Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2022, 2022:82
https://doi.org/10.28919/cmbn/7518
ISSN: 2052-2541

A THEORETICAL DISCUSSION ON MODELING THE NUMBER OF COVID-19 DEATH CASES USING PENALIZED SPLINE NEGATIVE BINOMIAL REGRESSION APPROACH

NUR CHAMIDAH[1,2,*], MARISA RIFADA[1,2], DITA AMELIA[1,2]

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

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

Copyright © 2022 the author(s). This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract: The Covid-19 pandemic that has occurred since the end of 2019 has changed almost the entire order of the world community, including Indonesia, in terms of health, economic, social and cultural arrangements. Based on the initial study, it is known that the number of Covid-19 deaths in East Java in 2020 has a high variance in each district/city which will cause an over dispersion problem, to overcome this, regression can be used assuming the response variable has a negative binomial distribution. Therefore, in this study we determine theoretically a model estimate of the number of cases of Covid-19 deaths in East Java due to comorbidities using a nonparametric negative binomial regression (NNBR) model approach based on a penalized spline estimator which is applied to generalized additive model (GAM). In this study, we provided steps for a local scoring algorithm to estimate NNBR model based on penalized spline estimator. In the future, the theoretical results of this study can be applied to the real data namely

[*]Corresponding author
E-mail address: nur-c@fst.unair.ac.id
Received May 21, 2022

the number of Covid-19 death cases affected by comorbidities such as percentage of diabetes mellitus patients, percentage of hypertension over 15 years old patients, and percentage of tuberculosis patients.

Keywords: comorbidities; Covid-19; nonparametric negative binomial regression; penalized spline estimator.

2010 AMS Subject Classification: 62F10, 62G05, 62G08.

1. INTRODUCTION

The Covid-19 Handling Task Force stated that the percentage of deaths in Indonesia of 3.24 percent was still relatively high when compared to the global percentage of 2.08 percent [1]. East Java Province is one of the most densely populated provinces in Indonesia which was once the province with the highest mortality rate of 7.2% because it was above the global death rate of 2.1% [2]. Therefore, it is hoped that the Indonesian government will prioritize suppression of Covid-19 death cases, especially in East Java Province. According to the Deputy Chairperson of the IDI Executive Board, Adib Khumaidi, the high number of Covid-19 deaths was caused by co-morbidities [3]. Based on the type of comorbidities, hypertension and diabetes mellitus are among the highest ranks in Covid-19 cases in Indonesia with percentages of 50.1% and 34.8%, respectively [4]. The results of the study [5] showed that co-morbidities such as Tuberculosis can cause the risk of death in Covid-19 patients so that it has an impact on increasing cases of Covid-19 deaths. Therefore, to determine how much influence comorbidities have on the number of Covid-19 deaths in East Java, it is necessary to model using a regression model approach.

There are two approaches in regression modeling, namely the parametric regression approach (global) and the nonparametric regression approach (local). Estimators in nonparametric regression that are often used include local spline and linear. Cases of Covid-19 deaths in 38 districts/cities in East Java vary widely, so it is more appropriate to use a nonparametric regression approach based on a spline estimator that accommodates the locality of the regression function so that it is more flexible, while a local linear estimator can accommodate model variations in each district/city. However, if a parametric approach is used, only one model is generated (global). Based on the initial study, it is known that the number of Covid-19 deaths in East Java in 2020 has a high variance in each district/city which will cause an over dispersion problem, to overcome this, regression can be used assuming the response variable has a negative binomial distribution. The researchers who have used the negative binomial distribution in their researches, for examples, [6].

Based on these facts and conditions, in this study we determine a model estimate of the number of cases of Covid-19 deaths in East Java due to comorbidities using a negative binomial nonparametric regression model approach based on a penalized spline estimator.

2. PRELIMINARIES

In this section, we provide brief overview of Covid-19 and comorbidities, negative binomial regression, and penalized spline estimator.

2.1. Covid-19 and Comorbidities

Corona viruses (Covid-19) are a large family of viruses that can cause disease in animals or humans. In humans, several coronaviruses are known to cause respiratory infections ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). In addition, at the end of 2019 a new Corona Virus case was found which was later identified as coronavirus disease (COVID-19) [6,7]. In the health sector, the Covid-19 pandemic is the cause of the health crisis with the high number of deaths caused by the Covid-19 virus. The Covid-19 Handling Task Force stated that the percentage of deaths in Indonesia of 3.24 percent was still relatively high when compared to the global percentage of 2.08 percent [1]. As of December 31, 2020, there were 743,198 confirmed COVID-19 patients, of which 22,138 were declared dead [8]. From these data, East Java is the province with the most deaths in Indonesia, with 5,827 people. The corona virus or Covid-19 continues to spread and move up significantly, especially in East Java which shows exponential movements from day to day [9].

