International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 9 Issue 12 December - 2025, Pages: 80-88

Application of Truncated Spline Nonparametric Regression For Modeling Income Inequality in Yogyakarta: A Panel Data Approach

Nashwa Carista[1], Toha Saifudin[2]*, Dita Amelia[3], M. Fariz Fadillah Mardianto[4]

[1,2,3,4]Statistics Study Program, Department of Mathematics, Faculty of Sains and Technology, Universitas Airlangga, Surabaya, Indonesia

Corresponding author: tohasaifudin@fst.unair.ac.id

Abstract: *Income inequality remains a persistent development issue in Indonesia, including Yogyakarta, and is commonly assessed using the Gini Ratio as key indicator. This research aims to model income inequality in Yogyakarta by analyzing the Gini Ratio using truncated spline nonparametric regression within a panel data framework. The analysis utilizes secondary data from the Central Bureau of Statistics for the 2020-2025 period, incorporating several socio-economic indicators as predictor variables. The truncated spline method is selected for its ability to accommodate both linear and nonlinear patterns without imposing strict functional assumptions. Model selection relies on the Generalized Cross Validation (GCV) criterion, with the optimal model obtained using three knots. The best model yields GCV value of 0.00045 and R-Square value of 99.53%, indicating an excellent fit. The findings show that key economic indicators significantly explain variations in Gini Ratio, demonstrating the effectiveness of nonparametric methods in capturing the complex structure of income distribution. These insights support evidence-based policymaking aligned with Sustainable Development Goals (SDGs) 10, which emphasizes reducing inequality within regions. Despite limitations related to aggregated panel data and the short observation period, this research provides methodological value by offering a flexible analytical framework that can complement traditional parametric approaches.***

Keywords— Gini Ratio, Income Inequality, Panel Data, Nonparametric Regression, Truncated Spline

1. INTRODUCTION

Income inequality remains a persistent socioeconomic challenge in many countries, including Indonesia. The Gini Ratio is commonly employed to measure disparities in income distribution, where a higher value indicates greater inequality [1]. According to the latest data released by Statistics Indonesia (BPS), the national Gini Ratio reached 0.381 in 2024, reflecting a slight improvement from the previous year yet still highlighting structural disparities in income distribution [2]. At the provincial level, the Special Region of Yogyakarta has consistently recorded the highest inequality in Indonesia from 2020 to 2023, before becoming the second highest in 2024 with a Gini Ratio of 0.428. These figures underscore the persistent income gap within the province and reinforce the urgency of analyzing the underlying factors contributing to this inequality [3].

Existing policies such as the Family Hope Program (PKH), Non-Cash Food Assistance (BPNT), and the Village Fund have been implemented to alleviate inequality and improve welfare. However, their effectiveness remains limited due to issues of unequal distribution, leakages, and short-term relief mechanisms [4][5]. These policy constraints underscore the need for empirical approaches that more accurately capture the complex determinants of inequality at the regional level. Recent research on income inequality across Java and Indonesia has utilized parametric panel regression methods, identifying variables such as the Human Development Index (HDI), provincial minimum wages, labor productivity, and GRDP as significant determinants of inequality [6][7]. While relevant, these studies predominantly rely on linear parametric models, which may not sufficiently accommodate nonlinear relationships embedded in socio-economic data.

Advancements in nonparametric regression have introduced more flexible analytical tools for modeling complex and nonlinear patterns without assuming a predetermined functional form. Spline-based estimators, particularly truncated splines offer advantages in capturing local fluctuations through optimal knot selection and ensuring smooth, continuous curve estimation [8]-[10]. Evidence from recent studies suggests that truncated spline estimators outperform several alternative methods, including Fourier series, in terms of modeling accuracy and interpretability [11]. Applications of nonparametric truncated spline regression have also been extended to longitudinal and panel datasets, as demonstrated in studies examining inequality determinants in West Java. Amelia applied truncated nonparametric spline regression to analyze the Gini Ratio and indicated that HDI, Labor Force Participation Rate (LFPR), and GRDP had significant negative effects on inequality, while poverty percentage and the health index showed positive contributions to the increase in the Gini Ratio [12]. Nonetheless, research applying truncated spline nonparametric regression to model income inequality in Yogyakarta remains scarce.

This gap highlights the need for an analytical framework capable of capturing the nonlinear and dynamic behavior of socio-economic indicators influencing inequality in Yogyakarta. Unlike previous studies that rely heavily on linear parametric approaches, the present research introduces a truncated spline nonparametric regression model within a panel data setting to more accurately identify the underlying structure of inequality determinants. This approach strengthens methodological novelty by combining spline-based flexibility with the expanded information offered by panel data. The objective of this research is to model income inequality in Yogyakarta using a truncated spline nonparametric regression approach based on panel data from 2020-2025, incorporating HDI, LFPR, poverty rate, open unemployment rate, and district/city minimum wages as predictors. In addition, this research directly aligns with Sustainable Development Goals (SDGs) 10, which emphasizes reducing inequality through evidence-based, inclusive, and regionally adaptive policy strategies. This research contributes to the literature by presenting a flexible modeling framework capable of capturing nonlinear variations in inequality and offering empirical insights that can support more effective regional policy formulation.

2. Materials And Methods

2.1 Research Data

This research employs a quantitative approach using secondary data obtained from official institutional publications in Indonesia. The research focuses on Yogyakarta Province, covering all five districts/cities as the analysis units. The dataset consists of panel data spanning five years, from 2019 to 2023, resulting in a total of 25 observational units. All data were sourced from the Yogyakarta Central Statistics Agency through its official website. The data used encompass a set of economic, social, and labor-related indicators empirically associated with income inequality. The research incorporates one response variable and six predictor variables, as shown in Table 1.

