STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **16**, October 2026, pp 3030–3049.
Published online in International Academic Press (www.IAPress.org)

A New Nonparametric Regression Approach Using a Three-Mixed Estimator (T-MENR) with Application in Public Health Data: A Case Study of Prevalence of Hypertension

Andrea Tri Rian Dani [1], Meylita Sari [2∗], Syaiful Anam [2], Hilmi Aziz Bukhori [2], Ahmad Hakiim Jamaluddin [3], Yossy Candra [4], Nur Chamidah [5], I Nyoman Budiantara [6], Vita Ratnasari [7]

[1] *Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Mulawarman, Samarinda 75119, Indonesia*
[2] *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Malang 65145, Indonesia*
[3] *Department of Mathematics & Statistics, Universiti Putra Malaysia, Selangor 43400, Malaysia*
[4] *Center for Religious Harmony, Secretariat General, Ministry of Religious Affairs, Jakarta 10710, Indonesia*
[5] *Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya 60115, Indonesia*
[6] *Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia*

Abstract This study proposes a novel Three-Mixed Estimator Nonparametric Regression (T-MENR) model to address the limitations of conventional single and two-mixed estimator approaches in capturing complex nonlinear relationships. The proposed framework integrates truncated spline, Gaussian kernel, and Fourier series estimators within a unified additive structure, allowing each predictor variable to be modeled according to its underlying pattern. The optimal model configuration is determined using the Generalized Cross-Validation (GCV) criterion by evaluating 54 possible estimator combinations. The empirical application to hypertension prevalence data in Indonesia shows that the best model (C27) achieves strong performance, with a minimum GCV value of 0.1972, RMSE of 0.4902, and coefficient of determination (R^2) of approximately 0.9536. These results indicate that the T-MENR model is highly effective in explaining variability in hypertension prevalence. Compared to single nonparametric models, the proposed approach consistently produces lower prediction error and higher explanatory power, demonstrating the advantage of combining multiple estimators within a single framework. The results also reveal that different health behavioral factors follow distinct relationship patterns, which are better represented when modeled using appropriate estimators. Despite its strong performance, the model requires substantial computational effort due to the large number of parameter combinations. Future research may focus on incorporating statistical inference, such as hypothesis testing, to further improve model interpretability.

Keywords Nonparametric Regression, Truncated Spline, Gaussian Kernel, Fourier Series, Three-Mixed Estimators, Health Modeling

AMS 2010 subject classifications: 62G08, 62J05, 62P10

DOI: 10.19139/soic-2310-5070-3951

1. Introduction

In the process of modeling for one or more variables, then of course the first thing that needs to be done is to explore the pattern of the relationship. Then the variables are divided into independent variables and dependent variables. The most used analysis method is regression analysis [1]. Regression is one of technique in statistics used to model the relationships from independent and dependent variables [2]. Identification used to check the

∗Correspondence to: Meylita Sari (meylitasari@ub.ac.id). Department of Mathematics, Faculty of Science, Technology, and Mathematics, Universitas Brawijaya, Malang 65145, Indonesia

ISSN 2310-5070 (online) ISSN 2311-004X (print)

pattern of relationship from dependent and independent variables can be accomplished by make a scatter plot [3]. If the shape of the functional relationship pattern is clearly, used is a parametric approach. But, when the shape of the pattern of functional relationships is not clearly known, the approach that can be used is nonparametric [4], [5]. The limited information, the form of the function, and the unclear pattern of the relationship are important considerations, recommended to do modeling with a nonparametric regression approach.

In the real cases, of course, a parametric approach regression model is not suitable for use in this state, as the shape of the functional relationship pattern from dependent and independent variables is often unknown. With the advancement of computing era today, the nonparametric regression approach, which usually has a complex computational complexity, is becoming simpler and more widely used [6]. Nonparametric regression approaches are known to be very well used for unknown data patterns because they do not depend on the assumptions of a specific regression curve, thus providing a high degree of flexibility when the data is expected to adapt to its intended shape [7], [8]. A regression curve without the influence of the researcher's subjectivity [9], [10]. An important point when performing regression analysis is to find a suitable form for estimating the regression curve [10].

There are several estimators in the nonparametric regression that have been developed by researcher's, namely Spline by [11], [12], [13], [14], [15], [16]. Kernel by [11], [17], [18], [19], [20]. The Fourier series by [21], [22], [23], [24], [25]. Currently, nonparametric regression models are limited to only one form of estimation for all independent variables used. In the real case, each independent variables may have a different pattern, so that [17] develop a mixed estimator model where the independent variables will be approached by different estimator forms according to the characteristics of the relationship pattern [26], [27], [28], [29].

The definition of the mixed estimator in the nonparametric regression is a multi-estimator model whose regression curve is assumed to be additive, and the curve regression will be approached by two or more types of estimators [17]. The form of mixed estimator that is currently developing is limited to only using two estimator forms, including the combination of the Spline-Kernel by [11], [17], [27], [29], [30]. Spline-Fourier Series by [31], [32], [33]. Kernel-Fourier Series by [7], [34], [35], and etc.

The limitation when using a nonparametric two mixed estimator is that the independent variable shows three different pattern characteristics. For example, if there are three independent variables $(x_1 - x_3)$, where each pattern of the relationship between the independent variables and the dependent variables is different. Based on the framework of thinking and problems that occur in the real cases, it is necessary to study the Three-Mixed Estimator Model.

Now, we will develop a Three-Mixed Estimator Nonparametric Regression (T-MENR) model. The T-MENR model, which in this study is mixed of Truncated Spline, Fourier Series, and Gaussian Kernel. In the Truncated Spline estimator, determining the number of the optimal knot points is necessary. In the Kernel estimator, which must determine the right bandwidth, while in the Fourier series it is necessary to determine the optimal number of oscillations. It becomes an interesting problem and needs to be studied when using a T-MENR model. Where, in this study, it will combine the Truncated Spline, Fourier Series, and Gaussian Kernel, so must be determine the point of knots, bandwidth, and the number of oscillations, hereinafter referred to as smoothing parameters, simultaneously and optimally. To select the optimal smoothing parameter in the T-MENR model, one can use the Generalized Cross-Validation (GCV) value [36].

Hypertension prevalence has become a major public health concern, particularly in developing countries, including Indonesia [37]. According to Badan Pusat Statistik (BPS), health indicators are increasingly recognized as essential components of societal welfare alongside economic measures. Hypertension is widely referred to as a silent killer because it often develops without noticeable symptoms yet significantly increases the risk of severe complications such as cardiovascular disease, stroke, and kidney failure.

At the global level, hypertension is closely linked to the Sustainable Development Goals (SDGs), particularly Goal 3 (Good Health and Well-being), which emphasizes reducing premature mortality from non-communicable diseases. In the modern era, characterized by rapid urbanization, lifestyle changes, and demographic transitions, the prevalence of hypertension continues to rise [38]. This condition is influenced by a complex interaction of multiple factors, including the prevalence of heart disease, diabetes mellitus, smoking behavior, physical inactivity, dietary habits, and access to health services [39], [40], [41]. The multidimensional nature of these determinants implies that modeling hypertension prevalence requires flexible statistical approaches capable of capturing nonlinear relationships and mixed patterns among predictors. Previous studies have commonly employed parametric or single nonparametric regression approaches. However, such models often face limitations when dealing with heterogeneous data structures, where some variables exhibit linear trends while others follow nonlinear or oscillatory patterns.

To address these limitations, this study proposes the Three-Mixed Estimator Nonparametric Regression (T-MENR) model, which integrates parametric, truncated spline, and Fourier components within a unified framework. This hybrid approach allows the model to simultaneously accommodate linear effects, local nonlinearities, and periodic or fluctuating patterns in the data. The novelty of this study lies in the development and application of the T-MENR model to analyze hypertension prevalence by incorporating multiple health-related risk factors within a single flexible modeling framework. Unlike conventional approaches, the proposed method can adapt to diverse data characteristics, thereby improving estimation accuracy and interpretability. The results are expected to provide a more comprehensive analytical basis for designing targeted, evidence-based, and sustainable health policies, in line with the achievement of SDGs Goal 3.

