Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:79
https://doi.org/10.28919/cmbn/8582
ISSN: 2052-2541

MODELLING THE NUMBER OF HIV CASES IN INDONESIA USING NEGATIVE BINOMIAL REGRESSION BASED ON LEAST SQUARE SPLINE ESTIMATOR

ARIP RAMADAN[1,2], NUR CHAMIDAH[3,4,*], I NYOMAN BUDIANTARA[5]

[1]Doctoral Study Program of Mathematics and Natural Sciences, Faculty of Science and Technology, Airlangga University, Surabaya 60115, Indonesia

[2]Information System Study Program, Department of Industrial and System Engineering, Telkom University Surabaya Campus, Indonesia

[3]Department of Mathematics, Faculty of Science and Technology, Airlangga University, Surabaya 60115, Indonesia

[4]Research Group of Statistical Modeling in Life Science, Faculty of Science and Technology, Airlangga University, Surabaya 60115, Indonesia

[5]Department of Statistics, Fac. Sci. and Data Analytics, Sepuluh Nopember Institute of Technology, Indonesia

Abstract: HIV is a quite dangerous infectious disease. The HIV transmission can be controlled through several factors including the number of prisoners, early-age marriages, and contraceptive users. In order to control these factors, proper judgment must be used. To analyze the association between HIV and these factors, an appropriate statistical model approach, such as, a least square spline nonparametric negative binomial regression model, is proposed in this research. The results of the analysis show that the proposed model approach is better than the classical model approach,

*Corresponding author
E-mail address: nur-c@fst.unair.ac.id
Received April 09, 2024

namely, parametric negative binomial regression model. This is supported by a deviance value of the proposed model of 15.6552 which is less than a deviance value of the parametric negative binomial regression model that is 32.2954. Also, the number of HIV cases will decrease when the number of prisoners is less than 768. In addition, the number of the HIV cases will increase only if the percentage of early-ager marriages is between 11.155% and 3.2475%, and apart from this range the number of HIV cases will decrease. Additionally, the number of HIV cases will increase if contraceptive users are between 42.085% and 48.610%, apart from this range the number of HIV cases will decrease.

Keywords: HIV; prisoners; early-age marriages; contraceptive users; least square spline.

2010 AMS Subject Classification: 62G08, 62P10, 65D07, 65D10.

1. INTRODUCTION

The HIV is a quite dangerous infectious disease [1]. Factors that spread HIV can come from humans themselves or environmental factors. Several important factors that have quite an influence on the spread of the HIV are the number of prisoners, the number of early-age marriages, and the number of contraceptive users [2]. In this research, we analyze the relationship between the number of HIV cases and several factors mentioned previously. Because these factors come from outside the individual, they can be controlled by a policy [3]. We use a nonparametric regression for modeling and analyzing data of HIV cases, because the relationship between HIV and these factors has no special pattern [4]. The estimation method used is a least square spline, because it has capability for modeling and analyzing the data that has change patterns in certain sub-intervals [5]. This is important, so that the estimation results become more precise.

Researches on the HIV has been carried out by several researchers previously. Tohari et al. [6] analyzed data of HIV and AIDS by using bi-response negative binomial regression based on local linear estimator. In their research, it was found that the nonparametric regression model is better than the parametric regression model. Next, Zhang and Yuan [7], examined HIV using semiparametric regression with kernel smoothing estimation. For data that has two responses or what is called bi-response, it has been applied to HIV and AIDS data. Tohari et al. [8] used the bi-response nonparametric negative binomial regression based on local linear estimator to model and analyze the HIV and AIDS cases. On the other hand, parametric regression models have also been carried out on HIV data by Hossain et al. [9], namely, they used a parametric survival model to model the lifetime of HIV patients. The use of the negative binomial regression has also been applied in other fields by Chamidah et al. [10]. They applied the negative binomial regression to data on the number of speed violations on toll roads.

The studies mentioned previously have slight weaknesses when applied to our data. Our data shows that the relationship between HIV, prisoners, early-age marriages, and contraceptive users does not have a specific pattern. The local linear estimator is not suitable for our data which has patterns in sub-intervals [11–13]. The appropriate estimator is the least square spline. To compare model goodness-of-fit, we use deviance values. This research aims to model and analyze the relationship between HIV and prisoners, early-age marriages, and contraceptive users. This research uses nonparametric regression approach because it is suitable for data that does not have a certain relationship pattern. The estimator used is a least square spline because our data has change patterns in certain sub-intervals. We use this method to minimize errors in estimation because this method is appropriate to the conditions of our research data.

2. MATERIALS AND METHODS

In this section, we provide materials and methods used for modeling and analyzing the HIV cases in Indonesia.

2.1. Dataset

In the following table, we provide the number of the HIV cases in 30 provinces of Indonesia, and the spreading factors of HIV.

Table 1. Dataset of the Number of HIV Cases in Indonesia in 2023.

