Volume. 7 No 1 June (2024)

e-ISSN: **2615-3270**
p-ISSN: **2615-3599**

Eigen Mathematics Journal

Homepage jurnal: https://eigen.unram.ac.id/index.php/eigen

Modeling of Economic Growth Rate in West Nusa Tenggara Province with Longitudinal Kernel Nonparametric Regression

***Muhammad Rizaldi* [a], *Nurul Fitriyani* [b], *Zulhan Widya Baskara* [,*]**

[a.] Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Mataram, Indonesia. Email: mrizaldi787@gmail.com

[b.] Department of Statistics, Faculty of Mathematics and Natural Sciences, University of Mataram, Indonesia. Email: nurul.fitriyani@unram.ac.id

[c.] Department of Statistics, Faculty of Mathematics and Natural Sciences, University of Mataram, Indonesia. Email: zulhan.wb@unram.ac.id

ABSTRACT

Economic growth can indicate the success of economic development in people's lives, so it is essential to study the relationship between economic growth and factors that affect economic growth. Regression analysis is one of the most widely used statistical data analysis methods to determine the relationship pattern between the independent and dependent variables. Three methods can be used to estimate the regression curve, one of which is nonparametric regression. Economic growth data is one form of longitudinal data, with observations of independent subjects, with each subject being observed repeatedly over different periods. Kernel nonparametric regression model applications can be used for longitudinal data. This research aims to estimate the curve and get the best regression model. In this research, the smoothing technique chosen to estimate the nonparametric regression model for longitudinal data is the kernel triangle estimator, which can be obtained by minimizing the square of error using Weighted Least Squares (WLS) and selecting the optimum bandwidth using the Generalized Cross Validation (GCV) method. This study uses the economic growth rate in West Nusa Tenggara as the dependent variable and the human development index, population density, general allocation funds, local revenue, and labor force participation as independent variables. The result showed that the model is less accurate because of the low value of the coefficient for determination and the high value of the mean absolute percentage error (MAPE). This can be caused by the selection of bandwidth intervals that are too small.

Keywords: Generalized Cross Validation, Kernel Estimation, Longitudinal Data, Nonparametric Regression, Rate of Economic Growth

Submitted: 22-10-2023; Accepted: 14-06-2024

Doi: https://doi.org/10.29303/emj.v7i1.188

1. Introduction

Regression analysis is one of the most widely used statistical data analysis methods to determine the relationship pattern between the independent variable and the dependent variable. According to Hardle (1994), three approaches can be used to estimate the regression curve: parametric, semiparametric, and nonparametric approaches. In the parametric approach, the relationship between the variables is known or estimated from the shape of the regression curve; for example, it is assumed to form a linear, quadratic, exponential and polynomial pattern. While the semiparametric approach is used if the relationship pattern between a set of independent variables and the dependent variable is known and some are not, the shape of the regression curve. Suppose there is no information whatsoever on the form of the function and it does not meet the assumptions of normality and homogeneity of the variance of the data error. In that case, a nonparametric approach is used.

A nonparametric approach is a regression approach suitable for data whose curve shape is unknown, thus providing great flexibility (Budiantara, 2009). According to Sukarsa and Srinadi (2012), the estimation of the nonparametric regression function is based on data using smoothing techniques such as histograms, kernel estimators, k-nearest neighbours, orthogonal series, spline estimators, Fourier series, and wavelets. Each of these techniques has advantages in parameter estimation. One of the most widely used methods of estimating nonparametric regression parameters is the kernel approach, which has a flexible form, easy-to-adjust mathematical calculations, and relatively fast average convergence. The kernel approach function can also be used as an alternative to solving fluctuating data because kernel nonparametric regression does not require particular assumptions to be met (Wolberg, 2000).

Several kernel functions can approximate data distribution patterns, including uniform, triangle, epanechnikov, gaussian, quadratic, and cosine kernels. Commonly used kernel functions are the Gaussian, epanechnikov, and triangle kernels. Triangle kernels are often used because they are easier and faster in calculations and accurate in modeling fluctuating data (Sukarsa & Srinadi, 2012). According to Puspitasari et al. (2012), the triangle kernel function has a smaller MSE value than others, so the model obtained is better.

In the regression analysis, a method is needed to estimate parameters so that the estimate is a Best Linear Unbiased Estimator (BLUE), one of which is the Weighted Least Squares (WLS) method. The WLS method is very good at overcoming heteroscedasticity (Arifin, 2018). WLS can maintain the efficiency of its estimator without losing its unbiased and consistent properties.

