

Pemodelan Regresi Nonparametrik Spline Truncated pada Data Longitudinal

(Studi Kasus: Persentase Penduduk Miskin Provinsi Papua Tahun 2016-2019)

Andrea Tri Rian Dani^{1*}, Ludia Ni'matuzzahroh², Vita Ratnasari³, dan I Nyoman Budiantara⁴

Received: 24 February 2021

Accepted: 27 March 2021

Published: 31 March 2021

^{1,2,3,4}Departemen Statistika, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111, Indonesia

*Corresponding author: andreatriandani98@gmail.com

ABSTRAK – Saat ini pendekatan regresi nonparametrik banyak mendapat perhatian dari para peneliti, dikarenakan memiliki fleksibilitas yang tinggi dan tidak tergantung pada asumsi bentuk kurva regresi. Spline truncated adalah salah satu model dalam regresi nonparametrik yang sering digunakan, karena mampu menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu. Pada analisis regresi, data yang seringkali digunakan adalah data *cross-section*, namun yang sebenarnya adalah analisis regresi juga dapat diterapkan pada data longitudinal, khususnya dengan menggunakan pendekatan regresi nonparametrik. Data longitudinal merupakan gabungan antara data *cross-section* dan *time series*. Penelitian ini bertujuan untuk memodelkan data persentase penduduk miskin di Provinsi Papua Tahun 2016 hingga Tahun 2019 menggunakan model regresi nonparametrik spline truncated. Metode estimasi parameter yang digunakan adalah *Weighted Least Squares* (WLS). Banyaknya titik knot yang dicobakan adalah 1 hingga 3 titik knot. Berdasarkan hasil analisis, model regresi nonparametrik spline truncated terbaik adalah model yang menggunakan 1 titik knot, dengan nilai GCV yang paling minimum yaitu sebesar 8,05 dan Koefisien Determinasi (R^2) sebesar 99,98%.

Kata Kunci– Kemiskinan, Longitudinal, Spline Truncated

I. PENDAHULUAN

Salah satu analisis dalam statistika yang dapat digunakan untuk menyelidiki pola hubungan antara satu atau lebih variabel, yaitu analisis regresi. Variabel yang digunakan dalam pemodelan regresi yaitu variabel respon dan variabel prediktor. Tujuan utama dalam melakukan analisis regresi yaitu mencari bentuk estimasi kurva regresi yang tepat [1]. Terdapat tiga pendekatan dalam melakukan estimasi kurva regresi, yaitu pendekatan regresi parametrik, regresi semiparametrik, dan regresi nonparametrik. Pendekatan regresi parametrik dapat digunakan apabila bentuk kurva regresi diketahui. Pendekatan regresi semiparametrik dapat digunakan apabila bentuk kurva regresi sebagian diketahui dan sebagian tidak diketahui, sedangkan pada pendekatan regresi nonparametrik bentuk kurva regresi diasumsikan tidak diketahui. Dengan demikian, pendekatan regresi nonparametrik tidak tergantung pada asumsi bentuk kurva regresi tertentu, sehingga memiliki fleksibilitas yang tinggi [2]. Pada pemodelan dengan menggunakan pendekatan regresi nonparametrik, terdapat beberapa estimator yang telah dikembangkan oleh para peneliti, diantaranya adalah Spline, Kernel, dan Deret Fourier.

Salah satu model regresi nonparametrik yang sering digunakan yaitu spline. Kelebihan spline adalah mampu mengestimasi pola data sesuai pergerakannya. Hal tersebut terjadi karena dalam regresi nonparametrik spline terdapat titik-titik knot, yaitu titik perpaduan bersama yang menunjukkan terjadinya perubahan pola perilaku data [2]. Salah satu basis fungsi yang biasa digunakan dalam pendekatan spline, yaitu spline truncated. Adapun beberapa penelitian sebelumnya yang menggunakan pendekatan regresi nonparametrik spline truncated diantaranya [3]-[4].

Pada analisis regresi, biasanya data yang digunakan merupakan data *cross-section*. Namun saat ini, analisis regresi juga dapat diterapkan pada data longitudinal, khususnya dengan pendekatan regresi nonparametrik. Data longitudinal merupakan gabungan antara data *cross-section* dengan data *time series* [5]. Beberapa keuntungan dalam penggunaan data longitudinal yaitu dapat mengurangi kolinieritas antar variabel prediktor sehingga mampu menghasilkan estimasi yang lebih efisien [6]. Penerapan proses estimasi model regresi pada data longitudinal, mampu mengestimasi karakteristik subyek secara individu dengan memperhatikan adanya dinamika antar waktu dan masing-masing variabel dalam penelitian secara bersama-sama. Adapun beberapa penelitian sebelumnya yang menggunakan pendekatan regresi nonparametrik pada data longitudinal diantaranya [7]-[9].

Pendekatan regresi nonparametrik dapat diterapkan di berbagai bidang keilmuan, salah satunya yang berkaitan dengan masalah sosial. Salah satu masalah yang sering dijumpai yaitu tingkat pendidikan yang rendah dapat memicu timbulnya kemiskinan. Berdasarkan data BPS pada Tahun 2020, Provinsi Papua memiliki persentase penduduk miskin tertinggi di Indonesia. Badan Pusat Statistik (BPS) mendefinisikan kemiskinan sebagai ketidakmampuan individu dalam memenuhi kebutuhan dasar minimal untuk hidup layak.

Berdasarkan uraian diatas, maka pada penelitian ini akan dimodelkan persentase kemiskinan Kabupaten/Kota di Provinsi Papua dengan menggunakan regresi nonparametrik spline truncated pada data longitudinal, dimana periode

tahun yang digunakan dimulai dari Tahun 2016 hingga Tahun 2019. Diharapkan hasil penelitian ini dapat bermanfaat dalam pengembangan ilmu statistika khususnya pada bidang regresi nonparametrik.

