Andi_Nurhayati_Latif.pdf

by

The Implementation of Hierarchical Linear Spline Regression Model to Maternal Mortality Rate Data in Indonesia

A.Nurhayati Latief[a], Anna Islamiyati[a], Nurtiti Sunusi[a], Hedi Kuswanto[a]

[a]Universitas Hasanuddin, Jl. Perintis Kemerdekaan Km 10, Makassar 90245, Indonesia

Abstract

Hierarchical spline nonparametric regression analysis is one regression method to identify the correlation pattern among the variables with unidentified curves and nested data structure using a spline truncated regression approach on its function.

On the maternal mortality rate, there is nested data structure between the regional/ city level and provincial level. This study aimed to identify the correlation between the response variable and predictor variable from the maternal mortality rate data in 2020. There are two analyzed levels consisting of the city/ regional level (level-1) with 256 cities/ regencies and the provincial level (level-2). The level-1 predictor variable is total pregnant women taking blood booster supplements and total pregnant women attending K4 antenatal care and the level-2 predictor variable is poverty rates and school enrollment rates. A spline truncated approach has high flexibility and this study employs Maximum Likelihood as its estimator.

The finding suggested that if a province has a lower poverty rate than 7.24%, maternal mortality rates tend to increase. If a province has lower enrolment rates than 67.8 %, the maternal mortality rate tends to decrease. If a regency or a city has fewer pregnant women taking blood booster supplements than 2258, maternal mortality rates tend to decrease. If a regency or a city has fewer pregnant women attending K4 antenatal care than 2258, the maternal mortality rate tends to decrease.

Keywords: Hierarchical Model, Maximum Likelihood, Spline Truncated, Maternal Mortality Rate;

1. Introduction

Regression analysis is one method for analyzing and modeling correlation patterns between the dependent variables and one or more predictor variables. (Montgomery et al., 2012) Some regression analysis method includes parametric regression analysis and non-parametric regression analysis. On the parametric regression, there is a rigid and solid assumption that regression curves can be identified as, linear, quadratic, cubic, degree of the polynomial, exponential, and so on. Meanwhile, if the regression curve is unknown, the non-parametric regression analysis will be applied. The method of the non-parametric regression model is highly flexible and objective because they are not attached to the researcher's assumption and subjectivity. (Budiantara, 2011)

Some of the commonly used non-parametric regression models are kernel smoothing (Bowman et al., 1998), spline regression (Lawrence, 2021), smoothing spline (Wang, 1998), fourier series, and local polynomials. One of the commonly used non-parametric regression models is spline regression, a spline is a special function defined piecewise by segmented polynomials joined at points called knots. Spline is a local model with local compatibility between knots enabling us to estimate the function and predict the data's functional form (Lawrence, 2021). Spline is commonly used due to its advantages, such as an effective and detailed statistical and visual interpretation, high flexibility, ability to handle data/function smoothly, ability to control unstable data at a certain sub-interval, and good capacity to be generalized in complex and complicated statistical modeling. (Budiantara, 2011) Spline regression in research has been extensively performed in cubic spline (Islamiyati, 2019), penalized spline (Islamiyati et al., 2019), spline truncated (Rahim et al., 2019), (Kuswanto H. et al), and smoothing spline.

In the expansion of non-parametric regression, the researchers have developed a hierarchical regression spline model. Hierarchical modeling is a statistical model to analyze data on a different level (Groenewegen, n.d.). Hox (2002) reported that the term 'multilevel' refers to the hierarchical or nested data structure, for example, individuals in an organization, repeated action by individuals in a cluster on a cluster sampling. A data set can be said as structured hierarchically when there is lower-level research nested in the higher-level research (Alkharusi, 2011). Problems due to interpretation errors are highly possible to occur when ignoring structure in a data set. (Hox, 2002) (Goldstein, 2007)

In its development, research concerning the spline multilevel model has been extensively implemented (He et al., 2017; Howe et al., 2016; Macdonald-Wallis et al., 2012; Tilling et al., 2014). From the research, it was identified that the splines provide better interpretation in analyzing data with a hierarchical structure. In Indonesia, a research model with a multilevel spline has been extensively performed. Kuswanto, et al. (2022) researched a multilevel spline on National Exam Scores using the restricted maximum likelihood method.

