ISSN 2229-5518

Confidence Intervals for Parameters of Nonparametric Regression Spline Truncated Model for Longitudinal Data Using Pivotal Quantity

Khaerul Umam[1], Ismaini Zain[2], I Nyoman Budiantara[3]

Abstract—The relationship pattern between the response variable with predictor variable can be obtained by using regression curve. The approach in regression curve that often used is a parametric regression approach, which assumed the form of regression curve is known. However, not all pattern of regression curve is known that there are no information about the kind of relationship between the response variable and predictor variable. If the regression curve is unknown then we can predict the model using nonparametric regression model approach. Nonparametric function estimation is a major research area at the present time. One of the nonparametric regression approaches is the spline truncated, which has the advantage of knot points. With the point of knots, the resulting model will follow the form of changes in data behavior patterns. Data obtained from the repeated observation of each object at different time intervals is called longitudinal data. Studies of spline-truncated nonparametric regression using longitudinal data have been limited to obtaining point estimation. While the point estimation has a weakness that is the probability of error in guessing the true value of a parameter θ is greater. It therefore needs a range of probability values for θ with a certain degree of confidence called the interval estimation or confidence interval. Confidence interval is one of the most important parts of statistical inference. The confidence interval confirms that the true value of parameter value will be in the range interval. In constructing the shortest interval for the parameters of spline truncated nonparametric regression model of longitudinal data using pivotal quantity. From the result obtained confidence intervals for parameters of nonparametric spline truncated regression for longitudinal data using pivotal quantity which is distributed student-t.

Index Terms— confidence interval, parameters, nonparametric regression, spline truncated, optimum knot points, longitudinal data, pivotal quantity.

[1] Student of master degree program in Department of Statistics in Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. E-mail: khaerulumam9100@mail.com
[2] Doctor, Lecturer in Department of Statistics in Institut Teknologi Sepuluh Nopember,Surabaya, Indonesia. E-mail: ismainizain @mail.com
[3] Professor, Lecturer in Department of Statistics in Institut Teknologi Sepuluh Nopember,Surabaya, Indonesia. E-mail: i_nyoman_b @its.ac.id

ISSN 2229-5518

1 Introduction

The relationship pattern between the response variable with predictor variable can be obtained by using regression curve. The approach in regression curve that often used is a parametric regression approach, which assumed the form of regression curve is known. However, not all pattern of regression curve is known that there are no information about the kind of relationship between the response variable and predictor variable. If the regression curve is unknown then we can predict the model using nonparametric regression model approach. Nonparametric function estimation is a major research area at the present time and we just mention representative example of modern techniques for multivariate function estimation in several dimension such as ACE [1], MARS [2], Additive Models [3] and Regression Spline [4], [5], and [6]. The curve simply assumed to be smooth in the sense contained in a particular function space. If the form of regression curve is not known pattern, then nonparametric regression analysis is more recommended for use [7]. One of the nonparametric function estimation is regression spline. Spline has a high flexibility and ability to estimate data behavior that tends to differ at different intervals and spline is a model providing superior and very respectable visually statistical interpretation [8]. Some research using spline functions are [9], [10], [11] and [12]. One of the family of spline function is spline truncated [13] and [14]. Spline truncated has knot points. Knot points will connect pieces polynomial such that the spline truncated has good flexibility for nonlinearity relationship between predictor and response variables [13].

Changes in data behavior patterns if observed year by year can provide more complete information about the dynamics of changes in the behavioral patterns of the data. Data obtained from the repeated observation of each object at different time intervals is called longitudinal data. Longitudinal data modelling with nonparametric regression has been developed by [15], [16] and [17]. Research of longitudinal data modelling with nonparametric regression spline truncated has been proposed by [18], but limited to obtaining point estimation. While the point estimation has a weakness that is the probability of error in guessing the true value of a parameter (θ) is greater. It therefore needs a range of probability values for (θ) with a certain degree of confidence called the interval estimation or confidence interval [19].

Confidence interval is one of the most important parts of statistical inference. The confidence interval confirms that the true value of parameter value will be in the range interval. Confidence interval with smoothing spline has been studied by [20], but not using longitudinal data nor spline truncated. In constructing the shortest interval for the parameters of spline truncated nonparametric regression model of longitudinal data using pivotal quantity. The pivotal quantity method is mainly due to George Bernard and David Fraser of The University of Waterloo, and this method is perhaps one of the most elegant methods of constructing confidence intervals for unknown parameters [19]. Confidence intervals for parameters of nonparametric regression can be used to determine predictor variables that significantly influence response variables. If the confidence interval contains a zero value, then the predictor variable has no significant effect on the response variable. In this research developed the shortest confidence intervals of spline truncated nonparametric regression model parameters for longitudinal data.

