International Journal of Data Science
Vol. 6, No. 1, June 2025, pp. 54-69

Parameters and Knot Points Estimation for Spline Methods Applied in Time Series Data

Amjed Mohammed Sadek [a,1,*], Asmaa Salah Aladdin Sulaiman [a,2]

[a] Department of Mathematics, Faculty of Basic Education, Telafer University, Mosul, Telafer, Iraq
[1] amjedmamed@gmail.com, [2]asmaa.s.sulaiman@uotelafer.edu.iq
* corresponding author

ARTICLE INFO

Article history
Received February 23, 2025
Revised April 17, 2025
Accepted June 26, 2025

Keywords
Knot points
Parameter Estimation
B-Spline
Smoothing Spline
Penalized Spline

ABSTRACT

The purpose of this study is to investigate and compare several nonparametric regression approaches, including penalized spline methods, B-splines, and smoothing splines. Applying these techniques to simulated and real datasets, such as Iraqi oil export data, focuses on parameter estimation and identifying optimal knot points for predicting periodic and nonlinear trends. The knot points are selected using generalized cross-validation (GCV) to ensure an accurate fit to the data. For time-series data with nonlinearities and periodic patterns in the response variable, this research employs nonparametric regression with sequential explanatory variables. We research simulated data that exhibit periodic patterns similar to economic cycles, as well as nonlinear data that employs complex equations to model interactions among variables. Simulations were conducted across a range of standard deviations and sample sizes. The efficiency of parameter estimation in these synthetic datasets was quantified using the mean absolute average error (MAME). For the empirical application, the parameters of the nonparametric regression models were estimated using monthly Iraqi oil export data, with the MAME employed as the evaluation metric. The effectiveness of these techniques is further evaluated in forecasting future values by calculating the mean absolute percentage error (MAPE). Among the approaches, the penalized spline consistently achieves the lowest average mean squared error across all standard deviation levels and sample sizes in the simulated data, while also demonstrating robust forecasting performance. In contrast, the smoothing spline outperforms the other methods in terms of parameter estimation accuracy.

1. Introduction

A statistical method known as regression analysis is used to identify the relationship between an explanatory variable and a response variable. This method enables predictions of the dependent variable from the independent variable. The assumptions underlying regression analysis are often relevant only to specific variables in particular contexts. When a parametric model is inaccurate, it can lead to misinterpretations that can significantly misguide decision-making. Moreover, there are instances in which an appropriate parametric model does not exist [1], [2]. To overcome these limitations, employing nonparametric regression techniques becomes a compelling solution. These methods effectively estimate parameters in cases where data exhibit nonlinear relationships. The technique of constructing a smoothing curve from available data is known as smoothing. This approach is an excellent alternative when conventional parametric models fall short or when the assumptions underlying regression analysis are not met, ensuring reliable results in challenging scenarios.[3].

Nonparametric regression is a valuable technique across many research and data analysis fields due to its advantages. The model can capture complex, nonlinear relationships that parametric approaches would miss due to its flexibility—observed that nonlinear models facilitated accurate fitting of generalized additive models using nonparametric regression. To handle symmetric error distributions, robust nonparametric regression techniques were developed. These approaches are especially helpful in situations where the relationships in the data are poorly characterized or change across sections. This method permits precise forecasts that, without imposing strict limitations, capture complex patterns and variations in the data.[4]

In nonparametric regression, a response variable and one or more explanatory variables are typically used. Instead of estimating regression coefficients, it mainly focuses on assessing a smoothing function that gives a more accurate representation of the data. This smoothing function helps identify the underlying trend between one or more explanatory variables and the response variable. The scatterplot smoothing approach, which is used when there is only one explanatory variable, improves the scatterplot's visual clarity and facilitates the identification of patterns in the relationship between the descriptive and response variables.[5] Nonparametric regression is employed to determine the relationship between variables without assuming a specific functional form, and the estimated covariates are then incorporated into models in which multiple equations are solved concurrently.[6]

In the context of nonparametric regression, several methods are used to estimate nonparametric regression models, including local polynomial regression, smoothing splines, regression splines, kernel smoothing, and penalized splines.[7] In addition, nonparametric regression models were specifically modified for use in time-series analysis, enabling the depiction of nonlinear interactions. Furthermore, time series analysis has adopted nonparametric regression models, which permit the modelling of possible nonlinear relationships. For assessing smooth structural changes in time series models, Chen and Hong [8] suggested nonparametric estimate methods.

The principal aim of this study is to conduct a systematic evaluation and comparative analysis of several nonparametric regression techniques—namely, smoothing splines, B-splines, and penalized splines—within the context of time-series data characterized by cyclical patterns and nonlinear dynamics. By applying these methods, the study seeks to improve the accuracy of both forecasting and parameter estimation, particularly in settings where conventional parametric models are insufficient to capture the underlying complexity of the data. This study's significant contribution is that it applies these nonparametric approaches to real-world and simulated datasets, focusing on Iraq's monthly oil exports. This method demonstrates the importance of nonparametric regression in addressing real-world challenges in energy forecasting. To illuminate the best approaches for various types of data, the study compares and contrasts these methods using performance metrics, including mean absolute average error (MAAE) and mean absolute percentage error (MAPE).In addition to providing a thorough analysis that examines how the selection of knots and smoothing parameters may affect prediction accuracy, this study advances the discipline of nonparametric regression by demonstrating the adaptability of these models in capturing intricate nonlinear relationships. These approaches are essential across sectors such as engineering, economics, and environmental research. The results have important implications for future work in time series analysis, particularly in settings where data do not conform to parametric assumptions.

The structure of this paper is described as follows: Section 2 presents the related work. Section 3 presents the methods and procedures used in this study: regression spline, B-spline, smoothing spline, and penalized spline, as well as the estimation of smoothing parameters. Section 4 presents the analysis of simulated data and reports the results. The final section presents a reasoned conclusion based on the findings obtained from the simulation study and the real-data application.

2. Materials And Methods

2.1 Related Work

Nonparametric regression estimates typically exhibit visible divergence from their parametric counterparts due to fundamental differences in their underlying assumptions. Unlike parametric methods that impose strong a priori assumptions regarding the functional form of the relationship between variables, nonparametric regression employs flexible models. This inherent flexibility enables nonparametric estimates to effectively capture intricate patterns and local variations present within the data [9]. Consequently, nonparametric approaches demonstrate greater capacity to adapt to the inherent structure of the data, potentially yielding more accurate and reliable predictions than parametric methods constrained by prespecified functional forms. [5]

In contrast, parametric approaches often presuppose a specific distribution for the data. EL-Morshedy et al.[10] highlighted the significance of parameter estimation in regression models by introducing the discrete Burr–Hatke distribution. Nevertheless, nonparametric regression is appropriate for analyzing data with uncertain or nonstandard distributions because it does not require such assumptions. Gal et al. [11] suggested a technique for estimating parameters in nonparametric regression using residuals based on symmetric and nonsymmetric distributions. In addition, many parametric approaches are less effective than nonparametric regression in handling outliers and influential data. The model is more resistant to extreme values because it focuses on local data points, so outliers have less impact on the overall fit. To reduce the possibility of model misspecification, Cizek and Sadikoglu [12] studied nonparametric regression and found that it needs just modest identification assumptions.