In the study of Hogan et al [5] in their article, it was stated that co-morbidities such as tuberculosis can cause the risk of death in Covid-19 patients so that this has an impact on increasing cases of death from this virus. Research results [4] showed that hypertension and diabetes mellitus are included in the highest ranking in the number of Covid-19 cases based on the type of comorbidities in Indonesia with a percentage of 50.1% and 34.8%, respectively. Therefore, in this study the indicators used were the percentage of people with diabetes mellitus, the percentage of patients with hypertension and patients with pulmonary disease (Tuberculosis) in each district/city in East Java.

2.2. Negative Binomial Regression

A negative binomial regression is one approach to overcome the response variable in the form of count data whose variance value is greater than the mean or also called over-dispersion. Y response variables in the form of count data are usually analyzed using Poisson regression, but in practice cases of over-dispersion are often found [10]. Negative binomial regression is much more flexible when compared to Poisson regression, because negative binomial regression does not have certain assumptions [6,10].

According to [11], the regression function of a negative binomial regression model is as follows:

(1) $$\boldsymbol{f(y_i|x_i)} = \left(\frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)}{y_i\Gamma\left(\frac{1}{\alpha}\right)}\right)\left(\frac{1}{1+\alpha\mu(x_i)}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu(x_i)}{1+\alpha\mu(x_i)}\right)^{y_i}$$

where $y_i = 0, 1, 2, \ldots, \alpha > 0$ and $m(x_i)$ is defined as follows:

(2) $$\mu(\boldsymbol{x_i}) = \boldsymbol{E}(\boldsymbol{y_i}|\boldsymbol{x_i}) = \mathbf{exp}\,(\boldsymbol{m}(\boldsymbol{x_i}))$$

where y_i is count data of response variable which cannot be negative. Hence, the expected value of y_i is also non-negative. However, this is not appropriate because the m_i value space is in the interval $(-\infty, \infty)$.

This problem can be solved by applying the link function between the $m(x_i)$ and $\mu(x_i)$ namely by using logarithm or log link which is usually used in the negative binomial regression [11] such that we get:

(3) $$\mathbf{ln}\left(\mu(\boldsymbol{x_i})\right) = \boldsymbol{m}(\boldsymbol{x_i})$$

Equation (3) becomes more precise because it is in accordance with the definition in the interval

$(0, \infty)$ so that interpretation becomes easier. After obtaining the connecting function, we have the following equation:

(4) $$\mu(\boldsymbol{x_i}) = \mathbf{exp}(\boldsymbol{m}(\boldsymbol{x_i}))$$

where $m(x_i)$ is a regression function in parametric regression which is defined as $x_i\beta$.

2.3. Additive Model

Additive model is a model of response variable y which depends on the sum of several predictor variable functions of x. According to [11], an additive model can be expressed as follows:

(5) $$\boldsymbol{y_i = \sum_{j=1}^{p} m_j(x_{ji}) + \varepsilon_i\,, i = 1, 2, \dots, n}$$

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

2.4. Penalized Spline Estimator

The form of the model used in the penalized spline estimator is an additive model as follows:

(6) $$\boldsymbol{y_i = \sum_{j=1}^{p} m_j(x_{ji}) + \varepsilon_i\,,\ i = 1, 2, \dots, n}$$

where y_i is the *i*-th observation value of response variable; m_j is the *j*-th function of predictor variable; x_{ji} is the *i*-th observation value of the *j*-th predictor; and ε_i is the *i*-th value of random error that has zero mean and variance σ^2.

Next, the spline functions with order d and knot points $\xi_1, \xi_2, \dots, \xi_k$ are as follows:

(7) $$\boldsymbol{m_j(x_{ji}) = \sum_{h=0}^{d_j+k_j} \beta_{jh}\phi_h(x_{ji})\,;\ i = 1, 2, \dots, n\,;\ j = 1, 2, \dots, p}$$

where k represents the number of knot points; $\beta_j = \left(\beta_{j0}, \beta_{j1}, \dots, \beta_{pj+k_j}\right)^T$ is a vector of coefficients; and $\phi_{j0}, \phi_{j1}, \dots, \phi_{p_j+k_j}$ are short interval function bases namely $\phi_h(x_{ji})$ represents a function defined as follows:

(8) $$\boldsymbol{\phi_h(x_{ji}) = \begin{cases} x_{ji}^{h} & \textbf{for } 0 \le h \le d_j \\ \left(x_{ji} - \xi_{i(h-dj)}\right)_+^{d_j} & \textbf{for } d_j + 1 \le h \le d_j + k_j \end{cases}}$$

where d_j is polynomial order of j-th predictor; k_j is the number of knots for j-th predictor; and h is a index of basis function which is a positive integer such that:

$$(x_{ji} - \xi_{j(h-d_j)})_+^{d_j} = \begin{cases} (x_{ji} - \xi_{j(h-d_j)})_+^{d_j} & , x \geq \xi_{j(h-d_j)} \\ 0 & , x < \xi_{j(h-d_j)} \end{cases} \quad (9)$$

3. MAIN RESULTS

In this section, we present the results of this study including characteristics of model variables, estimating the nonparametric negative binomial regression based on penalized spline estimator, and selecting optimal smoothing spline.

3.1. Characteristics of Model Variables

In general, the penalized spline nonparametric negative binomial regression model can be applied to more than one predictor variables or multipredictor. For example we can apply the model to the number of Covid-19 death cases affected by comorbidities such as percentage of diabetes mellitus patients, percentage of hypertension over 15 years old patients, and percentage of tuberculosis patients. For this example case, we have a multipredictor (i.e., three predictors) nonparametric negative binomial regression (NNBR) model where percentage of diabetes mellitus patients (x_1), percentage of hypertension over 15 years old patients (x_2), and percentage of tuberculosis patients (x_3) as predictor variables, and the number of Covid-19 death cases (y) as response variable.

In this study, for predicting the number of covid-19 death cases we use a nonparametric negative binomial regression (NNBR) model approach. Generally, the main problem if we use regression models is estimating the regression functions of the regression models [12–31]. For this purpose, in the following section we discuss the estimation of the nonparametric negative binomial regression (NNBR) model that is started by estimating the regression function using penalized spline estimator.

3.2. Estimating the Nonparametric Negative Binomial Regression Model

Suppose we have a paired observations $\{y_i, \boldsymbol{x}_i\}$, $i = 1,2,\ldots,n$ which follows the following nonparametric negative binomial regression (NNBR) model:

$$\mu(\boldsymbol{x_i}) = \boldsymbol{E}(\boldsymbol{y_i}|\boldsymbol{x_i}) = \mathbf{exp}\left(\sum_{j=1}^{p} m_j\left(x_{ji}\right)\right) \quad , \boldsymbol{i = 1, 2, \ldots, n} \tag{10}$$

where y_i is the *i*-th observation value of response variable; f_j is the *j*-th function of predictor variable; x_{ji} is the *i*-th observation value of the *j*-th predictor; and ε_i is the *i*-th value of random error that has zero mean and variance σ^2.

Furthermore, by considering the NNBR model in (10), we can express the spline functions with order d and knot points $\xi_1, \xi_2, \ldots, \xi_k$ of NNBR model as follows:

$$\boldsymbol{m_j(x_{ji}) = \sum_{h=0}^{d_j+k_j} \beta_{jh}\phi_h(x_{ji})\,, i = 1, 2, \ldots, n \quad , \ j = 1, 2, \ldots, p} \tag{11}$$

where k represents the number of knot points; $\beta_j = \left(\beta_{j0}, \beta_{j1}, \ldots, \beta_{pj+k_j}\right)^T$ is a vector of coefficients of the NNBR model; and $\phi_{j0}, \phi_{j1}, \ldots, \phi_{p_j+k_j}$ are short interval function bases of the NNBR model namely $\phi_h(x_{ji})$ represents a function which is defined as follows:

$$\boldsymbol{\phi_h(x_{ji})} = \begin{cases} \boldsymbol{x_{ji}^h} & \mathbf{for}\ \boldsymbol{0 \le h \le d_j} \\ \left(\boldsymbol{x_{ji} - \xi_{i(h-dj)}}\right)_+^{\boldsymbol{d_j}} & \mathbf{for}\ \boldsymbol{d_j + 1 \le h \le d_j + k_j} \end{cases} \tag{12}$$

where d_j is polynomial order of *j*-th predictor; k_j is the number of knots for *j*-th predictor; and h is a index of basis function which is a positive integer such that:

$$(\boldsymbol{x_{ji} - \xi_{j(h-d_j)}})_+^{\boldsymbol{d_j}} = \begin{cases} (\boldsymbol{x_{ji} - \xi_{j(h-d_j)}})_+^{\boldsymbol{d_j}} & , \boldsymbol{x \ge \xi_{j(h-d_j)}} \\ \mathbf{0} & , \boldsymbol{x < \xi_{j(h-d_j)}} \end{cases} \tag{13}$$

