JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI e-ISSN: 2614-8811
Published by Departement of Mathematics, Hasanuddin University, Indonesia
https://journal.unhas.ac.id/index.php/jmsk/index
Vol. 19, No. 3, May 2023, pp. 498-505
DOI: 10.20956/j.v19i3.25260

Modeling the Hotel Tax Revenue in Central Lombok using Nonparametric Regression

Joji Ardian Pembargi[1], Ena Setiawana[2], Rosanti Wahidatussolihah[3], Nurul Fitriyani[4*]

[1, 2, 3]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Mataram, Mataram, INDONESIA*
[4]*Department of Statistics, Faculty of Mathematics and Natural Sciences, University of Mataram, Mataram, INDONESIA*

Email address: *jojiardian15@gmail.com[1], enasetiawana@gmail.com[2], rosantiws29@gmail.com[3], nurul.fitriyani@unram.ac.id[4]*
[*]*Corresponding author*
Received: 30 January 2023; Accepted: 5 March 2023; Published: 5 May 2023

Abstract

In Central Lombok Regency, the hotel tax is one of the highest incomes contributing to Regional Original Revenue. A hotel tax is a tax on services provided by the hotel. This research aims to estimate the nonparametric kernel regression curve on hotel tax revenue data in Central Lombok. The method used is nonparametric kernel regression analysis with the seven kernel functions. The results of the analysis with the Generalized Cross Validation (GCV) criteria, the optimal bandwidth values generated by the seven kernel functions have varying values. Although the bandwidth values vary, the resulting estimation results are similar, and the comparison of the Mean Square Error (MSE) values of the seven kernel functions is not significantly different.

Keywords: bandwidth, curve estimation, hotel tax revenue, kernel functions, nonparametric kernel.

1. INTRODUCTION AND PRELIMINARIES

The establishment of several regions as Kawasan Ekonomi Khusus (KEK), or Special Economic Zones, is the latest method used by the government to accelerate equitable economic development in various areas in Indonesia. One of the targets for developing a Special Economic Area is increasing investment in arranging regions with geo-economics and geostrategic superiorities with diverse business activities. In 2022, there were eighteen Special Economic Areas (KEK) in Indonesia, including KEK Mandalika, located in Central Lombok, West Nusa Tenggara Province.

Based on Government Regulation No. 52 of 2014, the KEK Mandalika is appointed a Tourism KEK/Special Economic Area [14]. Tourism has a huge opportunity to take advantage of the regional economy, increase cultural values, and create new economic activities such as hotels, restaurants, and transportation. These various economic activities have positively impacted profits for business entities, local community economic movements, and increased Original Local Government Revenue [20].

Based on the law of the Republic of Indonesia Number 28 of 2009, regional taxes are divided into two types, namely provincial taxes and regency/city taxes. One of the regional taxes collected by district/city governments is hotel taxes [12]. Hotel tax is a tax on services provided by hotels, where hotels provide lodging services by charging a fee. Hotel facilities include motels, inns, tourist lodges, guesthouses, and lodging houses. Hotel tax is a source of local taxes with a high potential to increase local revenue [13].

Based on the Regional Original Revenue of Central Lombok Regency in 2020, that hotel taxes provide the third largest percentage after Street Lighting Tax and Land and Building Acquisition Fees Tax, which is 17 % sourced from 1 to 4 stars hotel tax, jasmine hotel tax, cottage tax, and lodging/ inn tax. However, it was observed that the government's goal in 2020 was that the hotel tax only reached 11.12 % and 38.90 % in 2021 [16]. The goal that the government has set is related to the arrangement of regional budget policies. If hotel tax revenues in the following year still do not reach the government's goal, it will cause a discrepancy in regional expenditures and original local government revenue. Therefore, a solution is needed to support the government in determining the goal of Original Local Government Revenue and the arrangement of the annual regional budget, namely by determining the forecast of an appropriate amount of hotel tax for the future. One of the forecasting methods that could be used for hotel tax calculations is nonparametric regression analysis with a kernel estimator.