Number	Province	The Number of HIV Cases (Y)	The Number of Prisoners (X_1)	Percentage of Early-Age Marriages (X_2)	Percentage of Contraceptive Users (X_3)
1	Aceh	175	5,325	4.60	40.54
2	Sumatera Utara	1,904	19,088	4.82	38.27
3	Sumatera Barat	364	3,607	3.48	41.98
⋮	⋮	⋮	⋮	⋮	⋮
30	Papua	1,790	869	13.21	19.33

Table 1 presents a dataset that consists of the number of HIV cases as a response variable, and the number of prisoners, the percentage of early-age marriages, and the percentage of contraceptive users as predictor variables that are thought to influence the number of HIV cases.

2.2. Overdispersion Test

According to Winter and Bürkner [14], theoretically the count data can be modeled with a Poisson distribution. However, the Poisson distribution has a special characteristic, namely, the mean value is as same as the variance value [15]. If the data obtained turns out to have a variance value that is greater than the mean, then overdispersion occurs [16]. So, the appropriate distribution for overdispersed count data is a negative binomial distribution [17]. To test whether the data has overdispersion or not, the score statistic test is used. The statistic test is given by the following equation [18]:

$$\begin{aligned} U_u &= \sum_{k=1}^{K} \frac{\partial}{\partial_{\tau_u}} l_k\left(y_k;\beta,\mathbf{0}\right) \\ &= \frac{1}{2}\sum_{k=1}^{K}\left(\left[\frac{\partial \ln f_k\left(y_k;\beta\right)}{\partial \beta_u}\right]^2 + \left[\frac{\partial^2 \ln f_k\left(y_k;\beta\right)}{\partial \beta_u^2}\right]\right), \qquad u = 1,\cdots,q \end{aligned} \tag{1}$$

2.3. Negative Binomial Regression

Suppose that we have a paired data $(\mathbf{x}_i, y_i)$, where $\mathbf{x}_i = \left(x_{1i}, x_{2i}, \cdots x_{pi}\right)$, $i = 1, 2, \cdots, n$ and n is the number of observations. The paired data $(\mathbf{x}_i, y_i)$ satisfies a negative binomial regression model if it meets the following functional relationship [6,8]:

$$f\left(y_i \mid \mathbf{x}_i\right) = \frac{\Gamma\left(y_i + \frac{1}{\alpha}\right)}{y_i!\,\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{1}{1+\alpha\boldsymbol{\mu}\left(\mathbf{x}_i\right)}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha\boldsymbol{\mu}\left(\mathbf{x}_i\right)}{1+\alpha\boldsymbol{\mu}\left(\mathbf{x}_i\right)}\right)^{y_i} \tag{2}$$

where $\alpha > 0$; $y_i = 0, 1, 2, \cdots, n$; $\boldsymbol{\mu}\left(\mathbf{x}_i\right)$ is defined as follows:

$$\boldsymbol{\mu}\left(\mathbf{x}_i\right) = E\left(y_i \mid \mathbf{x}_i\right) = \exp\left(\mathbf{s}\left(\mathbf{x}_i\right)\right) = \mathbf{y} \tag{3}$$

and $\mathbf{s}(\mathbf{x}_i)$ is the logarithm or log link which is usually used in the negative binomial model.

2.4. Least Square Spline Nonparametric Regression

By applying least square spline method, the negative binomial regression model can be used to predict the number of HIV cases based on the number of prisoners and the number of early-age marriages. Here, we consider an additive model that is a model of response variable y which depends on the sum of the several predictor variable functions of x. The additive model can be expressed as follows:

$$y_i = \sum_{j=1}^{p} s_j\left(x_{ji}\right) + \varepsilon_i, i = 1, 2, \cdots, n$$

where ε_i is a zero mean random error with variance σ^2, and s is an unknown regression function that is a smooth function in a Sobolev space.

Next, the least square spline as one of estimators to estimate the $s(x_i)$ function in the nonparametric regression model is a polynomial cut with different polynomial segments joined by vertices [19]. The segmented nature of the least square spline provides greater flexibility than ordinary polynomials. The least square spline regression model can adapt to the characteristics of the data. The least square spline function with one predictor variable of order q and vertex $\tau_1, \tau_2, \cdots, \tau_m$ can be expressed in the following form:

(4) $$s(x_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_q x_i^q + \sum_{k=1}^{m} \beta_{q+k} \left(x_i - \tau_k\right)_+^q$$

where

(5) $$\left(x_i - \tau_k\right)_+^q = \begin{cases} \left(x_i - \tau_k\right)^q, & x \geq \tau_k \\ 0 & , x < \tau_k \end{cases}$$

and β is a real constant parameter [19–21].