Hence, because of equations (10)–(13), penalized splines are polynomial slices in which different segments are joined together at knot points $\xi_1, \xi_2, \ldots, \xi_k$ in a way that ensures a certain continuity. The knot points contained in the penalized spline are determined based on a quantile sample of a single value of independent variables $\{x_i\}_{i=1}^n$. So, the multipredictor penalized spline functions for *n* observations are as follows:

$$m(x_{j1}) = \beta_{j0} + \beta_{j1}x_{j1}{}^1 + \cdots + \beta_{jd_j}x_{j1}{}^{d_j}\beta_{j(d_j+1)}(x_{j1} - \xi_{j1})_+^{d_j} + \cdots + \beta_{j(d_j+k_j)}\left(x_{j1} - \xi_{jk_j}\right)_+^{d_j}$$

$$m_j(x_{j2}) = \beta_{j0} + \beta_{j1}x_{j2}{}^1 + \cdots + \beta_{jd_j}x_{j2}{}^{d_j}\beta_{j(d_j+1)}(x_{j2} - \xi_{j1})_+^{d_j} + \cdots + \beta_{j(d_j+k_j)}\left(x_{j2} - \xi_{jk_j}\right)_+^{d_j}$$

$$\vdots$$

$$m_j(x_{jn}) = \beta_{j0} + \beta_{j1}x_{jn}{}^1 + \cdots + \beta_{jd_j}x_{jn}{}^{d_j}\beta_{j(d_j+1)}(x_{jn} - \xi_{j1})_+^{d_j} + \cdots + \beta_{j(d_j+k_j)}\left(x_{jn} - \xi_{jk_j}\right)_+^{d_j} \quad .$$

Hence, we can express the penalized spline function into a matrix notation as follows:

(14) $$\boldsymbol{m_j}(\boldsymbol{X_j}) = \boldsymbol{X_j\beta_j}$$

where

$$\boldsymbol{m_j}(\boldsymbol{X_j}) = \begin{bmatrix} m_j(x_{j1}) \\ m_j(x_{j2}) \\ \vdots \\ m_j(x_j) \end{bmatrix} ; \ \boldsymbol{X_j} = \begin{bmatrix} 1 & x_{j1}{}^1 & x_{j1}{}^2 & \cdots & x_{j1}{}^{d_j} & (x_{j1} - \xi_{j1})_+^{d_j} & \cdots & \left(x_{j1} - \xi_{jk_j}\right)_+^{d_j} \\ 1 & x_{j2}{}^1 & x_{j2}{}^2 & \cdots & x_{j2}{}^{d_j} & (x_{j2} - \xi_{j1})_+^{d_j} & \cdots & \left(x_{j2} - \xi_{jk_j}\right)_+^{d_j} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{jn}{}^1 & x_{jn}{}^2 & \cdots & x_{jn}{}^{d_j} & (x_{jn} - \xi_{j1})_+^{d_j} & \cdots & \left(x_{jn} - \xi_{jk_j}\right)_+^{d_j} \end{bmatrix} .$$

Estimate value $\hat{\boldsymbol{\beta}}_j$ can be obtained by minimizing a penalized least square (PLS) function of predictor variable x_j as follows:

(15) $$\boldsymbol{L_j = n^{-1}\sum_{i=1}^{n}\left(y_i - m_j(x_{ji})\right)^2 + \lambda_j \sum_{h=1}^{k_j} \beta_{j(d_j+h)}^2 \ , \ j = 1, 2, \ldots, p}$$

where λ_j is smoothing parameter of predictor variable x_j; d_j is order of polynomial for *j*-th predictor; and k_j is the number of knots for *j*-th predictor.

Next, we express the goodness of fit function namely $n^{-1}\sum_{i=1}^{n}\left(y_i - m_j(x_{ji})\right)^2$ of PLS function in (15) into a matrix notation as follows:

$$\boldsymbol{n^{-1}\sum_{i=1}^{n}\left(y_i - m_j(x_{ji})\right)^2 = n^{-1}(\mathbf{y} - X_j\beta_j)^T(\mathbf{y} - X_j\beta_j)}$$

$$= n^{-1}\left(\mathbf{y}^T - \boldsymbol{X}_j{}^T\boldsymbol{\beta_j}{}^T\right)\left(\mathbf{y} - \boldsymbol{X}_j\boldsymbol{\beta_j}\right)$$

$$= n^{-1}\left(\mathbf{y}^T\mathbf{y} - \mathbf{y}^T\boldsymbol{X}_j\boldsymbol{\beta_j} - \boldsymbol{\beta_j}{}^T\boldsymbol{X}_j{}^T\mathbf{y} + \boldsymbol{\beta_j}{}^T\boldsymbol{X}_j{}^T\boldsymbol{X}_j\boldsymbol{\beta_j}\right)$$

$$\boldsymbol{= n^{-1}\left(\mathbf{y}^T\mathbf{y} - 2\beta_j{}^TX_j{}^T\mathbf{y} + \beta_j{}^TX_j{}^TX_j\beta_j\right).}$$