Maximum likelihood has a number of preferred statistical properties, including consistent, asymptotic (approaching a value or curve arbitrarily closely), efficient (variance is lower than other predictors), and invariance parameterization (prediction does not vary when measurement or parameter vary in a permissible method). Maximum Likelihood is applied extensively in all statistical applications, including categorical data analysis, logistic regression, and structural equation modeling. (Psy et al., 2019) Generally, The Maximum Likelihood estimator have better statistical properties than the ordinary least square estimator. The maximum likelihood estimator is unbiased and has minimum variance compared to other unbiased estimator. Maximum likelihood is a consistent estimator. (Montgomery et al., 2012).

The application of the hierarchical spline model will be shown in the Maternal Mortality Rate Data of 2020 in Indonesia. The maternal mortality rate is the number of maternal deaths resulting from pregnancy, childbirth and postpartum which is used as an indicator of women's health status. The maternal mortality rate (MMR) is also one of the main indicators of the success of the maternal health program. In addition, this indicator is also able to assess the degree of public health, because of its sensitivity to improve the health services, both in terms of accessibility and quality (Ministry of Health RI, 2020). There have been many studies on truncated spline nonparametric regression modelling of maternal mortality rates. Putri (2018) used truncated spline regression to model new postpartum and post-miscarriage family planning participants in East Java, Arfan (2014) who examined factors that influenced maternal mortality in East Java with linear spline regression. In addition, Rahim, et al (2019) examined the factors that influence maternal mortality using truncated spline nonparametric regression model.

This study aims to identify the application of a hierarchical spline regression model using maximum likelihood as the estimator. This model is applied to the Indonesian maternal mortality rate in 2020. This study improves understanding about the hierarchical spline regression model and the application of maximum likelihood as the estimator.

2. Methodology

2.1. Data Sources

City/ regional level predictor variable level was collected from the Ministry of Health Database with total samples of 256 regencies/ cities. The data also included pregnant women taking blood booster supplements (X_1) and pregnant women attending K4 antenatal care (X_2). Predictor variable data on the provincial level presenting 16 provinces including the poverty rates (Z_1) and school enrollment rates (Z_2). The employed response variable in this study is the maternal mortality rate.

2.2. Research Method

The steps of data analysis carried out in this study are as follows:

1. Modelling Maternal Mortality Rate using Hierarchical Spline Linear Truncated Estimator on the city/ regional level with knot point.
2. Estimating parameters using Hierarchical Spline Linear Truncated Estimator on the city/ regional level with knot point. Hierarchical Spline Linear Truncated Estimator Model

   $$\widehat{Y_{ij}} = \widehat{\beta_{0j}} + \widehat{\beta_{1j1}} x_{1ij} + \widehat{\beta_{1jp}}(x_{1ij} - k_{1h})_+ + \widehat{\beta_{2j1}} x_{2ij} + \widehat{\beta_{2jp}}(x_{2ij} - k_{2h})_+ + e_{ij}$$

3. Estimating parameters using Hierarchical Spline Linear Truncated Estimator on the provincial level with Maximum Likelihood. Hierarchical Spline Linear Truncated Estimator Model

   $$\widehat{\beta_{0j}} = \widehat{\gamma_{01}} + \widehat{\gamma_{011}} Z_{1j} + \widehat{\gamma_{01p}}(Z_{1j} - k_{4h})_+ + \widehat{\gamma_{021}} Z_{2j} + \widehat{\gamma_{02p}}(Z_{2j} - k_{4h})_+ + u_{0j}$$

   From the regional, city, and, provincial levels, a mixed model was obtained.

   $$\hat{Y}_{ij} = \Big[\widehat{\gamma_{01}} + \widehat{\gamma_{011}} Z_{1j} + \widehat{\gamma_{01h}}(Z_{1j} - k_{4h})_+ + \widehat{\gamma_{021}} Z_{2j} + \widehat{\gamma_{02h}}(Z_{2j} - k_{4h})_+ + \widehat{\beta_{1j1}} x_{1ij} + \widehat{\beta_{1jh}}(x_{1ij} - k_{1h})_+ + \widehat{\beta_{2j1}} x_{2ij} + \widehat{\beta_{2jh}}(x_{2ij} - k_{2h})_+ + \widehat{\beta_{3j1}} x_{3ij} + \widehat{\beta_{3jh}}(x_{3ij} - k_{3h})_+\Big] + [u_{0j} + e_{ij}]$$

4. Obtaining the most optimal Maternal Mortality Rate model according to minimum GCV value.
5. Performing model interpretation on maternal mortality rate.