2.5. Maximum Likelihood Cross-Validation

To determine the best model, a maximum likelihood cross-validation (MLCV) is used. The selection of smoothing parameters in nonparametric regression analysis with the spline approach is conducted to obtain good modeling [22–34]. One method that is often used to determine the optimum smoothing parameter is using the maximum likelihood cross-validation (MLCV) method

[34]. The general form of the MLCV can be expressed as follows:

$$\text{MLCV} = \sum_{i=1}^{n} \ln f\left(y_i, \hat{s}_{-i}(x)\right) \tag{6}$$

where $\hat{s}_{-i}(x)$ is an estimate of the regression function at the point x_i without including the i-th data [8,35]. The best model is the model that has the smallest value of the MLCV [34].

3. RESULTS AND DISCUSSIONS

In the following, we present the scatterplots for the number of prisoners against with the number of HIV cases, for the percentage of early-age marriages against with the number of HIV cases, and for the percentage of contraceptive users against with the number of HIV cases. All these plots are given in the Figures 1–3.

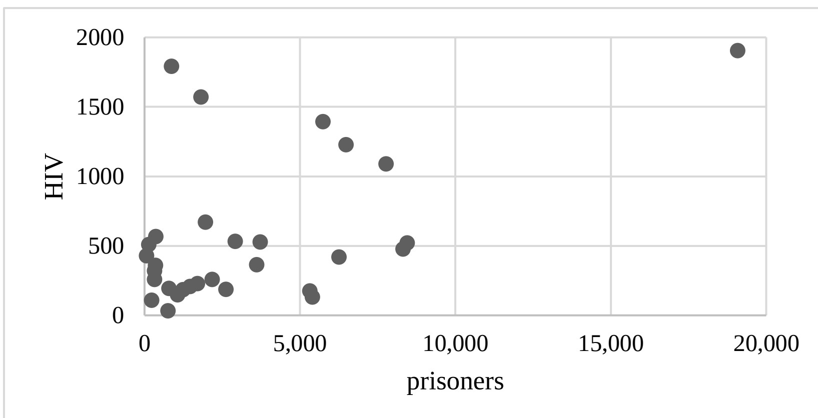

Figure 1. Scatterplot for the number of prisoners versus the number of HIV cases.

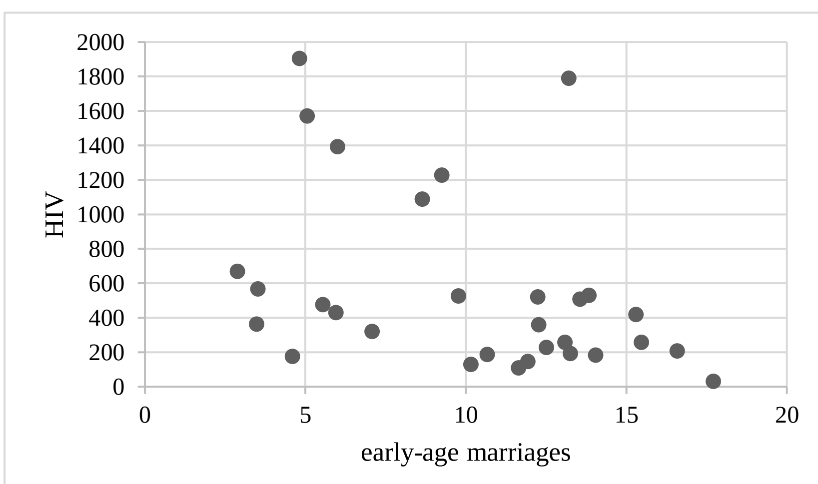

Figure 2. Scatterplot for the percentage of early-age marriages versus the number of HIV cases.

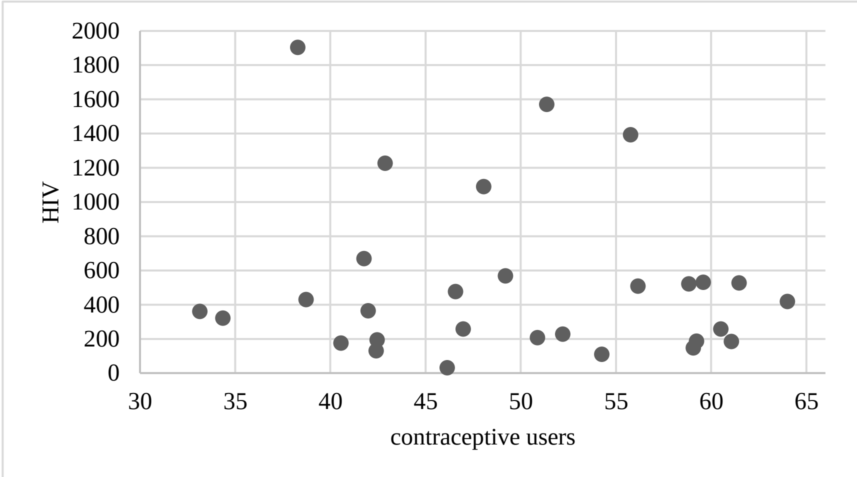

Figure 3. Scatterplot for the percentage of contraceptive users versus the number of HIV cases.

