CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (1) (2026), Pages 194-207
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Modelling Factors Affecting the Middle Income Trap in Indonesia Using Generalized Additive Models (GAM)

Dita Amelia*, Suliyanto, Azizah Atsariyyah Zhafira, Aulia Ramadhanti, Billy Christandy Suyono, and Firqa Aqila Hizbullah

Department of Mathematics, Statistics Study Program, Faculty of Science and Technology, Airlangga University, Surabaya, Indonesia

Abstract

Indonesia is currently facing the risk of the Middle Income Trap (MIT), a condition in which economic growth stagnates after reaching middle-income status. This study aims to identify and model socio–economic factors affecting MIT at the provincial level in Indonesia during 2020–2023. The Generalized Additive Model (GAM) is employed to capture nonlinear and heterogeneous relationships between predictors and GRDP per capita with complex patterns that conventional linear or parametric models often fail to detect. The use of GAM in this context represents a methodological contribution, as studies applying GAM for MIT analysis in Indonesia remain very limited. This research therefore introduces a novel analytical approach by demonstrating how GAM can reveal flexible functional relationships and uncover nonlinear effects that are overlooked by traditional panel regression. GRDP per capita is modeled using six predictors: life expectancy, poverty rate, informal employment share, upper secondary education completion, food insecurity prevalence, and population density. The best model is obtained using the Gaussian family with an identity link, with five predictors showing nonlinear effects and food insecurity exhibiting a negative linear influence. The selected model demonstrates strong performance, indicated by an AIC value of 2743.279 and a R^2 of 98.6%, suggesting a very high explanatory power. In addition, the model achieves good predictive accuracy, with a MAPE of 8.04%. The findings support evidence-based policies aligned with Sustainable Development Goal (SDG) 8, promoting inclusive and sustainable economic growth. .

Keywords: Generalized Additive Model; Goodness of Fit; GRDP per capita; Middle Income Trap.

1. Introduction

Indonesia, as the fourth most populous country in the world, envisions becoming a high-income nation by 2045, as outlined in its national long-term agenda known as The Golden Indonesia 2045. Achieving this vision requires not only sustained and inclusive economic growth but also equitable development across all provinces. One of the most pressing economic challenges facing Indonesia is the risk of falling into the Middle Income Trap (MIT), a condition where a country stagnates economically after reaching a middle-income level, making it difficult to transition to a high-income status [1, 2]. According to projections from Indonesia's Badan Perencanaan

*Corresponding author. E-mail: ditaamelia@fst.unair.ac.id

Submitted: July 01, 2025 Reviewed: August 08, 2025 Accepted: January 08, 2026
DOI: https://doi.org/10.18860/cauchy.v11i1.35119

Pembangunan Nasional (Bappenas), the country must sustain annual GDP growth of 6–8% to avoid MIT and fulfill its development goals. However, when compared to other large population countries such as China and India, Indonesia has not achieved 8% economic growth in the past 16 years. Furthermore, persistent structural disparities among provinces particularly in gross regional domestic product (GRDP) per capita raise concerns about uneven economic resilience and inclusive development.

Based on the World Bank classification, Indonesia is currently categorized as an upper middle-income country and, like other countries in this category, faces complex challenges in avoiding MIT, especially at the provincial level [3]. One of the key indicators in identifying potential MIT is the GRDP per capita, which reflects the average income generated by a region and serves as a proxy in measuring the economic well-being of a region. In 2024, Indonesia recorded a per capita income of USD 4,960.3, still far from the developed country threshold of USD 12,535, and reflecting a position that is vulnerable to MIT [4]. The large inequality between provinces in terms of GRDP per capita indicates that not all regions have equal competitiveness and growth capacity, which in aggregate may slow down Indonesia's transition to a high-income country. Therefore, the use of provincial GRDP per capita data as a MIT indicator is crucial in understanding regional economic contribution and resilience. This study focuses on the period 2020 to 2023 as it reflects the significant dynamics of the COVID-19 pandemic and the uneven economic recovery process between regions, providing important context in analyzing regional resilience to MIT risks [5].

Previous studies have addressed various economic and social factors contributing to MIT in Indonesia using different statistical and econometric approaches. For instance, Ratnasari et al. [6] employed panel regression and identified life expectancy, gross enrollment ratio, and gross fixed capital formation as significant predictors. Dewi et al. [7] using a Vector Error Correction Model (VECM), found that exchange rates, investment, and foreign direct investment positively affect long-term GDP, while inflation had a negative impact. Other studies by Malale and Sutikno [8] revealed the negative effects of exports, agricultural value-added, and foreign aid on gross national income per capita. While these works have laid a valuable foundation for understanding MIT in Indonesia, many still rely on parametric methods that assume linear relationships and fixed functional forms. These assumptions may limit the ability to capture complex and nonlinear interactions between economic variables, particularly at the provincial level where structural differences are pronounced.

To address this gap, the present study proposes a nonparametric modeling approach to investigate the determinants of MIT in Indonesia using the Generalized Additive Model (GAM). GAM is well suited for economic data that exhibit nonlinear and regionally diverse patterns because it is very flexible, does not impose a predetermined functional form and allows the estimation of a smooth relationship between predictors and response variables, and the smoothing function in GAM helps reduce overfitting and provides more accurate estimates, especially in provincial panel data that have significant variation between regions [9–11]. By modeling GRDP per capita across Indonesia's 34 provinces from 2020 to 2023, this study captures not only spatial variations but also dynamic effects in the post-pandemic recovery period an era marked by structural transformation and socioeconomic realignment across regions.

