RESEARCH ARTICLE • OPEN ACCESS

A Comparative Study of Linear and Quadratic Spline Regression Models for Predicting HbA1c Levels in Patients with Diabetes Mellitus

Samsul Arifin, Dewi Anggraini, and Nur Salam

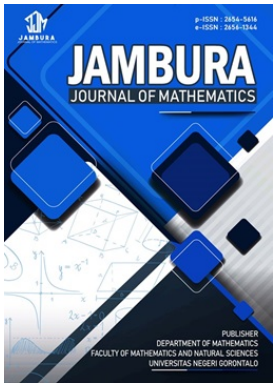
JAMBURA JOURNAL OF MATHEMATICS

Volume 7, Issue 2, Pages 183–188, August 2025

Received 7 July 2025, Revised 25 August 2025, Accepted 27 August 2025, Published 29 August 2025

To Cite this Article : S. Arifin, D. Anggraini, and N. Salam,"A Comparative Study of Linear and Quadratic Spline Regression Models for Predicting HbA1c Levels in Patients with Diabetes Mellitus", *Jambura J. Math*, vol. 7, no. 2, pp. 183–188, 2025, *https://doi.org/10.37905/jjom.v7i2.33292*

JOURNAL INFO • JAMBURA JOURNAL OF MATHEMATICS



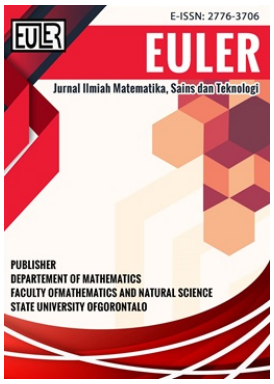
Homepage	:	http://ejurnal.ung.ac.id/index.php/jjom/index
Journal Abbreviation	:	Jambura J. Math.
Frequency	:	Biannual (February and August)
Publication Language	:	English (preferable), Indonesia
DOI	:	https://doi.org/10.37905/jjom
Online ISSN	:	2656-1344
Editor-in-Chief	:	Hasan S. Panigoro
Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
Country	:	Indonesia
OAI Address	:	http://ejurnal.ung.ac.id/index.php/jjom/oai
Google Scholar ID	:	iWLjgaUAAAAJ
Email	:	info.jjom@ung.ac.id

JAMBURA JOURNAL • FIND OUR OTHER JOURNALS

Jambura Journal of Biomathematics

Jambura Journal of Mathematics Education

Jambura Journal of Probability and Statistics



EULER : Jurnal Ilmiah Matematika, Sains, dan Teknologi

Jambura Journal of Mathematics, Volume 7, Issue 2, Pages 183–188, August 2025 https://doi.org/10.37905/jjom.v7i2.33292

Research Article

Check for updates

A Comparative Study of Linear and Quadratic Spline Regression Models for Predicting HbA1c Levels in Patients with Diabetes Mellitus

Samsul Arifin[1], **Dewi Anggraini**[1], **Nur Salam**[1]

[1]*Department of Statistics, Universitas Lambung Mangkurat, Banjarbaru 70714, Indonesia*

ARTICLE HISTORY
Received 7 July 2025
Revised 25 August 2025
Accepted 27 August 2025
Published 29 August 2025

KEYWORDS
Spline Regression
HbA1c
Fasting Blood Glucose
LDL Cholesterol

ABSTRACT. *HbA1c is widely recognized as a key clinical indicator for monitoring and controlling diabetes, as it reflects average blood glucose levels over the preceding 2–3 months and is closely linked to the risk of complications. This study compares linear and quadratic truncated spline regression models for predicting HbA1c levels in patients with diabetes mellitus. The analysis used retrospective medical record data from a hospital in Makassar, Indonesia, collected in 2023. Fasting blood glucose and LDL cholesterol were included as predictors, with HbA1c as the response variable. Truncated spline regression was applied to capture nonlinear associations between predictors and HbA1c, and the comparison focused on linear versus quadratic specifications. The selection of the best model was based on the minimum GCV value. The model selection process indicated that the best specification was the linear truncated spline regression with two knot points. For FBG, the optimal knots were located at 159 mg/dL and 368 mg/dL, yielding the lowest GCV value of 3.5798. For LDL cholesterol, the best fit was achieved with knots at 183 mg/dL and 191 mg/dL, resulting in a GCV value of 4.3325. The predictive performance of this model was further supported by an R^2 value of 0.3861, indicating that the linear spline with two knots provides a better fit compared with the quadratic spline alternative. The spline approach showed a better fit based on GCV in depicting the changes in the influence of predictors on HbA1c, suggesting its potential as a more accurate predictive model for clinical and epidemiological purposes.*

1. Introduction

Diabetes Mellitus (DM) is a chronic disease characterised by persistently high blood glucose levels [1]. This occurs because the body is unable to produce enough insulin or cannot use insulin effectively [2]. The prevalence of DM in Indonesia continues to increase each year [3]. In Makassar, the number of DM cases is relatively high and has become a significant concern in various healthcare facilities. One of the key indicators for monitoring and controlling this disease is the HbA1c level, which is a measure of glycated hemoglobin [4]. HbA1c reflects the average blood glucose level over the past 2–3 months and serves as the primary reference for assessing diabetes control in patients [5].