II. TINJAUAN PUSTAKA

Pada bagian ini, akan dibahas mengenai beberapa landasan teori yang terkait:

A. Regresi Nonparametrik

Salah satu pendekatan dalam melakukan estimasi kurva regresi yaitu dengan menggunakan pendekatan regresi nonparametrik. Pendekatan regresi nonparametrik digunakan apabila bentuk kurva regresi atau pola data tidak diketahui. Regresi nonparametrik memiliki fleksibilitas dalam mencari sendiri bentuk kurva regresinya tanpa dipengaruhi subjektivitas peneliti [2]. Secara umum, model regresi nonparametrik dapat dituliskan seperti pada Persamaan (1).

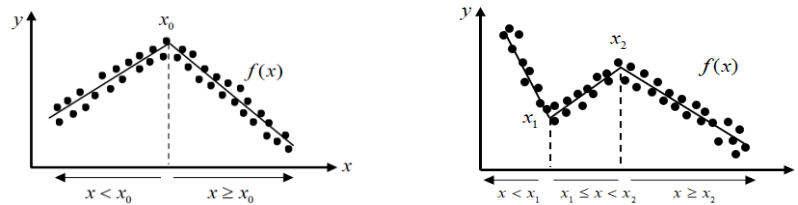
$$y_i = f(x_i) + \varepsilon_i \quad ; i = 1, 2, \dots, n \tag{1}$$

dimana:

- y_i : variabel respon
- $f(x_i)$: kurva regresi yang akan dihampiri
- x_i : variabel prediktor
- ε_i : *error random* yang diasumsikan identic, independen, dan berdistribusi normal dengan *mean* bernilai nol dan variansi σ^2 .

B. Regresi Nonparametrik Spline Truncated pada Data Longitudinal

Spline truncated mampu menangani data yang memiliki perilaku yang berubah-ubah pada sub-sub interval tertentu dan cenderung mencari sendiri estimasi data kemanapun pola data bergerak. Kelebihan dari spline truncated ini terjadi karena dalam fungsi terdapat adanya titik-titik knot, yang merupakan titik perpaduan bersama dimana terdapat perubahan perilaku data atau fungsi yang polanya berubah [1]. Pola perubahan pola data dapat berbentuk beraneka ragam, sebagai contoh disajikan pada Gambar 1.



Gambar 1 Ilustrasi Pola Perubahan Data berdasarkan Titik Knot

Secara umum, model regresi nonparametrik spline truncated multivariabel pada data longitudinal dapat dituliskan seperti pada Persamaan (2).

$$y_{ij} = f(x_{ijq}) + \varepsilon_{ij} \quad ; i = 1, 2, \dots, n ; j = 1, 2, \dots, t ; q = 1, 2, \dots, Q \tag{2}$$

dengan,

$$f(x_{ijq}) = \sum_{q=1}^Q \left(\sum_{m=1}^M \theta_{miq} x_{ijq}^m + \sum_{l=1}^R \delta_{liq} (x_{ijq} - K_{liq})_+^M \right)$$

$$(x_{ijq} - K_{liq})_+^M = \begin{cases} (x_{ijq} - K_{liq})^M & ; x_{ijq} \geq K_{(l+M)i} \\ 0 & ; x_{ijq} < K_{(l+M)i} \end{cases}$$

dimana:

- y_{ij} : variabel respon
- $f(x_{ijq})$: fungsi regresi nonparametrik spline truncated
- x_{ijq} : variabel prediktor
- ε_{ij} : *error random* yang diasumsikan identic, independen, dan berdistribusi normal dengan *mean* bernilai nol dan variansi σ^2
- $\sum_{m=1}^M \theta_{miq} x_{ijq}^m$: komponen polinomial
- $\sum_{l=1}^R \delta_{li} (x_{ijq} - K_{liq})_+^M$: komponen truncated

Persamaan (2) dapat ditulis kedalam bentuk notasi matriks sebagai berikut:

$$\tilde{y} = X\tilde{\beta} + \tilde{\varepsilon} \tag{3}$$

Dimana dapat dijabarkan:

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_n \end{bmatrix} = \begin{bmatrix} X_1(K) & 0 & \dots & 0 \\ 0 & X_2(K) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_n(K) \end{bmatrix} \begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \vdots \\ \tilde{\beta}_n \end{bmatrix} + \begin{bmatrix} \tilde{\varepsilon}_1 \\ \tilde{\varepsilon}_2 \\ \vdots \\ \tilde{\varepsilon}_n \end{bmatrix}$$

Dengan vektor respon nya sebagai berikut:

$$\begin{aligned} \tilde{y}_1 &= (y_{11} \ y_{12} \ \dots \ y_{1t})^T \\ \tilde{y}_2 &= (y_{21} \ y_{22} \ \dots \ y_{2t})^T \\ &\vdots \\ \tilde{y}_n &= (y_{n1} \ y_{n2} \ \dots \ y_{nt})^T \end{aligned}$$