Regression analysis is a method or technique for examining research hypotheses to determine whether there is an impact between one variable and another, which is stated in mathematical equation form (regression). Based on its application, regression analysis is classified into two types: parametric regression and nonparametric regression [2],[8]. This study will use nonparametric regression analysis to model hotel tax revenues in Central Lombok. Nonparametric regression is used since it does not depend on a particular model; hence it can be claimed that nonparametric regression offers high data flexibility. In addition, nonparametric regression can also be used to model the existence of nonlinearity in a model [10],[11]. The formulation of nonparametric regression is as follows:

$$y_i = f(x_i) + \varepsilon_i \quad ; i = 1, 2, 3, \dots, n \tag{1.1}$$

with:

y_i = estimated value of the regression
$f(x_i)$ = unknown nonparametric regression function
ε_i = residual
x_i = value of the independent variable

Smoothing techniques that can be applied in nonparametric methods include kernel estimators, histograms, wavelets, Fourier series, splines, orthogonal series, and others [3],[19]. One of the smoothing techniques that can be used in nonparametric regression is the kernel estimator. Kernel estimator can be applied to estimate regression function $f(x_i)$. Generally, the kernel density function is as follows:

$$K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right), \qquad untuk -\infty < x < \infty, \qquad h > 0 \tag{1.2}$$

$K(x)$ represents the kernel function, and h indicates the bandwidth.

The seven (7) kernel functions are presented as follows [10]:

Table 2.1. Kernel functions

Kernel functions	Equations for $K(x)$
Quartic	$\frac{15}{16}(1-x^2)^2$: $\lvert x\rvert \leq 1, 0$ others
Triweight	$\frac{35}{32}(1-x^2)^3$: $\lvert x\rvert \leq 1, 0$ others
Gaussian	$\frac{1}{\sqrt{2\pi}} exp\left(\frac{1}{2}(-x^2)\right)$: $-\infty \leq x \leq \infty$
Cosine	$\frac{\pi}{4} cos\left(\frac{\pi}{2}x\right)$: $\lvert x\rvert \leq 1, 0$ others
Epanechnikov	$\frac{3}{4}(1-x^2)$: $\lvert x\rvert \leq 1, 0$ others
Triangle	$(1-\lvert x\rvert)$: $\lvert x\rvert \leq 1, 0$ others
Uniform	$\frac{1}{2}$: $\lvert x\rvert \leq 1, 0$ others

Former studies have used several types of kernel functions in kernel nonparametric regression. Some used the quartic kernel function, such as in modeling the water discharge in the Jangkok Watershed, Lombok Island, by autoregressive pre-whitening [6]; also in predicting the daily rainfall and simulating monthly rainfall data of Dodokan Watershed in the statistical downscaling model [8],[5]. Other studies also used the Gaussian kernel function in the studies, such as expecting the temperature in Selaparang Lombok and predicting the monthly rainfall of Sembalun Station in a statistical downscaling model [7],[9]; also in modeling crude birth rate in West Nusa Tenggara Province [15]. This study estimated and determined the kernel regression curves using seven (7) types of kernel functions.

One form of the kernel estimator is the Nadaraya-Watson estimator [17]. The form of the Nadaraya-Watson estimator is as follows:

$$\hat{y}_i = \frac{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)} y_i = \sum_{i=1}^{n} W_{hi}(x)\ y_i$$

with,

$$W_{hi}(x) = \frac{K\left(\frac{x_i - x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)} \quad 1.3$$

The equation shows how bandwidth and the kernel function K affect the function $\widehat{y_i}$. Bandwidth is used to modify the smoothness of the estimated curve [1],[18]. The form of addition $\widehat{y_i}$ is,