Thus, we have:

$$n^{-1}\sum_{i=1}^{n}\left(y_i - m_j(x_{ji})\right)^2 = n^{-1}\left(\mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}_j^T \mathbf{X}_j^T \mathbf{y} + \boldsymbol{\beta}_j^T \mathbf{X}_j^T \mathbf{X}_j \boldsymbol{\beta}_j\right). \tag{16}$$

Also, we express the penalized function namely $\sum_{h=1}^{k_j} \beta_{j(d_{j+h})}^2$ of PLS function in (15) into an equation as follows:

$$\sum_{h=1}^{k_j} \boldsymbol{\beta}_{j(d_{j+h})}^2 = \boldsymbol{\beta}_{j(d_{j+1})}^2 + \boldsymbol{\beta}_{j(d_{j+2})}^2 + \cdots + \boldsymbol{\beta}_{j(d_{j+k_j})}^2 \tag{17}$$

Next, we define a diagonal matrix $\boldsymbol{D}_j$ as follows:

$$\boldsymbol{D}_j = \begin{pmatrix} a_{11} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & 0 & \cdots & 0 \\ 0 & \vdots & 0 & a_{(d_j+1)(d_j+1)} & 0 & 0 \\ \vdots & 0 & \cdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & a_{(d_j+k_j+1)(d_j+k_j+1)} \end{pmatrix} \tag{18}$$

where $a_{11}, a_{22} = \cdots = a_{(d_j+1)(d+1)} = 0$ dan $a_{(d_j+2)(d_j+2)} = a_{(d_j+k_j+1)(d_j+k_j+1)} = 1$.

Hence, the penalized function in (17) can be expressed in a matrix notation as follows:

$$\sum_{h=1}^{k_j} \boldsymbol{\beta}_{j(d_{j+h})}^2 = \tag{19}$$

$$\begin{pmatrix} \beta_{j0} & \beta_{j1} & \cdots & \beta_{(d_j+k_j)} \end{pmatrix} \begin{pmatrix} a_{11} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & 0 & \cdots & 0 \\ 0 & \vdots & 0 & a_{(d_j+1)(d_j+1)} & 0 & 0 \\ \vdots & 0 & \cdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & a_{(d_j+k_j+1)(d_j+k_j+1)} \end{pmatrix} \begin{pmatrix} \beta_{j0} \\ \beta_{j1} \\ \vdots \\ \beta_{(d_j+k_j)} \end{pmatrix}$$

$$= \boldsymbol{\beta}_j^T \boldsymbol{D}_j \boldsymbol{\beta}_j \;, \quad j = 1,2,\ldots,p.$$

Based on equations (16) and (19) we can express the PLS function in (15) into a matrix notation as follows:

$$\boldsymbol{L}_j = n^{-1}\left(\mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}_j^T \mathbf{X}_j^T \mathbf{y} + \boldsymbol{\beta}_j^T \mathbf{X}_j^T \mathbf{X}_j \boldsymbol{\beta}_j\right) + \lambda_j \boldsymbol{\beta}_j^T \boldsymbol{D}_j \boldsymbol{\beta}_j, \;\; j = 1,2,\ldots,p. \tag{20}$$

Furthermore, we can obtain $\widehat{\boldsymbol{\beta}}_j$ by applying a partial differentiation as follows:

$$\frac{\partial \boldsymbol{L}_j}{\partial \boldsymbol{\beta}_j} = \mathbf{0}$$

$$n^{-1}\left(0 - 2\boldsymbol{X}_j^T \mathbf{y} + 2\boldsymbol{X}_j^T \boldsymbol{X}_j \boldsymbol{\beta}_j\right) + 2\lambda_j \boldsymbol{D}_j \boldsymbol{\beta}_j = \mathbf{0}$$

$$-2\boldsymbol{X}_j^{\,T}\mathbf{y} + 2\boldsymbol{X}_j^{\,T}\boldsymbol{X}_j\boldsymbol{\beta}_{\boldsymbol{j}} + 2n\lambda_j\boldsymbol{D}_j\boldsymbol{\beta}_j = \mathbf{0}$$

