CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 10 (2) (2025), Pages 1277-1292
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Sensitivity of Bayesian Truncated Spline Regression to Prior and Knot Configuration in Stunting Models

Septi Nafisa Ulluya Zahra[1,3], Adji Achmad Rinaldo Fernandes[1,3*], Achmad Efendi[1], Solimun[1,3], Alfiyah Hanun Nasywa[1,3], and Fachira Haneinanda Junianto[2,3]

[1] *Department of Statistics, Faculty of Mathematics and Natural Sciences, Brawijaya University, Indonesia*
[2] *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Brawijaya University, Indonesia*
[3] *Data Science and Statistics for Business Analytic Center, Faculty of Mathematics and Natural Sciences, Brawijaya University, Indonesia*

Abstract

Childhood stunting is a persistent health problem, with determining factors that often follow complex and non-linear patterns. To capture these patterns more accurately, this study developed a Bayesian bi-response regression model with a truncated spline approach to analyze the non-linear effects of economic factors, dietary patterns, and the environment on nutritional stunting and physical stunting. Sensitivity analysis was conducted to assess the effect of variations in prior type and number of knots on model performance using Deviance Information Criterion (DIC), Root Mean Square Error (RMSE), and prediction accuracy. The results showed that the combination of Normal–Gamma informative prior was the most stable and reliable, characterized by the lowest DIC and RMSE values, while the non-informative Uniform prior caused serious model instability, especially when the number of knots increased. The environmental variable (Y_3) showed strong parameter unidentifiability, reflected in a very wide credible interval, indicating the limitations of the truncated power basis in handling certain data variations. Overall, these findings indicate that prior specification has a much greater influence than the number of knots in determining model robustness, and emphasize the importance of using informative priors and more stable spline bases in modeling the non-linear determinants of stunting in children.

Keywords: Sensitivity Analysis, Bayesian Regression, Nonparametric Regression, Truncated Spline, Stunting

1 Introduction

Childhood stunting remains a persistent global health issue, reflecting chronic nutritional deprivation and environmental adversity during the critical growth period. Despite significant progress in reducing malnutrition worldwide, approximately 149 million children under five remain stunted, predominantly in low- and middle-income countries [1]. Stunting is multifactorial, influenced by economic hardship, inadequate dietary intake, and poor environmental sanitation [2], [3].

*Corresponding author. E-mail: fernandes@ub.ac.id

Submitted: October 11, 2025 Reviewed: November 17, 2025 Accepted: November 26, 2025
DOI: https://doi.org/10.18860/cauchy.v10i2.37381

Understanding these complex interactions requires analytical methods capable of capturing nonlinear and heterogeneous relationships.

Traditional parametric regression approaches often assume linear or low-order polynomial relationships between predictors and outcomes, which may not reflect the complex dynamics inherent in stunting data [4]. Consequently, nonparametric regression models, especially those based on truncated splines, have gained traction because they can flexibly approximate unknown functional forms without restrictive assumptions [5], [6]. The truncated spline method divides the covariate domain into intervals defined by knots and fits piecewise polynomial functions, allowing localized modeling of curvature and interactions [7].

Recent developments extend truncated spline regression into the Bayesian framework, which allows parameter uncertainty to be fully characterized via posterior distributions and enables the inclusion of prior knowledge about smoothness or model complexity [8], [9]. The Bayesian nonparametric spline approach is particularly advantageous in health applications, where data structures are complex and prior information, such as expected smoothness or variance, is available from earlier studies [10]. However, one of the critical challenges in such modeling lies in sensitivity to the choice of prior distributions and the number of knots used in the spline basis.

The choice of prior in Bayesian nonparametric regression, whether for spline coefficients, variance components, or smoothing parameters, can substantially influence posterior inference, especially when sample sizes are moderate or covariate distributions are irregular [11], [12]. Overly informative priors may constrain the estimated regression function, while weakly informative or improper priors may lead to overfitting or numerical instability [13]. Similarly, the number and placement of knots directly affect model flexibility: too few knots may underfit, missing local variations, whereas too many knots may overfit, increasing variance and computational cost [6].

In practice, these two elements, prior specification and knot configuration, are intertwined, jointly determining model smoothness and inference robustness. Sensitivity analysis is therefore essential to assess how alternative prior assumptions or knot selections influence posterior estimates and predictive performance [14]. In this study, sensitivity is examined through a direct model comparison framework using Deviance Information Criterion (DIC) and Root Mean Square Error (RMSE), allowing the effects of different priors and knot counts to be evaluated without relying on advanced sensitivity techniques. This approach provides a consistent and transparent assessment of model stability across all prior–knot configurations.

However, existing studies have not examined how the stability of Bayesian truncated spline models depends on prior specification and knot configuration. This gap is important because both elements directly control the smoothness and flexibility of the estimated non-linear function. To address this gap, this study provides the first systematic evaluation of prior–and–knot sensitivity within a Bayesian truncated spline framework applied to stunting analysis.

In the context of global stunting analysis, where data often exhibit nonlinear and region-specific relationships, the combination of Bayesian truncated spline regression and sensitivity assessment provides a powerful modeling framework. It enables flexible estimation of nonlinear effects of economic, dietary, and environmental factors on multiple stunting indicators (e.g., nutritional and physical stunting), while explicitly quantifying how results depend on prior and model configuration choices. Such robust modeling can inform more reliable and context-sensitive policy interventions aimed at achieving the Sustainable Development Goal (SDG) 2, which targets the elimination of all forms of malnutrition by 2030.

Accordingly, this study aims to assess the sensitivity of Bayesian truncated spline regression to variations in prior specification and knot configuration when modeling nutritional and physical stunting. The analysis highlights which combinations produce stable estimates and which lead to instability, thereby providing methodological guidance for future stunting research.

2 Literature Review

This literature review is designed to provide a structured understanding of the key concepts underlying this study. The discussion begins with an overview of stunting as a public health issue, followed by a review of the socioeconomic, nutritional, and environmental factors that contribute to the risk of stunting. Subsequent sections examine relevant methodological perspectives, including nonlinear modeling approaches and the use of Bayesian frameworks. Together, these sections form the conceptual and analytical foundation for the empirical investigation presented in this study.

