PARAMETER
JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA

e-ISSN. 2829-6303
Vol. 05, No. 1, September 2026
Pages 403-420

Nonlinear Small Area Poverty Estimation with MARS-Based Feature Reduction

Ajriansyah[1], Bambang Widjanarko Otok[1]*, Sutikno[1]

[1]Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, East Java, Indonesia

E-mail Correspondence Author: *bambang_wo@its.ac.id

Abstract

The use of small area estimation combined with multivariate adaptive regression splines (SAE-MARS) is intended to address the lack of estimation precision in cases with nonlinear patterns. Compared to previous studies, which generally treated MARS solely as a tool to capture nonlinear functions without feature/variable selection, the novelty of this study lies in introducing a stepwise variable reduction procedure based on GCV to eliminate uninformative predictors. Two modeling schemes were constructed: using all variables (SAE-MARS-full) and reducing insignificant variables (SAE-MARS-reduced). The SAE-MARS method was implemented using two-stage modeling. The first stage formed the MARS model, and then proceeded to the second stage to construct the SAE. SAE-MARS can deliver a sharp reduction in standard error; however, this does not apply globally. The SAE-MARS-reduced model performed slightly better because of its simpler structure and high level of model accuracy. Although the decrease in average global RRMSE between the two SAE-MARS models appears modest (13.02% for SAE-MARS-reduced compared to 13.37% for SAE-MARS-full), feature reduction yields a very consistent precision advantage across all regions. Thus, the proposed SAE-MARS-reduced model is empirically proven to effectively handle nonlinear patterns, produce highly precise estimates, and demonstrate that eliminating unimportant variables in the MARS process is crucial.

Keywords: *SAE-MARS, feature reduction, nonlinear, poverty, improving estimation precision.*

: https://doi.org/10.30598/parameterv5i2pp403-420

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright © 2026 The Author(s).
Research Ariticle ***Open Acces***
Email: jurnalparameter@gmail.com
Homepage: https://ojs3.unpatti.ac.id/index.php/parameter

Submitted: June 2026
Accepted: September 2026

1. INTRODUCTION

Based on the latest data, across all regions of Indonesia, the island of Java accounts for more than half of the national poor population [1], making it one of the regions that requires special attention. The mandate of the Sustainable Development Goals (SDGs), particularly the first goal, and the aspirations of "Indonesia Emas 2045" are realized through efforts to eradicate poverty; therefore, the Indonesian government must play a significant role in addressing this problem. As policymakers, the government must formulate various policies to control poverty, and this formulation process must be grounded in field data. Thus, the availability of up-to-date and accurate data, especially those that can provide details down to small-area levels, is crucial [2]–[4].

Surveys conducted by Statistics Indonesia (BPS), including the national socioeconomic survey, often produce accurate estimates at the national or provincial levels. However, when disaggregated to lower levels, such as regencies/municipalities, the accuracy of these surveys declines, and for certain indicators, survey results may even be unsuitable for publication due to significant bias. At more granular levels, such as subdistricts and villages/urban villages, the sample sizes tend to be small, and the variances are high owing to various constraints [5], [6]. In this condition, subpopulation parameter estimates using direct estimation have large sampling variance and high standard error and are not sufficiently reliable for decision making [3], [7]. Therefore, the small area estimation (SAE) method was developed to address such cases [8]. This method reduces the standard error (SE) of small area parameter estimates using the "borrowing strength" concept from two directions, namely, from other areas and auxiliary variables, thereby enabling the correction of previously less reliable estimates to become more accurate [6], [8], [9].

The use of standard SAE models, such as the empirical best linear unbiased predictor (EBLUP), in poverty cases cannot guarantee effective estimation results because these models require linearity assumptions to be met, whereas the relationships between poverty parameters and their auxiliary variables are often found to be nonlinear [7], [10]–[12]. To address this issue, SAE models can be combined with nonlinear models, such as multivariate adaptive regression splines (MARS), to form a nonlinear SAE-MARS model [13]. Although the use of the SAE-MARS model was introduced in [13], this study specifically focuses on proposing a framework within SAE-MARS by first implementing MARS-based variable reduction before proceeding to the SAE stage. Compared to previous research on SAE-MARS, this study aims to address a crucial challenge regarding the practice of retaining all auxiliary variables during the bootstrap process to obtain the mean squared error (MSE), even though some variables are already known not to contribute significantly to the model.

This practice may introduce unnecessary complexity and increase model uncertainty during the resampling process. Instead of using all variables in the model, this study investigated how the gradual elimination of irrelevant auxiliary variables prior to SAE integration can optimize hyperparameter tuning and stabilize the bootstrap estimation process. Thus, the main contribution of this study is to provide evidence that reducing insignificant variables in SAE-MARS not only simplifies the model but can also systematically improve estimation precision in the observed small areas. When uninformative or redundant predictors are kept in the candidate set, the adaptive forward-pass of MARS during repeated parametric bootstrap draws tends to fit spurious knot points and interaction terms on random sampling noise. This high model-selection instability across resampling iterations artificially inflates the variability of synthetic predictions, leading to overestimated bootstrap MSE. To resolve this issue, the proposed

framework introduces a systematic MARS-based feature reduction procedure prior to SAE integration. By pre-filtering variables with 0% importance, the hyperparameter search space for knot placement during the 10,000 bootstrap replications is effectively constrained.

This step prevents noise propagation, stabilizes basis-function formation across resampling draws, and ultimately yields more reliable, low-variance MSE estimates across areas. Accordingly, the use of MARS in auxiliary variable reduction and nonlinear SAE modeling to obtain more precise estimates compared to direct estimation becomes an intriguing approach. Furthermore, in previous papers, only 78 areas had actual relative standard error (RSE) values out of a total of 119 areas **[13]**, whereas in this study, each observation area is equipped with its own actual RSE value, allowing model comparisons to involve all observation areas. This study is expected to broaden the application of MARS in SAE modeling, particularly by incorporating the reduction of unimportant variables to improve the precision of poverty rate estimates on Java Island, Indonesia. This study is also expected to contribute to efforts to provide valid micro-level data to support data-driven policy formulation in addressing poverty in Java and Indonesia.

2. MATERIALS AND METHODS

2.1. Study Area and Dataset

This study focuses on improving the precision of estimating the percentage of the poor population in each Regency/City on the island of Java, totaling 119 regions, along with several auxiliary variables. The total of 119 regions comprises 6 provinces, including DKI Jakarta (6 regions), Banten (8 regions), DI Yogyakarta (5 regions), West Java (27 regions), Central Java (35 regions), and East Java (38 regions). All datasets used are sourced from publications by Statistics Indonesia for the year 2024, accessed via the BPS website for each province (for example, for the Regency/City dataset in DKI Jakarta: https://jakarta.bps.go.id/id/query-builder). Each provincial BPS portal provides complete information on the variables used for each region in this study. The research variables are presented in **Table 1**.