Nonparametric regression is a well-known smoothing approach that has been used recently in several different fields of study. Demir and Toktamis [13] investigated the adaptive kernel estimator for long-tailed and multimodal distributions and the nonparametric kernel estimator with fixed bandwidth. Shang and Cheng [14] addressed essential issues in the use of distribution algorithms by developing a smoothing spline approach and computational trade-offs. To predict the yield curve, Feng and Qian [15] presented a natural cubic spline model that is dynamic and uses a two-stage process. Among the many uses of B-spline functions that Than and Tjahjowidodo [16] brought to light were their implementations in CAD, numerical control systems, and computer graphics. Xiao [17] investigated penalized splines extensively, including B-splines and an integrated squared derivative penalty, in the context of large-sample asymptotic theory.

2.2 Regression Spline

The estimate of the relationship between the function of explanatory variables (m (x_i)) and response variables (y_i) is the procedure that is involved in nonparametric regression. In this paper, we will offer an overview of some of the more common techniques that are used in nonparametric regression models:

$$y_i = m(z_i) + \varepsilon_i, i = 1,2,\ldots,n \tag{1}$$

where ε_i represents the error for each observation.

The smoothing technique is the basis of nonparametric regression, yielding a smoother. It is a technique for predicting the function of predictor variables, as can be used to improve the appearance of trends in the plot, with the support of a smoother. According to Eubank [18], who first proposed the regression spline concept, a set of locations defines neighborhoods:

$$\xi_0, \xi_1, \xi_2, \ldots\ldots, \xi_m, \xi_{m+1} \tag{2}$$

In the range of an interval $[a, b]$, where $a = \xi_0 < \xi_1 < \cdots < \xi_m < \xi_{m+1} < b$. The term for these specific locations is denoted as knots, and $\xi_r, r = 1,2,..,m$ are called interior knots. Therefore, A regression spline may be formed with the m-th degree truncated power basis with K knots $\xi_1, \xi_2, \ldots\ldots, \xi_m$:

$$1, z_i, \cdots, z_i^m, (z_i - \xi_1)_+^m, \ldots., (z_i - \xi_m)_+^m, p = M + m + 1 \tag{3}$$

where u_+^m refer to the m-th power of the positive part of u , where $u_+ = max(0, u)$.The first $(m + 1)$ basis functions of the truncated power basis (3) are polynomials of degree up to m, and the others are all the truncated power functions of degree m. Therefore, A regression spline can be described as

$$m(x_i) \sum_{r=0}^{p} \varphi_r z_i^r + \sum_{j=1}^{p} \varphi_{k+j} (z_i - \xi)_+^p \tag{4}$$

where $\varphi_0, \varphi_1, \ldots, \varphi_{k+K}$ is the unknown regression coefficient that need to be determined with an appropriate loss minimization method [19], [20]

2.3 B-Spline Regression Method

The spline model is a piecewise polynomial with piecewise-defined characteristics at intervals k, defined by knot points. Points that represent changes in the data within subintervals are called knot points. When the spline order is high, multiple knots or knots that are too close together will generate a matrix that is practically singular in computation, which means that normal equations cannot be solved. This is the most significant limitation of the spline method. The problem with B-splines is that they cannot be assessed directly since their basis can only be defined recursively[21] .Therefore, The B-spline basis function may be defined recursively as follows:

$$B_s^u(z_i) = \begin{cases} 1 \ , \xi_s < z < \xi_{s+1} \\ 0 \quad \text{otherwise} \end{cases} \tag{5}$$

where $B_s^u(z_i)$ is the sth of the B-spline basis function of the order u for the knot points sequence ξ.[22] For the piecewise polynomial function, Liu et al. [23] computed B-splines of any degree using an algorithm. Evaluating the function of B-splines at the uth degree from the (u − 1) th degree can be described as

$$B_s^u(z_i) = \frac{z_i - \xi_s}{\xi_{s+u-1} - \xi_s} B_s^{u-1} + \frac{\xi_{s+u} - z_i}{\xi_{s+u} - \xi s + 1} B_{s+1}^{u-1} \tag{6}$$

Where the basis of order u with knot points $\{B_s^u | s = 1,2, \ldots, K + u + 1\}$.A B-spline representation of the nonparametric regression model can be described as follows:

$$y_i = \sum_{s=1}^{k} B_s^m(z_i) \ \gamma_s + \varepsilon_i, i = 1,2, \ldots, n. \tag{7}$$

Hence, the following is the fitting of the function of B-splines assessed at the knots ξ_s , where s = 1..., K:

$$\widehat{m}(z_i) = \sum_{s=1}^{k} B_s^m(z_i) \gamma_s \tag{8}$$

Moreover, the criteria for penalized least squares are as follows:

$$PLS = (y - B\gamma)^T (y - B\gamma) + \lambda \gamma \Omega_k \gamma$$

where$\{B\}_{is} = B_s^m(z_i), \{\Omega_k\}_{is} = \int B_i''(z_i) B_s''(z_i) dx, \gamma = (\gamma_1, \ldots, \gamma)^T$is the coefficient regression vector of the B-spline. Therefore, the solution of the function of the B-splines, denoted as $\widehat{m}_\lambda$, to the problem of minimizing the PLS involves the following:

$$\widehat{m}_\lambda = (B^T B + \lambda \Omega_K)^{-1} B^T y \tag{9}$$

2.4 Smoothing Spline Regression Method

The smoothing spline method's approximated process involves fitting a function of predictor variables $(m(z_i))$ by minimizing the penalized least squares criterion, which is expressed by

$$PLS = RSS + \lambda \int_a^b \{m''(z_i)\}^2 \, dx \tag{10}$$

where the first part RSS = $\sum_{i=1}^{n}\{y_i - m(z_i\}^2$ is the residual of the square, and the second part $\lambda\int_a^b\{m''(z_i)\}^2\,dx$ is the roughness penalty in the interval [a,b], This is a curve metric known as the smoothing parameter (λ).Therefore, the second part (roughness penalty) can be written in matrix form

$$\lambda\int_a^b\{m''(z_i)\}^2\,dx = m^T H\ m \tag{11}$$

where $m = (m_1, m_2, \ldots, m_k)^T, m_r = m(\xi_r), r = 1,2,..,k$. [24] generally, k refers to the number of knots, and $\xi_1, \ldots, \xi_k$ are all the knot points of the smoothing spline such that may be arranged in ascending order as

$$-\infty \le a < \xi_1 < \xi_2 < \cdots < \xi_k < b \le \infty$$

Therefore, the matrix H can be written as follows

$$H = C\,D^{-1}C^T \tag{12}$$

where C is a matrix as a $p \times (p-2)$matrix, and D is a matrix as a $(p-2)\times(p-2)$ Therefore, from (11) and (12), the penalized least square criterion can be described as

$$\|y - wm\|^2 + \lambda m^T H m \tag{13}$$

where $y = (y_1, y_2, \ldots, y_n)^T$ are the response variable, and $W = (w_{ir})$ is an $n \times p$ incidence matrix with w_{ir}=1 if $z_i = \xi_r$and otherwise, and $\|y - mw\|^2 = \sum_{i=1}^{n}\{y_i - f(x_i)\}^2$. [25] Consequently, the smoothing spline function $(\hat{m}_\lambda)$ evaluated at knots ξ_r $r = 1,2,\ldots,k$ may be expressed explicitly as follows:

$$\hat{m}_\lambda = (WW^T + \lambda H)^{-1}W^T y. \tag{14}$$

2.5 Penalized Spline Regression Method

The smoothing spline approach requires computing an integral that quantifies the function's roughness, whereas the penalized spline method addresses this issue by employing a truncated power basis, as shown in Equation (3).