In understanding the factors that influence HbA1c levels or in predicting HbA1c levels, parametric approaches, such as linear regression, are still commonly used [6]. However, the relationship between HbA1c levels and variables such as body mass index, fasting blood glucose, and low-density lipoprotein (LDL) cholesterol is often nonlinear [7]. When the relationships between variables do not follow a linear pattern, the use of conventional linear regression models can yield inaccurate estimates or fail to accurately represent the data structure. Therefore, nonparametric approaches such as spline regression offer a more flexible alternative, as they can model relationships in the form of curve segments [8]. Using splines, nonlinear or varying relationships across the data range can be captured more effectively [9].

Several previous studies have shown that spline regression outperforms ordinary linear regression in modelling health data, particularly when the relationships between variables are nonlinear [10, 11]. Lu et al. [12] used a nonparametric approach to predict HbA1c levels. Islamiyati et al. [13] also used a nonparametric bivariate spline approach to examine factors influencing HbA1c in DM patients. In Indonesia, the application of spline regression is still limited, and few studies specifically compare linear splines and quadratic splines, particularly in the context of predicting HbA1c in DM patients.

The data for this study were obtained from the 2023 medical records of patients with diabetes mellitus in hospitals in Makassar, Indonesia. Based on this background, the study aims to compare linear and quadratic spline regression models in predicting HbA1c levels among diabetic patients. The results of this study will provide a clearer understanding of which model is more accurate and suitable for clinical use. They will also serve as a reference for further research in the fields of epidemiology and health statistics.

2. Methods

This study is a retrospective study conducted at a hospital in Makassar in 2023 using medical record data from 87 patients with diabetes mellitus. The response variable (y) is the

*Corresponding Author.

Email : *samsularrr@ulm.ac.id* (S. Arifin)
Homepage : http://ejurnal.ung.ac.id/index.php/jjom/index / E-ISSN : 2656-1344

HbA1c level, while the predictor variables include FBG (x_1) and LDL cholesterol (x_2). The modeling approach applied is a simultaneous additive regression model. To capture potential nonlinearity, the truncated spline regression method was employed to compare linear and quadratic basis functions. The selection of knot points was carried out numerically and evaluated using the GCV method. The number of knots was restricted to one or two, mainly due to computational limitations and to maintain model parsimony. The choice of truncated spline over other spline approaches was motivated by its computational simplicity, interpretability, and suitability for the moderate sample size used in this study. Data analysis was performed using RStudio software version 2024.12.0.

2.1. Parametric Regression

Parametric regression analysis is used when the form of the relationship between the response variable and the predictor variables is known beforehand, such as linear, quadratic, or cubic relationships [14]. If the assumptions about the form of this parametric model are correct, the parameter estimates generated will be highly efficient. In general, the parametric linear regression model can be expressed as follows in eq. (1):

$$y_i = \vartheta_0 + \sum_{j=1}^{p} \vartheta_j x_{ji} + \varepsilon_i, \tag{1}$$

where y_i is the response variable, x_{ji} is the predictor variable, ϑ_0 is the intercept, ϑ_j is the regression parameter, and ε_i is the residual, assumed to be independent and normally distributed. Eq. (1) can be expressed in matrix form as shown in eq. (2):

$$y = \boldsymbol{X}\vartheta + \varepsilon, \tag{2}$$

where:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{bmatrix};$$

$$\vartheta = \begin{bmatrix} \vartheta_0 \\ \vartheta_1 \\ \vdots \\ \vartheta_p \end{bmatrix}; \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

The parameter ϑ can be estimated using the ordinary least squares (OLS) method by minimising the sum of squared errors

$$\begin{aligned} \sum_{i=1}^{n} \varepsilon^2 &= \varepsilon'\varepsilon \\ &= (y - \boldsymbol{X}\vartheta)'(y - \boldsymbol{X}\vartheta) \\ &= (y' - \boldsymbol{X}'\vartheta')(y - \boldsymbol{X}\vartheta) \\ &= y'y - y'\boldsymbol{X}\vartheta - \vartheta'\boldsymbol{X}'y + \vartheta'\boldsymbol{X}'\boldsymbol{X}\vartheta \\ &= y'y - 2\vartheta'\boldsymbol{X}'y + \vartheta'\boldsymbol{X}'\boldsymbol{X}\vartheta. \end{aligned} \tag{3}$$