2.1 Stunting

Stunting is a condition in which children fail to attain expected growth because of prolonged nutritional deficits, often exhibited as a height that falls below the -2 standard deviation threshold for children of the same age. It affects both the nutritional status and physical development of children. In terms of nutrition, chronic lack of essential nutrients, including proteins, vitamins, and minerals, impairs the formation of body tissues and organ systems, limiting physical growth and cognitive maturation. Malnourished children are also more vulnerable to infections due to compromised immune defenses [15].

Physically, stunting is assessed using the height-for-age metric compared to standard growth curves; children whose height is more than two standard deviations below the reference median are classified as stunted. Continuous monitoring of children's growth enables early detection of growth faltering, allowing nutritional and environmental interventions to be implemented promptly to mitigate long-term developmental damage [16].

This phenomenon arises from a combination of interrelated risk factors, including family economic status, dietary patterns, environmental hygiene, and inadequate nutrient intake. Household economic constraints restrict access to nutritious foods and essential health services, limiting consumption options necessary for proper development. Dietary adequacy plays a crucial role: a balanced diet inclusive of diverse food groups such as fruits, vegetables, proteins, and micronutrient-rich items supports both physical growth and immune function in young children [17]. Environmental conditions are also key: poor water quality, insufficient sanitation facilities, and low hygiene practices increase the incidence of infections (e.g., diarrhea), which in turn hinder nutrient absorption and impair growth [18].

2.2 Bayesian Nonparametric Regression

In Bayesian inference, all model parameters are treated as random variables characterized by probability distributions rather than fixed quantities. Prior distributions represent initial beliefs about the parameters before observing the data, while the likelihood function captures the information provided by the observed data. By combining both components, the posterior distribution provides an updated and comprehensive understanding of the parameters after data observation [19], [20].

The response variable is assumed to follow a normal distribution conditional on the predictor variables, regression coefficients, and variance, denoted as $(Y|X, \beta, \sigma^2) \sim N(X\beta, \sigma^2)$. Based on this assumption, the probability density and likelihood functions can be expressed as shown in Equations (1) and (2). These functions describe how the data are generated according to the model and provide the foundation for posterior estimation [21]. In Bayesian regression, variables have distributions and probability density functions as in Equation (1).

$$p(Y|X, \beta, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta)\right\} \tag{1}$$

Based on the probability density function (1), the likelihood function can be defined as shown in Equation (2).

$$
\begin{aligned}
p(Y \mid X, \beta, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_i - X_i\beta)^2}{2\sigma^2}\right) \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta)\right).
\end{aligned} \tag{2}
$$

Meanwhile, the posterior distribution can be expressed as shown in Equation (3).

$$
\begin{aligned}
&\text{Posterior} \propto \text{Likelihood} \times \text{Prior} \\
&p(\beta, \sigma^2|Y, X) \propto p(Y|X, \beta, \sigma^2)p(\sigma^2)p(\beta|\sigma^2) \\
&p(\beta, \sigma^2|Y, X) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta)\right\} \\
&\qquad \times (\sigma^2)^{-(\frac{v}{2}+1)} \exp\left\{-\frac{vs^2}{2\sigma^2}\right\} \times (\sigma^2)^{-\frac{k}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\beta - \mu)^T\Lambda(\beta - \mu)\right\}
\end{aligned} \tag{3}
$$

The posterior distribution, as formulated in Equation (3), is proportional to the product of the likelihood and the priors of the parameters. This posterior serves as the basis for Bayesian estimation and inference, enabling parameter estimation and uncertainty quantification simultaneously. Building upon this foundation, the Bayesian nonparametric regression approach introduces greater flexibility by modeling the regression function using spline-based or kernel-based techniques. This allows the model to capture nonlinear and complex relationships between variables without assuming a specific parametric form [22], [23].

Determination of the optimal knot point in spline regression can be done using the Generalized Cross Validation (GCV) method, which selects the node point with the minimum GCV value to produce the best model [24].In spline regression, GCV is often used to select the optimal number of knots.

$$
\text{GCV}(\mathbf{K}) = \frac{n^{-1}\sum_{i=1}^{n}[y_i - \hat{f}(x_i)]^2}{[n^{-1}\text{trace}(I - A(\mathbf{K}))]^2} \tag{4}
$$

where I is the identity matrix and $A = X(X^TX)^{-1}X^T$.

2.3 Sensitivity Analysis

Sensitivity analysis is a rigorous evaluation procedure undertaken to assess how variation in input parameters, model assumptions, or structure affects the resulting outputs. Through this approach, researchers can identify which parameters or model components exert the greatest influence on results, enabling focused attention on the critical aspects. Moreover, sensitivity analysis plays a crucial role in testing model robustness, namely, how consistent results remain when faced with uncertainties, specification errors, or unexpected changes. In this way, it helps not only to understand internal model dynamics but also to increase confidence in conclusions drawn from complex model-based decisions [25].

In the context of a Bayesian nonparametric bi-response regression using spline basis functions, sensitivity analysis becomes especially important because the model has several properties that can make its inference sensitive to assumptions and choices. First, Bayesian models heavily depend on the choice of priors; altering the prior distributions can meaningfully change posterior parameter estimates or predictions [19]. Second, nonparametric spline methods (for example, B-splines, truncated splines, or other spline bases) are flexible in structure; the number of knots (and their placement) can substantially influence the shape of the estimated functions. For truncated splines, different counts of knots (e.g., $K = 1, 2, 3$) may lead to different amounts of local curvature or flexibility, affecting how well the model fits nonlinear patterns. In this study, the configuration with three knots was selected based on the smallest GCV value, indicating the optimal balance between model fit and complexity.

Truncated power splines were used in this study because of their simple form and ease of interpretation. However, this approach is known to have the disadvantage of potential multicollinearity between bases, especially when the number of knots increases [26]. Alternatives such as B-splines and penalized splines offer better numerical stability, but are less transparent in interpretation. Considering the research objective of explicitly evaluating the sensitivity of the prior and knots, the truncated power spline is a suitable choice despite these limitations.

This study uses a truncated spline of order 1 because its functional form is more stable for moderate sample sizes, easier to interpret, and less prone to overfitting than higher-order spline bases [26]. After selecting the optimal knot configuration through GCV, the chosen knot locations were held *fixed* for all subsequent model fits; only the prior specification (Normal, Uniform, Jeffreys, Gamma, etc.) was varied. This design isolates the effects of prior choice from changes in knot placement and ensures a fair and consistent comparison across prior scenarios.

