**NONPARAMETRIC SPLINE TRUNCATED REGRESSION WITH KNOT POINT SELECTION METHOD GENERALIZED CROSS VALIDATION AND UNBIASED RISK**

**Tutik Handayani1, Sifriyani2, Andrea Tri Rian Dani3**

1,2,3 Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia

tutik.handayani.ttk.110301@gmail.com1, sifriyani@fmipa.unmul.ac.id2, andreatririandani@fmipa.unmul.ac.id3

|  |  |  |
| --- | --- | --- |
|  |  | **ABSTRACT** |
| **Article History:**Received : D-M-20XXRevised : D-M-20XXAccepted : D-M-20XXOnline : D-M-20XX |  | Nonparametric regression approaches are used when the shape of the regression curve between the response variable and the predictor variable is assumed to be unknown. Nonparametric excess regression has high flexibility. A frequently used nonparametric regression approach is a truncated spline that has excellent ability to handle data whose behavior is variable at certain sub-intervals. The aim of this study was to obtain the best model of multivariable nonparametric regression with linear and quadratic truncated spline approaches using Generalized Cross Validation (GCV) and Unbiased Risk (UBR) methods and to find out the factors influencing stunting prevalence in Indonesia in 2021. The data used are the prevalence of stunting as a response variable and the predictor variable used by the percentage of infants receiving Exclusive breastfeeding for 6 months, the percentage of households with proper sanitation, the percentage of toddlers receiving Early Childhood Cultivation (IMD), the percentage of the poor population, and the percentage of pregnant womenIt's a risk. Results show that the best linear and quadratic nonparametric spline truncated regression model in modeling the stunting prevalence is linear truncated spline using the GCV method with three knot points. This model has the minimum GCV value of with MSE value of and a value of . Factors influencing the incidence of stunting in Indonesia in 2021 include the percentage variable of infants receiving Exclusive breastfeeding for 6 months, the percentage of households with proper sanitation, the percentage of poor people, and the percentage of pregnant women at risk of KEK. |
| **Keyword:**GCV; Nonparametric Regression; Linear Truncated SplineStunting; UBR. |
|  |
| <https://doi.org/10.31764/jtam.vXiY.ZZZ> | **This is an open access article under the CC–BY-SA license** |

—————————— ◆ ——————————

1. **INTRODUCTION**

Regression analysis is a statistical method that studies a model form of a set of data that relates between variables in explaining causal relationships or relationships(Padatuan et al., 2021). Regression analysis methods are used in analyzing data and drawing conclusions formed as a result of the relationship of dependence of variables to other variables(Zarkasi et al., 2021). Nonparametric regression is a model approach method used when the shape of the regression curve between the response variable and the predictor variable is assumed to be of unknown shape or pattern(Pasarella et al., 2022). Nonparametric regression curves are only assumed to be smooth, meaning they are contained within a certain function space so that nonparametric regression has high flexibility because the data is expected to find for itself the form of the regression curve estimate without being influenced by the researcher's subjectivity factor(Sifriyani et al., 2022).

A truncated spline is a piece of polynomial that has a continuously segmented property so that it effectively explains the local characteristics of the data function. The spline truncated approach is an approach towards matching data while still taking into account the smoothness of the curve. One of the advantages of truncated splines is that they are flexible, meaning that this model tends to look for data estimates wherever the data pattern moves (Sifriyani et al., 2021)

The application of multivariable nonparametric regression with a truncated spline approach with the GCV and UBR knot point methods can be applied to health data, one of which is Stunting data(Purnaraga et al., 2020). Stunting is an indication of chronic malnutrition as a result of poor interaction of various determinants of child nutrition. A child is stunted and severely stunted when the length or height is less than two and three standard deviations respectively of the average height of a child of the same age(Mediani, 2020). Furthermore, PSG revealed that the incidence of stunting was higher in children under five (29.6%) compared to the younger group under two years old (20.1%) The high incidence of stunting, more than 50% in a number of districts and the lack of information about children's short posture due to health problems made people pressed to resuscitate to stop blaming heredity as the cause of children not meeting height standards. The lack of information about child nutrition and the abandonment of risk factors and the impact of stunting, require immediate remedial action, especially from mothers and caregivers, so that our next generation can survive in a competitive global economy (Liem et al., 2019).

Unlike weight, the height of a toddler is often ignored by his parents. It is easier and more noticeable to notice the size of the child compared to other children of the same age or even different ages. Often, we hear parents complain "my daughter is too thin" or "she's too big". Both refer to body size or weight. On the other hand, unless it is excessive and affects the functioning of vital organs, being overweight to some extent is associated as healthy(Almoosawi et al., 2016).

Parents with overweight or obese children tend to underestimate their children's weight, they do not refer to a standard weight at an appropriate age but rather seek to maintain their own perception of 'healthy weight (Park, 2017). In everyday life, weight indicates whether a child is healthy, while height depends on heredity and is not related to a person's health status(Cheng et al., 2016). A child whose posture is less than average looks like it was born that way, has nothing to do with nutritional status(Remmers et al., 2014).

Although it indicates malnutrition, stunting is not only caused by poor nutrition. Conducted a study using data from 137 developing countries, that poor environmental health results in children under the age of five years having a greater risk of stunting through various environmental-based diseases including diarrhea(Danaei et al., 2016). According to Checkley et al. (2008), 25% of stunting cases in children under two years of age are associated with diarrhea. The same conclusion was confirmed by a study conducted in East Java in which it was found that diarrhea and poor hygienic behavior were significantly associated with stunting(Desyanti & Nindya, 2017). When repeatedly experiencing diarrhea, a child has a higher probability of stunting.

