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

NIDHOMUDDIN\textsuperscript{1}, NUR CHAMIDAH\textsuperscript{2,*}, ARDI KURNIAWAN\textsuperscript{2}

\textsuperscript{1}Doctoral Study Program of Mathematics and Natural Science, Faculty of Science and Technology, Airlangga University, Surabaya 60115, Indonesia

\textsuperscript{2}Department of Mathematics, 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: In this paper, we describe a theoretical discussion about confidence intervals for longitudinal data based on local linear estimator. The confidence interval represents the range of possible values in the estimating process. The confidence intervals for the parameter in nonparametric regression can be used to determine the predictor variables that have a significant effect on the response variable. In this research, we theoretically discuss estimation of the confidence interval of the parameter on multipredictor bireponse nonparametric regression model for longitudinal data based on local linear estimator which is applied to data of the case increase and case fatality rates COVID-19 in Indonesia. The estimation result can be used to determine the predictor variable, e.g. temperature which has a significant effect on the case increase and case fatality rates COVID-19 in Indonesia so that it can be advised to the ministry of health to control the case increase and case fatality rates COVID-19 in Indonesia.

Keywords: Confidence interval, local linear, case increase, case fatality, COVID-19.

2010 AMS Subject Classification: 62G05, 62G08, 62E15.

*Corresponding author
E-mail address: nur-c@fst.unair.ac.id
Received October 17, 2021
1. INTRODUCTION

Regression analysis is a method that can be used to explain the relationship pattern between response variables and predictor variables. If the pattern does not follow a certain pattern, such as linear, quadratic, then modeling using nonparametric regression approach is appropriate to solve it [1]. Several researchers who have studied some estimators in nonparametric regression analysis are [2–11] used splines estimators; [12,13] used both smoothing spline and kernel estimators; [14–18] used local polynomial estimators; [19–25] used local linear estimators, and [26] used local linear estimator and stated that it is better than kernel estimator because it has mean square prediction less than that kernel estimator.

In real life, there are usually many cases involving regression models with two response variables that are correlated with each other and influenced by more than one predictor variable, so that the problem can be solved by using multipredictor biresponse nonparametric regression. Several researchers who have studied biresponse nonparametric regression, for examples, [27,28] studied splines estimators for biresponse nonparametric regression applied to hormone data using spline estimator; [29,30] studied biresponse nonparametric regression using local linear estimator; [31] studied multipredictor biresponse nonparametric regression using local linear estimator. However, these researchers are only limited to study point estimation.

One of the most important parts of statistical inference is the confidence interval. The confidence interval represents the range of possible values in which, with some certainty, we can find the statistical size of the population [32]. Confidence intervals for parameters in nonparametric regression can be used to determine the predictor variables that have a significant effect on the response variable. The conclusion is decided by looking at whether the parameter confidence interval contains a zero value. If the confidence interval contains a value of zero, then the predictor variable has no significant effect on the response variable. There are several researchers who have discussed confidence intervals. They are, for examples, [33] used spline estimator; [34] used local polynomial estimator for estimating confidence intervals.

Although much research has been studied on confidence intervals however the research was
applied to cross-section data. Therefore, the aim of this research is to create a confidence interval of parameter for nonparametric regression model with two response variables and more than one predictor variable which is applied to longitudinal data, namely data of case increase and case fatality rates COVID-19 in Indonesia. Hence, in this research, we discuss theoretically how to estimate the confidence interval of the parameter on nonparametric regression model by using a local linear estimator. The linear local estimator has the advantage of estimating the function at each point so that the model we get is closer to the actual data pattern then can be used to construct a confidence interval of case increase and case fatality rates COVID-19 in Indonesia.

