable y, f is an unknown smoothing function, X_i is the observation values 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 $t_1 < . . . < t_n$ (time) and ε_i is assumed to be correlated. Diggle and Hutchinson (1989). Therefore, this research shall consider the spline smoothing for non-parametric estimation of a regression function in a time-series context with the model;

$$y_i = f\left(t_i\right) + \varepsilon_{ti} \tag{1.2}$$

where; Y_i = observation values of the response variable y, f = an unknown smoothing function, t_i = time for i = 1 . . . n, e_{ti} = 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 minimizes the penalized sum of squares

$$SS^*(h) = \sum_{i=1}^{n} \{yi - f(x_i)\}^2 + \lambda \int_{x_{\min}}^{x_{\max}} \{f''(x)\}^2 dx \tag{1.3}$$

where; λ is a smoothing constant, the first term in the equation is the residual sum of square, the second term is a roughness penalty, which is large when the integrated second derivative of regression function $f''(x)$ is large when f(x) is rough (i.e. with 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 $\hat{f}(x)$ simply interpolates the data, if λ is very large, then $\hat{f}$ will be selected so that $\hat{f}''(x)$ is everywhere 0, which implies a globally linear least-squares fit to all data. There is the need to tackle the problem associated with estimating the best spline smoothing methods for time series observation in the presence of correlational error, Diggle and Hutchinson (1989).

There are vast literatures on Spline Smoothing modeling of time series data in the presence autocorrelated error; Diggle and Hutchinson (1989), Yuedong (1998), Yuedong et. al. (2000), Opsomer, Yuedong and Yang (2001), Wahba et. al. (1995), Carew et. al (2002), Hall and Keilegom (2003), Francisco-Fernandez and Opsomer (2005), Hart and Lee (2005), Krivobokova and Kauermann (2007), Shen (2008), Kim, Park, Moon, and Kim (2009), Morton et.al. (2009), Wang, Meyer and Opsomer (2013), Adams, Ipinyomi and Yahaya (2017) Chen and Huang (2011).

The objective of this study is to propose a new smoothing method (PSM) by modifying two of the existing spline smoothing methods (i.e. the Generalized Cross Validation (GCV) and Unbiased Risk (UBR)) and compare it with three existing estimation methods namely; Generalized Maximum Likelihood (GML), Generalized Cross Validation (GCV) and Unbiased Risk (UBR) for time series observations in the presence of autocorrelated error.

Spline smoothing estimation methods for time series observations in the present of autocorrelation error were discussed in section one. Section two reviews the existing spline smoothing method and the proposed selection method, Section 3 presents the Monte Carlo simulation study, equation used for generating values in simulation and experimental design and data generation, section four compares the four methods via a simulation study, and finally, the result discussion and conclusion were presented in last section.

II. 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;

$$GCV = \frac{\frac{1}{n}\left\|(I - A_\lambda)\lambda\right\|^2}{\left[\frac{1}{n} Trace \quad (I - A_\lambda)\right]^2} \tag{1.4}$$

Where; n is Pairs of measurement/observations {xi,yi}, λ is Smoothing parameters, A_λ is the ith diagonal element of smoother matrix.

a. 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

$$M(\lambda) = \frac{y'(I - A(\lambda))y}{\left[\det{}^+(I - A(\lambda))\right]^{1/(n-m)}} \tag{1.5}$$

Where; $\det^+(I - A(\lambda))$ is the product of the n – m nonzero eigenvalues of [I – A(λ)], y is Smoothing parameter, W is the correlation structure, A is the diagonal element of smoother matrix, n is the pairs of measurement/observations and m is number of zero eigenvalues, Wahba (1985).

b. 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.

