Jurnal Ilmu Ekonomi dan Pembangunan
P-ISSN 1412 – 2200 | E-ISSN 2548 – 1851
Vol. 23 No. 1, March, 2023, Page 1-8

NONPARAMETRIC TRUNCATED SPLINE REGRESSION MODELING ON POVERTY RATES IN NORTH SUMATRA

Hasrat Ifolala Zebua*
Badan Pusat Statistik, Jakarta, Indonesia
*Corresponding author: hasrat20001@mail.unpad.ac.id

ARTICLE INFO

Article history
Received : 18 February 2021
Revised : 20 February 2023
Accepted : 1 March 2023

Keywords
Poverty Rate;
HDI;
Trunated Spline;
GCV

JEL classification
I32; C14; O15

ABSTRACT

This study aims to investigate the relationship between the Human Development Index (HDI) and poverty levels in North Sumatra. Using data from Central Bureau of Statistics (BPS), the study employs a nonparametric truncated spline regression model to analyze the relationship. The findings reveal that HDI significantly impacts poverty levels, with higher HDI associated with lower poverty rates. The model used in this study offers a robust approach to understanding the dynamics between HDI and poverty, and the results underscore the importance of improving HDI to reduce poverty. The research highlights an R-Squared value of 82.35%, indicating a strong correlation between HDI and poverty in the region.

1. INTRODUCTION

Poverty eradication remains a pressing global concern, reflected in the first goal of the Sustainable Development Goals (SDGs), which aims to eliminate poverty in all its forms everywhere (Zuhdiyaty & Kaluge, 2017). Achieving this goal is a critical priority for governments worldwide, as it directly impacts the economic and social well-being of populations. In response, the Indonesian government has implemented various poverty alleviation and eradication programs aimed at improving basic living conditions and, ultimately, reducing the poverty rate (Murdiyana & Mulyana, 2017).

North Sumatra, the most populous province on the island of Sumatra and the fourth-most populous in Indonesia, presents a unique case in the national context of poverty (Sari, 2019). In 2020, the province's poverty rate stood at 8.75 percent, slightly lower than the national average of 9.78 percent. Despite this, North Sumatra's poverty rate is the fifth highest among the provinces in Sumatra, indicating a significant challenge that requires sustained government attention. Given that more than 5 percent of Indonesia's population resides in North Sumatra, it is imperative that poverty alleviation efforts in the province are intensified and not solely concentrated in select regions.

The persistent poverty in North Sumatra highlights the need for a nuanced understanding of the factors contributing to poverty and the effectiveness of existing government interventions (Ardianto, 2016). A comprehensive analysis of the socio-economic conditions in the province is necessary to identify the gaps in current strategies and to develop targeted measures that address the root causes of poverty. This study aims to contribute to this understanding by examining the socio-economic dynamics in North Sumatra, with a particular focus on the effectiveness of poverty alleviation programs and the distribution of poverty across different regions within the province.

In this context, the study will explore the demographic and economic factors that influence poverty in North Sumatra, including the role of employment, education, and access to basic services. The research will also assess the impact of government programs on poverty reduction and identify areas where further intervention is needed. By providing a detailed analysis of the poverty landscape in North Sumatra, this study aims to inform policy decisions and contribute to the development of more effective poverty alleviation strategies in the province.

The significance of this research lies in its potential to influence policy-making at both the provincial and national levels. By shedding light on the specific challenges faced by North Sumatra, the study will help policymakers design more targeted and effective interventions that can accelerate poverty reduction in the province and contribute to the broader goal of eradicating poverty in Indonesia. Additionally, the findings of this research may have implications for other provinces in Indonesia that face similar challenges, thereby contributing to a more comprehensive national strategy for poverty eradication.

In summary, this study seeks to provide a detailed examination of poverty in North Sumatra, with a focus on the effectiveness of government interventions and the socio-economic factors that contribute to poverty in the province. The findings will inform policy decisions and contribute to the development of more effective poverty alleviation strategies, ultimately supporting the achievement of the SDGs in Indonesia.

2. RESEARCH METHODS

The study utilizes 2020 data from Central Bureau of Statistics (BPS) to analyze the relationship between poverty and the Human Development Index (HDI) across 33 districts/cities in North Sumatra (Badan Pusat Statistik Provinsi Sumatera Utara, 2021). The poverty rate, represented by the Head Count Index (HCI-P0), indicates the percentage of people living below the poverty line (Balai Diklat Keuangan Makassar, 2021). The HDI, a composite index introduced by UNDP, measures achievements in longevity, education, and standard of living (Wiguna & Iskandar, 2021). The poverty rate is the dependent variable, while HDI is the independent variable. Data analysis was performed using R version 4.0.2.