According to a report by the World Health Organization (2020), Stunting is caused by various child developmental disorders and has affected 21.3% of children under the age of five globally with a total of 144 million cases. Some studies show that this condition is more common in Asian and African countries(Schmidt, 2014). The 2018 Basic Health Research (Riskesdas) and the 2019 Indonesian Toddler Nutritional Status Survey (SSGBI) reported a decrease in stunting rates from 30.8% to 27.7%. However, it is still a health problem because its prevalence is above the WHO standard of 20% (Latuihamallo et al., 2022).

Nutrient intake plays an important role in supporting child development. Thus, low energy and protein consumption leads to growth failure(Drennen et al., 2019). This was evident in 45.7% of the sample population who had an energy intake of <70% AKE, while 36.1% was <80% RDA(Mehedi Hasan Abid et al., 2021). Furthermore, previous research revealed that toddlers with low consumption are at 7.5 times higher risk of stunting. A Brazilian study also reported that protein intake should meet a child's nutritional needs. Toddlers are at 1.59 times the risk of stunting when their protein intake is below the level of nutritional adequacy.

Child growth is the main indicator for assessing nutritional status in children under five years of age (under five years old) and is also one of the 6 global nutrition targets set by WHO in 2012 (da Silva et al., 2018) and is a leading indicator of the Sustainable Development Goals (SDGs) of 2030(Dewi et al., 2019). This can be interpreted to mean that nutrition has an important role in the growth period of toddlers. Adequate nutrition, health conditions, protection, and safety factors play an important role in child development, especially at an early age The occurrence of stunting in this period can affect the structure and function of the brain where a reduced number of cells causes growth delays. A survey by the Indonesian Ministry of Health revealed that 16% of children under five years old have impaired fine and gross motor development, hearing loss, decreased intelligence and speech delay with a total of 0.4 million cases(Padatuan et al., 2021).

If at the age of growth the toddler is already stunted, it will have a bad impact on the future of the child. The risk of obesity occurring in children will be very high considering that short children have a low ideal weight. An increase in body weight alone can cause the child's Body Mass Index (BMI) to rise beyond the normal limit. This situation will continue for a long time until the risk of degenerative diseases occurs. Children are considered stunted and severely stunted if body length and height based on age range are less than the WHO-Multicentre Growth Reference Median Standard Study (MGRS)(Hendraswari et al., 2021).

According to the World Health Organization (WHO) an area can be said to have stunting problems if the stunting percentage is above 20%. Referring to WHO standards, it can be known that Indonesia has a fairly high stunting problem of 31.8% so it is necessary to pay special attention from the government to this incident to anticipate the increasing prevalence of stunting in the future. One of the special concerns that can be done is to pay attention to the factors that cause stunting in Indonesia(Indanah et al., 2022).

The Sustainable Development Goals (SDGs) one of the goals in the health sector mentions the target of community nutrition (Singh et al., 2017), namely by 2030, ending all forms of malnutrition, including the achievement of the 2025 international target to reduce stunting and wasting in children under five, reducing stunting programs globally The government pays appropriate attention to stunting in children under the age of 2-3 years through the National and Internal Nutrition international movement namely the Scaling Up Nutrition (SUN) movement with a concentration system to border areas(Yulianti et al., 2022).

The causes of stunting are multifactorial, which include genetic, socio-demographic, economic status, as well as cultural and environmental factors and other health-related variables(Geberselassie et al., 2018). This is in line with research conducted in Bangladesh in 2018 where the results showed that stunting factors are caused by parental education factors. In addition, sanitation facilities in households also determine the incidence of stunting in Indonesia, unkempt toilets and unprocessed drinking water provide three times greater chance of stunting(Torlesse et al., 2016).

Previous research has shown that shaky growth before birth and 18 months after pregnancy is associated with poor language and motor development (Sudfeld et al., 2015). Stunted children aged 2, 5, and 9 years had verbal scores and a lower IQ of 4.6 points compared to others(Koshy et al., 2022). Some studies have also revealed that they have lower scores in all aspects of development. A study in Kalasan showed that stunted children were 3.9 times more at risk of suspicion than others with normal growth(Nahar et al., 2020).

1. **METHODS**

The data used in this study are secondary data, namely data on the prevalence of stunting, the percentage of households that have access to proper sanitation, the percentage of toddlers who get IMD, the percentage of pregnant women at risk of SEZ obtained from the Ministry of Health of the Republic of Indonesia Study of Nutritional Status of Indonesia (SSGI). Meanwhile, data on the percentage of babies aged less than 6 months who get exclusive breastfeeding and the number of poor people are obtained from the Central Statistics Agency (BPS).