Since $y'\boldsymbol{X}\vartheta$ Is a scalar, its transpose matrix is $(y'\boldsymbol{X}\vartheta)' = \vartheta'\boldsymbol{X}'y$. Next, the sum of squared errors is differentiated concerning ϑ. Then the derivative is set equal to zero to obtain the estimated result of ϑ As shown in eq. (4):

$$\begin{aligned} \left.\frac{\partial \varepsilon'\varepsilon}{\partial \vartheta}\right|_{\vartheta=\widehat{\vartheta}} &= \frac{\partial[y'y - 2\vartheta'\boldsymbol{X}'y + \vartheta'\boldsymbol{X}'\boldsymbol{X}\vartheta]}{\partial\vartheta} \\ 0 - 2\boldsymbol{X}'y + 2\boldsymbol{X}'\boldsymbol{X}\widehat{\vartheta} &= 0 \\ -2\boldsymbol{X}'y + 2\boldsymbol{X}'\boldsymbol{X}\widehat{\vartheta} &= 0 \\ 2\boldsymbol{X}'\boldsymbol{X}\widehat{\vartheta} &= 2\boldsymbol{X}'y \\ \boldsymbol{X}'\boldsymbol{X}\widehat{\vartheta} &= \boldsymbol{X}'y \\ \widehat{\vartheta} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'y, \end{aligned} \tag{4}$$

where $\widehat{\vartheta} = [\widehat{\vartheta}_0, \widehat{\vartheta}_1, \widehat{\vartheta}_2, \dots, \widehat{\vartheta}_p]'$.

2.2. Nonparametric Regression

Nonparametric regression is an alternative method that can be used when the form of the relationship between the response variable and the predictor variables is not known precisely [15]. Unlike parametric regression, this approach does not require specific assumptions about the form of the regression function, making it more flexible in capturing complex data patterns. Nonparametric regression with p predictors involving n observations can be modelled in the following eq. (5):

$$y_i = f\,(x_{1i}, x_{2i},\ \dots,\ x_{pi}) + \varepsilon_i;\ i = 1, 2, \dots\ , n, \tag{5}$$

where f is the nonparametric regression function. The form of the nonparametric regression function in eq. (5) is not known, and it is only assumed to be contained within a specific function space. The choice of function space is based on the smoothness properties of the function, f. One approach to nonparametric regression is spline regression. Splines have the advantage of handling data patterns that exhibit upward or downward trends, thanks to the use of knot points, resulting in a relatively smooth curve.

2.3. Spline Regression

Spline is a segmented polynomial known for its flexible properties. This main characteristic distinguishes splines from regular polynomials, as they can adapt the curve shape more effectively to variations in data patterns or specific functions. In general, a spline function of order q with knot points at $k_1, k_2,\ \dots, k_r$ can be expressed in the following eq. (6):

$$f\,(x_i) = \sum_{l=0}^{q} \beta_l x_i^l + \sum_{h=1}^{r} \beta_{q+h}(x_i - k_h)_+^q, \tag{6}$$

where $\vartheta_0, \vartheta_1, \dots, \vartheta_q, \vartheta_{q+1}, \dots,\ \vartheta_{q+r}$ are the regression parameters, k_h is the h-th knot point, and $(x_i - k_h)_+^q$ is the polynomial function expressed as shown in eq. (7):

$$(x_i - k_h)_+^q = \begin{cases} (x_i - k_h)^q, & \text{if } x_i \geq k_h, \\ 0, & \text{if } x_i < k_h. \end{cases} \tag{7}$$

If in eq. (6), the value of q is substituted with 1, 2, or 3, the resulting spline functions are called linear spline, quadratic spline, and cubic spline, respectively.

The function that represents the relationship between the p-th predictor and a single response, when approximated with the spline function $f(x_i)$ in eq. (6), can be expressed in the following eq. (8):

$$\begin{aligned} y_i &= f(x_{1i}) + f(x_{2i}) + \cdots + f(x_{pi}) + \varepsilon_i \\ &= \sum_{j=1}^{p} f(x_{ji}) + \varepsilon_i; i = 1, 2, \ldots, n, \end{aligned} \tag{8}$$

where:

$$f(x_{ji}) = \vartheta_{0j} + \sum_{l=1}^{q} \vartheta_{jl} x_{ji}^{l} + \sum_{h=1}^{r} \vartheta_{j(q+h)} (x_{ji} - k_{jh})_{+}^{q};$$
$$j = 1, 2, \ldots, p,$$

k_{jh} is the value of the knot point, r is the number of knot points, q is the order of the spline polynomial, and p is the number of predictor variables. Eq. (8) can be expressed in matrix form as shown in eq. (9):