2. Materials and Methods

2.1. *Truncated Spline*

The truncated spline model on paired data is a flexible nonparametric regression approach that accommodates potential nonlinear relationships between predictor and response variables by incorporating piecewise polynomial functions with specified knot points [42]. The Truncated Spline model on paired data $(x_{1i}, x_{2i}, ..., x_{Pi}, y_i)$, where it is assumed that the relationship between the dependent and independent variables as follows:

$$y_i = f(x_{1i}, x_{2i}, ..., x_{Pi}) + \epsilon_i \tag{1}$$

The regression curve $f(x_{1i}, x_{2i}, ..., x_{Pi})$ considered to follow the additive model:

$$f(x_{1i}, x_{2i}, ..., x_{Pi}) = \sum_{p=1}^{P} f_p(x_{pi}) \tag{2}$$

The components of the Truncated Spline for each independent variable can typically be written as:

$$\hat{f}_p(x_p) = \hat{\gamma}_{0p} + \sum_{j=1}^{m}\sum_{p=1}^{P} \hat{\gamma}_{jp} x_p^j + \sum_{k=1}^{r}\sum_{p=1}^{P} \hat{\xi}_{(m+k)p}(x_p - K_{kp})_+^m \tag{3}$$

Can be summarized in Equation (4).

$$\begin{aligned}\hat{f}(x) &= X_0\hat{\gamma} + X_1(K_1)\hat{\xi}_1 + \cdots + X_p(K_p)\hat{\xi}_p \\ &= X(\psi)\hat{\beta}\end{aligned} \tag{4}$$

with:
$X(\psi) = [\, X_0 \;\; X_1(K_1) \;\; \cdots \;\; X_p(K_p) \,]$

$$\hat{\beta} = \begin{bmatrix} \hat{\gamma} \\ \hat{\xi}_1 \\ \vdots \\ \hat{\xi}_p \end{bmatrix}$$

The vector $f(x)$ has dimension $(n1)$, the vector γ has dimension $((p+1)1)$, and the vectors $\xi_1, \xi_2, \ldots, \xi_p$ each have dimension $(r1)$. The matrices $X_1(K_1), X_2(K_2), \ldots, X_p(K_p)$ have dimension (nr), while the matrix X_0 has dimension $(n(p+1))$.

2.2. *Gaussian Kernel*

The nonparametric regression model with the Gaussian Kernel estimator as follows:

$$y_i = h(w_i) + \epsilon_i \tag{5}$$

Where w_i is independent variable, whose pattern is assumed to follow the characteristics of the Kernel and y_i is dependent variable [43], [44]. The regression curve of $h(w_i)$ is a curve with unknown pattern, which can be approached by the Gaussian kernel estimator.

$$\hat{h}_\lambda(w) = n^{-1}\left[\sum_{i=1}^{n} G_{\lambda i}(w)\right] y_i \tag{6}$$

with:

$G_{\lambda i}(w) = \frac{B_\lambda(w-w_i)}{n^{-1}\sum_{i=1}^{n} B_\lambda(w-w_i)}$

$B_\lambda(w-w_i) = \frac{1}{\lambda} B\left(\frac{w-w_i}{\lambda}\right)$

Where the Kernel function in Equation (7).

$$B(w) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}(-w^2)\right); I_{[-\infty,\infty]}(w) \tag{7}$$

Based on Equation (6), we can write the matrix form in Equation (8):

$$\hat{h}(w) = G(\lambda)y \tag{8}$$

where:

$$G(\lambda) = \begin{bmatrix} n^{-1}G_{\lambda 1}(w_1) & n^{-1}G_{\lambda 2}(w_1) & \cdots & n^{-1}G_{\lambda n}(w_1) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1}G_{\lambda 1}(w_n) & n^{-1}G_{\lambda 2}(w_n) & \cdots & n^{-1}G_{\lambda n}(w_n) \end{bmatrix}; y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

2.3. *Fourier Series*

The Fourier estimator in nonparametric regression model as follows:

$$y_i = v(z_i) + \epsilon_i \tag{9}$$

Where z_i is independent variable, whose pattern is assumed to follow the characteristics of the Fourier series. The regression curve of $v(z_i)$ is a curve with sine or cosine waves [45], [46].

$$\hat{v}(z_i) = \hat{\theta} z_i + \frac{1}{2}\hat{\alpha}_0 + \sum_{q}^{Q} \hat{\alpha}_q \cos q(z_i) \tag{10}$$

Based on Equation (10), we can write the matrix as:

$$\hat{v}(z) = Z(\Omega)\hat{\delta} \tag{11}$$

with:

$$Z(\Omega) = \begin{bmatrix} z_1 & \frac{1}{2} & \cos z_1 & \cos 2z_1 & \cdots & \cos Qz_1 \\ z_2 & \frac{1}{2} & \cos z_2 & \cos 2z_2 & \cdots & \cos Qz_2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ z_n & \frac{1}{2} & \cos z_n & \cos 2z_n & \cdots & \cos Qz_n \end{bmatrix}; \hat{\delta} = \begin{bmatrix} \theta \\ \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_Q \end{bmatrix}$$

Based on Equation (10), $Z(\Omega)$ has dimension $n(q+2)$, $\hat{\delta}$ has dimension $(q+2)1$, and the vector $\hat{v}(z)$ has dimension $(n1)$.

3. Result and Discussion

In this section, we will conduct a study and explain the form of the Three-Mixed Estimator Nonparametric Regression (T-MENR) model and its data application.

3.1. *Three-Mixed Estimator Nonparametric Regression (T-MENR)*

Suppose that paired data are given, with one independent variable modeled using a truncated spline estimator and one independent variable for each of the other components. The model can be written as follows:

$$y_i = f(x_i) + h(w_i) + v(z_i) + \epsilon_i \tag{12}$$

The Three-Mixed Estimator Nonparametric Regression (T-MENR) model with Truncated Spline, Gaussian Kernel, and Fourier series can be written:

$$y_i = \eta(x_i, w_i, z_i) + \epsilon_i \tag{13}$$

with $\eta(x_i, w_i, z_i)$ is $f(x_i) + h(w_i) + v(z_i)$.

$\epsilon = (\epsilon_1, \epsilon_2, ..., \epsilon_n)^T$ denotes the random error vector, assumed to satisfy $E(\epsilon) = \mathbf{0}, Var(\epsilon) = \sigma^2 \boldsymbol{I}_n$ where $\boldsymbol{I}$ is the $n \times n$ identity matrix. Thus, the error terms are assumed to be independent and identically distributed with zero mean and constant variance.

The T-MENR model in Equation (13) can be written as:

$$y = X(\psi)\beta + G(\lambda)y + Z(\Omega)\delta + \epsilon \tag{14}$$

We can write the error as follows:

$$\epsilon = [I - G(\lambda)]y - X(\psi)\beta - Z(\Omega)\delta \tag{15}$$

Then it can be summarized $[I - G(\lambda)]y$ to y^*.

$$\epsilon = y^* - X(\psi)\beta - Z(\Omega)\delta \tag{16}$$

Furthermore, the Sum of Squares Error can be written as follows:

$$\begin{aligned} A(\beta, \delta) &= \epsilon^T \epsilon \\ &= (y^* - X(\psi)\beta - Z(\Omega)\delta)^T (y^* - X(\psi)\beta - Z(\Omega)\delta) \end{aligned} \tag{17}$$

Equation (17) can be simplified to:

$$\begin{aligned} A(\beta, \delta) =& y^* y^* - 2\beta^T X(\psi)^T y^* - 2\delta^T Z(\Omega)^T y^* + 2\beta^T X(\psi)^T Z(\Omega)\delta + \beta^T X(\psi)^T X(\psi)\beta + \\ & \delta^T Z(\Omega)^T Z(\Omega)\delta \end{aligned} \tag{18}$$

To get an estimate of β and δ, then it can be done with a partial derivative of Equation (18) on β and δ.

$$\frac{\partial A(\beta, \delta)}{\partial \beta} = -2X(\psi)^T y^* + 2X(\psi)^T Z(\Omega)\delta + 2X(\psi)^T X(\psi)\beta \tag{19}$$

$$\frac{\partial A(\beta, \delta)}{\partial \beta} = -2Z(\Omega)^T y^* + 2Z(\Omega)^T X(\psi)\beta + 2Z(\Omega)^T Z(\Omega)\delta \tag{20}$$

Equation (19) and (20) is equal to zero, so we will get:

$$\hat{\beta} = \left[X(\psi)^T X(\psi)\right]^{-1} \left(X(\psi)^T y^* - X(\psi)^T Z(\Omega)\hat{\delta}\right) \tag{21}$$

$$\hat{\delta} = \left[Z(\Omega)^T Z(\Omega)\right]^{-1} \left(Z(\Omega)^T y^* - Z(\Omega)^T X(\psi)\hat{\beta}\right) \tag{22}$$

Based on Equations (21) and (22), it is known that the estimation results of $\hat{\beta}$ and $\hat{\delta}$ still contain parameters. To solve it, can use the Substitution Method in the process. Equations (21) and (22) can be summarized as:

$$\hat{\beta} = Cy^* - CZ(\Omega)\hat{\delta} \tag{23}$$

and

$$\hat{\beta} = Cy^* - CZ(\Omega)\hat{\delta} \tag{24}$$

where:
$C = \left[X(\psi)^T X(\psi)\right]^{-1} X(\psi)^T$
$D = \left[Z(\Omega)^T Z(\Omega)\right]^{-1} Z(\Omega)^T$

The initial step is substitute Equation (23) to Equation (24), where:

$$\begin{aligned} \hat{\delta} &= Dy^* - DX(\psi)\hat{\beta} \\ &= Dy^* - DX(\psi)(Cy^* - CZ(\Omega)\hat{\delta}) \end{aligned} \tag{25}$$

So that:

$$\hat{\delta} = [I - DX(\psi)CZ(\Omega)]^{-1} [D - DX(\psi)C]\, y^* \tag{26}$$

with $y^* = [I - G(\lambda)]\, y$.

