OPEN ACCESS
Journal of Multidisciplinary Applied Natural Science Vol. 6 No. 1 (2026) https://doi.org/10.47352/jmans.2774-3047.329
Check for updates
Research Article

Modeling of Platelet and Hematocrit in Dengue Hemorrhagic Fever (DHF) Patients Using Semiparametric Bi-response Regression Approach Based on Local Polynomial Estimator for Longitudinal Data

Tiani Wahyu Utami, Nur Chamidah*, Toha Saifudin, Budi Lestari, and Norma Latif Fitriyani

Received : August 22, 2025	Revised : September 30, 2025	Accepted : October 9, 2025	Online : December 22, 2025

Abstract
Dengue hematrocit fever (DHF) is a health problem in Indonesia, which tends to cause an increase in the number of sufferers and is becoming more widespread. Semarang City is an endemic region in Indonesia. Platelet and hematocrit modeling are required to diagnose DHF. The independent variables in this model were hemoglobin (Hb) level and examination time, whereas platelets and hematocrit were the dependent variables. This study used secondary data obtained from the Roemani Hospital in Semarang, Indonesia. The study included 13 patients who met the criteria for Grade 2 DHF, and their blood samples were collected once daily for 6 days during their hospitalization to form longitudinal data. This study aimed to model hematocrit and platelet counts using semiparametric bi-response regression with local polynomial estimators. The GCV method was used to select the optimal combination of bandwidth and polynomial order. The results obtained for platelets were a polynomial order of 2 and a bandwidth of 0.1, while hematocrit was selected with a polynomial order of 1 and a bandwidth of 0.8. Platelet and hematocrit modeling using the semiparametric bi-response regression applied to the in-sample data resulted in a coefficient of determination (R^2) of 90.12%. The model results can be used to predict platelets and hematocrit with high accuracy, yielding an MAPE of 4.84%. Based on the analysis results, the increase in Hb and hematocrit has a unidirectional relationship (both increase) and is in the opposite direction to the number of platelets, which usually decreases (thrombocytopenia). Platelet dynamics in patients with Grade 2 DHF who were hospitalized for 6 days showed that on the 3rd or 4th day, the patient experienced thrombocytopenia and an increase in hematocrit above normal, which is a sign of plasma leakage; therefore, it is necessary to be aware that this patient's condition requires more intensive care to stabilize platelets and hematocrit.

Keywords: DHF, hematocrit, local polynomial estimator, longitudinal data, platelets, semiparametric bi-response regression

1. INTRODUCTION

The dengue virus, spread by the Aedes aegypti mosquito, causes dengue hemorrhagic fever (DHF), an acute febrile illness characterized by the transmission of the virus through human blood and the development of acute fever for 2–7 days [1]. This disease should not be underestimated because it can result in death. According to estimates by the World Health Organization (WHO), one person dies from DHF every 20 min worldwide. DHF often occurs in tropical areas, especially in Indonesia [2]. DHF is an infectious illness with significant morbidity and mortality rates that causes outbreaks in different parts of Indonesia [3]. Semarang City is one of the areas endemic to DHF, with the second-highest number of DHF cases in Indonesia [4][5]. The Semarang City Health Service reported that, as of 2022, the number of DHF cases in Semarang City had reached 857, according to the Head of the Semarang City Health Service. The number of DHF cases in Semarang in 2022 increased by two to three times compared to that in 2021 [6].

According to the WHO, DHF is characterized by the presence of thrombocytopenia, specifically a platelet count below 100000/μL [7]. Platelet values can help describe the course of DHF because the average measurement is performed once daily while the patient is hospitalized and treated until they are cured [8]. In addition to examining platelet count, other tests are necessary, including the evaluation of hematocrit levels. Based on previous studies, there is a significant relationship related to the interaction between hematocrit and platelet count [9][10]. The four grades are always accompanied by laboratory parameters of platelets <100000/μL, and there is evidence of plasma leakage (increased hematocrit) [11]. Hemoglobin (Hb) plays a crucial role in diagnosing DHF. In the early stages, Hb levels are

Publisher's Note:
Pandawa Institute stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Table 1. Research variables.

Variable	Operational Definition
$y_{ij}^{(1)}$	Platelets in subject-*i* observation-*j*
$y_{ij}^{(2)}$	Hematocrit in subject-*i* observation-*j*
t_{ij}	Time of examination in subject-*i* observation-*j*
z_{ij}	Hemoglobin in subject-*i* observation-*j*

Table 2. MAPE criteria.

MAPE	Accurate Criteria
< 10%	Highly
10%–20%	Good
20%–50%	Reasonable
> 50%	Weak

Table 3. Characteristics of Grade 2 DHF patients.

Gender	Patient
Adult Male	Patient 8, Patient 9, Patient 11
Adult Female	Patient 2, Patient 5, Patient 6, Patient 10
Children	Patient 1, Patient 3, Patient 4, Patient 7, Patient 12, Patient 13

usually normal; however, over time, they increase, followed by an increase in hemoconcentration. Fluids that enter the extravascular space and protein leakage will have an impact on increasing Hb levels, which in turn increases hematocrit [12]. Hematocrit levels increase with increasing hemoglobin levels, whereas platelet levels decrease when Hb levels rise [13]. Therefore, this study aimed to model the patterns of platelet and hematocrit levels in patients with DHF, which were measured at least once a day while they were being treated in the hospital. This type of data is called longitudinal data, which refers to the collection of data from *n* subjects who are mutually independent, and each subject is observed repeatedly at various times. When the number of individuals is the same, longitudinal data have the benefit of providing a more accurate estimate of the treatment impact than cross-sectional data owing to the reduction in observation error [14]. A regression analysis approach is more suitable for modeling longitudinal data.

There are two types of regression models in regression analysis: parametric and non-parametric [15]. The examination time was assumed to be unknown in the form of a relationship pattern, which made it a nonparametric component. In contrast, Hb is considered to follow a linear relationship, making it a parametric component. Under these conditions, a suggested semiparametric regression model is proposed. Previous research suggests that in patients with DHF, platelet and Hb levels are correlated, making semiparametric bi-response regression modeling the appropriate approach for modeling [16]. Several semiparametric regressions have been studied in previous research, including the use of the Fourier Series estimator [17], Local Polynomial estimator using single response [18], mixed estimator (Spline Truncated, Fourier series, and Kernel) [19], and spline estimator [20]-[23]. By using the C-Square test showed that platelets and hematocrit are correlated in patients with DHF; therefore, appropriate modeling utilizes bi-response regression modeling [24].

The goal of semiparametric bi-response regression modeling is to estimate the regression curves. The local polynomial is one of several estimators that can be used to estimate regression curves. Local polynomials have several advantages, including the ability to reduce asymptotic bias and produce good estimations [25]. The nonparametric curve estimated in this model uses Local polynomial estimator because the data curve implemented in the model, namely platelet and hematocrit levels, showed neither monotonically increasing nor monotonically decreasing trends; therefore, the appropriate estimator uses a local polynomial compared to the spline estimator, Fourier Series, or other estimators. Previous research used a local polynomial estimator in a semiparametric bi-response regression model on longitudinal data but was applied to simulations and Grade 1 DHF data, therefore this research was conducted on Grade 2 DHF data [14]. This modeling uses Grade 2 DHF data because it occurs in a critical phase which is the most dangerous period in DHF disease. This phase is characterized by a decrease in body temperature after a high fever, but actually raises the risk of blood plasma leakage, bleeding, and the potential for life-threatening shock, so it is necessary to create a modeling that can be analyzed further. This study aimed to model semiparametric regression using a local polynomial estimator on platelet and hematocrit levels of patients with DHF Grade 2 during hospitalization at Roemani Hospital Semarang in 2022, based on examination time and Hb levels. Based on the results of this study, the public can obtain information more easily and simply in the form of modeling using a semiparametric bi-response regression approach for longitudinal data based on a local polynomial estimator. Through modeling, the dynamics of changes over time can be understood, and it can also be determined when platelet and hematocrit levels decrease and increase, as well as when patients require more intensive treatment, thereby reducing the mortality rate of patients with DHF. This study will be conducted at Roemani Hospital in Semarang, one of the largest referral hospitals in Semarang City.

2. MATERIALS AND METHODS

2.1. Local Polynomial Estimators for Semiparametric Regression Models on Longitudinal Data

The first step in modeling longitudinal data is to estimate the parameters of the semiparametric one-response regression using local polynomial approach. The weighted least squares (WLS) estimation method was used for nonparametric parameter estimation, whereas the ordinary least squares (OLS) method was used for parametric parameter estimation. The following semiparametric one-response regression model (Equation (1)) was applied to the longitudinal data [14].

$$y_{ij} = x_{ij}{}^T \beta + \eta(t_{ij}) + e_{ij}; i = 1, 2, ..., n; j = 1, 2, ...m \quad \textbf{(1)}$$

where $e_{ij} \sim N(0,s^2)$ is the observation error, and the regression function , h is an unknown function estimated by the local polynomial. The parametric component is $x_{ij}{}^T b$, while $h(t_{ij})$ is nonparametric component. Equation (1) can be expressed in matrix form of Equation (2).