**Table 1.** Research Variables

|  |  |
| --- | --- |
| **Variable Name** | **Variable Definition** |
| Prevalence of Stunting Toddlers | The condition of growth failure in children under five years old (infants under five years old) due to chronic malnutrition so that the child is too short for his age. |
| Percentage of Babies Getting Exclusive Breastfeeding for 6 Months | Babies who only get breast milk from birth to 6 months of age in one working area at a certain period of time. |
| Percentage of Households That Have Access to Proper Sanitation | Intentional behavior in the cultivation of clean living with the intention of preventing humans from coming into direct contact with dirt or other harmful waste materials in the hope that this effort will maintain and improve human health |
| Percentage of Toddlers Who Get Early Breastfeeding Initiation (IMD) | is the beginning of a mother giving breast milk to her baby when the baby is born into the world, namely in the first hours or 1 hour after giving birth. |
| Percentage of Poor People | The percentage of poor people who are below the poverty line. |
| Percentage of Pregnant Women at Risk of SEZ | SEZ is a condition where the mother experiences malnutrition. |

Research steps can be written in each stage to achieve research objectives, so the steps taken are as follows:

1. Collecting secondary data on stunting prevalence cases in Indonesia in 2021 and the factors that influence it.
2. Conducting descriptive statistical analysis on response variables and predictor variables which aims to determine the characteristics of stunting prevalence cases in Indonesia in 2021.
3. Identify data patterns formed from stunting prevalence cases in Indonesia in 2021 with every factor that influences them using scatterplots.
4. Performs multicollinearity detection on predictor variables.
5. Selecting the optimal knot point using the GCV method on linear truncated spline nonparametric regression.
6. Selecting the optimal knot point using the UBR method on linear truncated spline nonparametric regression.
7. Performs parameter estimation of truncated spline nonparametric regression models.
8. Comparing GCV and UBR methods in optimal knot point selection on linear truncated spline nonparametric regression.
9. Selects the best nonparametric spline truncated regression model with optimal knot points obtained from GCV and UBR methods based on MSE and R2.
10. Test the error normality assumption.
11. Conduct a simultaneous parameter significance test and proceed to perform a partial test if a decision is obtained to reject H0 on the simultaneous test.
12. Interpret the model obtained and make conclusions.
13. **RESULT AND DISCUSSION**
	* + 1. **Descriptive Statistical Analysis**

Descriptive statistics for response variable data and predictor variables from observational data are shown in Table 2 below:

**Table 2.** Descriptive Statistics of Response Variables and Predictors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **Minimum** | **Maximum** | **Variance** |
| Prevalence of Stunting Toddlers | 25,21 | 10,90 | 37,80 | 31,70 |
| Percentage of Babies Getting Exclusive Breastfeeding for 6 Months | 68,88 | 52,75 | 81,46 | 60,24 |
| Percentage of Households That Have Access to Proper Sanitation | 80,97 | 40,81 | 97,12 | 98,51 |
| Percentage of Toddlers Who Get Early Breastfeeding Initiation (IMD) | 49,23 | 32,50 | 62,70 | 57,60 |
| Percentage of Poor People | 10,76 | 4,53 | 26,86 | 29,19 |
| Percentage of Households That Have Access to Proper Sanitation | 12,10 | 3,10 | 40,70 | 50,94 |

Based on Table 2, it can be seen that the case of stunting prevalence in Indonesia in 2021 obtained an average value of 25.21%, with the lowest value of 10.90% and the highest of 37.80%. The variable percentage of babies getting exclusive breastfeeding for 6 months in Indonesia in 2021, obtained an average value of 68.88%, the lowest value of 52.75% and the highest of 81.46%. The variable percentage of households that have access to proper sanitation in Indonesia in 2021 obtained an average value of 80.97%, with the lowest value of 40.81% and the highest of 97.12%. The variable percentage of toddlers who get IMD in Indonesia in 2021, obtained an average value of 49.23%, with the lowest value of 32.50% and the highest of 62.70%. The variable percentage of poor people in Indonesia in 2021, obtained an average value of 10.76%, the province that has a percentage of poor people, namely Bali at 4.53% and the highest is Papua at 26.86%. The variable percentage of pregnant women at SEZ risk in Indonesia in 2021 obtained an average value of 12.1%, with the lowest value of 3.10% and the highest of 40.70%.

* + - 1. **Relationship Patterns Between Predictor Variables and Response Variables**

The relationship between the response variable and the five variables that are thought to have an effect can be seen in the scatterplot in Figure 1.





**Figure 1.** Scatterplot Between Response Variables and 5 Variables Suspected of Influence

Based on Figure 1, it can be seen that the distribution between the Stunting Percentage data and the five variables that are suspected to have an effect does not form a certain pattern, the data pattern spreads and some data that is far from other data distributions, so the spline truncated nonparametric regression method can be used on the data because the shape of the data pattern is unknown.

* + - 1. **Selection of Optimal Knot Points Using the GCV Method**

The first step taken before modeling using truncated spline nonparametric regression is to determine the number of knot points used. In this study, the knot points tried were 3 knots, after which one optimal knot, two optimal knots, and three optimal knots will be sought. Here is the selection of the optimal knot point using the GCV method.

The nonparametric spline truncated regression model on stunting prevalence data in Indonesia in 2021 with one knot point is as follows.

|  |  |
| --- | --- |
|  | (1) |

After 50 knot point experiments to obtain the optimal knot point, 5 smallest CV values with one knot are obtained as shown by table 3 as follows.

**Table 3.** GCV Values with One Point Knots on Linear Truncated Spline Nonparametric Regression

|  |  |
| --- | --- |
| **Knot Point** | **GCV** |
|  |  |  |  |  |
| 71,50 | 77,58 | 52,22 | 19,11 | 27,65 | 8,74 |
| 72,08 | 78,73 | 52,84 | 19,57 | 28,42 | 9,20 |
| 70,91 | 76,43 | 51,61 | 18,66 | 26,89 | 9,53 |
| 72,67 | 79,88 | 53,45 | 20,02 | 29,19 | 10,09 |
| 73,26 | 81,03 | 54,07 | 20,48 | 29,96 | 10,34 |

Table 3 shows the minimum GCV values for multivariable nonparametric regression models on linear truncated splines with a single knot point of 8.74. with the optimal knot point in variable X1 which is 71.50, in variable X2 which is 77.58, in variable X3 which is 52.22, in variable X4 which is 19.11, and in variable X5 which is 27.65. A truncated spline nonparametric regression model using a minimum GCV value with one knot can be written down in equation (2).

|  |  |
| --- | --- |
|  | (2) |

Furthermore, the selection of knot points is carried out using two knot points. The nonparametric spline truncated regression model on stunting prevalence data in Indonesia in 2021 with two knots is as follows:

|  |  |
| --- | --- |
|   | (3) |

50 knot point experiments were conducted to obtain the optimal knot point. Table 4 shows the 10 smallest CV values with two knots dots.

