Multivariable Semiparametric Regression Used Priestley-Chao Estimators

Makkulau[1], Andi Tenri Ampa[2], Baharuddin[3], Mukhsar[4], Agusrawati[5], Andi Tenri Pannangngareng Makkulau[6], I Nyoman Sudiasa[7]

[1,2,3,4,5,6] Statistics Department, Universitas Halu Oleo, Indonesia
[7] Civil Engineering Department, Institut Teknologi Nasional Malang,Indonesia

makkulaufmipa@uho.ac.id

Abstract. Semiparametric regression combines the goodness of parametric regression estimators and Kernel regression. In this research, a new method of semiparametric regression model was developed which contains parametric and non-parametric components, the second being multivariable, where the data contains outlier data. Here we propose a multivariable Kernel method approach for data that does not follow a certain pattern and has outliers. The kernel function used is a multivariable Gaussian Kernel function with a Priestley-Chao estimators approach. The goodness of the Kernel estimator depends on the value of the bandwidth parameter, to get the optimal bandwidth parameter using the Generalized Cross Validation (GCV) method. The results of the theoretical study obtained from the Mixed Kernel Regression Curve and multivariable Fourier Series estimators in this semiparametric regression, which is a combination of multivariable parametric and multivariable nonparametric estimators. The estimators obtained are biased estimators but are a class of linear estimators.

Keywords: GCV, semiparametric, kernel.tatistical Modeling, School Dropout, Indonesia

N. Djam'an et al. (eds.), *Proceedings of the 5th International Conference on Statistics, Mathematics, Teaching, and Research 2023 (ICSMTR 2023)*, Advances in Computer Science Research 109,
https://doi.org/10.2991/978-94-6463-332-0_14

1 Introduction

Semiparametric regression is a form of regression analysis method, apart from parametric and nonparametric regression. Semiparametric regression contains parametric and nonparametric components, the regression curve is estimated using a semiparametric estimator. As is the case in parametric regression and nonparametric regression, semiparametric regression, especially for multivariable cases [1], allows for several types of estimators.

Several estimators in nonparametric regression, including the Kernel [2] are used for unpatterned data. Spline Smoothing [3] is used for data with changing patterns that do not depend on knot points (smooth). Truncated Splines are used for data with changing patterns in certain sub-intervals and Fourier Series [4] used on data that tends to repeat patterns. Among these estimators, the Kernel estimator is more often used in nonparametric regression. This is because the Kernel estimator is simpler [5]. In addition, the Kernel estimator has a relatively faster convergence speed compared to the Local Polynomial, Fourier Series or Spline estimators [6].

Meanwhile, the semiparametric regression estimator is used to estimate the regression model where the predictor variable has a parametric pattern and other predictor variables whose shape is unknown have been developed by several researchers. Semiparametric regression estimators without using mixed estimators have been developed by researchers including [7]. [8] developed Partial Spline Smoothing estimation, developed the selection of smoothing parameters for Spline Smoothing by using Risk estimation minimization,while [6]; [9] develops a Fourier Series approach to semiparametric regression.

These researchers design non-parametric or semi-parametric regression models without outlier data. In the actual situation there is an outlier data pattern, the researchers estimate the model by paying attention to the presence of outlier data. This was developed by Prestly_Chao in 1972.

To estimate semiparametric regression curves according to data patterns that contain outliers in their nonparametric components, both nonparametric and semiparametric regression were developed. Research on the Prestly_Chao estimator in semiparametric regression using the Kernel function approach. Therefore, a method for selecting the optimal bandwidth was developed. Next, the properties of the estimator obtained are shown.

2 Research Methods

To achieve the research objectives, the following steps were made.

1 Obtaining Estimator of Kernel Gaussian Multivariable Semiparametric Regression with Nadaraya_Watson Estimation
2 Selecting Bandwidth Parameters in Multivariable Kernel Estimator Semiparametric Regression with GCV.
3 Investigate the properties of the obtained estimates.