2 Literature Review

2.1 Nonparametric Regression

The regression curve between predictor and response variables is not always known. If forced to use parametric regression then the resulting model is not in accordance with the form of relationship pattern which will ultimately produce a large error. Nonparametric regression is one of the approaches used to determine the relationship pattern between predictor variables and the unknown response of the regression curve or no complete past information about the shape of the data pattern [2].

Some approaches in nonparametric regression include: spline, kernel, fourier series, wavelet, etc. Spline is an approach often used in nonparametric regression. Spline regression has a functional basis which in its parameter optimization process uses optimization. Spline regression has the advantage of adjusting data patterns that change sharply with knots.

In general, nonparametric regression model:

$$y_i = f(z_i) + \varepsilon_i, i = 1,2,\dots,n, \tag{1}$$

With y_i is the i-th response variable, while the function $f(z_i)$ is the regression curve, with z_i as the predictor variable and ε_i is the random error assumed to be independent normal distribution with mean zero and variance σ^2 [4].

2.2 Spline Truncated Nonparametric Regression for Longitudinal Data

Spline truncated nonparametric regression model on longitudinal data can be written in the form:

$$y_{iw} = f(z_{iw}) + \varepsilon_{iw}, i = 1,2,\dots,n; w = 1,2,\dots,t \tag{2}$$

With

$$f(z_{iw}) = \sum_{j=0}^{m} \beta_{ji} z_{iw}^j + \sum_{k=1}^{r} \gamma_{ki}(z_{iw} - K_{ki})_+^m, \tag{3}$$

where n is the number of observed objects and t is the amount of time of the object being observed, while

$$\sum_{j=0}^{m} \beta_{ji} z_{iw}^j$$

is polynomial components and

$$\sum_{k=1}^{r} \gamma_{ki}(z_{iw} - K_{ki})_{+}^{m}$$

is truncated with:

$$(z_{iw} - K_{ki})_{+}^{m} = \begin{cases} (z_{iw} - K_{ki})^{m}, & z_{iw} \geq K_{ki} \\ 0 & , z_{iw} < K_{ki} \end{cases} \quad (4)$$

Equation (1) is a nonparametric regression spline-truncated form in longitudinal data with one nonparametric predictor variable. If the nonparametric spline truncated regression in the longitudinal data consists of one response variable with a nonparametric predictor variable of q, then the spline truncated regression curve for longitudinal data with m = 1 can be expressed in terms of the following equation

$$f(z_{iwl}) = \beta_{0i} + \sum_{l=1}^{q}\left(\beta_{li} z_{iwl} + \sum_{k=1}^{r} \gamma_{kli}(z_{iwl} - K_{kli})_{+}^{1}\right). \quad (5)$$

So, equation (2) becomes

$$y_{iw} = \beta_{0i} + \sum_{l=1}^{q}\left(\beta_{li} z_{iwl} + \sum_{k=1}^{r} \gamma_{kli}(z_{iwl} - K_{kli})_{+}^{1}\right) + \varepsilon_{iw}, \quad (6)$$

with $i = 1, 2, \ldots, n; w = 1, 2, \ldots, t$.

In longitudinal data, parameter estimates were obtained using Weighted Least Square (WLS) to overcome correlations in the same observational subjects. Then write equation (6) in matrix notation as follows

$$\tilde{y} = \boldsymbol{Z}[\mathbf{K}]\tilde{B} + \tilde{\varepsilon} \quad (7)$$

with $K = K_1, K_2, \ldots, K_r$.

The $\tilde{B}$ estimator is obtained by completing the WLS optimization as follows:

$$\min_{B \in R^{nt(1+q(r+1))}} \{(\tilde{y} - \boldsymbol{Z}[\mathbf{K}]\tilde{B})^{\mathrm{T}} \mathbf{W}^{-1} (\tilde{y} - \boldsymbol{Z}[\mathbf{K}]\tilde{B})\}, \quad (8)$$

With matrix $\boldsymbol{W}$ is given by

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{W_1} & 0 & \cdots & 0 \\ 0 & \boldsymbol{W_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{W_n} \end{bmatrix}. \quad (9)$$

3 Methods

3.1 Generalized Cross Validation (GCV)

One of the most commonly used methods of choosing an optimum knot point is the Generalized Cross Validation (GCV). Compared with other methods, such as Cross Validation (CV) and Unbiased Risk (UBR) or Generalized Maximum Likelihood (GML) methods, GCV has theoretically optimal asymptotic properties [3]. GCV method also has advantages that do not require knowledge of the population variance σ^2 and GCV invariance method of transformation [3].