**Table 4**.GCV Values with Two Point Knots on Linear Truncated Spline Nonparametric Regression

|  |  |
| --- | --- |
| **Knot Point** | **GCV** |
|  |  |  |  |  |
| 71,50 | 77,58 | 52,22 | 19,11 | 27,65 | 9,69 |
| 74,43 | 83,33 | 55,30 | 21,39 | 31,49 |
| 71,50 | 77,58 | 52,22 | 19,11 | 27,65 | 9,71 |
| 75,01 | 84,48 | 55,92 | 21,85 | 32,26 |
| 71,50 | 77,58 | 52,22 | 19,11 | 27,65 | 9,71 |
| 72,67 | 79,88 | 53,45 | 20,02 | 29,19 |
| 71,50 | 77,58 | 52,22 | 19,11 | 27,65 | 9,74 |
| 76,77 | 87,93 | 57,77 | 23,21 | 34,56 |
| 71,50 | 77,58 | 52,22 | 19,11 | 27,65 | 9,84 |
| 77,34 | 89,07 | 58,38 | 23,67 | 35,33 |

Table 4. shows the minimum CV value for two knots of 9.69 with optimal knot points in variables X1 of 71.50 and 74.43, in variable X2 of 77.58 and 83.33, in variable X3 of 52.22 and 55.30, in variables X4 of 19.11 and 21.39, and in variables X5 of 27.65 and 31.49. A truncated spline nonparametric regression model using a minimum GCV value with two knots can be written down in equation (4).

|  |  |
| --- | --- |
|   | (4) |

The next step is to select knot points using three knots. The spline truncated nonparametric regression model on stunting prevalence data in Indonesia in 2021 with three knots based on equation (5) is as follows.

|  |  |
| --- | --- |
|   | (5) |

50 knot point experiments were conducted to obtain the optimal knot point. Table 5 shows the 5 smallest GCV values with three knots.

**Table 5.** GCV Values with Three Point Knots on Linear Truncated Spline Nonparametric Regression

|  |  |
| --- | --- |
| **Knot Point** | **GCV** |
|  |  |  |  |  |
| 71,50 | 77,58 | 52,22 | 19,11 | 27,65 | 7,29 |
| 77,36 | 89,07 | 58,38 | 23,67 | 35,33 |
| 80,87 | 95,97 | 62,08 | 26,40 | 39,93 |
| 71,50 | 77,58 | 52,22 | 19,11 | 27,65 | 7,32 |
| 76,77 | 87,93 | 57,77 | 23,21 | 34,56 |
| 80,87 | 95,97 | 62,08 | 26,40 | 39,93 |
| 71,50 | 77,58 | 52,22 | 19,11 | 27,65 | 7,42 |
| 77,94 | 90,22 | 59,00 | 24,12 | 36,09 |
| 80,29 | 94,82 | 61,47 | 25,95 | 39,16 |
| 71,50 | 77,58 | 52,22 | 19,11 | 27,65 | 7,66 |
| 77,36 | 89,07 | 58,38 | 23,67 | 35,33 |
| 80,29 | 94,82 | 61,47 | 25,95 | 39,16 |
| 71,50 | 77,58 | 52,22 | 19,11 | 27,65 | 7,79 |
| 77,94 | 90,22 | 59,00 | 24,12 | 36,09 |
| 80,87 | 95,97 | 62,08 | 26,40 | 39,93 |

Table 5 shows the minimum CV value for three knots of 7.29 with the optimal knot point at variable X1 which is 71.50; 77,36; and 80.87, at variable X2 which is 77.58; 89,07; and 95.97, at variable X3 which is 52.22; 58,38; and 62.08, at variable X4 which is 19.11; 23,67; and 26.40, and at variable X5 which is 27.65; 35,33; and 39.93. A truncated spline nonparametric regression model using a minimum GCV value of three knots can be written down in equation (6).

|  |  |
| --- | --- |
|   | (6) |

The results of optimal knot point selection using GCV with one knot point, two knots point, and three knots point can be seen in Table 6.

**Table 6**.GCV Method Knot Point Comparison

|  |  |
| --- | --- |
| **Knot Point** | **GCV Value** |
| 1 Knot Point | 8,74 |
| 2 Knot Point | 9,69 |
| 3 Knot Point | 7,29 |

Table 6 shows a minimum GCV value of 7.29 in models with three knots. Parameter estimation of a truncated spline nonparametric regression model using the GCV method with three knots is as follows.

|  |  |
| --- | --- |
|    | (7) |

The truncated spline nonparametric regression model with three knots in equation (7) yielded an R2 value of 99.57, this suggests the model can explain the Prevelence data for stunting in 2021 of 94.07%.