Table 1. Research Variables

Variables	Scale (Unit)	Mean	Min	Max	Std. Dev.
Percentage of Poor Population (Y)	Ratio (%)	9.07	2.34	20.83	3.70
Percentage of Households with Adequate Sanitation (X1)	Ratio (%)	84.75	45.88	99.34	13.37
Percentage of Households with Access to Safe Drinking Water (X2)	Ratio (%)	96.05	77.92	100	4.56
Labor Force Participation Rate (X3)	Ratio (%)	71.58	58.13	86.62	4.65
Open Unemployment Rate (X4)	Ratio (%)	5.03	1.56	9.18	1.79
Gross Regional Domestic Product per Capita (Constant Prices) (X5)	Ratio (Thousand Rupiah)	48174.94	14132.83	510015	62769.53
Average Years of Schooling (X6)	Ratio (Year)	8.77	5.08	12.12	1.66
Percentage of Per Capita Expenditure on Food (X7)	Ratio (%)	50.88	36.8	65.01	6.45

Variables	Scale (Unit)	Mean	Min	Max	Std. Dev.
Population Density (X8)	Ratio (Population/ km²)	3268.36	385	21955.16	4402.78
Percentage of Population with Health Insurance (X9)	Ratio (%)	72.33	35.25	97.71	13.82

The author ensured that administrative and temporal consistency was strictly maintained by aligning the list of regions used with the administrative data of each province, particularly for 2024, to avoid the exclusion of any regions and ensured that all variables referred to the 2024 calendar year. As shown in **Table 1**, the dataset exhibits substantial cross-regional variation, particularly in economic capacity (GRDP per capita, X5) and demographic pressure (population density, X8), which reflects the structural socioeconomic diversity across urban municipalities and rural regencies in Java.

2.2. Direct Estimation and Sampling Variance

The percentage of the poor population in each regency/city on the island of Java in this study was the result of direct estimation in the national socioeconomic survey by the BPS. This variable was treated as the response in both MARS and SAE-MARS modeling. Each region that has a direct estimation value in this study is accompanied by a relative standard error (RSE) value, and the RSE itself is a tool to measure how precise an indicator's estimate is, resulting in 119 observed regions, each complete with the precision measure of its estimated value. The BPS has established standards concerning the relative standard error (RSE) in direct estimation. There are three categories of an estimate's value based on the resulting RSE: First, estimates with RSE < 25% are categorized as accurate; second, estimates with RSE between 25% and 50% fall into the category that requires caution in their use; and third, if the RSE > 50%, then the estimate is considered inaccurate **[14]**. The dataset used in this study only had RSE values falling into the first or second categories. Specifically, for direct estimation values in the second category, the small area estimation (SAE) method can be used to improve precision **[14].** SAE modeling does not require the RSE value; rather, sampling variance is needed. Because there is a mathematical relationship between RSE and sampling variance, the estimator of sampling variance $\left(\hat{\sigma}_i^2\right)$ can be obtained by the **Equation (1) [13]**.

$$\hat{\sigma}_i^2 = \left(SE\left(\hat{\theta}_i\right)\right)^2 = \left(\frac{RSE\left(\hat{\theta}_i\right)}{100\%} \times \hat{\eta}_i\right)^2 \quad (1)$$

with $\hat{\sigma}_i^2$ is the estimator of sampling variance, and SE is standard error. The application of this formulation rests on three standard design assumptions: (i) direct survey estimators are design-unbiased for the true domain parameter θ_i; (ii) sampling errors are normally distributed as $e_i \sim N\left(0, \sigma_i^2\right)$; and (iii) the estimated sampling variances $\hat{\sigma}_i^2$ are treated as known reference parameters. It should be noted that, theoretically, Fay-Herriot modeling assumes sampling variance σ_i^2 is constant and known. However, in practice, this value may contain estimation uncertainty because it is estimated from survey sample data, particularly if estimated for areas with small sample sizes. According to standard conventions in the small area estimation literature based on area **[8]**, **[15]** the value of $\hat{\sigma}_i^2$ in this study is treated as a known reference value in the development of the SAE-MARS model and the shrinkage factor. Although treating $\hat{\sigma}_i^2$ as

fixed ignores potential sampling variability in microdata—which directly governs the shrinkage factor—the substantial sample sizes per district in Susenas 2024 ensure highly stable variance estimates across Java. Moreover, the 10,000-replication parametric bootstrap procedure conditions on σ_i^2 while explicitly propagating the estimation uncertainty associated with the area-level random effect variance ($\hat{\sigma}_u^2$). Regarding uncertainty in auxiliary variables, all auxiliary variables are reliably and completely available for all 119 districts/cities in Java Island with no missing data. In terms of data source independence, the research variables can be divided into three categories: first, non-survey data based on registration and macro balance, namely GRDP per capita (X5) and population density (X8), which are free from household sampling error; second, independent survey data originating from the 2024 Sakernas (national labor force survey), namely labor force participation rate (X3) and open unemployment rate (X4), which have a sampling frame separate from Susenas (national socio-economic survey); and third, welfare data based on the 2024 Susenas (X1, X2, X6, X7, X9). Auxiliary variables originating from the same survey have the potential to introduce sampling error covariance or measurement uncertainty. Nevertheless, according to standard practice in applied Fay-Herriot modeling **[15]**, officially published area-level datasets are treated as fixed covariates.

2.3. Nonlinear Small Area Estimation

Small Area Estimation (SAE), originally developed by Fay and Herriot **[8]**, is a model-based indirect estimation method designed to estimate subpopulation parameters with very small or even zero sample sizes **[8], [16]–[18]**. The key to SAE in performing its task lies in extracting information ("borrowing strength") and using it to form weights (shrinkage factor) that improve the variance of the estimated values. The process of "borrowing strength" is conducted from two sources: areas with similar conditions and auxiliary variables **[5], [6], [19]**. The standard Fay-Herriot SAE model consists of two model hierarchies. The first is called the sampling model: $\hat{\theta}_i = \theta_i + e_i$, and the second is the linking model: $\theta_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + u_i$. The combination of these two hierarchical models forms the Fay-Herriot model, as shown in **Equation (2)**.

$$\hat{\theta}_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + u_i + e_i \tag{2}$$

with, $\hat{\theta}_i$ is the value of the direct estimation (response); $i = 1, 2, \dots, n$; e_i is a sampling error for the i-th area; $e_i \sim N(0, \sigma_i^2)$, where σ_i^2 is sampling variance; $\boldsymbol{x}_i^T$ is an auxiliary variable vector $(1, x_{1i}, x_{2i}, \dots, x_{pi})$; $\boldsymbol{\beta}$ is the regression coefficient vector $(\beta_0, \beta_1, \dots, \beta_p)^T$; and u_i is the random effect for the i-th region (describes interregional variability that cannot be explained by $\boldsymbol{x}_i^T$ as well as being independent from e_i), $u_i \sim N(0, \sigma_u^2)$, where σ_u^2 is random effects variance (representing the remaining inter-regional variation after accounting for the linking model).