Kemudian matriks komponen nonparametrik spline truncated:

$$\begin{aligned} X_1(K) &= \begin{bmatrix} x_{111} & \dots & x_{111}^M & \dots & x_{11Q} & \dots & x_{11Q}^M & (x_{111} - K_{111})_+^M & \dots & (x_{111} - K_{R11})_+^M & \dots & (x_{11Q} - K_{11Q})_+^M & \dots & (x_{11Q} - K_{R1Q})_+^M \\ x_{121} & \dots & x_{121}^M & \dots & x_{12Q} & \dots & x_{12Q}^M & (x_{121} - K_{111})_+^M & \dots & (x_{121} - K_{R11})_+^M & \dots & (x_{12Q} - K_{11Q})_+^M & \dots & (x_{12Q} - K_{R1Q})_+^M \\ \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{1t1} & \dots & x_{1t1}^M & \dots & x_{1tQ} & \dots & x_{1tQ}^M & (x_{1t1} - K_{111})_+^M & \dots & (x_{1t1} - K_{R11})_+^M & \dots & (x_{1tQ} - K_{11Q})_+^M & \dots & (x_{1tQ} - K_{R1Q})_+^M \end{bmatrix} \\ \\ X_2(K) &= \begin{bmatrix} x_{211} & \dots & x_{211}^M & \dots & x_{21Q} & \dots & x_{21Q}^M & (x_{211} - K_{121})_+^M & \dots & (x_{211} - K_{R21})_+^M & \dots & (x_{21Q} - K_{12Q})_+^M & \dots & (x_{21Q} - K_{R2Q})_+^M \\ x_{221} & \dots & x_{221}^M & \dots & x_{22Q} & \dots & x_{22Q}^M & (x_{221} - K_{121})_+^M & \dots & (x_{221} - K_{R21})_+^M & \dots & (x_{22Q} - K_{12Q})_+^M & \dots & (x_{22Q} - K_{R2Q})_+^M \\ \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{2t1} & \dots & x_{2t1}^M & \dots & x_{2tQ} & \dots & x_{2tQ}^M & (x_{2t1} - K_{121})_+^M & \dots & (x_{2t1} - K_{R21})_+^M & \dots & (x_{2tQ} - K_{12Q})_+^M & \dots & (x_{2tQ} - K_{R2Q})_+^M \end{bmatrix} \\ \\ &\vdots \\ \\ X_n(K) &= \begin{bmatrix} x_{n11} & \dots & x_{n11}^M & \dots & x_{n1Q} & \dots & x_{n1Q}^M & (x_{n11} - K_{1n1})_+^M & \dots & (x_{n11} - K_{Rn1})_+^M & \dots & (x_{n1Q} - K_{1nQ})_+^M & \dots & (x_{n1Q} - K_{RnQ})_+^M \\ x_{n21} & \dots & x_{n21}^M & \dots & x_{n2Q} & \dots & x_{n2Q}^M & (x_{n21} - K_{1n1})_+^M & \dots & (x_{n21} - K_{Rn1})_+^M & \dots & (x_{n2Q} - K_{1nQ})_+^M & \dots & (x_{n2Q} - K_{RnQ})_+^M \\ \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{nt1} & \dots & x_{nt1}^M & \dots & x_{ntQ} & \dots & x_{ntQ}^M & (x_{nt1} - K_{1n1})_+^M & \dots & (x_{nt1} - K_{Rn1})_+^M & \dots & (x_{ntQ} - K_{1nQ})_+^M & \dots & (x_{ntQ} - K_{RnQ})_+^M \end{bmatrix} \end{aligned}$$

Dan vektor parameter:

$$\begin{aligned} \tilde{\beta}_1 &= (\theta_{111} \ \dots \ \theta_{m11} \ \dots \ \theta_{11Q} \ \dots \ \theta_{m1Q} \ \delta_{111} \ \dots \ \delta_{R11} \ \dots \ \delta_{11Q} \ \dots \ \delta_{R1Q})^T \\ \tilde{\beta}_2 &= (\theta_{121} \ \dots \ \theta_{m21} \ \dots \ \theta_{12Q} \ \dots \ \theta_{m2Q} \ \delta_{121} \ \dots \ \delta_{R21} \ \dots \ \delta_{12Q} \ \dots \ \delta_{R2Q})^T \\ &\vdots \\ \tilde{\beta}_n &= (\theta_{1n1} \ \dots \ \theta_{mn1} \ \dots \ \theta_{1nQ} \ \dots \ \theta_{mnQ} \ \delta_{1n1} \ \dots \ \delta_{Rn1} \ \dots \ \delta_{1nQ} \ \dots \ \delta_{RnQ})^T \end{aligned}$$

Serta vektor errornya sebagai berikut:

$$\begin{aligned} \tilde{\varepsilon}_1 &= (\varepsilon_{11} \ \varepsilon_{12} \ \dots \ \varepsilon_{1t})^T \\ \tilde{\varepsilon}_2 &= (\varepsilon_{21} \ \varepsilon_{22} \ \dots \ \varepsilon_{2t})^T \\ &\vdots \\ \tilde{\varepsilon}_n &= (\varepsilon_{n1} \ \varepsilon_{n2} \ \dots \ \varepsilon_{nt})^T \end{aligned}$$

Dimana \tilde{y} berupa vektor kolom berukuran $nt \times 1$, matriks X merupakan matriks berukuran $nt \times (M + R)nQ$, $\tilde{\beta}$ merupakan vektor kolom berukuran $(M + R)nQ \times 1$, dan $\tilde{\varepsilon}$ berupa vektor kolom berukuran $nt \times 1$. Metode estimasi parameter pada penelitian ini adalah Weighted Least Squares (WLS), karena metode estimasi ini mampu mengatasi korelasi dalam subyek pengamatan yang sama. Berdasarkan Persamaan (3), maka estimator $\tilde{\beta}$ dapat diperoleh dengan menyelesaikan optimasi WLS sebagai berikut:

$$\min_{\tilde{\beta} \in R^{(M+R)nQ}} \{(\tilde{y} - X\tilde{\beta})^T W(\tilde{y} - X\tilde{\beta})\} \tag{4}$$