Nonparametric regression is employed in this study to explore the relationship between the response and predictor variables without assuming a specific parametric form for the underlying function (Sugiyono, 2017). This approach is particularly useful for modeling complex relationships where traditional parametric methods may be inadequate. The general form of the nonparametric regression model used is:

$$y_i = f(x_i) + \varepsilon_i; i = 1,2, \dots, n \quad (1)$$

Where y_i represents the response variable, $f(x_i)$ is an unknown non-parametric function, and ε_i is the random error term, assumed to be identically and independently distributed.

This study specifically utilizes the nonparametric truncated spline regression model due to its flexibility in handling data with varying behavior across different intervals. Splines, particularly truncated splines, are advantageous in situations where the data exhibit changes in patterns at different sub-intervals, which are marked by knot points (Tripena, 2011; Wahba, 1990). These knot points allow the model to adapt to changes in the relationship between variables across different ranges of the predictor variable.

The spline function, characterized by its order p and knot points $k_1, k_2, \dots, k_r$, can be expressed as:

$$f(x_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_i^j + \sum_{l=1}^{r} \beta_{P+l} (x_i - k_l)_+^P \quad (2)$$

Where p is a polynomial order or degree and kl is a knot point indicating a data pattern change.

So a nonparametric regression model with a spline function can be written as follows (Ardiansyah, 2019):

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_i^j + \sum_{l=1}^{r} \beta_{P+l}(x_i - k_l)_+^P + \varepsilon_i \quad (3)$$

With truncated functions as follows:

$$(x_i - k_l)_+^P = \begin{cases} (x_i - k_l)^P & , x_i \geq k_l \\ 0 & , x_i < k_l \end{cases} \quad (4)$$

One simple illustration of a truncated spline function with three knots at $t = k_1 \leq t = k_2 \leq t = k_3$ is as follows (Budiantara, 2011):

$$S_3(t) = \beta_1 t + \beta_2(t - k_1)_+^1 + \beta_3(t - k_2)_+^1 + \beta_4(t - k_l)_+^1 \quad (5)$$

Or can be presented in the form:

$$S_3(t) \begin{cases} \beta_1 t & , t < k_1 \\ \beta_1 t + \beta_2(t - k_1)_+^1 & , k_1 \leq t < k_2 \\ \beta_1 t + \beta_2(t - k_1)_+^1 + \beta_3(t - k_2)_+^1 & , k_2 \leq t < k_3 \\ \beta_1 t + \beta_2(t - k_1)_+^1 + \beta_3(t - k_2)_+^1 + \beta_4(t - k_l)_+^1 & , t \geq k_3 \end{cases} \quad (6)$$

The selection of optimal knot points, crucial for the truncated spline model, is performed using the Generalized Cross Validation (GCV) method. GCV, proposed by Craven & Wahba (1979), is known for its asymptotic optimal properties and is effective in determining the model that best fits the data. The optimal knot points are those that minimize the GCV value, ensuring a balance between model complexity and goodness of fit.

$$GCV(k_1, k_2, \dots k_r) = \frac{MSE(k_1, k_2, \dots k_r)}{(\frac{1}{n} trace[I - A(k_1, k_2, \dots k_r])^2} \quad (7)$$

Where I is the identity matrix, n is the number of samples, $A(k_1, k_2, \dots k_r)$ is the matrix $X(X^T X)^{-1} X^T$ and $MSE(k_1, k_2, \dots k_r)$ with the formula by Hardle (2004).

The formula by Hardle (2004) is as follows:

$$MSE(k_1, k_2, \dots k_r) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (8)$$

The data analysis techniques involves several stages: 1) Descriptive Analysis to understand district/city characteristics in North Sumatra; 2) Exploratory Data Analysis using scatterplots to visualize variable relationships; 3) Optimal Knot Point Selection minimizing GCV and MSE; 4) Model Estimation using truncated spline regression; 5) Statistical Testing through simultaneous and partial tests for model significance; and 6) Interpretation to analyze the relationship between HDI and poverty in North Sumatra.

3. RESULTS AND DISCUSSION

Descriptive analysis

Characteristic poverty levels and HDI in North Sumatra are shown in Table 1 below.