The limitation of the linearity assumption in the Fay-Herriot SAE can be addressed by combining MARS and SAE. MARS, first introduced by Friedman **[20]**, is a nonparametric regression model capable of handling high-dimensional and nonlinear patterned data because it is data-driven and can detect patterns even when the functional form of the data is unknown **[20]–[23]**. The general form of the MARS model is presented in the following **Equations**:

$$f(\boldsymbol{x}_i) = \beta_0 + \sum_{m=1}^{M} \beta_m \prod_{q=1}^{Q_m} \left(s_{qm} \left(x_{iv(q,m)} - k_{qm} \right) \right)_+ \tag{3}$$

$$\theta_i = \beta_0 + \sum_{m=1}^{M} \beta_m \prod_{q=1}^{Q_m} \left(s_{qm} \left(x_{iv(q,m)} - k_{qm} \right) \right)_+ + \varepsilon_i \tag{4}$$

$$\theta_i = \beta_0 + \sum_{m=1}^{M} \beta_m B_m(\boldsymbol{x}_i) + \varepsilon_i \tag{5}$$

In summary, this can be written as follows:

$$\theta_i = f(\boldsymbol{x}_i) + \varepsilon_i \tag{6}$$

with β_0 = intercept; β_m = coefficient of the m-th basis function; $B_m(\boldsymbol{x_i})$ = m-th basis function; $\boldsymbol{x}_i$ = predictor variable vector; and ε_i = random error of the i-th observation, where $\varepsilon_i \sim iid\ N(0, \sigma^2)$. Specifically for basis function $B_m(\boldsymbol{x_i})$, at each knot point, a pair of basis functions in the form of hinge functions operating in opposite directions is provided, which can be written as:

$$\begin{aligned} \left(+\left(x_{iv(q,m)} - k_{qm} \right) \right)_+ &= \max\left(0, x_{iv(q,m)} - k_{qm}\right) \\ &= \begin{cases} x_{iv(q,m)} - k_{qm}, & if\ x_{iv(q,m)} > k_{qm} \\ 0, & if\ others \end{cases} \end{aligned} \tag{7}$$

$$\begin{aligned} \left(-\left(x_{iv(q,m)} - k_{qm} \right) \right)_+ &= \max\left(0, k_{qm} - x_{iv(q,m)}\right) \\ &= \begin{cases} k_{qm} - x_{iv(q,m)}, & if\ x_{iv(q,m)} < k_{qm} \\ 0, & if\ others \end{cases} \end{aligned} \tag{8}$$

with $x_{iv(q,m)} \in \{x_j\}_{j=1}^{p}$ = predictor variable number-v; $k_{qm} \in \{x_{iv(q,m)}\}_{i=1}^{n}$ = knot point.

The process of constructing the SAE-MARS model, which is designed to address nonlinear cases, is carried out by substituting the nonlinear equation $f(\boldsymbol{x}_i)$ MARS in **Equation (6)** into the SAE model replaces the linear equation $\boldsymbol{x}_i^T \boldsymbol{\beta}$ in **Equation (2)**, and forms the following **Equation [13]**:

$$\hat{\theta}_i = f(\boldsymbol{x}_i) + u_i + e_i \tag{9}$$

In this case, the estimator $\hat{f}(\boldsymbol{x}_i)$ is estimated from the best MARS model through a forward-backward process, then this estimator is treated as a synthetic estimator. The transformation of the synthetic estimator, which was previously $\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}$, into $\hat{f}(\boldsymbol{x}_i)$ causes the form of the empirical estimator to also adjust to:

$$\hat{\theta}_i^{SAE-MARS} = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) \hat{f}(\boldsymbol{x}_i) \tag{10}$$

with shrinkage factor ($\hat{\gamma}_i$):

$$\hat{\gamma}_i = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \sigma_i^2} \tag{11}$$

Although the empirical estimator form of the SAE-MARS model follows the EBLUP estimator of Fay-Herriot, this does not make it an EBLUP estimator because it employs

a nonlinear equation in the linking model and the selection of the spline function is performed adaptively and is data-driven. The shrinkage factor serves to balance the estimation results between an estimate tending closer to either the direct estimator or the synthetic estimator, depending on which has relatively better precision. Meanwhile, on the other hand, because $\hat{f}(\boldsymbol{x}_i)$ is treated as a fixed value that is assumed to be known and ignores the variance of its estimate, a parametric bootstrap is used to estimate the MSE.

2.4. MARS-Based Feature Reduction

The MARS model can flexibly accommodate data shapes. The key to MARS's flexibility lies in the ability of the basis functions to adapt to changes in the data slope. In summary, the MARS algorithm finds the optimal equation through two stages: the forward iteration, which uses a greedy search system to find the model equation with the smallest error but is overfitted, and the backward iteration, which removes insignificant basis functions from the overfitted model until it achieves the optimal model condition based on the smallest general cross-validation (GCV) criterion [20]. Similar to other machine learning models, MARS is constructed by performing hyperparameter tuning to determine the best combination of model hyperparameters. In this study, a total of 3×3×5×1 = 45 model scenarios were constructed based on combinations of hyperparameters, including the maximum number of basis functions (Max BF) by setting it at 2, 3, and 4 times the number of auxiliary variables; maximum interaction (MI) with interaction degrees of 1, 2, and 3; minimum observation (MO) by setting it at 0, 1, 2, 3, and 5 observations; and penalty = 3. The rationale for determining the boundaries in hyperparameter tuning in this study is to follow the conventions recommended by Friedman [20]. MARS modeling is conducted before being combined with the SAE model in order to first obtain $f(\boldsymbol{x}_i)$ from the best MARS model [13]. The modeling workflow for variable reduction based on MARS is carried out stepwise as follows:

1. Conducting MARS modeling based on the forward-backward algorithm by following hyperparameter tuning, which will produce a number of the best candidate MARS models
2. One of these candidate models will be selected as the best model based on the criterion of having the smallest GCV value
3. The best MARS model produces output in the form of variable importance levels (based on GCV or residual sum of squares) presented as percentages, where a predictor variable is considered unimportant when its importance level is 0%
4. If there is one or more predictor variable with an importance level of 0%, then those variables are subsequently excluded and the modeling is repeated from step 1
5. The iterative modeling process was stopped if there were no more predictor variables with an importance level of 0%. If, in the initial modeling, all predictor variables have an importance level >0%, then it can proceed directly to the integration stage with the SAE model.