W merupakan matriks pembobot berukuran $nt \times nt$ dan berisi diagonal (W_1, W_2, \dots, W_n) . Penyelesaian optimasi pada Persamaan (4) sebagai berikut:

$$(\tilde{y} - X\tilde{\beta})^T W(\tilde{y} - X\tilde{\beta}) = (\tilde{y}^T - \tilde{\beta}^T X^T)W(\tilde{y} - X\tilde{\beta})$$

$$\begin{aligned}
 &= \tilde{y}^T \mathbf{W} \tilde{y} - \tilde{\beta}^T \mathbf{X}^T \mathbf{W} \tilde{y} - \tilde{y}^T \mathbf{W} \mathbf{X} \tilde{\beta} + \tilde{\beta}^T \mathbf{X}^T \mathbf{W} \mathbf{X} \tilde{\beta} \\
 &= \tilde{y}^T \mathbf{W} \tilde{y} - 2\tilde{\beta}^T \mathbf{X}^T \mathbf{W} \tilde{y} + \tilde{\beta}^T \mathbf{X}^T \mathbf{W} \mathbf{X} \tilde{\beta} \\
 &= P(\tilde{\beta})
 \end{aligned}$$

Untuk mendapatkan estimasi $\tilde{\beta}$, maka dapat melakukan derivatif parsial $P(\tilde{\beta})$ terhadap $\tilde{\beta}$, sehingga:

$$\begin{aligned}
 \frac{\partial P(\tilde{\beta})}{\partial \tilde{\beta}} &= \frac{\partial (\tilde{y}^T \mathbf{W} \tilde{y} - 2\tilde{\beta}^T \mathbf{X}^T \mathbf{W} \tilde{y} + \tilde{\beta}^T \mathbf{X}^T \mathbf{W} \mathbf{X} \tilde{\beta})}{\partial \tilde{\beta}} \\
 &= -2\mathbf{X}^T \mathbf{W} \tilde{y} + 2\mathbf{X}^T \mathbf{W} \mathbf{X} \tilde{\beta}
 \end{aligned}$$

Kemudian penurunan secara parsial diatas disama dengankan dengan nol, sehingga dapat dituliskan:

$$-2\mathbf{X}^T \mathbf{W} \tilde{y} + 2\mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\beta} = 0 \tag{5}$$

Persamaan (5) dapat ditulis kedalam bentuk

$$\mathbf{X}^T \mathbf{W} \tilde{y} = \mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\beta} \tag{6}$$

Sehingga berdasarkan Persamaan (6), dapat diperoleh estimator $\hat{\beta}$ sebagai berikut:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{W} \tilde{y})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{X} \tag{7}$$

Dengan menggunakan matriks pembobot \mathbf{W} :

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{W}_{n_t} \end{bmatrix}$$

C. Pemilihan Titik Knot Optimal

Pada regresi nonparametrik spline truncated, menentukan titik knot optimal menjadi hal yang sangat penting. Titik knot merupakan titik perpaduan bersama dimana terdapat perubahan perilaku pola data pada selang yang berbeda [10]. Salah satu metode yang dikembangkan oleh peneliti, untuk menentukan titik knot yang optimal yaitu dengan menggunakan *Generalized Cross-Validation* (GCV) [11]. Fungsi GCV untuk model regresi nonparametrik spline truncated pada data longitudinal ditampilkan pada Persamaan (8).

$$GCV(\tilde{k}) = \frac{(nt)^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{[(nt)^{-1} \text{trace}(\mathbf{I} - \mathbf{A}(\tilde{k}))]^2} \tag{8}$$

Dimana:

$GCV(\tilde{k})$: vektor yang memuat nilai GCV dari titik-titik knot

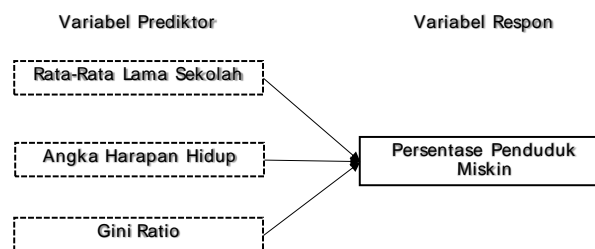
nt : banyaknya pengamatan sebanyak $n \times t$

$\mathbf{A}(\tilde{k})$: matriks yang memuat titik knot $\tilde{k} = (k_1, k_2, \dots, k_R)^T$ diperoleh $\mathbf{X}(\tilde{k})[\mathbf{X}(\tilde{k})^T \mathbf{X}(\tilde{k})]^{-1} \mathbf{X}(\tilde{k})^T$

III. METODOLOGI PENELITIAN

A. Variabel Penelitian

Berdasarkan penjelasan mengenai bagian kemiskinan pada bagian II, maka terdapat beberapa faktor yang diduga mempengaruhi kemiskinan. Adapun kerangka konseptual penelitian disajikan pada Gambar 2.



Gambar 2 Kerangka Konseptual Penelitian

Variabel yang digunakan pada penelitian ini terdiri atas variabel respon dan variabel prediktor.

Variabel Respon:

Y : Persentase Penduduk Miskin

Variabel Prediktor:

X_1 : Rata-Rata Lama Sekolah

X_2 : Angka Harapan Hidup

X_3 : Gini Ratio

B. Struktur Data Penelitian

Struktur data pada penelitian ini disajikan pada Tabel 1.

Tabel 1 Struktur Data Penelitian dengan Data Longitudinal

Kabupaten/Kota (i)	Tahun (j)	Y_{ij}	$X_{i,j,1}$	$X_{i,j,2}$	$X_{i,j,3}$
1	2016	$Y_{1,1}$	$X_{1,1,1}$	$X_{1,1,2}$	$X_{1,1,3}$
	2017	$Y_{1,2}$	$X_{1,2,1}$	$X_{1,2,2}$	$X_{1,2,3}$
	2018	$Y_{1,3}$	$X_{1,3,1}$	$X_{1,3,2}$	$X_{1,3,3}$
	2019	$Y_{1,4}$	$X_{1,4,1}$	$X_{1,4,2}$	$X_{1,4,3}$
⋮	⋮	⋮	⋮	⋮	⋮
29	2016	$Y_{29,1}$	$X_{29,1,1}$	$X_{29,1,2}$	$X_{29,1,3}$
	2017	$Y_{29,2}$	$X_{29,2,1}$	$X_{29,2,2}$	$X_{29,2,3}$
	2018	$Y_{29,3}$	$X_{29,3,1}$	$X_{29,3,2}$	$X_{29,3,3}$
	2019	$Y_{29,4}$	$X_{29,4,1}$	$X_{29,4,2}$	$X_{29,4,3}$

