Research Article GU J Sci 38(3): 1518-1538 (2025) DOI: 10.35378/gujs.1587357

Gazi University

Journal of Science

http://dergipark.org.tr/gujs

Some Paradoxical Perspectives on Approaches and Structures in Functional Data Analysis

Tuba SEKERCI[1,*], Mehmet GURCAN[2]

[1] Fırat University, Vocational School of Social Sciences, 23000, Elazig, Türkiye
[2] Fırat University, Faculty of Science, 23000, Elazig, Türkiye

Highlights

- This article presents an alternative perspective on functional regression analysis.
- Findings are supported by the literature and offer practical suggestions for existing studies.
- The advantages of using Bernstein polynomials in functional data analysis are thoroughly discussed.

Article Info

Received: 18 Nov 2024
Accepted: 17 Mar 2025

Keywords

Interpolator estimation
Kernel estimator
Longitudinal data
Bernstein polynomial

Abstract

Developments in functional data analysis have attracted considerable attention in recent literature. This study aims to provide general information about functional data analysis and demonstrate how it can be enriched with auxiliary tools. When the article is considered as a whole, the results predominantly pertain to functional linear models. The first section discusses the estimation of regression model parameters using the Least Squares method. It explains how the Least Squares method is applied within a functional framework and incorporates auxiliary calculations as part of the modeling process. At this stage, the Error Sum of Squares, which forms the basis of the Least Squares method, is represented as a vector field. The second section addresses the interim estimation problem. In this part, the Bernstein polynomial is combined with the wavelet transform to address the interim estimation challenge. The final section introduces various types of functional data analysis. Specifically, the Bernstein polynomial is used in estimating a functional linear model with functional coefficients. Employing the Bernstein polynomial as a model component in the linear model offers a simpler and more innovative approach compared to traditional functional linear model structures. The methods proposed in this study are generally practical and compatible with the classical framework of functional data analysis.

1. INTRODUCTION

Although the origins of data science are difficult to trace precisely, the field is generally considered to have emerged in the early eighteenth century, particularly with the development of probability theory. Ronald Aymler Fisher (1890-1962) is one of the pioneers of statistical science in the modern sense. He broke many grounds, including descriptive statistics for data and the relationship between data sets [1]. However, at the time, statistical science primarily served as a supporting tool in fields such as genetics, biology, demography, and public health. Consequently, its development was largely driven by the need to understand data structures. After reaching a certain level of maturity, the focus shifted from population characteristics to data sources, and from data structure to event structure. From this point onward, data analysis evolved into a functional data structure. One of the simplest examples of this is time series data. For a long time, time series data were often interpreted as being lagged in itself. Analysts searched for recurring periods within the data, with the assumption that these periods would continually repeat. Meteorological data, in particular, serve as a prime example of this due to their inherently cyclical structure [2]. Later, as time series data analysis expanded beyond weather forecasting, many techniques employed in Fourier analysis were found to yield significant results in this context [2-5].

Indeed, in some cases, a dataset should not be regarded as independent from its data source. Here, when we refer to the data source, we do not only mean the audience. By audience, we traditionally mean the community in which the data is observed. In contrast, when we discuss the data source, we are referring to the origin from which the data emerges, shaped by the functional structure of the event. Naturally, not all

*Corresponding author, e-mail: tsekerci@firat.edu.tr

datasets are influenced by this distinction. The same conceptual difference applies to the definitions of random variables and stochastic variables. Although many researchers consider both terms to have the same meaning, a stochastic variable carries more meaning than a random variable. A stochastic variable participates in a sequence of events and characterizes their structure. In this sense, it defines a more specific subset within the concept of a random variable [6].

In this context, the interpolation problem is directly related to the functional data structure. There are many interpolation estimators in use today. Examining these estimators reveals an effort to create a functional form suitable for the data structure within a localized region. It is challenging to categorize these methods under a single framework. However, one prominent method in this regard is kernel estimation. This method is used to estimate the density function. Instead of performing traditional predictions using the data histogram, researchers can conduct a more refined analysis through kernel density estimation. Unlike the histogram, the kernel technique provides a smooth estimate of the density function [7]. Two key concepts govern kernel estimation: the kernel function and the smoothing parameter. In general, terms such as smoothing function, smoothing method, and smoothness are commonly used in this context. To clarify, a scatter plot is typically considered rough, whereas a polynomial curve is perceived as smooth. In this sense, a scatter plot or histogram of the data represents a rough graph, while the probability density function of a polynomial or continuous distribution representing the data presents a smooth graph. Naturally, a function with an analytical expression can effectively convey comprehensive information. A classic example is the sensitivity of the arithmetic mean to extreme outliers in the data. Rather than calculating the mean directly from the raw data, it is often more practical to compute it from the existing density function.

As emphasized at the outset, the use of probability theory in data analysis has led to the development of numerous methods. There are many commonly used concepts in this field. For instance, when referring to the distribution of data, we often mean the percentiles of grouped data. The same applies to the density or distribution function. The distribution function expresses cumulative percentiles. If the distribution function has an analytical expression, the variable represents the numerical value of the data, while the function's output corresponds to the cumulative probability. In such cases, the inverse of the distribution function is effectively used as a data generator,

$$X = F^{-1}(p), \qquad 0 \leq p \leq 1.$$

As you can see, the most basic functional data structure is a data generator. Similarly, let us imagine that we are trying to extract data from normal distributions. There are certainly many ways to generate data from a normal distribution today [8]. Simply way we can also follow the following method,

$$p = exp\left\{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right\}$$

$$X = \mu \pm \sigma\sqrt{-2lnp}.$$

As can be seen here, the data source can be easily expressed with the help of a parameter. Table 1 presents the means and variances obtained from datasets generated for different means and standard deviations.

Table 1. *Means for datasets generated from a normal distribution*

n	μ	σ	$\bar{X}$	$\bar{X}(polynomial)$
50	10	5	9.999	9.879
50	10	10	9.999	9.597
50	10	15	9.999	9.693

The rightmost column of the table above is the average obtained from the polynomial approximation of the generated data. The materials in functional data analysis are representative even for small datasets that are randomly generated from normal distributions, as demonstrated by the straightforward example presented here.

The most fundamental principle in functional data analysis is to express the analytical structure of the data source using a functional form. The functional structure employed for the data source can adequately represent the data insofar as it aligns with the nature of the data source. A well-designed functional structure tailored to the data source undoubtedly ensures the accurate derivation of all statistics relevant to the research. However, not all variables and their corresponding data necessarily conform to the functional data structure. In reality, the functional structure of the data source may not always exist. Therefore, functional data analysis is applicable to specific data types. The independent variable space, as defined by the differential structure, naturally establishes the statistical model. Similarly, the differential characteristics provided by explanatory variables delineate the differential structure of the event. This study aims to examine several structures employed in functional data analysis. A historical review of the literature reveals that the most significant motivation during the formative phase of this theory was the advantages offered by basis functions within Hilbert space [9]. Notably, Ramsey, one of the pioneering researchers in this area, made significant contributions to this field. Commonly utilized basis functions in functional data analysis and functional linear models include trigonometric, exponential, and polynomial functions. Among these, Fourier-type bases—especially trigonometric and exponential—are particularly favored. In recent years, spline bases have also become increasingly popular, especially within the context of functional linear modeling. This study highlights the practical benefits of various auxiliary tools that can be interchangeably utilized in functional data analysis.