2.5. Experimental Design

This study compares the MARS model using all auxiliary variables with the MARS model conducted in stages to ensure that all auxiliary variables are important variables to be combined with the SAE model, which then forms the SAE-MARS model. Furthermore, the implementation of the SAE-MARS model was aimed at reducing the standard error (SE) to obtain more precise poverty estimation results. This design enabled a comparison between three estimation methods: direct estimation, SAE-MARS

with all variables, and SAE-MARS with reduced variables. The SAE-MARS algorithm consists of two stages: the first stage is MARS modeling, and the second stage is SAE-MARS modeling. In the first stage, MARS modeling is performed by including all variables or using a stepwise process to reduce uninformative variables. Subsequently, the best model from the two MARS model schemes was used in the SAE-MARS model until the model estimation results were obtained, along with the mean squared error (MSE) for metric calculation in comparing the performance of the methods. The performance of the three estimation methods was measured based on the evaluation metrics. Several evaluation metrics were used, namely RSE, relative root mean squared error (RRMSE), relative efficiency (RE), RSE reduction, and R-squared. Specifically regarding RSE and RRMSE, these two metrics were compared directly. The RSE measures the performance of direct estimation, whereas the RRMSE measures the performance of the SAE-MARS model. The form of the MSE in the SAE-MARS model cannot be derived analytically; therefore, it does not have a closed form; therefore, the bootstrap method is applied to obtain the value of the MSE. The type of bootstrap used was a parametric bootstrap with 10,000 replications, in which the sampling variance was retained and model selection uncertainty was included. The following is the bootstrap flow for obtaining the MSE of the SAE-MARS model.

1. Fitting SAE-MARS model:
   The final SAE-MARS model was obtained $\hat{f}(\boldsymbol{x}_i)$, σ_u^2, and $\hat{\theta}_i$
2. Generate random effects:
   For each region:
   $$u_i^{*(b)} \sim N(0, \sigma_u^2) \tag{12}$$
3. Generate sampling error:
   $$e_i^{*(b)} \sim N\left(0, \sigma_i^2\right) \tag{13}$$
4. Generate SAE-MARS bootstrap data:
   $$\hat{\theta}_i^{*(b)} = \hat{f}(\boldsymbol{x}_i) + u_i^{*(b)} + e_i^{*(b)} \tag{14}$$
   where the actual bootstrap response of SAE-MARS is
   $$\theta_i^{*(b)} = \hat{f}(\boldsymbol{x_i}) + u_i^{*(b)} \tag{15}$$
5. Fitting the SAE-MARS model again:
   Obtain $\hat{f}^{*(b)}(\boldsymbol{x}_i)$ and $\hat{\sigma}_u^{2\,*(b)}$ from data $\hat{\theta}_i^{*(b)}$ and $\boldsymbol{x_i}$ by performing the SAE-MARS modeling process
6. Calculate the predicted response using the modified estimator:
   The modified result estimators for SAE-MARS can be applied to the bootstrap in the following manner:
   $$\hat{\theta}_i^{SAE-MARS\,*(b)} = \hat{\gamma}_i^{*(b)} \hat{\theta}_i^{*(b)} + \left(1 - \hat{\gamma}_i^{*(b)}\right) \hat{f}^{*(b)}(\boldsymbol{x}_i) \tag{16}$$

where:

$$\hat{\gamma}_i^{*(b)} = \frac{\hat{\sigma}_u^{2\,*(b)}}{\hat{\sigma}_u^{2\,*(b)} + \sigma_i^2} \tag{17}$$

7. Squared error bootstrap SAE-MARS calculation:

$$squared\ error_i = \left(\hat{\theta}_i^{SAE-MARS*(b)} - \theta_i^{*(b)}\right)^2 \quad (18)$$

8. Repeat steps ii to vii B times, with $b = 1,2, \ldots, B$. The value of B is set up to 10,000
9. Get MSE SAE-MARS:

$$MSE\left(\hat{\theta}_i^{SAE-MARS}\right) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}_i^{SAE-MARS*(b)} - \theta_i^{*(b)}\right)^2 \quad (19)$$

3. RESULTS AND DISCUSSION

3.1. Exploratory Data Analysis (EDA)

In this section, an analysis is conducted to examine the condition of the dataset prior to modeling. The pattern of the relationship between the response and each auxiliary variable is shown in **Figure 1**.

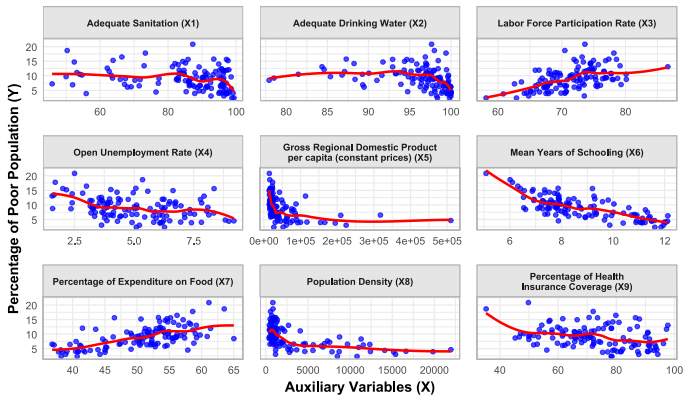


Figure 1. Pattern of the Relationship Between Poverty Percentage and Auxiliary Variables

Based on **Figure 1**, the relationship between the response variable and each auxiliary variable tended to follow a nonlinear pattern. GRDP per capita and population density visually exhibited the most nonlinear patterns relative to the response. Subsequent analysis measured the strength of the relationship between the response and auxiliary variables. **Figure 2** presents a visualization of the correlation strength.

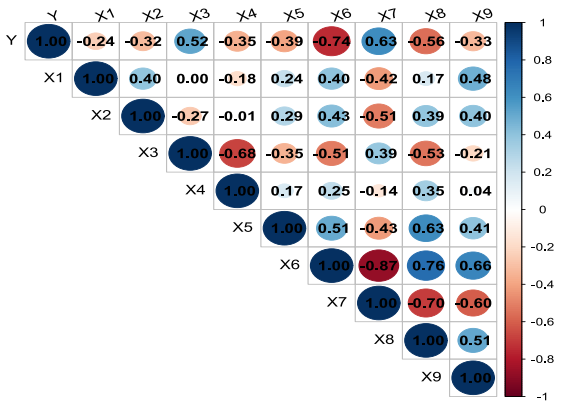


Figure 2. Strength of Correlation between Poverty Percentage and Auxiliary Variables

The correlation between the auxiliary variables and each response was statistically significant; therefore, all auxiliary variables were used as predictors. In **Figure 2**, the correlation between the average years of schooling and the response was the strongest, while the percentage of households with adequate sanitation had the lowest correlation, with a correlation value of only –0.24. Some variables showed indications of possible multicollinearity, for example, between the average years of schooling and the percentage of per-capita expenditure on food. Such cases can affect the performance of the MARS model. Although MARS can automatically select important variables, it may have difficulty determining more objective knot-point values. As empirical support, the following are the variance inflation factor (VIF) values for each auxiliary variable: (X1 = 1.63, X2 = 1.58, X3 = 2.68, X4 = 2.09, X5 = 1.74, X6 = 6.57, X7 = 4.89, X8 = 3.53, X9 = 2.05). Two variables had relatively high VIF values, namely X6 and X7; however, they did not exceed the VIF threshold of 10, so it can be concluded that none of the auxiliary variables exhibited multicollinearity.