3. Results and Discussion

3.1. The Application of Hierarchical Spline Linear Truncated Regression Model

Hierarchical Spline Linear Truncated Regression Model with Maximum Likelihood Method on the predictors is presented as follows, estimation of Hierarchical Spline Linear Truncated Regression Model with one-knot point approach. Selection of optimal knot point was performed by observing minimum CGV from each knot point on the predictor of total pregnant women taking blood booster supplements (X_1), total pregnant women attending K4 antenatal care (X_2), the total poverty rate (Z_1), and the total percentage of enrollment rates (Z_2). Hierarchical spline linear truncated regression model uses maximum likelihood method with one point knot.

Level-1:

$$\widehat{Y_{ij}} = \widehat{\beta_{0j}} + \widehat{\beta_{1j1}} x_{1ij} + \widehat{\beta_{1j2}}(x_{1ij} - k_{11})_+ + \widehat{\beta_{2j1}} x_{2ij} + \widehat{\beta_{2j2}}(x_{2ij} - k_{21})_+ + e_{ij}$$

Level-2:

$$\widehat{\beta_{0i}} = \widehat{\gamma_{01}} + \widehat{\gamma_{011}} Z_{1i} + \widehat{\gamma_{012}}(Z_{1i} - k_{41})_+ + \widehat{\gamma_{021}} Z_{2i} + \widehat{\gamma_{022}}(Z_{2i} - k_{51})_+ + u_{0i}$$

3.2. Minimum GCV value selection

GCV value from one-knot point modelling on the city/ regional level is presented in the following Table 1:

Table 1. GCV value with knot point for regional/ city level predictor

X_1 (k_{11})	X_2 (k_{21})	GCV
2258	2222	3541.25
2258	12768	3658.346
2654	15183	3712.997
2733	18587	3734.307
2911	14606	3754.22
3142	6467	3776.455
3423	11431	3815.348
3497	21576	3851.562
3530	40441	3869.136
3535	3579	3805.868

From Table 1, the minimum GCV value on the city/ regional level is 3541.25 in the knot point $k_{11} = 2258, k_{21} = 2222$.
GCV value from one-knot point modelling on the provincial level is presented in the following Table 2:

Table 2. GCV value with a one-knot point for provincial level predictor

Z_1 (k_{31})	Z_2 (k_{41})	GCV
4.69	72.4	182,6561
4.83	77.6	183,7274
7.24	67.8	146,7373
7.97	83	171,716
8.43	69.4	173,5294
8.43	72.4	171,3909
8.99	71.3	166,1671
9.14	75.9	179,5336
11.46	72.4	169,3347
12.76	78.2	184,2884

From table 2, the minimum GCV value on level 1 is 146.7373 in the knot point $k_{31} = 7.24, k_{41} = 67.8$.

3.3. Significance Testing of Hierarchical Model Parameter

Significance testing of hierarchical spline linear truncated regression model with maximum likelihood method was performed to identify the predictor variables that significantly affect the maternal mortality rate.

Significance Testing of City/Regional Level Parameter

The significant results of the hierarchical spline linear truncated regression model on the city/ regional level are presented in the following Table 3:

Table 3. Significance parameter results on the city/ regional level parameters

Variable	Parameter	Coefficient	t value	p-value	Result
	β_0	374,360	9.919	0.000	Significant
X_1	β_{11}	0.437	3.332	0.001	Significant
	β_{12}	-0.434	-3.305	0.001	Significant
X_2	β_{21}	-0.560	-4.183	0.000	Significant
	β_{22}	0.556	4.145	0.000	Significant

From table 3, it was identified that there are 5 parameters. The decision is that the parameter is significant to the model. Therefore, the predictor variable of the total pregnant women taking blood booster supplements (X_1) and attending K4 antenatal care (X_2) had a significant effect on the maternal mortality rate.

Significance Testing of Provincial level Parameter

Significance results of the hierarchical spline linear truncated regression model on the provincial level are presented in the following Table 4:

Table 4. Significance Parameter on the Provincial Level

Variable	Parameter	Coefficient	t value	p-value	Result
	γ_{01}	9744.714	1.269	0.231	Not Significant
Z_1	γ_{011}	1.542	0.532	0.605	Not Significant
	γ_{012}	-2.856	-0.761	0.463	Not Significant
Z_2	γ_{021}	-143.785	-1.269	0.230	Not Significant
	γ_{022}	143,075	1.261	0.233	Not Significant

From table 4, it was identified that there are 5 parameters. The decision implies significance of the model and therefore, the predictor variable of the poverty rate (Z_1) and the enrollment rates (Z_2) had a significant effect on the maternal mortality rate.