2. Preliminaries

Multi-predictor biresponse nonparametric regression model on longitudinal data is a nonparametric regression model that describes the relationship pattern of two correlated response variables and more than one predictor variable with a dataset collected from several subjects in a certain period. Suppose we have a paired longitudinal data \((x_{1ik}, x_{2ik}, \ldots, x_{pik}, y_{ik}^{(1)}, y_{ik}^{(2)})\), where \(i = 1, 2, \ldots, n\); \(k = 1, 2, \ldots, m_i\); and \(p\) is the number of predictor variables. The nonparametric regression model can be expressed as follows [28]:

\[
y_{ik} = f_j(x_{1ik}) + f_j(x_{2ik}) + \cdots + f_j(x_{pik}) + \varepsilon_{ik}
\]

\[
= \sum_{j=1}^{p} f_j(x_{ijk}) + \varepsilon_{ik}
\]

where \(\mathbf{y} = \begin{pmatrix} y^{(1)} \\
, y^{(2)} \end{pmatrix}^T\), \(f_j(x_{ijk}) = \begin{pmatrix} f_{(1),j}(x_{ijk}), f_{(2),j}(x_{ijk}) \end{pmatrix}^T\), and \(\varepsilon = \begin{pmatrix} \varepsilon^{(1)} \\
, \varepsilon^{(2)} \end{pmatrix}^T\). \(\varepsilon\) is random error which assumed to follow a normal distribution with mean \(\mathbf{0}\) and variance \(\Sigma\). Matrix \(\Sigma\) can be expressed as \(\Sigma = \sigma^2 \mathbf{W}^{-1}\), \(\mathbf{W}\) is a weighted matrix obtained from the inverse variance-covariance matrix of the first response variable error and second response variable error.

Regression function \(f_{\tau(r)}(x); r = 1, 2\) in equation (1) would be estimated using local linear estimator and it can be written as follows:
For two response variables, equation (2) can be written as follows:

\[
\begin{align*}
\beta_1(x) &= \beta_0^{(1)}(x_o) + (x_1 - x_{o1}) \beta_1^{(1)}(x_o) + (x_2 - x_{o2}) \beta_2^{(1)}(x_o) + \cdots + (x_p - x_{op}) \beta_p^{(1)}(x_o) \\
\beta_2(x) &= \beta_0^{(2)}(x_o) + (x_1 - x_{o1}) \beta_1^{(2)}(x_o) + (x_2 - x_{o2}) \beta_2^{(2)}(x_o) + \cdots + (x_p - x_{op}) \beta_p^{(2)}(x_o)
\end{align*}
\]

Hence, equation (3) can be expressed as follows:

\[
f(x) = \mathbf{X}^\top(x_o) \beta(x_o)
\]

where \( f = [f_1(x) f_2(x)]^\top \), \( \mathbf{X} = \begin{bmatrix} x_1^{(1)} & 0 \\ 0 & x_2^{(2)} \end{bmatrix} \) \( x^{(i)} = (1 \ (x_1 - x_{o1}) \ (x_2 - x_{o2}) \ \cdots \ (x_p - x_{op})) \)

\[
\beta = \begin{bmatrix} \beta_1^{(1)} \\ \beta_2^{(2)} \end{bmatrix}, \tilde{\beta}^{(i)} = \begin{bmatrix} \beta_0^{(i)}(x_o) \\ \beta_1^{(i)}(x_{o1}) \\ \cdots \\ \beta_p^{(i)}(x_{op}) \end{bmatrix}^\top
\]

Furthermore, the estimated parameter of multi-predictor biresponse nonparametric regression model in equation (4) is obtained using the weighted least square method, and therefore we have the estimated parameter as follows [30]:

\[
\tilde{\beta}(x_o) = [\mathbf{X}^\top(x_o) \mathbf{W} \mathbf{K}_h(x_o) \mathbf{X}(x_o)]^{-1} \mathbf{X}^\top(x_o) \mathbf{W} \mathbf{K}_h(x_o) y
\]