The Unbiased Risk is therefore given as;

$$V_k(\lambda) = \frac{\frac{1}{n}\left\|W^{\frac{K}{2}}(I - S_\lambda)\lambda\right\|^2}{\left[\frac{1}{n} tr\left(W^{k-1}(I - S_\lambda)\right)\right]^2} \qquad k = 0,1,2 \tag{1.6}$$

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. Yuedong (1998).

c. Proposed Smoothing Method (PSM)

A Spline Smoothing model is defined as

$$y_i = f(X_i) + \varepsilon_i \tag{1.7}$$

where; Y is the response variable, X is vector of the predictor variable, F is Regression function, and = error term ε

GCV becomes modified as

$$G(\lambda) = \frac{Y^I W(I - S_\lambda)Y}{\det\left[W(I - S_\lambda)\right]^{\frac{1}{n-m}}} \tag{1.8}$$

To extend GCV, unbiased Risk method was proposed and correlation structure was introduced as;

$$U(\lambda) = \frac{\frac{1}{n}\left\|W^{\frac{K}{2}}(I - S_\lambda)\right\|^2}{\left[\frac{1}{n} Trace\left(W^{k-1}(I - S_\lambda)\right)\right]^2} \tag{1.9}$$

where k = 1. The new Spline smoothing selection method is proposed to allow the presence of correlation structure.

The proposed smoothing method (PSM) derived is the minimizer of V(λ) given by

$$P(\lambda)=\frac{\frac{1}{n}\left\|(I-S_{\lambda})\lambda\right\|^{2}}{\frac{1}{n}\left\|W^{\frac{1}{2}}(I-S_{\lambda})\right\|^{2}} \tag{1.10}$$

where; n is Pairs of observations, λ is Smoothing parameter, W is the correlation structure, S_λ is the diagonal element of smoother matrix

III. Materials and Method

a. 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

$$y_{1i}=\frac{Sin\ \pi_{i}}{t}+\varepsilon_{t} \qquad i=1,2,\ldots n \qquad t=\in[0,2] \tag{1.11}$$

where; ε's are generated by a first-order autoregressive process AR (1) with mean 0, standard deviations 0.8 and 1.0 and first-order correlations (i.e. ρ = 0.2, 0.5 and 0.8) and its 95% Bayesian confidence interval. Wahba, (1983) and Diggle, (1989).

b. Experimental design and data generation

The experimental plan applied in this research work was designed to have three sample Sizes (n) of 20, 60 and 100, three autocorrelation levels, i.e. α = 0.2, 0.5 and 0.8, four smoothing functions were considered i.e. λ = 1, 2, 3 and 4, two standard deviation were considered, i.e. σ = 0.8 and 1.0. The data were generated for 1000 replications for each of the $3\times3\times4\times2=72$ combinations of cases n, α, λ, and σ. The criterion used is the PMSE values to evaluate $\hat{f}_{\lambda}$ computed according to each of the estimation given as;

$$PMSE(\lambda)=\tfrac{1}{n}E\left[\left(f(x_i)-\hat{f}(x_i)\right)^2\right],\ \left(\hat{f}_{\lambda}(x_i)=\left(\hat{f}_{\lambda}\right)i\right) \tag{1.12}$$

where $f(x_i)$ is the value at knots x_i of the appropriate function given as $x_i=\frac{i-0.05}{n}$ Aydin, Memmedli and Omay (2013). 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).

IV. Simulation Result

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.

a. Performance of the four smoothing methods based on predictive mean square error.

Criterion when σ = 0.8:

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 0.8 sigma level. It was discovered that for GCV and for sample size 20 the predictive mean square error of 4.938284 at λ = 1, decreases to 2.789043 at λ = 2 and further decreased to 2.018062 when λ = 4. The predictive mean square error increases as the level of autocorrelation increases from 4.938284 when α = 0.2 to 5.735483 when α = 0.5 and to 5.70041 when α = 0.8 for smoothing function (λ) = 1 and sample size = 20. It was also discovered that the predictive mean square error decreases as the sample size increases; at n = 20 the PMSE decreased from 4.938284 to 1.353605 at n = 60 and further deceases from 1.353605 to 0.394855 at n = 100 and for 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 20 the predictive mean square error is 3.902353, it decreased to 2.328352 as the sample size increased to 60 and further decreased to 2.314015 as the sample size increased to 100. It is noticed that the PMSE of GML increases from 2.638143 to 2.804273 as the autocorrelation error level increases of 0.2 to 0.5, but decreases from 2.804273 to 2.625861 as the autocorrelation level increases from 0.5 to 0.8. For all the other increase in autocorrelation error levels, the PMSE increased correspondingly, thus 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 20 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 0.2. PSM decreased from 4.188747 at sample size = 20 to 2.853925 at sample size 60 and further decreased to 2.287803 at sample size 100. The predictive mean square error of PSM increases from 2.853925 to 1.822216 as the autocorrelation error level increases of 0.2 to 0.5 for sample size is 60 and increases from 1.822216 and 1.812007 as the autocorrelation error level increases of 0.5 to 0.8 for sample size is 60.