GCV function is given by

$$GCV(\tilde{k}) = \frac{n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{[n^{-1} trace(\mathbf{I} - \boldsymbol{A}[\mathbf{K}])]^2}, \quad (10)$$

with GCV being a vector containing GCV values from knot points. Optimum knot points is obtained through optimization

$$\min_{k_1,k_2,\ldots,k_r}\{GCV(\tilde{k})\} = \min_{k_1,k_2,\ldots,k_r}\left\{\frac{n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{[n^{-1} trace(\mathbf{I} - \boldsymbol{A}[\mathbf{K}])]^2}\right\}, \quad (11)$$

with $\tilde{k} = (K_1, K_2, \ldots, K_r)$. $\boldsymbol{A}[\mathbf{K}]$ get from equation

$$\hat{\tilde{y}} = \boldsymbol{A}[\mathbf{K}]\tilde{y} \quad (12)$$

3.2 Pivotal Quantity

There are several methods for constructing confidence intervals for an uknown parameter (θ). In this research, using pivotal quantity method to constructing confidence intervals for unknown parameters. *Pivotal quantity* is a function of random sample of size n from a population and parameter (θ) whose probability distribution is independent of the parameter (θ).

4 RESULTS

Based on equation (7) and (8) we get estimator of parameter $\tilde{B}$ using Weighted Least Square (WLS).

Lemma 1.

If model of nonparametric spine truncated regression for longitudinal data on equation (7) with

$\tilde{\varepsilon} \sim N\left(\tilde{0}, \sigma^2 \boldsymbol{W}\right)$,

then the estimator parameter is

$$\hat{\tilde{B}} = \left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1} \mathbf{Z}[\mathbf{K}]\right)^{-1} \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1} \tilde{y}.$$

Proof

Based on equation (8) we obtain

$$\left(\tilde{y} - \mathbf{Z}[\mathbf{K}]\tilde{B}\right)^{\mathrm{T}} \mathbf{W}^{-1} \left(\tilde{y} - \mathbf{Z}[\mathbf{K}]\tilde{B}\right)$$

$$= \tilde{y}^{\mathrm{T}}\mathbf{W}^{-1}\tilde{y} - \tilde{B}^{\mathrm{T}}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\tilde{y} - \tilde{y}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\tilde{B} + \tilde{B}^{\mathrm{T}}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\tilde{B} \quad (13)$$

To find the value of $\tilde{B}$ that minimize the sum squares of the devations with Weigthed Least Square, we differentiate with respect to $\tilde{B}^{\mathrm{T}}$ and set the the results equal to 0:

$$\frac{\partial\left(\tilde{y}^{\mathrm{T}}\mathbf{W}^{-1}\tilde{y} - \tilde{B}^{\mathrm{T}}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\tilde{y} - \tilde{y}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\tilde{B} + \tilde{B}^{\mathrm{T}}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\tilde{B}\right)}{\partial\tilde{B}^{\mathrm{T}}}$$

$$= -\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\tilde{y} + \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\tilde{B} \quad (14)$$

Then set the results from (14) equal to 0

$$\Leftrightarrow -\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\tilde{y} + \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\hat{\tilde{B}} = 0$$

$$\Leftrightarrow \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\hat{\tilde{B}} = \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\tilde{y}$$

$$\hat{\tilde{B}} = \left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1} \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\tilde{y} \quad (15)$$

Then from equation (7) we get

$$\hat{\tilde{y}} = \mathbf{Z}[\mathbf{K}]\hat{\tilde{B}}$$

ISSN 2229-5518

$$\hat{\tilde{y}} = \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\tilde{y} \tag{16}$$

Equation (16) is equivalent with

$$\hat{\tilde{y}} = \mathbf{A}[\mathbf{K}]\,\tilde{y} \tag{17}$$

with

$$\mathbf{A}[\mathbf{K}] = \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1} \tag{18}$$

Next, if we assume

$\tilde{\varepsilon} \sim N\left(\tilde{0}, \sigma^2 \boldsymbol{W}\right)$ then $\tilde{y} \sim N\left(\mathbf{Z}[\mathbf{K}]\tilde{B}, \sigma^2 \boldsymbol{W}\right)$.

The following is given a Lemma about the distribution of $\hat{\tilde{B}}$.

Lemma 2.