3. **Main Results**

The first step to construct the confidence interval of the parameter on multi-predictor biresponse nonparametric regression model is by determining the mean and variance of \( y \). The obtained result would be used to determine the mean and variance of \( \tilde{\beta}(x_o) \). By assuming \( \varepsilon \sim N(0, \sigma^2 \mathbf{W}^{-1}) \), and since \( y \) is a linear combination of \( \varepsilon \), then \( y \) is also to follow a normal distribution. Hence, we have:

\[
E(y) = E(\mathbf{X}(x_o) \beta(x_o) + \varepsilon) = \mathbf{X}(x_o) \beta(x_o) + E(\varepsilon) = \mathbf{X}(x_o) \beta(x_o)
\]

\[
\text{var}(y) = \text{var}(\mathbf{X}(x_o) \beta(x_o) + \varepsilon) = 0 + \text{var}(\varepsilon) = \sigma^2 \mathbf{W}^{-1}
\]

Based on (6) and (7), we get \( y \) which follows a normal distribution with mean \( \mathbf{X}(x_o) \tilde{\beta}(x_o) \) and variance \( \sigma^2 \mathbf{W}^{-1} \). The mean and variance values of \( \tilde{\beta}(x_o) \) are as follows:
 escort abismo  acepción compendio distinto
The next step is determining the distribution of pivotal quantity. If the numerator and denominator are divided by the root of the population variance $\sigma^2$ then we will obtain the pivotal quantity which follows a t-student distribution with degree of freedom $2R-(2(p+1))$ and therefore it can be written as follows:

\[ U(x) = \frac{\frac{\hat{\beta}(x_0) - \beta(x_0)}{\sqrt{\sigma^2 v(x_0)}}}{\sqrt{\frac{y^T[I-A]y}{2R-(2(p+1))}}} \]

Furthermore, to prove equation (16), we must prove Theorem 1, Theorem 2., and Theorem 3 as follows:

**Theorem 1.** If $\hat{\beta}(x_0) - N(\beta(x_0), \sigma^2 v(x_0))$ then $B \sim N(0,1)$

**Proof.** Noted that $B = \frac{\hat{\beta}(x_0) - \beta(x_0)}{\sqrt{\sigma^2 v(x_0)}}$, since $B$ is a linear combination of $\hat{\beta}(x_0)$ then $B$ follows a normal distribution $(B \sim N(E(B), \text{var}(B)))$.

**Example:**

\[
\begin{align*}
E(B) &= E \left( \frac{\hat{\beta}(x_0) - \beta(x_0)}{\sqrt{\sigma^2 v(x_0)}} \right) = E \left( \frac{\hat{\beta}(x_0) - E(\beta(x_0))}{\sqrt{\sigma^2 v(x_0)}} \right) = \frac{E(\hat{\beta}(x_0)) - E(\beta(x_0))}{\sqrt{\sigma^2 v(x_0)}} = 0 \\
\text{Var}(B) &= \text{Var} \left( \frac{\hat{\beta}(x_0) - \beta(x_0)}{\sqrt{\sigma^2 v(x_0)}} \right) = \left( \frac{1}{\sqrt{\sigma^2 v(x_0)}} \right)^2 \text{Var} \left( \hat{\beta}(x_0) - \beta(x_0) \right) = \frac{\sigma^2 v(x_0)}{\sigma^2 v(x_0)} = 1
\end{align*}
\]

Based on equation (17) and (18), we get: $B \sim N(0,1)$
**Theorem 2.** If \( y \sim N(\mathbf{X}(\mathbf{x}_0)\beta(\mathbf{x}_0),\sigma^2\mathbf{W}^{-1}) \) then \( \frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\sigma^2} \) is \( \chi^2_{a, \frac{1}{2\sigma^2}(\mathbf{y}^T\mathbf{y}^T\mathbf{W}^{-1}\mathbf{x}_0)} \) if and only if \( \mathbf{B} \) is idempotent of rank \( a \).