$$\boldsymbol{X}_j^{\,T}\boldsymbol{X}_j\boldsymbol{\beta}_{\boldsymbol{j}} + n\lambda_j\boldsymbol{D}_j\boldsymbol{\beta}_j = \boldsymbol{X}_j^{\,T}\mathbf{y}$$

$$\left(\boldsymbol{X}_j^{\,T}\boldsymbol{X}_j + n\lambda_j\boldsymbol{D}_j\right)\boldsymbol{\beta}_{\boldsymbol{j}} = \boldsymbol{X}_j^{\,T}\mathbf{y}$$

Thus, we have the estimated parameter for the *j*-th predictor as follows:

(21) $$\widehat{\boldsymbol{\beta}}_j = \left(\boldsymbol{X}_j^{\,T}\boldsymbol{X}_j + n\lambda_j\boldsymbol{D}_j\right)^{-\mathbf{1}}\boldsymbol{X}_j^{\,T}\mathbf{y}\ ,\quad j = 1,2,\ldots,p.$$

Next, by substituting equation (21) into equation (14), we obtain the estimated regression function model for the *j*-th predictor as follows:

(22) $$\widehat{\boldsymbol{m}}_{\boldsymbol{j}}\left(\boldsymbol{X}_{\boldsymbol{j}}\right) = \boldsymbol{X}_{\boldsymbol{j}}\left(\boldsymbol{X}_{\boldsymbol{j}}^{\,\boldsymbol{T}}\boldsymbol{X}_{\boldsymbol{j}} + \boldsymbol{n\lambda_j D_j}\right)^{-\mathbf{1}}\boldsymbol{X}_{\boldsymbol{j}}^{\,\boldsymbol{T}}\mathbf{y}\ ,\quad \boldsymbol{j} = \mathbf{1},\mathbf{2},\ldots,\boldsymbol{p}.$$

We use equation (22) in the local scoring algorithm to obtain the regression function estimator for each additive nonparametric regression model whose response variables are from the exponential family. The local scoring algorithm consists of a scoring step (outer loop) which is iterated until the average deviance value converges and a weighted back-fitting step (inner loop) is iterated until the average Residual Sum of Square (RSS) value converges.

The local scoring algorithm follows the following steps are:

(a). Determining the optimal smoothing parameter value for *j*-th predictor based on Generalized Cross Validation (GCV): (i). determine the polynomial order, the number of knots and value of each knots; (ii).obtain matrix $\boldsymbol{X}_{\boldsymbol{j}}$ in accordance with polynomial order and knots; (iii). calculate $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$ in equation (21) and $\widehat{m}_j\left(\boldsymbol{X}_{\boldsymbol{j}}\right) = \boldsymbol{X}_{\boldsymbol{j}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$ in equation (22); (iv). calculate $\boldsymbol{H}\left(\lambda_j\right) = \boldsymbol{X}_{\boldsymbol{j}}\left(\boldsymbol{X}_{\boldsymbol{j}}^{\boldsymbol{T}}\boldsymbol{X}_{\boldsymbol{j}} + n\lambda_j\boldsymbol{D}_{\boldsymbol{j}}\right)^{-\mathbf{1}}\boldsymbol{X}_{\boldsymbol{j}}^{\boldsymbol{T}}$ as a smoother matrix; (v). calculate $GCV\left(\lambda_j\right) = \frac{n^{-1}\sum_{i=1}^{n}(y_i - m_{\lambda i})^2}{\left(n^{-1}trace\left[1-\boldsymbol{H}(\lambda_j)\right]\right)^2}$; (vi). iterate step (i) to step (vi) until minimum GCV value is obtained;