Provinsi Papua terdiri dari 28 Kabupaten dan 1 Kota. Pada penelitian ini tahun yang digunakan dimulai dari Tahun 2016 hingga Tahun 2019, dengan variabel prediktor yang digunakan sebanyak 3 (X_1 hingga X_3)

C. Langkah-Langkah Penelitian

Untuk menyelesaikan tujuan penelitian, maka perlu disusun langkah-langkah penelitian sebagai berikut:

- Melakukan analisis statistika deskriptif mengenai persentase penduduk miskin dan variabel yang diduga mempengaruhinya.
- Membuat diagram pencar (*scatter plot*) antara variabel respon dan variabel prediktor
- Banyaknya titik knot yang dicobakan pada penelitian ini dibatasi, yaitu 1, 2 hingga 3 titik knot.
- Memodelkan data persentase penduduk miskin Provinsi Papua di Indonesia Tahun 2016-2019 menggunakan model regresi nonparametrik spline truncated pada data longitudinal.
- Memilih titik knot yang optimal menggunakan metode *Generalized Cross-Validation* (GCV). Nilai GCV yang minimum akan memberikan titik knot yang optimal.
- Menetapkan model regresi nonparametrik spline truncated pada data longitudinal yang terbaik berdasarkan kriteria GCV yang minimum.
- Menghitung nilai Koefisien Determinasi (R^2) sebagai kriteria kebaikan model.
- Membuat interpretasi model.

IV. ANALISIS DAN PEMBAHASAN

A. Statistika Deskriptif

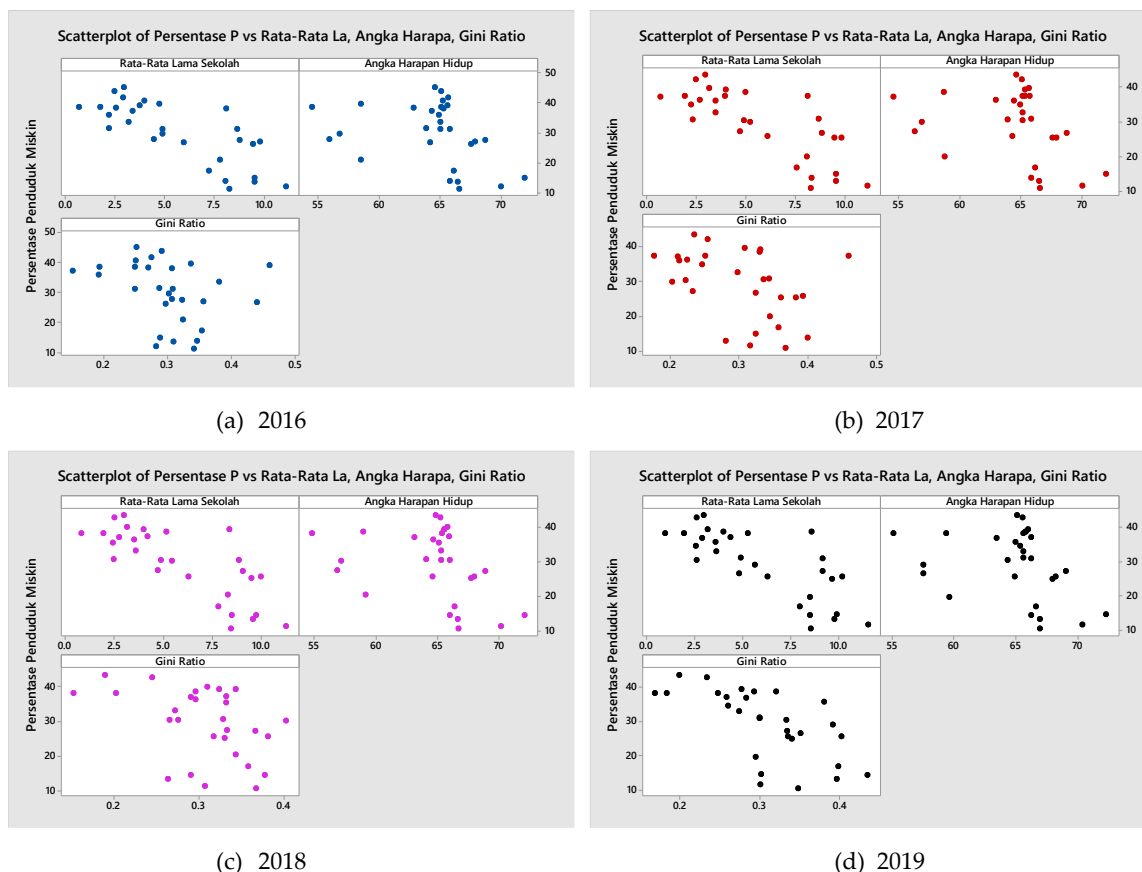
Akan ditampilkan statistika deskriptif untuk melihat gambaran secara umum untuk masing-masing variabel yang digunakan. Data yang digunakan untuk analisis statistika deskriptif adalah data Tahun 2019.