Table 1: Research Variables

Variable	Description	Unit	Scale
Y	Gini Ratio	Index (0–100)	Ratio
X_1	Human Development Index (HDI)	Index (0–100)	Ratio
X_2	District/City Minimum Wage (CMW)	Million IDR per month	Ratio
X_3	Percentage of Poor People (PPP)	Percent (%)	Ratio
X_4	Open Unemployment Rate (OUR)	Percent (%)	Ratio
X_5	Labor Force Participation Rate (LFPR)	Percent (%)	Ratio

The research recognizes employment conditions as an essential external economic factor influencing income inequality, captured through the Labor Force Participation Rate (LFPR). Other potential macroeconomic determinants, such as inflation or government policy interventions, were not included due to data availability constraints and the lack of strong empirical evidence supporting their direct association with the Gini Ratio within the provincial context.

2.2 Panel Data

Longitudinal data refers to repeated observations collected from the same units over multiple periods, allowing researchers to analyze changes in behavior or characteristics over time. When these repeated measurements are taken from multiple units and observed in the same time span, the structure becomes panel data, which combines both cross-sectional and time-series dimensions [13]. This dual dimensionality enables a more comprehensive examination of variations across units as well as changes across time. Through this structure, panel data supports the application of adaptive modeling techniques, including nonparametric regression, which can capture dynamic patterns that may evolve over time. Panel data involves repeatedly observing the same analysis units within a defined time horizon, allowing for a richer understanding of the relationship between response and predictor variables.

Diggle [14] explains that studies involving repeated measurements over time differ fundamentally from pure cross-sectional studies, which take observations only once at a specific moment. Data compilation in panel settings may follow prospective or retrospective approaches, depending on the availability and structure of the information. Within regression modeling, nonparametric regression offers greater flexibility for panel data compared to parametric models, as it does not impose strict functional form assumptions and can adapt to nonlinear relationships more effectively [15]. One of the primary advantages of panel data is its ability to reduce multicollinearity among predictors and increase estimation efficiency. Additionally, panel data enables researchers to address analytical questions that cannot be adequately explored using only cross-sectional or time-series data in isolation [16].

2.3 Truncated Spline Nonparametric Regression on Panel Data

In general, the multipredictor nonparametric regression model on panel data with y_{ji} is the response variable and $s = 1,2,\dots,p$ predictor variables where subject $j = 1,2,\dots,m$ and each subject is observed $i = 1,2,\dots,n$ times can be expressed in the form of Equation (2) as follows [17][18].

$$y_{ji} = f(x_{1ji}, x_{2ji}, \dots, x_{pji}) + \varepsilon_{ji} \tag{1}$$

where y_{ji} is response variable of the j-th subject and i-th observation, x_{sji} is s-th predictor variable, j-th subject, and i-th observation, f is an unknown function of the predictor x_{sji}, and ε_{ji} represents the residual of the j-th subject and i-th observation which assumed to be identical, independent distributed with a mean of 0 and constant variance.

According to Takezawa [19], the truncated power basis function used in the truncated spline nonparametric regression with polynomial degree K and R knots is $\{(x-\tau_1)_+^K, \dots, (x-\tau_R)_+^K\}$. In panel data with p predictor variables, the multipredictor truncated spline nonparametric regression model on panel data can be defined as a spline function f with polynomial degree K and R knots. By assuming each predictor variables are not correlated can be written as an additive model [20]. Then, Equation (1) can be expressed in Equation (2) as follows [21]-[24].

$$y_{ji} = \sum_{s=1}^{p} f(x_{sji}) + \varepsilon_{ji} \tag{2}$$

where

$$f(x_{sji}) = \sum_{k=0}^{K} \alpha_{sjk} x_{sji}^{k} + \sum_{r=1}^{R} \beta_{sjr} \left(x_{sji} - \tau_{sjr}\right)_+^K \tag{3}$$

$$\left(x_{sji} - \tau_{sjr}\right)_+^K = \begin{cases} \left(x_{sji} - \tau_{sjr}\right)^K & , x_{sji} \geq \tau_{sjr} \\ 0 & , x_{sji} < \tau_{sjr} \end{cases} \tag{4}$$

with α_{sjk} is polynomial coefficients of k-th degree for s-th predictor in j-th subject that representing the global component of the spline function, β_{sjr} is coefficients of s-th predictor for j-th subject at the r-th knot location that representing the truncated spline components, and τ_{sjr} is the r-th knot location for s-th predictor in j-th subject serving as a point where the function is allowed to change slope or curvature.

The truncated spline nonparametric function in Equation (3) can be described by each subject is observed $i = 1, 2, \dots, n$ as matrix in Equation (5) follows [25].

$$\boldsymbol{f}(\boldsymbol{x_{sj}}) = \boldsymbol{X_{sj}} \boldsymbol{\delta_{sj}} \tag{5}$$

where

$$\boldsymbol{f}(\boldsymbol{x_{sj}}) = \begin{bmatrix} f(x_{sj1}) \\ f(x_{sj2}) \\ \vdots \\ f(x_{sjn}) \end{bmatrix}, \boldsymbol{X_{sj}} =$$

$$\begin{bmatrix} 1 & x_{sj1} & \dots & x_{sj1}^K & \left(x_{sj1} - \tau_{sj1}\right)_+^K & \dots & \left(x_{sj1} - \tau_{sjR}\right)_+^K \\ 1 & x_{sj2} & \dots & x_{sj2}^K & \left(x_{sj2} - \tau_{sj1}\right)_+^K & \dots & \left(x_{sj2} - \tau_{sjR}\right)_+^K \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{sjn} & \dots & x_{sjn}^K & \left(x_{sj1} - \tau_{sj1}\right)_+^K & \dots & \left(x_{sjn} - \tau_{sjR}\right)_+^K \end{bmatrix}$$

and $\boldsymbol{\delta_{sj}} = \left[\alpha_{sj0}, \alpha_{sj1}, \dots, \alpha_{sjK}, \beta_{sj1}, \dots, \beta_{sjR}\right]'$.