Estimating hierarchical model parameters

From table 4, hierarchical spline linear truncated regression parameters with a one-point knot on the city/ regional level are presented as follows:

$$\widehat{Y_{ij}} = 374.360 + 0.437x_{1ij} - 0.434(x_{1ij} - 2258)_+ - 0.560x_{2ij} + 0.556(x_{2ij} - 2222)_+$$

It was identified that the applied response variable on level 2 is β_{0j}, where j is the province in this study representing 16 provinces. From table 4, the hierarchical spline linear truncated regression parameter with only one knot-point on the provincial level is presented as follows:

$$\widehat{\beta_{0j}} = 9744.714 + 1.542Z_{1ij} - 2.856(Z_{1ij} - 7.24)_+ - 143.785Z_{2ij} + 143.075(Z_{2ij} - 67.8)_+$$

The results of a mixed hierarchical linear model at the city/ regional and provincial levels can be presented as follows:

$$\widehat{Y_{ij}} = 9744.714 + 1.542Z_{1ij} - 2.856(Z_{1ij} - 7.24)_+ - 143.785Z_{2ij} + 143.075(Z_{2ij} - 67.8)_+ + 0.437x_{1ij} - 0.434(x_{1ij} - 2258)_+ - 0.560x_{2ij} + 0.556(x_{2ij} - 2222)_+$$

3.4. Interpretation of Hierarchical Spline Linear Truncated Regression Model

As can be seen from the scatterplot whether the data has a specific sub-interval pattern in each variable:

Interpretation of Pregnant Women Taking Blood Booster Supplements (X_1)

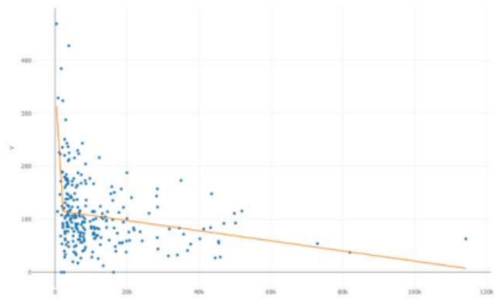

Fig. 1. Estimation Curve of Hierarchical Spline Linear Truncated Regression Model with Sub-Interval Pattern Line of X_1 Predictor Variable and Response Variable

In fig. 1, it was identified that the scatterplot between response variables or maternal mortality rates and X_1 predictor variables or total pregnant women taking blood booster supplements has a pattern change in a particular sub-interval. It was identified from figure 1 at the interval, it was found that in approximately 2222, maternal mortality rates tend to decrease.

Interpretation of Pregnant Women Attending K4 Antenatal Care (X_2)

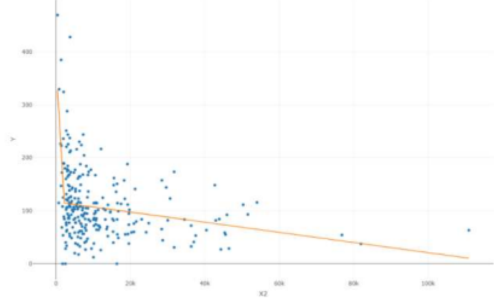

Figure 2. Estimation Curve of Hierarchical Spline Linear Truncated Regression Model with Sub-interval Pattern Line of X_2 Predictor Variable and Response Variable

From figure 2, it was identified that the scatterplot between the response variable or Maternal Mortality Rate and X_2 predictor variable or total pregnant women attending K4 antenatal care has a pattern change at a particular sub-interval. It was identified from figure 2 at the interval of approximately 2222, maternal mortality rates tend to decrease.

