Journal of Theoretical and Applied Information Technology
31st March 2023. Vol.101. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MULTIRESPONSE REGRESSION SEMIPARAMETRIC TRUNCATED SPLINE (CASE STUDY: SKIN DEFECT OF FARMING COWS)

YUMNA AQILA KALTSUM, NI WAYAN SURYA WARDHANI, ADJI ACHMAD RINALDO FERNANDES, SUCI ASTUTIK, RAHMA FITRIANI

Departement of Statistics, Faculty of Mathematics and Natural Sciences, Brawijaya University, 65154

Indonesia

Corresponding author: yumnaqila0999@gmail.com

ABSTRACT

This study aims to model a semiparametric multiresponse regression analysis model using the Truncated Spline approach as a solution in its nonparametric component to obtain the best model in the case study to help prevent and treat skin diseases in cattle. The solution is done using Weighted Least Square (WLS) to be able to capture multiresponse modeling where there are indications between parts of the same cow's body that have correlated wounds. The novelty of this research is on a semiparametric truncated spline multiresponse model using a non-uniform order and number of knots for each predictor variables.

Keywords: *Multiresponse Analysis, Semiparametric Regression Analysis, Truncated Spline, Farming Cows, Defect Skin Disease*

1. INTRODUCTION

Regression analysis is a statistical method where the method can describe the relationship between the predictor variable and the response variable and can predict the value of the response variable based on the known value of the predictor variable. The relationship between these variables can be written into a mathematical model. There are several models of approaches to regression analysis, one of which is by looking at the shape of the relationship curve between the predictor variables and the response variables. This approach by looking at the shape of the curve divides the regression analysis into three approach models, if the shape of the curve in the regression analysis is known then the regression model approach is called a parametric regression model while in nonparametric regression the shape of the curve is assumed to be unknown [1] If there is more than one predictor variable and only some of them have a known curve shape, the approach used in regression analysis is semiparametric regression[2]

One of the nonparametric regression approaches is spline. Spline is used in nonparametric regression because it can follow the pattern of relationships between predictor variables and response variables and is very flexible [3]. Spline has several models including smoothing spline and truncated spline. The smoothing spline predicts the function based on the model accuracy criteria and smoothing parameters that regulate the size of the curve smoothness [2], while the truncated spline considers the existence of knot points in determining the most optimal points. Not only splines, another nonparametric regression approach is to use kernel. Kernel nonparametric regression can be used to estimate the conditional expected value of a random variable so that it can find a nonlinear relationship between a pair of response random variables and predictors and obtain the appropriate weights.

A relationship in regression analysis does not always only exist between predictor variables and response variables. Multi-response regression is a regression approach model when one response variable has a relationship with other response variables. The description of the relationship between responses can be seen through the covariance variance matrix [4].

Skin disease in cattle is one of the most common diseases. The itching reaction in cattle causes the cattle to use objects around them to scratch their bodies which then irritates some parts of the cattle's skin [5]. Irritation that arises causes wounds so that there is a defect in the skin of cattle and can attract flies which can aggravate the wounds on the skin of cattle. In addition to ectoparasite factors, differences in regional conditions such as differences in temperature and humidity can affect wounds on cows skin [6f].

2. LITERATURE REVIEW

2.1 Semiparametric Regression

In practice, not all predictor variables are linearly related or the form of their relationship to the response variable can be known[8]. Semiparametric regression is a combination between parametric regression and nonparametric regression [7]. In this study, the parametric regression section will be completed first and followed by the nonparametric section. If it is known that it $x_{11}, x_{12}, \ldots, x_{1p}$ is a parametric component that is thought to influence y linearly and it $x_{21}, x_{22}, \ldots, x_{2q}$ is a nonparametric component that is thought to influence y not linearly or the form of the relationship that influences y is not known, then the semiparametric regression model is obtained as follows:

$$y_i = \beta_0 + \beta_1 x_{11i} + \ldots + \beta_p x_{1pi} + f(x_{21i}, x_{22i}, \ldots, x_{2qi}) + \varepsilon_i \quad (1)$$

Where:

y_i : Response variable at observation i

β_0 : intercept

β_p : Slope

x_{1pi} : The p-the parametric predictor variable on the i-th observation

x_{2q} : The q-the nonparametric predictor variable on the i-th observation

ε_i : distributed parametric model error $\varepsilon \sim N(0, \sigma_\varepsilon^2)$

2.2 Truncated Spline Regression

Spline is a segmented and continuous truncated polynomial slice. The advantage of truncated spline polynomial regression is that it can follows the regression curve[9]. This can happen because the spline has a common fusion point that shows a change in the pattern of data behavior called knot points. The knot point allows adjustment to local characteristics so that the spline has high flexibility.