Hence, the multipredictor truncated spline nonparametric regression model on panel data in Equation (2) for $i = 1, 2, \dots, n$ observation can be written in Equation (6) below.

$$\boldsymbol{y_j} = \sum_{s=1}^{p} \boldsymbol{f}(\boldsymbol{x_{sj}}) + \boldsymbol{\varepsilon_j} \tag{6}$$

Then Equation (6) also can be presented as matrix in Equation (7).

$$\boldsymbol{y_j} = \boldsymbol{X_j} \boldsymbol{\delta_j} + \boldsymbol{\varepsilon_j} \tag{7}$$

where

$$\boldsymbol{y_j} = \begin{bmatrix} y_{j1} \\ y_{j2} \\ \vdots \\ y_{jn} \end{bmatrix}, \boldsymbol{X_j} = \begin{bmatrix} \boldsymbol{X_{1j}} & \boldsymbol{X_{2j}} & \dots & \boldsymbol{X_{pj}} \end{bmatrix}, \boldsymbol{\delta_j} = \begin{bmatrix} \boldsymbol{\delta_{1j}} \\ \boldsymbol{\delta_{2j}} \\ \vdots \\ \boldsymbol{\delta_{pj}} \end{bmatrix}, \boldsymbol{\varepsilon_j} = \begin{bmatrix} \varepsilon_{j1} \\ \varepsilon_{j2} \\ \vdots \\ \varepsilon_{jn} \end{bmatrix}$$

Using Equation (7), the multipredictor truncated spline nonparametric regression model on panel data for $j = 1, 2, \dots, m$ subject can be expressed in matrix notation in Equation (8) as follows.

$$\boldsymbol{y} = \boldsymbol{X\delta} + \boldsymbol{\varepsilon} \tag{8}$$

where

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \vdots \\ \boldsymbol{y_m} \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X_2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{X_m} \end{bmatrix}, \quad \boldsymbol{\delta} = \begin{bmatrix} \boldsymbol{\delta_1} \\ \boldsymbol{\delta_2} \\ \vdots \\ \boldsymbol{\delta_m} \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon_1} \\ \boldsymbol{\varepsilon_2} \\ \vdots \\ \boldsymbol{\varepsilon_m} \end{bmatrix}$$

By using the weight $\boldsymbol{V} = cov(\boldsymbol{\varepsilon})$ written as follows [15].

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V_2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{V_m} \end{bmatrix}, \text{ where } \boldsymbol{V_j} = \mathrm{cov}(\boldsymbol{\varepsilon_j}).$$

The estimation of Equation (8) can be obtained by minimizing the Weighted Least Squares (WLS) function so that $\hat{\boldsymbol{\delta}}$ written in Equation (9) [26].

$$\hat{\boldsymbol{\delta}} = \left(\boldsymbol{X}(\boldsymbol{\tau})'\mathbf{V}^{-1}\mathbf{X}(\boldsymbol{\tau})\right)^{-1} \boldsymbol{X}(\boldsymbol{\tau})'\mathbf{V}^{-1}\mathbf{y} \tag{9}$$

where $\boldsymbol{\tau} = mp(1 + K + R)$. Substitution the Equation (9) to Equation (8), then obtained estimate the multipredictor truncated spline nonparametric regression model on panel data in Equation (10) as follows.

$$\hat{\boldsymbol{y}} = \boldsymbol{A}(\boldsymbol{\tau})\boldsymbol{y} \tag{10}$$

where $\boldsymbol{A}(\boldsymbol{\tau}) = \boldsymbol{X}(\boldsymbol{\tau})\left(\boldsymbol{X}(\boldsymbol{\tau})'\mathbf{V}^{-1}\mathbf{X}(\boldsymbol{\tau})\right)^{-1} \boldsymbol{X}(\boldsymbol{\tau})'\mathbf{V}^{-1}$.

Furthermore, to determine the optimal knot, the Generalized Cross Validation (GCV) method can be used. The GCV method is a development of the Cross Validation (CV) method where the difference lies in the factors that divide the residual. In the GCV method, the factor is the average value of these factors. The GCV value is then obtained by summing the quadratic residuals that have been corrected by the square of these factors. The optimal knot value is given by the smallest GCV value, The GCV function for truncated spline nonparametric regression model on panel data is given in Equation (11) as follows [27][28].

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 9 Issue 12 December - 2025, Pages: 80-88

$$GCV(\tau) = \frac{MSE(\tau)}{\left[\frac{1}{mn} trace\left(\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{\tau})\right)\right]^2} \quad (11)$$

Then, model evaluation using Mean Square Error (MSE) written in Equation (12) and R-Square written in Equation (13) below [29].

$$MSE(\tau) = \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n}\left(y_{ji} - \hat{y}_{ji}\right)^2 \quad (12)$$

$$R^2 = 1 - \frac{\sum_{j=1}^{m}\sum_{i=1}^{n}\left(y_{ji} - \hat{y}_{ji}\right)^2}{\sum_{j=1}^{m}\sum_{i=1}^{n}\left(y_{ji} - \bar{y}\right)^2} \quad (13)$$

where $\bar{y} = \sum_{j=1}^{m}\sum_{i=1}^{n} y_{ji}$.

2.4 Research Stages

The stages of data analysis carried out with the truncated spline nonparametric regression approach using R software in this research are as follows.

1. Perform descriptive statistical analysis including the mean, minimum, and maximum values to describe the characteristics of the research variables
2. Truncated spline nonparametric regression modeling with the following steps
   a. Create a scatterplot between the response variable and each predictor variable to understand the relationship pattern between the variables.
   b. Selecting the optimal knot points based on the smallest GCV value using the formula in Equation (11) to determine the best truncated spline nonparametric regression model
   c. Estimating the parameters of the best truncated spline model using Equation (9)
   d. Constructing the best truncated spline nonparametric regression model based on the estimated parameters in the form shown in Equation (10)
   e. Calculating model evaluation the MSE value using Equation (12) and R-Square value using Equation (13)
3. Interpret the best truncated spline nonparametric regression model

3. Results and Discussion

3.1 Descriptive Statistics

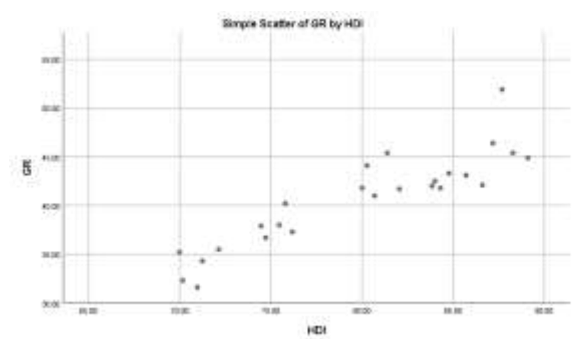
The descriptive statistics of the research variables for 25 panel observations (2020-2024) are presented in Table 2.