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} n^{-1}W_{h1}(x_1)y_1 + n^{-1}W_{h2}(x_1)y_2 + \cdots + n^{-1}W_{hn}(x_1)y_n \\ n^{-1}W_{h1}(x_2)y_1 + n^{-1}W_{h2}(x_2)y_2 + \cdots + n^{-1}W_{hn}(x_2)y_n \\ \vdots \\ n^{-1}W_{h1}(x_n)y_1 + n^{-1}W_{h2}(x_n)y_2 + \cdots + n^{-1}W_{hn}(x_n)y_n \end{bmatrix}$$

One method to get the optimum bandwidth value is using the GCV criteria, which is defined as follows [4]:

$$GCV(h) = \frac{MSE}{\left(n^{-1}tr(\mathbf{I} - \mathbf{H}(h))\right)^2} \quad (1.4)$$

with,

$$MSE = n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (1.5)$$

I is the identity matrix and the matrix $\mathbf{H}(h)$ contains elements of equation (1.3).

Time series modeling in determining the independent variables can be formulated by $\{Z_i, i > 1\}$, by defining the lag value Z_{i-1} as the independent variable (X_i) and lag value Z_i as dependent variable (Y_i). The problem can be considered as a form of smoothing regression as follows [10]:

$$\{(X_i, Y_i)\} = \{\, (Z_{i-1}, Z_i)\}, i = 1, 2, \ldots, n \quad (1.6)$$

Moreover, The data used in this study is secondary data, namely monthly data on hotel tax revenues in Central Lombok from January 2016 to December 2020. The study object used as an independent variable is monthly data on hotel tax revenues in Central Lombok from January 2016 to 2020. Meanwhile, the dependent variable in this study is the rupiah exchange rate against the United States dollar from February 2016 to December 2020.

This study estimated the regression curve using a kernel estimator with seven (7) types of kernel functions. The following steps are used in this investigation with the help of R software: (1) Determining the independent and dependent variables for monthly hotel tax revenues in Central Lombok from January 2016 to 2020 based on equation (1.6); (2) Determining the optimal bandwidth of the seven kernel functions; (3) Using optimal bandwidth value, estimate the nonparametric regression curve of each kernel function; and (4) Examining the Mean Square Error (MSE) value provided by each kernel function to determine the quality of the model using equation (1.5).

2. MAIN RESULTS

The research object used as the independent variable is hotel tax revenue data for the period January 2016 to November 2020. Meanwhile, the dependent variable used is hotel tax revenue data in Central Lombok for the period February 2016 to December 2020. The following is a plot between the independent (X) and dependent (Y) variables on hotel tax revenue data in Central Lombok.

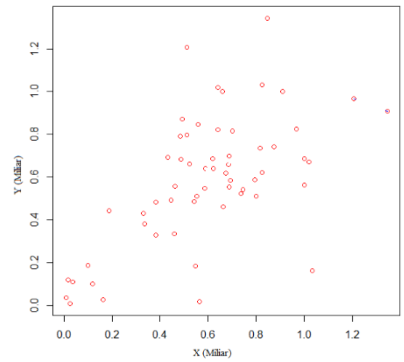


Figure 2.1. Plot of monthly hotel tax data for the period January 2016-December 2020

Figure 2.1 shows that the data pattern plot does not follow any particular pattern, such as linear, quadratic, or other. As a result, parametric regression is not appropriate for use in this study. Based on this, hotel tax revenue data in Central Lombok will be estimated using a nonparametric approach in this study. Then, estimate the regression curve using the optimal bandwidth value for each kernel function. The optimal bandwidth value is determined based on Generalized Cross Validation (GCV) criteria by minimizing the GCV value. The GCV value is determined by the value of Mean Squared Error (MSE), tr matrix $\mathbf{H}(h)$, and bandwidth. Based on the analysis results, the optimal bandwidth values for the seven kernel functions are presented in Table 2.2 as follows:

Table 2.2. Bandwidth values and GCV

Kernel function	Optimal bandwidth value	GCV value
Quartic	0.0505671	0.05572975
Triweight	0.0673276	0.05610093
Gaussian	0.0245646	0.05538893
Cosine	0.0373963	0.05500873
Epanechnikov	0.0278746	0.05475462
Triangle	0.0370763	0.05563462
Uniform	0.2499495	0.05632089