$$y = X\boldsymbol{\vartheta} + \varepsilon, \tag{9}$$

and

$$y = (y_1, y_2, \ldots, y_n)',$$
$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{p1}^{q} & (x_{11} - k_{11})_{+}^{q} & \cdots & (x_{p1} - k_{pr})_{+}^{q} \\ 1 & x_{12} & \cdots & x_{p2}^{q} & (x_{12} - k_{11})_{+}^{q} & \cdots & (x_{p2} - k_{pr})_{+}^{q} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{pn}^{q} & (x_{1n} - k_{11})_{+}^{q} & \cdots & (x_{pn} - k_{pr})_{+}^{q} \end{bmatrix},$$
$$\vartheta = (\vartheta_{01}, \vartheta_{p1}, \ldots, \vartheta_{pq}, \vartheta_{1(q+1)}, \ldots, \vartheta_{p(q+r)})',$$
$$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'.$$

2.4. Knot Point Selection

The selection of the optimal knot points can be done using the Generalized Cross Validation (GCV) method. GCV theoretically has asymptotic optimality properties. The GCV method also has the advantage of not requiring knowledge of the population variance, and it is invariant to transformations. The GCV method is typically expressed in eq. (10):

$$\text{GCV}(k_h) = \frac{MSE(k_h)}{\left[n^{-1} trace(\boldsymbol{I} - \boldsymbol{A}(k_h))\right]^2}, \tag{10}$$

where $MSE(k_h) = n^{-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$, k is the knot point, the matrix $\boldsymbol{A}$ is $X(X^tX)^{-1}X^t$, and $\boldsymbol{I}$ is the identity matrix. The minimum GCV value gives the optimal knot point value.

2.5. Diabetes Mellitus

DM is a chronic metabolic disorder characterized by hyperglycemia due to abnormalities in insulin secretion, insulin action, or both [16]. This disease has the potential to cause various serious complications, both microvascular, such as nephropathy and retinopathy, and macrovascular, such as coronary heart disease. Therefore, monitoring specific biochemical parameters is essential in the management of DM. One of the most commonly used parameters is HbA1c levels, which reflect the average blood glucose level over the past 8–12 weeks [17]. HbA1c is considered a key indicator for evaluating long-term glycemic control and is closely related to the risk of chronic complications in diabetic patients.

One of the factors that influences HbA1c values is the blood glucose level measured after fasting for at least eight hours [18]. Fasting blood glucose provides a direct indication of the glycemic status at the time and empirically shows a positive correlation with HbA1c, particularly in patients with unstable glucose control. In other words, high fasting blood glucose over time can lead to the accumulation of glucose that binds to hemoglobin, thereby increasing the HbA1c value [19]. Therefore, fasting blood glucose is not only important as a screening tool and for early diagnosis but also relevant for monitoring the success of therapy periodically.

LDL cholesterol also plays a crucial role in the pathophysiology of DM, particularly in indirectly impairing glycemic control [20]. LDL cholesterol is commonly referred to as "bad cholesterol" because it contributes to the formation of atherosclerotic plaques, which can impair vascular function. Several studies have shown that high LDL cholesterol is associated with increased insulin resistance and systemic inflammation, which can ultimately lead to worsened glucose control and higher HbA1c levels. Thus, HbA1c, fasting blood glucose, and LDL cholesterol levels are three interrelated indicators that are important to analyse in understanding the metabolic condition of diabetic patients.

3. Results and Discussion

3.1. Scatterplot

The relationship pattern between the response variable and the predictor variables is examined to understand the characteristics of the connection between them. One of the simplest ways to evaluate this relationship pattern is through visualisation using a scatterplot. A scatterplot displays the relationship between two different variables and allows for the identification of any patterns that may form, such as linear, quadratic, or other forms. If the scatterplot reveals a clear pattern, a parametric approach can be used in modelling. On the other hand, if no structured pattern is found, a nonparametric approach is more appropriate to apply. The scatterplot between HbA1c levels and the predictor variables is presented in Figure 1.

Based on the scatterplot between HbA1c levels and the predictor variables, no clear pattern, such as a straight line or specific curve, is observed. This irregularity is likely due to individual characteristic differences and the influence of other variables that have not yet been analysed. Therefore, the relationship between fasting blood glucose and LDL cholesterol with HbA1c levels is not suitable for modelling using conventional parametric regression and is more appropriately analysed using a nonparametric approach [13].

3.2. Selection of Optimal Knot Points

A commonly used approach to determine the optimal knot points is the GCV, where the smallest GCV value indicates the best model with the minimum prediction error. In this study, the selection of knot points is limited to one or two. The GCV values obtained for the nonparametric spline regression model on the fasting blood glucose variable are presented in Table 1.