2.MATERIAL-METHOD

2.1. An Auxiliary Material in Functional Data Structure: Bernstein Polynomial

There are multiple ways to build a functional framework for data. However, in this paper, we will only be interested in longitudinal data. The first explanatory variable of a longitudinal observed dataset is the time parameter. The source of the observed data is therefore a parametric function. In general, equally spaced observations are an important advantage for longitudinal data. Thus, a family of Berstein-type polynomials can be easily used to represent the data source. An n-th order Bernstein polynomial is defined as follows [9-10]

$$B_n(t) = \sum_{j=0}^{n} \binom{n}{j} X_j t^j (1-t)^{n-j}, \quad 0 \le t \le 1. \tag{1}$$

This polynomial has many important applications in the literature [11] used the polynomial for a growth curve model [12] used the approximation properties of the Bernstein polynomial to estimate the pause moments in a growth curve model. The approximation properties of the polynomial can be found mainly in [13].

Different types of polynomials can also be used, although in general a continuous data structure. Naturally, these polynomials can be compared in terms of fit. However, it is natural that the Bernstein polynomial is preferred in many places for ease of use.

2.2. Differential Forms

A continuous random variable has a continuous distribution function, even if not explicitly stated. This crucial property ensures that the distribution function can be derived as the integral of another function. At first glance, this might appear as an unnecessary detail. Indeed, by differentiating the distribution function, we obtain the density function, and conversely, by integrating the density function, we retrieve the distribution function. However, there is a significant paradox related to the normal distribution. If we only possess the density function without the distribution function, it may seem that the normal distribution lacks a distribution function altogether—let alone an absolutely continuous distribution function. However, this assertion is inaccurate. The fact that we cannot obtain the distribution function analytically does not imply its non-existence. Therefore, the probabilities associated with the normal distribution are calculated numerically by integrating the density function and then tabulated. In summary, we can state that the derivative of any known elementary function does not yield the density function of the normal distribution.

A similar non-solution is found in regression analysis. Even in a simple linear model, there are as many equations for two unknowns as there are data to solve. However, it is not possible to satisfy all of these equations with only two constants. Ideally, we assume that such an equation exists. We just have to capture this equation with a small error. The construction and solution of this problem are discussed in detail in [14].

In this case, suppose that our ideal model is as follows,

$$z = f\left(x_1, \cdots, x_p\right). \tag{2}$$

Here $\left(x_1, \cdots, x_p\right)$ a vector of our explanatory variables z is the response variable. Obtained from the response variable

$$\nabla z = \left(\frac{\partial z}{\partial x_1}, \cdots, \frac{\partial z}{\partial x_p}\right). \tag{3}$$

The line integral of the vector field over the parametric vector of explanatory variables will give us the model equaiton we are looking for,

$$z = \oint_0^t \nabla z d\vec{r}(t). \tag{4}$$

Here $\vec{r}(t)$ has the following parametric form,

$$\vec{r}(t) = \left(x_1(t), \cdots, x_p(t)\right). \tag{5}$$

For example $z = c_1 x_1 + \cdots + c_p x_p$ model. In this case $\nabla z = \left(c_1, \cdots, c_p\right)$ and we can easily write the following equation

$$z = \oint_0^t \left(c_1, \cdots, c_p\right) d\left(x_1(t), \cdots, x_p(t)\right) = c_1 x_1(t) + \cdots + c_p x_p(t).$$

This equation gives us a parametric form of the linear regression model. The calculation of the coefficients can be done by the least squares method or by using the constants required by the differential structure.

2.3. Estimation of Regression Parameters: Least Squares Method

The least squares method (LSM), which is considered one of the most significant methods developed in the last century, is an optimization problem. Finding the regression parameters that minimize the loss function generated through the regression equation enables the construction of the estimation equation. Generally, the estimation of the utility model can serve several purposes. Two primary applications include: approximating a transcendental equation with a polynomial over a closed interval, and modeling a dataset with an uncertain transcendental structure using a linear model. The least squares method has been successfully applied to both problems. In general, the loss function for these two problems can be expressed as follows [14-17]

$$U = \int_{-\infty}^{\infty} (Y - P_n(X))^2 dx. \tag{6}$$

Regression coefficients are obtained from the solution of the system of equations (system of normal equations) formed by setting the first derivatives of the loss function concerning the regression coefficients equal to zero.

Illustrative Example 1: $Y = expx$ function $[2,3]$ in the range of the predictor $P(x) = a + bx + cx^2$ polynomial. For this, let us construct the loss function given by Equation (6) and its derivatives

$$U = \int_2^3 (expx - (a + bx + cx^2)\,)^2 dx$$

$$\frac{\partial U}{\partial a} = \int_2^3 (expx - a - bx - cx^2) dx = 0$$

$$\frac{\partial U}{\partial b} = \int_2^3 x(expx - a - bx - cx^2) dx = 0$$

$$\frac{\partial U}{\partial c} = \int_2^3 x^2(e^x - a - bx - cx^2) dx = 0.$$

In this way, the normal equations can be obtained as follows,

$$12.7 = a + 2.5b + 6.33c$$

$$32.78 = 2.5a + 6.33b + 16.25c$$

$$85.65 = 6.33a + 16.25b + 42.2c.$$

From the solution of these equations $a = -19.04$, $b = 12.535$, $c = 0.064$ coefficients are obtained.

Table 2 below $\hat{y} = -19.04 + 12.535x + 0.064x^2$ prediction equation and the actual values.

Table 2. *Estimated and error values of the function*

x	y	$\hat{y}$	$Error = y - \hat{y}$
2	7.38	6.286	1.094
2.1	8.16	7.565	0.595
2.2	9.02	8.847	0.173
2.3	9.97	10.129	-0.159
2.4	11.02	11.412	-0.392
2.5	12.18	12.69	-0.51
2.6	13.46	13.983	-0.523
2.7	14.87	15.271	-0.401
2.8	16.44	16.559	-0.119
2.9	18.17	17.849	0.321
3	20.08	19.141	0.939