Therefore, the sensitivity analysis should consider:

1. **Variations in prior specification:** for regression coefficients, one might compare informative priors (normal distribution with a certain mean and precision) versus non-informative (flat or wide uniform) priors. For residual variance or error variance, one may compare informative priors (gamma for precision/inverse gamma for variance) versus non-informative (Jeffreys prior). This yields multiple combinations (four combinations of prior choices for regression coefficients and error variance).
2. **Variations in knot counts:** evaluate how changing the number of knots (e.g., $K = 1, 2, 3$) influences the estimated response functions. More knots allow more flexibility or local wiggles; fewer knots enforce smoother, more global shapes.

To evaluate the sensitivity, one can compare model outputs (posterior estimates, predictive curves) across these combinations. Performance complexity can be compared using model comparison or fit metrics (Deviance Information Criterion, DIC), predictive error metrics (Root Mean Squared Error, RMSE), and Prediction sensitivity (difference between expected posterior mean and true value or reference). By doing so, the researcher can check if the conclusions remain stable across prior and knot settings, demonstrating that inferences are not unduly dependent on particular modeling choices.

From the literature, such approaches have been explored in the adaptive Bayesian nonparametric regression context. For example, recent work proposes adaptive smoothing priors that automatically adjust smoothness in response to curvature of the underlying function [27]. Moreover, fully Bayesian knotted spline methods (with priors on knot number and location, using reversible-jump MCMC) have been advanced to infer the number and placements of knots, and to penalize overcomplex models via priors on model complexity [28]. These approaches show that prior choice on the knot number can impact resulting function estimates significantly; therefore, embedding sensitivity checks is valuable for ensuring robust inference [27], [28].

In the Bayesian framework, several measures are commonly used to evaluate the goodness of fit and predictive performance of a model. One widely used criterion is the Deviance Information Criterion (DIC), which balances model fit and complexity. For a Bayesian bi-response regression model, the deviance is defined based on the full parameter vector $\theta = (\beta_1, \beta_2, \Sigma)$, not only on β. Accordingly, the deviance is given by

$$D(\theta) = -2 \log p(Y \mid \theta), \tag{5}$$

where $p(Y \mid \theta)$ is the full likelihood of the bivariate normal model, which includes both regression coefficients and the error covariance matrix Σ.

The DIC is then computed as

$$\text{DIC} = \bar{D} + p_D, \tag{6}$$

where $\bar{D}$ is the posterior mean deviance and p_D is the effective number of parameters. These terms are defined as

$$\bar{D} = E_{\theta|Y}[D(\theta)], \tag{7}$$

$$p_D = \bar{D} - D(\bar{\theta}), \tag{8}$$

with $\bar{\theta}$ denoting the posterior mean of the full parameter vector. This formulation ensures that the DIC correctly accounts for all components of the bi-response model, including the regression parameters and the error covariance structure. A smaller DIC value indicates a better trade-off between model fit and parsimony [29].

In addition to DIC, the Root Mean Square Error (RMSE) is employed to assess the predictive accuracy of the model. For a bivariate response regression, RMSE is formulated as in Equation (9).

$$\text{RMSE} = \sqrt{\frac{1}{2n}\sum_{i=1}^{n}\sum_{l=1}^{2}(y_{li} - \hat{y}_{li})^2} \tag{9}$$

Predictive sensitivity is used to evaluate the stability of model predictions under different prior specifications. It measures how much the predicted values change when the prior is altered from the reference (baseline) prior. A model is considered stable if changes in priors lead to only small variations in predictions.

$$\text{PS}(s) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i^{(s)} - \hat{y}_i^{(\text{ref})}\right)^2} \tag{10}$$

where $\hat{y}_i^{(s)}$ denotes the predicted value for the i-th observation under prior scenario s, $\hat{y}_i^{(\text{ref})}$ denotes the predicted value under the reference prior, and n is the total number of observations. A smaller value of $\text{PS}(s)$ indicates higher predictive stability of the model under prior perturbations.

By comparing these metrics across different prior specifications and the number of knots, researchers can determine which Bayesian model configuration yields the most stable and accurate inference.

3 Research Methods

This section explains the methodological framework adopted in this study. It begins with a description of the data sources and variable construction, followed by the stages of analysis used to assess nonlinearity, estimate the Bayesian truncated spline model, and evaluate model performance under different prior specifications and knot configurations. Together, these steps form the systematic procedure used to address the research objectives.

3.1 Data Sources and Research Variables

This study utilizes primary data from a research grant project conducted by [30]. Data collection took place in Sumberputih Village, Wajak District, Malang Regency, Indonesia, an area designated as a priority location in the regional stunting prevention acceleration program. The village was selected due to its classification as a high-risk area, making it relevant to national efforts addressing stunting as a major public health concern.

A total of 100 respondents were surveyed, following the minimum sample size guidelines proposed by [31] for models containing fewer than seven variables. This sample size is also appropriate for the nonparametric Bayesian framework, as prior regularization supports stable estimation even under moderate sample sizes [19]. The study involves five variables in total:

three predictors Economic Level (X_1), Children's Diet (X_2), and Environment (X_3) and two response variables, Nutritional Stunting (Y_1) and Physical Stunting (Y_2).

Both Y_1 and Y_2 are constructed as composite indices derived from several Likert-scale questionnaire items. Each index represents an aggregated measure of nutritional and physical stunting risk, respectively, obtained by averaging the corresponding item-level scores so that higher values indicate greater risk. This mean-based construction results in approximately continuous response scales, making the indices suitable for analysis using a Gaussian-based Bayesian bi-response regression model.