The GCV values generated by the seven kernel functions with the optimal bandwidth values are nearly identical. The optimal bandwidth values for the seven kernel functions in Table 2.2 are used to obtain the kernel nonparametric regression curve estimate results. The following are the model for estimating the curve of each kernel function:

a. Model with quartic kernel function (blue dash in Figure 2.2(a)):

$$\hat{y}_i = \frac{\sum_1^{59}\left(1-\left(\frac{x_i - x}{0.0505671}\right)^2\right)^2 y_i}{\sum_1^{59}\left(1-\left(\frac{x_i - x}{0.0505671}\right)^2\right)^2}$$

b. Model with triweight kernel function (blue dash in Figure 2.2(b)):

$$\hat{y}_i = \frac{\sum_1^{59}\left(1-\left(\frac{x_i - x}{0.0673276}\right)^2\right)^3 y_i}{\sum_1^{59}\left(1-\left(\frac{x_i - x}{0.0673276}\right)^2\right)^3}$$

c. Model with Gaussian kernel function (blue dash in Figure 2.2(c)):

$$\hat{y}_i = \frac{\sum_{i=1}^{59} exp\left(-\frac{1}{2}\left(\frac{x_i - x}{0.0245646}\right)^2\right) y_i}{\sum_{i=1}^{59} exp\left(-\frac{1}{2}\left(\frac{x_i - x}{0.0245646}\right)^2\right)}$$

d. Model with Cosine kernel function (blue dash in Figure 2.2(d)):

$$\hat{y}_i = \frac{\sum_1^{59} cos\left(\frac{\pi}{2}\left(\frac{x_i - x}{0.0373963}\right)\right) y_i}{\sum_1^{59} cos\left(\frac{\pi}{2}\left(\frac{x_i - x}{0.0373963}\right)\right)}$$

e. Model with epanechnikov kernel function (blue dash in Figure 2.2(e)):

$$\hat{y}_i = \frac{\sum_1^{59}\left(1-\left(\frac{x_i - x}{0.0278746}\right)^2\right) y_i}{\sum_1^{59}\left(1-\left(\frac{x_i - x}{0.0278746}\right)^2\right)}$$

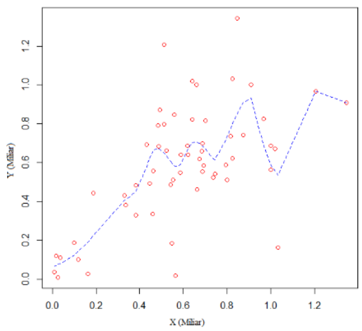
f. Model with triangle kernel function (blue dash in Figure 2.2(f)):

$$\hat{y}_i = \frac{\sum_1^{59}\left(1-\left|\frac{x_i - x}{0.0370763}\right|\right) y_i}{\sum_1^{59}\left(1-\left|\frac{x_i - x}{0.0370763}\right|\right)}$$

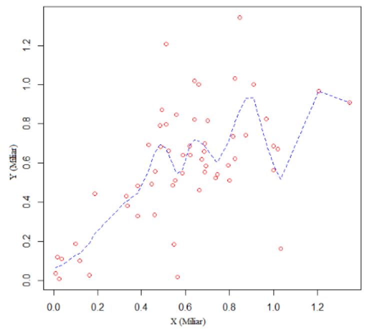
g. Model with uniform kernel function (blue dash in Figure 2.2(g)):

$$\hat{y}_i = \frac{1}{2}$$

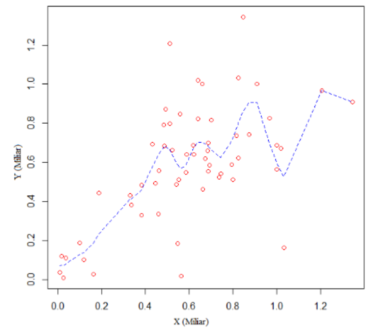
Based on the model obtained for seven (7) types of kernel function, the curve estimation results are shown in Figure 2.2 as follows:



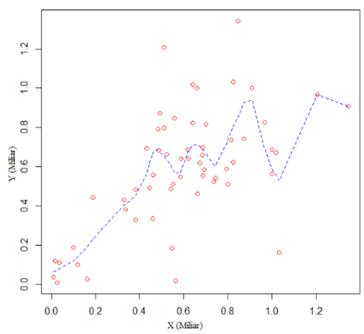
(a) Quartic



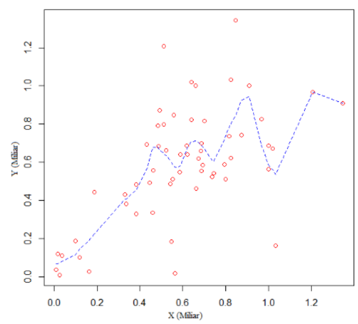
(b) Triweight



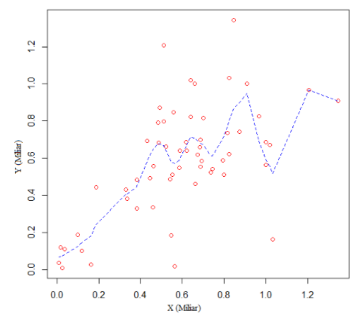
(c) Gaussian



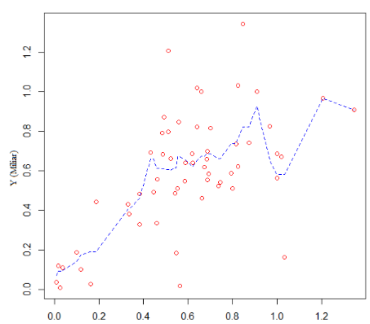
(d) Cosine



(e) Epanechnikov



(f) Triangle



(g) Uniform

Figure 2.2. The estimation results of each kernel function with optimal bandwidth (red dots represent the actual data and blue dash represents the prediction data)

Based on Figure 2.2, each kernel function with the optimal bandwidth generates estimation results that are almost identical to and follow the actual data pattern. Furthermore, the MSE values obtained in Table 2.3 can be used to determine the accuracy of the estimation results of each kernel function:

Table 2.3. MSE value of each kernel function

Kernel Function	MSE Value
Quartic	0.02554211
Triweight	0.02797275
Gaussian	0.02783461
Cosine	0.02320782
Epanechnikov	0.01868983
Triangle	0.01989417
Uniform	0.05034266

The MSE value indicates the accuracy of the estimate generated by the kernel estimator. The lower the MSE value, the better the estimation results. Therefore, it is expected to get the smallest possible MSE value. Based on the research results, the lowest MSE value obtained from epanechnikov kernel function and the highest MSE value obtained from the uniform kernel function. However, the MSE values for the seven (7) types of kernel functions are not significantly different. Therefore, the estimated nonparametric kernel regression curve with the seven (7) kernel functions applied to hotel tax revenue data in Central Lombok is not significantly different. Based on the independent and dependent variables used, the Central Lombok government should pay more attention to hotel tax revenue data, especially in the last month.

3. CONCLUSION

The nonparametric regression approach with the kernel estimator using the quartic, triweight, gaussian, cosine, epanechnikov, triangle, and uniform kernel functions on hotel tax revenue data in Central Lombok produces a relatively similar and good estimation value, based on the small MSE values for the seven (7) types of kernel functions. The effect of the optimum bandwidth value and MSE value also determines the estimation results obtained. The results show that the MSE values for different types of kernel functions are not significantly different, so the estimated nonparametric kernel regression curve applied to hotel tax revenue data in Central Lombok is not significantly different.

