*Parameter: Journal of Statistics*
Available at https://bestjournal.untad.ac.id/index.php/parameter
Vol. 2 No. 3 2022, 11 - 16
DOI: *https://doi.org/10.22487/27765660.2022.v2.i3.15743*
eISSN: 2776-5660

# MODELING OF POVERTY LEVEL IN CENTRAL SULAWESI USING NONPARAMETRIC KERNEL REGRESSION ANALYSIS APPROACH

**Nur Sakinah[1*], Nurfitra[2], Nurmasyita Ihlasia[3], Lilies Handayani[4]**
[1,2,3,4]Department of Statistics, Faculty of Mathematics and Natural Sciences, Tadulako University, Soekarno-Hatta Street, Palu 94118, Central Sulawesi, Indonesia.

***e-mail**: nursakinah895@gmail.com

## ABSTRACT

*Poverty is defined as a person's inability to meet their basic needs. The level of poverty that exists can be used to assess the good or bad of a country's economy. The kernel regression method is used in this study to model the poverty rate in Central Sulawesi in 2020. According to the findings of this study, comparing poverty rate predictions for the Gaussian Kernel function and the Epanechnikov Kernel function with optimal bandwidth can be said to use different kernel functions with optimal bandwidth for each - each of these kernel functions will produce the same curve estimate. So, in kernel regression, the selection of the optimal bandwidth value is more important than the selection of the kernel function. Because of the use of various kernels functions with optimal bandwidth values results in almost the same curve estimation.*

**Keywords:** *Kernel Regression, Poverty Level, Bandwidth*

**Cite:** Sakinah, N., Nurfitra, Ihlasia, N., & Handayani, L. (2022). Modeling of Poverty Level in Central Sulawesi Using Nonparametric Kernel Regression Analysis Approach. *Parameter: Journal of Statistics*, *2*(3), 11-16. https://doi.org/10.22487/27765660.2022.v2.i3.15743



[*]Received: Dec 2021; Reviewed: Dec 2022; Published: Dec 2022

## INTRODUCTION

Poverty is a socioeconomic issue that is always present in society. The level of poverty that exists can be used to assess the good or bad of a country's economy (Windra, et al., 2016). Poverty is defined as a person's inability to meet the bare necessities of life (Kuncoro, 2000). This disability is distinguished by low income capabilities. Low income ability means a lower ability to meet the average standard of living. Poverty is also a multifaceted and complex issue. This means that the problem of poverty is linked to problems in other areas of life. As a result, in order to alleviate poverty, it is necessary to examine various factors that may influence the occurrence of poverty.

In general, in the period March 2013-September 2019 the poverty rate in Central Sulawesi decreased both in terms of number and percentage, but since 2015 the poverty rate has shown fluctuations that tend to increase until 2017 before finally showing a downward trend again until 2019. After the covid pandemic -19 in 2020, the number of poor people and the percentage of poor people will increase again (BPS, 2021). The number of poor people in Central Sulawesi in September 2020 reached 403.74 thousand people. There was an increase in the number of poor people by five thousand people compared to March 2020. Meanwhile, compared to September 2019, the number of poor people decreased by 0.29 thousand people. Based on area of residence, in the period March 2020-September 2020, the number of poor people in urban areas increased by 6.7 thousand people and rural areas decreased by 1.7 thousand people. The percentage of poverty in urban areas rose from 8.76 percent to 9.21 percent, while in rural areas it rose from 14.69 percent to 14.76 percent (BPS, 2021).

Regression analysis is a statistical technique for determining the pattern of the relationship between response and predictor variables. Regression analysis is therefore a statistical inference method for a regression function or regression curve (Eubank, 1999). Because not all cases of data in a regression analysis have parametric patterns such as linear, quadratic, or cubic patterns, other regression approaches such as nonparametric and semiparametric regression approaches are required. Hardle (1990) states that if the data pattern tends to follow a linear, quadratic, or cubic model, the approach is appropriate. A parametric regression approach is used when the data pattern fits a regression curve, whereas a nonparametric regression approach is used when the data pattern does not fit a regression curve (Budiantara, *et al*., 2010).