If $\hat{\tilde{B}}$ is given by the equation (15) then

$$\hat{\tilde{B}} \sim N\left(\tilde{B}, \sigma^2\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\right)$$

Proof Since $\tilde{y}$ is normally distributed, equation (12) which is a linear combination of $\tilde{y}$ is also normally distributed with the expected value and its variance.

$$\begin{aligned} E\left(\hat{\tilde{B}}\right) &= E\left(\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\tilde{y}\right) \\ &= \left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}E\left(\tilde{y}\right) \\ &= \left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\tilde{B} \\ &= \tilde{B} \end{aligned} \tag{19}$$

$$\begin{aligned} \mathrm{Var}\left(\hat{\tilde{B}}\right) &= \mathrm{Var}\left(\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\tilde{y}\right) \\ &= \left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathrm{Var}\left(\tilde{y}\right) \\ &\quad \left(\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\right)^{\mathrm{T}} \\ &= \sigma^2\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1} \end{aligned} \tag{20}$$

Based on the results of equations (19) and (2), it is proven that $\hat{\tilde{B}} \sim N\left(\tilde{B}, \sigma^2\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\right)$.

Next we construct confidence intervals $(1-\alpha)100\%$ for estimator parameter $\tilde{B}_v$, $v = 1,2,\ldots,rqn$ using pivotal quantity method with unkown variance σ^2.

We will using Mean Square Error (MSE) to expected the variance of the estimator $\hat{\tilde{B}}_v$. Now using pivotal quantity as follows

$$\begin{aligned} T_v\left(z_1, z_2, \ldots, z_q, y\right) &= \frac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{\mathrm{Var}\left(\hat{\tilde{B}}_v\right)}} \\ &= \frac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{\sigma^2\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}} \\ &= \frac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{MSE\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}} \end{aligned} \tag{21}$$

$T_v\left(z_1, z_2, \ldots, z_q, y\right)$ is a pivotal quantity for parameters $\tilde{B}_v$.

With

$$\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}$$

is the vv-th diagonal element of the matrix

$$\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}$$

and MSE for unkown variance σ^2 is given by

$$MSE = \frac{(\tilde{y} - \hat{\tilde{y}})^T(\tilde{y} - \hat{\tilde{y}})}{nt - n(1 + q(r+1))}$$

$$= \frac{\left(\tilde{y} - \mathbf{Z}[\mathbf{K}]\hat{\tilde{B}}\right)^{\mathrm{T}}\left(\tilde{y} - \mathbf{Z}[\mathbf{K}]\hat{\tilde{B}}\right)}{nt - n(1 + q(\mathrm{r}+1))} \tag{22}$$

Then from equation (22) we get

$$\mathrm{MSE} = \frac{\tilde{y}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\right)\tilde{y}}{nt\text{ - }n(1 + q(r+1))} \tag{23}$$

The following is given a Lemma about the distribution of $T_v\left(z_1, z_2, \ldots, z_q, y\right)$.

Lemma 3.

If $T_v\left(z_1, z_2, \ldots, z_q, y\right)$ is given by the equation (21) then

$$T_v\left(z_1, z_2, \ldots, z_q, y\right) \sim \mathrm{t}_{(nt - n(1+q(r+1)))}$$

Proof

To facilitate the proof, we do alittle translation

$$T_v\left(z_1, z_2, \ldots, z_q, y\right) = \frac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{MSE\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}}$$

$$= \frac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{\dfrac{\tilde{y}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\right)\tilde{y}}{nt\text{-}n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}} \tag{24}$$

$$= \frac{\dfrac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}}}{\sqrt{\dfrac{\tilde{y}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\right)\tilde{y}}{nt\text{ - }n(1 + q(r+1))}}} \tag{25}$$

Equation (25) equivalent with

ISSN 2229-5518

$$T_v\left(z_1, z_2, \ldots, z_q, y\right) = \frac{\omega_1}{\sqrt{\dfrac{\omega_2}{nt - n(1+q(r+1))}}} \tag{26}$$

where

$$\omega_1 = \frac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}} \tag{27}$$

and

$$\omega_2 = \tilde{y}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1} \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\right)\tilde{y} \tag{28}$$

Next performed the following proofing steps:

i. The first step proves $\omega_1 \sim N(0,1)$.

The following is given a Lemma the distribution of ω_1.

Lemma 4.

If $\omega_1 = \dfrac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}}$**,** then $\omega_1 \sim N(0,1)$.