The next step is to substitute Equation (26) to Equation (23), where:

$$\hat{\beta} = \left(C - CZ(\Omega)\left([I - DX(\psi)CZ(\Omega)]^{-1} [D - DX(\psi)C]\right)\right) [I - G(\lambda)]\, y \tag{27}$$

Equation (27) and (30) can be summarized as:

$$\hat{\delta} = Ly \tag{28}$$

$$\hat{\beta} = Uy \tag{29}$$

with:

$L = [I - DX(\psi)CZ(\Omega)]^{-1} [D - DX(\psi)C] [I - G(\lambda)]$

$U = C - CZ(\Omega)\left([I - DX(\psi)CZ(\Omega)]^{-1} [D - DX(\psi)C]\right) [I - G(\lambda)]$

Based on the results of parameter estimation from $\hat{\beta}$ and $\hat{\delta}$ in Equation (31) and (28), and the form of the Gaussian Kernel estimator in Equation 8. The T-MENR model will be obtained as follows:

$$\begin{aligned} \hat{\eta}_{\psi,\lambda,\Omega}(x, w, z) &= \hat{f}(x) + \hat{h}(w) + \hat{v}(z) \\ &= X(\psi)\hat{\beta} + G(\lambda)y + Z(\Omega)\hat{\delta} \\ &= [X(\psi)U + G(\lambda) + Z(\Omega)L]\, y \end{aligned} \tag{30}$$

Equation (30) can be written briefly as:

$$\hat{\eta}_{\psi,\lambda,\Omega}(x,w,z) = R(\psi,\lambda,\Omega)\, y \tag{31}$$

Vector $\hat{\eta}_{\psi,\lambda,\Omega}(x,w,z)$ is very and highly dependent on the number of knot points, the location where the knots points, the bandwidth value, and the number of oscillations which will then be referred to as smoothing parameters.

To find the best T-MENR model, you need to choose the best optimal smoothing parameters with Generalized Cross-Validation (GCV) method. In this case, the formula of the GCV method will be modified according to the T-MENR model, which can then be written in Equation (32).

$$GCV = \left[\frac{MSE(\psi,\lambda,\Omega)}{\left(n^{-1}\, trace\,[I - R(\psi,\lambda,\Omega)]\right)^2}\right]^2 \tag{32}$$

where:

$MSE(\psi,\lambda,\Omega)$: $\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}$

I : Identity matrix

3.2. Estimation Algorithm of the Three-Mixed Estimator Nonparametric Regression (T-MENR)

The estimation of the proposed T-MENR model combines three nonparametric smoothing techniques, namely truncated spline, Gaussian kernel, and Fourier series. Since the model involves several smoothing parameters, including knot locations, bandwidths, and Fourier oscillations, the estimation is performed through an iterative optimization procedure. The optimal combination of smoothing parameters is selected by minimizing the Generalized Cross Validation (GCV) criterion, which provides a balance between model goodness-of-fit and complexity. The overall estimation procedure is summarized in Algorithm 1.

Algorithm 1. T-MENR Estimation

Input: y, X_S, X_K, X_F
Generate candidate knots.
Generate candidate bandwidths.
Generate Fourier oscillation candidates.
for each knot **do**
  for each bandwidth **do**
    Construct spline matrix $X(\psi)$.
    Compute Gaussian kernel matrix $G(\lambda)$.
    Construct Fourier matrix $Z(\Omega)$.
    Estimate Fourier coefficients $\hat{\delta}$.
    Estimate spline coefficients $\hat{\beta}$.
    Compute $\hat{y}$.
    Evaluate RMSE, R^2, and GCV.
  end for
end for
Select parameters with minimum GCV.
Refit T-MENR using optimal parameters.
Compute final prediction and model accuracy.
Output: $\hat{\delta}$, $\hat{\beta}$, optimal knot, bandwidth, oscillation, $\hat{y}$

The proposed T-MENR algorithm was implemented in RStudio (Version 2025.09.1+401) using custom R scripts. The estimation procedure iteratively evaluates all candidate combinations of spline knots, Gaussian kernel bandwidths, and Fourier oscillations based on the GCV criterion to determine the optimal model. All computations were performed using **R version 4.5.x** in **RStudio** (Version 2025.09.1+401) on a MacBook Air M3 equipped with an Apple M3 processor and 16 GB of unified memory.

3.3. Data Application

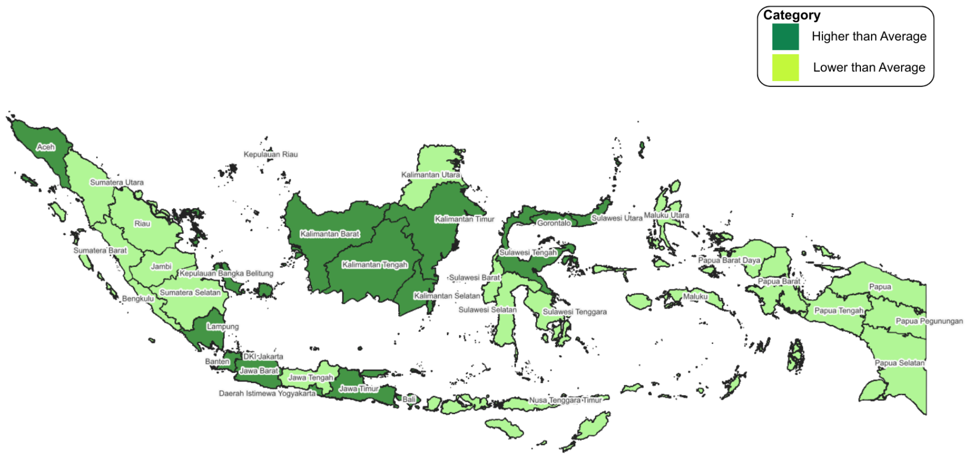
This study employs secondary data derived from the 2023 Indonesian Health Survey (Survei Kesehatan Indonesia (SKI)) conducted by the Ministry of Health of the Republic of Indonesia. The unit of analysis comprises 38 provinces, enabling the data to reflect regional-level health conditions across the country. A purposive sampling strategy is applied, considering factors such as data availability, data recency, and alignment with the research objectives. The SKI 2023 dataset is considered suitable due to its comprehensive coverage and up-to-date information on health indicators, particularly those related to disease prevalence and its associated determinants. All variables in this study are measured on a ratio scale. Detailed descriptions of variable notation, names, types, and operational definitions are provided in Table 1.

Table 1. Description of Research Variables

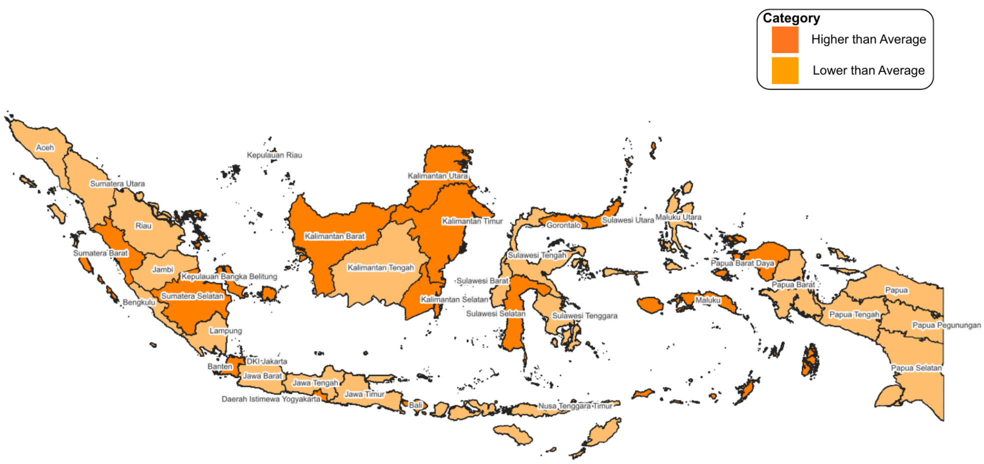
Variable	Description	Operational Definition
Response Variable	Prevalence of Hypertension (y)	Calculated as the proportion of individuals with hypertension relative to the total population in a given region and year.
Predictor Variable	Proportion of Hypertension Follow-up/ Check-ups at Health Facilities (x_1)	Calculated as the proportion of hypertension patients who routinely attend follow-up visits compared to the total number of hypertension patients.
Predictor Variable	Proportion of Insufficient Physical Activity among Population Aged $\geq$ 10 Years (x_2)	Defined as the proportion of individuals aged $\geq$ 10 years with low levels of physical activity within a region.
Predictor Variable	Proportion of Habitual Consumption of Fatty/ Cholesterol/ Fried Foods $\geq$ 1 Time (x_3)	Calculated as the proportion of the population who consume fatty, high-cholesterol, or fried foods at least once within the observation period.