3.2. MARS Modeling

Before the modeling was performed, there were 45 hyperparameter combinations, and each of these combinations was used to build a model. In addition, the modeling process was divided into two types: first, the MARS model with all predictor variables, and second, the MARS model that reduced the unimportant variables.

3.2.1. MARS with full variables

In this section, MARS modeling was conducted using all auxiliary variables. For ease of identification, this type of model is referred to as the MARS-full model. After 45 model scenarios were constructed, each model's GCV value was compared, and the MARS-full model with the smallest GCV (the best model) was obtained, with a GCV value of 4.87, resulting from a combination of hyperparameters, including maximum basis function (Max BF) = 18, maximum interaction (MI) = 2, minimum observation (MO) = 3, and penalty = 3. The model also outputs the variable importance levels. **Table 2** presents the importance of the auxiliary variables.

Table 2. Importance Level of MARS-full Auxiliary Variables

Variables	Importance	
	GCV	RSS
Average Years of Schooling (X6)	100	100
Percentage of Households with Access to Safe Drinking Water (X2)	53.5	60.7
Percentage of Population with Health Insurance (X9)	35.2	46.3
GRDP per Capita (X5)	32.3	42.7
Percentage of Households with Adequate Sanitation (X1)	19.5	33.3
Population Density (X8)	19.5	33.3
Labor Force Participation Rate (X3)	19.4	27.4
Open Unemployment Rate (X4) – unused	0	0
Percentage of Per Capita Expenditure on Food (X7) – unused	0	0

Based on **Table 2**, the average years of schooling variable is the most important in this model because it is the most dominant in explaining the variation in the percentage of poor population, while there are two variables that do not contribute, namely the open unemployment rate and the percentage of per capita expenditure on food. Although there are two auxiliary variables that are not important, $f(\boldsymbol{x}_i)$ function from the best MARS-full model will be used in the SAE-MARS modeling.

3.2.2. MARS with variable reduction

The MARS model at this stage can be referred to as MARS-reduced, since the modeling was carried out in stages to obtain a model in which each predictor is an important variable. The best model was also built based on the existing 45 scenarios, where the best model had the smallest GCV value of 4.87, with a combination of hyperparameters including maximum basis function (Max BF) = 14, maximum interaction (MI) = 2, minimum observation (MO) = 3, and penalty = 3. In this section, two rounds of MARS modeling were performed to obtain the final MARS-reduced model. The first MARS model used all variables; however, two variables were found to be unimportant. Therefore, these variables were not included in the second MARS. This subsequently produced the variable importance levels for the auxiliary variables for the MARS-reduced model, as presented in **Table 3**.

Table 3. Importance Level of MARS-reduced Auxiliary Variables

Variables	Importance	
	GCV	RSS
Average Years of Schooling (X6)	100	100
Percentage of Households with Access to Safe Drinking Water (X2)	53.5	60.7
Percentage of Population with Health Insurance (X9)	35.2	46.3
GRDP per Capita (X5)	32.3	42.7
Percentage of Households with Adequate Sanitation (X1)	19.5	33.3
Population Density (X8)	19.5	33.3
Labor Force Participation Rate (X3)	19.4	27.4

The optimal model function $f(\boldsymbol{x}_i)$ from the MARS-reduced will also be used in SAE-MARS. Based on **Table 3**, the magnitude and order of the contributions of each auxiliary variable are exactly the same as those shown in **Table 2**; the only difference is that variables deemed unimportant in the MARS-full are excluded from the MARS-reduced model.

3.3. SAE-MARS Modeling

The SAE-MARS modeling process was divided into two stages: first, obtaining the best MARS model, either from the MARS-full or MARS-reduced. Then, modeling nonlinear SAE by replacing the linear elements of the SAE EBLUP with $\hat{f}(\boldsymbol{x}_i)$ functions from the best MARS model in the first stage. The $\hat{f}(\boldsymbol{x}_i)$ equation MARS-full and MARS-reduced are sequentially written in **Equations (20)-(21)**.

$$\begin{aligned}\hat{f}(\boldsymbol{x}_i)^{MARS-full} &= 8.34 + 0.14(BF_1) - 0.16(BF_2) - 1.33(BF_3) \\ &+ 0.096(BF_4) + 0.061(BF_5BF_6) + 0.0002(BF_7BF_8) \\ &+ 0.15(BF_9BF_3) + 0.000007(BF_{10}BF_4) \\ &- 0.0005(BF_{11}BF_4)\end{aligned} \tag{20}$$

$$\begin{aligned}\hat{f}(\boldsymbol{x}_i)^{MARS-reduced} &= 8.34 + 0.14(BF_1) - 0.16(BF_2) - 1.33(BF_3) \\ &+ 0.096(BF_4) + 0.061(BF_5BF_6) + 0.0002(BF_7BF_8) \\ &+ 0.15(BF_9BF_3) + 0.000007(BF_{10}BF_4) \\ &- 0.0005(BF_{11}BF_4)\end{aligned} \tag{21}$$

The nonlinear equations produced by the best MARS-full and MARS-reduced models were identical. **Equations (20)** and **(21)** each contain 11 BFs, which result in the greatest

reduction in the residual sum of squares (RSS). Because the BFs in **Equations (20) and (21)** are of the same form, the structure of each BF is presented in **Table 4**.

Table 4. MARS-full and MARS-reduced Basis Function Forms

Basis functions		
$BF_1 = h(67.46 - X1)$	$BF_5 = X2$	$BF_9 = h(96.96 - X2)$
$BF_2 = h(74.55 - X3)$	$BF_6 = h(7.16 - X6)$	$BF_{10} = h(31901 - X5)$
$BF_3 = h(X6 - 7.16)$	$BF_7 = h(X1 - 67.46)$	$BF_{11} = h(847 - X8)$
$BF_4 = h(X9 - 52.84)$	$BF_8 = h(1141 - X8)$	

Based on the modeling results of SAE-MARS-full and SAE-MARS-reduced, no striking differences were observed; however, differences in the combination of hyperparameters, particularly the maximum number of basis functions and the number of auxiliary variables used, open up the possibility of differences in the estimation results at the SAE-MARS modeling stage, especially in the process of estimating MSE.