From Figures 1–3, it can be seen that there is no pattern of relationship between the number of prisoners, the percentage of early-age marriages, and the percentage of contraceptive users versus the number of HIV cases. In other words, based on these scatterplots there is no indication of the certain form of these relationship presented by them. Therefore, a possible alternative model approach to solve this case of the HIV data is a nonparametric regression model approach.

Next, the results of the dispersion test on the HIV data are presented in Table 2 that consists of a dispersion ratio value, a Pearson's Chi-Squared value, a P-value, and detection result of the presence of overdispersion.

Table 2. Dispersion test result

Dispersion Ratio	400.228
Pearson's Chi-Squared	10,806.145
P-value	<0.001
Overdispersion detected?	Yes

From the Table 2, it can be seen that the results of the overdispersion test are the dispersion ratio value of 400.228, the Pearson's Chi-Squared value of 10,806.145, and the P-value that is less than 0.001 where it is less than a significance level (α) of 0.05. This means that the overdispersion was detected in the response variable, namely, HIV cases. So, the suitable method approach to analyze the HIV cases data is nonparametric negative binomial regression based on least square spline, because it can accommodate response variable that experiences the overdispersion. In addition, if the response variable experiences overdispersion and continues using a Poisson analysis, it will result an overestimate where the predictor variables which should not have a significant effect on the response variable could actually be the opposite. Next, in Table 3 we present the calculation results for each predictor variable using R-code including the number of knot point, the optimal knot points, and the MLCV value.

Table 3. Knot point for each predictor variable

Predictor	**Number of Knot Point**	**Knot Point**	**MLCV**
x_1	3	768; 1887.5; 5382.5	-214.2355
x_2	3	5.9625; 11.155; 3.2475	
x_3	3	42.085; 48.610; 58.145	

Based on Table 3, the optimal number of knot point for each predictor variable is 3 knots. This is based on the optimal MLCV value of -214.2355.

Hereinafter, based on the results presented in the Table 3, we obtained the parameter estimation results of the parametric regression model approach, and the least square spline nonparametric regression model approach. These results are presented in Table 4.

Table 4. Parameter estimation results for parametric and nonparametric regression model

Parameters Estimate	**Parametric Regression**	**Parameter Estimate**	**Nonparametric Regression**
$\hat{\beta}_0$	7.6003	$\hat{\beta}_0$	103.8540
$\hat{\beta}_1$	7.28×10^{-5}	$\hat{\beta}_1$	– 0.0010
$\hat{\beta}_2$	– 0.0626	$\hat{\beta}_{1,1}$	0.0014
$\hat{\beta}_3$	– 0.0209	$\hat{\beta}_{1,2}$	– 0.0002
$\hat{\alpha}$	2.1142	$\hat{\beta}_{1,3}$	– 0.0002
		$\hat{\beta}_2$	– 0.0927
		$\hat{\beta}_{2,1}$	– 0.1036
		$\hat{\beta}_{2,2}$	0.5928
		$\hat{\beta}_{2,3}$	– 0.7325
		$\hat{\beta}_3$	– 0.0891
		$\hat{\beta}_{3,1}$	0.2020
		$\hat{\beta}_{3,2}$	– 0.1394
		$\hat{\beta}_{3,3}$	0.0264
		$\hat{\alpha}$	0.2821

In Table 4 are all the parameters that have been estimated for the parametric negative binomial regression, and least square spline nonparametric negative binomial regression.

Furthermore, based on the results presented in the Table 3 and Table 4, we obtained the parameter estimation results of the parametric regression model approach, and the least square spline nonparametric regression model approach compared with the actual value, for the HIV cases in provinces of Indonesia. These results are presented in Table 5.

Table 5. Estimation results of parametric and nonparametric regression models for the HIV cases in provinces of Indonesia

Province	Actual Data	Parametric Estimation	Nonparametric Estimation
Aceh	175	945.0000	846.3455
Sumatera Utara	1,904	2,661.0080	2281.1840
Sumatera Barat	364	867.5774	567.6777
Riau	476	975.5444	1,346.4160
Jambi	187	358.9324	232.5722
Sumatera Selatan	521	501.3992	709.2849
Bengkulu	146	297.0766	176.0366
Lampung	526	392.5854	352.6551
Kepulauan Bangka Belitung	184	252.7211	243.3178
Kepulauan Riau	669	802.3044	427.1283
DI Yogyakarta	567	587.7138	774.1340
Banten	1,392	648.4070	1153.9420
Bali	1,571	566.7699	630.2961
Nusa Tenggara Barat	207	271.3914	137.7277
Nusa Tenggara Timur	429	615.2234	552.7388
Kalimantan Barat	531	298.5275	423.7473
Kalimantan Tengah	257	250.4896	208.4816
Kalimantan Selatan	419	316.5397	467.0339
Kalimantan Timur	1,089	749.0155	876.2324
Kalimantan Utara	130	645.1357	299.6686
Sulawesi Utara	508	266.6347	481.5056
Sulawesi Tengah	227	346.5531	334.8827
Sulawesi Selatan	1,227	731.9944	402.4650
Sulawesi Tenggara	193	379.4769	169.6305
Gorontalo	109	314.8329	271.9056
Sulawesi Barat	31	265.3735	57.5937
Maluku	320	639.8888	499.2163
Maluku Utara	257	337.2308	421.3916
Papua Barat	360	475.4867	374.8846
Papua	1,790	621.3128	1,263.4800