There are several nonparametric regression modeling methods, including Kernel, K-Nearest, MARS Fourier Series, Orthogonal Series, Wavelets, and Neural Networks. The application of kernel functions to estimate the regression function has proven to be quite beneficial. A "coarse" function, i.e. a random function, should be a "smoothed out" function. Data analysis with the kernel enables the use of multiple bandwidths and the selection of the final estimator based on a qualitative evaluation of the estimation results. Purwanti (2019) used the Kernel Epanechnikov and Kernel Gaussian regressions on the 2016-2018 Indonesian Sharia Stock Index in her research on kernel regression. The Poverty Level Modeling in Central Sulawesi will be used in this study, which will employ a nonparametric kernel regression approach with the Nadaraya-Watson estimation and the Kernel Epanechnikov and Kernel Gaussian regression functions. It is hoped that the model developed will be used to develop intervention measures and monetary policy decisions. More effective and efficient.

## MATERIALS AND METHODS

### 1. Data Sources

The data used in this study are secondary data obtained from the Central Statistics Agency (BPS) in the publication of Central Sulawesi in 2021. The objects in this study consist of 13 districts/cities in Central Sulawesi Province.

### 2. Research Variables

The variables used are the poverty rate (Y), the labor force participation rate ($X_1$), and the open unemployment rate ($X_2$).

## 3. Methods

The method used in this research is Nonparametric Kernel Regression. Kernel regression is one of the methods in non-statistics parametrics method for estimating conditional expectations of random variables using the kernel function. This will perform kernel smoothing using a weighted average of the data on kernel regression. As for the kernel regression models (Zulfikar, 2008), they are as follows:

$$Y_i = m(x_i) + \varepsilon_i \qquad , i = 1,2,3 \dots, n \tag{1}$$

where:
$Y_i$ : Independent variable
$\varepsilon_i$ : Error
$m(x_i)$ : Regression curve

The use of a kernel function to estimate the regression function can be very useful. A "rough" function, or a random function that has been "smoothed out," must be a random function. Kernel functions in regression include Uniform, Triangle, Epanechnikov, Quartic, Triweight, Cosine, and Gaussian kernel functions.

## 4. Data Analysis

The method used in this research is Kernel Regression Analysis, namely the Epanechnikov Kernel Function and the Gaussian Kernel. The stages of analysis in this research are:

1. Descriptive statistical analysis
2. Create a scatter diagram between the response variable (Y) and each predictor variable (X)
3. Determine the bandwidth value to be used for each predictor variable
4. Calculate the MSE value of the two predictor variables
   For the MSE values of the tho predictor variables, can use the following equation:
   $$MSE(\hat{m}) = \frac{1}{n}\ (y_i - \hat{m}(x_i))^2 \tag{2}$$
5. Choose the optimum bandwidth based on the smallest MSE value
   The approach method in selecting the optimum bandwith is the Cross Validation (CV). The CV function is the following equation (Hardle, 1990):
   $$CV\ (h) = \ n^{-1} \sum_{j=1}^{n} \left[Y_j - \hat{m}_{h,j}(X_j)\right]^2 \tag{3}$$
6. Make a graph of the prediction results with two CV methods.
7. Comparing the Epanechnikov kernel function and the Gaussian kernel function based on the graph of the prediction results
8. Conclusion.

## RESULTS AND DISCUSSION

### Descriptive Statistical Analysis

Descriptive statistics of each research variable can be seen in Table 1.