Proof

Since $\hat{\tilde{B}}$ is normally distributed, and statistic ω_1 is a linear combination of $\hat{\tilde{B}}$ then $\omega_1 \sim N\left(E(\omega_1), \mathrm{Var}(\omega_1)\right)$ where

$$\begin{aligned} E(\omega_1) &= E\left(\frac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}}\right) \\ &= \frac{1}{\sqrt{\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}} E\left(\hat{\tilde{B}}_v - \tilde{B}_v\right) \\ &= 0, \end{aligned} \tag{29}$$

and

$$\begin{aligned} Var(\omega_1) &= Var\left(\frac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}}\right) \\ &= \frac{1}{\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}} Var\left(\hat{\tilde{B}}_v - \tilde{B}_v\right) \\ &= \frac{1}{\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}} \left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv} + 0 \\ &= 1. \end{aligned} \tag{30}$$

Based on the results of equations (29) and (30) it is proven that $\omega_1 \sim N(0,1)$.

ii. The second step proves $\omega_2 \sim \chi^2_{(nt - n(1+q(r+1)))}$.

The following is given a Lemma about the distribution of ω_2.

Lemma 5.

If $\omega_2 = \tilde{y}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1} \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\right)\tilde{y}$,

then $\omega_2 \sim \chi^2_{(nt - n(1+q(r+1)))}$

Proof

$$\begin{aligned} \omega_2 &= \tilde{y}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1} \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\right)\tilde{y} \\ &= \tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y} \end{aligned}$$

To prove $\omega_2 \sim \chi^2_{(nt - n(1+q(r+1)))}$ it is necessary to show that the matrix

$$\mathbf{A} = \mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1} \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}$$

is symmetrical and idempotent.

.The matrix $\mathbf{A}$ is said to be symmetrical if $\mathbf{A}^{\mathrm{T}} = \mathbf{A}$. The following is a proof that $\mathbf{A}$ is symmetrical

$$\mathbf{A} = \mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1} \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}$$

Let

$$\mathbf{D} = \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1} \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}$$

Which is a symmetric and idempotent matrix, then $\mathbf{D}^{\mathrm{T}} = \mathbf{D}$**,** so that

$$\mathbf{A}^{\mathrm{T}} = (\mathbf{I} - \mathbf{D})^{\mathrm{T}} = (\mathbf{I} - \mathbf{D}) = \mathbf{A} \tag{31}$$

From the above description shows that the matrix $\mathbf{A}$ is symmetrical with the size $nt \times nt$. Further proved also that the matrix $\mathbf{A}$ is idempotent. Matrix $\mathbf{A}$ is said to be idempotent if $\mathbf{A}^2 = \mathbf{A}$. Here is a proof that $\mathbf{A}$ is idempotent.

$$\begin{aligned} \mathbf{A}^2 &= \left(\mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1} \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\right) \\ &\quad \left(\mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1} \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\right) \\ &= \mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1} \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1} \\ &= \mathbf{A} \end{aligned} \tag{32}$$

Based on the results of equations (31) and (32) it is proven that the matrix A is symmetric idempotent, then

$$\omega_2 \sim \chi^2_{\left(rank(\mathbf{A}),\, \tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}/2\sigma^2\right)}.$$

Then next step is to get the rank of matrix A and the value of $\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}$.

First step, since the matrix **A** is symmetric and idempotent then rank (**A**) = trace (**A**).

$$\begin{aligned} \text{trace}\,\mathbf{A} &= \text{trace}\left(\mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1} \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\right) \\ &= \text{trace}\left(\mathbf{I}_{nt}\right) - \text{trace}\left(\mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1} \mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\right) \\ &= \text{trace}\left(\mathbf{I}_{nt}\right) - \text{trace}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}} \mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\right) \end{aligned}$$

ISSN 2229-5518

$$= \text{trace}\left(\mathbf{I}_{nt}\right) - \text{trace}\left(\mathbf{I}_{n(1+q(r+1))}\right)$$
$$= nt - n(1+q(r+1)) \tag{33}$$

Second step is calculate the value of $\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}$ as follows

$$\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y} = \left(\mathbf{Z}[\mathbf{K}]\tilde{B}\right)^{\mathrm{T}}\mathbf{A}\left(\mathbf{Z}[\mathbf{K}]\tilde{B}\right)$$
$$= \left(\mathbf{Z}[\mathbf{K}]\tilde{B}\right)^{\mathrm{T}}\left(\mathbf{I}-\mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\right)\left(\mathbf{Z}[\mathbf{K}]\tilde{B}\right)$$
$$= \tilde{B}^{\mathrm{T}}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{Z}[\mathbf{K}]\tilde{B} - \tilde{B}^{\mathrm{T}}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{Z}[\mathbf{K}]\tilde{B}$$
$$= 0 \tag{34}$$

From the results of equations (33) and (34) it can be concluded that:

$$\omega_2 \sim \chi^2_{(nt-n(1+q(r+1))} \tag{35}$$

iii. The last step proves ω_1 and ω_2 are independent.