Interpretation of Poverty Rates (Z_1)

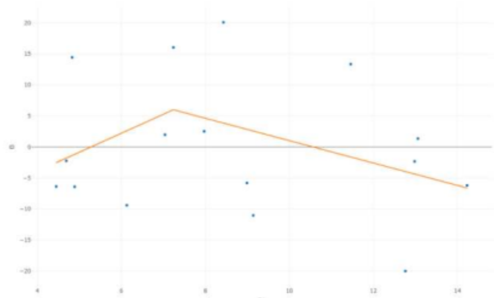

Figure 3 Estimation Curve of Hierarchical Spline Linear Truncated Regression Model with Sub-interval Pattern Line of Z_1 predictor variable and response variable

From figure 3, it was identified that the scatterplot between the response variable or Maternal Mortality Rate and the Z_1 predictor variable or total pregnant women attending K4 antenatal care has a pattern change at a particular sub-interval. From figure 3, at an interval less than 7.24%, the maternal mortality rate has a rising tendency and at an interval more than 7.24%, the maternal mortality rate has a decreasing tendency. Interpretation of Enrollment Rates (Z_2)

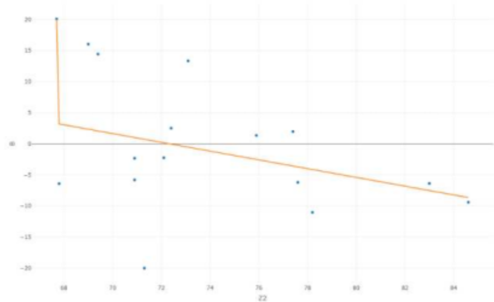

Figure 4 Estimation Curve of Hierarchical Spline Linear Truncated Regression Model with Sub-interval Pattern Line of Z_2 predictor variable and response variable

From figure 4, it was identified that the scatterplot between the response variable or Maternal Mortality Rate and the Z_2 predictor variable or enrolment rates has a pattern change at a particular sub-interval. As can be seen from figure 4, at an interval of approximately 67.8%, the maternal mortality rate pattern tends to decrease.

4. CONCLUSION

Hierarchical Spline Linear Truncated Regression Model with one-knot point is presented as follows:

$$\widehat{Y_{ij}} = 9744.714 + 1.542Z_{1ij} - 2.856(Z_{1ij} - 7.24)_+ - 143.785Z_{2ij} + 143.075(Z_{2ij} - 67.8)_+ + 0.437x_{1ij} - 0.434(x_{1ij} - 2258)_+ - 0.560x_{2ij} + 0.556(x_{2ij} - 2222)_+$$

From the estimation model, it can be concluded that when the province has a percentage of poverty rate less than 7.24, the maternal mortality rate tends to increase. If a province has lower enrolment rates than 67.8 %, the maternal mortality rate tends to decrease. If a regency or a city has fewer pregnant women taking blood booster supplements than 2258, maternal mortality rates tend to decrease. If a regency or a city has fewer pregnant women attending K4 antenatal care than 2258, the maternal mortality rate tends to decrease.

References

Alkharusi, H. (2011). Hierarchical Linear Models: Applications in Educational Assessment Research. Educational Research Journal, 26(1).

Arfan, N., & Budiantara, I. N., 2014. Pendekatan Spline untuk Estimasi Kurva Regresi Nonparametrik (Studi Kasus pada Data Angka Kematian Maternal di Jawa Timur). Jurnal Sains dan Seni Pomits Vol. 3, No. 1, (2014) 2337-3520

Bowman, A. W., Wand, M. P., & Jones, M. C. (1998). Kernel Smoothing. Biometrics, 54(1), 393. https://doi.org/10.2307/2534029

Budiantara, I. N. (2011). Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika Yang Mandiri Dan Berkarakter. Prosiding Seminar Nasional FMIPA Undiksha, 9–28.

Goldstein, H. (2007). Handbook of Multilevel Analysis | Springer. 504. http://www.springer.com/br/book/9780387731834

Groenewegen, P. P. (n.d.). Multilevel Modeling for Public Health.

He, L., Lu, C. T., Ding, H., Wang, S., Shen, L., Yu, P. S., & Ragin, A. B. (2017). Multi-way multi-level Kernel modeling for neuroimaging classification. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-Janua, 6846–6854. https://doi.org/10.1109/CVPR.2017.724

Howe, L. D., Tilling, K., Matijasevich, A., Petherick, E. S., Santos, A. C., Fairley, L., Wright, J., Santos, I. S., Barros, A. J. D., Martin, R. M., Kramer, M. S., Bogdanovich, N., Matush, L., Barros, H., & Lawlor, D. A. (2016). Linear spline multilevel models for summarising childhood growth trajectories: A guide to their application using examples from five birth cohorts. Statistical Methods in Medical Research, 25(5), 1854–1874. https://doi.org/10.1177/0962280213503925

