Jurnal Matematika UNAND
Vol. **15** No. **3** Hal. 406 – 417
Edisi Juli 2026
ISSN : 2303–291X
e-ISSN : 2721–9410

https://doi.org/10.25077/jmua.15.3.406-417.2026

PERFORMANCE COMPARISON OF SEMIPARAMETRIC KERNEL REGRESSION USING EPANECHNIKOV, GAUSSIAN, AND BIWEIGHT KERNELS IN MODELING LIFE EXPECTANCY IN INDONESIA

CINTA RIZKI OKTARINA[a], SIGIT NUGROHO[b],* IDHIA SRILIANA[b], PUTRI SUCI ARIA[a]
[a] *Department of Mathematics,*
Faculty of Information Technology, Institut Teknologi Batam, Indonesia,
[b] *Department of Statistics,*
Faculty of Mathematics and Natural Sciences, Universitas Bengkulu, Indonesia,
email : cinta@iteba.ac.id, snugroho@unib.ac.id,
idhiasriliana@unib.ac.id, 2424001@student.iteba.ac.id

Received February 3, 2026 Received in revised form May 20, 2026
Accepted June 27, 2026 Available online July 31, 2026

Abstract. Life expectancy is an important indicator of population health and socioeconomic development, and its relationship with explanatory variables often exhibits both linear and nonlinear patterns. This study applies a semiparametric kernel regression approach to model life expectancy in Indonesia by combining a parametric component and a nonparametric kernel-based component. Undernutrition is modeled parametrically, while access to improved drinking water and the illiteracy rate are treated as nonparametric predictors to flexibly capture their nonlinear effects. The nonparametric component is estimated using the Nadaraya-Watson kernel estimator, while parameter estimation is carried out using the Ordinary Least Squares method. Several kernel functions are evaluated, and the optimal kernel and bandwidth are selected based on the minimum Generalized Cross-Validation (GCV) criterion. The results show that the Gaussian kernel with an optimal bandwidth of $(100, 100)$ provides the best performance. The proposed model achieves excellent accuracy, as indicated by a coefficient of determination of 99.99% and a very small Root Mean Square Error (RMSE). These findings demonstrate that semiparametric kernel regression is a flexible and reliable approach for modeling life expectancy and other public health indicators characterized by complex relationships.

Keywords: bandwidth selection, life expectancy, semiparametric kernel regression

1. Introduction

Regression analysis is a statistical method used to model the relationship between response variables and predictor variables [1]. The regression function

*Corresponding author

can be estimated using three main approaches, namely parametric, nonparametric, and semiparametric methods [2]. Parametric regression assumes a specific functional form, nonparametric regression allows flexible functional shapes, while semiparametric regression combines both approaches by modeling known components parametrically and unknown components nonparametrically [3].

Semiparametric regression combines parametric and nonparametric components to model complex relationships between response and predictor variables. This approach allows part of the regression function to be specified parametrically, while the remaining part is left unspecified and estimated nonparametrically, thereby providing greater flexibility than fully parametric models [4]. In practice, the nonparametric component is commonly estimated using kernel-based methods due to their ability to capture nonlinear patterns in the data without imposing a predetermined functional form, as demonstrated in kernel and spline-based applications.

The kernel function plays a fundamental role in kernel-based nonparametric regression by determining the weighting structure of observations in the neighborhood of the estimation point, where observations closer to the target point are assigned larger weights. The smoothness of the resulting regression curve is jointly governed by the choice of kernel function and the bandwidth parameter, which together control the bias–variance trade-off of the estimator [5]. Commonly adopted kernel functions include Uniform, Epanechnikov, Gaussian, and biweight kernels, each exhibiting distinct smoothing characteristics and theoretical properties [6].

Previous studies indicate that semiparametric kernel regression has been widely applied in modeling the Composite Stock Price Index (IDX), which involves a combination of linear and nonlinear relationships among variables. Several studies employ the Nadaraya-Watson estimator with various kernel functions, such as Gaussian and triangular kernels, to accommodate differences in patterns between parametric and nonparametric predictor variables. The results demonstrate excellent model performance, as indicated by a high coefficient of determination, namely $R^2 = 97.15\%$, along with satisfactory forecasting accuracy reflected by MAPE values below 10% [7,8].