$$\omega_1 = \frac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}}$$
$$= \frac{\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\tilde{y} - \tilde{B}_v}{\sqrt{\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}}$$
$$= \frac{\mathbf{C}\tilde{y} - \tilde{B}_v}{\sqrt{\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}}$$

and

$$\omega_2 = \tilde{y}^{\mathrm{T}}\left(\mathbf{I}-\mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\right)\tilde{y}$$
$$= \tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}$$

then ω_1 and ω_2 are independent if and only if $\mathbf{CA} = 0$.

$$\mathbf{CA} = \left[\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\right] \times \left[\mathbf{I}-\mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\right]$$
$$= \left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1} + -\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}$$
$$= 0 \tag{36}$$

From the equation of (36) it is proven that $\mathbf{CA} = 0$ then ω_1 and ω_2 are independent.

Based on the description of points (i), (ii), and (iii) it can be deduced that statistics from (21)

$$T_v\left(z_1, z_2, \ldots, z_q, y\right) = \frac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{MSE\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}}$$
$$= \frac{\omega_1}{\sqrt{\dfrac{\omega_2}{nt - n(1+q(r+1))}}}$$
$$= \frac{N(0,1)}{\sqrt{\dfrac{\chi^2_{(nt-n(1+q(r+1))}}{nt - n(1+q(r+1))}}} \sim \mathrm{t}_{(nt-n(q(r+1)))}$$

So it is proven that

$$T_v\left(z_1, z_2, \ldots, z_q, y\right) \sim \mathrm{t}_{(nt-n(1+q(r+1)))} \tag{37}$$

is the pivotal quantity for the regression parameter $\tilde{B}_v$ when variance $\left(\sigma^2\right)$ is unknown. Further confidence intervals $(1-\alpha)$ can be obtained by solving the equation in probability

$$P\left(L_v \le T_v\left(z_1, z_2, \ldots, z_q, y\right) \le U_v\right) = 1-\alpha \tag{38}$$

With L_v and U_v being elements of real numbers, where $L_v < U_v$.

If equation (24)

$$T_v\left(z_1, z_2, \ldots, z_q, y\right) = \frac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{\dfrac{\tilde{y}^{\mathrm{T}}\left(\mathbf{I}-\mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\right)\tilde{y}}{nt-n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}}$$

Let $\mathbf{A} = \mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}$

Then

$$T_v\left(z_1, z_2, \ldots, z_q, y\right) = \frac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{\dfrac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt-n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}} \tag{39}$$

Then from substituted equation (39) to equation (38), it becomes

$$P\left(L_v \le \frac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{\dfrac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}} \le U_v\right) = 1-\alpha, \tag{39}$$

Equation (39) is equivalent to

$$\tilde{B}_v \le \hat{\tilde{B}}_v - L_v\sqrt{\frac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}} \tag{40}$$

and

$$\tilde{B}_v \ge \hat{\tilde{B}}_v - U_v\sqrt{\frac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}} \tag{41}$$

Based on equation (40) and (41) there is a confidence interval for parameters of nonparameteric regression spline truncated for longitudinal data when variance unknown is

$$P\left(\hat{\tilde{B}}_v - L_v\sqrt{\frac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)_{vv}^{-1}} \le \tilde{B}_v \le \hat{\tilde{B}}_v - U_v\sqrt{\frac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)_{vv}^{-1}}\right) = 1-\alpha \tag{42}$$

Next find the values of $L_v \in R$ and $U_v \in R$, so the length of the interval in equation (42) is the shortest. Let $\ell(L_v, U_v)$ be the length of the confidence interval above then:

$$\begin{aligned}\ell(L_v,U_v) &= \left(\hat{\tilde{B}}_v - L_v\sqrt{\frac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)_{vv}^{-1}}\right) + \\ &\quad -\left(\hat{\tilde{B}}_v - U_v\sqrt{\frac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)_{vv}^{-1}}\right) \\ &= \left((U_v - L_v)\sqrt{\frac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)_{vv}^{-1}}\right)\end{aligned} \tag{43}$$

As a result the shortest $(1-\alpha)$ confidence interval is obtained from completing conditional optimization:

$$\min_{L_v,U_v \in R}\left\{\ell(L_v,U_v)\right\} = \min_{L_v,U_v \in R}\left\{\sqrt{\frac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)_{vv}^{-1}}\right\} \tag{44}$$

with conditions

$$\int_{L_v}^{U_v} \psi(w)\,dw = 1-\alpha \tag{45}$$

or

$$\varphi(U_v) - \varphi(L_v) - (1-\alpha) = 0 \tag{46}$$

The function ψ is the probability distribution $\mathrm{t}_{(nt-n(1+q(r+1)))}$ and φ is the cumulative probability distribution $\mathrm{t}_{(nt-n(1+q(r+1)))}$.