Tabel 2 Statistika Deskriptif

Variabel	Definisi Variabel	Minimum	Maksimum	Rata-Rata	Standar Deviasi
Y	Persentase Penduduk Miskin	10,35	43,65	29,22	10,02
X_1	Rata-Rata Lama Sekolah	0,97	11,55	6,01	3,08
X_2	Angka Harapan Hidup	55,12	72,27	64,98	3,82
X_3	Gini Ratio	0,17	0,44	0,31	0,06

B. Diagram Pencar

Selanjutnya untuk melihat pola hubungan antara variabel prediktor terhadap variabel respon dapat menggunakan diagram pencar. Ditampilkan diagram pencar dari persentase penduduk miskin dan variabel yang diduga mempengaruhinya untuk setiap tahun-nya sebagai berikut:



Gambar 3 Diagram Pencar antara Variabel Respon dan Prediktor

Berdasarkan gambar diatas, diagram pencar antara persentase penduduk miskin sebagai variabel respon dengan masing-masing variabel yang diduga mempengaruhi terlihat memiliki pola yang berubah-ubah pada sub-sub interval tertentu. Bentuk pola data yang demikian merupakan karakteristik dari estimator spline truncated, sedemikian sehingga pada penelitian ini akan di gunakan pendekatan regresi nonparametrik spline truncated untuk memodelkan permasalahan kemiskinan di Provinsi Papua.

C. Pemodelan Persentase Penduduk Miskin dengan Model Regresi Nonparametrik Spline Truncated Data Longitudinal

Pada bagian ini, akan dilakukan pemodelan kasus persentase penduduk miskin Kabupaten/Kota di Provinsi Papua Tahun 2016 hingga 2019 menggunakan model regresi nonparametrik spline truncated. Metode estimasi parameter yang digunakan adalah Weighted Least Squares (WLS). Pada estimasi WLS perlu adanya matriks pembobot. Pada penelitian ini matrik pembobot yang digunakan adalah

$$W_i = n^{-1}I_{n_i} \tag{9}$$

Tabel 3 Nilai GCV dengan Satu Titik Knot

Knot $k_{l,i,q}; l = 1; i = 1,2, \dots, 29; q = 1,2,3$					GCV
$k_{1,1,1}$	$k_{1,1,2}$...	$k_{1,29,2}$	$k_{1,29,3}$	
8,54	66,91	...	70,36	0,31	8,05
8,26	66,53	...	69,99	0,28	8,07
8,54	66,90	...	70,35	0,31	20,64
8,53	66,88	...	70,34	0,31	30,50

Bobot yang digunakan akan memberikan perlakuan yang sama untuk setiap pengamatan dalam subyek penelitian. Pada penelitian ini akan ditampilkan 4 posisi titik knot yang berbeda yang diduga merupakan titik knot yang optimal.

1) Pemodelan Spline Truncated Data Longitudinal dengan Menggunakan 1 Titik Knot

Pada data longitudinal, titik knot yang dihasilkan pada setiap Kabupaten/Kota akan berbeda sesuai dengan perilaku atau pola data. Nilai GCV terkecil menggunakan 1 titik knot pada data persentase penduduk miskin Kabupaten/Kota Provinsi Papua ditampilkan pada Tabel 3.

Berdasarkan Tabel 3, diperoleh nilai GCV minimum sebesar 8,05. Model regresi nonparametrik spline truncated dengan data longitudinal yang terbentuk menggunakan 1 titik knot sebagai berikut:

$$\hat{y}_{i,j} = \hat{\theta}_{1i1}X_{ij1} + \hat{\delta}_{1i1}(X_{ij1} - K_{1i1})_+ + \hat{\theta}_{1i2}X_{ij2} + \hat{\delta}_{1i2}(X_{ij2} - K_{1i2})_+ + \hat{\theta}_{1i3}X_{ij3} + \hat{\delta}_{1i3}(X_{ij3} - K_{1i3})_+$$

Dimana $i = 1,2, \dots, 29$ dan $j = 1,2, \dots, 4$.

Nilai koefisien determinasi (R^2) dari model regresi nonparametrik spline truncated 1 titik knot yang optimal yaitu sebesar 99,98%.

2) Pemodelan Spline Truncated Data Longitudinal dengan Menggunakan 2 Titik Knot

Nilai GCV yang minimum menggunakan 2 titik knot pada data persentase penduduk miskin Kabupaten/Kota Provinsi Papua ditampilkan pada Tabel 4.

Tabel 4 Nilai GCV dengan Dua Titik Knot

Knot $k_{l,i,q}; l = 1,2 ; i = 1,2, \dots, 29 ; q = 1,2,3$					GCV
$k_{1,1,1}$	$k_{2,1,1}$...	$k_{1,29,3}$	$k_{2,29,3}$	8,07
8,26	8,56	...	0,28	0,32	20,61
8,54	8,56	...	0,31	0,32	30,50
8,53	8,56	...	0,31	0,32	30,51
8,53	8,54	...	0,31	0,31	

Berdasarkan Tabel 4, diperoleh nilai GCV minimum sebesar 8,07. Model regresi nonparametrik spline truncated dengan data longitudinal yang terbentuk menggunakan 2 titik knot sebagai berikut:

$$\hat{y}_{i,j} = \hat{\theta}_{1i1}X_{ij1} + \hat{\delta}_{1i1}(X_{ij1} - K_{1i1})_+ + \hat{\delta}_{2i1}(X_{ij1} - K_{2i1})_+ + \hat{\theta}_{1i2}X_{ij2} + \hat{\delta}_{1i2}(X_{ij2} - K_{1i2})_+ + \hat{\delta}_{2i2}(X_{ij2} - K_{2i2})_+ + \hat{\theta}_{1i3}X_{ij3} + \hat{\delta}_{1i3}(X_{ij3} - K_{1i3})_+ + \hat{\delta}_{2i3}(X_{ij3} - K_{2i3})_+$$

Dimana $i = 1,2, \dots, 29$ dan $j = 1,2, \dots, 4$.

Nilai koefisien determinasi (R^2) dari model regresi nonparametrik spline truncated 2 titik knot yang optimal yaitu sebesar 99,99%.