Nonparametric Regression with a truncated spline approach has the form of a function where in the function there is a separator knot

The general form of a function estimator for one predictor variable:

$$f(x_i) = \alpha_0 + \alpha_1 x_i + \ldots + \alpha_r x_i^r + \alpha_{rk}(x_i - ks_1)_+^r + \ldots + \alpha_{rk}(x_i - k_k)_+^r \quad (2)$$

dengan,

$$(x_i - k_k)_+ = \begin{cases} (x_i - k_k) & ; \; x_i \geq k_k \\ 0 & ; \; x_i < k_k \end{cases} \quad (3)$$

The general form for more than one predictor variable:

$$\begin{aligned} f(x_{1i}, x_{2i}, \ldots, x_{pi}) &= f(x_{1i}) + f(x_{2i}) + \ldots + f(x_{pi}) \\ &= \alpha_{01} + \alpha_{11}x_{1i} + \ldots + \alpha_{r1}x_{1i}^r + \alpha_{r11} \\ &\quad (x_{1i} - k_{11})_+^r + \ldots + \alpha_{rk1}(x_{1i} - k_{k1})_+^r \\ &\quad + \alpha_{02} + \alpha_{12}x_{2i} + \ldots + \alpha_{r2}x_{2i}^r + \alpha_{r12} \\ &\quad (x_{2i} - k_{12})_+^r + \ldots + \alpha_{rk2}(x_{2i} - k_{k2})_+^r \\ &\quad + \ldots + \alpha_{0p} + \alpha_{1p}x_{pi} + \ldots + \alpha_{rp}x_{pi}^r \\ &\quad + \alpha_{r1p}(x_{p2i} - k_{1p})_+^r + \ldots + \alpha_{rkp}(x_{p2i} - k_{kp})_+^r \end{aligned} \quad (4)$$

Equation (4) can be written as:

$$\tilde{f}(x) = \mathbf{X}[K_1, K_2, \ldots, K_k]\tilde{\alpha} \quad (5)$$

2.3 Multiresponse Analysis

In regression modeling, a predictor variable does not always only have a relationship with one response variable. Multiresponse regression is an approach model when a response variable has a relationship with more than one response variable. The description of the relationship between responses can be seen through the variance matrix [4].

$$\hat{\rho}_{ij} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)(Y_j - \bar{Y}_j)}{\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2 \sum_{i=1}^{n}(Y_j - \bar{Y}_j)^2}} \quad (6)$$

To accommodate the correlation between the response variables, weighting is used using a variance matrix measuring M N × M N on the matrix and submatrix.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 & \rho_{12}\sigma_1\sigma_2 & 0 & \cdots & 0 & \cdots & \rho_{1m}\sigma_1\sigma_m & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \cdots & 0 & 0 & \rho_{12}\sigma_1\sigma_2 & \cdots & 0 & \cdots & 0 & \rho_{1m}\sigma_1\sigma_m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_1^2 & 0 & 0 & \cdots & \rho_{12}\sigma_1\sigma_2 & \cdots & 0 & 0 & \cdots & \rho_{1m}\sigma_1\sigma_m \\ \rho_{12}\sigma_1\sigma_2 & 0 & \cdots & 0 & \sigma_2^2 & 0 & \cdots & 0 & \cdots & \rho_{2m}\sigma_2\sigma_m & 0 & \cdots & 0 \\ 0 & \rho_{12}\sigma_1\sigma_2 & \cdots & 0 & 0 & \sigma_2^2 & \cdots & 0 & \cdots & 0 & \rho_{2m}\sigma_2\sigma_m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_{12}\sigma_1\sigma_2 & 0 & 0 & \cdots & \sigma_2^2 & \cdots & 0 & 0 & \cdots & \rho_{2m}\sigma_2\sigma_m \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \rho_{1m}\sigma_1\sigma_m & 0 & \cdots & 0 & \rho_{2m}\sigma_2\sigma_m & 0 & \cdots & 0 & \cdots & \sigma_m^2 & 0 & \cdots & 0 \\ 0 & \rho_{1m}\sigma_1\sigma_m & \cdots & 0 & 0 & \rho_{2m}\sigma_2\sigma_m & \cdots & 0 & \cdots & 0 & \sigma_m^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_{1m}\sigma_1\sigma_m & 0 & 0 & \cdots & \rho_{2m}\sigma_2\sigma_m & \cdots & 0 & 0 & \cdots & \sigma_m^2 \end{bmatrix}_{MN \times MN} \quad (7)$$