Hox, J. (2002). Multilevel Analysis Techniques and Applications. Lawrence Erlbaum Associates, Publishers.
Islamiyati, A. (2019). Model Spline Kubik Dengan Titik-Titik Knots Dalam Regresi Nonparametrik. http://dx.doi.org/10.31227/osf.io/z3dte
Islamiyati, A., Raupong, & Anisa. (2019). Use of penalized spline linear to identify change in pattern of blood sugar based on the weight of diabetes patients. International Journal of Academic and Applied Research, 3(12), 75–78.
Kemenkes, 2020. Profil Kesehatan Indonesia Tahun 2020, Jakarta: Pusat Data dan Informasi. Diakses dari https://pusdatin.kemkes.go.id/
Kuswanto, H., Islamiyati, A. & and Ilyas, N., 2022. Multilevel Nonparametrik Regression Model with Truncated Linear Spline Estimator on Students' National Examination Scores. International Journal of Academic and Applied Research (IJAAR).
Lawrence, C. H. (2021). Th E Q Uest Fo R Th E Prim Itive. In Medieval Monasticism. https://doi.org/10.4324/9781315838588-16
Macdonald-Wallis, C., Lawlor, D. A., Palmer, T., & Tilling, K. (2012). Multivariate multilevel spline models for parallel growth processes: Application to weight and mean arterial pressure in pregnancy. Statistics in Medicine, 31(26), 3147–3164. https://doi.org/10.1002/sim.5385
Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis (Fifth Edit). A JOHN WILEY & SONS INC., PUBLICATION.
Psy, N., Regression, M., & York, N. (2019). The idea is similar to the idea of ordinary least squares (OLS) in regression in which the squared errors or residuals are minimized to obtain the best fit of the regression line to the data and the regression coefficients. 1–3.
Putri, I.M.E., Budiantara, I. N., 2018. Regresi Nonparametrik Spline Truncated untuk Memodelkan Persentase Peserta KB Baru Pasca Persalinan dan Pasca Keguguran di Jawa Timut Tahun 2016. Surabaya : Institut Teknologi Sepuluh Nopember
Rahim, F., Budiantara, I. N., & Permatasari, E. O. (2019). Spline Truncated Nonparametric Regression Modeling for Maternal Mortality Rate in East Java. Inferensi, 2(1), 39. https://doi.org/10.12962/j27213862.v2i1.6812
Tilling, K., MacDonald-Wallis, C., Lawlor, D. A., Hughes, R. A., & Howe, L. D. (2014). Modelling childhood growth using fractional polynomials and linear splines. Annals of Nutrition and Metabolism, 65, 129–138. https://doi.org/10.1159/000362695
Wang, Y. (1998). Mixed effects smoothing spline ANOVA. Journal of The Royal Statistical Society, Series B, 60, 159–174.

Andi_Nurhayati_Latif.pdf

ORIGINALITY REPORT

12%	11%	4%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ijeais.org Internet Source	3%
2	iptek.its.ac.id Internet Source	2%
3	ecampus.poltekkes-medan.ac.id Internet Source	1%
4	Submitted to Universitas Airlangga Student Paper	1%
5	web.pdx.edu Internet Source	1%
6	repository.its.ac.id Internet Source	1%
7	Mutia Maharani, Dewi Retno Sari Saputro, Putranto Hadi Utomo. "Application of the cubic spline nonparametric regression model to human development index data in Indonesia", AIP Publishing, 2022 Publication	1%
8	www.frontiersin.org Internet Source	

1 %

9	Mariana Ngundju Awang, Diyan Maria Kristin. "Quality of Visit and Fulfillment Needs after Childbirth: Study on Postpartum Maternal Mortality", JURNAL INFO KESEHATAN, 2020 Publication	<1 %
10	journal.unhas.ac.id Internet Source	<1 %
11	A T R Dani, V Ratnasari, I N Budiantara. "Optimal Knots Point and Bandwidth Selection in Modeling Mixed Estimator Nonparametric Regression", IOP Conference Series: Materials Science and Engineering, 2021 Publication	<1 %
12	gala.gre.ac.uk Internet Source	<1 %
13	ejurnal.ung.ac.id Internet Source	<1 %
14	eudl.eu Internet Source	<1 %
15	jurnal.untagsmg.ac.id Internet Source	<1 %