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 20 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 = 20 to 1.88788 at sample size 60 and further decreased to 1.431244 at sample size 100. The predictive mean square error of UBR increases from 3.416732 to 3.526772 as the autocorrelation error level increases of 0.2 to 0.5 for sample size is 20 and increases from 3.526772 and 3.611808 as the autocorrelation error level increases of 0.5 to 0.8 for sample size the same sample size.

Table 1: The MSE result of the simulated study for GML, GCV, PSM and UBR in the presence of autocorrelation (α) = 0.3, 0.5 and 0.8 for n = 20, 60 and 100 when standard deviation (σ) = 0.8.

λ	Smoothing Methods	PMSE N = 20 ρ= 0.2	ρ= 0.5	ρ= 0.8	N = 60 ρ= 0.2	ρ= 0.5	ρ= 0.8	N = 100 ρ= 0.2	ρ= 0.5	ρ= 0.8
λ = 1	GCV	4.938284	5.735483	5.700411	1.353605	3.179886	5.817303	0.394855	4.190077	4.753061
	GML	3.788134	3.902353	4.557857	2.328352	2.429546	2.625861	2.314015	2.836043	2.438085
	PSM(k=1)	4.188747	1.977449	2.05909	2.853925	1.822216	1.812007	2.287803	1.573442	1.605743
	UBR	3.777261	2.810875	1.449087	2.101405	2.317046	1.118518	1.913073	2.079789	0.841755
λ = 2	GCV	2.789043	3.755684	5.368908	1.123143	1.374032	4.406313	0.341562	2.96876	3.188995
	GML	2.638143	2.804237	1.300494	2.19448	2.018002	1.027948	2.040446	1.334802	0.171129
	PSM(k=1)	4.208498	2.018938	2.105152	2.823294	1.879530	1.778426	2.287803	1.573403	1.200836
	UBR	3.469432	2.506771	1.017353	1.88788	1.616574	1.230349	1.431244	0.220508	1.532589
λ = 3	GCV	3.175146	3.507623	4.218419	2.472227	1.730359	1.456264	0.334902	0.815361	1.992452
	GML	3.624478	3.802802	4.263339	2.094332	2.958588	2.996486	1.990265	2.22264	0.8030926
	PSM(k=1)	4.202272	2.025768	2.112142	1.816911	0.175471	1.765224	1.531958	0.467133	0.124897
	UBR	3.416732	3.526772	3.611808	1.857928	2.525618	2.564013	1.361115	1.866935	3.321139
λ = 4	GCV	2.018062	3.42688	2.169436	1.094332	0.173144	2.74644	0.332736	2.765412	2.928445
	GML	3.615946	2.800514	1.250932	2.175146	1.938749	5.985579	1.973208	1.984518	5.983278
	PSM(k=1)	4.11762	2.028096	2.114477	1.814626	1.701375	1.760514	1.500005	1.430172	1.098286
	UBR	3.398581	3.512612	4.927715	1.857928	1.94582	3.615934	1.337717	1.815722	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 1.0 sigma level. It was discovered that for GCV, at α = 0.5 and sample size 20 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 decreased to 0.873763 when λ = 4. The predictive mean square error increases as the level of autocorrelation increases from 2.217985 when α = 0.2 to

4.652218 when α = 0.5 and to 5.219997 when α = 0.8 for smoothing function (λ) = 1 and sample size = 20. It was also discovered that for smoothing function (λ) = 2, the predictive mean square error decreases as the sample size increases; at n = 20 the PMSE decreased from 2.038837 to 1.036064 at n = 60 and further deceased to 0.106917 at n = 100.

The predictive mean square error (PMSE) of GML decreases as the smoothing parameter increases. For small sample size and at α = 0.8, the predictive mean square error decreased from 1.460676 at λ = 1 to 1.191663 at λ = 2 then decreases 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 as the sample size increases.