2.4 Multiresponse Semiparametric Truncated Spline

When an observation has more than one response variable and between responses, variables have a relationship with each other and the observation has more than two predictor variables with some of the predictor variable relationships with the response variables known and partly unknown[4], then modeling with Multiresponse Semiparametric Truncated Spline is used. as follows:

$$\underset{\sim}{y} = \mathbf{X}_1\underset{\sim}{\beta} + \mathbf{X}_2[K]\underset{\sim}{\alpha} + \underset{\sim}{\varepsilon} \qquad (8)$$

Where,

$$\underset{\sim}{y} = \left(\underset{\sim}{y}_1{}^T, \underset{\sim}{y}_2{}^T, \ldots, \underset{\sim}{y}_m{}^T\right)^T$$

$$\underset{\sim}{y}_1 = (y_{11}, \ldots, y_{1n})^T$$

$$\underset{\sim}{y}_2 = (y_{21}, \ldots, y_{2n})^T$$

$$\underset{\sim}{y}_m = (y_{31}, \ldots, y_{mn})^T$$

$$\underset{\sim}{\beta} = (\underset{\sim}{\beta}_1{}^T, \underset{\sim}{\beta}_2{}^T, \ldots, \underset{\sim}{\beta}_m{}^T)^T$$

$$\underset{\sim}{\beta}_i^T = (\beta_{i0}, \ldots, \beta_{im})$$

$$\underset{\sim}{\alpha} = (\underset{\sim}{\alpha}_1{}^T, \underset{\sim}{\alpha}_2{}^T, \ldots, \underset{\sim}{\alpha}_m{}^T)^T$$

$$\underset{\sim}{\alpha}_i^T = (\alpha_{i0}, \ldots, \alpha_{im})$$

$$\underset{\sim}{\varepsilon} = \left(\varepsilon_{11}, \ldots, \varepsilon_{1n}, \varepsilon_{21}, \ldots, \varepsilon_{2n}, \ldots, \varepsilon_{m1}, \ldots, \varepsilon_{mn}\right)^T$$

m = many response variables
n = many observations
and

$$\mathbf{X}_1 = \begin{bmatrix} & x_{1111} & x_{1211} & \cdots & x_{1q11} \\ & x_{1112} & x_{1212} & \cdots & x_{1q12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ & x_{111n} & x_{121n} & \cdots & x_{1q1n} \\ & x_{1121} & x_{1221} & \cdots & x_{1q21} \\ & x_{1122} & x_{1222} & \cdots & x_{1q22} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ & x_{112n} & x_{122n} & \cdots & x_{1q2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ & x_{11m1} & x_{12m1} & \cdots & x_{1qm1} \\ & x_{11m2} & x_{12m2} & \cdots & x_{1qm2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ & x_{11mn} & x_{12mn} & \cdots & x_{1qmn} \end{bmatrix},$$

$$\mathbf{X}_2[K] = \begin{bmatrix} \mathbf{X}_{21}[K] & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{22}[K] & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_{2m}[K] \end{bmatrix}_{QM\times(1+r+k)M}$$

$$\mathbf{X}_{21}[K] = \begin{bmatrix} 1 & x_{11} & \cdots & x_{11}^r\,(x_{11}-k_{111})_+^r & \cdots & (x_{11}-k_{k11})_+^r \\ 1 & x_{221} & \cdots & x_{221}^r\,(x_{221}-k_{121})_+^r & \cdots & (x_{221}-k_{k21})_+^r \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{2q1} & \cdots & x_{2q1}^r\,(x_{2q1}-k_{1q1})_+^r & \cdots & (x_{2q1}-k_{kq1})_+^r \end{bmatrix}_{Q\times(1+r+k)}$$

$$\mathbf{X}_{22}[K] = \begin{bmatrix} 1 & x_{212} & \cdots & x_{212}^r\,(x_{212}-k_{112})_+^r & \cdots & (x_{212}-k_{k12})_+^r \\ 1 & x_{222} & \cdots & x_{222}^r\,(x_{222}-k_{122})_+^r & \cdots & (x_{222}-k_{k22})_+^r \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{2q2} & \cdots & x_{2q2}^r\,(x_{2q2}-k_{1q2})_+^r & \cdots & (x_{2q2}-k_{kq2})_+^r \end{bmatrix}_{Q\times(1+r+k)}$$

$$\mathbf{X}_{2m}[\mathrm{K}] = \begin{bmatrix} 1 & x_{21m} & \cdots & x_{21m}^r\,(x_{21m}-k_{1pm})_+^r & \cdots & (x_{21m}-k_{kpm})_+^r \\ 1 & x_{22m} & \cdots & x_{22m}^r\,(x_{22m}-k_{1pm})_+^r & \cdots & (x_{22m}-k_{kpm})_+^r \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{2qm} & \cdots & x_{2qm}^r\,(x_{2qm}-k_{1qm})_+^r & \cdots & (x_{2qm}-k_{kqm})_+^r \end{bmatrix}_{Q\times(1+r+k)}$$