Table 2: Descriptive Statistics

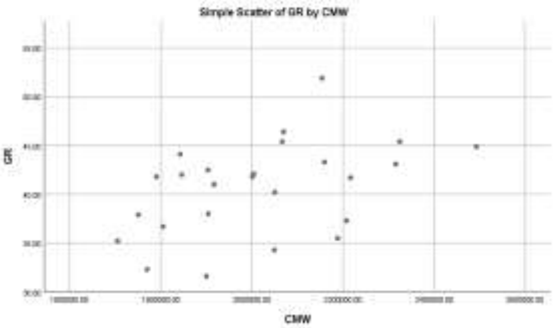
Variable	Mean	Minimum	Maximum
Gini Ratio	40.6	31.6	51.9
HDI	79.88	69.98	89.1
CMW	2.017369	1.705000	2.492997
PPP	12.11	6.26	18.38
OUR	4.23	2.01	9.1561
LFPR	73.60	65.3	78.83

Based on Table 2, the average Gini Ratio in Yogyakarta is 40.6, indicating a moderate level of income inequality. Gunung Kidul records the lowest inequality of 31.6, while Yogyakarta City shows the highest of 51.9. The Human Development Index (HDI) has an average value of 79.88 categorized as high although disparities remain across districts, with Gunung Kidul at the lowest level of 69.98 and Yogyakarta City at the highest of 89.1. The average City/Regency Minimum Wage (CMW) is 2,017,369 rupiah reflecting differences in regional economic capacity, with the lowest value found in Gunung Kidul and the highest in Yogyakarta City. The average Percentage of Poor People (PPP) is 12.11, showing substantial variation from 6.26 in Yogyakarta City to 18.38 in Kulon Progo. The average Open Unemployment Rate (OUR) is 4.23%, with Kulon Progo having the lowest unemployment and Yogyakarta City the highest. Meanwhile, the Labor Force Participation Rate (LFPR) averages 73.60, with Gunung Kidul recording the lowest participation rate and Kulon Progo the highest. These patterns collectively highlight significant socio-economic disparities across districts.

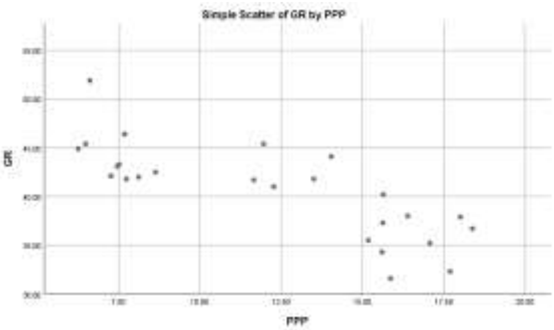
Initial exploratory analysis was conducted using scatterplots. A scatterplot aims to illustrate the relationship between the response and predictor variables. If the scatterplot shows no discernible pattern, a nonparametric approach may be employed. The scatterplot of the response variable and five predictor variables is presented in Fig. 1.



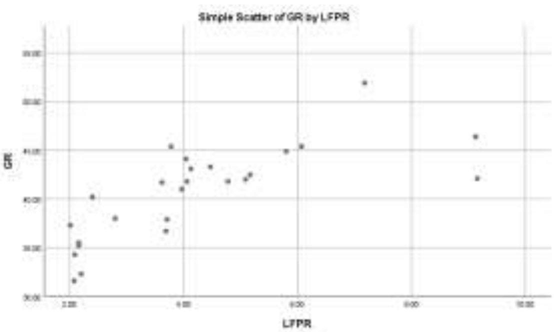
(a)



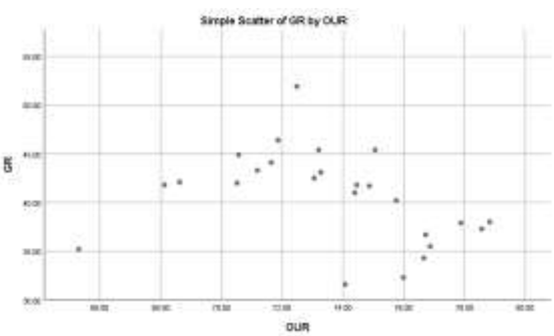
(b)



(c)



(d)



(e)

Fig. 1. Scatterplot of Gini Ratio Against (a) HDI, (b) CMW, (c) PPP, (d) LFPR, (e) OUR

3.2 Selection of Optimal Knot Points

The selection of knot point values is based on the alpha-percent quantile method, which identifies changes in data patterns in truncated spline nonparametric regression model. The optimal number and position of knots are determined using the Generalized Cross Validation (GCV) criterion, where the model with the smallest GCV value is selected. In this research, knot configurations of one, two, and three knots were evaluated through extensive trials consisting of 48 trials for one knot, 1128 trials for two knots, and 4715 trials for three knots. The smallest GCV values obtained from each knot specification are presented in Table 3.

Table 3: Smallest GCV Value Based on Each Knot

Knot	GCV	R-Square
One Knot	0.11858	86.812
Two Knots	0.01088	95.042
Three Knots	0.00045	99.532

Based on Table 3, the smallest GCV value at three knots with the smallest GCV value of 0.00045 and R-Square value of 99.53% that explains the variability of Gini Ratio values that can be explained by predictor variables including HDI, CMW, PPP, LFPR, and OUR. The MSE value of the model is 0.102, which shows that the model has a low prediction error in predicting the Gini Ratio in Yogyakarta. The number of knot points and the optimal knot point value are used to estimate the best truncated spline nonparametric regression model.