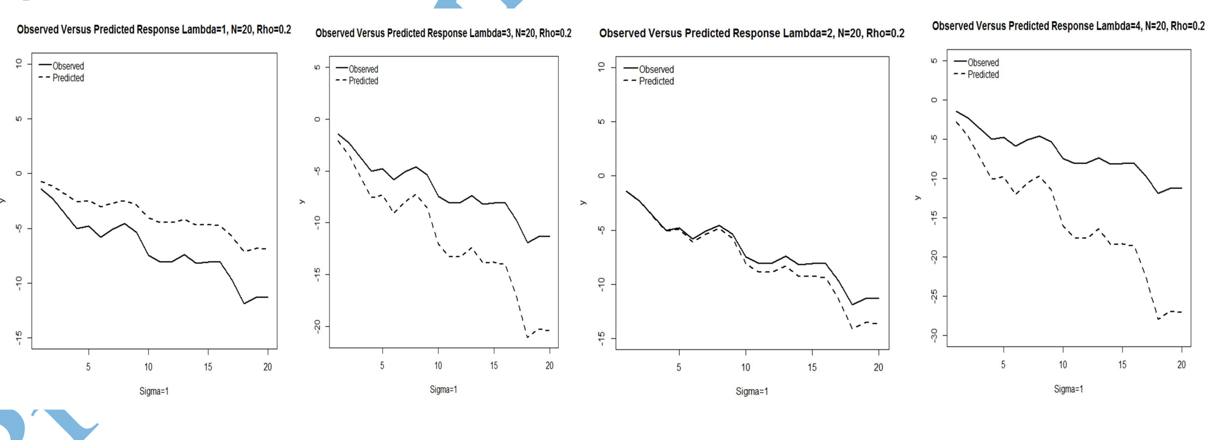
At sample size 20 the predictive mean square error is 1.402249, it decreased to 1.285324 as the sample size increased to 60 and further decreased to 0.917754 as the sample size increased to 100. It is noticed that the predictive mean square error of GML increases from 1.344602 to 2.150393 as the autocorrelation error level increases of 0.2 to 0.5, and increases from 2.150393 to 2.723054 as the autocorrelation level increases from 0.5 to 0.8. Thus there is efficiency in GML, but it was observed that predictive mean square error decreased as the autocorrelation error level increases.

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 20 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 decreases 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 0.2. PSM decreased from 1.706005 at sample size = 20 to 1.337262 at sample size 60 and further decreased to 1.111343 at sample size 100. The predictive mean square error of PSM decreases from 1.9762941 to 1.878994 as the autocorrelation error level increases of 0.2 to 0.5 for sample size is 20 and further decreases from 1.878994 to 1.62727 as the autocorrelation error level increases of 0.5 to 0.8 for sample size is 20.

The predictive mean square error for UBR increases as the autocorrelation level decreases as the smoothing level and sample size increases.

At sample size 20 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 0.8, UBR decreases from 2.549091 at sample size = 20 to 2.412688 at sample size 60 and further decreased to 1.540203 at sample size 100.

The predictive mean square error of UBR increases from 2.166318 to 2.202126 as the autocorrelation error level increases of 0.2 to 0.5 for sample size is 20 and increases from 2.202126 to 2.563679 as the autocorrelation error level increases of 0.5 to 0.8 for sample size the same sample size, but it was observed that predictive mean square error decreased as the autocorrelation error level increases.