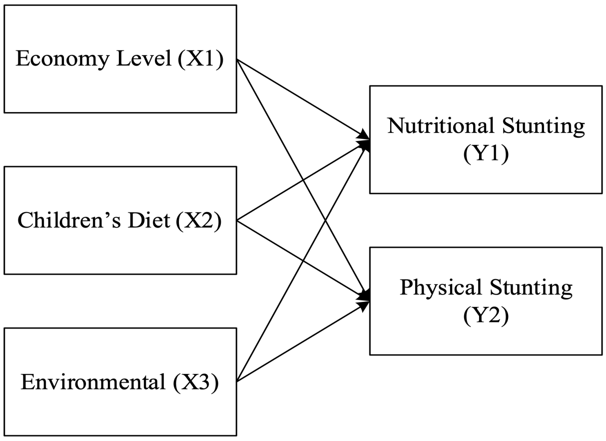

Figure 1: Research Model

3.2 Research Stages

The stages of research in this study are as follows:

- Conduct a linearity test using Ramsey's RESET.
- Define the base model using truncated spline nonparametric regression of order 1 (linear) with a maximum of 3 knots and a selected prior distribution.
- Vary the number of knots (1, 2, and 3) to observe the effect on model estimation and goodness-of-fit.
- Test different types of prior distributions — informative (Normal and Gamma distributions) and non-informative (Uniform and Jeffreys distributions) — to examine their influence on parameter stability.
- Perform parameter estimation using MCMC with Gibbs Sampling, running 10000 iterations for each model configuration.
- Evaluate changes in estimated coefficients, model fit statistics (DIC, RMSE, and predictive sensitivity) for each scenario.
- Compare results across all configurations to assess the model's robustness to variations in the choice of priors and the number of knots.
- Identify the optimal combination of prior type and number of knots that produces the most stable and accurate model performance.

4 Results

This section presents the research results in a structured manner to provide a comprehensive understanding of the findings. The presentation begins with the validity and reliability results of the questionnaire used, followed by the linearity test results as a preliminary step in assessing the relationship patterns between variables. Next, the results of the Bayesian truncated spline bi-response model estimation, prior sensitivity analysis and knot configuration, as well as substantive interpretations of non-linear patterns are explained step by step. This arrangement ensures that readers can follow the flow of results and discussions sequentially before entering into more detailed tables and analyses.

Table 1: Validity and Reliability of the Questionnaire

Research Variables	Items	Corrected Item Total	Information	Cronbach's Alpha	Information
Economy Level (X_1)	$X_{1.1.1}$	0.56	Valid	0.75	Reliable
	$X_{1.1.2}$	0.69	Valid		
	$X_{1.2.1}$	0.44	Valid		
	$X_{1.2.2}$	0.43	Valid		
	$X_{1.2.3}$	0.53	Valid		
Children's Diet (X_2)	$X_{2.1.1}$	0.45	Valid	0.78	Reliable
	$X_{2.1.2}$	0.52	Valid		
	$X_{2.1.3}$	0.39	Valid		
	$X_{2.2.1}$	0.70	Valid		
	$X_{2.2.2}$	0.69	Valid		
	$X_{2.2.3}$	0.62	Valid		
	$X_{2.3.1}$	0.51	Valid		
	$X_{2.3.2}$	0.42	Valid		
	$X_{2.3.3}$	0.34	Valid		
Environment (X_3)	$X_{3.1.1}$	0.42	Valid	0.73	Reliable
	$X_{3.1.2}$	0.40	Valid		
	$X_{3.1.3}$	0.40	Valid		
	$X_{3.2.1}$	0.36	Valid		
	$X_{3.2.2}$	0.40	Valid		
	$X_{3.2.3}$	0.47	Valid		
	$X_{3.3.1}$	0.40	Valid		
	$X_{3.3.2}$	0.41	Valid		
	$X_{3.3.3}$	0.43	Valid		
Nutritional Stunting (Y_1)	$Y_{1.1.1}$	0.41	Valid	0.73	Reliable
	$Y_{1.1.2}$	0.53	Valid		
	$Y_{1.1.3}$	0.61	Valid		
	$Y_{1.2.1}$	0.46	Valid		
	$Y_{1.2.2}$	0.35	Valid		
	$Y_{1.2.3}$	0.48	Valid		

Based on the analysis of Corrected Item Total Correlation and Cronbach's Alpha, all variables showed adequate reliability and validity. It is assessed through a Corrected Item Total Correlation of more than 0.30 and Cronbach's Alpha of more than 0.60. Therefore, the variables of Economic Level, Children's Diet, Environment, and Nutritional Stunting show that the variables are valid and reliable.

Table 2: Linearity Test Results

No.	Relationships Between Variables	F Test Statistics	p-value	Result
1	Economic Level (X_1) $\rightarrow$ Nutritional Stunting (Y_1)	27.151	<0.001	Not Linear
2	Children's Diet (X_2) $\rightarrow$ Nutritional Stunting (Y_1)	14.910	<0.001	Not Linear
3	Environment (X_3) $\rightarrow$ Nutritional Stunting (Y_1)	3.333	<0.037	Not Linear
4	Economic Level (X_1) $\rightarrow$ Physical Stunting (Y_2)	23.687	<0.001	Not Linear
5	Children's Diet (X_2) $\rightarrow$ Physical Stunting (Y_2)	52.915	<0.001	Not Linear
6	Environment (X_3) $\rightarrow$ Physical Stunting (Y_2)	61.524	<0.001	Not Linear

Ramsey's Regression Specification Error Test (RESET) was employed to assess the adequacy of linearity between each predictor and response variable prior to applying the spline-based model. The test was conducted by estimating a simple baseline linear model for each predictor–response relationship:

$$Y_{ki} = \alpha_k + \beta X_{ji} + \varepsilon_{ki},$$

where $k \in \{1, 2\}$ corresponds to the two response variables (Y_1 and Y_2), and $j \in \{1, 2, 3\}$ corresponds to the predictors X_1, X_2, and X_3.

For each baseline model, the RESET procedure augments the specification with powers of the fitted values (e.g., $\hat{Y}^2$, $\hat{Y}^3$) and performs an F-test on their joint significance. The null hypothesis tested is:

$$H_0 : \text{The linear model is correctly specified (no omitted non-linearity).}$$

A small p-value indicates rejection of H_0, suggesting that the linear specification does not adequately capture the relationship between the predictor and the response. As reported in Table 2, all six relationships yielded statistically significant p-values, indicating non-linear patterns in each predictor–response pair. These findings provide the empirical basis for adopting the truncated spline approach in the subsequent Bayesian multivariate analysis.