3.3 Parameter Estimation

The parameter estimation of the truncated spline nonparametric regression model was conducted for each district/city in Yogyakarta. Based on Table 3, the optimal model was obtained using three knots, the parameter estimation for each district/city follows this three-knot configuration. Prior to estimating the regression coefficients, the optimal knot locations for each predictor variable were identified. Table 4 presents the optimal knot points for Yogyakarta City, which recorded the highest Gini Ratio among all districts/cities in the province. These knot points form the basis for constructing the truncated spline components of the model.

Table 4: Optimal Knot Points for Yogyakarta City

Para-meter	Knot Point	Para-meter	Knot Point	Para-meter	Knot Point
β_{111}	86.66	β_{213}	23.33	β_{412}	6.28
β_{112}	86.97	β_{311}	6.29	β_{413}	8.06
β_{113}	88.29	β_{312}	6.46	β_{511}	68.71
β_{211}	20.14	β_{313}	7.22	β_{512}	69.27
β_{212}	20.74	β_{411}	5.87	β_{513}	71.70

After determining the knot positions, parameter estimation was performed for the truncated spline regression model.

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 9 Issue 12 December - 2025, Pages: 80-88

While the overall model structure is consistent across districts/cities, the coefficient values and knot locations differ according to regional characteristics. As an illustration, Table 5 presents the parameter estimation results for Yogyakarta City.

Table 5: Parameter Estimation Results for Yogyakarta City

Para-meter	Estimation	Para-meter	Estimation	Para-meter	Estimation
α_{110}	-0.011	β_{112}	-0.545	β_{313}	0.094
α_{111}	0.396	β_{113}	-1.358	β_{411}	-0.233
α_{211}	-0.119	β_{211}	-1.199	β_{412}	0.783
α_{311}	1.4	β_{212}	-2.601	β_{413}	2.545
α_{411}	0.414	β_{213}	2.667	β_{511}	1.339
α_{511}	0.78	β_{311}	1.280	β_{512}	0.025
β_{111}	-1.259	β_{312}	0.985	β_{513}	3.758

3.4 Truncated Spline Nonparametric Regression Model

The truncated spline nonparametric regression model based on the knot value in Table 4 and parameter estimates in Table 5 for Yogyakarta City is expressed in the Equation (14).

$$\hat{y}_{1i} = -0.011 + 0.396x_{11i} - 1.259(x_{11i} - 86.66)_+ - 0.545(x_{11i} - 86.97)_+ - 1.358(x_{11i} - 88.29)_+ - 0.119x_{21i} - 1.199(x_{21i} - 2.014)_+ - 2.601(x_{21i} - 2.074)_+ + 2.667(x_{21i} - 2.333)_+ + 1.4x_{31i} + 1.280(x_{31i} - 6.29)_+ + 0.985(x_{31i} - 6.46)_+ + 0.094(x_{31i} - 7.22)_+ + 0.414x_{41i} - 0.233(x_{41i} - 5.87)_+ + 0.783(x_{41i} - 6.28)_+ + 2.545(x_{41i} - 8.06)_+ + 0.78x_{51i} + 1.339(x_{51i} - 68.71)_+ + 0.025(x_{51i} - 69.27)_+ + 3.758(x_{51i} - 71.7)_+ \quad (14)$$

With segmentation can be formed for each predictor variable. If $x_{21i} \leq 2.014$; $x_{31i} \leq 6.29$; $x_{41i} \leq 5.87$; $x_{51i} \leq 68.71$ then the prediction of Gini Ratio in Yogyakarta City is presented in Equation (15).

$$\hat{y}_{1i} = \begin{cases} -0.011 + 0.396x_{11i} + g(x) & ; \quad x_{11i} \leq 86.66 \\ 109.11 - 0.863x_{11i} + g(x) & ; \quad 86.66 < x_{11i} \leq 86.97 \\ 156.491 - 1.408x_{11i} + g(x) & ; \quad 86.97 < x_{11i} \leq 88.29 \\ 276.367 - 2.766x_{11i} + g(x) & ; \quad x_{11i} > 88.29 \end{cases} \quad (15)$$

where $g(x) = -0.119x_{21i} + 1.4x_{31i} + 0.414x_{41i} + 0.78x_{51i}$.

Based on Equation (15), the interpretation is obtained when HDI is less than 86.66 if the HDI increase by 1 unit, it will increases the Gini Ratio index by 0.396. When HDI lies between 86.66 to 86.97, every one point increase in HDI tends to decrease the Gini Ratio index by 0.863. When HDI lies between 86.97 to 88.29 then every 1 increase in HDI tends to decrease the Gini Ratio index by 1.408. Meanwhile, when HDI is more than 88.29, every 1 increase in HDI tends to decrease the Gini Ratio index by 2.766. Equation (15) shows that HDI has a consistently negative effect on the Gini Ratio across all segments, although the magnitude varies depending on the HDI interval. This pattern indicates that improvements in human development through education, health, and living standards play a substantial role in reducing inequality. The findings align with [30], who demonstrate that variations in human development strongly affect inequality, particularly in lower- and middle-income regions where disparities in education dominate. Rachmawatie [31] further confirms the relevance of HDI by showing that it significantly contributes to changes in inequality across districts in Indonesia. The consistency of these results across studies underscores the importance of strengthening human development policies as an inequality-reduction strategy.