3 Results and Discussion

3.1 Multivariable Kernel Semiparametric Model

Given paired data $(x_{1i}, x_{2i}, ..., x_{pi}, t_{1i}, t_{2i}, ..., t_{qi}, y_i)$ which follows a semiparametric regression model $y_i = \mu(x_{1i}, x_{2i}, ..., x_{pi}, t_{1i}, t_{2i}, ..., t_{qi}) + \varepsilon_i$, $i = 1, 2, ..., n,$ μ is a regression curve, ε ais a random error that is assumed to be independent and identically normally distributed with zero mean and variance σ^2. Regression curve μ assumed to be additive, so it can be written as:

$$y_i = \sum_{j=1}^{p} m_j(x_{ji}) + \sum_{k=1}^{q} g_k(t_{ki}) + \varepsilon_i,\ i = 1, 2, \ldots, n, \tag{1}$$

where $\sum_{j=1}^{p} m_j(x_{ji}),\ i = 1, 2, \ldots, n$ is a parametric component and $\sum_{k=1}^{q} g_k(t_{ki})$ is a nonparametric component, whice for $i = 1, 2, \ldots, n$, obtained, the equation becomes:

$$y_1 = \sum_{j=1}^{p} m_j(x_{j1}) + \sum_{k=1}^{q} g_k(t_{k1}) + \varepsilon_1$$

$$y_n = \sum_{j=1}^{p} m_j(x_{jn}) + \sum_{k=1}^{q} g_k(t_{kn}) + \varepsilon_n$$

Next can be written:

$$\begin{gathered} y_1 = m_1(x_{11}) + \cdots + m_p(x_{p1}) + g_1(t_{11}) + \cdots + g_q(t_{q1}) + \varepsilon_1 \\ \vdots \\ y_n = m_1(x_{1n}) + \cdots + m_p(x_{pn}) + g_1(t_{1n}) + \cdots + g_q(t_{qn}) + \varepsilon_n \end{gathered}$$

As a result, the regression model in equation (1) can be written as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{p} m_j(x_{j1}) \\ \sum_{j=1}^{p} m_j(x_{j2}) \\ \vdots \\ \sum_{j=1}^{p} m_j(x_{jn}) \end{bmatrix} + \begin{bmatrix} \sum_{k=1}^{q} g_k(t_{k1}) \\ \sum_{k=1}^{q} g_k(t_{k2}) \\ \vdots \\ \sum_{k=1}^{q} g_k(t_{kn}) \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} m(x_1) \\ m(x_2) \\ \vdots \\ m(x_n) \end{bmatrix} + \begin{bmatrix} g(x_1) \\ g(x_2) \\ \vdots \\ g(x_n) \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

In matrix form it can be presented as follows:

$$\mathbf{y} = \mathbf{m_j} + \mathbf{g_k} + \boldsymbol{\varepsilon} \tag{2}$$

Next is the regression curve $\sum_{j=1}^{p} m_j(x_{jn})$ in equation (1) is a parametric component approximated by a linear function of the form:

$$m(x_i) = \beta_0 + \beta_1 x_i \tag{3}$$

Kernels. Parametric components so that equation (2) can be written as:

$$\begin{bmatrix} m(x_1) \\ m(x_2) \\ \vdots \\ m(x_n) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{p} m_j(x_{j1}) \\ \sum_{j=1}^{p} m_j(x_{j2}) \\ \vdots \\ \sum_{j=1}^{p} m_j(x_{jn}) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{p} \left(\beta_{0j} + \beta_{1j} x_{j1}\right) \\ \sum_{j=1}^{p} \left(\beta_{0j} + \beta_{1j} x_{j2}\right) \\ \vdots \\ \sum_{j=1}^{p} \left(\beta_{0j} + \beta_{1j} x_{jn}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \beta_{01} + \beta_{11} x_{11} + \beta_{02} + \beta_{12} x_{21} + \ldots + \beta_{0p} + \beta_{1p} x_{p1} \\ \beta_{01} + \beta_{11} x_{12} + \beta_{02} + \beta_{12} x_{22} + \ldots + \beta_{0p} + \beta_{1p} x_{p2} \\ \vdots \\ \beta_{01} + \beta_{11} x_{1n} + \beta_{02} + \beta_{12} x_{2n} + \ldots + \beta_{0p} + \beta_{1p} x_{pn} \end{bmatrix}$$

where: $\beta_0^* = \beta_{01} + \beta_{02} + \ldots + \beta_{0p} = \begin{bmatrix} 1 & x_{11} & x_{21} & \ldots & x_{p1} \\ 1 & x_{12} & x_{22} & \ldots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_0^* \\ \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1p} \end{bmatrix}$ (4)