The above example is quite classic for functional regression. It should be noted that the coefficients are directly dependent on the range of integration. If the model equation can be constructed, this model is quite open to be estimated with the help of a simpler model. In the classical linear model, the normal equations are obtained by sums instead of integrals. The classical regression equation is given by the following expected value

$$Model := E(Y|X = x). \tag{7}$$

The above model has both a probabilistic and a functional structure. To understand these structures more easily, let us give two important properties of the expected value

$$EE(Y|X=x) = EY \tag{8}$$

$$E\big(XE(Y|X=x)\big) = EXY. \tag{9}$$

Proof. (for Equation (9))

$$E\big(XE(Y|X=x)\big) = \iint x\,E(Y|X=x) f\big(x, E(Y|X=x)\big)dxdt$$

$$= \iint x(\int yf(y|x)dy)\, f\big(x, E(Y|X=x)\big)dxdt$$

$$= \iiint xy\frac{f(x,y)}{f(x)}\, f\big(x, E(Y|X=x)\big)dxdydt$$

$$= \iint xy\frac{f(x,y)}{f(x)}\,(\int f\big(x, E(Y|X=x)\big)dt)dxdy$$

$$= \iint xy\frac{f(x,y)}{f(x)}\, f(x)dxdy$$

$$= \iint xyf(x,y)dxdy$$

$$= EXY$$

Now assume that the model equation is simply linear. In this case, Equation (7) is as follows,

$$E(Y|X=x) = a + bX$$

$$EE(Y|X=x) = EY = a + bEX \tag{10}$$

$$E\big(XE(Y|X=x)\big) = EXY = aEX + bEX^2. \tag{11}$$

The last two equations obtained are normal least squares equations. Indeed, considering the right-hand sides, we can easily write the following

$$\begin{bmatrix} 1 & EX \\ EX & EX^2 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} EY \\ EXY \end{bmatrix}.$$

If both sides of the equation are multiplied by the number of observations, the normal least squares equations are obtained

$$\begin{bmatrix} n & \sum X \\ \sum X & \sum X^2 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}.$$

In normal equations, $E(Y|X=x)$ the fact that it is obtained from the conditional expected value emphasizes the importance of using Kernel estimators in model estimation [18,19]. It is important to note that the

conditional expected value is a random variable. The functional structure of the conditional expected value corresponds to the functional structure of the regression source.

2.4. Two Important Tools in the Forecasting Problem: Taylor Series and Bernstein Polynomial

Why does the researcher seek to determine the equation representing the relationship between two correlated or regressive variables? The answer is straightforward: the values in the dataset are inherently related to one another. However, predicting a value for any explanatory variable not observed in the dataset can only be achieved by deriving the model equation. Therefore, the interpolation problem is a critical issue in regression analysis. Two key constraints arise when constructing the model equation: the first is "minimum error," and the second is "minimum variance." It is nearly impossible to satisfy both simultaneously. As such, we must determine which of these two constraints holds greater priority in our analysis. For interpolation to be meaningful, the first requirement is minimizing variance. While achieving minimum error is possible, we cannot guarantee that the error for intermediate values will remain minimal. Thus, minimizing variance should be our primary goal. This explains why linear models are so widely used in regression.

Data with a functional source should follow an acceptable curve between two consecutive observation values. As a classic example t_i instantly $y_i = 3$ and t_{i+1} instantly $y_{i+1} = 4$. It is quite unreasonable for the intermediate value of functional data to be infinite. Therefore, it is quite common to use kernels such as B-splines or Bernstein polynomials that provide smooth transitions when connecting two nodes, i.e.two consecutive data. Let us briefly emphasize here that the Word “smoothness” stands for what we are describing.

As the degree of a polynomial increases, the variance in polynomial models also increases, making the error described above quite common. Polynomial models exhibit a phenomenon similar to inflation. These models minimize error and may initially appear visually appealing, but they tend to produce significant errors, particularly at intermediate values. To address this issue, one effective solution is the Taylor polynomial. To estimate any intermediate value using the Taylor polynomial, the polynomial should be expanded around a point close to this value. This approach reduces the estimation error. For any given $c \in R$ the Taylor polynomial in the vicinity of the point is defined as follows

$$f(t) = f(c) + f^{(1)}(c)(t-c) + f^{(2)}(c)\frac{(t-c)^2}{2!} + \cdots \quad . \tag{12}$$

Here $c \in R$ value of $t \in R$ is important to minimize the error of the prediction.

This polynomial is named Taylor’s not because of the construction of the polynomial, but because it can bound the error between the predicted value and the true value. This is because the construction of the polynomial is quite simple and anonymous. In some studies, it is not enough to bind the error with the prediction value, but it is also necessary to bind the derivatives of the prediction operator. However, this is not possible for every estimation polynomial. In particular, Bernstein and Feller polynomials can easily provide this property [12-20]. Therefore, it is an important tool for finding intermediate values. Let us now note the following illustrative example of the Taylor polynomial.

Illustrative Example 2: In the Table 3, $y = y(x)$ Let’s examine the data with the functional structure $(y(x) = x^4/4 + x^3/3 + x^2/2)$,

Table 3. *First derivative values obtained from the data*

x	y	$y^{(1)}, h = 0.02$	$y^{(1)}, h = 0.04$	$y^{(1)}, h = 0.06$
1.8	6.1884			
1.82	6.408	10.98		
1.84	6.6348	11.34	11.16	
1.86	6.8669	11.605	11.4725	11.3083

1.88	7.105	11.905	11.755	11.6166
1.9	7.349	12.2	12.0525	11.9033
1.92	7.5998	12.54	12.37	12.215
1.94	7.8567	12.845	12.6925	12.5283
1.96	8.120	13.165	13.005	12.85
1.98	8.390	13.5	13.3325	13.17
2	8.666	13.8	13.65	13.4883

The second derivative values obtained from the dataset are presented in Table 4.

Table 4. *Second derivative values obtained from the data*

x	y	$y^{(2)}, h = 0.02$	$y^{(2)}, h = 0.04$	$y^{(2)}, h = 0.06$
1.8	6.1884			
1.82	6.408			
1.84	6.6348	18		
1.86	6.8669	13.25		
1.88	7.105	15	14.875	
1.9	7.349	14.75	14.5	
1.92	7.5998	17	15.375	
1.94	7.8567	15.25	16	15.195
1.96	8.120	16	15.875	15.7783
1.98	8.390	16.75	16	15.9166
2	8.666	15	16.125	16

Using the first and second derivative values above $y(x)$ we can only do a trinomial Taylor expansion of the functional data. $y(x)$ of the data source $x = 2$ the series expansion in the vicinity is as follows (for respectively $h = 0.02, 0.04, 0.06$)

$$y(x) \approx 8.666 + 13.8(x-2) + 15\frac{(x-2)^2}{2} \tag{13}$$

$$y(x) \approx 8.666 + 13.65(x-2) + 16.125\frac{(x-2)^2}{2} \tag{14}$$

$$y(x) \approx 8.666 + 13.4883(x-2) + 16\frac{(x-2)^2}{2}. \tag{15}$$

Using the above equations $x = 1.98$ forecasts are respectively $y = 8.393, 8.396, 8.399$ is obtained as. Note that the shorter the step interval, the more accurate the predictions are. In Table 5, $y(x)$ observation and prediction values for the data source are shown.