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\eta} + \underset{\sim}{\varepsilon} \quad \textbf{(2)}$$

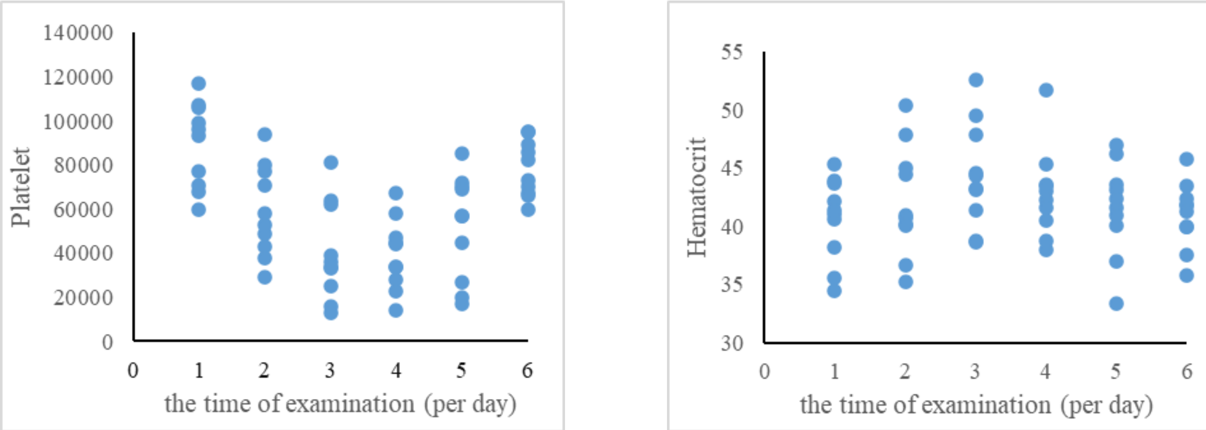
For example, given a known estimate $\hat{\underset{\sim}{\beta}}$, the residual can be calculated $\underset{\sim}{y}^* = \underset{\sim}{y} - X\hat{\underset{\sim}{b}}$, and the residual can be used as the response variable to estimate the estimator $\underset{\sim}{\eta}$ to obtain Equation (3).

$$\underset{\sim}{y}^* = \underset{\sim}{\eta} + \underset{\sim}{\varepsilon} \quad \textbf{(3)}$$

The Taylor series $\eta(t_{ij})$ in Equation (3) is approximated by a polynomial of degree p at point t around point t_0 as stated by Equations (4) and (5) [18].

Table 4. Descriptive statistics on Grade 2 DHF patients.

Variable	Mean	Standard Deviation	Minimum	Maximum
Platelet (/µL)	63115.38	27086.67	13000	152000
Hematocrit (%)	41.83	3.95	33.4	52.6
Haemoglobin (g/dL)	14.22	1.27	11.6	18.4
Examination Time (days)	3.5	1.72	1	6



(a) **(b)**

Figure 1. Scatterplot of (a) platelets with examination time; (b) hematocrit with examination time in Grade 2 DHF patients.

$$\eta(t_{ij}) \approx \eta(t_0) + (t_{ij}-t_0)\eta^{(1)}(t_0) + ... + (t_{ij}-t_0)^p \eta^{(p)}(t_0)/p! \quad \textbf{(4)}$$

$$t_{ij} \in [t_0 - h, t_0 + h]$$

Let $\theta_k(t_0) = \eta^{(k)}(t_0)/k!; k = 1,2,...,p$, then: $\eta(t_{ij}) \approx \underset{\sim}{z}_{ij}^T \underset{\sim}{\theta}(t_0)$ **(5)**

$$z_{ij} = [1,(t_{ij}-t_0),(t_{ij}-t_0)^2,...,(t_{ij}-t_0)^p]^T,\ \underset{\sim}{\theta}(t_0) = [\theta_0(t_0),\theta_1(t_0),\theta_2(t_0),...,\theta_p(t_0)]^T$$

The parameters $\hat{\underset{\sim}{\theta}}$ be an estimator of $\underset{\sim}{\theta}$ obtained by minimizing the WLS criterion [14] in Equation (6);

$$WLS = (\underset{\sim}{y}^* - Z_{t_0}\underset{\sim}{\theta}(t_0))^T K_h (\underset{\sim}{y}^* - Z_{t_0}\underset{\sim}{\theta}(t_0)) \quad \textbf{(6)}$$

Where $Z_i = [z_{i1}, z_{i2}, ..., z_{im}]^T$, $Z_{t_0} = [Z_1^T, Z_2^T, ..., Z_n^T]$, $K_h = diag(K_{1h},....,K_{nh}); K_h(.) = \frac{1}{h}K(\frac{.}{h})$, K is the Kernel function, and h is the bandwidth. Based on calculations using the WLS method, estimator of θ is obtained by Equation (7).

$$\hat{\underset{\sim}{\theta}}(t_0) = (Z_{t_0}{}^T K_h Z_{t_0})^{-1} Z_{t_0} K_h \underset{\sim}{y}^* \quad \textbf{(7)}$$

Next, $\hat{\underset{\sim}{\beta}}$ is sought by minimizing OLS, Equation (8) could be used [14]:

$$\Omega(\underset{\sim}{\beta}) = ((I-B)y - (I-B)X\underset{\sim}{\beta})^T((I-B)y - (I-B)X\underset{\sim}{\beta}) \quad \textbf{(8)}$$

with $B = Z_{t_0}(Z_{t_0}{}^T K_h Z_{t_0})^{-1} Z_{t_0} K_h$

2.2. Methods

2.2.1. Data Sources and Research Variables

This study discusses the modeling of platelet and hematocrit patterns in patients with DHF, measured at least once daily during hospitalization. This research utilized secondary data obtained from Roemani Hospital, Semarang, Indonesia, between March and October 2022. The collection of data from separate subjects observed repeatedly at various times is known as longitudinal data [26]. The variables studied were platelet and hematocrit levels in patients with DHF (Table 1). This study included a sample of 13 patients with DHF Grade 2, marked by Grade 1 and accompanied by viral infection with spontaneous bleeding manifestations in the form of nosebleeds [27]. The data used in this research were from 13 patients whose blood samples were taken once a day for six days during hospitalization, resulting in a total of 78 datasets. The longitudinal data used in this research per patient had the same treatment which was measured repeatedly for 6 days so that the measurement interval was regular and without missing data.

This study used longitudinal data with a limited number of subjects. Despite the limited number of subjects, longitudinal data still have the advantage that each subject contributes numerous observations over time. This rich information from each subject helps build a deeper understanding of the dynamics at play, even when the number of subjects is small in the study. The ability of longitudinal data to measure changes over time within the same subjects provides more statistical power and precise estimates of effects than collecting data from a large, single-point cross-sectional study would [14]. The data used in the study were divided into two groups, namely 80% in-sample data and 20% out-sample data, so that the number of data points as in-sample data was 10 patients, while the out-sample data were three patients. The dependent variables in this study were hematocrit and platelet counts in patients with DHF during hospital treatment, whereas the independent variables were Hb and observation time (per day) during hospitalization.

2.2.2. Research Steps

The first step in modeling the platelet and hematocrit data was to estimate the parameters of the bi-response semiparametric model using local polynomial regression. The estimation method used was the WLS method for nonparametric parameter estimation, while the OLS method was used for parametric parameter estimation. Next, we estimated the regression model, which was then implemented in the data of DHF patients during hospitalization at Roemani Hospital, Semarang City in 2022. The effects of Hb level and examination duration on platelet and hematocrit data were modeled using local polynomial semiparametric bi-response regression. The steps taken are testing the correlation between dependent variables, namely, platelet and hematocrit levels. The Pearson correlation test is one of the available correlation tests. This test was performed when the data were continuous data. The data were divided into in-sample and out-sample data, with in-sample data from 10 DHF patients and out-sample data from three DHF patients.

Selection of optimum polynomial order and bandwidth using the GCV method. Calculate the optimal order and bandwidth (h) without weighting ($W=V^{-1}$) using the following steps. Determine the set of polynomial orders and the lower bound bandwidth, upper bound, $h=\{h(lower),\ldots, h(upper)\}$. Determination of GCV. The choice of polynomial order and bandwidth (h) corresponds to the lowest GCV value (Eq. 9) [14][28].

$$GCV(h)=\frac{2N^{-1}(\underset{\sim}{y}^{T}[I-B^{*}(h)]^{T})([I-B^{*}(h)]\underset{\sim}{y})}{(2N^{-1}trace[I-B^{*}(h)])^{2}} \quad \textbf{(9)}$$

Determining parameters of the semiparametric bi -response regression model ($\hat{\beta}$ and $\hat{\theta}$) without using the weighting matrix (W). A heteroscedasticity test was conducted using the Box-M test. Determining the weighting matrix (***W***). The weighting matrix, which V^{1} is the matrix of the error variance covariance of response variables 1 and 2 ($W=V^{-1}$) [28]. A weighting matrix W is needed to estimate a semiparametric bi-response regression model with heteroscedastic data. Obtaining the weighting matrix ($W=V^{-1}$) by calculating the $\sigma^2{}_{11}$(error variance in response 1), is σ_{22}^2 the error variance in response 2, and $\sigma_{12}$is the covariance between responses 1 and 2 in Equation (10) [14].

$$W=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}^{-1}$$

$$\Sigma_{rr}=diag(\sigma_1^{2(r)},\sigma_2^{2(r)},\ldots,\sigma_n^{2(r)})\ ,\Sigma_{sr}=diag(\sigma_1^{(s)(r)},\sigma_2^{(s)(r)},\ldots,\sigma_n^{(s)(r)})\ \textbf{(10)}$$

W is a $2N$x$2N$ matrix, ($N=n\times m_i$). Calculate the optimal order and bandwidth (h) using weighting ($W=V^{-1}$), as in step 3(a and b). Determine the semiparametric bi-response regression estimation model ($\hat{\beta}$ and $\hat{\theta}$) based on Equations (7) and (10)