Kernel nonparametric regression model applications can be used for longitudinal data. Longitudinal data is observations of n independent subjects, each observed repeatedly over time (Liang and Zeger, 1986). One form of longitudinal data is economic growth data. According to the Central Bureau of Statistics (2014), economic growth is a process of continuous growth in a country's economic conditions towards better conditions over a certain period. Economic growth can indicate the success of economic development in people's lives, so it is very important to research economic growth.

Former studies have utilized nonparametric kernel regression to model time series data, such as in modeling climate data in Lombok Island by autoregressive pre-whitening. The approach was also used in the statistical downscaling model in several data (Hadijati et al., 2016a, 2016b, 2017, 2021, 2022). Other researchers also used the kernel approach, such as modeling crude birth rate and malnourished children in West Nusa Tenggara Province (Pratiwi et al., 2020, Sauri et al., 2020), modeling hotel tax revenue, and forecasting local original income in Central Lombok (Pembargi et al., 2023a, 2023b). Research on economic growth modeling with longitudinal kernel nonparametric regression, especially in West Nusa Tenggara Province, has not been widely explored.

Based on the description that has been explained previously, in this study, an estimation of the nonparametric regression curve of the kernel function was carried out using the Weighted Least Squares (WLS) method. Then, the results of this curve estimation will be applied to longitudinal data with a case study of the economic growth rate of West Nusa Tenggara Province (NTB) 2016-2020.

2. Method

The data used in this study is longitudinal, namely the economic growth rate of districts/cities in the Province of NTB in 2016-2020. This data is secondary data sourced from the district/city Central Statistics Agency (BPS) publications in NTB and the Indonesian Ministry of Finance. In this study, the data is divided into dependent variables and several independent variables. These variables are presented in Table 1 as follows:

Table 1. The research variables used

Variable	Information	Unit
y	Economic Growth Rate of Regency/City in West Nusa Tenggara	$percent$ (%)
x_1	Human Development Index (HDI)	$percent$ (%)
x_2	Population Index	$peoples/km^2$
x_3	General Allocation Fund	$billion\ Rp$
x_4	Original Local Government Revenue	$billion\ Rp$
x_5	Labor Force Participation Rate	$percent$ (%)

This research was conducted in the following stages:

a. a literature study;
b. collecting data;
c. determining the estimated curve;
d. identifying data patterns;
e. performing a multicollinearity test, detected via Variance Inflation Factor (VIF):
$$VIF_j = \frac{1}{1 - {R_j}^2}$$
${R_j}^2$ is the coefficient of determination between j-th independent variable and other independent variables;
f. determining the optimal bandwidth by utilizing Generalized Cross Validation (GCV) measures and the best model;
g. performing residual assumption test, i.e. constant variance and independence with Durbin-Watson test:
$$dW = \frac{\sum_{i=1}^{m} \sum_{t=2}^{n} (\varepsilon_{it} - \varepsilon_{it-1})^2}{\sum_{i=1}^{m} \sum_{t=1}^{n} {\varepsilon_{it}}^2}$$
ε_{it} is the i-th residual data during time t;
h. testing the goodness of the model and calculating the accuracy of the prediction with,
$$MAPE = \frac{1}{mn} \sum_{i=1}^{10} \sum_{t=1}^{5} \left| \frac{y_{it} - \hat{y}_{it}}{y_{it}} \right| \times 100\%$$
y_{it} represents the i-th actual data during time t and $\hat{y}_{it}$ represents the i-th predicted data during time t; and
i. drawing conclusions.

3. Results

The results and research begin with the estimation of the nonparametric kernel triangle regression curve to obtain the following equation:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T W}(\boldsymbol{X_{it}})\boldsymbol{X})^{-1}\boldsymbol{X^T W}(\boldsymbol{X_{it}})\boldsymbol{Y}$$

Then, the resulting estimation model is:

$$\widehat{\boldsymbol{Y}} = [1]\widehat{\boldsymbol{\beta}}$$

and $(\boldsymbol{X^T W}(\boldsymbol{X_{it}})\boldsymbol{X})^{-1}\boldsymbol{X^T W}(\boldsymbol{X_{it}}) = \widehat{\boldsymbol{H}}$.