3) Pemodelan Spline Truncated Data Longitudinal dengan Menggunakan 3 Titik Knot

Nilai GCV yang minimum menggunakan 3 titik knot pada data persentase penduduk miskin Kabupaten/Kota Provinsi Papua ditampilkan pada Tabel 5.

Tabel 5 Nilai GCV dengan Tiga Titik Knot

Knot $k_{l,i,q}; l = 1,2,3 ; i = 1,2, \dots, 29 ; q = 1,2,3$					GCV
$k_{1,1,1}$	$k_{2,1,1}$...	$k_{2,29,3}$	$k_{3,29,3}$	30,52
8,49	8,52	...	0,31	0,32	30,53
8,46	8,53	...	0,31	0,32	30,53
8,46	8,49	...	0,30	0,32	30,54
8,46	8,49	...	0,30	0,31	

Berdasarkan Tabel 5, diperoleh nilai GCV minimum sebesar 30,52. Model regresi nonparametrik spline truncated dengan data longitudinal yang terbentuk menggunakan 3 titik knot sebagai berikut:

$$\hat{y}_{i,j} = \hat{\theta}_{1i1}X_{ij1} + \hat{\delta}_{1i1}(X_{ij1} - K_{1i1})_+ + \hat{\delta}_{2i1}(X_{ij1} - K_{2i1})_+ + \hat{\delta}_{3i1}(X_{ij1} - K_{3i1})_+ + \hat{\theta}_{1i2}X_{ij2} + \hat{\delta}_{1i2}(X_{ij2} - K_{1i2})_+ + \hat{\delta}_{2i2}(X_{ij2} - K_{2i2})_+ + \hat{\delta}_{3i2}(X_{ij2} - K_{3i2})_+ + \hat{\theta}_{1i3}X_{ij3} + \hat{\delta}_{1i3}(X_{ij3} - K_{1i3})_+ + \hat{\delta}_{2i3}(X_{ij3} - K_{2i3})_+ + \hat{\delta}_{3i3}(X_{ij3} - K_{3i3})_+$$

Dimana $i = 1,2, \dots, 29$ dan $j = 1,2, \dots, 4$.

Nilai koefisien determinasi (R^2) dari model regresi nonparametrik spline truncated 3 titik knot yang optimal yaitu sebesar 99,99%.

4) Pemilihan Model Regresi Nonparametrik Spline Truncated Terbaik

Berikut ditampilkan hasil ringkasan model regresi nonparametrik spline truncated dalam proses pemilihan titik knot yang optimal sebagai berikut:

Tabel 6 Ringkasan Hasil

Titik Knot	GCV	Koefisien Determinasi (R ²)
1 titik knot	8.05	99,98%
2 titik knot	8.07	99,99%
3 titik knot	30.52	99,99%

Penentuan model regresi nonparametrik terbaik dilakukan dengan membandingkan nilai GCV yang minimum diantara beberapa titik knot yang dicobakan. Pada penelitian ini titik knot yang dicobakan adalah satu hingga tiga titik knot. Berdasarkan Tabel 6, dapat diketahui jika model regresi nonparametrik spline truncated pada data longitudinal yang terbaik adalah model dengan 1 titik knot, dengan nilai GCV yang paling minimum sebesar 8,05 dan nilai R² sebesar 99,98%.

Secara umum, model regresi nonparametrik spline truncated pada data longitudinal dengan 1 titik knot dapat dituliskan sebagai berikut:

$$\hat{y}_{i,j} = \hat{\theta}_{1i1}X_{ij1} + \hat{\delta}_{1i1}(X_{ij1} - K_{1i1})_+ + \hat{\theta}_{1i2}X_{ij2} + \hat{\delta}_{1i2}(X_{ij2} - K_{1i2})_+ + \hat{\theta}_{1i3}X_{ij3} + \hat{\delta}_{1i3}(X_{ij3} - K_{1i3})_+$$

Dimana $i = 1,2, \dots, 29$ dan $j = 1,2, \dots, 4$.

Model regresi nonparametrik yang terbaik, dapat dituliskan di setiap Kabupaten/Kota sebagai berikut:

1. Kabupaten Merauke

$$\hat{y}_{1,j} = -1,50X_{1j1} - 13,21(X_{1j1} - K_{1,1,1})_+ + 0,40X_{1j2} - 17,61(X_{1j2} - K_{1,1,2})_+ - 10,26X_{1j3} - 0,13(X_{1j3} - K_{1,1,3})_+$$

2. Kabupaten Jayawijaya

$$\hat{y}_{2,j} = -5,51X_{2j1} - 14,59(X_{2j1} - K_{1,2,1})_+ + 1,23X_{2j2} - 23,71(X_{2j2} - K_{1,2,2})_+ - 18,76X_{2j3} + 2,68(X_{2j3} - K_{1,2,3})_+$$