The novelty of this research lies in the application of GAM to panel data at the provincial level for MIT analysis an approach that remains underexplored in Indonesian economic literature. This method allows for flexible modeling of nonlinear effects from key social and economic indicators such as life expectancy, poverty rate, informal labor share, secondary education completion rate, food insecurity prevalence, and population density. Unlike previous studies that treated these variables linearly or assumed constant effects, this work reveals how their influence on GRDP per capita varies across regions and over time. The findings offer valuable insights for policymakers aiming to design targeted interventions that prevent MIT and promote sustainable regional growth.

The objectives of this research are (1) to identify the most significant social and economic factors affecting MIT risk with the GAM model (3) by considering these various factors, it is expected to contribute to efforts to achieve Sustainable Development Goal (SDG) point 8, namely Decent Work and Economic Growth, (3) can be the basis for formulating more effective policies in various aspects, including economic, social, and environmental, as well as increasing regional competitiveness, reducing social inequality, and accelerating Indonesia's transition to a high-income country.

2. Methods

2.1. Data Source

This study uses secondary data obtained from the official website of the Central Bureau of Statistics (BPS) at `https://www.bps.go.id/id`. The dataset covers the period from 2020 to 2023 across 34 provinces in Indonesia, resulting in a total of 136 observations.

The response variable (Y) in this study is the Gross Regional Domestic Product (GRDP) per capita at constant prices, which is used as a proxy for the risk of Middle Income Trap (MIT). Conceptually, MIT refers to a situation where a region or country experiences stagnation in income growth, preventing it from reaching high-income status. In this study, lower or stagnant GRDP per capita indicates a higher risk of MIT, while higher growth in GRDP per capita suggests lower risk. Thus, GRDP per capita effectively captures the economic performance related to MIT at the provincial level.

The predictor variables (x_1 to x_6) include: life expectancy, percentage of poor population, proportion of informal employment, completion rate of upper secondary education, prevalence of food insecurity, and population density. These variables are selected to represent socio-economic and demographic factors potentially associated with the risk of MIT.

2.2. Longitudinal Data Analysis

Longitudinal data is a data obtained through repeated observations at different times of several objects [12]. Longitudinal data can be used to see changes and variations in changes between individuals. Unlike panel data, the time of repeated observations made on longitudinal data does not have to be the same. Suppose there are $i = 1, 2, \ldots, m$ individuals who each have repeated observations $j = 1, 2, \ldots, n_i$ with observation time t_{ij}. Then the longitudinal data regression model is generally expressed as follows.

$$Y_{it} = \beta_{i0} + \beta_{i1}X_{1it} + \beta_{i2}X_{2it} + \cdots + \beta_{ij}X_{ijt} + \varepsilon_{it}$$

Where :

- Y_{it} is response variable of j-th subject and t-th time;
- X_{ijt} is predictor variable of i-th subject from j-th observations and t-th time;
- β_{ij} is regression coefficient of i-th subject from j-th observation.

The model can be written simpler in a matrix form below.

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

Where $\mathbf{Y}_i$ is a matrix with $1 \times n_i$ size; $\mathbf{X}_i$ is a matrix with $n_i \times p$ size; $\boldsymbol{\beta}$ is a vector with $1 \times p$ size; $\boldsymbol{\varepsilon}_i$ is error vector for individual i.

In longitudinal modelling, $\boldsymbol{\varepsilon}_i$ is typically assumed to follow a multivariate normal distribution with mean zero and a covariance structure that captures within-subject correlation, such as independent, autoregressive (AR), compound symmetry, or unstructured covariance. Additionally, the variance of $\boldsymbol{\varepsilon}_i$ may be assumed to be homoskedastic or heteroskedastic across measurement times. These assumptions are essential because misspecifying the correlation or variance structure may lead to inefficient parameter estimates and invalid statistical inference.

2.3. Generalized Additive Model

The Generalized Additive Model (GAM) is an extension of the additive model. This model allows the distribution of the response variable to come from the exponential family. It is also referred to as an extension of the generalised linear model, as it replaces linear predictors with additive ones, making the generalised additive model more flexible than both the generalised linear model and the additive model [13].

The generalized additive model consists of several random components, namely the response variable, fixed components represented by the additive predictors, and a link function that connects these two components. In the context of this study, GAM is particularly relevant because socio-economic indicators related to the Middle Income Trap such as life expectancy, poverty rate, education attainment, and informal employment often exhibit nonlinear and regionally heterogeneous relationships with GDRP per capita. The flexibility of GAM allows these complex patterns to be captured accurately, making it suitable for modeling provincial disparities in Indonesia. The response variable as a random component is assumed to have an exponential family density function [14].

This structure enables Generalized Additive Model not only to model nonlinear relationships between the predictors and the response variable while preserving interpretability, but also to provide greater flexibility compared to the Generalized Linear Model (GLM) or multiple linear regression, which assume strictly linear relationships. By incorporating smooth functions, GAM allows the data to determine the functional form, thereby reducing the risk of model misspecification and improving the model's ability to capture complex economic patterns. The general form of this additive model is as follows [15]:

$$g(u_i) = \boldsymbol{X_i\theta} + f_1(X_{1i}) + f_2(X_{2i}) + \cdots + f_j(X_{ji})$$

where $u_i = E(Y_i)$ and $Y_i \sim$ exponential distribution family.
with:

- g is a monotonic smoothing link function;
- $\boldsymbol{X_i}$ is the i-th row of the predictor variable value matrix for the parametric component;
- $\boldsymbol{\theta}$ is the vector of parameters for the parametric model;
- f_j is the smoothing function of X_{ji}.