Table 5 shows the comparison between the actual HIV cases data and HIV cases estimates using the parametric negative binomial regression approach, and the HIV cases estimates using the least square spline nonparametric negative binomial regression approach.

Next, in the Table 6 we present the deviance values for the parametric negative binomial regression model approach and for the least square spline nonparametric negative binomial regression model approach.

Table 6. Deviance Values for parametric regression and nonparametric regression

Deviance Value	
Parametric Regression	**Nonparametric Regression**
32.2954	15.6552

Table 6 shows the comparison of deviance values between parametric regression model approach and nonparametric regression model approach where the deviance value of the nonparametric regression model approach is smaller than that of the parametric regression model approach. This means that in this case, the nonparametric regression model approach is better than the parametric regression model approach for analyze the HIV cases data.

In addition, in the following we provide plots of actual data, parametric regression estimate, and nonparametric regression estimate. These plots of the actual data, parametric regression estimate, and nonparametric regression estimate are given in Figure 4.

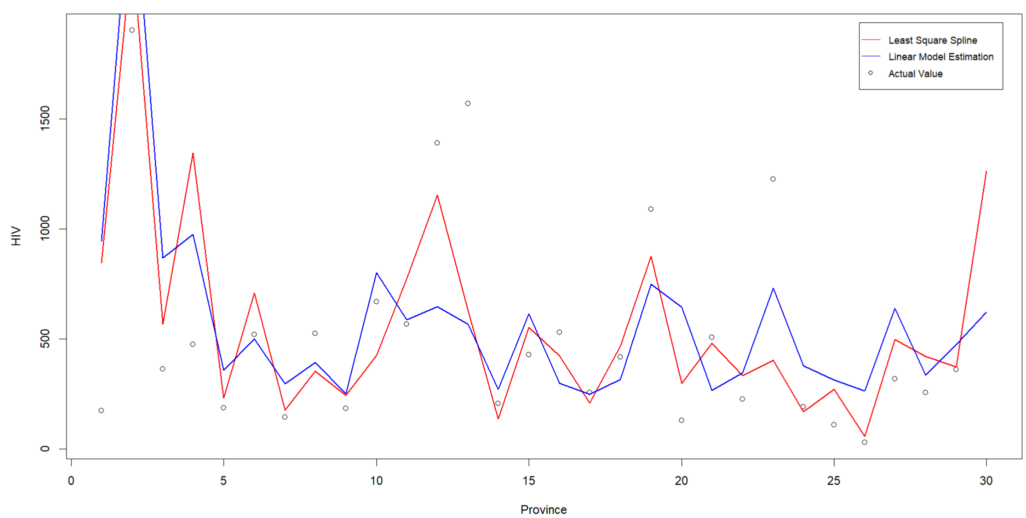

Figure 4. Plots of actual data, parametric regression estimate, and nonparametric regression estimate

The Figure 4 is a plot between actual data, parametric estimation results and nonparametric estimation results. It can be seen that the nonparametric estimation results are closer to the actual data. This is in accordance with the comparison of the coefficient of determination and deviance values from the two regression models. Next, by applying Equation (4), the nonparametric model can be explained as follows:

(7)
$$\hat{y} = \exp\{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_{1,1}(x_1 - \tau_{1,1})_+ + \hat{\beta}_{1,2}(x_1 - \tau_{1,2})_+ + \hat{\beta}_{1,3}(x_1 - \tau_{1,3})_+ + \hat{\beta}_2 x_2 + \hat{\beta}_{2,1}(x_2 - \tau_{2,1})_+ + \hat{\beta}_{2,2}(x_2 - \tau_{2,2})_+ + \hat{\beta}_{2,3}(x_2 - \tau_{2,3})_+ + \hat{\beta}_3 x_3 + \hat{\beta}_{3,1}(x_3 - \tau_{3,1})_+ + \hat{\beta}_{3,2}(x_3 - \tau_{3,2})_+ + \hat{\beta}_{3,3}(x_3 - \tau_{3,3})_+\}$$
$$= \exp\{103.8540 - 0.0010x_1 + 0.0014(x_1 - 768)_+ - 0.0002(x_1 - 1887.5)_+ + -0.0002(x_1 - 5382.5)_+ - 0.0927x_2 - 0.1036(x_2 - 5.9625)_+ + 0.5928(x_2 - 11.155)_+ - 0.7325(x_2 - 13.2475)_+ - 0.0891x_3 + 0.2020(x_3 - 42.085)_+ - 0.1394(x_3 - 48.610)_+ + 0.0264(x_3 - 58.145)_+\}.$$