3.4. Comparison of Estimation Method Performance

In this section, the performance of three methods—direct estimation, MARS-full, and MARS-reduced — are compared. This comparison uses the main metrics, namely, RSE and RRMSE. RSE represents the precision measure for the direct estimation, whereas RRMSE reflects the precision produced by SAE-MARS-full and SAE-MARS-reduced. Although these two metrics are distinguished in terms of their context of use, they can nevertheless be directly compared because, mathematically, both function as measures of relative uncertainty normalized on the same scale. In the direct estimator $\hat{\theta}_i$, based on the bias design in the survey it is assumed that $(Bias\ (\hat{\theta}_i) \approx 0)$, can mean that $MSE\ (\hat{\theta}_i) = Var(\hat{\theta}_i) + [Bias\ (\hat{\theta}_i)]^2 = [SE(\hat{\theta}_i)]^2$, thus $RMSE\ (\hat{\theta}_i) = SE(\hat{\theta}_i)$, so that $RRMSE\ (\hat{\theta}_i) = RSE(\hat{\theta}_i)$. Meanwhile, in the SAE method, it is necessary to measure model uncertainty through MSE in order to form:

$$RRMSE\ (\hat{\theta}_i^{SAE-MARS}) = \frac{\sqrt{\widehat{MSE}(\hat{\theta}_i^{SAE-MARS})}}{\hat{\theta}_i^{SAE-MARS}} \times 100\% \quad (22)$$

In addition, the two metrics differ in their estimation procedures, with the RSE directly reflecting the sampling design variance through complex survey mechanisms (stratification and clustering), whereas the RRMSE is obtained through a parametric bootstrap process to capture both area-level random effects and the variability of the MARS model selection. Therefore, evaluating the RSE against the RRMSE provides an established and mathematically grounded benchmark for precision improvement in small area applications. Other metrics to be used include relative efficiency (RE), RSE reduction, and R-squared; however, these metrics are not used in the context of comparing direct estimation with SAE-MARS, but rather to compare the two SAE-MARS models—SAE-MARS-full and SAE-MARS-reduced— to determine which model is better at reducing the standard error in the estimation of the percentage of poor population on Java island. A comparison of the RSE and RRMSE is presented in **Figure 3**.

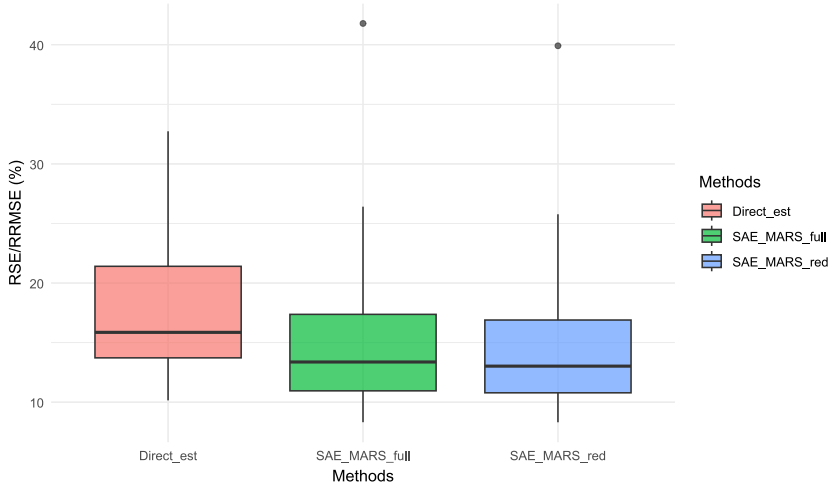


Figure 3. Boxplot comparison of RSE and RRMSE

The results shown in **Figure 3** indicate that the SAE-MARS model can drastically improve the average estimation precision. However, one area stands out significantly, yielding an RRMSE value much higher than the others, and far exceeding the RSE results in the direct estimation. The area identified as an outlier was South Tangerang City in Banten Province. In the case of South Tangerang City—a highly urbanized metropolitan area with extreme feature values (e.g., Average Years of Schooling > 11.8 years) and an already precise direct estimate—the SAE-MARS model yields a higher RRMSE owing to synthetic extrapolation bias and shrinkage distortion against the low-variance direct survey estimate. While a dedicated, granular case study of metropolitan outliers is beyond the primary scope of this paper, this instance highlights the operational boundaries of non-parametric SAE when direct survey estimates are already highly reliable. On average, the SAE-MARS model—particularly the SAE-MARS-full and SAE-MARS-reduced—manages to sharply reduce the standard error, although not uniformly across all regions. When comparing the models, SAE-MARS-reduced performed slightly better, with an average RRMSE of 13.02%, compared to SAE-MARS-full's average RRMSE of 13.37%. Further comparisons between the SAE-MARS models are listed in **Table 5**.

Table 5. SAE-MARS-full vs SAE-MARS-reduced

Metrics	Models	
	SAE-MARS-full	SAE-MARS-reduced
Relative Efficiency (RE)	1.48	1.54
Average RSE Reduction (%)	17.16	18.70
R-Squared (%)	89.67	89.67

Table 5 shows that SAE-MARS-reduced is more efficient than SAE-MARS-full. Moreover, this model also performs better in terms of average RSE reduction percentage; although the difference is not very large, there is still a distinction between the MARS model with all auxiliary variables and the MARS model that first reduces unimportant variables. For the R-squared metric, both models yield the same value, indicating that both models explain the response variable equally well. The difference observed arises from technical differences in obtaining the MSE. Thus, overall, SAE-MARS-reduced becomes an estimation method that, at least slightly, outperforms in aggregate evaluation metric comparisons. To provide a more objective perspective, detailed

evaluation by region consistently shows better performance. Specifically, SAE-MARS-reduced achieves lower MSE in 107 out of 119 regencies/cities (89.92%), while SAE-MARS-full achieves lower MSE in only 12 regencies (10.08%). The compression of this global average difference is primarily driven by a small number of influential urban outliers (notably South Tangerang City). From the model-building process to the estimation value acquisition, there was no significant difference between the two types of SAE-MARS. However, once the bootstrap stage was entered to calculate the MSE, differences began to emerge. The MSE for 119 areas produced by SAE-MARS-reduced was mostly lower than the MSE produced by SAE-MARS-full, as evidenced by the fact that only 12 areas showed SAE-MARS-full outperforming SAE-MARS-reduced in terms of MSE reduction. In the remaining 107 areas, SAE-MARS-reduced was the winner. Although not globally superior, SAE-MARS-reduced remains more advantageous. Thus, even though MARS is classified as a model that automatically eliminates unimportant variables, the decision to retain these unimportant variables in the model, especially during the bootstrap stage, can result in higher MSE values. One possible reason is that the MARS-full model is more complex and contains more noise during the model reformation process when bootstrap is performed, compared to the simpler MARS-reduced model, where it is clear that the variables used do indeed contribute.