The smoothing function used in this model is a cubic regression spline. However, it is not mandatory for every independent variable to use the same spline order. In GAM, the choice of smoothing basis, whether cubic splines, thin plate splines, B splines, or other variants, can differ for each predictor depending on its underlying data pattern and required flexibility. The decision to use cubic splines in this study was based on their stability, interpretability, and common usage in GAM applications.

The construction of spline smoothers also requires specifying the number and placement of knots, which determine the flexibility of the smoothing curve. A greater number of knots allows the function to capture more complex nonlinear patterns, whereas fewer knots produce smoother, more general structures. The role of knots is therefore central in shaping the smooth term within GAM. Detailed discussions on the concept of knots and their influence on spline-based smoothing [15].

2.4. Classical Assumption Test

Before constructing the GAM, classical assumption tests are carried out to ensure that the selected model specification is appropriate for the characteristics of the data. These tests are important because they help determine whether the relationship among variables is linear or nonlinear and whether the predictors exhibit multicollinearity, with two aspects influencing whether a smoothing function is needed in GAM and how stable the parameter estimates will be.

The linearity test is conducted to assess whether the relationship between the independent and dependent variables follows a linear pattern. The Ramsey RESET test is commonly used for this purpose, with the hypothesis:
H_0 : Model is correctly specified and the relationship is linear
H_1 : Model presence of nonlinearity

A p-value greater than 0.05 leads to the acceptance of H_0, suggesting linearity, whereas a p-value less than 0.05 indicates nonlinearity [16]. In the context of GAM, evidence of nonlinearity provides the justification for applying smoothing functions to model flexible relationships between predictors and GRDP per capita.

The multicollinearity test is performed to evaluate the degree of intercorrelation among predictor variables, as high multicollinearity can lead to unstable estimates and unreliable interpretations. This is assessed using the Variance Inflation Factor (VIF), with values below 10 generally indicating that multicollinearity is not problematic [17]. Ensuring the absence of multicollinearity is essential so that each predictor contributes uniquely to the model and the smoothing components in GAM can be interpreted properly.

2.5. Model Goodness Test

The goodness-of-fit test is essential for evaluating whether the selected model accurately represents the observed data. This process includes two main evaluations: testing model coefficient determination (R^2) and assessing the smoothing basis dimension.

The smoothing basis dimension, on the other hand, determines the degree of flexibility in the model's smoothing function. An underspecified basis dimension may lead to excessive smoothing, masking essential data patterns and introducing structure into the residuals. The adequacy of the basis can be assessed using the k-index, calculated as the ratio of the residual variance estimate to its expected value. A k-index greater than 1 suggests that the model has sufficient flexibility to capture underlying data structures [18]. In general, the formula to calculate the R^2 value is given as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \hat{y}_{it})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \bar{y})^2}$$

The AIC value is calculated using the following formula:

$$\text{AIC} = 2k - 2\ln(\hat{L})$$

where k is the number of estimated parameters in the model and $\hat{L}$ is the MLE for the model. Meanwhile, the MAPE formula for the GAM model is as follows:

$$\text{MAPE} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{|y_{it} - \hat{y}_{it}|}{y_{it}}\right]$$

where y_{it} is the actual value of the response variable (panel individual time), $\hat{y}_{it}$ is the predicted value from the GAM model, and N is the total number of observations. A model with the minimum AIC and MAPE, but the largest coefficient of determination, can be considered the best model [15].

2.6. Research Stage

The steps for analyzing middle income trap data using the GAM approach in this research are as follows:

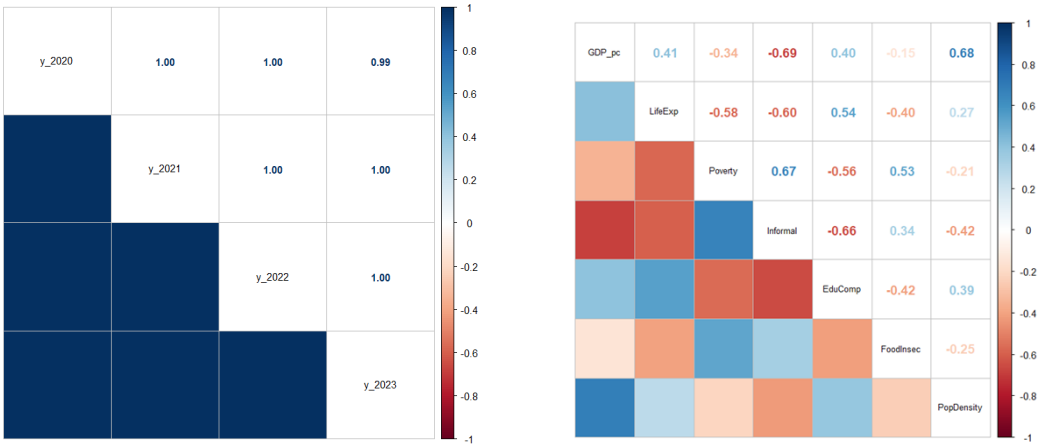
1. Identify the variables that contribute to the middle income trap in Indonesia.
2. Create descriptive statistics of the middle income trap data in Indonesia that has been obtained to analyse the description or characteristics of the response variables and predictor variables.

3. Conduct linearity tests using Ramsey's RESET test and multicollinearity tests.
4. Model the middle income trap data in Indonesia using the Generalized Additive Model (GAM) method.
5. Conduct model feasibility tests to evaluate whether the Generalized Additive Model is appropriate.
6. Interpret the results of the Generalized Additive Model applied to the Middle Income Trap data in Indonesia.

3. Results and Discussion

3.1. Exploratory Data Analysis

This section provides an overview of the dataset through descriptive and visual exploratory techniques. It aims to assess the temporal consistency of GDP per capita (2020–2023) and examine the correlation structure among key explanatory variables to identify underlying patterns and potential multicollinearity.