So equation (4) can be written in matrix form, as follows:

$$\mathbf{m}_j = \mathbf{X}\boldsymbol{\beta} \tag{5}$$

where:

$$\mathbf{X}=\begin{bmatrix}1 & x_{11} & x_{21} & \dots & x_{p1}\\ 1 & x_{12} & x_{22} & \dots & x_{p2}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & x_{1n} & x_{2n} & \dots & x_{pn}\end{bmatrix} \text{dan}\, \boldsymbol{\beta}=\begin{bmatrix}\beta_0^* & \beta_{11} & \beta_{12} & & \beta_{1p}\end{bmatrix}^T$$

3.2 Multivariable Semiparametric Kernel Model Using Priestley and Chao

For regression curve $\sum_{k=1}^{q} g_k(t_{ki})$ in equation (1) is a nonparametric component of the Kernel $g_k(t_{ki}),\ k=1,2,...,q$ approached by a Kernel function, then by [10], estimate $\hat{g}$ is

$$\hat{g}_k(t_{ki})=\frac{1}{n\varphi_k}\sum_{i=1}^{n}K\left(\frac{t_k - t_{ki}}{\varphi_k}\right)y_i$$

then to look for an estimator $\sum_{k=1}^{q} g_k(t_{ki})$ in equation (1) then first look for the estimate $\hat{g}_k(t_{ki})$ with $k=k1$ up to you $k=kn$ For $g_k(t_{ki})$ which is approximated by the Kernel, then the estimator of $\hat{g}_k(t_{ki})$ is

$$:\hat{g}_k(t_{ki})=\frac{1}{n\varphi_k}\sum_{i=1}^{n}K\left(\frac{t_k - t_{ki}}{\varphi_k}\right)y_i \tag{6}$$

Based on equation (6), equation (7) can be written as:

$$\begin{bmatrix}\hat{g}(t_{k1})\\ \hat{g}(t_{k2})\\ \vdots \\ \hat{g}(t_{kn})\end{bmatrix}=\begin{bmatrix}\frac{1}{n\varphi_k}\sum_{i=1}^{n}K\left(\frac{t_{k1}-t_{k1}}{\varphi_k}\right)y_1+\dots+\frac{1}{n\varphi_k}\sum_{i=1}^{n}K\left(\frac{t_{k1}-t_{kn}}{\varphi_k}\right)y_n\\ \frac{1}{n\varphi_k}\sum_{i=1}^{n}K\left(\frac{t_{k2}-t_{k1}}{\varphi_k}\right)y_1+\dots+\frac{1}{n\varphi_k}\sum_{i=1}^{n}K\left(\frac{t_{k2}-t_{kn}}{\varphi_k}\right)y_n\\ \vdots \\ \frac{1}{n\varphi_k}\sum_{i=1}^{n}K\left(\frac{t_{kq}-t_{k1}}{\varphi_k}\right)y_1+\dots+\frac{1}{n\varphi_k}\sum_{i=1}^{n}K\left(\frac{t_{kq}-t_{kn}}{\varphi_k}\right)y_n\end{bmatrix} \tag{7}$$

So equation (7) can be written in matrix form, as follows:

$$\hat{g}_k(t_{ki})=\mathbf{\Omega}_k(\varphi_k)\mathbf{y} \tag{8}$$

As a result of equation (7), then the estimate for the regression curve $\sum_{k=1}^{q} g_k(t_{ki})$ obtained, as follows:

$$\sum_{k=1}^{q}\hat{g}_k(t_{ki})=\sum_{k=1}^{q}\mathbf{\Omega}_k(\varphi_k)\mathbf{y}=\mathbf{\Omega}_1(\varphi_k)\mathbf{y}+\ldots+\mathbf{\Omega}_g(\varphi_k)\mathbf{y}=\mathbf{P}_k\mathbf{y} \tag{9}$$

dimana : $\mathbf{P}_k=\mathbf{\Omega}_1(\varphi_k)+\ldots+\mathbf{\Omega}_g(\varphi_k)$