Table 5. *Error values obtained from the estimation polynomial (13)*

x	y	$\hat{y}$	$Error = y - \hat{y}$
1.8	6.1884	6.206	-0.0176
1.82	6.408	6.425	-0.017
1.84	6.6348	6.65	-0.0152
1.86	6.8669	6.881	-0.0141
1.88	7.105	7.118	-0.013
1.9	7.349	7.361	-0.012
1.92	7.5998	7.61	-0.0102
1.94	7.8567	7.865	-0.0083
1.96	8.120	8.126	-0.006
1.98	8.390	8.393	-0.003
2	8.666	8.666	0

Errors in the table above $x = 2$ tends to decrease as it approaches its value. Now, using the same data, let's construct the square spline functions with the help of Bernstein polynomial as follows,

$$y(x) = 6.1884(1-x)^2 + 2(6.408)x(1-x) + 6.6348x^2, \quad 1.8 \le x \le 1.84$$

$$y(x) = 6.6348(1-x)^2 + 2(6.8669)x(1-x) + 7.105x^2, \quad 1.84 \le x \le 1.88$$

$$y(x) = 7.105(1-x)^2 + 2(7.349)x(1-x) + 7.5998x^2, \quad 1.88 \le x \le 1.92$$

$$y(x) = 7.5998(1-x)^2 + 2(7.8567)x(1-x) + 8.12x^2, \quad 1.92 \le x \le 1.96$$

$$y(x) = 8.12(1-x)^2 + 2(8.39)x(1-x) + 8.666x^2, \ 1.96 \le x \le 2.$$

Table 6. *Prediction values calculated using Spline polynomials*

x	y	$\hat{y}$	$Error = y - \hat{y}$
1.81	6.2978	6.2986	-0.0008
1.82	6.408	6.4098	-0.8818
1.83	6.521	6.5218	-0.0008
1.85	6.7501	6.7512	-0.0011
1.86	6.8669	6.8684	-0.0015
1.87	6.9852	6.9863	-0.0011
1.89	7.2264	7.2274	-0.001
1.9	7.349	7.3507	-0.0017
1.91	7.4738	7.4748	-0.001
1.93	7.7275	7.7286	-0.0011
1.94	7.8567	7.8583	-0.0016
1.95	7.9876	7.9887	-0.0011
1.97	8.2542	8.2553	-0.0011
1.98	8.39	8.3915	-0.0015
1.99	8.5275	8.5228	-0.0047

It is evident from Table 6 above that the errors are close to zero when quadratic spline functions are used. As the quadratic spline functions obtained using Bernstein polynomials match the observed values at the extreme points, these estimated values are excluded from the table above. In general, the spline method yields more consistent estimates in interpolation. Naturally, like any method, it has both advantages and certain limitations. For this method to be effective, it is crucial that the data is appropriate for the method and that the intervals between data points are small.

2.5. Bernstein Window and Windowed Regression

The interpolation problem is more practical than the regression equation. Linear models used to avoid the problem of high variance do not minimize the amount of error as much as desired. Instead, interpolation can provide a better fit and avoid the problem of high variance since it is not attributed to the whole data. However, in this case, we may need to transfer the interpolation problem from a single variable to more dimensions. In this case, we can develop a useful method called "*windowed interpolation*". Let's call the method "BW" for short since we will do this with Bernstein-type polynomials. The basic BW function will be defined as follows

$$BW = a(1-t_1)(1-t_2) + b(1-t_1)t_2 + ct_1(1-t_2) + dt_1t_2, \qquad 0 \le t_1, t_2 \le 1. \tag{16}$$

Here the coefficients are the observation values at the corner points of the window. Let's take a look at the following short example to see how to apply this method in detail.

Illustrative Example 3: Consider the following data

The observation values obtained from the data source are listed in Table 7.

Table 7. *Observations from the source*

Sequence number:	X_1	X_2	X_3	Code
1	1.25	1	1	(0,0)
2	1.5	1	2	(0,1)
3	4.75	2	3	(1,0)
4	9.75	3	3	(1,1)
5	17.25	4	5	
6	37.75	6	7	

An arbitrary intermediate value $X_2 = 2.5$ and $X_3 = 2.5$ as the first four values. Since these values are between the first four values for the four corners where we will create the window, the corners should be chosen as the first four values. In this case BW should be written as follows

$$BW(1) = 1.25(1-t_1)(1-t_2) + 1.5(1-t_1)t_2 + 4.75t_1(1-t_2) + 9.75t_1t_2.$$

The interpolation results for selected intermediate values using the BW(1) method are shown in Table 8.

Table 8. *BW(1) prediction values for some selected intermediate values*

X_2	X_3	t_1	t_2	$\widehat{X_1}$	X_1	Error
2.5	2.5	0.75	0.75	6.734	6.875	0.141
1.6	2.8	0.3	0.9	3.807	3.26	-0.547
1.06	2.64	0.03	0.82	1.6768	1.7836	0.1068
1.24	2.1	0.12	0.55	2.121	2.062	-0.059

Now let the intermediate values be chosen to be within the last four data. In this case the BW can be constructed as follows

$$BW(2) = 4.75(1-t_1)(1-t_2) + 9.75(1-t_1)t_2 + 17.25t_1(1-t_2) + 37.75t_1t_2.$$

The interpolation results for selected intermediate values using the BW(2) method are presented in Table 9.

Table 9. *BW(2) prediction values for some selected intermediate values*

X_2	X_3	t_1	t_2	$\widehat{X_1}$	X_1	Error
3.5	5.74	0.375	0.685	16.844	13.685	-3.159
2.98	6.56	0.245	0.89	15.642	10.52	-5.121
5.832	6.912	0.958	0.978	36.137	35.74	-0.396
2.008	3.012	0.002	0.003	4.79	4.785	-0.0049

It is possible to increase the intermediate values used in the tables above. When we pay attention to the interpolation estimates, the estimates have less error at the extremes. This is because the Bernstein polynomial gives more accurate estimates at the extremes. The EKK estimate of the functional data of interest in the example is obtained as follows

$$X_1 = -10.189 + 7.006X_2 + 0.25X_3.$$

The errors of some of the values obtained from this estimation equation are given in the Table 10.

Table 10. *EKK estimation values*

X_2	X_3	$\widehat{X_1}$	X_1	Error
2.5	2.5	7.951	6.875	-1.076
1.6	2.8	1.72	3.26	1.54
1.06	2.64	-2.533	1.7836	4.3166
1.24	2.1	-1.25	2.062	3.312

As can be seen from the table above, the errors of the EKK estimates are higher than the errors obtained from the BW(1) interpolation. The interval length chosen while applying the EKK also effects this result. The variance will increase as the interval length increases. When the interval length is narrowed, EKK is already reduced to solving the interpolation problem. In this case, the linear model has no meaning.