To obtain parameter and function estimators from semiparametric regression by looking at the weights, the $\mathbf{\Sigma}^{-1}$Weighted Least Square (WLS) optimization is carried out by solving the following equation,parametric component completion :

$$\min\left\{\underset{\sim}{\varepsilon}_1^{\mathrm{T}}\mathbf{\Sigma}^{-1}\underset{\sim}{\varepsilon}_1\right\} = \min\left\{(\underset{\sim}{y} - \mathbf{X}_1\underset{\sim}{\beta})^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\underset{\sim}{y} - \mathbf{X}_1\underset{\sim}{\beta})\right\} \qquad (9)$$

To estimate the value $\underset{\sim}{\beta}$ that minimizes JKG ($\underset{\sim}{\varepsilon}^T\underset{\sim}{\varepsilon}$), then a partial derivative is performed $\underset{\sim}{\beta}$.

$$\frac{\partial(\underset{\sim}{\varepsilon}_1{}^T\mathbf{\Sigma}^{-1}\underset{\sim}{\varepsilon}_1)}{\partial\underset{\sim}{\beta}} = 0 \qquad (10)$$

So that the parameter estimator value is obtained $\hat{\underset{\sim}{\beta}}$

$$\hat{\underset{\sim}{\beta}} = (\mathbf{X}_1{}^T\mathbf{\Sigma}^{-1}\mathbf{X}_1)^{-1}(\mathbf{X}_1{}^T\mathbf{\Sigma}^{-1}\underset{\sim}{y}) \qquad (11)$$

solve the non-parametric components by estimation using the Weight Least Square which is a derivative of the nonparametric Ordinary Least Square

$$\min\left\{\underset{\sim}{\varepsilon}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\underset{\sim}{\varepsilon}\right\} = \min\left\{(\underset{\sim}{y}_2 - \mathbf{X}_2[\mathrm{K}]\underset{\sim}{\alpha})^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\underset{\sim}{y}_2 - \mathbf{X}_2[\mathrm{K}]\underset{\sim}{\alpha})\right\} \qquad (12)$$

To estimate the value $\underset{\sim}{\alpha}$ can be estimated by minimizing the Number of Squares of Errors ($\underset{\sim}{\varepsilon}^T\underset{\sim}{\varepsilon}$), then a partial derivative of. is performed $\underset{\sim}{\alpha}$.

$$\frac{\partial(\underset{\sim}{\varepsilon}^T\mathbf{\Sigma}^{-1}\underset{\sim}{\varepsilon})}{\partial\underset{\sim}{\alpha}} = 0 \qquad (13)$$

So for the coefficient $\hat{\underset{\sim}{\alpha}}$ estimator of the function:

$$\hat{\underset{\sim}{\alpha}} = (\mathbf{X}_2[K]^T\mathbf{\Sigma}^{-1}\mathbf{X}_2[K])^{-1}(\mathbf{X}_2[K]^T\mathbf{\Sigma}^{-1}\underset{\sim}{y}_2) \qquad (14)$$

Then the estimated function is obtained:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\hat{f}(x_2) = \mathbf{X}_2[\mathbf{K}]\tilde{\alpha} \quad (15)$$
$$\hat{f}(x_2) = \mathbf{X}_2[\mathbf{K}](\mathbf{X}_2[K]^T\,\mathbf{\Sigma}^{-1}\mathbf{X}_2[K])^{-1}\mathbf{X}_2[K]^T\,\mathbf{\Sigma}^{-1}\tilde{y}$$

This study uses the order of linear and quadratic polynomials using a maximum of 2 knots so that 4 equation models are obtained:

1. Linear 1 Knot

$$\tilde{y}=\tilde{\beta}_0+\tilde{\beta}_1\tilde{x}_{11}+\dots+\tilde{\beta}_p\tilde{x}_{1p}+\tilde{\alpha}_{01}+\dots+\tilde{\alpha}_{0q}+\tilde{\alpha}_{11}\tilde{x}_{21}+\dots+\tilde{\alpha}_{1q}\tilde{x}_{2q}+\tilde{\alpha}_{111}(\tilde{x}_{21}-k_{11})_+ +\dots+\tilde{\alpha}_{11q}(\tilde{x}_{2q}-k_{1q})_+ +\varepsilon_i \quad (16)$$