Let $\delta_r(i) = (\delta_1(i), \cdots, \delta_r(i))^T$ represent the degree k truncated power basis with K knots $\xi_1, \xi_2, \cdots, \xi_K$. subsequently, we may articulate $m(z_i)$ in equation (1) as $\delta(i)_r\theta$,where $\theta = [\theta_0, \theta_1, \cdots, \theta_{k+k}]^T$ is the vector that represents the corresponding coefficient. [26] Let H be a $p \times p$ diagonal matrix, where the first k + 1 diagonal elements are set to zero and the remaining diagonal entries are set to one. Therefore, the matrix H can be given as

$$H = \begin{bmatrix} 0 & 0 \\ 0 & I_r \end{bmatrix}$$

Moreover, the penalized smoothing spline is denoted as $\hat{m}_\lambda = \delta_r(i)^T\hat{\theta}$, where the value of $\hat{\theta}$ is the PLS criteria that minimizes the following:

Penalized least squares (PLS) = $(y - W\theta)^T(y - W\theta) + \lambda\theta H\theta$

Where $W = (\delta_r(z_1), \ldots, \delta_r(z_n))^T$,and $\theta H\theta = \sum_{r=1}^{k}\theta_{k+r}^2$

The penalized spline smoother is described as

$$\hat{m}_\lambda = W(W^T W + \lambda H)^{-1}W^T y \tag{15}$$

2.6 Estimation of Smoothing Parameters

Specifically, the generalized cross-validation (GCV) that was proposed by Wahba [27] and Craven [28] and Wahba is the primary focus of the smoothing parameter selection that is being discussed in this study. The generalized cross-validation (GCV), which optimizes a smoothness selection criterion, is the optimal value for the smoothing parameter. Minimizing the GCV function it facilitates the selection of smoothing parameters. The function employs the following formula:

$$GCV\ (\lambda) = \sum_{t=1}^{n} \left(\frac{y_t - \hat{y}_t}{n - tr(H)} \right) \quad (16)$$

where H of smoothing spline is $I + \lambda K$, The B-spline is $(B^T B + \lambda \Omega_K)^{-1} B^T$, and penalized spline is $F(F^T F + \lambda^3 D)^{-1} F^T$

3. Results and discussion

3.1 Simulation Study

This section presents a Monte Carlo simulation conducted in R to estimate the response variable and evaluate the performance of Spline methods, including B-spline, smoothing spline, and penalized spline. Accordingly, the explanatory response variable in the time-series data exhibits periodic patterns and nonlinear features. Therefore, the periodic patterns observed in time series data are simulated by using the following function:

$$z_t = \sqrt{m_t}\, cos\big(2\pi\big[1 + \sqrt{m_t}\big]\big) + \varepsilon_t\ t = 1,2,3,..,n \quad (17)$$

where ε_t denotes an error term that follows a normal distribution with a mean of zero and standard deviations 1,3, and 5, as demonstrated in Figures1-3.

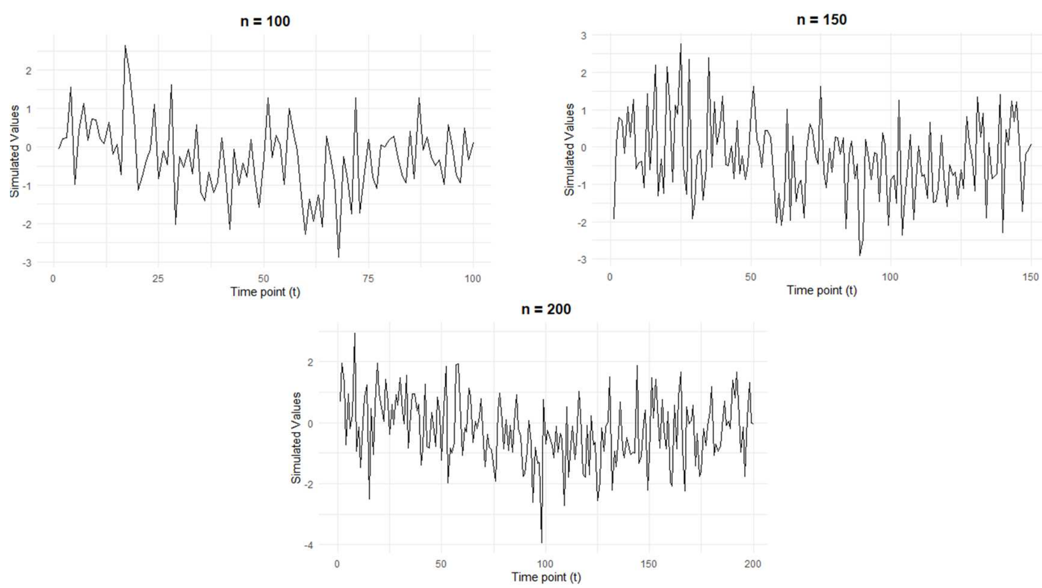


Figure 1. The Plot of Periodic Patterns of Time Series for Different Sample Sizes With $\sigma = 1$