(a) Correlation matrix of GRDP per capita across Indonesian provinces from 2020 to 2023.

(b) Correlation matrix between GRDP per capita and predictors.

Fig. 1: (a) and (b) Visual exploratory analysis of GRDP and explanatory variables across Indonesian provinces.

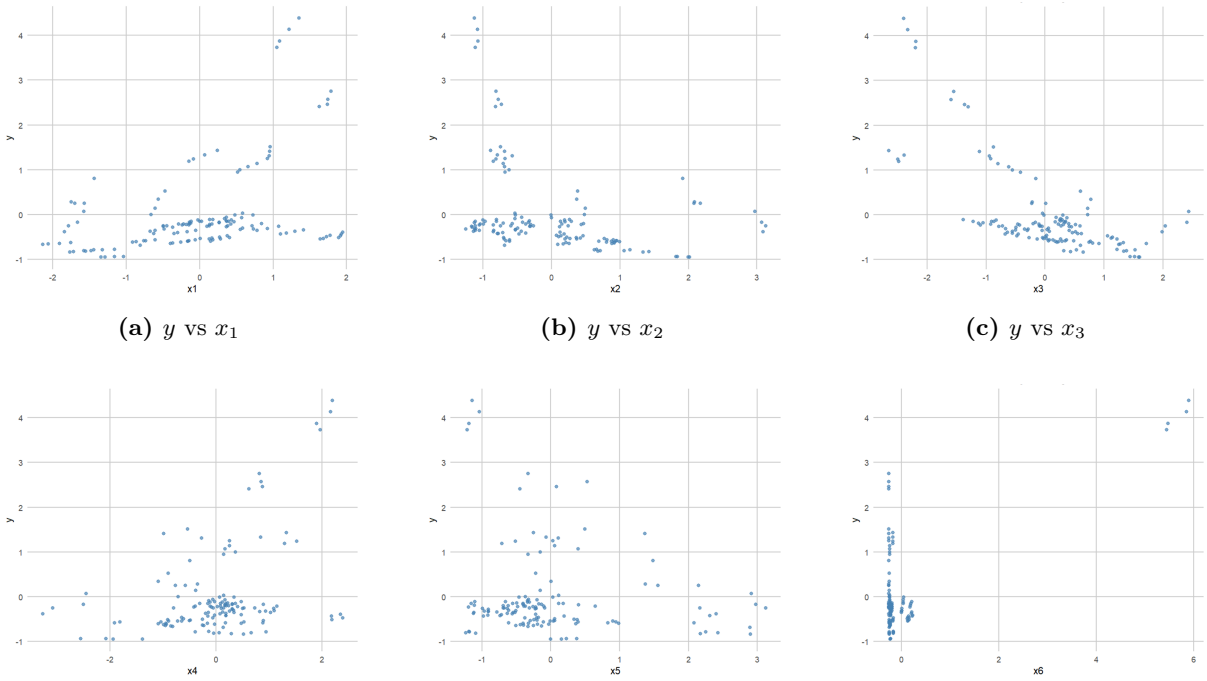
To evaluate the temporal consistency of the response variable (Fig. 1a), a pairwise correlation analysis of GDP per capita from 2020 to 2023 was conducted across Indonesian provinces. The results show very high correlations ($r = 0.99$–1.00), indicating remarkable stability in provincial economic performance over time. While this suggests strong temporal dependence, it may limit the role of time in longitudinal modeling unless addressed through differencing or fixed-effects methods.

Subsequently, the correlation matrix (Fig. 1b) was examined to identify redundancy and interaction patterns among the explanatory variables. The matrix incorporates both Pearson correlation coefficients and color gradients to visually represent the strength and direction of relationships.

Key findings reveal that GDP per capita is strongly negatively correlated with informal employment ($r = -0.69$), and positively associated with population density ($r = 0.68$) and life expectancy ($r = 0.41$), suggesting that more developed provinces tend to be more urbanized, healthier, and less dependent on informal labor. Life expectancy also shows a positive relationship with educational attainment ($r = 0.54$), and negative associations with poverty ($r = -0.58$) and informal employment ($r = -0.60$), reflecting the interconnected nature of health, education, and

labor formality.

Poverty is positively correlated with informal employment ($r = 0.67$), and negatively associated with education and life expectancy, indicating structural socioeconomic disadvantages. Informality itself exhibits strong negative relationships with GDP per capita and education, reinforcing its role as an indicator of underdevelopment. Furthermore, food insecurity is moderately associated with poverty and informality, while being inversely related to education and life expectancy. Lastly, population density correlates positively with GDP per capita and education, and negatively with informality, underscoring disparities between urban and rural regions.



(a) y vs x_1 **(b)** y vs x_2 **(c)** y vs x_3

(d) y vs x_4 **(e)** y vs x_5 **(f)** y vs x_6

Fig. 2: (a)–(f) Scatter plots between response variable y and explanatory variables x_1 to x_6.

Fig. 2 presents scatter plots illustrating the relationship between GRDP per capita (Y) and six explanatory variables. Among these, population density (x_6) shows the strongest positive linear association with GRDP, followed by life expectancy (x_1) and secondary education completion (x_4), indicating that more developed regions tend to be more urbanized, healthier, and better educated. Conversely, informal employment (x_3) exhibits a strong negative relationship, suggesting that economic informality constrains per capita output. Poverty rate (x_2) shows a weak negative trend, while food insecurity (x_5) appears to have little association with GRDP. Given that some variables do not exhibit clear linearity, a Generalized Additive Model (GAM) is employed to flexibly capture potential nonlinear effects in the subsequent analysis.