(b). Estimating penalized spline estimator using local scoring algorithm: (i). Obtain initial value of $\widehat{m}_j^{(r)}\left(X_j\right)$ based on equation (22) at iteration 0 (r = 0) from step (a); (ii). Calculate initial $m_i^{(0)} = \sum_{j=1}^{p}\widehat{m}_j^{(0)}(X_j)$; (iii). Calculate initial weighting matrix $diag(W^{(0)}) = \mu^{(0)} = \exp(\eta^{(0)})$; (iv). Calculate initial adjusted dependent vector $z^{(0)} = \eta^{(0)} + \left(Y - \mu^{(0)}\right)/diag(W^{(0)})$; (v). Calculate partial residual $R_{ij}^{(r+1)} = z_i - \sum_{r=1}^{j-1} m_r^{(r)}(X_{ir}) - \sum_{r=j+1}^{p} m_r^{(r)}(X_{ir})$; (vi). Calculate $\widehat{m}_j^{(r+1)}(X_j) = H(\lambda_j) R_{ij}^{(r+1)}$; (vii). Calculate $\hat{\mu}_i = \exp\left(\sum_{j=i}^{p} \widehat{m}_j^{(r+1)}(X_j)\right)$; (viii). Calculate Mean Square Error or $MSE^{(s+1)} = \frac{1}{n}\left(Y-\mu^{(r+1)}\right)^T\left(Y-\mu^{(r+1)}\right)$; (ix). Iterate step (i) to step (vii) until the value of $MSE^{(r+1)}$ convergences to a small value of > 0 ; (x). Recalculate $\eta^{(r+1)}, z^{(r+1)}, W^{(r+1)}$ and $\mu^{(r+1)}$; (xi). Calculate $avg(Deviance) = \frac{2}{n}\sum_{i=1}^{n}\left\{y_i ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i + \gamma^{-1}) ln\left[\frac{y_i+\gamma^{-1}}{\hat{\mu}_i+\gamma^{-1}}\right]\right\}$; and (xii). Iterate step (i) to step (x) until mean of deviance convergences to a small value of $\varepsilon > 0$ as follows:

$$|Avg(Deviance)^{(r+1)} - Avg(Deviance)^{(r)}| < \varepsilon$$

4. CONCLUSIONS

The estimated regression function of the NNBR model based on penelized spline estimator which is applied to GAM is obtained by using local scoring algorithm. In the future, this estimated model can be applied to the real data namely the number of Covid-19 death cases affected by comorbidities such as percentage of diabetes mellitus patients, percentage of hypertension over 15 years old patients, and percentage of tuberculosis patients.

ACKNOWLEDGEMENTS

The authors are very grateful to Faculty of Science and Technology, Airlangga University, Surabaya, Indonesia, for funding this research through the PUF (*Penelitian Unggulan Fakultas*) grant with contract number 1437/UN3.1.8/PT/2022. Also, the authors thank Dr. Drs. Budi Lestari, PG.Dip.Sc., M.Si., a proof-reader from the University of Jember, Indonesia who has provided corrections, criticisms, and suggestions that were useful for improving the quality of this manuscript.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] Satgas Covid-19, Persentase Kematian Covid RI Konsisten di Atas Dunia, CNN Indonesia, Jakarta (2021), https://www.cnnindonesia.com/nasional/20210831192832-20-688013/satgas-persentase-kematian-covid-ri-konsisten-di-atas-dunia. [Access 18 September 2021].

[2] G.D. Rembulan, T. Wijaya, D. Palullungan, et al. Kebijakan Pemerintah Mengenai Coronavirus Disease (COVID-19) di Setiap Provinsi di Indonesia Berdasarkan Analisis Klaster, J. Ind. Eng. Manage. Syst. 13 (2020), 74-86. https://doi.org/10.30813/jiems.v13i2.2280.

[3] R.A. Putra, Faktor Pengaruh Tingginya Rasio Kematian Akibat COVID-19 di Indonesia, (2020), https://www.dw.com/id/tingginya-rasio-kematian-corona-di-indonesia/a-52835267. [Access 18 September 2021].

[4] E. Purwaningsih, Analisis Kebijakan Kesehatan Berdasarkan Analisis Kelompok Risiko Terhadap Persebaran Kasus Covid-19 di Indonesia Tahun 2020, J. Kebijakan Kesehatan Indonesia, 10 (2021), 86–93. https://doi.org/10.22146/jkki.61663.

[5] A.B. Hogan, B.L. Jewell, E. Sherrard-Smith, et al. Potential impact of the COVID-19 pandemic on HIV, tuberculosis, and malaria in low-income and middle-income countries: a modelling study, Lancet Glob. Health. 8 (2020), e1132–e1141. https://doi.org/10.1016/s2214-109x(20)30288-6.

[6] Nidhomuddin, N. Chamidah, A. Kurniawan, Confidence interval of the parameter on multipredictor biresponse longitudinal data analysis using local linear estimator for modeling of case increase and case fatality rates COVID-19 in Indonesia: A theoretical discussion, Commun. Math. Biol. Neurosci., 2022 (2022), Article ID 23. https://doi.org/10.28919/cmbn/6900.

[7] D. Handayani, D.R. Hadi, F. Isbaniah, E. Burhan, H. Agustin, Corona Virus Disease 2019, J. Respirol. Indones. 40 (2020), 119–129. https://doi.org/10.36497/jri.v40i2.101.