2. Linear 2 Knot

$$\tilde{y}=\tilde{\beta}_0+\tilde{\beta}_1\tilde{x}_{11}+\dots+\tilde{\beta}_p\tilde{x}_{1p}+\tilde{\alpha}_{01}+\dots+\tilde{\alpha}_{0q}+\tilde{\alpha}_{11}\tilde{x}_{21}+\dots+\tilde{\alpha}_{1q}\tilde{x}_{2q}+\tilde{\alpha}_{111}(\tilde{x}_{21}-k_{11})_+ +\tilde{\alpha}_{121}(\tilde{x}_{21}-k_{21})_+ +\dots+\tilde{\alpha}_{11q}(\tilde{x}_{2q}-k_{1q})_+ +\tilde{\alpha}_{12q}(\tilde{x}_{2q}-k_{2q})_+ +\varepsilon \quad (17)$$

3. Quadratic 1 Knot

$$\tilde{y}=\tilde{\beta}_0+\tilde{\beta}_1\tilde{x}_{11}+\dots+\tilde{\beta}_p\tilde{x}_{1p}+\tilde{\alpha}_{01}+\dots+\tilde{\alpha}_{0q}+\tilde{\alpha}_{11}\tilde{x}_{21}+\dots+\tilde{\alpha}_{1q}\tilde{x}_{2q}+\tilde{\alpha}_{21}(\tilde{x}_{21})^2+\dots+\tilde{\alpha}_{2q}(\tilde{x}_{2q})^2+\tilde{\alpha}_{211}(\tilde{x}_{21}-k_{11})_+^{\,2}+\dots+\tilde{\alpha}_{21q}(\tilde{x}_{2q}-k_{1q})_+^{\,2}+\tilde{\varepsilon} \quad (18)$$

4. Quadratic 2 Knot

$$\tilde{y}=\tilde{\beta}_0+\tilde{\beta}_1\tilde{x}_{11}+\dots+\tilde{\beta}_p\tilde{x}_{1p}+\tilde{\alpha}_{01}+\dots+\tilde{\alpha}_{0q}+\tilde{\alpha}_{11}\tilde{x}_{21}+\dots+\tilde{\alpha}_{1q}\tilde{x}_{2q}+\tilde{\alpha}_{21}(\tilde{x}_{21})^2+\dots+\tilde{\alpha}_{2q}(\tilde{x}_{2q})^2+\tilde{\alpha}_{211}(\tilde{x}_{21}-k_{11})_+^{\,2}+\tilde{\alpha}_{221}(\tilde{x}_{21}-k_{21})_+^{\,2}+\dots+\tilde{\alpha}_{21q}(\tilde{x}_{2q}-k_{1q})_+^{\,2}+\tilde{\alpha}_{22q}(\tilde{x}_{2q}-k_{2q})_+^{\,2}+\varepsilon \quad (19)$$

3. RESEARCH METHODS

3.1 Multiresponse Semiparametric Truncated Spline

The data used in this study is secondary data taken from the Annual Report on the Prevention and Management of Pests and Diseases in Cattle Livestock, Department of Agriculture and Livestock as many as 90 cows as samples. The predictor variables in the data above are Chloropidae (X1), Muscidae (X2), Humidity (X3), and Temperature (X4) as well as the response variable which is the area of the skin defect (cm ^{2}) in each part of the cow's body where the part is the Cheek (Y1) and Belly (Y2).

3.2 Multiresponse Semiparametric Truncated Spline

The variables used are composed of four predictor variables and two response variables. Predictor variables include Chloropidae (X1), Muscidae (X2), Humidity (X3), and Temperature (X4). The research model diagram can be seen in ... The response variable is the area of the skin defect (cm ^{2}) on each part of the cow's body skin where the parts are the Cheeks (Y1) and Belly (Y2). The research model diagram can be seen in Figure 1.

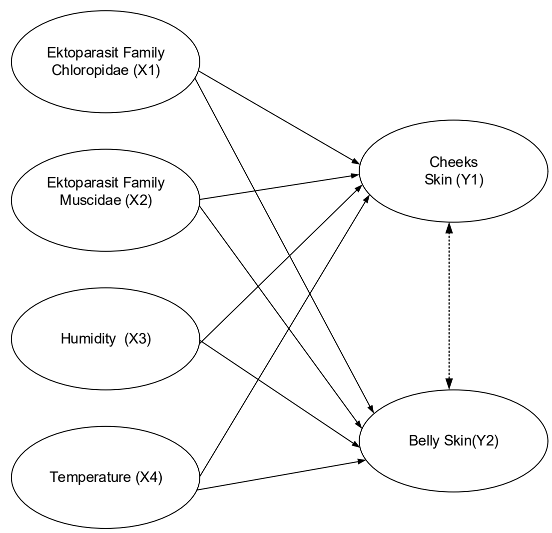

Figure 1 Diagram Model

3.3 Research Procedures

Steps used in this study:

1. Making Model Diagram research.
2. Testing the Linearity Assumption with Ramsey RESET
3. Getting Varian Covariance Matrix between Response Variables
4. Get parameter estimates for Parametric components and function estimates for nonparametric components
5. Choose the best Model by looking at the largest Adjusted R Square
6. Interpret the results of the best obtained model.