Based on Table 1, the nonparametric Spline regression model with a linear order and two knot points at 159 and 368

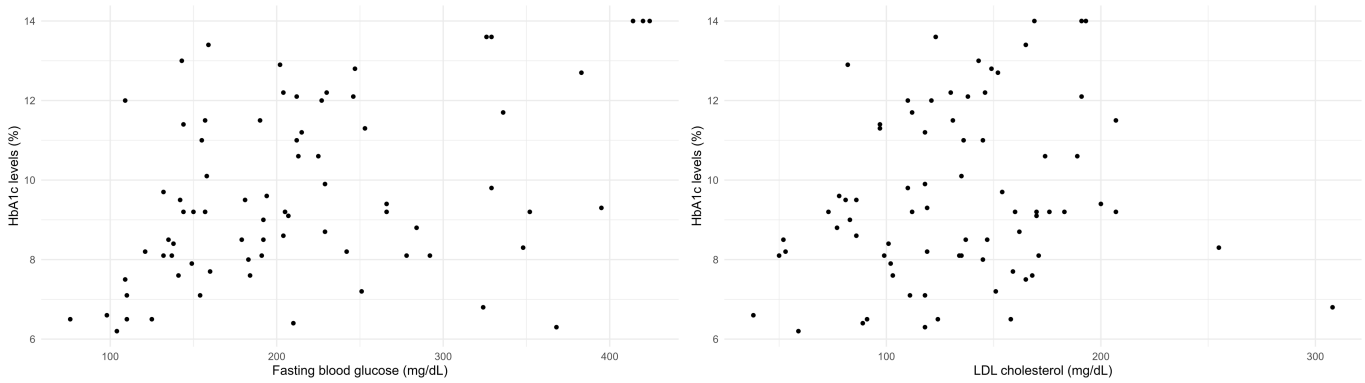


Figure 1. Scatterplot of HbA1c levels and predictor variables

Table 1. Knot selection for the fasting blood glucose variable

Order	Knot Selection			
	1 Knot	GCV	2 Knot	GCV
Linear	**395**	**3.8413**	**159; 368**	**3.5797**
	368	3.8861	158; 368	3.5799
	383	3.8862	157; 368	3.5808
	143	3.9742	157; 395	3.5808
	144	3.9746	158; 395	3.5810
	352	3.9757	159; 395	3.5819
	142	3.9782	155; 395	3.5827
	⋮	⋮	⋮	⋮
	253	4.1666	98; 253	4.2422
Quadratic	**383**	**3.9672**	**98; 352**	**3.6352**
	368	3.9677	104; 352	3.6365
	395	3.9820	109; 352	3.6384
	352	3.9893	98; 348	3.6386
	348	3.9983	110; 352	3.6389
	336	4.0317	104; 348	3.6399
	329	4.0524	109; 348	3.6418
	⋮	⋮	⋮	⋮
	205	4.2095	225; 420	4.2824

Table 2. Knot selection for the LDL cholesterol variable

Order	Knot Selection			
	1 Knot	GCV	2 Knot	GCV
Linear	**193**	**4.4482**	**183; 191**	**4.3325**
	191	4.4535	183; 189	4.3561
	200	4.4716	176; 191	4.3755
	189	4.4820	183; 193	4.3767
	207	4.4881	174; 191	4.3943
	183	4.5648	176; 193	4.3967
	176	4.6441	176; 189	4.3993
	⋮	⋮	⋮	⋮
	255	4.6473	174; 193	4.4114
Quadratic	**77**	**4.5473**	**191; 200**	**4.5989**
	78	4.5480	193; 200	4.6018
	81	4.5500	189; 200	4.6021
	82	4.5507	189; 193	4.6053
	83	4.5513	189; 207	4.6100
	86	4.5526	191; 207	4.6103
	89	4.5532	193; 207	4.6148
	⋮	⋮	⋮	⋮
	91	4.5533	189; 191	4.6170

was selected as the best model for the variable x_1, which is fasting blood glucose. This selection is based on the smallest GCV value of 3.5797, indicating that this model has the lowest prediction error compared to other knot point combinations. A lower GCV value indicates that the linear order model with two knot points can represent the relationship between fasting blood glucose and HbA1c levels more optimally. Although the quadratic order model was also evaluated, its lowest GCV value of 3.6352 at knot points 98 and 352 is still higher than that of the best linear model. Therefore, the approach using the linear order Spline with two knot points is considered the most appropriate for modelling the relationship between fasting blood glucose and HbA1c levels.