* + - 1. **Optimal Knot Point Using UBR Method**

The first step taken before modeling using truncated spline nonparametric regression is to determine the number of knot points used. In this study, the knot points tried were 3 knots, after which one optimal knot, two optimal knots, and three optimal knots will be sought. Here is the selection of the optimal knot point using the UBR method. The spline truncated nonparametric regression model on stunting prevalence data in Indonesia in 2021 with three knots based on equation (8) is as follows.

|  |  |
| --- | --- |
|   | (8) |

After 50 knot point experiments to obtain the optimal knot point, 5 smallest UBR values with one knot are obtained as shown by table 7 as follows.

**Tabel 7.** UBR Values with One Point Knots on Linear Truncated Spline Nonparametric Regression

|  |  |  |
| --- | --- | --- |
| **Knot Point** |  | **UBR** |
|  |  |  |  |  |
| **73,84** | **82,18** | **54,69** | **20,93** | **30,72** |  |  |
| 62,71 | 60,35 | 42,98 | 12,28 | 16,14 |  |  |
| 70,33 | 75,28 | 50,99 | 18,20 | 26,12 |  |  |
| 77,94 | 90,22 | 59,00 | 24,12 | 36,09 |  |  |
| 65,64 | 66,09 | 46,06 | 14,55 | 19,98 |  |  |

Table 7 shows the minimum UBR values for multivariable nonparametric regression models on linear truncated splines with one point of knots of. with the optimal knot point in variable X1 which is 73.84, in variable X2 which is 82.18, in variable X3 which is 54.69, in variable X4 which is 20.93, and in variable X5 which is 30.72. A truncated spline nonparametric regression model using a minimum UBR value with one knot can be written down in equation (9).

|  |  |
| --- | --- |
|   | (9) |

Furthermore, the selection of knot points is carried out using two knot points. The nonparametric spline truncated regression model on stunting prevalence data in Indonesia in 2021 with two knots is as follows:

|  |  |
| --- | --- |
|   | (10) |

50 knot point experiments were conducted to obtain the optimal knot point. Table 8 shows the 5 smallest UBRs with two knots.

**Tabel 8**.UBR Values with Two Point Knots on Linear Truncated Spline Nonparametric Regression

|  |  |  |
| --- | --- | --- |
| **Titik Knot** |  | **UBR** |
|  |  |  |  |  |
| **77,94** | **90,22** | **59,00** | **24,12** | **39,09** |  |  |
| **80,87** | **95,97** | **62,08** | **26,40** | **39,93** |
| 79,12 | 92,52 | 60,23 | 25,04 | 37,63 |  |  |
| 80,87 | 95,97 | 62,08 | 26,40 | 39,93 |
| 75,60 | 85,63 | 56,54 | 22,30 | 33,03 |  |  |
| 79,11 | 92,52 | 60,23 | 25,03 | 37,63 |
| 75,60 | 85,63 | 56,54 | 22,30 | 33,03 |  |  |
| 79,70 | 93,67 | 60,85 | 25,49 | 38,40 |
| 77,35 | 89,07 | 58,38 | 23,67 | 35,33 |  |  |
| 78,53 | 91,37 | 59,62 | 24,58 | 36,86 |

Table 8. shows the minimum UBR value for two knots of  with optimal knot points in variables X1 of 77.94 and 80.87, in variables X2 of 90.22 and 95.97, in variables X3 of 59.00 and 62.08, in variables X4 of 24.12 and 26.40, and in variables X5 of 39.09 and 39.93. A truncated spline nonparametric regression model using a minimum GCV value with two knots can be written down in equation (11).

|  |  |
| --- | --- |
|   | (11) |

The next step is to select knot points using three knots. A nonparametric spline truncated regression model on stunting prevalence data in Indonesia in 2021 with three knots based on equation (12) is as follows.

|  |  |
| --- | --- |
|   | (12) |

50 knot point experiments were conducted to obtain the optimal knot point. Table 5 shows the 5 smallest UBR values with three knots.

**Tabel 9.** UBR Values with Three Point Knots on Linear Truncated Spline Nonparametric Regression

|  |  |  |
| --- | --- | --- |
| **Titik Knot** |  | **UBR** |
|  |  |  |  |  |
| **66,23** | **67,24** | **46,67** | **15,01** | **20,74** |  |  |
| **76,77** | **87,93** | **57,77** | **23,21** | **34,56** |
| **77,94** | **90,22** | **59,00** | **24,12** | **36,09** |
| 66,23 | 67,24 | 46,67 | 15,01 | 20,74 |  |  |
| 72,67 | 79,88 | 53,45 | 20,02 | 29,19 |
| 76,77 | 87,93 | 57,77 | 23,21 | 34,56 |
| 66,81 | 68,39 | 47,29 | 15,47 | 21,52 |  |  |
| 77,94 | 90,22 | 59,00 | 24,12 | 36,09 |
| 79,70 | 93,67 | 60,85 | 25,49 | 38,40 |
| 66,81 | 68,39 | 47,29 | 15,47 | 21,52 |  |  |
| 75,01 | 84,48 | 55,92 | 21,84 | 32,26 |
| 77,36 | 89,07 | 58,38 | 23,67 | 35,32 |
| 65,64 | 66,09 | 46,06 | 14,55 | 19,98 |  |  |
| 77,36 | 89,07 | 58,38 | 23,67 | 35,33 |
| 80,87 | 95,97 | 62,08 | 26,40 | 39,93 |

Table 9 shows the minimum UBR value for three knots of  with the optimal knot point on variable X1 is 66.23; 76,77; and 77.94, at variable X2 which is 67.24; 87,93; and 90.22, at variable X3 which is 46.67; 57,77; and 59.00, at variable X4 which is 15.01; 23,21; and 24.12, and at variable X5 which is 20.74; 34,56; and 36.09. A truncated spline nonparametric regression model using a minimum UBR value of three knots can be written down in equation (13).

|  |  |
| --- | --- |
|   | (13) |

The results of optimal knot point selection using UBR with one knot point, two knots point, and three knot points can be seen in Table 10.