Table 1. Descriptive Statistics

Variable	Min	Max	Mean	Var
Poverty Levels	3.88	25.69	10.80	22.167
HDI	61.51	80.98	70.75	21.109

Source: Processed data (2021)

Table 1 on the previous page provides a descriptive overview of the poverty levels and Human Development Index (HDI) across districts/cities in North Sumatra in 2020 (BPS, 2021). The average poverty level is 10.80%, with significant variability across the region, as indicated by the variance of 22.167. The lowest poverty rate is found in Deli Serdang at 3.88%, suggesting better living standards or access to resources, while West Nias shows the highest poverty rate at 25.69%, indicating severe deprivation. Similarly, the HDI values vary, with West Nias again at the lower end with an HDI of 61.51, signaling poor access to education, health, and income, whereas Medan stands out with the highest HDI of 80.98, reflecting better socio-economic conditions. These variations suggest a substantial disparity in development and poverty levels across different regions of North Sumatra.

Exploratory Data Analysis

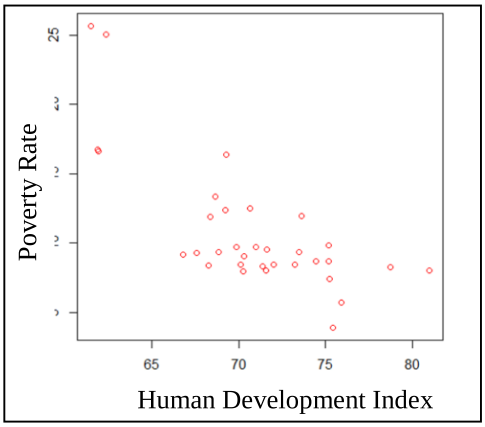

Figure 1. Relationship patterns of poverty levels and HDI of districts/cities in North Sumatra in 2020

Source: Processed data (2021)

Figure 1 illustrates the scatterplot between the poverty levels and HDI across districts/cities in North Sumatra, showing an unclear, non-linear relationship. The data clusters around the lower-middle HDI range, suggesting that most regions have moderate levels of human development. This scattered, non-uniform pattern necessitates the use of nonparametric regression methods like the truncated spline, which can flexibly model complex, non-linear relationships between variables without assuming a specific functional form.

Optimal Node Point Selection

Identifying the optimal knot points in the truncated spline regression is crucial for accurately modeling the relationship between poverty levels and HDI. Knot points indicate where the relationship between variables changes, allowing for a more nuanced model. The selection process involves minimizing the Generalized Cross Validation (GCV) and Mean Squared Error (MSE) values across different model orders and knot combinations, as shown in Table 2 on the next page. The best model is achieved with a second-order polynomial and three knot points, yielding the lowest GCV (5.270220) and MSE (3.794171). These optimal knot points at HDI values of 61.93, 62.36, and 66.79, indicate significant shifts in the relationship between HDI and poverty, suggesting different poverty dynamics across these HDI thresholds.

Table 2. Comparison of GCV and Minimum MSE values

Orde	Sum of *Knot*	GCV	MSE
2	1	7.423502	6.135125
2	2	7.111582	5.492048
2	3	5.270220	3.794171
3	1	7.960632	6.147743
3	2	7.478741	5.775594
3	3	6.843040	4.926488
4	1	8.260476	6.378922
4	2	7.436390	6.013795
4	3	6.217188	5.892889

Souce: Processed data (2021)

Based on Table 2, the model with the minimum GCV and MSE values is the model with order two and three knot points, with the value GCV of 5.270220 and the MSE of 3.794171. The nonparametric model truncated spline is the best model with order two and three knot points. The optimal knot points for the HDI variable are 61.93, 62.36, and 66.79.

Modeling

Estimates of models with optimal knot points are obtained as follows:

Table 3. Estimate Regression Coefficient

Coefficient	Estimate
β_0	1403.98000
β_1	-22.40819
β_2	42.94134
β_3	-2.70357
β_4	2.85437

Source: Processed data (2021)

From the estimates obtained, the regression model of the truncated spline can be formed:

$$\hat{y} = 1403.98 - 22.40819x_1 + 42.94134(x_1 - 61.93)^1_+ - 23.70357(x_1 - 62.36)^1_+ \\ + 2.854375(x_1 - 66.79)^1_+$$

Figure 2 below shows the shape of the data spread pattern between the response variable and the predictor variable in the truncated spline function.