4.1 MCMC Specification and Diagnostics

To ensure transparency and reproducibility, the full MCMC configuration used in estimating the multivariate Bayesian truncated spline model is reported in Table 3. The model was fitted using four parallel chains, each run for 10,000 iterations with a burn-in period of 2,000 iterations. No thinning was applied, and a total of 32,000 posterior samples were retained for inference.

Table 3: MCMC Specification

Setting	**Value**
Number of chains	4 parallel chains
Total iterations per chain	10,000 iterations
Burn-in length	2,000 iterations
Thinning	None applied
Posterior samples retained	8,000 per chain (32,000 total)
Convergence diagnostics	$\hat{R} < 1.01$, ESS > 400 for all parameters
Additional checks	Trace plots and autocorrelation plots inspected

All posterior summaries, hypothesis testing results, and DIC calculations in this manuscript are based exclusively on post–burn-in samples that passed the convergence diagnostics. The Gelman–Rubin statistic ($\hat{R}$) indicated excellent convergence across all parameters, while effective sample size (ESS) values exceeded 400, confirming efficient mixing of the Markov chains. Trace plots and autocorrelation functions further supported the stability of the posterior draws.

4.2 Bayesian Nonparametric Regression

A truncated spline nonparametric regression model on a linear order with three point knots for three predictor variables and two response variables with Bayesian calculation are presented as follows.

$$\begin{aligned}\hat{f}_{1i} = {} & 1.278\, X_{1i} + 0.859\,(X_{1i} - 1.324)_+ - 0.886\,(X_{1i} - 2.574)_+ + 0.063\,(X_{1i} - 2.887)_+ \\ & + 0.962\, X_{2i} + 1.375\,(X_{2i} - 2.060)_+ - 0.848\,(X_{2i} - 3.407)_+ - 0.862\,(X_{2i} - 3.946)_+ \\ & 3.255\, X_{3i} - 1.258\,(X_{3i} - 2.239)_+ + 2.111\,(X_{3i} - 2.538)_+ - 1.546\,(X_{3i} - 2.836)_+ .\end{aligned} \quad (11)$$

$$\begin{aligned}\hat{f}_{2i} = 0.911\,X_{1i} - 0.699\,(X_{1i} - 1.324)_+ - 0.539\,(X_{1i} - 2.574)_+ + 0.364\,(X_{1i} - 2.887)_+ \\ 2.667\,X_{2i} - 4.260\,(X_{2i} - 2.060)_+ - 1.052\,(X_{2i} - 3.407)_+ + 0.865\,(X_{2i} - 3.946)_+ \\ + 2.292\,X_{3i} + 3.100\,(X_{3i} - 2.239)_+ + 2.949\,(X_{3i} - 2.538)_+ - 6.901\,(X_{3i} - 2.836)_+.\end{aligned} \tag{12}$$

Table 4: Hypothesis Testing with Bayesian Approach

Relation	$\hat{\beta}_i$	**Mean Posterior**	**Credible Interval (Lower)**	**Credible Interval (Upper)**	**Significance**
$X_1 \to Y_1$	$\beta_1 X_1$	1.278	-0.071	2.674	No
$X_1 \to Y_1$	$\beta_2(X_1 - K_{11})_+$	0.859	0.138	1.581	Yes
$X_1 \to Y_1$	$\beta_3(X_1 - K_{12})_+$	-0.886	-1.631	-0.161	Yes
$X_1 \to Y_1$	$\beta_4(X_1 - K_{13})_+$	0.063	-0.731	0.838	No
$X_2 \to Y_1$	$\beta_5 X_2$	0.962	-0.064	1.974	No
$X_2 \to Y_1$	$\beta_6(X_2 - K_{21})_+$	1.375	0.665	2.129	Yes
$X_2 \to Y_1$	$\beta_7(X_2 - K_{22})_+$	-0.848	-1.608	-0.122	Yes
$X_2 \to Y_1$	$\beta_8(X_2 - K_{23})_+$	-0.862	-1.579	-0.140	Yes
$X_3 \to Y_1$	$\beta_9 X_3$	3.255	-9.926	9.613	No
$X_3 \to Y_1$	$\beta_{10}(X_3 - K_{31})_+$	-1.258	-2.173	-0.498	Yes
$X_3 \to Y_1$	$\beta_{11}(X_3 - K_{32})_+$	2.111	0.575	3.567	Yes
$X_3 \to Y_1$	$\beta_{12}(X_3 - K_{33})_+$	-1.546	-2.952	0.001	No
$X_1 \to Y_2$	$\beta_1 X_1$	0.911	0.102	1.702	Yes
$X_1 \to Y_2$	$\beta_2(X_1 - K_{11})_+$	-0.699	-3.076	1.742	No
$X_1 \to Y_2$	$\beta_3(X_1 - K_{12})_+$	-0.539	-1.740	0.617	No
$X_1 \to Y_2$	$\beta_4(X_1 - K_{13})_+$	0.364	-0.825	1.639	No
$X_2 \to Y_2$	$\beta_5 X_2$	2.667	1.180	4.122	Yes
$X_2 \to Y_2$	$\beta_6(X_2 - K_{21})_+$	-4.260	-6.159	-2.317	Yes
$X_2 \to Y_2$	$\beta_7(X_2 - K_{22})_+$	-1.052	-2.607	0.623	No
$X_2 \to Y_2$	$\beta_8(X_2 - K_{23})_+$	0.865	-0.797	2.390	No
$X_3 \to Y_2$	$\beta_9 X_3$	2.292	0.890	3.673	Yes
$X_3 \to Y_2$	$\beta_{10}(X_3 - K_{31})_+$	3.100	-9.292	9.309	No
$X_3 \to Y_2$	$\beta_{11}(X_3 - K_{32})_+$	2.949	1.322	4.609	Yes
$X_3 \to Y_2$	$\beta_{12}(X_3 - K_{33})_+$	-6.901	-9.785	-3.892	Yes

Based on the Bayesian estimation results of the bi-response truncated spline model, several coefficients were found to significantly influence nutritional stunting (Y_1) and physical stunting (Y_2). For nutritional stunting, the significant parameters were β_2 and β_3 for the spline components of X_1, indicating that the economic level affects Y_1 at specific knot locations rather than through a simple linear trend. In addition, all spline components of children's diet (X_2)—namely β_6, β_7, and β_8—were significant, showing strong nonlinear effects on nutritional stunting. The environmental variable (X_3) also exhibited significant nonlinear influences through β_{10} and β_{11}, while its linear component β_9 remained insignificant.