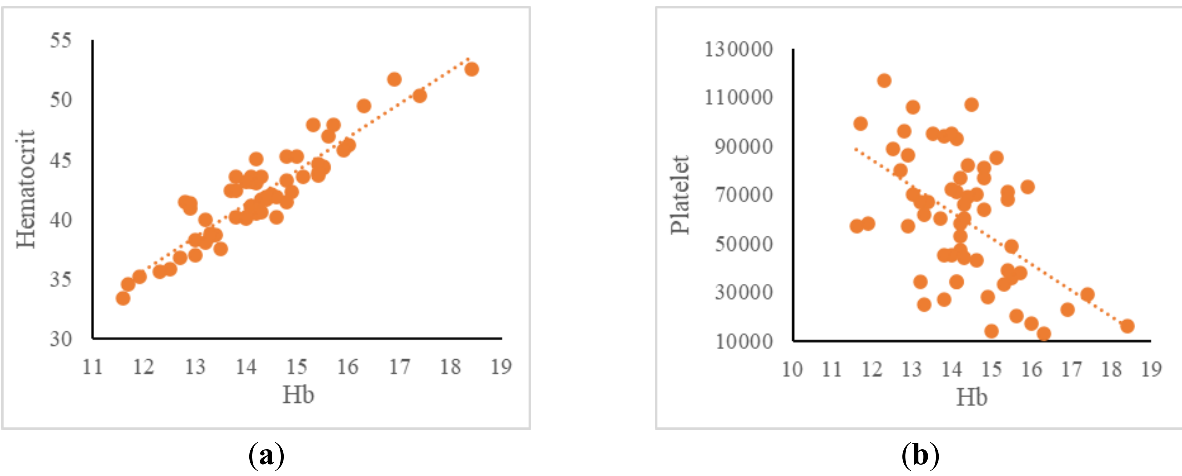


Figure 2. Scatterplot of (a) Hematocrit with Hb and (b) Platelets with Hb in DHF Grade 2 patients.

Table 5. Combination of polynomial order and bandwidth using GCV.

Polynomial Order		Bandwidth		GCV
Response 1 (*p1*)	**Response 2 (*p2*)**	**Response 1 (*h1*)**	**Response 2 (*h2*)**	
1	1	0.2	0.2	64.8066
1	2	0.2	0.7	64.8065
1	2	0.1	0.1	64.8000
2	1	0.1	0.8	64.7999*

*The Smallest GCV

using matrix W obtained previously based on the polynomial order and optimum bandwidth. The MSE, coefficient of determination (R^2), and MAPE were then calculated. MAPE values are categorized into 4 (Table 2) [29].

$$MSE(h) = 2N^{-1}(\underset{\sim}{y}^T[I - B^*(h)]^T)([I - B^*(h)]\underset{\sim}{y}) \text{ , } R^2 = \frac{\sum_{i=1}^{2}\sum_{j=1}^{m}\sum_{k}^{n}\left(\hat{y}_{jk}^{(i)} - \bar{y}\right)^2}{\sum_{i=1}^{2}\sum_{j=1}^{m}\sum_{k}^{n}\left(y_{jk}^{(i)} - \bar{y}\right)^2}$$

$$;0 \le R^2 \le 1$$

$$MAPE = \sum_{i=1}^{2}\sum_{j=1}^{m}\sum_{k}^{n}\left|\frac{y_{jk}^{(i)} - \hat{y}_{jk}^{(i)}}{\hat{y}_{jk}^{(i)}}\right| \times 100\%$$

3. RESULTS AND DISCUSSIONS

3.1. Descriptive Statistics

A semiparametric bi-response regression model estimation of longitudinal data was implemented in patients with Grade 2 DHF treated at Roemani Hospital using a local polynomial estimator. This study was performed to determine the relationship between platelets. A semiparametric bi-response regression model estimation of longitudinal data was implemented in patients with Grade 2 DHF treated at Roemani Hospital using a local polynomial estimator. This study investigated the relationship between platelets (y_1) , and the number of hematocrit cells (y_2) against the time of examination (per day) (t) and the number of hemoglobin (Hb) cells (x). This study utilized two response variables: the number of platelets and hematocrit in patients with Grade 2 DHF treated in the hospital.

The criteria for Grade 2 DHF include fever symptoms, followed by a positive tourniquet test, evidence of bleeding, indications of a viral infection, and additional manifestations of viral infection, including spontaneous bleeding, such as nosebleeds and red spots. Based on the data obtained, it is presented in Table 3, which shows the characteristics of patients with Grade 2 DHF patients, totalling 13 patients who were hospitalized for 6 days, based on the gender of the patients, consisting of three adult male patients, four adult female patients and six children.

Before conducting the regression analysis, it was necessary to understand the initial data overview using descriptive statistics. Table 4 presents the descriptive statistics of 13 patients hospitalized for 6 days with Grade 2 DHF. The independent variables consisted of parametric and nonparametric components. Hb was the parametric component, whereas the time of evaluation for hospitalized patients with Grade 2 DHF was a non-parametric component. Scatterplot of platelets and hematocrit against examination time in patients with Grade 2 DHF. Figure 1 shows that the data are spread out, do not follow a specific pattern, or have an unknown pattern. Therefore, the predictor variable, as a nonparametric component, was the examination time. The following are the results of the scatterplot between the two response variables, namely hematocrit and platelets, and Hb in patients with DHF Grade 2. Based on the results of the scatter plot, the hematocrit level data on Hb have a pattern that is suspected to be known; therefore, the predictor variable for Hb is a parametric component. Parametric components can be linear, quadratic, or cubic; however, Figure 2 shows a linear relationship pattern.

3.2. The Correlation Test

This study employed a local polynomial estimator to model the bi-response semiparametric regression curve. The kernel function used was the Gaussian Kernel. The advantages of the Gaussian Kernel function are its flexible nature, ability to capture complex relationships in data, ability to produce efficient and accurate estimates, and relatively easy mathematical calculations [30]. The goodness criteria used were the MSE and the coefficient of determination. The best model is the one with the highest value, approaching 100%. The calculation of local polynomial birespon regression is quite complicated and involves many processes; therefore, software is needed to solve it, one of which is R software. One of the developments in this research is the creation of Graphical User Interface (GUI) for the R program to facilitate data analysis. The predictors used in this study were examination time and Hb level. Before estimating the parameters of the second-response model, a correlation test was conducted. Correlation test was conducted to determine the correlation between the two response variables, platelet count and hematocrit level, in patients with DHF Grade 2. This test was used to determine the correlations between the two responses used in this study. The following are the results of the Pearson correlation test. The Pearson correlation test output results obtained the correlation value and p-value for patients with Grade 2.

The hypothesis are H_0: $\rho = 0$ (No correlation between platelets and hematocrit) and H_1: $\rho \neq 0$ (There is a significant correlation between platelets and hematocrit). The rejection criterion for the above hypothesis is that, if the value $|t_{value}| > t_{(\frac{\alpha}{2}, n-2)}$ rejects H_0, there is a significant correlation between the two variables [31]. Based on the output results of the Pearson correlation test on Grade 2 DHF data, the p-values were 9.378×10^{-7} and t_{value} = -5.4873. Based on the test criteria, using α = 5%, so that $t_{(\frac{\alpha=0.05}{2}, n-2)} = t_{table} = 1.992$ so that the p-value < 0.05= α and $|t_{value}| > t_{table}$ then the decision to reject Ho, it can be interpreted as indicating a significant correlation between platelets and hematocrit in DHF Grade 2 patients.

3.3. Estimation of the Semiparametric Bi-response Regression Model

The study included 13 patients, who were split into two groups: in-sample and out-of-sample data. The estimation process was first performed on the in-sample data of 10 patients (patients 1–10). Patients 11–13 were used as out-of-sample data. Before estimating the model parameters using in-sample data, we first determined the optimal polynomial order and bandwidth without using the weighting matrix V^{-1}. The initial step determines the combination of polynomial orders for both responses without using the W weighting matrix based on in-sample data. Next, we determine which bandwidth, based on the GCV approach, can be attained concurrently, that is, the bandwidth with the minimum GCV. Parameter estimation on Grade 2 DHF patient data uses the optimal bandwidth value without considering the weighting W. In this estimation, the first response error value (platelet) and the second (hematocrit) in patients with DHF Grade 2 were obtained, which were then tested for heteroscedasticity. The semiparametric bi-response model can be written as follows [14]. The predictor variables consist of x_{ij} and t_{ij}, while $y_{ij}^{(1)}$ is the 1st response, $y_{ij}^{(2)}$ is the 2nd response. The relationship between $y_{ij}^{(k)}$ and t_{ij} is unknown in its functional form, while between $y_{ij}^{(k)}$ and x_{ij} its functional form is known linearly, then the relationship between $y_{ij}^{(k)}$, x_{ij} and t_{ij} uses the semiparametric bi-response model as expressed in Equation (11).

$$y_{ij}^{(k)} = \beta_{0.ij}^{(k)} + \beta_{1.ij}^{(k)} x_{ij} + \eta^{(k)}(t_{ij}) + e_{ij}^{(k)}; i = 1,2,3,...,13; j = 1,2,3,...,6; k = 1,2 \quad \textbf{(11)}$$

Equation (11) can be expressed in matrix form in Equation (12).