These findings confirm that semiparametric kernel regression is a flexible and reliable approach for modeling economic and financial data characterized by high volatility and complex relationships that do not fully satisfy classical parametric regression assumptions. Given its strong empirical performance, this approach is also well suited for modeling life expectancy, where socioeconomic determinants often exhibit both linear and nonlinear effects.

2. Methods

2.1. *Regression Analysis*

Regression analysis is a statistical method used to investigate and model the relationship between a response variable and one or more predictor variables [9]. Over time, regression analysis has been widely applied across various scientific disciplines and has become a fundamental tool in data modeling. According to [2], there are three main approaches for estimating regression curves, namely parametric, nonparametric, and semiparametric regression.

2.1.1. *Parametric Regression*

Parametric regression is applied when the functional form of the relationship between the response variable and the predictors is known a priori based on theoretical considerations or prior information [2]. The most commonly used parametric regression model is linear regression, which can generally be expressed as follows [10]:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i, \tag{2.1}$$

Parameter estimation in this model is typically performed using the *Ordinary Least Squares* (OLS) method under the assumptions stated in the Gauss-Markov theorem [11].

2.1.2. *Nonparametric Regression*

Nonparametric regression is a flexible modeling approach that does not assume a specific functional form for the regression function, allowing it to adapt to the underlying data structure [12]. The general form of a nonparametric regression model can be written as follows [4]:

$$y_i = f(t_i) + \varepsilon_i, \tag{2.2}$$

This approach is particularly suitable when the relationship between variables is unknown and is determined entirely by the observed data patterns.

2.1.3. *Semiparametric Regression*

Semiparametric regression combines parametric and nonparametric components, where part of the model is specified parametrically while the remaining part is modeled flexibly [2]. This approach is appropriate when certain aspects of the relationship structure are known, whereas others are not. The general form of a semiparametric regression model is given by [13]:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + f(t_i) + \varepsilon_i. \tag{2.3}$$

In this model, the parametric component is estimated using linear regression techniques, while the nonparametric component $f(t_i)$ is commonly estimated using kernel-based or spline-based approaches.

2.2. ***Kernel Nadaraya Watson***

Kernel nonparametric regression, also known as local mean regression, is an approach commonly used when data points are unevenly spaced or when predictor variables are stochastic, as it estimates the regression function without imposing linearity or a specific parametric form [12]. This method is based on a weighted average of the response variable, where the weights are determined by a kernel function and a bandwidth parameter (h). Kernel regression can be viewed as a special case of local polynomial regression of degree zero (local constant), in which the regression function is locally approximated by a constant and greater weight is assigned to observations closer to the estimation point [5]. A widely used estimator in this framework is the Nadaraya-Watson kernel estimator, where the bandwidth serves as the main smoothing parameter, with larger values of h producing smoother estimated curves [14,15]. The selection of an optimal bandwidth, for instance via the Generalized Cross-Validation (GCV) method, is therefore crucial for improving estimation accuracy and stability.

Several kernel functions commonly used as weighting functions in nonparametric regression include the Epanechnikov, Gaussian, and Biweight kernels [16]. Let $K(t_i)$ denote a kernel function used as a weighting function in kernel-based estimation involving a bandwidth parameter [17]. In general, a kernel function is real-valued, continuous, bounded, and symmetric [18], and must satisfy the normalization and zero first-moment properties [19].