If $x_{11i} \leq 86.66$; $x_{31i} \leq 6.29$; $x_{41i} \leq 5.87$; $x_{51i} \leq 68.71$ then the prediction of Gini Ratio in Yogyakarta City is presented in Equation (16).

$$\hat{y}_{1i} = \begin{cases} -0.011 - 0.119x_{21i} + g(x) & ; \quad x_{21i} \leq 2.014 \\ 24.14 - 0.239x_{21i} + g(x) & ; \quad 2.014 < x_{21i} \leq 2.074 \\ 78.076 - 0.499x_{21i} + g(x) & ; \quad 2.074 < x_{21i} \leq 2.333 \\ 15.858 - 0.232x_{21i} + g(x) & ; \quad x_{21i} > 2.333 \end{cases} \quad (16)$$

where $g(x) = 0.396x_{11i} + 1.4x_{31i} + 0.414x_{41i} + 0.78x_{51i}$.

Based on Equation (16), the interpretation is obtained when CMW is less than 2.014 if the CMW increase by 1 unit, it will decrease the Gini Ratio index by 0.119. When CMW lies between 2.014 to 2.074, every one point increase in CMW tends to decrease the Gini Ratio index by 0.239. When CMW lies between 2.074 to 2.333 then every 1 increase in CMW tends to decrease the Gini Ratio index by 0.499. Meanwhile, when CMW is more than 2.333, every 1 increase in CMW tends to decrease the Gini Ratio index by 0.232. Equation (16) highlights that increases in CMW also reduce income inequality across all wage intervals. These results support the idea that minimum wage policies have redistributive effects that benefit lower-income workers. Li et al. [32] similarly report robust evidence that increases in minimum wages significantly reduce inequality among migrant populations in China. Barford et al. [33] emphasize that living wages also contribute to poverty alleviation and inequality reduction. Overall, the findings reaffirm the central role of wage regulation in improving economic equity.

If $x_{11i} \leq 86.66$; $x_{21i} \leq 2.014$; $x_{41i} \leq 5.87$; $x_{51i} \leq 68.71$ then the prediction of Gini Ratio in Yogyakarta City is presented in Equation (17).

$$\hat{y}_{1i} = \begin{cases} -0.011 + 1.4x_{31i} + g(x) & ; \quad x_{31i} \leq 6.29 \\ -8.059 + 2.697x_{31i} + g(x) & ; \quad 6.29 < x_{31i} \leq 6.46 \\ -14.427 + 3.664x_{31i} + g(x) & ; \quad 6.46 < x_{31i} \leq 7.22 \\ -15.105 + 3.758x_{31i} + g(x) & ; \quad x_{31i} > 7.22 \end{cases} \quad (17)$$

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 9 Issue 12 December - 2025, Pages: 80-88

where $g(x) = 0.396x_{11i} - 0.119x_{21i} + 0.414x_{41i} + 0.78x_{51i}$.
Based on Equation (17), the interpretation is obtained when PPP is less than 6.29 if the PPP increase by 1 unit, it will increase the Gini Ratio index by 1.4. When PPP lies between 6.29 to 6.46, every one point increase in PPP tends to increase the Gini Ratio index by 2.697. When PPP lies between 6.46 to 7.22 then every 1 increase in PPP tends to increase the Gini Ratio index by 3.664. Meanwhile, when PPP is more than 7.22, every 1 increase in PPP tends to increase the Gini Ratio index by 3.758. Equation (17) reveals a positive relationship between PPP and the Gini Ratio. Across all intervals, higher rates of poverty are associated with increases in income inequality. The influence becomes stronger as the proportion of poor individuals increases, demonstrating that regions with higher poverty burdens experience disproportionately higher inequality. This is consistent with Min et al. [34], who discuss how poverty, inequality, and income growth are intertwined. Their work illustrates that high inequality exacerbates the challenges of poverty eradication, particularly in countries with lower GDP per capita. The results of the present research emphasize the need for targeted poverty reduction programs as an integral part of inequality mitigation.

If $x_{11i} \leq 86.66$; $x_{21i} \leq 2.014$; $x_{31i} \leq 6.29$; $x_{51i} \leq 68.71$ then the prediction of Gini Ratio in Yogyakarta City is presented in Equation (18).

$$\hat{y}_{1i} = \begin{cases} -0.011 + 0.414x_{41i} + g(x) & ; \quad x_{41i} \leq 5.87 \\ 1.359 + 0.181x_{41i} + g(x) & ; \quad 5.87 < x_{41i} \leq 6.28 \\ -3.561 + 0.964x_{41i} + g(x) & ; \quad 6.28 < x_{41i} \leq 8.06 \\ -24.080 + 3.509x_{41i} + g(x) & ; \quad x_{41i} > 8.06 \end{cases} \quad (18)$$

where $g(x) = 0.396x_{11i} - 0.119x_{21i} + 1.4x_{31i} + 0.78x_{51i}$.
Based on Equation (18), the interpretation is obtained when OUR is less than 5.87 if the OUR increase by 1 unit, it will increase the Gini Ratio index by 0.414. When OUR lies between 5.87 to 6.28, every one point increase in OUR tends to increase the Gini Ratio index by 0.181. When OUR lies between 6.28 to 8.06 then every 1 increase in OUR tends to increase the Gini Ratio index by 0.964. Meanwhile, when OUR is more than 8.06, every 1 increase in OUR tends to increase the Gini Ratio index by 3.509. Equation (18) indicates that higher levels of open unemployment are consistently associated with increases in income inequality across all knot segments. Rolim et al. [35] highlight that in certain labor market conditions, rising unemployment tends to reinforce disparities because job losses are disproportionately concentrated among low-skilled and low-income workers, while higher-income groups remain largely insulated from labor market shocks. Consequently, unemployment amplifies pre-existing structural disadvantages and widens distributional gaps rather than narrowing them. This alignment with inequality-augmented Phillips curve dynamics underscores how labor market imbalances can produce inequality-increasing effects even in contexts where inflationary pressures are subdued.