For physical stunting (Y_2), the estimation results showed a significant effect from the linear component of X_1 (β_1), whereas its spline components (β_2, β_3, β_4) were not significant. The dietary variable (X_2) significantly affected Y_2 through its linear term β_5 and the first spline component β_6, while the remaining spline terms were not significant. Regarding environmental conditions (X_3), both the linear effect β_9 and the third spline component β_{12} were significant. Meanwhile, β_{10} was not significant, and β_{11} showed a strong positive effect.

These patterns confirm that the relationships between the predictors and both stunting outcomes exhibit nonlinear behavior across specific knot points. Significant spline coefficients indicate that the effect of each predictor changes depending on its value relative to the knot locations, reflecting heterogeneous influences within different ranges of the data. Conversely, insignificant coefficients imply that variations in certain predictor ranges do not exert substantial influence. Overall, the truncated spline approach within a Bayesian framework effectively captures these localized nonlinearities in the predictor–response relationship.

4.3 Sensitivity Analysis

Sensitivity analysis on the choice of prior type and the number of knots has the potential to influence Bayesian inference. The purpose of the sensitivity analysis is to examine how robust the posterior results (posterior mean and credible interval) are to variations in prior selection (informative vs. non-informative) and the number of knots in the truncated spline. In practice, the sensitivity analysis will be conducted by specifying several systematic alternative scenarios: (i) variation of priors for β_l (Normal/Uniform), (ii) variation of priors for σ_l^2 or τ_l (Gamma/Jeffreys), and (iii) variation in the number of knots (1, 2, and 3). For each scenario, posterior samples will be generated using Gibbs Sampling until sufficient burn-in is achieved, followed by comparing the posterior summaries and model fit metrics. The results of the sensitivity analysis for the four criteria are presented in Table 5.

Table 5: Sensitivity Analysis

Prior Distribution for Coefficient	Prior Distribution for Error	Knot	DIC	RMSE	Predictive Sensitivity
Normal (Informative)	Gamma (Informative)	1	357.8	0.621	0.0001
Normal (Informative)	Gamma (Informative)	2	357.4	0.611	0.0005
Normal (Informative)	Gamma (Informative)	3	356.9	0.595	0.0001
Normal (Informative)	Jeffreys (Non-Informative)	1	357.8	0.621	0.0003
Normal (Informative)	Jeffreys (Non-Informative)	2	357.7	0.611	0.0001
Normal (Informative)	Jeffreys (Non-Informative)	3	361.8	0.596	0.0002
Uniform (Non-Informative)	Gamma (Informative)	1	464.1	0.664	0.0002
Uniform (Non-Informative)	Gamma (Informative)	2	367.0	0.611	0.0030
Uniform (Non-Informative)	Gamma (Informative)	3	1593.0	3.011	0.0010
Uniform (Non-Informative)	Jeffreys (Non-Informative)	1	369.0	0.625	0.0003
Uniform (Non-Informative)	Jeffreys (Non-Informative)	2	368.0	0.611	0.0006
Uniform (Non-Informative)	Jeffreys (Non-Informative)	3	13479.0	4.666	0.0009

Table 5 shows that the combination of an informative Normal prior for the spline coefficients together with either an informative Gamma prior or the non-informative Jeffreys prior for the error variance produces the most stable in-sample performance. This is reflected in the lowest DIC values and consistently smaller RMSE across all knot configurations. Models with three knots achieve the best in-sample fit, as indicated by reduced RMSE, although the DIC differences among the 1, 2, and 3 knot specifications remain small when informative priors are used.

In contrast, models employing a non-informative Uniform prior display substantial deterioration in performance, especially as the number of knots increases. Under the Uniform–Gamma and Uniform–Jeffreys combinations, both DIC and RMSE increase sharply at three knots, indicating numerical instability and sensitivity to model complexity. This pattern demonstrates that the model is considerably more sensitive to prior specification than to knot placement.

The predictive sensitivity measures reinforce this conclusion. Under the Normal–Gamma and Normal–Jeffreys priors, predictive sensitivity remains near zero, which indicates that fitted values are robust to small perturbations in the prior. Conversely, Uniform priors generate much larger predictive sensitivity—particularly in the three-knot model—showing that small changes in prior assumptions lead to substantial shifts in predictions. This confirms that moderately informative priors are essential for maintaining robustness in the bi-response spline setting.

The superior performance of the Normal–Gamma prior arises because the informative Normal prior provides regularization for the spline coefficients, preventing excessive fluctuation of the truncated basis functions, while the Gamma prior places controlled structure on the error variance. These priors enhance posterior stability and reduce multicollinearity among spline terms, which is important in flexible models that include multiple knots. In contrast, the Uniform prior offers no regularization, resulting in inflated posterior variance and a tendency to overfit random noise rather than capturing the underlying nonlinear structure.

From a substantive perspective, these results underscore the importance of stable nonlinear estimation in the analysis of nutritional and physical stunting. The Normal–Gamma prior yields reliable nonlinear patterns—such as threshold effects in dietary indicators and diminishing influence of economic conditions—ensuring that these effects reflect meaningful relationships rather than artifacts of model instability. Conversely, the poor performance of the Uniform prior demonstrates that insufficient regularization can lead to misleading conclusions, especially in spline models with higher flexibility.

Overall, Table 5 confirms that prior specification plays a more critical role than the number of knots. When informative priors are used, increasing the number of knots from one to three results in only minor changes in DIC and RMSE and maintains low predictive sensitivity, indicating stable model behavior. In contrast, non-informative Uniform priors become increasingly unstable as knot complexity increases, highlighting the necessity of appropriate prior selection for robust inference in the study of child stunting.