Based on Equation (7) and by applying Equation (5), we can obtain the relationship between HIV, the number of prisoners, the percentage of early-age marriages, and the percentage of contraceptive users that are respectively presented in Equations (8)–(10).

$$\hat{s}(x_1) = 103.8540 - 0.0010x_1 + 0.0014(x_1 - 768)_+ - 0.0002(x_1 - 1887.5)_+ - 0.0002(x_1 - 5382.5)_+$$

(8)
$$\hat{s}(x_1) = \begin{cases} 103.8540 - 0.0010x_1 & \text{for} \quad x_1 < 768 \\ 93.102 + 0.013x_1 & \text{for} \quad 768 \le x_1 < 1887.5 \\ 93.4795 + 0.0128x_1 & \text{for} \quad 1887.5 \le x_1 < 5382.5 \\ 94.556 + 0.0126x_1 & \text{for} \quad x_1 \ge 5382.5 \end{cases}.$$

$$\hat{s}(x_2) = 103.8540 - 0.0927x_2 - 0.1036(x_2 - 5.9625)_+ + 0.5928(x_2 - 11.155)_+ - 0.7325(x_2 - 13.2475)_+$$

(9)
$$\hat{s}(x_2) = \begin{cases} 103.8540 - 0.0927x_2 & \text{for} \quad x_2 < 5.9625 \\ 103.236 - 0.1963x_2 & \text{for} \quad 5.9625 \le x_2 < 11.155 \\ 96.6236 + 0.3965x_2 & \text{for} \quad 11.155 \le x_2 < 13.2475 \\ 106.327 - 0.336x_2 & \text{for} \quad x_2 \ge 13.2475 \end{cases}.$$

$$\hat{s}(x_3) = 103.8540 - 0.0891x_3 + 0.2020(x_3 - 42.085)_+ - 0.1394(x_3 - 48.610)_+ + 0.0264(x_3 - 58.145)_+$$

(10)
$$\hat{s}(x_3) = \begin{cases} 103.8540 - 0.0891x_3 & \text{for} \quad x_3 < 42.085 \\ 95.3528 + 0.1129x_3 & \text{for} \quad 42.085 \le x_3 < 48.610 \\ 102.129 - 0.0265x_3 & \text{for} \quad 48.610 \le x_3 < 58.145 \\ 100.594 - 0.0001x_3 & \text{for} \quad x_3 \ge 58.145 \end{cases}.$$

Based on Equation (8), it shows that: (a). for the number of prisoners is less than 768, if the number of prisoners increases by one person, then the number of HIV cases will decrease by 0.0010; (b). If the number of prisoners is between 768 and 1887.5, then an increase of one prisoner will increase the number of HIV cases by 0.013; (c). If the number of prisoners is between 1887.5 and 5382.5, then an increase of one prisoner will increase the number of HIV cases by 0.0128; (d). If the number of prisoners is more than 5382.5 then adding one prisoner will increase the number of HIV cases by 0.0126.

Next, Equation (9) presents a relationship between the percentage of early-age marriages and the number of HIV cases. It shows that: (a). If early-age marriages are less than 5.9625%, then adding one case of early-age marriages will reduce the number of HIV cases by 0.0927; (b). If early-age marriages are between 5.9625% and 11.155%, then the addition of one case of early-age marriages will reduce the number of HIV cases by 0.1963; (c). If early-age marriages are between 11.155% and 13.2475%, then the addition of one case of early-age marriages will increase the number of HIV cases by 0.3965; (d). If early-age marriages are more than 13.2475%, then the number of HIV cases will decrease by 0.336.

Similarly to the interpretation of Equation (8) and Equation (9), the Equation (10) can also be interpreted as follows: (a). If there are less than 42.085% contraceptive users, then the addition of one contraceptive user will reduce the number of HIV cases by 0.0891; (b). If contraceptive users are between 42.085% and 48.610%, then the addition of one contraceptive user will increase the number of HIV cases by 0.1129; (c). If contraceptive users are between 48.610% and 58.145%, then the addition of one contraceptive user will reduce the number of HIV cases by 0.0265; (d). If contraceptive users are more than 58.145%, then the addition of one contraceptive user will reduce the number of HIV cases by 0.0001.

4. CONCLUSIONS

In this research, two conclusions were obtained. The first conclusion is that the best model for analyzing HIV case data with the number of prisoners, percentage of early-age marriages, and percentage of contraceptive users is a least square spline nonparametric negative binomial regression model. In this case, the proposed model is better than a parametric negative binomial regression model. This can be observed based on the deviance values of these two models, namely, the least square spline nonparametric negative binomial regression model has a smaller deviance value than the parametric negative binomial regression model.