Based on Table 2, the nonparametric Truncated Spline regression model with a linear order and two knot points at 183 and 191 was selected as the best model for the variable x_2, which is LDL cholesterol. This selection is based on the smallest GCV value of 4.3325, indicating that this model has the lowest prediction error compared to other knot point combinations. A lower GCV value indicates that the linear order model with two knot points can represent the relationship between LDL cholesterol and HbA1c levels more optimally. Although the quadratic order model was also evaluated, its lowest GCV value of 4.5989 at knot points 191 and 200 is still higher than that of the best linear model. Therefore, the approach using the Truncated Spline with a linear order and two knot points is considered the most appropriate for modelling the relationship between LDL cholesterol and HbA1c levels.

3.3. *Estimation of the Truncated Spline Model*

After selecting the model based on the GCV value, it was found that the best model for modelling the relationship between the predictor variables, and the response variable is the nonparametric linear Spline regression model with two knot points for the variables GDP and LDL cholesterol. This model was chosen because it has the lowest GCV value among the other models evaluated. The estimated parameter results of the model are shown in the following Equation:

$$y = 2.7625 + 0.0449x_1 - 0.0461\,(x_1 - 159) + 0.0826\,(x_1 - 368) + 0.0123x_2 + 0.2407\,(x_2 - 183) - 0.3061(x_2 - 191).$$

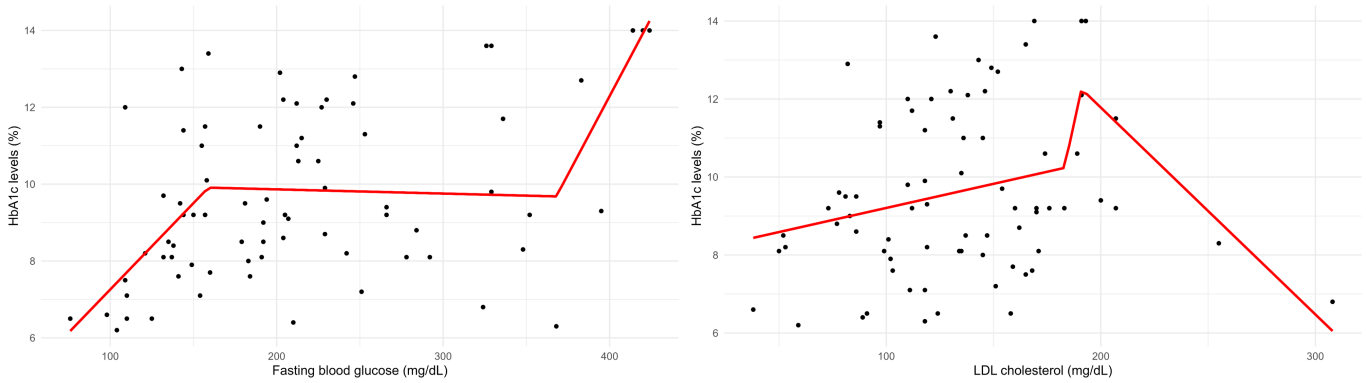


Figure 2. Estimation of the linear truncated spline curve with 2 knot points

Table 3. Hypothesis testing for predictor variables

Variable	Parameter	Coefficient	SE	t-Value	Sig.	R-Squared
Fasting blood glucose (x_1)	β_{01}	2.7626	1.9392	1.4246	0.1585	0.3861
	β_{11}	0.0449	0.0135	3.3307	0.0014*	
	β_{12}	-0.0460	0.0158	-2.9191	0.0046*	
	β_{13}	0.0826	0.0248	3.3323	0.0013*	
LDL cholesterol (x_2)	β_{02}	7.9692	0.8859	8.9960	0.0000*	
	β_{21}	0.0124	0.0069	1.7893	0.0777	
	β_{22}	0.2407	0.1184	2.0337	0.0456*	
	β_{23}	-0.3061	0.1241	-2.4676	0.0159*	

Based on the parameter estimation results from the linear order truncated spline model with two knot points, it was found that the relationship between fasting blood glucose and HbA1c levels experienced a change in slope at two knot points, namely 159 and 368. Before the first knot point ($x_1<159$), each increase in fasting blood glucose of 1 mg/dL was associated with a 0.0449 unit increase in HbA1c levels. After the point of 159, this effect decreased by 0.0461, causing the slope to change to a negative value of -0.0012. However, after the second knot point ($x_1>368$), the effect increased again by 0.0826, making the slope positive at 0.0814. This pattern shows that the relationship between X_1 and Y is not a constant linear relationship, but rather changes at specific value ranges.

A similar pattern was found in the relationship between LDL cholesterol and HbA1c levels. At $x_2<183$, every 1 mg/dL increase in LDL cholesterol was associated with a 0.0123 unit increase in HbA1c. After the first knot point (183), the effect increased by 0.2407 to 0.2530. However, after crossing the second knot point (191), the effect decreased by 0.3061, causing the slope to become negative at -0.0531. This indicates that the effect of LDL on HbA1c increases sharply in the range of 183–191, but then decreases once the LDL level exceeds 191 mg/dL. These patterns are visualised in the regression curve shown in Figure 2.