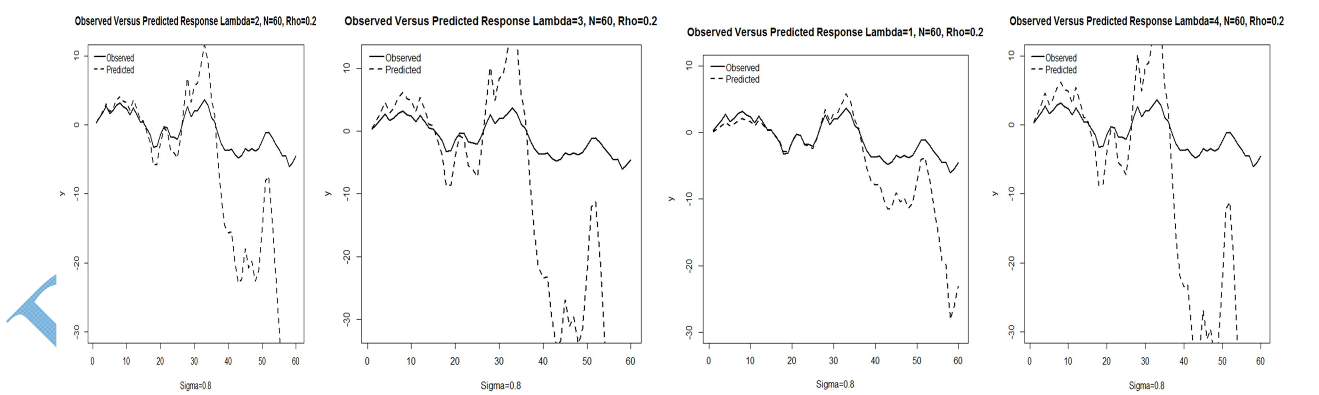


(a) (b) (c) (d)

Figure 1: Plots of the observations* (. . .) *and Estimates* (---) *With Smoothing Parameters Chosen by GCV (a), GML (b), PSM (c), and UBR (d) for n = 20

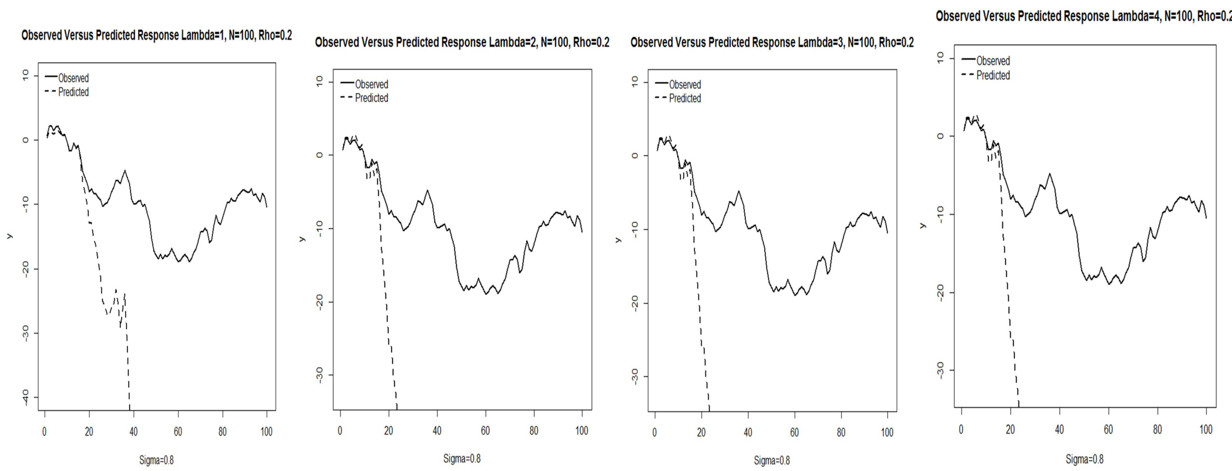
Table 2: The MSE result of the simulated study for GML, GCV, PSM and UBR in the presence of autocorrelation (α) = 0.3, 0.5 and 0.8 for n = 20, 60 and 100 when standard deviation (σ) = 1.0

		PMSE								
		N = 20			N = 60			N = 100		
λ	Smoothing Methods	α = 0.2	α = 0.5	α = 0.8	α = 0.2	α = 0.5	α = 0.8	α = 0.2	α = 0.5	α = 0.8
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	1.2739570	1.659382	1.159813	1.104291	1.454671	1.259721
	UBR	1.259853	2.014616	2.549091	1.221922	1.578077	2.412688	0.715468	0.387835	1.540203



(a) (b) (c) (d)

Figure 2: Plots of the Observations (. . .) and Estimates (---) With Smoothing Parameters Chosen by GCV (a), GML (b), PSM (c),and UBR (d) for n = 60



(a) (b) (c) (d)

Figure 3: Plots of the Observations (. . .) and Estimates (---) With Smoothing Parameters Chosen by GCV (a), GML (b), PSM (c),and UBR (d) for n = 100