Based on these properties, the Nadaraya-Watson kernel regression estimator constructs the estimated regression function as a weighted average of the response variable. At point t_i, the estimator is defined as:

$$\hat{m}(t_i) = \frac{K\left(\sum_{g=1}^{G} \frac{t_g - t_{gi}}{h_g}\right) y_i}{\frac{1}{n}\sum_{i=1}^{n} K\left(\sum_{g=1}^{G} \frac{t_g - t_{gi}}{h_g}\right)}, \tag{2.4}$$

$$= W_h(t_i)\, y_i, \tag{2.5}$$

where:

$$W_h(t_i) = \frac{K\left(\sum_{g=1}^{G} \frac{t_g - t_{gi}}{h_g}\right)}{\frac{1}{n}\sum_{i=1}^{n} K\left(\sum_{g=1}^{G} \frac{t_g - t_{gi}}{h_g}\right)}. \tag{2.6}$$

The weight function $W_h(t_i)$ can be expressed in matrix form as follows:

$$\mathbf{W}_h = \begin{bmatrix} W_h(t_1) & 0 & \cdots & 0 \\ 0 & W_h(t_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_h(t_n) \end{bmatrix}. \tag{2.7}$$

2.3. *Semiparametric Kernel Regression*

In the semiparametric regression model, the unknown function $m(\cdot)$ and the regression parameters $\boldsymbol{\beta}$ are estimated from the data. The function $m(\cdot)$ is estimated using a kernel estimator, whereas $\boldsymbol{\beta}$ is estimated using the ordinary least squares method. Suppose there are n observations with p parametric predictors x and one nonparametric predictor t. The semiparametric model can be written as:

$$y_i = \sum_{k=1}^{p} \beta_k x_{ik} + m(t_i) + \varepsilon_i, \tag{2.8}$$

for $i = 1, 2, \cdots, n$.

Rearranging Equation (2.8) yields:

$$y_i - \sum_{k=1}^{p} \beta_k x_{ik} = m(t_i) + \varepsilon_i. \tag{2.9}$$

Define the adjusted response variable:

$$y_i^* = y_i - \sum_{k=1}^{p} \beta_k x_{ik}. \tag{2.10}$$

Then, Equation (2.9) becomes:

$$y_i^* = m(t_i) + \varepsilon_i. \tag{2.11}$$

The nonparametric component $m(t)$ is interpreted as the conditional expectation of y^* given $T = t$, namely:

$$m(t) = E(y^* \mid T = t) = \int_{-\infty}^{\infty} y^* f(y^* \mid t)\, dy^* = \frac{\displaystyle\int_{-\infty}^{\infty} y^* f(t, y^*)\, dy^*}{f(t)}. \tag{2.12}$$

Using the Nadaraya-Watson kernel estimator, $m(t)$ can be estimated as:

$$\hat{m}(t) = \frac{K\left(\sum_{g=1}^{G} \frac{t_g - t_{gi}}{h_g}\right)}{\displaystyle\frac{1}{n} \sum_{i=1}^{n} K\left(\sum_{g=1}^{G} \frac{t_g - t_{gi}}{h_g}\right)} = \sum_{i=1}^{n} W_i(t)\, y_i^*, \tag{2.13}$$

where the kernel weights are defined by:

$$W_i(t) = \frac{K\left(\sum_{g=1}^{G} \frac{t_g - t_{gi}}{h_g}\right)}{\displaystyle\frac{1}{n} \sum_{i=1}^{n} K\left(\sum_{g=1}^{G} \frac{t_g - t_{gi}}{h_g}\right)}, \qquad i = 1, 2, \cdots, n. \tag{2.14}$$

Substituting y_i^* from Equation (2.10) into Equation (2.13) gives:

$$\hat{m}(t) = \sum_{i=1}^{n} W_i(t) \left(y_i - \sum_{k=1}^{p} \beta_k x_{ik} \right). \tag{2.15}$$

Finally, the estimated semiparametric kernel regression model is obtained by replacing $m(t_i)$ in Equation (2.8) with $\hat{m}(t_i)$:

$$y_i = \sum_{k=1}^{p} \beta_k x_{ik} + \sum_{i=1}^{n} W_i(t) \left(y_i - \sum_{k=1}^{p} \beta_k x_{ik} \right) + \varepsilon_i. \tag{2.16}$$