3.2. Asumption Test

To determine whether each predictor variable exhibits a linear or nonlinear relationship with GRDP per capita, the Ramsey RESET test was applied. The results are summarized in Table 1, indicating which variables are best modeled using nonparametric smooth functions within the Generalized Additive Model (GAM) framework.

Table 1: Linearity test result between each predictor and response

Relationship	*P*-Value	Decision
X_1 with Y	0.03111	Model nonlinear
X_2 with Y	0.001297	Model nonlinear
X_3 with Y	5.119×10^{-10}	Model nonlinear
X_4 with Y	0.007392	Model nonlinear
X_5 with Y	0.2279	Model linear
X_6 with Y	0.002458	Model nonlinear

Based on Table 1, variables X_1, X_2, X_3, X_4, and X_6 have p-values less than 0.05, suggesting significant deviations from linearity. Consequently, these variables are more appropriately modeled using nonparametric smoothing functions. In contrast, variable X_5 has a p-value above 0.05, indicating a linear relationship with the response variable and thus can be modeled parametrically.

Before proceeding with the modeling process, it is essential to ensure that the predictor variables do not suffer from multicollinearity, which can distort the estimates and interpretation of regression coefficients. Multicollinearity is assessed using the Variance Inflation Factor (VIF), where a value greater than 10 indicates a potential issue. Table 2 presents the VIF values for each predictor variable to evaluate the presence of multicollinearity in the model.

Table 2: Variance Inflation Factor (VIF) values for each predictor variable

Variable	VIF	Result
X_1	1.809488	non-multicollinearity
X_2	2.419267	non-multicollinearity
X_3	2.704469	non-multicollinearity
X_4	2.014042	non-multicollinearity
X_5	1.504933	non-multicollinearity
X_6	1.307959	non-multicollinearity

Based on Table 2, all VIF values from the multicollinearity assumption test are less than 10, indicating that there is no high correlation among the predictor variables in the regression model. This suggests that each variable contributes uniquely to the model without redundancy, ensuring stable coefficient estimates.

Following this, a residual normality test was conducted to examine whether the residuals from the regression model are normally distributed. The Kolmogorov–Smirnov test was employed, where the data is considered to follow a normal distribution if the null hypothesis (H_0) is accepted and the p-value is greater than 0.05. The test result shows a p-value of 0.1341, indicating that H_0 is accepted and the residuals are normally distributed.

3.3. Generalized Additive Model Spesification

Based on the Ramsey RESET test, the results indicate that only X_5 exhibits a linear relationship with the response variable, while the remaining predictors show nonlinear patterns. Therefore, the Generalized Additive Model (GAM) is appropriate, as it allows linear predictors to be replaced with additive smooth components, making the generalized additive framework more flexible in capturing complex relationships. Based on Fig. 2, the relationships between y and x_1, x_2, x_3, x_4, and x_6 display nonlinear patterns; consequently, these variables are modeled using smooth functions $s(.)$. The selection of interaction terms in the GAM is not arbitrary but is grounded in the observed relationships among the variables.

The scatter plots show that both X_1 and X_4 have positive relationships with y, although the degree of dispersion and slope varies across certain ranges of values. The nonlinear interaction te(x_1, x_4) is therefore included to flexibly capture these varying effects. Meanwhile, the scatter plots between y and X_2 as well as x_3 indicate unstable negative relationships with high variability.

The interaction te(x_2, x_3) is employed to model their simultaneous effects, which cannot be adequately represented by purely additive components. Furthermore, the nonlinear interaction te(x_6, x_4) is incorporated to allow the effect of X_6 to vary flexibly across different levels of X_4. Based on these considerations, the GAM specification is formulated as follows:

$$\hat{y} = \beta_0 + \beta_5 x_5 + f_1(x_1) + f_2(x_2) + f_3(x_3) + f_4(x_4) + f_6(x_6) + f_{14}(x_1, x_4) + f_{23}(x_2, x_3) + f_{64}(x_6, x_4)$$

The model is specified based on explanatory data analysis and theoretically motivated relationships among variables. Variable x_5 is modeled linearly due to its simple association. Variables x_1, x_2, x_3, x_4, x_6 are smoothed to capture nonlinear effects, and (x_1, x_4), (x_2, x_3), (x_6, x_4) are modeled as multivariate nonlinear interactions due to their strong interdependence. Parameter estimation is conducted using a cubic regression spline basis with penalized likelihood, which controls the smoothness of the estimated functions through roughness penalties.

The results of the Generalized Additive Model are presented in Table 3 below.

Table 3: Summary of parameter estimation in the Generalized Additive Model

Parametric Coefficients				
Variable	**Estimate**	**Std. Error**	**t-value**	**p-value**
Intercept	49531.6	1950.1	25.400	$< 2 \times 10^{-16}$
x_5	-384.5	160.2	-2.401	0.0186
Approximate significance of smooth terms				
Variable	**edf**	**Ref.df**	**F**	**p-value**
$s(x_1)$	1.000	1.000	0.582	0.447662
$s(x_2)$	6.857	7.498	0.420	0.698518
$s(x_3)$	3.783	4.531	0.252	0.940453
$s(x_4)$	1.000	1.000	0.233	0.630288
$s(x_6)$	4.355	4.978	8.918	6.44×10^{-7}
te(x_1, x_4)	11.142	13.349	3.685	0.000129
te(x_2, x_3)	14.291	14.917	6.202	$< 2 \times 10^{-16}$
te(x_6, x_4)	7.444	8.105	2.199	0.034351

Based on the estimation results, the parametric components indicate that the intercept is statistically significant because p-value $< \alpha = 5\%$, representing the baseline level of the response variable. The variable x_5 has a negative estimated coefficient of -384.5 and is statistically significant, indicating that an increase in x_5 is associated with a decrease in the response variable, holding other covariates constant.