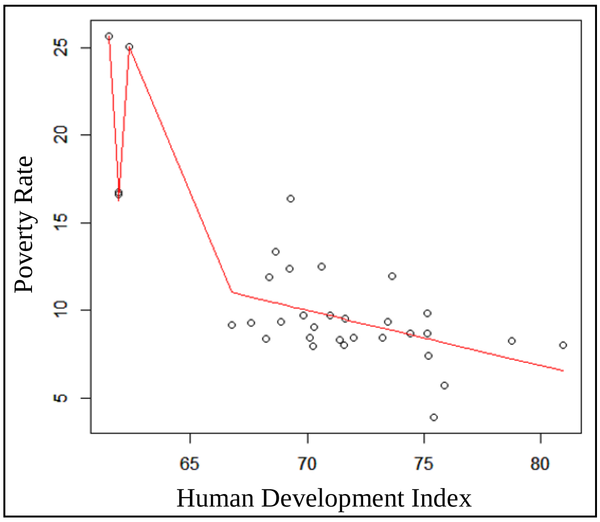

Figure 2. Graph of nonparametric truncated spline functions
Source: Processed data (2021)

Model Parameter Significance Testing

The results of the simultaneous test can be seen in Table 4 below.

Table 4. ANOVA Nonparametric Regression Model Truncated Spline

Source	df	SS	MS	F
Regression	4	584.14	146.03	32.66
Error	8	125.20	4.47	P-Val
Total	32	709.34	-	(0.000)

Source: Processed data (2021)

The results of the simultaneous test, as shown in Table 4 above, yield an F-statistic of 32.66 with a p-value of 0.000. Since the F-statistic significantly exceeds the critical value at a 5% significance level (F-table = 2.93), we reject the null hypothesis that all regression coefficients are zero. This indicates that the truncated spline model as a whole is statistically significant. The partial significance tests (Table 5) further confirm that each knot point contributes uniquely to the model, with all p-values being well below the 5% significance threshold. This underscores the importance of accounting for different HDI segments in explaining the variability in poverty levels.

So the conclusion is that the model has at least one significant parameter. To find out which parameter is substantial, it is necessary to test the significance of the parameter partially. The results of partial tests can be seen in Table 5 below.

Table 5. Partial Test Results of Nonparametric Model

Coef.	Estimate	p-value
-22.408	-3.771	0.0007*
42.941	4.468	0.0001*
-23.703	-3.948	0.0004*
2.8543	5.650	0.0000*

Note: *Significant at α=5%
Source: Processed data (2021)

Based on Table 5 above, the entire parameter is significant at a degree of significance (α) 5 percent. This is seen from a smaller p-value when compared to a value α. So it is concluded that the HDI variable significantly influences North Sumatera's poverty rate. Beside that, the R-Squared ratio was 82.35 percent. This means that the variation in the poverty rate described by the HDI is 82.35 percent, while the remaining 17.65 percent is explained by other variables outside the model.

Model Interpretation

The interpretation of the model reveals the nuanced impact of HDI on poverty across different segments. For HDI values below 61.93, a one-unit increase in HDI results in a substantial decrease in poverty by approximately 22.41%. This effect is particularly strong in regions like West Nias and South Nias, which are characterized by low HDI. However, in the HDI range of 61.93 to 62.36, a paradoxical increase in poverty by about 20.53% with rising HDI is observed, indicating that marginal improvements in HDI within this range may not be sufficient to overcome existing socio-economic challenges. For HDI between 62.36 and 66.79, the poverty-reducing effect of HDI resumes but at a much lower rate (3.17%). Above 66.79, further increases in HDI continue to reduce poverty, albeit very slightly (0.32%), suggesting that regions with relatively high HDI have already addressed the most critical factors affecting poverty.

Meanwhile, when the HDI ranges from 61.93 to 62.36 when the HDI increases by 1 unit, the poverty level tends to increase by 20.533 percent. The district/city area included in this category is Nias district. The HDI ranges from 62.36 to 66.79, and if the HDI increases by 1 unit, the poverty level tends to fall by 3.170 percent. The district/city area included in this category is North Nias district. When the HDI is greater than 66.79, and when the HDI increases by 1 unit, the poverty rate tends to decrease by 0.316045 percent. The areas included in this category are 29 other districts/cities, assuming other variables are constant.