4. RESULT AND DISCUSSION

4.1 Linearity Test

In statistical modeling, it is necessary to have information about the relationship between the predictor variable and the response variable so that it can help determine the approach used between

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

parametric and nonparametric. The following are the results of the Linearity Test using the Ramsey RESET

Table 1: Center Table Captions Above The Tables.

Variable	P Value	Relation
X_1 to Y_1	0.585	Linear
X_2 to Y_1	0.002	Non-Linear
X_3 to Y_1	0.026	Non-Linear
X_4 to Y_1	0.082	Non-Linear
X_1 to Y_2	0.000	Non-Linear
X_2 to Y_2	0.033	Non-Linear
X_3 to Y_2	0.057	Non-Linear
X_4 to Y_2	0.092	Non-Linear

Based on the results of the linearity test above, it is found that pairs of variables are linear or non-linear so for linear pairs, a parametric approach will be used and a non-linear approach will be used.

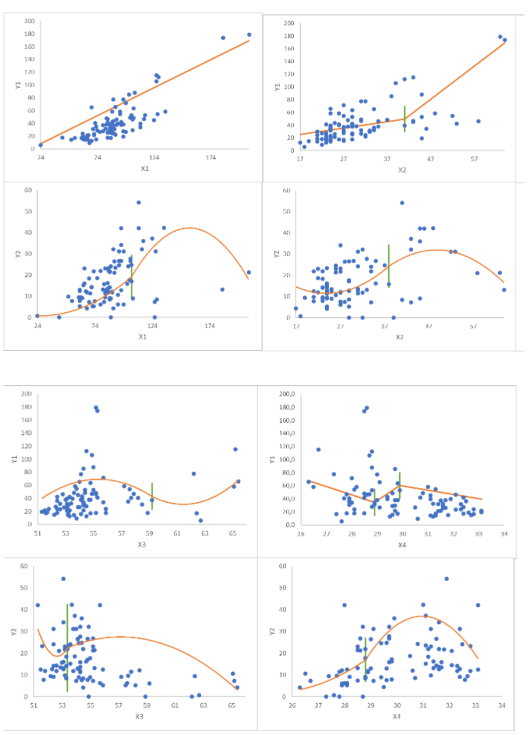

Figure 2 Scatter Plot Model

In this study, the selection of knots for the truncated spline was done intuitively based on the results of the linearity test and the shape of the curve obtained from the distribution of research data. The model for each response is obtained as follows:

The semiparametric multiresponse regression analysis model uses a truncated spline as an approach to solving it by limiting modeling to the second order (quadratic) with a maximum number of knots of 2 knots. The results of the linearity test and the selection of the best knots on the relationship between the predictor variables and the response variables are obtained by the model:

$$\hat{f}_1(x_1,x_2,x_3,x_4) = 0,003 + 0,82x_1 - 0,057x_2 + 1,22(x_2 - 40)_+ + 0,183x_3 - 0,003x_3^2 + 0,352(x_3 - 60,7)_+^2 - 1,613x_4 - 1,892(x_4 - 29)_+ - 1,51(x_4 - 30)_+$$

$$\hat{f}_2(x_1,x_2,x_3,x_4) = 0,001 + 0,05x_1 + 0,001x_1^2 - 0,002(x_1 - 106)_+^2 + 0,023x_2 + 0,006x_2^2 - 0,032(x_2 - 38)^2 + 0,04x_3 - 0,012x_3^2 + 0,067(x_3 - 53,4)_+^2 + 0,031x_4 + 0,04x_4^2 - 0,558(x_4 - 29)_+^2$$

In truncated spline modeling, the model is divided into several areas bounded by knots, these areas are called regimes. When the model lies in a regime either on the right or on the left it will produce a different model.