The effective degrees of freedom (edf) provide insight into the complexity of the smooth functions. The smooth terms $s(x_1)$ and $s(x_4)$ have edf values close to one, suggesting relationships that are approximately linear. In contrast, $s(x_2)$ and $s(x_3)$ exhibit edf values greater than one, suggesting the presence of nonlinear tendencies; however, these effects are not statistically significant at the 5% significance level. The smooth term $s(x_6)$ shows a significant nonlinear effect with an edf of 4.355, demonstrating a strong nonlinear association with the response variable.

The nonlinear interaction terms modeled using tensor product smooths play a substantial role in the model. The interactions te(x_1, x_4) and te(x_2, x_3) are statistically significant at the 5% level, with edf values of 11.142 and 14.291, respectively, indicating complex and non-additive interaction effects. The interaction te(x_6, x_4) is also significant with an edf of 7.444, suggesting that the nonlinear effect of x_6 on the response variable is moderated by the level of x_4.

The resulting model equation is as follows:

$$\begin{aligned} y = 49531.6 - 384.5x_5 + f_1(x_1) + f_2(x_2) + f_3(x_3) + f_4(x_4) \\ + f_6(x_6) + f_{14}(x_1, x_4) + f_{23}(x_2, x_3) + f_{64}(x_6, x_4) \end{aligned} \tag{1}$$

3.4. Goodness of Fit Test

3.4.1. Basis Dimension Adequacy (k-index)

The adequacy of the basis dimension was evaluated using the k-index obtained from the `gam.check()` procedure to ensure that the selected basis dimensions were sufficiently flexible to capture the underlying data patterns. A k-index value close to or greater than 1, accompanied by a non-significant p-value ($p > 0.05$), indicates that the chosen basis dimension (k) is adequate and does not suffer from undersmoothing.

Table 4: Basis Dimension Adequacy Result Test

Variable	k'	edf	k-index	p-value
$s(x_1)$	9.00	1.00	0.90	0.12
$s(x_2)$	9.00	6.86	0.89	0.10
$s(x_3)$	9.00	3.78	1.07	0.74
$s(x_4)$	9.00	1.00	1.04	0.66
$s(x_6)$	9.00	4.36	1.08	0.77
$te(x_1, x_4)$	22.00	11.14	0.94	0.20
$te(x_2, x_3)$	22.00	14.29	1.07	0.80
$te(x_6, x_4)$	22.00	7.44	1.02	0.54

Based on Table 4, all smooth terms in the selected GAM model exhibit k-index values close to or exceeding 1, with corresponding p-values greater than 0.05. Specifically, the smooth terms $s(x_1)$, $s(x_2)$, $s(x_3)$, $s(x_4)$, and $s(x_6)$ show k-index values ranging from 0.89 to 1.08, while their effective degrees of freedom (edf) remain well below the maximum basis dimension ($k' = 9$). This indicates that the smoothing functions are neither over-restricted nor overly flexible.

Similarly, the tensor product interaction terms te(x_1, x_4), te(x_2, x_3), and te(x_6, x_4) present k-index values between 0.94 and 1.07, with no significant evidence of insufficient basis dimensions. The edf values for these interaction terms are also substantially lower than their corresponding basis dimensions ($k' = 22$), suggesting that the model effectively captures nonlinear interaction structures without inducing excessive smoothness.

Overall, these results confirm that the selected basis dimensions are appropriate for all smooth and interaction terms in the model. Therefore, the GAM specification possesses adequate flexibility to represent the complex nonlinear relationships between socio-economic predictors and GRDP per capita, and no increase in the basis dimension is required.

3.4.2. Akaike Information Criterion (AIC)

Based on the estimation results, the selected GAM yields an AIC value of 2743.279, which is substantially lower than that of the linear GAM model (AIC = 3074.118). This indicates that the selected model provides a better balance between goodness of fit and model complexity. The improvement reflects the ability of the model to capture nonlinear relationships and interaction effects through smooth functions and tensor product terms, leading to enhanced explanatory power without excessively increasing model complexity.

3.4.3. Coefficient of Determination

Based on the model estimation results, the coefficient of determination calculated from the deviance reduction yields an R^2 value of 0.986. This indicates that the selected Generalized Additive Model is able to explain approximately 98.6% of the variability in GRDP per capita across provinces during the 2020–2023 period, while the remaining 1.4% is influenced by factors outside the model specification. This high R^2 value demonstrates that the model has a very strong explanatory capability in capturing the socio-economic determinants associated with Middle Income Trap dynamics at the provincial level.