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\eta} + \underset{\sim}{\varepsilon} \quad \textbf{(12)}$$

where $\underset{\sim}{y} = (\underset{\sim}{y}^{(1)}, \underset{\sim}{y}^{(2)})^T$ consists of two response variables that are correlated with each other. $X\underset{\sim}{\beta} = (X^{(1)}\underset{\sim}{\beta}^{(1)}, X^{(2)}\underset{\sim}{\beta}^{(2)})^T$ is the parametric component, while $\underset{\sim}{\eta} = (\underset{\sim}{\eta}^{(1)}, \underset{\sim}{\eta}^{(2)})^T$ is the nonparametric component, and $\boldsymbol{\varepsilon} = (\underset{\sim}{\varepsilon}^{(1)}, \underset{\sim}{\varepsilon}^{(2)})^T$ is a subject error with a variance-covariance matrix $\boldsymbol{\Sigma_i}$, with index i indicating the subject [14]. Subsequently, $\underset{\sim}{\eta} = (\underset{\sim}{\eta}^{(1)}, \underset{\sim}{\eta}^{(2)})^T$ is approximated by Taylor Series of polynomials of degree- p at point t around point t_0 [14].

$$\eta^{(k)}(t_{ij}) \approx \eta_0^{(k)}(t_0) + (t_{ij} - t_0)\eta_1^{(k)}(t_0) + ... + (t_{ij} - t_0)^p \eta_p^{(k)}(t_0) / p!; k = 1,2.$$
$$\eta^{(k)}(t_{ij}) \approx \theta_0^{(k)}(t_0) + (t_{ij} - t_0)\theta_1^{(k)}(t_0) + ... + (t_{ij} - t_0)^p \theta_p^{(k)}(t_0) / p!; k = 1,2.$$

with $\theta_r^{(k)}(t_0) = \eta_r^{(k)}(t_0) / r!; r = 0,1,2,..., p$

Defining $\eta^{(k)}(t_{ij}) \approx (z_{ij}^{(k)})^T \underset{\sim}{\theta}^{(k)}, i = 1,2,...n$

where $z_{ij}^{(k)} = [1, (t_{ij} - t_0), (t_{ij} - t_0)^2, ..., (t_{ij} - t_0)^p]^T$,

$\underset{\sim}{\theta}^{(k)} = [\theta_0^{(k)}(t_0), \theta_1^{(k)}(t_0), \theta_2^{(k)}(t_0), ..., \theta_p^{(k)}(t_0)]^T$,

so that equation 14 becomes: $\underset{\sim}{y} = X\underset{\sim}{\beta} + Z_{t_0}\underset{\sim}{\theta}(t_0) + \underset{\sim}{\varepsilon}$

$$\underset{\sim}{\mathbf{y}} = \begin{bmatrix} \mathbf{y}^{(1)} \\ --- \\ \mathbf{y}^{(2)} \end{bmatrix} = \begin{bmatrix} y_{11}^{(1)} \\ y_{12}^{(1)} \\ \vdots \\ y_{1m_1}^{(1)} \\ y_{21}^{(1)} \\ \vdots \\ y_{2m_2}^{(1)} \\ \vdots \\ y_{n1}^{(1)} \\ \vdots \\ y_{nm_n}^{(1)} \\ --- \\ y_{11}^{(2)} \\ y_{12}^{(2)} \\ \vdots \\ y_{1m_1}^{(2)} \\ y_{21}^{(2)} \\ \vdots \\ y_{2m_2}^{(2)} \\ \vdots \\ y_{n1}^{(2)} \\ \vdots \\ y_{nm_n}^{(2)} \end{bmatrix} ; \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & 0 & 0 \\ 1 & x_{12} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1m_1} & 0 & 0 \\ 1 & x_{21} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{2m_2} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{nm_n} & 0 & 0 \\ & --- & & \\ 0 & 0 & 1 & x_{11} \\ 0 & 0 & 1 & x_{12} \\ \vdots & & \vdots & \\ 0 & 0 & 1 & x_{1m_1} \\ 0 & 0 & 1 & x_{21} \\ \vdots & & \vdots & \\ 0 & 0 & 1 & x_{2m_2} \\ \vdots & & \vdots & \\ 0 & 0 & 1 & x_{n1} \\ \vdots & & \vdots & \\ 0 & 0 & 1 & x_{nm_n} \end{bmatrix}$$

$$\underset{\sim}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0^{(1)} \\ \beta_1^{(1)} \\ --- \\ \beta_0^{(2)} \\ \beta_1^{(2)} \end{bmatrix} ; \quad \mathbf{X}\underset{\sim}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0^{(1)} + \beta_1^{(1)} x_{11} \\ \beta_0^{(1)} + \beta_1^{(1)} x_{12} \\ \vdots \\ \beta_0^{(1)} + \beta_1^{(1)} x_{1m_1} \\ \beta_0^{(1)} + \beta_1^{(1)} x_{21} \\ \vdots \\ \beta_0^{(1)} + \beta_1^{(1)} x_{2m_2} \\ \vdots \\ \beta_0^{(1)} + \beta_1^{(1)} x_{n1} \\ \vdots \\ \beta_0^{(1)} + \beta_1^{(1)} x_{nm_n} \\ ---- \\ \beta_0^{(2)} + \beta_1^{(2)} x_{11} \\ \beta_0^{(2)} + \beta_1^{(2)} x_{12} \\ \vdots \\ \beta_0^{(2)} + \beta_1^{(2)} x_{1m_1} \\ \beta_0^{(2)} + \beta_1^{(2)} x_{21} \\ \vdots \\ \beta_0^{(2)} + \beta_1^{(2)} x_{2m_2} \\ \vdots \\ \beta_0^{(2)} + \beta_1^{(2)} x_{n1} \\ \vdots \\ \beta_0^{(2)} + \beta_1^{(2)} x_{nm_n} \end{bmatrix}_{2N\times 1}$$

$$Z_{t_0} = \begin{bmatrix} Z^{(1)}(t_0) & 0 \\ 0 & Z^{(2)}(t_0) \end{bmatrix} ;$$

$$Z^{(k)}(t_0) = \begin{bmatrix} 1 & (t_{11}-t_0) & (t_{11}-t_0)^2 & \cdots & (t_{11}-t_0)^p \\ 1 & (t_{12}-t_0) & (t_{12}-t_0)^2 & \cdots & (t_{12}-t_0)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (t_{1m_1}-t_0) & (t_{1m_1}-t_0)^2 & \cdots & (t_{1m}-t_0)^p \\ 1 & (t_{21}-t_0) & (t_{21}-t_0)^2 & \cdots & (t_{21}-t_0)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (t_{2m_2}-t_0) & (t_{2m_2}-t_0)^2 & \cdots & (t_{2m}-t_0)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (t_{n1}-t_0) & (t_{n1}-t_0)^2 & \cdots & (t_{n1}-t_0)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (t_{nm_n}-t_0) & (t_{nm_n}-t_0)^2 & \cdots & (t_{nm_n}-t_0)^p \end{bmatrix}_{Nx(p+1)} ;$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_0^{(1)}(t_0) \\ \theta_1^{(1)}(t_0) \\ \theta_2^{(1)}(t_0) \\ \vdots \\ \theta_p^{(1)}(t_0) \\ --- \\ \theta_0^{(2)}(t_0) \\ \theta_1^{(2)}(t_0) \\ \theta_2^{(2)}(t_0) \\ \vdots \\ \theta_p^{(2)}(t_0) \end{bmatrix}_{2(p+1)\times 1}$$

For example, given a known estimate , $\underset{\sim}{\hat{\beta}}$ the residual can be calculated $\underset{\sim}{y}* = \underset{\sim}{y} - X\underset{\sim}{\hat{\beta}}$ and the residual can be used as the response variable to estimate the estimator $\underset{\sim}{\theta}$ (Eq. 13).

$$\underset{\sim}{y}^* = Z_{t_0}\underset{\sim}{\theta}(t_0) + \underset{\sim}{\varepsilon} \tag{13}$$

The parameter $\hat{\theta}$ is an estimator of θ obtained by minimizing the WLS criterion (Eq. 14) [14].

$$WLS = (\underset{\sim}{y}^* - Z_{t_0}\underset{\sim}{\theta}(t_0))^T W K_h(t_0)(\underset{\sim}{y}^* - Z_{t_0}\underset{\sim}{\theta}(t_0)) \tag{14}$$

where K is the Kernel function, and h is the bandwidth so

$$K_h(t_0) = \begin{pmatrix} diag(K_{1h}^{(1)}(t_0), K_{2h}^{(1)}(t_0), ..., K_{nh}^{(1)}(t_0)) & 0 \\ 0 & diag(K_{1h}^{(2)}(t_0), K_{2h}^{(2)}(t_0), ..., K_{nh}^{(2)}(t_0)) \end{pmatrix}$$

$$K_{ih}^{(r)}(t_0) = diag(K_{i1}^{(r)}(t_{il}-t_0), K_{i2}^{(r)}(t_{i2}-t_0), ..., K_{im_i}^{(r)}(t_{im_i}-t_0))$$

The Kernel function used was the Gaussian Kernel. Gaussian Kernel: $K(t)=\frac{1}{\sqrt{2\pi}}exp\left(\frac{1}{2}(-t^2)\right)$, $-\infty < t < \infty$.