Then, data patterns were identified using a scatterplot to see the relationship between the dependent variable and each independent variable. The following scatterplot was obtained:

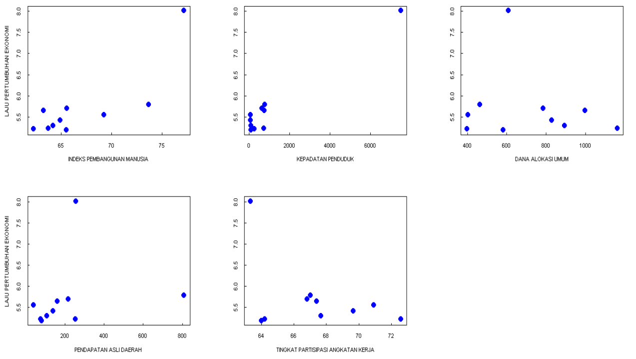

Figure 2. *Scatterplot* of the Economic Growth Rate variable with other independent variables in 2016

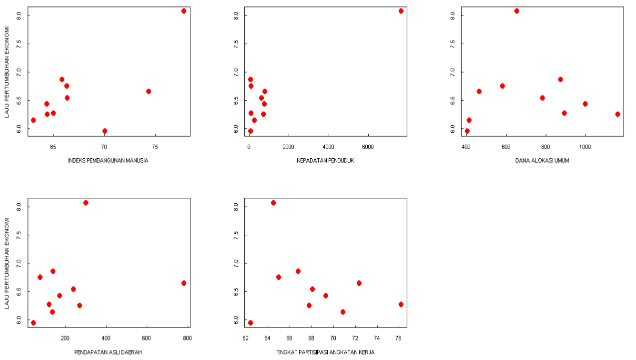

Figure 3. *Scatterplot* of the Economic Growth Rate variable with other independent variables in 2017

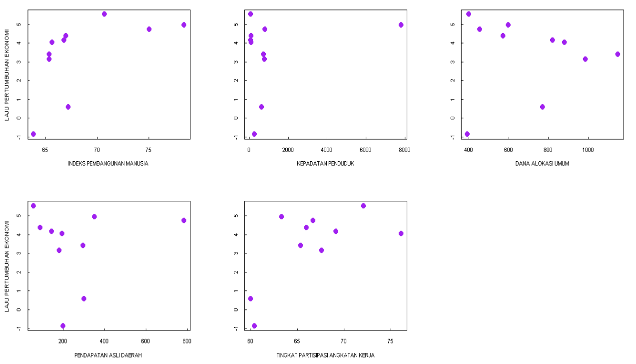

Figure 4. *Scatterplot* of the Economic Growth Rate variable with other independent variables in 2018

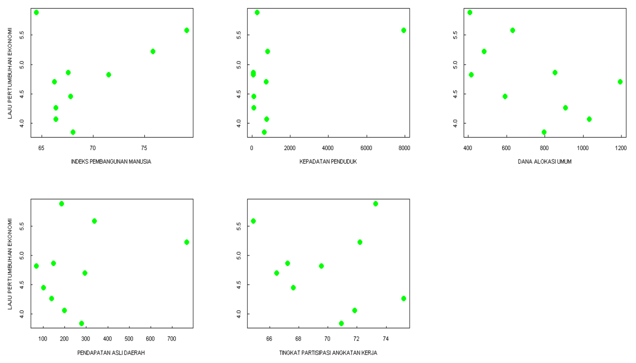

Figure 5 *Scatterplot* of the Economic Growth Rate variable with other independent variables in 2019

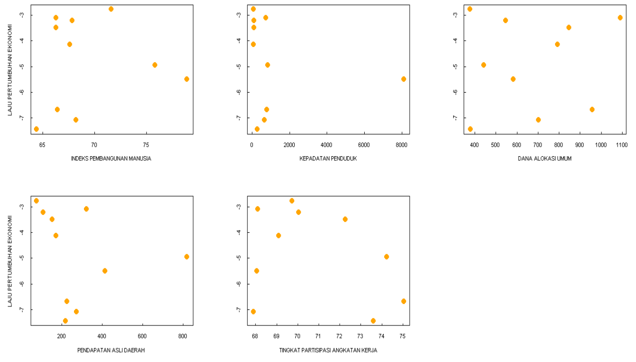

Figure 6. *Scatterplot* of the Economic Growth Rate variable with other independent variables in 2020

Based on Figures 3.1 to 3.5 above, it can be seen that the pattern of relationship between the dependent variable Economic Growth Rate and each independent variable Human Development Index, Population Index, General Allocation Fund Original Local Government Revenue, and Labor Force Participation Rate does not follow a particular pattern, so the estimated model used is nonparametric regression (Hardle, 1994; Budiantara, 2009). Then, a multicollinearity test is carried out to find out the variables that will be used in the next stage.