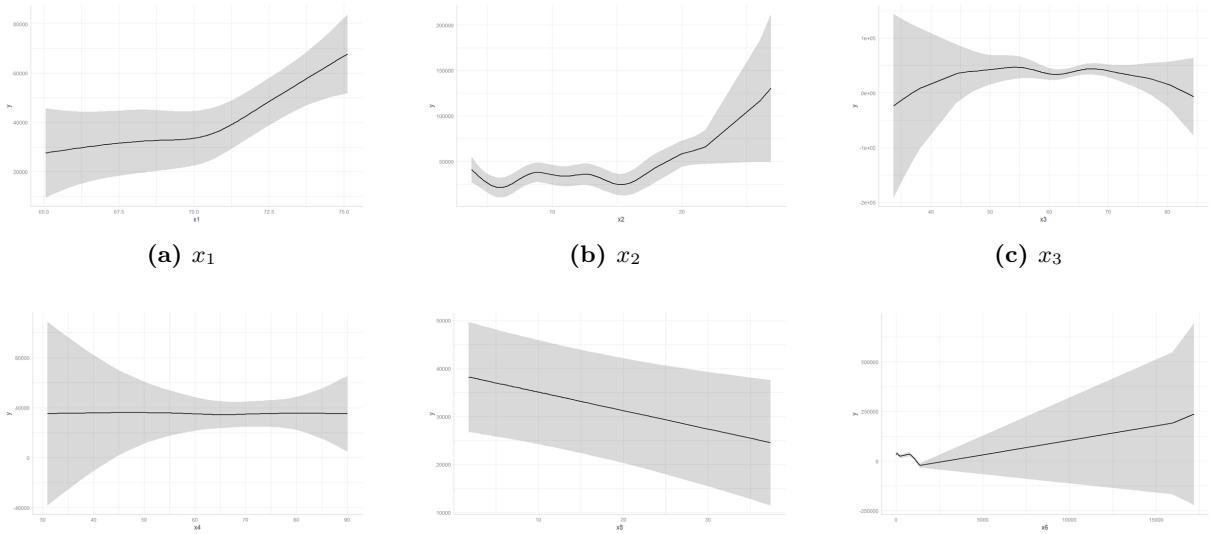
3.4.4. Prediction Accuracy (MAPE)

The prediction accuracy of the selected GAM was evaluated using the Mean Absolute Percentage Error (MAPE). Based on the estimation results, the model produces a MAPE value of 8.04%, indicating that, on average, the model's predictions deviate from the observed GRDP per capita values by approximately 8%. This level of prediction error suggests that the GAM model demonstrates good predictive performance, as MAPE values below 10% are generally classified as highly accurate in applied economic and regional modeling studies. The relatively low MAPE reflects the model's ability to capture nonlinear relationships and interaction effects among socio-economic variables, thereby producing reliable predictions across provinces with heterogeneous economic characteristics.

Overall, the obtained MAPE value confirms that the selected GAM is not only statistically well-fitted but also effective in terms of predictive accuracy, supporting its suitability for policy-relevant analysis of Middle Income Trap dynamics in Indonesia.

3.5. Interpretation of the Selected Model

Based on the results of the generalized additive model analysis for the middle income trap data in Indonesia, the following plot illustrates the estimated smooth effects of each predictor on GRDP per capita.



(a) x_1 **(b)** x_2 **(c)** x_3

(d) x_4 **(e)** x_5 **(f)** x_6

Fig. 3: Partial effect plots of explanatory variables x_1 to x_6 based on the fitted GAM model.

Based on Fig. 3 and Eq. (1), the fitted Generalized Additive Model (GAM) identifies both linear and nonlinear components in the relationship between socioeconomic factors and GRDP per capita, which is used in this study as a proxy for the Middle Income Trap (MIT). Statistical significance varies across predictors, so interpretations focus on effects supported by model output.

Among the smooth terms, only population density $s(x_6)$ exhibits a statistically significant nonlinear effect (edf $= 4.355$, $p < 0.05$), indicating a complex, non-monotonic relationship with GRDP per capita. This suggests that spatial concentration of population shapes regional economic performance and, therefore, influences the risk of MIT. Provinces with either very low or very high population density may face barriers to economic growth, either due to insufficient agglomeration or infrastructure and congestion constraints.

The parametric term x_5 (prevalence of food inadequacy) is statistically significant and enters the model linearly with a negative coefficient. This indicates that higher levels of food insecurity are associated with lower GRDP per capita, increasing the likelihood that provinces remain in

the middle-income category—directly linking a socioeconomic vulnerability to the risk of MIT.

The interaction terms between life expectancy and secondary education completion ($te(x_1, x_4)$), poverty and informal employment ($te(x_2, x_3)$), and population density and education ($te(x_6, x_4)$) are all statistically significant ($p < 0.05$). These results imply that the impact of individual socioeconomic factors on GRDP per capita—and therefore on the risk of MIT—depends on the levels of other variables, highlighting structural complementarities in provincial development.

Overall, the evidence suggests that nonlinearities in this model are primarily expressed through population density and through interactions among socioeconomic variables, rather than through multiple significant nonlinear main effects. The GAM framework is particularly useful for capturing complex determinants of MIT, showing that risk is influenced not only by single factors but by their interactions.

4. Conclusion

Based on the results of this study, it can be concluded that the Generalized Additive Model (GAM) provides a flexible and informative framework for analyzing the Middle Income Trap (MIT) at the provincial level in Indonesia during 2020–2023. The selected model, estimated under a Gaussian family with an identity link function and using automatically determined basis dimensions (effective knots: 9 for univariate smooths, 22 for bivariate interaction smooths), outperforms the linear GAM as indicated by a substantially lower AIC (2743.279 versus 3074.118). The high in-sample R^2 (0.986) and MAPE (8.04%) indicate strong goodness-of-fit, although these reflect in-sample performance only.

Among explanatory variables, prevalence of food inadequacy (x_5) exhibits a statistically significant main effect, showing a negative linear association with GRDP per capita. This implies that food insecurity directly increases the risk of provinces being trapped in the middle-income category.

With respect to nonlinear main effects, population density ($s(x_6)$) is statistically significant, indicating a complex, nonlinear relationship with GRDP per capita and hence the risk of MIT. Other smooth terms are not significant, suggesting weaker evidence for independent nonlinear effects of life expectancy, poverty, informal employment, or secondary education.

Significant nonlinear interaction effects between life expectancy and education, poverty and informal employment, and population density and education suggest that the influence of socioeconomic factors on the MIT is conditional on the levels of other variables. This emphasizes the importance of coordinated policy efforts targeting multiple dimensions of development to mitigate the risk of MIT.