Then, the parameter estimator of $\underset{\sim}{\theta}$ is obtained (Eq. 15).

$$\hat{\underset{\sim}{\theta}}(t_0)=\left(Z_{t_0}{}^T WK_h(t_0)Z_{t_0}\right)^{-1} Z_{t_0}{}^T WK_h(t_0)\underset{\sim}{y}^* \quad \textbf{(15)}$$

The Local Polynomial estimator [18] is $\hat{\underset{\sim}{\eta}}(t_{ij})=Z_{t_0}\left(Z_{t_0}{}^T WK_h(t_0)Z_{t_0}\right)^{-1} Z_{t_0}{}^T WK_h(t_0)\underset{\sim}{y}^*$

$$\hat{\underset{\sim}{\eta}}(t_{ij})=Z_{t_0}\left(Z_{t_0}{}^T WK_h(t_0)Z_{t_0}\right)^{-1} Z_{t_0}{}^T WK_h(t_0)(\underset{\sim}{y}-X\underset{\sim}{\beta}) \quad \textbf{(16)}$$

Thus, based on Equation 16, η can be stated as follows: $\hat{\underset{\sim}{\eta}}(t_{ij})=B\underset{\sim}{y}^*=B(\underset{\sim}{y}-X\underset{\sim}{\beta})$
with $B=Z_{t_0}(Z_{t_0}{}^T K_h Z_{t_0})^{-1} Z_{t_0}{}^T K_h$

Next, $\hat{\beta}$ is sought by minimizing OLS as follows [14]:

$$OLS(\underset{\sim}{\beta})=((I-B)\underset{\sim}{y}-(I-B)X\underset{\sim}{\beta})^T((I-B)\underset{\sim}{y}-(I-B)X\underset{\sim}{\beta})$$

so that we obtain Eq. 17.

$$\hat{\underset{\sim}{\beta}}=\left(X^T(I-B)^T(I-B)X\right)^{-1}X^T(I-B)^T(I-B)\underset{\sim}{y} \quad \textbf{(17)}$$

The next stage is the estimation of the bi-response semiparametric regression model using a local polynomial based on the weighting matrix obtained in the previous step, which is estimated using the following Equation 18.

$$\hat{\underset{\sim}{y}}=X\hat{\underset{\sim}{\beta}}+Z_{t_0}\hat{\underset{\sim}{\theta}}(t_0) \quad \textbf{(18)}$$

$\hat{\underset{\sim}{y}}=\left(B_{par}+B(I-B_{par})\right)\underset{\sim}{y}$ with $B_{par}=X\left(X^T(I-B)^T(I-B)X\right)^{-1}X^T(I-B)^T(I-B)$

$\hat{\underset{\sim}{y}}=B_{sempar}(h)\underset{\sim}{y}$, with $B_{sempar}(h)=B_{par}+B(I-B_{par})$

3.4. Heteroscedasticity Analysis

Heteroscedasticity analysis was carried out using the Box-M test, and the rejection criterion for the

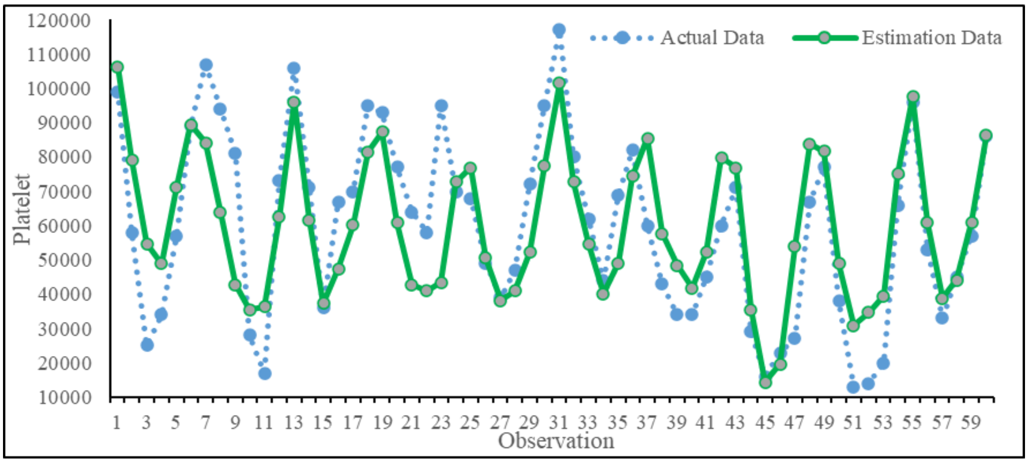


Figure 3. Scatterplot of the estimated in-sample data results and the actual platelet values for DHF Grade 2 patients.

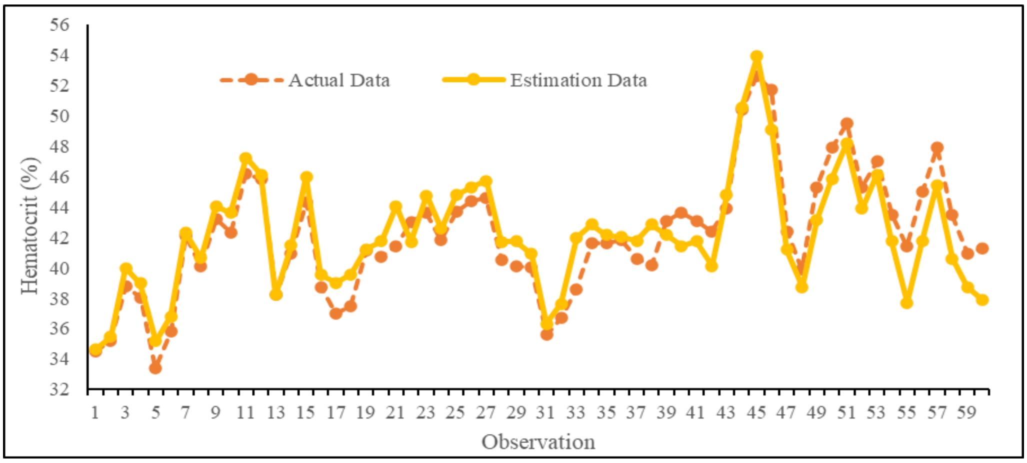


Figure 4. Scatterplot of the estimated in-sample data results and the actual hematocrit values for DHF Grade 2 patients.

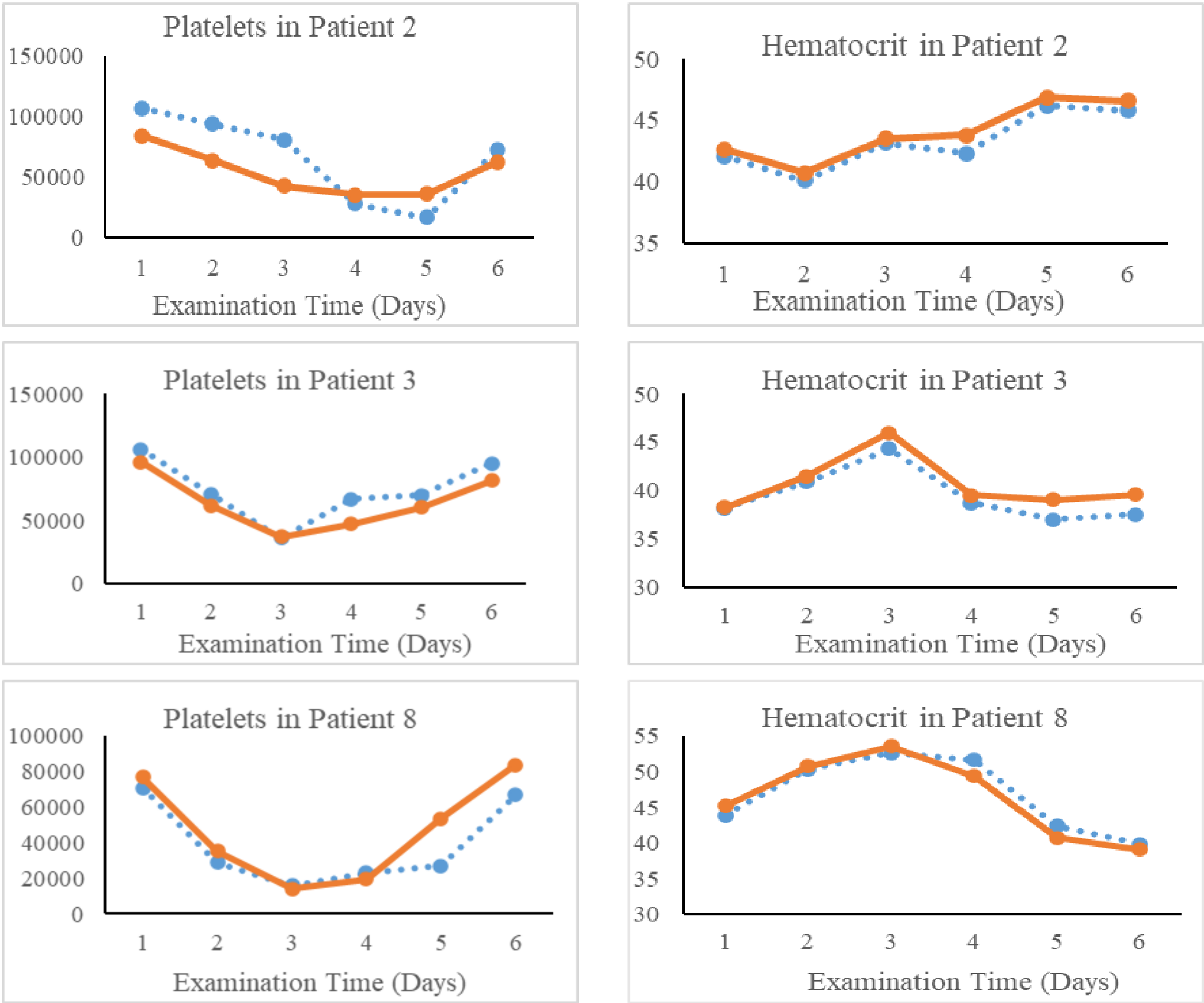


Figure 5. Plot of platelets and hematocrit in patients 2, 3 and 8 Grade 2 DHF.

Box-M test was that if the p-value was < 5%, then the decision was to reject Ho. Heteroscedasticity analysis was performed using the Box-M test with the following hypotheses.

$$H_0 : \Sigma_1 = \Sigma_2 = ... = \Sigma_{10}$$
$$H_1 : \Sigma_i \neq \Sigma_j ; i \neq j ; i, j = 1,2,...,10 \text{ (at least one } \Sigma_i \neq \Sigma_j \text{)}$$

Based on the Box M test output, the p-value < 2.2×10^{-16}, using alpha 5%, then reject Ho, so it can be concluded that there is heteroscedasticity between the error in response 1 (platelet) and the error in response 2 (hematocrit) in DHF Grade 2 patients based on in-sample data. After conducting the heteroscedasticity test, the semiparametric bi-response regression model was estimated on the longitudinal data using a weighted local polynomial estimator with matrix W [14]. Nonparametric curves $\eta(t_i)$ were estimated by local polynomial estimators using kernel weighting matrix K_h and W. Based on the results of the heteroscedasticity test, the W weighting matrix is then formed. The next step is to form the weighting matrix ($W = V^{-1}$) according to equation 12 by calculating the σ^2_{11} (error variance in response 1), σ^2_{22} is the error variance in response 2, and σ_{12} is the covariance between responses 1 and 2.