Based on Figure 2, the curve represents the relationship between fasting blood glucose, LDL cholesterol levels, and HbA1c levels. For fasting blood glucose, it forms linear segments with a change in the slope direction at the knot points 159 and 368. Meanwhile, for LDL cholesterol, the curve changes curvature at points 183 and 191. This visualisation highlights that the truncated spline approach is effective in capturing the dynamics of the nonlinear relationship globally, while remaining linear in specific local segments.

3.4. *Hypothesis Testing*

The purpose of hypothesis testing is to determine whether each predictor variable, including its spline components, significantly contributes to the response variable. The test is performed using a t-test for each regression coefficient, comparing the calculated t-value against the t-distribution with a specific degree of freedom, or by examining the resulting p-value. If the p-value is smaller than the significance level ($\alpha = 0.05$), the parameter is considered statistically significant. The results of the hypothesis testing are shown in Table 3.

Based on Table 3, it can be seen that most parameters have a p-value<0.05, which means they are statistically significant. For variable x_1, three parameters, β_{11}, β_{12}, and β_{13}, are significant, indicating that both the linear relationship and the change in the pattern after the knot points have an impact on HbA1c levels. However, the constant parameter β_{01} is not significant, meaning it has a negligible effect on the model. For variable x_2, the constant parameter β_{02} and two spline components (β_{22} and β_{23}) are also significant. Meanwhile, the linear parameter β_{21} has a p-value slightly above 0.05 (0.0777), indicating that it is not significant at the 5% level, but it can still be considered in the interpretation. These results suggest that the relationship between GDP and LDL cholesterol, as well as HbA1c, does not follow a specific pattern; thus, using a spline model provides more accurate results.

4. Conclusion

Based on the analysis results, the relationship between HbA1c levels and fasting blood glucose, as well as LDL cholesterol, does not form a clear linear pattern. Therefore, nonparametric approaches such as spline regression are more suitable than ordinary linear regression. The best model obtained is the

truncated spline regression model of linear order with two knot points. For GDP, the optimal knot points are at values 159 and 368, while for LDL cholesterol, the knot points are at 183 and 191. This selection is based on the lowest GCV value, which indicates the smallest prediction error. The estimation results and hypothesis tests show that most of the model parameters are statistically significant. The coefficient of determination (R^2) of the final model is 0.3861, indicating that approximately 38.61% of the variation in HbA1c levels can be explained by fasting blood glucose and LDL cholesterol through the spline regression model. This suggests that both GDP and LDL cholesterol have an impact on HbA1c levels, with the relationship pattern shifting at specific points. The spline approach proves to be effective in describing the nonlinear relationship between variables.

Author Contributions. **S. Arifin**: Conceptualisation, methodology, formal analysis, writing, original draft preparation, visualisation, and software. **D. Anggraini**: Supervision, editing, validation, and project administration. **N. Salam**: Editing, methodology, and validation. All authors have read and agreed to the published version of the manuscript.

Acknowledgement. The authors would like to express their gratitude to all parties who have contributed to this research. We greatly appreciate the editor and reviewers for their valuable feedback and support in improving this work.

Funding. This research did not receive any external funding.

Conflict of interest. The authors declare that there are no conflicts of interest related to this article.

Data availability. Not applicable.

References

[1] D. Purnamasari, E. N. Tetrasiwi, G. J. Kartiko, C. Astrella, K. Husam, and P. W. Laksmi, "Sarcopenia and chronic complications of type 2 diabetes mellitus," *Rev. Diabet. Stud.*, vol. 18, no. 3, pp. 157–165, 2022, doi: 10.1900/RDS.2022.18.157.

[2] J. Burillo, P. Marqués, B. Jiménez, C. González-Blanco, M. Benito, and C. Guillén, "Insulin resistance and diabetes mellitus in Alzheimer's disease," *Cells*, vol. 10, no. 5, 2021, doi: 10.3390/cells10051236.

[3] S. I. Oktora and D. B. Butar, "Determinants of diabetes mellitus prevalence in Indonesia," *Kemas*, vol. 18, no. 2, pp. 266–273, 2022, doi: 10.15294/kemas.v18i2.31880.

[4] A. Chaib *et al.*, "Effects of nutrition education on metabolic profiles of patients with type 2 diabetes mellitus to improve glycated hemoglobin and body mass index," *J. King Saud Univ. - Sci.*, vol. 35, no. 1, p. 102437, 2023, doi: 10.1016/j.jksus.2022.102437.

[5] N. N. Sreedevi, G. V. Kumar, B. Chendar, M. V. Bhaskar, K. S. S. S. Baba, and I. K. Mohan, "Unexpected detection of abnormal hemoglobin variants during routine HbA1c analysis by HPLC technique: Challenges and opportunities in HbA1c testing," *Avicenna J. Med. Biochem.*, vol. 9, no. 1, pp. 15–21, 2021, doi: 10.34172/ajmb.2021.03.