The following values are obtained using the Variance Inflation Factor (VIF).

Table 2. Multicollinearity test

Variables	VIF
x_1	2,096
x_2	1,432
x_3	1,144
x_4	1,973
x_5	1,306

Based on the table above, it can be seen that the VIF value for each independent variable is less than 10. This indicates that there is no multicollinearity between the independent variables. Based on the multicollinearity test, this research can be continued using all independent variables, which are x_1, x_2, x_3, x_4, and x_5. Modeling data with nonparametric kernel regression begins with selecting the optimal bandwidth using a genetic algorithm. In this case, this algorithm is used to help minimize the Generalized Cross-Validation (GCV) value to determine the optimal bandwidth. The experimental results show that the minimum GCV value is 41.8016269, and each bandwidth optimum value $h_1 = 0{,}6139987$; $h_2 = 1{,}0184682$; $h_3 = 1{,}0135395$; $h_4 = 0{,}9359553$; and $h_5 = 0{,}8486045$.

After obtaining the bandwidth value, the estimated value of the model is determined using the following Nadaraya-Watson equation below.

$$\hat{y}_{it} = \frac{\sum_{i=1}^{10}\sum_{t=1}^{5}\left(\prod_{j=1}^{5}\frac{1}{h_j}\left(1-\left|\frac{x_{itj}-x_j}{h_j}\right|\right)\right)y_{it}}{\sum_{i=1}^{10}\sum_{t=1}^{5}\left(\prod_{j=1}^{5}\frac{1}{h_j}\left(1-\left|\frac{x_{itj}-x_j}{h_j}\right|\right)\right)}$$

The estimated data value with optimal bandwidth is presented in the following table.

Table 3 Estimated values

y_{it}	y_{it}	$\hat{y}_{it}$
y_{11}	5,7	3,187096166
y_{12}	6,54	6,106168006
y_{13}	0,57	0,209718408
y_{14}	3,84	2,893048101
y_{15}	−7,08	4,824901532
y_{21}	5,65	6,030458722
y_{22}	6,43	5,510353137
y_{23}	3,14	6,453493918
y_{24}	4,06	3,776216692
y_{25}	−6,68	2,682989515
y_{31}	5,23	−2,026033837
y_{32}	6,25	6,116198542
y_{33}	3,4	0,212748454
y_{34}	4,7	2,719763586
y_{35}	−3,1	7,136935694
y_{41}	5,42	4,43769073
y_{42}	6,86	2,86354171
y_{43}	4,16	4,627242318
y_{44}	4,86	3,67840117
y_{45}	−4,13	4,359244727
y_{51}	5,19	0,896979987
y_{52}	6,75	0,633664417
y_{53}	4,38	1,821277732
y_{54}	4,45	2,699855995
y_{55}	−3,21	5,289798474
y_{61}	5,3	4,930364092
y_{62}	6,27	3,608907095
y_{63}	4,04	3,192894253
y_{64}	4,26	3,664153916
y_{65}	−3,49	3,828674127
y_{71}	5,55	5,06736775
y_{72}	5,95	0,375863749
y_{73}	5,53	4,625053664
y_{74}	4,82	5,133887662
y_{75}	−2,77	5,028783776
y_{81}	5,22	4,715315836
y_{82}	6,14	4,988681054
y_{83}	−0,87	0,770374512
y_{84}	5,88	4,170747807
y_{85}	−7,44	3,702716386
y_{91}	8,01	3,692315209
y_{92}	8,07	3,336256749
y_{93}	4,95	3,276428194
y_{94}	5,58	3,098038246
y_{95}	−5,5	2,687686937
y_{101}	5,79	4,149805474
y_{102}	6,65	5,040530824
y_{103}	4,74	3,602288439
y_{104}	5,22	4,658927433
y_{105}	−4,95	4,388688330

Furthermore, the residual assumption test is carried out. The residual assumption test was carried out to test the feasibility of the model obtained. A model is feasible if it meets normal, identical, and independent assumptions. Because it is assumed that the residuals are normally distributed in nonparametric kernel regression, an identical and independent assumption test will be carried out.