Table 1. Descriptive Statistics

| Variables | Mean | Variance | Minimum | Maximum |
|---|---|---|---|---|
| Y | 13 436 | 9.88 | 6.80 | 17:39 |
| $X_1$ | 69.34 | 13.66 | 63.93 | 75.33 |
| $X_2$ | 3746 | 2923 | 2:39 | 8:38 |

Based on the above output, to the data rate of poverty (Y) has the average - average of 13 436 percent. The lowest poverty rate is Palu City, which is 6.80 percent, while the highest poverty rate is Donggala Regency, which is 17.39 percent. In the LFPR data ($X_1$) the average LFPR value is 69.34%. The lowest percentage of LFPR value is in Donggala County, which is 63.93%, while the highest

LFPR is Tojo Una-Una Regency, which is 75.33%, for TPT data ($X_2$) for Central Sulawesi province in 2020 the average is 3.746%, with the lowest TPT being Poso at 2.39%, while the highest TPT is Palu city at 8.38%.

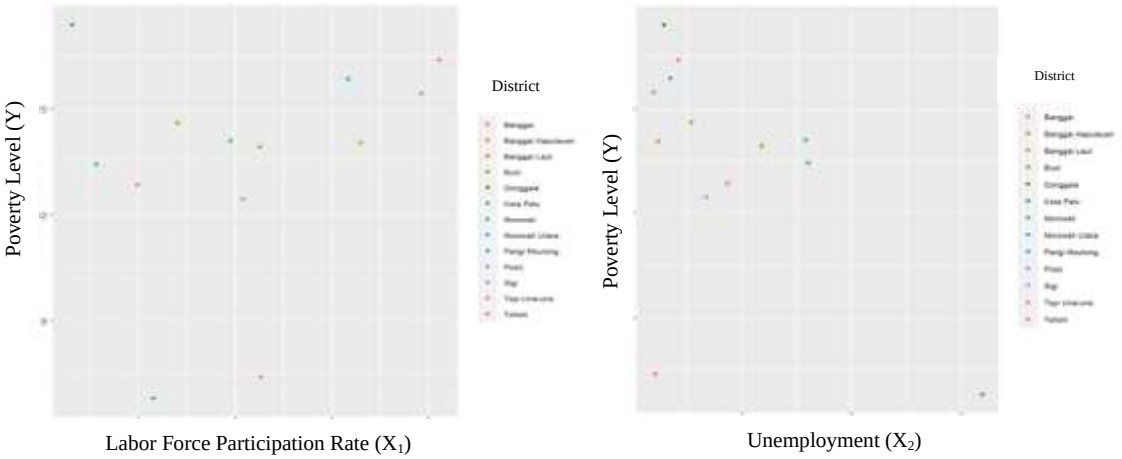


Figure 1. Plot of Distribution between Poverty Levels (Y): with Labor Force Participation Rate ($X_1$), Unemployment ($X_2$)

Figure 1 depicts a pattern of relationships that do not form a certain pattern based on the relationship pattern formed between the poverty level, which is the response variable, and each predictor variable, including the level of labor force participation (X1) and unemployment (X2). This demonstrates that there is a nonparametric component in which the function of the regression curve is unknown, so the model is estimated using nonparametric regression.

**Selection of Optimum Bandwidth**

Table 2. Optimum Bandwidth Value

| Variable | Type of Method CV | Bandwidth | R-Square | MSE |
|---|---|---|---|---|
| $X_1$ | CV.AIC | 74553205 | 0.043658 | 2.226314 |
| | CV.LS | 2.731169 | 0.254288 | 1.376219 |
| $X_2$ | CV.AIC | 1.786362 | 0.396701 | 2.281878 |
| | CV.LS | 0.200959 | 0.558526 | 1.989131 |

According to Table 2, the MSE value in $X_1$ and $X_2$ using the CV.AIC method is 2.226314 and 2.281878, respectively, while the MSE value in $X_1$ and $X_2$ using the CV.LS method is 1.376219 and 1.989131. The smallest MSE value can be used to determine the optimal bandwidth for $X_1$ and $X_2$. According to the table above, the CV.LS method has the smallest MSE value for $X_1$, and the CV.LS method has the smallest MSE value for $X_2$. As a result, CV.LS is the best model in this analysis, with a bandwidth of 2.731169 for X1 and 0.200959 for $X_2$.