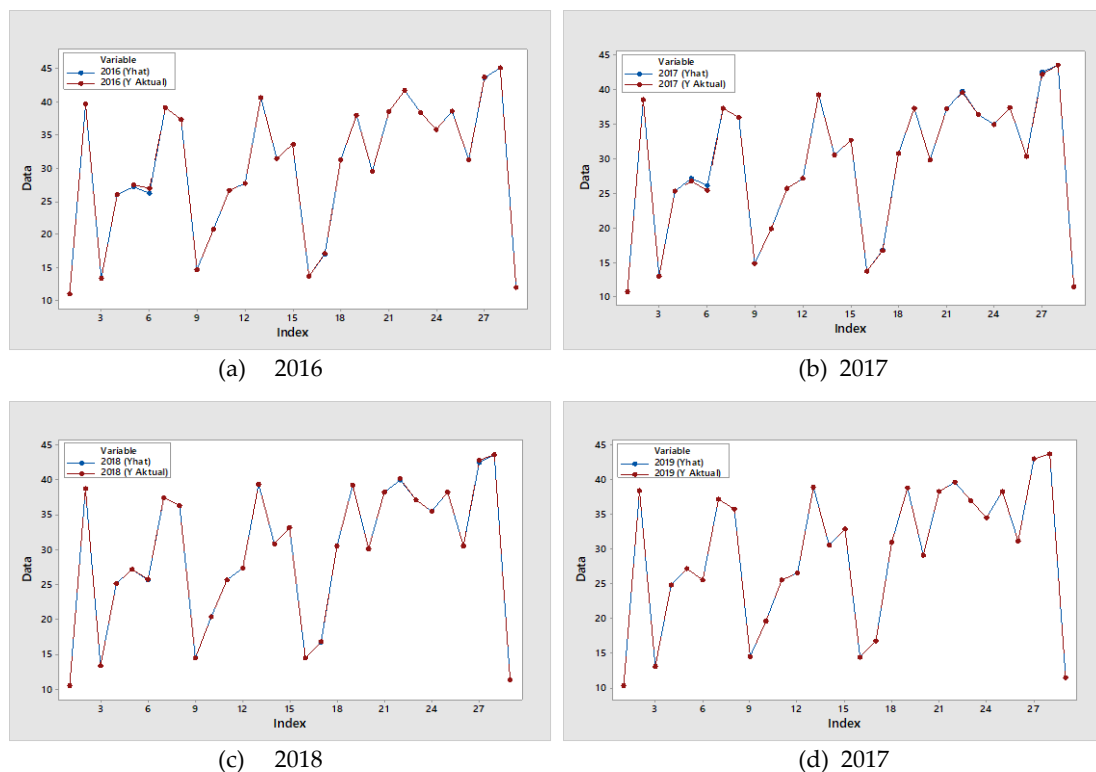
Hingga

.

29. Kota Jaya Pura

$$\hat{y}_{29,j} = -2,21X_{29j1} + 15,49(X_{29j1} - K_{1,29,1})_+ + 0,59X_{29j2} + 14,73(X_{29j2} - K_{1,29,2})_+ - 17,21X_{29j3} + 1,24(X_{29j3} - K_{1,29,3})_+$$

Selanjutnya di visualisasikan nilai y dan \hat{y} dari hasil pemodelan regresi nonparametrik spline truncated menggunakan data longitudinal untuk setiap tahunnya sebagai berikut:



Gambar 4 Visualisasi nilai y dan \hat{y} dari hasil pemodelan

Terlihat pada Gambar 4, jika visualisasi nilai y dan \hat{y} cenderung mengikuti data aktual.

V. KESIMPULAN

Model regresi nonparametrik spline truncated data longitudinal yang diaplikasikan pada data persentase penduduk miskin di Provinsi Papua, Indonesia dengan bobot $W_i = n^{-1}I_{n_i}$ serta jumlah titik knot yang dicobakan adalah 1 hingga 3 titik knot. Model terbaik yang dihasilkan berdasarkan kriteria GCV minimum adalah model regresi nonparametrik spline truncated dengan 1 titik knot. Secara umum, modelnya dapat dituliskan sebagai berikut:

$$\hat{y}_{i,j} = \hat{\theta}_{1i1}X_{ij1} + \hat{\delta}_{1i1}(X_{ij1} - K_{1i1})_+ + \hat{\theta}_{1i2}X_{ij2} + \hat{\delta}_{1i2}(X_{ij2} - K_{1i2})_+ + \hat{\theta}_{1i3}X_{ij3} + \hat{\delta}_{1i3}(X_{ij3} - K_{1i3})_+$$

Dimana $i = 1, 2, \dots, 29$ dan $j = 1, 2, \dots, 4$, dengan nilai GCV yang paling minimum sebesar 8,05 dan nilai R^2 sebesar 99,98%.

REFERENSI

- [1] I. N. Budiantara, Regresi Nonparametrik Spline Truncated, Surabaya: ITS Press, 2019.
- [2] R. L. Eubank, Nonparametric Regression and Spline Smoothing, New York: Marcel Dekker, 1999.
- [3] D. R. S. Saputro, K. R. Demu, dan P. Widyaningsih, "Nonparametric Truncated Spline Regression Model on Data of Human Development Index (HDI) in Indonesia," *IOP Conf. Series: Journal of Physics*, vol. 1188, pp. 01-01, 2018.
- [4] A. T. R. Dani, N. Y. Adrianingsih, dan A. Ainurrochmah, "Pengujian Hipotesis Simultan Model Regresi Nonparametrik Spline Truncated dalam Pemodelan Kasus Ekonomi," *JAMBURA Journal of Probability and Statistics*, vol. 01, pp. 98-106, 2020.
- [5] D. N. Gujarati, Basic Econometrics, 4th Edition, New York: McGraww-Hill Companies Inc., 2013.
- [6] C. Hsiao, Analysis of Panel Data, New York: Cambridge University Press, 2003.
- [7] M. Durban, J. Harezlak, M. P. Wand, dan R. J. Carrol, "Simple Fitting of Subject-Specific Curves for Longitudinal Data," *Statistics in Medicine*, pp. 01-24, 2004.
- [8] L. Laome, "Model Regresi Semiparametrik Spline untuk Data Longitudinal pada Kasus Kadar CD4 Penderita HIV," *Paradigma*, vol. 13, pp. 189-194, 2009.
- [9] A. Prahutama, "Model Regresi Nonparametrik Polinomial Lokal Birespon pada Data Longitudinal," Tesis, Surabaya: Institut Teknologi Sepuluh Nopember, 2013.
- [10] W. Hardle, Applied Nonparametric Regression, Berlin: Humboldt-Universitat zu Berlin, 1994.
- [11] G. Wahba, Spline Models for Observational Data, Pennsylvania: SIAM, 1990.