**Tabel 10**.Perbandingan Titik Knot Metode UBR

|  |  |
| --- | --- |
| **Knot Point** | **UBR Value** |
| 1 Knot Point |  |
| 2 Knot Point |  |
| 3 Knot Point |  |

Table 10 shows the minimum UBR value of i.e. in models with two knots. The parameter estimation of a truncated spline nonparametric regression model using the UBR method with two knots is as follows.

|  |  |
| --- | --- |
|   | (14) |

The truncated spline nonparametric regression model with three knots in equation (14) yielded an R2 value of 0.8467, this suggests the model can explain the 2021 Prevelence stunting data of 84.67%.

* + - 1. **Comparison of GCV Method and UBR Method in selection of Optimal Knot Point**

The following is shown a comparison table of GCV and UBR methods in the selection of optimal knot points in stunting prevalence data in Indonesia in 2021.

**Table 11**.Comparison of GCV and UBR Methods

|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | **Knot Points** | **MSE** | **R2** |
| GCV | Three Knot Points | 1,82 | 94,07 |
| UBR | Two Knot Points | 4,87 | 84,16 |

Table 11 shows that the MSE value of 1.82 obtained in nonparametric spline truncated regression modeling with three knots using the GCV method is better than the UBR method with an MSE value of 4.87. Clarified by the R2 value produced by the GCV method of 94.07 is greater than the R2 value produced by the UBR method of 84.16. So it can be concluded that the GCV method with three knots is a more appropriate method in selecting optimal knot points than the UBR method in Indonesia's stunting prevalence data in 2021.

* + - 1. **Simultaneous testing of parameter significance**

Simultaneous testing is performed to test whether all the parameters in the regression model have a significant effect. The results of simultaneous testing are as follows.



there is at least one 

The ANOVA results can be seen in Table 12 as follows.

**Table 12**.ANOVA

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sources of Variance** | **Degree of freedom****(df)** | **Sum of Square****(SS)** | **Mean Square****(MS)** | **F** |
| Regression | 20 | 861,09 | 43,05 | 7,92 |
| Error | 13 | 184,69 | 5,43 |
| Total | 33 | 1045,78 |  |

Based on Table 12 it can be seen that the calculated F value of 7.92 when compared to the value of F(20,13,0.05) of 2.46, it is decided that H0 is rejected. So it can be concluded that there is at least one significant parameter to the prevalence of stunting in Indonesia in 2021.

* + - 1. **Partial Parameter Significance Testing**

Testing was performed to determine which parameters had a significant effect on the regression model. Partial test results are as follows:





The results of the test can be partially seen in Table 13 as follows.

**Table 13.** Partial Parameter Significance Test Results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Parameter** | **Estimation** | **p-value** | **t-value** | **Decision** |
| Constant | $$β\_{0}$$ | 45,23 | 0,00\* | 5,37 | Significant |
|  | $$β\_{11}$$ | -0,09 | 0,19 | -1,37 | Insignificant |
| $$β\_{12}$$ | -0,16 | 0,11 | -1,69 | Insignificant |
| $$β\_{13}$$ | -0,16 | 0,04\* | -2,25 | Significant |
| $$β\_{14}$$ | 0,03 | 0,81 | -0,24 | Insignificant |
|  | $$β\_{21}$$ | 0,70 | 0,00\* | 5,39 | Significant |
| $$β\_{22}$$ | -0,22 | 0,96 | -0,06 | Insignificant |
| $$β\_{23}$$ | 2,73 | 0,30 | 1,09 | Insignificant |
| $$β\_{24}$$ | 5,68 | 0,52 | 0,65 | Insignificant |
|  | $$β\_{31}$$ | 1,97 | 0,50 | 0,69 | Insignificant |
| $$β\_{32}$$ | -9,23 | 0,11 | -1,72 | Insignificant |
| $$β\_{33}$$ | 6,17 | 0,07 | 1,96 | Insignificant |
| $$β\_{34}$$ | -3,37 | 0,24 | -1,23 | Insignificant |
|  | $$β\_{41}$$ | 6,62 | 0,07 | 1,96 | Insignificant |
| $$β\_{42}$$ | 3,31 | 0,07 | 1,96 | Insignificant |
| $$β\_{43}$$ | -5,64 | 0,01\* | -2,79 | Significant |
| $$β\_{44}$$ | -3,76 | 0,01\* | -2,79 | Significant |
|  | $$β\_{51}$$ | -1,88 | 0,01\* | -2,79 | Significant |
| $$β\_{52}$$ | -6,14 | 0,00\* | -5,37 | Significant |
| $$β\_{53}$$ | -4,09 | 0,00\* | -5,37 | Significant |
| $$β\_{54}$$ | -2,04 | 0,00\* | -5,37 | Significant |

Based on Table 13, there are 21 parameters in the truncated spline nonparametric regression model formed with a 95% confidence level and a significance level of 5% there are 9 significant parameters and 12 insignificant parameters. The parameter is declared significant if the p-value is less than the level of significance. Partially variable, the percentage of babies getting exclusive breastfeeding for 6 months (x1), the percentage of households that have proper sanitation (x2), the percentage of poor people (x4), and the percentage of pregnant women at risk of SEZ (x5) affect the prevalence of stunting (y), while the variable percentage of toddlers who get IMD (x3) has no effect on the prevalence of stunting (y) in Indonesia in 2021.