Based on the selected predictor variable knot, the smallest knot value is taken as the lower limit for the model in regime 1. The following equation is obtained when all predictor variables are located in regime 1 or to the left of the knots. So in the secondary data for x_1 <106, x_2<38, x_3<53.4, and x_4<29 we get:

$$\hat{f}_1(x_1,x_2,x_3,x_4) = 0,003 + 0,82x_1 - 0,057x_2 + 1,22(0) + 0,183x_3 - 0,003x_3^2 + 0,352(0)^2 - 1,613x_4 - 1,892(0) - 1,51(0)$$

$$\hat{f}_2(x_1,x_2,x_3,x_4) = 0,001 + 0,05x_1 + 0,001x_1^2 - 0,002(0)^2 + 0,023x_2 + 0,006x_2^2 - 0,032(0)^2 + 0,04x_3 - 0,012x_3^2 + 0,067(0)^2 + 0,031x_4 + 0,04x_4^2 - 0,558(0)^2$$

simplified to:

$$\hat{f}_1(x_1,x_2,x_3,x_4) = 0,003 + 0,82x_1 - 0,057x_2 + 0,183x_3 - 0,003x_3^2 - 1,613x_4$$

$$\hat{f}_2(x_1,x_2,x_3,x_4) = 0,001 + 0,05x_1 + 0,001x_1^2 + 0,023x_2 + 0,006x_2^2 + 0,04x_3 - 0,012x_3^2 + 0,031x_4 + 0,04x_4^2$$

The interpretation of the model is: when fewer than 106 pests from the Chloropidae Family (x_1) arrive, each arrival of one pest will increase the defect area by 0.82 cm^2 on the cheek (y_1) and expand 0.05 cm^2 on the belly (y_2). In the arrival of pests from the Muscidae Family (x_2) when there are less than 32 individuals, each arrival of one tail will expand the defect by 0.023 cm^2 on the belly (y_2) and reduce the defect by 0.057 cm^2 on the cheek (y_1). When the humidity (x_3) of the cage is below 53.4% and every 1% increase in humidity (x_3) of the cage can increase the defect area on the cheek (y_1) by 0.18 cm^2. At a temperature (x_4) below 29°C, every 1°C increase in temperature (x_4) can also reduce the defect on the cheek (y_1) by 1.613 cm^2 and can expand the defect on the belly (y_2) of 0.031 cm^2.

For conditions when all predictor variables are in the final regime, namely Regime 2 for 1 knot and Regime 3 for 2 knots. So based on the model from the secondary data for x_1>106, x_2>40, x_3>60.7, and x_4>30. are obtained.

Based on the selected predictor variable knot, the largest knot value is taken as the lower limit on the model in regime 2. For conditions when all predictor variables are in the final regime, namely Regime 2 for 1 knot and Regime 3 for 2 knots. Then based on the model from the secondary data for x_1>106, x_2> 40, x_3> 60.7, and x_4> 30.5 we get:

$$\hat{f}_1(x_1,x_2,x_3,x_4) = 1293.44 + 0,82x_1 - 1,163x_2 - 42,5498x_3 + 0,349x_3^2 - 3,123x_4$$

$$\hat{f}_3(x_1,x_2,x_3,x_4) = -344,975 - 0,374x_1 + 0,003x_1^2 + 2,455x_2 - 0,026x_2^2 - 7.115x_3 + 0,055x_3^2 + 33,845x_4 - 0,518x_4^2$$

When there are more than 106 pests from the Chloropidae Family (x_1), each arrival of one tail will increase the area of the defect by 0.82 cm^2 on the cheeks (y_1) and reduce 0.371 cm^2 on the belly (y_2). The arrival of pests from the Muscidae Family (x_2) when more than 40 tails will expand the defect by 2.429 cm^2 on the belly (y_2) for each arrival of one tail and reduce by 1.163 cm^2 on the cheeks (y_1). When the humidity (x_3) of the cage is above 60.7%, and every 1% increase in humidity (x_3) of the cage can reduce the defect area on the cheek (y_1) by 42.2 cm^2 and belly (y_2) by 7.05 cm^2. Temperature (x_4) above 30.5°C for every 1°C increase in temperature (x_4) can reduce the defect on the cheek (y_1) by 3.12 cm^2 and can expand the defect on the belly(y_2) of 33.2 cm^2.

Figure 2 shows the distribution of data for each form of relationship between predictor variables and response variables for the best model obtained. Defects on the skin of the cheeks (Y_1), each time a pest of the Chloropidae Family (X_1) arrives the defects on the cheeks (Y_1) become wider. Arrival of Muscidae (X_2) when there were less than 41 tails had a smaller effect than when there were more than 41 individuals. Humidity (X_3) maintained in the range of 51%-53% and 61%-63% will reduce defects on the cheeks. Temperature (X_4) which is in conditions of 28-29°C and 33-34°C can reduce defects on the cheeks.