To determine an appropriate smoothing level, the bandwidth h is selected by minimizing the Generalized Cross-Validation (GCV) criterion [20]. For a given h, let $\hat{\mathbf{y}}(h) = (\hat{y}_1(h), \cdots, \hat{y}_n(h))^\top$ denote the fitted values obtained from the semi-parametric kernel regression model, and let $\mathbf{S}(h)$ be the corresponding smoothing (hat) matrix such that $\hat{\mathbf{y}}(h) = \mathbf{S}(h)\mathbf{y}$. The GCV function is defined as:

$$\mathrm{GCV}(h) = \frac{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\big(y_i - \hat{y}_i(h)\big)^2}{\left(1 - \dfrac{1}{n}\mathrm{tr}\big(\mathbf{S}(h)\big)\right)^2}. \tag{2.17}$$

The optimal bandwidth is then chosen as:

$$h_{\text{opt}} = \arg\min_{h>0}\ \mathrm{GCV}(h). \tag{2.18}$$

In this study, the bandwidth that yields the smallest GCV value is adopted to produce a stable and accurate estimate of the nonparametric component.

3. Result and Discussion

3.1. *Research Design, Data Source, and Variables*

This study employs an applied quantitative research design using a semiparametric regression approach to analyze factors influencing life expectancy in Indonesia. The data used are cross-sectional data for the year 2024, with provinces in Indonesia as the unit of observation. The data were obtained from official publications of Statistics Indonesia (BPS) and other relevant government institutions.

The response variable in this study is life expectancy (Y), which represents the overall level of population health and quality of life. The explanatory variables consist of three social and public health indicators, namely X_1 undernutrition (percentage of the population experiencing poor nutritional status) [21], X_2 access to improved drinking water (percentage of households with access to safe drinking water) [22], and X_3 the illiteracy rate (percentage of the population unable to read and write) [23]. All variables are expressed as percentages and are assumed to influence life expectancy through both linear and nonlinear relationships. Therefore, a semiparametric regression approach is adopted to flexibly capture the complex relationships between the response variable and its predictors, which may not be adequately modeled using fully parametric regression methods.

3.2. *Analysis Procedure*

The analysis in this study was conducted through a comparative estimation process using the Nadaraya–Watson kernel with Epanechnikov, Gaussian, and biweight kernel functions:

(1) Collecting life expectancy data along with factors suspected to influence it.
(2) Visualizing the data using scatter plots to identify relationship patterns between the response variable and predictor variables.
(3) Determining the general form of the semiparametric kernel regression model.
(4) Selecting the kernel function and determining the optimal bandwidth using the Generalized Cross Validation (GCV) method.
(5) Estimating the parameters and regression function of the semiparametric kernel model and visualizing the estimation results.

(6) Evaluating model performance using the coefficient of determination (R^2) and Root Mean Square Error (RMSE).

3.3. *Visualization of the Relationship Between the Response and Predictor Variables*

Visualizing the relationship between the response variable and the predictor variables is an essential preliminary step in regression analysis to identify the direction, strength, and underlying pattern of the relationships. Scatter plots allow the distribution of the data and potential linear or nonlinear trends to be examined visually, providing initial insight into the structure of the relationships among variables. This information serves as an important basis for selecting an appropriate modeling approach, particularly in determining whether the relationships can be adequately captured by parametric models or require more flexible methods such as semiparametric kernel regression.

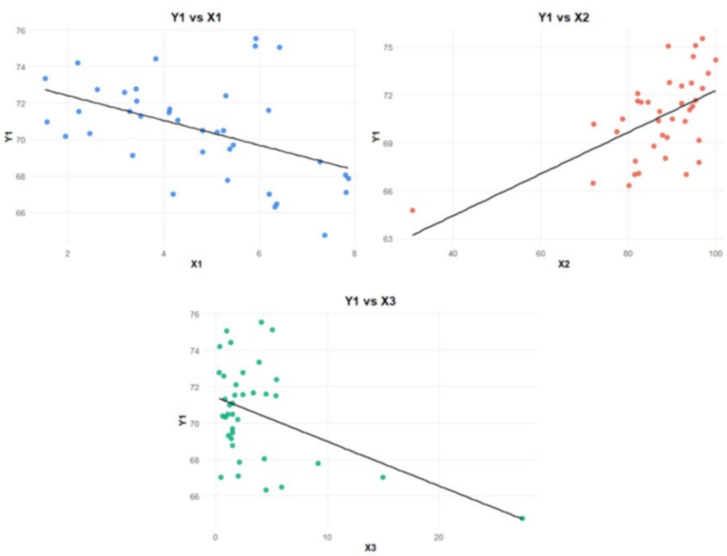


Figure. 1: Scatter plots of the response variable Y against predictor variables X_1, X_2, and X_3.