1) Exploratory Data Analysis

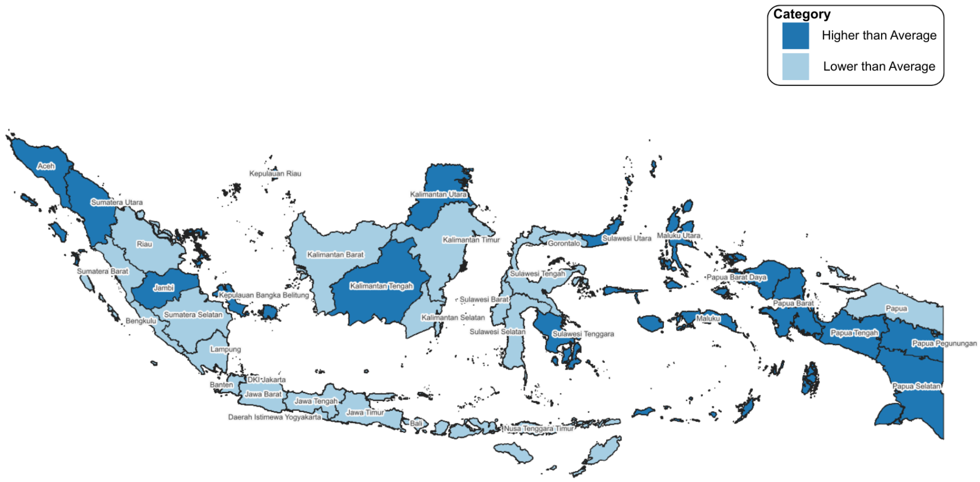
In this study, an exploratory analysis was conducted prior to model estimation using spatial mapping for all variables by classifying their values into above- and below-average categories. The mean value for each variable was calculated and used as the threshold for classification. This visualization aims to provide an initial overview of the spatial distribution patterns and variability of each variable across regions. The results of the spatial mapping are presented in Figure 1.



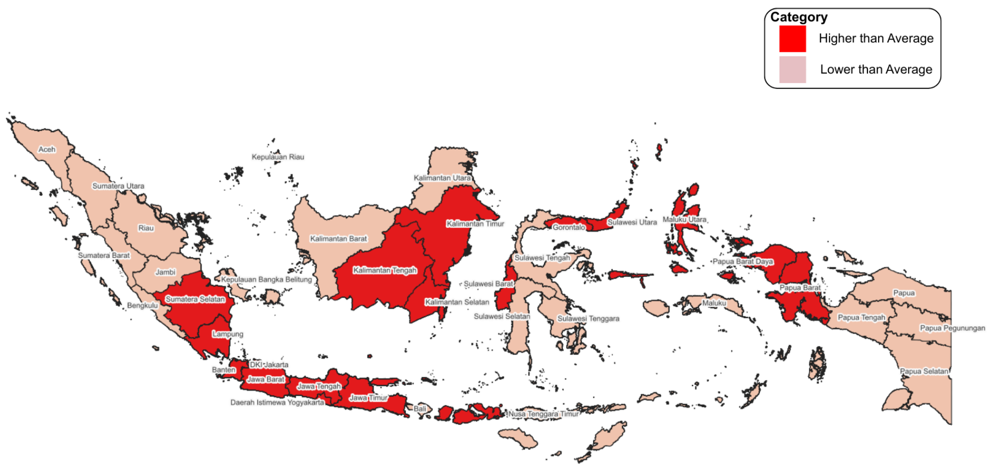
(a) Spatial Mapping of Prevalence of Hypertension



(b) Spatial Mapping of Proportion of Hypertension Follow-up/ Check-ups at Health Facilities



(c) Spatial Mapping of Proportion of Insufficient Physical Activity among Population Aged $\geq$ 10 Years



(d) Spatial Mapping of Proportion of Habitual Consumption of Fatty/ Cholesterol/ Fried Foods $\geq$ 1 Time

Figure 1. Spatial Mapping of Each Variable

Based on Figure 1a, the spatial mapping of prevalence of hypertension shows a clear regional clustering pattern. Areas with prevalence above the average are concentrated in Jawa, Bali, and Kalimantan, as well as parts of Sulawesi and Sumatra. In contrast, eastern Indonesia, particularly Nusa Tenggara, Maluku, and Papua, tends to fall below the average, indicating relatively better hypertension conditions compared to the western and central regions. This pattern highlights a noticeable disparity in hypertension distribution across regions.

Based on Figure 1b, the spatial mapping of proportion of hypertension follow-up or check-ups at health facilities shows considerable variation across regions. Areas with values above the average are widely observed in Jawa and Bali, almost all of Kalimantan, as well as parts of Sulawesi and island regions. In contrast, most areas in eastern Indonesia remain below the average. When related to Figure 1a, this suggests that regions in the east with relatively lower hypertension prevalence are also characterized by lower utilization of follow-up health services, indicating differences between disease occurrence and service utilization across regions.

Based on Figure 1c, the spatial mapping of proportion of insufficient physical activity among population aged $\geq$ 10 years exhibits a more dispersed pattern. Values above the average are observed not only in metropolitan areas such as Jakarta but also in several regions of Sumatra, Kalimantan, Sulawesi, and eastern Indonesia, including Maluku and Papua. Meanwhile, areas below the average, indicating relatively better physical activity levels, are generally found across most parts of Jawa (excluding Jakarta), Bali, Nusa Tenggara, and parts of Sulawesi and Kalimantan. This pattern indicates that the distribution of physical activity is relatively widespread rather than concentrated in specific regions.

Based on Figure 1d, the spatial mapping of proportion of habitual consumption of fatty, cholesterol-rich, or fried foods ≥ 1 time also shows a varied spatial pattern. Areas with values above the average are concentrated in Jawa, parts of Kalimantan, and southern Sumatra, as well as several regions such as Sulawesi and North Maluku. In contrast, areas below the average are commonly found in northern Sumatra, Bali, Nusa Tenggara Timur, and parts of Papua. This pattern reflects regional variation in dietary habits across Indonesia.

2) Linearity Assessment Using the Teräsvirta Test

The preliminary exploration indicates that the data exhibit considerable variability across observations, suggesting the potential presence of complex relationships that may not be adequately captured by a purely linear model. Therefore, a nonlinearity test using the Teräsvirta test was conducted to evaluate the suitability of both linear and nonlinear modeling approaches. This step is essential for determining the most appropriate model specification. The results of the Teräsvirta test are presented in Table 2.

Table 2. Test for Nonlinearity

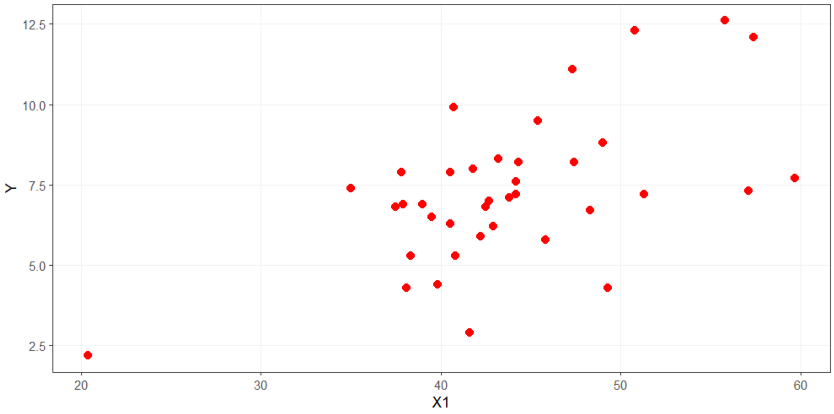
	Teräsvirta test
p-value	0.0079

Based on the results of the Teräsvirta test presented in Table 2, the *p-value* of 0.0079 is lower than the significance level of 0.05, leading to the rejection of H_0. In this test, H_0 assumes that the relationship between the independent and dependent variables is linear. The rejection of H_0 indicates the presence of a nonlinear relationship in the data, implying that the linearity assumption is not satisfied. Therefore, parametric approaches such as linear regression may be less appropriate, and nonlinear modeling approaches are better suited to capture the underlying relationships among variables, particularly in health-related data that tend to be complex.