$\hat{\mathbf{g}}_k$ vektor berukuran nx1, $\mathbf{P}_k$ matriks berukuran nxn, dan $\mathbf{y}$ vektor berukuran nx1

$$\mathbf{y}=\left[y1\ \ y2\ \ldots yn\right]^T$$

$$\mathbf{P}_k(\varphi_k)=\begin{bmatrix}\frac{1}{n\varphi_k}\sum_{i=1}^{n}K\left(\frac{t_{k1}-t_{k1}}{\varphi_k}\right)\mathrm{y}_1+\ldots+\frac{1}{n\varphi_k}\sum_{i=1}^{n}K\left(\frac{t_{k1}-t_{kn}}{\varphi_k}\right)\mathrm{y}_n\\ \frac{1}{n\varphi_k}\sum_{i=1}^{n}K\left(\frac{t_{k2}-t_{k1}}{\varphi_k}\right)\mathrm{y}_1+\ldots+\frac{1}{n\varphi_k}\sum_{i=1}^{n}K\left(\frac{t_{k2}-t_{kn}}{\varphi_k}\right)\mathrm{y}_n\\ \vdots\\ \frac{1}{n\varphi_k}\sum_{i=1}^{n}K\left(\frac{t_{kq}-t_{k1}}{\varphi_k}\right)\mathrm{y}_1+\ldots+\frac{1}{n\varphi_k}\sum_{i=1}^{n}K\left(\frac{t_{kq}-t_{kn}}{\varphi_k}\right)\mathrm{y}_n\end{bmatrix}$$

K is Kernel

φ is bandwidthd

Next, to find the parametric component estimator, it can be obtained from minimizing $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ from the equation:

$$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}=(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{P}_k\mathbf{y})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{P}_k\mathbf{y})$$

$$\frac{\partial(\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon})}{\partial\boldsymbol{\beta}}=-2\mathbf{X}^T\mathbf{y}+2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}-2\mathbf{X}^T\mathbf{P}_k\mathbf{y}=0$$

$$\hat{\boldsymbol{\beta}}=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{P}_k+\mathbf{I})\mathbf{y} \tag{10}$$

Based on equations (9) and (10), the estimates for equation (1) are obtained as follows:

$$\hat{\mathbf{y}}=\mathbf{X}\hat{\boldsymbol{\beta}}-\hat{\mathbf{g}}_k$$

$$\hat{\mathbf{y}}=\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{P}_k+\mathbf{I})\mathbf{y}+\mathbf{P}_k\mathbf{y}$$

$$=\boldsymbol{\lambda}^*\mathbf{Y} \tag{11}$$

Where $\boldsymbol{\lambda}^*=\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(2\mathbf{P}_k+\mathbf{I})$

Multivariable Kernel mixture estimator in semiparametric regression, very dependent on optimal smoothing parameters, optimal bandwidth. One method that can be used is the GCV method [10].

If the Multivariable Kernel mixture estimator in semiparametric regression is presented in equation (11), it is as follows:

$$\hat{\boldsymbol{\mu}}_{(\boldsymbol{\beta},\boldsymbol{\varphi})}(\mathbf{x},\mathbf{t},\mathbf{z})=\boldsymbol{\lambda}^*\mathbf{y}$$

then the mean square error (MSE) is obtained:

$$\text{MSE}(\boldsymbol{\varphi})=n^{-1}\left\|\left(\mathbf{I}-\boldsymbol{\lambda}^*\right)\mathbf{y}\right\|^2,$$

so that the GCV of the Multivariable Kernel estimator in semiparametric regression:

$$GCV(\boldsymbol{\varphi}) = \frac{n^{-1}\left\|\left(\mathbf{I}-\boldsymbol{\lambda}^{*}\right)\mathbf{y}\right\|^{2}}{(n^{-1}trace(\mathbf{I}-\boldsymbol{\lambda}^{*})^{2}} \tag{11}$$

The optimal smoothing parameters, optimal oscillation parameters and optimal bandwidth are obtained in $GCV(\boldsymbol{\varphi})$ thesmallest.

3.3 Properties of the Regression Curve Estimator

In this subchapter we will discuss the properties of mixed Kernel and multivariable Fourier Series estimators in semiparametric regression. Before investigating the properties of the estimator, first investigate the properties of the estimator $\hat{\boldsymbol{\gamma}}$, estimator $\hat{\boldsymbol{\beta}}$, parametric regression curve estimator $\hat{\mathbf{m}}_{(\boldsymbol{\beta},\boldsymbol{\varphi})}(\mathbf{x},\mathbf{t})$, kernel regression curve estimator $\hat{\mathbf{g}}_{(\boldsymbol{\beta},\boldsymbol{\varphi})}(\mathbf{x},\mathbf{t})$, and Semiparametric regression curve estimator $\hat{\boldsymbol{\mu}}_{(\boldsymbol{\beta},\boldsymbol{\varphi})}(\mathbf{x},\mathbf{t})$. Then proceed with the properties of mixed Kernel and Fourier Series estimators $\hat{\boldsymbol{\mu}}_{(\boldsymbol{\beta},\boldsymbol{\varphi})}(\mathbf{x},\mathbf{t})$. This property is investigated from the estimator's expected value.