Despite its strong in-sample performance, the study is limited by the short observation period (2020–2023) and potential model complexity, which could introduce overfitting. Future research should extend the time horizon, include additional structural variables, and apply out-of-sample validation or alternative models to strengthen empirical foundations for policies aimed at sustainable and inclusive economic growth consistent with SDG 8, ultimately reducing the risk of MIT across Indonesian provinces.

CRediT Authorship Contribution Statement

Dita Amelia: Conceptualization, Methodology, Writing–Original Draft. **Suliyanto:** Formal Analysis, Writing–Review, Supervision. **Azizah Atsariyyah Zhafira:** Formal Analysis, Visualization, Project Administration, Writing–Review & Editing. **Aulia Ramadhanti:** Software, Validation, Methodology. **Billy Christandy Suyono:** Data Curation, Investigation, Resources. **Firqa Aqila Hizbullah:** Visualization, Validation, Literature Review.

Declaration of Generative AI and AI-assisted technologies

In the writing of this work, Generative Artificial Intelligence (AI) tools were used solely to assist in the linguistic conversion and grammar refinement role. All major components of this work, including analysis, interpretation, and conclusions, were written independently by the authors without AI involvement in the substance of the content

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

We sincerely thank the Department of Mathematics, Universitas Airlangga, for offering the resources and assistance required to carry out this research. Their support played a crucial role in ensuring the successful completion of the project. This research would not have been accomplished without their meaningful contribution.

Data Availability

The dataset utilized in this study can be accessed publicly through the official website of Badan Pusat Statistik (BPS) Indonesia.

References

[1] S. Maryanti, P. Widayat, and N. Lubis. "Economic transformation to get out of the middle income trap condition to reach Indonesia Gold 2045". In: *ADPEBI International Journal of Business and Social Science* 3.1 (2023), pp. 63–78. DOI: 10.54099/aijbs.v3i1.356.

[2] L. Glawe and H. Wagner. "The middle-income trap: Definitions, theories and countries concerned—A literature survey". In: *Comparative Economic Studies* 58 (2016), pp. 507–538. DOI: 10.1057/s41294-016-0014-0.

[3] K. Metreau, E. Young, and S. G. Eapen. *World Bank country classifications by income level for 2024–2025*. [Online; accessed 23-March-2025]. 2024. https://blogs.worldbank.org/en/opendata/world-bank-country-classifications-by-income-level-for-2024-2025.

[4] D. N. Prasetiani, H. F. Anindya, and A. S. Yoshe. "Strategi menghadapi middle income trap: Dampak hilirisasi mineral terhadap pendapatan negara Indonesia era Joko Widodo". In: *Indonesia Foreign Policy Review* 11.1 (2024). DOI: 10.5281/zenodo.14562414.

[5] V. Wanggai, M. Delanova, and Y. M. Yani. "Stabilitas Ekonomi Indonesia Dalam Pandemi Covid-19 Dan Potensi Indonesia Untuk Terjebak Middle Income Trap". In: *Jurnal Academia Praja : Jurnal Magister Ilmu Pemerintahan* 6.1 (2023), pp. 146–165. DOI: 10.36859/jap.v6i1.1424.

[6] V. Ratnasari, S. H. Audha, and A. T. R. Dani. "Statistical modeling to analyze factors affecting the middle-income trap in Indonesia using panel data regression". In: *MethodsX* 11.102379 (2023). DOI: 10.1016/j.mex.2023.102379.

[7] R. K. Dewi, D. E. Sari, and D. Wahyuningsih. "Analisis Makro Ekonomi Sebagai Langkah Indonesia Keluar Dari Middle Income Trap". In: *Inspire Journal: Economics and Development Analysis* 1.1 (2021), pp. 99–111.

[8] A. W. Malale and M. A. Sutikno. "Analisis Middle-Income Trap di Indonesia". In: *Jurnal Bppk* 7.2 (2014), pp. 91–110.

[9] K. Larsen. "GAM: the predictive modeling silver bullet". In: *Multithreaded Stitch Fix* 30 (2015), pp. 1–27. https://multithreaded.stitchfix.com/blog/2015/07/30/gam/.

[10] S. K. Sapra. "Generalized additive models in business and economics". In: *International Journal of Advanced Statistics and Probability* 1.3 (2013), pp. 64–81. DOI: 10.14419/ijasp.v1i3.1022.

[11] N. Beck and S. Jackman. "Beyond linearity by default: Generalized additive models". In: *American Journal of Political Science* 42.2 (1998), pp. 596–627. DOI: 10.2307/2991772.

[12] H. Yozza, Siswadi, and B. Suharjo. "Analisis Data Longitudinal dengan Metode Regresi Berstruktur Pohon (Kasus Penyakit Kencing Manis)". In: *Indonesian Journal of Statistics and Its Applications* 6.1 (2001), pp. 14–21.

[13] Trevor Hastie and Robert Tibshirani. *Generalized Additive Models*. London: Chapman and Hall, 1990.

[14] Ludwig Fahrmeir, Thomas Kneib, Stefan Lang, and Brian D. Marx. *Regression: Models, Methods and Applications*. Berlin: Springer-Verlag, 2013.

[15] Simon Wood. *Generalized Additive Models: An Introduction with R*. Boca Raton: Chapman & Hall, 2006.

[16] Damodar Gujarati and Dawn Porter. *Basic Econometrics*. 5th ed. New York: McGraw Hill Inc., 2009.

[17] Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach*. 8th ed. Mason: South-Western Cengage Learning, 2016.

[18] Simon N. Wood. *Generalized Additive Models: An Introduction with R*. 2nd ed. Boca Raton: CRC Press, 2017.