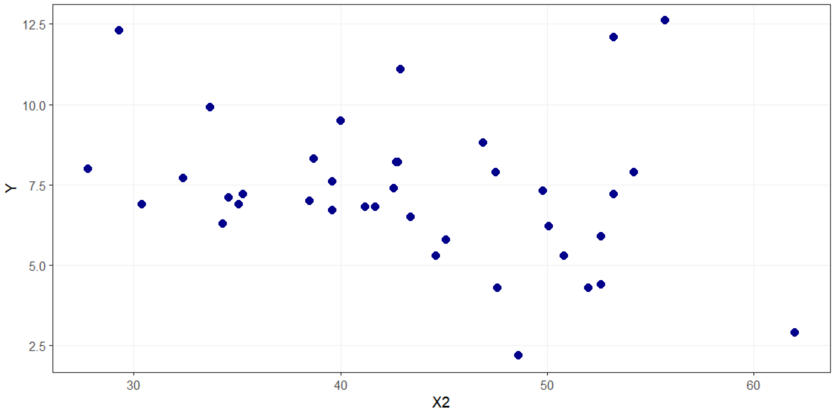
3) Exploring Relationship Patterns Using Scatter Plots

To reinforce the results of the previous test, a visual analysis was conducted using scatter plots to examine the relationship between the response variable, namely hypertension prevalence, and each predictor variable. A scatter plot is a graphical tool that displays the relationship between two variables by plotting paired observations, allowing patterns, trends, and potential nonlinear relationships to be visually identified. This approach is particularly useful for detecting the form and direction of relationships before model specification. The visualization results are presented in Figure 2.

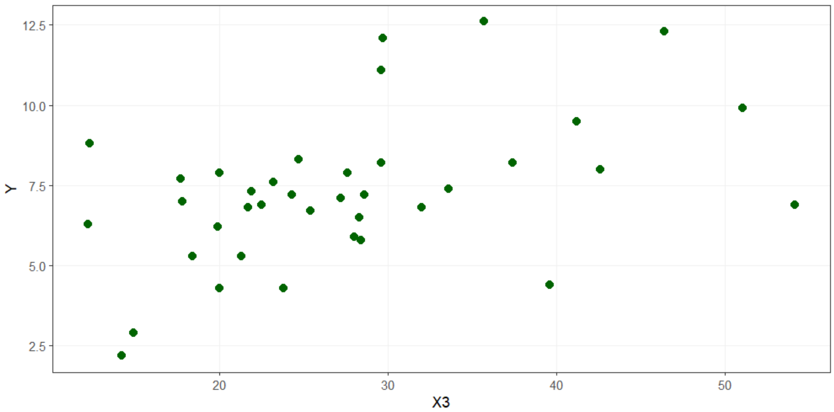
Based on the scatter plot visualization in Figure 2, there is clear heterogeneity in the relationship between hypertension prevalence and the three predictor variables, indicating that each relationship follows a different functional pattern. Figure 2a shows local changes in the slope across different ranges of the data, suggesting that the effect of the predictor varies over intervals and is therefore more appropriately modeled using a truncated spline approach with knot points. Figure 2b exhibits a fluctuating or oscillatory pattern characterized by repeated increases and decreases, reflecting a cyclical structure that is suitably captured using a Fourier series representation. Meanwhile, Figure 2c displays an irregular and highly dispersed pattern without a clear global trend, indicating a complex local structure that is better accommodated by a Gaussian kernel estimator due to its flexibility in capturing smooth variations in the data.



(a) Scatter Plot of Prevalence of Hypertension versus Proportion of Hypertension Follow-up/ Check-ups at Health Facilities



(b) Scatter Plot of Prevalence of Hypertension versus Proportion of Insufficient Physical Activity among Population Aged ≥ 10 Years



(c) Scatter Plot of Prevalence of Hypertension versus Proportion of Habitual Consumption of Fatty/ Cholesterol/ Fried Foods ≥ 1 Time

Figure 2. Scatter Plots of Prevalence of Hypertension and Predictor Variables

However, determining the type of estimator solely based on the visual interpretation of scatter plots may introduce subjectivity in the model selection process. Therefore, this study does not rely exclusively on visual inspection in specifying the estimator for each predictor variable. Instead, a more comprehensive evaluation is conducted by considering all possible combinations of estimator structures, namely Truncated Spline, Fourier Series, and Gaussian Kernel for each predictor variable. This approach is expected to yield a more objective and optimal model structure based on the evaluation criteria used, thereby improving the accuracy in capturing the relationships among variables. The combination schemes of estimator structures for each predictor variable are presented in Table 3.

Table 3. Combination Schemes of Truncated Spline, Fourier Series, and Gaussian Kernel Estimator Structures for Each Predictor Variable

Scheme	Estimator Structure			Combination Labels
	x_1	x_2	x_3	
1	Truncated Spline	Gaussian Kernel	Fourier Series	C1 – C9
2	Truncated Spline	Fourier Series	Gaussian Kernel	C10 – C18
3	Gaussian Kernel	Truncated Spline	Fourier Series	C19 – C27
4	Gaussian Kernel	Fourier Series	Truncated Spline	C28 – C36
5	Fourier Series	Truncated Spline	Gaussian Kernel	C37 – C45
6	Fourier Series	Gaussian Kernel	Truncated Spline	Truncated Spline

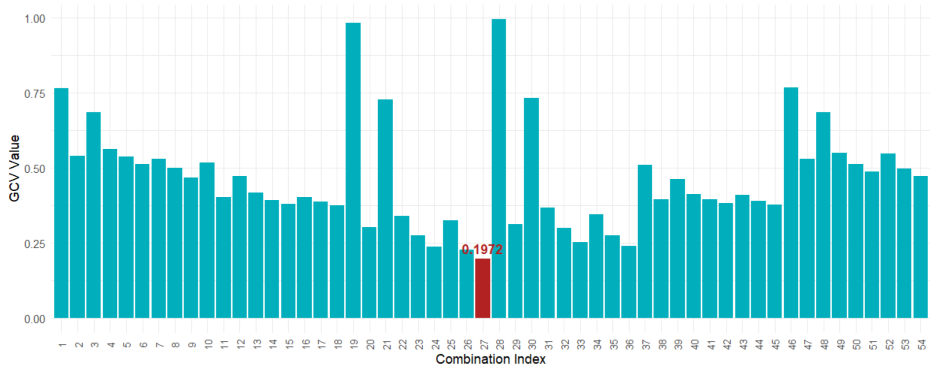
Based on Table 3, each estimator structure yields nine parameter combinations derived from variations in the number of knots for the Truncated Spline (1 to 3 knots) and the number of oscillations for the Fourier series (1 to 3 oscillations), while a single optimal bandwidth is used for the Gaussian kernel to ensure the best modeling performance. Based on this scheme, a total of 54 model combinations are evaluated in this study, which are sequentially labeled from C1 to C54 to maintain transparency in the model evaluation process.

All model combinations are then systematically assessed using the Generalized Cross Validation (GCV) criterion. This criterion is employed to evaluate the balance between model goodness of fit and model complexity. In this way, the selection of the best model can be carried out objectively, ensuring that the resulting model not only achieves high accuracy but also avoids overfitting. The comparison of GCV values for all 54 model combinations is presented in Figure 3.

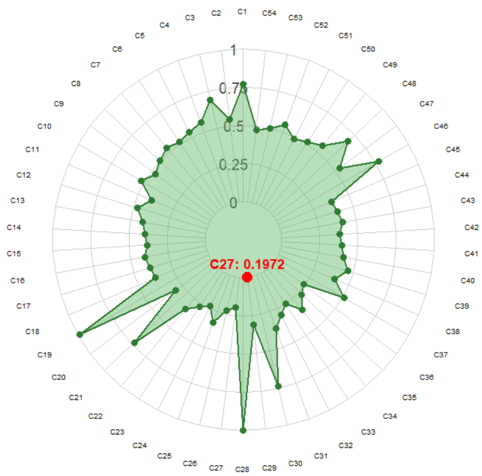
Based on Figure 3, there is a notable fluctuation in GCV values across the 54 evaluated model combinations. The bar plot in Figure 3a clearly illustrates pronounced spikes in GCV values for certain combinations, such as C19 and C28, while simultaneously highlighting combination C27 as the most prominent minimum point compared to all other bars. This observation is further supported by the radar plot in Figure 3b, where C27 is identified as the point closest to the center of the circle, providing strong visual evidence of the location of the most optimal model convergence. The C27 configuration belongs to the third estimator structure scheme with parameters of three knots and three oscillations, yielding the lowest GCV value of 0.1972.

4) Estimation of the T-MENR Model Parameters for Hypertension Prevalence

The parameter estimation results for the six model structure combinations indicate that the best-performing model is achieved under the third estimator structure scheme. In this configuration, the Follow up Care Proportion variable is modeled using a Gaussian kernel estimator, the Insufficient Physical Activity Proportion variable is approximated using a Truncated Spline estimator, and the Frequent Consumption of Fatty Foods (≥ 1) variable is modeled using a Fourier series estimator. The selection of the third estimator structure scheme as the optimal model is based on its goodness of fit criteria compared to all other configurations evaluated in the previous stage.



(a) Bar Plot of GCV Values



(b) Radar Plot of GCV Values

Figure 3. Comparison of GCV Values for All Model Combinations

The detailed goodness of fit measures for this third estimator scheme, including variations in the number of knots and oscillations, are presented in Table 4.