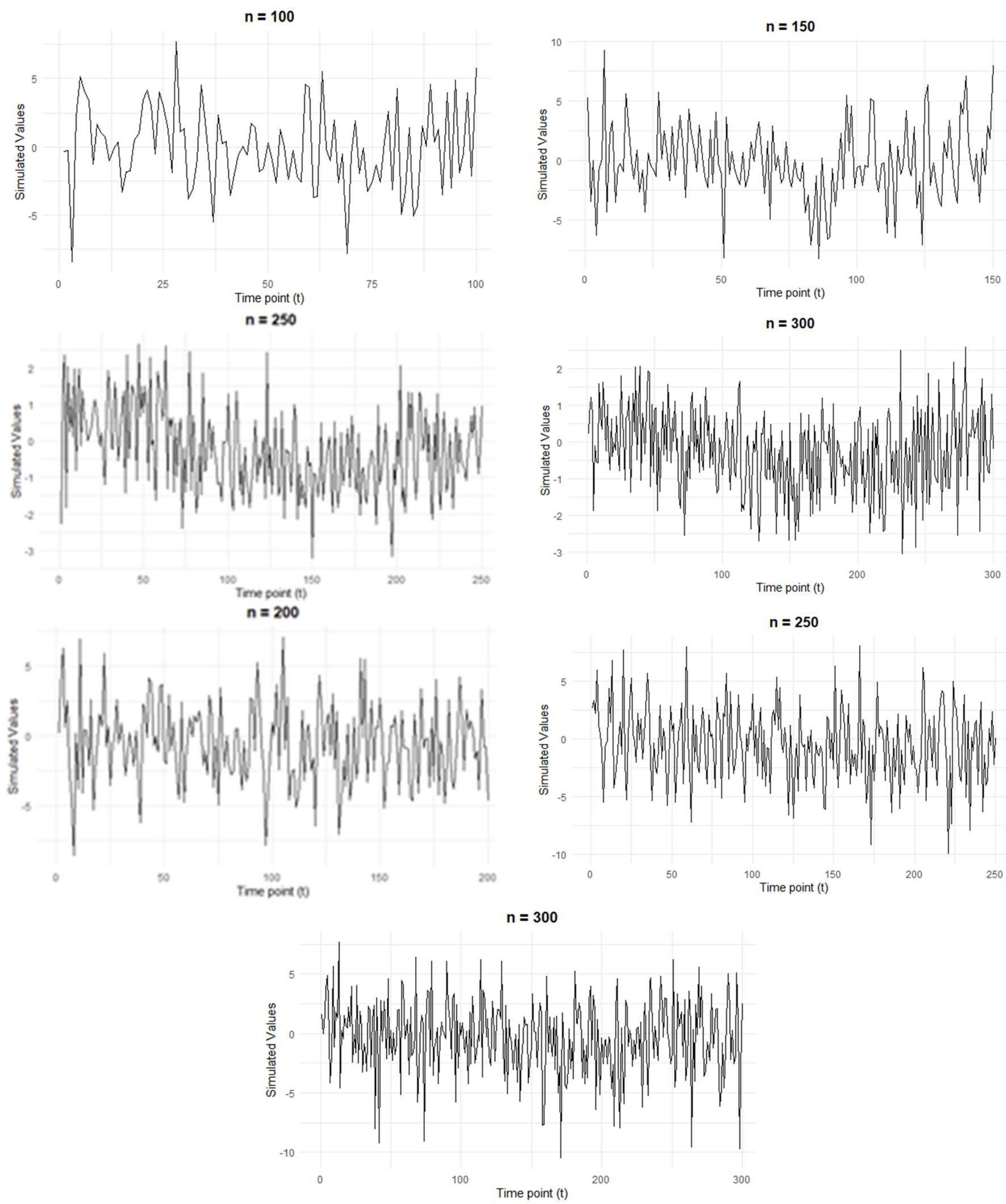
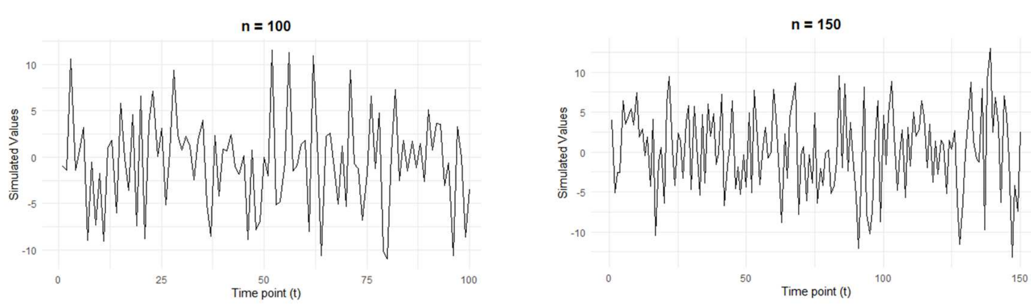


Figure 2. The Plot of Periodic Patterns of Time Series for Different Sample Sizes With $\sigma = 3$



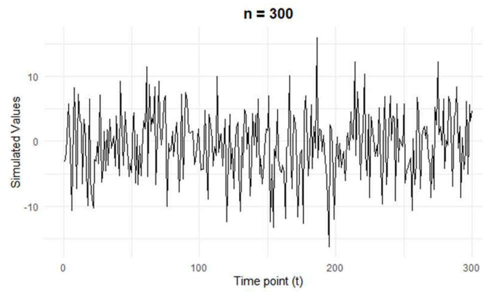


Figure 3. The Plot of Periodic Patterns of Time Series for Different Sample Sizes With $\sigma = 5$

Furthermore, the response variables with utilized of nonlinear shapes are simulated by using the following function:

$$z_t = 2\sqrt{m_t}\, sin\left(2\pi\left[\frac{1+20}{m_t+20}\right]\right) + \varepsilon_t, t = 1,2,\ldots.,n \tag{18}$$

where ε_t denotes an error term that follows a normal distribution with a mean of zero and standard deviations 1,3, and 5, as demonstrated in Figures4-6.

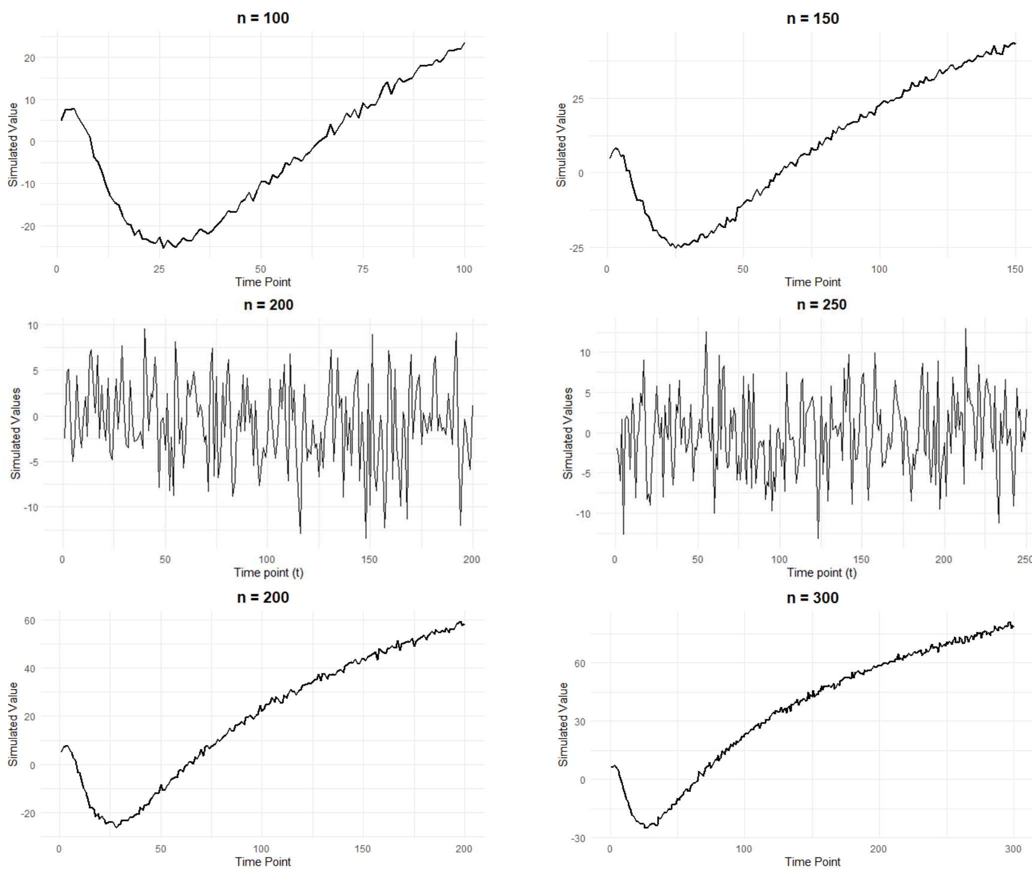


Figure 4. The Plot of Nonlinear Shapes of Time Series for Different Sample Sizes With $\sigma = 1$