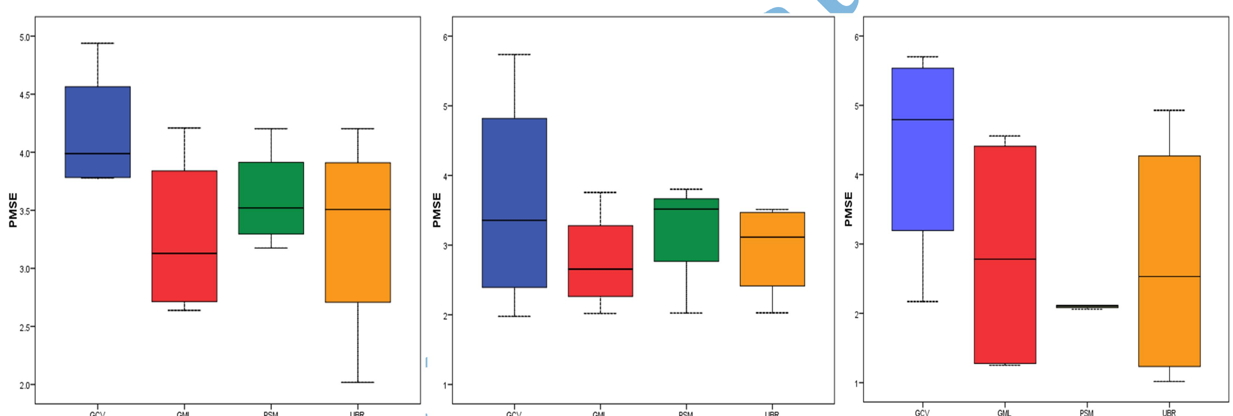


Figure 4: The plot of the GML, GCV, PSM and UBR of the MSE of the simulated study in the presence of autocorrelation when σ = 1, ρ = 0.2 and n = 20

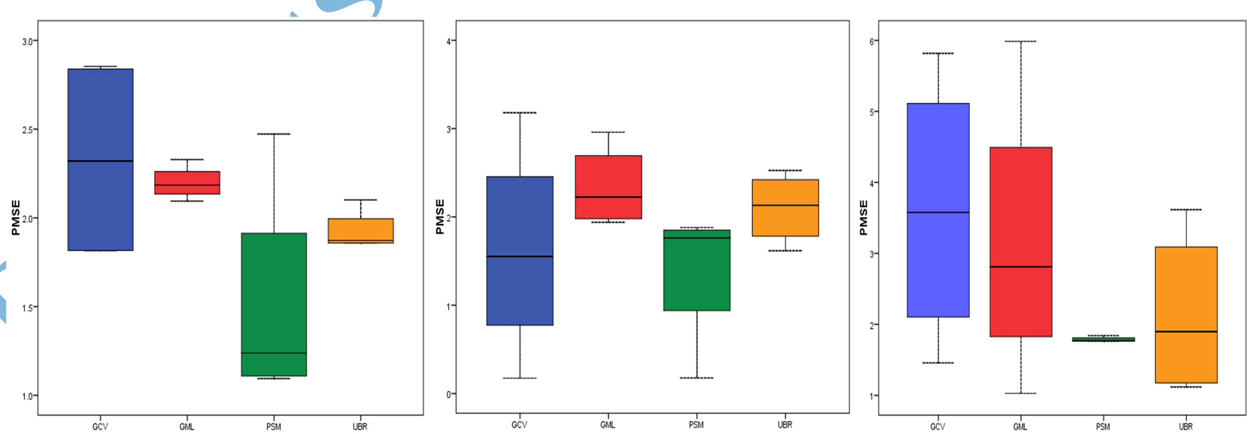


Figure 5: The plot of the GML, GCV, PSM and UBR of the MSE of the simulated study in the presence of autocorrelation when σ = 1, ρ = 0.2 and n = 60