Table 4. Goodness of Fit Measures for the Optimal T-MENR Model Scheme

Combination	Estimator Structure (Third Scheme)			Evaluation Values				Computational Time (s)
	Optimal Bandwidth	Knots	Oscillations	GCV	RMSE	R^2	Bandwidth	
C19			1	0.9834	0.9181	83.7136	0.2551	128
C20		1	2	0.3018	0.5414	94.3360	0.1125	146
C21			3	0.7282	0.8111	87.2874	0.2031	171
C22			1	0.3394	0.5972	93.1095	0.1269	182
C23	1	2	2	0.2755	0.5471	94.2169	0.1125	196
C24			3	0.2371	0.5061	95.0511	0.1010	214
C25			1	0.3255	0.5966	93.1229	0.1269	276
C26		3	2	0.2278	0.5376	94.4156	0.1125	294
C27			**3**	**0.1972**	**0.4902**	**95.3573**	**0.1010**	**317**

Based on the results presented in Table 4, the best model combination is obtained using three knot points and three oscillations, as represented by combination C27, which yields the minimum GCV value of 0.1972. The low GCV value at C27 indicates that the model achieves an optimal balance between goodness of fit and model complexity, thereby minimizing the risk of overfitting. In addition, configuration C27 also produces the lowest RMSE value of 0.4902, reflecting the smallest prediction error, and is supported by a bandwidth value of 0.1010, indicating an appropriate level of smoothness in capturing the underlying data pattern. Furthermore, the high coefficient of determination suggests that approximately 95.3573 percent of the variation in the response variable can be simultaneously explained by the model. Therefore, it can be concluded that the T-MENR model at point C27 represents the most accurate and efficient structure among all evaluated combinations, as further specified in Equation (33).

$$\begin{aligned}\hat{y}_i =& \frac{1}{38}\sum_{i=1}^{n}\left[\frac{\frac{1}{0.1010}K\left(\frac{x_1-x_{1i}}{0.1010}\right)}{\frac{1}{38}\sum_{i=1}^{n}\frac{1}{0.1010}K\left(\frac{x_1-x_{1i}}{0.1010}\right)}\right]y_i + 8.9617 - 0.532x_2 + 0.706(x_2-30.13)_+ - \\ & 0.277(x_2-37.12)_+ + 0.095(x_2-40.24)_+ + 0.0021 + 0.2294x_3 - 0.5894\cos(x_3) - \\ & 0.2377\cos(2x_3) + 4.6681\cos(3x_3)\end{aligned} \tag{33}$$

The estimated T-MENR model can be decomposed into three main components, namely the Gaussian kernel component for variable x_1, the truncated spline component for variable x_2, and the Fourier series component for variable x_3. Each component captures a distinct pattern of relationship between the predictor variables and the response variable.

a. Gaussian Kernel Component (x_1)
The kernel component for x_1 is expressed as a weighted average of the response values, where the weights are determined by the proximity between observations. In this study, a Gaussian kernel with a specific bandwidth is applied. This component does not produce an explicit parametric form, meaning that the relationship between x_1 and the response variable is modeled flexibly. The estimated value at a given point depends primarily on nearby observations, allowing the model to capture smooth local variations in the data. Therefore, the effect of x_1 cannot be interpreted through a single global coefficient, but rather through its local influence across different regions of the data.

b. Truncated Spline Component (x_2)
The truncated spline component for variable x_2 can be expressed as follows:

$$\hat{y}_i = 8.9617 - 0.532x_2 + 0.706(x_2-30.13)_+ - 0.277(x_2-37.12)_+ + 0.095(x_2-40.24)_+$$

This form indicates that the relationship between x_2 and the response variable, namely hypertension prevalence, is piecewise linear, with structural changes occurring at three knot points, namely 30.13, 37.12, and 40.24.

- For $x_2 \leq 30.13$, all truncated components are assumed to be zero, so the model becomes:

$$\hat{y}_i = 8.9617 - 0.532x_2$$

- For $30.13 < x_2 \leq 37.12$, the slope changes to: $-0.532 + 0.706 = 0.174$
This indicates a change in the direction of the relationship to positive after passing the first knot point, suggesting the presence of a threshold at that value.
- For $37.12 < x_2 \leq 40.24$, the slope becomes: $-0.532 + 0.706 - 0.277 = -0.103$
The relationship becomes negative again, indicating a structural change at the second knot.
- For $x_2 > 40.24$, the slope becomes: $-0.532 + 0.706 - 0.277 + 0.095 = -0.008$
This value is close to zero, indicating that the effect of x_2 on hypertension prevalence becomes very small.

c. Fourier series Component (x_3)
The Fourier component for variable x_3 can be expressed as follows:

$$\hat{y}_i = 0.0021 + 0.2294x_3 - 0.5894\cos(x_3) - 0.2377\cos(2x_3) + 4.6681\cos(3x_3)$$

This component indicates that the relationship between x_3 and the response variable is fluctuating or non-monotonic. The presence of multiple cosine functions with different frequencies suggests that the effect of x_3 varies periodically. The relatively large coefficient associated with $\cos(3x_3)$indicates that higher-frequency fluctuations play a dominant role in shaping the relationship pattern.
Overall, the proposed T-MENR model provides not only accurate prediction but also meaningful insights into the complex relationships between hypertension prevalence and its associated risk factors. By integrating truncated spline, Gaussian kernel, and Fourier series estimators, the model is able to identify different relationship patterns, including local variations, threshold effects, and fluctuating behaviors that cannot be adequately represented by a single global functional form. These findings indicate that the influence of hypertension risk factors is heterogeneous across their observed ranges, implying that public health interventions should be tailored to the characteristics of each determinant. For example, risk factors exhibiting threshold effects require priority intervention once critical levels are exceeded, whereas predictors with gradual nonlinear relationships are better addressed through continuous monitoring and long-term preventive programs. Likewise, factors displaying fluctuating patterns require adaptive intervention strategies that can respond to changing conditions rather than fixed policy targets. Consequently, the proposed model provides evidence that can assist policymakers in prioritizing high-risk populations, allocating healthcare resources more efficiently, and designing targeted hypertension prevention and control programs based on the specific behavior of individual risk factors.

5) Performance Evaluation of the Optimal T-MENR Model

Based on the best T-MENR model in Equation (1), a visualization was conducted to evaluate the agreement between the predicted and observed values. The results of this visualization, presented in the form of a scatter plot, are shown in Figure 4.

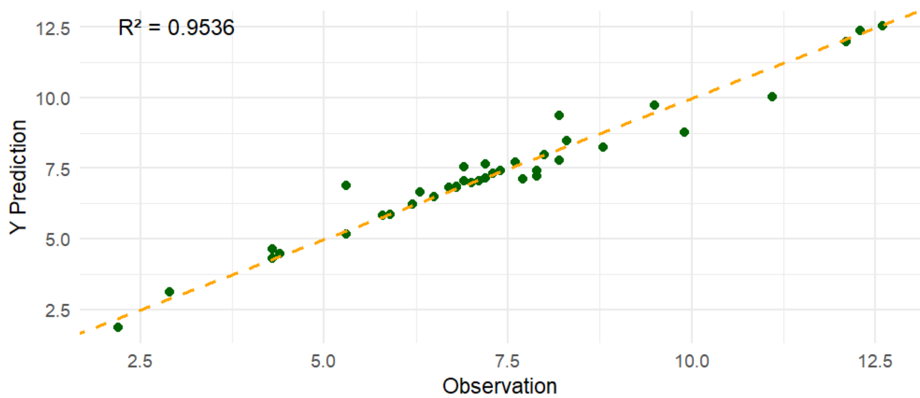


Figure 4. Scatter Plot of Observed and Predicted Values with the Fit Line of the T-MENR Model

Based on Figure 4, the relationship between the observed and predicted values generated by the model is illustrated. It can be seen that most points are distributed closely around the diagonal line, which represents perfect agreement. This indicates that the model achieves a high level of accuracy in predicting the response values. The high coefficient of determination ($R^2 \approx 0.9536$) further supports that the model is able to explain a substantial proportion of the variability in the data effectively. In addition, a further visualization was conducted to compare the patterns of the observed and predicted values based on the order of observations, as presented in Figure 5.