[8] Kemenkes RI, Situasi Terkini Perkembangan Novel Coronavirus (Covid-19): Data dilaporkan sampai 31 Desember 2020, Kementerian Kesehatan Republik Indonesia, (2021). https://infeksiemerging.kemkes.go.id/situasi-infeksi-emerging/situasi-terkini-perkembangan-coronavirus-disease-covid-19-01-januari-2021 [Access 12 October 2021].

[9] W. Yustanti, N. Rahmawati, Y. Yamasari, Klastering Wilayah Kota/Kabupaten Berdasarkan Data Persebaran Covid-19 Di Propinsi Jawa Timur dengan Metode K-Means, J. Inform. Eng. Educ. Technol. 4 (2020), 1–9. https://doi.org/10.26740/jieet.v4n1.p1-9.

[10] P. McCullagh, A. Nelder, Generalized linier model, 2nd Edition, Champman and Hall, London, (1989).

[11] J. M. Hilbe, Binomial negative regression, Cambridge University Press, New York, (2011).

[12] T.J. Hastie, R.J. Tibshirani, Generalized additif models, Chapman and Hall, London, (1990).

[13] 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.

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

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

[16] 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.

[17] N. Chamidah, B. Lestari, Estimation of covariance matrix using multi-response local polynomial estimator for designing children growth charts: A theoretically discussion, J. Phys.: Conf. Ser. 1397 (2019), 012072. https://doi.org/10.1088/1742-6596/1397/1/012072.

[18] B. Lestari, Fatmawati, I.N. Budiantara, et al. 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.

[19] 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. http://repository.unair.ac.id/id/eprint/114276.

[20] 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. Create. Change. 13 (2020), 45–67.

[21] N. Chamidah, Y.S. Yonani, E. Ana, B. Lestari, Identification the number of Mycobacterium tuberculosis based on sputum image using local linear estimator, Bull. Electric. Eng. Inform. 9 (2020), 2109–2116. https://doi.org/10.11591/eei.v9i5.2021.

[22] 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.

[23] 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.

[24] N. Chamidah, K.H. Gusti, E. Tjahjono, et al. Improving of classification accuracy of cyst and tumor using local polynomial estimator, Telkomnika. 17 (2019), 1492-1500. https://doi.org/10.12928/telkomnika.v17i3.12240.

[25] N. Chamidah, E. Tjahjono, A.R. Fadilah, et al. Standard growth charts for weight of children in East Java using local linear estimator, J. Phys.: Conf. Ser. 1097 (2018) 012092. https://doi.org/10.1088/1742-6596/1097/1/012092.

[26] B. Lestari, Fatmawati, I.N. Budiantara, et al. Estimation of regression function in multi-response nonparametric regression model using smoothing spline and kernel estimators, J. Phys.: Conf. Ser. 1097 (2018), 012091. https://doi.org/10.1088/1742-6596/1097/1/012091.

[27] B. Lestari, I.N. Budiantara, S. Sunaryo, M. Mashuri, Spline smoothing for multi-response nonparametric regression model in case of heteroscedasticity of variance, J. Math. Stat. 8(3) (2012), 377–384. https://doi.org/10.3844/jmssp.2012.377.384.

[28] B. Lestari, I.N. Budiantara, S. Sunaryo, et al. Spline estimator in multi-response nonparametric regression model. Jurnal ILMU DASAR. 11 (2010), 17–22.

[29] B. Lestari, I.N. Budiantara, S. Sunaryo, et al. Matriks Kovariansi dalam Regresi Nonparametrik Multirespon Pada Kasus Korelasi Sama dan Korelasi Tidak Sama. Jurnal Ilmiah Matematika dan Pendidikan Matematika. 4 (2012), 161–171. https://doi.org/10.20884/1.jmp.2012.4.1.2950.

[30] B. Lestari, Semiparametric modeling of consumer price index. Majalah Ilmiah Matematika dan Statistika, 16 (2016), 79–90.

[31] B. Lestari, N. Chamidah, T. Saifudin, Estimasi Fungsi Regresi Dalam Model Regresi Nonparametrik Birespon Menggunakan Estimator Smoothing Spline dan Estimator Kernel. Jurnal Matematika, Statistika & Komputasi. 15 (2019), 20–24. https://doi.org/10.20956/jmsk.v15i2.5710.

[32] B. Lestari, N. Chamidah, Estimating regression function of multi-response semiparametric regression model using smoothing spline. J. Southwest Jiaotong Univ. 55 (2020), 1–9. https://doi.org/10.35741/issn.0258-2724.55.5.26.