3.5. Selection of Polynomial Order and Optimal Bandwidth

After conducting the heteroscedasticity test and forming the W weighting matrix, the next step is to determine the polynomial order and optimal bandwidth using the W weighting matrix based on in-sample data. The local polynomial order used to estimate the model was p = {1, 2} because if the polynomial order is large it will cause overfitting so it needs to be limited. The bandwidths tested were set to the lower bound of 0.1 and the upper bound of 6.8, with a range of values 0.1. Based on the combination of polynomial order and bandwidth, the GCV value is calculated using Equation 9. The selection of the optimal polynomial order and bandwidth(h) corresponds to the smallest GCV value. The results of the combination of polynomial order and bandwidth using the GCV method are presented in Table 5, which was obtained from the four smallest GCV values.

Based on Table 5, the minimum (smallest) GCV value is 64.7999, which is located at the combined polynomial order $p1$=2 and $p2$=1, while the bandwidths *are* h1=0.1 and h2=0.8. Therefore, the optimal bandwidth for Grade 2 DHF patient data was at response 1 (platelets) of 0.1 and at response 2 (hematocrit) of 0.8, located at the combined polynomial order p1=2 and p2=1.

3.6. Modeling of Platelets and Hematocrit Levels in Patients with DHF

The next process is determining the parameters of the regression model using the weighting matrix W and the kernel function K_h applied to the in-sample data. The parameters of the semiparametric bi-response regression model were estimated using a local polynomial estimator on the Grade 2 DHF patient data. The optimal bandwidth values for the first and second responses were 0.1 and 0.8, respectively. The MSE value was 121948569, resulting from a polynomial order of two for the first and one for the second response. The actual data and estimated in-sample data using the local polynomial estimator in response 1 (platelets), as shown in Figure 3. A comparison graph between the actual data and estimated in-sample data using the local polynomial estimator for response 2 (hematocrit) is shown in Figure 4.

Based on Figure 3, the broken lines represent the actual data for response 1 (platelets), and the continuous lines represent the estimated data for response 1 (platelets). In Figure 4, the broken lines represent the actual data, while the continuous lines represent the estimated in-sample data in the bi-response semiparametric regression model for response 2. Both graphs show the bi-response

Table 6. Analysis of the condition of several Grade 2 DHF patients.

Patient	Gender	Analysis of the Condition
Patient 2	Female (Adult)	The patient experienced Thrombocytopenia, with the lowest platelet count on the 4th day, namely 35429.22/µL. Based on calculations with 95% confidence intervals, the average platelet count of the patient was between 25160.55/µL and 83221.27/µL. The highest hematocrit (46.88%) was observed on the 5th day of hospitalisation, indicating hematocrit above normal, which necessitated monitoring for hemoconcentration. Based calculations with 95% confidence intervals, the average hematocrit of the patient was between 42.16% and 45.93%.
Patient 3	Children	The patient experienced Thrombocytopenia, with the lowest platelet count occurring on day 3, at 37282.21/µL. Based calculations with 95% confidence intervals, the average platelet count of the patient was between 4444.71/µL and 83529.4/µL. The highest hematocrit (45.49%) was observed on day 3, with hematocrit above normal and hemoconcentration requiring intensive treatment on that day. Based calculations with 95% confidence intervals, the average hematocrit of the patient was between 38.47% and 42.71%.
Patient 8	Male (Adults)	The patient experienced thrombocytopenia and the lowest platelet count occurred on day 3. at 14251.55/µL Based calculations with 95% confidence intervals, the average platelet count of the patient was between 23952.84/µL and 70663.52/µL. The highest hematocrit (53.6%) was observed on day 3, indicating a hematocrit above normal and hemoconcentration, which required intensive treatment starting on day 3. Based calculations with 95% confidence intervals, the average hematocrit of the patient was between 42.49% and 51.14%.

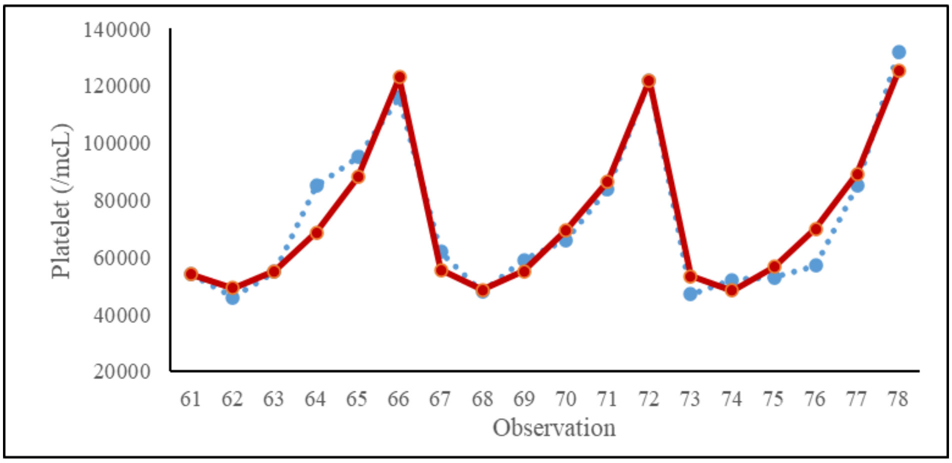


Figure 6. Plot of the results of the out-sample data estimation of predicted platelet values for patients with Grade 2 DHF.

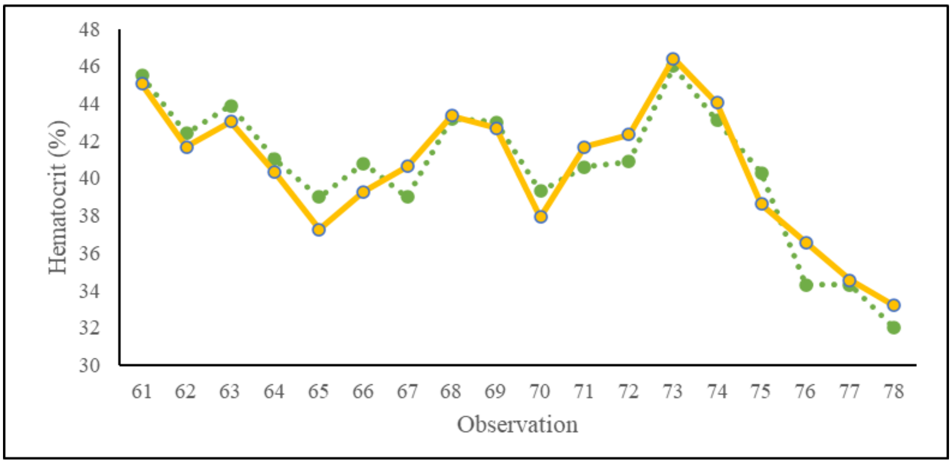


Figure 7. Plot of the results of the out-sample data estimation of the predicted hematocrit values for patients with Grade 2 DHF.

semiparametric regression model estimation results, using the local polynomial estimator, continuous lines show that the estimated data follows the pattern of the actual data. Therefore, the evaluation results of the semiparametric bi-response regression model, based on a coefficient of determination (R^2) of 90.12%, indicate that examination time and Hb have an influence of 90.12% on platelets in patients with DHF Grade 2, while the remaining 9.88% is affected by other factors. The MAPE value obtained from this model was 15.92%, indicating a good level of accuracy. The platelet and hematocrit modeling estimation results obtained using local polynomial bi-response regression on longitudinal data yielded an R^2 value greater than 90%, indicating that the model accurately predicted the response variables. The following is a plot of the dynamics of platelet and hematocrit changes in data from several patients with Grade 2 DHF hospitalized for 6 days using a semiparametric biresponse regression model with a polynomial estimator.

Based on Figure 5, the blue dotted line shows the actual data, and the orange connected line shows the estimated data. Dengue fever is characterized by thrombocytopenia, which is characterized by a decrease in the number of platelets to below 100000/µL. In addition, the presence of an increased hematocrit above normal levels in patients with dengue is a sign of plasma leakage and can be fatal. The following are the normal hematocrit values: 33–38% in children, 40–50% in adult males, and 36–44% in adult females [32]. Some conditions experienced by patients with Grade 2 DHF are based on age and sex, as shown in Figure 5, so that each patient can be analyzed.

After obtaining the optimal polynomial order and bandwidth values, the parameter estimation was performed using a local polynomial estimator. Platelet and hematocrit modeling equations for patients with DHF can be formulated based on the model parameter estimation results. The following shows platelet modeling of Grade 2 DHF in the 3rd patient at the 5th examination time (Equation (19)).

$$\hat{y}_{35}^{(1)} = \hat{\beta}_{0.35}^{(1)} + \hat{\beta}_{1.35}^{(1)} x_{35} + \hat{\theta}_{0.35}^{(1)} + (t_{35}-t)\hat{\theta}_{1.35}^{(1)} + (t_{35}-t)^2 \hat{\theta}_{2.35}^{(1)}$$
$$\hat{y}_{35}^{(1)} = 162166.6 - 7853.7x_{35} + (5-t)^2 6.97\times 10^{-16}; 4.9 < t < 5.1 \quad \textbf{(19)}$$

The hematocrit modeling of Grade 2 DHF in the 3rd patient at the 5th examination time is presented in Equation (20).

$$\hat{y}_{35}^{(2)} = \hat{\beta}_{0.35}^{(2)} + \hat{\beta}_{1.35}^{(2)} x_{35} + \hat{\theta}_{0.35}^{(2)} + (t_{35}-t)\hat{\theta}_{3.25}^{(2)}$$
$$\hat{y}_{35}^{(2)} = 16.03 + 2.78x_{35} - (5-t)8.55; 4.2 < t < 5.8 \quad \textbf{(20)}$$

Based on Equation (19), platelet modeling Grade 2 DHF disease in the 3rd patient at the 5th examination time, if there is an increase in the amount of Hb of 1 g/dL, it will reduce the number of platelets in DHF patients by 7853.7/µL, assuming that the observation time variable in the model is constant. Equation (20), which models hematocrit in Grade 2 DHF in the 3rd patient at the 5 th examination time, indicates that a 1-g/dL increase in Hb will increase the hematocrit of patients with DHF by 2.78%, assuming that the observation time variable in the model remains constant.