Based on Figure 5, it can be observed that the pattern of the predicted values closely follows the movement of the observed values in a consistent manner. The model is able to capture data fluctuations, including relatively sharp

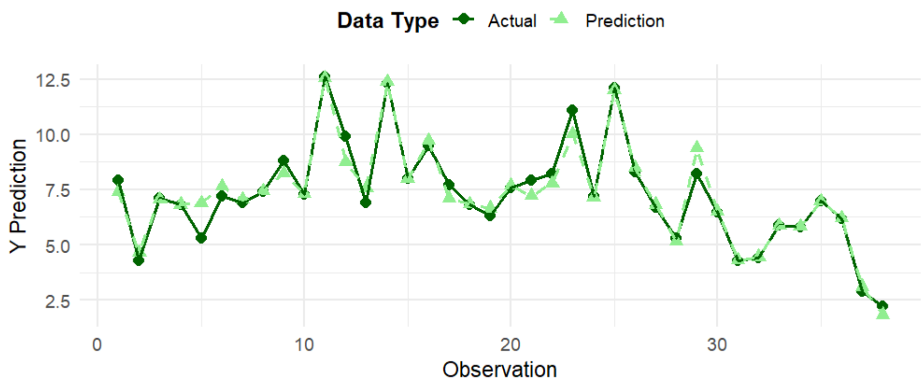


Figure 5. Comparison of Observed and Predicted Value Patterns Based on Observation Order in the T-MENR Model

changes at certain observation points. This indicates that the model not only achieves good overall accuracy, but also effectively represents local variations in the data. Therefore, the obtained T-MENR model can be considered to have optimal predictive performance in terms of both accuracy and pattern conformity. To strengthen the model evaluation results, a performance comparison was conducted between the T-MENR model and single nonparametric models, namely the Truncated Spline, Gaussian Kernel, and Fourier Series models. This comparison is important to objectively assess the superiority of the multi-estimator approach in the T-MENR framework over single-estimator approaches in modeling nonlinear relationships.

To ensure an objective and unbiased comparison, each single model was estimated using its optimal parameter configuration obtained through optimization based on the GCV criterion. In this process, GCV was used as the basis for selecting the optimal parameters for each model, namely three knot points for the Truncated Spline, three oscillations for the Fourier series, and an optimal bandwidth for the Gaussian kernel. Meanwhile, the best T-MENR model was obtained under combination C27, which integrates three knot points, three oscillations, and an optimally estimated bandwidth within a unified modeling framework. Subsequently, the performance of all models was compared using the GCV, RMSE, and R^2 criteria to assess the balance between model fit, prediction error, and the ability to explain data variability. The comparison results of model performance are presented in Table 5.

Table 5. Performance Comparison of the T-MENR Model and Single Nonparametric Models

Evaluation Metric	T-MENR	Single Nonparametric Models		
		Truncated Spline	Fourier Series	Gaussian Kernel
CGV	**0.1972**	2.6549	17.7331	3.7784
RMSE	**0.4902**	1.0719	1.3298	1.6230
R^2	**0.9536**	0.7779	0.6583	0.4910
Computational Time (s)	**317**	189	160	147

Based on Table 5, the T-MENR model (C27) demonstrates overall good performance compared to the single nonparametric models. This is reflected in its lowest GCV and RMSE values, as well as its highest coefficient of determination (R^2). These findings indicate that the T-MENR model achieves the best balance between model fit and complexity, while simultaneously producing minimal prediction error and a higher capacity to explain data variability. This improved predictive performance is accompanied by an increase in computational time, requiring 317 seconds compared with 147–189 seconds for the single nonparametric models. This additional computational cost is an expected consequence of simultaneously optimizing multiple smoothing parameters from the truncated spline, Gaussian kernel, and Fourier series components. Nevertheless, the increase in computation time is compensated by a substantial improvement in model accuracy, suggesting that the proposed T-MENR framework provides a favorable trade-off between computational efficiency and predictive performance.

This good performance indicates that the multi-estimator approach in the T-MENR model is more effective in capturing the complexity of nonlinear relationships compared to single-estimator approaches. By integrating multiple estimators within a unified modeling framework, the T-MENR model can represent data patterns more comprehensively, both in terms of global trends and local variations. Overall, these evaluation results indicate that the obtained T-MENR model exhibits strong predictive performance. The model is not only able to produce estimates that closely approximate the observed values, but also consistently follows the underlying patterns and variations in the data. Therefore, this model is suitable to be used as a predictive tool for estimating hypertension prevalence and can serve as a robust basis for data driven decision making in the health sector.

4. Conclusion

This study demonstrates that the Three-Mixed Estimator Nonparametric Regression (T-MENR) approach is a flexible and effective framework for modeling complex nonlinear relationships between health behavioral factors and hypertension prevalence in Indonesia. By integrating multiple nonparametric estimators within a single model, T-MENR can accommodate heterogeneous data characteristics and capture diverse relationship patterns across predictor variables.

The empirical results show that the optimal model configuration, selected using the Generalized Cross Validation (GCV) criterion, achieves strong predictive performance with a minimum GCV value of 0.1972, RMSE of 0.4902, and coefficient of determination of approximately 0.9536. These findings indicate that the T-MENR model can explain variations in hypertension prevalence with high accuracy. Overall, the multi-estimator approach in T-MENR not only improves predictive performance but also provides a more realistic representation of complex relationships in health data. Despite these promising results, several limitations should be acknowledged. First, the current study focuses on the development and estimation of the T-MENR model, and its theoretical statistical properties, including consistency, asymptotic bias, variance, and efficiency, have not yet been investigated. Second, the performance evaluation is limited to comparisons with individual nonparametric estimators, whereas comparisons with modern machine learning approaches have not been conducted. Third, the model selection procedure requires substantial computational effort because it involves simultaneous optimization of multiple smoothing parameters across numerous candidate model combinations.

Future research should therefore establish the asymptotic properties of the proposed estimator, develop statistical inference procedures such as hypothesis testing and confidence intervals, investigate more computationally efficient optimization strategies, and conduct comprehensive benchmarking against state-of-the-art machine learning methods to further evaluate the predictive capability and practical applicability of the T-MENR framework.

Acknowledgments

The authors would like to express their sincere gratitude to Universitas Brawijaya for providing financial support for this research and the publication of this paper. The authors also extend their appreciation to the reviewers for their valuable comments and constructive suggestions, which have greatly improved the quality of this work.

FUNDING STATEMENT

This research was funded through a research grant provided by Visiting Lecturer Program Batch 2, Universitas Brawijaya. The funding support was part of the university's internal research scheme aimed at encouraging academic research and innovation. The authors gratefully acknowledge this support, which made the completion of this study possible.