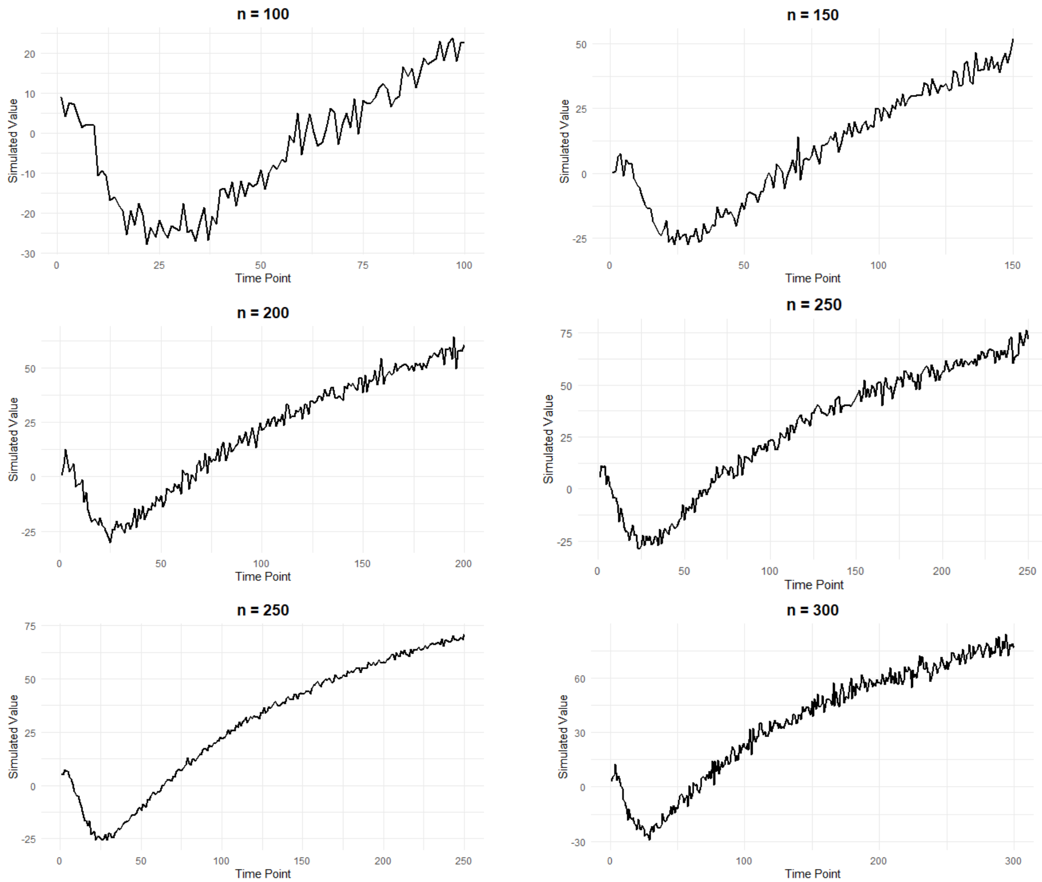
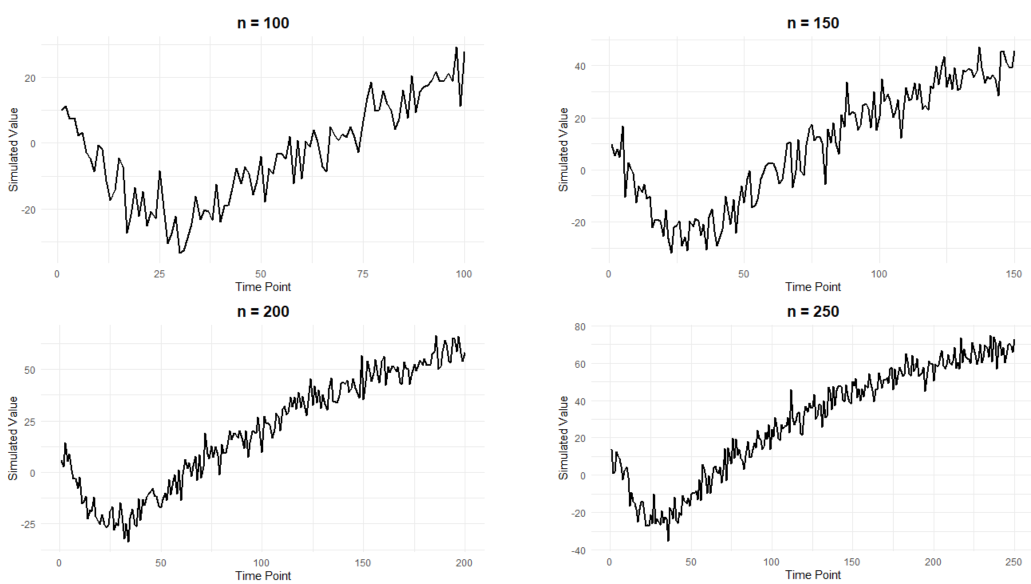


Figure 5. The Plot of Nonlinear Shapes of Time Series for Different Sample Sizes With $\sigma = 3$



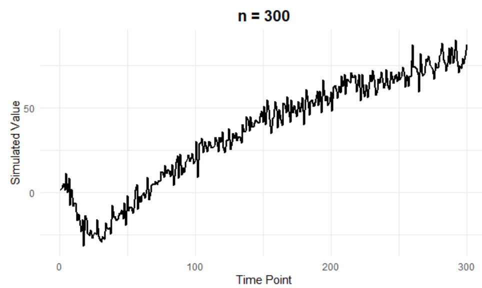


Figure 6. The Plot of Nonlinear Shapes of Time Series for Different Sample Sizes With $\sigma = 5$

3.2 Simulation Design

This simulations study utilized five different sample sizes: n=100,150,200,250, and 300 with three different standard deviations values: $\sigma = 1$,3,and 5. Furthermore, the data generated were replicated 1000 times for each sample sizes to determine the best spline nonparametric regression method was applied to predict time series data characterized by periodic patterns and nonlinear shapes in the response variable, with explanatory variable considered as sequential data.