5 Discussion

The results of the Bayesian bi-response truncated spline regression reveal nuanced and distinctly nonlinear influences of economic, dietary, and environmental factors on both nutritional stunting (Y_1) and physical stunting (Y_2). The significance of specific spline components indicates that the effects of predictors vary across the covariate range rather than remaining constant. For instance, the effect of economic level becomes more pronounced only within certain spline segments, whereas dietary quality exhibits a consistently strong association with nutritional stunting across all spline components. These patterns support existing evidence that nonparametric spline approaches are effective for capturing localized nonlinearities that linear models would overlook [32]. From a substantive perspective, such nonlinear shapes align with well-established mechanisms in stunting research, such as diminishing returns in economic improvements and threshold effects in dietary adequacy.

The sensitivity analysis further highlights the critical role of prior specification in Bayesian spline modeling. The combination of an informative Normal prior for regression coefficients and a Gamma prior for error variance outperformed other configurations in terms of DIC, RMSE, and predictive sensitivity, and maintained stable performance even as the number of knots increased. This is consistent with recent literature showing that informative priors help regulate coefficient magnitudes and stabilize posterior estimation in flexible Bayesian smoothing models [24]. In contrast, the use of non-informative Uniform priors resulted in substantial performance deterioration, particularly when three knots were included. The marked increase in DIC and RMSE suggests that without sufficient regularization, the truncated spline basis becomes overly flexible and susceptible to overfitting or numerical instability [28]. These findings reinforce the idea that prior selection has a stronger influence on model robustness than knot configuration alone.

Differences between the Gamma and Jeffreys priors for the error variance also illustrate how hyperprior structure affects estimation. The Gamma prior imposed clearer constraints on the variance component, improving convergence and reducing posterior uncertainty, whereas the Jeffreys prior, being flat and fully data-driven, produced greater variability especially under higher spline flexibility. This is aligned with recent Bayesian smoothing studies emphasizing the importance of variance hyperpriors in managing model complexity [32]. Taken together, the results indicate that while increasing spline complexity through additional knots can reduce prediction error, the type of prior plays a more decisive role in ensuring estimation stability and producing interpretable nonlinear effects.

An important empirical finding is the instability of the environmental factor (X_3), which shows wide credible intervals and inconsistent significance across prior knot combinations. This pattern likely reflects the inherently variable and long-term nature of environmental exposures, which are difficult to capture precisely in a cross-sectional spline framework. The instability also suggests potential parameter unidentifiability stemming from the truncated power basis, particularly when combined with weak priors. Given this limitation, more numerically stable spline bases such as B-splines or penalized splines may offer improved reliability for modeling environmental determinants of stunting. Future research should also consider alternative model specifications or longitudinal designs to better capture cumulative environmental effects.

Overall, these findings underscore that robust and interpretable estimation of nonlinear relationships in stunting research depends critically on appropriate prior specification. The Normal Gamma prior combination consistently delivers stable, accurate, and theoretically plausible results, making it a suitable choice for applied analyses of nutritional and physical stunting where predictor effects are expected to vary across the covariate space.

6 Conclusion

This study shows that the Bayesian bi-response truncated spline regression model is capable of capturing non-linear relationship patterns between economic factors, dietary patterns, and the environment on nutritional stunting and physical stunting. Sensitivity analysis confirms that model performance is much more influenced by prior specifications than by the number of knots used. Informative priors, particularly the Normal Gamma combination, consistently produce more stable estimates, lower DIC and RMSE, and Predictive sensitivity, while Uniform priors tend to cause estimation instability, especially when the number of knots increases. These findings emphasize the importance of regularization through appropriate priors in Bayesian spline modeling.

However, the results also reveal limitations in modeling environmental factors (X_3), as evidenced by very wide credible intervals and unstable inferences across various prior and knot configurations. This indicates potential parameter unidentifiability associated with the use of truncated spline bases, particularly in cross-sectional data designs. Thus, further research is recommended to consider the use of more stable spline bases, such as B-splines or penalized splines, or to apply alternative model structures to produce more reliable estimates.

From an application perspective, these results emphasize the importance of robust statistical approaches in analyzing the determinants of stunting, given that the relationships between variables are nonlinear and sensitive to model specifications. The use of appropriate informative priors can help produce more accurate inferences and support the development of more targeted interventions in efforts to reduce stunting. Further research could also be directed towards the development of adaptive priors, hierarchical Bayesian frameworks, or the application of models to spatial and longitudinal data to capture the more complex dynamics of stunting.

CRediT Authorship Contribution Statement

Septi Nafisa Ulluya Zahra: Conceptualization, Methodology, Software, Formal Analysis, Writing–Original Draft Preparation, Visualization. **Adji Achmad Rinaldo Fernandes:** Conceptualization, Methodology, Supervision, Validation, Writing–Review & Editing, Project Administration, Funding Acquisition. **Achmad Efendi:** Data Curation, Formal Analysis, Writing–Review & Editing. **Solimun:** Conceptualization, Methodology, Supervision, Validation. **Alfiyah Hanun Nasywa:** Data Collection, Resources, Investigation, Writing–Original Draft Preparation. **Fachira Haneinanda Junianto:** Software, Data Curation, Visualization, Writing–Review & Editing.

Declaration of Generative AI and AI-assisted technologies

Generative AI tools (ChatGPT, OpenAI) were used only for language editing (grammar, spelling, and clarity) and minor support in idea formulation. All scientific content and conclusions are the sole responsibility of the authors.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

This research received no external funding.

Data and Code Availability

The research data used in this study consist of primary survey responses collected as part of a research grant project and contain sensitive information related to household conditions and child stunting risk. Due to ethical, legal, and confidentiality restrictions, the dataset cannot be made publicly available. Access to the data may be granted upon reasonable request to the corresponding author and is subject to approval by the project's ethical oversight committee as well as the signing of a confidentiality agreement.

The analysis code used to implement the Bayesian bi-response spline models can be shared upon reasonable request. While the code may require minor modification to run without access to the confidential dataset, all model specifications and computational procedures described in the manuscript can be fully reproduced using the provided scripts.

References

[1] D. Headey, A. Palloni, and N. Prasada Rao, "Global trends and inequalities in childhood stunting," *The Lancet Global Health*, vol. 11, no. 4, e551–e563, 2023. DOI: 10.1016/S2214-109X(23)00045-7.

[2] T. U. Chikako, A.-A. Seidu, J. E. J. Hagan, and B. O. Ahinkorah, "Complex multilevel modelling of undernutrition in ethiopia," *Nutrients*, vol. 13, no. 9, p. 3018, 2021. DOI: 10.3390/nu13093018.