2.6. Some Important Results of the Bernstein Polynomial

The primary objective of regression analysis is to develop a model equation that most accurately represents the data source. As the number of variables increases, it becomes more challenging to achieve this across the entire dataset, or more precisely, across the defined ranges of the variables. In functional data analysis, the main objective is to represent the data source with minimal error. In this respect, it is also important to determine in which type of analysis the functional data will be applied. If our goal is to obtain the regression model, it is not sufficient to identify the data sources of the variables individually. It is also essential to determine the functional relationship between the response variable and the explanatory variables. Under these circumstances, it is natural to incorporate differential structures into the analysis. However, obtaining an unpredictable differential structure based on the data structure is challenging. In such cases, the data collection process should be well understood, and the differential structure between variables should align with the narrative of the data. Many interpolation methods can provide accurate predictions based on the data range. However, if the data gap is not sufficiently small, the derivative of the estimators may not align with the derivative of the data source. Naturally, there is no one-size-fits-all approach for the researcher to make the optimal choice in these cases. The use of Bernstein polynomials in interpolation estimation is advantageous because the mean of the estimator coincides with the mean of the variable. Let us now demonstrate how to average a Bernstein polynomial

$$EX = \int_0^1 B_n X(t) dt. \tag{17}$$

Proof.

$$\int_0^1 B_n X(t)dt = \int_0^1 \sum_{j=0}^{n} X(j/n)\binom{n}{j} t^j (1-t)^{n-j} dt$$

$$= \sum_{j=0}^{n} X(j/n)\binom{n}{j} \int_0^1 t^j (1-t)^{n-j} dt$$

$$= \sum_{j=0}^{n} X(j/n)\binom{n}{j} B(j+1, n-j+1)$$

$$= \sum_{j=0}^{n} X(j/n)/n$$

$$= \frac{1}{n}\{X(0) + \cdots X(1)\}$$

$$= EX$$

Here $B(n,m)$ is denoted by the beta function.

To make the above proof understandable, we can easily give the following example. A X let the random variable take the values -2, -1, 1, 2 with equal probability. It is clear that the mean of the variable will be zero. Now let us take the integral of the polynomial representing this variable below,

$$\int_0^1 B_n X(t)dt = \int_0^1 (-2(1-t)^3 - 3t(1-t)^2 + 3t^2(1-t) + 2t^3)dt = 0.$$

In such examples, it is important to pay attention to the number of repetitions of the values if the probabilities are not equal. This result allows us to calculate the other moments in this way. In this case, we can easily write the following equation for the variance,

$$VarX = \int_0^1 B_n^2 X dt - E^2 X. \tag{18}$$

Proof.

$$VarX = E(X - EX)^2$$

$$= E(B_n X - EX)^2$$

$$= \int_0^1 \{B_n^2 X - 2B_n X + E^2 X\}dt$$

$$= \int_0^1 B_n^2 X dt - E^2 X$$

However, when calculating the moment of multiplication, the operation should be performed on the prediction polynomial of the new variable obtained from the product of the variables, not the product of the prediction polynomials

$$EXY = EB_n(XY)(t) = \int_0^1 B_n(XY)(t)dt. \tag{19}$$

The condition of independence here $B_n(XY) = B_nXB_nY$ condition can be met. This condition can only be met by using probability values and not by multiplying polynomial terms.

2.7. Regression of Bernstein Polynomials

Using least squares regression of time-dependent functional structures instead of least squares regression using measured values of variables in longitudinal observations provides more statistical information to researchers. However, it should be noted that the functional structure of the variable should be estimated accurately. When all variables in the model are observed longitudinally, the response variable $Y(t)$ and the explanatory variables $X_1(t), \cdots, X_p(t)$ have a functional structure. In this case, the linear model structure can be assumed to be as follows

$$Y(t) = a_0 + a_1X_1(t) + \cdots + a_pX_p(t) + \varepsilon. \tag{20}$$

Based on this model, EKK estimation of the model coefficients is equivalent to the procedure applied in multiple regression. The most important contribution of the model is that, unlike the multiple regression model, when any observation value is taken from outside the dataset, the corresponding value is used for estimation. $t_0 > 0$ parameter can be estimated by finding the parameter. This contribution can provide more accurate answers to researchers when the data source is selected correctly. This is explained in the example in the BW model. The important point we would like to emphasize in this subsection is the case where the explanatory variables are chosen in an orthonormal way. When we pay attention to the matrix form of the EKK normal equations, the coefficients matrix of the parameter vector consists of the variances on the prime diagonal and the covariances of the explanatory variables on the diagonal. In the ideal case, when a standardization and steepening operation is performed on the explanatory variables, the coefficients matrix turns into a unit matrix. In this case, since there is an ineffective multiplication factor, the vector of coefficients is directly X^tY multiplier. In such a case, the regression constant directly depends on the mean of the responsive variable, and the other coefficients $\langle Y, X_j \rangle$ will be equal to its product

$$a_0 = \int_0^1 y(t)dt = Ey(t) \tag{21}$$

$$a_j = \int_0^1 y(t)x_j(t)dt. \tag{22}$$

As we said at the beginning, this structure constitutes the basic logic of functional data analysis. What is difficult to recognize here is that once the functional forms of the explanatory variables have been determined, these variables must satisfy the following orthogonality condition

$$\langle X_i, X_j \rangle = 0\,, \qquad i \neq j\,. \tag{23}$$

Different kernels are used when estimating the functional forms of explanatory variables. The popular one in the literature is the Fourier Kernel given below

$$\{e^{int} : n \geq 0\}. \tag{24}$$

However, in the case of a polynomial Kernel used in the B-Spline method, the condition that the two polynomials satisfy the orthogonality condition is not mentioned. Instead of this condition, the condition

that they are uncorrelated is preferred. As a result, both cases lead to the same door. While the orthogonality condition is given by (23), the uncorrelatedness condition is given by the following expression

$$\langle X_i, X_j \rangle = \int_0^1 x_i(t)x_j(t)dt \sim EXY = 0. \tag{25}$$

The equivalence in the above form is used because it corresponds to the numerical calculation of the integral. Spline methods are very similar to the Bernstein Kernel. The ease of use that the Bernstein Kernel offers to researchers in terms of ease of use can make the use of this polynomial widespread. The simplest definition of interval $[0,1]$ integrals are treated over this unit interval and division by the interval length is ineffective when calculating averages.

2.8. Functional Linear Model and Different Types

A linear model structure is briefly expressed by the following Equation [21]

$$Y = a_0 + \langle X, A \rangle + \varepsilon\,. \tag{26}$$

Here A denotes the vector of regression coefficients excluding the constant term. Here, the coefficients $\langle X, A \rangle$ the inner product can be expressed in Equation (27) in a suitably chosen functional space

$$\langle X, A \rangle = \int_0^1 X(t)A(t)dw(t). \tag{27}$$

Here w is a real measure valid in the domain of definition. Due to this property, the model equation expressed Equation (26) can be conveniently written Equation (28)

$$Y = a_0 + \int_0^1 X(t)A(t)dw(t) + \varepsilon\,. \tag{28}$$

In the context of the generalized functional linear model proposed in this study, the model equation is defined as the conditional mean of the response variable [22]

$$\hat{Y} = E\{Y|X(t)\}\,. \tag{29}$$

There are many later studies in the literature on how to choose the expressions in the integral in Equation (28) above. For example, B-spline bases are an important option adapted to this structure [22]. It is worth emphasizing that the Fourier basis is the most preferred choice in this type of study. Let us now assume that the basis system that best represents the explanatory variables in the set of definitions is the following

$$\{\phi_j : j \geq 0\}. \tag{30}$$

In this case $X(t)$ and $A(t)$ functions can be written in terms of this base system

$$X(t) = \sum\nolimits_j x_j\,\phi_j \; and \; A(t) = \sum\nolimits_j a_j\,\phi_j. \tag{31}$$