The optimization of equations (44) and (45) or (46) using the Langrange Multiple method. Lagrange function is then formed as follows.

$$G(L_v,U_v,\lambda) = (U_v - L_v)\sqrt{\frac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)_{vv}^{-1}} + \lambda\left[\varphi(U_v) - \varphi(L_v) - (1-\alpha)\right] \tag{47}$$

where λ is the Lagrange constant. The next step is to do a partial derivative of functions $G(L_v, U_v, \lambda)$ against L_v, U_v, and λ. So obtained:

$$\frac{\partial G(L_v,U_v,\lambda)}{\partial L_v} = -\sqrt{\frac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)_{vv}^{-1}} + -\lambda\left(\varphi'(L_v)\right) = 0 \tag{48}$$

$$\frac{\partial G(L_v,U_v,\lambda)}{\partial U_v} = \sqrt{\frac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)_{vv}^{-1}} + \lambda\left(\varphi'(U_v)\right) = 0 \tag{49}$$

$$\frac{\partial G(L_v,U_v,\lambda)}{\partial \lambda} = \left[\varphi(U_v) - \varphi(L_v) - (1-\alpha)\right] = 0 \tag{50}$$

Based on equation (48) and (49) obtained

$$\psi(L_v) = \psi(U_v) \tag{51}$$

then the solution of equation (51) is

$L_v = U_v$ or $L_v = -U_v$.

But the suit solution of equation (51) is $L_v = -U_v$. So to obtain the shortest confidence interval must take the values L_v and U_v that suit the equation :

$$\int_{-\infty}^{L_v} \psi(w)\,dw = \int_{U_v}^{\infty} \psi(w)\,dw = \frac{\alpha}{2} \tag{52}$$

If the level of significance $(1-\alpha)$ is determined, then the values of L_v and U_v that conform can be seen in the student-*t* distribution table with $\mathrm{t}_{(nt-n(1+q(r+1)))}$. Thus the shortest confidence intervals for parameter of the nonparametric regression spline truncated model for longitudinal data is given by:

$$P\left(\hat{\tilde{B}}_v - U_v\sqrt{\frac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)_{vv}^{-1}} \le \tilde{B}_v \le \hat{\tilde{B}}_v + U_v\sqrt{\frac{\tilde{y}^{\mathrm{T}}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{\mathrm{T}}\,\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)_{vv}^{-1}}\right) = 1-\alpha \tag{53}$$

where U_v is derived from $\int_{U_v}^{\infty} \psi(w)\,dw = \frac{\alpha}{2}$.

Since the student-t distribution already exists, the shortest confidence intervals for parameter $\hat{\tilde{B}}_v$ can be written as:

$$P\left(\hat{\tilde{B}}_v - t_{\left(\frac{\alpha}{2}, nt-n(1+q(r+1))\right)}\sqrt{\frac{\tilde{y}^{T}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{T}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}} \le \tilde{B}_v \le \hat{\tilde{B}}_v + t_{\left(\frac{\alpha}{2}, nt-n(q(r+1))\right)}\sqrt{\frac{\tilde{y}^{T}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{T}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}\right) = 1-\alpha \tag{54}$$

with $\mathbf{A} = \mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{T}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{T}\mathbf{W}^{-1}$

5 Conclusion

In this research was constructed the confidence intervals using the concept of pivotal quantity and the shortest confidence interval. The shortest confidence interval for parameters of nonparametric spline truncated regression for longitudinal data using pivotal quantity when variance (σ^2) is unknown, the pivotal quantity is

$$T_v(z_1, z_2, \ldots, z_q, y) = \frac{\hat{\tilde{B}}_v - \tilde{B}_v}{\sqrt{\frac{\tilde{y}^{T}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{T}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}}$$

With $\mathbf{A} = \mathbf{I} - \mathbf{Z}[\mathbf{K}]\left(\mathbf{Z}[\mathbf{K}]^{T}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}\mathbf{Z}[\mathbf{K}]^{T}\mathbf{W}^{-1}$ and