First of all, look for the expected value of**estimator** $\hat{\boldsymbol{\beta}}$ is:

$E(\hat{\boldsymbol{\beta}}) \neq \boldsymbol{\beta}$ Andobtained: $E(\hat{\boldsymbol{\gamma}}) \neq \boldsymbol{\gamma}$

Next, look for the expected value of the regression curve estimator $\sum_{j=1}^{p}\hat{\mathbf{m}}_{j}(x_{j})$ is:

$$E\left[\sum_{j=1}^{p}\hat{\mathbf{m}}_{j}(x_{j})\right] = E\left[\mathbf{X}\mathbf{T}^{*}\mathbf{y}\right] = \mathbf{X}\mathbf{T}^{*}E\left[\mathbf{y}\right]$$

$$= \mathbf{X}\mathbf{T}^{*}E\left[\sum_{j=1}^{p}m_{j}(x_{ji}) + \sum_{k=1}^{q}g_{k}(t_{ki})\right]$$

$$E\left[\sum_{j=1}^{p}\hat{\mathbf{m}}_{j}(x_{j})\right] \neq \sum_{j=1}^{p}m_{j}(x_{ji}) \tag{12}$$

Then look for the expected value of the regression curve estimator $\sum_{k=1}^{q}\hat{\mathbf{g}}_{k}(t_{k})$ is:

$$E\left[\sum_{j=1}^{p}\hat{\mathbf{g}}_{j}(x_{j})\right] \neq \sum_{k=1}^{q}g_{k}(t_{ki}) \tag{13}$$

Next is the expected value of the regression curve estimator $\hat{\boldsymbol{\mu}}_{(\boldsymbol{\beta},\boldsymbol{\varphi})}(\mathbf{x},\mathbf{t})$, following:

$$E\left[\hat{\boldsymbol{\mu}}_{(\boldsymbol{\beta},\boldsymbol{\varphi},\boldsymbol{\lambda},\mathbf{s})}(\mathbf{x},\mathbf{t},\mathbf{z})\right] \neq \mathbf{X}(\mathbf{X}^{T}\mathbf{X})^{-1}\mathbf{X}^{T}(\mathbf{P}_{k}+\mathbf{I})\mathbf{y} + \mathbf{P}_{k}\mathbf{y}$$

obtained:

$$E\left[\hat{\boldsymbol{\mu}}_{(\boldsymbol{\beta},\boldsymbol{\varphi})}(\mathbf{x},\mathbf{t})\right]\neq \boldsymbol{\mu}_{(\boldsymbol{\beta},\boldsymbol{\varphi})}(\mathbf{x},\mathbf{t}) \quad (14)$$

Estimators $\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}$, $\hat{\mathbf{m}}_{(\boldsymbol{\beta},\boldsymbol{\varphi})}(\mathbf{x},\mathbf{t}), \hat{\mathbf{g}}_{(\boldsymbol{\beta},\boldsymbol{\varphi})}(\mathbf{x},\mathbf{t})$, and estimator $\hat{\boldsymbol{\mu}}_{(\boldsymbol{\beta},\boldsymbol{\varphi})}(\mathbf{x},\mathbf{t})$ biased from view(12), (13), and (14). like other semipametric regression curve estimators. Nevertheless, this estimator is linear in y. This can be seen in the discussion (12), (13), and (14), which produces $\hat{\mathbf{m}}_{(\boldsymbol{\beta},\boldsymbol{\varphi})}(\mathbf{x},\mathbf{t})=\mathbf{X}\mathbf{T}^{*}\mathbf{y},\ \hat{\mathbf{g}}_{(\boldsymbol{\beta},\boldsymbol{\varphi})}(\mathbf{x},\mathbf{t})=\mathbf{P}\mathbf{y},$ And $\hat{\boldsymbol{\mu}}_{(\boldsymbol{\beta},\boldsymbol{\varphi})}(\mathbf{x},\mathbf{t})=\boldsymbol{\lambda}^{*}\mathbf{y}$.