**Proof.** Noted that \( A = \frac{\mathbf{y}^T[\mathbf{I}-\mathbf{A}]\mathbf{y}}{\sigma^2} = \frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\sigma^2} \) and \( \mathbf{A} = \mathbf{X}(\mathbf{X}^T\mathbf{K}_h\mathbf{W})^{-1}\mathbf{X}^T\mathbf{K}_h\mathbf{W} \). Hence, we get \( \mathbf{B} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{K}_h\mathbf{W})^{-1}\mathbf{X}^T\mathbf{K}_h\mathbf{W} \). Furthermore, it will be proved that \( \mathbf{B} \) is idempotent matrix as follows:

\[
\begin{align*}
\mathbf{BB} &= \left( \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{K}_h\mathbf{W})^{-1}\mathbf{X}^T\mathbf{K}_h\mathbf{W} \right) \left( \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{K}_h\mathbf{W})^{-1}\mathbf{X}^T\mathbf{K}_h\mathbf{W} \right) \\
&= \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{K}_h\mathbf{W})^{-1}\mathbf{X}^T\mathbf{K}_h\mathbf{W} \\
&= \mathbf{B}
\end{align*}
\]

Equation (19) shows that \( \mathbf{B} \) is idempotent. The next step is to calculate the rank \( (\mathbf{B}) \).

\[
\begin{align*}
a &= \text{rank} \left( \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{K}_h\mathbf{W})^{-1}\mathbf{X}^T\mathbf{K}_h\mathbf{W} \right) \\
&= \text{trace} \left( \mathbf{I} \right) - \text{trace} \left( (\mathbf{X}^T\mathbf{K}_h\mathbf{W})^{-1}\mathbf{X}^T\mathbf{K}_h\mathbf{W} \right) \\
&= \text{trace} \left( \mathbf{I}_{2R} \right) - \text{trace} \left( \mathbf{I}_{2(p+1)} \right) \\
&= 2R - (2(p+1))
\end{align*}
\]

Then, we calculate the value of \( \lambda \) as follows:

\[
\begin{align*}
\lambda &= \frac{1}{2\sigma^2} \left( \mathbf{X}\hat{\mathbf{f}} \right)^T \left( \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{K}_h\mathbf{W})^{-1}\mathbf{X}^T\mathbf{K}_h\mathbf{W} \right) \left( \mathbf{X}\hat{\mathbf{f}} \right) = 0
\end{align*}
\]

Based on equation (19), (20) and (21) we then have \( A = \frac{\mathbf{y}^T[\mathbf{I}-\mathbf{A}]\mathbf{y}}{\sigma^2} \sim \chi^2_{(2R-(2(p+1)))} \).

**Theorem 3.** If \( y \sim N(\mathbf{X}(\mathbf{x}_0)\beta(\mathbf{x}_0),\sigma^2\mathbf{W}^{-1}) \) then \( \mathbf{D}_y \) and \( \mathbf{y}^T\mathbf{B}\mathbf{y} \) are independent if and only if \( \mathbf{D}_y\mathbf{B} = 0 \).

**Proof.** Noted that \( \mathbf{D}_y = \mathbf{X}\hat{\mathbf{f}} = \mathbf{X}(\mathbf{X}^T\mathbf{K}_h\mathbf{W})^{-1}\mathbf{X}^T\mathbf{K}_h\mathbf{W} \) then \( \mathbf{D} = \mathbf{X}(\mathbf{X}^T\mathbf{K}_h\mathbf{W})^{-1}\mathbf{X}^T\mathbf{K}_h\mathbf{W} \) and \( \mathbf{B} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{K}_h\mathbf{W})^{-1}\mathbf{X}^T\mathbf{K}_h\mathbf{W} \). It will show that \( \mathbf{D}\mathbf{B} \) is as follows:

\[
\begin{align*}
\mathbf{DB} &= \left( \mathbf{X}(\mathbf{X}^T\mathbf{K}_h\mathbf{W})^{-1}\mathbf{X}^T\mathbf{K}_h\mathbf{W} \right) \left( \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{K}_h\mathbf{W})^{-1}\mathbf{X}^T\mathbf{K}_h\mathbf{W} \right) = 0
\end{align*}
\]
It means that equation (17) is proven. So, equation (17) can be written as follows:

$$\hat{\beta}(x_0)_g - \beta(x_0)_g$$

(23) \[ U(x) = \frac{\hat{\beta}(x_0)_g - \beta(x_0)_g}{\sqrt{\sigma^2 v(x_0)_{gg}}} \sim t_{(2R-(2p+1))} \]