If $x_{11i} \leq 86.66$; $x_{21i} \leq 2.014$; $x_{31i} \leq 6.29$; $x_{41i} \leq 5.87$ then the prediction of Gini Ratio in Yogyakarta City is presented in Equation (19).

$$\hat{y}_{1i} = \begin{cases} -0.011 + 0.78x_{51i} + g(x) & ; \quad x_{51i} \leq 68.71 \\ -92.026 + 2.119x_{51i} + g(x) & ; \quad 68.71 < x_{51i} \leq 69.27 \\ -93.766 + 2.144x_{51i} + g(x) & ; \quad 69.27 < x_{51i} \leq 71.7 \\ 175.667 + 5.902x_{51i} + g(x) & ; \quad x_{51i} > 71.7 \end{cases} \quad (19)$$

where $g(x) = 0.396x_{11i} - 0.119x_{21i} + 1.4x_{31i} + 0.414x_{41i}$.

Based on Equation (19), the interpretation is obtained when LFPR is less than 68.71 if the LFPR increase by 1 unit, it will increase the Gini Ratio index by 0.78. When LFPR lies between 68.71 to 69.27, every one point increase in LFPR tends to increase the Gini Ratio index by 2.119. When LFPR lies between 69.27 to 71.7 then every 1 increase in LFPR tends to increase the Gini Ratio index by 2.144. Meanwhile, when LFPR is more than 71.7, every 1 increase in LFPR tends to increase the Gini Ratio index by 5.902. Equation (19) indicates that LFPR has a positive association with income inequality. This finding may reflect structural labor market characteristics, such as disparities in job opportunities or productivity across sectors. Clark [36] shows that inequality is shaped by factors such as sectoral employment composition, labor supply dynamics, gender participation, government size, and trade flows. The present research supports these insights by illustrating that higher labor force participation does not automatically translate into equitable income distribution without parallel improvements in job quality and wage structures.

4. Conclusions

This research analyzes income inequality in Yogyakarta from 2020 to 2024 using truncated spline nonparametric regression. The results demonstrate that key socio-economic factors, such as the Human Development Index (HDI), minimum wage, poverty rate, labor force participation, and unemployment significantly influence the Gini Ratio. The optimal model selected based on Generalized Cross Validation (GCV), employs three knots, yielding the lowest GCV value of 0.00045, MSE value of 0.102, and a high determination coefficient R-Square of 99.53%. This finding indicates that the truncated spline method effectively captures the nonlinear relationship in income distribution, providing a robust analytical framework for economic inequality assessment. Notably, Yogyakarta City exhibits the highest Gini Ratio, indicating severe economic disparity. Therefore, targeted policy interventions are crucial to fostering more equitable economic growth.

5. References

[1] T. Haruyama, “A Schumpeterian Exploration of Gini and Top/Bottom Income Shares,” Macroeconomic Dynamics, vol. 29, pp. 1–33, 2024, doi: 10.1017/S1365100524000403

[2] Badan Pusat Statistik, Statistik Indonesia Statistical Yearbook of Indonesia 2024. Jakarta: Badan Pusat Statistik, 2024.

[3] Badan Pusat Statistik Provinsi DI Yogyakarta, Statistik Daerah Daerah Istimewa Yogyakarta 2024. Bantul: BPS Provinsi DI Yogyakarta, 2024.

[4] P. De Agostini, A. Paulus, H. Sutherland, and I. Tasseva, The effect of tax-benefit changes on the income distribution in EU countries since the beginning of the economic crisis, EUROMOD Working Paper Series, 2017.

[5] D. Susilo and Z. Arifin, “Pengaruh program keluarga harapan (PKH) terhadap kesejahteraan masyarakat di Kabupaten Sleman,” Jurnal Ilmu Sosial dan Politik, vol. 23, no. 3, pp. 198–208, 2020.

[6] R. D. Andina, Jajang, and Supriyanto, “Analisis faktor-faktor yang memengaruhi ketimpangan distribusi pendapatan di Pulau Jawa tahun 2014–2020,” Jurnal Ilmiah Matematika dan Pendidikan Matematika, vol. 13, no. 1, pp. 1–14, 2021.

[7] M. A. Wibowo and S. P. Pangestuty, “Analisis pengaruh program bantuan pemerintah terhadap ketimpangan pendapatan di Yogyakarta menggunakan pendekatan multivariate adaptive regression spline (MARS),” Jurnal Ekonomi dan Kebijakan Publik, vol. 10, no. 2, pp. 145–160, 2023.

[8] R. L. Eubank, C. Huang, Y. M. Maldonado, N. Wang, S. Wang, and R. J. Buchanan, “Smoothing spline estimation in varying-coefficient models,” Journal of the Royal Statistical Society, ser. B, vol. 66, no. 3, pp. 653–667, 2004, doi: 10.1111/j.1467-9868.2004.B5595.x

[9] G. Nurhuda, Wasono, and D. A. Nohe, “Nonparametric regression modeling based on spline truncated estimator on simulation data,” Jurnal Matematika, Statistika dan Komputasi, vol. 19, no. 1, pp. 172–182, 2022, doi: 10.20956/jv19i1.21534

[10] N. P. A. M. Mariati, I. W. Sudiarsa, and G. A. M. A. Putri, “Estimator truncated spline dan sifat liniernya dalam regresi nonparametrik multivariabel,” Emasains, vol. 13, no. 2, pp. 19–24, 2024, doi: 10.59672/emasains.v13i2.4048

[11] A. T. R. Dani and N. Y. Adrianingsih, “Pemodelan regresi nonparametrik dengan estimator spline truncated vs deret Fourier,” JJOM, vol. 3, no. 1, pp. 26–36, 2021, doi: 10.34312/jjom.v3i1.7713

[12] P. D. Amelia, “Analisis faktor-faktor yang memengaruhi gini rasio di Provinsi Jawa Barat menggunakan regresi nonparametrik spline truncated,” Master’s thesis, Institut Teknologi Sepuluh Nopember, Surabaya, 2023.

[13] A. Cascarano, J. Mur-Petit, J. H. Gonzalez, M. Camacho, N. T. Eadie, P. Gkontra, M. C. Hyam, J. Vitria, and K. Lekadir, “Machine and deep learning for longitudinal biomedical data: A review of methods and applications,” Springer, vol. 56, pp. 1711–1771, 2023, doi: 10.1007/s10462-023-10561-w

[14] P. J. Diggle, P. J. Heagerty, K. Y. Liang, and S. L. Zeger, Analysis of Longitudinal Data. New York: Oxford University Press, 2002.