4 Conclusion

Given paired data $(x_{1i},\ x_{2i},\ ...,\ x_{pi},\ t_{1i},\ t_{2i},\ ...,\ t_{qi},\ y_i)$ which follows a semiparametric regression model $y_i=\mu(x_{1i},\ x_{2i},\ ...,\ x_{pi},\ t_{1i},\ t_{2i},\ ...,\ t_{qi})+\varepsilon_i$, $i=1,2,...,n,$ μ is a regression curve, ε ais a random error that is assumed to be independent and identically normally distributed with zero mean and variance σ^2. Regression curve μ assumed to be additive, so it can be written as:

$$y_i=\sum_{j=1}^{p}m_j(x_{ji})+\sum_{k=1}^{q}g_k(t_{ki})+\varepsilon_i,\ i=1,2,\ldots,n$$

where $\sum_{j=1}^{p}m_j(x_{ji}),\ i=1,2,\ldots,n$ is a parametric component and $\sum_{k=1}^{q}g_k(t_{ki})$ is a nonparametric component, so that:

1. $\hat{\mathbf{y}}=\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{P}_k+\mathbf{I})\mathbf{y}+\mathbf{P}_k\mathbf{y}$

where:

$$\mathbf{P}_k=\boldsymbol{\Omega}_1(\varphi_k)+\ldots+\boldsymbol{\Omega}_g(\varphi_k)$$

$$\mathbf{X}=\begin{bmatrix}1 & x_{11} & x_{21} & \cdots & x_{p1}\\ 1 & x_{12} & x_{22} & \cdots & x_{p2}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn}\end{bmatrix}$$

2. The method for selecting optimal bandwidth parameters using GCV produces the following results.

$$GCV(\boldsymbol{\varphi})=\frac{n^{-1}\left\|\left(\mathbf{I}-\boldsymbol{\lambda}^{*}\right)\ \mathbf{y}\ \right\|^2}{(n^{-1}trace(\mathbf{I}-\boldsymbol{\lambda}^{*})^2}$$

where:

$$\boldsymbol{\lambda}^* = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(2\mathbf{P}_k + \mathbf{I})$$

3. The results obtained show that the estimator obtained is a biased but linear estimator

The results of the theoretical study obtained from the Mixed Kernel Regression Curve and multivariable Fourier Series estimators in this semiparametric regression, which is a combination of multivariable parametric and multivariable nonparametric estimators. The estimators obtained are biased estimators but are a class of linear estimators.

References

1. Makkulau, Gusti N.A. Wibawa, Andi Tenri Ampa, Andi Tenri Pannangngareng Makkulau, Sri Harini, Angga Dwi Mulyanto. Discriminant Analysis for Determination of Early Childhood Education Accreditation In Southeast Sulawesi Province, CAUCHY –Jurnal Matematika Murni dan Aplikasi Volume 8(2) (2023), Pages 7-15 p-ISSN: 2086-0382; e-ISSN: 2477-3344
2. Nadaraya and Watson: On Estimating Regression. Theori Of Probability and ITS Application (1964).
3. Eubank, R.L: Spline Smoothing and Nonparametrik Regression, Mercel Dekker, New York,(1988).
4. Bilodeau, M.: Fourier Smoother and Additive Models, *The Canadion Journal of Statistical,* 3: 257-269 (1992).
5. Okumura, H. and Naito, K.: Nonparametric Kernel Regression for Multinominal Data, *Journal of Statistics Planning and Inference,* 14 (2006).
6. Cheng, M.Y., Paige, R.L., Sun, S., and Yan, K.: Variance Reduction for Kernel Estimatiors in Clustered/Longitudinal Data Analysis, *Journal of Statistical Planning and Inference,* 140: 1389-1397 (2010).
7. Wahba, G.: *Spline Models for Observasion Data*, SIAM Pensylvania (1990).
8. Budiantara, I N. and Mulianah: Pemilihan Bandwidth Optimal dalam Regresi Semiparametrik Kernel dan Aplikasinya, *Sigma*, 10, 2: 159-166 (2007).
9. Ampa, AT., Budiantara,IN., Zain, I.:Modeling the Level of Drinking Water Clarity in Surabaya City Drinking Water Regional Company Using Combined Estimation of Multivariable Fourier Series and Kernel, *Sustainability* **2022**, *14*(20), 13663; https://doi.org/10.3390/su142013663(2022)
10. Ampa, AT., Budiantara,IN., Zain, I.: Selection of Optimal Smoothing Parameters in Mixed Estimator of Kernel and Fourier Series in Semiparametric Regression, J. Phys.: Conf. Ser. 2123 012035 (2021).