[6] Z. Zhu, Y. Mao, and G. Chen, "Predictive value of HbA1c for in-hospital adverse prognosis in COVID-19: A systematic review and meta-analysis," *Prim. Care Diabetes*, vol. 15, no. 6, pp. 910–917, 2021, doi: 10.1016/j.pcd.2021.07.013.

[7] G. K. Jakubiak, A. Chwalba, A. Basek, G. Cieslar, and N. Pawlas, "Glycated hemoglobin and cardiovascular disease in patients without diabetes," *J. Clin. Med.*, vol. 14, no. 1, pp. 1–14, 2025, doi: 10.3390/jcm14010053.

[8] J. L. Kirkby, Á. Leitao, and D. Nguyen, "Spline local basis methods for non-parametric density estimation," *Stat. Surv.*, vol. 17, pp. 75–118, 2023, doi: 10.1214/23-SS142.

[9] S. Arifin, A. Islamiyati, and R. Raupong, "Kemampuan estimator spline linear dalam analisis komponen utama," *ESTIMASI J. Stat. Its Appl.*, vol. 1, no. 1, p. 40, 2020, doi: 10.20956/ejsa.v1i1.9262.

[10] N. C. Chesnaye, M. Van Diepen, F. Dekker, C. Zoccali, K. J. Jager, and V. S. Stel, "Non-linear relationships in clinical research," *Nephrol. Dial. Transplant.*, vol. 40, no. 2, pp. 244–254, 2025, doi: 10.1093/ndt/gfae187.

[11] Sifriyani, A. R. M. Sari, A. T. R. Dani, and S. Jalaluddin, "Bi-response truncated spline nonparametric regression with optimal knot point selection using generalized cross-validation in diabetes mellitus patient's blood sugar levels," *Commun. Math. Biol. Neurosci.*, vol. 2023, pp. 1–18, 2023, doi: 10.28919/cmbn/7903.

[12] R. Lu *et al.*, "The application of multivariate adaptive regression splines in exploring the influencing factors and predicting the prevalence of HbA1c improvement," *Ann. Palliat. Med.*, vol. 10, no. 2, pp. 1296–1303, 2021, doi: 10.21037/apm-19-406.

[13] A. Islamiyati, A. Kalondeng, N. Sunusi, M. Zakir, and A. K. Amir, "Biresponse nonparametric regression model in principal component analysis with truncated spline estimator," *J. King Saud Univ. - Sci.*, vol. 34, no. 3, p. 101892, 2022, doi: 10.1016/j.jksus.2022.101892.

[14] E. A. Geissinger, C. L. L. Khoo, I. C. Richmond, S. J. M. Faulkner, and D. C. Schneider, "A case for beta regression in the natural sciences," *Ecosphere*, vol. 13, no. 2, pp. 1–16, 2022, doi: 10.1002/ecs2.3940.

[15] M. Salibian-Barrera, "Robust nonparametric regression: Review and practical considerations," *Econom. Stat.*, In Press, Corrected Proof, Apr. 2023. doi: 10.1016/j.ecosta.2023.04.004.

[16] S. Alam, M. K. Hasan, S. Neaz, N. Hussain, M. F. Hossain, and T. Rahman, "Diabetes mellitus: Insights from epidemiology, biochemistry, risk factors, diagnosis, complications and comprehensive management," *Diabetology*, vol. 2, no. 2, pp. 36–50, 2021, doi: 10.3390/diabetology2020004.

[17] M. Wang and T. Hng, "Standards of medical care for patients with diabetes mellitus," *Diabetes Care*, vol. 12, no. 5, pp. 365–368, 1989, doi: 10.2337/diacare.12.5.365.

[18] K. Sung and S. H. Lee, "Social determinants of health and type 2 diabetes in Asia," *J. Diabetes Investig.*, vol. 16, no. 6, pp. 971–983, 2025, doi: 10.1111/jdi.70024.

[19] K. S. Boye, V. T. Thieu, M. J. Lage, H. Miller, and R. Paczkowski, "The association between sustained HbA1c control and long-term complications among individuals with type 2 diabetes: A retrospective study," *Adv. Ther.*, vol. 39, no. 5, pp. 2208–2221, 2022, doi: 10.1007/s12325-022-02106-4.

[20] I. Bonilha, E. Hajduch, B. Luchiari, W. Nadruz, W. Le Goff, and A. C. Sposito, "The reciprocal relationship between LDL metabolism and type 2 diabetes mellitus," *Metabolites*, vol. 11, no. 12, 2021, doi: 10.3390/metabo11120807.