REFERENCES

[1] Chacon J. E., and Duong T., 2018. *Multivariate Kernel Smoothing and its Applications*, Taylor and Francis Group, New York.

[2] Denis D. J., 2016. *Applied Univariate, Bivariate, and Multivariate Statistics.* John Wiley and Sons Inc., Canada.

[3] Eubank R. L., 1999. *Nonparametric Regression and Spline Smoothing.* Marcel Dekker, New York.

[4] Fitriyani, N. and Budiantara, I.N., 2014. Metode Cross Validation dan Generalized Cross Validation dalam Regresi Nonparametrik Spline: Studi Kasus Data Fertilitas di Jawa Timur (translate: CV and GCV Methods in Spline Nonparametric Regression: A Case Study of Fertility Data in East Java). Proceeding of *Seminar Nasional Pendidikan Sains*, 1089–1095. Universitas Negeri Surabaya.

[5] Hadijati, M., Irwansyah and Fitriyani, N., 2021. Prediction of Daily Rainfall in Dodokan Watershed Based on Statistical Downscaling Model: An Effort to Manage Watershed Ecosystems. *Eastern Journal of Agricultural and Biological Sciences (EJABS)*, Vol. 1, No. 2, 1–7.

[6] Hadijati, M., Irwansyah and Fitriyani, N., 2022. Autoregressive Prewhitening on the Nonparametric Regression Model of Water Discharge in the Jangkok Watershed, Lombok Island. *Global Journal of Pure and Applied Mathematics*, Vol. 18, No. 1, 307–318.

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

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

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

[10] Hardle W., 1994. *Applied Nonparametric Regression.* Humboldt Universität zu Berlin, Berlin.

[11] Hollander M., Wolfe D. A., and Chicken E., 2014. *Nonparametric Statistical Methods*, Third Edition. John Wiley and Sons Inc., Canada.

[12] Law of The Republic of Indonesia Number 28 of 2009 Concerning of Regional Taxes and Regional Levies. https://jdih.kemenkeu.go.id/fulltext/2009/28tahun2009uu.html *[accessed on 9 Oktober 2022]*.

[13] Mahyuliza, Rahmadana M. F., and Nugrahadi E. W., 2021. Analysis of Factors Affecting Taxpayer Comliance Paying Hotel Tax, Restaurant Tax and Entertainment Tax in Medan City. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, Vol. 4, No. 1, 563-574.

[14] National Council for Special Economy of the Republic of Indonesia, KEK Mandalika, https://kek.go.id/kawasan/KEK-Mandalika *[accessed on 3 Oktober 2022]*.

[15] Pratiwi, D., Mursy, L.A.A., Rizaldi, M., and Fitriyani, F., 2020. Regresi Nonparametrik Kernel Gaussian pada Pemodelan Angka Kelahiran Kasar di Provinsi Nusa Tenggara Barat (translate: Nonparametric Gaussian Kernel Regression in Modeling Crude Birth Rate in West Nusa Tenggara Province). *Eigen Mathematics Journal*, Vol. 3, No. 2, 100-105.

[16] Regional Revenue Management Agency, Central Lombok, 2021. *Unpublished.*

[17] Scott D. W., 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley and Sons Inc., Canada.

[18] Silverman B. W., 1998. *Density Estimation for Statistics and Analysis.* Chapman and Hall, London.

[19] Sukarsa I. K. G., and Srinadi I. G. A. M., 2012. Estimator Kernel dalam Model Regresi Nonparametrik (translate: Kernel Estimators in Nonparametric Regression Models), *Jurnal Matematika*, Vol. 2, No. 1, 19-30.

[20] Zulkarnaen N, Sayuti M., dan Fajariah F., 2022. Konsep Pengembangan Kawasan Ekonomi Khusus (KEK) Mandalika Berorientasi Kesejahteraan Masyarakat Lokal (translate: The Mandalika Special Economic Zone (SEZ) Development Concept Oriented Welfare of Local Communities). *Ganec Swara*, Vol. 16, No. 1, 1362-1369.