The results of the identical assumption test can be seen in the following figure.

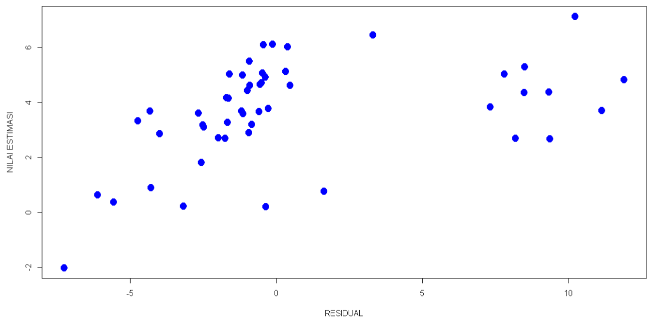

Figure 7. *Scatter plot* ε and $\hat{y}$

Based on the figure above, it can be seen that the data spreads in all directions and does not form a pattern. This shows that there is no heteroscedasticity, meaning that the identical assumption is met.

Then, the Durbin-Watson equation is used to test the independent assumptions. The following are the results obtained:

$$dW = 2{,}250155945$$

From the calculation above, a dW value of 2.250155945 is obtained, and the dL and dU values are known respectively based on the Durbin-Watson table with 50 data and 5 independent variables, 1.3346 and 1.7708. Because the value of dW is greater than dU, the fourth option in Table 3.2 is used for decision-making, namely $4 - dU < dW < 4 - dL$. Based on this, the dW value is between the dL and dU values, namely $2.2292 < 2.250155945 < 2.6654$. This shows that the dW value is between the dU and dL values. So there is no autocorrelation between residuals or the model residuals fulfill the independent assumptions.

The coefficient of determination (R^2) and Mean Absolute Percentage Error (MAPE) can be used to measure the feasibility of the model that has been obtained. After doing the calculations, the R^2 value is 0.192045. The independent variables can explain the dependent variable by 19.2%, and other variables explain the rest. This is supported by the MAPE values obtained for each location, which are as follows:

$MAPE_1 = 61{,}35\%$ $MAPE_6 = 58{,}81\%$
$MAPE_2 = 54{,}74\%$ $MAPE_7 = 81{,}36\%$
$MAPE_3 = 81{,}39\%$ $MAPE_8 = 58{,}81\%$
$MAPE_4 = 63{,}49\%$ $MAPE_9 = 79{,}16\%$
$MAPE_5 = 87{,}17\%$ $MAPE_{10} = 67{,}94\%$

Based on the calculation results above, the MAPE value is obtained for more than 50% of each location, meaning that the model obtained is less accurate. So it can be concluded that the estimation model obtained is less suitable for forecasting or prediction. This can be caused by the selection of bandwidth intervals that are too small and narrow. These results can be a basis for consideration for further research, especially in selecting bandwidth intervals.

4. Conclusion

Based on the analysis that has been done, the nonparametric kernel regression model is obtained as follows:

$$\hat{y} = \frac{\sum_{i=1}^{10}\sum_{t=1}^{5}\left(\prod_{j=1}^{5}\frac{1}{h_j}\left(1-\left|\frac{x_{itj}-x_j}{h_j}\right|\right)\right)y_{it}}{\sum_{i=1}^{10}\sum_{t=1}^{5}\left(\prod_{j=1}^{5}\frac{1}{h_j}\left(\left(1-\left|\frac{x_{itj}-x_j}{h_j}\right|\right)\right)\right)}$$

with h_j, $j = 1,2,3,4,5$ and each value $h_1 = 0{,}6139987$; $h_2 = 1{,}0184682$; $h_3 = 1{,}0135395$; $h_4 = 0{,}9359553$; and $h_5 = 0{,}8486045$.

From this model, R^2 value of 19.2% and a MAPE of more than 50% were obtained for each observation location, meaning that the model obtained was less accurate for the data, which can be caused by the selection of bandwidth intervals that are too narrow. It is hoped that these results can be a basis for consideration for further research, especially in selecting bandwidth intervals.

REFERENCES

Anisa, N., Debataraja, N. N., and Martha, S., 2019, Estimating Kernel Nonparametric Regression Models using the Nadaraya-Watson Estimator, *Bimaster Journal*, 4(08), 633-638.

Arifin, M. K., 2018, Analysis of Old School Expectancy Rates in Eastern Indonesia Using Weighted Least Square Regression, *Journal of Mathematics* "MANTIK", 1(04), 32-41.