4. CONCLUSION

The study demonstrates the effectiveness of using a nonparametric regression model with a truncated spline for handling data across different intervals. Specifically, the second-order model with knot points at 61.93, 62.36, and 66.79 was found to be the most optimal for this approach. This model reveals that the Human Development Index (HDI) variable significantly impacts poverty levels in North Sumatra, explaining 82.35% of the variation within the model. These findings underscore the importance of HDI as a key determinant of poverty, providing valuable insights for policymakers aiming to alleviate poverty in the region. Future research should explore the integration of spatial aspects and additional predictor variables to further refine the model's accuracy. Comparing this method with alternative approaches could also offer deeper insights and validate the robustness of the truncated spline model. Additionally, extending the analysis to other regions could help determine the generalizability of these findings across different socio-economic contexts.

5. REFERENCES

Ardiansyah. (2019). *Pemodelan Faktor yang Mempengaruhi Kemiskinan di Provinsi Sulawesi Selatan dengan Regresi Nonparametri Spline* [Universitas Islam Negeri Alauddin Makassar]. https://repositori.uin-alauddin.ac.id/14909/

Ardianto, T. (2016). Analisis Faktor-Faktor Yang Mempengaruhi Tingkat Kemiskinan Kabupaten/Kota di Provinsi Jawa Timur Tahun 2009-2013. *Jurnal Ilmiah Mahasiswa FEB, 4*(2). https://jimfeb.ub.ac.id/index.php/jimfeb/article/view/2932

Badan Pusat Statistik. (2021). *Indeks Pembangunan Manusia 2020*. Jakarta: Badan Pusat Statistik. https://www.bps.go.id/id/publication/2021/04/30/8e777ce2d7570ced44197a37/indeks-pembangunan-manusia-2020.html

Badan Pusat Statistik Provinsi Sumatera Utara. (2021). *Indikator Kesejahteraan Rakyat Provinsi Sumatera Utara 2020*. Medan: Badan Pusat Statistik Provinsi Sumatera Utara. https://sumut.bps.go.id/id/publication/2021/01/29/eca629e4d550758eedff416c/indikator-kesejahteraan-rakyat-provinsi-sumatera-utara-2020.html

Balai Diklat Keuangan Makassar. (2021). *Menuju Desa Mandiri dan Sejahtera: Dari Mana Dimulai?* Kementerian Keuangan: Badan Pendidikan Dan Pelatihan Keuangan. https://bppk.kemenkeu.go.id/balai-diklat-keuangan-makassar/artikel/menuju-desa-mandiri-dan-sejahtera-dari-mana-dimulai-014554

Budiantara. (2011). Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter. *PROSIDING SEMINAR NASIONAL MIPA UNDIKSHA 2011*. https://ejournal.undiksha.ac.id/index.php/semnasmipa/article/view/2736

Craven, P., & Wahba, G. (1978). Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by The Method of Generalized Cross Validation. *Numerische Mathematik, 31*, 377–403. https://doi.org/10.1007/BF01404567

Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge: Cambridge University Press.

Murdiyana, M., & Mulyana, M. (2017). Analisis Kebijakan Pengentasan Kemiskinan di Indonesia. *Jurnal Politik Pemerintahan, 10*(1), 73–96. https://doi.org/10.33701/jppdp.v10i1.384

Sari, J. (2019). Analisis Tingkat Kemiskinan Masyarakat Di Provinsi Sumatera Utara. *Al-Masharif: Jurnal Ilmu Ekonomi Dan Kesilaman, 7*(2), 290–304. https://jurnal.uinsyahada.ac.id/index.php/Al-masharif/article/view/2190

Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.

Tripena, A. (2011). Penentuan Model Regresi Spline Terbaik. *PROSIDING SEMINAR NASIONAL STATISTIKA UNIVERSITAS DIPONEGORO 2011*, 92–102. http://eprints.undip.ac.id/33908/

Wahba, G. (1990). *Spline Models for Observational Data*. Madison: University of Wisconsin-Madison.

Wiguna, A. D. A., & Iskandar, D. D. (2021). Analisis Pengaruh PDB Perkapita, Pengeluaran Pemerintah dan Pengangguran Terhadap IPM di Negara-Negara ASEAN Tahun 2009-2018. *Diponegoro Journal of Economics*, *10*(3), 148–161. https://ejournal3.undip.ac.id/index.php/jme/article/view/39683

Zuhdiyaty, N., & Kaluge, D. (2017). Analisis Faktor - Faktor Yang Mempengaruhi Kemiskinan di Indonesia Selama Lima Tahun Terakhir. *Jurnal Ilmiah Dan Ekonomi Asia*, *11*(2), 27–31. https://doi.org/10.32812/jibeka.v11i2.42.