Based on the scatter plots, distinct relationship patterns are observed between the response variable Y and each predictor. The relationship between Y and X_1 shows an approximately linear and negative trend, indicating that the effect of X_1 can be adequately represented using a parametric component. In contrast, the relationships between Y and X_2 as well as Y and X_3 exhibit more complex patterns with greater dispersion, suggesting the presence of nonlinear behavior that cannot be fully captured by a simple linear model. Although X_2 tends to have a positive association with Y and X_3 shows a negative association, the variability of the data across their ranges supports the use of flexible modeling. Therefore, X_2 and X_3 are treated as nonparametric components, while X_1 is modeled parametrically within a semiparametric kernel regression framework to effectively capture both linear and nonlinear effects on life expectancy.

3.4. *General Semiparametric Kernel Regression Model*

The general semiparametric kernel regression model combines parametric and nonparametric components to capture both linear and nonlinear relationships between the response variable and predictor variables. This approach enables clear interpretation of predictors with linear effects through the parametric component, while simultaneously accommodating complex or unknown functional relationships using kernel-based nonparametric components.

Suppose there is one response variable y, one predictor with a parametric component x, and two predictors with nonparametric components $\mathbf{t} = (t_1, t_2)$. The parametric effect of x is modeled linearly, whereas the joint effect of (t_1, t_2) is estimated nonparametrically using a kernel estimator. The general form of the semiparametric kernel regression model is given by:

$$\hat{y}_i = \beta_0 + \beta_1 x_i + \frac{K\left(\frac{t_1 - t_{1i}}{h_1} + \frac{t_2 - t_{2i}}{h_2}\right)[y_i - (\beta_0 + \beta_1 x_i)]}{\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{t_1 - t_{1i}}{h_1} + \frac{t_2 - t_{2i}}{h_2}\right)}, \tag{3.1}$$

where h_1 and h_2 are the bandwidth parameters associated with t_1 and t_2, respectively, and β_0 and β_1 are regression parameters corresponding to the parametric component.

3.5. *Selection of the Optimal Kernel Function and Bandwidth*

The determination of the optimal kernel function and bandwidth is a crucial step in semiparametric kernel regression, as the bandwidth parameter strongly influences the smoothness and accuracy of the estimated regression function. In this study, the optimal bandwidth is selected using the Generalized Cross-Validation (GCV) criterion by evaluating a grid of candidate bandwidth values. Specifically, the bandwidth is searched within a predefined range with a lower bound of 5 and an upper bound of 100, using an increment of 5. For each candidate bandwidth, the corresponding GCV value is computed, and the bandwidth that yields the minimum GCV value is selected as the optimal bandwidth. The following section presents the results of the data analysis used to determine the optimal kernel function and bandwidth.

The results indicate that among the kernel functions considered, the Gaussian kernel yields the smallest GCV value, suggesting the best estimation performance compared to the Epanechnikov and Biweight kernels. This finding implies that the Gaussian kernel provides a better balance between smoothness and goodness of fit for the data under study. The optimal bandwidth combination for all kernel functions is obtained at $(h_1, h_2) = (100, 100)$, indicating that a relatively smooth estimation is required to capture the underlying relationship between the

Table 1: Optimal Kernel Function and Bandwidth Based on GCV

Kernel Function	**GCV**	**Combination**	h
Epanechnikov	1.43×10^{-4}	400	(100, 100)
Gaussian	3.19×10^{-5}	400	(100, 100)
Biweight	4.62×10^{-4}	400	(100, 100)

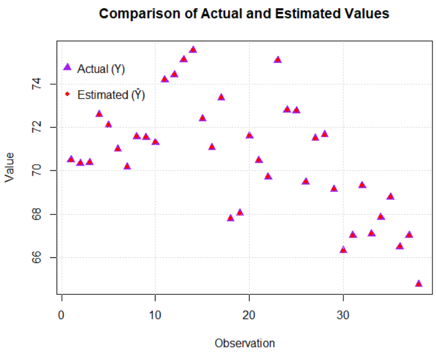
response variable and the predictors. Therefore, the Gaussian kernel with bandwidth $(100, 100)$ is selected as the optimal kernel-bandwidth combination for subsequent analysis.