The model estimation of the platelet count of Grade 2 DHF in the 6th patient at the 2nd examination time is presented in Equation (21).

$$\hat{y}_{62}^{(1)} = \hat{\beta}_{0.62}^{(1)} + \hat{\beta}_{1.62}^{(1)} x_{62} + \hat{\theta}_{0.62}^{(1)} + (t_{62}-t)\hat{\theta}_{1.62}^{(1)} + (t_{62}-t)^2 \hat{\theta}_{2.62}^{(1)}$$
$$\hat{y}_{62}^{(1)} = 171236 - 7774.88x_{62} + (2-t)^2 1.21\times 10^{-15}; 1.9 < t < 2.1 \quad \textbf{(21)}$$

Meanwhile, the hematocrit modeling of Grade 2 DHF in the 6th patient at the 2nd examination time is presented in Equation (22).

$$\hat{y}_{62}^{(2)} = \hat{\beta}_{0.62}^{(2)} + \hat{\beta}_{1.62}^{(2)} x_{62} + \hat{\theta}_{0.62}^{(2)} + (t_{62}-t)\hat{\theta}_{3.62}^{(2)}$$
$$\hat{y}_{62}^{(2)} = 40.69 + 3.45x_{62} - (2-t)6.81; 1.2 < t < 2.8 \quad \textbf{(22)}$$

Based on Equation (20), for the 6th patient at the time of examination on the 2nd day, a 1-gram/dL increase in Hb will reduce the number of platelets in the DHF patient by 7774.88/µL. Based on Equation (21), for the 6th patient at the time of examination on the 2nd day, a 1-gram/dL increase in Hb will increase the hematocrit number of the DHF patient by 3.45%.

Based on the analysis results, it can be concluded that Hb in patients with Grade 2 DHF has a positive effect on hematocrit, as reported earlier [32]. Conversely, Hb has a negative effect on platelets. This means that in cases of DHF, the increase in Hb and hematocrit has a unidirectional relationship (both increase) and is in the opposite direction to the number of platelets, which usually decreases (thrombocytopenia), especially when hemoconcentration occurs due to blood plasma leakage through damaged capillary walls [12]. If there are indications of severe plasma leakage in patients with DHF, it can lead to shock and death if not treated properly.

Based on Figure 5 and Table 6, the dynamics of platelet and hematocrit changes in each Grade 2 DHF patient during the 6-day hospitalization period were observed. On average, patients experienced a decrease in platelet count from the first to the 4th day, followed by an increase from the 5th to the 6th day. On the 3rd or 4th day, the average Grade 2 DHF patient experienced thrombocytopenia with the lowest decrease in platelets. At that time, the patient's hematocrit was not stable, as patients with Grade 2 DHF typically experience hematocrit levels above normal, and 40% also experience hemoconcentration. This is a marker of plasma leakage and can be fatal; therefore, it is necessary to be aware that the patient's condition requires more intensive care to stabilize the platelet count and hematocrit. Several studies have reported that in DHF in adults without shock, the platelet count

Table 7. Model accuracy values for in-sample and out-of-sample data.

Model Accuracy	*In-sample*	*Out-sample*
MSE	121948569	20820591
R^2	90.12%	98.77%
MAPE	15.92%	4.84%

begins to decrease on the 3rd day and shows a significant decrease on the 4th day of illness [33]. Although there is no definite evidence of the cause of thrombocytopenia in dengue fever, several mechanisms are thought to play a role, including decreased platelet production due to bone marrow suppression, increased platelet destruction, and excessive use of platelets.

After obtaining the estimated platelet and hematocrit values, predictions were made based on the out-sample data. Predictor variables were used to predict the values of the platelet and hematocrit response variables by estimating the parameters of the out-of-sample data. The data were then entered into the regression model based on the results of the in-sample model. Patients 11–13, namely observations 61–78, were used as out-sample data, resulting in out-sample data estimates in the form of predicted platelet values. The out-sample data produced an estimate of the predicted platelet values (Figure 6). The actual plot and estimated results of the out-sample data in the form of the predicted hematocrit values are shown in Figure 7.

Figure 6 shows that the estimated out-sample data for platelet prediction, indicated by the red line, followed the pattern of the actual values indicated by the blue dotted line. Figure 7 shows the same thing; the estimated out-sample data for hematocrit prediction, indicated by the yellow line, follows the pattern of the actual values indicated by the green dotted line. After the estimation process of the in-sample data, the accuracy of the model was evaluated using the criteria of the mean squared error (MSE), coefficient of determination (R^2), and mean absolute percentage error (MAPE). Table 7, which displays the MAPE, R^2, and MSE values for both the in-sample and out-of-sample data, summarizes the estimation findings for the optimal combination based on the program's execution.

Based on the MAPE values in Table 7, the in-sample data of 60 observations (patients 1–10) yielded a MAPE of 15.92% and R^2 of 90.12%, indicating that the resulting model had accurate predictions. Meanwhile, the out-sample data of 18 observations (patients 11–13) yielded an MAPE of 4.84% and R^2 of 98.77%, indicating that the prediction model was highly accurate. Based on the results obtained at 95% confidence level, the model mean error (MAPE) is between 3.21% and 6.48%. This shows different results from the previous research , who modeled platelets producing an $R^2$of 84.25% [18]. It can be seen that the R^2 values in this study are still better, so this study improves on previous research. Previous research conducted used a spline penalty applied to crime cases and obtained a regression model with an R^2 value of 83.18%, showing different results [22], so this research with a local polynomial estimator is better.

4. CONCLUSIONS

Semiparametric bi-response regression modeling of local polynomial estimators was applied to data on patients with Grade 2 DHF treated at Roemani Hospital, Semarang City, with predictor variables in the form of examination time and Hb level, while the response variables were platelet and hematocrit levels. Hb is a parametric component, while the time of evaluation for hospitalized Grade 2 DHF patients is the nonparametric component. The estimation of the bi-response semiparametric regression model by the local polynomial applied to estimate platelets and hematocrit in patients with Grade 2 DHF produced a coefficient of determination of 90.12%, which can be used to predict platelets and hematocrit very accurately with a prediction inaccuracy (MAPE) of 4.84%. The results of this modeling will be very useful because they produce specific laboratory parameters (platelets and hematocrit) that can predict the course of Grade 2 DHF, allowing for early detection. This will impact the progression of patients who initially appear well and may progress to a state of shock. Prompt diagnosis, close monitoring, and supervision are key to successful management of dengue fever. If the platelet count decreases (thrombocytopenia) on the 3rd or 4th day of hospitalization and hemoconcentration occurs due to blood plasma leakage through damaged capillary walls in Grade 2 DHF patients, this can cause shock and death if not treated properly. Patients with Grade 2 DHF experienced thrombocytopenia and the lowest platelet count, followed by an increase in the hematocrit above normal levels, and 40% also experienced hemoconcentration. This is a marker of plasma leakage and can be fatal; therefore, it is necessary to be aware that the patient's condition requires more intensive care to stabilize the platelet count and hematocrit.

AUTHOR INFORMATION

Corresponding Author

Nur Chamidah — Department of Mathematics, Airlangga University, Surabaya-60115 (Indonesia); Research Group of Statistical Modeling in Life Science, Airlangga University, Surabaya-60115 (Indonesia);
orcid.org/0000-0003-1592-4671
Email: nur-c@fst.unair.ac.id

Authors

Tiani Wahyu Utami — Doctoral Study Program of Mathematics and Natural Sciences, Airlangga University, Surabaya-60115 (Indonesia); Statistics Study Program, Universitas Muhammadiyah Semarang, Semarang-50273 (Indonesia);
orcid.org/0009-0008-5278-5037

Toha Saifudin — Department of Mathematics, Airlangga University, Surabaya-60115 (Indonesia); Research Group of Statistical Modeling in Life Science, Airlangga University, Surabaya-60115 (Indonesia);
orcid.org/0000-0002-6716-3096

Budi Lestari — Research Group of Statistical Modeling in Life Science, Airlangga University, Surabaya-60115 (Indonesia); Department of Mathematics, The University of Jember, Jember-68121 (Indonesia);
orcid.org/0009-0005-1978-8613

Norma Latif Fitriyani — Research Group of Statistical Modeling in Life Science, Airlangga University, Surabaya-60115 (Indonesia); Department of Artificial Intelligence Data Science, Sejong University, Seoul-05006 (Republic of Korea);
orcid.org/0000-0002-1133-3965

Author Contributions

Conceptualization, Software, Visualization, Supervision, Methodology, and Writing – Original Draft Preparation, T. W. U., N. C., T. S., and B. L.; Validation and Resources, N. C., T. S., and B. L.; Formal Analysis and Writing – Review & Editing, T. W. U., N. C., T. S., B. L., D. A.; Investigation and Data Curation, T. W. U., and N. C.; Project Administration and Funding Acquisition, T. W. U.

Conflicts of Interest

The authors declare no conflict of interest.