Therefore, the generalized cross-validation (GCV) is used to choose the optimal smoothing parameter estimation as well as the cross-validation method (CV), while the number of knots is controlled and specified using cross-validation procedures, which ensure the curve suitably fits the data points.

3.3 Simulation Result

Tables 1, 2, and 3 present the values of the Mean Average Absolute Error (MAAE) and the number of knot points for spline methods that were applied in the periodic patterns and nonlinear time series data for all sample sizes 100, 150, 200, 250, and 300 under different values of standard deviation of error as 1, 3, and 5.

Table 1. The Values of Mean Average Absolute Error (MAAE) and the Mean of the Knot Point for Different Sample Sizes with $\sigma = 1$

n	Methods	Nonlinear		Periodic patterns	
		MAAE	**No. of knots**	**MAAE**	**No. of knots**
	B-Spline	0.623547	99	0.277846	99
100	Penalized spline	**0.256635**	100	**0.231000**	100
	Smoothing Spline	0.325871	65	0.463913	65
	B-Spline	0.836252	140	0.629568	140
150	Penalized spline	**0.478959**	145	**0.418975**	145
	Smoothing Spline	0.562514	90	0.547623	90
	B-Spline	0.992571	193	0.780542	193
200	Penalized spline	0.505439	197	**0.671000**	197
	Smoothing Spline	0.987871	120	0.717160	121
	B-Spline	1.094318	240	1.186803	240
250	Penalized spline	**0.820787**	235	**0.632654**	235
	Smoothing Spline	0.976725	180	1.212508	180
	B-Spline	1.720780	283	1.182154	283
300	Penalized spline	**0.948713**	291	**0.876321**	291
	Smoothing Spline	1.070494	225	0.996325	225

Table 2. The Values of Mean Average Absolute Error (Maae) and the Mean of the Knot Points for Different Sample Sizes With $\sigma = 3$

n	Methods	Nonlinear		Periodic patterns	
		MAAE	**No. of knots**	**MAAE**	**No. of knots**
100	B-Spline	0.743992	99	0.554278	99
	Penalized spline	**0.474869**	100	**0.399641**	100
	Smoothing Spline	0.599272	65	0.722549	65
150	B-Spline	0.508108	140	0.588418	140
	Penalized spline	**0.456979**	145	**0.433687**	145
	Smoothing Spline	0.642273	90	0.6774215	90
200	B-Spline	0.556112	193	1.02537778	193
	Penalized spline	**0.495113**	197	**0.744865**	197
	Smoothing Spline	0.936218	120	1.188651	121
250	B-Spline	1.093273	240	1.100456	240
	Penalized spline	**0.507539**	235	**0.782214**	235
	Smoothing Spline	0.898173	180	0.978214	180
300	B-Spline	1.451603	283	1.187922	283
	Penalized spline	**0.675219**	291	**0.822169**	291
	Smoothing Spline	0.906713	225	0.922314	225

Table 3. The Values of Mean Average Absolute Error (MAAE) and the Mean of the Knot Points for Different Sample Sizes with $\sigma = 5$

n	Methods	Nonlinear		Periodic patterns	
		MAAE	**No. of knots**	**MAAE**	**No. of knots**
100	B-Spline	0.854213	99	0.622154	99
	Penalized spline	**0.317659**	100	**0.6188974**	100
	Smoothing Spline	0.862231	65	0.9123541	65
150	B-Spline	1.022845	140	0.922514	140
	Penalized spline	**0.725146**	145	**0.811236**	145
	Smoothing Spline	0.933126	90	0.988745	90
200	B-Spline	0.900326	193	0.778965	193
	Penalized spline	**0.890148**	197	**0.633145**	197
	Smoothing Spline	0.978641	120	0.855263	121
250	B-Spline	0.455623	240	1.200354	240
	Penalized spline	**0.811879**	235	**0.844567**	235
	Smoothing Spline	1.003265	180	1.188976	180
300	B-Spline	1.233654	283	0.665532	283
	Penalized spline	**0.974561**	291	**0.447158**	291
	Smoothing Spline	1.122302	225	0.881135	225

As shown in Tables 1, 2, and 3, the mean absolute error (MAAE) for the periodic data differs slightly from that for the nonlinear data. Therefore, the mean average absolute error (MAAE) for the smoothing spline method is higher than that of other methods, and there are fewer knots utilized. Furthermore, the increase in standard deviation corresponds to a rise in the mean absolute error (MAE), indicating an effect on the model's fitness. Despite the increase in sample size, parameter estimates remained consistent, indicating robustness to variations in sample size. Therefore, the penalized spline method consistently outperformed the other nonparametric regression models.

3.4 Real Data Application

Since the beginning, Iraqi oil exports have significantly contributed to the country's economy. This is because oil exports contribute to energy security, primary energy production, industrial usage, human development, and other areas of economic growth. The Iraqi economy is extremely dependent on oil exports. This study used a dataset of Iraq's oil exports, comprising 228 monthly records from January 2005 to December 2024. The dataset shows a nonlinear trend and periodic patterns with a component of seasons, as Figure 7 illustrates.

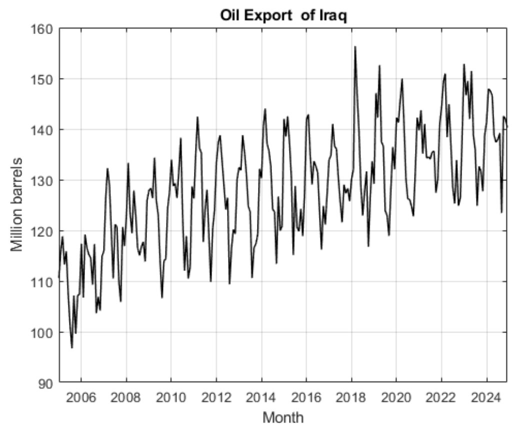


Figure 7. The Time Series Plot of Oil Export of Iraq

The real data analysis employed three nonparametric approaches to determine the smoothing function for Iraq's oil exports. These were the smoothing spline, the B-spline, and the penalized spline. A sequence of 228 months refers to the explanatory variable, while the monthly oil export volume (Million barrels) serves as the response variable. The Mean Absolute Error (MAE) is a metric used to evaluate a model's precision, defined as the average of absolute differences between expected and actual values. The accuracy of the prediction is evaluated in percentage terms by the Mean Absolute Percentage Error (MAPE). MAAE and MAPE are used to assess forecasting and estimate precision. Therefore, the following are the equations used to calculate MAAE and MAPE:

$$MAAE = \frac{1}{228}\sum_{i=1}^{228}|y_i - \hat{y}_i|, i = 1,2,\ldots.,228 \quad (19)$$

$$MAPE = \frac{1}{228}\sum_{t=1}^{228}\left|\frac{(y_i-\hat{y}_i)}{y_i}\right| \times 100, t = 1,2,3,\ldots,228 \quad (20)$$

Moreover, the Mean Average Absolute Error (MAAE) and knot points are approximated from the spline methods as: smoothing spline, B-spline, and penalized spline as shown in Figure 8.