1. **CONCLUSION AND SUGGESTIONS**

The best multivariable nonparametric regression model in the case of stunting prevalence in Indonesia year obtained from a linear truncated spline approach using the GCV method with three knots resulting in an MSE value of 1.82 with an R2 value of 94.07%, so it can be concluded that the GCV method with three knots is a more appropriate method used for optimal knot point selection compared to using the UBR method in the case of stunting prevalence in Indonesia year 2021. Based on testing the significance of partial parameters of factors that influence stunting prevalence cases in Indonesia in 2021 based on the best model, namely the variable percentage of babies getting exclusive breastfeeding for 6 months (x1), the percentage of households that have proper sanitation (x2), the percentage of poor people (x4), and the percentage of pregnant women at risk of SEZ (x5).

**REFERENCES**

Almoosawi, S., Jones, A. R., Parkinson, K. N., Pearce, M. S., Collins, H., & Adamson, A. J. (2016). Parental Perception of Weight Status: Influence on Children’s Diet in the Gateshead Millennium Study. *PLOS ONE*, *11*(2), e0144931. https://doi.org/10.1371/journal.pone.0144931

Cheng, T. S., Loy, S. L., Cheung, Y. B., Chan, J. K. Y., Tint, M. T., Godfrey, K. M., Gluckman, P. D., Kwek, K., Saw, S. M., Chong, Y.-S., Lee, Y. S., Yap, F., & Lek, N. (2016). Singaporean Mothers’ Perception of Their Three-year-old Child’s Weight Status: A Cross-Sectional Study. *PLOS ONE*, *11*(1), e0147563. https://doi.org/10.1371/journal.pone.0147563

da Silva, I. C. M., França, G. V, Barros, A. J., Amouzou, A., Krasevec, J., & Victora, C. G. (2018). Socioeconomic Inequalities Persist Despite Declining Stunting Prevalence in Low- and Middle-Income Countries. *The Journal of Nutrition*, *148*(2), 254–258. https://doi.org/10.1093/jn/nxx050

Danaei, G., Andrews, K. G., Sudfeld, C. R., Fink, G., McCoy, D. C., Peet, E., Sania, A., Smith Fawzi, M. C., Ezzati, M., & Fawzi, W. W. (2016). Risk Factors for Childhood Stunting in 137 Developing Countries: A Comparative Risk Assessment Analysis at Global, Regional, and Country Levels. *PLOS Medicine*, *13*(11), e1002164. https://doi.org/10.1371/journal.pmed.1002164

Desyanti, C., & Nindya, T. S. (2017). Hubungan Riwayat Penyakit Diare dan Praktik Higiene dengan Kejadian Stunting pada Balita Usia 24-59 Bulan di Wilayah Kerja Puskesmas Simolawang, Surabaya. *Amerta Nutrition*, *1*(3), 243. https://doi.org/10.20473/amnt.v1i3.2017.243-251

Dewi, A. R., Dewi, Y. L. R., & Murti, B. (2019). Life Course Factors Associated with Stunting in Children Aged 2-5 Years: A Path Analysis. *Journal of Maternal and Child Health*, *4*(5), 348–357. https://doi.org/10.26911/thejmch.2019.04.05.09

Drennen, C. R., Coleman, S. M., Ettinger de Cuba, S., Frank, D. A., Chilton, M., Cook, J. T., Cutts, D. B., Heeren, T., Casey, P. H., & Black, M. M. (2019). Food Insecurity, Health, and Development in Children Under Age Four Years. *Pediatrics*, *144*(4). https://doi.org/10.1542/peds.2019-0824

Geberselassie, S. B., Abebe, S. M., Melsew, Y. A., Mutuku, S. M., & Wassie, M. M. (2018). Prevalence of stunting and its associated factors among children 6-59 months of age in Libo-Kemekem district, Northwest Ethiopia; A community based cross sectional study. *PLOS ONE*, *13*(5), e0195361. https://doi.org/10.1371/journal.pone.0195361

Hendraswari, C. A., Purnamaningrum, Y. E., Maryani, T., Widyastuti, Y., & Harith, S. (2021). The Determinants of Stunting in 24-59 Month-Old Children in Kulon Progo District 2019. *Kesmas: National Public Health Journal*, *16*(2). https://doi.org/10.21109/kesmas.v16i2.3305

Indanah, I., Wanda, D., & Nurhaeni, N. (2022). Empowerment of Parents with Stunting Children. *Jurnal Aisyah : Jurnal Ilmu Kesehatan*, *7*(4). https://doi.org/10.30604/jika.v7i4.1183

Koshy, B., Srinivasan, M., Gopalakrishnan, S., Mohan, V. R., Scharf, R., Murray-Kolb, L., John, S., Beulah, R., Muliyil, J., & Kang, G. (2022). Are early childhood stunting and catch-up growth associated with school age cognition?—Evidence from an Indian birth cohort. *PLOS ONE*, *17*(3), e0264010. https://doi.org/10.1371/journal.pone.0264010

Latuihamallo, A., Margawati, A., Mexitalia, M., Ediati, A., & Syauqy, A. (2022). Differences in Development and Diet of Stunting and Non-Stunting Children in the Rowosari Health Center Work Area, Semarang, Indonesia. *Jurnal Gizi Indonesia (The Indonesian Journal of Nutrition)*, *10*(2), 161–167. https://doi.org/10.14710/jgi.10.2.161-167