3.6. *Parameter Estimation and Model Visualization*

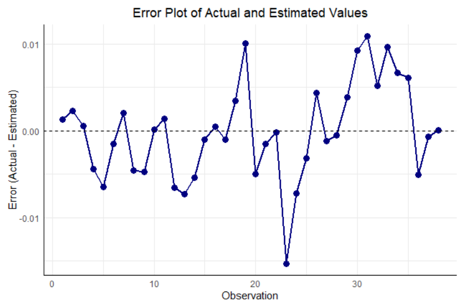
This subsection presents the estimation results of the parameters in the semiparametric kernel regression model. The parametric component is estimated using regression techniques, yielding the estimates of the intercept and slope parameters, while the nonparametric component is estimated using a kernel-based approach with the optimal bandwidth values $h_1 = h_2 = 100$. The estimated parameters describe the linear effect of the parametric predictor, whereas the kernel-based component captures the nonlinear effects associated with the nonparametric predictors. Based on the estimation results, the semiparametric kernel regression model is expressed as follows:

$$\hat{y}_i = 61.6096 + 0.1028\, x_i + \frac{K\left(\frac{t_1 - t_{1i}}{100} + \frac{t_2 - t_{2i}}{100}\right)[y_i - (61.6096 + 0.1028\, x_i)]}{\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{t_1 - t_{1i}}{100} + \frac{t_2 - t_{2i}}{100}\right)}. \quad (3.2)$$

The estimation results indicate that the semiparametric kernel regression model exhibits an excellent level of performance. The coefficient of determination reaches $R^2 = 99.99\%$, indicating that almost all variability in the response variable is explained by the model. In addition, the Root Mean Square Error (RMSE) for Y is equal to 0.00550, which reflects a very small prediction error. These results confirm that the proposed model provides an accurate and reliable representation of the relationship between the response variable and its predictors. Figure 2a presents a visualization comparing the observed values of Y and the corresponding estimated values obtained from the semiparametric kernel regression model.



(a) Comparison between observed and estimated values of Y.



(b) Error plot of the semiparametric kernel regression model.

Figure. 2: Visualization of model performance: comparison between observed and estimated values and the corresponding error plot.

The comparison plot shows that the estimated values closely follow the observed values of Y, indicating a very strong agreement between the model predictions and the actual data. This result confirms that the semiparametric kernel regression model is able to capture the underlying data pattern accurately. Furthermore, the error plot demonstrates that the residuals are small and fluctuate around zero without exhibiting any systematic pattern, suggesting that the model is free from significant bias and provides stable and reliable predictions across observations.

4. Conclusion

This study demonstrates that semiparametric kernel regression is an effective and flexible approach for modeling life expectancy in Indonesia by accommodating both linear and nonlinear relationships among socioeconomic factors. Undernutrition is adequately captured through the parametric component, while access to improved drinking water and illiteracy rate exhibit nonlinear effects that are well modeled using kernel-based estimation. Among the kernel functions evaluated, the Gaussian kernel provides the best performance, as indicated by the minimum Generalized Cross-Validation (GCV) value, with an optimal bandwidth of $(100, 100)$. The resulting model achieves excellent accuracy, reflected by a coefficient of determination of 99.99% and a very small RMSE value, confirming the reliability of the proposed approach. These findings suggest that semiparametric kernel regression, particularly with a Gaussian kernel, is well suited for analyzing life expectancy data and other public health indicators characterized by complex relationships.

Bibliography

[1] R. Pahlepi, I. Sriliana, W. Agwil, and C. R. Oktarina. Biresponse spline truncated nonparametric regression modeling for longitudinal data on monthly

stock prices of three private banks in indonesia. *Barekeng*, 19(4):2467–2480, 2025.