Where

$$T_v(z_1, z_2, \ldots, z_q, y) \sim t_{nt-n(1+q(r+1)}$$

and the shortest confidence intervals for parameter $\tilde{B}_v$ as follows

$$P\left(\hat{\tilde{B}}_v - t_{\left(\frac{\alpha}{2}, nt-n(1+q(r+1))\right)}\sqrt{\frac{\tilde{y}^{T}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{T}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}} \le \tilde{B}_v \le \hat{\tilde{B}}_v + t_{\left(\frac{\alpha}{2}, nt-n(q(r+1))\right)}\sqrt{\frac{\tilde{y}^{T}\mathbf{A}\tilde{y}}{nt - n(1+q(r+1))}\left(\mathbf{Z}[\mathbf{K}]^{T}\mathbf{W}^{-1}\mathbf{Z}[\mathbf{K}]\right)^{-1}_{vv}}\right) = 1-\alpha$$

References

[1] L. Brieman and J. H. Friedman, "Estimting Optimal Transformation for Multiple Regression and Correlation," *Journal of the American Statistical Association,* vol. 80, no. 391, pp. 580-598, 1985.

[2] J. H. Friedman, "Multivariate Adaptive Regression Splines," *The Annals of Statistics,* vol. 19, no. 1, pp. 1-67, 1991.

[3] A. Buja, T. Hastie and R. Tibshirani, "Linear Smoothers and Additive Models," *The Annals of Statistics,* vol. 17, no. 2, pp. 453-555, 1989.

[4] C. J. Stone, "Additive Regression and Other Nonparametric Models," *The Annals of Statistics,* vol. 13, no. 2, pp. 689-705, 1985.

[5] L.-A. Chen, "Multivariate Regression Splines," *Computational Statistics and Data Analysis,* vol. 26, pp. 71-82, 1997.

[6] C. d. Visser, Q.P.Chu and J.A.Murder, "A New Approach To Linier Regression With Multivariate Splines," *Automatica,* vol. 45, pp. 2903-2909, 2009.

[7] I. N. Budiantara, "Spline Dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang," in *Pidato Pengukuhan Untuk Jabatan Guru Besar pada Jurusan Statistika*, Institut Teknologi Sepuluh Nopember (ITS) Surabaya, 2009.

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

[9] R. Diana, I. N. Budiantara, Purhadi and S. Darmesto, "Smoothing Spline in Semiparametric Additive Regression Model With Bayesian Approach," *Journal of Mathematics and Statistics,* vol. 9, no. 3, pp. 161-168, 2013.

[10] B. Lestari, I. N. Budiantara, S. Sunaryo and M. Mashuri, "Spline Smoothing for Multi-Response Nonparametric Regression Model in Case of Heteroscedasticity of Variance," *Journal of Mathematics and Statistics,* vol. 8, no. 3, pp. 377-384, 2012.

[11] T. C. Lee, "Smoothing Parameter Selection for Smoothing Splines: A Simulation Study," *Computational Statistics and Data Analysis,* vol. 42, pp. 139-148, 2003.

[12] G. Kauermann, "A note on smoothing parameter selection for penalized spline smoothing," *Journal of Statistical Planning and Inference,* vol. 127, pp. 53-69, 2005.

[13] W. Wibowo, C. Wene, I. N. Budiantara and E. O. Permatsari, "Multiresponse semiparametric regression for modelling the effect of regional socio-economic variables on the use of information technology," in *AIP Conference Proceedings 1825*, Makassar, 2017.

[14] Rismal, I. N. Budiantara and D. D. Prastyo, "Mixture model of spline truncated and kernel in multivariable nonparametric regression," in *AIP Conference Proceedings*, Kuala Lumpur, 2016.

[15] H. Wu and J.-T. Zhang, Nonparametric Regression Methods for Longitudinal Data Analysis, New Jersey: A John-Wiley and Sons Inc. Publication, 2006.

[16] P. C. Besse, H. Cardot and F. Ferraty, "Simultaneous non-parametric regrees," *Computational Statistics and Data Analysis,* vol. 24, pp. 255-270, 1997.

[17] C. M. Crainiceanu, "Nonparametric Regression Methods for Longitudinal Data Analysis. Mixed-effects Modelling Approaches," *Journal of the American Statistical Association,* vol. 102, p. 761, 20122.

[18] H. Jacqmin-Gadda, P. Joly, D. Commenges, C. Binquet and G. Chene, "Penalized likelihood approach to estimate a smooth mean curve on longitudinal data," *Statistics in Medicine,* vol. 21, pp. 2391-2409, 2002.

[19] P. Sahoo, Probability and Mathematical Statistics, Lousville, USA: University of Louisville, 2013.

[20] C. Gu and G. Wahba, "Smoothing Spline ANNOVA with Component-Wise Bayesian "Confidence Intervals"," *Journal of Computational and Graphical Statistics,* vol. 2, no. 1, pp. 97-117, 1993.