The advantages of the SAE-MARS-reduced model are further reinforced by the comparative performance results of the two types of SAE-MARS models against the direct estimation method. It is evident that the improvement in precision over direct estimation by the SAE-MARS-reduced model is sharper than the precision improvement provided by the SAE-MARS-full model. This advantage is, of course, due to a smaller MSE. In terms of efficiency, both SAE-MARS models are much more efficient and precise than direct estimation. Furthermore, regarding the main metrics, namely RSE and RRMSE, **Figure 3** shows that, on average, the RRMSE of the SAE-MARS models, particularly the SAE-MARS-reduced, is lower than the average RSE of the direct estimation method. However, the aspect of model uncertainty in estimating MSE with bootstrap does not always proceed smoothly. A case in point is that there are several areas that yield an RRMSE greater than the RSE from direct estimation. SAE-MARS-full improved precision for 103 out of 119 areas (86.55%), while SAE-MARS-reduced performed slightly better with precision improvements in 89.92% of the total areas. Interestingly, the 12 areas whose precision was not improved by SAE-MARS-reduced all fall within the 16 areas whose precision was not improved by SAE-MARS-full. Therefore, there are four areas that were improved by SAE-MARS-reduced, which could not be improved by SAE-MARS-full. Although SAE-MARS-reduced does not guarantee an increase in precision in every area, eliminating irrelevant variables in MARS beforehand can increase the benefits yielded by the SAE-MARS model. To clarify the performance divergence between SAE-MARS-full and SAE-MARS-reduced, three distinct modeling mechanisms must be distinguished: **(1) Hyperparameter Tuning (Controlled Constant)**: Both models were evaluated over the identical 45-scenario tuning grid, selecting the same knot configuration and yielding identical baseline GCV values (4.87) and fitted functions $\hat{f}(x)$. Thus, tuning does not account for MSE differences. **(2) Feature Reduction**: Eliminating uninformative predictors (X4 and X7, both with 0% importance) restricts the candidate space from 9 to 7 variables without sacrificing explanatory power (R2 = 89.67%). **(3) Bootstrap Variability (Source of Precision Gains)**: The precision gain originates during the 10,000 parametric bootstrap replications. Retaining X4 and X7 in SAE-MARS-full allows MARS to occasionally fit spurious knots on synthetic sampling noise across draws. By contrast, SAE-MARS-

reduced excludes these variables prior to resampling, preventing noise-fitting during refitting and yielding lower MSEs in 107 out of 119 districts (89.92%). Although explicit sensitivity benchmarking across alternative random seeds and spatial cross-validation schemes was not conducted—representing a recognized study limitation—the B = 10,000 bootstrap setup ensures high internal numerical stability. The main results of this study reinforce the principle of parsimony in the MARS model [20], [24] and support the use of the two-stage MARS model [21]. Furthermore, this study makes a new contribution to the Small Area Estimation (SAE) literature by showing that refitting on the set of important variables selected by MARS elimination significantly yields a lower MSE than models using all variables, as demonstrated directly by the lower bootstrap Mean Squared Error (MSE) values achieved in 107 out of 119 districts/cities (89.92%). By eliminating uninformative predictors (X4 and X7), the search space for knot placement during 10,000 bootstrap replications becomes more restricted, thereby effectively reducing irrelevant disturbances and resulting in a more consistent reduction in the prediction error. Although B = 10,000 bootstrap samples significantly reduced the variance of the Monte Carlo simulation, extensive sensitivity evaluations of alternative seed configurations, variations in bootstrap sample size, and spatial cross-validation schemes were not performed explicitly.

4. CONCLUSION

This study strengthens the position of the nonparametric MARS method for use in SAE modeling. The implementation of the SAE-MARS, particularly with the SAE-MARS-reduced concept, is important because, in addition to accommodating nonlinear forms, the model also delivers a sharp decrease in the standard error of the estimated values, even though the reduction achieved is not comprehensive. In a small number of areas that are difficult to explain by the model, the use of SAE-MARS has not yet succeeded in improving the precision. Furthermore, there is one extreme case, namely in South Tangerang City, where the precision of both the SAE-MARS models is quite high compared to the initial condition under direct estimation. This condition objectively demonstrates the limitations of the SAE-MARS model, so that subsequent aligned research can examine it more rationally and in greater depth. Despite its various limitations, the SAE-MARS-reduced model performed slightly better than the SAE-MARS-full model, and both outperformed direct estimation across many evaluation metrics, indicating the success of the proposed model in achieving the research objectives. Nine auxiliary variables were used in the study, but only seven were important: average years of schooling, percentage of households with improved drinking water, percentage with health insurance, gross regional domestic product per capita at constant prices, percentage of households with improved sanitation, population density, and labor force participation rate. According to the GCV-based variable importance criteria, these variables exhibited strong empirical predictive utility in explaining the variation in poverty rates across regions on the island of Java. Average years of schooling was the most dominant variable in this study, indicating that poverty in each region of Java was related to education levels and human resources. It should be noted that in this study, the relationships presented between the auxiliary variables and poverty were not causal but rather associative, to help improve the estimation precision. This study has several limitations that need to be addressed. First, the empirical evaluation relies on a cross-sectional survey and administrative data from a single reference year (2024), which captures spatial variation across districts but cannot account for the longitudinal dynamics of poverty. Second, although the parametric bootstrap

procedure reflects the variability of model selection and resampling in obtaining the MSE, its validity depends on standard distributional assumptions regarding area-level random effects and sampling errors. Third, it has not yet explicitly conducted a sensitivity analysis, determined the variation in bootstrap sample size, or established a spatial cross-validation scheme. Further research could extend this framework to spatiotemporal models and explore nonparametric bootstrap variants to further relax distributional assumptions, as well as examine formal sensitivity testing and spatially structured resampling configurations to better validate the stability of nonparametric feature reduction in small area estimation. Despite existing limitations, overall, the use of MARS—particularly MARS-reduced—in constructing nonlinear SAE models successfully increased estimation precision in poverty cases across most observed regions, and this study may serve as a sufficiently informative reference for nonlinear SAE modeling, especially using the MARS approach for poverty cases on Java Island.

Acknowledgment

Through this writing, the authors would like to express their deepest gratitude to the Indonesia Endowment Fund for Education (LPDP) and the Department of Statistics, Institut Teknologi Sepuluh Nopember.

Funding Information

This research was funded by the Indonesia Endowment Fund for Education (LPDP).

Author Contributions Statement

Ajriansyah: Conceptualization, writing-original draft, software, data curation, and formal analysis. **Bambang Widjanarko Otok**: Writing-review & editing, methodology, validation, and draft preparation. **Sutikno**: Writing-review & editing, resources, software, and visualization. All authors discussed the results and contributed to the final manuscript.

Conflict of Interest Statement

Authors state no conflict of interest.

Data Availability

The data supporting the findings of this study are available from the corresponding author, [Bambang Widjanarko Otok], upon reasonable request.

REFERENCES

[1] BPS RI, "Jumlah Penduduk Miskin (Ribu Jiwa) Menurut Provinsi dan Daerah," *Badan Pusat Statistik (BPS) Republik Indonesia*, 2026. https://www.bps.go.id/id/statistics-table/2/MTg1IzI=/jumlah-penduduk-miskin-ribu-jiwa-menurut-provinsi-dan-daerah.html (accessed Jul. 19, 2026).