In this case, Equation (27) is reduced to the following form,

$$\langle X, A\rangle = \int_0^1 \sum_j x_j\,\phi_j(t) \sum_i a_i\,\phi_i(t)dt$$

$$= \sum_j x_j \sum_i a_i \int_0^1 \phi_j(t)\phi_i(t)dt$$

$$= \sum_j x_j\,a_j.$$

If this product is used in Equation (28), it takes the following form,

$$y_j = a_0 + x_j a_j + \varepsilon_j$$

$$a_j = \frac{y_j - a_0}{x_j}\ , \qquad j = 1, \cdots, n.$$

In short, the model equation takes the Equation (32)

$$y(t) = a_0 + A(t)x(t) + \ \varepsilon(t). \tag{32}$$

In this model $a_0 + \varepsilon(t) = a(t)$ to use the base system. In this case, however, the error term of the model will consist only of errors due to the fit of the chosen base system to the variable. It is also possible to write Equation (33) by assigning a functional structure to all these practically used loops [23].

$$y(t) = a_0 + \int_0^1 X(s)A(t,s)dw(s) + \varepsilon\,. \tag{33}$$

As a result, the development of functional linear models or functional regression structures in the literature has been driven by the development of software techniques and algorithms. In this context, the concept of appropriate basis, represented by (30), is of great importance. Basis functions are categorized into three main groups exponential, trigonometric, or polynomial, and different versions of these three main groups have emerged thanks to the possibility of calculating in infinitely small intervals provided by the form. In particular, while it was not possible to find polynomial bases, wavelet or window bases can now offer us this possibility.

3. THE RESEARCH FINDINGS

This study, which examines the use of functional structures mainly in regression models under the framework of linear models in general, has attempted to address much information in the literature at the same time in the findings and conclusion sections. Since the findings in the study are presented together with their equivalents in the literature, they are both supported by their equivalents in the literature and provide practical suggestions for the information in the literature. In particular, BW regression adapts window-based analyses to the regression framework, which is an important contribution to the improvement in both fit and prediction. At this stage, the findings and conclusions of the study need to be briefly discussed and explained. First, let us talk about the “mother wavelet” adaptations of the mother wavelet transform.

3.1. Basic Base Function and Its Versions: Infinitesimal Intervals

Without loss of generality $t \geq 0$ the function defining the unit interval on the axis can be given Equation (34)

$$\psi(t) = \begin{cases} 1 & : \ 0 \le t \le 1 \\ 0 & : \ d.d. \end{cases} . \tag{34}$$

This transform is called the mother wavelet. Graphically, it consists of a unit cube placed at the beginning of the axis. Instead of the expression "unit cube", it could also be called "window into the region of interest" or "focused region of interest". These expressions serve the purpose better. Therefore, all foreign sources use the term "window" as in the international scientific literature. Let us also visualize this moving along the axis. In this case, a wave moving from left to right along the axis gives the appearance of a "wave". As a result, we should not forget that the still state is a window opened in space and the moving state is a wave.

Now let's narrow the window given by Equation (34) a bit,

$$\psi(nt) = \begin{cases} 1 & : \ 0 \le nt \le 1 \\ 0 & : \ d.d. \end{cases}$$

$$= \begin{cases} 1 & : \ 0 \le t \le 1/n \\ 0 & : \ d.d. \end{cases} = \phi_n(t). \tag{35}$$

Now let's move the window given by Equation (34) a little bit,

$$\psi(t+a) = \begin{cases} 1 & : \ 0 \le t+a \le 1 \\ 0 & : \ d.d. \end{cases}$$

$$= \begin{cases} 1 & : \ -a \le t \le a \\ 0 & : \ d.d. \end{cases} = \phi^a(t). \tag{36}$$

Now let's free the window given by Equation (34) from the center,

$$\psi_{ab}(t) = \begin{cases} 1 & : \ a \le t \le b \\ 0 & : \ d.d. \end{cases} . \tag{37}$$

Now let's do all these things together

$$\psi_{abn}(t) = \psi\left(\frac{2^j(t-a)}{b-a} - n\right). \tag{38}$$

Here $[a,b]$ of the range 2^j is assumed to be divided into subintervals. $n = 0, \cdots, 2^j$ to be $\psi_{abn}(t)$ of this range $\left[a + n((b-a)/2^j, a + (n+1)((b-a)/2^j\right]$ is in the lower interval. $[a,b] = [0,1]$ then Equation (38) becomes the following simplification,

$$\psi_n(t) = \psi\left(\frac{t - n2^{-j}}{2^{-j}}\right). \tag{39}$$

It is clear that both Equation (38) and (39) are orthagonal,

$$\int_0^1 \psi_n(t)\psi_m(t)\, dt = 0\,, \qquad n \neq m.$$

In addition, they are ensured to be unit-sized, that is, orthanormal, as follows,

$$\psi_{abn}(t) = \sqrt{\frac{2^j}{b-a}}\, \psi\left(\frac{2^j(t-a)}{b-a} - n\right)$$

$$\psi_n(t) = \sqrt{2^j}\,\psi\left(\frac{t - n2^{-j}}{2^{-j}}\right).$$

If we pay attention to the form of the Bernstein polynomial, the kernels that make up the sum are basically of the following form

$$b_k(t) = \binom{n}{k} t^k (1-t)^{n-k}, \qquad k = 0, \cdots, n.$$

It is not possible to write system orthagonally. However $n+1$ obtained from the data $B_n X(t)$ functional structure and infinitesimal range $\psi_k(t),\ k = 0, \cdots, 2^j$ the bases allow us to create the following functional form

$$\xi(t) = \sum_{k=0}^{2^j} B_n X(t) \psi_k(t). \tag{40}$$

Since this type of analysis is very convenient for researchers, it has been used in various fields in the literature [24-26]. The part that interests us is the detection of similarity between objects by the program, which is the basis of many learning algorithms. Most nonparametric methods treat similarity as a constant, but much recent work has shown the advantages of learning it, in particular exploiting local invariances in the data or capturing the partially nonlinear manifold on which most of the data lie [26]. The secret lies in this last sentence of Vincent and Bengio. Given a linear prediction model, not all data are included in the linear regression. The closed model can be in linear form. However, it is important for the accuracy of the prediction to examine the non-linear structure in narrow regions. BW regression provides an important convenience in this idea.

4. RESULTS AND DISCUSSION

The first analysis that comes to mind when analyzing time-dependent data is naturally time series. However, this idea is fundamentally wrong. In time series analysis, the data must be periodic or have the ability to be periodic to be analyzed. Buys Ballot, the inventor of this work, used the following sentence in his memoirs: "Every numerical series must have a period. Only if the observation is long enough to catch this period." [27]. If not? The sequential correlation model used as the basic understanding of time series remains a wrongly chosen model. In this case, functional data analysis can best analyze longitudinal data. The first applications of functional data analysis were already designed for time series data. Although Ramsay did not include such an understanding when he published his first major work on the subject, the application of the theory naturally shifted to this type of analysis [28]. Indeed, Ramsay is the most successful functional techniques researcher of the last century. The main objective of the FDA is to see the projection of the response variable onto the space of explanatory variables using the bases that best represent the data source. The resulting relationship is the functional equation of the regression. Therefore, the main direction of the FDA in the literature is functional linear modeling and estimation [29].