Therefore, the pivotal quantity in (23) can be written as follows:

(24) \[ U(x) = \frac{\hat{\beta}(x_0)_g - \beta(x_0)_g}{\sqrt{\frac{y^T [I - A] y}{\sigma^2 v(x_0)_{gg}}} \sim t_{(2R-(2p+1))} \]

The next step after obtaining the pivotal quantity is to construct the confidence interval for parameters of multi-predictor biresponse nonparametric regression model based on local linear estimator for longitudinal data. The confidence interval can be obtained by solving the probability as follows.

(25) \[ P(a_{jk} \leq U_{jk}(x) \leq b_{jk}) = 1 - \alpha \quad i = 1, 2, ..., n; j = 1, 2, ..., p; k = 1, 2, ..., m_i \]

where \( a_{jk} \) and \( b_{jk} \) are constants, \( a_{jk} < b_{jk} \) and \( \alpha \) is a margin of error. If equation (24) is substituted into equation (25), then we get

(26) \[ P \left( a_{jk} \leq \frac{\hat{\beta}(x_0)_g - \beta(x_0)_g}{\sqrt{\frac{y^T [I - A] y}{\sigma^2 v(x_0)_{gg}}} \leq b_{jk} \right) = 1 - \alpha \]

Equation (26) can be simplified into the form as follows:

(27) \[ P \left( \beta(x_0)_g - b_{jk} L \leq \beta(x_0)_g \leq \hat{\beta}(x_0)_g - a_{jk} L \right) = 1 - \alpha \]

where \( L = \sqrt{\frac{y^T [I - A] y}{\sigma^2 v(x_0)_{gg}}} \)

Furthermore, \( a_{jk} = -b_{jk} \) and \( b_{jk} \) is obtained from the t-student distribution for \( \frac{\alpha}{2} \) and degree of freedom of \( 2R-(2(p+1)) \). Hence, the confidence interval for the parameter of multi-predictor
bireponse nonparametric regression model based on local linear estimator for longitudinal data is as follows:

\[(28) \quad P\left( \hat{\beta}\left(x_0\right) - t_{\alpha/2R(2(p+1))}L \leq \beta\left(x_0\right) \leq \hat{\beta}\left(x_0\right) + t_{\alpha/2R(2(p+1))}L \right) = 1 - \alpha \]

The confidence interval for parameter of nonparametric regression model in equation (28) can be used to determine the predictor variables that have a significant effect on the response variable which is applied to data of the case increase and case fatality rates COVID-19 in Indonesia.

4. CONCLUSIONS

Theoretically, based on equation (28) we can obtain the confidence interval of case increase and case fatality rates COVID-19 in Indonesia and then we can determine the predictor variables that have a significant effect on the case increase and case fatality rates COVID-19 in Indonesia. The estimation result can be advised to the ministry of health to control the case increase and case fatality rates COVID-19 in Indonesia.

ACKNOWLEDGMENT

The authors thank the Directorate of Research, Technology and Community Service (DRTPM), Directorate General of Higher Education, Research and Technology, the Ministry of Education, Culture, Research and Technology, The Republic of Indonesia for funding this research. Additionally, the authors are grateful to Dr. Drs. Budi Lestari, PG.Dip.Sc., M.Si., from The University of Jember, who 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


A. Baíllo, A. Grané, Local linear regression for functional predictor and scalar response, J. Multivar. Anal. 100