REFERENCES

1. N. R. Drapper and H. Smith, *Applied Regression Analysis, 3rd ed.*, Canada: John Wiley and Sons Inc., 1998.
2. A. T. R. Dani *et al.*, "New Statistical Approach to Forecasting Earth's Skin Temperature from MERRA-2 Satellite Using Semiparametric Time Series Regression with Mixed Additive Spline Fourier (STSR-MASF)," *Forecasting*, vol. 8, no. 1, p. 6, Jan. 2026. doi: 10.3390/forecast8010006.
3. W. Hardle, *Applied Nonparametric Regression*, Berlin: Humboldt-Universitat zu Berlin, 1994. doi: 10.2307/2348990.
4. H. Chen, "Polynomial splines and nonparametric regression," *J. Nonparametr. Stat.*, vol. 1, no. 1–2, pp. 143–156, Jan. 1991. doi: 10.1080/10485259108832516.
5. N. R. A. A. Siregar, N. Chamidah, M. Rifada, B. Lestari, and D. Aydin, "Penalized Spline Semiparametric Logistic Regression for Modelling Coronary Heart Disease Risk Based on Demographic and Lifestyle Factors," *Statistics, Optimization and Information Computing*, vol. 15, no. 6, pp. 4736–4756, May 2026. doi: 10.19139/soic-2310-5070-3334.
6. K. Takezawa, *Introduction to Regression Methods*, Canada, 2006. doi: 10.1201/9780367813161-12.
7. A. T. Ampa *et al.*, "Selection of Optimal Smoothing Parameters in Mixed Estimator of Kernel and Fourier Series in Semiparametric Regression," *Journal of Physics: Conference Series*, vol. 2123, no. 1, p. 012035, 2021. doi: 10.1088/1742-6596/2123/1/012035.
8. L. R. Cheruiyot, "Local linear regression estimator on the boundary correction in nonparametric regression estimation," *Journal of Statistical Theory and Applications*, vol. 19, no. 3, pp. 460–471, 2020. doi: 10.2991/jsta.d.201016.001.
9. M. A. D. Octavanny *et al.*, "A New Mixed Estimator in Nonparametric Regression for Longitudinal Data," *Journal of Mathematics*, vol. 2021, 2021. doi: 10.1155/2021/3909401.
10. N. Y. Adrianingsih *et al.*, "Mixture Model Nonparametric Regression and Its Application," *Journal of Physics: Conference Series*, vol. 1842, no. 1, 2021. doi: 10.1088/1742-6596/1842/1/012044.
11. R. Hidayat *et al.*, "Kernel-Spline Estimation of Additive Nonparametric Regression Model," *IOP Conference Series: Materials Science and Engineering*, vol. 546, no. 5, 2019. doi: 10.1088/1757-899X/546/5/052028.
12. H. Nurcahayani *et al.*, "Nonparametric truncated spline regression on modelling mean years schooling of regencies in Java," *AIP Conference Proceedings*, vol. 2194, 2019. doi: 10.1063/1.5139805.
13. B. Lestari *et al.*, "Spline estimator and its asymptotic properties in multiresponse nonparametric regression model," *Songklanakarin Journal of Science and Technology*, vol. 42, no. 3, pp. 533–548, 2020. doi: 10.14456/sjst-psu.2020.68.
14. N. Chamidah *et al.*, "Estimating Mean Arterial Pressure Affected by Stress Scores Using Spline Nonparametric Regression Model Approach," *Communications in Mathematical Biology and Neuroscience*, vol. 2020, no. 72, pp. 1–12, 2020.
15. Sifriyani *et al.*, "Application of Nonparametric Geographically Weighted Spline Regression Model for Spatial Mapping of Open Unemployment Rate in Kalimantan," *Journal of Physics: Conference Series*, vol. 2123, no. 1, 2021. doi: 10.1088/1742-6596/2123/1/012038.
16. A. T. Fitriyah, N. Chamidah, T. Saifudin, B. Lestari, and D. Aydin, "Least Squares Spline Estimation Method in Semiparametric Time Series Regression for Predicting Indonesia Composite Index," *Statistics, Optimization and Information Computing*, vol. 14, no. 6, pp. 3776–3803, Nov. 2025. doi: 10.19139/soic-2310-5070-2704.
17. I. N. Budiantara *et al.*, "The Combination of Spline and Kernel Estimator for Nonparametric Regression and its Properties," *Applied Mathematical Sciences*, vol. 9, no. 122, pp. 6083–6094, 2015. doi: 10.12988/ams.2015.58517.
18. K. H. Aljuhani and L. I. Al turk, "Modification of the adaptive Nadaraya-Watson kernel regression estimator," *Scientific Research and Essays*, vol. 9, no. 22, pp. 966–971, 2014. doi: 10.5897/SRE2014.6121.
19. A. T. Ampa *et al.*, "Kernel multivariable semiparametric regression model in estimating the level of open unemption in East Java Province," *Journal of Physics: Conference Series*, vol. 1899, no. 1, 2021. doi: 10.1088/1742-6596/1899/1/012127.
20. B. C. Kouassi *et al.*, "Nadaraya-Watson estimation of a nonparametric autoregressive model," *Malaya Journal of Matematik*, vol. 9, no. 4, pp. 251–258, 2021. doi: 10.26637/mjm904/009.
21. M. Bilodeau, "Fourier smoother and additive models," *Canadian Journal of Statistics*, vol. 20, no. 3, pp. 257–269, 1992. doi: 10.2307/3315313.
22. A. Prahutama *et al.*, "Modelling fourier regression for time series data - A case study: Modelling inflation in foods sector in Indonesia," *Journal of Physics: Conference Series*, vol. 974, no. 1, 2018. doi: 10.1088/1742-6596/974/1/012067.
23. Suparti *et al.*, "Modeling longitudinal data based on Fourier regression," *Journal of Physics: Conference Series*, vol. 1217, no. 1, 2019. doi: 10.1088/1742-6596/1217/1/012105.
24. N. P. A. M. Mariati *et al.*, "Modeling Poverty Percentages in the Papua Islands using Fourier Series in Nonparametric Regression Multivariable," *Journal of Physics: Conference Series*, vol. 1397, no. 1, 2019. doi: 10.1088/1742-6596/1397/1/012071.
25. N. Chamidah *et al.*, "Fourier series estimator for predicting international market price of white sugar," *AIP Conference Proceedings*, vol. 2329, 2021. doi: 10.1063/5.0042287.
26. H. Nurcahayani *et al.*, "The Curve Estimation of Combined Truncated Spline and Fourier Series Estimators for Multiresponse Nonparametric Regression," *Mathematics*, vol. 9, no. 10, p. 1141, 2021.
27. R. Hidayat *et al.*, "The regression curve estimation by using mixed smoothing spline and kernel (MsS-K) model," *Communications in Statistics - Theory and Methods*, vol. 0, no. 0, pp. 1–12, 2020. doi: 10.1080/03610926.2019.1710201.
28. N. P. A. M. Mariati *et al.*, "The application of mixed smoothing spline and fourier series model in nonparametric regression," *Symmetry*, vol. 13, no. 11, 2021. doi: 10.3390/sym13112094.
29. D. P. Rahmawati *et al.*, "Mixed Spline Smoothing and Kernel Estimator in Biresponse Nonparametric Regression," *International Journal of Mathematics and Mathematical Sciences*, vol. 2021, 2021. doi: 10.1155/2021/6611084.
30. A. T. R. Dani *et al.*, "Optimal Knots Point and Bandwidth Selection in Modeling Mixed Estimator Nonparametric Regression," *IOP Conference Series: Materials Science and Engineering*, vol. 1115, no. 1, p. 012020, 2021. doi: 10.1088/1757-899x/1115/1/012020.
31. A. W. Wening *et al.*, "Semiparametric regression curve estimation for longitudinal data using mixed spline truncated and fourier series estimator," *Journal of Physics: Conference Series*, vol. 1538, no. 1, 2020. doi: 10.1088/1742-6596/1538/1/012061.
32. N. P. A. M. Mariati *et al.*, "Combination Estimation of Smoothing Spline and Fourier Series in Nonparametric Regression," *Journal of Mathematics*, vol. 2020, 2020. doi: 10.1155/2020/4712531.

33. I. W. Sudiarsa, "Simulations Study Combined Estimator Fourier Series and Spline Truncated in Multivariable Nonparametric Regression," *IOP Conference Series: Materials Science and Engineering*, vol. 546, no. 5, 2019. doi: 10.1088/1757-899X/546/5/052074.
34. N. Afifah *et al.*, "Mixed Estimator of Kernel and Fourier Series in Semiparametric Regression," *Journal of Physics: Conference Series*, vol. 855, no. 1, 2017. doi: 10.1088/1742-6596/855/1/012002.
35. I. N. Budiantara *et al.*, "Modeling percentage of poor people in Indonesia using kernel and Fourier series mixed estimator in nonparametric regression," *Investigacion Operacional*, vol. 40, no. 4, pp. 538–550, 2019. doi: 10.5281/zenodo.3721293.
36. V. Ratnasari *et al.*, "Nonparametric Regression Mixed Estimators of Truncated Spline and Gaussian Kernel based on Cross-Validation (CV), Generalized Cross-Validation (GCV), and Unbiased Risk (UBR) Methods," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 11, no. 6, pp. 2400–2406, 2021.
37. E. Ana *et al.*, "Modeling of hypertension risk factors using local linear of additive nonparametric logistic regression," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, 2019. doi: 10.1088/1742-6596/1397/1/012067.
38. T. M. Kaombe *et al.*, "Bivariate logistic regression model diagnostics applied to analysis of outlier cancer patients with comorbid diabetes and hypertension in Malawi," *Scientific Reports*, vol. 13, no. 1, 2023. doi: 10.1038/s41598-023-35475-z.
39. W. Xu *et al.*, "Optimal systolic blood pressure target, time to intensification, and time to follow-up in treatment of hypertension: Population based retrospective cohort study," *BMJ*, vol. 350, 2015. doi: 10.1136/bmj.h158.
40. I. Tasić *et al.*, "Predictors of cardiovascular events in hypertensive patients with high cardiovascular risk," *Medicina*, vol. 56, no. 4, 2020. doi: 10.3390/medicina56040182.
41. P. Degoulet *et al.*, "Factors predictive of attendance at clinic and blood pressure control in hypertensive patients," *British Medical Journal*, vol. 287, pp. 88–93, 1983. doi: 10.1136/bmj.287.6385.88.
42. V. Ratnasari *et al.*, "Nonparametric Regression Mixed Estimators of Truncated Spline and Gaussian Kernel based on Cross-Validation (CV), Generalized Cross-Validation (GCV), and Unbiased Risk (UBR) Methods," *International Journal on Advanced Science Engineering Information Technology*, vol. 11, no. 6, pp. 2400–2406, 2021.
43. C.-K. Chu and J. S. Marron, "Choosing a Kernel Regression Estimator," *Statistical Science*, vol. 6, no. 4, pp. 404–436, 1991.
44. W. Cui and M. Wei, "Strong Consistency of Kernel Regression Estimate," *Open Journal of Statistics*, vol. 3, no. 3, pp. 179–182, 2013. doi: 10.4236/ojs.2013.33020.
45. H. Alabdally *et al.*, "Improving Dry-Bulb Air Temperature Prediction Using a Hybrid Model Integrating Genetic Algorithms with a Fourier–Bessel Series Expansion-Based LSTM Model," *Forecasting*, vol. 7, no. 3, 2025. doi: 10.3390/forecast7030046.
46. W. Y. Ming and L.-J. Huang, "Fourier Series Neural Networks for Regression," in *Proceedings of IEEE International Conference on Applied System Innovation*, IEEE, 2018, pp. 716–719.