**Comparison of Kernel Function Epanechnikov and Gaussian Kernel Functions**

Table 3. Estimated Epanechnikov Kernel Function and Gaussian kernel function

| Variables | Epanechnikov | Gaussian |
|---|---|---|
| $X_1$ | 5.924369 | 2.676103 |
| $X_2$ | 1.255207 | 0.566991 |

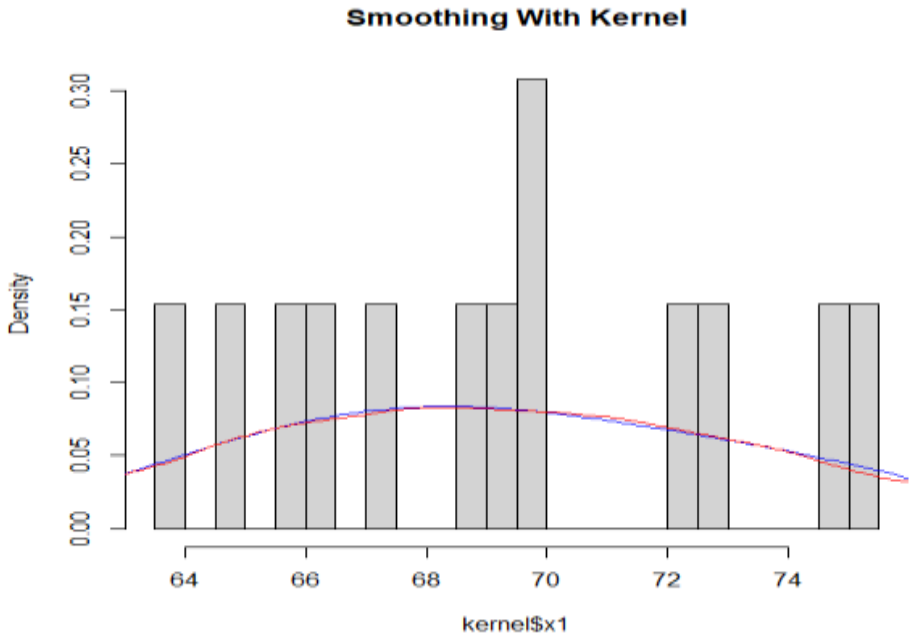


Figure 2. Histogram function Kernel In $X_1$

Based on the estimated results of a variable plot Labor Force Participation Rate (LFPR) for Gaussian Kernel function and Kernel Epanechnikov function using optimal bandwidth, the lines are very close together, implying that using different kernel functions with optimal bandwidth for each kernel function will result in the same curve estimation.