The second conclusion is relationship between the number of HIV cases and the number of prisoners is directly proportional when the number of prisoners is from 768 to 5382.5. Then, the relationship is inversely proportional when the number of prisoners is less than 768. The relationship between the number of HIV cases and the percentage of early-age marriages is directly proportional when early-age marriages are between 11.155% and 13.2475%. Then, the relationship is inversely proportional when early-age marriages are less than 5.9625%, between 1.9625% and 11.155%, and more than 13.2475%. The relationship between the number of HIV and the percentage of contraceptive users is directly proportional when contraceptive users are between 42.085% and 48.610%. Also, it will be inversely proportional to other than that range.

ACKNOWLEDGMENT

The authors thank Dr. Budi Lestari, Drs., PG.Dip.Sc., M.Si. for providing pre-review, constructive comments and suggestions for improving the quality of this paper.

CONFLICT OF INTERESTS

The authors confirm that there is no conflict of interests.

REFERENCES

[1] S. Shao, Y. Shao, From a deadly disease to a manageable chronic disease, HIV/AIDS remains a challenge for mankind, Infect. Microb. Dis. 5 (2023), 41–43. https://doi.org/10.1097/im9.0000000000000121.

[2] T. Farley, C. Mwachari, C. Cohen, et al. Effect of contraceptive methods on natural history of HIV: Studies from

the mombasa cohort, J. Acquir. Immune Defic. Syndr. 38 (2005), S20–S21. https://doi.org/10.1097/01.qai.0000167031.25901.ec.

[3] G. Bruce, Link, J. Phelan, Social conditions as fundamental causes of disease, J. Health Soc. Behav. 35 (1995) 80–94. https://doi.org/10.2307/2626958.

[4] P. Čížek, S. Sadıkoğlu, Robust nonparametric regression: A review, WIREs Comput. Stat. 12 (2019), e1492. https://doi.org/10.1002/wics.1492.

[5] N. Chamidah, B. Lestari, A. Massaid, et al. Estimating mean arterial pressure affected by stress scores using spline nonparametric regression model approach, Commun. Math. Biol. Neurosci. 2020 (2020), 72. https://doi.org/10.28919/cmbn/4963.

[6] A. Tohari, N. Chamidah, F. Fatmawati, Modelling of HIV and AIDS cases in indonesia using bi-response negative binomial regression approach based on local linear estimator, Ann. Biol. 36 (2020), 215–219.

[7] J. Zhang, A. Yuan, Semiparametric current status regression model using kernel smoothing with application to HIV study, Biostat. Epidemiol. 7 (2022), e2140226. https://doi.org/10.1080/24709360.2022.2140226.

[8] A. Tohari, N. Chamidah, Fatmawati, et al. Modelling the number of HIV and AIDS cases in East Java using biresponse multipredictor negative binomial regression based on local linear estimator, Commun. Math. Biol. Neurosci. 2021 (2021), 73. https://doi.org/10.28919/cmbn/5652.

[9] A. Hossain, F. Islam, H. Chakraborty, Parametric regression model based on reversed hazard rate: An application to left censored heavy tailed HIV viral load data, Bull. Malays. Math. Sci. Soc. 45 (2022), 567–598. https://doi.org/10.1007/s40840-022-01360-7.

[10] N. Chamidah, A. Widyanti, F. Trapsilawati, et al. Local linear negative binomial nonparametric regression for predicting the number of speed violations on toll road: A theoretical discussion, Commun. Math. Biol. Neurosci. 2021 (2021), 10. https://doi.org/10.28919/cmbn/5282.

[11] E. Ana, N. Chamidah, P. Andriani, et al. Modeling of hypertension risk factors using local linear of additive nonparametric logistic regression, J. Phys.: Conf. Ser. 1397 (2019), 012067. https://doi.org/10.1088/1742-6596/1397/1/012067.

[12] N. Chamidah, B. Zaman, L. Muniroh, et al. Designing local standard growth charts of children in East Java Province using a local linear estimator, Int. J. Innov., Creat. Change 13 (2020), 45-67.

[13] N. Chamidah, Y.S. Yonani, E. Ana, et al. Identification the number of Mycobacterium tuberculosis based on

sputum image using local linear estimator, Bull. Electr. Eng. Inform. 9 (2020), 2109–2116. https://doi.org/10.11591/eei.v9i5.2021.

[14] B. Winter, P. Bürkner, Poisson regression for linguists: A tutorial introduction to modelling count data with brms, Lang. Linguist. Compass. 15 (2021), e12439. https://doi.org/10.1111/lnc3.12439.

[15] S. Saidi, N. Herawati, K. Nisa, Modeling with Generalized Linear Model on COVID-19: Cases in Indonesia, Int. J. Electron. Commun. Syst. 1 (2021), 25–33.

[16] D. Cahoy, E. Di Nardo, F. Polito, Flexible models for overdispersed and underdispersed count data, Stat. Papers. 62 (2021), 2969–2990. https://doi.org/10.1007/s00362-021-01222-7.