ACKNOWLEDGEMENT

This research was supported by Muhammadiyah University of Semarang, Indonesia. The authors thank Roemani Hospital in Semarang City, Indonesia for providing data used in this research.

DECLARATION OF GENERATIVE AI

Not applicable.

REFERENCES

[1] N. Hidayah, I. Iskandar, and Z. Abidin. (2017). "Prevention of dengue hemorrhagic fever (DHF) associated with the Aedes aegypti larvae presence based on the type of water source". *Journal of Tropical Life Science.* **7** (2): 115-120. 10.11594/jtls.07.02.05.

[2] I. N. Faridah, H. Dania, Y. H. Chen, W. Supadmi, B. D. Purwanto, M. J. Heriyanto, M. A. Aufa, W. C. Chang, and D. A. Perwitasari. (2022). "Dynamic changes of platelet and factors related dengue haemorrhagic fever: A retrospective study in Indonesian". *Diagnostics.* **12** (4): 1-12. 10.3390/diagnostics12040950.

[3] M. A. Hartini and E. T. Pawenang. (2023). "The distribution of dengue fever case based on environmental factors using spatial analysis". *Jurnal Presipitasi: Media Komunikasi dan Pengembangan Teknik Lingkungan.* **20** (2): 345-355. 10.14710/presipitasi.v20i2.345-355.

[4] C. Hermania and W. Cahyati. (2023). "Kejadian DBD di Kota Semarang tahun 2019-2021". *Journal of Public Health Research and Development.* 376-385. 10.15294/higeia.v7i3.65192.

[5] F. A. Ciptono, D. S. Lintang, Y. Sri, and Martini. (2021). "Gambaran demam berdarah

dengue Kota Semarang tahun 2014-2019". *Jurnal Ilmiah Mahasiswa.* **2** (4): 1147-1152.

[6] H. Hariyanto, D. Sutiningsih, and M. Sakundarno. (2024). "Determinants of the incidences of dengue hemorrhagic fever in 2019-2021". *Indonesian Journal of Global Health Research.* **2** (4): 2325-2332. 10.37287/ijghr.v2i4.250.

[7] S. Das, C. Abreu, M. Harris, J. Shrader, and S. Sarvepalli. (2022). "Severe thrombocytopenia associated with dengue fever: An evidence-based approach to management of thrombocytopenia". *Journal of Tropical Medicine.* 1-7. 10.1155/2022/3358325.

[8] A. A. Jacob, E. Joseph, and U. Poothiode. (2022). "Significance of platelet indices in dengue fever patients". *International Journal of Medical Science in Clinical Research and Review.* **5** (5): 531-539. 10.5281/zenodo.7085729.

[9] S. Wulandari and F. Zulaikha. (2021). "Hubungan laboratorium kadar trombosit dan status gizi sebagai faktor risiko DHF pada pasien anak di Puskesmas Mangkurawang Tenggarong". *Borneo Student Research.* **2** (3): 1752-1758.

[10] M. Y. Sifriyani, Diu, Z. Mar'ah, D. Anggraini, and S. Jalaluddin. (2023). "Modeling of dengue hemorrhagic fever cases in AWS Hospital Samarinda using bi-responses nonparametric regression with estimator spline truncated". *Communications in Mathematical Biology and Neuroscience.* 1-12. 10.28919/cmbn/7888.

[11] N. Made, D. Handayani, D. Putu, C. Udiyani, N. Putu, and A. Mahayani. (2022). "Hubungan kadar trombosit, hematokrit, dan hemoglobin dengan derajat demam berdarah dengue pada pasien anak rawat inap di BRSU Tabanan". *Aesculapius Medical Journal.* **2** (2): 130-136.

[12] A. D. Ratulangi, H. Djohan, and L. Kamilla. (2025). "The relationship between NS1 examination and the examination of hemoglobin, hematocrit, leukocytes, platelets, and erythrocytes in dengue fever patients in the pediatric ward". *International Medical Scientific Journal.* **7** (1): 32-39. 10.53770/medica.v7i1.486.

[13] D. D. Tirtadevi, S. N. Riyanti, and R. Wisudanti. (2021). "Korelasi jumlah trombosit dan kadar hematokrit terhadap tingkat keparahan pasien demam berdarah dengue di RSD dr. Soebandi Jember". *Journal of Agromedicine and Medical Sciences.* **7** (3): 156-151. 10.19184/ams.v7i3.24362.

[14] T. W. Utami, N. Chamidah, T. Saifudin, B. Lestari, and D. Aydin. (2025). "Estimation of biresponse semiparametric regression model for longitudinal data using local polynomial kernel estimator". *Symmetry.* **17** (3): 1-22. 10.3390/sym17030392.

[15] H. Husain, P. I. Rahayu, M. R. Nisardi, M. A. Al-Fadhilah, and A. Husain. (2025). "Semiparametric biresponse regression modeling mixed spline truncated, Fourier series, and kernel in predicting rainfall and sunshine". *Statistics, Optimization and Information Computing.* **14** (1): 62-76. 10.19139/soic-2310-5070-2166.

[16] A. F. Salsabila, Juniastuti, D. Husada, and D. Puspitasari. (2024). "Hematology profiles and disease severity of pediatric dengue virus infection at a tertiary hospital in Surabaya, Indonesia". *Majalah Biomorfologi.* **34** (1): 1-9. 10.20473/mbiom.v34i1.2024.1-9.

[17] Kuzairi, Miswanto, and M. F. F. Mardianto. (2021). "Semiparametric regression based on Fourier series for longitudinal data with weighted least square optimization". *Journal of Physics: Conference Series.* **1836** (1). 10.1088/1742-6596/1836/1/012038.

[18] T. W. Utami, N. Chamidah, and T. Saifudin. (2024). "Platelet modeling in DHF patients using local polynomial semiparametric regression on longitudinal data". *AIP Conference Proceedings.* **3302** (1): 1-10. 10.1063/5.0262183.

[19] N. Chamidah and B. Lestari. (2019). "Estimation of covariance matrix using multi -response local polynomial estimator for designing children growth charts: A theoretical discussion". *Journal of Physics: Conference Series.* **1397** (1). 10.1088/1742-6596/1397/1/012072.

[20] W. Ramadan, N. Chamidah, B. Zaman, L. Muniroh, and B. Lestari. (2019). "Standard growth chart of weight for height to determine wasting nutritional status in East Java based on semiparametric least square spline estimator". *IOP Conference Series: Materials Science and Engineering.* **546** (5). 10.1088/1757-899X/546/5/052063.

[21] F. Jiang, Y. Ma, and R. J. Carroll. (2023). "A spline-assisted semiparametric approach to nonparametric measurement error models". *Econometrics and Statistics.* 10.1016/j.ecosta.2023.05.001.

[22] Kosmaryati and M. D. Kartikasari. (2020). "Analysis of penalized semiparametric regression model on bi-response longitudinal data". *Advances in Social Science, Education and Humanities Research.* **474** : 172-178. 10.2991/assehr.k.201010.025.

[23] M. Setyawati, N. Chamidah, and A. Kurniawan. (2021). "Modelling scholastic aptitude test of State Islamic Colleges in Indonesia using least square spline estimator in longitudinal semiparametric regression". *Journal of Physics: Conference Series.* **1764** (1). 10.1088/1742-6596/1764/1/012077.

[24] C. H. Tampubolon, E. Manalu, and T. P. Dewi. (2019). "Relationship between increased hematocrit levels and lowest platelet counts with outcomes of dengue fever patients in children". *Indonesian Journal of Global Health Research.* **2** (4): 993-1000.

[25] A. H. Welsh and T. W. Yee. (2006). "Local regression for vector responses". *Journal of Statistical Planning and Inference.* **136** (9): 3007-3031. 10.1016/j.jspi.2004.01.024.

[26] S. Side, W. Sanusi, and M. W. Maksum. (2020). "Model regresi semiparametrik spline untuk data longitudinal pada kasus demam berdarah dengue di Kota Makassar". *Journal of Mathematics, Computations and Statistics.* **3** (1): 20. 10.35580/jmathcos.v3i1.19181.

[27] S. Efendi, N. Sriyanah, H. R. Buntu, I. Syam, S. Suarni, and D. Djunaedi. (2022). "The grade of dengue hemorrhagic fever in children". *Indonesian Journal of Global Health Research.* **4** (2): 411-418.

[28] H. Wu and J. T. Zhang.(2006)." Nonparametric regression methods for longitudinal data analysis. 10.5860/CHOICE.44-0978.

[29] E. Vivas, H. Allende-Cid, and R. Salas. (2020). "A systematic review of statistical and machine learning methods for electrical power forecasting with reported MAPE score". *Entropy.* **22** (12): 1-24. 10.3390/e22121412.

[30] L. R. Cheruiyot, G. O. Orwa, and O. R. Otieno. (2020). "Kernel function and nonparametric regression estimation: Which function is appropriate?". *African Journal of Mathematics and Statistics Studies.* **3** (1): 51-59.

[31] M. Zhang, W. Li, L. Zhang, H. Jin, Y. Mu, and L. Wang. (2023). "A Pearson correlation-based adaptive variable grouping method for large-scale multi-objective optimization". *Information Sciences.* **639**. 10.1016/j.ins.2023.02.055.

[32] D. Sahassananda, V. Thanachartwet, P. Chonsawat, B. Wongphan, S. Chamnanchanunt, M. Surabotsophon, and V. Desakorn. (2021). "Evaluation of hematocrit in adults with dengue by a laboratory information system". *Journal of Tropical Medicine.* 1-9. 10.1155/2021/8852031.

[33] M. Raadsen, J. du Toit, T. Langerak, B. van Bussel, E. van Gorp, and M. Goeijenbier. (2021). "Thrombocytopenia in virus infections". *Journal of Clinical Medicine.* **10** (4): 1-33. 10.3390/jcm10040877.