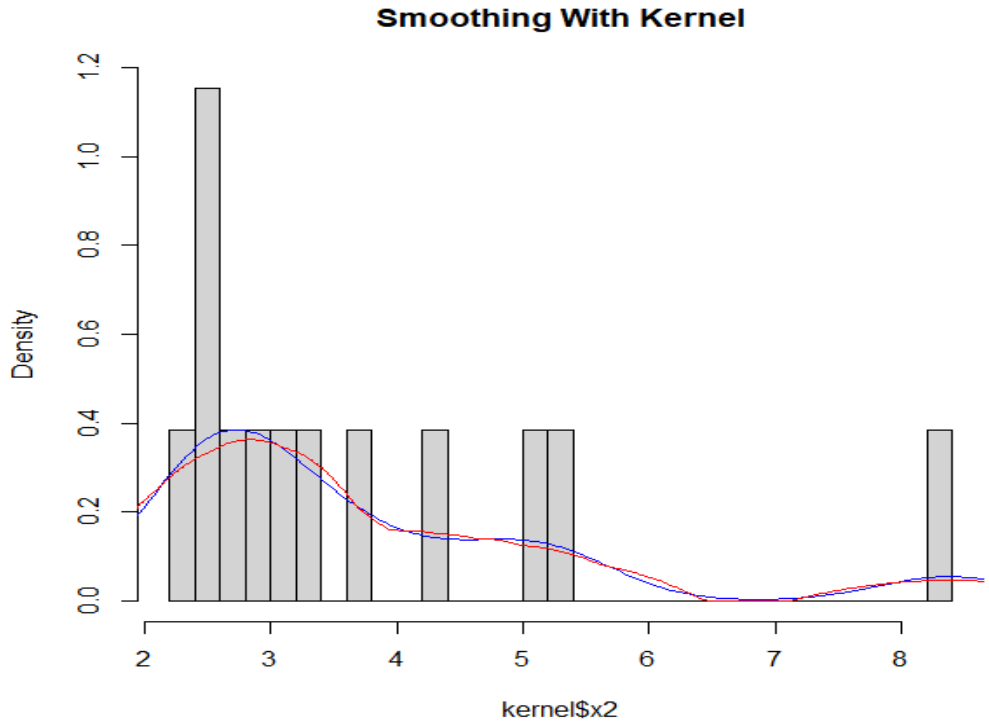


Figure 3. Histogram of Kernel Functions at $X_2$

The lines are very close on the plot of the estimation results of the Open Unemployment Rate (TPT) variable for the Gaussian Kernel function and the Epanechnikov Kernel function using the optimal bandwidth, so it can be said that using a different kernel function with the optimal bandwidth for each of these kernel functions will produce the same curve estimate.

## CONCLUSION

Based on the findings of the analysis, which included a comparison of poverty rate predictions for the Gaussian Kernel function and the Epanechnikov Kernel function using optimal bandwidth, it is possible to conclude that using different kernel functions with optimal bandwidth for each kernel function will result in the same curve estimation. As a result, in kernel regression, the selection of the optimal bandwidth value is more important than the selection of the kernel function. Because using different kernel functions with optimal bandwidth values yields nearly identical curve estimation.

## REFERENCE

BPS. (2021, November 15). *Kemiskinan Dan Ketimpangan*. Retrieved from Badan Pusat Statistika https://www.bps.go.id/subject/23/kemiskinan-dan-ketimpangan.html.

Budiantara, I. N., Lestari, B., & Islamiyati, A. (2010). *Estimator Spline Terbobot dalam Regresi Nonparametrik dan Semiparametrik Heteroskedastisitas untuk Data Longitudinal*. Hibah Penelitian Kompetensi. LPPM Institut Teknologi Sepuluh Nopember, Surabaya.

Eubank. (1999). *Nonparametric Regression and Spline Smoothing 2nd Edition*. Marcel Deker, New York.

Hardle, W. (1990). *Regresi Nonparametrik Terapan* No 19. Pers Universitas Cambridge, AS.

Purwanti, I. (2019). *Regresi Nonparametrik Kernel Menggunakan Estimator Nadaraya-Watson Dalam Data Time Series (Studi Kasus : Tingkat Suku Bunga, Kurs, Inflasi, Jumlah Uang*. Universitas Islam Negeri Sunan Ampel Surabaya.

Kuncoro, M. (2000). *Ekonomi Pembangunan: Teori, Masalah dan Kebijakan*. UPP AMP YKPN.

Windra., Marwoto, P.B., Rafani,Y. (2016). Analisis Pengaruh Inflasi, Pertumbuhan Ekonomi, Dan Tingkat Pengangguran Terhadap Kemiskinan Di Indonesia. *Jurnal Progresif Manajemen Bisnis*, *14* (2), 19-27.

Zulfikar. (2008). Koreksi Bias Estimator Kernel dengan Bootstrap. *Jurnal Sains Dan Teknologi, 1*(2), 38–44.