[2] I. N. Budiantara. *Regresi Nonparametrik Spline Truncated*. ITS Press, Surabaya, 2019.

[3] S. Yang, J. Liu, P. Du, J. Wu, and L. Wan. Semiparametric approach for dual-response robust design with complex two-factor interactions. *Quality Engineering*, pages 1–23, 2025.

[4] D. Ruppert, M.P. Wand, and R.J. Carroll. *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2003.

[5] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1996.

[6] C.R Oktarina, S. Nugroho, I. Sriliana, P. Novianti, E. Sunandi, and R. Pahlepi. Estimation of stunting and wasting in sumatra 2022 with nadaraya–watson kernel and penalized spline. *Inferensi*, 8(3):197–?, 2025. Online ISSN: 2721-3862.

[7] A. M. Kahar, Suparti, and A. R. Hakim. Analisis indeks harga saham gabungan dan faktor pengaruhnya menggunakan pemodelan regresi semiparametrik kernel dilengkapi gui-r. *Jurnal Gaussian*, 12(1):30–41, 2023.

[8] E. Era, S.W. Rizki, and Y. Yundari. Pemodelan indeks harga saham gabungan dengan model regresi semiparametrik kernel. *Equator: Journal of Mathematical and Statistical Sciences*, 1(1), 2022.

[9] C. M. Douglas, E.A. Peck, and G. G. Vining. *Introduction to Linear Regression Analysis*. John Wiley & Sons, Hoboken, NJ, 6 edition, 2021.

[10] R. J. Freund and W. J. Wilson. *Statistical Methods*. Academic Press, San Diego, 2nd edition, 2003. An imprint of Elsevier Science.

[11] D. N. Gujarati. *Econometrics by Example*. Palgrave Macmillan, London, second edition, 2015.

[12] R. L. Eubank. *Nonparametric Regression and Spline Smoothing*. Marcel Dekker, New York, second edition, 1999.

[13] Y. Wang. *Smoothing Splines: Methods and Applications*. CRC Press, Boca Raton, 2011.

[14] T. H. Ali. Modification of the adaptive nadaraya–watson kernel method for nonparametric regression. *Communications in Statistics – Simulation and Computation*, 51(2):391–403, 2019.

[15] T. H. Ali, H. A. Hayawi, and D.S.I. Botani. Estimation of the bandwidth parameter in nadaraya–watson kernel nonparametric regression based on universal threshold level. *Communications in Statistics – Simulation and Computation*, 52(4):1476–1489, 2021.

[16] A. Guidoum. Kernel estimator and bandwidth selection for density and its derivatives: The kedd package. *arXiv preprint arXiv:2012.06102*, 3(1):1–22, 2020.

[17] I. Sriliana, I. Budiantara, and V. Ratnasari. The performance of mixed truncated spline–local linear nonparametric regression model for longitudinal data. *MethodsX*, 12:102652, 2024.

[18] W. Hardle and D. Barry. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, 1994.

[19] Suparti, R. Santoso, A. Prahutama, et al. *Regresi Nonparametrik*. Wade Group, Ponorogo, 1 edition, 2017.

[20] C.R. Oktarina, I. Sriliana, and S. Nugroho. Penalized spline semiparametric regression for bivariate response in modeling macro poverty indicators. *Indonesian Journal of Applied Statistics*, 8(1):13–23, 2025.

[21] W. Alwi, A. Sauddin, and N. M. Islamiah. Faktor-faktor yang mempengaruhi angka harapan hidup di sulawesi selatan menggunakan analisis regresi. *Jurnal Matematika dan Statistika serta Aplikasinya*, 11(1):72–80, 2023.

[22] A. Tanadjaja, I. Zain, and W. Wibowo. Pemodelan angka harapan hidup di papua dengan pendekatan geographically weighted regression. *Jurnal Sains dan Seni ITS*, 6(1):D82–D87, 2017.

[23] N. Amalia and Mahmudah. Faktor yang mempengaruhi angka harapan hidup di provinsi jawa timur tahun 2014 dengan melihat nilai statistik Cp mallows. *Jurnal Wiyata*, 7(1):13–14, 2020.