Approximately ninety relevant studies exist in the literature, with over three-quarters of them conducted after 2005. Overall, there are over seventy publications on base selection. A common view is that the B-spline method is the most popular of the bases used. Indeed, if one looks at the form or the application package of the theory, the B-spline basis is at the forefront. However, functional principal component analysis and linear models are also included in the literature.

Nearly half of the studies reviewed focus on applications in biostatistics, which justifies a dedicated discussion of this field. In particular, functional linear models have been employed in diffusion tensor imaging and fiber imaging, with an emphasis on basis function selection tailored to the data characteristics [30]. Functional principal component analysis (FPCA) and functional clustering techniques have been applied to gene expression and microarray data [31]. Locally weighted regression and functional clustering

were used to analyze the activity of posterior horn neurons in the spinal cord [32]. FPCA has also been used to study age-specific mortality [33], and a combination of functional principal components and splines has been applied to age-specific breast cancer mortality rates [34]. Furthermore, FDA methods have been effectively applied to a wide range of data types, including mortality, fertility, and migration rates; time-dependent gene expression data; colon cancer incidence; speech variability in individuals with lisps; and emotional responses to music. Despite the breadth of FDA applications, studies employing Bernstein polynomials in functional regression analysis remain relatively limited. A notable example is a 2023 study [35], which introduced a new method for estimating shape-constrained functional regression coefficients based on the Bernstein polynomial approach. It is important to reiterate that the foundation of FDA lies in representing elements within a functional space using appropriate basis functions. Once the functional spaces of the explanatory and response variables are defined, the response variable is projected onto the space of explanatory variables through these basis functions—most commonly via the least squares method. In this context, the Bernstein positive operator serves as an effective auxiliary tool, owing to its structural properties, and forms the foundation of the methodological framework adopted in this study.

Why Bernstein polynomial (*Advisor's comment*)? Weierstrass' brilliant proof is that "*the space of continuous functions is a dense subspace of polynomials*". An extension of this proof is that we can represent every continuous function by a polynomial of unrestricted degree. For example, as mentioned at the beginning of the results section, the Taylor polynomial has this representation capability. However, what is important is the method Weierstrass used to make this proof and Bernstein polynomial, which is a natural output of this method. This reason alone is enough to answer the question. However, there is additional information. Bernstein polynomial is simply an operator that converts a function into a polynomial. From the point of view of probability, its structure is taken from the binomial distribution. It also represents the core part of the beta distribution. There is no need to talk at length here about the wide range of applications and properties of the beta distribution. When we add all these features together, the richness of the interpretation of an estimate using the Bernstein polynomial is also the richness of the interpretation obtained from the estimate. For example, when we look carefully at the model structure given by Equation (32) in the results section and before, we see that the ratio of the response variable to the explanatory variable has an index value. In a linear model structure, we expect this index value to be constant. In fact, in general, the group of constant index values in regression models are linear models. Otherwise, the linearity of the equation never requires that the data model be linear. But for some reason, this subtle point never gets noticed. Another similar example is that in a regression model, the error term is a random variable. Every researcher of all time says this. However, the error term is a secondary random variable. The primary random variable is the explanatory and response variable. In this case, one can argue that the error term is stochastic. The observation values of the error term can be calculated firstly after the data are collected and secondly, after the regression model is decided and established. This means that it is somewhat implausible to calculate the probability of an event occurring after the experiment has been conducted. Probability and randomness exist for experiments that are designed but not implemented. Naturally, the randomness of the error term is also highly controversial. This idea can easily be used as a secondary parameter in the calculation of the term excluding the explanatory variables of the regression model given by Equation (38) in the conclusion. This means replacing the constant term with the variable in the functional regression model. After the exit of the FDA, the understanding of the constant term in the linear model should be used as the term excluding the explanatory variable. Finally, I would like to close this commentary with a rather regrettable piece of literature. Some studies in the literature have attempted to deform the Kernel part of the Bernstein polynomial slightly to make it orthonormal [36]. I am sure that Sergey Natanovic Bernstein's bones ache as the man who solved Hilbert's ninth problem. The nature of some things must be preserved as it is. It never occurred to Bernstein to adapt the construction of a polynomial instead of the orthanormal basis of a functional space. It's really funny. One should not overdo things for the sake of doing them. Let's pay a little attention to the form below because of the nature of both constructions

$$x_k = X(k/n) = \langle X(t), \psi_k(t) = \psi(\frac{t - k/n}{1/n}) \rangle.$$

The coefficients here are already obtained from the basis of the functional space and are used as multipliers of the kernels forming the polynomial. It is obvious that the polynomial cannot be written in the following from,

$$B_n X \neq \sum_k \langle X(t), \psi_k(t) \rangle b_k.$$

With Spline b_k when we put the Kernels together, it is easily seen that it is already B-spline. This study is presented from a critical perspective, aiming to contribute new insights to the literature. One of the aspects that has not been sufficiently addressed is the use of functional basis components. It is deemed important to explore this area further and bring it to the attention of the academic community. In the context of functional data analysis, basis components play a crucial role in modeling the relationships among variables. Therefore, a comprehensive examination of this approach may offer significant contributions to the field. Let's take a brief look: we define explanatory variables in general terms as "X" and the response variable "Y" let's denote it by. For these two to be regressive, we need $X = T\xi^t$ and $Y = T\eta^t$ in the form of at least one common T stakeholder. We aim $X \to Y$ is to find the model $\xi^t \to \eta^t$ operation. To be able to capture this operation representation here is the job of the functional basis components. It is considered that the Bernstein operator could be a suitable method for achieving this representation.

ACKNOWLEDGEMENTS

This article is derived from the doctoral dissertation titled 'The Use of the Bernstein Approach in Functional Data Analysis: An Examination of Variance Analysis and Regression,' completed at Fırat University and numbered 870240.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Fisher, R.A., Owen, A.R.G., "An Appreciation of the Life and Work of Sir Ronald Aylmer Fisher: FRS, FSS Sc. D.", Journal of the Royal Statistical Society, Series D (The Statistician), 12(4): 313-319, (1962). DOI: https://doi.org/10.2307/2986951

[2] Cittert-Eymers, V., "CHD Buys Ballot (1817-1890)", Scientiarum Historia: Tijdschrift voor de Geschiedenis van de Wetenschappen en de Geneeskunde, 10(1): 145-153, (1968).