[2] H. Chandra, N. Salvati, and R. Chambers, "Small area estimation under a spatially non-linear model," *Comput. Stat. Data Anal.*, vol. 126, pp. 19–38, 2018, doi: 10.1016/j.csda.2018.04.002.

[3] S. Marchetti, N. Salvati, E. Fabrizi, and N. Tzavidis, "Small area estimation of equivalized income for local labour systems in Italy via M-quantile area-level models," *Stat. Methods Appt.*, vol. 34, no. 3, pp. 449–470, 2025, doi: 10.1007/s10260-025-00791-3.

[4] U. Maulana, M. Y. Darsyah, and T. W. Utami, "Small area estimation untuk pendugaan jumlah penduduk miskin di kota semarang dengan pendekatan kernel-bootstrap," *Statistika*, vol. 2, no. 2, 2014, [Online]. Available: https://garuda.kemdiktisaintek.go.id/documents/detail/1767627

[5] L. F. Trihandika, A. Ubaidillah, A. Az-Zahra, A. Amirudin, M. Rusydiana, and Z. Maharani, "Small Area Estimation Anak Tidak Sekolah di Pulau Kalimantan Tahun 2023," *Limits J.*

Math. Its Appl., vol. 21, no. 2, pp. 273–288, 2024, doi: 10.12962/limits.v21i2.21497.
[6] S. A. Yilema, Y. A. Shiferaw, H. M. Fenta, and A. T. Belay, "Estimating the local - level child full polio vaccination rates in Ethiopia using a hierarchical Bayes small area estimation approach," *Discov. Public Heal.*, vol. 22, no. 292, 2025, doi: 10.1186/s12982-025-00695-3.
[7] I. H. Jin, F. Liu, J. Park, E. Eugenio, and S. Liu, "Bayesian hierarchical spatial model for small-area estimation with non-ignorable nonresponses and its application to the NHANES dental caries data," *J. Korean Stat. Soc.*, vol. 53, no. 4, pp. 949–969, 2024, doi: 10.1007/s42952-024-00274-3.
[8] R. E. Fay and R. A. Herriot, "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," *J. ofthe Am. Stat. Assoc.*, vol. 74, no. 366, pp. 269–277, 1979, doi: 10.1080/01621459.1979.10482505.
[9] N. Permatasari and A. Ubaidillah, "Small Area Estimation of poverty using remote sensing data," *Stat. J. IAOS*, vol. 41, no. 1, pp. 180–190, 2025, doi: 10.1177/18747655241308390.
[10] B. Xu and R. Xu, "How can government expenditure effectively achieve energy poverty reduction? A non-linear perspective," *Energy*, vol. 335, p. 137760, 2025, doi: 10.1016/j.energy.2025.137760.
[11] M. S. Amin Megat Ali *et al.*, "Short-term Gini coefficient estimation using nonlinear autoregressive multilayer perceptron model," *Heliyon*, vol. 10, no. 4, p. e26438, 2024, doi: 10.1016/j.heliyon.2024.e26438.
[12] S. Rutten *et al.*, "Penalized distributed lag non-linear models for small area data using Laplacian-P-splines," *Stat. Comput.*, vol. 36, no. 38, 2025, doi: 10.1007/s11222-025-10790-9.
[13] Ajriansyah, B. W. Otok, and Sutikno, "Integrating Multivariate Adaptive Regression Splines into Small Area Estimation for Nonlinear Poverty Modeling in Java Island," *arXiv Prepr.*, pp. 1–20, 2026, doi: https://doi.org/10.48550/arXiv.2609.21987.
[14] BPS RI, "Governance and Framework for Small Area Estimation," 2025. [Online]. Available: https://www.bps.go.id/en/publication/2025/12/05/ae0cd2085208074ac09c65eb/governance-and-framework-for-small-area-estimation.html
[15] J. N. K. Rao and I. Molina, *Small Area Estimation*. John Wiley & Sons, 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118735855
[16] D. Gartina and L. Khikmah, "Pendugaan Kemiskinan Menggunakan Small Area Estimation Dengan Pendekatan Emperical Best Linier Unbiased Prediction (EBLUP)," *Statistika*, vol. 8, no. 2, 2020, doi: 10.26714/jsunimus.8.2.2020.159-165.
[17] Priatmadani, P. P. Sari, E. N. Rahmat, P. B. Aji, F. A. Nafiis, and N. Istiana, "Small Area Estimation of Maluku and Papua Island Child Poverty Levels in 2023," *J. Mat. Stat. dan Komputasi*, vol. 21, no. 1, pp. 46–61, 2024, doi: 10.20956/j.v21i1.35293.
[18] D. Dedianto and I. Y. Wulansari, "Aplikasi Small Area Estimation (SAE) Metode Pseudo-EBLUP Pada Official Statistics Di Indonesia Studi Kasus : Estimasi Pengeluaran Rumah Tangga di Provinsi Jawa Timur Tahun 2016," *J. Apl. Stat. Komputasi Stat.*, vol. 10, no. 2, 2018, doi: 10.34123/jurnalasks.v10i2.89.
[19] S. E. Hosseini, D. Shahsavani, M. R. Rabiei, and M. Arashi, "Small area estimation with partially linear mixed-t model with measurement error," *J. Comput. Appl. Math.*, vol. 446, no. September 2023, p. 115871, 2024, doi: 10.1016/j.cam.2024.115871.
[20] J. H. Friedman, "Multivariate Adaptive Regression Splines," *Ann. Stat.*, vol. 19, no. 1, pp. 1–141, 1991, doi: 10.1214/aos/1176347963.
[21] F. López and K. Kholodilin, "Putting MARS into space. Non-linearities and spatial effects in hedonic models," *Reg. Sci. Assoc. Int.*, pp. 871–896, 2023, doi: 10.1111/pirs.12738.
[22] A. Nayeem, H. Bonakdari, S. Khalloufi, and M. Aider, "Application of multivariate adaptive regression splines (MARS) to study the colorization occurring in the process of lactulose production following lactose electro-activation," *Int. Dairy J.*, vol. 168, p. 106291, 2025, doi: 10.1016/j.idairyj.2025.106291.
[23] M. E. Seno, H. A. Zeini, H. Imran, M. Noori, S. N. Henedy, and N. M. Ghazaly, "Advancing in creep index of soil prediction: A groundbreaking machine learning approach with Multivariate Adaptive Regression Splines," *Results Mater.*, vol. 24, p. 100641, 2024, doi: 10.1016/j.rinma.2024.100641.

[24] M. Bekar Adiguzel and M. A. Cengiz, “Model selection in multivariate adaptive regressions splines (MARS) using alternative information criteria,” *Heliyon*, vol. 9, no. 9, p. e19964, 2023, doi: 10.1016/j.heliyon.2023.e19964.