[15] I. Sriliana, I. N. Budiantara, and V. Ratnasari, “A truncated spline and local linear mixed estimator in nonparametric regression for longitudinal data and its application,” Symmetry, vol. 14, no. 12, p. 2687, 2022, doi: 10.3390/sym14122687

[16] A. Ebru, “Using machine learning to analyze longitudinal data: A tutorial guide and best-practice recommendations for social science researchers,” Applied Psychology, vol. 72, no. 3, pp. 1339–1364, 2022, doi: 10.1111/apps.12435

[17] M. A. D. Octavanny, I. N. Budiantara, H. Kuswanto, and D. P. Rahmawati, “Nonparametric regression model for longitudinal data with mixed truncated spline and Fourier series,” Abstract and Applied Analysis, vol. 2020, p. 4710745, 2020, doi: 10.1155/2020/4710745

[18] N. Chamidah, B. Lestari, I. N. Budiantara, and D. Aydin, “Estimation of multiresponse multipredictor nonparametric regression model using mixed estimator,” Symmetry, vol. 16, no. 4, p. 386, 2024, doi: 10.3390/sym16040386

[19] K. Takezawa, Introduction to Nonparametric Regression. New Jersey: John Wiley & Sons, 2006.

[20] M. Ramli, I. N. Budiantara, and V. Ratnasari, “Statistical inference on nonparametric regression model with approximation of Fourier series function: Estimation and hypothesis testing,” MethodsX, vol. 13, p. 102922, 2024, doi: 10.1016/j.mex.2024.102922

[21] R. Putra and M. G. Fadhlurrahman, “Determination of the best knot and bandwidth in geographically weighted truncated spline nonparametric regression using generalized cross validation,” MethodsX, vol. 10, p. 101994, 2023, doi: 10.1016/j.mex.2022.101994

[22] M. F. F. Mardianto, Gunardi, and H. Utami, “An analysis about Fourier series estimator in nonparametric regression for longitudinal data,” Mathematics and Statistics, vol. 9, no. 4, pp. 501–510, 2021, doi: 10.13189/ms.2021.090409

[23] T. Utami, E. Maharani, and A. Fadlurohman, “Modeling DHF in Central Java using hybrid nonparametric spline truncated-Fourier series approach,” Barekeng: Jurnal Ilmu Matematika dan Terapan, vol. 18, no. 3, pp. 1459–1470, 2024, doi: 10.30598/barekengvol18iss3pp1459-1470

[24] I. Sriliana, I. N. Budiantara, and V. Ratnasari, “The performance of mixed truncated spline-local linear nonparametric regression model for longitudinal data," MethodsX, vol. 12, p. 102652, 2024, doi: 10.1016/j.mex.2024.102652

[25] D. A. Widyastuti, A. A. R. Fernandes, and H. Pramoedyo, "Spline estimation method in nonparametric regression using truncated spline approach," Journal of Physics, vol. 1872, p. 012027, 2020, doi: 10.1088/1742-6596/1872/1/012027

[26] M. Ramli, V. Ratnasari, and I. N. Budiantara, "Estimation of matrix variance-covariance on nonparametric regression spline truncated for longitudinal data," Journal of Physics, vol. 1562, p. 012014, 2019, doi: 10.1088/1742-6596/1562/1/012014

[27] M. A. D. Octavanny, I. N. Budiantara, H. Kuswanto, and D. P. Rahmawati, "Modeling of children ever born in Indonesia using Fourier series nonparametric regression," Journal of Physics, vol. 1752, p. 012019, 2019, doi: 10.1088/1742-6596/1752/1/012019

[28] A. Areveeporn, "The estimating parameter and number of knots for nonparametric regression methods in modelling time series data," Modelling, vol. 5, pp. 1413–1434, 2024, doi: 10.3390/modelling5040073

[29] S. Suliyanto, T. Saifudin, M. Rifada, and D. Amelia, "Statistical inferences and applications of nonparametric regression models based on Fourier series," MethodsX, vol. 14, p. 103217, 2025, doi: 10.1016/j.mex.2025.103217

[30] I. Permanyer and J. Smits, "Inequality in human development across the globe," Population and Development Review, vol. 46, no. 3, pp. 583–601, 2020, doi: 10.1111/padr.12343

[31] D. Rachmawatie, "The influence of human development index on the income distribution inequality in Yogyakarta," Jambura Agribusiness Journal, vol. 4, no. 2, pp. 67–75, 2023, doi: 10.37046/jaj.v4i2.13911

[32] Y. Li, J. Qi, and S. Wu, "Minimum wage and income inequality among the migrant population in China," Labor History, vol. 66, no. 2, pp. 175–189, 2024, doi: 10.1080/0023656x.2024.2371126

[33] A. Barford, A. Beales, and M. Zorila, "An expert study of systemic influences on progress towards living wages: A key to unlock the Sustainable Development Goals," Business Strategy and Development, vol. 8, no. 1, pp. 1–19, 2025, doi: 10.1002/bsd2.70048

[34] J. Min, B. Soergel, J. Kikstra, J. Koch, and B. van Ruijven, "Income and inequality pathways consistent with eradicating poverty," Environmental Research Letters, vol. 19, no. 11, p. 114041, 2024, doi: 10.1088/1748-9326/ad7b5d

[35] L. Rolim, L. B. de Carvalho, and D. Lang, "Monetary policy rules and the inequality-augmented Phillips curve," Economic Modelling, vol. 139, p. 106780, 2024, doi: 10.1016/j.econmod.2024.106780

[36] R. Clark, "Income inequality in the post-2000 era: Development, globalization, and the state," International Sociology, vol. 35, no. 3, pp. 260–283, 2020, doi: 10.1177/0268580920905464