[17] J. Stoklosa, R.V. Blakey, F.K.C. Hui, An overview of modern applications of negative binomial modelling in ecology and biodiversity, Diversity. 14 (2022), 320. https://doi.org/10.3390/d14050320.

[18] R.J. Ohara Hines, A comparison of tests for overdispersion in generalized linear models, J. Stat. Comput. Simul. 58 (1997), 323–342. https://doi.org/10.1080/00949659708811838.

[19] N. Chamidah, B. Lestari, A.Y. Wulandari, et al. Z-score standard growth chart design of toddler weight using least square spline semiparametric regression, AIP Conf. Proc. 2329 (2021), 060031. https://doi.org/10.1063/5.0042285.

[20] N. Chamidah, B. Lestari, T. Saifudin, Modeling of blood pressures based on stress score using least square spline estimator in bi-response non-parametric regression, Int. J. Innov., Creat. Change 5 (2019), 1200–1216.

[21] Fatmawati, I. N. Budiantara, B. Lestari, Comparison of smoothing and truncated splines estimators in estimating blood pressure models, Int. J. Innov., Creat. Change 5 (2019), 1177–1199.

[22] N. Chamidah, B. Lestari, A. Massaid, et al. Estimating mean arterial pressure affected by stress scores using spline nonparametric regression model approach, Commun. Math. Biol. Neurosci. 2020 (2020), 72. https://doi.org/10.28919/cmbn/4963.

[23] W. Ramadan, N. Chamidah, B. Zaman, et al. Standard growth chart of weight for height to determine wasting nutritional status in East Java based on semiparametric least square spline estimator, IOP Conf. Ser.: Mater. Sci. Eng. 546 (2019), 052063. https://doi.org/10.1088/1757-899x/546/5/052063.

[24] B. Lestari, Fatmawati, I.N. Budiantara, N. Chamidah, Smoothing parameter selection method for multiresponse nonparametric regression model using smoothing spline and Kernel estimators approaches, J. Phys.: Conf. Ser. 1397 (2019), 012064. https://doi.org/10.1088/1742-6596/1397/1/012064.

[25] B. Lestari, Fatmawati, I.N. Budiantara, Spline estimator and its asymptotic properties in multiresponse nonparametric regression model, Songklanakarin J. Sci. Technol. 42 (2020), 533–548.

[26] N. Chamidah, B. Lestari, I.N. Budiantara, et al. Consistency and asymptotic normality of estimator for parameters in multiresponse multipredictor semiparametric regression model, Symmetry 14 (2022), 336. https://doi.org/10.3390/sym14020336.

[27] B. Lestari, N. Chamidah, D. Aydin, et al. Reproducing kernel hilbert space approach to multiresponse smoothing spline regression function, Symmetry 14 (2022), 2227. https://doi.org/10.3390/sym14112227.

[28] N. Chamidah, B. Zaman, L. Muniroh, et al. Multiresponse semiparametric regression model approach to standard growth charts design for assessing nutritional status of East Java toddlers, Commun. Math. Biol. Neurosci. 2023 (2023), 30. https://doi.org/10.28919/cmbn/7814.

[29] B. Lestari, N. Chamidah, I.N. Budiantara, et al. Determining confidence interval and asymptotic distribution for parameters of multiresponse semiparametric regression model using smoothing spline estimator, J. King Saud Univ. - Sci. 35 (2023), 102664. https://doi.org/10.1016/j.jksus.2023.102664.

[30] D. Aydın, E. Yılmaz, N. Chamidah, et al. Right-censored partially linear regression model with error in variables: application with carotid endarterectomy dataset, Int. J. Biostat. 20 (2023), 245–278. https://doi.org/10.1515/ijb-2022-0044.

[31] D. Aydın, E. Yılmaz, N. Chamidah, et al. Right-censored nonparametric regression with measurement error, Metrika (2024). https://doi.org/10.1007/s00184-024-00953-5.

[32] N. Chamidah, B. Lestari, I.N. Budiantara, et al. Estimation of multiresponse multipredictor nonparametric regression model using mixed estimator, Symmetry 16 (2024), 386. https://doi.org/10.3390/sym16040386.

[33] D. Aydin, M. Memmedli, R.E. Omay, smoothing parameter selection for nonparametric regression using smoothing spline, Eur. J. Pure Appl. Math. 6 (2013), 65–60.

[34] D.J. Cummins, T.G. Filloon, D. Nychka, Confidence intervals for nonparametric curve estimates, J. Amer. Stat. Assoc. 96 (2001), 233–246. https://doi.org/10.1198/016214501750332811.

[35] P. Chaudhuri, A. Dewanji, On a likelihood-based approach in nonparametric smoothing and cross-validation, Stat. Probab. Lett. 22 (1995), 7–15. https://doi.org/10.1016/0167-7152(94)00040-f.