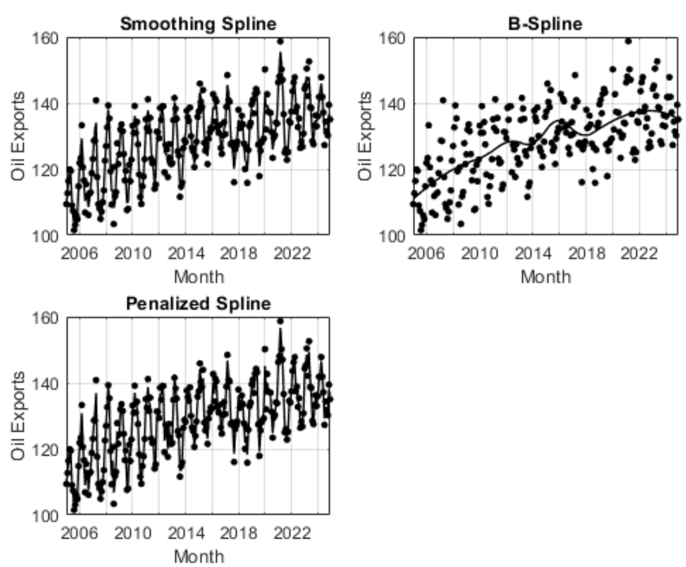


Figure 8. The Fitted Nonparametric Regression Model of Iraq's Oil Export

The fitted nonparametric regression models of all methods, the smoothing spline, B-splines, and penalized spline, make it hard to pick the best method, as shown in the figure above. The outperforming method is then investigated using mean average absolute error (MAAE). Therefore, Table 4 presents the MAAE values and the number of knot points.

Table 4. The MAAE Values and Knot Points for Estimating the Nonparametric Regression Spline

B-spline		Penalized spline		Smoothing spline	
Knots	MAAE	Knots	MAAE	Knots	MAAE
205	16893.221	220	15487.221	185	19845.554

The results in the table above indicate that the best knot points and the mean absolute error (MAE) for nonparametric smoothing methods are achieved by B-spline, smoothing spline, and penalized spline. The penalized spline method yields the most accurate estimates for this dataset, as it has the lowest mean absolute error (MAE).

Furthermore, the estimate is followed by the use of these nonparametric regression models to forecast values for the next 12 months. The MAPE is then calculated to evaluate accuracy over the specified period. All methods are shown in Table 5, which includes the actual data, expected values, and the MAPE.

Table 5. The Amount of Oil Export Per Month, Forecasting Values for 12 Months, and MAPE.

Months	Oil export per month	B-spline	Smoothing spline	Penalized spline
January	3365123.78	3164123.18	3154122.20	3465140.08
February	3275148.65	3163120.05	3155110.85	3385048.75
March	3250698.22	3150597.32	3140580.25	3360799.22
April	3350862.98	3150761.99	3147675.88	3460963.18
May	3150963.11	3050883.23	3040873.20	3260973.10
June	3450899.88	3440889.45	3439779.95	3480998.95
July	3516981.85	3514861.80	3513850.70	3618991.99
August	3475187.66	3455170.55	3450175.99	3495199.86
September	3514189.47	3513186.35	3512155.05	3519396.97
October	3315264.87	3313340.60	3312541.75	3418274.99
November	3400145.96	3400125.75	3400120.50	3500199.86
December	3375487.33	3365477.20	3335455.25	3498697.93
January	3250142.27	3240130.15	3241125.23	3390182.87
	MAPE	11.4897	9.8865	5.7996

Based on the table above, the most suitable method for estimating the actual data is penalized spline nonparametric regression. It outperformed the other techniques in predicting future values, achieving the lowest mean absolute percentage error (MAPE) of 5.7996. This indicates that the penalized spline method achieves high accuracy in future predictions. Additionally, the B-spline method outperformed the smoothing spline in prediction accuracy, with an MAPE of 9.8865.

Therefore, Figure 9 compares three non-parametric regression methods—B-splines, smoothing splines, and penalized splines—for modelling Iraq's oil exports over 12 months. Both B-splines and smoothing splines closely follow the data points, whereas the penalized spline also performs well but yields a smoother curve. Notably, the smoothing spline exhibits greater variation and deviates from the other methods, especially around months 10 and 11. Overall, B-splines and penalized splines demonstrate the best fit for accurately forecasting oil exports.

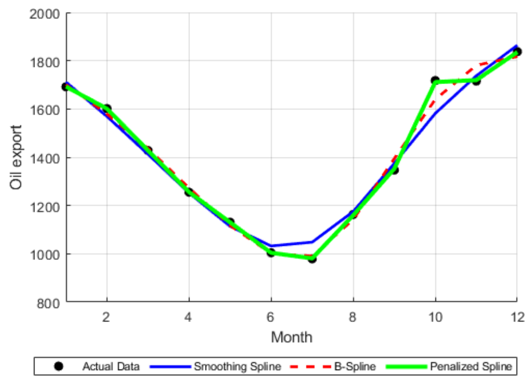


Figure 9. Plot of Actual Data and Predictions Over 12 Future Months

As shown in Table 6, the three non-parametric regression techniques require more knots to predict future data than to fit the actual datasets. Nevertheless, finding the best knots may not always be a significant undertaking, and increasing the number of knots does not necessarily indicate the best method. Based on this study, smoothing splines and B-splines use the same knot approach as [29]; however, penalized splines are better at predicting future values. The relationship between the explanatory variables and the response variables may vary at particular points in the space of the explanatory variables, referred to as knots. They are frequently employed in spline-based nonparametric regression methods, such as cubic splines and piecewise linear regression, thereby enhancing model flexibility and accuracy.

Table 6. The Mean Average Absolute Error (MAAE), Mean Absolute Percentage Error (MAPE), and The Number of Knot Points for Iraq's Oil Export

Knot	B-spline		Smoothing spline		Penalized spline	
	MAAE	MAPE	MAAE	MAPE	MAAE	MAPE
50	75,8545.1	35.986	95,4658.1	30.963	60,8865.2	30.554
100	60,3567.4	33.265	90,4625.7	27.125	53,4469.5	26.145
150	57,1548.6	33.154	88,9875.2	22.189	47,4458.1	21.112
200	55,5241.9	32.758	88,6532.1	17.332	41,8874.2	15.789

4. Conclusion

This study is significant because it compares popular nonparametric regression methods on simulated and real-world data, including smoothing splines, B-splines, and penalized splines. Standard deviations and sample sizes are used to simulate periodic patterns and nonlinear forms. In addition, using a real dataset, such as Iraq's oil exports, yielded fitted model results that were similar to those derived from the simulated data. As noted, penalized splines perform well for predicting future values. Although these are advantages, nonparametric regression also entails difficulties, including the risk of overfitting, the need for larger sample sizes, and greater analytical complexity. Future studies should investigate other nonparametric regression methods, such as kernel smoothing or local polynomial regression, and increase the number of knots to improve model accuracy. It would be possible to test these methods on datasets of varying complexity. To conduct a comprehensive assessment of forecasting accuracy, it will also be necessary to consider the computational efficiency of larger datasets and to use alternative error metrics, such as root mean square error (RMSE) or mean absolute error (MAE).