Budiantara, I. N., 2009, *Spline in Nonparametric and Semiparametric Regression: A Statistical Modeling of the Present and the Future*, ITS Press, Surabaya.

Central Statistics Agency, 2014, *District/City Government Financial Statistics*, BPS, Jakarta.

Hadijati, M., Irwansyah, & Fitriyani, N., 2021, Prediction of Daily Rainfall in Dodokan Watershed Based on Statistical Downscaling Model: An Effort to Manage Watershed Ecosystems. *Eastern Journal of Agricultural and Biological Sciences* (EJABS), 1(2), 1–7.

Hadijati, M., Irwansyah, & Fitriyani, N., 2022, Autoregressive Prewhitening on the Nonparametric Regression Model of Water Discharge in the Jangkok Watershed, Lombok Island. *Global Journal of Pure and Applied Mathematics*, 18(1), 307–318.

Hadijati, M., Komalasari, D. and Fitriyani, N., 2016a, Statistical Downscaling Model Using Nonparametric Regression to Predict Temperature in Selaparang Lombok. *Proceeding of the First International Conference on Science and Technology*, 129-132. Universitas Mataram.

Hadijati, M., Komalasari, D., & Fitriyani, N., 2017, Simulation of Monthly Rainfall Data of Dodokan Watershed Using Nonparametric Statistical Downscaling Model. *Proceeding of the Second International Conference on Science and Technology*, 213–219.

Hadijati, M., Komalasari, D., and Fitriyani, N., 2016b, Statistical Downscaling Regresi Nonparametrik Kernel Untuk Prediksi Curah Hujan Bulanan Stasiun Sembalun (translate: Statistical Downscaling Kernel Nonparametric Regression For Prediction of Monthly Rainfall Sembalun Station). *Proceeding of Seminar Nasional Matematika II*, 186-196. Universitas Udayana.

Halim, S. and Bisono, I., 2012, Kernel Functions in Nonparametric Regression Methods and Their Applications to Priest River Experimental Forest's Data, *Industrial Engineering Journal*, 8(1), 73-81.

Hardle, W., 1994, *Applied Nonparametric Regression*, Cambridge University Press, New York.

Liang, K. Y. and Zeger, S. L., 1986, Longitudinal Data Analysis using Generalized Linear Models, *Biometrics Journal*, 1(73), 13-22.

Maulidia, M. J., Budiantara, I. N., and Purnomo, J. D. T., 2019, Nonparametric Regression Curve Estimation using Mixed Spline Truncated and Kernel Estimator for Longitudinal Data, *AIP Conference Proceedings*, 2194, 020063-1-020063-8.

Pembargi, J. A., Hadijati, M., & Fitriyani, N., 2023a, Kernel Nonparametric Regression for Forecasting Local Original Income. *Jurnal Varian*, 6(2), 119-126.

Pembargi, J. A., Setiawana, E., Wahidatussolihah, R., & Fitriyani, N., 2023b, Modeling the Hotel Tax Revenue in Central Lombok using Nonparametric Regression. *Jurnal Matematika, Statistika Dan Komputasi*, 19(3), 498-505.

Pratiwi, D., Mursy, L.A.A., Rizaldi, M., and Fitriyani, F., 2020, Regresi Nonparametrik Kernel Gaussian pada Pemodelan Angka Kelahiran Kasar di Provinsi Nusa Tenggara Barat (translate: Nonparametric Gaussian Kernel Regression in Modeling Crude Birth Rate in West Nusa Tenggara Province). *Eigen Mathematics Journal*, 3(2), 100-105.

Puspitasari, I., Suparti, and Wilandari, Y., 2012, Analysis of the Composite Stock Price Index (IHSG) Using the Kernel Regression Model, *Gaussian Journal*, 1(1), 93-102.

Sauri, M. S., Hadijati, M., & Fitriyani, N., 2021, Spline and Kernel Mixed Nonparametric Regression for Malnourished Children Model in West Nusa Tenggara. *Jurnal Varian*, 4(2), 99–108.

Sukarsa, I. K. G. and Srinadi, I. G. A. M., 2012, Kernel Estimator in Nonparametric Regression Models, *Journal of Mathematics* 2(1), 19-30.

Wu, H. and Zhang, J. T., 2006, *Nonparametric Regression Methods for Longitudinal Data Analysis*, John Willey & Sons, New York.