Liem, S., Marta, Dr. R. F., & Panggabean, Prof. Dr. phil. H. (2019). Sanitation Behavior and Risk of Stunting: Understanding the Discourse of a Public Service Announcement. *Jurnal The Messenger*, *11*(2), 168. https://doi.org/10.26623/themessenger.v11i2.1317

Mediani, H. S. (2020). Predictors of Stunting Among Children Under Five Year of Age in Indonesia: A Scoping Review. *Global Journal of Health Science*, *12*(8), 83. https://doi.org/10.5539/gjhs.v12n8p83

Mehedi Hasan Abid, Dm., Haque, A., & Kamrul Hossain, Md. (2021). *Factors Causing Stunting Among Under-Five Children in Bangladesh* (pp. 45–53). https://doi.org/10.1007/978-981-15-5856-6\_5

Nahar, B., Hossain, M., Mahfuz, M., Islam, M. M., Hossain, M. I., Murray‐Kolb, L. E., Seidman, J. C., & Ahmed, T. (2020). Early childhood development and stunting: Findings from the MAL‐ED birth cohort study in Bangladesh. *Maternal & Child Nutrition*, *16*(1). https://doi.org/10.1111/mcn.12864

Padatuan, A. B., Sifriyani, S., & Prangga, S. (2021). PEMODELAN ANGKA HARAPAN HIDUP DAN ANGKA KEMATIAN BAYI DI KALIMANTAN DENGAN REGRESI NONPARAMETRIK SPLINE BIRESPON. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *15*(2), 283–296. https://doi.org/10.30598/barekengvol15iss2pp283-296

Park, S. H. (2017). Asian Parents’ Perception of Child Weight Status: A Systematic Review. *Journal of Child and Family Studies*, *26*(9), 2363–2373. https://doi.org/10.1007/s10826-017-0765-x

Pasarella, M. D., Sifriyani, S., & Amijaya, F. D. T. (2022). NONPARAMETRIK REGRESSION MODEL ESTIMATION WITH THE FOURIER SERIES THE FOURIER SERIES APPROACH AND ITS APPLICATION TO THE ACCUMULATIVE COVID-19 DATA IN INDONESIA. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *16*(4), 1167–1174. https://doi.org/10.30598/barekengvol16iss4pp1167-1174

Purnaraga, T., Sifriyani, S., & Prangga, S. (2020). REGRESI NONPARAMAETRIK SPLINE PADA DATA LAJU PERTUMBUHAN EKONOMI DI KALIMANTAN. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *14*(3), 343–356. https://doi.org/10.30598/barekengvol14iss3pp343-356

Remmers, T., van Grieken, A., Renders, C. M., Hirasing, R. A., Broeren, S. M. L., & Raat, H. (2014). Correlates of Parental Misperception of Their Child’s Weight Status: The ‘Be Active, Eat Right’ Study. *PLoS ONE*, *9*(2), e88931. https://doi.org/10.1371/journal.pone.0088931

Schmidt, C. W. (2014). Beyond Malnutrition: The Role of Sanitation in Stunted Growth. *Environmental Health Perspectives*, *122*(11). https://doi.org/10.1289/ehp.122-A298

Sifriyani, Ilmi, H., & Mar’ah, Z. (2021). Application of Nonparametric Geographically Weighted Spline Regression Model for Spatial Mapping of Open Unemployment Rate in Kalimantan. *Journal of Physics: Conference Series*, *2123*(1), 012038. https://doi.org/10.1088/1742-6596/2123/1/012038

Sifriyani, S., Rasjid, M., Rosadi, D., Anwar, S., Wahyuni, R. D., & Jalaluddin, S. (2022). Spatial-Temporal Epidemiology of COVID-19 Using a Geographically and Temporally Weighted Regression Model. *Symmetry*, *14*(4), 742. https://doi.org/10.3390/sym14040742

Singh, A., Upadhyay, A. K., & Kumar, K. (2017). Birth Size, Stunting and Recovery from Stunting in Andhra Pradesh, India: Evidence from the Young Lives Study. *Maternal and Child Health Journal*, *21*(3), 492–508. https://doi.org/10.1007/s10995-016-2132-8

Sudfeld, C. R., Charles McCoy, D., Danaei, G., Fink, G., Ezzati, M., Andrews, K. G., & Fawzi, W. W. (2015). Linear Growth and Child Development in Low- and Middle-Income Countries: A Meta-Analysis. *Pediatrics*, *135*(5), e1266–e1275. https://doi.org/10.1542/peds.2014-3111

Torlesse, H., Cronin, A. A., Sebayang, S. K., & Nandy, R. (2016). Determinants of stunting in Indonesian children: evidence from a cross-sectional survey indicate a prominent role for the water, sanitation and hygiene sector in stunting reduction. *BMC Public Health*, *16*(1), 669. https://doi.org/10.1186/s12889-016-3339-8

Yulianti, I., Nurasmi, N., & Padlilah, R. (2022). Stunting Risk Factors In Children Under 5 Years Old In Indonesian Border Regions (North Kalimantan-Malaysia). *International Journal of Health and Pharmaceutical (IJHP)*, *2*(2), 275–283. https://doi.org/10.51601/ijhp.v2i1.42

Zarkasi, R. N., Sifriyani, S., & Prangga, S. (2021). IDENTIFIKASI FAKTOR-FAKTOR YANG MEMPENGARUHI INDEKS PEMBANGUNAN MANUSIA DI KALIMANTAN MENGGUNAKAN REGRESI PANEL. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *15*(2), 277–282. https://doi.org/10.30598/barekengvol15iss2pp277-282