Abdominal skin defects (Y_2) can decrease if fewer than 60 Chloropidae pests (X_1) arrive. Pests of the Muscidae family are less than 27 individuals. The humidity level (X_3) is in the range of 50%-52% (k). The constant temperature (X_4) at 24-25°C can also reduce the occurrence of defects in the belly skin (Y_2) of the cow.

The results of the function estimator for the semiparametric truncated spline model show that when fewer than 81 pests of the Chloropidae family (X_1) arrive, it causes an increase in the area of the defect on the cheek skin (Y_1) and belly skin (Y_2) for each arrival of a unit of Chloropidae pests (X_1). The knot point in the model shows changes in the influence characteristics of each predictor variable on the response variable.

Apart from pests belonging to the Chloropidae family (X_1), nuisance pests belonging to the Muscidae family (X_2) also affect defects in cowhide. Based on the results of the analysis that has been carried out, the effect of Muscidae (X_2) has a different behavior from Chloropidae (X_1). Muscidae did not attack much on the cheeks (Y_1) and belly (Y_2).

The influence of humidity (X_3) affects the damage to cowhide. Based on the results of the analysis of decreasing humidity (X_3) in cowsheds can reduce the risk of defects in cowhide, humidity (X_3) remains within 50-52%.

Temperature (X_4) in the range of 27-28℃ can suppress the expansion of defects in cowhide. An increase in temperature (X_4) in the cowshed can expand the defects in the skin. On the cheeks (Y_1) and belly (Y_2), changes in temperature (X_4) at 33-32℃ can also reduce the extent of skin defects.

The truncated spline semiparametric regression model used in this study has a goodness of fit of R^2_{adj} of 0.92 where this value indicates the best model that can describe the real situation by 92%.

5. CONCLUSIONS

Based on the results of the analysis and discussion, it can be concluded that:

1. In the semiparametric truncated spline multiresponse model, produces a different form of function estimator for each predictor variable so that this model produces an Adjusted R^2 value of 85,4%
2. The presence of knots in the estimation of the multiresponse truncated spline model shows that there is a change in the nature or behavior of the data so that the shape of the relationship pattern can change according to the shape of the actual nature.
3. Modeling using multi-response semiparametric truncated spline can help the government and cattle breeders prevent and minimize damage to cowhide.

Suggestions:

1. This study only estimates semiparametric truncated spline functions with linear and quadratic orders. For further research, hypothesis and significance tests can be carried out.
2. This study only uses two response variables. For further research can use more than two response variables.
3. This study uses the intuitiveness of the plot to select the best knots. Future research can compare the model with the selection of knots intuitively and mathematically.

6. ACKNOWLEDGEMENT

Thank you to all parties for their support and input in the preparation of this research. All the support and input given is very useful for the perfection of this research.

REFERENCES:

[1] Budiantara, IN (2006). Nonparametric regression in statistics. Mathematics National Seminar.

[2]Budiantara, IN (2005). TRUNCATED POLYNOMIAL SPLINE FAMILY MODEL IN SEMIPARAMETRIC REGRESSION. 8.

[3] Eubank, RL, Huang, C., Maldonado, YM, Wang, N., Wang, S., & Buchanan, RJ (2004). Smoothing spline estimation in varying-coefficient models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66 (3), 653–667. https://doi.org/10.1111/j.1467-9868.2004.B5595.x

[4] Fernandes, A. A. R., Budiantara, I. N. I., Otok, B. W., & Suhartono. (2014). Reproducing Kernel Hilbert space for penalized regression multi-predictors: Case in longitudinal data. International Journal of Mathematical Analysis, 8(40), 1951-1961.

[5] Djenaan, F., Assa, G., Poli, Z., & Lomboan, A. (2018). Types and populations of flies in cattle in Tolok Village, Tompaso District, Minahasa Regency. ZOOTEC , 39 (1), 51–56.

[6] Septianingsih, NLPD, Dwinata, IM, & Oka, IBM (2019). Hippobosca Sp. Fly Infestation Survey. On Bali Cattle in Badung Regency.

[7] Ruppert, D., Wand, MP, & Carroll, RJ (2003). Semiparametric regression . Cambridge university press.

[8] Fernandes, A. A. R. (2017). Investigation of instrument validity: Investigate the consistency between criterion and unidimensional in instrument validity (case study in management research). International journal of law and management.

[9] Murbarani, N., Swastika, Y., Dwi, A., Aris, B., & Chamidah, N. (2019). MODELING OF THE PERCENTAGE OF AIDS SUFFERERS IN EAST JAVA PROVINCE WITH NONPARAMETRIC REGRESSION APPROACH BASED ON SPLINE TRUNCATED ESTIMATOR. *Indonesian Journal of Statistics and Its Applications*, *3*(2), 139-147.