[3] Dozie Kelechukwu, C.N., Christian, C.I., "Decomposition with the Mixed Model in Time Series Analysis using Buys-Ballot Procedure", Asian Journal of Advanced Research and Reports, 17(2): 8-18, (2023). DOI: 10.9734/ajarr/2023/v17i2465

[4] Iwueze, I.S., Nwogu. E.C., "Buys–Ballot Estimates for time series decomposition", Global Journal of Mathematical Sciences, 3(2): 83-98, (2004). DOI: 10.4314/gjmas.v3i2.21356

[5] Iwueza, I.S., and Ohakwe, J., "Buys-Ballot estimates when stochastic trend is quadratic", Journal of the Nigerian Association of Mathematical Physics, 8: 311-318, (2004). DOI: https://doi.org/10.4314/jonamp.v8i1.40020

[6] Doob Joseph, L., "What is a stochastic process?", The American Mathematical Monthly, 49(10): 648-653, (1942). DOI: https://doi.org/10.1080/00029890.1942.11991300

[7] Węglarczyk, S., "Kernel density estimation and its application", ITM Web of Conferences 23, EDP Sciences, (2018). DOI: https://doi.org/10.1051/itmconf/20182300037

[8] Leva, J.L., "A fast normal random number generator", ACM Transactions on Mathematical Software (TOMS), 18(4): 449-453, (1992). DOI: https://doi.org/10.1145/138351.138364

[9] Joy, K.I., "On-line geometric modeling notes", Computer Science Department, University of California, Davis. http://graphics. cs. UC Davis. Edu/CAGDNotes, (1996).

[10] Lorentz, G.G., "Bernstein polynomials", American Mathematical Society, (2012).

[11] Gürcan, M., Colak, C., Orman. M.N., "Bernstein polynomial approach against to some frequently used growth curve models on animal data", Pakistan Journal of Statistics, 26(3): 509-516, (2010).

[12] Gürcan, M., Colak, C., "Generalization of korovkin type approximation by appropriate random variables & moments and an application in medicine", Pakistan Journal of Statistics, 27(3): 283-297, (2011).

[13] Totik, V., "Approximation by Bernstein polynomials", American Journal of Mathematics, 116(4): 995-1018, (1994). DOI: https://doi.org/10.2307/2375007

[14] Güral, Y., Demirelli, A., Gürcan, M., "Gölgelerin oyunu: İzdüşümlerin istatistiksel çıkarsamaları ve türkiye'de döviz kurlarını etkileyen makroekonomik göstergeler üzerine bir uygulama", Avrupa Bilim ve Teknoloji Dergisi, 38: 341-351, (2022). DOI: https://doi.org/10.31590/ejosat.1039913

[15] Aitken, A.C., "IV.—On least squares and linear combination of observations", Proceedings of the Royal Society of Edinburgh, 55: 42-48, (1936). DOI: https://doi.org/10.1017/S0370164600014346

[16] Markov, A.A., "Wahrscheinlichkeitsrechnung", Leipzig [1287], (1912).

[17] Hansen, B.E., "A modern Gauss–Markov theorem", Econometrica, 90(3): 1283-1294, (2022). DOI: https://doi.org/10.3982/ECTA19255

[18] Gasser, T., Müller, H.G., "Kernel estimation of regression functions", Smoothing Techniques for Curve Estimation: Proceedings of a Workshop held in Heidelberg, April 2–4, 1979, Springer Berlin Heidelberg, (1979). DOI: https://doi.org/10.1007/BFb0098489

[19] Hill, P.D., "Kernel estimation of a distribution function", Communications in Statistics-Theory and Methods, 14(3): 605-620, (1985). DOI: https://doi.org/10.1080/03610928508828937

[20] Gürcan, M., Çalık, S., "Generalized convergence characterizations of Feller operatör", Pakistan Journal of Statistics, 27(3): 213-219, (2011).

[21] Rao, C.R., "Estimation of parameters in a linear model", The Annals of Statistics, 4(6): 1023-1037, (1976). DOI: 10.1214/aos/1176343639

[22] Müller, H.G., Stadtmüller, U., "Generalized functional linear models", The Annals of Statistics, 33: 774-805, (2005). DOI: 10.1214/009053604000001156

[23] Cardot, H., Ferraty, F., Sarda, P., "Spline estimators for the functional linear model", Statistica Sinica, 13: 571-591, (2003). DOI: https://doi.org/10.1016/S0167-7152(99)00036-X

[24] Kwak, N., Choi, C.H., "Input feature selection by mutual information based on Parzen window", IEEE Transactions On Pattern Analysis and Machine Intelligence, 24(12): 1667-1671, (2002). DOI: https://doi.org/10.1109/TPAMI.2002.1114861

[25] Babich, G.A., Camps, O.I., "Weighted Parzen windows for pattern classification", IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(5): 567-570, (1996). DOI: https://doi.org/10.1109/34.494647

[26] Vincent, P., Bengio, Y., "Manifold parzen Windows", Advances in Neural Information Processing Systems, 15: (2002).

[27] Köse, H., Gürcan, M., "Zamana bağlı gözlem serilerinin tarihsel başlangıcı ve Buys Ballot'un incelemeleri", Selçuk Üniversitesi Fen Fakültesi Fen Dergisi, 1(18): 33-38, (2001).

[28] Ramsay, J.O., Dalzell, C.J., "Some tools for functional data analysis", Journal of the Royal Statistical Society Series B: Statistical Methodology, 53(3): 539-561, (1991). DOI: https://doi.org/10.1111/j.2517-6161.1991.tb01844.x

[29] Ullah, S., Caroline, F.F., "Applications of functional data analysis: A systematic review", BMC Medical Research Methodology, 13(2013): 1-12. DOI: https://doi.org/10.1186/1471-2288-13-43

[30] Zhu, H., Li, S., Shen, D., Guo, Y., Liu, T., and Chou, Y.C., "FRATS: Functional regression analysis of DTI tract statistics", IEEE Transactions on Medical Imaging, 29(4): 1039-1049, (2010). DOI: https://doi.org/10.1109/TMI.2010.2040625

[31] Wu, P.S., Müller, H.G., "Functional embedding for the classification of gene expression profiles", Bioinformatics, 26(4): 509-517, (2010). DOI: https://doi.org/10.1093/bioinformatics/btp711

[32] Kim, S.B., Rattakorn, P., Peng, Y.B., "An effective clustering procedure of neuronal response profiles in graded thermal stimulation", Expert Systems with Applications, 37(8): 5818-5826, (2010). DOI: https://doi.org/10.1016/j.eswa.2010.02.025

[33] Hyndman, R.J., Shang, H.L., "Rainbow plots, bagplots, and boxplots for functional data", Journal of Computational and Graphical Statistics, 19(1): 29-45, (2010). DOI: https://doi.org/10.1198/jcgs.2009.08158

[34] Erbas, B., Sato, T., Yamaguchi, M., Inoue, K., and Fujita, M., "Using functional data analysis models to estimate future time trends in age-specific breast cancer mortality for the United States and England–Wales", Journal of Epidemiology, 20(2): 159-165, (2010). DOI: https://doi.org/10.2188/jea.JE20090072

[35] Ghosal, R., Chen, Y., Zhou, Y., and Wang, X., "Shape-constrained estimation in functional regression with Bernstein polynomials", Computational Statistics & Data Analysis, 178: 107614, (2023). DOI: https://doi.org/10.1016/j.csda.2022.107614

[36] Bellucci, M.A., "On the explicit representation of orthonormal Bernstein polynomials", arXiv preprint arXiv:1404.2293, (2014). DOI: https://doi.org/10.48550/arXiv.1404.2293