[3] D. Dwomoh, C. Sewor, S. K. Annim, S. Stranges, N.-B. Kandala, and A. K. Amegah, "Do dietary practices and household environmental quality mediate socio-economic inequalities in child undernutrition risk in west africa?" *Public Health Nutrition*, vol. 26, no. 5, pp. 1022–1033, 2022. DOI: 10.1017/S1368980022002269.

[4] Q. Li and J. S. Racine, *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, 2007. Available online.

[5] R. Eubank, *Nonparametric Regression and Spline Smoothing*. Marcel Dekker, 1999. Available online.

[6] Y. Zhao, Y. Li, and D. Xu, "Robust bayesian spline regression under prior and knot uncertainty," *Statistics and Computing*, vol. 33, no. 4, pp. 82–97, 2023. DOI: 10.1007/s11222-022-10184-y.

[7] D. Ruppert, M. P. Wand, and R. J. Carroll, *Semiparametric Regression*. Cambridge University Press, 2003. Available online.

[8] S. Lang and A. Brezger, "Bayesian p-splines," *Journal of Statistical Software*, vol. 7, pp. 1–25, 2012. DOI: 10.18637/jss.v007.i01.

[9] L. Fahrmeir, T. Kneib, S. Lang, and B. Marx, *Regression: Models, Methods and Applications*. Springer, 2021. Available online.

[10] M. C. Edwards, R. Meyer, and N. Christensen, "Bayesian nonparametric spectral density estimation using B-spline priors," *Statistics and Computing*, vol. 29, pp. 67–78, 2019. DOI: 10.1007/s11222-017-9796-9.

[11] P. Gustafson, "Bayesian sensitivity analysis: Theoretical developments and applications," *Statistical Science*, vol. 35, no. 2, pp. 235–256, 2020. DOI: 10.1214/19-STS742.

[12] T. Ohigashi and S. Sugasawa, "Efficient prior sensitivity and tipping-point analysis for medical research," *arXiv*, 2025, Art. no. arXiv:2510.10034. DOI: 10.48550/arXiv.2510.10034.

[13] J. J. Abellan, R. G. Cowell, and S. Moral, "Robustness and sensitivity analysis in bayesian networks revisited," *Artificial Intelligence*, vol. 330, p. 104 052, 2024. DOI: 10.1016/j.artint.2023.104052.

[14] H. Jia, T. Xu, S. Liang, P. Zhao, and C. Xu, "Parameter sensitivity, uncertainty, and identifiability analysis," *Environmental Modelling & Software*, vol. 109, pp. 324–335, 2018. DOI: 10.1016/j.envsoft.2018.03.001.

[15] D. K. Kinyoki, J. A. Berkley, G. M. Moloney, et al., "Environmental predictors of stunting among children under-five in somalia," *BMC Public Health*, vol. 16, p. 654, 2016. DOI: 10.1186/s12889-016-3320-6.

[16] S. Adi, I. Krisnana, P. D. Rahmawati, and U. Maghfiroh, "Environmental factors that affect the incidence of stunting in under-five children: A literature review," *Pediomaternal Nursing Journal*, vol. 9, no. 1, pp. 42–44, 2023. DOI: 10.20473/pmnj.v9i1.43863.

[17] E. Purwita, "Determinants of stunting in children under five in rural areas," *Midwifery Journal*, vol. 10, no. 4, 2022.

[18] P. Y. Tan et al., "Prevalence and key determinants of the triple burden of childhood malnutrition," *Critical Reviews in Food Science and Nutrition*, vol. 65, no. 28, pp. 5683–5697, 2024. DOI: 10.1080/10408398.2024.2419539.

[19] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd. Chapman and Hall/CRC, 2014. Available online.

[20] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. Wiley, 2022. Available online.

[21] P. D. Hoff, *A First Course in Bayesian Statistical Methods*. Springer, 2021. Available online.

[22] O. Bilal, A. Hekmat, I. Shahzad, et al., "Boosting machine learning accuracy for cardiac disease prediction," *Review of Socionetwork Strategies*, vol. 19, pp. 271–300, 2025. DOI: 10.1007/s12626-025-00190-w.

[23] Y. Wang and D. M. Blei, "The blessings of multiple causes," *Journal of the American Statistical Association*, vol. 114, no. 528, pp. 1574–1596, 2019. DOI: 10.1080/01621459.2019.1604364.

[24] D. Wu, H. G. Park, C. R. Grudzen, and K. S. Goldfeld, "Bayesian hierarchical penalized spline models," *Statistics in Medicine*, vol. 44, e10304, 2025. DOI: 10.1002/sim.10304.

[25] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, *Global Sensitivity Analysis: The Primer*. John Wiley & Sons, 2008. Available online.

[26] A. A. R. Fernandes and S. Solimun, *Applied Multivariate Statistics for Social and Management Sciences*. Springer, 2022.

[27] B. Tobias, L. Kuhlmann, and L. Fahrmeir, "Adaptive bayesian nonparametric regression with dynamic smoothness priors," *Journal of Computational and Graphical Statistics*, vol. 34, no. 2, pp. 456–474, 2025. DOI: 10.1080/10618600.2025.2450458.

[28] A. Thielmann, T. Kneib, and B. Säfken, "Enhancing adaptive spline regression," *Journal of Computational and Graphical Statistics*, vol. 34, no. 4, pp. 1397–1409, 2025. DOI: 10.1080/10618600.2025.2450458.

[29] E. Edeh, X. Liang, and C. Cao, "Probing beyond: The impact of model size and prior informativeness," *Behavior Research Methods*, vol. 57, p. 108, 2025. DOI: 10.3758/s13428-025-02609-2.

[30] A. A. R. Fernandes, Solimun, and S. N. U. Zahra, *Bayesian Truncated Spline Regression for Stunting Models*. Malang, Indonesia: Faculty of Mathematics and Natural Sciences, Brawijaya University, 2025.

[31] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 8th. Cengage Learning, 2019. Available online.

[32] P. Bach and N. Klein, "Anisotropic multidimensional smoothing using bayesian tensor product p-splines," *Statistics and Computing*, vol. 35, p. 43, 2025. DOI: 10.1007/s11222-025-10569-y.