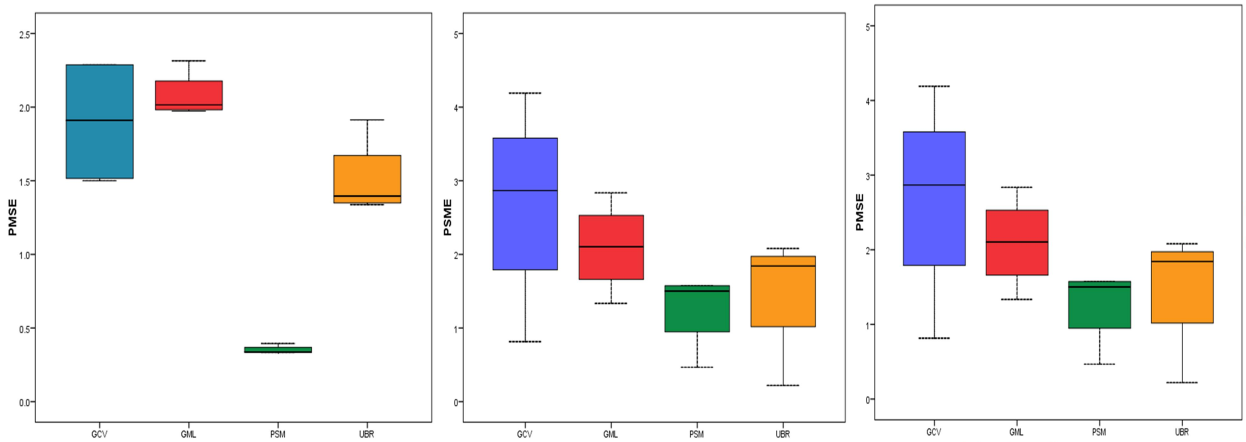


Figure 6: The plot of the GML, GCV, PSM and UBR of the MSE of the simulated study in the presence of autocorrelation when σ = 1, ρ = 0.2 and n = 100

Figures 1 and 5 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 compare with GCV and GML. We conclude that all four methods estimate the smoothing parameters and the functions well but the PSM and UBR provide better estimates than GCV and GML in terms of mean-square error.

The PSM method is more stable when the sample size is small, such as when N = 20 while UBR method performs slightly better when N = 60. In this case there were several replications where GCV and GML providing more estimates of smoothing parameters which lead to under smoothing of the data. This behavior of the GCV method was investigated in Wahba and Wang (1993) and Wang (1998).

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 = 0.2	1.08	1.39	1.47	1.63
α = 0.5	1.89	1.71	1.66	1.48
α = 0.8	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 = 20	2.434	2.179	1.711	2.326
n = 60	2.041	1.900	1.549	1.921
n = 100	1.124	1.047	1.145	0.951
Grand mean	1.867	1.709	1.468	1.732
Ranks	4	2	1	3

V. Discussion Of Result

In this study, we presented Spline smoothing estimation method for time series observations in the presence of auto correlated 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 increase in sample size. The simulation result under the finite sampling properties of PMSE criterion shows that all estimators are consistent and adversely affected by auto correlated error the estimators' ranks as follows, PSM, GML, UBR and GCV. The result suggested that PSM should be preferred when autocorrelation level is mild and high (α = 0.5 – 0.8). If there is low autocorrelation in the observations, (i.e. α = 0.2 – 0.5) the unbiased Risk (UBR) should be considered. It was observed that GCV and GML were mostly affected by the presence of auto correlation and therefore had an asymptotically similar behavioural 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 at all the smoothing parameters.

VI. Conclusion

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 undersmooth relative to the other smoothing method especially for small sample size i.e. n = 20 and 60.(see figure 1 and 2). The result of this experiment with n = 20 and n = 60 is in slight agreement with the monte-carlo results from Barry (1983) and Wahba (1985).

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).

References

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.

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

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

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.

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.*

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

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.

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.

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

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.

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

Morton, R., Kang, E.L. and Henderson, B.L. (2009), Smoothing Spline for trend estimation and prediction in time series. *Environmetrics*, 20(3): 249 – 259.

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

Shen, H. (2008). On Modeling and Forecasting Time Series on Smooth Curve, *Technometric,* 51(3): 227 – 238.

Wahba, G. (1975). Optimal convergence properties of variable knot kernel and orthogonalseries methods for density estimation. *Annals of Statistics*, 3: 15 – 29.

Wahba, G. (1979). Convergence rates of thin plate smoothing splines when the data are Noisy in T. Gasser and M. Rosenblatt (eds), Smoothing techniques for curve estimation, *Springer-Verlag*.

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

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.

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.

Wang, H., Meyer M. C. and Opsomer J.D. (2013). Constrained spline regression in the presence of AR(p) errors. *Journal of Nonparametric Statistics,* 25:809 – 827.

Yuedong, W. and Brown M.B. (1996). A Flexible Model for Human Circadian Rhytmns, *Biometric,* 52: 588 – 596

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

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