References

[1] A. M. Sadek and L. M. Ali, "Developing a mixed nonparametric regression modeling (simulation study)," *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 21, pp. 6988–7000, 2023.

[2] H. M. Khalaf and A. S. A. Sulaiman, "New method for finding the weight distribution and spectrum by testing the optimal solution of the points in PG(3,2)," *J. Theor. Appl. Inf. Technol.*, vol. 102, no. 9, pp. 3725–3729, 2024.

[3] M. Salibian-Barrera, "Robust nonparametric regression: Review and practical considerations," *Econom. Stat.*, vol. 31, pp. 363–379, Jan. 2024, doi: 10.1016/j.ecosta.2023.04.004.

[4] S. Amodio, M. Aria, and A. D'Ambrosio, "On concavity in nonlinear and nonparametric regression models," *Statistica*, vol. 74, no. 1, pp. 85–98, 2014.

[5] A. M. Sadek and L. A. Mohammed, "Evaluation of the performance of kernel non-parametric regression and ordinary least squares regression," *JOIV: Int. J. Inform. Vis.*, vol. 8, no. 3, pp. 1352–1359, Sep. 2024, doi:10.62527/joiv.8.3.2430.

[6] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, 2nd ed. Boca Raton, FL, USA: CRC Press, 1999. doi: 10.1201/9781482273144.

[7] W. Härdle, *Applied Nonparametric Regression*. Cambridge, U.K.: Cambridge Univ. Press, 1990. doi:10.1017/CCOL0521382483.

[8] B. Chen and Y. Hong, "Testing for smooth structural changes in time series models via nonparametric regression," *Econometrica*, vol. 80, no. 3, pp. 1157–1183, May 2012, doi: 10.3982/ECTA7990.

[9] J. Kloke and J. McKean, *Nonparametric Statistical Methods Using R*. Boca Raton, FL, USA: Chapman and Hall/CRC, 2024. doi: 10.1201/9781003039617.

[10] M. El-Morshedy, M. S. Eliwa, and E. Altun, "Discrete Burr-Hatke distribution with properties, estimation methods and regression model," *IEEE Access*, vol. 8, pp. 74359–74370, 2020, doi: 10.1109/ACCESS.2020.2988431.

[11] Y. Gai, X. Zhu, and J. Zhang, "Testing symmetry of model errors for nonparametric regression models by using correlation coefficient," *Commun. Stat. Simul. Comput.*, vol. 51, no. 4, pp. 1436–1453, 2022, doi:10.1080/03610918.2019.1670844.

[12] P. Čížek and S. Sadıkoğlu, "Robust nonparametric regression: A review," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 12, no. 3, p. e1492, May 2020, doi: 10.1002/wics.1492.

[13] S. Demir and Ö. Toktamiş, "On the adaptive Nadaraya-Watson kernel regression estimators," *Hacet. J. Math. Stat.*, vol. 39, no. 3, pp. 429–437, 2010.

[14] Z. Shang and G. Cheng, "Computational limits of a distributed algorithm for smoothing spline," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 3809–3845, 2017.

[15] P. Feng and J. Qian, "Forecasting the yield curve using a dynamic natural cubic spline model," *Econ. Lett.*, vol. 168, pp. 73–76, Jul. 2018, doi: 10.1016/j.econlet.2018.04.009.

[16] D. Van Than and T. Tjahjowidodo, "A direct method to solve optimal knots of B-spline curves: An application for non-uniform B-spline curves fitting," *PLoS ONE*, vol. 12, no. 3, p. e0173857, Mar. 2017, doi:10.1371/journal.pone.0173857.

[17] L. Xiao, "Asymptotic theory of penalized splines," *Electron. J. Stat.*, vol. 13, no. 1, pp. 747–794, 2019, doi:10.1214/19-EJS1541.

[18] D. Ruppert, "Nonparametric regression and spline smoothing," *J. Amer. Stat. Assoc.*, vol. 96, no. 456, pp. 1387–1388, Dec. 2001, doi: 10.1198/016214501753382417.

[19] A. Perperoglou, W. Sauerbrei, M. Abrahamowicz, and M. Schmid, "A review of spline function procedures in R," *BMC Med. Res. Methodol.*, vol. 19, no. 1, p. 46, Mar. 2019, doi: 10.1186/s12874-019-0666-3.

[20] L. C. Marsh and D. R. Cormier, *Spline Regression Models*. Thousand Oaks, CA, USA: SAGE Publications, 2001. doi: 10.4135/9781412985901.

[21] S. N. Wood, *Generalized Additive Models: An Introduction with R*, 2nd ed. Boca Raton, FL, USA: Chapman and Hall/CRC, 2017. doi: 10.1201/9781315370279.

[22] T. Lyche, C. Manni, and H. Speleers, "Foundations of spline theory: B-splines, spline approximation, and hierarchical refinement," in *Splines and PDEs: From Approximation Theory to Numerical Linear Algebra*, M. Bozzini, T. Lyche, and G. Monegato, Eds. Cham, Switzerland: Springer, 2018, pp. 1–76. doi: 10.1007/978-3-319-94911-6_1.

[23] X. Liu, X. Wang, Z. Wu, D. Zhang, and X. Liu, "Extending Ball B-spline by B-spline," *Comput. Aided Geom. Des.*, vol. 82, p. 101926, Sep. 2020, doi: 10.1016/j.cagd.2020.101926.

[24] Y. Wang, *Smoothing Splines: Methods and Applications*. Boca Raton, FL, USA: CRC Press, 2011. doi:10.1201/b10954.

[25] C. Ke and Y. Wang, "Smoothing spline nonlinear nonparametric regression models," *J. Amer. Stat. Assoc.*, vol. 99, no. 468, pp. 1166–1175, Dec. 2004, doi: 10.1198/016214504000000755.

[26] G. Kauermann, "A note on smoothing parameter selection for penalized spline smoothing," *J. Stat. Plan. Inference*, vol. 127, no. 1-2, pp. 53–69, Jan. 2005, doi: 10.1016/j.jspi.2003.09.023.

[27] G. Wahba, *Spline Models for Observational Data*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics (SIAM), 1990. doi: 10.1137/1.9781611970128.

[28] P. Craven and G. Wahba, "Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation," *Numer. Math.*, vol. 31, no. 4, pp. 377–403, Dec. 1978, doi:10.1007/BF01404567.

[29] A. Wytyczak-Partyka, "Organ surface reconstruction using B-splines and Hu moments," *Acta Polytech. Hung.*, vol. 11, no. 10, pp. 163–179, 2014. doi: 10.12700/APH.11.10.2014.10.9.